ANÁLISIS MULTIVARIADO

CIENCIAS BÁSICAS

textosuniversitarios

Javier Bech Vertti

UNIVERSIDAD AUTÓNOMA DE AGUASCALIENTES

Análisis Multivariado

Primera edición 2019

D.R. © Universidad Autónoma de Aguascalientes
Av. Universidad No. 940
Ciudad Universitaria
C.P. 20131, Aguascalientes, Ags.
www.uaa.mx/direcciones/dgdv/editorial/

© Javier Bech Vertti

ISBN 978-607-8652-68-6

Hecho en México / *Made in Mexico*

PRÓLOGO

El "Cuaderno de Trabajo**: ANÁLISIS MULTIVARIADO",** es una obra elaborada durante mi jubilación para **apoyar en particular al estudiante de la carrera de Mercadotecnia y a otros estudiantes de carreras afines adscritas al Centro de Ciencias Económicas y Administrativas de la U.A.A**., en el estudio de los conceptos estadísticos durante el desarrollo de su **curso "Análisis Multivariado",** o bien como fuente de consulta general, por medio de una gran variedad de **ejemplos ilustrativos resueltos en detalle, actividades de aprendizaje, autoevaluaciones, ejercicios de refuerzo y sesiones de inducción a los software estadísticos Excel y Minitab**, así como un sin número de **ayudas didácticas** que incluyen ejercicios complementarios, autoevaluaciones con reactivos de falso o verdadero y de opción múltiple, un amplio glosario de términos, simbología utilizada, fórmulas claves utilizadas en cada capítulo. Las respuestas a gran parte de los ejercicios planteados en **el "C**uaderno de Trabajo**: Análisis multivariado"** se desarrolará por separado como guía adicional.

Este **"C**uaderno de Trabajo**: Análisis Multivariado",** fue diseñado en principio como **auxiliar para el estudiante** y en forma **opcional para el profesor** que lo desee utilizar como **guía del curso** en el planteamiento y resolución de los ejercicios aquí planteados **coadyuvando con esto en el proceso enseñanza-aprendizaje** durante el desarrollo del curso de **"Análisis Multivariado"** o como fuente de consulta general**.** Debo hacer hincapié en que **NO pretende suplir en forma alguna la labor del docente en el aula**, sin embargo la forma de abordar los contenidos de la materia permite **ampliar y reforzar los temas tratados en el curso así como estimular nuevas prácticas pedagógicas y herramientas**, para construir un **aprendizaje significativo en el aula de clase** y que en forma muy puntual **debe llevar al estudiante a ser capaz de**: comprender **los principios del análisis multivariado** y aplicar algunos **métodos de reducción y de clasificación** en **problemas de su área de estudio** en el ámbito de los fenómenos económicos,

financieros, comerciales y administrativos, **demostrando capacidad para analizar e interpretar resultados numéricos estadísticos en contextos específicos.**

El "**C**uaderno de Trabajo**: ANÁLISIS MULTIVARIADO"** está **organizado** en **tres capítulos** en principio **para la carrera de Mercadotecnia**: **Introducción y análisis previo de los datos multivariados, Métodos de reducción y Métodos de clasificación**. En cada uno de ellos se tratan contenidos relevantes del tema y por eso, **todos se inician con la descripción de los aprendizajes esperados que debe lograr el estudiante**. Cada contenido a partir del segundo capítulo se estructura en las siguientes secciones:

1. **Síntesis de los conceptos básicos**: es un resumen de los conceptos centrales involucrados en los aprendizajes en el aula de clase. Asimismo, se encuentran las principales fórmulas y relaciones numéricas que sustentan la Estadística.
2. **Ejemplos ilustrativos resueltos**: en esta sección se plantean ejercicios representativos de la clase y se resuelven en detalle.
3. **Actividades de aprendizaje**: esta sección le permitirá al estudiante fijar las ideas y están diseñados para ser resueltos primero en forma convencional, es decir a mano, y posteriormente utilizando un software estadístico para comparar sus resultados, fomentando así la retroalimentación correspondiente.
4. **Autoevaluaciones**: esta sección le permitirá al estudiante ejercitar los aprendizajes en el aula de clase y podrá autoevaluar su desempeño y darse cuenta en que puntos o áreas se encuentra más débil y en cuales más fuerte para así enfocar sus esfuerzos en los puntos más débiles, ahorrándole tiempo en la preparación de su examen departamental.
5. **Ejercicios de refuerzo**: en esta sección, se presentan varios ejercicios para su resolución, orientados a la preparación del estudiante para el examen departamental del capítulo.
6. **Ejemplos ilustrativos resueltos en Excel y/o Minitab**: esta sección utiliza pantallas de captura, cuadros de dialogo, gráficos y salidas de resultados que le permitirá al estudiante familiarizarse con los comandos básicos necesarios para lograr buenos resultados con estos software estadísticos.
7. **Notas al margen izquierdo**: esta sección le permitirá advertir al estudiante de algún aspecto a remarcar o alertar de lo que se dice en el texto desarrollando su capacidad de análisis al tener que comprender y examinar el texto minuciosamente.

Uso de calculadora y software estadístico: Para trabajar con el presente **"Cuaderno de trabajo: Análisis Multivariado",** el estudiante debe **usar calculadora y algún software estadístico**. En cuanto al **uso de la calculadora**, se tomó como referencia **la calculadora Casio fx.82MS**, El uso de un **software estadístico como Excel y/o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben comprender primero los pasos del proceso. Por lo mismo es **muy importante que primero se resuelva el ejercicio en forma manual** y posteriormente se utilice un software para comparar los resultados. Es importante mencionar **que pueden existir diferencias en las respuestas debido a la cantidad de dígitos** que se utilizan en los cálculos manuales. Dado que en algunos contenidos se utiliza álgebra matricial, **se sugiere utilizar en general aproximaciones de al menos 5 dígitos.**

Al final de cada capítulo se presenta la siguiente **ayuda didáctica**:

- **Autoevaluaciones con reactivos de falso o verdadero y de opción múltiple** que permitirá reforzar el aprendizaje del alumno.

Al final del **"Cuaderno de trabajo: Análisi Multivariado",** se presenta:

- **Un glosario de términos** de los conceptos expuestos.
- **Simbología** utilizada.
- **Fórmulas clave** que apoyan la resolución de los ejercicios propuestos en cada sección.
- **Apéndice con Tablas.**
- **Bibliografía.**

Y en un **anexo** aparte al **"Cuaderno de trabajo: Análisis Multivariado",** se presentan las:

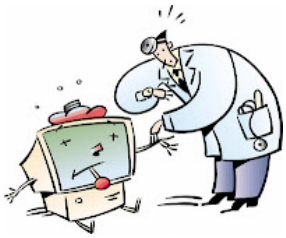
- **Respuestas a los ejercicios** planteados en las actividades de aprendizaje, autoevaluaciones y ejercicios, así como de las autoevaluaciones con reactivos de de falso o verdadero y de opción múltiple y/o llenado del enunciado de cada capítulo.

Finalmente**...** el viejo refrán que dice que **"La práctica hace al maestro"** tiende a ser mas cierto de lo que muchas veces pensamos. **La práctica, y no la genética**, es lo que hace **al maestro**. Y es **la perseverancia**, a lo largo del tiempo, lo que verdaderamente saca adelante a los **triunfadores**, por eso este "**C**uaderno de Trabajo**: ANÁLISIS MULTIVARIADO"** se ha elaborado con la finalidad de que **el estudiante ejercite los procedimientos que se sugieren a lo largo del curso de "Análisis Multivariado",** pensando que entre **más se practique**, mucho mayor será **la comprensión que se tenga de ellos.** Este material pude ser de **estudio independiente o combinado** con las instrucciones del docente en caso de que se opte por utilizarlo como guía de estudio. Se recomienda **utilizarlo como material de apoyo, para reforzar conocimientos, para autoevaluarse, o como preparación para la evaluación presencial**. Lo importante es que **le sirva al estudiante para que identifique cuáles son los temas que necesita reforzar y para darse cuenta de los logros que ha alcanzado. Espero que sea de utilidad.**

PROYECTO DE JUBILACIÓN 2019

ARQ. Y M. EN ADMÓN. JAVIER BECH VERTTI

P.1 PROBLEMÁTICA Y NECESIDADES.



Ante la carencia de un **Cuaderno de Trabajo** para el **estudio y la enseñanza de las metodologías estadísticas** necesarias para entender **el comportamiento de algunos fenómenos aleatorios relacionados con el área de Mercadotecnia**, que permita **desarrollar en el alumno habilidades** para **aplicar dichas metodologías** asociadas al análisis de datos y dado que se requiere de la **consulta de una vasta bibliografía** , se ha observado, a través de 30 años de experiencia docente, que los **alumnos suelen limitarse a los apuntes que el profesor les ofrece**, **eventualmente consultan algún texto** y en **pocas ocasiones acuden a asesoría**; lo que **redunda en un bajo nivel de aprovechamiento**, evidente en los trabajos y exámenes que realizan.

Por lo que hace **a los profesores**, también se ha observado que **eligen como apoyo algún texto base**, que **puede no apegarse completamente al programa de la materia** y **complementan con otros textos** los temas faltantes **para la elaboración de apuntes y trabajos**, con lo que dejan **puntos del programa parcialmente cubiertos o sin cubrir** además de la **falta de uniformidad en la nomenclatura** al utilizar varios textos **lo que crea conflicto en los alumnos** que por alguna razón, deben repetir el curso.

En la **mayoría de los textos no se incluye ningún apartado para realizar actividades extra clase ni de autoevaluación**, con el que **los alumnos** puedan en **forma personal reafirmar y evaluar sus conocimientos a medida que avanza el semestre y retroalimentarse para preparar sus exámenes**, por lo que **no se puede fomentar en forma efectiva el proceso de auto aprendizaje** que marca tanto el programa de la materia como el **modelo educativo adoptado por la Institución** donde **el alumno debe ser un agente activo con una orientación constructivista.**

P.2

JUSTIFICACIÓN DEL PROYECTO

De acuerdo a la **problemática y necesidades planteadas**, se justifica el esfuerzo para remediarlas mediante **la elaboración de un "Cuaderno de Trabajo para la Materia de Análisis Multivariado"** principalmente para la **carrera de Mercadotecnia,** con posibilidad de utilizarse **en carreras afines adscritas al Centro de Ciencias Aconómicas y Administrativas.** La materia de **Análisis Multivariado en la carrera de Mercadotecnia** se cursa hasta la elaboración del presente trabajo en el **quinto semestre** y el **cuaderno de trabajo conducirá a los alumnos en su estudio** permitiéndoles **ampliar de forma importante su horizonte de conocimientos.** Asimismo, **servirá de apoyo para el desarrollo del trabajo de los profesores** uniformizando por un lado tanto los **criterios, niveles y alcances establecidos por la Academia correspondiente al curso por el Departamento de Estadística** para cada contenido, **como la nomenclatura que pudiera utilizar el profesor**, **evitando así la confusión de los alumnos que por algún motivo tuvieran que recursar la materia**; además **el profesor podrá contar como apoyo** con una **serie de actividades individuales y/o grupales** así como **reactivos o ítems** para que los alumnos **desarrollen y/o realicen autoevaluaciones** ya sea **en forma presencial en el aula de clase o bien mediante alguna plataforma ó Software de cómputo** estadístico como pudieran ser **Moodle, Minitab y/o Excel**, que ya tiene habilitado el Departamento de Estadística como apoyo para algunos cursos.

P.3

OBJETIVO GENERAL Y OBJETIVOS PARTICULARES

a) General:

Elaborar un **documento** que sirva como **"C**uaderno de Trabajo**: ANÁLISIS MULTIVARIADO"** para la materia de **"Análisis Multivariado" de la Carrera de Licenciado en Mercadotecnia y/o carreras afines adscritas al Centro de Ciencias Administrativas de la U.A.A**., o como de consulta general, que contenga material útil para estudiar y entender el comportamiento de fenómenos aleatorios relacionados con el área de la Mercadotecnia **dentro del programa de Análisis Multivariado** para la carrera de **Mercadotecnia**, ofreciendo **a alumnos y profesores** información **diversa y actualizada**, así como **herramientas para construir un aprendizaje significativo en el aula de clase**, que contribuya a **mejorar la calidad de la educación**, cercanos a la realidad que viven los estudiantes de nuestro Estado y por ende del país.

b) Específicos:

- Que el **"C**uaderno de Trabajo**: ANÁLISIS MULTIVARIADO"** impacte en los **procesos educativos y de enseñanza-aprendizaje** por medio de la interacción de los alumnos con los contenidos pedagógicos incorporados en el mismo.
- Que el **"C**uaderno de Trabajo**: ANÁLISIS MULTIVARIADO"** se pueda **utilizar como una herramienta de apoyo docente** en el tratamiento de los temas y contenidos de los libros de Texto, **con base en el programa de la materia de Análisis Multivariado de Mercadotecnia**, o como de consulta general, con la finalidad de **ampliar y reforzar los temas que en ellos se traten**, así como **estimular nuevas prácticas pedagógicas y herramientas**, para construir un **aprendizaje significativo en el aula de clase.**
- Que el **"C**uaderno de Trabajo**: ANÁLISIS MULTIVARIADO" defina al profesor como guía y mediador del proceso de debate, reflexión y participación** que se genere en el aula y le **sugiera estrategias didácticas e innovadoras para el tratamiento de los**

contenidos curriculares, a fin de integrarlas a sus experiencias y métodos propios.

- Que el **"C**uaderno de Trabajo**: ANÁLISIS MULTIVARIADO"** indique **la posible incorporación de las TIC´S en los procesos educativos**, a fin de establecer un **puente natural entre la forma tradicional de presentar los contenidos curriculares y las posibilidades que brindan las nuevas tecnologías.**

P.4 METAS Y/O ACTIVIDADES

Desarrollar contenidos marcados en el programa de la **materia de Análisis Multivariado de la carrera de Mercadotecnia**, mediante:

a) **La revisión bibliográfica** para la selección y compilación del material idóneo para el programa de la materia.

b) **Estructuración del material.**
- Breve introducción de conceptos básicos.
- Elaboración de ejemplos ilustrativos.
- Actividades para desarrollar en clase y/o laboratorio, en forma individual y/o grupal.
- Sesiones de inducción y/ o prácticas de laboratorio para ser realizadas con algún paquete de cómputo, con ejemplos ilustrativos.
- Propuestas de problemas, ejercicios y/o tareas para reforzar los temas con y sin respuestas.

c) Autoevaluaciones para resolver en clase o para que el profesor pueda capturarlas y aplicarlas en algún medio digital como la plataforma Moodle que el Departamento de Estadística ya tiene a disposición de los docentes.

d) **Apéndices.**

e) **Glosario de términos.**

f) **Formularios.**

P.5

OBSERVACIONES GENERALES

El "Cuaderno de Trabajo**: ANÁLISIS MULTIVARIADO"** aunque va dirigido específicamente a la **carrera de Mercadotecnia** podrá ser utilizado perfectamente como apoyo o consulta general durante el desarrollo de la materia de "**Análisis Multivariado"** en **carreras afines adscritas al Centro de Ciencias Económicas y Administrativas u otros Centros Académicos de la U.A.A** y/o equivalentes a la **Academia correspondiente del Departamento de estadística de la U.A.A.**

P.6

AGRADECIMIENTOS

A **IMELDA**, mi esposa, por la paciencia y apoyo tan grande que ha tenido siempre conmigo y en particular, durante el desarrollo de este trabajo.

A **ANGÉLICA,** mi hija, por la ardua labor de transcribir el presente trabajo de la forma manuscrita a la forma digital.

A **CRISTINA VALDERRAMA,** mi exalumna, por su gran ayuda durante su servicio social al desarrollar manualmente los ejercicios que se presentan en este trabajo.

A la **U.A.A.,** mi «*alma mater*» ,a los **CENTROS BÁSICO Y ECONÓMICO ADMINISTRATIVO, DEPARTAMENTOS DE ESTADÍSTICA Y MERCADOTECNIA**, , por todas las facilidades que me otorgaron para poder llegar hasta este momento.

CONTENIDO

CAPÍTULO 1. INTRODUCCIÓN AL ANÁLISIS MULTIVARIANTE

Capítulo 1. Introducción al Análisis Multivariante 22
Objetivo 1.1 22
Antecedentes..... 22
Introducción al análisis multivariante..... 23
Conceptos Básicos. Análisis Multivariante..... 23
Objetivo 1.2..... 27
Antecedentes..... 27
1.2.1. Valor teórico. Escalas de medición y su relación con las técnicas multivariantes..... 27
Conceptos básicos. Valor Teórico..... 28
1.2.2. Base de Datos. Formato de Encuesta sobre Bienes Raíces. Código para la computadora. Construcción de la Base de datos "BIENES1"..... 32
Conceptos Básicos. Base de Datos..... 32
1.2.2.1. Ejemplo Ilustrativo en Minitab 17. Construcción de la Base de datos "BIENES1"..... 40
Objetivo 1.3..... 44
Antecedentes..... 45
1.3.1. Análisis Previo de los datos multivariantes. Examen gráfico de los datos multivariantes. Datos Ausentes. Casos Atípicos..... 45
Conceptos Básicos. Examen gráfico de los datos multivariantes. Datos Ausentes. Datos ó Casos Atípicos..... 45
1.3.1.1. Ejemplo Ilustrativo. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 57
1.3.1.1 Actividad de Aprendizaje. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 60
1.3.1.1 Ejercicio de Autoevaluación. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 61
1.3.1.1 Ejercicios de Refuerzo. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 62
1.3.1.1.1 Ejercicios de Refuerzo. Datos Atípicos. Diagrama de caja y Brazos Simple..... 62
1.3.1.1.2 Ejercicios de Refuerzo. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 62
1.3.1.1.3 Ejercicios de Refuerzo. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 63

1.3.1.2. D.A. Ejemplo Ilustrativo en Minitab 17. Datos Atípicos..... 63
1.3.1.2 Actividad de Aprendizaje en Minitab 17. Datos Atípicos..... 65
1.3.1.2 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos..... 65
1.3.1.2 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos..... 66
1.3.1.2.1 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos..... 66
1.3.1.2.2 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos..... 66
1.3.1.2.3 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos..... 66
1.3.1.3. D.A. Ejemplo Ilustrativo en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 67
1.3.1.3 Actividad de Aprendizaje en Minitab 17 . Datos Atípicos. Un Nivel de Factor..... 71
1.3.1.3 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 72
1.3.1.3 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 72
1.3.1.3.1. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 72
1.3.1.3.2. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 73
1.3.1.3.3. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 73
1.3.1.4. D.A. Ejemplo Ilustrativo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 74
1.3.1.4 Actividad de Aprendizaje en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 77
1.3.1.4 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 77
1.3.1.4 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 78
1.3.1.4.1. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 78
1.3.1.4.2. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 78
1.3.1.4.3. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 79
1.3.1.5. D.A. Ejemplo Ilustrativo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 80
1.3.1.5 Actividad de Aprendizaje en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 87
1.3.1.5 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 87
1.3.1.5 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 88
1.3.1.5.1. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 88
1.3.1.5.2. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 88
1.3.1.5.3. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 89
1.3.2. Análisis Previo de los datos multivariantes. Verificación de los supuestos del Análisis Multivariante. Normalidad Homocedasticidad y linealidad..... 89
1.3.2.1. P.N. Ejemplo Ilustrativo en Minitab 17. Prueba de Normalidad. Prueba de Anderson

Darling…. 94
1.3.2.1 Actividad de Aprendizaje en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling…. 98
1.3.2.1 Ejercicio de Autoevaluación en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling…. 99
1.3.2.1 Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling…. 99
1.3.2.1.1. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling…. 99
1.3.2.1.2. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling…. 100
1.3.2.1.3. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling…. 100
1.3.2.2. P.H. Ejemplo Ilustrativo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos…. 100
1.3.2.2 Actividad de Aprendizaje en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos…. 100
1.3.2.2 Ejercicio de Autoevaluación en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos…. 117
1.3.2.2 Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos…. 118
1.3.2.2.1. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos…. 119
1.3.2.2.2. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos…. 119
1.3.2.2.3. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos…. 120
1.3.2.3. P.H. Ejemplo Ilustrativo en Minitab 17. Prueba de Homocedasticidad. Prueba de Barlett…. 121
1.3.2.3 Actividad de Aprendizaje en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett…. 128
1.3.2.3 Ejercicio de Autoevaluación en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett…. 128
1.3.2.3 Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett…. 129
1.3.2.3.1. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett…. 129
1.3.2.3.2. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett…. 129
1.3.2.3.3. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett…. 130
Objetivo 1.4 Sistema de Ecuaciones Lineales…. 130
Antecedentes…. 130

1.4.1 Sistemas matriciales de resolución de Sistemas de Ecuaciones Lineales..... 131
Conceptos Básicos. Sistemas matriciales de resolución de sistemas lineales de ecuaciones..... 131
1.4.1.1. Ejemplo ilustrativo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 133
1.4.1.1 Actividad de Aprendizaje. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 135
1.4.1.1 Ejercicio de Autoevaluación. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 135
1.4.1.1 Ejercicios de Refuerzo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 136
1.4.1.1.1. Ejercicios de Refuerzo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 136
1.4.1.1.2. Ejercicios de Refuerzo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 136
1.4.1.1.3. Ejercicios de Refuerzo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 136
1.4.1.2. Ejemplo Ilustrativo en Excel. Resolución de un sistema lineal de ecuaciones mediante el método de matriz inversa..... 137
1.4.1.2 Actividad de Aprendizaje en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 145
1.4.1.2 Ejercicio de Autoevaluación en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 145
1.4.1.2 Ejercicios de Refuerzo en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 147
1.4.1.2.1. Ejercicios de Refuerzo en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 147
1.4.1.2.2. Ejercicios de Refuerzo en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 147
1.4.1.2.3. Ejercicios de Refuerzo en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 147
Autoevaluación con reactivos de Falso ó Verdadero..... 148
Autoevaluación con reactivos de Opción Múltiple..... 151

CAPÍTULO 2. MÉTODOS DE REDUCCIÓN

Capítulo 2. Métodos de Reducción....... 155
Objetivo 2.1 Componentes Prtincipales, Factor Análisis..... 155
Antecedentes..... 155
2.1.1. Factor, Carga de Factor, Porcentaje de variación explicada, Comunalidad..... 155
Conceptos Básicos. Análisis de Factores..... 155

2.1.1.1. Ejemplo Ilustrativo..... 162
Ejemplo Ilustrativo 2.1.1.1. Factor Análisis..... 162
2.1.1.1. Actividad de Aprendizaje..... 168
Actividad de Aprendizaje 2.1.1.1. Factor Análisis..... 168
2.1.1.1. Ejercicio de Autoevaluación. 171
Autoevaluación 2.1.1.1. Factor Análisis..... 171
2.1.1.1. Ejemplo Ilustrativo en Minitab 17..... 172
Ejemplo ilustrativo Integral en Minitab 17 2.1.1.1. Factor Análisis..... 172
2.1.1.1. Ejercicios de Refuerzo..... 185
Ejercicios de Refuerzo 2.1.1.1. Factor Análisis..... 185
2.1.1.1.1. Ejercicios de Refuerzo. Factor Análisis..... 185
2.1.1.1.2. Ejercicios de Refuerzo. Factor Análisis..... 186
2.1.1.1.3. Ejercicios de Refuerzo. Factor Análisis..... 186
Objetivo 2.2. Análisis de Conglomerados. (Cluster Análisis)..... 187
Antecedentes..... 187
2.2.1 Análisis de Conglomerados. Medidas de distancia y similitud. Método jerárquico aglomerativo..... 187
Conceptos básicos. Análisis de Conglomerados..... 187
2.2.1.1. Ejemplo ilustrativo..... 194
Ejemplo ilustrativo 2.2.1.1. Análisis de Conglomerados. Matriz de distancia..... 194
2.2.1.1. Actividad de Aprendizaje..... 225
Actividad de Aprendizaje 2.2.1.1. Análisis de Conglomerados. Matriz de distancia..... 225
2.2.1.1. Ejercicio de Autoevaluación..... 226
Autoevaluación 2.2.1.1. Análisis de Conglomerados. Matriz de distancia..... 226
2.2.1.1. Ejemplo ilustrativo Minitab 17..... 227
Ejemplo ilustrativo integral en Minitab 17. Análisis de Conglomerados. Medida de distancia..... 227
2.2.1.1. Ejercicios de Refuerzo..... 251
Ejercicios de Refuerzo 2.2.1.1. Análisis de Conglomerados. Matriz de distancia..... 251
2.2.1.1.1. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de distancia..... 251
2.2.1.1.2. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de distancia..... 251
2.2.1.2. Ejemplo ilustrativo..... 252
Ejemplo ilustrativo 2.2.1.2. Análisis de Conglomerados. Matriz de similitud..... 252
2.2.1.2. Actividad de aprendizaje..... 294
Actividad de aprendizaje 2.2.1.2. Análisis de Conglomerados. Matriz de similitud..... 294
2.2.1.2. Ejercicio de autoevaluación..... 295
Autoevaluación 2.2.1.2. Análisis de Conglomerados. Matriz de similitud..... 295
2.2.1.2. Ejemplo ilustrativo en Minitab 17..... 297
Ejemplo ilustrativo integral en Minitab 17. Análisis de Conglomerados. Medida de similitud..... 297
2.2.1.2. Ejercicios de refuerzo. 323
Ejercicios de refuerzo 2.2.1.2. Análisis de Conglomerados. Matriz de similitud..... 323
2.2.1.2.1. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de similitud..... 323
2.2.1.2.2. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de similitud..... 324
2.2.1.2.3. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de similitud..... 325
Autoevaluación con reactivos de Falso ó Verdadero..... 326
Autoevaluación con reactivos de Opción Múltiple..... 330

CAPÍTULO 3. MÉTODOS DE CLASIFICACIÓN

Capítulo 3. Métodos de Clasificación..... 333
Objetivo 3.1. Análisis discriminante..... 333
Antecedentes..... 333
3.1.1. El análisis discriminante..... 333
Conceptos Básicos. Análisis Discriminante..... 334
3.1.1.1. Ejemplo Ilustrativo..... 343
Ejemplo Ilustrativo. 3.1.1.1. Análisis Discriminante..... 343
3.1.1.1. Actividad de Aprendizaje..... 357
Actividad de Aprendizaje. 3.1.1.1. Análisis discriminante..... 357
3.1.1.1. Ejercicio de Autoevaluación..... 358
Autoevaluación. 3.1.1.1. Análisis discriminante..... 359
3.1.1.1. Ejercicios de Refuerzo..... 361
Ejercicios de Refuerzo. 3.1.1.1. Análisis discriminante..... 361
3.1.1.1.1. Ejercicios de Refuerzo. Análisis discriminante..... 361
3.1.1.1.2. Ejercicios de Refuerzo. Análisis discriminante..... 362
3.1.1.1.3. Ejercicios de Refuerzo. Análisis discriminante..... 363
Objetivo 3.2. Pruebas de hipótesis con datos multivariados..... 365
Antecedentes..... 365
3.2.1. Pruebas sobre varias variables. (Vector de Medias)..... 365
Conceptos básicos. Inferencia para un vector de medias..... 366
3.2.1.1. Ejemplo Ilustrativo..... 368
Ejemplo Ilustrativo- 3.2.1.1. Vector de Medias..... 368
3.2.1.1. Actividad de Aprendizaje..... 376
Actividad de Aprendizaje. 3.2.1.1. Vector de Medias..... 376
3.2.1.1. Ejercicio de Autoevaluación..... 377
Autoevaluación. 3.2.1.1. Vector de Medias..... 377
3.2.1.1. Ejercicios de Refuerzo..... 378
Ejercicios de Refuerzo 3.2.1.1. Vector de Medias..... 378
3.2.1.1.1. Ejercicios de Refuerzo. Vector de Medias..... 378
3.2.1.1.2. Ejercicios de Refuerzo. Vector de Medias..... 379
3.2.1.1.3. Ejercicios de Refuerzo. Vector de Medias..... 380
3.2.2. Pruebas de hipótesis para una comparación de medias para varias muestras..... 382
Conceptos básicos. Diferencia de medias multivariada..... 382
3.2.2.1. Ejemplo Ilustrativo..... 386
Ejemplo Ilustrativo. 3.2.2.1. Diferencia de medias multivariada..... 386
3.2.2.1. Actividad de Aprendizaje..... 408
Actividad de Aprendizaje. 3.2.2.1. Diferencia de medias multivariada..... 408
3.2.2.1. Ejercicio de Autoevaluación..... 409
Autoevaluación. 3.2.2.1. Diferencia de medias multivariada..... 409
3.2.2.1. Ejercicios de Refuerzo..... 411
Ejercicios de Refuerzo 3.2.2.1. Diferencia de medias multivariada..... 411
3.2.2.1.1. Ejercicios de Refuerzo. Diferencia de medias multivariada..... 411

3.2.2.1.2. Ejercicios de Refuerzo. Diferencia de medias multivariada..... 412
3.2.2.1.3. Ejercicios de Refuerzo. Diferencia de medias multivariada..... 413
3.2.3. Análisis de Varianza Multivariada (MANOVA)..... 415
Conceptos básicos. Manova..... 415
3.2.3.1. Ejemplo Ilustrativo..... 421
Ejemplo Ilustrativo. 3.2.3.1. Manova..... 421
3.2.3.1. Actividad de Aprendizaje..... 446
Actividad de Aprendizaje. 3.2.3.1. Manova..... 446
3.2.3.1. Ejercicio de Autoevaluación..... 447
Autoevaluación. 3.2.3.1. Manova..... 447
3.2.3.1. Ejemplo Ilustrativo en Minitab 17..... 449
Ejemplo Ilustrativo Integral en Minitab 17. 3.2.3.1. Manova..... 449
3.2.3.1. Ejercicios de Refuerzo..... 470
Ejercicios de Refuerzo. 3.2.3.1. Manova..... 470
3.2.3.1.1. Ejercicios de Refuerzo. Manova..... 470
3.2.3.1.2. Ejercicios de Refuerzo. Manova..... 471
3.2.3.1.3. Ejercicios de Refuerzo. Manova..... 472
Autoevaluación con reactivos de Falso y Verdadero..... 474
Autoevaluación con reactivos de Opción Múltiple..... 477
Glosario General..... 482
Tabla A1. Puntos porcentuales del rango estudentizado..... 492
Tabla A2. Valores críticos de t..... 498
Tabla A3. Valores críticos de F α=0.05..... 504
Tabla A4. Valores críticos de F α=0.01..... 510
Bibliografía..... 516

ANEXO DE RESPUESTAS CAPÍTULO 1. INTRODUCCIÓN AL ANÁLISIS MULTIVARIANTE

1.3.1.1 Actividad de Aprendizaje. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 520
1.3.1.2 Actividad de Aprendizaje en Minitab 17. Datos Atípicos..... 523
1.3.1.3 Actividad de Aprendizaje en Minitab 17 . Datos Atípicos. Un Nivel de Factor..... 525
1.3.1.4 Actividad de Aprendizaje en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 530
1.3.1.5 Actividad de Aprendizaje en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 533
1.3.2.1 Actividad de Aprendizaje en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling..... 540
1.3.2.2 Actividad de Aprendizaje en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos..... 543
1.3.2.3 Actividad de Aprendizaje en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett..... 558
1.4.1.1 Actividad de Aprendizaje. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 564
1.4.1.2 Actividad de Aprendizaje en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 566
1.3.1.1 Ejercicio de Autoevaluación. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 574
1.3.1.2 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos..... 578
1.3.1.3 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 580
1.3.1.4 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 584
1.3.1.5 Ejercicio de Autoevaluación en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 587
1.3.2.1 Ejercicio de Autoevaluación en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling..... 593
1.3.2.2 Ejercicio de Autoevaluación en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos..... 596
1.3.2.3 Ejercicio de Autoevaluación en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett..... 610
1.4.1.1 Ejercicio de Autoevaluación. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 615
1.4.1.2 Ejercicio de Autoevaluación en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 618
1.3.1.1.1 Ejercicios de Refuerzo. Datos Atípicos. Diagrama de caja y Brazos Simple..... 625
1.3.1.1.2 Ejercicios de Refuerzo. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 629
1.3.1.1.3 Ejercicios de Refuerzo. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 632
1.3.1.2.1 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos..... 635
1.3.1.2.2 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos..... 637
1.3.1.2.3 Ejercicios de Refuerzo en Minitab 17. Datos Atípicos..... 638
1.3.1.3.1. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 641

1.3.1.3.2. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Un Nivel de Factor.....
1.3.1.3.3. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Un Nivel de Factor..... 645
1.3.1.4.1. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 649
1.3.1.4.2. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 653
1.3.1.4.3. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Simple..... 655
1.3.1.5.1. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 658
1.3.1.5.2. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 661
1.3.1.5.3. Ejercicios de Refuerzo en Minitab 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple..... 667
1.3.2.1.1. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling..... 674
1.3.2.1.2. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling..... 680
1.3.2.1.3. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Normalidad. Prueba de Anderson Darling..... 683
1.3.2.2.1. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos..... 685
1.3.2.2.2. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos..... 687
1.3.2.2.3. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos..... 702
1.3.2.3.1. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett..... 716
1.3.2.3.2. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett..... 731
1.3.2.3.3. Ejercicios de Refuerzo en Minitab 17. Verificación de Supuestos. Prueba de Homocedasticidad. Prueba de Barlett..... 736
1.4.1.1.1. Ejercicios de Refuerzo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 741
1.4.1.1.2. Ejercicios de Refuerzo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 746
1.4.1.1.3. Ejercicios de Refuerzo. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 749
1.4.1.2.1. Ejercicios de Refuerzo en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 751
1.4.1.2.2. Ejercicios de Refuerzo en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 753
760

1.4.1.2.3. Ejercicios de Refuerzo en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa..... 766
Autoevaluación con reactivos de Falso ó Verdadero..... 774
Autoevaluación con reactivos de Opción Múltiple..... 774

ANEXO DE RESPUESTAS. CAPÍTULO 2. MÉTODOS DE REDUCCIÓN

2.1.1.1. Actividad de Aprendizaje. Factor Análisis..... 775
2.2.1.1. Actividad de Aprendizaje. Análisis de Conglomerados. Matriz de distancia..... 780
2.2.1.2. Actividad de aprendizaje. Análisis de Conglomerados- Matriz de Similitud..... 805
2.1.1.1. Ejercicio de Autoevaluación. Factor Análisis..... 824
2.2.1.1. Ejercicio de Autoevaluación. Análisis de Conglomerados. Matriz de distancia..... 828
2.2.1.2. Ejercicio de autoevaluación. Análisis de Conglomerados. Matriz de similitud..... 856
2.1.1.1.1. Ejercicios de Refuerzo. Factor Análisis..... 874
2.1.1.1.2. Ejercicios de Refuerzo. Factor Análisis..... 876
2.1.1.1.3. Ejercicios de Refuerzo. Factor Análisis..... 878
2.2.1.1.1. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de distancia..... 882
2.2.1.1.2. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de distancia..... 908
2.2.1.2.1. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de similitud..... 936
2.2.1.2.2. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de similitud..... 949
2.2.1.2.3. Ejercicio de refuerzo. Análisis de Conglomerados. Matriz de similitud..... 966
Autoevaluación con reactivos de Falso ó Verdadero..... 983
Autoevaluación con reactivos de Opción Múltiple..... 983

ANEXO DE RESPUESTAS. CAPÍTULO 3. MÉTODOS DE CLASIFICACIÓN

3.1.1.1. Actividad de Aprendizaje. Análisis Discriminante..... 984
3.2.1.1. Actividad de Aprendizaje. Vector de medias..... 994
3.2.2.1. Actividad de Aprendizaje. Diferencia de medias Multivariada..... 1000
3.2.3.1. Actividad de Aprendizaje. Manova..... 1019
3.1.1.1. Ejercicio de Autoevaluación. Análisis Discriminante..... 1041
3.2.1.1. Ejercicio de Autoevaluación. Vector de Medias..... 1052
3.2.2.1. Ejercicio de Autoevaluación. Diferencia de Medias Multivariada..... 1059
3.2.3.1. Ejercicio de Autoevaluación. Manova..... 1078
3.1.1.1.1. Ejercicios de Refuerzo. Análisis discriminante..... 1101
3.1.1.1.2. Ejercicios de Refuerzo. Análisis discriminante..... 1111
3.1.1.1.3. Ejercicios de Refuerzo. Análisis discriminante..... 1122

3.2.1.1.1. Ejercicios de Refuerzo. Vector de Medias..... 1131
3.2.1.1.2. Ejercicios de Refuerzo. Vector de Medias..... 1139
3.2.1.1.3. Ejercicios de Refuerzo. Vector de Medias..... 1146
3.2.2.1.1. Ejercicios de Refuerzo. Diferencia de medias multivariada..... 1153
3.2.2.1.2. Ejercicios de Refuerzo. Diferencia de medias multivariada..... 1171
3.2.2.1.3. Ejercicios de Refuerzo. Diferencia de medias multivariada..... 1188
3.2.3.1.1. Ejercicios de Refuerzo. Manova..... 1205
3.2.3.1.1. Ejercicios de Refuerzo. Manova con Minitab 17..... 1223
3.2.3.1.2. Ejercicios de Refuerzo. Manova..... 1243
3.2.3.1.2. Ejercicios de Refuerzo. Manova con Minitab 17..... 1260
3.2.3.1.3. Ejercicios de Refuerzo. Manova..... 1279
3.2.3.1.3. Ejercicios de Refuerzo. Manova con Minitab 17..... 1298
Autoevaluación con reactivos de Falso y Verdadero..... 1319
Autoevaluación con reactivos de Opción Múltiple..... 1319

CAPÍTULO 1. INTRODUCCIÓN AL ANÁLISIS MULTIVARIANTE

OBJETIVO 1.1. El alumno entenderá los conceptos del análisis multivariante en situaciones del área administrativa así como los objetivos del mismo y los tipos de técnicas multivariantes existentes.

ANTECEDENTES	**CONCEPTOS DE:** Método Científico**,** protocolo de investigación, recolección y compendio de datos, organización de conjuntos de datos, diseño de experimentos y reconocimientos, medición de la variación, tanto de datos experimentales como de reconocimiento, estimación de parámetros de población y suministro de varias medidas de la exactitud y precisión de esas estimaciones, ensayo de hipótesis respecto a poblaciones, relación entre dos variables.

1.1.1

INTRODUCCIÓN AL ANÁLISIS MULTIVARIANTE.

CONCEPTOS
BÁSICOS
ANÁLISIS
MULTIVARIANTE

En el análisis multivariado tanto la variable dependiente como las independientes pueden ser categóricas o cuantitativas. Las variables categóricas deben ser dicotómicas o ser transformadas en dicotómicas. Las variables independientes pueden ser:

Todas cuantitativas.

Todas categóricas.

Algunas cuantitativas y otras categóricas

Variables dummy Dijimos más arriba que las variables cualitativas deben ser dicotómicas o deben transformarse en

1.- DEFINICIONES:

1.- Conjunto de métodos ó técnicas diseñados con el fin de examinar e interpretar la información contenida en un conjunto de variables, sin perder la interacción ó grado en que se afectan unas con otras. [1]

2.- Es el conjunto de métodos estadísticos cuya finalidad es analizar simultáneamente conjuntos de datos multivariantes en el sentido de que hay varias variables medidas para cada individuo u objeto estudiado.[2]

3.- Todos los métodos estadísticos que analizan simultáneamente medidas múltiples de cada individuo u objeto sometido a investigación.[3]

2.- OBJETIVOS DEL ANÁLISIS MULTIVARIANTE:

1.- Proporcionar métodos cuya finalidad es el estudio conjunto de datos multivariantes que el análisis estadístico uni y bidimensional es incapaz de conseguir.

2.- Ayudar al analista o investigador a tomar decisiones óptimas en el contexto en el que se encuentra teniendo en cuenta la información disponible por el conjunto de datos analizado.

3.- TIPOS DE TÉCNICAS MULTIVARIANTES:

En general se pueden clasificar en dos grandes grupos:

1) Métodos de dependencia.
2) Métodos de Interdependencia.

1.- Métodos de dependencia:

Supone que las variables analizadas están divididas en dos grupos: *las variables dependientes y las variables independientes.* El objetivo de los métodos de dependencia consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma y cuya relación causa-efecto se puede representar por medio de una función ó ecuación.

Dentro de los métodos de dependencia se pueden ubicar las siguientes técnicas:

[1] Jorge de la Garza García (1995) Análisis de la información Mercadológica a través de la estadística multivariante. Ed. Alambra Mexicana.
[2] Manuel Salvador Figueras. Introducción al Analisis Multivariante
[3] Hair, Anderson, Tatham y Black (1999) Análisis Multivarriante. Prentice Hall

dicotómicas. Para las variables independientes esta transformación se efectúa creando las así llamadas variables dummy, que son variables dicotómicas inexistentes pero que las creamos para poder efectuar el análisis estadístico.

El Análiis de regresión genera una ecuación para describir la relación estadística entre uno o más predictores y la variable de respuesta y para predecir nuevas observaciones. La regresión generalmente utiliza el método de mínimos cuadrados ordinarios, del cual se obtiene la ecuación al minimizar la suma de los residuos cuadrados.

Los resultados de regresión indican la dirección, el tamaño y la significancia estadística de la relación entre un predictor y una respuesta.

· El signo de cada coeficiente indica la dirección de la relación.

· Los coeficientes representan el cambio de la media en la respuesta para una unidad de cambio en el predictor mientras mantiene constantes otros predictores en el modelo.

· El valor p de cada coeficiente prueba la hipótesis nula de que el coeficiente es igual a cero (no tiene efecto). Por lo tanto, los valores p bajos sugieren que el predictor es una adición significativa

1.1. Análisis de regresión y correlación múltiple.

Su objetivo consiste en obtener, a partir de los datos, una ecuación que permita conocer la relación causa-efecto que se presenta entre las variables independientes y una dependiente y cuantificar el grado de relación que existe entre las variables independientes con respecta a la dependiente.

Las variables que intervienen en ambas técnicas son de naturaleza *cuantitativa* es decir pueden medirse en forma numérica.

Ejemplo: Predecir el gasto anual en el cine de una persona a partir de su nivel de ingresos, nivel educativo, sexo y edad.

1.2. Análisis discriminante.

Esta técnica permite explicar las diferencias entre grupos de personas u objetos, por medio de variables independientes, con el objeto de llegar a clasificar nuevas observaciones en los grupos establecidos.

El análisis discriminante maneja "n" variables independientes *cuantitativas* todas ellas y *una* variable dependiente *cualitativa* dicotómica ó que tenga como máximo cuatro categorías.

Ejemplo: Distinción entre usuarios habituales u ocasionales de un producto con base en el precio, calidad, etc.

1.3. Análisis de varianza y covarianza.

La técnica de análisis de varianza, conocida también como ANOVA, se utiliza básicamente para la experimentación. Aquí se pueden manejar un conjunto de variables independientes de naturaleza *cualitativa*, para explicar otro conjunto de variables dependientes de naturaleza *cuantitativa.*

Ejemplo: Determinar si la posición en que se puede colocar un producto en un estante (alta, media ó baja) afecta las ventas de dicho producto.

En el ejemplo anterior la posición es la variable independiente cualitativa y las ventas la dependiente cuantitativa.

El análisis de covarianza, conocido como ANCOVA, se aplica en conjunto con ANOVA para experimentos en los que existen variables que quedan fuera de control del investigador y que afectan directamente las mediciones de las variables dependientes. Con ANCOVA se trata de encontrar como afecta la variable incontrolable a la variable dependiente, para aislar dicho efecto y por decirlo así controlarlo en cierto modo, después de llevar acabo el experimento.

1.4. Análisis de correlación canónica.

Esta técnica consiste en determinar el grado de relación entre dos conjuntos de variables.

Su objetivo es relacionar simultáneamente varias variables métricas dependientes e independientes calculando combinaciones lineales de cada conjunto de variables que maximicen la correlación existente entre los dos conjuntos de

a su modelo.

· La ecuación predice nuevas observaciones dados valores predictores específicos.

Utilice el Análisis discriminante para clasificar las observaciones en dos o más grupos si usted tiene una muestra con grupos conocidos. El Análisis discriminante también se puede utilizar para investigar en que manera las variables contribuyen a la separación de grupos.

ANCOVA o análisis de covarianza es una extensión del análisis de varianza (ANOVA) que permite modelar y realizar ajustes para las variables de entrada que se midieron, pero que no se aleatorizaron o controlaron en el experimento. ANCOVA prueba si los factores tienen un efecto después de eliminar la varianza debido a las covarianzas.

Se utiliza el Análisis de Componentes Principales como una ayuda para comprender la estructura de datos subyacente y/o conformar un número menor de variables no correlacionadas (por ejemplo, para evitar la multicolinealidad en la regresión).

Se utiliza un Análisis de Factoresal igual que el análisis de componentes principales, a fin de resumir la estructura de covarianzas de los datos en algunas dimensiones de los

variables.

Ejemplo: Analizar como están relacionadas el tiempo dedicado al trabajo y el ocio de una persona con su nivel de ingresos, su edad y su nivel de educación.

Casos particulares de él son el coeficiente de correlación y el coeficiente de determinación; la diferencia es que éstos se refieren a las relaciones de variables de "una a una" y de " una a muchas", respectivamente y en la correlación canónica se busca la relación de "muchas a muchas".

2.- Métodos de interdependencia ó estructurales:

Estos métodos no distinguen entre variables dependientes e independientes y su objetivo principal es resumir información. Consiste en identificar que variables están relacionadas, como lo están y porqué.

Ejemplo: Describir el comportamiento del mercado con un menor número de variables.

En estas técnicas, todas las variables se manejan como independientes, pues a diferencia de las técnicas dependientes, no trata de buscar relaciones de causa-efecto entre las variables, sino que busca algo común entre ellas para unirlas y así resumir la información que se tiene.

Algunas de estas técnicas son las siguientes:

2.1. Análisis de componentes principales o de factor común.

Conocida como análisis de factores, se utiliza para analizar las relaciones entre un gran número de variables métricas y así explicarlas solamente por medio de factores (también llamados atributos, características ó dimensiones), los cuales describen al grupo.

Ejemplo:

Si se tratara de evaluar el desempeño en la operación de un restaurante y se estuvieran manejando las variables sabor, cantidad servida, rapidez, amabilidad, variedad del platillo, limpieza, iluminación, ambiente, temperatura y precio, éstas mediante el análisis de factores, tal vez quedarían resumidas bajo las siguientes dimensiones:

- SERVICIO: rapidez y amabilidad.
- LOCAL: limpieza, iluminación, ambiente y temperatura.
- CALIDAD DEL PRODUCTO: sabor, cantidad servida, variedad del platillo.
- PRECIO.

2.2. Análisis de conglomerados o de agrupamiento o análisis cluster.

Esta técnica se usa principalmente en la segmentación de mercados, con base en similitudes; y consiste en formar grupos conteniendo elementos lo más parecido posibles entre sí y lo más diferente posible al compararse con otros. Puede servir como entrada al análisis discriminante, porque los grupos se forman con el análisis de agrupamiento y la diferencia entre ellos se encuentra con el discriminante, o sea,

datos. Sin embargo, el énfasis en el análisis factorial es la identificación de los "factores" subyacentes que pudieran explicar las dimensiones asociadas con una variabilidad grande de los datos.

Se utiliza el Análsiis de conglomerados o Cluster Analisis para clasificar las observaciones en grupos cuando los grupos se desconocen inicialmente.

Este procedimiento utiliza un método jerárquico aglomerativo que comienza con todas las observaciones separadas, cada una de ellas formando su propio conglomerado. En el primer paso, las dos observaciones más cercanas entre sí se unen. En el siguiente paso, una tercera observación se une a las dos primeras u otras dos observaciones se unen para formar un conglomerado diferente. Este proceso continuará hasta que todos los conglomerados se unan en un solo conglomerado; sin embargo, este conglomerado individual no es útil para propósitos de clasificación Por lo tanto, usted debe decidir cuántos grupos son lógicos para sus datos y clasificarlos adecuadamente.

que este último comprueba si en realidad segmentados correctamente en el agrupamiento.

La diferencia de esta técnica con el análisis de factores consiste en que se trata de agrupar a objetos ó individuos con base en sus actitudes o comportamientos comunes, es decir, se busca agrupar a las personas según sus respuestas que demuestren comportamientos comunes y *no* trata de agrupar variables como el análisis de factores. El de agrupamiento tiene como objetivo clasificar una muestra de entidades en grupos mutuamente excluyentes con base en la similitud de dichas entidades.

Por otra parte, a diferencia del análisis discriminante, en el de agrupamiento los grupos no se encuentran predefinidos, sino que se busca identificarlos.

El análisis de agrupamiento generalmente involucra dos etapas. En la primera se busca medir la similitud o asociación entre las entidades, para determinar en cuantos grupos se hará la clasificación. En una segunda etapa se trata de definir el perfil de las personas; esto se puede hacer aplicando el discriminante a los grupos identificados en el agrupamiento.

Ejemplo: Clasificar grupos de alimentos (pescados, carnes, vegetales y leche) en función de sus valores nutritivos.

BIBLIOGRAFIA:

1. Hair, Anderson, Tatham y Black (1999) Análisis Multivariante. Prentice Hall
2. Jorge de la Garza García (1995) Análisis de la información Mercadológica a través de la estadística multivariante. Ed. Alhambra Mexicana.
3. Kinner, C.T. y Taylor (1993) Investigación de mercados. Mc Graw hill, México.

OBJETIVO 1.2. El alumno entenderá el significado del valor teórico como elemento esencial del análisis multivariante así como las escalas de medición y su relación con las técnicas multivariantes.

ANTECEDENTES

CONCEPTOS DE:

Experimento. Unidad experimental. Medidas. Variable de respuesta. Ensayos ó réplicas. Aleatorización. Agrupamiento. Bloqueo. Balanceo. Factores controlados. Factores no controlados. Tratamientos ó niveles de un factor. Error experimental. Efectos del tratamiento. Variación total. Variación entre tratamientos. Variación dentro de tratamientos. Análisis de varianza (ANOVA).

1.2.1 VALOR TEÓRICO, ESCALAS DE MEDICIÓN Y SU RELACIÓN CON LAS TÉCNICAS MULTIVARIANTES.

CONCEPTOS BÁSICOS

VALOR TEÓRICO

VALOR TEÓRICO, ESCALAS DE MEDICIÓN Y SU RELACIÓN CON LAS TÉCNICAS MULTIVARIANTES.

Aunque el análisis multivariante tiene sus orígenes en la estadística uni y bivariante, la extensión al dominio multivariante introduce conceptos y cuestiones adicionales. Estos conceptos van desde la necesidad de entender el concepto del elemento básico del análisis multivariante - el valor teórico - a las cuestiones específicas acerca de los tipos de escalas de medida utilizadas y los resultados estadísticos de los tests de significación y los intervalos de confianza. Cada concepto juega un papel importante en la correcta aplicación de cualquier técnica

Aunque el análisis multivariante tiene sus raíces en la estadística univariante y bivariante, la extensión al dominio multivariante introduce conceptos y cuestiones adicionales, que van desde el "valor teórico" hasta las escalas de medida utilizadas, los errores de medición, los resultados estadísticos de las pruebas de significación y los intervalos de confianza. La utilización de un modelo multivariante conlleva la elaboración de un plan de investigación bien definido que incluye los objetivos analíticos en términos conceptuales, la selección de la técnica, la evaluación de los supuestos básicos de dicha técnica, la estimación del modelo y su interpretación, para finalizar con la aplicación de las técnicas de validación para determinar la estabilidad de los resultados obtenidos

Como posible clasificación, según el grado de información que contienen unas variables, se pueden dividir a éstas en: (i) Variables Nominales:Sólo

multivariante.

El valor teórico.

El elemento esencial del análisis multivariante es el *valor teórico*, una combinación lineal de variables con ponderaciones determinadas empíricamente. El investigador especifica las variables, mientras que las ponderaciones son objeto específico de determinación por parte de la técnica multivariante. Un valor teórico de "*n*" variables ponderadas (X_1 a X_n) puede expresarse matemáticamente así:

$$\text{Valor teórico} = w_1X_1 + w_2X_2 + w_3X_3 + \ldots + w_nX_n$$

Donde X_n es la variable observada y w_n es la ponderación determinada por la técnica multivariante.

El resultado es un valor único que indica una combinación de *todo el conjunto* de variables que mejor se adaptan al objeto del análisis multivariante específico. En **regresiones múltiples**, el valor teórico se determina de tal forma que guarde la mejor correlación con la variable que se está prediciendo. En el **análisis discriminante**, el valor teórico se forma de tal manera que produzca resultados para cada observación que diferencien de forma máxima entre grupos de observaciones y en el **análisis factorial**, los valores teóricos se forman para representar mejor las estructuras subyacentes o la dimensionalidad de las variables tal y como se representan en sus intercorrelaciones.

El valor teórico capta el carácter multivariante del análisis por lo que es el punto central del análisis ya que se debe entender no solo su impacto conjunto para lograr el objetivo de cada técnica, sino también la contribución de cada variable separada al efecto del valor teórico en su conjunto.

Escalas de medida.

La *medida* es importante para representar con precisión el concepto de interés y es vital en la selección del método de análisis multivariante apropiado. El análisis de los datos trae consigo la separación, identificación y medida de la variación en un conjunto de variables, tanto entre ellas mismas como entre una variable dependiente y una o más variables independientes por lo tanto se requiere entender cómo se miden las variables para determinar qué técnicas multivariantes pueden aplicarse a las investigaciones.

Existen dos tipos de variables:

Variables cualitativas o no métricas. Estas variables describen o identifican a un objeto o individuo tomando un atributo, una característica o propiedad categórica, es decir *no* numérica.

Estas variables se pueden medir con base en dos tipos de *escalas de*

distinguen entre varias categorías, sin que exista ninguna jerarquía entre ellas; (ii) Variables Ordinales Además de distinguir distintas categorías para una variable, se puede distinguir una relación de orden entre ellas; (iii) Variables de Intervalo Además de contener las características de las dos anteriores (distingue entre valores y entre la distinta jerarquía de valores) añade el hecho de dotar de sentido a la diferencia entre los valores de la variable. Es decir, la distancia o diferencia entre dos valores consecutivos de la variable es siempre el mismo; (iv) Variables de razón. Son idénticas a las anteriores salvo que presentan un origen absoluto de medida. En estas variables tiene sentido tomar fracciones de sus valores o razones. Se puede decir que un valor es el doble que otro.

medición:

1.- *Nominal.* Esta escala asigna un número a una característica ó atributo del objeto o persona, con el fin de manejarla en forma cuantitativa.

Por ejemplo:

Género: M___ F___

En el caso de la variable "género", los valores que se toman no son cuantitativos sino cualitativos (femenino, masculino). Para manejarla de manera cuantitativa se asigna un número, por ejemplo, el 1 a *m* (masculino) y el 2 a *f* (femenino).

2.- *Ordinal.* Esta escala de medición gradúa los datos con base en la jerarquización de las marcas ó clases evaluadas, según la cantidad de atributos que el entrevistado otorga al objeto ó persona.

Ejemplo:

Organice los siguientes productos por orden de importancia, de acuerdo con su gusto por las siguientes marcas de cigarrillos (1º = me gusta más que todos, 5º = que gusta menos que todos):

___ Marlboro

___ Raleigh

___Vantage

___Viceroy

___Montana

Los métodos estadísticos multivariados se pueden seleccionar o clasificar teniendo en cuenta varios aspectos
pero todos ellos deben incluir: a) la estructura de la matriz de datos, b) el objetivo perseguido, y c) la naturaleza de esos datos (Dagnelie, 1981).

Variables cuantitativas o métricas. Estas variables identifican a un objeto o individuo en forma cuantitativa o numérica, pues se puede decir que es diferente de otro en cantidad o grado.

Estas variables se pueden medir con las siguientes *escalas de medición:*

1.- *Intervalo.* A diferencia de la escala ordinal, permite diferenciar a un objeto o persona de otra en cantidad o grado, indicando que al pasar de un valor a otro de la escala, el atributo está cambiando en la misma proporción. En la escala de intervalo, el cero no es absoluto, no significa " ausencia de algo ".

a) Según la estructura de la matriz de datos, los métodos pueden clasificarse según sean las

Las escalas de intervalo más familiares son las escalas de temperatura Celsius y Fahrenheit. Ambas tienen un punto de cero arbitrario, pero ese cero no indica una cantidad cero o ausencia de temperatura, dado que podemos registrar

variables o los individuos de la matriz de datos de base. La estructura se refiere a si las variables o los individuos son diferentes, o si pertenecen a un grupo o a más grupos de variables o de individuos:
- sin ninguna estructura en particular, (análisis de componentes principales y análisis factorial; conglomerados)
- una estructura entre variables, (métodos de regresión múltiple ó análisis de correlación canónica)
- una estructura entre individuos, (análisis discriminante)
- ambas estructuras (análisis de correspondencias múltiples)

b) Según el objetivo perseguido, los métodos son muy difíciles de clasificar, pues puede haber muchos y muy diferentes, pero los agruparemos en dos grandes grupos: los descriptivos, y los inferenciales.

c) Según la naturaleza de los datos
- En el caso de los métodos descriptivos:
- si las "p" variables son cuantitativas (Análisis Factorial Clásico (Análisis de Componentes Principales y Análisis Factorial común)
- si las "p" variables son cualitativas y/o cuantitativas (Métodos de Conglomerados (clusters) y

temperaturas por debajo del punto cero de esa escala.

2.- *Razón.* Esta escala es muy similar a la de intervalo, con la diferencia de que el cero sí tiene un significado absoluto de "ausencia de". Por lo tanto, no se tiene que establecer un significado para cada número y la escala es comparable entre sí, ya que el cero es absoluto.

Por ejemplo:

Estatura ____

Clasificación de las técnicas del análisis multivariante .[4]

Existen muy diversas técnicas multivariantes para analizar información mercadológica. Para llegar a clasificar una técnica y así saber cuál es factible para ser utilizada, se deben considerar los siguientes aspectos:

1.- Primero, debe verificarse si, de acuerdo con la naturaleza de la investigación, el conjunto de variables puede ser dividido en independientes y dependientes.

2.- Si la respuesta fue afirmativa, debe cuestionarse cuántas variables dependientes serán manejadas

3.- Por último, debe identificarse el tipo de escala de medición en el que se están manejando las variables.

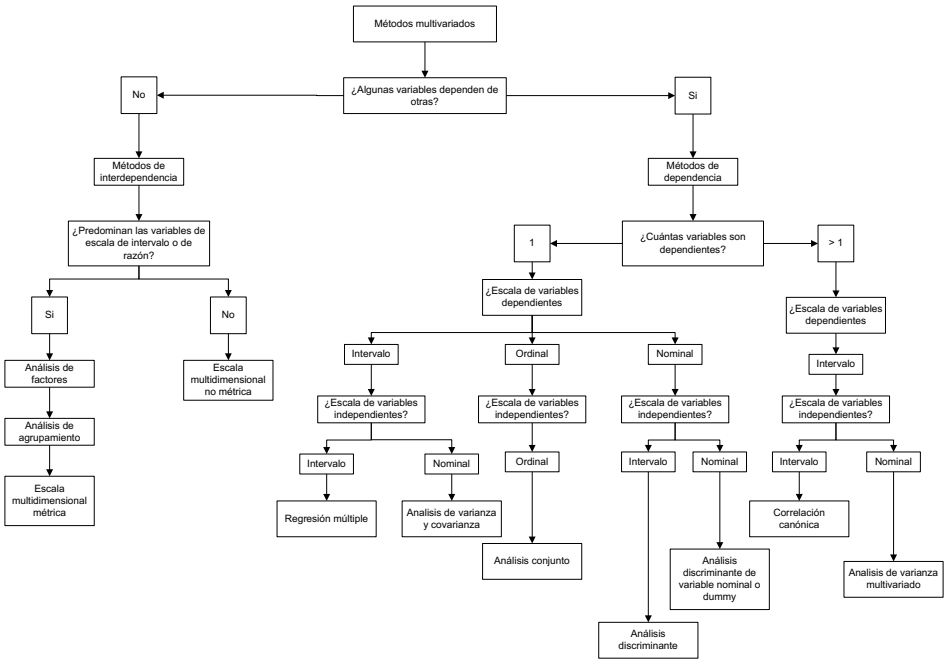
A continuación se presenta la clasificación de las técnicas de multivariante, considerando los tres aspectos explicados anteriormente.

[4] Jorge de la Garza García (1995) Análisis de la información Mercadológica a través de la estadística multivariante. Ed. Alambra Mexicana.

Análisis de Correspondencias (Simple y Múltiple)

- En el caso de los métodos inferenciales:

En este caso, siempre hay dos grupos de variables y casi siempre se reconocen como variables independientes y variables dependientes . Por esto, tenemos que tener en cuenta la naturaleza y la cantidad de variables de cada uno de los grupos.



BIBLIOGRAFIA:

1. Hair, Anderson, Tatham y Black (1999) Análisis Multivariante. Prentice Hall
2. Jorge de la Garza García (1995) Análisis de la información Mercadológica a través de la estadística multivariante. Ed. Alhambra Mexicana.
3. Kinner, C.T. y Taylor (1993) Investigación de mercados. Mc Graw hill, México.

1.2.2 BASE DE DATOS. FORMATO DE ENCUESTA SOBRE BIENES RAICES. CÓDIGO PARA LA COMPUTADORA. CONSTRUCCIÓN DE LA BASE DE DATOS "BIENES1".

CONCEPTOS BÁSICOS BASE DE DATOS. 1.2.2.

Una base de datos es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso. En este sentido; una biblioteca puede considerarse una base de datos compuesta en su mayoría por documentos y textos impresos en papel e indexados para su consulta.

Una **BASE DE DATOS** es una colección de información organizada de forma que un programa de ordenador pueda seleccionar rápidamente los fragmentos de datos que necesite. Una base de datos es un sistema de archivos electrónico.

Las bases de datos tradicionales se organizan por campos, registros y archivos. Un **campo** es una pieza única de información; un **registro** es un sistema completo de campos; y un **archivo** es una colección de registros. Por ejemplo, una guía de teléfono es análoga a un archivo. Contiene una lista de registros, cada uno de los cuales consiste en tres campos: nombre, dirección, y número de teléfono.
A veces se utiliza DB, de *database* en inglés, para referirse a las bases de datosActualmente, y debido al desarrollo tecnológico de campos como la informática y la electrónica, la mayoría de las bases de datos están en formato digital, siendo este un componente electrónico, por tanto se ha desarrollado y se ofrece un amplio rango de soluciones al problema del almacenamiento de datos.

Las bases de datos se utilizan primordialmente para almacenar datos históricos o actuales que posteriormente se pueden utilizar para estudiar el comportamiento de un conjunto de datos a través del tiempo, realizar proyecciones, tomar decisiones y realizar análisis de datos para inteligencia empresarial o de tipo científico en investigaciones.

Entre las principales características de los sistemas de base de datos podemos mencionar:

- Independencia lógica y física de los datos.
- Redundancia mínima.
- Acceso concurrente por parte de múltiples usuarios.
- Integridad de los datos.
- Consultas complejas optimizadas.
- Seguridad de acceso y auditoría.
- Respaldo y recuperación.
- Acceso a través de lenguajes de programación estándar.

Una base de datos posee el siguiente orden jerárquico:

- Tablas
- Campos
- Registros
- Lenguaje

FORMATO DE ENCUESTA DE BIENES RAICES DENOMINADA "BIENES1"

CODIGO (Por favor, **INSERTE** en los espacios correspondientes el número ó valor apropiado)

_ _ _ . _ 1. Valor de avalúo (Dlls. $ 000) _ _ _ . _
1 2 3 4 5

_ _ . _ _ 2. Tamaño del lote (00 de pies cuadrados) _ _ . _ _
7 8 9 10 11

_ 3. Número de recámaras
13

_ . _ 4. Número de baños _ . _
15 16 17

_ _ 5. Número de habitaciones _ _
19 20

_ _ 6. Antiguedad de la casa (en años) _ _
22 23

_ _ _ _ 7. Impuestos anuales. ($) _ _ _ _
25 26 27 28

_ 8. Tipo de estacionamiento interior: (0) Ninguno (1) Cochera de un automóvil
30 (2) Cochera de dos automoviles (3) Cochera de tres automóviles.

_ 9. Ubicación geográfica de la casa: (1) Norte (2) Sur (3) Este
32

_ 10. Estilo arquitectónico: (1) Moderno (2) Hacienda
34 (3) Colonial (4) Granja (5) Con desniveles.

_ 11. Tipo de combustible para la calefacción: (1) Gas (2) Petroleo
36

_ 12. Tipo de sistema de calefacción: (1) Aire caliente
38 (2) Agua caliente (3) Otro

_ 13. Tipo de alberca ubicada en la propiedad: (1) Ninguna
40 (2) Sobre el nivel del piso (3) En el piso

_ 14. Cocina con antecomedor: (0) Ausente (1) Presente.
42

_ 15. Aire acondicionado: (0) Ausente (1) Presente
44

_ 16. Chimenea: (0) Ausente (1) Presente.
46

_ 17. Drenaje municipal: (0) Ausente (1) Presente
48

_ 18. Sótano (0) Ausente (1) Presente
50

_ 19. Cocina integral: (0) Ausente (1) Presente
52

_ 20. Tina de hidromasaje: (0) Ausente (1) Presente
54

_ _ _ 21. Código del entrevistado.
56 57 58

EJEMPLO DE CODIFICACIÓN DE LAS RESPUESTAS DEL ENCUESTADO CODIFICADO 001

PREGUNTA	TIPO DE PREGUNTA	CODIGO PARA LA COMPUTADORA	COLUMNA ASIGNADA PARA LA CAPTURA DE DATOS	RESPUESTA DEL ENCUESTADO	RESPUESTA CODIFICADA
1	Valor del Avalúo	**VALOR**	1-5	$ 190,000	190.0
2	Tamaño del Lote	**TAMAÑO**	7-11	690 Pies Cuadrados	6.90
3	Recámaras	**RECAMARAS**	13	4	4
4	Baños	**BAÑOS**	15-17	2	2.0
5	Habitaciones	**HABITACIONES**	19-20	8	8
6	Antigüedad	**ANTIGÜEDAD**	22-23	38	38
7	Impuestos	**IMPUESTOS**	25-28	$ 3,750	3750
8	Estacionamiento	**ESTACIONAMIENTO**	30	Un Auto	1
9	Ubicación	**UBICACIÓN**	32	Este	3
10	Estilo	**ESTILO**	34	Colonial	3
11	Combustible Calefacción	**COMBCALEF**	36	Petroleo	2
12	Sistema Calefacción	**SISTCALEF**	38	Agua Caliente	2
13	Alberca	**ALBERCA**	40	Ninguna	1
14	Cocina con antecomedor	**ANTECOMEDOR**	42	Presente	1
15	Aire Acondicionado	**AIREACOND**	44	Ausente	0
16	Chimenea	**CHIMENEA**	46	Ausente	0
17	Drenaje	**DRENAJEMUN**	48	Presente	1
18	Sótano	**SOTANO**	50	Ausente	0
19	Cocina Integral	**COCINAINTE**	52	Ausente	0
20	Hidromasaje	**HIDROMASAJ**	54	Ausente	0
21	Código entrevistado	**CODIGO**	56-58	001	001

BASE DE DATOS "BIENES 1"

OBS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	190	6.9	4	2	8	38	3750	1	3	3	2	2	1	1	0	0	1	0	0	0	1
2	215	6	2	2	7	30	2856	1	1	4	2	2	1	1	1	0	1	1	0	0	2
3	160	6	3	2	6	35	3240	0	3	1	2	2	1	1	0	1	1	0	1	1	3
4	195	6	5	2	8	35	4000	1	1	4	2	1	1	1	0	0	1	0	1	1	4
5	163	7	3	1	6	39	2700	1	3	1	2	1	1	1	0	0	1	0	0	0	5
6	181	7	4	1.5	7	32	3959	1	2	4	2	2	2	1	0	0	1	1	0	0	6
7	220	9.75	4	1.5	7	60	3142	2	2	3	2	2	1	1	0	0	0	1	0	0	7
8	159.9	6	4	1	7	38	2800	1	3	1	1	2	1	1	0	0	1	0	1	1	8
9	160	6	2	1	7	35	2250	1	1	1	2	1	1	1	0	0	1	1	1	1	9
10	195	6	3	2	7	38	2995	1	3	4	2	2	1	1	0	1	1	0	1	1	10
11	165	9	4	1	6	32	2500	1	3	4	2	2	1	1	0	1	1	0	0	0	11
12	190	7	4	2	6	27	2600	1	2	4	1	1	1	1	0	0	1	1	1	1	12
13	180	11.2	4	1	9	32	2800	1	3	1	2	2	1	1	0	0	1	0	0	0	13
14	181	6	5	2	10	35	2500	0	3	1	2	2	1	1	0	0	1	0	0	0	14
15	160	5.6	3	1	6	37	2027	0	3	1	2	2	1	1	0	1	1	0	1	1	15
16	160	6.5	2	1	5	35	2250	1	2	1	2	2	1	0	0	0	1	1	0	0	16
17	185	6	3	1.5	8	37	2500	1	3	2	2	2	1	1	0	1	1	0	1	1	17
18	160	6	4	1	6	39	2200	1	3	1	2	2	1	1	0	0	1	0	1	1	18
19	176	6	4	2	7	37	3156	0	3	1	2	2	1	1	0	1	1	0	1	1	19
20	179.9	8.7	4	1	6	36	2741	1	3	1	2	2	1	1	0	1	1	0	0	0	20
21	147	6	4	1	6	39	2100	0	3	1	2	2	1	1	0	0	1	0	0	0	21
22	189	6.5	3	2	7	36	2400	0	3	1	2	2	2	1	0	0	1	0	1	1	22
23	176	6	4	1.5	10	30	2857	1	2	1	1	1	1	1	0	1	1	0	1	1	23
24	181	7	3	1.5	7	28	3959	1	2	3	2	2	1	1	0	0	1	1	0	1	24
25	151	6.5	3	1	6	35	2280	0	3	1	2	2	1	0	0	0	1	0	0	1	25
26	189	9.41	3	1.5	8	24	2900	1	1	4	1	1	1	1	0	0	0	1	1	1	26
27	170	6	3	2	8	38	2700	1	3	1	2	2	1	1	0	0	1	0	1	1	27
28	249	11.25	4	2.5	9	22	4600	1	1	3	2	2	1	1	0	0	1	1	1	1	28
29	267	9.88	6	2.5	6	33	4300	0	1	3	2	2	1	1	0	1	1	1	0	0	29
30	185	5.82	3	1	8	27	3700	1	2	4	2	2	1	1	0	0	0	1	0	0	30
31	269.9	10	4	2	6	28	4500	2	2	4	2	2	1	1	1	0	1	1	1	1	31
32	199.9	6	4	2	6	38	2700	1	1	1	2	2	1	1	0	0	0	1	1	1	32
33	180	6	3	1.5	7	35	2300	1	3	4	2	2	1	1	0	1	1	0	1	1	33
34	189.9	7.2	3	2	8	35	2860	1	2	1	2	2	3	1	0	1	0	1	0	0	34
35	299.9	7.5	4	2.5	7	14	4343	1	2	3	2	2	1	1	0	0	0	1	0	0	35
36	179.9	8	3	1	7	35	2911	1	2	5	1	1	2	1	0	0	1	1	1	1	36
37	210	10	3	1.5	6	30	3930	1	2	5	2	2	1	1	1	0	1	1	1	1	37
38	163	6	3	1	7	36	1889	0	3	1	2	2	1	1	0	0	1	0	1	1	38
39	159.9	6	3	1	6	39	2378	0	2	1	1	1	2	1	0	0	0	1	0	0	39
40	165	6.25	3	1	8	30	1800	0	2	4	2	1	1	1	0	1	1	0	1	1	40
41	180	6	4	1	6	35	2900	0	3	4	2	2	1	1	0	1	1	0	1	1	41
42	169	6	3	1	6	37	2671	1	1	4	2	1	1	1	0	0	1	0	1	1	42
43	179	6	3	1	7	35	3060	1	1	4	2	2	1	1	0	0	1	1	0	0	43
44	180	6	4	1	7	39	2400	0	3	1	2	1	1	0	0	0	1	0	1	1	44
45	218	6	3	2	9	31	3100	1	1	5	2	2	1	1	0	0	1	1	0	0	45
46	150	6	5	2	8	39	3177	1	3	1	2	2	2	1	0	0	0	0	1	1	46
47	247	8	4	2	6	30	3100	1	1	1	2	3	2	1	1	1	1	1	1	1	47
48	173.5	6	4	1	7	39	3000	1	3	1	2	2	1	1	0	1	1	0	1	1	48
49	218	7.4	3	2	6	30	3000	1	1	1	2	2	2	1	0	0	1	1	1	1	49
50	152	6	4	2	8	36	2700	1	3	1	2	2	1	1	0	0	0	0	0	1	50

OBS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
51	300	5.18	4	3.5	6	15	4762	2	1	3	2	2	1	1	1	0	1	1	1	1	51
52	185	7	1	1	8	28	2700	1	2	4	2	2	1	1	0	0	1	1	1	1	52
53	200	7.5	4	2	6	36	3606	1	3	2	2	2	1	1	0	1	1	0	1	1	53
54	172	6	3	1	7	35	2752	1	1	4	2	2	1	1	0	0	1	0	1	1	54
55	230	7.2	3	2.5	7	31	3400	1	1	5	2	2	1	1	0	0	1	1	1	1	55
56	235	5	3	1.5	7	34	3148	1	1	1	2	2	1	1	0	0	0	1	1	1	56
57	186	6	5	2	7	30	3344	1	3	3	2	2	1	1	0	0	1	0	1	1	57
58	200	9.9	3	1.5	8	30	3871	1	2	5	2	2	1	1	0	0	1	1	1	1	58
59	190	6	4	1.5	7	44	3300	1	1	1	2	2	1	1	0	0	1	1	1	1	59
60	140	6	3	1	5	38	2450	1	3	1	2	2	1	0	0	0	0	0	1	1	60
61	180	7.2	3	1	8	40	1900	0	1	4	2	3	1	1	0	0	1	0	0	0	61
62	213	7	4	2	6	35	3600	0	1	4	2	2	1	1	0	0	1	1	1	1	62
63	115	18.9	3	1	10	16	2400	0	2	4	2	2	1	1	0	0	0	0	1	1	63
64	200	10.9	5	2	6	28	2540	1	2	4	2	2	1	1	0	0	1	1	0	0	64
65	153	6.18	4	1	8	39	2158	0	3	1	2	2	1	1	0	1	1	0	1	1	65
66	180	9.1	4	1.5	6	33	2853	1	3	2	2	2	1	1	0	1	1	0	1	1	66
67	163	6	4	1	6	38	2160	0	3	1	2	2	1	1	0	1	0	0	1	0	67
68	160	6.5	3	1	6	33	2500	1	3	4	2	2	1	0	0	1	1	0	0	0	68
69	175	6	4	1	6	38	2594	2	1	1	2	2	1	1	0	0	1	1	0	0	69
70	170	5	3	2	6	41	2300	1	1	1	2	3	1	1	0	0	1	1	1	1	70
71	155	6	4	1	6	35	2436	1	3	1	2	2	1	1	0	0	1	0	0	1	71
72	163	8.4	4	1	6	37	2400	1	3	2	2	2	1	1	0	1	1	0	0	1	72
73	160	6	4	2	7	38	2500	1	3	4	2	2	1	1	0	1	1	0	1	1	73
74	175	6	5	2	7	36	3024	1	3	1	2	2	1	1	0	0	1	1	1	1	74
75	177	6.42	4	2	9	39	3100	0	3	4	2	2	2	1	0	1	1	0	1	1	75
76	190	31.5	5	2	6	31	4200	0	2	5	2	2	1	1	0	0	1	1	1	1	76
77	158	6	4	1	7	35	1900	0	3	1	2	2	1	1	0	0	1	0	0	0	77
78	179	6	3	1	7	37	2950	2	3	2	2	2	1	1	0	1	1	0	1	1	78
79	215	7	4	2	7	35	3860	1	1	1	2	1	1	1	0	0	1	1	0	0	79
80	190	8.25	3	2	8	31	3600	0	2	4	2	2	1	1	0	0	1	1	1	1	80
81	170	11.8	4	2	7	34	3500	2	3	3	2	2	1	1	1	0	1	0	1	1	81
82	230	7	3	1.5	10	28	3500	1	1	5	1	1	1	1	0	0	1	1	1	1	82
83	185	16	5	1.5	8	38	3200	2	3	2	2	2	2	1	0	1	1	0	1	1	83
84	230	10	4	1.5	6	61	3638	2	2	3	2	2	1	1	0	1	1	1	0	0	84
85	180	7.2	3	2	8	14	1582	1	1	1	2	2	1	1	0	0	1	1	1	1	85
86	200	10.3	4	1.5	7	37	4079	1	3	3	2	2	1	1	0	1	1	0	1	1	86
87	180	11	4	1	8	40	2500	1	3	1	2	2	2	1	0	0	1	0	1	1	87
88	190	6	4	2	6	38	2890	0	3	1	2	2	1	1	0	0	1	0	1	1	88
89	225	6.86	3	2	5	30	3000	0	1	4	2	2	1	1	0	1	1	1	1	0	89
90	155	6.48	3	1	6	35	2483	1	1	4	2	1	1	1	0	0	0	0	0	1	90
91	179	7	4	1	8	37	2800	1	1	2	2	2	1	1	0	1	0	0	1	1	91
92	218	6	4	2	7	37	3460	0	3	2	2	2	3	1	0	1	1	0	1	1	92
93	173	10.6	4	2	6	36	2634	1	3	1	2	2	1	1	0	0	1	0	1	1	93
94	183	6	3	2	7	39	2708	1	2	4	2	2	2	1	0	0	1	1	1	1	94
95	215	7.7	4	1	6	31	2989	1	2	5	1	1	2	1	0	0	1	1	1	1	95
96	159	7.5	3	1	7	55	1814	1	2	4	2	2	1	0	0	0	1	0	1	1	96
97	195	7.75	3	1.5	6	34	3175	1	1	5	2	2	1	1	0	0	1	1	1	1	97
98	180	6.95	3	2	6	32	2800	1	2	4	2	2	2	1	0	0	1	1	1	1	98
99	161	6	4	1.5	6	36	2400	1	3	4	2	2	1	1	0	1	1	0	1	1	99

100	180	6	4	1	9	25	2750	1	1	4	2	2	1	1	0	0	0	1	0	0	100

OBS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
101	175	6	4	1	7	34	2500	1	2	1	2	2	1	1	0	0	1	1	1	1	101
102	225	6	3	2	10	33	3701	0	1	5	2	2	1	1	0	0	1	1	1	1	102
103	100	12	7	1	8	21	1775	0	2	4	2	2	1	1	0	0	1	1	1	1	103
104	165	6	4	1.5	6	36	2640	0	3	4	2	2	2	1	0	1	0	0	1	1	104
105	173	8	4	1.5	7	36	2700	1	3	4	2	1	1	1	0	1	1	0	1	1	105
106	175	10	4	2	7	34	3000	1	2	1	2	2	1	1	0	0	1	1	1	1	106
107	176	6	3	1.5	8	31	3276	0	3	1	2	2	1	1	0	0	1	0	1	1	107
108	198	6.3	4	1	7	34	3454	1	1	1	2	2	1	1	0	0	1	1	1	1	108
109	190	5	4	2	10	38	2386	1	2	1	2	2	1	1	0	0	1	1	1	1	109
110	185	28.1	6	2	8	44	4000	1	3	3	2	3	1	0	0	1	0	1	0	1	110
111	195	7.75	4	2.5	8	22	3000	1	1	4	2	2	1	1	0	0	1	1	1	1	111
112	224	6	4	2	7	37	3200	2	3	2	2	2	3	1	0	1	1	0	1	1	112
113	190	7	4	2	7	35	3465	1	1	1	2	1	1	1	0	0	1	1	1	1	113
114	245	6	3	1.5	10	30	3900	1	1	5	2	2	1	1	1	0	1	1	0	0	114
115	207	11.7	3	2	9	28	4200	1	1	5	2	1	1	1	0	0	1	1	0	0	115
116	210	6.9	4	1.5	8	24	3500	1	2	5	2	2	2	1	0	0	1	1	0	0	116
117	170	8	4	1	7	35	3000	2	3	4	2	2	1	1	0	1	1	0	0	0	117
118	185	9.01	3	1.5	6	25	2400	1	2	4	2	2	1	1	0	1	0	1	1	1	118
119	170	6	3	2	6	39	3040	2	3	4	2	2	2	1	0	1	1	0	1	1	119
120	199	10	3	2	7	38	3512	2	1	1	2	3	2	1	0	1	1	1	1	1	120
121	175	6	3	2	6	39	2800	1	3	1	2	2	1	1	0	0	1	0	1	1	121
122	190	3.55	4	2	7	35	2300	0	1	1	2	1	1	1	0	0	1	1	0	0	122
123	153	7.2	4	1	7	39	2513	1	3	1	2	2	1	0	0	0	1	0	1	1	123
124	190	5	3	1.5	7	85	1481	1	2	3	2	2	1	1	0	0	1	0	1	1	124
125	170	7.5	4	2	7	35	3616	1	3	1	2	1	1	0	0	0	1	1	1	1	125
126	180	7	4	2.5	6	32	2500	0	2	1	2	2	1	1	0	0	1	1	1	1	126
127	165	6	3	1.5	7	39	2100	0	3	1	2	2	1	1	0	0	1	0	0	0	127
128	193	6	4	2	7	35	3000	0	3	3	2	2	1	1	0	0	1	0	1	1	128
129	170	6	4	2	7	38	2500	0	3	2	2	2	1	1	0	1	1	0	1	1	129
130	188	11.1	4	2	6	36	4089	1	3	1	2	3	1	1	0	0	1	0	1	1	130
131	167	8.2	4	1	9	39	2225	0	3	4	2	2	1	1	0	1	0	0	1	1	131
132	298	12.2	5	2.5	8	24	5200	1	1	3	2	2	1	1	0	0	1	1	1	0	132
133	178	6	4	2	6	36	3250	1	3	4	2	2	3	1	0	1	1	0	1	1	133
134	160	7.8	3	1	7	20	3300	1	2	4	2	2	1	1	0	0	1	1	1	1	134
135	229	7	3	2	7	35	4100	1	1	3	1	2	1	1	0	0	1	0	0	0	135
136	215	7.5	4	2	8	35	3100	1	1	1	2	1	1	1	0	0	1	1	0	0	136
137	180	6	5	2	6	19	3500	0	3	4	2	2	1	1	0	1	1	0	1	1	137
138	190	6.78	4	2	6	39	2500	1	1	1	2	3	1	1	0	0	1	1	1	1	138
139	151	6	3	1	8	37	2074	0	3	1	2	2	1	1	0	0	1	0	1	1	139
140	240	6.48	4	2.5	9	16	4909	2	3	3	2	2	1	1	1	1	0	1	1	1	140
141	290	10.5	4	2.5	9	29	5000	2	1	5	2	2	1	1	0	1	1	1	0	0	141
142	305	12.1	3	2	7	46	3640	0	2	2	2	2	1	1	1	1	1	1	1	0	142
143	190	7	3	1.5	11	28	2948	1	2	5	1	1	2	1	0	0	1	1	1	1	143
144	310	7.32	4	2.5	9	32	3717	1	1	5	2	2	1	1	0	0	1	1	1	1	144
145	220	7	3	1.5	7	30	3200	1	2	5	1	2	1	1	0	0	1	1	0	0	145
146	160	7.85	4	2	6	28	2850	0	3	3	2	2	1	1	0	0	1	0	0	0	146

147	150	6	4	1	9	38	2400	1	3	1	2	2	1	1	0	0	1	0	1	1	147
148	190	7	5	1.5	6	28	4100	1	2	5	1	1	1	1	0	1	1	1	0	0	148
149	165	6	3	1	6	34	2836	1	3	1	2	2	1	1	0	0	1	0	0	0	149
150	160	6	3	1	7	39	2650	1	3	1	2	2	1	1	0	0	1	0	1	1	150
OBS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
151	187	6	3	1	8	36	2700	1	3	1	2	2	2	1	0	0	1	0	0	0	151
152	160	12.1	5	1	7	37	2800	1	3	4	2	2	1	1	0	1	1	0	1	1	152
153	206	11	4	1	6	35	2886	1	1	1	2	2	1	1	0	0	1	1	0	0	153
154	180	8.4	3	2	6	36	2900	0	3	4	2	2	1	1	0	1	1	0	1	1	154
155	163	8.87	3	1	6	37	1900	1	2	4	2	3	1	1	0	1	0	1	0	0	155
156	170	6	4	2	7	32	3000	1	3	4	2	1	1	1	0	1	1	0	0	0	156
157	190	7	4	2	7	28	2935	1	2	1	2	2	1	1	0	0	1	1	1	1	157
158	175	6	4	2	7	37	3000	1	3	2	2	2	1	1	0	1	1	0	1	1	158
159	225	10.4	3	2	6	15	3500	2	1	4	2	2	1	1	0	0	0	1	1	1	159
160	160	10	3	1	7	40	2307	1	3	1	2	2	1	1	0	0	1	0	1	1	160
161	175	7	3	1.5	8	31	2850	1	2	5	2	2	1	1	0	0	1	1	1	1	161
162	177	6	5	2	7	39	3200	0	3	4	2	2	1	1	0	0	1	0	1	1	162
163	220	6	3	1.5	7	31	3000	1	1	5	2	2	1	1	0	0	1	1	1	1	163
164	168	6	3	1	6	35	3000	0	3	1	2	2	1	1	0	1	1	0	1	1	164
165	160	6	4	1	7	35	2400	0	3	1	2	2	1	1	0	0	1	0	1	1	165
166	270	10.3	3	2.5	6	26	5400	2	1	5	2	2	1	1	0	0	1	1	0	0	166
167	185	6	3	1	8	30	2500	1	2	4	2	2	1	1	0	0	1	1	1	1	167
168	205	6	4	1	7	36	2827	0	1	1	2	2	1	0	0	0	1	1	0	0	168
169	174	9.12	4	1	7	36	2800	0	3	2	2	2	1	1	0	1	1	0	1	1	169
170	169	6	4	2	6	37	2911	0	3	4	2	2	1	1	0	1	1	0	1	1	170
171	175	7.63	4	2	7	37	2800	1	3	2	2	2	1	1	0	1	1	0	1	1	171
172	190	9	3	2	6	39	2945	1	2	1	2	2	1	1	0	0	1	1	0	0	172
173	167	6	4	1.5	7	36	2400	1	3	4	2	2	1	1	0	1	1	0	1	1	173
174	193	6	3	2.5	6	27	3450	1	3	5	2	2	1	1	0	1	1	1	0	1	174
175	155	6	3	1	8	35	2200	1	3	1	2	2	1	1	0	0	1	1	1	1	175
176	240	7	4	2.5	8	29	4100	1	1	5	2	2	1	1	1	0	1	1	1	1	176
177	166	7.3	4	1	7	35	2850	0	3	1	2	2	1	1	0	0	1	0	1	1	177
178	210	12.2	5	2	7	30	4400	1	1	1	2	2	3	1	0	0	1	1	0	0	178
179	180	6	4	2	6	38	3048	1	3	4	2	2	1	1	0	1	1	0	1	1	179
180	154	8.4	4	2	7	40	2550	0	3	1	2	2	1	1	0	0	1	0	0	0	180
181	177	7.5	3	1	9	65	1300	1	2	3	2	3	1	1	0	0	1	1	0	0	181
182	214	5.5	4	2	6	23	3800	1	3	5	2	2	1	1	0	1	1	1	0	0	182
183	165	6	3	1	7	39	2500	1	3	1	2	2	1	1	0	0	1	0	1	1	183
184	184	9.1	3	1.5	5	30	4134	1	2	5	2	2	1	1	0	1	1	1	0	0	184
185	149	6	3	1	7	37	2000	0	3	1	2	2	1	1	0	0	1	0	0	0	185
186	215	7.48	3	2	6	35	3200	1	1	1	2	2	1	1	0	0	1	1	0	0	186
187	159	6	4	2	10	30	2400	1	3	4	2	2	1	1	0	1	1	0	0	0	187
188	205	8	5	2	8	36	2900	0	3	1	2	2	3	1	0	1	0	1	1	1	188
189	210	7.2	4	2	7	30	3600	1	1	1	2	2	1	1	0	0	1	1	0	0	189
190	171	10	3	1.5	6	30	3115	1	2	5	2	1	2	1	0	0	1	1	1	0	190
191	185	6.25	3	1	9	80	1000	0	2	3	1	2	1	1	0	0	1	1	1	1	191
192	189	8	3	1.5	7	32	3698	1	2	5	2	2	2	1	0	1	1	1	0	0	192
193	176	6	3	2	6	34	2800	1	3	1	2	2	1	1	0	1	1	0	1	1	193
194	229	10.8	4	1	8	36	3100	0	1	1	2	2	1	1	0	1	1	0	1	1	194

195	188	7	4	1.5	7	30	3815	1	2	5	2	2	1	1	0	0	1	1	1	0	195
196	230	7.3	3	1.5	8	31	3814	1	2	5	2	2	2	1	0	0	1	1	1	1	196
197	280	8	3	3	9	32	3200	1	1	5	2	2	1	1	1	0	1	1	1	1	197
198	198	6	5	3	6	33	2800	0	2	1	2	2	1	0	0	0	1	1	1	1	198
199	153	6	3	1.5	10	36	2650	0	3	1	2	2	1	1	0	1	1	0	1	1	199
200	200	11	3	2	7	33	2767	0	2	4	2	1	2	1	1	1	1	1	0	0	200
OBS	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
201	280	6	3	2.5	6	30	4058	2	1	5	2	2	1	1	0	0	1	1	1	1	201
202	175	6	3	1	8	35	2200	1	1	4	2	1	1	1	0	1	0	0	1	1	202
203	195	6.5	4	2	6	30	3700	0	2	1	2	2	2	1	0	1	1	1	1	1	203
204	185	4.1	3	1	7	38	2589	1	1	1	2	3	1	1	0	0	1	1	1	0	204
205	180	7.2	4	1	7	30	2900	1	1	1	2	2	1	0	0	0	1	1	0	0	205
206	205	7.5	4	2	6	26	3235	0	1	1	2	2	1	1	0	0	1	1	0	0	206
207	180	6	4	2	7	35	2784	1	2	1	2	1	1	1	0	0	1	1	0	0	207
208	190	6	4	2	9	39	3480	1	1	3	2	1	1	1	0	0	1	1	0	0	208
209	225	6.8	4	2.5	7	23	3600	1	2	5	2	2	1	1	0	0	1	1	1	1	209
210	160	6	4	1	8	35	2800	1	3	1	2	2	2	1	0	0	1	0	1	1	210
211	200	7	4	2	5	33	4163	2	2	5	2	1	3	1	0	1	1	1	1	1	211
212	180	8.5	3	1	7	35	2725	1	2	4	2	2	2	1	0	0	1	1	0	0	212
213	195	6	4	2	7	32	3000	0	1	1	2	2	1	1	0	0	1	1	0	0	213
214	170	6	4	2	10	38	2200	1	3	1	2	2	1	1	0	0	1	0	0	0	214
215	260	6.5	3	2.5	7	31	4400	0	1	5	2	2	1	1	1	0	1	1	1	1	215
216	175	6.3	3	1	5	39	3000	1	3	4	2	2	1	1	0	1	0	0	1	1	216
217	168	6	2	1	6	38	2400	1	1	1	2	2	1	1	0	0	1	1	1	1	217
218	180	6	4	1	7	50	2419	1	1	1	2	1	1	1	1	1	1	1	0	0	218
219	185	6	4	1	11	35	2200	1	1	1	1	2	1	1	0	0	1	1	1	1	219
220	287	6	5	2.5	6	32	4000	1	1	5	2	2	3	1	0	0	1	1	1	1	220
221	166	6	3	1	6	45	1600	0	1	1	2	2	1	1	0	0	1	1	1	1	221
222	265	6	3	1	7	31	2900	1	1	5	1	1	1	1	0	0	1	1	1	0	222
223	182	9.97	4	2	7	30	3739	0	2	4	2	2	1	1	0	1	0	1	0	0	223
224	230	6	4	2	6	35	3558	1	1	1	2	2	1	1	0	0	1	1	0	0	224
225	194	8.3	2	2	6	35	3440	1	1	4	2	2	3	1	1	1	1	1	1	1	225
226	295	12.2	5	3	8	28	6470	2	1	2	2	2	1	1	0	1	1	1	0	0	226
227	170	6.5	3	2	6	35	2303	1	1	4	2	2	1	1	0	1	1	0	0	0	227
228	175	5.4	5	1.5	8	37	3500	2	3	3	2	2	1	1	0	0	1	0	1	1	228
229	275	6	3	2	9	32	4400	2	1	5	2	2	1	1	1	1	1	1	1	1	229
230	150	6	3	1	6	37	2300	2	3	1	2	2	1	0	0	0	0	0	0	1	230
231	225	7.7	3	2.5	10	34	4100	1	1	5	2	2	1	1	0	0	1	1	0	0	231
232	178	7	4	2	7	35	3100	1	2	1	2	2	1	0	0	1	1	1	0	0	232
233	232	6.8	4	2.5	8	10	4980	1	2	3	2	2	1	1	0	0	0	1	1	1	233

1.2.2.1.B.D. EJEMPLO ILUSTRATIVO EN MINITAB 17. Base de Datos

EJEMPLO ILUSTRATIVO INTEGRAL EN MINITAB 17. 1.2.2.1. CREACIÓN DE UNA BASE DE DATOS

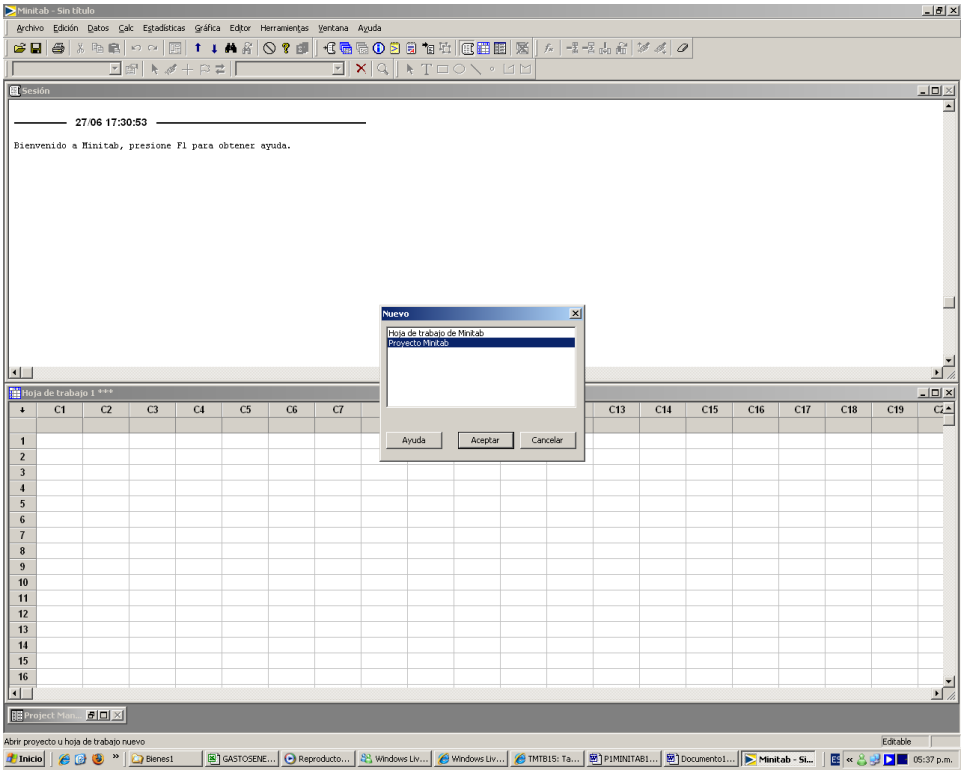
Utilice el menú Archivo para abrir, cerrar, guardar, imprimir o ejecutar diversos tipos de archivo que Minitab puede utilizar.

Tipos de archivo de Minitab que contienen datos

Dos de los tipos de archivos principales que usted utilizará contienen datos:

- Proyectos – contienen hojas de trabajo, junto con la salida de la ventana Sesión y gráficas
- Hojas de trabajo –

Para crear una base de datos en **MINITAB17**, accede a **MINITAB VERSIÓN 17.** Cuando aparezca la pantalla da un clic en la opción **Archivo** de la barra menú y selecciona **Nuevo**. Aparecerá la ventana **Nuevo. Selecciona la opción Proyecto Minitab y Aceptar**



Procede a editar la base de datos de la siguiente manera:

Coloca los nombres de las variables debajo de cada número de variable conforme al código para la computadora de la encuesta:

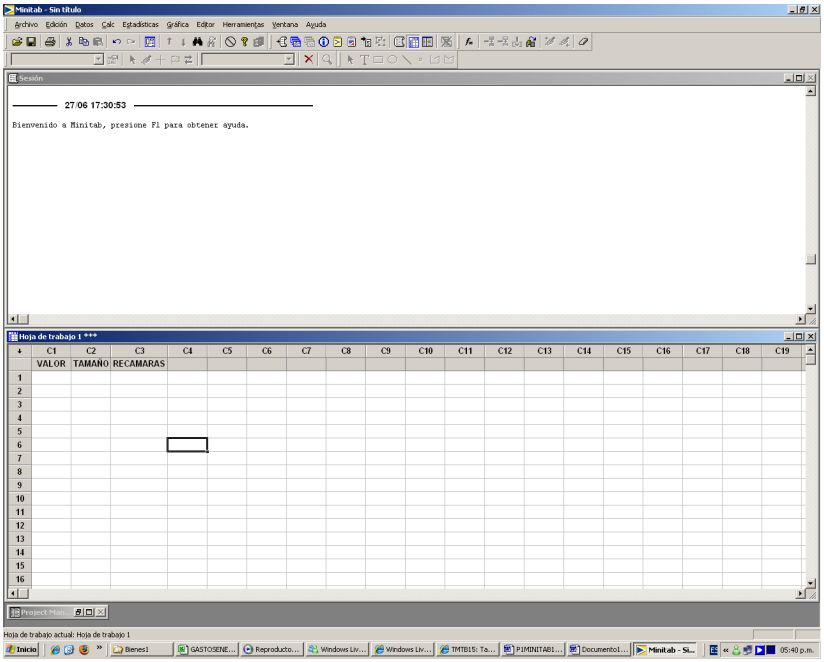
contienen todos sus datos: columnas, constantes y matrices.

Usted puede tener múltiples hojas de trabajo abiertas en un proyecto. Cuando usted abre un archivo de proyecto, todas las hojas de trabajo que estaban dentro de ese proyecto cuando las guardó la última vez están disponibles para usted. Cuando usted guarda un proyecto, las hojas de trabajo se guardan dentro de ese archivo de proyecto.

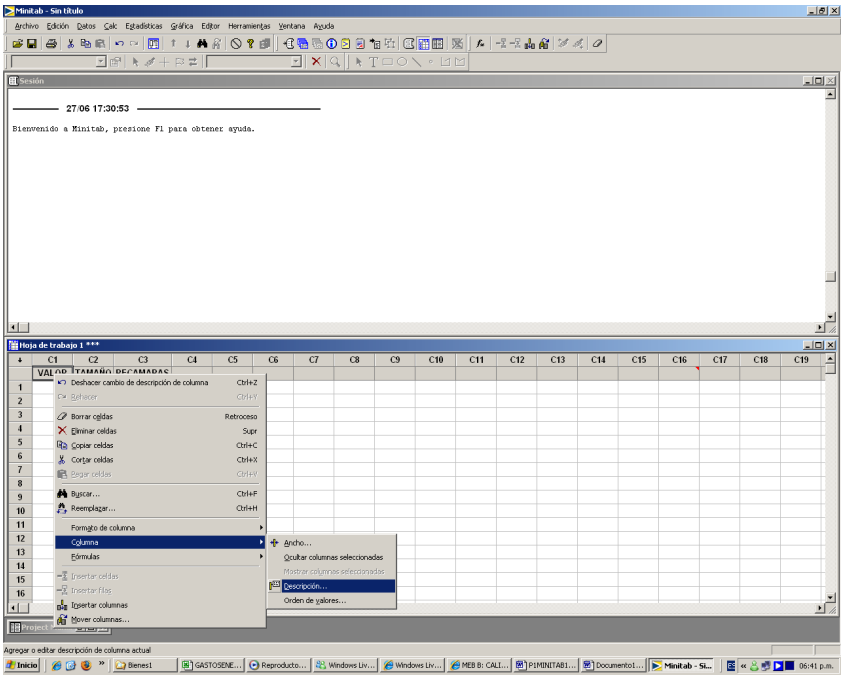
Usted puede agregar hojas de trabajo a su proyecto al copiar datos de un archivo utilizando Archivo > Abrir hoja de trabajo. Observe que se copian los datos. Esto quiere decir que cuando usted cambia los datos dentro de un proyecto, no está afectando al archivo original. Usted puede abrir un proyecto, cambiar datos y guardar el proyecto una y otra vez sin afectar al archivo original.

Usted también puede guardar una hoja de trabajo como un archivo separado que puede utilizarse en otros proyectos de Minitab o en otras aplicaciones. Véase Archivo > Guardar hoja de trabajo actual (como).

Generalmente, los archivos de datos que usted incluye y guarda en su proyecto serán hojas de trabajo de Minitab. Dichas hojas de trabajo pueden ser archivos autónomos (archivos con la extensión MTW) o partes de un archivo de proyecto (MPJ). Usted puede tener una vista preliminar para ver una lista de todas las hojas de trabajo en el archivo y

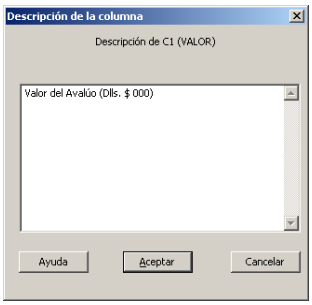


Una vez termines de colocar los nombres de cada **variable (CÓDIGO PARA LA COMPUTADORA)**, coloca el cursor en el primer nombre de la variable y da clic con el botón derecho del mouse. Aparecerá un sub-menú con varias opciones. Selecciona la opción **columna**. Aparece otro sub-menú con varias opciones. Selecciona la opción **descripción**.

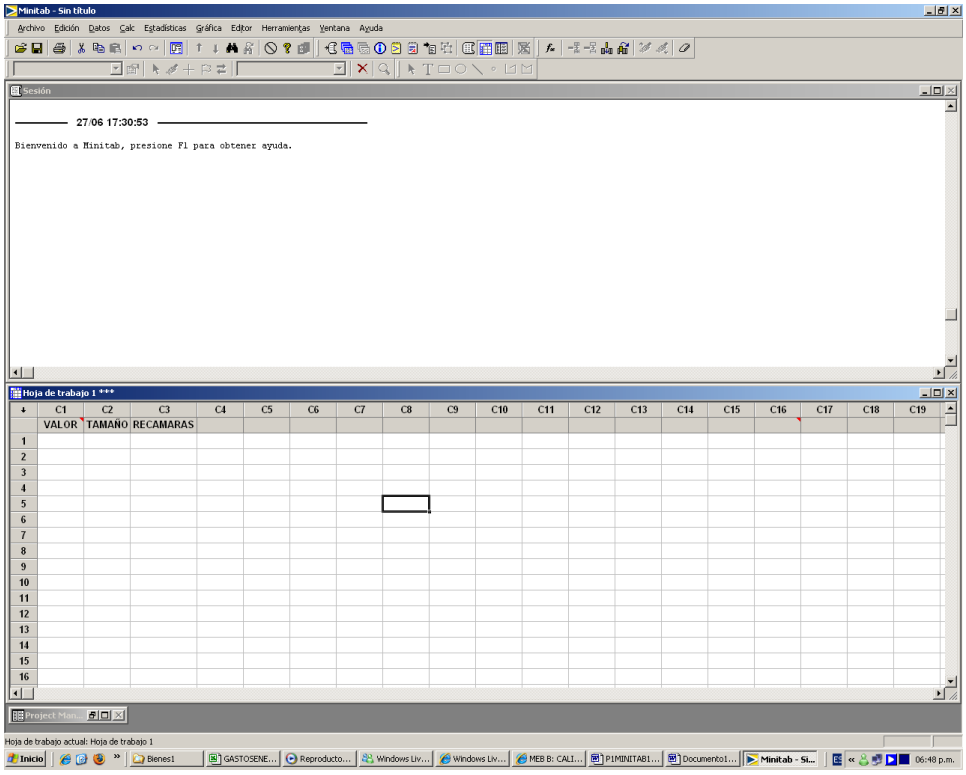


puede utilizar Archivo > Abrir hoja de trabajo para abrir una o más hojas de trabajo de ese proyecto para agregar a su proyecto actual.

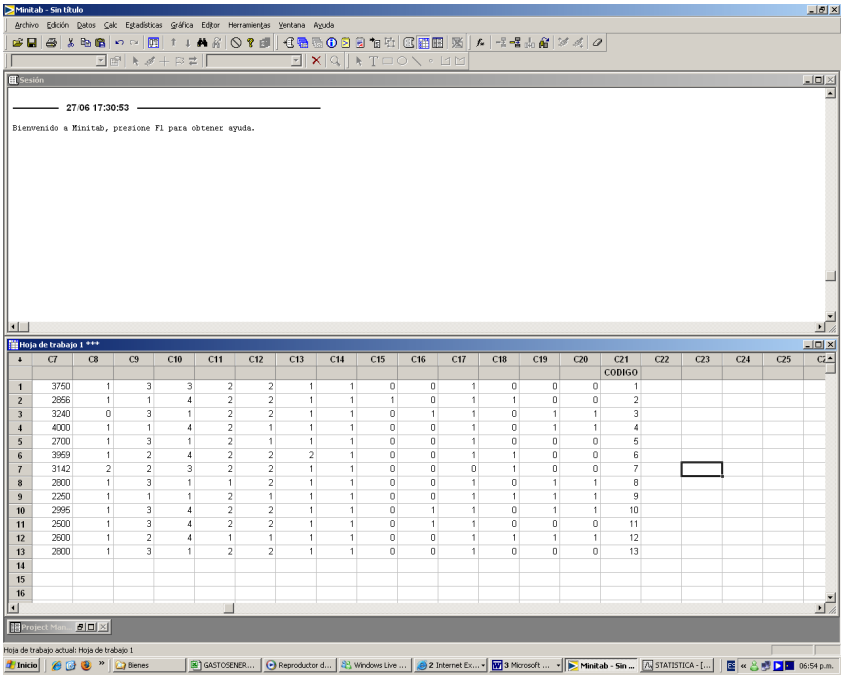
En la ventana de cuadro que aparece escribe la pregunta tal como viene en el cuestionario precodificado.



Una vez escrita la pregunta da **Aceptar** y se colocará un pequeño recuadro rojo que servirá posteriormente para leer el texto de la pregunta cuando pongas el cursor en dicho triángulo.



Repite el procedimiento anterior con todas las variables. Una vez terminado de completar la información en todas las variables, introduce en cada celda el valor correspondiente.

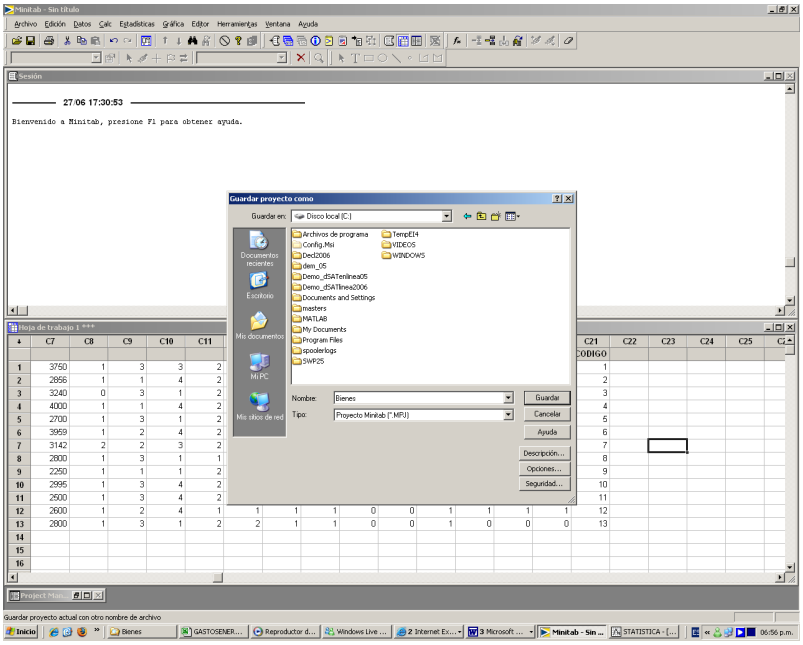


Cuando termines de introducir todos los datos selecciona con el puntero del mouse en el menú de **ARCHIVO** de la barra principal, la opción **Guardar proyecto como** y da un clic con el botón izquierdo del mouse.

El sistema presenta la pantalla **Guardar proyecto como:**

Tipos de archivos que no son de Minitab que contienen datos

Usted también puede abrir y guardar archivos de datos en los formatos de muchas aplicaciones, como Excel y Lotus 1-2-3. Véase Abrir archivos con otras aplicaciones. Para intercambiar datos con otras aplicaciones, tales como programas de computación centrales, usted puede abrir y guardar archivos de texto. Finalmente, usted puede intercambiar datos con las versiones de Minitab en otras plataformas utilizando las hojas de trabajo portátiles de Minitab (archivos MTP). Véase Abrir y guardar con versiones anteriores de Minitab u otras plataformas.



Coloca en el campo **Nombre** el nombre del archivo, en este caso **BIENES1** y la extensión correspondiente que te indique el sistema. En el campo **Guardar en,** selecciona en que disco deseas guardar la información y con el puntero del mouse selecciona el botón **Guardar** de la ventana y da un clic con el botón izquierdo del mouse.

Con lo anterior queda terminada la creación de la base de datos en el paquete **MINITAB17**. Para salir del sistema selecciona con el puntero del mouse en el menú **Archivo** de la barra la opción **Salir** y da un clic con el botón izquierdo del mouse. El sistema te mandará al escritorio de Windows en la pantalla.

OBJETIVO 1.3. El alumno entenderá la importancia de aplicar las técnicas de examen de datos al proceso estadístico multivariado que requiere el análisis de datos ausentes así como la comprobación de los supuestos subyacentes en todos los métodos multivariados.

ANTECEDENTES

CONCEPTOS DE:

Datos estadísticos, concepto de variable y su clasificación, escalas de medición de una variable, muestreo estadístico, error de muestreo, diseño de una muestra, tipos de muestreo, análisis descriptivo, presentación de los datos con una y/o dos variables, estadísticos o estimadores más usuales, medidas de tendencia central, medidas de dispersión, medidas de posición, medidas de sesgo, medidas de curtosis, diagrama de caja y brazos.

1.3.1 ANÁLISIS PREVIO DE LOS DATOS MULTIVARIADOS.EXAMEN GRÁFICO DE LOS DATOS. DATOS AUSENTES. CASOS ATÍPICOS

CONCEPTOS BÁSICOS
EXAMEN GRAFICO DE LOS DATOS.
DATOS AUSENTES.
CASOS ATÍPICOS

ANÁLISIS PREVIO DE LOS DATOS:

Introducción:

La potencia estadística de las técnicas multivariadas requiere grandes conjuntos de datos supuestos más complejos que los que encontramos en el análisis univariante, por ejemplo, los efectos de los datos ausentes, los cuales por definición no se representan directamente en los resultados pueden ser sustanciales por el impacto que producen sobre la naturaleza y carácter de los resultados.

El propósito de este contenido es proporcionar una visión general de las técnicas de examen de datos, que van desde el simple proceso de inspección visual de los gráficos al proceso estadístico Multivariado que requiere el análisis de datos ausentes y la comprobación de los supuestos subyacentes en todos los métodos multivariados.

Aunque puede parecer que es malgastar tiempo, esfuerzo y recursos al proceso de examinar los datos previamente, los problemas "ocultos" que surgen del análisis previo de los datos puede conducir a problemas potencialmente catastróficos que

se pueden evitar al seguir estos análisis previos cada vez que se aplique una técnica multivariada evitando cometer errores serios.

Podemos distinguir cuatro fases distintas en el examen de datos. La primera incluye un examen gráfico de la naturaleza de las variables a analizar y las relaciones que forman las bases del análisis multivariado. La segunda, un proceso de evaluación para entender el impacto que pueden tener los datos ausentes sobre el análisis y alternativas que se pueden usar cuando se presentan en forma relativa. Una tercera fase la forman las técnicas que mejor se ajustan para la identificación de casos atípicos que puedan distorsionar las relaciones sobre una o más variables estudiadas y finalmente una cuarta que consiste en los métodos analíticos necesarios para evaluar la capacidad de los datos para cumplir con los supuestos estadísticos específicos de muchas técnicas multivariadas.

1. EXAMEN GRÁFICO DE LOS DATOS:

Este examen ayuda a la comprensión de las características básicas de los datos y sus relaciones subyacentes, cuando se considera el análisis univariado, el nivel de comprensión puede ser muy simple, pero a medida que el investigador se introduce en los más complejos análisis multivariados, la necesidad y el nivel de comprensión aumentan considerablemente, por lo que es necesario revisar los métodos gráficos básicos que existen para ayudar a la obtención de una comprensión de las características de los datos, particularmente en un sentido multivariado.

La mayoría del software estadístico tiene módulos particularmente comprensivos de técnicas graficas preparadas para el examen previo de los datos que se ven mejorada muchas veces con medidas estadísticas más detalladas de la descripción de los datos.

La naturaleza de la variable: Análisis de la forma de la distribución:

Para entender la naturaleza de cualquier variable, el punto de inicio es caracterizar la forma de su distribución. Aunque algunas medidas descriptivas que se analizarán paralelamente para analizar la normalidad, el investigador puede obtener una perspectiva adecuada de la variable a través de un **histograma** que representa gráficamente la frecuencia de los datos en cada categoría. Si el examen de la distribución tiene como objetivo evaluar su normalidad, se puede superponer la curva normal sobre la distribución.

Utilice un histograma para evaluar la forma y la tendencia central de sus datos y para evaluar si sus datos siguen o no una distribución específica, tal como la distribución normal .

Las barras representan el número de observaciones que se ubican dentro de intervalos consecutivos o secciones. Debido a que cada barra representa muchas observaciones, un histograma es más útil cuando se tiene una gran cantidad de datos.

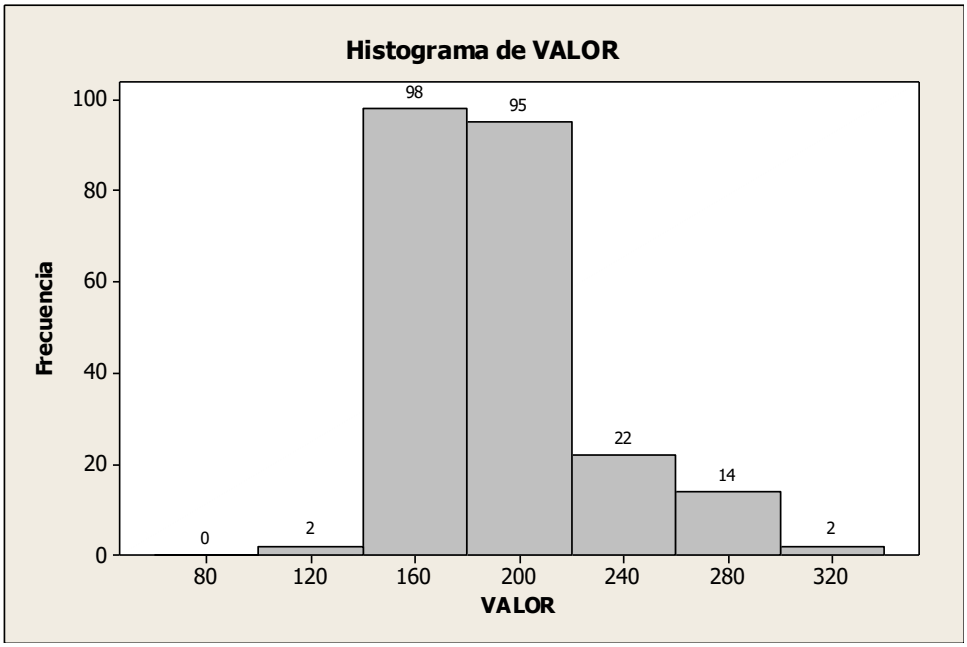
Utilice una gráfica de puntos para evaluar la forma y la tendencia central de sus datos. Igual que un histograma, una gráfica de puntos se divide en secciones. Sin embargo, una gráfica de puntos puede ser más útil que un histograma cuando se tiene una cantidad pequeña de datos porque:

· Por opción predeterminada, una gráfica de puntos tiene muchas más secciones que un

histograma.

· Cada punto representa una observación (o un número pequeño de observaciones).

Las gráficas de puntos también son útiles para comparar grupos de datos.



Utilice una gráfica de tallo y hoja para mostrar los valores de datos reales en un formato de secciones. Aunque es similar a una gráfica de puntos, una gráfica de tallo y hoja:

· Se proyecta hacia un lado.

· Utiliza las cifras iniciales de los valores de muestra para determinar las secciones (por ejemplo, una sección puede tener valores entre 0 y 9, otra sección pudiera tener valores entre 10 y 19, y así sucesivamente).

· Muestra las cifras de los valores individuales en lugar de puntos.

Una variante del histograma es el **diagrama de tallo y hojas** que presenta el mismo cuadro de gráfico, pero que también proporciona una enumeración de los valores de los datos reales. El tallo es la raíz al cual se añaden las hojas.

Diseño de tallo y hoja: VALOR

Tallo y hoja de VALOR N = 233

Unidad de hoja = 10

2 1 01

2 1

25 1 4445555555555555555555

100 1
66666666666666666666666666666666777777777777777777777777777+

(73) 1
88888888888888888888888888888888888888888888999999999999999999+

60 2 00000001111111111111

38 2 222222222223333333

20 2 4444

16 2 666677

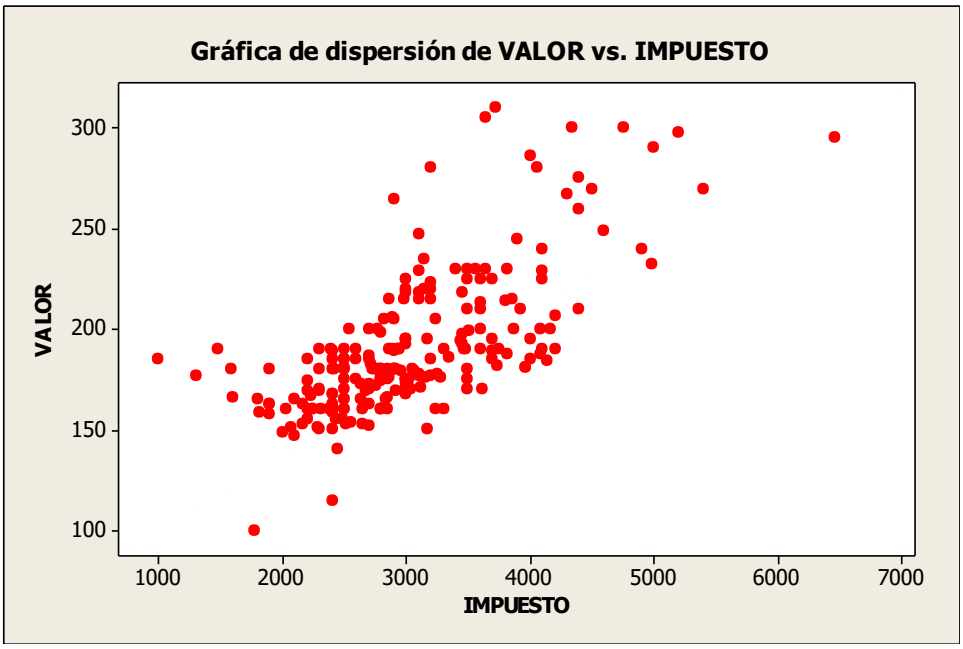
10 2 88899999

2 3 01

Análisis de la relación entre variables:

Utilice una gráfica de dispersión para evaluar la relación entre dos variables. Los valores de las dos variables sirven como las coordenadas de x y las coordenadas de y para graficar cada observación.

Una vez realizado un meticuloso examen de la distribución de una variable, el investigador está interesado en examinar las relaciones entre dos o más variables. El método más común de análisis bivariante es el **gráfico de dispersión,** un gráfico de puntos de datos basados en dos variables. Se representa una variable en el eje horizontal y la otra en vertical. Las variables pueden ser valores observados, valores esperados o incluso residuos. Los puntos del gráfico representan los correspondientes valores conjuntos de las variables para cualquier caso dado.



Cuando los puntos se organizan a lo largo de una línea recta, tendremos una relación lineal de correlación. Un conjunto de puntos curveados puede indicar una relación no lineal, que se puede tratar de varias formas. También puede suceder que no existan patrones tratándose de puntos aparentemente aleatorios, en este caso no existe relación.

Existen diversos tipos de gráficos de dispersión, sin enbargo un formato que se adapta particularmente a las técnicas multivariadas, es la matriz del grafico de dispersión que muestra todas las combinaciones de variables en la posición inferior de la matriz y la diagonal muestra los histogramas de las variables. En la parte superior de la matriz se incluyen las correlaciones correspondientes para que el investigador pueda valorar la correlación representada en cada gráfico.

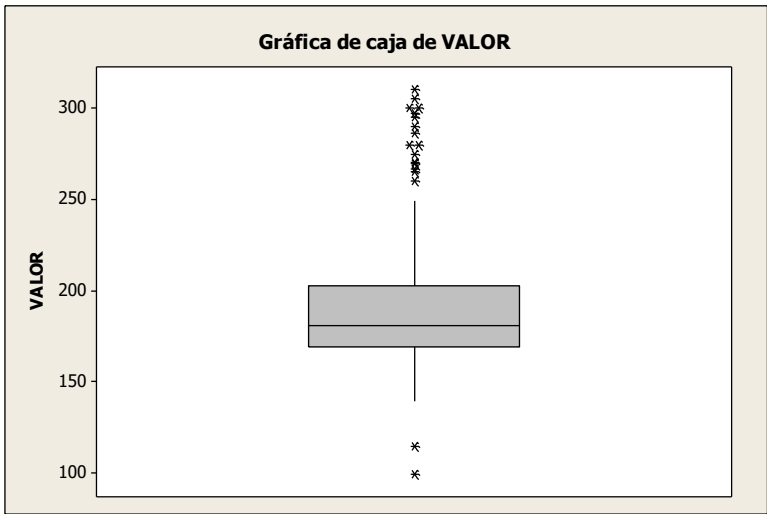
Análisis de las diferencias entre grupos:

El investigador se enfrenta también con la tarea de entender el carácter y la diferencia entre dos o más grupos de una variable para dos o más variables métricas, tal como ocurre en el análisis discriminante, análisis de varianza y MANOVA. En estos casos se necesita entender cómo se distribuyen los valores para cada grupo para tener significancia estadística, además de otro aspecto importante como la identificar los **casos atípicos**, que pueden resultar ser aparentes sólo cuando los valores de los datos se separan en grupos.

El método que se utiliza para esta tarea es el grafico de **caja y brazos múltiple.** Los límites superior e inferior de la caja marcan los cuartiles superior e inferior de la distribución de los datos, es decir la longitud de la caja es la distancia entre el primer y el tercer cuartil conocida como el Rango intercuartílico (RIQ) y contiene el 50% de los datos centrales de la distribución. La línea dentro de la caja representa la posición de la mediana y conforme a si ésta cae cerca del final de la caja, se indica la presencia de la asimetría o de sesgo hacia alguno de los lados. Las líneas que se extienden desde la caja (llamadas bigotes o brazos) representan la distancia entre la mayor y la menor de las observaciones que están a menos de un cuartil de la caja. Los casos atípicos son observaciones que se sitúan entre 1.5 y 3.0 RIQ fuera de la caja.

Utilice una gráfica de caja para evaluar y comparar las características de distribución, tales como mediana, rango y simetría, así como para identificar valores atípicos.

Las opciones de presentación incluyen símbolos para la media y cajas para los intervalos de confianza de la mediana.



Un valor ausente puede tener distintos significados. Es posible que el campo no fuera aplicable, que el evento no se produjera o que los datos no estuvieran disponibles. También puede deberse a que la persona que escribió los datos no conocía el valor correcto, o no se preocupó de rellenar un campo.

Sin embargo, hay muchos escenarios de análisis de datos en los que los valores ausentes proporcionan información importante. El significado de los valores ausentes depende en gran parte del contexto. Por ejemplo, un valor de fecha ausente en una lista de facturas tiene un significado sustancialmente diferente de la ausencia de una fecha en la columna que indica la fecha de contratación de un empleado.

2. DATOS AUSENTES:

Introducción:

Los **datos ausentes** son algo generalizado en el análisis multivariado. Por esta razón, el desafío consiste en enfrentarse a los resultados producidos por los datos ausentes en los procesos de estimación y que afectan a la generalidad de los resultados. La ocupación primaria del investigador es determinar las razones que subyacen en el dato ausente, debiendo entender el proceso principal de esta ausencia de datos para seleccionar el curso de acción apropiado.

Un proceso de datos ausentes es cualquier evento sistemático externo al encuestador (errores en la captura de los datos o problemas de su recolección) o acción por parte del encuestado (tales como rehusar a contestar) que da lugar a la ausencia de datos.

El impacto de los datos ausentes es perjudicial, no sólo por sus potenciales sesgos "escondidos", sino también por su efecto en el tamaño de la muestra disponible en el análisis. En muchos análisis multivariantes, la ausencia de datos elimina tantas observaciones, que lo que en principio era una muestra adecuada, se reduce a una inadecuada. En estas circunstancias, el investigador debe buscar observaciones adicionales o encontrar una solución para la ausencia de datos en la muestra original.

Razones causantes de datos ausentes:

La ausencia de datos puede ocurrir por muchas razones y en muchas situaciones. Uno se debe a factores de procedimiento tales como errores en la entrada de datos que crean códigos inválidos, restricciones de representatividad o fallos al completar el cuestionario. En estas situaciones, el investigador tiene poco o nulo control sobre los procesos de ausencia de datos. Otro tipo de procesos de pérdida de datos tiene lugar cuando la respuesta es inaplicable, como las preguntas en relación a los años de matrimonios para adultos que nunca han estado casados.

Otros tipos de procesos de datos ausentes tiene que ver con el encuestado, que no desea contestar tal pregunta. El investigador debe preveer estos problemas e intentar minimizarlos en el diseño de la investigación y en los niveles de recolección de los datos.

Tratamiento de datos ausentes:

Las aproximaciones o soluciones que tratan los datos ausentes pueden clasificarse en tres categorías basadas en la aleatoriedad de los procesos de datos ausentes,

en función del método empleados para eliminarlos:

a. Utilizar sólo aquellas observaciones con datos completos.

La aproximación más sencilla y directa para tratar con datos ausentes es incluir sólo aquellas observaciones con datos completos. Este método está incluido en casi todo el software estadístico y es el método por defecto en muchos programas. No obstante, este método sólo es recomendable cuando los valores ausentes se presentan en forma aleatoria y no forman un patrón, ya que podrían sesgarse los resultados. Incluso aunque fueran aleatorios los datos ausentes, podía la muestra resultante quedar reducida a una muestra inadecuada por su tamaño para los propósitos del análisis.

b. Supresión de caso(s) y/o variable(s).

Otra solución simple para los datos ausentes es suprimir el caso(s) y/o variable(s) que peor se comporta(n) respecto a los datos ausentes. El investigador determina la extensión de los datos ausentes sobre cada caso y variable, y entonces suprime los casos y variables que excedan el nivel especificado. Cuando se presenta un patrón de datos ausentes puede ser una solución viable este método. Si una variable, que no sea la variable dependiente tiene valores ausentes y es una candidata de eliminación, el investigador debe asegurarse de que existan variables alternativas, que se espera que estén altamente correlacionadas, para representar a la variable original.

c. Métodos de imputación.

Una tercera forma para tratar con datos ausentes, es usar alguno de los **muchos métodos de imputación.** La imputación es el proceso de estimación de valores ausentes basado en valores de otras variables y/o casos de la muestra. La finalidad es emplear relaciones conocidas que puedan identificarse en los valores válidos de la muestra que ayuden a estimar valores ausentes. El investigador debe considerar el uso de la imputación en cada instancia, dados sus potenciales impactos sobre el análisis.

En los métodos de imputación suelen utilizarse básicamente variables métricas, ya que se pueden hacer estimaciones de los datos ausentes. Mientras que en las variables no métricas requieren una estimación de un valor específico en lugar de una estimación en una escala continua, como por ejemplo la estimación del género del encuestado, cuando este dato este ausente.

La sustitución de datos ausentes consiste en sustituir datos ausentes por valores estimados sobre la base de otra información existente en la muestra. Esto se puede llevar a cabo de distintas maneras:

- *Sustitución de caso:* en este método, las observaciones con datos ausentes se sustituyen con otras observaciones no muestrales y es el que más se

utiliza cuando existen datos ausentes en un caso particular.

- *Sustitución por la media:* es tal vez el método más empleado y consiste en sustituir los valores ausentes por una variable cuyo valor medio se calcula sobre todas las respuestas válidas. La lógica de esta aproximación, es que la media es el mejor valor de sustitución aunque tiene tres desventajas: la primera es que invalida las estimaciones de la varianza derivada de las fórmulas estándar de la varianza para conocer la verdadera varianza de los datos. La segunda, la distribución real de los valores se encuentra distorsionada por la sustitución de los datos ausentes por la media. La tercera es que el método modifica la correlación observada porque todos los datos ausentes tendrán un valor único constante. Por otro lado tiene la ventaja de ser fácil de realizar y proporcionar una información completa para todos los casos.

- *Sustitución por valor constante:* el investigador utiliza este método para sustituir los valores ausentes por un valor constante derivado de fuentes externas o de una investigación previa por considerar que el valor de sustitución de una fuente externa es más valido que el valor generado internamente por la media.

- *Imputación por regresión:* el análisis de regresión se usa en este método para predecir valores ausentes de una variable basándose en su relación con las otras variables del conjunto de datos. Este método tiene el atractivo de usar las relaciones preexistentes en la muestra como base de predicción, sin embargo, puede que refuerce las relaciones ya existentes en la muestra y hacer que el análisis sea menos generalizable. Otra desventaja es que se subestima la varianza de la distribución, además de suponer a priori que la variable con datos ausentes tiene correlaciones sustanciales con otras variables. Si estas correlaciones no son buenas para una estimación significativa es mejor utilizar otros métodos, como la sustitución por la media.

3. CASOS O VALORES ATÍPICOS:

Si una muestra contiene más de un supuesto valor atípico, entonces la prueba de Grubbs y la relación Q de Dixon pudieran no ser efectivas. Las otras pruebas de Dixon se diseñaron para superar el efecto de ocultación que pueden ocasionar múltiples valores atípicos potenciales.

No debe utilizar las pruebas de valor atípico de Minitab más de una vez en la misma muestra. Si elimina un valor atípico de su muestra y

Los casos atípicos son observaciones que presentan una combinación de características que pueden ser identificadas que hace que diferencien claramente de las otras observaciones. Pueden ser benéficos o problemáticos y deben ser analizados dentro del contexto del análisis y, su evaluación está en función de la información que pueden proporcionar. Cuando son benéficos pueden ser indicativos de ciertas características de un segmento de la población que se llegaría a descubrir en el curso normal del análisis. Si son problemáticos no son representativos de la población y están en contra de los objetivos del análisis y pueden distorsionar seriamente las pruebas estadísticas. Se sugiere realizar un análisis meticuloso de influencia para averiguar el tipo de influencia que ejercen.

Los casos atípicos se pueden clasificar en cuatro categorías: la primera los que

luego vuelve probar, arriesga eliminar valores que no son en realidad atípicos.

surgen de un error de procedimiento, tales como la captura de datos o un error en la codificación que puede, en algunos casos, evitarse con un adecuado filtro de datos. La segunda es la observación que ocurre como consecuencia de un acontecimiento extraordinario y existe una aplicación por la unidad de la observación y el investigador debe decidir si es representativo de la muestra y debe ser retenido en el análisis o si hay que suprimirlo. La tercera categoría comprende las observaciones extraordinarias para las que el investigador no tiene explicación y, a menos que el investigador crea que representan un segmento válido de la población, deben ser eliminados. La cuarta y última contiene las observaciones que se sitúan fuera del rango ordinario de valores de cada variable, pero que son únicos en su combinación de valores entre las variables. En este caso el atípico no es un miembro válido de la población.

Valor atípico - una observación inusualmente grande o pequeña. Los valores ubicados más allá de los bigotes o brazos son valores atípicos.

Detección de casos atípicos:

Pueden identificarse los casos atípicos desde un enfoque univariado, divariado o multivariado, y el investigador debe utilizar cualquiera de ellos buscando una consistencia entre los métodos de identificación de casos atípicos.

Por opción predeterminada, el bigote o brazo superior en Minitab se extiende hasta el valor de datos más alto dentro del límite superior.

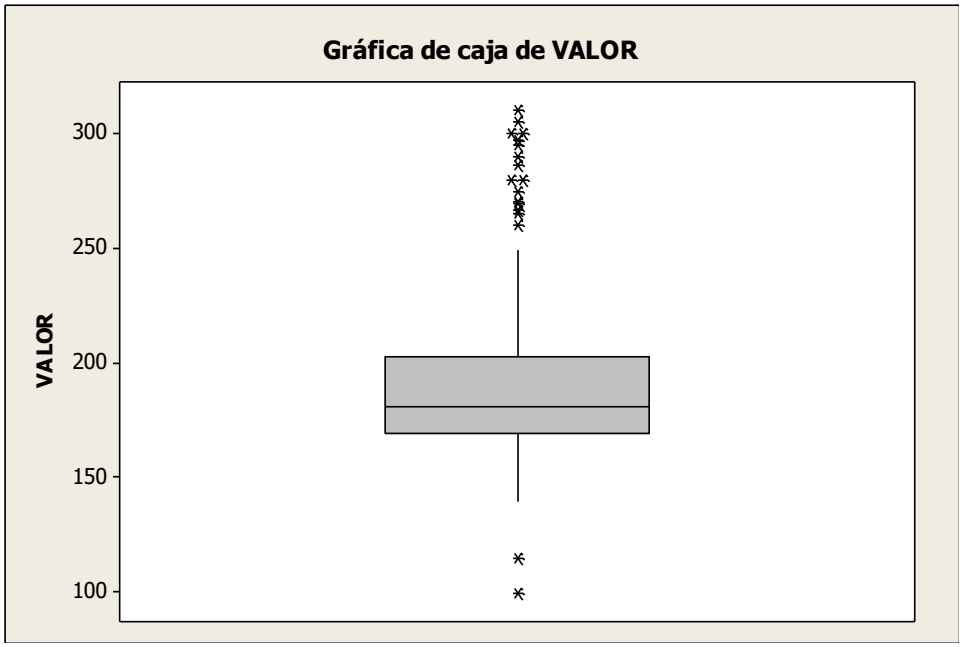
Límite superior = Q3 + 1.5 (Q3 - Q1)

Por opción predeterminada, el bigote o brazo inferior en Minitab se extiende hasta el valor más bajo dentro del límite inferior.

Límite inferior = Q1- 1.5 (Q3 - Q1)

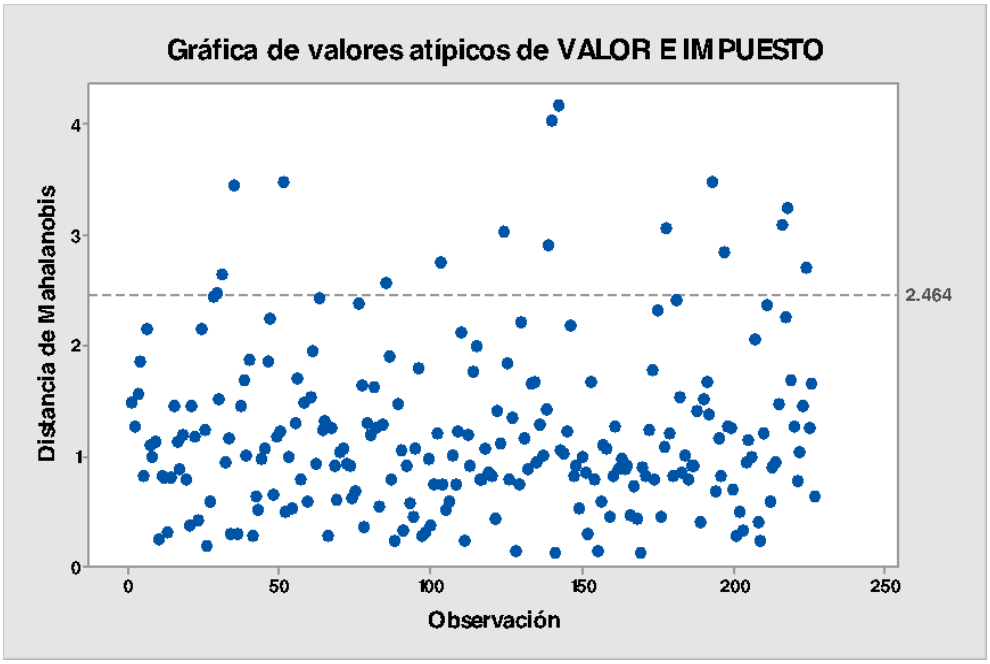
- *Detección univariante:* examina la distribución de las observaciones, seleccionando como casos atípicos aquellos casos que caigan fuera de los rangos de la distribución, estableciendo un umbral para la designación como caso atípico. Un enfoque típico convierte en primer lugar los valores de los datos en valores estándar que tienen media cero con varianza uno. En muestras pequeñas (de 80 o incluso menos), las pautas sugeridas identifican como atípicos aquellos casos con valores estándar de 2.5 o superiores, y para muestras más grandes las pautas sugieren que el valor del umbral estandarizado se sitúe entre 3 y 4.

A menudo es más fácil identificar gráficamente los valores atípicos, identificando dichos valores en gráficas de cajas, al etiquetar las observaciones que son por menos 1.5 veces el rango intercuartílico (RIQ) (Q3-Q1) desde el borde de la caja.



- *Detección bivariante:* Pueden evaluarse conjuntamente pares de variables mediante un gráfico de dispersión. Casos que caigan fuera del rango del resto de las observaciones pueden identificarse como puntos atípicos en el gráfico de dispersión. Para ayudar a determinar el rango esperado de las observaciones, el investigador necesita una forma de medición objetiva de la posición multidimensional de cada observación relativa a un punto común. La medida D^2 de Mahalanobis puede usarse para este fin. Es una medida de la distancia de cada observación en un espacio multidimensional respecto al centro de las observaciones. Dada la naturaleza de las pruebas estadísticas, se sugiere se use un nivel muy conservador, quizás 0.01, como un umbral para la designación como caso atípico.

La Distancia de Mahalanobis es la distancia entre un punto de datos y el centroide de un espacio multivariado (media general). Utilice la distancia de Mahalanobis en Análisis de componentes principales para identificar valores atípicos. Es un método multivariado más poderoso para detectar valores atípicos que el procedimiento de examinar una variable a la vez, porque toma en cuenta las diferentes escalas entre variables y las correlaciones entre éstas.



Es importante verificar la presencia de **valores atípicos** ya que éstos pueden influir significativamente en sus resultados. La gráfica de valores atípicos muestra las distancias de Mahalanobis para cada una de las observaciones en su hoja de trabajo de Minitab. Los puntos que se encuentran por encima de la línea de referencia del eje Y son valores atípicos.

La distancia de Cook, o D, es una medida general del impacto combinado de cada observación sobre los valores ajustados. Dado que D se calcula utilizando valores con apalancamiento y residuos estandarizados, considera la posibilidad de que una observación sea poco común con respecto tanto a los valores x como a los valores y. Geométricamente, la la distancia de Cook es una medida de la distancia entre los valores ajustados calculados con y sin la observación i^{ésima}. Los valores grandes, los cuales son indicativos de observaciones poco comunes, pueden producirse

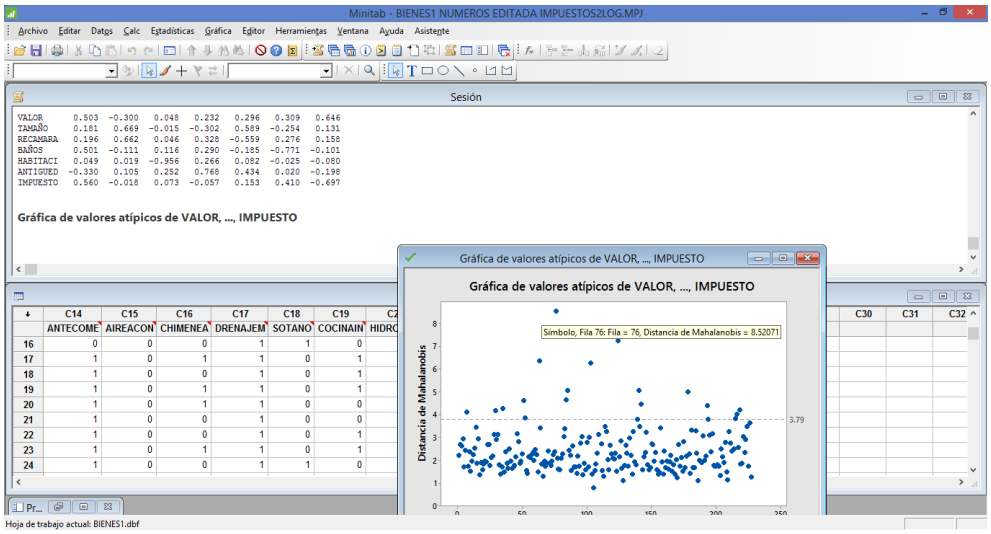
El análisis de influencia es otra variante. Utilice medidas de diagnóstico, tales como distancia de Cook o DFITS para determinar si el valor atípico es una observación influyente. Para determinar el efecto del valor atípico sobre sus resultados, realice el análisis con y sin la observación para ver cómo cambia el modelo. Note que una observación pudiera ser un valor atípico en un modelo lineal, pero se explica mejor mediante un modelo no lineal.

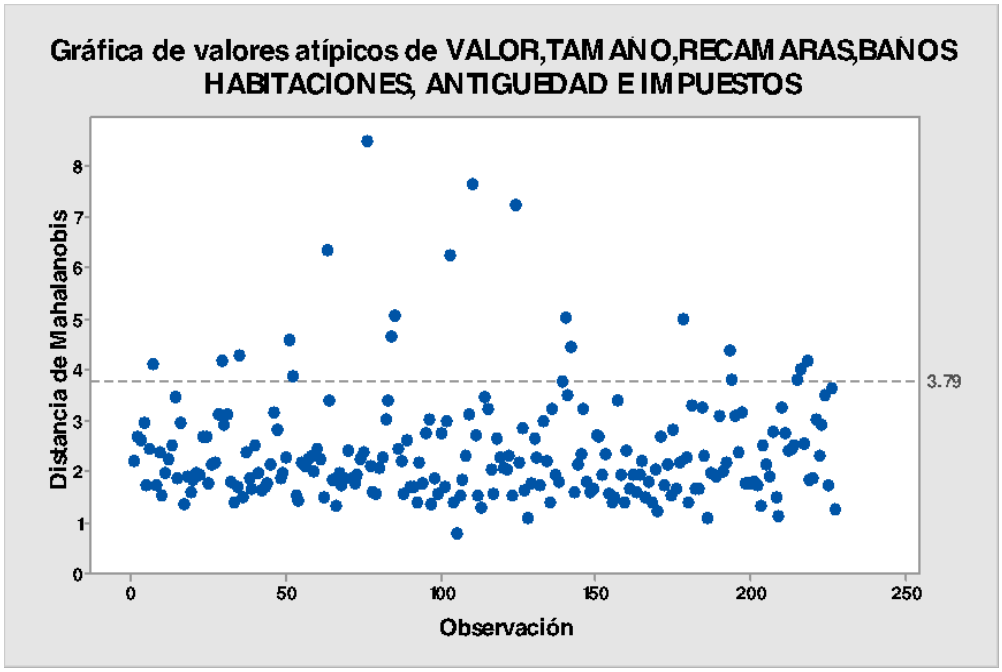
- *Detección multivariada:* además de la evaluación univariada y divariada, es conveniente realizar una evaluación multivariada de cada observación para identificar casos atípicos a lo largo de un conjunto de variables. Dado que la mayoría de los análisis multivariados tienen más de dos variables, el investigador necesita una forma de medición objetiva de la posición multidimensional de cada observación relativa a un punto común. La medida D^2 de Mahalanobis puede usarse para este fin. Es una medida de la distancia de cada observación en un espacio multidimensional respecto al centro de las observaciones. Dada la naturaleza de las pruebas estadísticas, se sugiere se use un nivel muy conservador, quizás 0.01, como un umbral para la designación como caso atípico.

La distancia de Mahalanobis es la distancia entre un punto de datos y el centroide de un espacio multivariado (media general). Utilice la distancia de Mahalanobis en

porque la observación posee 1) un residuo grande y un apalancamiento moderado, 2) un apalancamiento grande y un residuo moderado o 3) un residuo y un apalancamiento grandes. Algunos estadísticos recomiendan comparar D con la distribución F (p, n-p). Si D es mayor que el valor F en el percentil 50, entonces D es considerado extremo y se debe examinar. Otros estadísticos recomiendan comparar las estadísticas D entre sí, identificando valores que sean extremadamente grandes en relación con los demás valores. Una sencilla manera de comparar los valores D consiste en graficarlos utilizando graficas de series de tiempo, donde el eje x representa las observaciones, no un índice ni un período de tiempo.

Análisis de componentes principales para identificar valores atípicos. Es un método multivariado más poderoso para detectar valores atípicos que el procedimiento de examinar una variable a la vez, porque toma en cuenta las diferentes escalas entre variables y las correlaciones entre éstas.





Es importante verificar la presencia de **valores atípicos** ya que éstos pueden influir significativamente en sus resultados. La gráfica de valores atípicos muestra las distancias de Mahalanobis para cada una de las observaciones en su hoja de trabajo de Minitab. Los puntos que se encuentran por encima de la línea de referencia del eje Y son valores atípicos.

1.3.1.1 EJEMPLO ILUSTRATIVO. Datos atípicos. Diagrama de Caja y Brazos Simple.

EJEMPLO ILUSTRATIVO 1.3.1.1 DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE

Utilice una gráfica de puntos para evaluar la forma y la tendencia central de sus datos. Igual que un histograma, una gráfica de puntos se divide en secciones. Sin embargo, una gráfica de puntos puede ser más útil que un histograma cuando se tiene una cantidad pequeña de datos porque:

- Por opción predeterminada, una gráfica de puntos tiene muchas más secciones que un histograma.
- Cada punto representa una observación (o un número pequeño de observaciones).

Una agencia de Bienes Raíces desea examinar el **VALOR** general de las casas que forman su cartera de ventas. Las casas se encuentran ubicadas en diferentes orientaciones de la ciudad. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

215.0	210.0	174.5	177.5	297.5	180.0
181.0	205.0	260.0	173.0	290.0	160.0
189.0	154.0	175.0	174.9	219.9	176.5
176.0	215.0	286.5	185.0	189.9	189.9
267.0	280.0	169.5	164.9	159.9	185.0

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas sin recortar los datos o valores atípicos generado por Minitab:

Tallo y hoja de VALOR N = 30
Unidad de hoja = 1.0

```
 2   15  49
 5   16  049
 12  17  3445667
(7)  18  0155999
 11  19
 11  20  5
 10  21  0559
 6   22
 6   23
 6   24
 6   25
 6   26  07
 4   27
 4   28  06
```

Las gráficas de puntos también son útiles para comparar grupos de datos.

Utilice una gráfica de tallo y hoja para mostrar los valores de datos reales en un formato de secciones. Aunque es similar a una gráfica de puntos, una gráfica de tallo y hoja:

· Se proyecta hacia un lado.

· Utiliza las cifras iniciales de los valores de muestra para determinar las secciones (por ejemplo, una sección puede tener valores entre 0 y 9, otra sección pudiera tener valores entre 10 y 19, y así sucesivamente).

· Muestra las cifras de los valores individuales en lugar de puntos.

2 29 07

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas recortando los datos o valores atípicos generado por Minitab:

Tallo y hoja de VALOR N = 30
Unidad de hoja = 1.0

```
 1   15  4
 2   15  9
 4   16  04
 5   16  9
 8   17  344
 12  17  5667
 14  18  01
(5)  18  55999
 11  19
 11  19
 11  20
 11  20  5
 10  21  0
 9   21  559
 6   22
 6   22
 6   23
 6   23
 6   24
 6   24
 6   25
 6   25
 6   26  0
 5   26  7
```

HI 280, 286, 290, 297

Otra forma de ordenarlos es colocarlos en forma ascendente:

154.0	173.0	176.5	185.0	210.0	267.0
159.9	174.5	177.5	189.0	215.0	280.0
160.0	174.9	180.0	189.9	215.0	286.5
164.9	175.0	181.0	189.9	219.9	290.0
169.5	176.0	185.0	205.0	260.0	297.5

Resumen de 5 Números: Resultados aproximados a las décimas.

1. Valor Mínimo: **154.0**

2. Cuartil 1 Q_1=Dato 7 + 0.75 (Dato 8-Dato 7)= 174.5 + 0.75 (174.9-174.5)=**174.8**

   P(Q_1)= 1 $\frac{30+1}{4}$= 7.75

3. Mediana $\tilde{X}$= $\frac{Dato\ 16-Dato15}{2}$=$\frac{185.0+185.0}{2}$=**185.5**

   P($\tilde{X}$)=$\frac{n+1}{2}$=$\frac{30+1}{2}$=15.5

4. Cuartil 3 Q_3=Dato 23+ 0.25(Dato 24- Dato 23)=215.0+0.25(219.9-215.0)=**216.2**

   P(Q_3)= 3 $\frac{30+1}{4}$= 23.25

5. Valor Máximo= **297.5**

Esquema de vallas: Resultados aproximados a las décimas

VEI (Valla Externa Inferior)= Q_1-3(RIQ)=174.8-3(216.2-174.8)= 174.8-3(41.4)=**50.6**

VII (Valla Interna Inferior)= Q_1-1.5(RIQ)=174.8-1.5(216.2-174.8)= 174.8-1.5(41.4)=**112.7**

VIS (Valla Interna Superior)= Q_1+1.5(RIQ)=216.2+1.5(216.2-174.8)= 216.2+1.5(41.4)=**278.3**

VES (Valla Externa Superior)= Q_1-3(RIQ)=216.2+3(216.2-174.8)= 216.2+3(41.4)=**340.4**

Por opción predeterminada, en Minitab, la Valla Externa Inferior se extiende hasta el valor más bajo dentro del límite inferior.

Límite inferior = Q1- 3.0 (Q3 - Q1)

Por opción predeterminada, en Minitab, la Valla Interna Inferior se extiende hasta el valor más bajo dentro del límite inferior.

Límite inferior = Q1- 1.5 (Q3 - Q1)

Por opción predeterminada, en Minitab, la parte inferior de la caja es el primer cuartil (Q1) - 25% de los valores de datos son menores que o iguales a este valor.

Por opción predeterminada,en Minitab, la parte superior de la caja es el tercer cuartil (Q3) - 75% de los valores de datos son menores que o iguales a este valor.

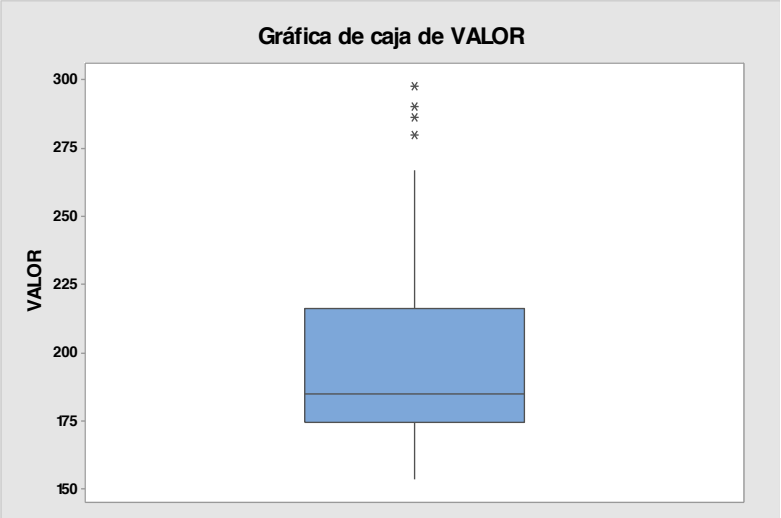
Por opción predeterminada, en Minitab, La Valla Interna Superior se extiende hasta el valor de datos más alto dentro del límite superior.

Límite superior = Q3 + 1.5 (Q3 - Q1)

Por opción predeterminada, en Minitab, La Valla Externa Superior se extiende hasta el valor de datos más alto dentro del límite superior.

Límite superior = Q3 + 3.0 (Q3 - Q1)

Diagrama de Caja y Brazos



Interpretación: En el diagrama anterior se puede observar que existen 4 valores atípicos: **280, 286.5, 290.0 y 297.5** entre las **VIS** y la **VES (**a los que les denominamos comúnmente valores aberrantes) que deben ser estudiados con detalle para ver el tratamiento que se le van a dar en estudios subsecuentes ya que es posible que provoquen un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la izquierda de la caja, sugiere un sesgo hacia la derecha producto tal vez de los valores o casos atípicos).

1.3.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE APRENDIZAJE 1.3.1.1 DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **NORTE** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante ordenando los datos en forma ascendente ya sea en forma manual o mediante un diagrama de tallo y hojas, presente el resumen de cinco números, construya el esquema de vallas correspondiente y el diagrama de caja y brazos con su debida interpretación.

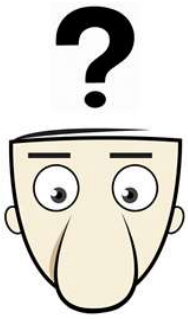
Los datos se muestran a continuación:

215.0	267.0	247.0	235.0	170.0	155.0
195.0	199.9	218.0	190.0	215.0	179.0
160.0	169.0	299.9	180.0	229.5	195.0
189.0	179.0	172.0	213.0	180.0	179.9
249.0	218.0	230.0	175.0	225.0	225.0

1.3.1.1 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material estudiado anterioriormente. La respuesta a este ejercicio de Autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 1.3.1.1
DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.



Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante ordenando los datos en forma ascendente ya sea en forma manual o mediante un diagrama de tallo y hojas, presente el resumen de cinco números, construya el esquema de vallas correspondiente y el diagrama de caja y brazos con su debida interpretación.

Los datos se muestran a continuación:

181.0	181.0	179.9	200.0	230.0	174.9
220.0	185.0	210.0	115.0	183.0	100.0
190.0	269.9	159.9	199.9	215.0	174.9
160.0	189.9	165.0	189.9	159.0	189.9
176.0	299.9	185.0	189.9	180.0	210.0

1.3.1.1 EJERCICIOS DE **REFUERZO**

EJERCICIOS DE REFUERZO 1.3.1.1 DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

1.3.1.1.1 Una agencia de Bienes Raíces desea examinar el **IMPUESTO** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **IMPUESTO** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante ordenando los datos en forma ascendente ya sea en forma manual o mediante un diagrama de tallo y hojas, presente el resumen de cinco números, construya el esquema de vallas correspondiente y el diagrama de caja y brazos con su debida interpretación.

Los datos se muestran a continuación:

3959	3959	2911	3871	3638	2500
3142	3700	3930	2400	2708	1775
2600	4500	2378	2540	2989	3000
2250	2860	1800	4200	1814	2386
2857	4343	2700	3600	2800	3500

1.3.1.1.2 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **NORTE** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas el **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante ordenando los datos en forma ascendente ya sea en forma manual o mediante un diagrama de tallo y hojas, presente el resumen de cinco números, construya el esquema de vallas correspondiente y el diagrama de caja y brazos con su debida interpretación.

Los datos se muestran a continuación:

30	33	30	34	41	35
35	38	30	44	35	37
35	37	15	40	28	34
24	35	35	35	14	25
22	31	31	38	30	33

1.3.1.1.3 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante ordenando los datos en forma ascendente ya sea en forma manual o mediante un diagrama de tallo y hojas, presente el resumen de cinco números, construya el esquema de vallas correspondiente y el diagrama de caja y brazos con su debida interpretación.

Los datos se muestran a continuación:

32	28	35	30	61	34
60	27	30	16	39	21
27	28	39	28	31	34
35	35	30	31	55	38
30	14	28	31	32	24

1.3.1.2.D.A. EJEMPLO ILUSTRATIVO EN MINITAB 17. Datos Atípicos

EJEMPLO ILUSTRATIVO 1.3.1.2. INTEGRAL EN MINITAB 17. DATOS ATÍPICOS

Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

1 Abra la base de datos que creó **"BIENES1".** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***VALOR***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

Utilice esta opción de valores atípicos para identificar un valor atípico individual en una muestra. Las hipótesis son las siguientes:

- H_0 (la hipótesis nula): Todos los valores de la muestra corresponden a la misma población normalmente distribuida.
- H_1 (la hipótesis alternativa): Uno de los valores de la muestra no corresponde a la misma población normalmente distribuida.

Las opciones de las relaciones de Dixon le ayudan a detectar un valor atípico en presencia de otros supuestos valores atípicos. Para una explicación más detallada, véase Elección de la prueba adecuada de valores atípicos.

No debe utilizar las pruebas de valor atípico de Minitab más de una vez en la misma muestra. Si elimina un valor atípico de su muestra y luego vuelve probar, arriesga eliminar valores que no son en realidad atípicos.

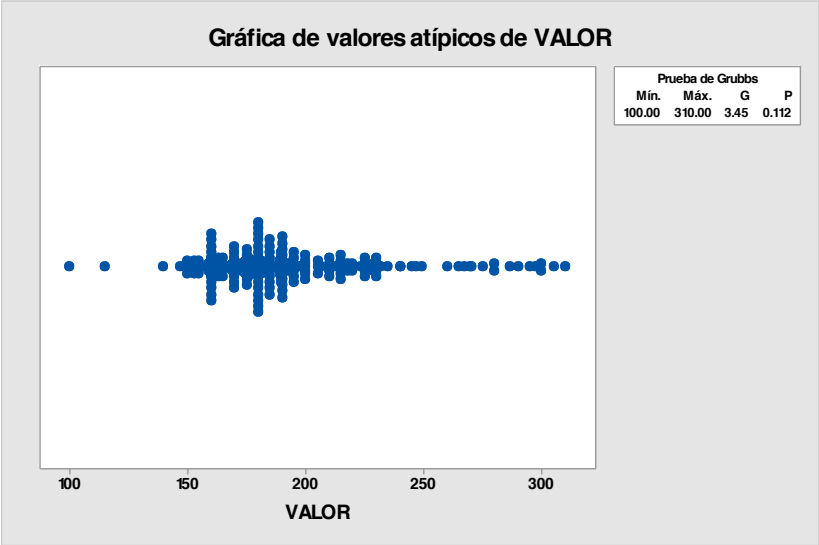
Salida de la ventana Sesión

Prueba de valores atípicos: VALOR

```
Método

Hipótesis nula          Todos los valores de los datos provienen de
la misma población normal
Hipótesis alterna       El valor más pequeño o más grande de los
datos es un valor atípico
Nivel de significancia  α = 0.05


Prueba de Grubbs

Variable    N   Media  Desv.Est.    Mín.    Máx.     G      P
VALOR     233  190.85      34.55  100.00  310.00  3.45  0.112

* NOTA * No hay valor atípico en el nivel de significancia de 5%
```

Gráfica de valores atípicos de VALOR

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **VALOR** muestran el valor más pequeño es de 100 y el más grande de 310. La estadística de G de 3.45 indica que el valor de datos más pequeño ó más grande de 100 ó 310, es 3.45

desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.112. Este valor p es mayor que el nivel de significancia que se eligió, de modo que no se puede rechazar la hipótesis nula y concluir que el valor más pequeño ó el más grande no es un valor atípico.

1.3.1.2 ACTIVIDAD DE **APRENDIZAJE EN MINITAB 17. Datos Atipicos.**

ACTIVIDAD DE APRENDIZAJE EN MINITAB 17
1.3.1.2
DATOS ATÍPICOS

Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **TAMAÑO DEL LOTE** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.**

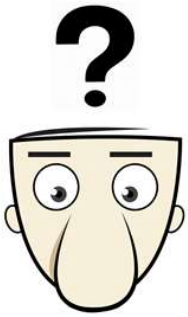
1.3.1.2 EJERCICIO DE **AUTOEVALUACIÓN. Datos Atipicos.**

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN
1.3.1.2

Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **ANTIGÜEDAD DE LAS CASAS** parecen inusualmente

DATOS ATÍPICOS



pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver esta Autoevaluación.**

1.3.1.2 EJERCICIOS DE REFUERZO. Datos Atípicos

EJERCICIOS DE REFUERZO 1.3.1.2 DATOS ATÍPICOS

1.3.1.2.1. Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **NÚMERO DE RECÁMARAS** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.**

1.3.1.2.2. Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **NÚMERO DE BAÑOS** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.**

1.3.1.2.3. Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **IMPUESTOS ANUALES** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.**

1.3.1.3.D.A. EJEMPLO ILUSTRATIVO EN MINITAB 17.Datos Atipicos. Un nivel de Factor

EJEMPLO **ILUSTRATIVO I**NTEGRAL EN **MINITAB 17. D**ATOS ATÍPICOS PARA UN NIVEL DE UN **FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO.**

Utilice esta opción de crear un subconjunto de hoja de trabajo para copiar las filas especificadas de la hoja de trabajo activa en una nueva hoja de trabajo. Usted puede especificar el subconjunto basándose en números de fila, puntos destacados en una gráfica, o una condición, como hombres no casados menores de 50 años.

Dividir hoja de trabajo divide, o desapila, la hoja de trabajo activa en dos o más hojas de trabajo nuevas basándose en una o más "Por" variables. Crear subconjunto de hoja de trabajo y Dividir hoja de

Se reciben datos sobre una encuesta realizada a las casas ubicadas en la zona **ESTE** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en esta zona **ESTE** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. Este ejemplo se debe resolver en dos pasos:

1. Crear un subconjunto de hoja de trabajo
2. Realizar la Prueba de valores atípicos de Minitab

Crear un subconjunto de hoja de trabajo.

Nota importante:

Al **crear un subconjunto de hoja de trabajo** se divide, o desapila, la hoja de trabajo activa en hojas de trabajo nuevas basándose en una o más "Por" variables. Crear subconjunto de hoja de trabajo siempre copian los datos en hojas de trabajo **nuevas**. Como la base de datos **"BIENES 1"** contiene todas las casas ubicadas en la zona **NORTE, SUR y ESTE** debemos crear **un subconjunto de hoja de trabajo** que solo incluya las casas ubicadas en el **ESTE** Esto nos permite analizar los diferentes niveles de un factor por separado en caso de que la prueba de **ANOVA** haya sido significativa y exista la posibilidad de que las medias de la variable **VALOR** no sea la misma en los tres niveles del Factor **UBICACIÓN**

1 Abra la base de datos que creó **"BIENES1"**

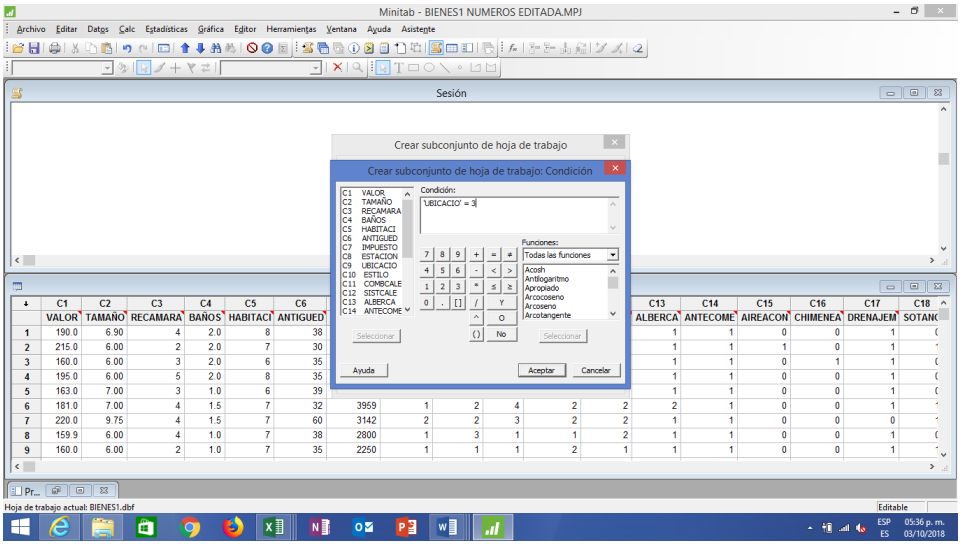
2 Elija **Datos > Crear subconjunto de hoja de trabajo**.

3 En **Incluir o excluir**, elija **Especificar las filas que se incluirán.**

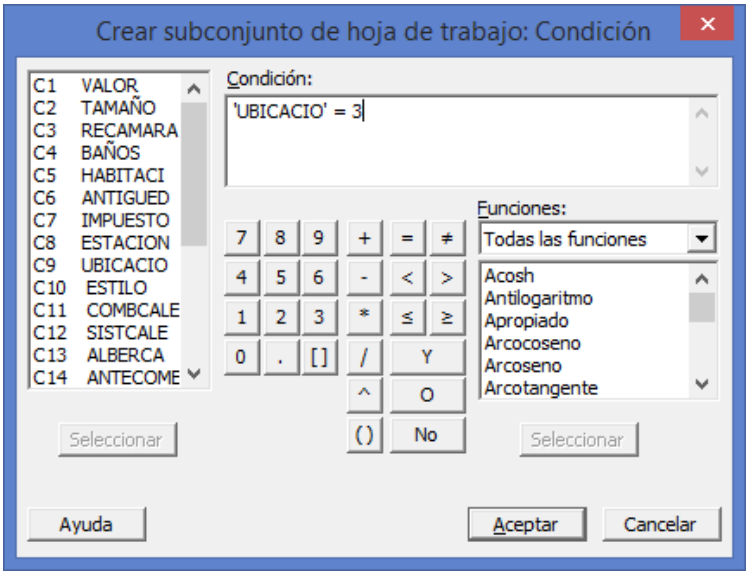
4 En **Especificar las filas que se incluirán**, elija **Filas que coincidan**. Haga clic en **Condición**.

5 En **Condición**, ingrese ***UBICACIÓN=3.*** Haga clic en **Aceptar** en cada cuadro de diálogo.

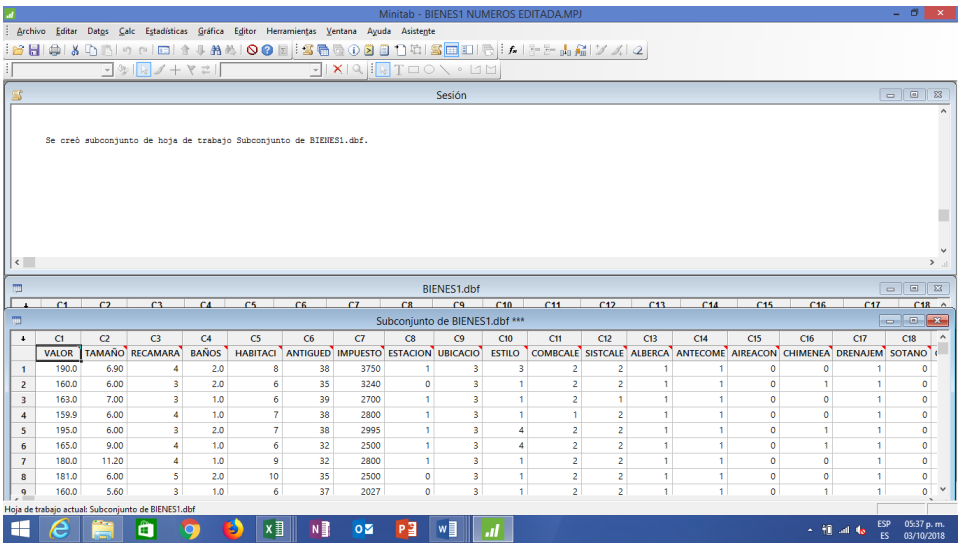
trabajo siempre copian los datos en hojas de trabajo **nuevas**. Usted puede utilizar Copiar columnas para reemplazar datos en la hoja de trabajo **actual** con un subconjunto.



Nota importante: Minitab es sensible al nombre que se le asignó a la variable por lo cual es conveniente no escribir el nombre de la variable sino seleccionarla de la parte izquierda de la ventana y solo añadir **=3.**



Minitab Crea un subconjunto de hoja nueva que incluye las casas ubicadas en la zona **ESTE**



Utilice esta opción de prueba de valores atípicos para identificar un valor atípico individual en una muestra. Las hipótesis son las siguientes:

- H_0 (la hipótesis nula): Todos los valores de la muestra corresponden a la misma población normalmente distribuida.
- H_1 (la hipótesis alternativa): Uno de los valores de la muestra no corresponde a la misma población normalmente distribuida.

Las opciones de las relaciones de Dixon le ayudan a detectar un valor atípico en presencia de otros supuestos valores atípicos. Para una explicación más detallada, véase Elección de la prueba adecuada de valores atípicos.

No debe utilizar las pruebas de valor atípico de Minitab más de una vez en la misma muestra. Si elimina un valor atípico de su muestra y luego vuelve probar, arriesga

Realizar la Prueba de valores atípicos de Minitab

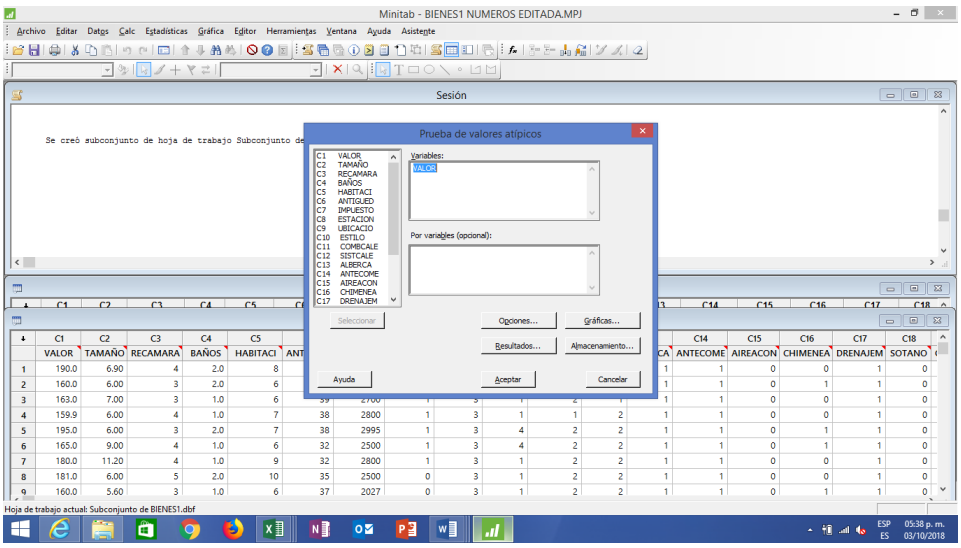
1 En la nueva hoja de trabajo que Minitab generó de la base de datos **"BIENES1"** que incluye solo las casas ubicadas en la zona **ESTE** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***Valor***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.



eliminar valores que no son en realidad atípicos.

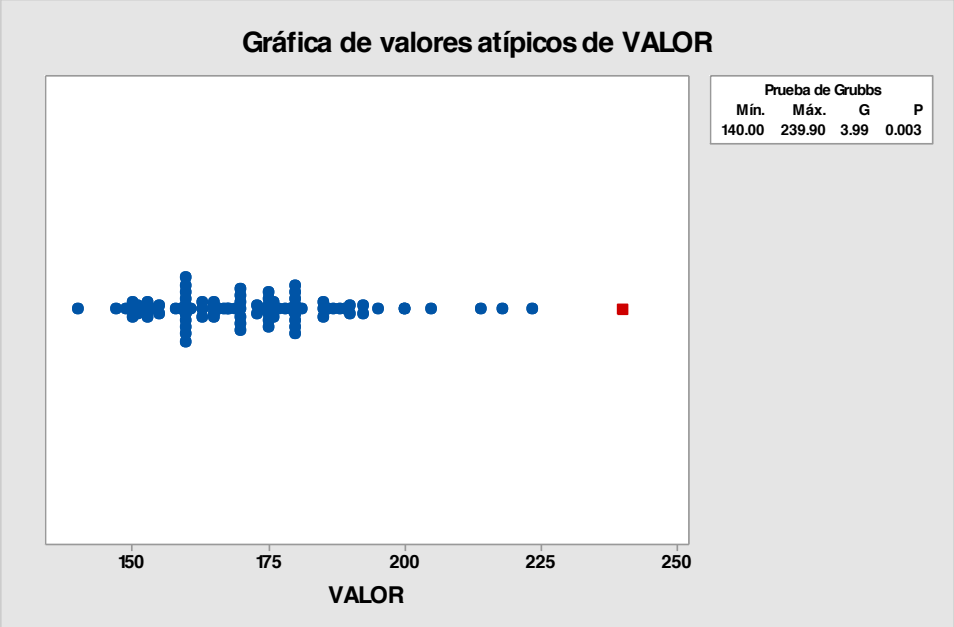
Salida de la ventana Sesión

Prueba de valores atípicos: VALOR

```
Método

Hipótesis nula          Todos los valores de los datos provienen de
la misma población normal
Hipótesis alterna       El valor más pequeño o más grande de los
datos es un valor atípico
Nivel de significancia  α = 0.05


Prueba de Grubbs

Variable   N   Media  Desv.Est.    Mín.    Máx.      G      P
VALOR     99  172.34      16.92  140.00  239.90  3.99  0.003


Valor atípico

                 Valor
Variable  Fila  atípico
VALOR       65    239.9
```

Gráfica de valores atípicos de VALOR

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **VALOR** muestran el valor más pequeño es de 140 y el más grande de 239. La estadística de G de 3.49 indica que el valor de datos más pequeño ó más grande de 100 ó 305, es 3.99 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.003. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico. En este caso corresponde al valor 305 que se observa en rojo en la gráfica y que pertenece al caso o **fila 65 del nuevo subconjunto de hoja de trabajo**.

Es importante aclarar que hay que ubicar dicho caso en la base de datos "BIENES1" original por si es el caso en que haya que eliminirlo para depurar dicha base de datos.

1.3.1.3 ACTIVIDAD DE **APRENDIZAJE EN MINITAB 17. Datos Atipicos. Un nivel de Factor**

ACTIVIDAD DE
APRENDIZAJE 1.3.1.3
DATOS ATÍPICOS PARA UN NIVEL DE UN **FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO**

Se reciben datos sobre una encuesta realizada a las casas ubicadas en la zona **SUR** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en esta zona **SUR** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

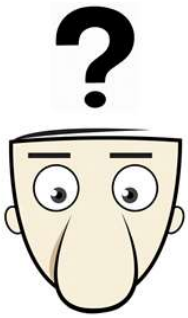
a) Crear un subconjunto de hoja de trabajo.
b) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

1.3.1.3 EJERCICIO DE AUTOEVALUACIÓN EN MINITAB 17. Datos Atipicos. Un nivel de Factor

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 1.3.1.3
DATOS ATÍPICOS PARA UN NIVEL DE UN FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO



Se reciben datos sobre una encuesta realizada a las casas ubicadas en la zona **NORTE** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en esta zona **NORTE** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

a) Crear un subconjunto de hoja de trabajo.
b) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver esta Autoevaluación.

1.3.1.3 EJERCICIOS DE REFUERZO EN MINITAB 17. Datos Atipicos. Un nivel de Factor

EJERCICIOS DE REFUERZO

1.3.1.3.1 Se reciben datos sobre una encuesta realizada a las casas con estilo **MODERNO** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la

1.3.1.3 DATOS ATÍPICOS PARA UN NIVEL DE UN FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

variable **VALOR** en las casas con estilo **MODERNO** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

a) Crear un subconjunto de hoja de trabajo.
b) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

1.3.1.3.2 Se reciben datos sobre una encuesta realizada a las casas con estilo **HACIENDA** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en las casas con estilo **HACIENDA** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

a) Crear un subconjunto de hoja de trabajo.
b) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

1.3.1.3.3 Se reciben datos sobre una encuesta realizada a las casas con estilo **GRANJA** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en las casas con estilo **GRANJA** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

a) Crear un subconjunto de hoja de trabajo.
b) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

1.3.1.4.D.A. EJEMPLO ILUSTRATIVO EN MINITAB 17. Datos Atípicos. Diagrama de Caja y Brazos Simple.

EJEMPLO ILUSTRATIVO INTEGRAL EN MINITAB 17. DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

Por opción predeterminada, en Minitab, la Valla Externa Inferior se extiende hasta el valor más bajo dentro del límite inferior.

Límite inferior = Q1- 3.0 (Q3 - Q1)

Por opción predeterminada, en Minitab, la Valla Interna Inferior se extiende hasta el valor más bajo dentro del límite inferior.

Límite inferior = Q1- 1.5 (Q3 - Q1)

Una agencia de Bienes Raíces desea examinar el **VALOR** general de las casas que forman su cartera de ventas. Las casas se encuentran ubicadas en diferentes orientaciones de la ciudad. De la base de datos **"BIENES1"** se eligió una muestra aleatorioa de 30 de ellas que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de la misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

215.0	210.0	174.5	177.5	297.5	180.0
181.0	205.0	260.0	173.0	290.0	160.0
189.0	154.0	175.0	174.9	219.9	176.5
176.0	215.0	286.5	185.0	189.9	189.9
267.0	280.0	169.5	164.9	159.9	185.0

Utilice el paquete Minitab para resolver esta Autoevaluación.

Para introducir los datos agrega una hoja de trabajo vacía al proyecto actual:

1 Elija **Archivo > Nuevo**.

2 Seleccione **Hoja de trabajo de Minitab** y haga clic en **Aceptar**.

3 En **C1** coloca el nombre de la variable **VALOR**.

4 Teclea los datos de la tabla en forma vertical.

Por opción predeterminada, en Minitab, la parte inferior de la caja es el primer cuartil (Q1) - 25% de los valores de datos son menores que o iguales a este valor.

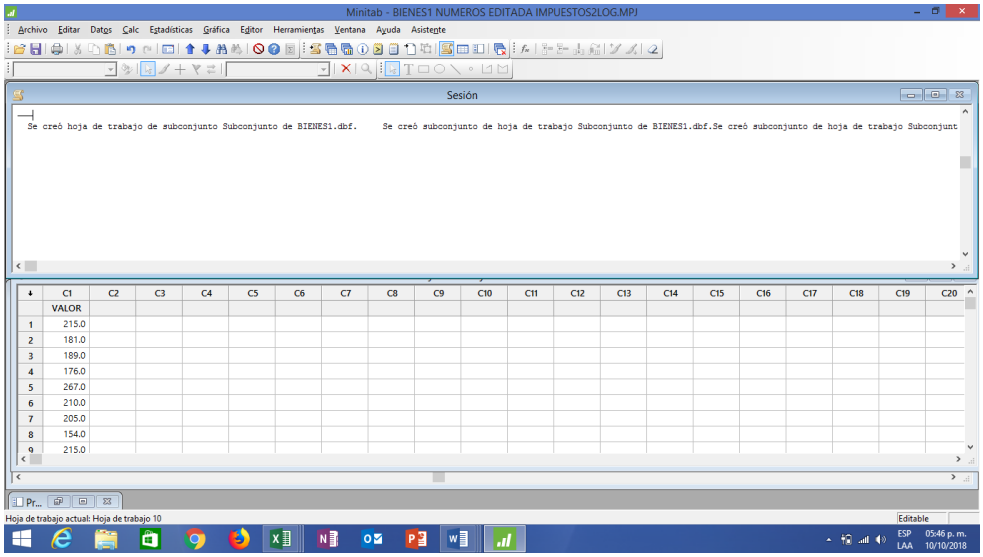
Por opción predeterminada,en Minitab, la parte superior de la caja es el tercer cuartil (Q3) - 75% de los valores de datos son menores que o iguales a este valor.

Por opción predeterminada, en Minitab, La Valla Interna Superior se extiende hasta el valor de datos más alto dentro del límite superior.

Límite superior = Q3 + 1.5 (Q3 - Q1)

Por opción predeterminada, en Minitab, La Valla Externa Superior se extiende hasta el valor de datos más alto dentro del límite superior.

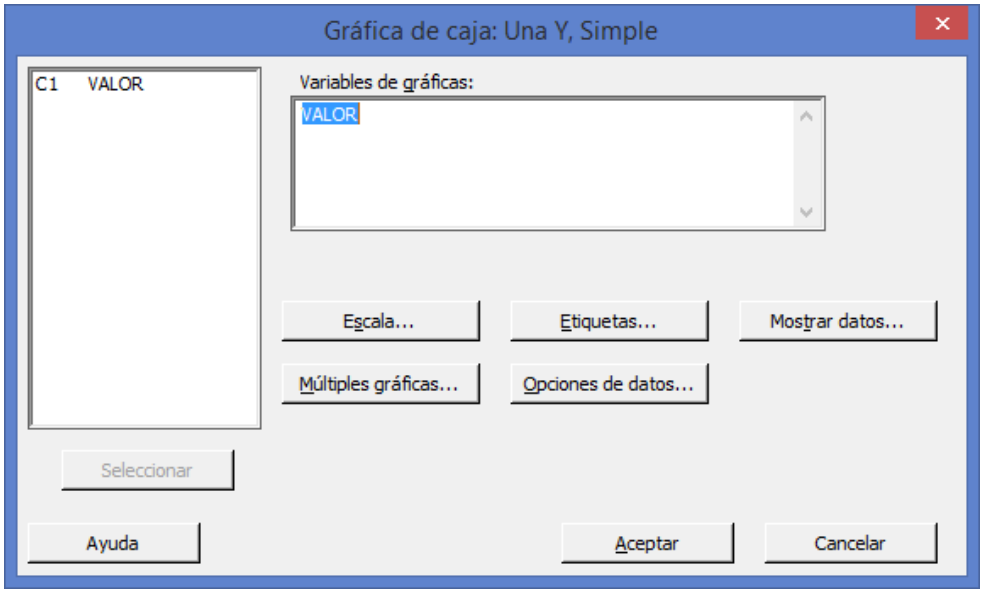
Límite superior = Q3 + 3.0 (Q3 - Q1)



5 Elija **Gráfica > Gráfica de caja**.

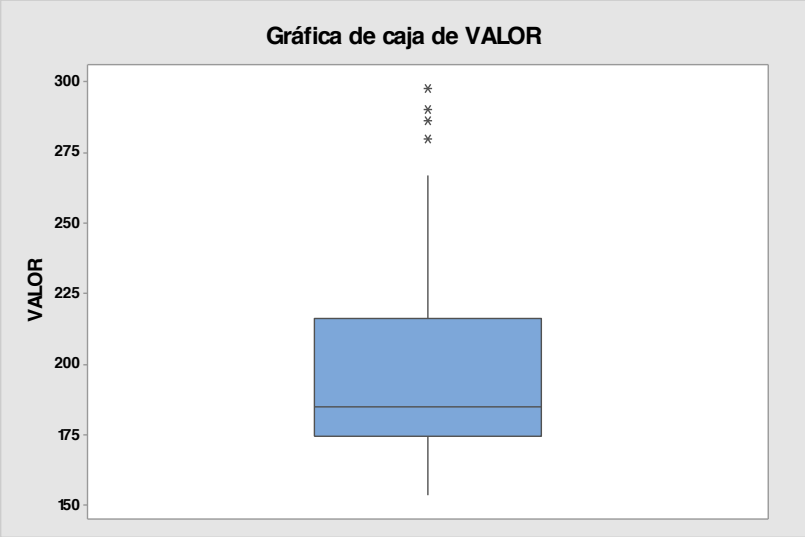
6 En **Una Y**, elija **Simple**. Haga clic en **Aceptar**.

7 En **Variables de graficas**, ingrese ***VALOR***.

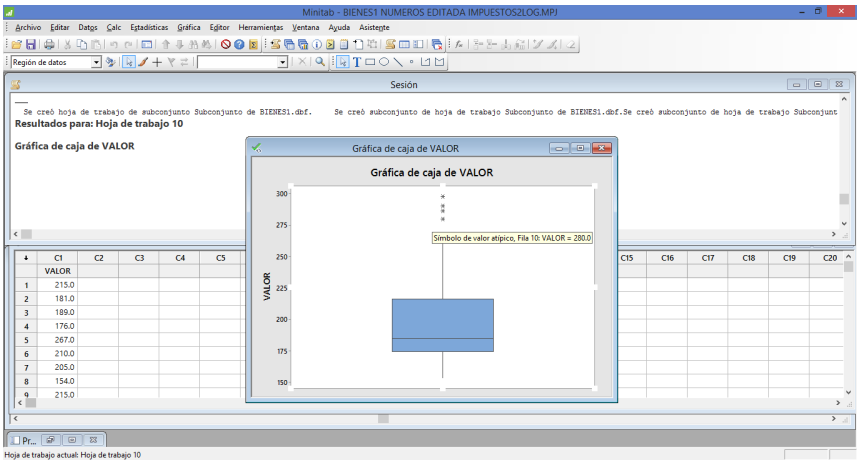


8 Haga clic en **Aceptar**.

Salida de la ventana Gráfica



Interpretación: En el diagrama anterior se puede observar que existen 4 valores atípicos que pueden ser identificados colocando el puntero del mouse en cada uno de ellos proporcionandonos la siguiente información:



Fila 10 Valor 280.0

Fila 14 Valor 286.5

Fila 22 Valor 290.0

Fila 21 Valor 297.5

Estos valores ó datos ó casos atípicos deben ser estudiados con detalle para ver el tratamiento que se le van a dar en estudios subsecuentes ya que es posible que provoquen un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la izquierda de la caja sugiere un sesgo hacia la derecha producto tal vez de los valores o casos atípicos.

1.3.1.4 ACTIVIDAD DE **APRENDIZAJE. Datos Atípicos. Diagrama de Caja y Brazos Simple.**

ACTIVIDAD DE APRENDIZAJE 1.3.1.4 DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE

Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **NORTE** que forman su cartera de ventas. De la base de datos **"BIENES1"** se eligieron aleatoriamente a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

215.0	267.0	247.0	235.0	170.0	155.0
195.0	199.9	218.0	190.0	215.0	179.0
160.0	169.0	299.9	180.0	229.5	195.0
189.0	179.0	172.0	213.0	180.0	179.9
249.0	218.0	230.0	175.0	225.0	225.0

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

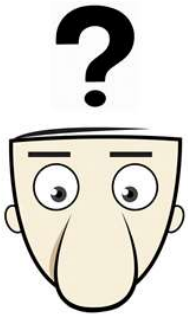
1.3.1.4 EJERCICIO **DE AUTOEVALUACIÓN. Datos Atipicos. Diagrama de Caja y Brazos Simple.**

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 1.3.1.4 DATOS ATÍPICOS. DIAGRAMA DE CAJA

Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de

Y BRAZOS SIMPLE



las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

181.0	181.0	179.9	200.0	230.0	174.9
220.0	185.0	210.0	115.0	183.0	100.0
190.0	269.9	159.9	199.9	215.0	174.9
160.0	189.9	165.0	189.9	159.0	189.9
176.0	299.9	185.0	189.9	180.0	210.0

Utilice el pauete Minitab para resolver esta Autoevaluación.

1.3.1.4 EJERCICIOS DE **REFUERZO. Datos Atípicos. Diagrama de Caja y Brazos Simple.**

EJERCICIOS DE REFUERZO 1.3.1.4 DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

1.3.1.4.1 Una agencia de Bienes Raíces desea examinar el **IMPUESTO** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. De la base de datos **"BIENES1"** se seleccionan aleatoriamente a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **IMPUESTO** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

3959	3959	2911	3871	3638	2500
3142	3700	3930	2400	2708	1775
2600	4500	2378	2540	2989	3000
2250	2860	1800	4200	1814	2386
2857	4343	2700	3600	2800	3500

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo

1.3.1.4.2 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **NORTE** que forman su cartera de ventas. De la base de datos **"BIENES1"** se seleccionan aleatoriamente a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas la **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin

embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

Los datos se muestran a continuación:

30	33	30	34	41	35
35	38	30	44	35	37
35	37	15	40	28	34
24	35	35	35	14	25
22	31	31	38	30	33

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo

1.3.1.4.3 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. De la base de datos **"BIENES1"** se seleccionan aleatoriamente a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas la **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

32	28	35	30	61	34
60	27	30	16	39	21
27	28	39	28	31	34
35	35	30	31	55	38
30	14	28	31	32	24

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo

1.3.1.5.D.A. EJEMPLO ILUSTRATIVO EN MINITAB 17. Datos Atípicos. Diagrama de Caja y Brazos Múltiple.

EJEMPLO ILUSTRATIVO 1.3.1.5. INTEGRAL EN MINITAB 17. DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE

Por opción predeterminada, en Minitab, la Valla Externa Inferior se extiende hasta el valor más bajo dentro del límite inferior.

Límite inferior = Q1- 3.0 (Q3 - Q1)

Por opción predeterminada, en Minitab, la Valla Interna Inferior se extiende hasta el valor más bajo dentro del límite inferior.

Límite inferior = Q1- 1.5 (Q3 - Q1)

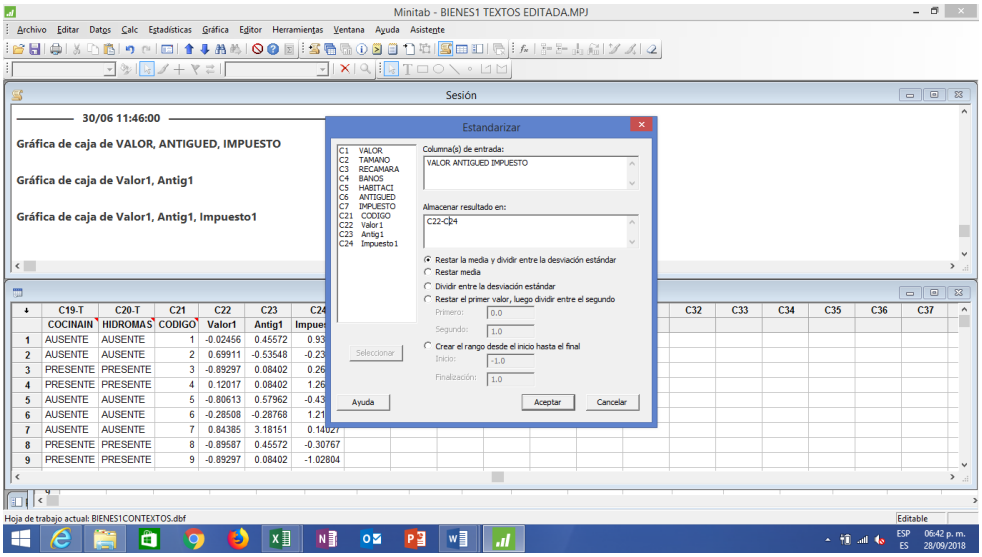
Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

Nota importante: Como las variables VALOR, ANTIGÜEDAD E IMPUESTOS están medidas en diferentes escalas primero se deben estandarizar las variables de la siguiente manera:

1 Abra la base de datos que creó **"BIENES1".** Elija **Calc > Estandarizar**.

2 En **Columnas de entrada**, ingrese las columnas que desea estandarizar. En este caso ingrese **VALOR, ANTIGUEDAD E IMPUESTOS.**

3 En **Almacenar resultados en**, ingrese las columnas en las que desea almacenar los datos estandarizados. En este caso escriba **C22-C24**, haga clic en **Aceptar**



Por opción predeterminada, en Minitab, la parte inferior de la caja es el primer cuartil (Q1) - 25% de los valores de datos son menores que o iguales a este valor.

Por opción predeterminada,en Minitab, la parte superior de la caja es el tercer cuartil (Q3) - 75% de los valores de datos son menores que o iguales a este valor.

Por opción predeterminada, en Minitab, La Valla Interna Superior se extiende hasta el valor de datos más alto dentro del límite superior.

Límite superior = Q3 + 1.5 (Q3 - Q1)

Por opción predeterminada, en Minitab, La Valla Externa Superior se extiende hasta el valor de datos más alto dentro del límite superior.

Límite superior = Q3 + 3.0 (Q3 - Q1)

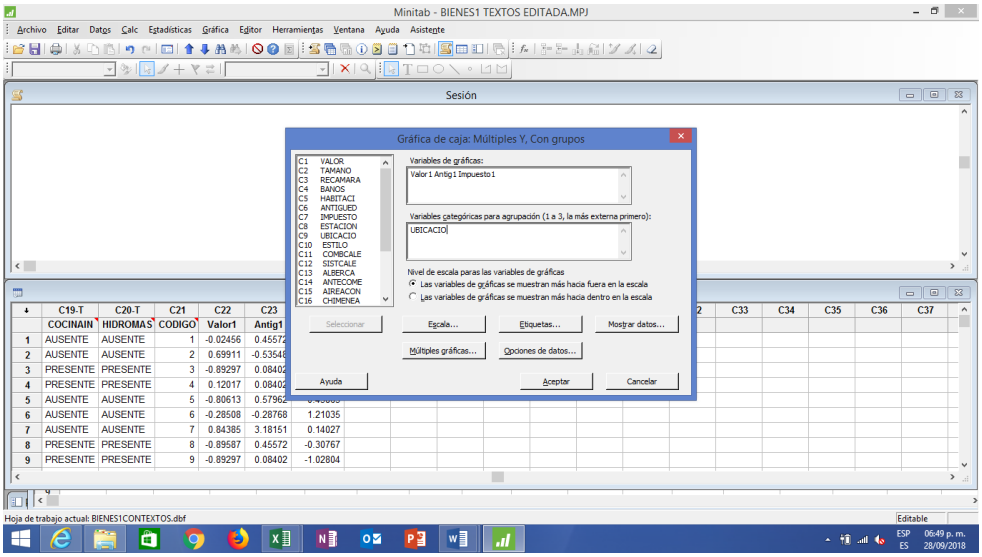
4 Los datos estandarizados aparecerán en las columas 22,23 y 24 de la base original **"BIENES1"**. Colocar el nombre de las variable estantarazadas como **Valor 1, Antig1 e Impuesto 1** para identificarlas.

1 Elija **Gráfica > Gráfica de caja**.

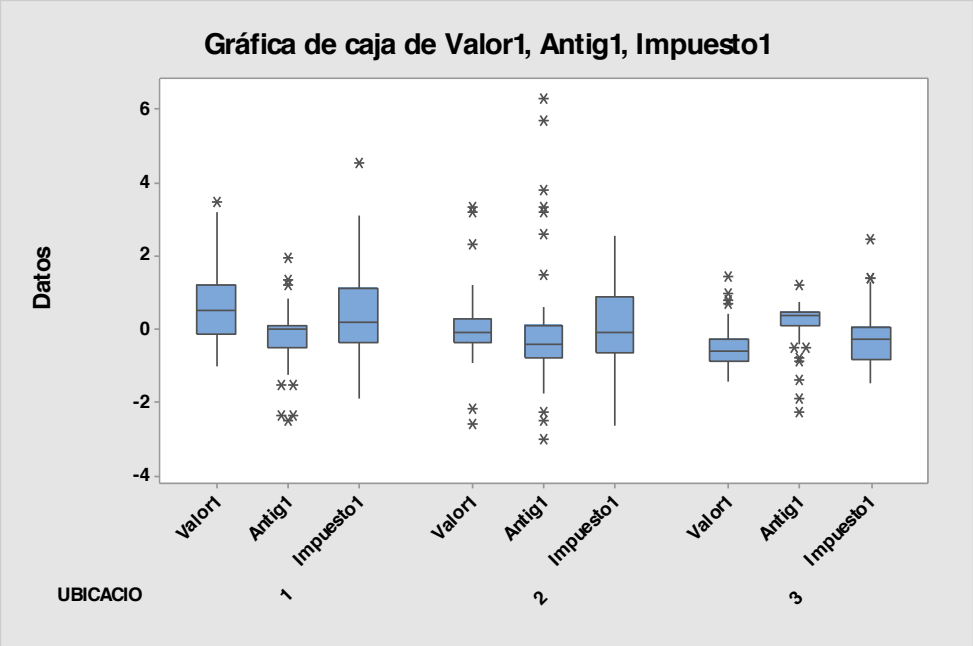
2 En **Múltiples Y**, elija **Con grupos**. Haga clic en **Aceptar**.

3 En **Variables de gráficas**, ingrese ***Valor1, Antig1 e Impuesto1***

4 En **Variables categóricas para agrupación (1 a 3, la más externa primero)**, ingrese ***Ubicación***



5 En **Nivel de escala para variables de gráficas**, elija **Las variables de gráficas se muestran más hacia dentro en la escala**. Haga clic en **Aceptar**.



Interpretación de los resultados

La gráfica de caja muestra:

· En general en todos los niveles de **UBICACIÓN** de las casas las tres variables muestran valores atípicos que hay que estudiar con excepción de **Valor1 e Impuestos1** en la parte inferior de la zona **NORTE, Impuestos1** en ambas partes de la zona **SUR** y **Valor1** e **Impuestos1** en la parte inferior de la zona **ESTE**.

· Para las casas ubicadas en la zona **SUR**, en la variable **Antig1** se detectan la mayor cantidad de valores atípicos con valores sumamente extremos.

Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

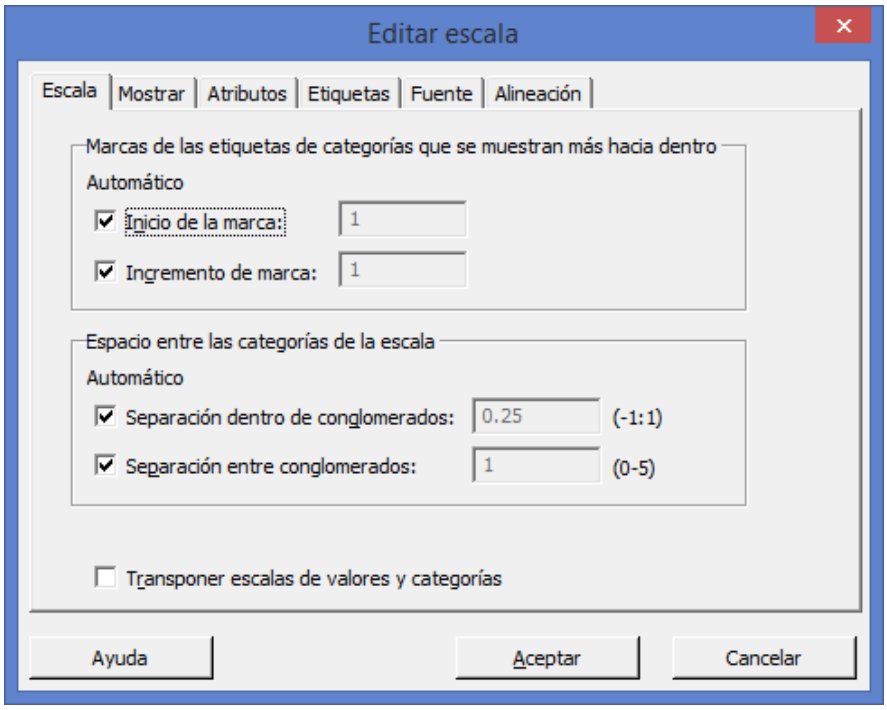
Asimismo se recomienda realizar análisis estadísticos, tales como **MANOVA balanceado**, para examinar más a fondo la relación entre los factores.

Para obtener información precisa de Q1, mediana, Q3, rango intercuartil, bigotes o brazos y N, coloque el cursor sobre cualquier parte de la gráfica de caja. Para ver el valor del eje Y y el número de fila de un valor atípico, coloque el cursor sobre el valor atípico.

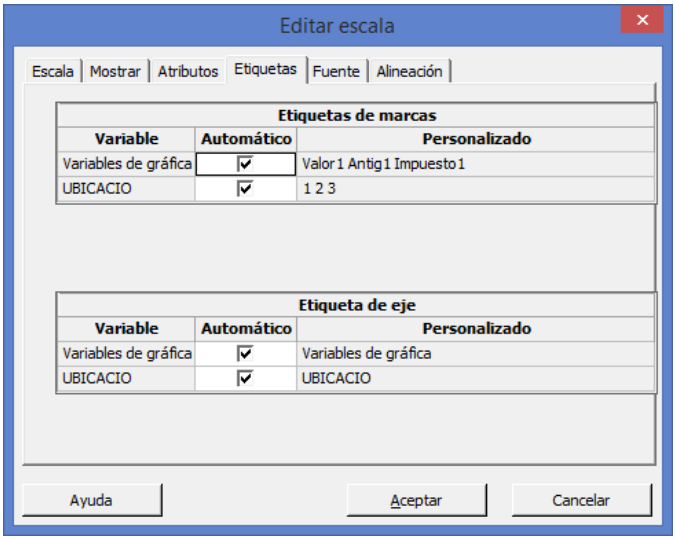
Etiquetado de los niveles de la variable o "Factor" UBICACIÓN:

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN Y Etiquetado del subtitulo CASAS DE "BIENES1":

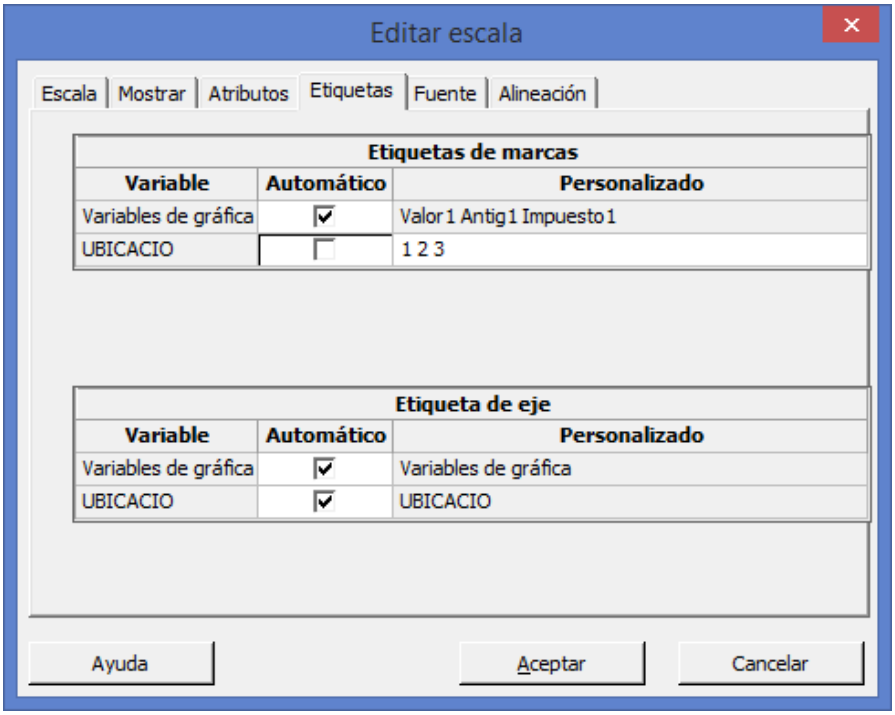
1 Haga doble clic en el número 1 en el renglón de UBICACIÓN para que Minitab despliegue la ventana ***Editar Escala.***



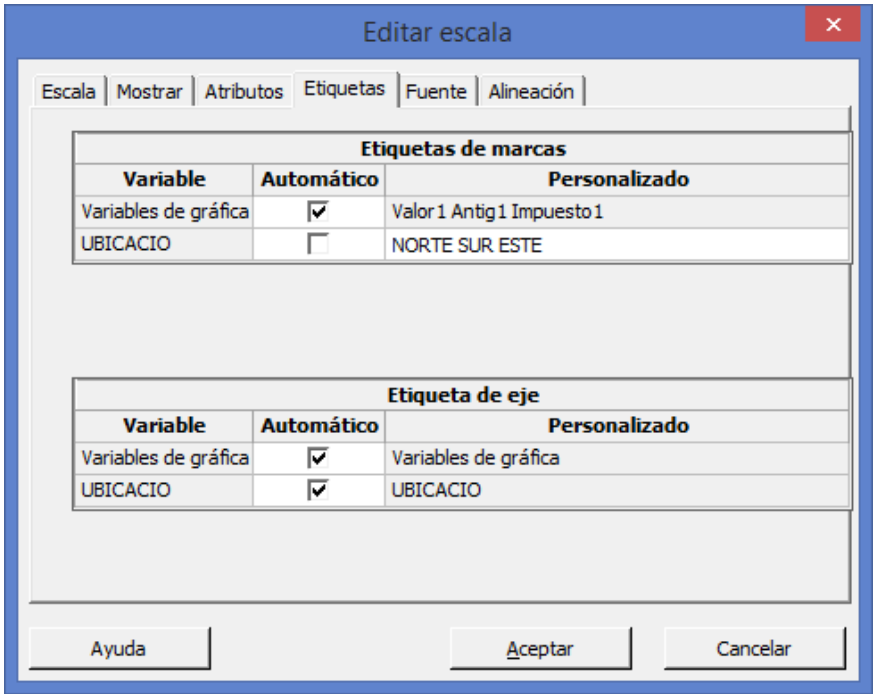
2 Seleccionar la pestaña **Etiquetas**



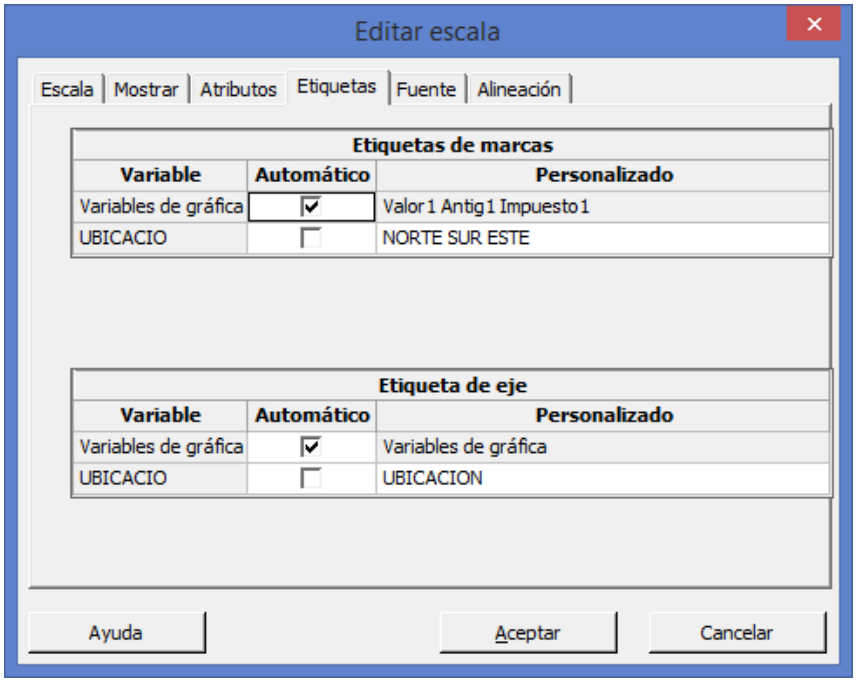
3 Desactivar la casilla **UBICACIÓN**



4 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE.**

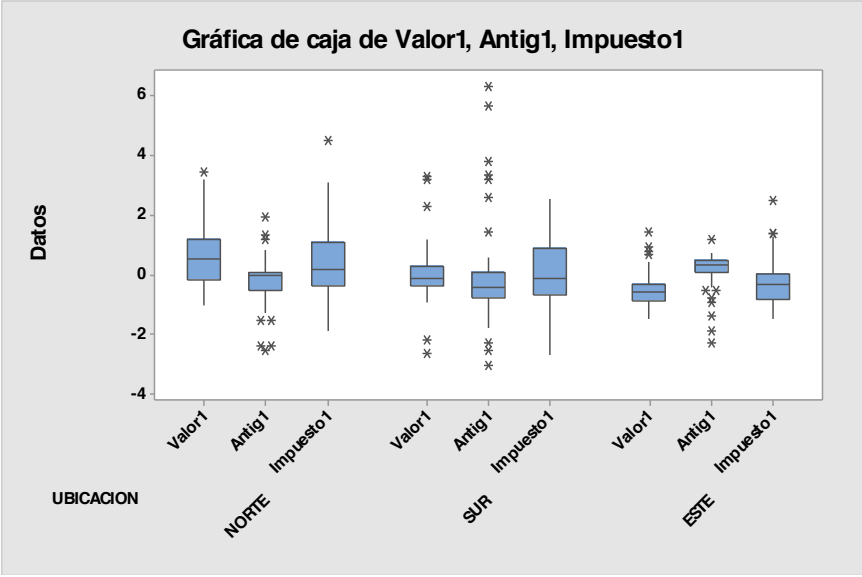


5 En **Etiqueta de eje** , si es el caso, desactivar la casilla de **UBICACIO** y escribir **UBICACIÓN.**

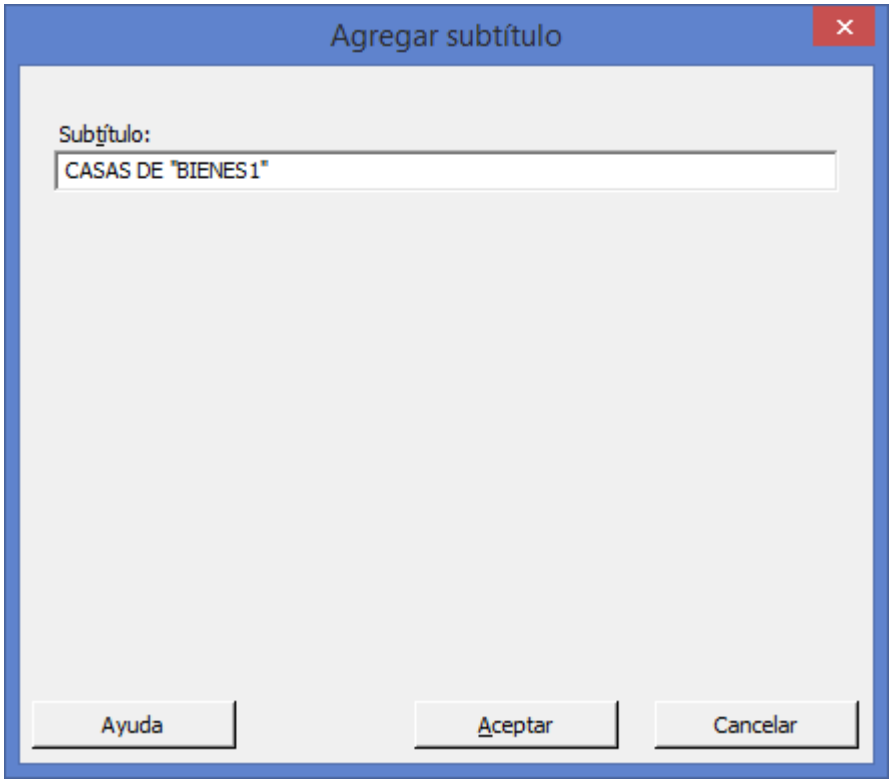


6 Haga clic en **Aceptar**.

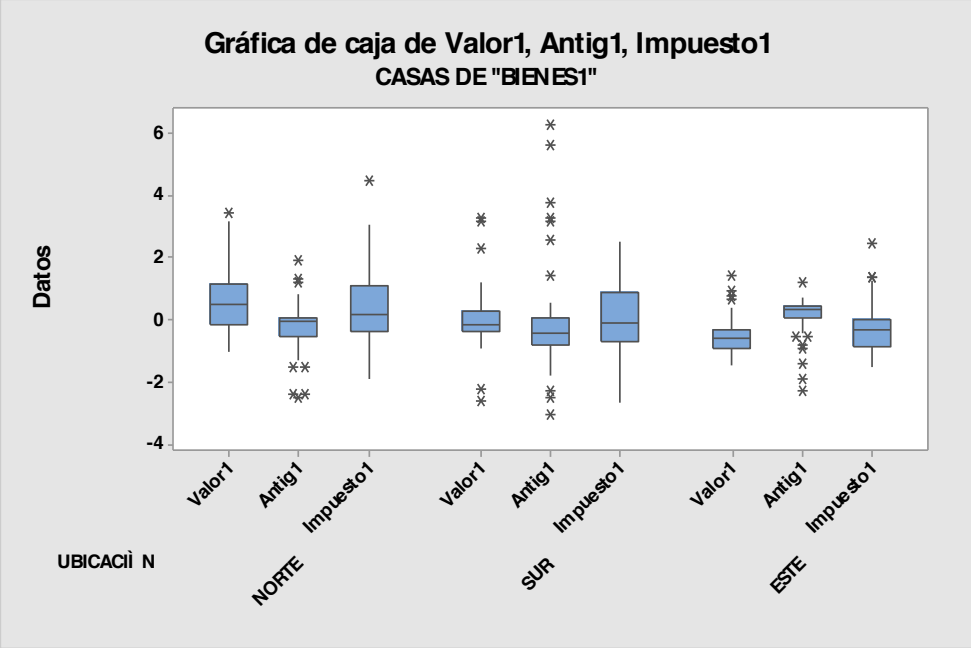
7 Aparece la gráfica debidamente etiquetada.



8 Coloque el mouse en el título de la grafica y presione el botón derecho del mouse. Seleccione la opción **Agregar** y la opción **Subtitulo. S**e escribe **CASAS DE "BIENES1"** en la ventana **Agregar subtítulo**.



9 Haga clic en **Aceptar.**
10 Aparece la gráfica debidamente etiquetada.



1.3.1.5 ACTIVIDAD DE **APRENDIZAJE. Datos Atípicos. Diagrama de Caja y Brazos Múltiple**

ACTIVIDAD DE APRENDIZAJE 1.3.1.5 DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE

Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas con estilo **MODERNO** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con el estilo MODERNO** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables.

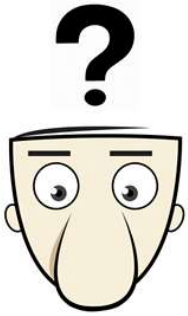
1.3.1.5 EJERCICIO **DE AUTOEVALUACIÓN. Datos Atipicos. Diagrama de Caja y Brazos Múltiple.**

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 1.3.1.5 DATOS ATÍPICOS.

Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio.

DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE



Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **CON AIRE ACONDICIONADO** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

Utilice el paquete Minitab para resolver esta Autoevaluación.

Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las caas que tienen AIRE ACONDICIONADO** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en

1.3.1.5 EJERCICIOS DE **REFUERZO. Datos Atípicos. Diagrama de Caja y Brazos Múltiple.**

EJERCICIOS DE REFUERZO 1.3.1.5 DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE.

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de

1.3.1.5.1 Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **SIN SÓTANO** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las caas que no tienen SOTANO** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables.

1.3.1.5.2 Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **SIN CHIMENEA** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las casas que no tienen CHIMENEA** y después, como las variables **VALOR,**

cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

ANTIGÜEDAD E IMPUESTOS están medidas en diferentes escalas se deben **estandariza**r las variables.

1.3.1.5.3 Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **SIN TINA DE HIDROMASAJE** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las casas que no tienen TINA DE HIDROMASAJE** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables.

1.3.2 ANÁLISIS PREVIO DE LOS DATOS MULTIVARIADOS. VERIFICACIÓN DE LOS SUPUESTOS DEL ANÁLISIS MULTIVARIADO. NORMALIDAD. HOMOCEDASTICIDAD.LINEALIDAD.

CONCEPTOS **BÁSICOS**
VERIFICACION DE LOS SUPUESTOS DEL ANÁLISIS **MULTIVARIADO**

VERIFICACION DE LOS SUPUESTOS DEL ANÁLISIS MULTIVARIADO.

La última etapa del examen de los datos consiste en la comprobación de los supuestos subyacentes básicos del análisis multivariado. Esta necesidad de comprobar los supuestos estadísticos aumenta en las aplicaciones multivariadas como consecuencia de la complejidad de las relaciones, por la gran cantidad de variables que hace que distorsiones y los sesgos potenciales sean más potentes cuando se incumplen los supuestos, por lo tanto, el investigador debe estar atento a cualquier incumplimiento de los supuestos y a las implicaciones que puedan tener para el proceso de estimación o interpretación de resultados.

El análisis multivariante requiere que los supuestos subyacentes a las técnicas multivariadas sean contrastados dos veces, primero las variables aisladas, similar a las pruebas de los supuestos univariados y posteriormente para el **valor teórico** del modelo multivariante, que actúa colectivamente sobre las

variables a analizar y por lo tanto, debe cumplir los mismos supuestos que las variables individuales.

1. NORMALIDAD:

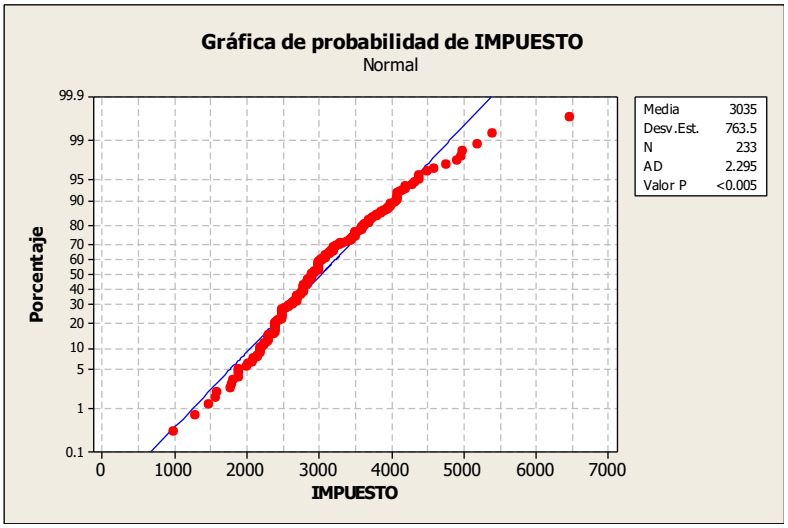
En el análisis multivariado, el supuesto fundamental es la **normalidad** de los datos, en referencia al perfil de la distribución de los datos para una única variable métrica y su correspondencia con una **distribución normal**. Si la variación con respecto a la normal es amplia, todas las pruebas estadísticas resultantes no son válidas, dado que se requiere normalidad para el uso de los estadísticos t y F. La normalidad multivariante, es decir la combinación de dos o más variables, implica que las variables individuales sean normales en un sentido univariado y que sus combinaciones también sean normales. La normalidad multivariante es más difícil de probar aunque existen varias pruebas para situaciones en las que la técnica multivariada se ve particularmente afectada por una violación de los supuestos.

Análisis gráfico de la normalidad:

La prueba más simple para diagnosticar la normalidad es una comprobación visual del histograma que compare los valores de los datos observados con una distribución aproximada de la distribución normal. Aunque este método es simple, también es problemático para muestras pequeñas, donde la construcción de un histograma puede distorsionar la representación visual de tal forma que el análisis sea inútil. Una forma de mayor confianza es el **gráfico de distribución normal** que compara la distribución acumulada de los valores reales de los datos con la distribución acumulada de una distribución normal. La distribución normal sigue una línea recta en diagonal. Si una distribución es normal, la línea que representa a la distribución real de los datos sigue de cerca de la diagonal.

La prueba de normalidad es una prueba de hipótesis para determinar si la población de la cual extrajo su muestra es no normal. Muchos procedimientos estadísticos dependen de la normalidad de la población, de modo que recurrir a una prueba de normalidad para determinar si se rechaza este supuesto pudiera ser un paso importante en su análisis. La hipótesis nula para una prueba de normalidad establece que la población es normal. La hipótesis alternativa establece que la población es no normal. Para determinar si los datos de su muestra provienen de una población no normal, usted puede elegir en general entre tres pruebas:

La prueba de **Anderson-Darling** es una prueba que compara la función de distribución acumulada empírica de los datos de su muestra con la distribución esperada si los datos son normales. Si esta diferencia observada es suficientemente grande, la prueba rechazará la hipótesis nula de normalidad en la población.



Pruebas estadísticas de normalidad:

Pueden utilizarse pruebas estadísticas para evaluar la normalidad. Una de ellas es simple y es una regla basada en el valor de la asimetría. El valor estadístico (z) se calcula como:

$$Z_{simetría} = \frac{simetría}{\sqrt{\frac{6}{n}}}$$

Donde n es el tamaño de la muestra.

Un valor "z" que también puede ser calculado para el valor de la curtósis es:

$$Z_{curtósis} = \frac{curtósis}{\sqrt{\frac{24}{n}}}$$

Si el valor calculado de "z" excede un valor crítico, entonces la distribución real de los datos no es normal. Un valor calculado que exceda ± 2.58, indica que podemos rechazar el supuesto sobre la normalidad de la distribución con un nivel de significancia de 0.01. Otro valor crítico habitualmente utilizado es ± 1.96, que corresponde a un nivel de significancia de 0.05.

Hay pruebas estadísticas que se encuentran en los programas computacionales. Los dos más comunes son las pruebas de Shapiro-Wilks y una modificación de la prueba de Kolmogorov-Smornov. Cada uno determina el nivel de significancia para las diferencias respecto a una distribución normal. Estas pruebas son poco útiles para muestras pequeñas (menores de 30) y muy sensibles para grandes muestras (mayores a 1000 observaciones), por lo que el investigador debe siempre utilizar tanto los gráficos como cualquier comprobación estadística para evaluar el grado de desviación de la normalidad.

La prueba de normalidad de **Ryan-Joiber** es una prueba que evalúa la normalidad calculando la correlación entre sus datos y las puntuaciones normales de sus datos. Si el coeficiente de correlación se encuentra cerca de 1, es probable que la población sea normal. La estadística de Ryan-Joiner evalúa la solidez de esta correlación; si se encuentra por debajo del valor crítico apropiado, usted rechazará la hipótesis nula de normalidad en la población. Esta prueba es similar a la prueba de normalidad de Shapiro-Wilk.

La prueba de normalidad de **Kolmogorov-Smirnov** es una prueba que compara la función de distribución acumulada empírica de los datos de su muestra con la distribución esperada si los datos son normales. Si esta diferencia observada es suficientemente grande, la prueba rechazará la hipótesis nula de normalidad en la población.

Si el valor p de esta prueba es menor que su nivel a elegido, usted puede rechazar su hipótesis nula y concluir que la población es no normal.

Estadístico de Anderson-Darling

Mide si los datos siguen una distribución particular. Mientras mejor se ajusta la distribución a los datos, menor será este estadístico. Utilice el estadístico de Anderson-Darling para comparar el ajuste de varias distribuciones, para ver cuál es el mejor, o para probar si una muestra de datos proviene de una población con una distribución específica. Por ejemplo, puede utilizar el estadístico Anderson-Darling para elegir entre la distribución de Weibull y la normal, para un análisis de datos de confiabilidad o para probar si los datos cumplen con el supuesto de normalidad de una prueba t.

La prueba de **Anderson-Darling** es una prueba que compara la función de distribución acumulada empírica de los datos de su muestra con la distribución esperada si los datos son normales. Si esta diferencia observada es suficientemente grande, la prueba rechazará la hipótesis nula de normalidad en la población.

Las hipótesis para la prueba de Anderson-Darling son:

H_0: Los datos siguen una distribución específica

H_1: Los datos no siguen una distribucion específica

Si el valor p (al estar disponible) para la prueba de Anderson-Darling, es inferior al nivel de significancia seleccionado (generalmente 0.05 ó 0.01), concluya que los datos no siguen la distribución específica.

En muchas ocasiones en que se indica no normalidad es en realidad un resultado de otras violaciones de los supuestos, por lo tanto remediando los otros incumplimientos, eliminamos el problema de la normalidad, y por esto debemos realizar pruebas de normalidad después o a la par con los análisis y soluciones de otras violaciones.

2. HOMOCEDASTICIDAD:

Es un supuesto que se refiere a las relaciones de dependencia entre las variables, es decir, que las variables dependientes tengan iguales niveles de varianza a lo largo del rango del predictor de las variables. El efecto de la heterocedasticidad está relacionado con el tamaño de la muestra, especialmente cuando examinamos la dispersión de la varianza entre grupos. Por ejemplo, en ANOVA o MANOVA, el tamaño de la heterocedasticidad de las pruebas estadísticas depende de los tamaños de la muestra asociada con los grupos de menor o mayor varianza. En el análisis de regresión múltiple, ocurrirán efectos similares en distribuciones altamente simétricas donde existan un número desproporcionado de encuestados en ciertos rangos de la variable independiente.

Existen dos pruebas de homocedasticidad: la prueba de **Bonett** y la prueba de **Levene**. Para cada prueba, la hipótesis nula plantea que las dos varianzas son iguales (H_0: $s^2_1 / s^2_2 = 1$). La hipótesis alternativa puede ser de cola izquierda (H_1: $s^2_1 / s^2_2 < 1$), de cola derecha (H_1: $s^2_1 / s^2_2 > 1$), o de dos colas (H_1: $s^2_1 / s^2_2 \neq 1$). De forma opcional, las relaciones de prueba diferentes de 1 (igualdad) se pueden especificar.

Pruebas estadísticas de homocedasticidad:

Las pruebas estadísticas de igual dispersión de la varianza, se refieren a la varianza en grupos formados por variables métricas. La prueba más común es el test de **Levene** que puede usarse para evaluar si las varianzas de una única variable métrica son iguales a lo largo de cualquier cantidad de grupos y cuando los datos provengan de distribuciones continuas, **pero no necesariamente distribuciones normales**. Este método considera las distancias de las observaciones con respecto a la mediana de la muestra, en lugar de la media de la muestra, esto hace que la prueba sea más sólida para las muestras más pequeñas. Cuando los datos provengan de **distribuciones normales**, la prueba de **Barlett** es un poco más sólida que la de Levene. Una prueba *F* reemplaza a la prueba de Barlett cuando solo se tiene dos niveles.

En estadística, la **prueba de Bartlett** se utiliza para probar si *k* muestras provienen de poblaciones con la misma varianza. A las varianzas iguales a través de las muestras se llama homocedasticidad u homogeneidad de varianzas. Algunas pruebas estadísticas, por ejemplo, el análisis de la varianza ANOVA, suponen que las varianzas son iguales en todos los grupos o muestras. La prueba de Bartlett se puede utilizar para verificar esa suposición.

La prueba de Bartlett es sensible a las desviaciones de la normalidad. Es decir, si las muestras provienen de distribuciones no normales, entonces la prueba de Bartlett puede ser simplemente para probar la no normalidad.
La Prueba de Levene y la de Bonett son alternativas a la prueba de Bartlett que son menos sensibles a las desviaciones de la normalidad.[1]

Soluciones para la heterocedasticidad:

Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no

normalidad resuelve igualmente la dispersión de la varianza.

3. LINEALIDAD:

Es un supuesto implícito en las técnicas multivariadas basadas en medidas de correlación, como la regresión múltiple, logística, factorial, etc., debido a que las correlaciones representan solo la asociación lineal entre variables, los efectos no lineales no estarán representados en el valor de la correlación. Por eso es adecuado examinar las relaciones para identificar cualquier desplazamiento de la linealidad que pueda influir en la correlación.

Identificación de relaciones no lineales:

Una forma común de evaluar la linealidad es examinar los gráficos de dispersión de las variables e identificar cualquier patrón no lineal en los datos. Alternativamente se puede usar el análisis de residuos de la regresión múltiple, ya que éstos reflejan la parte no explicada de la variable dependiente y por lo tanto, cualquier parte no lineal de la relación quedará reflejada en los residuos.

Soluciones para la no normalidad:

La aproximación más directa es transformar una o ambas variables para conseguir la normalidad. Una alternativa a la transformación de los datos es crear una nueva variable que represente la parte no lineal de la relación.

En estadística, la transformación de datos se efectúa para asegurarse que estos tienen una distribución normal (lo que puede ser un remedio para analizar datos que tienen otros tipos de distribución, falta de normalidad, linealidad, y homocedasticidad). Cuando se usan cierto tipo de pruebas estadísticas la transformación de datos puede ser necesaria ya que este tipo de análisis asumen que los datos deben ser lineales, normales y tienen homocedasticidad en su varianza. Si el tamaño de la muestra es demasiado pequeño o los datos están sesgados hacia una distribución no normal se puede intentar hacer una de las siguientes transformaciones: logarítmica, raíz

TRANSFORMACIONES DE LOS DATOS:

Transformaciones de los datos para conseguir la normalidad y la homocedasticidad:

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva. Cuando se examinan los residuos de la regresión múltiple buscando la heterocedasticidad, se observa que un indicio de varianzas desiguales es una distribución con perfil de cono. Si el cono se abre a la derecha, la mejor transformación es la inversa; si se abre a la izquierda, elegimos la raíz cuadrada.

Transformaciones para conseguir la linealidad:

cuadrada o inversa.

Hay numerosos procedimientos para conseguir la linealidad entre dos variables, sin embargo, las más simples incluyen la aplicación de la raíz cuadrada, el logaritmo y la inversa. Se prueban cada una de las transformaciones hasta conseguir la linealidad o bien, crear una nueva variable, denominada polinómica, que represente la parte no lineal de la relación.

1.3.2.1.P.N. EJEMPLO ILUSTRATIVO EN MINITAB 17. Prueba de Normalidad. Anderson Darling

EJEMPLO ILUSTRATIVO 1.3.2.1. INTEGRAL EN MINITAB 17. PRUEBA DE NORMALIDAD

Estadísticas > Estadísticas básicas > Prueba de normalidad

Genera una gráfica de probabilidad normal y realiza una prueba de hipótesis para examinar si las observaciones siguen o no una distribución normal . Para la prueba de normalidad , las hipótesis son,

H_0: los datos siguen una distribución normal vs. H_1: los datos no siguen una distribución normal

La escala vertical de la gráfica se asemeja a la escala vertical del papel de probabilidad normal. El eje horizontal es una escala lineal. La línea forma un estimado de la función de distribución acumulada para la población de la cual se extrajeron los datos. Con la gráfica se muestran estimaciones numéricas de los parámetros de la población , m y s, el valor de la prueba de normalidad y el valor p asociado.

Elementos del cuadro de diálogo

Variable : Ingrese la columna que se utilizará para el eje x. Minitab calcula la probabilidad de ocurrencia de cada observación en la columna (presuponiendo una distribución normal) y utiliza el logaritmo de las probabilidades calculadas como valores y.

Líneas percentiles: Minitab marca cada uno de los porcentajes en la columna con una línea de referencia horizontal en la gráfica y marca cada línea con el valor porcentual. Minitab dibuja una línea de referencia vertical, en la que la línea de

referencia horizontal intersecta el ajuste de la línea a los datos, y marca esta línea con el valor de datos estimado.

Ninguno: Elija esta opción para no mostrar ninguna línea de percentil.

En los valores de Y: Elija esta opción para ingresar valores de escala Y para colocar las líneas de percentiles. Ingrese valores entre 0 y 100 cuando los porcentajes se utilicen como el tipo de escala Y, o 0 a 1 cuando la probabilidad es el tipo de escala Y.

En los valores de datos: Elija esta opción para ingresar valores de datos para colocar líneas de percentiles.

Pruebas de normalidad: Para obtener explicaciones acerca de las pruebas de normalidad, véase el capítulo correspondiente.

Anderson-Darling: Elija esta opción para realizar una prueba de Anderson-Darling de normalidad, que es una prueba basada en la ECDF (función de distribución acumulada empírica).

Ryan-Joiner: Elija esta opción para realizar una prueba de Ryan-Joiner , que es similar a la prueba de Shapiro-Wilk . La prueba de Ryan-Joiner es una prueba basada en correlaciones.

Kolmogorov-Smirnov: Elija esta opción para realizar una prueba de Kolmogorov-Smirnov de normalidad, que es una prueba basada en la ECDF.

Título: Para reemplazar el título predeterminado por su propio título personalizado, escriba el texto que desee en este cuadro.

La prueba de **Anderson-Darling** es una prueba que compara la función de distribución acumulada empírica de los datos de su muestra con la distribución esperada si los datos son normales. Si esta diferencia observada es suficientemente grande, la prueba rechazará la hipótesis nula de normalidad en la población.

PRUEBA DE NORMALIDAD. Prueba de Anderson Darling.

Se midió el **VALOR** de las casas en la base de datos "**BIENES1"** que se construyó anteriormente.

Se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad.

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > Estadísticas básicas > Prueba de normalidad**.

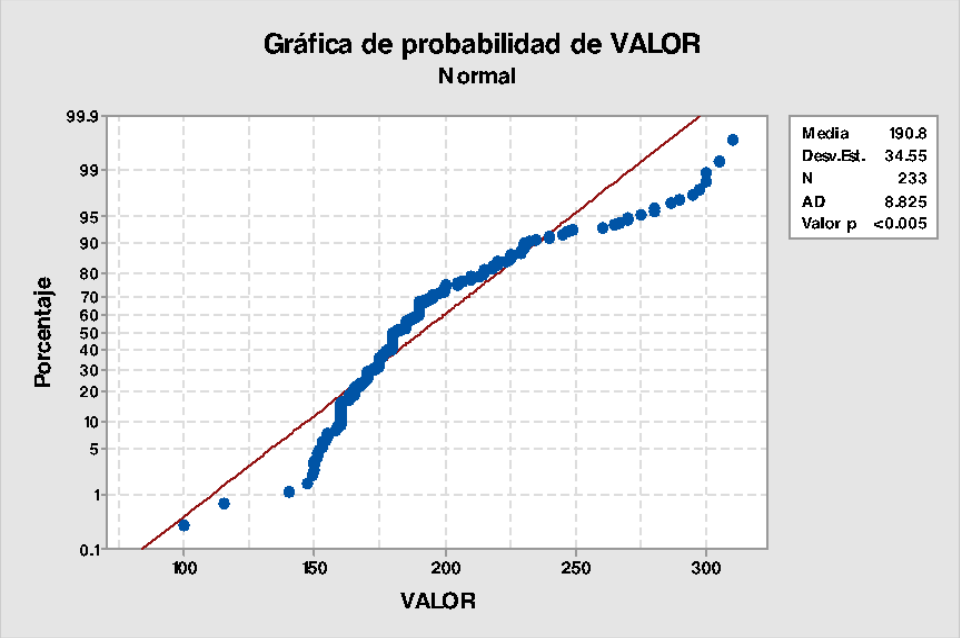
3 En **Variable**, ingrese ***Valor.*** Haga clic en **Aceptar**.

Salida de la ventana Gráfica

Genera una **gráfica de probabilidad** normal y realiza una **prueba de hipótesis** para examinar si las observaciones siguen o no una **distribución normal** . Para la **prueba de normalidad** , las hipótesis son,

H_0: los datos siguen una distribución normal vs. H_1: los datos no siguen una distribución normal

La escala vertical de la gráfica se asemeja a la escala vertical del papel de probabilidad normal. El eje horizontal es una escala lineal. La línea forma un estimado de la función de distribución acumulada para la **población** de la cual se extrajeron los datos. Con la gráfica se muestran **estimaciones numéricas de los parámetros de la población** , Media y Desviación Estándar, el tamaño de la muestra ó población, el valor de la prueba de normalidad (Estadístico de Anderson Darling) y el **valor p** asociado.



Interpretación de los resultados

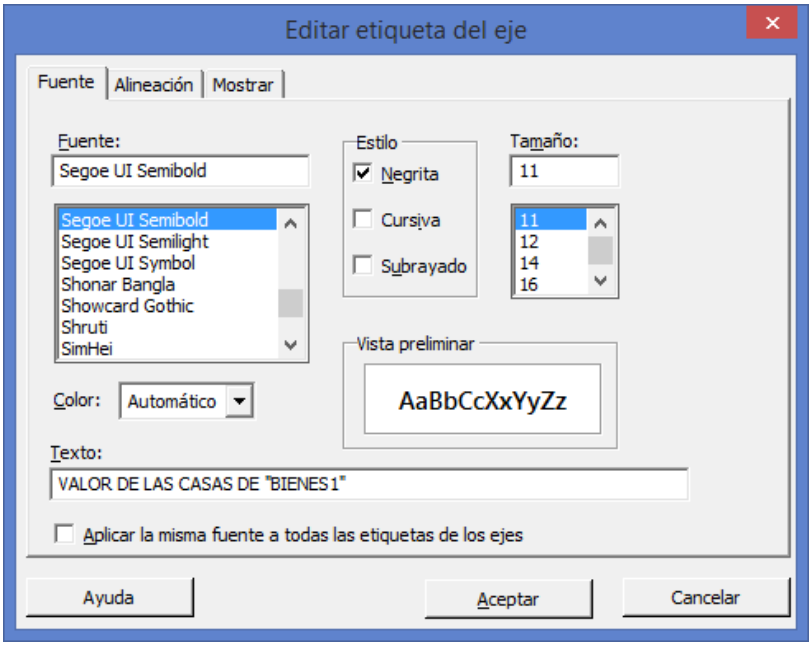
La salida gráfica es una gráfica de probabilidades normales versus los datos. Los datos se alejan de la línea ajustada de una manera más evidente en los extremos, o colas de la distribución. El valor p de la prueba de Anderson-Darling indica que, en niveles p menores a 0.005, hay evidencia de que los datos no siguen una distribución normal.

Los problemas de no normalidad puede solucionarse por medio de **transformaciones de datos** y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una o varias de las variables por lo que también se debe verificar el supuesto de homocedasticidad y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

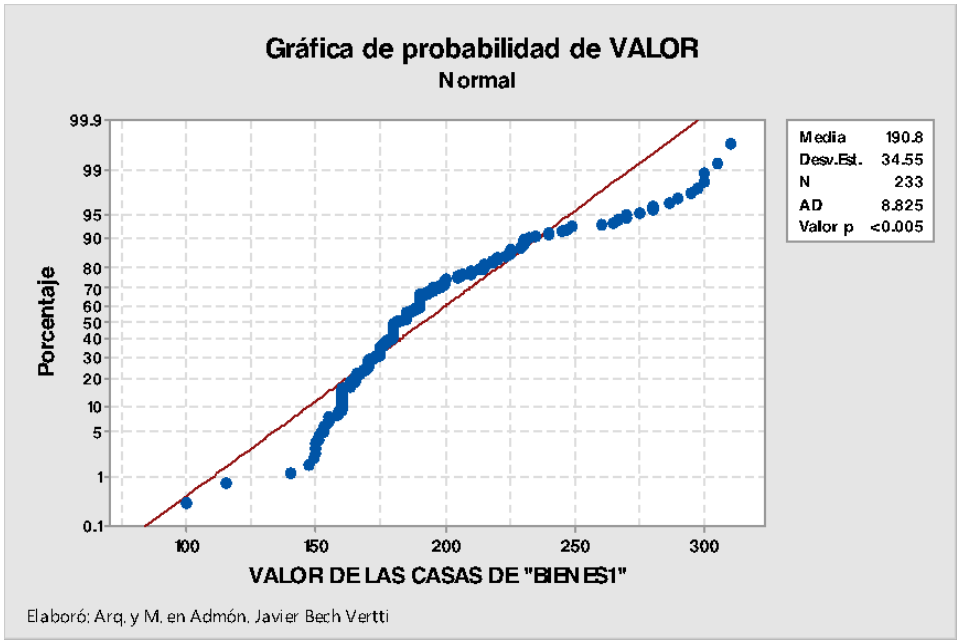
Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

Nota: Etiquete la gráfica dando doble clic en **VALOR** para que aparezca la ventana ***Editar etiqueta de eje*** y en la parte inferior donde dice ***Texto*** escriba **VALOR DE LAS CASAS DE "BIENES1"**. Haga clic en **Aceptar.**



Haga clic en **Aceptar.**



1.3.2.1 ACTIVIDAD DE **APRENDIZAJE. Prueba de Normalidad. Prueba de Anderson Darling.**

ACTIVIDAD DE APRENDIZAJE 1.3.2.1 PRUEBA DE NORMALIDAD. PRUEBA DE ANDERSON DARLING

Se midió el **VALOR** de las casas en la zona **ESTE** en la base de datos "**BIENES1"** que se construyó anteriormente.

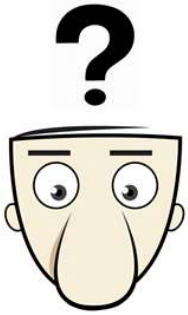
Se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas ubicadas en el **ESTE** e interprete los resultados y etiquetando correctamente todas las gráficas.

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

1.3.2.1 EJERCICIO DE AUTOEVALUACIÓN. Prueba de Normalidad. Prueba de Anderson Darling.

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 1.3.2.1 PRUEBA DE NORMALIDAD. PRUEBA DE ANDERSON DARLING



Se midió el **VALOR** de las casas en la zona **SUR** en la base de datos "**BIENES1"** que se construyó anteriormente.

Se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas ubicadas en el **SUR** e interprete los resultados etiquetando correctamente todas las gráficas.

Utilice el paquete Minitab para resolver esta Autoevaluación.

1.3.2.1 EJERCICIOS DE REFUERZO. Prueba de Normalidad. Prueba de Anderson Darling.

.

EJERCICIOS DE **REFUERZO 1.3.2.1 P**RUEBA **DE NORMALIDAD. PRUEBA DE ANDERSON DARLING.**

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

1.3.2.1.1 Se midió el **VALOR** de las casas en la zona **NORTE** en la base de datos "**BIENES1"** que se construyó con anterioridad.

Utilizando el paquete **Minitab** se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas en la zona **NORTE** e interprete los resultados y etiquetando correctamente todas las gráficas.

1.3.2.1.2 Se midió el **VALOR** de las casas con estilo **MODERNO** en la base de datos "**BIENES1"** que se construyó con anterioridad.

Utilizando el paquete **Minitab** se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas con estilo **MODERNO** e interprete los resultados y etiquetando correctamente todas las gráficas.

1.3.2.1.3 Se midió el **VALOR** de las casas con estilo **GRANJA** en la base de datos "**BIENES1"** que se construyó con anterioridad.

Utilizando el paquete **Minitab** se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas con estilo **GRANJA** e interprete los resultados y etiquetando correctamente todas las gráficas.

1.3.2.2 P.H.

EJEMPLO ILUSTRATIVO EN MINITAB 17. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos

EJEMPLO **ILUSTRATIVO I**NTEGRAL EN **MINITAB 17. 1.3.2.2. I**GUALDAD DE VARIANZAS. **(HOMOCEDASTICIDAD). PRUEBA DE LEVENE. TRANSFORMA-CIONES. ETIQUETADO DE GRÁFICOS**

Revisión general del análisis de la varianza:

El análisis de la varianza (ANOVA) es similar a la regresión en el sentido de que se utiliza para investigar y modelar la relación entre una variable de respuesta y una o más variables predictoras. Sin embargo, el análisis de la varianza se diferencia de la regresión, en dos aspectos: las variables predictoras son cualitativas (categóricas), y no se parte de ninguna suposición acerca de la naturaleza de la relación (es decir, el modelo no incluye coeficientes para las variables). En efecto, el análisis de la varianza amplia la prueba t de dos muestras para probar la igualdad de las medias de dos poblaciones de modo que la convierta en una hipótesis nula en la que se comparan más de dos medias y determinar si, o bien todas son iguales o no todas son iguales. Sin embargo, varios de los procedimientos de ANOVA de Minitab, permiten modelos en los que se incluyen tanto variables cualitativas, como cuantitativas.

Entre las capacidades para ANOVA que puede realizar Minitab figuran procedimientos para ajustar modelos ANOVA a datos recolectados en diseños de diversa naturaleza, para ajustar modelos MANOVA a diseños con respuestas múltiples, para ajustar modelos ANOM (análisis de medias), y gráficas para probar la igualdad de las varianzas, para gráficas de intervalos de confianza y gráficas de efectos principales e interacciones.

Se utilizará el paquete Minitab para resolver este Ejemplo Ilustrativo de Igualdad de Varianzar, realizar las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal y etiquetar debidamente todos los gráficos de presentación.

Prueba de igualdad de varianzas

Estadísticas > ANOVA > Prueba de igualdad de varianzas.

Utilice la prueba de varianzas para realizar **pruebas de hipótesis** para detectar la igualdad u homogeneidad de la **varianza** utilizando las **pruebas de Bartlett y de Levene**. Una P**rueba F** reemplaza a la prueba de Bartlett cuando usted sólo tiene dos niveles.

Muchos procedimientos estadísticos, incluyendo los **análisis de varianza**, parten del supuesto de que aunque es posible que haya diferentes muestras

provenientes de **poblaciones** con medias diferentes, éstas tienen la misma varianza. El efecto que ejerce la desigualdad de las varianzas en las inferencias depende en parte de si su modelo incluye **factores fijos o aleatorios**, disparidades en el tamaño de las muestras, y la elección del procedimiento de comparación múltiple. La prueba F de ANOVA es ligeramente afectada por la desigualdad de las varianzas si el modelo contiene sólo factores fijos y tiene tamaños de muestras iguales o casi iguales. Sin embargo, las pruebas F que incluyen efectos aleatorios pudieran ser sustancialmente afectadas. Utilice el procedimiento de prueba de varianza para probar la validez de la suposición de igualdad en las varianzas.

Elementos del cuadro de diálogo

Respuesta: Ingrese la columna que contiene la variable de respuesta.

Factores: Ingrese las columnas que contienen los **factores** en el modelo.

Nivel de confianza: Ingrese un valor de 0 a 100 como nivel de confianza deseado para los intervalos de confianza que se muestran en la gráfica. El nivel predeterminado es 95. Minitab utiliza el **método de Bonferroni** para calcular los **intervalos de confianza simultáneos.**

Título: Escriba en este cuadro el texto deseado para sustituir el título predeterminado por su propio título personalizado.

Comparación entre las pruebas de Bartlett y las pruebas de Levene

Minitab calcula y muestra una estadística de prueba y un **valor p** tanto para la prueba de Bartlett, como para la prueba de Levene donde, respectivamente, la hipótesis nula corresponde o bien a varianzas iguales, o a varianzas que en no todos los casos son iguales. Si sólo hay dos niveles, se realiza una prueba F en lugar de la prueba de Bartlett.

· Utilice la prueba de Bartlett cuando los datos provengan de distribuciones normales, la prueba de Bartlett no es sólida cuando los datos se apartan de la normalidad.

· Utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente distribuciones normales. Este método considera las distancias de las observaciones con respecto a la mediana de la muestra en lugar de la media de la muestra, esto hace que la prueba sea más sólida para las muestras más pequeñas.

Existen dos pruebas de homocedasticidad: la prueba de **Bonett** y la prueba de **Levene**. Para cada prueba, la hipótesis nula plantea que las dos varianzas son iguales (H_0: $s^2_1 / s^2_2 = 1$). La hipótesis alternativa puede ser de cola izquierda (H_1: $s^2_1 / s^2_2 < 1$), de cola derecha

PRUEBA DE LEVENE

Ejemplo de realización de prueba de igualdad de varianzas suponiendo que los datos provengan de distribuciones continuas, pero no necesariamente normales (Prueba de **Levene**) con el paquete Mininat 17**:**

(H_1: $s^2_1 / s^2_2 > 1$), o de dos colas (H_1: $s^2_1 / s^2_2 \neq 1$). De forma opcional, las relaciones de prueba diferentes de 1 (igualdad) se pueden especificar.

Usted pretende diseñar un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad en cada una de las casas con **ESTILO** diferente. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos. **Sin embargo, aquí se seleccionan dos métodos para demostrar las capacidades de Minitab para probar primeramente que se cumpla el supuesto de igualdad de varianzas (Homocedasticidad).**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**)**:**

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***VALOR***

4 En **Factores**, ingrese ***ESTILO.*** Haga clic en **Aceptar**.

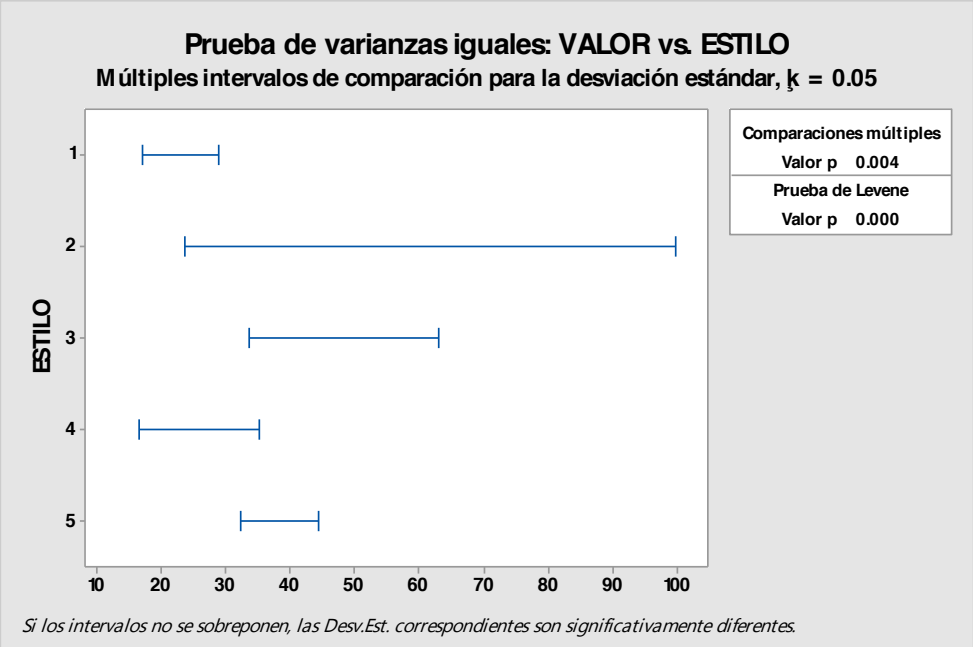
Salida de la ventana Sesión

Prueba de varianzas iguales: VALOR vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05


Intervalos de confianza de Bonferroni de 95% para desviaciones
estándar

ESTILO   N  Desv.Est.          IC
     1  97    21.7286  (17.7196,  27.371)
     2  15    43.7908  (17.5418, 131.983)
     3  23    43.1440  (29.0550,  72.145)
     4  60    23.4876  (15.1276,  38.104)
     5  38    36.4499  (28.1542,  50.621)

Nivel de confianza individual = 99%
```

```
Pruebas

                            Estadística
Método                        de prueba  Valor p
Comparaciones múltiples                —    0.004
Levene                              5.48    0.000
```

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de las pruebas **Levene** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **VALOR** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**,

así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la **inversa** (es decir, 1/Y o **1/X**). Las distribuciones asimétricas, pueden ser transformadas empleando la **raíz cuadrada, logaritmos o incluso la inversa** de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la **raíz cuadrada**, mientras que por lo regular, el l**ogaritmo** funciona mejor para la asimetría positiva

En estadística, la transformación de datos se efectúa para asegurarse que estos tienen una distribución normal (lo que puede ser un remedio para analizar datos que tienen otros tipos de distribución, falta de normalidad, linealidad, y homocedasticidad). Cuando se usan cierto tipo de pruebas estadísticas la transformación de datos puede ser necesaria ya que este tipo de análisis asumen que los datos deben ser lineales, normales y tienen homocedasticidad en su varianza. Si el tamaño de la muestra es demasiado pequeño o los datos están sesgados hacia una distribución no normal se puede intentar hacer una de las siguientes transformaciones: logarítmica, raíz cuadrada o inversa.

EJEMPLO DE TRANSFORMACIONES EN MINITAB 17 Y PRUEBA DE LEVENE

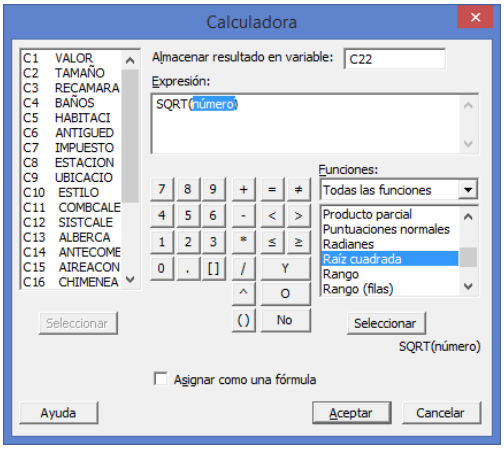
Transformaciónes Raiz Cuadrada, 1/X y Log de la variable VALOR con Minitab:

1 Abra la base de datos que creó **"BIENES1"**

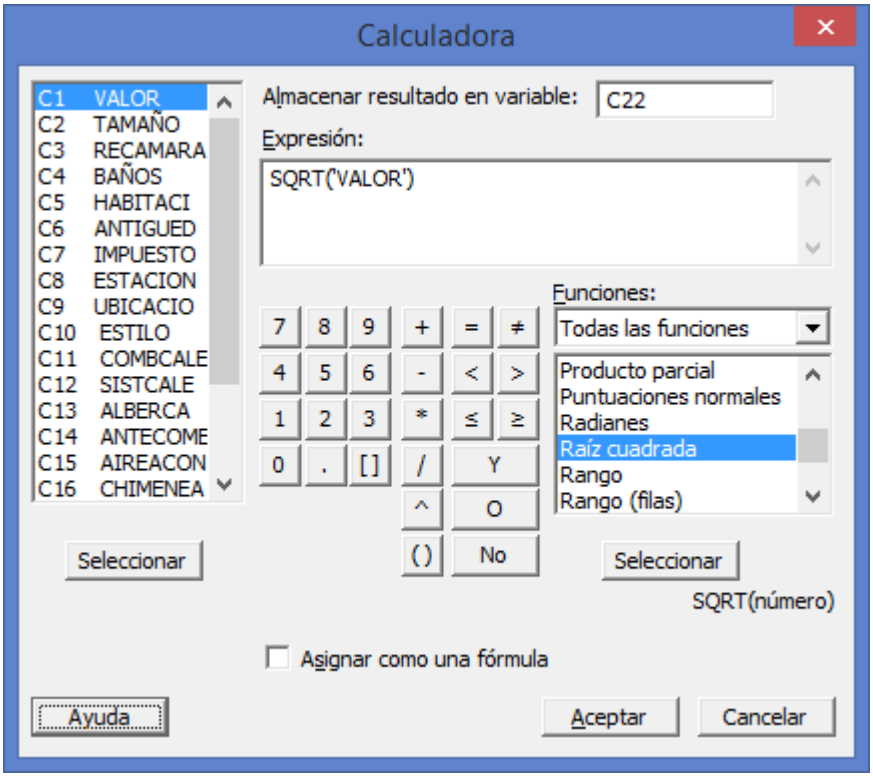
2 Elija **Calc > Calculadora**.

3 En **Almacenar resultado en variable**, ingrese ***C22***

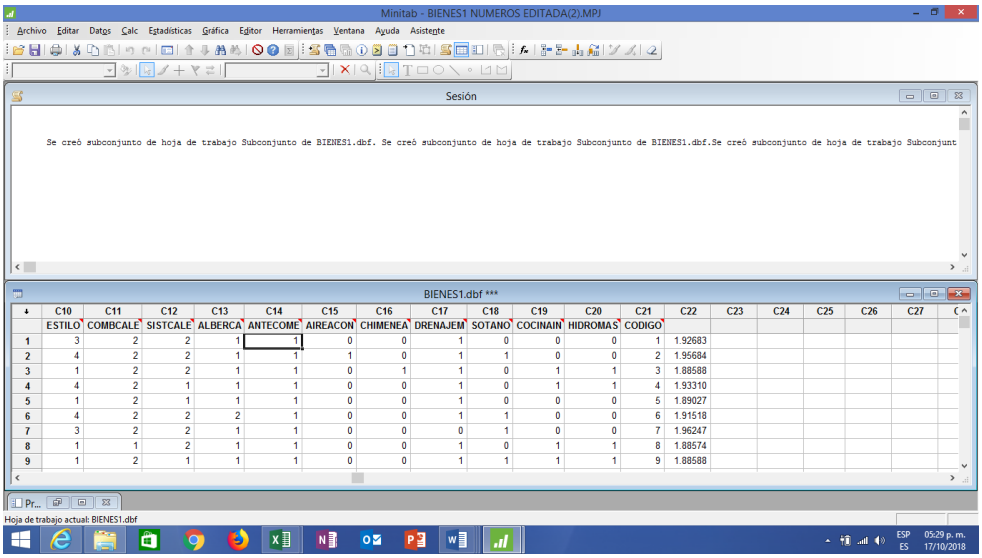
4 En **Expresión**, busque en **Funciones**: Raíz Cuadrada y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



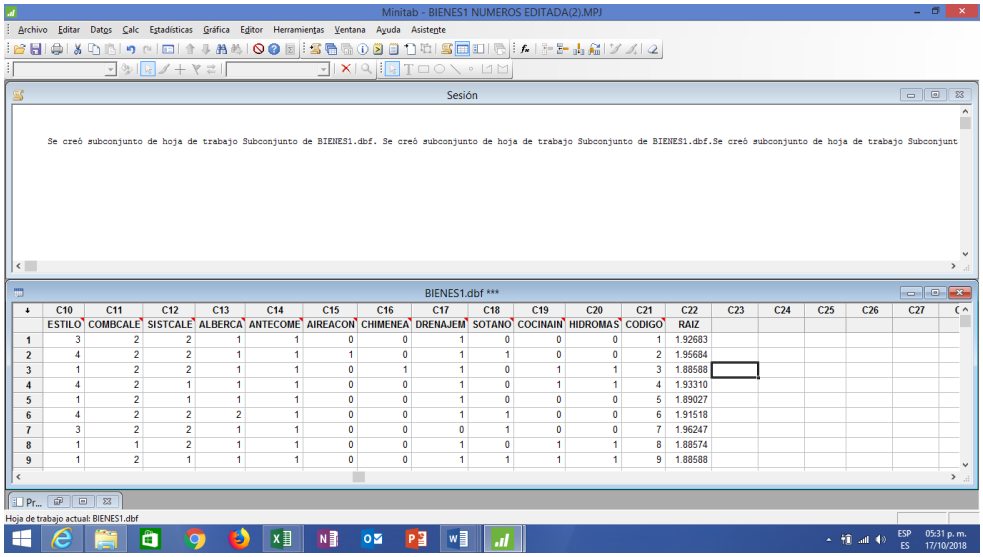
5 En **Expresión** aparece **SQRT(número).** Seleccione la variable **VALOR** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.



6 Haga clic en **Aceptar**.



7 En C22 teclee el nombre **RAÍZ**

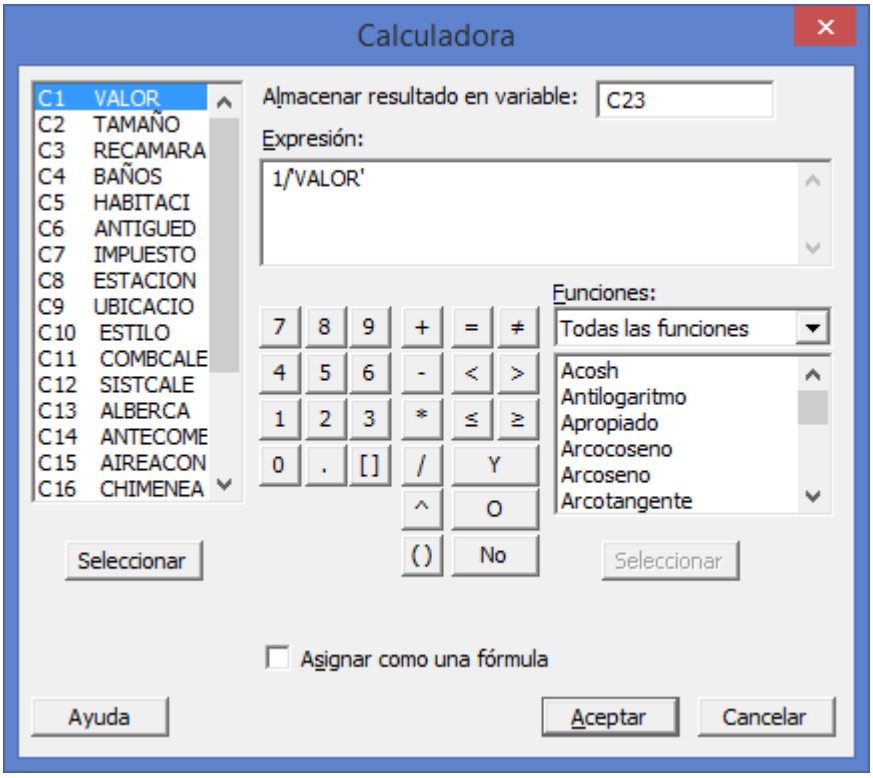


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAIZ					
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683					
2	4	2	2	1	1	1	0	1	1	0	0	2	1.95684					
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588					
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310					
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027					
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518					
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247					
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574					
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588					

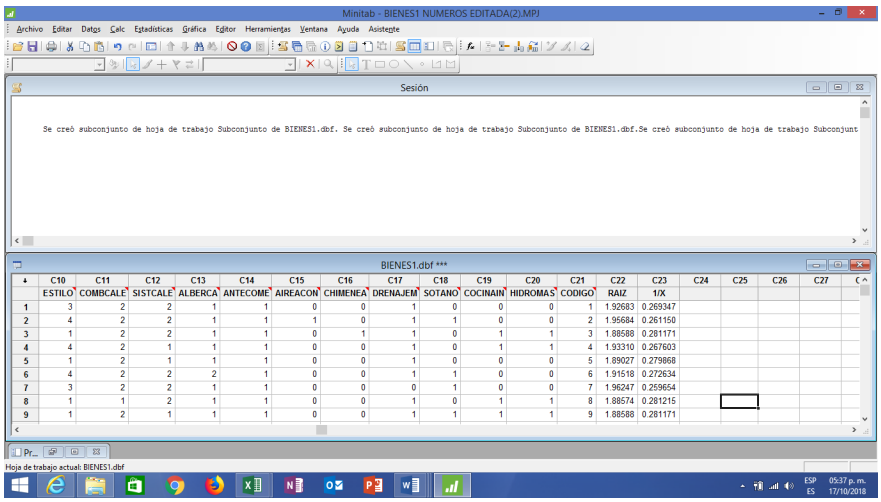
8 Elija **Calc > Calculadora**.

9 En **Almacenar resultado en variable**, ingrese ***C23***

10 En **Expresión**, para obtener la Inversa, **Teclee 1/** y seleccione la variable **VALOR** del lado izquierdo o doble clic.



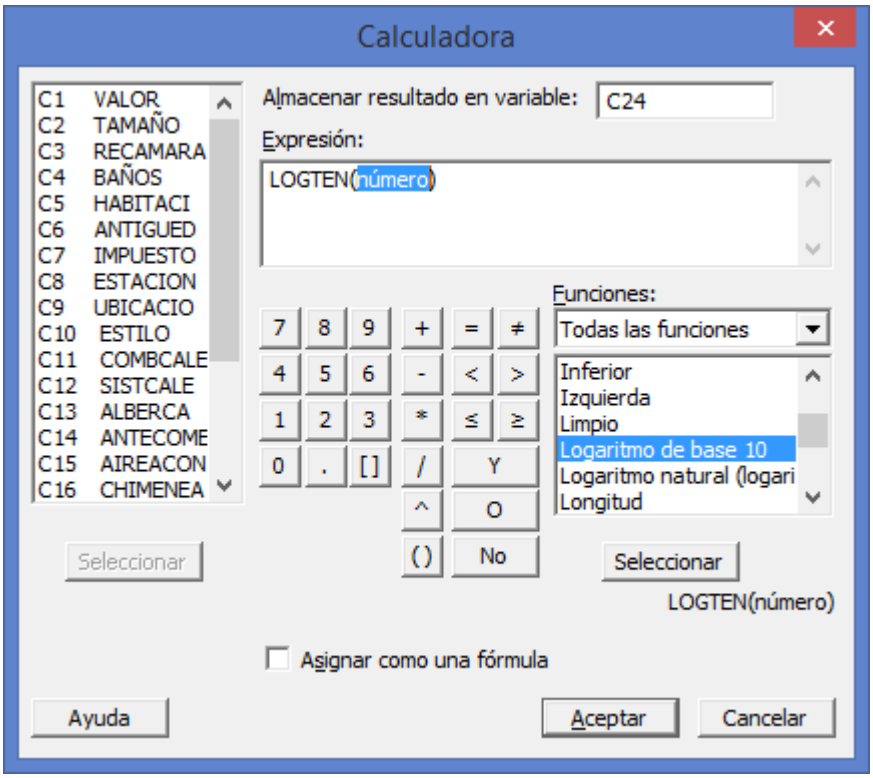
11 Haga clic en **Aceptar**. En **C23** escriba el nombre **1/X**



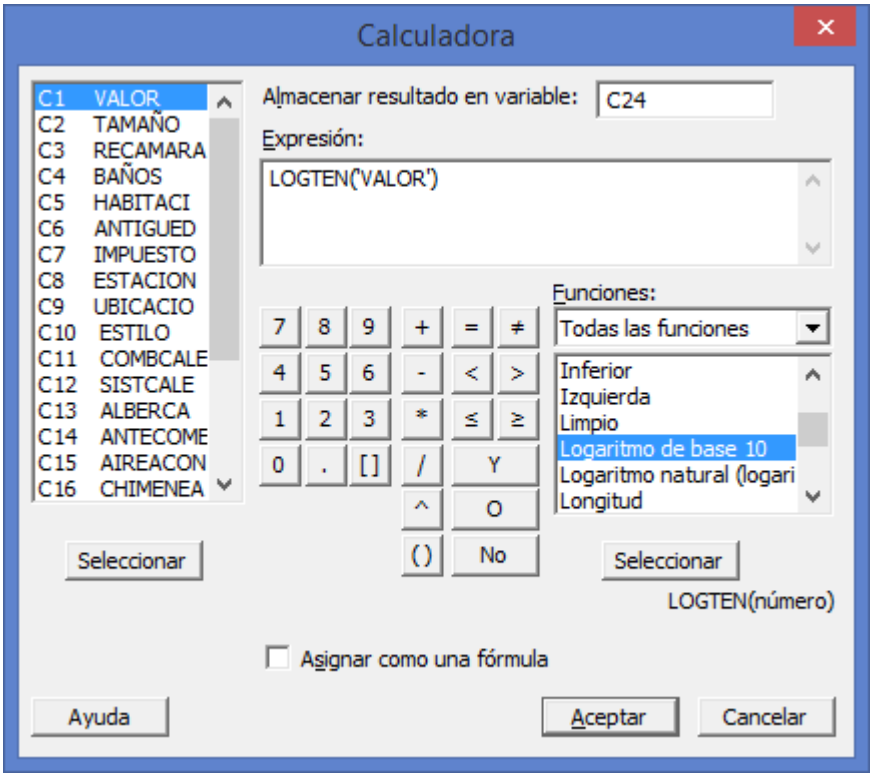
12 Elija **Calc > Calculadora**.

13 En **Almacenar resultado en variable**, ingrese ***C24***

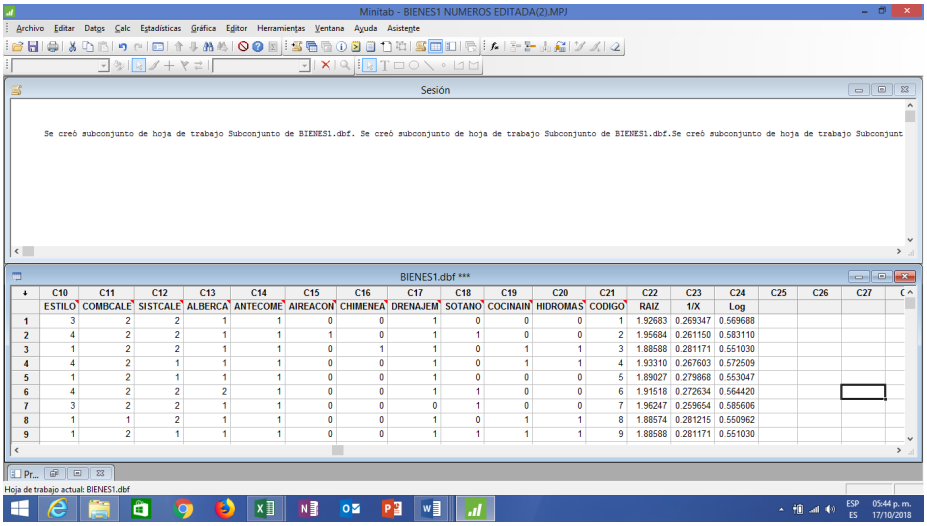
14 En **Expresión**, busque en **Funciones**: **Logaritmo base 10** y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



15 En **Expresión** aparece **LOGTEN(número).** Seleccione la variable **VALOR** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.



16 Haga clic en **Aceptar.** En **C24** escriba el nombre **Log**



Repita la prueba de Levene para las TRES transformaciones para ver si

se corrigió el problema de homocedasticidad y si fué así cual transformación resulto mejor.

A. CON LA RAÍZ CUADRADA DE VALOR

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***RAÍZ***

3 En **Factores**, ingrese ***ESTILO***. Haga clic en **Aceptar**.

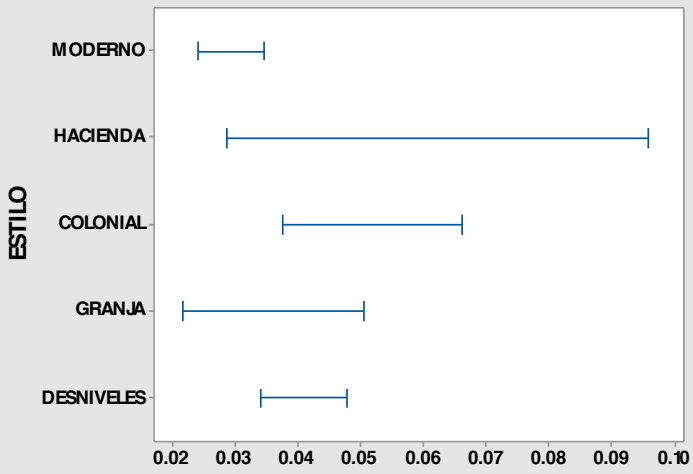
Salida de la ventana Sesión

Prueba de varianzas iguales: RAIZ vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05


Intervalos de confianza de Bonferroni de 95% para desviaciones
estándar

ESTILO   N  Desv.Est.            IC
     1  97  0.0283202  (0.0237186, 0.034737)
     2  15  0.0472208  (0.0204493, 0.131647)
     3  23  0.0467954  (0.0333582, 0.073924)
     4  60  0.0323497  (0.0202691, 0.053947)
     5  38  0.0388919  (0.0309284, 0.052462)

Nivel de confianza individual = 99%


Pruebas

                         Estadística
Método                     de prueba  Valor p
Comparaciones múltiples            —    0.008
Levene                          3.08    0.017
```

Prueba de varianzas iguales: RAIZ vs. ESTILO

Salida de la ventana Gráfica



Si los intervalos no se sobreponen, las Desv.Est. correspondientes son significativamente diferentes.
Elaboró: Arq. y M. en Admón. Javier Bech Vertti

Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **RAIZ CUADRADA** del **VALOR** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.017 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

B. CON LA LA INVERSA DE VALOR (1/X)

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***1/X***

3 En **Factores**, ingrese ***ESTILO***. Haga clic en **Aceptar**.

Salida de la ventana Sesión

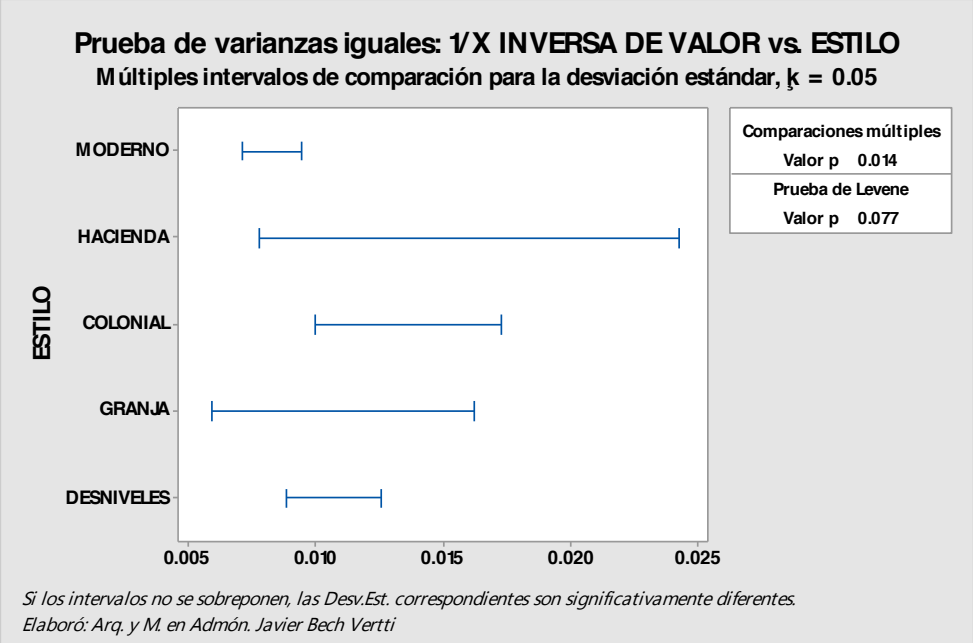
Prueba de varianzas iguales: 1/X vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05
```

```
Intervalos de confianza de Bonferroni de 95% para desviaciones
estándar

ESTILO   N  Desv.Est.            IC
     1  97  0.0080457  (0.0068001, 0.0097793)
     2  15  0.0124099  (0.0055708, 0.0333766)
     3  23  0.0123215  (0.0089739, 0.0190515)
     4  60  0.0095517  (0.0057398, 0.0166080)
     5  38  0.0101533  (0.0081479, 0.0135723)

Nivel de confianza individual = 99%


Pruebas

                          Estadística
Método                      de prueba  Valor p
Comparaciones múltiples             —    0.014
Levene                           2.14    0.077
```

Prueba de varianzas iguales: 1/X vs. ESTILO

Salida de la ventana Gráfica



Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **INVERSA** del **VALOR** de las

casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.077 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

C. CON EL LOGARITMO BASE 10 DE VALOR

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***Log***

3 En **Factores**, ingrese ***ESTILO.*** Haga clic en **Aceptar**.

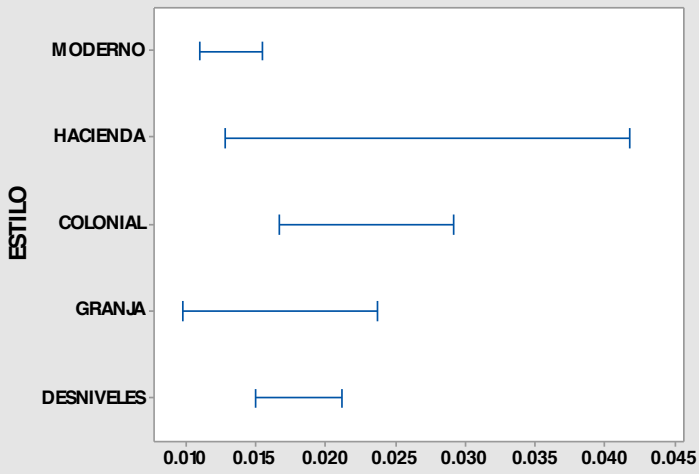
Salida de la ventana Sesión

Prueba de varianzas iguales: Log vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05


Intervalos de confianza de Bonferroni de 95% para desviaciones
estándar

ESTILO   N  Desv.Est.            IC
     1  97  0.0128313  (0.0107809, 0.0156884)
     2  15  0.0208416  (0.0091331, 0.0574205)
     3  23  0.0206678  (0.0148427, 0.0324086)
     4  60  0.0148313  (0.0091792, 0.0250387)
     5  38  0.0171300  (0.0136672, 0.0230313)

Nivel de confianza individual = 99%


Pruebas

                         Estadística
Método                     de prueba  Valor p
Comparaciones múltiples            —    0.010
Levene                          2.75    0.029
```

Prueba de varianzas iguales: Log vs. ESTILO

Salida de la ventana Gráfica



Si los intervalos no se sobreponen, las Desv.Est. correspondientes son significativamente diferentes.
Elaboró: Arq. y M. en Arq. Javier Bech Vertti

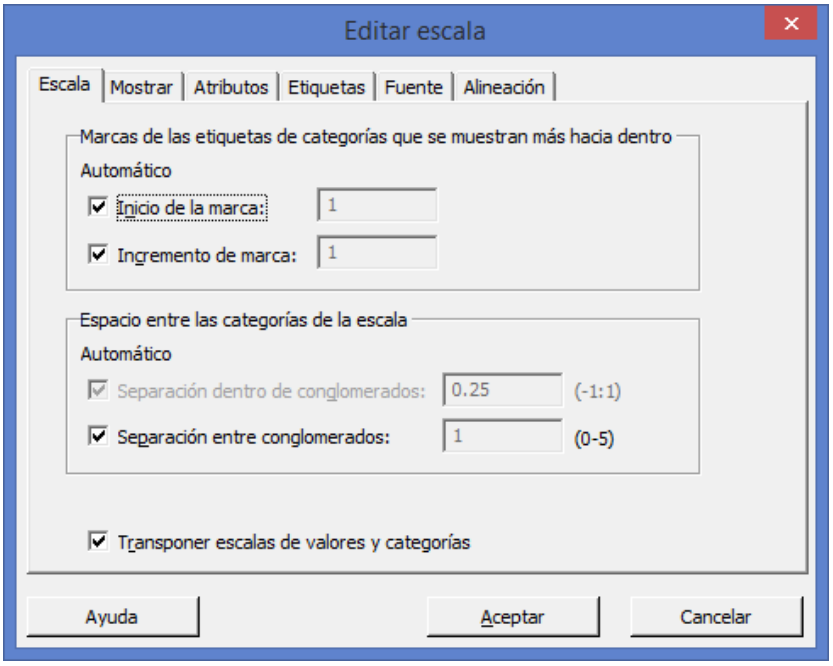
Interpretación de los resultados

Para el ejemplo para evaluar la transformación deL **LOGARITMO DECIMAL** del **VALOR** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.029 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

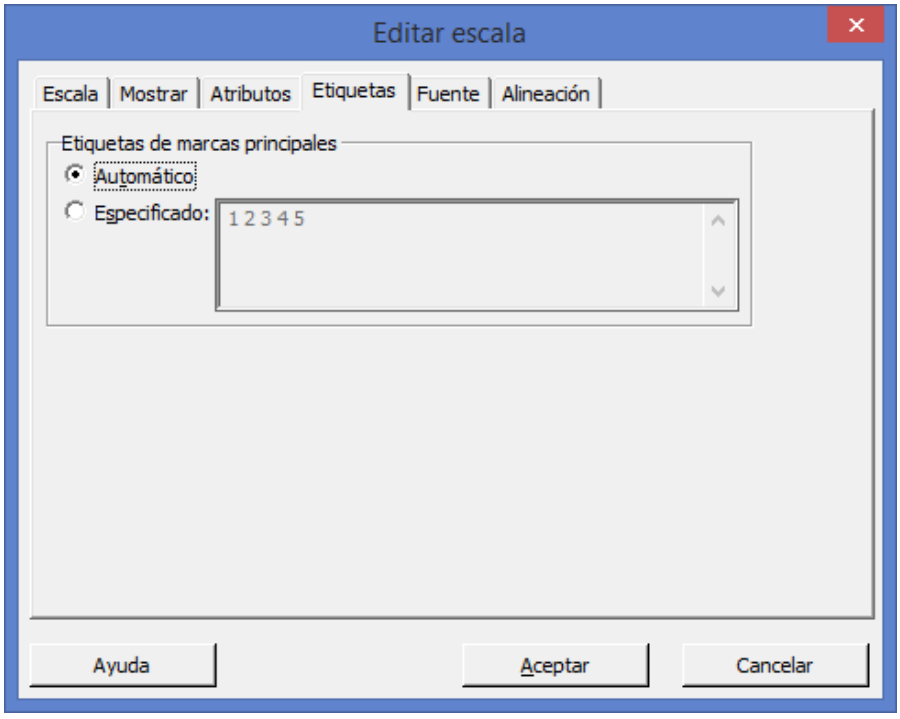
Nota Importante: Como puede observarse con la INVERSA de VALOR se logra corregir la falta de Homocedasticidad. No es del alcance de este trabajo tomar la decisión de si se debe o no trabajar con los datos transformados o sin transformar para usar alguna técnica multivariada sino solamente mostrar la forma de hacer dichas transformaciones.

EJEMPLO DE ETIQUETADO DE LOS NIVELES DE LA VARIABLE O "FACTOR" ESTILO:

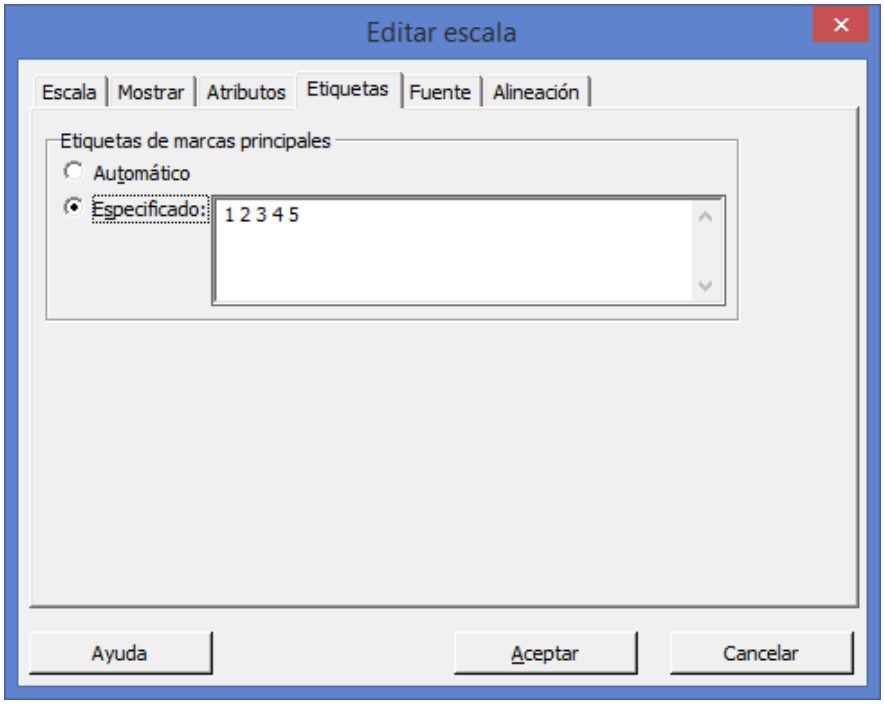
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **ESTILO** para que Minitab despliegue la ventana ***Editar Escala.***



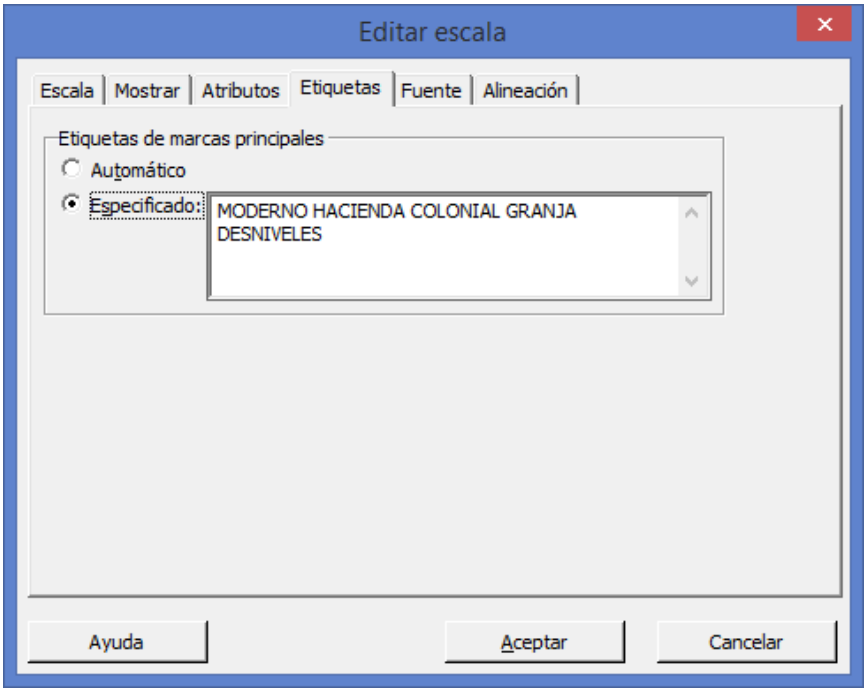
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

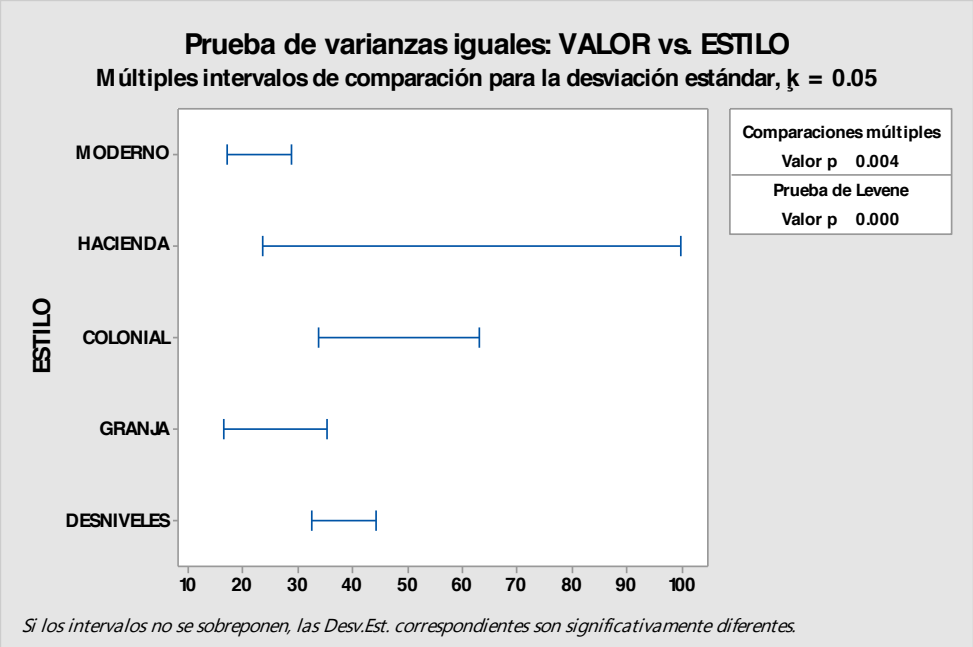


4 Borrar los números 1 2 3 4 5 y escribir **MODERNO HACIENDA COLONIAL GRANJA DESNIVELES**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



Con lo anterior queda terminado el ejemplo ilustrativo de **Igualdad de varianzas (Prueba de homocedasticidad). Prueba de Levene** en el paquete **MINITAB17**. Para salir del sistema selecciona con el puntero del mouse en el menú **Archivo** de la barra la opción **Salir** y da un clic con el botón izquierdo del mouse. El sistema te mandará al escritorio de Windows en la pantalla. **Aplica el procedimiento para salir de Windows**

1.3.2.2 ACTIVIDAD DE **APRENDIZAJE EN MINITAB 17. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos.**

ACTIVIDAD DE APRENDIZAJE EN

Usted pretende diseñar un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad en cada una de las casas con

MINITAB 17
1.3.2.2
PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE LEVENE. TRANSFORMA-CIONES. ETIQUETADO DE GRÁFICOS.

UBICACIÓN diferente. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos. **Sin embargo, aquí se seleccionan dos métodos para demostrar las capacidades de Minitab para probar primeramente que se cumpla el supuesto de igualdad de varianzas (Homocedasticidad).**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**)**:**

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

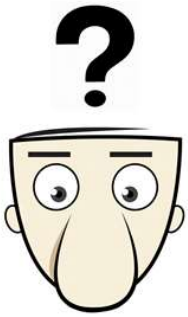
1.3.2.2 EJERCICIO DE AUTOEVALUACIÓN. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos.

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN
1.3.2.2
PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE

Se pretende diseñar un experimento para evaluar el valor de 233 casas de diferentes estilos. Se mide el **VALOR** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad con base en **si tienen o no CHIMENEA**. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de

LEVENE. TRANSFORMA-CIONES. ETIQUETADO DE GRÁFICOS.



comparación múltiple apropiado para sus datos.

Antes de realizar el **análisis de varianza,** se verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**)**:**

Utilice el paquete Minitab para resolver este Ejercicio de Autoevaluación de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

1.3.2.2 EJERCICIOS DE **REFUERZO. Prueba de Homocedasticidad. Prueba de Levene. Transformaciones. Etiquetado de Gráficos.**

EJERCICIOS DE **REFUERZO 1.3.2.2 P**RUEBA **DE HOMOCEDASTICIDAD. PRUEBA DE LEVENE. TRANSFORMA-CIONES. ETIQUETADO DE GRÁFICOS.**

1.3.2.2.1 Se pretende diseñar un experimento para evaluar el **IMPUESTO** que pagan 233 casas en diferentes zonas de la ciudad. Se mide el **IMPUESTO** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad en cada una de las casas con **ESTILO** diferente. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de comparación múltiple apropiado para los datos.

Antes de realizar el **análisis de varianza,** verifica la suposición de que la varianza es igual en todos los niveles del Factor **ESTILO** utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**) utilizando el paquete de cómputo **Minitab:**

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

presentación.

1.3.2.2.2 Se pretende diseñar un experimento para evaluar el **IMPUESTO** que pagan 233 casas en diferentes zonas de la ciudad. Se mide el **IMPUESTO** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad con respecto a su **UBICACIÓN** diferente. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de comparación múltiple apropiado para los datos.

Antes de realizar el **análisis de varianza,** verifica la suposición de que la varianza es igual en todos los niveles del Factor **UBICACIÓN** utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**) utilizando el paquete de cómputo **Minitab:**

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

1.3.2.2.3 Se pretende diseñar un experimento para evaluar el **IMPUESTO** que pagan 233 casas en diferentes zonas de la ciudad. Se mide el **IMPUESTO** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad con respecto a si tienen o no **CHIMENEA**. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de comparación múltiple apropiado para los datos.

Antes de realizar el **análisis de varianza,** verifica la suposición de que la varianza es igual en todos los niveles del Factor **CHIMENEA** utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**) utilizando el paquete de cómputo **Minitab:**

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

1.3.2.3.P.H. EJEMPLO ILUSTRATIVO EN MINITAB 17. Prueba de Homocedasticidad. Prueba de Barlett

EJEMPLO ILUSTRATIVO 1.3.2.3. INTEGRAL EN MINITAB 17. IGUALDAD DE VARIANZAS (HOMOCEDASTICIDAD). PRUEBA DE BARLETT

Revisión general del análisis de la varianza:

El análisis de la varianza (ANOVA) es similar a la regresión en el sentido de que se utiliza para investigar y modelar la relación entre una variable de respuesta y una o más variables predictoras. Sin embargo, el análisis de la varianza se diferencia de la regresión, en dos aspectos: las variables predictoras son cualitativas (categóricas), y no se parte de ninguna suposición acerca de la naturaleza de la relación (es decir, el modelo no incluye coeficientes para las variables). En efecto, el análisis de la varianza amplia la prueba t de dos muestras para probar la igualdad de las medias de dos poblaciones de modo que la convierta en una hipótesis nula en la que se comparan más de dos medias y determinar si, o bien todas son iguales o no todas son iguales. Sin embargo, varios de los procedimientos de ANOVA de Minitab, permiten modelos en los que se incluyen tanto variables cualitativas, como cuantitativas.

Entre las capacidades para ANOVA que puede realizar Minitab figuran procedimientos para ajustar modelos ANOVA a datos recolectados en diseños de diversa naturaleza, para ajustar modelos MANOVA a diseños con respuestas múltiples, para ajustar modelos ANOM (análisis de medias), y gráficas para probar la igualdad de las varianzas, para gráficas de intervalos de confianza y gráficas de efectos principales e interacciones.

Prueba de igualdad de varianzas

Estadísticas > ANOVA > Prueba de igualdad de varianzas.

Utilice la prueba de varianzas para realizar **pruebas de hipótesis** para detectar la igualdad u homogeneidad de la **varianza** utilizando las **pruebas de Bartlett y de Levene**. Una P**rueba F** reemplaza a la prueba de Bartlett cuando usted sólo tiene dos niveles.

Muchos procedimientos estadísticos, incluyendo los **análisis de varianza**, parten del supuesto de que aunque es posible que haya diferentes muestras provenientes de **poblaciones** con medias diferentes, éstas tienen la misma varianza. El efecto que ejerce la desigualdad de las varianzas en las inferencias depende en parte de si su modelo incluye **factores fijos o aleatorios**, disparidades en el tamaño de las muestras, y la elección del procedimiento de comparación múltiple. La prueba F de ANOVA es ligeramente afectada por la

desigualdad de las varianzas si el modelo contiene sólo factores fijos y tiene tamaños de muestras iguales o casi iguales. Sin embargo, las pruebas F que incluyen efectos aleatorios pudieran ser sustancialmente afectadas. Utilice el procedimiento de prueba de varianza para probar la validez de la suposición de igualdad en las varianzas.

Elementos del cuadro de diálogo

Respuesta: Ingrese la columna que contiene la variable de respuesta.

Factores: Ingrese las columnas que contienen los **factores** en el modelo.

Nivel de confianza: Ingrese un valor de 0 a 100 como nivel de confianza deseado para los intervalos de confianza que se muestran en la gráfica. El nivel predeterminado es 95. Minitab utiliza el **método de Bonferroni** para calcular los **intervalos de confianza simultáneos.**

Título: Escriba en este cuadro el texto deseado para sustituir el título predeterminado por su propio título personalizado.

Comparación entre las pruebas de Bartlett y las pruebas de Levene

Minitab calcula y muestra una estadística de prueba y un **valor p** tanto para la prueba de Bartlett, como para la prueba de Levene donde, respectivamente, la hipótesis nula corresponde o bien a varianzas iguales, o a varianzas que en no todos los casos son iguales. Si sólo hay dos niveles, se realiza una prueba F en lugar de la prueba de Bartlett.

· Utilice la prueba de Bartlett cuando los datos provengan de distribuciones normales, la prueba de Bartlett no es sólida cuando los datos se apartan de la normalidad.

· Utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente distribuciones normales. Este método considera las distancias de las observaciones con respecto a la mediana de la muestra en lugar de la media de la muestra, esto hace que la prueba sea más sólida para las muestras más pequeñas.

En estadística, la **prueba de Bartlett** se utiliza para probar si *k* muestras provienen de poblaciones con la misma varianza. A las varianzas iguales a través de las muestras se llama homocedasticidad u homogeneidad de varianzas. Algunas pruebas estadísticas, por ejemplo, el análisis de la varianza ANOVA, suponen

EJEMPLO DE UNA PRUEBA DE BARLETT CON MINITAB 17

Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con **ESTILO** diferente. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

que las varianzas son iguales en todos los grupos o muestras. La prueba de Bartlett se puede utilizar para verificar esa suposición.

La prueba de Bartlett es sensible a las desviaciones de la normalidad. Es decir, si las muestras provienen de distribuciones no normales, entonces la prueba de Bartlett puede ser simplemente para probar la no normalidad. La Prueba de Levene y la de Bonett son alternativas a la prueba de Bartlett que son menos sensibles a las desviaciones de la normalidad.[1]

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas.

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***VALOR***

4 En **Factores**, ingrese ***ESTILO***

5 En el botón **Opciones** active la casilla que dice: ***Utilice la prueba basándose en la distribución normal.*** Haga clic en **Aceptar** y nuevamente **Aceptar**

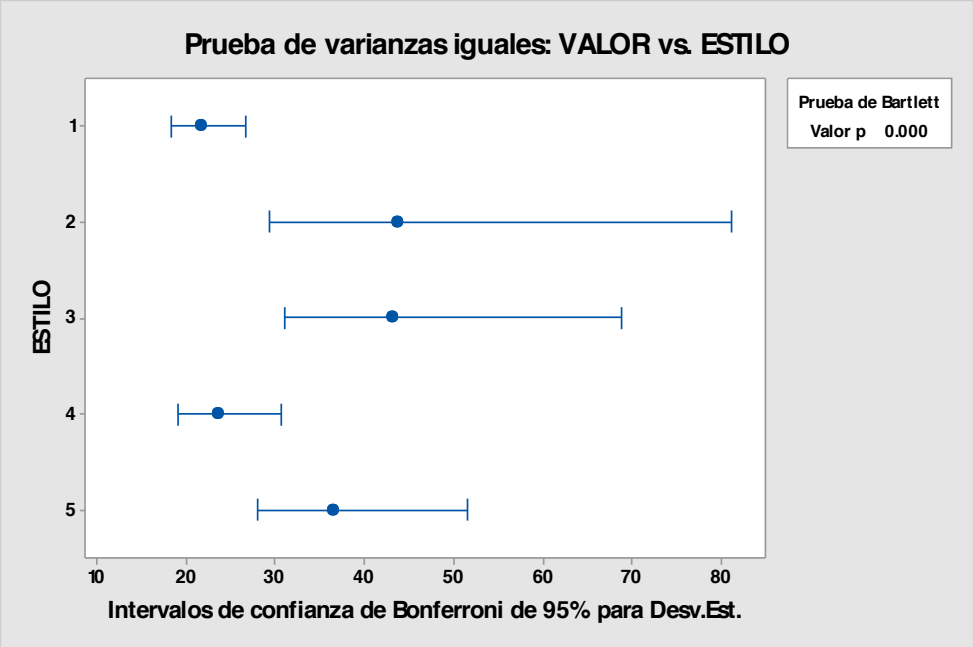
Salida de la ventana Sesión

Prueba de varianzas iguales: VALOR vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05

Se utiliza el método de Bartlett. Este método es exacto sólo para
datos
     normales.


Intervalos de confianza de Bonferroni de 95% para desviaciones
estándar

ESTILO   N  Desv.Est.          IC
     1  97    21.7286  (18.2938, 26.5988)
     2  15    43.7908  (29.2780, 81.1710)
     3  23    43.1440  (30.9337, 68.8346)
     4  60    23.4876  (18.9420, 30.5957)
     5  38    36.4499  (27.9595, 51.4287)

Nivel de confianza individual = 99%


Pruebas

          Estadística
Método      de prueba  Valor p
Bartlett        37.48    0.000
```

Prueba de varianzas iguales: VALOR vs. ESTILO

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de la prueba de **Barlett** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **VALOR** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

Los problemas de heterocedasticidad pueden solucionarse por medio de

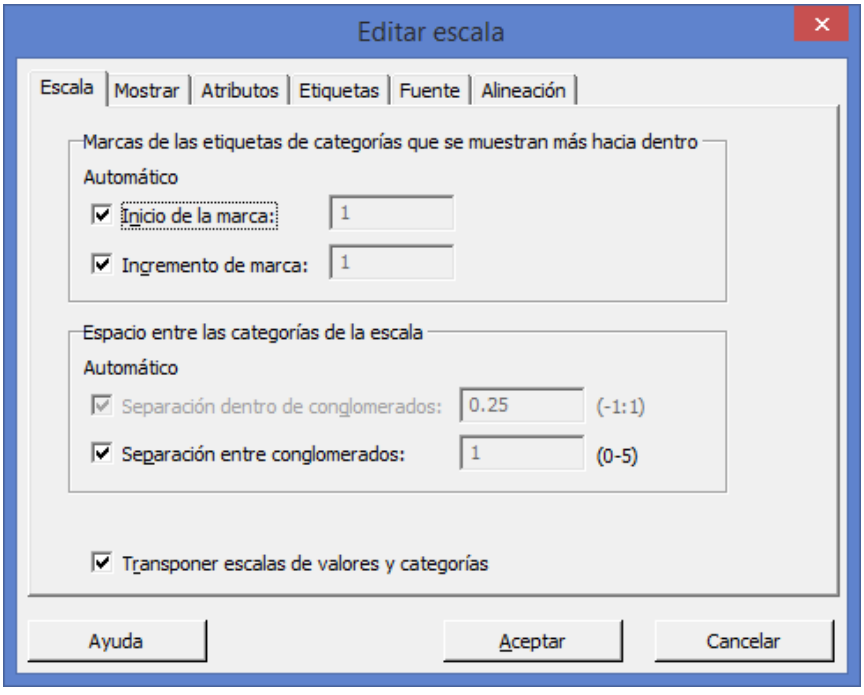
transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

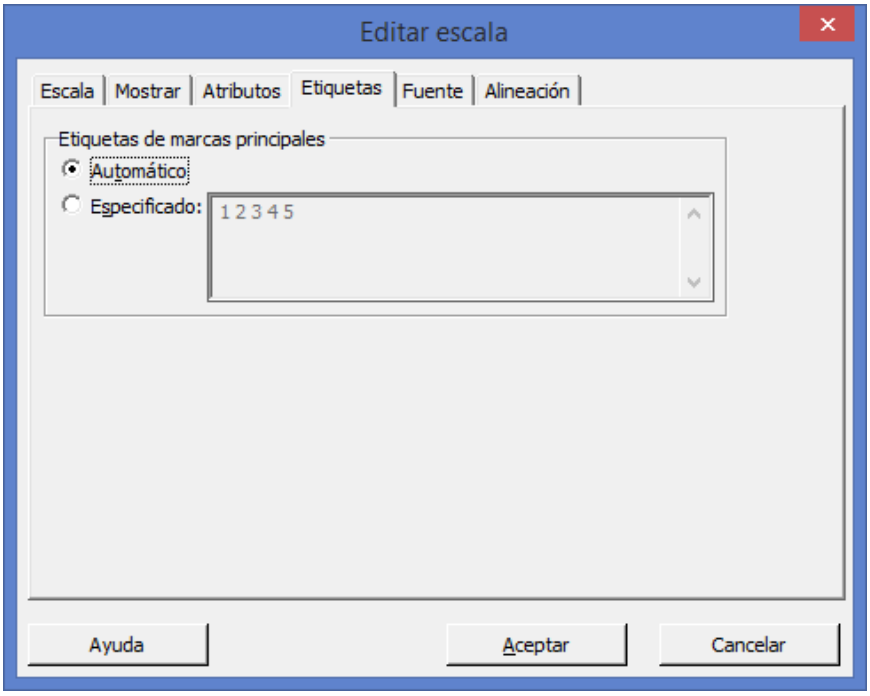
Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución y en la falta de homocedasticidad.

Etiquetado de los niveles de la variable o "Factor" ESTILO:

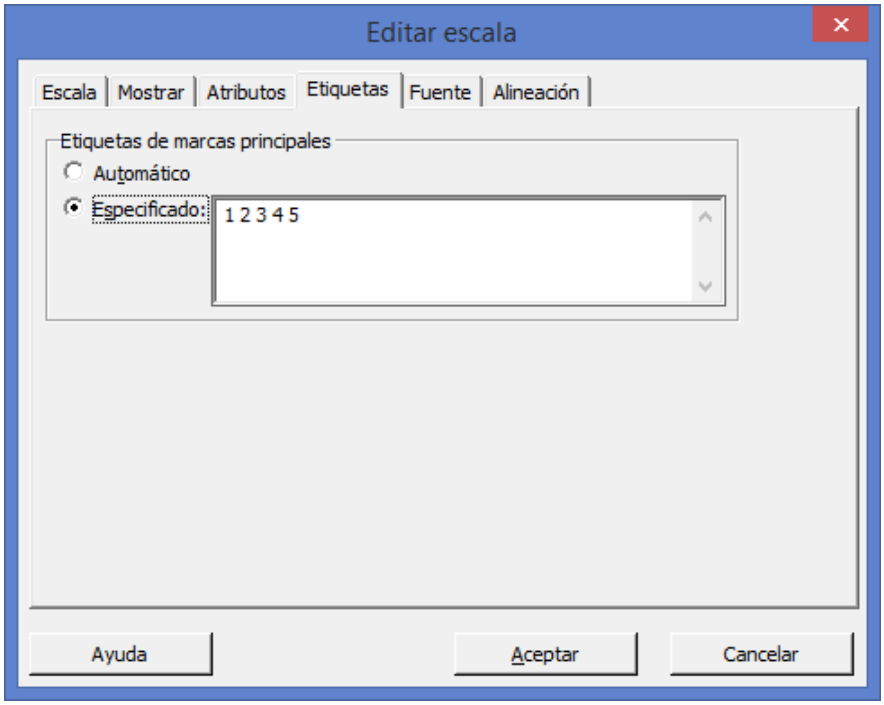
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **ESTILO** para que Minitab despliegue la ventana ***Editar Escala.***



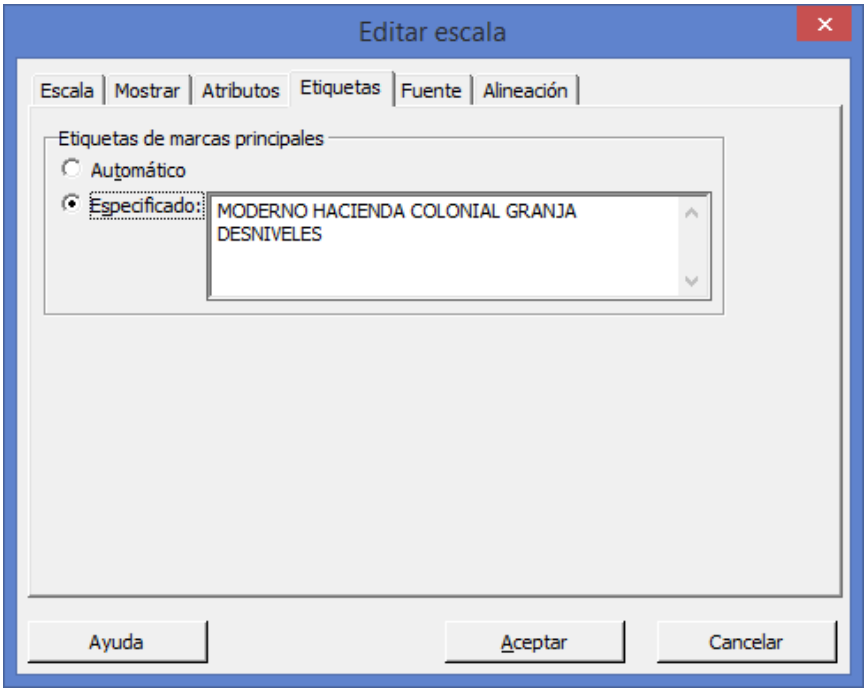
2 Seleccionar la pestaña **Etiquetas**



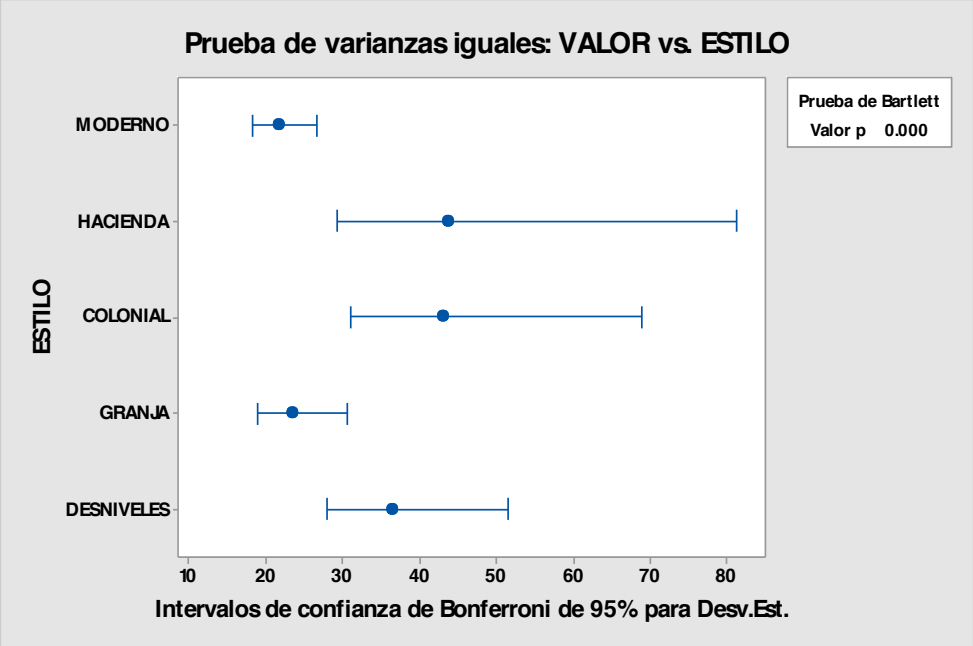
3 Activar la casilla **Especificado**



4 Borrar los números 1 2 3 4 5 y escribir **MODERNO HACIENDA COLONIAL GRANJA DESNIVELES**



5 Haga clic en **Aceptar**.
6 Aparece la gráfica debidamente etiquetada.



Con lo anterior queda terminado el ejemplo ilustrativo de **Igualdad de varianzas (Prueba de homocedasticidad) Prueba de Barlett** en el paquete **MINITAB17**. Para salir del sistema selecciona con el puntero del mouse en el menú **Archivo** de la barra la opción **Salir** y da un clic con el botón izquierdo del mouse. El sistema te mandará al escritorio de Windows en la pantalla. **Aplica el procedimiento para salir de Windows**

1.3.2.3 ACTIVIDAD DE **APRENDIZAJE. Prueba de Homocedasticidad. Prueba de Barlett**

ACTIVIDAD DE APRENDIZAJE 1.3.2.3 PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT

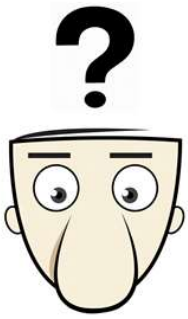
Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con base en la **UBICACIÓN** de las casas. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**.**

1.3.2.3 EJERCICIO **DE AUTOEVALUACIÓN. Prueba de Homocedasticidad. Prueba de Barlett.**

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 1.3.2.3 PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT



Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con respecto al tipo de **ALBERCA** ubicada en la propiedad. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

1.3.2.3 EJERCICIOS DE **REFUERZO. Prueba de Homocedasticidad. Prueba de Barlett.**

EJERCICIOS DE REFUERZO 1.3.2.3 PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT

1.3.2.3.1 Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **IMPUESTO** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con **ESTILO** diferente. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

1.3.2.3.2 Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **IMPUESTO** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con base en el **IMPUESTO** que pagan las casas. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

1.3.2.3.3 Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **IMPUESTO** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con respecto al tipo de **ALBERCA** ubicada en la propiedad. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

OBJETIVO 1.4. El alumno entenderá y aplicará un sistema de ecuaciones lineales representado como un producto de una matriz formada por los coeficientes de las incógnitas por un VECTOR COLUMNA, formado por las incognitas y este producto siendo igual a un VECTOR COLUMA formado por los términos independientes

ANTECEDENTES

CONCEPTOS DE:

Ecuación lineal, sistema de ecuaciones lineales,Matriz aumentada, representación matricial, operaciones elementales, ecuación algebraica, tamaño de una matriz, tipo de matrices, igualdad de matrices, operaciones con matrices, matriz de cofactores, matriz inversa, determinantes

1.4.1 SISTEMAS MATRICIALES DE RESOLUCION DE SISTEMAS DE ECUACIONES LINEALES

CONCEPTOS BÁSICOS SISTEMAS MATRICIALES DE RESOLUCIÓN DE SISTEMAS DE ECUACIONES

LINEALES

En matemáticas y álgebra lineal, un sistema de ecuaciones lineales, también conocido como sistema lineal de ecuaciones o simplemente sistema lineal, es un conjunto de ecuaciones lineales (es decir, un sistema de ecuaciones en donde cada ecuación es de primer grado), definidas sobre un cuerpo o un anillo conmutativo. Un ejemplo de sistema lineal de ecuaciones sería el siguiente:

2x+3y-4z=-4

x-y+3z=8

2x+5y-z=9

El problema consiste en encontrar los valores desconocidos de las variables x_1, x_2 y x_3 que satisfacen las tres ecuaciones.

SISTEMAS MATRICIALES DE RESOLUCIÓN DE SISTEMAS DE ECUACIONES LINEALES

Un sistema de **ecuaciones lineales** se puede representar como el producto de **una matriz** formada por los coeficientes de las incógnitas **por un VECTOR COLUMNA**, formado por las incógnitas, y este **producto** siendo **igual** a un **vector columna** formado por los términos independientes.

EJEMPLO: El sistema:

$$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_3z = d_2 \\ a_3x + b_3y + c_3z = d_3 \end{cases}$$

puede ser representado como:

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix}$$

$$A \times X = K$$

RESOLUCION DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE LA MATRIZ INVERSA

Para un sistema de ecuaciones lineales heterogéneas es válida la expresión:

A x X = K (**)

Donde:

A= Matriz de coeficientes

El problema de los sistemas lineales de ecuaciones es uno de los más antiguos de la matemática y tiene una infinidad de aplicaciones, como en procesamiento digital de señales, análisis estructural, estimación, predicción y más generalmente en programación lineal así como en la aproximación de problemas no lineales de análisis numérico.

X= Incógnitas

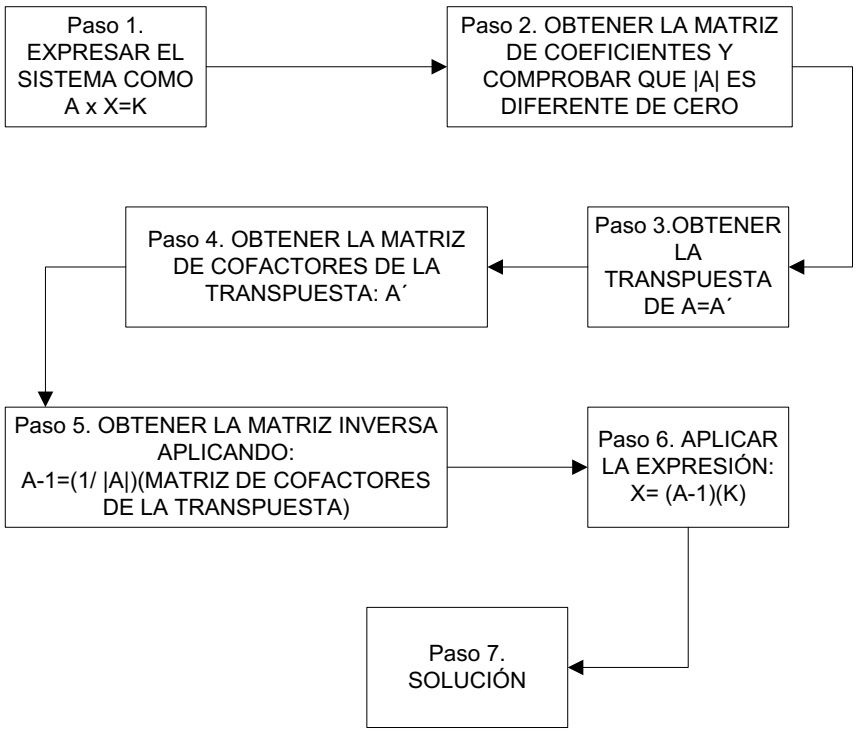
K= Términos independientes

Si ambos miembros de (**) los multiplicamos por A^{-1}, matriz inversa de A, obtenemos:

A^{-1} x (A x X)= A^{-1} x K o sea (A^{-1} x A) x X = A^{-1} x K y como, por definición, A^{-1} x A = A_u=1, se tiene:

$$1(X)= A^{-1}K, \text{ que es } X= A^{-1} \times K$$

Esquemáticamente el procedimiento es:



1.4.1.1 EJEMPLO ILUSTRATIVO. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.

EJEMPLO ILUSTRATIVO 1.4.1.1 RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ INVERSA

Resolver el siguiente sistema, empleando el método de la MATRIZ INVERSA:

$$\begin{aligned} 2x + 3y - 4z &= -4 \\ x - y + 3z &= 8 \\ 2x + 5y - z &= 9 \end{aligned}$$

Paso 1. Expresar el sistema como A x X = K

$$\underbrace{\begin{bmatrix} 2 & 3 & -4 \\ 1 & -1 & 3 \\ 2 & 5 & -1 \end{bmatrix}}_{A} \text{ x } \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} -4 \\ 8 \\ 9 \end{bmatrix}}_{K}$$

Paso 2. Buscar el valor del determinante $|A|$=

$$|A| = \begin{vmatrix} 2 & 3 & -4 \\ 1 & -1 & 3 \\ 2 & 5 & -1 \end{vmatrix} = +2 \begin{vmatrix} -1 & 3 \\ 5 & -1 \end{vmatrix} - 3 \begin{vmatrix} 1 & 3 \\ 2 & -1 \end{vmatrix} + (-4) \begin{vmatrix} 1 & -1 \\ 2 & 5 \end{vmatrix} =$$

Por el método de los menores correspondientes de una matriz:

$= 2(-1)^{1+1}(1-15) + 3(-1)^{1+2}(-1-6) - 4(-1)^{1+3}(5+2) = -28+21-28 = 35 \neq 0$ (tiene inversa)
Nota: Cuidar los signos de los cofactores + - +. Cofactor=expresión: $(-1)^{i+j}\ |\text{Mai}\mathbf{j}|$

Aplicando la regla de Sarrus:

$$|A| = \begin{vmatrix} 2 & 3 & -4 \\ 1 & -1 & 3 \\ 2 & 5 & -1 \\ 2 & 3 & -4 \\ 1 & -1 & 3 \end{vmatrix} = (2^{*}\text{-}1^{*}\text{-}1) + (1^{*}5^{*}\text{-}4) + (2^{*}3^{*}3) - (2^{*}\text{-}1^{*}\text{-}4) - (2^{*}5^{*}\text{-}3) - (1^{*}3^{*}\text{-}1) = -35 \neq 0$$

Paso 3. Obtener la matriz transpuesta de la matriz A, que es:

$$A^{T} = A' = \begin{bmatrix} 2 & 1 & 2 \\ 3 & -1 & 5 \\ -4 & 3 & -1 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A´, es decir formar α A´

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} -1 & 5 \\ 3 & -1 \end{vmatrix} & -\begin{vmatrix} 3 & 5 \\ -4 & -1 \end{vmatrix} & +\begin{vmatrix} 3 & -1 \\ -4 & 3 \end{vmatrix} \\ -\begin{vmatrix} 1 & 2 \\ 3 & -1 \end{vmatrix} & +\begin{vmatrix} 2 & 2 \\ -4 & -1 \end{vmatrix} & -\begin{vmatrix} 2 & 1 \\ -4 & 3 \end{vmatrix} \\ +\begin{vmatrix} 1 & 2 \\ -1 & 5 \end{vmatrix} & -\begin{vmatrix} 2 & 2 \\ 3 & 5 \end{vmatrix} & +\begin{vmatrix} 2 & 1 \\ 3 & -1 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(1-15) & -(-3-20) & +(9-4) \\ -(-1-6) & +(-2+8) & -(6+4) \\ +((5+2) & -(10-6) & +(-2-3) \end{bmatrix} = \begin{bmatrix} -14 & -17 & 5 \\ 7 & 6 & 10 \\ 7 & -4 & -5 \end{bmatrix}$$

Paso 5. Calcular la inversa de la matriz A´= $\frac{1}{|A|}$ α A´

$$\frac{1}{-35}\begin{bmatrix} -14 & -17 & 5 \\ 7 & 6 & -10 \\ 7 & -4 & -5 \end{bmatrix} = \begin{bmatrix} 14/35 & 17/35 & -5/35 \\ -7/35 & -6/35 & 10/35 \\ -7/35 & 4/35 & 5/35 \end{bmatrix}$$

Paso 6. Llevar este valor a la expresión $X=A^{-1}$ x K y tenemos:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 14/35 & 17/35 & -5/35 \\ -7/35 & -6/35 & 10/35 \\ -7/35 & 4/35 & 5/35 \end{bmatrix} x \begin{bmatrix} -4 \\ 8 \\ 9 \end{bmatrix} =$$

$$\begin{bmatrix} \left(\frac{14}{35} & \frac{17}{35} & -\frac{5}{35}\right) & * & \begin{bmatrix} -4 \\ 8 \\ 9 \end{bmatrix} \\ \left(-\frac{7}{35} & -\frac{6}{35} & \frac{10}{35}\right) & * & \begin{bmatrix} -4 \\ 8 \\ 9 \end{bmatrix} \\ \left(-\frac{7}{35} & \frac{4}{35} & \frac{5}{35}\right) & * & \begin{bmatrix} -4 \\ 8 \\ 9 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} -56/35 & +136/35 & -45/35 \\ 28/35 & -48/35 & +90/35 \\ 28/35 & +32/35 & +45/35 \end{bmatrix}$$

$$= \begin{bmatrix} 35/35 \\ 70/35 \\ 105/35 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Paso 7. La solución es x=1, y=2, z=3

1.4.1.1 ACTIVIDAD DE **APRENDIZAJE. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.**

ACTIVIDAD DE **APRENDIZAJE 1.4.1.1** **R**ESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**

Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$x + y + z = 11$$
$$2x - y + z = 5$$
$$3x + 2y + z = 24$$

1.4.1.1 EJERCICIO **DE AUTOEVALUACIÓN. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.**

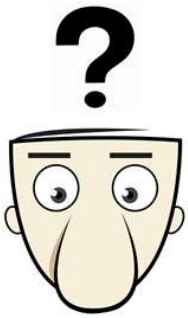
A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTO**EVALUACIÓN 1.4.1.1** **R**ESOLUCIÓN DE UN SISTEMA LINEAL DE

Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$4x + 2y + 3z = 4$$
$$-3x - y - 2z = -6$$

ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**



-2x – y – z = -5

1.4.1.1 EJERCICIOS DE **REFUERZO. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.**

EJERCICIOS DE REFUERZO 1.4.1.1 RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ INVERSA

1.4.1.1.1 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

3x – 5y + 7z =28
2x + 6y – 9z =-23
4x – 2y – 5z = 9

1.4.1.1.2 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

2x + 3y + z = 4
x + 2y + 2z = 6
5x + y + 4x = 21

1.4.1.1.3 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

2x + 5y – 3z = 17
6x – 2y -5z = -3
3x + 7y + 4z = -18

1.4.1.2 EJEMPLO ILUSTRATIVO EN EXCEL. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.

EJEMPLO ILUSTRATIVO INTEGRAL EN EXCEL 1.4.1.2 RESOLUCIÓN SISTEMA LINEAL ECUACIONES. MÉTODO MATRIZ INVERSA

Resolver el siguiente sistema, empleando el método de la MATRIZ INVERSA con Excel:

$$2x + 3y - 4z = -4$$
$$x - y + 3z = 8$$
$$2x + 5y - z = 9$$

Cómo resolver un sistema de ecuaciones con Excel

Uno de los cálculos que constituyen un verdadero reto para los estudiantes es la resolución de sistemas de ecuaciones.

Existen muchos tipos de **sistemas de ecuaciones**, pero hoy se va a **resolver con Excel** uno de los más sencillos: los sistemas de ecuaciones lineales:

$$2x + 3y - 4z = -4$$
$$x - y + 3z = 8$$
$$2x + 5y - z = 9$$

Para un sistema de ecuaciones lineales heterogéneas es válida la expresión:

$A \times X = K$ (**)

Donde:

A= Matriz de coeficientes
X= Incógnitas
K= Términos independientes
Si ambos miembros de (**) los multiplicamos por A^{-1}, matriz inversa de A, obtenemos:
$A^{-1} \times (A \times X) = A^{-1} \times K$ o sea $(A^{-1} \times A) \times X = A^{-1} \times K$ y como, por definición, $A^{-1} \times A = A_u = 1$, se tiene:

$$1(X) = A^{-1}K, \text{ que es } X = A^{-1} \times K$$

Para resolver un sistema de ecuaciones, debemos encontrar el valor de las variables que contiene:

X = 1

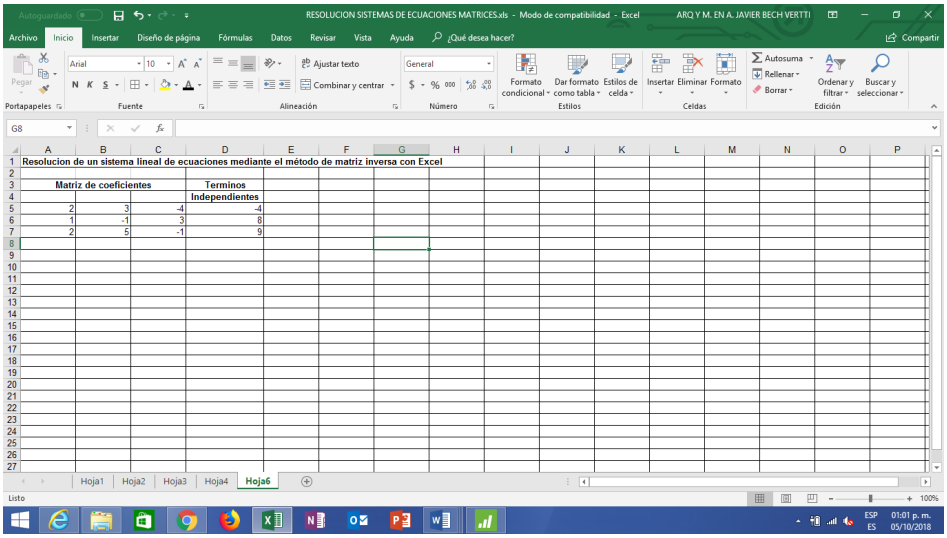
Y = 2

Z = 3

La única condición que debe darse para resolver este tipo de sistemas con **Excel** es que debe haber el mismo número de variables que de ecuaciones. El ejemplo anterior utiliza dos variables y dos ecuaciones. Tener tres ecuaciones significaría que deben existir también tres variables.

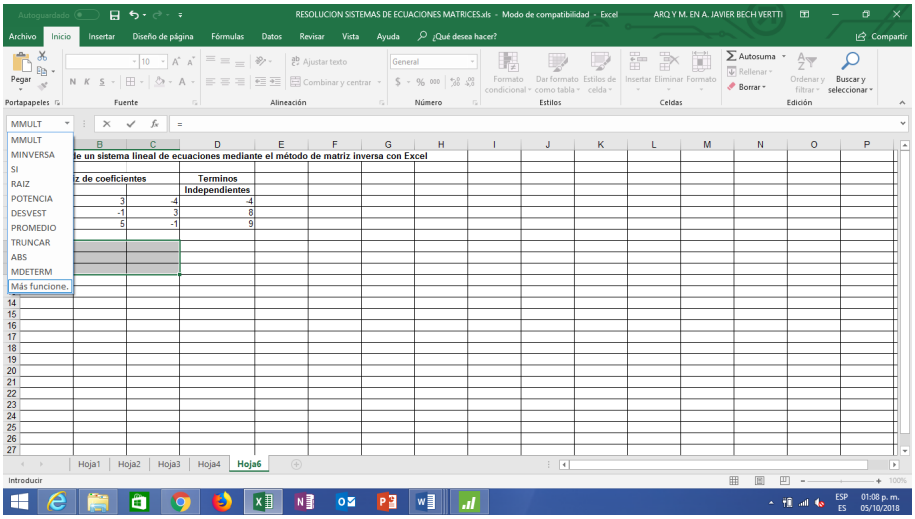
Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

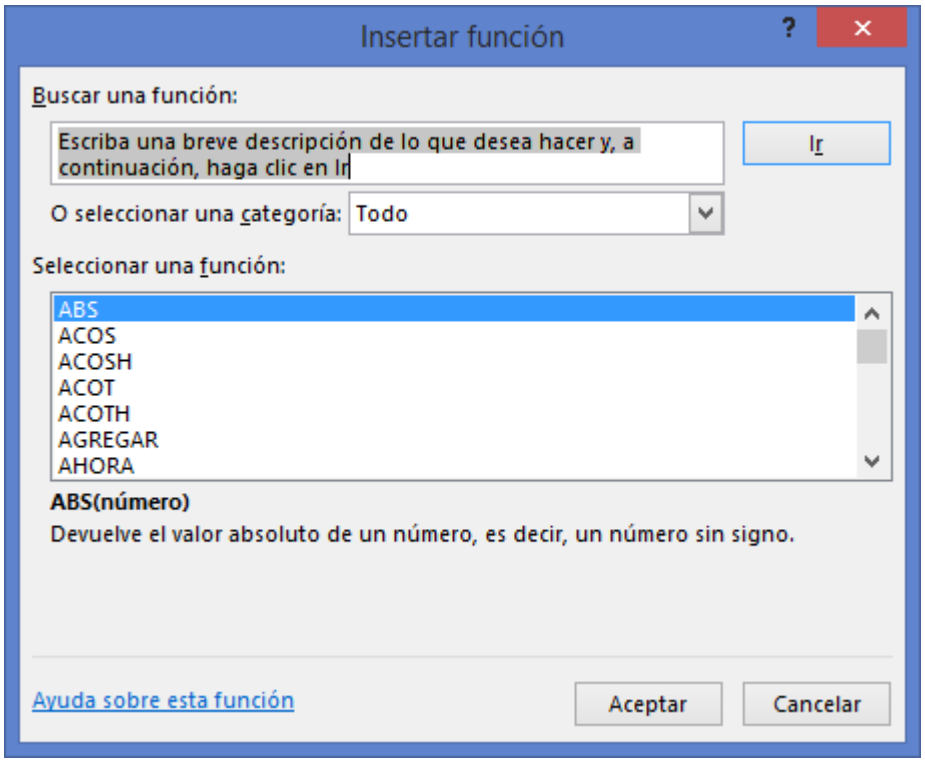
Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).
Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.
Introduce los valores independientes en otro rango al lado.



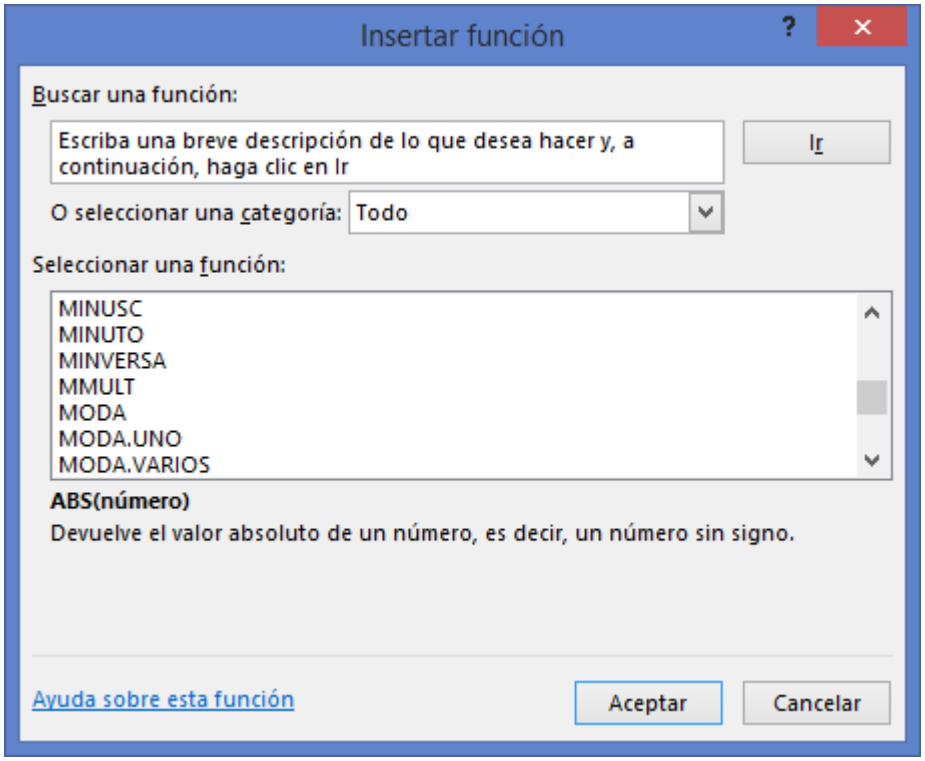
Utiliza la función matricial **MINVERSA ()** para calcular la matriz inversa de los coeficientes introducidos. (Recuerda que para introducir la función debes pulsar Ctrl + Mayús + Intro).

Como la matriz de coeficientes es de 3 x 3 se necesitas 9 celdas para mostrar los resultados de la matris Inversa para lo cual selecciona con el mouse 9 celdas debajo de la matriz de coeficientes dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MINVERSA** () seleccionando la opción **Más funciones** en la parte inferior

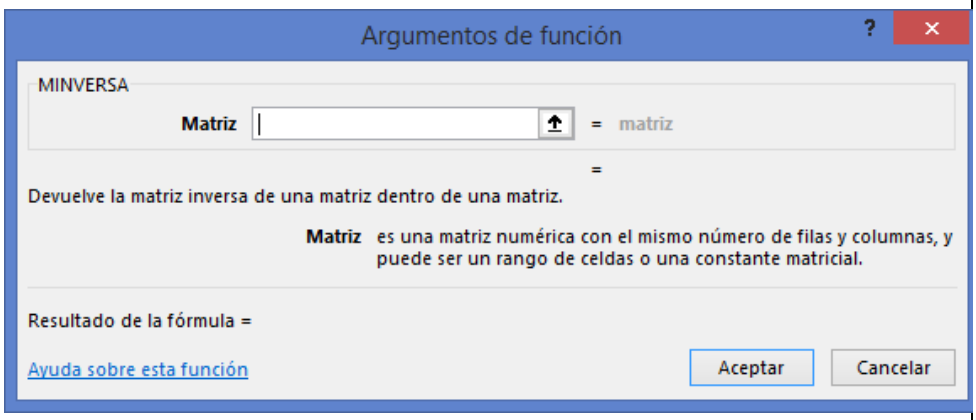

RESOLUCION SISTEMAS DE ECUACIONES MATRICES.xls - Modo de compatibilidad - Excel
ARQ Y M. EN A. JAVIER BECH VERTTI
MMULT
MINVERSA
SI
RAIZ
POTENCIA
DESVEST
PROMEDIO
TRUNCAR
ABS
MDETERM
Más funcione.
e un sistema lineal de ecuaciones mediante el método de matriz inversa con Excel
z de coeficientes
Terminos Independientes
Hoja1
Hoja2
Hoja3
Hoja4
Hoja6


Insertar función
Buscar una función:
Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir
Ir
O seleccionar una categoría: Todo
Seleccionar una función:
ABS
ACOS
ACOSH
ACOT
ACOTH
AGREGAR
AHORA
ABS(número)
Devuelve el valor absoluto de un número, es decir, un número sin signo.
Ayuda sobre esta función
Aceptar
Cancelar

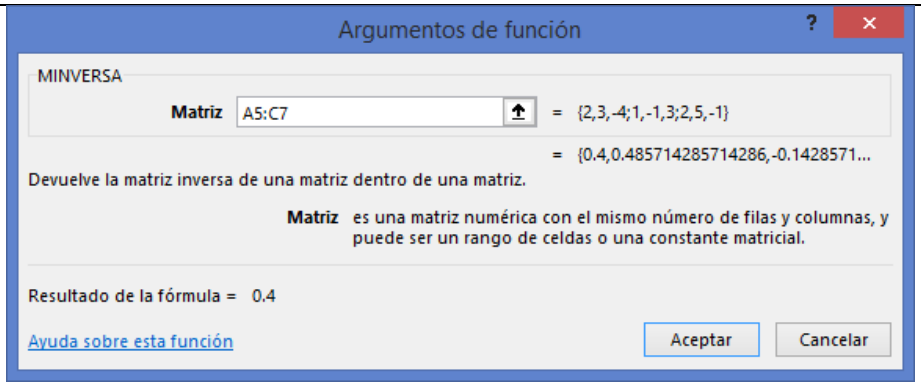
En Seleccionar una función busca **MINVERSA** y da un clic.



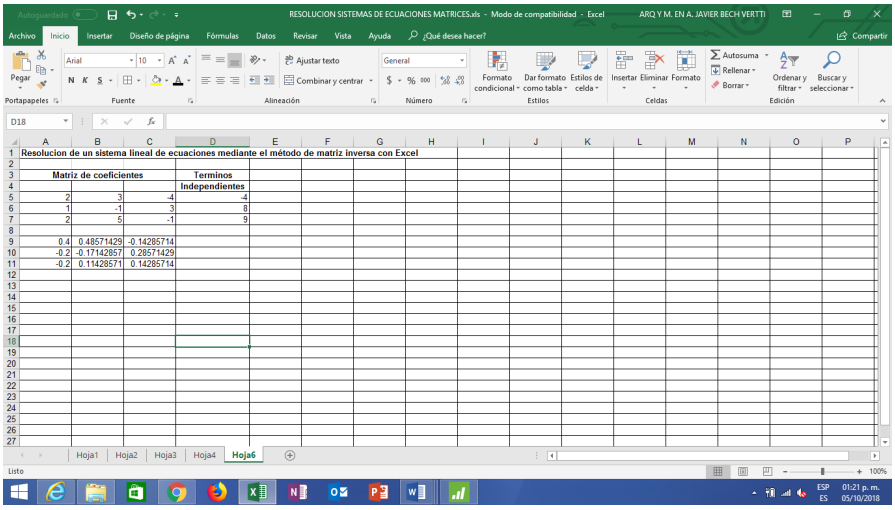
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



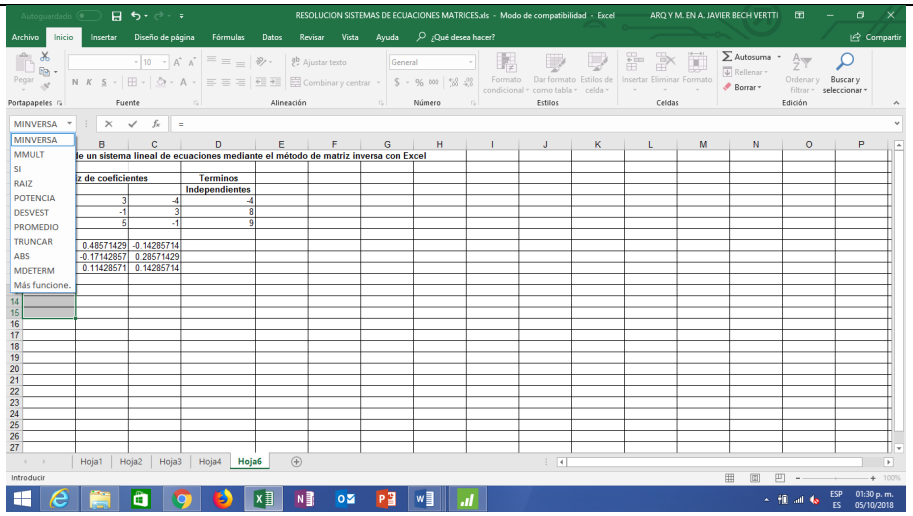
Selecciona las celdas de la matriz de coeficientes, en este caso A5:C7

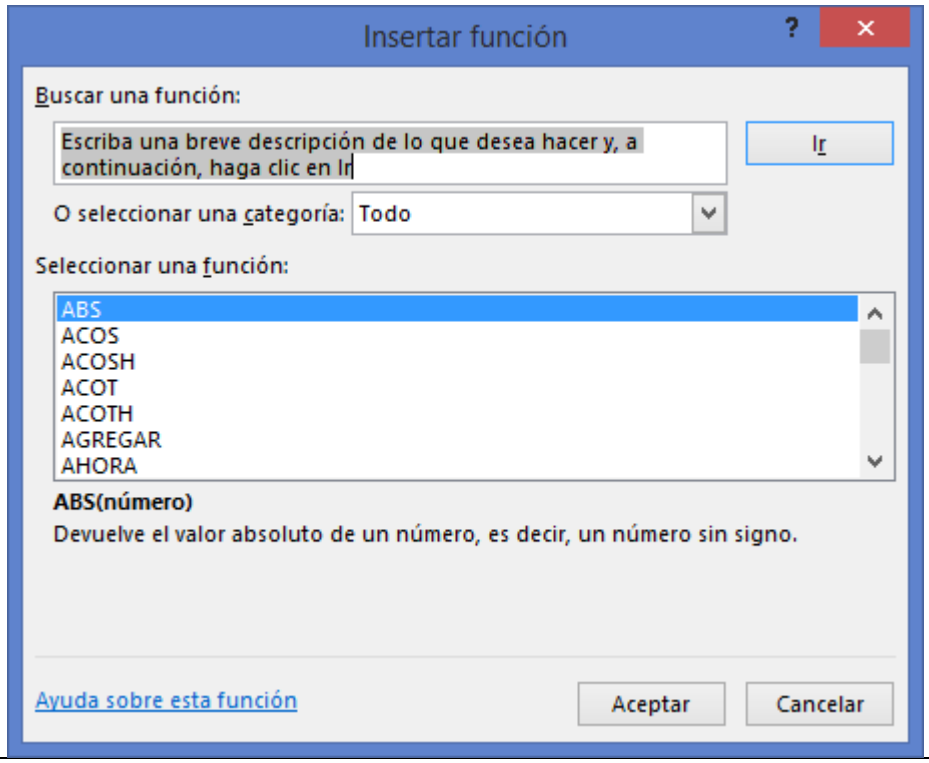


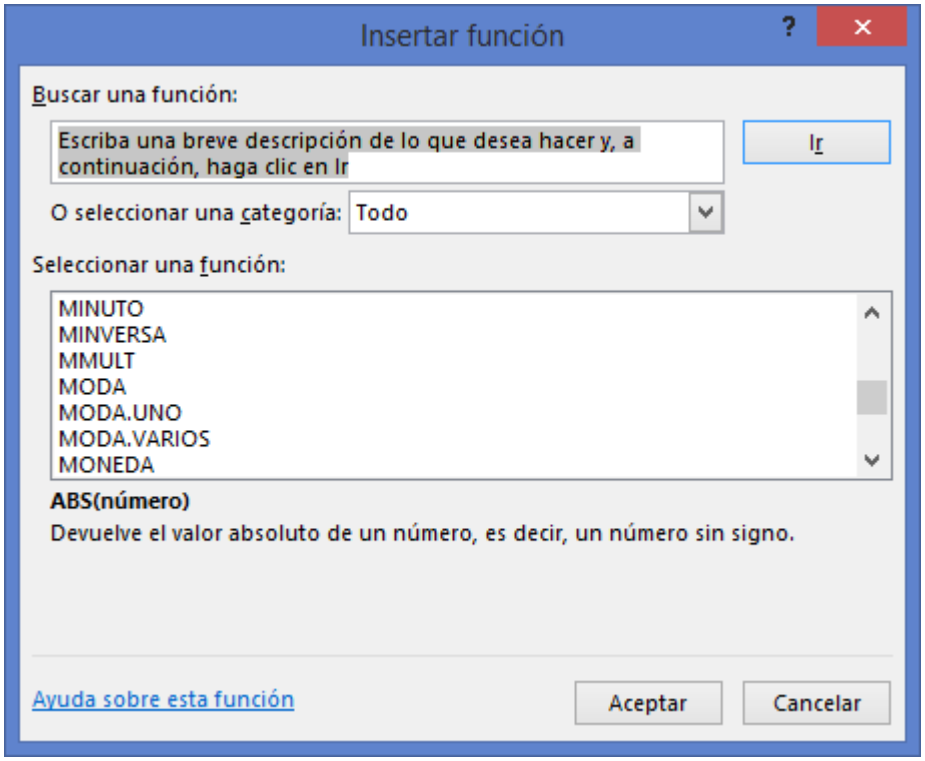
Ojo, no pulses Aceptar. La matriz inversa se determinará al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter)**.



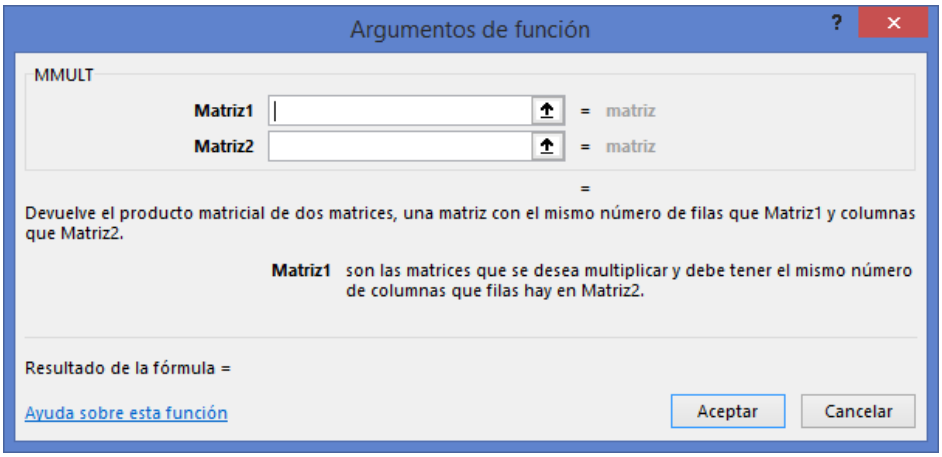
Como la matriz de los términos constantes es un vector de 1 x 3 se necesitas 3 celdas para mostrar los resultados de las Incógnitas **X, Y y Z** para lo cual selecciona con el mouse 3 celdas debajo de la **matriz inversa** dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MMULT** () seleccionando la opción **Más funciones** en la parte inferior


RESOLUCION SISTEMAS DE ECUACIONES MATRICES.xls - Modo de compatibilidad - Excel
ARQ.Y M. EN A. JAVIER BECH VERTTI
MINVERSA
MMULT
SI
RAIZ
POTENCIA
DESVEST
PROMEDIO
TRUNCAR
ABS
MDETERM
Más funcione...
Hoja1
Hoja2
Hoja3
Hoja4
Hoja6
Introducir

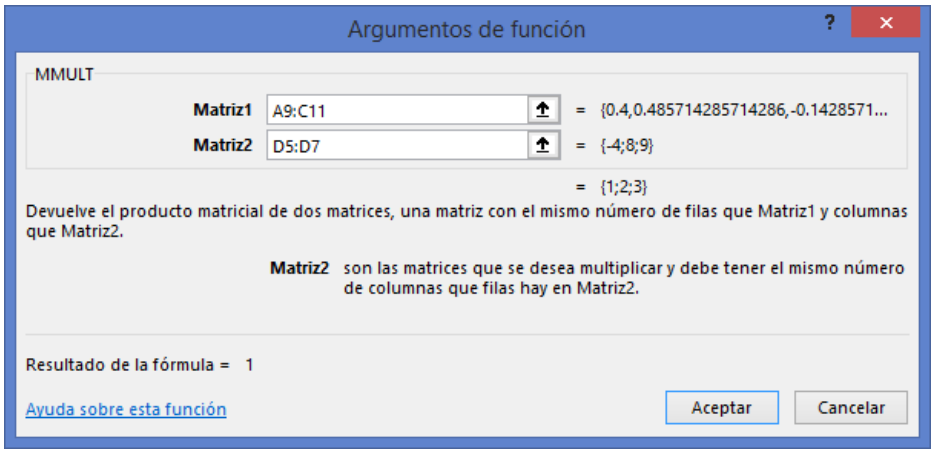

Insertar función
Buscar una función:
Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir
Ir
O seleccionar una categoría: Todo
Seleccionar una función:
ABS
ACOS
ACOSH
ACOT
ACOTH
AGREGAR
AHORA
ABS(número)
Devuelve el valor absoluto de un número, es decir, un número sin signo.
Ayuda sobre esta función
Aceptar
Cancelar



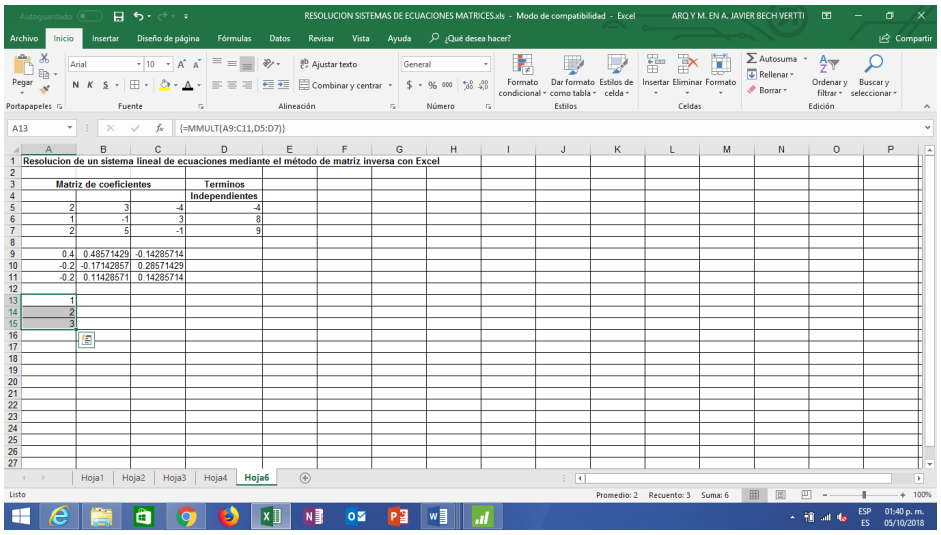
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



En el campo que dice Matriz 1 selecciona las celdas de la matriz de inversa, en este caso A9:C11 y en el campo que dice Matriz 2 selecciona las celdas de los términos constantes, en este caso D5:D7



Ojo, no pulses Aceptar. Los resultados de las incógnitas **X, Y, Z** se determinarán al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter), en este caso 1,2,3.** Por lo tanto las incognitas quedan **X=1, Y=2 y Z=3.**



Con lo anterior queda terminado el ejemplo ilustrativo de **resolución de un sistema lineal de ecuaciones mediante el método de matriz inversa** en el paquete **EXCEL**. **Aplica el procedimiento para salir de Windows**

1.4.1.2 ACTIVIDAD DE APRENDIZAJE EN EXCEL. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.

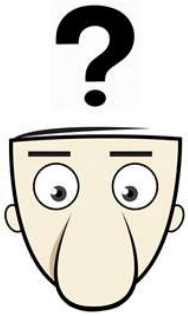
ACTIVIDAD DE APRENDIZAJE EN EXCEL
1.4.1.2
RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ INVERSA

Resolver el siguiente sistema, empleando el método de la MATRIZ INVERSA con Excel:

$$x + y + z = 11$$
$$2x - y + z = 5$$
$$3x + 2y + z = 24$$

1.4.1.2 EJERCICIO DE AUTOEVALUACIÓN en Excel. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra en el cuaderno de trabajo anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTO**EVALUACIÓN EN EXCEL** **1.4.1.2** **R**ESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA** 	Resolver en **Excel** el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $\|A\|$ $x + 2y - 3z = 5$ $3x - 22y + 6z = 4$ $7x - 6y - 3z = 15$

1.4.1.2 EJERCICIOS DE **REFUERZO. Resolución de un Sistema Lineal de Ecuaciones mediante el Método de matriz Inversa.**

EJERCICIOS DE REFUERZO EN EXCEL 1.4.1.2 RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ INVERSA EN EXCEL

1.4.1.2.1 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$x + y + 2z = 31$$
$$x + z + 2y = 32$$
$$y + z + 2x = 33$$

1.4.1.2.2 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$9x - 2y + 12z = 20$$
$$6x + 4y - 3z = 3$$
$$3x + 2y - 3z = -1$$

1.4.1.2.3 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$2x - y + 3z = 14$$
$$2y - z + 3x = 24$$
$$2z - x + 3y = 10$$

AUTOEVALUACIÓN CON REACTIVOS DE FALSO Ó VERDADERO

EN CADA UNO DE LOS REACTIVOS, CONTESTE CON UNA F SI CONSIDERA QUE LA AFIRMACIÓN ES FALSA Y CON UNA V SI CONSIDERA QUE LA AFIRMACIÓN ES VERDADERA.

1. La supresión de caso(s) y/o variable(s) se usa como tratamiento en caso de datos ausentes. ( )
2. La prueba de Barlett se utiliza para probar la homocedasticidad de varianzas de una variable. ( )
3. Si los datos provienen de distribuciones continuas, pero no necesariamente normales, la mejor prueba para checar la homocedasticidad es la prueba de Levene. ( )
4. En muestras grandes (mayores de 80) las pautas sugieren que el umbral de estandarizado para identificar los casos atípicos se sitúe entre 1.28 y 1.96. ( )
5. El método de transformación de los datos se usa para tratar los datos ausentes. ( )
6. Para la distribución de variables no normales asimétricas positivas, la transformación mas común para conseguir normalidad y homocedasticidad es el logaritmo decimal. ( )
7. La prueba de Anderson Darling se utiliza para probar la homocedasticidad de varianzas de una variable.

( )

8. Para evitar un efecto adverso en caso de datos ausentes se usa el método de impugnación.

( )

9. Para conseguir la linealidad entre dos variables, la transformación que se aplica es el logaritmo natural.

( )

10. La prueba de Barlett se utiliza para probar la normalidad de una distribución de datos.

( )

11. Si se sustituyen los valores ausentes en una base de datos por un único valor derivado de fuentes externas, la imputación fue por sustitución de caso.

( )

12. Para distribución de variables no normales planas, la trasformación mas común para conseguir normalidad y homocedasticidad es la raíz cuadrada.

( )

13. Para conseguir la linealidad entre dos variables, la transformación que se aplica es la inversa.

( )

14. Si se sustituyen valores ausentes en una base de datos por un único valor derivado de fuentes externas, la imputación fue por la media.

( )

15. En muestras pequeñas (de 80 incluso menos), las pautas sugeridas identifican como atípicos aquellos casos con valores estándar entre 1.96 y 2.58.

( )

16. La prueba de Shapiro-Wilks se utiliza para probarla normalidad de una distribución de datos.

( )

17. Si los datos provienen de distribuciones continuas, pero no necesariamente normales, la mejor prueba para checar la homocedasticidad es la prueba de Barlett.

( )

18. Si se sustituyen los valores ausentes en una base de datos por un único valor derivado de fuentes externas, la imputación fue por sustitución por valor constante.

( )

19. En muestras grandes (mayores de 80) para identificar los casos atípicos, las pautas sugieren que el umbral estandarizado se sitúe entre 3 y 4.

( )

20. Para distribución de variables no normales planas, la transformación mas usual para

conseguir normalidad y homocedasticidades la inversa. ( )

21. En muestras pequeñas (de 80 incluso menos), las pautas sugeridas identifican como atípicos aquellos casos con valores estándar de 2.58 ó superiores. ( )

22. Si los datos provienen de distribuciones continuas, con distribuciones normales, la mejor prueba para checarla homocedasticidad es la prueba de Shapiro-Wilks. ( )

23. Para distribución de variables no normales asimétricas negativas, la transformación mas común para conseguir normalidad y homocedasticidades el logaritmo natural. ( )

24. Para la distribución de variables no normales asimétricas positivas, la transformación mas usual para conseguir normalidad y homocedasticidad es la inversa. ( )

25. Si los datos provienen de distribuciones continuas, con distribuciones normales, la mejor prueba para checar la homocedasticidad es la prueba de Barlett. ( )

26. Para distribución de variables no normales asimétricas negativas, la transformación mas ususal para conseguir normalidad y homocedasticidades la raíz cuadrada. ( )

AUTOEVALUACIÓN CON REACTIVOS DE **OPCIÓN MÚLTIPLE**

EN CADA UNO DE LOS REACTIVOS SIGUIENTES, SELECCIONE LA OPCIÓN QUE CONSIDERE CORRECTA.

1.- Los siguientes tratamientos corresponden a datos ausentes excepto:

a) Utilizar sólo aquellas observaciones con datos completos.

b) Supresión de caso(s) y/o variable(s).

c) Método de impugnación

d) Método de transformación de los datos.

2.- Cuando en una base de datos se tiene datos ausentes, la imputación es el proceso de estimación de variables y/o casos de la muestra. Si se sustituyen los valores ausentes por una variable cuyo valor medio se calcula sobre todas las respuestas, la imputación fue por:

a) Sustitución de caso.

b) Sustitución por la media.

c) Sustitución por valor constante

d) Imputación por regresión.

3.- Si se sustituyen los valores ausentes en una base de datos por un único valor derivado de fuentes externas, la imputación fue por:

a) Sustitución de caso.

b) Sustitución por la media.

c) Sustitución por valor constante

d) Imputación por regresión.

4.- Si se predicen valores ausentes de una variable basándose en su relación con las otras variables del conjunto de datos, la imputación fue:

a) Sustitución de caso.

b) Sustitución por la media.

c) Sustitución por valor constante

d) Imputación por regresión.

5.- Pueden identificarse los casos atípicos desde un enfoque univariado seleccionando como casos atípicos aquellos casos que caigan fuera de los rangos de la distribución. En muestras pequeñas (de 80 incluso menos) las pautas sugeridas identifican como atípicos aquellos casos con valores estándar de:

a) 1.96

b) 2.58 ó superiores

c) entre 1.96 y 2,58

d) entre 3 y 4

6.- Pueden identificarse los casos atípicos desde un enfoque univariado seleccionando como casos atípicos aquellos casos que caigan fuera de los rangos de la distribución. En muestras grandes (mayores de 80) las pautas sugieren que el umbral de estandarizado se sitúe entre:

a) entre 1.28 y 1.96

b) entre 2.58 y 3

c) entre 1.96 y 2,58

d) entre 3 y 4

7.- Las siguientes pruebas sirven para probar la normalidad de una distribución de datos excepto:

a) Shapiro-Wilks

b) Kolmogorov-Smirnov

c) Anderson Darling

d) Barlett

8.- Las siguientes pruebas sirven para probar la homocedasticidad de varianzas de una variable para diferentes grupos excepto:

a) Barlett

b) Prueba F

c) Levene

d) Anderson Darling

9.- Si los datos provienen de distribuciones continuas, pero no necesariamente distribuciones normales, la mejor prueba para checar la homocedasticidad es:

a) Levene

b) Barlett

c) Shapiro-Wilks

d) Anderson Darling

10.- Si los datos provienen de distribuciones continuas, con distribuciones normales, la mejor prueba para checar la homocedasticidad es:

a) Levene

b) Barlett

c) Shapiro-Wilks

d) Anderson Darling

11.- Para las distribuciones de variables no normales planas, la transformación más común para conseguir normalidad y homocedasticidad es:

a) La inversa

b) La raíz cuadrada

c) El logaritmo decimal

d) El logaritmo natural

12.- Para las distribuciones de variables no normales asimétricas negativas, la transformación más común para conseguir normalidad y homocedasticidad es:

a) La inversa

b) La raíz cuadrada

c) El logaritmo decimal

d) El logaritmo natural

13.- Para las distribuciones de variables no normales asimétricas positivas, la transformación más común para conseguir normalidad y homocedasticidad es:

a) La inversa

b) La raíz cuadrada

c) El logaritmo decimal

d) El logaritmo natural

14.- Las siguientes transformaciones se aplican para conseguir la linealidad entre dos variables excepto:

a) La inversa

b) La raíz cuadrada

c) El logaritmo decimal

d) El logaritmo natural

CAPÍTULO 2. MÉTODOS DE REDUCCIÓN

OBJETIVO 2.1 El alumno aplicará la técnica de componentes principales (Factor Análisis ó Análisis Factorial ó de Factores) para identificar un conjunto de dimensiones o características latentes dentro de un conjunto grande de variables (factores)

ANTECEDENTES

CONCEPTOS DE:

Factor. Carga de Factor. Eigenvalor. Porcentaje de Variación Explicada. Comunalidad. Variación Total. Variación Común. Variación Específica. Variación Aleatoria.

2.1.1 FACTOR, CARGA DE FACTOR, PORCENTAJE DE VARIACIÓN EXPLICADA, COMUNALIDAD.

CONCEPTOS BÁSICOS ANALISIS DE FACTORES

El análisis por factores (FA: *factor analysis*) es una técnica que se emplea frecuentemente para crear nuevas variables que resumen toda la información de la que podría disponerse en las variables originales. Por ejemplo, considere una vez más la realización de exámenes de lectura, ortografía y ciencias a estudiantes del tercer año de primaria, en los que cada uno de los estudiantes puede obtener calificaciones altas, regulares o bajas en las cuatro áreas. Se esto ocurre en realidad, entonces se podría decir que estos resultados se están explicando por alguna característica o factor subyacente que es común a las cuatro pruebas. En este ejemplo, podría resultar razonable suponer que unan característica subyacente de ese tipo es la "inteligencia global".

El análisis por factores también se usa para estudiar las relaciones que podrían

existir entre las variables medidas en un conjunto de datos. Semejante al PCA, el FA es una técnica dirigida por las variables. Un objeto básico del FA es determinar si las variables respuesta exhiben patrones de relaciones entre sí, tales que esas variables se pueden dividir en subconjuntos de modo que las variables en un subconjunto estén fuertemente relacionadas una con cada una de las otras y que las variables en subconjuntos diferentes que tengan bajas correlaciones entre sí. Por lo tanto, el FA se usa con frecuencia para estudiar la estructura de correlación de las variables en un conjunto de datos. Una semejanza entre el FA y el PCA es que aquél también se puede usar para crear nuevas variables que no estén correlacionadas entre sí. Esas variables se llaman *clasificación de factores.*

Análisis de la estructura de datos

Existen dos procedimientos para analizar la estructura de covarianza de los datos:

· Componentes principales ayuda a comprender la estructura de covarianza en las variables originales y/o a crear un número menor de variables utilizando esta estructura.

· Análisis factorial, como los componentes principales, resume la estructura de covarianza de datos en un número más pequeño de dimensiones. El énfasis en el análisis factorial es la identificación de los "factores" subyacentes que podrían explicar las dimensiones asociadas con una gran variabilidad en los datos.

Una ventaja que parece tener el FA sobre el PCA, cuando se están creando nuevas variables es que, generalmente, las nuevas variables creadas por el FA son mucho más fáciles de interpretar que las creadas por el PCA. Si un investigador desea crear un conjunto más pequeño de nuevas variables que se pueden interpretar y que resuman la mayoría de la información existente en las variables medidas, entonces el FA debe considerarse seriamente.

Se considere una técnica estructural cuyo objetivo básico es el de resumir información donde sus variables son cuantitativas ò métricas y la mayoría de ellas deben estar medidas en escala de intervalo ò de razón.

Esta técnica analiza las relaciones entre un gran número de variables y trata de explicarlas por medio de un concepto (atributos, características o dimensiones) que las agrupe. Es decir, condensa la información de las variables originales mediante características comunes a dichas variables y que las representan lógicamente. Los factores obtenidos pueden utilizarse como variables en análisis posteriores, facilitando el manejo de la información ya condensada.

El análisis de factores no es una técnica exacta, sino a base de prueba y error y teniendo guías (criterios) que nos llevan hacia donde esta la respuesta.

OBJETIVOS ESPECÌFICOS DEL ANALISIS DE FACTORES

Utilice un análisis factorial , al igual que el análisis de componentes principales, a fin de resumir la estructura de covarianzas de los datos en algunas dimensiones de los datos. Sin embargo, el énfasis en el análisis factorial es la identificación de los "factores" subyacentes que pudieran explicar las dimensiones asociadas con una variabilidad grande de los datos.

1. Identificar un conjunto de dimensiones o características que se encuentran latentes (las cuales no se detectan fácilmente) dentro de un conjunto grande de variables. Cuando se está interesado en este objetivo se le llama *análisis de factores R.*
2. Encontrar características que describan a núcleos poblacionales (personas). Si se persigue este objetivo se llama *análisis de factores Q*. Es decir, a partir de variables se encontrarán conceptos que describan a personas; la diferencia con el análisis de factores anterior o R es que éste describe al mercado, *no* a las variables.
3. Utilizar a las nuevas variables generadas por la técnica para realizar estudios ò análisis posteriores como podrían ser entre otros la regresión, el discriminante, etc.
4. Crear datos para las nuevas variables encontradas, a partir de la información original.

TERMINOS IMPORTANTES EN EL ANÀLISIS DE FACTORES

Factor

Desde el punto de vista matemático un *factor* es una combinación lineal de variables originales de la investigación y está representado por la siguiente ecuación:

$$F_{ij} = A_{i1}X_{1j} + A_{i2}X_{2j} + A_{i3}X_{3j} + \ldots + A_{ik}X_{kj}$$

Donde:

i: número de factor
j: número de individuo o encuesta
k: número de variable

En esta ecuación las A representan la importancia que tiene cada variable, con respecto a la característica obtenida.

Asimismo es importante tener presente que las variables usadas en la técnica deben estar estandarizadas de tal manera que en términos de unidades puedan ser comparables, es decir su media es cero y su varianza igual a uno.

Carga de factor

Este término es el que se la da a la relación que llegan a tener cada una de las variables con cada uno de los factores. Es una medida de correlación entre una variable y un factor.

Eigenvalor

Se usa este término para denominar a la variación que cada uno de los factores logra explicar del conjunto de variables.

Para calcular el Eigenvalor, cada carga se eleva al cuadrado y se suma para cada factor.

Porcentaje de variación explicada

Representa la relación entre la variación explicada por el factor, con respecto a la variación máxima, la cual es igual al numero de variables, ya que cada una de ellas tiene una variación de uno por estar estandarizadas; por lo tanto, la mayor variación que se puede llegar a explicar sería la suma de las variaciones de cada variable.

A mayor porcentaje de varianza explicada quiere decir que el factor está explicando mejor a la variable.

Comunalidad

Se usa este término para nombrar a la variación explicada por la solución de factores para cada variable y por estar estandarizadas, el valor máximo posible de la comunalidad es de 1. Mientras más se aproxime el valor de la comunalidad a 1 quiere decir que se explico mejor la variable.

El valor de la comunalidad total se obtiene elevando al cuadrado cada carga y sumándolo por cada variable.

PASOS PARA LLEVAR A CABO UN ANÁLISIS DE FACTORES

El objetivo del análisis factorial es hallar un número pequeño de factores, o de variables no observables, que explique la mayor parte de la variabilidad de datos y aun así tenga sentido contextual. Es deseable que se decida cuántos factores desea utilizar y hallar las cargas que tengan más sentido para sus datos.

1. Identificar el problema
2. Generar la matriz de correlación
3. Seleccionar el modelo de factores que mejor convenga; puede ser de componentes principales o factor común.
4. Seleccionar un criterio para determinar qué características deberán presentarse en los factores, ortogonal u oblicuo.
5. Obtener una matriz de factores no rotada
6. Obtener una matriz de factores rotada.
7. Obtener los puntajes de los factores encontrados (opcional).

Identificar el problema

Se debe establecer claramente cuál es el problema que se tratará en la investigación y la información que se pretende resumir agrupando las variables en factores, para lo cual se debe considerar lo siguiente:

- La técnica no acepta variables medidas en escala nominal, es decir deben ser métricas, aunque permite el uso de variables *dummy* en menor número que las de intervalo ò razón.
- La técnica de factores no debe usarse para los casos en los que el número de entrevistas sea menor a 50 aunque se sugiere que sean más de 100 entrevistas. En general se considera que el tamaño de la muestra debe ser cuatro o cinco veces el número de variables que se pretenda agrupar con la técnica.

Generar la matriz de correlación

Proporciona las medidas de relación de las variables de interés, comparándolas por parejas y sirve en el análisis de factores cuando las variables se identifican con más de un factor o característica y se tiene que decidir con que factor se relaciona la variable.

Seleccionar el modelo de factores

En el análisis de factores se busca agrupar variables, buscando aspectos comunes entre ellas. La dispersión o variación de los datos indica posibles similitudes entre las variables; esta técnica usa la variación o varianza para medir la dispersión de la información, tomando como base el siguiente modelo:

Variación total= Variación comùn+Variación especifica + variación Aleatoria

Donde:

-*Variación total*: indica la variación o dispersión de cada variable.
-*Variación común*: representa la variación que hace parecidas a las variables, lo que las une.
-*Variación especifica*: lo que hace diferentes a las variables.
-*Variación aleatoria*: es el error, presente en todo modelo y como no se puede cuantificar, no se toma en cuenta.

En general la mayoría de los paquetes estadísticos computacionales usan dos tipos de modelos para identificar qué tanto se relacionan las variables con el factor:

1. *Componentes principales.* En este modelo se asume que la variación específica es tan pequeña que pudiéramos considerarla como cero y por lo tanto la variación total se considera igual a la variación común. Los paquetes computacionales si no se les especifica algún otro manejan este modelo.
2. *Factor común.* Asume que la variación específica es importante y que por lo tanto se deberá eliminar del modelo de variación y con el resto intentar agrupar las variables.
3.

Seleccionar el criterio para encontrar factores

Los factores pueden tener o no relación entre si. La decisión de obtener factores independientes o dependientes se hará seleccionando alguno de los criterios siguientes:

1. *Ortogonal* . Va a extraer características que sean independientes, es decir, que tienen una relación igual a cero.
2. *Oblicuo*. Extrae factores con cierto grado de relación entre ellos.

Obtener la matriz de factores no rotada

Esta matriz nos proporciona información en relación a que tanto se identifica cada variable con cada factor; estas medidas se llaman *cargas de factores*. Como en esta matriz casi todas las cargas resultan significativas por eso no se agrupan a las variables en este paso, solamente se determina el *número de factores* en los que quedarán agrupadas las variables de manera tentativa.

Para determinar el número de factores como solución inicial, la mayor parte de los paquetes computacionales utilizan el criterio denominado de la raíz latente. Mediante este criterio, se considera que un factor debe ser retenido en la solución si su Eigenvalue es mayor a 1. Si lo que queda por explicar es una cantidad menor a 1, no tendrá caso agregar un factor mas a la solución.

Asimismo se usa normalmente el criterio del porcentaje de variación explicada acumulada que establece que se deben considerar n factores como solución inicial, si el porcentaje de variación explicada acumulada se encuentra en un rango de 60% a 95%. Es decir, con este criterio se está dispuesto a perder

Número de factores

La selección del número de factores se basa con frecuencia en la proporción de varianza explicada por los factores, el conocimiento del tema y la sensatez de la solución. Inicialmente, se trata de utilizar el método de extracción de los componentes principales especificando el número de componentes igual al número de variables. Se examina la proporción de la variabilidad explicada por diferentes factores y se reduce su selección a cuántos factores se utilizarán. Una gráfica de sedimentación puede ser útil aquí para evaluar visualmente la importancia de los factores. Una vez que se haya reducido esta selección, se examina los ajustes de los diferentes análisis factorial. Los valores de comunalidad y la proporción de la variabilidad

de cada variable explicada por los factores que pueden ser de especial utilidad para comparar los ajustes. Se puede decidir agregar un factor si contribuye al ajuste de ciertas variables. Se puede tratar también con el método de extracción de máxima verosimilitud.

Rotación

Una vez que se haya seleccionado el número de factores, probablemente usted se querrá probar con diferentes rotaciones. Johnson y Wichern sugieren la rotación varimax. Un resultado similar de diferentes métodos puede dar crédito a la solución que se escogió. En este punto se pudiera interpretar los factores utilizando su conocimiento de los datos.

cuando mucho 40% y cuando menos 5% de información.

Cabe señalar que cada factor que se agrega a la solución en la matriz no rotada explica en una cantidad menor al conjunto de variables, puesto que cada uno que se agrega tiene menor variación por explicar.

Es en este paso donde se señala en forma inicial o tentativa el número de factores en que quedaràn agrupadas las variables pero no se agrupan ni concluyen las características que se obtienen.

Obtener la matriz de factores rotada

Esta matriz permite definir no solo el número de factores sino la agrupación de las variables en dichos factores, procurando obtener así una solución final lógica.

Mediante este proceso se ajustan los ejes coordenados o ejes factor en la dirección en que giran las manecillas del reloj con el fin de obtener una solución mas sencilla y con mayor significado teórico, es decir, hacer que una de las cargas de la variable sea lo mas alta posible para que se identifique perfectamente con un solo factor, mientras que sus demás cargas sean bajas de manera que su relación con otros factores se pierda.

En el proceso de rotar los factores se redistribuyen las cargas cambiando el valor de los Eigenvalue para cada factor sin que la variación explicada total y las comunalidades cambien, a menos que se eliminen o agreguen factores por lo que al rotar no se explica más o se pierde información solo se redistribuyen las cargas.

Al rotar los factores se trata de que las variables queden en un solo eje y que se agrupen las variables bajo una sola característica, lo que daría la mejor solución. Como este proceso es complejo para llevarlo a cabo manualmente se han creado paquetes computacionales que realizan rápidamente la rotación de factores, sin embargo hay que especificarles el criterio que deberá usar para realizar la rotación de factores y que en general es el criterio Varimax, mediante el cual se trata de identificar a un grupo de variables con un solo factor, es decir, simplifica por columna, buscando la máxima simplificación al generar tantos unos y ceros como sea posible en la matriz.

Una vez generada la matriz rotada, será necesario identificar en ella las cargas significativas que indiquen con qué factor se han identificado cada una de las variables, lo cual depende del tamaño de la muestra y el nivel de significancia manejado en la investigación, por lo que es necesario conocer el tamaño de la muestra ya que a medida que la muestra es mas grande existe mayor credibilidad o confianza en la información y se puede por lo tanto fijar limites cada vez menores en la carga del factor para considerarla significativa.

Criterios para identificar cargas significativas

1.Cuando *n* está entre 50 y 100:

a) Carga altamente significativa es la de $\pm$ 0.5 para arriba (prioridad 1).
b) Carga sólo significativa es entre $\pm$ 0.4 a $\pm$ 0.5 (prioridad 2).
c) Carga sòlo significativa es entre $\pm$ 0.3 a $\pm$ 0.4 (prioridad 3).
d) Si es menor a 0.3 se considera que no existen cargas significativas (prioridad 4).

2.Cuando *n* está entre 100 y 200:

a) Para un *alfa* igual a 5%, una carga significativa es aquella de $\pm$ 0.19 en adelante, habiendo aplicado primero las prioridades a,b,c y d del apartado anterior. Con una carga menor de 0.19 ya no se considera significativa.
b) Para un *alfa* igual a 1 % una carga es significativa de $\pm$ 0.26 en adelante habiendo aplicado las prioridades del apartado anterior.

3. Cuando *n* está entre 200 y 300:

a) Para un *alfa* igual a 5%, una carga significativa es aquella de $\pm$ 0.14 en adelante, habiendo aplicado primero las prioridades a,b,c y d del apartado anterior.
b) Para un *alfa* igual a 1 % una carga es significativa de $\pm$ 0.18 en adelante habiendo aplicado las prioridades del apartado anterior.

4. Cuando *n* es mayor o igual a 300:

a) Para un *alfa* igual a 5%, una carga significativa es aquella de $\pm$ 0.11 en adelante, habiendo aplicado primero las prioridades a,b,c y d del apartado anterior.
b) Para un *alfa* igual a 1 % una carga es significativa de $\pm$ 0.15 en adelante habiendo aplicado las prioridades del apartado anterior.

Cabe mencionar que a estos criterios se les puede dar continuidad, es decir, si no se cumple el criterio de la prioridad uno, se pasarà al criterio de la prioridad dos, y así sucesivamente, considerando el tamaño de la muestra y el nivel de significancia especificado en la investigación.

Por otro lado, si existe alguna variable que no tenga carga significativa, es decir, que no se haya identificado con ningún factor, puede deberse a que esa variable tenga una variación específica muy alta y es muy particular por lo que en ese caso existen dos posibles acciones:

- Se deja la solución tal cual, sin agrupar a esa variable y así se realiza el análisis.
- Se elimina la variable del banco de datos y se vuelve a correr de nuevo la solución.

Por último hay que considerar que la solución inicial no necesariamente es la mejor. Hay que analizar si las variables que quedan bajo cada factor se pueden agrupar lógicamente bajo una característica común ya que si no, no sería la mejor solución y se tendrían que probar otras soluciones con mas ò menos factores hasta que las variables queden agrupadas de manera lógica.

2.1.1.1 EJEMPLO ILUSTRATIVO

EJEMPLO ILUSTRATIVO 2.1.1.1 FACTOR ANÁLISIS

El Director de mercadotecnia de una compañía de Bienes Raíces desea conocer cuales dimensiones o indicadores básicos influyen al momento de pronosticar el valor de una casa o vivienda unifamiliar al momento en que un potencial cliente desea ya sea comprar o vender su casa con el objeto de brindar un servicio confiable y rápido. Para llevar a cabo la investigación dividió la Ciudad en tres grandes zonas: NORTE, SUR Y ESTE. En cada zona se aplicó una encuesta a 20 propietarios de casas unifamiliares. El cuestionario estaba conformado por 20 variables que se consideraba podían formar parte de dichos indicadores X_1= Tamaño del lote y X_2= Número de recámaras, , X_3= Número de baños, X_4= Número de habitaciones y X_5= Antigüedad de la casa, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico, X_9= Tipo de combustible para la calefacción, X_{10}= Tipo de sistema para la calefaccción, X_{11}= Tipo de alberca, X_{12}= Cocina con antecomedor y X_{13}= Aire acondicionado, X_{14}= Chimenea, X_{15}= Drenaje municipal X_{16}= Sótano, X_{17}= Cocina integral, X_{18}= Tina de hidromasaje.

Los resultados en la zona NORTE fueron los siguientes:

Casa No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	6.00	3	1.0	6	45	1600	0	1	2
2	7.00	3	1.5	10	28	3500	1	5	1
3	7.00	4	2.0	6	35	3600	0	4	2
4	7.00	3	2.0	7	35	4100	1	3	1
5	6.00	3	2.0	9	32	4400	2	5	2
6	7.75	4	2.5	8	22	3000	1	4	2
7	6.00	2	1.0	6	38	2400	1	1	2
8	10.32	3	2.5	6	26	5400	2	5	2
9	7.20	4	2.0	7	30	3600	1	1	2
10	9.41	3	1.5	8	24	2900	1	4	1
11	6.00	3	1.0	7	31	2900	1	5	1
12	7.50	4	2.0	8	35	3100	1	1	2
13	10.50	4	2.5	9	29	5000	2	5	2
14	6.00	3	2.5	6	30	4058	2	5	2
15	6.00	3	1.0	7	35	2752	1	4	2
16	7.20	4	1.0	7	30	2900	1	1	2
17	7.20	3	1.0	8	40	1900	0	4	2
18	6.00	5	2.5	6	32	4000	1	5	2
19	10.00	3	2.0	7	38	3512	2	1	2
20	10.80	4	1.0	8	36	3100	0	1	2

Casa No.	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
1	2	1	1	0	0	1	1	1	1
2	1	1	1	0	0	1	1	1	1
3	2	1	1	0	0	1	1	1	1
4	2	1	1	0	0	1	0	0	0
5	2	1	1	1	1	1	1	1	1
6	2	1	1	0	0	1	1	1	1
7	2	1	1	0	0	1	1	1	1
8	2	1	1	0	0	1	1	0	0
9	2	1	1	0	0	1	1	0	0
10	1	1	1	0	0	0	1	1	1
11	1	1	1	0	0	1	1	1	0
12	1	1	1	0	0	1	1	0	0
13	2	1	1	0	1	1	1	0	0
14	2	1	1	0	0	1	1	1	1
15	2	1	1	0	0	1	0	1	1
16	2	1	0	0	0	1	1	0	0
17	3	1	1	0	0	1	0	0	0
18	2	3	1	0	0	1	1	1	1
19	3	2	1	0	1	1	1	1	1
20	2	1	1	0	1	1	0	1	1

a) ¿Cuántas dimensiones o indicadores esperaría tener como solución inicial en esta investigación?. Explique su respuesta tomando en cuenta los criterios para analizar el número de factores en la solución inicial.
b) Teniendo como supuesto que la solución inicial es correcta, ¿Qué variables se identifican más con cada factor?. Explique su respuesta tomando en cuenta los criterios para identificar cargas significativas y sus prioridades.
c) Según la información de la matriz rotada, ¿con cuál de los factores se identificó mejor el conjunto de 18 variables?. Justifique.
d) ¿Cuál de las variables ha sido mejor explicada en esta investigación por la solución de factores?. Justifique.

<u>Respuesta al inciso a.</u>

Se supone que ya se realizaron todos los pasos necesarios hasta antes de generar la matriz de correlación, la cual es la siguiente:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1.000								
X_2	0.164	1.000							
X_3	0.220	0.401	1.000						
X_4	0.267	0.040	-0.025	1.000					
X_5	-0.260	-0.242	-0.497	-0.308	1.000				
X_6	0.371	0.230	0.777	0.091	-0.514	1.000			
X_7	0.194	-0.158	0.593	0.111	-0.462	0.713	1.000		
X_8	-0.107	**0.000**	0.403	0.215	-0.505	0.481	0.335	1.000	
X_9	0.058	0.302	0.187	-0.306	0.281	0.022	0.037	-0.289	1.000

	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
X_{10}	1.000								
X_{11}	0.253	1.000							
X_{12}	-0.043	0.072	1.000						
X_{13}	0.043	-0.772	0.053	1.000					
X_{14}	0.325	0.105	0.115	0.459	1.000				
X_{15}	0.383	0.072	-0.053	0.053	0.115	1.000			

Número de factores

La selección del número de factores se basa con frecuencia en la proporción de varianza explicada por los factores, el conocimiento del tema y la sensatez de la solución. Inicialmente, se trata de utilizar el método de extracción de los componentes principales especificando el número de componentes igual al número de variables. Se examina la proporción de la variabilidad explicada por diferentes

factores y se reduce su selección a cuántos factores se utilizarán. Una gráfica de sedimentación puede ser útil aquí para evaluar visualmente la importancia de los factores. Una vez que se haya reducido esta selección, se examina los ajustes de los diferentes análisis factorial. Los valores de comunalidad y la proporción de la variabilidad de cada variable explicada por los factores que pueden ser de especial utilidad para comparar los ajustes. Se puede decidir agregar un factor si contribuye al ajuste de ciertas variables. Se puede tratar también con el método de extracción de máxima verosimilitud.

X_{16}	-0.325	0.157	-0.115	0.115	-0.062	-0.115	1.000		
X_{17}	-0.136	0.231	0.313	0.168	0.105	-0.168	0.157	1.000	
X_{18}	0.038	0.257	0.281	0.187	0.153	-0.187	0.102	**0.899**	1.000

A partir de la matriz de correlación, se puede identificar que las variables con mayor relación son la 8 con la 18 es decir X_8= Estilo arquitectónico y X_{18}= Tina de hidromasaje, con una correlación de **0.899**, mientras las que tienen menor relación son la 2 con la 8 o sea X_2= Número de recámaras con X_8= Estilo arquitectónico, con una correlación de **0.000**. Sin embargo, a partir de la matriz de correlación, es muy difícil encontrar a las variables afines que se puedan agrupar en factores y resumir así la información. Se necesita verificar entonces la matriz de factores no rotada que proporcione mayor información por lo que se corrió un análisis de factores y se obtuvo la siguiente matriz:

Cargas de factores no rotados :

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
X_1	0.385	-0.210	0.258	-0.405	-0.645	-0.102	0.304	-0.053	-0.065
X_2	0.275	-0.368	0.154	0.392	-0.463	0.140	-0.582	-0.089	-0.033
X_3	0.821	-0.247	-0.091	0.303	0.008	-0.194	0.023	-0.175	-0.063
X_4	0.314	0.301	0.127	-0.639	-0.132	0.157	-0.436	-0.084	0.055
X_5	-0.727	-0.398	-0.240	-0.176	0.125	-0.122	0.029	-0.079	0.326
X_6	0.891	-0.199	0.056	-0.007	0.122	-0.162	0.042	0.028	0.037
X_7	0.779	-0.087	-0.087	-0.072	0.297	0.112	0.374	0.154	0.136
X_8	0.569	0.360	-0.174	0.082	0.259	-0.368	-0.282	0.233	-0.224
X_9	-0.050	-0.794	-0.217	0.081	0.029	0.216	0.037	-0.215	-0.368
X_{10}	-0.160	-0.772	-0.216	-0.166	0.040	-0.185	0.156	0.304	-0.215
X_{11}	0.175	-0.331	-0.422	0.462	-0.312	-0.086	-0.152	0.360	0.416
X_{12}	0.138	0.173	-0.453	-0.140	-0.027	-0.676	0.033	-0.444	0.119
X_{13}	0.304	0.004	-0.407	-0.438	0.437	0.358	-0.238	0.073	0.002
X_{14}	0.341	-0.383	-0.317	-0.685	-0.245	0.167	-0.013	-0.002	0.169
X_{15}	-0.095	-0.629	-0.043	0.043	0.466	-0.069	-0.297	-0.226	0.090
X_{16}	0.430	0.137	-0.135	0.440	0.062	0.538	0.267	-0.294	0.164
X_{17}	-0.139	0.383	-0.846	0.079	-0.159	0.092	-0.001	0.002	-0.063
X_{18}	-0.108	0.212	-0.868	0.041	-0.248	0.112	0.034	-0.007	-0.217
Eigenvalue	3.7604	2.7844	2.4448	1.9917	1.4936	1.2985	1.111	0.7414	0.6717
% Var	0.209	0.155	0.136	0.111	0.083	0.072	0.062	0.041	0.037

Variable	Factor 10	Factor 11	Factor 12	Factor 13	Factor 14	Factor 15	Factor 16	Factor 17	Factor 18
X_1	0.090	0.047	0.130	-0.015	-0.003	0.122	-0.083	0.021	0.003
X_2	-0.075	0.103	0.005	0.012	-0.034	-0.100	0.016	0.040	0.007
X_3	-0.123	-0.052	-0.167	-0.161	0.054	-0.087	-0.090	-0.026	-0.008
X_4	0.067	-0.348	-0.093	-0.004	0.092	0.033	0.019	0.006	-0.012
X_5	-0.108	0.050	0.118	-0.020	0.214	-0.067	-0.037	0.024	-0.006
X_6	0.072	0.288	-0.011	-0.057	0.106	0.081	0.097	0.007	-0.014
X_7	0.114	-0.090	-0.147	0.195	-0.002	-0.092	-0.017	0.037	0.006
X_8	0.055	-0.071	0.324	0.055	0.073	-0.034	-0.026	-0.003	0.007
X_9	-0.106	-0.067	0.048	0.247	0.050	0.042	0.008	-0.015	-0.010

X_{10}	-0.060	-0.229	0.023	-0.200	-0.074	-0.025	0.048	0.021	-0.002
X_{11}	-0.55	-0.119	-0.016	0.083	-0.015	0.117	-0.009	-0.019	-0.002
X_{12}	-0.181	-0.073	-0.006	0.068	-0.100	0.022	0.041	0.011	0.005
X_{13}	-0.356	0.150	-0.004	-0.047	-0.077	0.097	-0.049	0.012	0.003
X_{14}	0.085	0.099	0.104	0.005	-0.047	-0.138	0.031	-0.041	0.008
X_{15}	0.458	-0.005	-0.013	-0.049	-0.056	0.075	-0.033	0.001	0.008
X_{16}	-0.022	-0.181	0.240	-0.121	0.003	0.021	0.032	0.005	0.003
X_{17}	0.227	0.083	0.041	-0.002	-0.098	-0.049	-0.029	0.016	-0.027
X_{18}	0.106	0.026	-0.16	-0.059	0.151	0.046	0.019	0.003	0.022
Eigenvalue	0.5236	0.3840	0.2932	0.2063	0.1371	0.1101	0.0382	0.0080	0.0020
% Var	0.0290	0.021	0.016	0.011	0.008	0.006	0.002	0.000	0.000

Esta matriz no rotada muestra las cargas de factor que, como ya se mencionó, indican la relación de cada variable con cada factor .
Aquí la mejor solución en cuanto al número de factores sería un factor que agrupara a las diez y ocho variables. Todas tendrían algo en común entre sí, mientras que la peor solución serían diez y ocho factores, lo cual representaría que las variables no tuvieron nada en común y no se podría resumir nada la información.
Tanto por el criterio de la raíz latente como por el porcentaje de variación explicada acumulada, se tiene como solución inicial siete factores, ya que el *Eigenvalue* de los siete primeros es mayor a 1 y el porcentaje de variación explicada acumulada está entre 60 % y 95%, es de 82.70%
En este problema de diez y ocho variables la varianza ó variación máxima que se podría obtener, se tiene que F1 explica 3.7604, el F2 explica 2.7844, el F3 explica 2.4448, el F4 explica 1.9917, el F5 explica 1.4936, el F6 explica 1.2985 y el F7 explica 1.1111 de acuerdo con su *Eigenvalue*. El total de la variación es de 13.862 (total de los *Eigenvalue* de los 7 factores y se pierde 4.138 (18 – 13.862) de información.

Rotación

Una vez que se haya seleccionado el número de factores, probablemente usted se querrá probar con diferentes rotaciones. Johnson y Wichern sugieren la rotación varimax. Un resultado similar de diferentes métodos puede dar crédito a la solución que se escogió. En este punto se pudiera interpretar los factores utilizando su conocimiento de los datos.

Respuesta al inciso b.

Una vez se ha determinado el número de factores para la solución inicial se trata de agrupar a las variables en esos siete factores, lo cual se determina analizando la matriz de factores rotada, que es la siguiente:

Cargas de factores rotados y comunalidades

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Comunalidad
X_1	0.117	0.020	0.174	0.057	**-0.931**	-0.087	-0.053	0.943
X_2	0.078	-0.036	0.129	0.026	-0.063	**-0.955**	0.144	**0.962**
X_3	**0.853**	-0.072	0.019	-0.077	-0.062	-0.338	-0.130	0.874
X_4	-0.034	0.385	0.134	**0.783**	-0.219	-0.066	-0.113	0.846
X_5	**-0.583**	**-0.620**	-0.095	-0.077	0.107	0.197	-0.130	0.807
X_6	**0.865**	-0.026	0.204	0.160	-0.146	-0.140	-0.145	0.879
X_7	**0.865**	-0.055	0.010	0.159	-0.108	0.246	0.132	0.867
X_8	**0.532**	0.354	-0.065	0.180	0.305	-0.089	-0.477	0.772
X_9	0.072	**-0.783**	-0.046	-0.005	-0.038	-0.215	0.259	0.735
X_{10}	-0.024	**-0.837**	0.023	-0.53	-0.179	0.016	-0.136	0.756

X_{11}	0.214	-0.267	-0.394	-0.232	-0.001	**-0.573**	-0.066	0.659
X_{12}	0.169	0.014	-0.332	-0.045	-0.058	0.121	**-0.758**	0.733
X_{13}	0.262	-0.152	-0.214	**0.765**	0.248	0.188	0.080	0.825
X_{14}	0.167	-0.381	-0.175	**0.641**	**-0.551**	-0.013	-0.042	0.920
X_{15}	0.035	**-0.662**	0.268	0.124	0.408	-0.141	-0.071	0.719
X_{16}	**0.526**	0.190	-0.285	-0.115	0.110	-0.039	**0.599**	0.779
X_{17}	-0.103	0.114	**-0.932**	0.060	0.127	0.041	-0.096	0.922
X_{18}	-0.079	-0.029	**-0.933**	0.069	-0.000	-0.019	-0.067	0.886
Eigenvalue	3.3582	2.7105	2.3551	1.8038	1.6571	1.6134	1.3862	14.8844
% Var	0.187	0.151	0.131	0.100	0.092	0.090	0.077	0.827

Ni el *Eigenvalue* ni la comunalidad total, ni el porcentaje de variación explicada total cambian al rotar los factores, es decir, no cambian de la matriz no rotada a la rotada, **solo se distribuyeron las cargas de factor** y ahora sí algunas cargas tienen valores más grandes que en la no rotada como para poder determinar la forma como se agrupan las variables en cada factor.
Lo que si cambia es la importancia en que cada factor explica a las variables (*Eigenvalue* de cada factor) y en la matriz rotada es posible que cada factor hacia la derecha explique menos a las variables.
Para determinar en cual factor se agrupara cada variable debe analizarse la matriz rotada, buscando las cargas significativas. Al identificar las cargas significativas, no importa el signo que tengan, sino la magnitud de la relación, el signo negativo solo expresa una relación inversa. En principio se buscan cargas mayores a 0.50 utilizando la Prioridad 1, en caso de que una variable no tuviera una carga mayor a 0.5 se aplicarían las prioridades 2, 3 o 4, según sea el caso, de los criterios de cargas significativas en toda la matriz rotada. En este ejemplo se tiene:

X_1= Tamaño del lote se identifica con el F5 con una carga de 0.931
X_2= Número de recámaras se identifica con el F6 con una carga de 0.955
X_3= Número de baños se identifica con el F1 con una carga de 0.853
X_4= Número de habitaciones se identifica con el F4 con una carga de 0.783
X_5= Antigüedad de la casa se identifica con el **F1 y el F2** con unas cargas de **0.583 y 0.620**
X_6= Impuestos anuales se identifica con el F1 con una carga de 0.865
X_7= Tipo de estacionamiento interior se identifica con el F1 con una carga de 0.865
X_8= Estilo arquitectónico se identifica con el F1 con una carga de 0.532
X_9= Tipo de combustible para la calefacción se identifica con el F2 con una carga de 0.783
X_{10}= Tipo de sistema para la calefacción se identifica con el F2 con una carga de 0.837
X_{11}= Tipo de alberca se identifica con el F6 con una carga de 0.573
X_{12}= Cocina con antecomedor se identifica con el F7 con una carga de 0.758
X_{13}= Aire acondicionado se identifica con el F4 con una carga de 0.765
X_{14}= Chimenea se identifica con el **F4 y el F5** con unas cargas de **0.641 y 0.551**
X_{15}= Drenaje municipal se identifica con el F2 con una carga de 0.662

X_{16}= Sótano se identifica con el **F1 y el F7** con unas cargas de **0.526 y 0.599**
X_{17}= Cocina integral se identifica con el F3 con una carga de 0.932
X_{18}= Tina de hidromasaje se identifica con el F3 con una carga de 0.933

En caso de que alguna variable tuviera carga significativa con dos o más factores, se debe analizar en la matriz de correlación con cuales variables tiene mayor relación. En este caso **X_5= Antigüedad de la casa** se identifica con el **F1 y el F2** con unas cargas de 0.583 y 0.620, **X_{14}= Chimenea** se identifica con el **F4 y el F5** con unas cargas de 0.641 y 0.551 y **X_{16}= Sótano se identifica** con el **F1 y el F7** con unas cargas de 0.526 y 0.599; entonces se debe checar la correlación con todas las combinaciones que se pueden hacer con la variable X_5= Antigüedad, X_{14}= Chimenea y X_{16}= Sótano y las que forman con el F1 y F2, con el F4 y F5 y F1 y F7 respectivamente:

Para el caso X_5= Antigüedad de la casa se identifica con el F1 y el F2 con unas cargas de 0.583 y 0.620 tenemos:

X_5= Antigüedad VS X_3= Número de baños tienen una correlación de 0.497

X_5= Antigüedad VS X_6= Impuestos anuales tienen una correlación de **0.514**

X_5= Antigüedad VS X_7= Tipo de estacionamiento interior tienen una correlación de 0.462

X_5= Antigüedad VS X_8= Estilo arquitectónico tienen una correlación de 0.505

X_5= Antigüedad VS X_9= Tipo de combustible tienen una correlación de 0.281

X_5= Antigüedad VS X_{10}= Tipo de sistema para la calef. tienen una correlación de 0.429

X_5= Antigüedad VS X_{15}= Drenaje municipal tienen una correlación de 0.361

X_5= Antigüedad VS X_{16}= Sótano tienen una correlación de 0.364

Como la máxima correlación que se obtuvo es con X_6= Impuestos anuales, entonces la variable X_5= Antigüedad debe ir en el Factor 1.

Para el caso X_{14}= Chimenea se identifica con el F4 y el F5 con unas cargas de 0.641 y 0.551 tenemos:

X_{14}= Chimenea VS X_1= Tamaño del lote tienen una correlación de **0.539**

X_{14}= Chimenea VS X_4= Número de habitaciones tienen una correlación de 0.415

X_{14}= Chimenea VS X_{13}= Aire acondicionado tienen una correlación de 0.459

Como la máxima correlación que se obtuvo es con X_1= Tamaño del lote, entonces la variable X_{14}= Chimenea debe ir en el Factor 5.

Para el caso de X_{16}= Sótano se identifica con el F1 y el F7 con unas cargas de 0.526 y 0.599 tenemos:

X_{16}= Sótano VS X_3= Número de baños tienen una correlación de 0.395

X_{16}= Sótano VS X_4= Número de habitaciones tienen una correlación de 0.087

X_{16}= Sótano VS X_5= Antigüedad tienen una correlación de 0.364

X_{16}= Sótano VS X_6= Impuestos anuales tienen una correlación de 0.230

X_{16}= Sótano VS X_7= Tipo de estacionamiento interior tienen una correlación de **0.411**

X_{16}= Sótano VS X_8= Estilo arquitectónico tienen una correlación de 0.072
X_{16}= Sótano VS X_{12}= Cocina con antecomedor tienen una correlación de 0.115
Como la máxima correlación que se obtuvo es con X_7= Tipo de estacionamiento interior, X_{16}= Sótano debe ir en el Factor 1.
La solución final óptima sería:
F1: X_3= Número de baños, X_5= Antigüedad, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico y X_{16}= Sótano
F2: X_9= Tipo de combustible, X_{10}= Tipo de sistema para la calefacción y X_{15}= Drenaje municipal
F3: X_{17}= Cocina integral y X_{18}= Tina de hidromasaje
F4: X_4= Número de habitaciones y X_{13}= Aire acondicionado
F5: X_1= Tamaño del lote y X_{14}= Chimenea
F6: X_2= Número de recámaras y X_{11}= Tipo de alberca
F7: X_{12}= Cocina con antecomedor

Respuesta al inciso c.

Los valores de comunalidad y la proporción de la variabilidad de cada variable explicada por los factores que pueden ser de especial utilidad para comparar los ajustes. Se puede decidir agregar un factor si contribuye al ajuste de ciertas variables. Se puede tratar también con el método de extracción de máxima verosimilitud.

El porcentaje de varianza, que es el porcentaje de variación explicada con respecto a la máxima variación que se podría tener, se obtiene, por ejemplo para F1 dividiendo 3.7604/18, lo que es igual a 0.209 o 20.9 %. El Factor que ha explicado más al conjunto de variables es el 1, y después el 2 y así sucesivamente lo cual es lógico pues cada factor adicional explica menor cantidad, ya que cada vez queda menos por explicar, en este caso, por ejemplo, queda por explicar 4.138 de variación; si se agregan los 11 factores que faltan, ya cada uno explicaría muy poco y no convendría. El porcentaje de variación explicada acumulada total es de 0.827 o de 82.70 % de un máximo por explicar de 100%.

Respuesta al inciso d.

Se puede observar que la variable mejor explicada por el conjunto de factores es la variable X_2=Número de Recámaras, ya que es la que tiene la mayor comunalidad de 0.962

2.1.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE APRENDIZAJE 2.1.1.1 FACTOR ANÁLISIS

El Director de mercadotecnia de una compañía de Bienes Raíces desea conocer cuales dimensiones o indicadores básicos influyen al momento de pronosticar el valor de una casa o vivienda unifamiliar al momento en que un potencial cliente desea ya sea comprar o vender su casa con el objeto de brindar un servicio confiable y rápido. Para llevar a cabo la investigación dividió la Ciudad en tres grandes zonas: NORTE, SUR Y ESTE. En cada zona se aplicó una encuesta a 20

propietarios de casas unifamiliares. El cuestionario estaba conformado por 20 variables que se consideraba podían formar parte de dichos indicadores X_1= Tamaño del lote y X_2= Número de recámaras, , X_3= Número de baños, X_4= Número de habitaciones y X_5= Antigüedad de la casa, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico, X_9= Tipo de combustible para la calefacción, X_{10}= Tipo de sistema para la calefaccción, X_{11}= Tipo de alberca, X_{12}= Cocina con antecomedor y X_{13}= Aire acondicionado, X_{14}= Chimenea, X_{15}= Drenaje municipal X_{16}= Sótano, X_{17}= Cocina integral, X_{18}= Tina de hidromasaje.

Los resultados en la zona SUR fueron los siguientes:

Casa No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	31.5	5	2.0	6	31	4200	0
2	6.8	4	2.5	8	10	4980	1
3	6.0	4	1.5	10	30	2857	1
4	9.0	3	1.5	6	25	2400	1
5	7.0	3	1.5	7	28	3959	1
6	6.5	2	1.0	5	35	2250	1
7	9.1	3	1.5	5	30	4134	1
8	9.8	4	1.5	7	60	3142	2
9	6.0	3	1.0	8	30	2500	1
10	6.0	4	1.0	7	34	2500	1
11	10.0	4	2.0	7	34	3000	1
12	10.0	3	1.5	6	30	3930	1
13	7.0	4	2.0	7	35	3100	1
14	12.1	3	2.0	7	46	3640	0
15	8.5	3	1.0	7	35	2725	1
16	7.0	5	1.5	6	28	4100	1
17	6.0	3	1.0	6	39	2378	0
18	11.0	3	2.0	7	33	2767	0
19	10.0	4	2.0	6	28	4500	2
20	9.9	3	1.5	8	30	3871	1

Casa No.	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
1	2	1	1	0	0	1	1
2	2	1	1	0	0	0	1
3	1	1	1	0	1	1	0
4	2	1	1	0	1	0	1
5	2	1	1	0	0	1	1
6	2	1	0	0	0	1	1
7	2	1	1	0	1	1	1
8	2	1	1	0	0	0	1
9	2	1	1	0	0	1	1
10	2	1	1	0	0	1	1
11	2	1	1	0	0	1	1
12	2	1	1	1	0	1	1
13	2	1	0	0	1	1	1
14	2	1	1	1	1	1	1
15	2	2	1	0	0	1	1
16	1	1	1	0	1	1	1
17	1	2	1	0	0	0	1
18	1	2	1	1	1	1	1
19	2	1	1	1	0	1	1
20	2	1	1	0	0	1	1

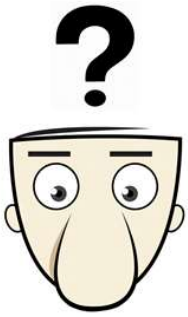
a) ¿Cuántas dimensiones o indicadores esperaría tener como solución inicial en esta investigación?. Explique su respuesta tomando en cuenta los criterios para analizar el número de factores en la solución inicial.

b) Teniendo como supuesto que la solución inicial es correcta, ¿Qué variables se identifican más con cada factor?. Explique su respuesta tomando en cuenta los criterios para identificar cargas significativas y sus prioridades.
c) Según la información de la matriz rotada, ¿con cuál de los factores se identificó mejor el conjunto de 18 variables?. Justifique.
d) ¿Cuál de las variables ha sido mejor explicada en esta investigación por la solución de factores?. Justifique.

2.1.1.1 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra al final del cuaderno de trabajo en el anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 2.1.1.1 FACTOR ANÁLISIS



El Director de mercadotecnia de una compañía de Bienes Raíces desea conocer cuales dimensiones o indicadores básicos influyen al momento de pronosticar el valor de una casa o vivienda unifamiliar al momento en que un potencial cliente desea ya sea comprar o vender su casa con el objeto de brindar un servicio confiable y rápido. Para llevar a cabo la investigación dividió la Ciudad en tres grandes zonas: NORTE, SUR Y ESTE. En cada zona se aplicó una encuesta a 20 propietarios de casas unifamiliares. El cuestionario estaba conformado por 20 variables que se consideraba podían formar parte de dichos indicadores X_1= Tamaño del lote y X_2= Número de recámaras, , X_3= Número de baños, X_4= Número de habitaciones y X_5= Antigüedad de la casa, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico, , X_9= Tipo de sistema para la calefaccción, X_{10}= Tipo de alberca, X_{11}= Cocina con antecomedor, X_{12}= Chimenea, X_{13}= Cocina integral, X_{14}= Tina de hidromasaje.
Los resultados en la zona ESTE fueron los siguientes:

Casa No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	6.9	4	2.0	8	38	3750	1
2	6.0	3	2.0	6	35	3240	0
3	7.0	3	1.0	6	39	2700	1
4	6.0	4	1.0	7	38	2800	1
5	6.0	3	2.0	7	38	2995	1
6	9.0	4	1.0	6	32	2500	1
7	11.2	4	1.0	9	32	2800	1
8	6.0	5	2.0	10	35	2500	0
9	5.6	3	1.0	6	37	2027	0

10	6.0	3	1.5	8	37	2500	1
11	6.0	4	1.0	6	39	2200	1
12	6.0	4	2.0	7	37	3156	0
13	8.7	4	1.0	6	36	2741	1
14	6.0	4	1.0	6	39	2100	0
15	6.5	3	2.0	7	36	2400	0
16	6.5	3	1.0	6	35	2280	0
17	6.0	3	2.0	8	38	2700	1
18	6.0	3	1.5	7	35	2300	1
19	6.0	3	1.0	7	36	1889	0
20	6.0	4	1.0	6	35	2900	0

Casa No.	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	3	2	1	1	0	0	0
2	1	2	1	1	1	1	1
3	1	1	1	1	0	0	0
4	1	2	1	1	0	1	1
5	4	2	1	1	1	1	1
6	4	2	1	1	1	0	0
7	1	2	1	1	0	0	0
8	1	2	1	1	0	0	0
9	1	2	1	1	1	1	1
10	2	2	1	1	1	1	1
11	1	2	1	1	0	1	1
12	1	2	1	1	1	1	1
13	1	2	1	1	1	0	0
14	1	2	1	1	0	0	0
15	1	2	2	1	0	1	1
16	1	2	1	0	0	0	1
17	1	2	1	1	0	1	1
18	4	2	1	1	1	1	1
19	1	2	1	1	0	1	1
20	4	2	1	1	1	1	1

a) ¿Cuántas dimensiones o indicadores esperaría tener como solución inicial en esta investigación?. Explique su respuesta tomando en cuenta los criterios para analizar el número de factores en la solución inicial.
b) Teniendo como supuesto que la solución inicial es correcta, ¿Qué variables se identifican más con cada factor?. Explique su respuesta tomando en cuenta los criterios para identificar cargas significativas y sus prioridades.
c) Según la información de la matriz rotada, ¿con cuál de los factores se identificó mejor el conjunto de 14 variables?. Justifique.
d) ¿Cuál de las variables ha sido mejor explicada en esta investigación por la solución de factores?. Justifique.

2.1.1.1.F.A. EJEMPLO ILUSTRATIVO EN MINITAB 17

EJEMPLO ILUSTRATIVO INTEGRAL EN MINITAB 17. 2.1.1.1. FACTOR ANÁLISIS.

El Director de mercadotecnia de una compañía de Bienes Raíces desea conocer cuales dimensiones o indicadores básicos influyen al momento de pronosticar el valor de una casa o vivienda unifamiliar al momento en que un potencial cliente desea ya sea comprar o vender su casa con el objeto de brindar un servicio confiable y rápido. Para llevar a cabo la investigación dividió la Ciudad en tres grandes zonas: NORTE, SUR Y ESTE. En cada zona se aplicó una encuesta a 20 propietarios de casas unifamiliares. El cuestionario estaba conformado por 20 variables que se consideraba podían formar parte de dichos indicadores X_1= Tamaño del lote y X_2= Número de recámaras, , X_3= Número de baños, X_4= Número de habitaciones y X_5= Antigüedad de la casa, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico, X_9= Tipo de combustible para la calefacción, X_{10}= Tipo de sistema para la calefaccción, X_{11}= Tipo de alberca, X_{12}= Cocina con antecomedor y X_{13}= Aire acondicionado, X_{14}= Chimenea, X_{15}= Drenaje municipal X_{16}= Sótano, X_{17}= Cocina integral, X_{18}= Tina de hidromasaje.

Los resultados en la zona NORTE fueron los siguientes:

Casa No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	6.00	3	1.0	6	45	1600	0	1	2
2	7.00	3	1.5	10	28	3500	1	5	1
3	7.00	4	2.0	6	35	3600	0	4	2
4	7.00	3	2.0	7	35	4100	1	3	1
5	6.00	3	2.0	9	32	4400	2	5	2
6	7.75	4	2.5	8	22	3000	1	4	2
7	6.00	2	1.0	6	38	2400	1	1	2
8	10.32	3	2.5	6	26	5400	2	5	2
9	7.20	4	2.0	7	30	3600	1	1	2
10	9.41	3	1.5	8	24	2900	1	4	1
11	6.00	3	1.0	7	31	2900	1	5	1
12	7.50	4	2.0	8	35	3100	1	1	2
13	10.50	4	2.5	9	29	5000	2	5	2
14	6.00	3	2.5	6	30	4058	2	5	2
15	6.00	3	1.0	7	35	2752	1	4	2
16	7.20	4	1.0	7	30	2900	1	1	2
17	7.20	3	1.0	8	40	1900	0	4	2
18	6.00	5	2.5	6	32	4000	1	5	2
19	10.00	3	2.0	7	38	3512	2	1	2
20	10.80	4	1.0	8	36	3100	0	1	2

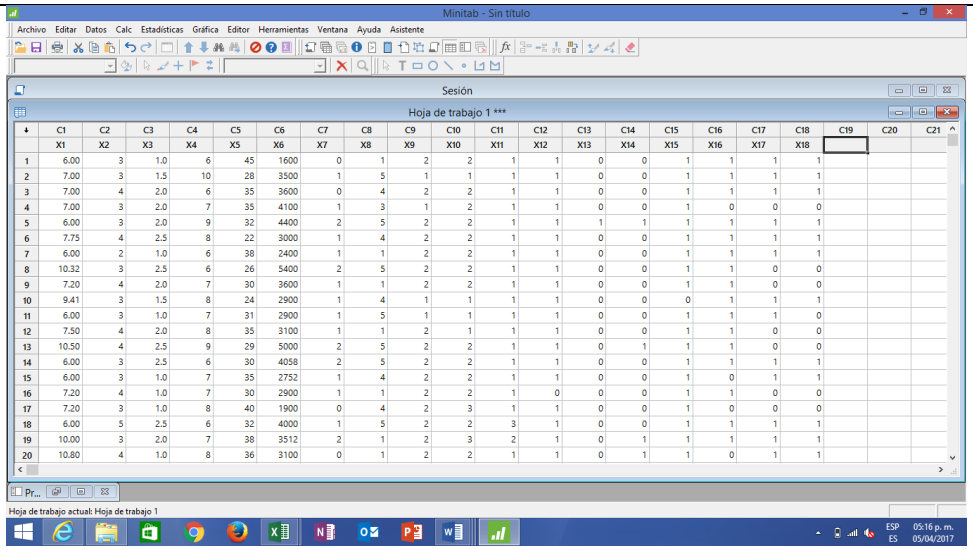
Casa No.	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
1	2	1	1	0	0	1	1	1	1
2	1	1	1	0	0	1	1	1	1
3	2	1	1	0	0	1	1	1	1
4	2	1	1	0	0	1	0	0	0
5	2	1	1	1	1	1	1	1	1
6	2	1	1	0	0	1	1	1	1
7	2	1	1	0	0	1	1	1	1
8	2	1	1	0	0	1	1	0	0
9	2	1	1	0	0	1	1	0	0
10	1	1	1	0	0	0	1	1	1
11	1	1	1	0	0	1	1	1	0
12	1	1	1	0	0	1	1	0	0
13	2	1	1	0	1	1	1	0	0
14	2	1	1	0	0	1	1	1	1
15	2	1	1	0	0	1	0	1	1
16	2	1	0	0	0	1	1	0	0
17	3	1	1	0	0	1	0	0	0
18	2	3	1	0	0	1	1	1	1
19	3	2	1	0	1	1	1	1	1
20	2	1	1	0	1	1	0	1	1

a) ¿Cuántas dimensiones o indicadores esperaría tener como solución inicial en esta investigación?. Explique su respuesta tomando en cuenta los criterios para analizar el número de factores en la solución inicial.
b) Teniendo como supuesto que la solución inicial es correcta, ¿Qué variables se identifican más con cada factor?. Explique su respuesta tomando en cuenta los criterios para identificar cargas significativas y sus prioridades.
c) Según la información de la matriz rotada, ¿con cuál de los factores se identificó mejor el conjunto de 18 variables?. Justifique.
d) ¿Cuál de las variables ha sido mejor explicada en esta investigación por la solución de factores?. Justifique.

Respuesta al inciso a)

Cuando el número de observaciones **es extenso y/o existen muchas variables,** los cálculos manuales son tediosos. Existen muchos paquetes de software que pueden mostrar los resultados entre ellos **Minitab.**

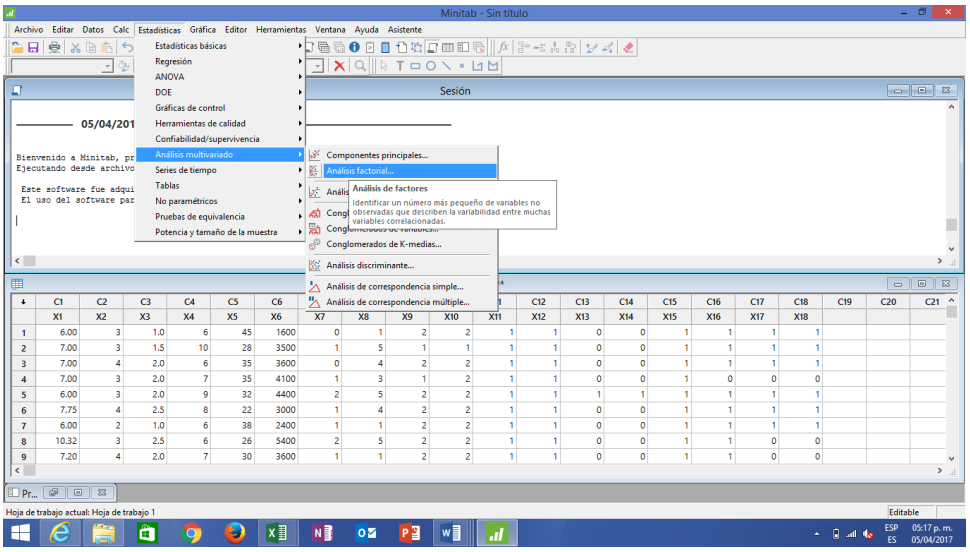
Comenzamos introduciendo los **datos** en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:



Hay tres maneras en las que usted puede llevar un análisis factorial en Minitab. La manera habitual, descrita a continuación, es ingresar columnas que contengan sus variables de medición.

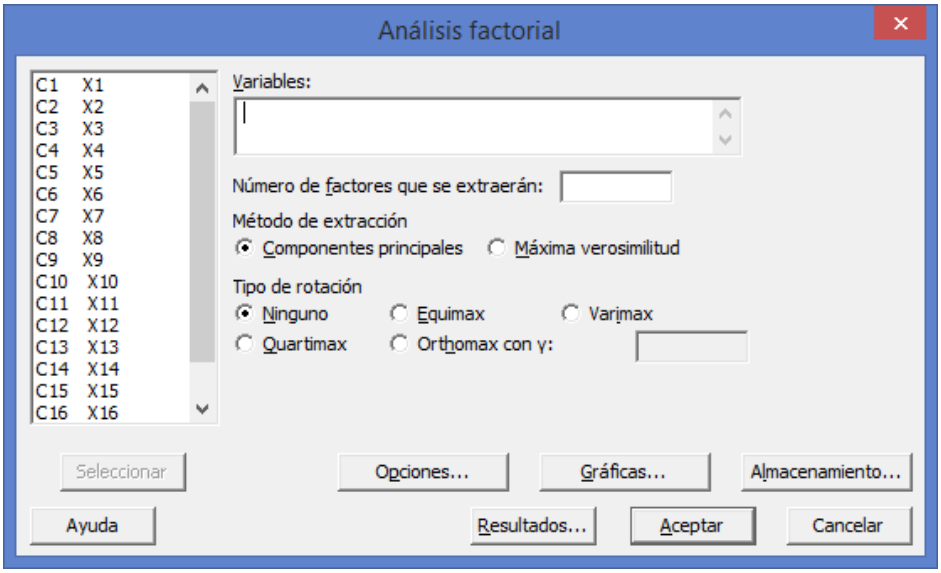
Para generar la matriz no rotada para visualizar la solución inicial elija **Estadísticas > Análisis multivariado > Análisis factorial**.

Aparece el siguiente cuadro de Diálogo:

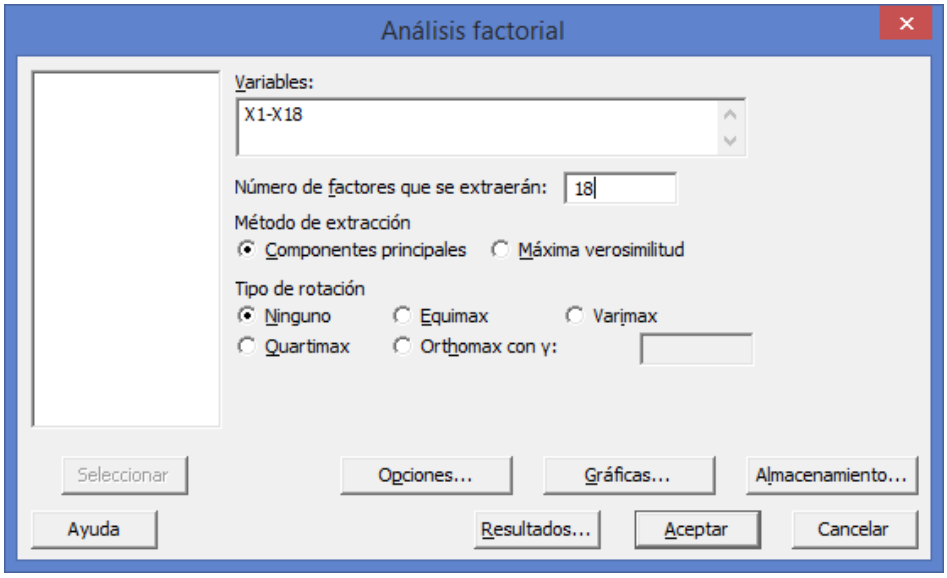


En **Variables**, ingrese las columnas que contienen los datos de medición seleccionando las 18 variables y oprimir el botón **Seleccionar**

Aparece el siguiente cuadro de Diálogo:



En **Número de factores que se extraerán** coloque **18** que corresponde al número de variables



En **Método de extracción** debe estar seleccionada por default la opción **Componentes principales** y en **Tipo de rotación** debe estar seleccionada por default la opción **Ninguno** y luego haga clic en **Aceptar**.

Salida de la ventana Sesión

Análisis factorial: X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14, X15, X16, X17, X18

Cargas de factores no rotados :

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
X_1	0.385	-0.210	0.258	-0.405	-0.645	-0.102	0.304	-0.053	-0.065
X_2	0.275	-0.368	0.154	0.392	-0.463	0.140	-0.582	-0.089	-0.033
X_3	0.821	-0.247	-0.091	0.303	0.008	-0.194	0.023	-0.175	-0.063
X_4	0.314	0.301	0.127	-0.639	-0.132	0.157	-0.436	-0.084	0.055
X_5	-0.727	-0.398	-0.240	-0.176	0.125	-0.122	0.029	-0.079	0.326
X_6	0.891	-0.199	0.056	-0.007	0.122	-0.162	0.042	0.028	0.037
X_7	0.779	-0.087	-0.087	-0.072	0.297	0.112	0.374	0.154	0.136
X_8	0.569	0.360	-0.174	0.082	0.259	-0.368	-0.282	0.233	-0.224
X_9	-0.050	-0.794	-0.217	0.081	0.029	0.216	0.037	-0.215	-0.368
X_{10}	-0.160	-0.772	-0.216	-0.166	0.040	-0.185	0.156	0.304	-0.215
X_{11}	0.175	-0.331	-0.422	0.462	-0.312	-0.086	-0.152	0.360	0.416
X_{12}	0.138	0.173	-0.453	-0.140	-0.027	-0.676	0.033	-0.444	0.119
X_{13}	0.304	0.004	-0.407	-0.438	0.437	0.358	-0.238	0.073	0.002
X_{14}	0.341	-0.383	-0.317	-0.685	-0.245	0.167	-0.013	-0.002	0.169
X_{15}	-0.095	-0.629	-0.043	0.043	0.466	-0.069	-0.297	-0.226	0.090
X_{16}	0.430	0.137	-0.135	0.440	0.062	0.538	0.267	-0.294	0.164
X_{17}	-0.139	0.383	-0.846	0.079	-0.159	0.092	-0.001	0.002	-0.063
X_{18}	-0.108	0.212	-0.868	0.041	-0.248	0.112	0.034	-0.007	-0.217
Eigenvalue	3.7604	2.7844	2.4448	1.9917	1.4936	1.2985	1.111	0.7414	0.6717
% Var	0.209	0.155	0.136	0.111	0.083	0.072	0.062	0.041	0.037

Variable	Factor 10	Factor 11	Factor 12	Factor 13	Factor 14	Factor 15	Factor 16	Factor 17	Factor 18
X_1	0.090	0.047	0.130	-0.015	-0.003	0.122	-0.083	0.021	0.003
X_2	-0.075	0.103	0.005	0.012	-0.034	-0.100	0.016	0.040	0.007
X_3	-0.123	-0.052	-0.167	-0.161	0.054	-0.087	-0.090	-0.026	-0.008
X_4	0.067	-0.348	-0.093	-0.004	0.092	0.033	0.019	0.006	-0.012
X_5	-0.108	0.050	0.118	-0.020	0.214	-0.067	-0.037	0.024	-0.006
X_6	0.072	0.288	-0.011	-0.057	0.106	0.081	0.097	0.007	-0.014
X_7	0.114	-0.090	-0.147	0.195	-0.002	-0.092	-0.017	0.037	0.006
X_8	0.055	-0.071	0.324	0.055	0.073	-0.034	-0.026	-0.003	0.007
X_9	-0.106	-0.067	0.048	0.247	0.050	0.042	0.008	-0.015	-0.010
X_{10}	-0.060	-0.229	0.023	-0.200	-0.074	-0.025	0.048	0.021	-0.002
X_{11}	-0.55	-0.119	-0.016	0.083	-0.015	0.117	-0.009	-0.019	-0.002
X_{12}	-0.181	-0.073	-0.006	0.068	-0.100	0.022	0.041	0.011	0.005
X_{13}	-0.356	0.150	-0.004	-0.047	-0.077	0.097	-0.049	0.012	0.003
X_{14}	0.085	0.099	0.104	0.005	-0.047	-0.138	0.031	-0.041	0.008
X_{15}	0.458	-0.005	-0.013	-0.049	-0.056	0.075	-0.033	0.001	0.008
X_{16}	-0.022	-0.181	0.240	-0.121	0.003	0.021	0.032	0.005	0.003
X_{17}	0.227	0.083	0.041	-0.002	-0.098	-0.049	-0.029	0.016	-0.027

Número de factores

La selección del número de factores se basa con frecuencia en la proporción de varianza explicada por los

factores, el conocimiento del tema y la sensatez de la solución. Inicialmente, se trata de utilizar el método de extracción de los componentes principales especificando el número de componentes igual al número de variables. Se examina la proporción de la variabilidad explicada por diferentes factores y se reduce su selección a cuántos factores se utilizarán. Una gráfica de sedimentación puede ser útil aquí para evaluar visualmente la importancia de los factores. Una vez que se haya reducido esta selección, se examina los ajustes de los diferentes análisis factorial. Los valores de comunalidad y la proporción de la variabilidad de cada variable explicada por los factores que pueden ser de especial utilidad para comparar los ajustes. Se puede decidir agregar un factor si contribuye al ajuste de ciertas variables. Se puede tratar también con el método de extracción de máxima verosimilitud.

X_{18}	0.106	0.026	-0.16	-0.059	0.151	0.046	0.019	0.003	0.022
Eigenvalue	0.5236	0.3840	0.2932	0.2063	0.1371	0.1101	0.0382	0.0080	0.0020
% Var	0.0290	0.021	0.016	0.011	0.008	0.006	0.002	0.000	0.000

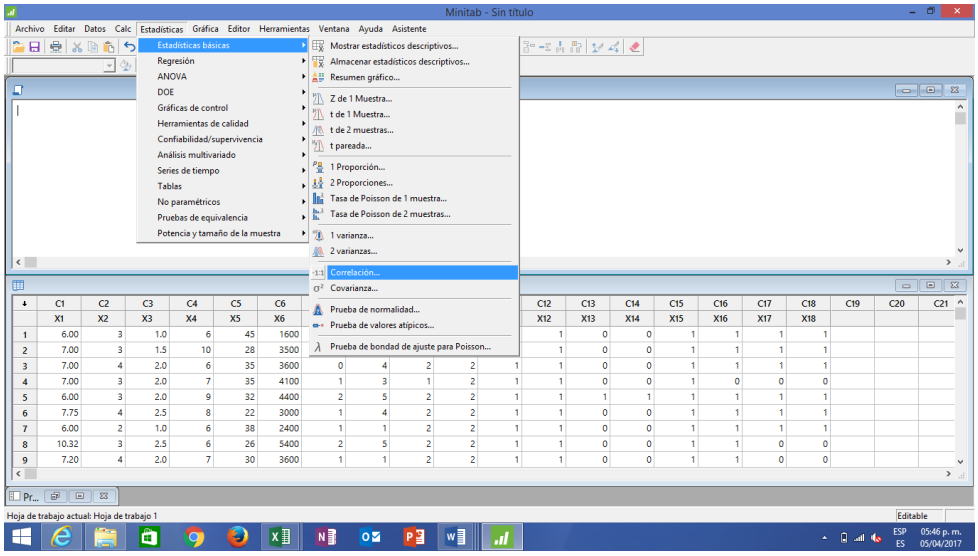
Esta matriz no rotada muestra las cargas de factor que, como ya se mencionó, indican la relación de cada variable con cada factor .
Aquí la mejor solución en cuanto al número de factores sería un factor que agrupara a las diez y ocho variables. Todas tendrían algo en común entre sí, mientras que la peor solución serian diez y ocho factores, lo cual representaría que las variables no tuvieron nada en común y no se podría resumir nada la información.

Tanto por el **criterio de la raíz latente como por el porcentaje de variación explicada acumulada**, se tiene **como solución inicial siete factores**, ya que el *Eigenvalue* de los siete primeros **es mayor a 1** y el porcentaje de variación explicada acumulada está entre 60 % y 95%, es de **82.70%**
En este problema de diez y ocho variables la varianza ó variación máxima que se podría obtener, se tiene que F1 explica 3.7604, el F2 explica 2.7844, el F3 explica 2.4448, el F4 explica 1.9917, el F5 explica 1.4936, el F6 explica 1.2985 y el F7 explica 1.1111 de acuerdo con su *Eigenvalue*. El total de la variación es de 13.862 (total de los *Eigenvalue* de los 7 factores y se pierde 4.138 (18 – 13.862) de información.

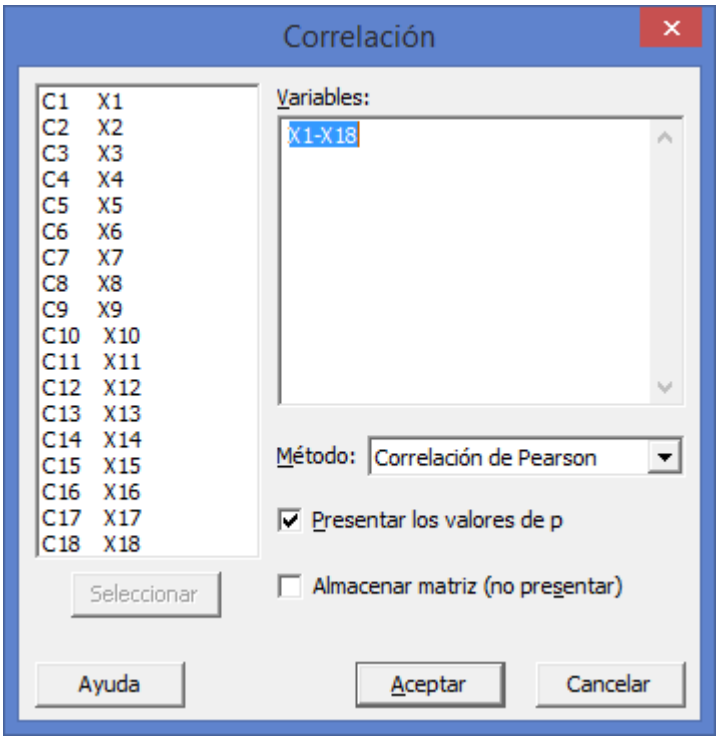
Respuesta al inciso b)

Para generar la **matriz de correlación** elija **Estadísticas > Estadísticas básicas > correlación**

Aparece el siguiente cuadro de Diálogo:



Aparece el siguiente cuadro de Diálogo:



Desactive las casilla **Presentar los valores de p** (quitar la palomita) y haga clic en **Aceptar**.

Salida de la ventana Sesión

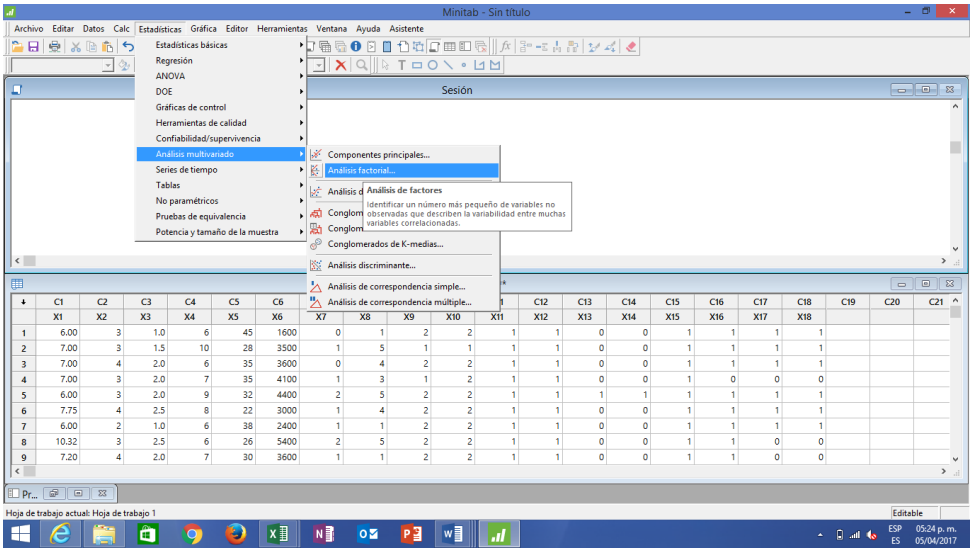
Matriz de correlación:

Correlación: X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14, X15, X16, X17, X18

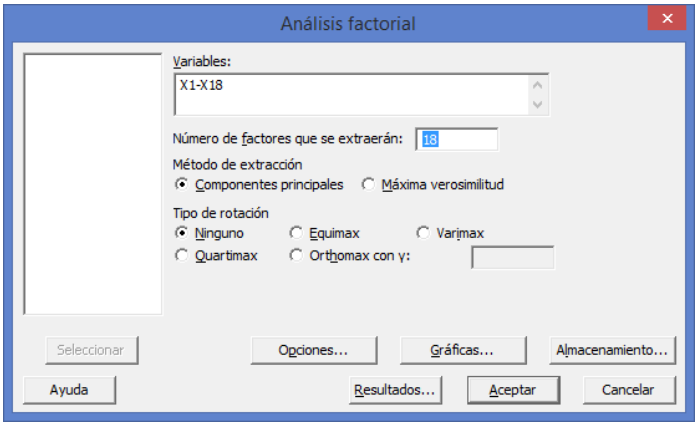
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1.000								
X_2	0.164	1.000							
X_3	0.220	0.401	1.000						
X_4	0.267	0.040	-0.025	1.000					
X_5	-0.260	-0.242	-0.497	-0.308	1.000				
X_6	0.371	0.230	0.777	0.091	-0.514	1.000			
X_7	0.194	-0.158	0.593	0.111	-0.462	0.713	1.000		
X_8	-0.107	**0.000**	0.403	0.215	-0.505	0.481	0.335	1.000	
X_9	0.058	0.302	0.187	-0.306	0.281	0.022	0.037	-0.289	1.000

	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
X_{10}	1.000								
X_{11}	0.253	1.000							
X_{12}	-0.043	0.072	1.000						
X_{13}	0.043	-0.772	0.053	1.000					
X_{14}	0.325	0.105	0.115	0.459	1.000				
X_{15}	0.383	0.072	-0.053	0.053	0.115	1.000			
X_{16}	-0.325	0.157	-0.115	0.115	-0.062	-0.115	1.000		
X_{17}	-0.136	0.231	0.313	0.168	0.105	-0.168	0.157	1.000	
X_{18}	0.038	0.257	0.281	0.187	0.153	-0.187	0.102	**0.899**	1.000

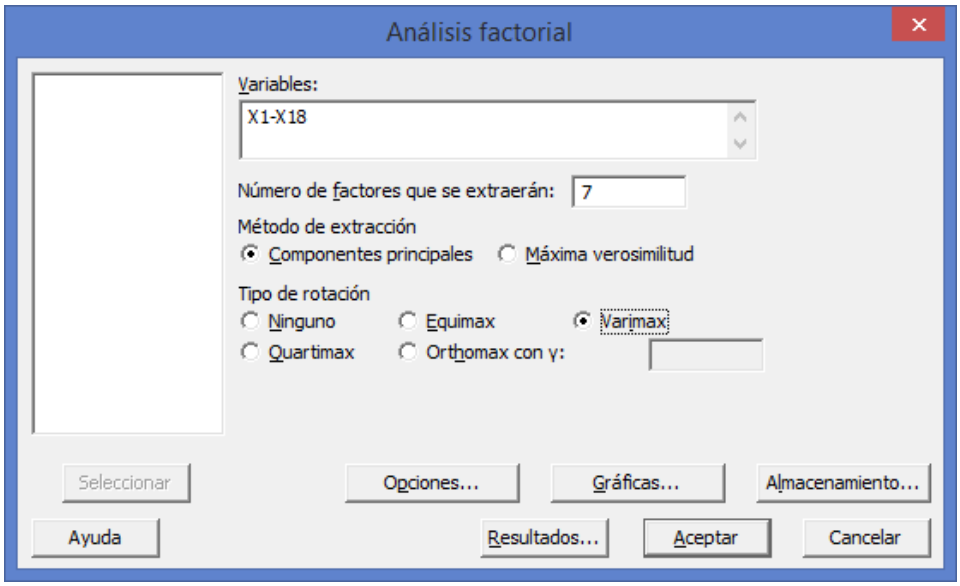
Para generar la **matriz rotada** elija **Estadísticas > Análisis Multivariado > Análisis factorial**



Aparece el siguiente cuadro de Diálogo:



En **Número de factores que se extraerán** colocar el número **7** y en **Tipo de rotación** seleccionar la opción **Varimax** y luego dar click en **Aceptar**



Salida de la ventana Sesión

Cargas de factores rotados y comunalidades

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Comunalidad
X_1	0.117	0.020	0.174	0.057	**-0.931**	-0.087	-0.053	0.943
X_2	0.078	-0.036	0.129	0.026	-0.063	**-0.955**	0.144	**0.962**
X_3	**0.853**	-0.072	0.019	-0.077	-0.062	-0.338	-0.130	0.874
X_4	-0.034	0.385	0.134	**0.783**	-0.219	-0.066	-0.113	0.846
X_5	**-0.583**	**-0.620**	-0.095	-0.077	0.107	0.197	-0.130	0.807
X_6	**0.865**	-0.026	0.204	0.160	-0.146	-0.140	-0.145	0.879
X_7	**0.865**	-0.055	0.010	0.159	-0.108	0.246	0.132	0.867
X_8	**0.532**	0.354	-0.065	0.180	0.305	-0.089	-0.477	0.772
X_9	0.072	**-0.783**	-0.046	-0.005	-0.038	-0.215	0.259	0.735
X_{10}	-0.024	**-0.837**	0.023	-0.53	-0.179	0.016	-0.136	0.756
X_{11}	0.214	-0.267	-0.394	-0.232	-0.001	**-0.573**	-0.066	0.659
X_{12}	0.169	0.014	-0.332	-0.045	-0.058	0.121	**-0.758**	0.733
X_{13}	0.262	-0.152	-0.214	**0.765**	0.248	0.188	0.080	0.825
X_{14}	0.167	-0.381	-0.175	**0.641**	**-0.551**	-0.013	-0.042	0.920
X_{15}	0.035	**-0.662**	0.268	0.124	0.408	-0.141	-0.071	0.719
X_{16}	**0.526**	0.190	-0.285	-0.115	0.110	-0.039	**0.599**	0.779
X_{17}	-0.103	0.114	**-0.932**	0.060	0.127	0.041	-0.096	0.922
X_{18}	-0.079	-0.029	**-0.933**	0.069	-0.000	-0.019	-0.067	0.886
Eigenvalue	3.3582	2.7105	2.3551	1.8038	1.6571	1.6134	1.3862	14.8844
% Var	0.187	0.151	0.131	0.100	0.092	0.090	0.077	0.827

Rotación

Una vez que se haya seleccionado el número de factores, probablemente usted se querrá probar con diferentes rotaciones. Johnson y Wichern sugieren la rotación varimax. Un resultado similar de diferentes métodos puede dar crédito a la solución que se escogió. En este punto se pudiera interpretar los factores utilizando su conocimiento de los datos.

Ni el *Eigenvalue* ni la comunalidad total, ni el porcentaje de variación explicada total cambian al rotar los factores, es decir, no cambian de la matriz no rotada a la rotada, **solo se distribuyeron las cargas de factor** y ahora sí algunas cargas tienen valores más grandes que en la no rotada como para poder determinar la forma como se agrupan las variables en cada factor.

Lo que si cambia es la importancia en que cada factor explica a las variables (*Eigenvalue* de cada factor) y en la matriz rotada es posible que cada factor hacia la derecha explique menos a las variables.

Para determinar en cual factor se agrupará cada variable debe analizarse la matriz rotada, buscando las cargas significativas. Al identificar las cargas significativas, no importa el signo que tengan, sino la magnitud de la relación, el signo negativo solo expresa una relación inversa. En principio se buscan cargas mayores a 0.50 utilizando la Prioridad 1, en caso de que una variable no tuviera una carga mayor a 0.5 se aplicarían las prioridades 2, 3 o 4, según sea el caso, de los criterios de cargas significativas en toda la matriz rotada. En este ejemplo se tiene:

X_1= Tamaño del lote se identifica con el F5 con una carga de 0.931
X_2= Número de recámaras se identifica con el F6 con una carga de 0.955
X_3= Número de baños se identifica con el F1 con una carga de 0.853
X_4= Número de habitaciones se identifica con el F4 con una carga de 0.783
X_5= Antigüedad de la casa se identifica con el **F1 y el F2** con unas cargas de **0.583 y 0.620**
X_6= Impuestos anuales se identifica con el F1 con una carga de 0.865
X_7= Tipo de estacionamiento interior se identifica con el F1 con una carga de 0.865
X_8= Estilo arquitectónico se identifica con el F1 con una carga de 0.532
X_9= Tipo de combustible para la calefacción se identifica con el F2 con una carga de 0.783
X_{10}= Tipo de sistema para la calefacción se identifica con el F2 con una carga de 0.837
X_{11}= Tipo de alberca se identifica con el F6 con una carga de 0.573
X_{12}= Cocina con antecomedor se identifica con el F7 con una carga de 0.758
X_{13}= Aire acondicionado se identifica con el F4 con una carga de 0.765
X_{14}= Chimenea se identifica con el **F4 y el F5** con unas cargas de **0.641 y 0.551**
X_{15}= Drenaje municipal se identifica con el F2 con una carga de 0.662
X_{16}= Sótano se identifica con el **F1 y el F7** con unas cargas de **0.526 y 0.599**
X_{17}= Cocina integral se identifica con el F3 con una carga de 0.932
X_{18}= Tina de hidromasaje se identifica con el F3 con una carga de 0.933

En caso de que alguna variable tuviera carga significativa con dos o más factores, se debe analizar en la matriz de correlación con cuales variables tiene mayor relación. En este caso **X_5= Antigüedad de la casa** se identifica con el **F1 y el F2** con unas cargas de 0.583 y 0.620, **X_{14}= Chimenea** se identifica con el **F4 y el F5** con unas cargas de 0.641 y 0.551 y **X_{16}= Sótano se identifica** con el **F1 y el F7** con unas cargas de 0.526 y 0.599; entonces se debe checar la correlación con todas las combinaciones que se pueden hacer con la variable X_5= Antigüedad, X_{14}= Chimenea y X_{16}= Sótano y las que forman con el F1 y F2, con el F4 y F5 y F1 y F7 respectivamente:

Para el caso X_5= Antigüedad de la casa se identifica con el F1 y el F2 con unas cargas de 0.583 y 0.620 tenemos:

X_5= Antigüedad VS X_3= Número de baños tienen una correlación de 0.497

X_5= Antigüedad VS X_6= Impuestos anuales tienen una correlación de **0.514**

X_5= Antigüedad VS X_7= Tipo de estacionamiento interior tienen una correlación de 0.462

X_5= Antigüedad VS X_8= Estilo arquitectónico tienen una correlación de 0.505

X_5= Antigüedad VS X_9= Tipo de combustible tienen una correlación de 0.281

X_5= Antigüedad VS X_{10}= Tipo de sistema para la calef. tienen una correlación de 0.429

X_5= Antigüedad VS X_{15}= Drenaje municipal tienen una correlación de 0.361

X_5= Antigüedad VS X_{16}= Sótano tienen una correlación de 0.364

Como la máxima correlación que se obtuvo es con X_6= Impuestos anuales, entonces la variable X_5= Antigüedad debe ir en el Factor 1.

Para el caso X_{14}= Chimenea se identifica con el F4 y el F5 con unas cargas de 0.641 y 0.551 tenemos:

X_{14}= Chimenea VS X_1= Tamaño del lote tienen una correlación de **0.539**

X_{14}= Chimenea VS X_4= Número de habitaciones tienen una correlación de 0.415

X_{14}= Chimenea VS X_{13}= Aire acondicionado tienen una correlación de 0.459

Como la máxima correlación que se obtuvo es con X_1= Tamaño del lote, entonces la variable X_{14}= Chimenea debe ir en el Factor 5.

Para el caso de X_{16}= Sótano se identifica con el F1 y el F7 con unas cargas de 0.526 y 0.599 tenemos:

X_{16}= Sótano VS X_3= Número de baños tienen una correlación de 0.395

X_{16}= Sótano VS X_4= Número de habitaciones tienen una correlación de 0.087

X_{16}= Sótano VS X_5= Antigüedad tienen una correlación de 0.364

X_{16}= Sótano VS X_6= Impuestos anuales tienen una correlación de 0.230

X_{16}= Sótano VS X_7= Tipo de estacionamiento interior tienen una correlación de **0.411**

X_{16}= Sótano VS X_8= Estilo arquitectónico tienen una correlación de 0.072

X_{16}= Sótano VS X_{12}= Cocina con antecomedor tienen una correlación de 0.115

Como la máxima correlación que se obtuvo es con X_7= Tipo de estacionamiento interior, X_{16}= Sótano debe ir en el Factor 1.

La solución final óptima sería:

F1: X_3= Número de baños, X_5= Antigüedad, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico y X_{16}= Sótano

F2: X_9= Tipo de combustible, X_{10}= Tipo de sistema para la calefacción y X_{15}= Drenaje municipal

F3: X_{17}= Cocina integral y X_{18}= Tina de hidromasaje

F4: X_4= Número de habitaciones y X_{13}= Aire acondicionado

F5: X_1= Tamaño del lote y X_{14}= Chimenea

F6: X_2= Número de recámaras y X_{11}= Tipo de alberca

F7: X_{12}= Cocina con antecomedor

Respuesta al inciso c)

Para contestar el inciso c) utilizaremos la salida de la ventana Sesión del inciso b):

Análisis factorial: X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14, X15, X16, X17, X18

Cargas de factores rotados y comunalidades
Rotación Varimax

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Comunalidad
X_1	0.117	0.020	0.174	0.057	**-0.931**	-0.087	-0.053	0.943
X_2	0.078	-0.036	0.129	0.026	-0.063	**-0.955**	0.144	**0.962**
X_3	**0.853**	-0.072	0.019	-0.077	-0.062	-0.338	-0.130	0.874
X_4	-0.034	0.385	0.134	**0.783**	-0.219	-0.066	-0.113	0.846
X_5	**-0.583**	**-0.620**	-0.095	-0.077	0.107	0.197	-0.130	0.807
X_6	**0.865**	-0.026	0.204	0.160	-0.146	-0.140	-0.145	0.879
X_7	**0.865**	-0.055	0.010	0.159	-0.108	0.246	0.132	0.867
X_8	**0.532**	0.354	-0.065	0.180	0.305	-0.089	-0.477	0.772
X_9	0.072	**-0.783**	-0.046	-0.005	-0.038	-0.215	0.259	0.735
X_{10}	-0.024	**-0.837**	0.023	-0.53	-0.179	0.016	-0.136	0.756
X_{11}	0.214	-0.267	-0.394	-0.232	-0.001	**-0.573**	-0.066	0.659
X_{12}	0.169	0.014	-0.332	-0.045	-0.058	0.121	**-0.758**	0.733
X_{13}	0.262	-0.152	-0.214	**0.765**	0.248	0.188	0.080	0.825
X_{14}	0.167	-0.381	-0.175	**0.641**	**-0.551**	-0.013	-0.042	0.920
X_{15}	0.035	**-0.662**	0.268	0.124	0.408	-0.141	-0.071	0.719
X_{16}	**0.526**	0.190	-0.285	-0.115	0.110	-0.039	**0.599**	0.779
X_{17}	-0.103	0.114	**-0.932**	0.060	0.127	0.041	-0.096	0.922
X_{18}	-0.079	-0.029	**-0.933**	0.069	-0.000	-0.019	-0.067	0.886
Eigenvalue	3.3582	2.7105	2.3551	1.8038	1.6571	1.6134	1.3862	14.8844
% Var	0.187	0.151	0.131	0.100	0.092	0.090	0.077	0.827

El porcentaje de varianza, que es el porcentaje de variación explicada con respecto a la máxima variación que se podría tener, se obtiene, por ejemplo para F1 dividiendo 3.7604/18, lo que es igual a 0.209 o 20.9 %. El Factor que ha explicado más al conjunto de variables es el 1, y después el 2 y así sucesivamente lo cual es lógico pues cada factor adicional explica menor cantidad, ya que cada vez queda menos por explicar, en este caso, por ejemplo, queda por explicar 4.138 de variación; si se agregan los 11 factores que faltan, ya cada uno explicaría muy poco y no convendría. El porcentaje de variación explicada acumulada total es de 0.827 o de 82.70 % de un máximo por explicar de 100%.

Respuesta al inciso d)

Para contestar el inciso d) utilizamos la salida de la ventana Sesión del inciso b):

Análisis factorial: X1, X2, X3, X4, X5, X6, X7, X8, X9, X10, X11, X12, X13, X14, X15, X16, X17, X18

Cargas de factores rotados y comunalidades
Rotación Varimax

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Comunalidad
X_1	0.117	0.020	0.174	0.057	**-0.931**	-0.087	-0.053	0.943
X_2	0.078	-0.036	0.129	0.026	-0.063	**-0.955**	0.144	**0.962**
X_3	**0.853**	-0.072	0.019	-0.077	-0.062	-0.338	-0.130	0.874
X_4	-0.034	0.385	0.134	**0.783**	-0.219	-0.066	-0.113	0.846
X_5	**-0.583**	**-0.620**	-0.095	-0.077	0.107	0.197	-0.130	0.807
X_6	**0.865**	-0.026	0.204	0.160	-0.146	-0.140	-0.145	0.879
X_7	**0.865**	-0.055	0.010	0.159	-0.108	0.246	0.132	0.867
X_8	**0.532**	0.354	-0.065	0.180	0.305	-0.089	-0.477	0.772
X_9	0.072	**-0.783**	-0.046	-0.005	-0.038	-0.215	0.259	0.735
X_{10}	-0.024	**-0.837**	0.023	-0.53	-0.179	0.016	-0.136	0.756
X_{11}	0.214	-0.267	-0.394	-0.232	-0.001	**-0.573**	-0.066	0.659
X_{12}	0.169	0.014	-0.332	-0.045	-0.058	0.121	**-0.758**	0.733
X_{13}	0.262	-0.152	-0.214	**0.765**	0.248	0.188	0.080	0.825
X_{14}	0.167	-0.381	-0.175	**0.641**	**-0.551**	-0.013	-0.042	0.920
X_{15}	0.035	**-0.662**	0.268	0.124	0.408	-0.141	-0.071	0.719
X_{16}	**0.526**	0.190	-0.285	-0.115	0.110	-0.039	**0.599**	0.779
X_{17}	-0.103	0.114	**-0.932**	0.060	0.127	0.041	-0.096	0.922
X_{18}	-0.079	-0.029	**-0.933**	0.069	-0.000	-0.019	-0.067	0.886
Eigenvalue	3.3582	2.7105	2.3551	1.8038	1.6571	1.6134	1.3862	14.8844
% Var	0.187	0.151	0.131	0.100	0.092	0.090	0.077	0.827

Se puede observar que la variable mejor explicada por el conjunto de factores es la variable X_2=Número de Recámaras, ya que es la que tiene la mayor comunalidad de 0.962

2.1.1.1 EJERCICIOS DE **REFUERZO**

EJERCICIOS DE REFUERZO 2.1.1.1 FACTOR ANÁLISIS

2.1.1.1.1 Se tiene la siguiente matriz de factores no rotada:

Variable	Factor 1	Factor 2	Factor 3	Factor 4
A	0.67	0.49	-0.41	-0.35
B	0.55	0.58	0.45	-0.39
C	0.45	0.39	0.31	0.33
D	0.53	0.41	0.36	0.25
E	0.70	0.35	-0.29	0.28
F	0.65	0.48	0.45	0.32

a) Obtenga las Comunalidades, el Eigenvalue y el Porcentaje de Variación Explicada:
b) A partir de estos resultados, ¿Se puede saber cómo agrupar a las variables?. Explique.
c) ¿Cuál es la variable que más se identifica con las características?. Explique el porqué de su respuesta.

2.1.1.1.2 Se tiene la siguiente matriz de factores no rotada.

	F1	F2
X_1	0.8386	-0.2196
X_2	-0.7387	-0.5203
X_3	0.5242	0.6748
X_4	0.5849	0.3674

a) Obtenga la variación explicada de cada factor y las comunalidades.
b) Se roto la matriz y generó la siguiente información:

	F1	F2
X_1	0.8050	0.3215
X_2	0.9033	-0.0202
X_3	-0.0208	0.8541
X_4	-0.2523	0.6430

Obtenga *Eigenvalue* y comunalidades.
c) ¿ Se ganó o perdió información con respecto a la matriz no rotada?
d) Agrupe o identifique a cada variable con el factor que le corresponda y concluya cómo quedaron agrupadas las variables.
e) ¿Cuál es la variable que más se identifica con las características?. Explique el porqué de su respuesta.

2.1.1.1.3 Una firma de investigación de mercados esta en el proceso de intentar relacionar pruebas de puntaje de actividad-interés a medidas de "éxito" de solicitantes de empleo como vendedores al detalle en un gran almacén de departamentos de cierta ciudad. Los puntajes de actividad-interés son: X_1=vivir en grupo; X_2afición por deportes al aire libre; X_3afición por la música y X_4=deseo de viajar.

Valores altos en X_i indican mayor interés o éxito respectivamente.

a) Utilice un programa de computo estadístico para correr un análisis de factores (Utilice el modelo de componentes principales y el criterio de rotación de factores Varimax)
b) Determine el número de factores óptimo y justifique su decisión.
c) Agrupe o identifique a cada variable con el factor que le corresponda y concluya cómo quedaron agrupadas las variables.
d) ¿Cuál es la variable que más se identifica con las características?. Explique el porqué de su respuesta.

Persona	X_1	X_2	X_3	X_4
1	21	26	7	8
2	22	16	11	7
3	16	28	11	7
4	17	30	9	13
5	12	26	12	7
6	25	10	18	14
7	18	21	14	16
8	15	17	5	11
9	14	23	13	8
10	18	20	10	5
11	14	29	14	11
12	15	23	16	7
13	25	21	14	12
14	15	20	3	10

OBJETIVO 2.2 El alumno interpretará los resultados de las técnicas de conglomerados (Cluster Análisis ó Análisis de Conglomerados)

ANTECEDENTES	**CONCEPTOS DE:**
	Medida de distancia, Medida de Similitud, Coeficiente de similitud, Agrupamiento Jerárquico., Dendograma.

2.2.1 ANÁLISIS DE CONGLOMERADOS, MEDIDAS DE DISTANCIA Y SIMILITUD, METODO JERÁRQUICO AGLOMERATIVO

CONCEPTOS BÁSICOS ANÁLISIS DE CONGLOMERADOS

Suponga que un arqueólogo descubre un gran escondite de fragmentos de alfarería y toma pequeñas muestras respectivas de cada fragmento, y cada muestra se puede analizar y se pueden averiguar las cantidades relativas de diferentes elementos químicos, como zinc, magnesio, hierro, etc. El arqueólogo quiere separar los fragmentos en montones distintos, de modo que los que queden en cada uno de esos montones provengan de la misma pieza de alfarería. Resulta claro que esto puede ser una tarea difícil porque el arqueólogo no sabe cuantos montones de fragmentos resultarán, cuántos fragmentos quedarán en cada montón o si hay algunos fragmentos que en realidad pertenezcan al mismo montón. El análisis por agrupación es un método multivariado que puede ayudar a resolver este problema.

El *análisis por agrupación* (CA: *cluster analysis*) es semejante al discriminante en el sentido de que se usa para clasificar individuos o unidades experimentales en subgrupos definidos de una manera única. Este análisis se puede emplear cuando el investigador cuenta con muestras aleatorias previamente obtenidas de cada uno de los subgrupos definidos de manera única. El análisis por agrupación trata de los problemas de clasificación cuando no se sabe de antemano de cuáles subgrupos se originan las observaciones.

El análisis de conglomerados es una técnica estructural que tiene por objeto resumir información, es decir, todas las variables son consideradas como independientes y en su mayoría deben estar medidas en una escala de intervalo o de razón. Este método se utiliza básicamente para la segmentación mercados en donde lo que se busca es formar grupos de objetos o personas lo mas homogéneos posible, al comparar los elementos que conforman un grupo, y lo mas heterogéneo, al comparar un elemento de un grupo con otro.

Esta técnica a veces se confunde con el análisis de factores pero la diferencia radica en que en factores se trata de agrupar variables para resumir información y en la técnica de conglomerados o agrupamiento se trata de reunir personas con base en las respuestas similares que se identifiquen.

El agrupamiento se puede realizar mediante el análisis de las similitudes o diferencias (distancias relativas) detectadas al comparar dos objetos o personas, es decir, se puede determinar lo parecido de las personas con base en las mayores similitudes, o en caso contrario, según las diferencias menores entre respuestas.

Para realizar un análisis de conglomerados o de agrupamiento se deben tomar en cuenta los siguientes requisitos:

1. Detectar si los objetos, personas o variables que se pretende agrupar lo pueden hacer en forma natural.
2. Las variables que son utilizadas en esta técnica, por regla general, son actitudes, opiniones, características demográficas o correlaciones entre variables y todas deben estar previamente estandarizadas.
3. Lo parecido de los elementos puede calcularse con una medida de distancia o bien con una de similitud.

Análisis de agrupamiento a través de la medida de distancia

Bajo este criterio se tendría como atención central agrupar aquellas personas u objetos con menores distancias entre si (lo mas parecido). Para cuantificar la distancia entre las personas u objetos se utiliza la siguiente **formula de la medida de distancia**:

$$d_{ij} = \left[\sum_{k=1}^{r} \left(X_{ik} - X_{jk} \right)^2 \right]^{1/2}$$

Donde:

d_{ij} : es la distancia de los perfiles de los objetos i y j .
X_{ik} : representa la característica k medida en el objeto i .
X_{jk} : representa la característica k medida en el objeto j.
r : es el número de características utilizadas para evaluar lo parecido.

Para realizar el análisis de agrupamiento los datos de las variables deben estar estandarizadas, si son métricas, es decir, eliminar sus unidades de manera que se encuentren todas en términos de *z*. Para lograrlo es necesario aplicar la siguiente formula:

$$Z = \frac{x - \mu}{\sigma}$$

Utilice el análisis de conglomerados de observaciones para clasificar las observaciones en grupos cuando los grupos se desconocen inicialmente.

Análisis de agrupamiento a través de la medida de similitud

La otra medida mediante la cual se puede evaluar lo parecido de las personas u objetos para después agruparlos es la medida de similitud, la cual se utiliza generalmente cuando se manejan características o variables cualitativas, indicando con un 1 la presencia del atributo y con 0 la ausencia del atributo...En esta técnica se manejan las características cualitativas por medio de variables *dummy* De acuerdo a la presencia o ausencia de atributos en las personas u objetos se establece la similitud existente.

Existen ocho coeficientes que permiten medir la similitud entre un objeto y otro que permiten evaluar el hecho de que se presente o no cierto atributo en ambas personas u objetos o solamente en uno de ellos de la siguiente manera:

Objeto i			Total
Objeto j	1	0	
1	a	b	
0	c	d	
Total			p

Donde:

a: numero de características presentes tanto en el objeto *i* como en *j* .
b: numero de características ausentes en el objeto *i* y presentes en el objeto *j.*
c: numero de características ausentes en el objeto *j* y presentes en el objeto *i.*
d: numero de ausencias de características en ambos objetos.
p: total de características a ser evaluadas (a+b+c+d).

La tabla anterior y las letras se pueden tomar como nomenclatura que se utilizara en los coeficientes de similitud que a continuación se tratarà para después aplicar el algoritmo de agrupamiento.

Coeficientes de similitud:

1. Del total de características, se evalúa cuantas están presentes y cuantas ausentes en ambas personas u objetos, es decir se trata de evaluar la proporción de atributos similares en ambos:

$$\frac{a+d}{p}$$

2. Se da importancia a la presencia y ausencia de las características en ambas personas u objetos, duplicándolas:

$$\frac{2(a+d)}{2(a+b)+b+c}$$

3. Se da importancia a las diferencias de la característica en las personas u objetos, duplicándola:

$$\frac{a+d}{a+d+2(b+c)}$$

4. Se calcula la proporción de las presencias en ambos, con respecto a todas las características:

$$\frac{a}{p}$$

5. Se determina la presencia de la característica en ambos, con respecto a *a,b* y *c* , ya que la ausencia en *d* no se puede evaluar:

$$\frac{a}{a+b+c}$$

6. Se evalúa solo la presencia al doble, sin poder evaluar la ausencia en ambos:

$$\frac{2a}{2a+b+c}$$

7. Se determina solo las diferencias en ellos al doble, sin poder evaluar la ausencia en ambos:

$$\frac{a}{a+2(b+c)}$$

8. Se evalúa cuantas características se encuentran presentes en ambas personas u objetos con respecto a las diferencias en ellos; es decir, se mide la relación de iguales a diferentes:

$$\frac{a}{b+c}$$

Nota: Los tres primeros coeficientes proporcionan resultados similares ya se basan en la presencia y ausencia de los atributos en ambas personas u objetos, mientras que los coeficientes 5, 6 y 7 dan entre si resultados similares por estar enfocándose a los atributos presentes en un objeto y diferentes entre ellos sin

evaluar la ausencia. El ultimo coeficiente es totalmente diferente a los dos grupos anteriores ya que mide la solo la relación de iguales a diferentes, dando como resultado una diagonal diferente a unos tendiendo a infinito porque al comparar a un objeto consigo mismo , *b* y *c* son igual a 0.

Este procedimiento utiliza un método jerárquico aglomerativo que comienza con todas las observaciones separadas, cada una de ellas formando su propio conglomerado. En el primer paso, las dos observaciones más cercanas entre sí se unen. En el siguiente paso, una tercera observación se une a las dos primeras u otras dos observaciones se unen para formar un conglomerado diferente. Este proceso continuará hasta que todos los conglomerados se unan en un solo conglomerado; sin embargo, este conglomerado individual no es útil para propósitos de clasificación. Por lo tanto, se debe decidir cuántos grupos son lógicos para los datos y clasificarlos adecuadamente.

METODO DE AGRUPAMIENTO JERARQUICO

Una vez calculado lo parecido de las personas u objetos con base en medidas de similitud y distancia se debe buscar el procedimiento a través del cual pueden agruparse. El algoritmo que se habrá de seguir para formar grupos de personas u objetos similares entre si se le conoce como **método de agrupamiento jerárquico.**

Este método se puede dividir en agrupamiento jerárquico aglomerativo y en agrupamiento jerárquico divisional, de acuerdo con las bases de similitud de que se cuente y en si en la secuencia de los pasos que se llevan a cabo. Lo que se espera es contar como solución mínima a todos los objetos o personas en un grupo y como solución máxima aun grupo (conglomerado) para cada elemento.

En el método **jerárquico aglomerativo** se comienza suponiendo que cada uno de los elementos que se pretende agrupar constituye un grupo o conglomerado por lo tanto existen tantos grupos como personas u objetos a ser agrupados. Con base en lo parecido de ellos, los elementos se van fusionando hasta terminar uniendo a todas las personas u objetos en un solo grupo o conglomerado y de ahí el nombre de aglomerativo, es decir de muchas partes se forma un todo, donde se van uniendo los elementos mas parecidos entre si.

La agrupación final de conglomerados (también denominada partición final) es la agrupación de conglomerados que, si los resultados son favorables, identificará grupos cuyas observaciones o variables compartan características comunes. La decisión acerca de la agrupación final también se denomina dendrograma . El dendrograma completo (diagrama de árbol) es una representación gráfica de la conformación en amalgama de observaciones o variables en un solo conglomerado. Cortar el dendrograma es similar a trazar una línea a lo largo del dendrograma para especificar la agrupación

En el método **jerárquico divisional** se trabaja al revés; supone que existe un solo grupo o conglomerado conteniendo a todas las personas u objetos y con base en diferencias los va dividiendo hasta terminar formando tantos grupos como objetos manejados en la técnica y de ahí su nombre de divisional, es decir de un todo de divide en partes.

Los resultados del agrupamiento mediante esta técnica u otra se representan en forma grafica mediante un diagrama llamado *dendograma*.

METODO JERARQUICO AGLOMERATIVO

En este método, el procedimiento consiste en ir agrupando los elementos parecidos por parejas y evaluar lo parecido que llega a ser una pareja o grupo con los restantes elementos que no se han fusionado. Para hacer esto se utilizan los llamados *criterios de eslabonamiento*.

final.

Se recomienda probar con varios métodos de eslabonamiento y comparar los resultados. Dependiendo de las características de los datos, algunos métodos pueden proveer "mejores" resultados que otros.

· Con el eslabonamiento simple, o del "vecino más cercano", la distancia entre dos conglomerados es la distancia mínima entre una observación en un conglomerado y una observación en el otro conglomerado. El eslabonamiento simple es una opción apropiada cuando los conglomerados están claramente separados. Cuando las observaciones están cerca entre sí, el eslabonamiento simple tiende a identificar conglomerados largos similares a cadenas que pueden tener una separación relativamente grande entre observaciones en cualquier extremo de la cadena. ·

· Con el eslabonamiento completo, o "del vecino más lejano", la distancia entre dos conglomerados es la distancia máxima entre una observación en un conglomerado y una observación en el otro conglomerado. Este método garantiza que todas las observaciones en un conglomerado se encuentren dentro de una distancia máxima y tiende a producir conglomerados con diámetros similares. Los resultados pueden ser sensibles a valores atípicos.

· Con el eslabonamiento promedio, la distancia entre

Tipos de criterios de eslabonamiento

Un eslabonamiento consiste en volver a establecer el grado de parecido de un nuevo grupo con el resto de los elementos que no habìan sido agrupados y existen básicamente tres formas o criterios para medir ese grado de parecido.

Nota: Por cuestiones de la terminología que se va a utilizar suponga que se utilizan los elementos *u,v* y *w* de los cuales ya se unieron (*uv*) y se pretende evaluar su parecido con *w*.

1. Eslabonamiento simple:

Este criterio asume que al momento de fusionarse dos personas u objetos, prevalecen las características de aquel elemento cuya distancia sea la mínima o, bien cuya similitud sea la máxima (mas parecido):

MIN { duw,dvw} MAX {Suw,Svw}

Este criterio supone que los elementos son tan parecidos como antes de fusionarse.

2. Eslabonamiento completo:

Este criterio considera que cuando dos elementos se unen, el grado de parecido de ese grupo a un tercer elemento va a estar dado por la máxima distancia o la mínima de las similitudes (menos parecido):

MAX { duw,dvw} MIN {Suw,Svw}

Este criterio supone que el nuevo grupo comparado con otro elemento no agrupado es tan diferente como los elementos antes de agruparse.

3. Eslabonamiento promedio:

Este criterio asume que al unirse dos elementos y formar un grupo, las características de ellos con respecto a un tercero estarán dadas por el promedio de las que originalmente poseían, lo cual puede representarse de la siguiente manera:

d(uv)w=(duw+dvw)/2 S(uv)w=(Suw+Svw)/2

dos conglomerados es la distancia media entre una observación en un conglomerado y una observación en el otro conglomerado. En vista de que los métodos de eslabonamiento simple o completos agrupan conglomerados sobre la base de distancias entre pares individuales, el eslabonamiento promedio utiliza una medida de ubicación más central.

Pasos para llevar a cabo un agrupamiento jerárquico aglomerativo:

1- Se necesita contar con la información en forma de matriz de distancia o similitud.

2- Se comienza con *n* grupos, cada uno de ellos formado por un elemento. Con base en la información manejada en la matriz de distancia o similitud se seleccionara a la pareja de elementos más parecida para formar un grupo.

3- Una vez ya fusionados los elementos y formado el grupo o conglomerado, se aplica alguno de los criterios de eslabonamiento para evaluar lo parecido que es este grupo con respecto a los demás elementos sin unirse.

4- Se repiten los pasos 2 y 3 un total de *n-1* veces hasta terminar fusionando a todas las personas u objetos en un solo grupo.

¿Cómo se sabe dónde cortar el dendrograma? Primero convendría que se ejecute un análisis del conglomerado sin especificar la partición final. Se examinen los resultados de los niveles de similitud y de distancia y en el dendrograma. El nivel de similitud en cualquier paso es el porcentaje de la distancia mínima en el paso, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud o de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendrograma, si esto tiene sentido para sus datos.

Después de elegir dónde se desea realizar la partición, se vuelve a ejecutar el procedimiento del conglomerado, utilizando, o bien el Número de conglomerados o Nivel de similitud para obtener un número establecido de

Dendrograma:

Es la representación grafica de los grupos o conglomerados formados donde en el eje horizontal se establecen los elementos que se están agrupando, de preferencia en el orden en que se fueron agrupando y en el eje vertical se manejan las medidas de distancia (o similitud) a las que se fueron uniendo los grupos. Si se trata de medidas de distancia el punto inicial es cero y si se trata de medidas de similitud el punto inicial es uno y hacia arriba irían disminuyendo los valores ya que serian menos parecidos los elementos que se van agrupando conforme se avanza en el proceso de agrupamiento.

El dendrograma esta formado por rectángulos que unen a los elementos agrupados con sus esquinas inferiores y su altura indica el grado de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los agrupamientos.

Mientras a mayor altura se realiza la unión de elementos para formar un grupo, es mayor la diferencia entre ellos (en similitud, es menor la similitud a mayor altura).

Mediante este dendrograma se puede determinar tentativamente en cuantos grupos o conglomerados se unieron los elementos.

Como criterio general si se decide dejar todos los elementos en un solo grupo no tiene caso, porque resume demasiado las características de los elementos. Si se decide dejar como solución a dos grupos, el agrupamiento se realizó a una

conjuntos o un Nivel de similitud para cortar el dendrograma.Se examinan los conglomerados resultantes en la partición final para observar si la agrupación parece lógica. La observación de dendrogramas para determinar diferentes agrupaciones finales puede ayudar a decidir cuáles de ellas tienen más sentido para los datos.

distancia muy alta, lo que indicaría es que en realidad los elementos que quedan en cada grupo no son muy similares.

2.2.1.1 EJEMPLO ILUSTRATIVO

EJEMPLO ILUSTRATIVO 2.2.1.1 ANÁLISIS DE CONGLOMERADOS. MATRIZ DE DISTANCIA.

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de los hombres mexicanos, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras . El estudio antropométrico contempló 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y la cintura de cinco personas del género **FEMENINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Cintura (cms)
1	1.52	57.12	84.10
2	1.67	76.85	86.20
3	1.54	67.62	85.22
4	1.56	49.35	83.60
5	1.63	87.22	87.52
Promedio:	**1.58**	**67.63**	**85.33**
Desviación estándar	**0.06**	**15.11**	**1.59**

a) Obtenga la pareja de personas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres

eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

c) Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

Resolución del inciso a)

Estandarización de variables

Antes que nada hay que se deben estandarizar las variables, es decir, eliminar el efecto de las unidades y hacerlas comparables, de manera que todas estén en términos de "Z" mediante la siguiente expresión:

$$Z = \frac{x - \bar{x}}{s}$$

$$Z_{11} = \frac{1.52 - 1.58}{.06} = -1.00 \qquad Z_{12} = \frac{57.12 - 67.63}{15.11} = -.70 \qquad Z_{13} = \frac{84.10 - 85.33}{1.59} = -.77$$

$$Z_{21} = \frac{1.67 - 1.58}{.06} = 1.50 \qquad Z_{22} = \frac{76.85 - 67.63}{15.11} = .61 \qquad Z_{23} = \frac{86.20 - 85.33}{1.59} = .55$$

$$Z_{31} = \frac{1.54 - 1.58}{.06} = -0.67 \qquad Z_{32} = \frac{67.62 - 67.63}{15.11} = 0.00 \qquad Z_{33} = \frac{85.22 - 85.33}{1.59} = -.07$$

$$Z_{41} = \frac{1.56 - 1.58}{.06} = -0.33 \qquad Z_{42} = \frac{49.35 - 67.63}{15.11} = -1.21 \qquad Z_{43} = \frac{83.60 - 85.33}{1.59} = -1.09$$

$$Z_{51} = \frac{1.63 - 1.58}{.06} = .83 \qquad Z_{52} = \frac{87.22 - 67.63}{15.11} = 1.30 \qquad Z_{53} = \frac{87.52 - 85.33}{1.59} = 1.38$$

Los datos estandarizados de las 5 personas, al considerar 3 variables, se presenta de la siguiente manera:

Persona	Altura estandarizada	Peso estandarizado	Cintura estandarizada
1	- 1.00	0.70	-0 .77
2	1.50	0.61	0.55
3	- 0.67	0.00	- 0.07
4	- 0.33	- 1.21	- 1.09
5	0.83	1.30	1.38
Promedio:	0.00	0.00	0.00
Desviación estándar	1.00	1.00	1.00

Cálculo de las distancias:

$$d_{ij} = \left[\sum_{K=1}^{r}(x_{ik} - x_{jk})^2\right]^{0.5}$$

$$d_{1,2} = [(-1.00 - 1.50)^2 + (-.70 - .61)^2 + (-.77 - .55)^2]^{0.5} = 3.12$$

$$d_{1,3} = [(-1.00 - (-0.67))^2 + (-.70 - 0.00)^2 + (-.77 - (-.07))^2]^{0.5} = 1.04$$

$$d_{1,4} = [(-1.00 - (-0.33))^2 + (-.70 - (-1.21))^2 + (-.77 - (-1.09))^2]^{0.5} = .90$$

$$d_{1,5} = [(-1.00 - .83)^2 + (-.70 - 1.30)^2 + (-.77 - 1.38)^2]^{0.5} = 3.46$$

$$d_{2,3} = [(1.50 - (-0.67))^2 + (.61 - 0.00)^2 + (.55 - (-07))^2]^{0.5} = 2.34$$

$$d_{2,4} = [(1.50 - (-0.33))^2 + (.61 - (-1.21))^2 + (.55 - (-1.09))^2]^{0.5} = 3.06$$

$$d_{2,5} = [(1.50 - .83)^2 + (.61 - 1.30)^2 + (.55 - 1.38)^2]^{0.5} = 1.27$$

$$d_{3,4} = [(-0.67 - (-0.33))^2 + (0.00 - (-1.21))^2 + (-.07 - (-1.09))^2]^{0.5} = 1.62$$

$$d_{3,5} = [(-0.67 - .83)^2 + (0.00 - 1.30)^2 + (-.07 - 1.38)^2]^{0.5} = 2.46$$

$$d_{4,5} = [(-0.33 - .83)^2 + (-1.21 - 1.30)^2 + (-1.09 - 1.38)^2]^{0.5} = 3.71$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5 personas a la que le nombraremos ***matriz de distancia*** la cual representa que tan parecidas son las personas con respecto a la distancia entre ellas.

	1	2	3	4	5
1	0	3.12	1.04	0.90	3.46
2	3.12	0	2.34	3.06	1.27
3	1.04	2.34	0	1.62	2.46
4	0.90	3.06	1.62	0	3.71
5	3.46	1.27	2.46	3.71	0

Como esta matriz es una matriz espejo, es decir la matriz superior es igual a la matriz inferior y la diagonal contiene ceros ya que la distancia que hay al comparar una persona consigo misma es nula, sólo será necesario utilizar la información de la matriz inferior como se muestra a continuación:

	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	0.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

Al observar la matriz de distancia se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya distancia entre sí es la menor que está formada por las personas 1 y 4, con una medida de distancia de 0.90 y la menos parecida será aquella cuya distancia entre sí es la mayor que está formada por las personas 4 y 5, con una medida de distancia de 3.71.

	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	0.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

Métodos de eslabonamiento

El método de eslabonamiento que se elija determinará cómo se define la distancia entre dos conglomerados. En cada etapa de la amalgamación, los dos conglomerados más cercanos se unen. Al comienzo, cuando cada observación constituye un conglomerado, la distancia entre los conglomerados es simplemente la distancia entre las observaciones. Posteriormente, después de que se unifican las observaciones, es necesaria una regla de eslabonamiento para calcular las distancias entre conglomerados, cuando hay múltiples observaciones en un conglomerado.

Se recomienda probar con varios métodos de eslabonamiento y comparar los resultados. Dependiendo de las características de los datos, algunos métodos pueden proveer "mejores" resultados que otros.

· Con el **eslabonamiento simple**, o del "vecino más cercano", la distancia entre dos conglomerados es la

distancia mínima entre una observación en un conglomerado y una observación en el otro conglomerado. El eslabonamiento simple es una opción apropiada cuando los conglomerados están claramente separados. Cuando las observaciones están cerca entre sí, el eslabonamiento simple tiende a identificar conglomerados largos similares a cadenas que pueden tener una separación relativamente grande entre observaciones en cualquier extremo de la cadena.

Resolución al inciso b)

Vinculación o eslabonamiento:

1. **Eslabonamiento Simple:**

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia menor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.90

	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	0.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 4, como se muestra en la siguiente matriz inferior:

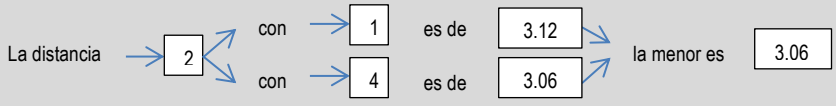
	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	0.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

El nuevo conglomerado (1,4) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,4)	2	3	5
(1,4)	0			
2		0		
3			0	
5				0

Paso 2. Eslabonar

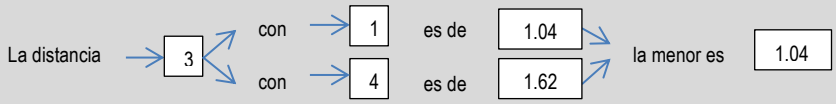
Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al nuevo conglomerado (1,4) con el resto de las personas; al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.11 y la de la persona 2 con la 4 que es 3.06, eligiendo entonces la distancia menor de las dos personas que en este caso es 3.06. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.06	0		
3			0	
5				0

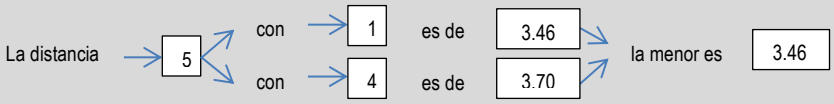
Repetimos el mismo criterio para eslabonar la persona 3 con la pareja de personas (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con 1 que es 1.04 y la de la personas 3 con 4 que es 1.62, eligiendo entonces la distancia menor de las dos personas que en este caso es 1.04. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.06	0		
3	1.04		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,4); al comparar (1,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 3.46 y la de la personas 5 con 4 que es 3.70, eligiendo entonces la distancia menor de las dos personas que en este caso es 3.46. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.06	0		
3	1.04		0	
5	3.46			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,4)	2	3	5
(1,4)	0			
2	3.06	0		
3	1.04	2.34	0	
5	3.46	1.27	2.46	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 3 que se eslabonaría al conglomerado formado por las personas (1,4) con la distancia menor de 1.04 como se muestra a continuación:

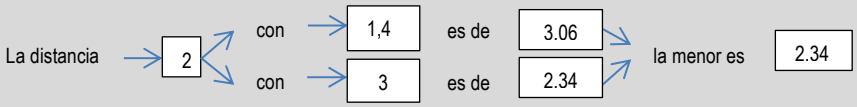
	(1,4)	2	3	5
(1,4)	0			
2	3.06	0		
3	1.04	2.34	0	
5	3.46	1.27	2.46	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,4,3)	2	5
(1,4,3)	0		
2		0	
5			0

Paso 2. Eslabonar

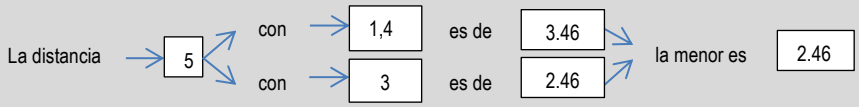
Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (1,4,3) con el resto de las personas; al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las 1 y 4 que es 3.06 y la de la persona 2 con la 3 que es 2.34, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.34. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4,3)	2	5
(1,4,3)	0		
2	2.34	0	
5			0

Repetimos el mismo criterio para eslabonar la persona 5 con el conglomerado (1,4,3); al comparar (1,4,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con las personas 1 y 4 que es 1.4 y la de la personas 5 con la 3 que es 2.46, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.46. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4,3)	2	5
(1,4,3)	0		
2	2.34	0	
5	2.46		0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,4,3)	2	5
(1,4,3)	0		
2	2.34	0	
5	2.46	1.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso serían las personas 2 y 5 para formar un nuevo conglomerado con la distancia menor de 1.27 como se muestra a continuación:

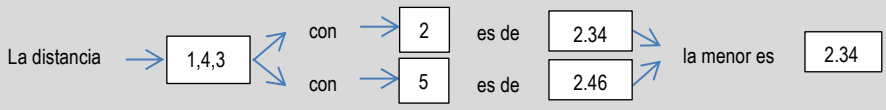
	(1,4,3)	2	5
(1,4,3)	0		
2	2.34	0	
5	2.46	1.27	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (1,4,3) con el resto de las personas; al comparar (1,4,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las personas (1,3,4) que es 2.34 y la de la persona 5 con las (1,3,4) que es 2.46, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.34. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)	2.34	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 2.34.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.90	1	4	1
2	3	1.04	1	3	1
3	2	1.27	2	5	2
4	1	2.34	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La segunda columna no muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 1 y la 4 y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 3 al primer conglomerado formado por las personas 1 y 4 como se muestra en la columna 4 a una distancia de 1.04. En la etapa 3 se forma un nuevo conglomerado como se aprecia en la columna 5 formado por las personas 2 y 5 a una distancia de 1.27 y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de 2.34. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

```
Variables estandarizadas, Distancia euclediana, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	**Nivel de distancia**	**Conglomerados incorporados**		Nuevo conglomerado
1	4	76.3492	**0.87219**	**1**	**4**	1
2	3	71.8157	**1.03937**	**1**	**3**	1
3	2	66.1312	**1.24900**	**2**	**5**	2
4	1	39.6809	**2.22443**	**1**	**2**	1

Dendograma

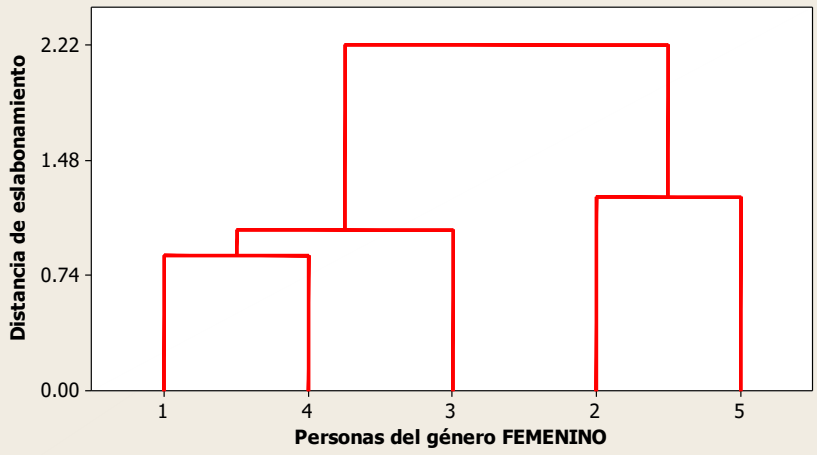
El dendrograma completo (diagrama de árbol) es una representación gráfica de la conformación en amalgama de observaciones o variables en un solo conglomerado. Cortar el dendrograma es similar a trazar una línea a lo largo del dendrograma para especificar la agrupación final.

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,4,3) y el conglomerado 2 por las personas (2,5).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Ejemplo ilustrativo 1. Dendograma de datos antropométricos



Elaboró: M. en Admón. Javier Bech Vertti

¿Cómo se sabe dónde cortar el dendrograma? Primero convendría que se ejecute un análisis del conglomerado sin especificar la partición final. Se examinen los resultados de los niveles de similitud y de distancia y en el dendrograma. El nivel de

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe

similitud en cualquier paso es el porcentaje de la distancia mínima en el paso, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud o de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendrograma, si esto tiene sentido para sus datos

· Con el eslabonamiento completo, o "del vecino más lejano", la distancia entre dos conglomerados es la distancia máxima entre una observación en un conglomerado y una observación en el otro conglomerado. Este método garantiza que todas las observaciones en un conglomerado se encuentren dentro de una distancia máxima y tiende a producir conglomerados con diámetros similares. Los resultados pueden ser sensibles a valores atípicos.

especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 3 y 4 los valores de distancia cambian abruptamente de 1.27 a 2.34 cuando se unen ambos grupos, por lo tanto, la solución más factible parece ser la de 2 conglomerados (1,4,3) y (2,5).

2. Eslabonamiento Completo:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia mayor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, las personas más parecidas se elegirán con la distancia menor, que en este caso es 0.90

	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 4, como se muestra en la siguiente matriz inferior:

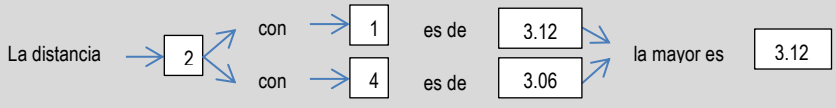
	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	0.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

El nuevo conglomerado (1,4) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,4)	2	3	5
(1,4)	0			
2		0		
3			0	
5				0

Paso 2. Eslabonar

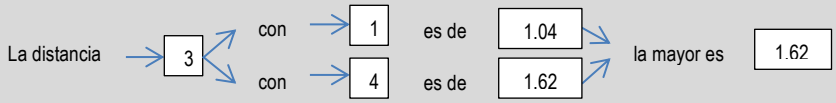
Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al nuevo conglomerado (1,4) con el resto de las personas; al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.12 y la de la persona 2 con la 4 que es 3.06, eligiendo entonces la distancia mayor de las dos personas que en este caso es 3.12. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.12	0		
3			0	
5				0

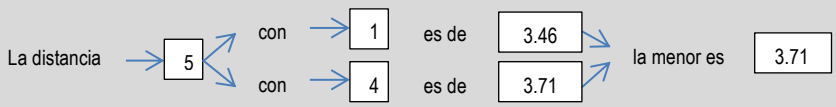
Repetimos el mismo criterio para eslabonar a la persona 3 con la pareja de personas (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre las personas 3 con 1 que es 1.04 y la de la personas 3 con 4 que es 1.62, eligiendo entonces la distancia mayor de las dos personas que en este caso es 1.62. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.12	0		
3	1.62		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,4); al comparar (1,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 3.46 y la de la personas 5 con 4 que es 3.71, eligiendo entonces la distancia mayor de las dos personas que en este caso es 3.71. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.12	0		
3	1.62		0	
5	3.71			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,4)	2	3	5
(1,4)	0			
2	3.12	0		
3	1.62	2.34	0	
5	3.71	1.27	2.46	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 2 y 5 con una distancia de 1.27 como se muestra a continuación:

	(1,4)	2	3	5
(1,4)	0			
2	3.12	0		
3	1.62	2.34	0	
5	3.71	1.27	2.46	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)		0	
3			0

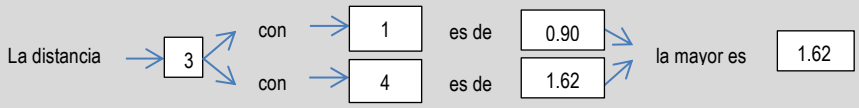
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (1,4) con el resto de las personas; al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre el conglomerado (1,4) con la persona 2 que es 3.12 y al conglomerado (1,4) con la de la persona 5 que es 3.71, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 3.71. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.71	0	
3			0

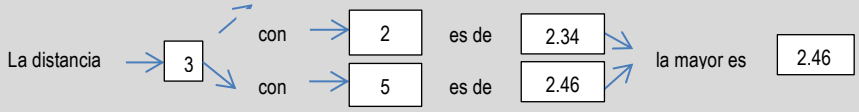
Repetimos el mismo criterio para eslabonar la persona 3 con el conglomerado (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre la persona 3 con la persona 1 que es 0.90 y la de la persona 3 con la 4 que es 1.62, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 1.62. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.71	0	
3	1.62		0

Repetimos el mismo criterio para eslabonar la persona 3 con el conglomerado (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre la persona 3 con la persona 2 que es 2.34 y la de la personas 3 con la 5 que es 2.46, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 2.46. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.71	0	
3	1.62	2.46	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona a un conglomerado ya formado, que en este caso sería la persona 3 que se eslabonaría al conglomerado formado por las personas (1,4) con la distancia menor de 1.62 como se muestra a continuación:

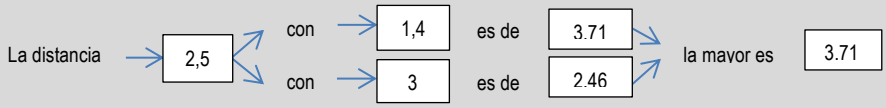
	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.71	0	
3	1.62	2.46	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (1,4,3) con el conglomerado (2,5); al comparar (1,4,3) con el conglomerado (2,5) las personas 2 y 5 se debe verificar en la matriz anterior las distancias entre las personas 2 y 5 con las personas (1,4) que es 3.71 y las personas 2 y 5 con la persona 3 que es 2.46, eligiendo entonces la distancia mayor de las dos combinaciones o cinco personas que en este caso es 3.71. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)	3.71	0

En este punto se detiene el procedimiento llegando a unir a las cinco personas en un solo grupo con una distancia de 3.71.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.90	1	4	1
2	3	1.27	2	5	2
3	2	1.62	1	3	1
4	1	3.71	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan **cuatro etapas**. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **columna 4** que en este caso fueron la 1 y la 4 a una distancia de **0.90** y en la **quinta columna** nos señala la formación del primer conglomerado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 5 a una distancia de **1.27** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 3 al primer conglomerado formado por las personas 1 y 4 a una distancia de **1.62** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **3.71**. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

```
Variables estandarizadas, Distancia euclediana, Enlace completo
Pasos de amalgamación


      Número                                                  Número de
          de    Nivel de   Nivel de     Grupos     Nuevo   obs. en el
Paso  grupos  semejanzal  distancia  incorporados  grupo  grupo nuevo
   1       4     76.3492    0.87219  1          4      1            2
   2       3     66.1312    1.24900  2          5      2            2
   3       2     56.2472    1.61350  1          3      1            3
   4       1      0.0000    3.68777  1          2      1            5
```

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

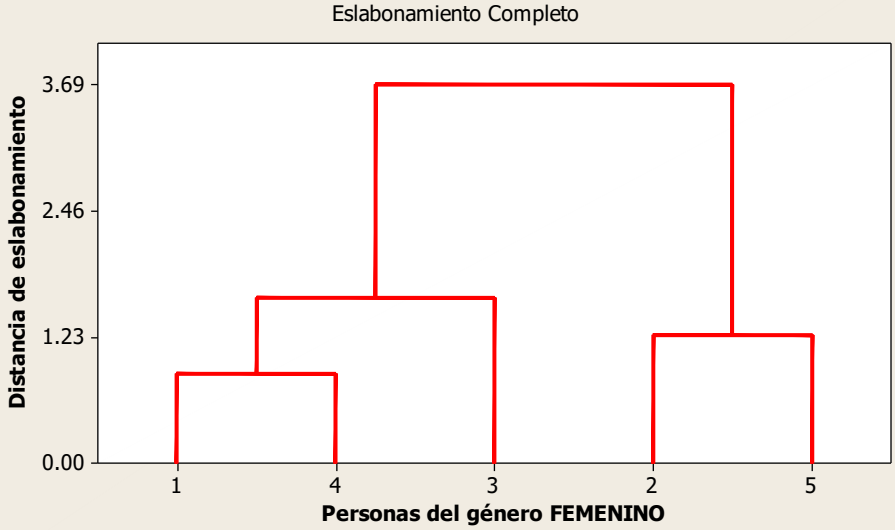
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a

los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,4,3) y el conglomerado 2 por las personas (2,5).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

El dendrograma completo (diagrama de árbol) es una representación gráfica de la conformación en amalgama de observaciones o variables en un solo conglomerado. Cortar el dendrograma es similar a trazar una línea a lo largo del dendrograma para especificar la agrupación final.

Ejemplo ilustrativo 1. Dendograma de datos antropométricos



Elaboró: M. en Admón. Javier Bech Vertti

¿Cómo se sabe dónde cortar el dendrograma? Primero convendría que se ejecute un análisis del conglomerado sin especificar la partición final. Se examinen los resultados de los niveles de similitud y de distancia y en el dendrograma. El nivel de similitud en cualquier paso es el porcentaje de la distancia mínima en el paso, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud o de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendrograma, si esto tiene sentido para sus datos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 3 y 4 los valores de distancia cambian abruptamente de 1.62 a 3.71 cuando se unen ambos grupos, por lo tanto, la solución más factible parece ser la de 2 conglomerados (1,4,3) y (2,5).

· Con el eslabonamiento promedio, la distancia entre dos conglomerados es la distancia media entre una observación en un conglomerado y una observación en el otro conglomerado. En vista de que los métodos de eslabonamiento simple o completos agrupan conglomerados sobre la base de distancias entre pares individuales, el eslabonamiento promedio utiliza una medida de ubicación más central.

3. Eslabonamiento Promedio:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia promedio, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.90

	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	0.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 4, como se muestra en la siguiente matriz inferior:

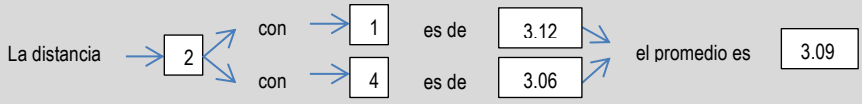
	1	2	3	4	5
1	0				
2	3.12	0			
3	1.04	2.34	0		
4	0.90	3.06	1.62	0	
5	3.46	1.27	2.46	3.71	0

El nuevo cluster (1,4) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,4)	2	3	5
(1,4)	0			
2		0		
3			0	
5				0

Paso 2. Eslabonar

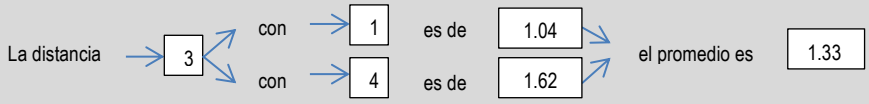
Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al nuevo conglomerado (1,4) con el resto de las personas; al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.12 y la de la persona 2 con la 4 que es 3.06, eligiendo entonces la distancia promedio de las dos personas que en este caso es 3.09. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.09	0		
3			0	
5				0

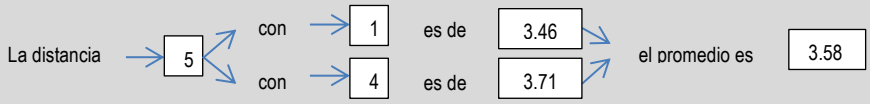
Repetimos el mismo criterio para eslabonar la persona 3 con la pareja de personas (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con 1 que es 1.04 y la de la personas 3 con 4 que es 1.62, eligiendo entonces la distancia promedio de las dos personas que en este caso es 1.33. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.09	0		
3	1.33		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,4); al comparar (1,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 3.46 y la de la personas 5 con 4 que es 3.71, eligiendo entonces la distancia promedio de las dos personas que en este caso es 3.58. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	2	3	5
(1,4)	0			
2	3.09	0		
3	1.33		0	
5	3.58			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,4)	2	3	5
(1,4)	0			
2	3.09	0		
3	1.33	2.34	0	
5	3.58	1.27	2.46	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 2 y 5 a una distancia menor de 1.27 como se muestra a continuación:

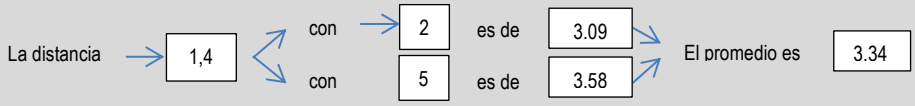
	(1,4)	2	3	5
(1,4)	0			
2	3.09	0		
3	1.33	2.34	0	
5	3.58	1.27	2.46	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)		0	
3			0

Paso 2. Eslabonar

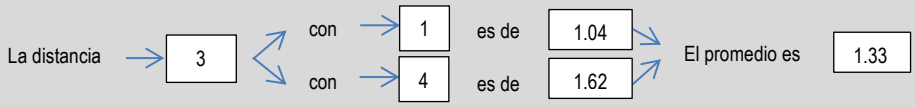
Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado(1,4) con las personas del conglomerado (2,5); al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre las persona (1,4) y la persona 2 que es 3.09 y la de las personas (1,4) con la persona 5 que es 3.58, eligiendo entonces la distancia promedio de las dos combinaciones o cuatro personas que en este caso es 3.34. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.34	0	
3			0

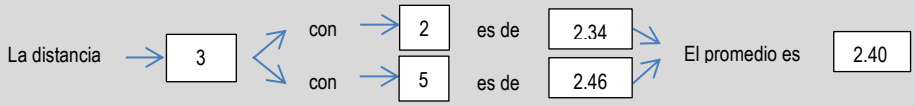
Repetimos el mismo criterio para eslabonar a la persona 3 con el conglomerado (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con las personas 1 y 4 que es 1.04 y la de la personas 3 con la 4 que es 1.62, eligiendo entonces la distancia promedio de las dos combinaciones o tres personas que en este caso es 1.33. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.34	0	
3	1.33		0

Repetimos el mismo criterio para eslabonar a la persona 3 con el conglomerado (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con la persona 2 que es 2.34 y la de la personas 3 con la 5 que es 2.45, eligiendo entonces la distancia promedio de las dos combinaciones o tres personas que en este caso es 2.40. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.34	0	
3	1.33	2.40	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 3 para eslabonarse al conglomerado formado por las personas (1,4) con la distancia menor de 1.33 como se muestra a continuación:

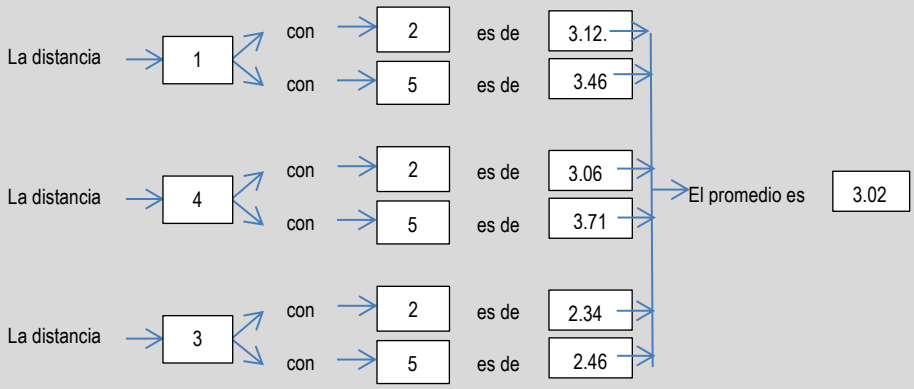
	(1,4)	(2,5)	3
(1,4)	0		
(2,5)	3.34	0	
3	1.33	2.40	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado (1,4,3) con el resto de las personas del conglomerado (2,5); al comparar (2,5) con las personas (1,4,3) se debe verificar **por única vez en la primera matriz de distancias** entre las personas 1,4 y 3 con las personas 2,5, eligiendo entonces la distancia promedio de las seis combinaciones o cinco personas que en este caso es 3.02. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)	3.02	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 2.87.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.90	1	4	1
2	3	1.27	2	5	2
3	2	1.33	1	3	1
4	1	3.02	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **cuarta columna** que en este caso fueron la 1 y la 4 a una distancia de **0.90** y en la **quinta columna** nos señala el número de conglomerado formado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 5 a una distancia de **1.27** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 3 al primer conglomerado formado por las personas 1 y 4 a una distancia de **1.33** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **3.02** Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

```
Variables estandarizadas, Distancia euclediana, Enlace de promedios
Pasos de amalgamación


      Número                                                  Número de
          de    Nivel de   Nivel de     Grupos     Nuevo   obs. en el
Paso  grupos  semejanzal  distancia  incorporados  grupo  grupo nuevo
   1       4     76.3492    0.87219  1          4      1            2
   2       3     66.1312    1.24900  2          5      2            2
   3       2     64.0314    1.32644  1          3      1            3
   4       1     19.8532    2.95563  1          2      1            5
```

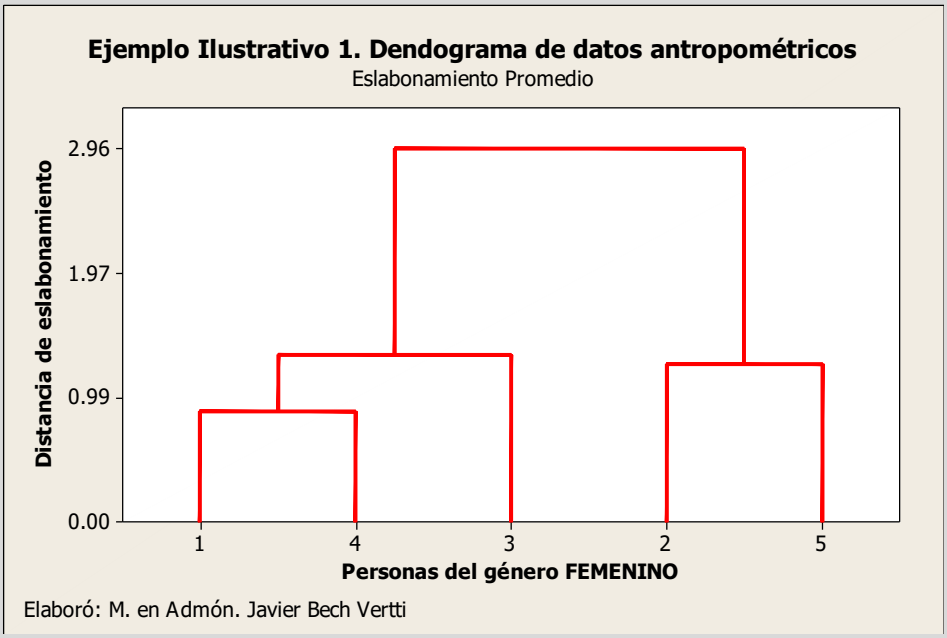
Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,4,3) y el conglomerado 2 por las personas

(2,5).

El dendrograma completo (diagrama de árbol) es una representación gráfica de la conformación en amalgama de observaciones o variables en un solo conglomerado. Cortar el dendrograma es similar a trazar una línea a lo largo del dendrograma para especificar la agrupación final.



Elaboró: M. en Admón. Javier Bech Vertti

¿Cómo se sabe dónde cortar el dendrograma? Primero convendría que se ejecute un análisis del conglomerado sin especificar la partición final. Se examinen los resultados de los niveles de similitud y de distancia y en el dendrograma. El nivel de similitud en cualquier paso es el porcentaje de la distancia mínima en el paso, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud o de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendrograma, si esto tiene sentido para sus datos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 3 y 4 los valores de distancia cambian abruptamente de 1.33 a 3.02 cuando se unen ambos grupos, por lo tanto, la solución más factible parece ser la de 2 conglomerados (1,4,3) y (2,5).

Respuesta al inciso c)

Nombre de los conglomerados

Una vez elegidos el número de conglomerados definitivo, para finalizar el proceso es deseable colocar un nombre apropiado a los conglomerados. En este caso el conglomerado 1 está formado por las personas 1,4 y 3 y el conglomerado 2 está formado por las personas 2 y 5. Para conocer el tipo de

personas que se encuentran en cada conglomerado se deben obtener la media de los datos originales de quienes la conforman variable por variable como se muestra a continuación:

Para el conglomerado 1:

- La media de la altura es de: (1.52+1.54+1.56)/3=1.54 Mts.
- La media del peso es de: (57.12+67.62+49.35)/3=58.03 Kg.
- La media de la cintura es de: (84.10+85.22+83.60)/3=84.30 Cms.

Para el conglomerado 2:

- La media de la altura es de: (1.67+1.63)/2=1.65 Mts.
- La media del peso es de: (76.85+87.22)/2=82.03 Kg.
- La media de la cintura es de: (86.20+87.52)/2=86.86 Cms.

Los resultados del conglomerado 1 nos indican que las mujeres de este grupo son más bajitas, más delgadas y de menor cintura que las mujeres del conglomerados 2, por lo que un posible nombre para el conglomerado 1 podría ser "talla pequeña" y para el conglomerado 2 "talla grande"

2.2.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE APRENDIZAJE 2.2.1.1 ANÁLISIS DE CONGLOMERADOS. MEDIDA DE DISTANCIA.

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de los hombres mexicanos, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras . El estudio antropométrico contempló 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y la cintura de cinco personas del género **MASCULINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Cintura (cms)
1	1.73	63.56	89.10
2	1.85	83.99	92.20
3	1.70	74.91	90.22
4	1.63	54.48	88.60

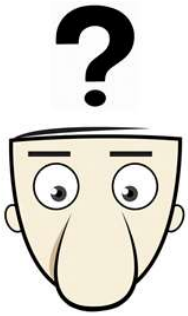
5	1.93	95.34	93.50
Promedio:	**1.77**	**74.46**	**90.72**
Desviación estándar	**0.12**	**16.16**	**2.08**

a) Obtenga la pareja de personas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.
c) Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

2.2.1.1 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra al final del cuaderno de trabajo en el anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 2.2.1.1 ANÁLISIS DE CONGLOMERADOS. MATRIZ DE DISTANCIA



Suponga que se desea segmentar a los clientes de un supermercado con base a la actitud que tienen cuando salen de compras. De acuerdo a una investigación previa, se identificaron tres variables de actitud y se pidió a los cinco entrevistados que expresaran su grado de acuerdo con tres afirmaciones en una escala de siete puntos en la cual 1 significa en desacuerdo y 7 significa de acuerdo.

Las afirmaciones podrían ser las siguientes:

V1 Salir de compras es divertido
V2 Puedo ahorrar mucho dinero si comparo precios
V3 Prefiero salir de compras con mi familia

Si los resultados para cada una de las afirmaciones se considera que son variables, se tienen tres, las cuales se representarán con V1, V2 y V3; por otra parte, las respuestas de cada uno de los entrevistados constituyen los cinco casos, que se representarán con C1, C2,…,C5. Los valores de las variables para

los cinco casos, se presentan a continuación:

ENTREVISTADO	VARIABLE		
	V1	V2	V3
C1	1	2	1
C2	5	7	5
C3	2	1	1
C4	7	7	5
C5	3	2	2
Promedio:	3.6	3.8	2.8
Desviación estándar	2.41	2.95	2.049

a) Obtenga la pareja de personas entrevistadas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.
c) Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

C.A. EJEMPLO ILUSTRATIVO EN MINITAB 17

EJEMPLO ILUSTRATIVO INTEGRAL EN MINITAB 17. ANÁLISIS DE CONGLOMERADOS. MEDIDA DE DISTANCIA

MINITAB 17

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de los hombres mexicanos, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras. El estudio antropométrico contempló 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y la cintura de cinco personas del género **FEMENINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Cintura (cms)
1	1.52	57.12	84.10
2	1.67	76.85	86.20
3	1.54	67.62	85.22
4	1.56	49.35	83.60
5	1.63	87.22	87.52

Promedio:	**1.58**	**67.63**	**85.33**
Desviación estándar	**0.06**	**15.11**	**1.59**

Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

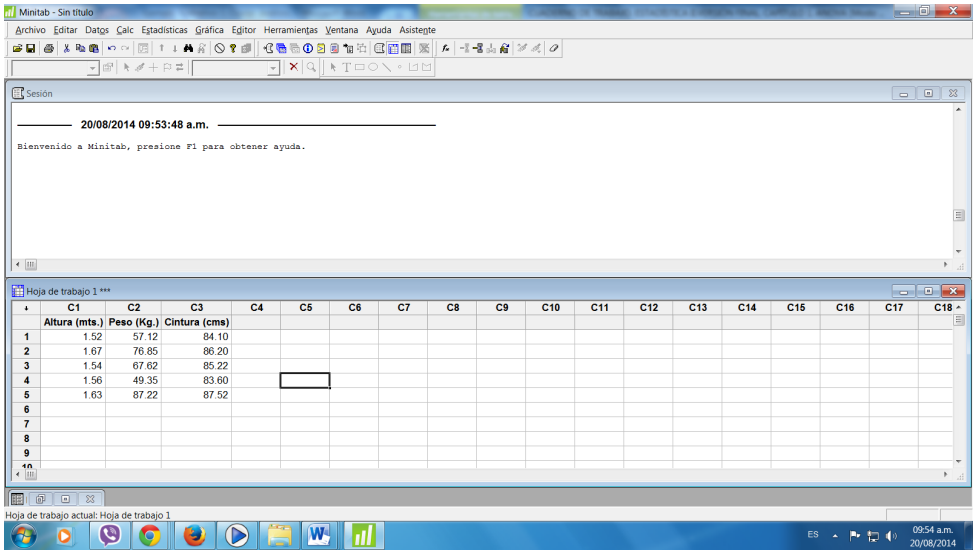
Solución. Eslabonamiento Simple o Individual:

Método	La distancia entre dos conglomerados es...	Razones para utilizar este método
Simple o Individual	La distancia mínima entre un elemento en un conglomerado y un elemento en el otro conglomerado. Se conoce también como el método del "vecino más cercano"	Más apropiado para observaciones que están claramente separadas. Cuando están cerca, el enlace simple o individual tiende a identificar conglomerados largos similares a cadenas que pueden tener una separación relativamente grande entre elementos en cualquier extremo de la cadena.

· Con el eslabonamiento simple, o del "vecino más cercano", la distancia entre dos conglomerados es la distancia mínima entre una observación en un conglomerado y una observación en el otro conglomerado. El eslabonamiento simple es una opción apropiada cuando los conglomerados están claramente separados. Cuando las observaciones están cerca entre sí, el eslabonamiento simple tiende a identificar conglomerados largos similares a cadenas que pueden tener una separación relativamente grande entre observaciones en cualquier extremo de la cadena. ·

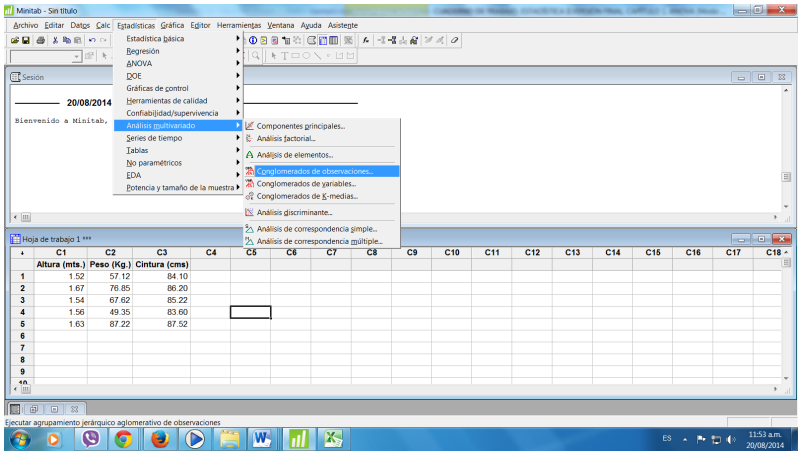
Cuando el número de observaciones **es extenso y/o existen muchas variables,** los cálculos manuales son tediosos. Existen muchos paquetes de software que pueden mostrar los resultados entre ellos **Minitab.**

Comenzamos introduciendo los datos en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:

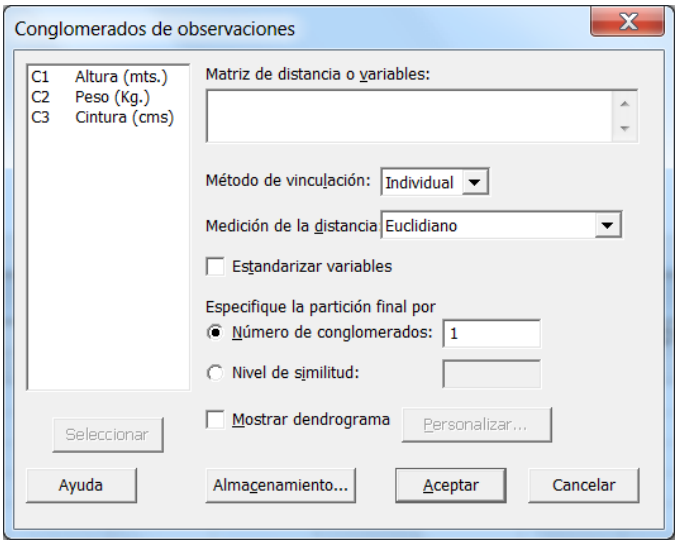


Elija de la barra menú ***Estadísticas> Análisis Multivariado>Conglomerado de observaciones.***

La opción de ***conglomerados de observaciones*** se utiliza para clasificar los elementos en grupos cuando éstos se desconocen inicialmente. En **Minitab** este procedimiento utiliza **el método jerárquico aglomerativo** que comienza con todos los elementos separados, cada uno de ellos formando su propio conglomerado. En el primer paso, los dos elementos más cercanos se unen entre sí. En el siguiente paso, un tercer elemento se une a los dos primeros u otros dos elementos se unen para formar un conglomerado diferente. Este proceso continúa hasta que todos los conglomerados se unan en un solo conglomerado; sin embargo, este conglomerado individual no es útil para propósitos de clasificación, por lo tanto, usted debe decidir cuántos grupos son lógicos para sus datos y clasificarlos adecuadamente



Aparece el siguiente cuadro de Diálogo:



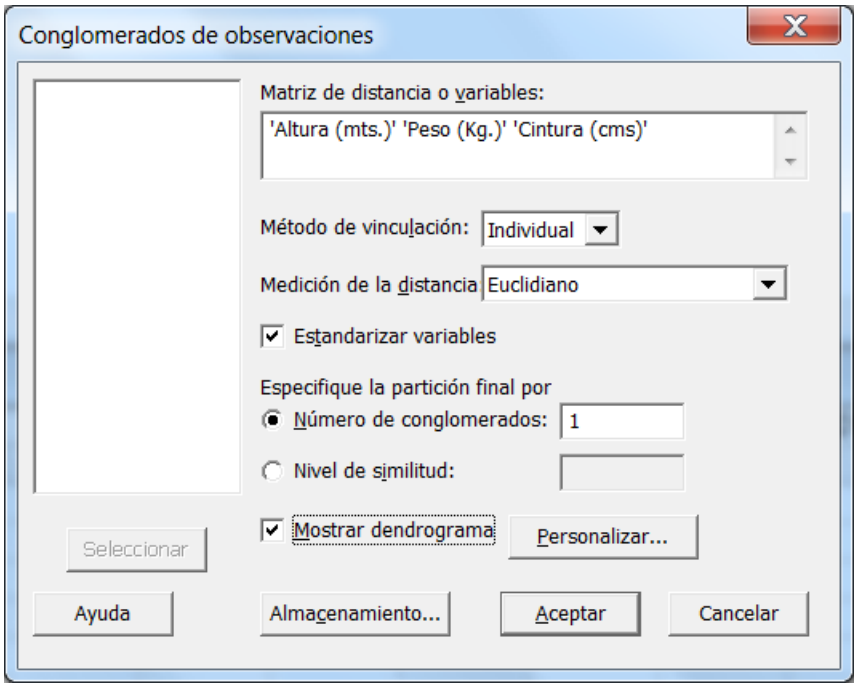
En ***Matriz de distancia o variables***, ingrese las tres variables dando doble click sobre cada una de ellas o seleccione las tres y oprima el botón Seleccionar.

En ***Método de vinculación***, elija ***Individual*** en ***Medición de la distancia*** elija ***Euclidiano***.

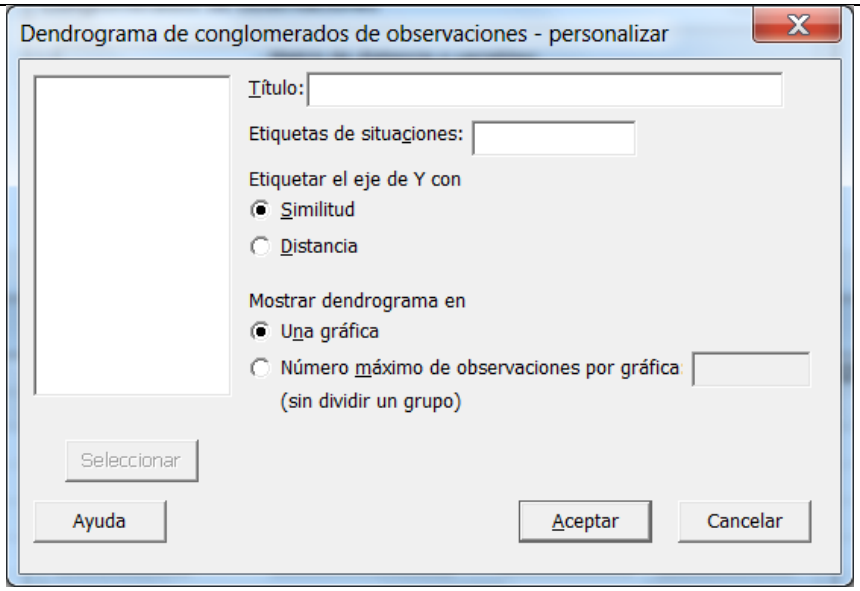
Marque la opción ***Estandarizar variables***.

Nota: Esta opción permite convertir todas las variables a una escala común al sustraer las medias y dividir entre la desviación estándar antes del cálculo de la matriz de distancia sobretodo si las variables están en unidades diferentes y usted desea minimizar el efecto de las diferencias de escala.

Marque la opción ***Mostrar dendograma***.

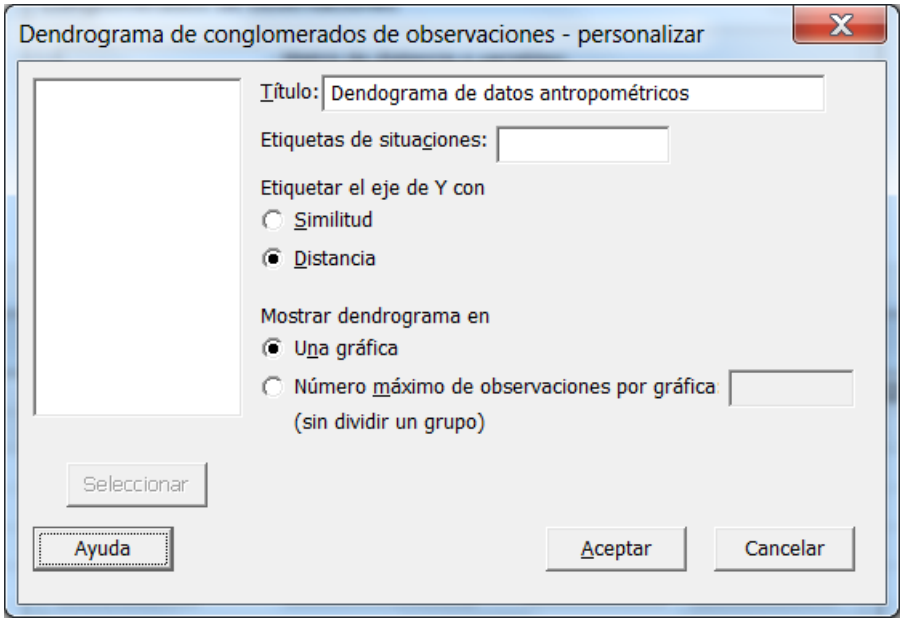


Haga clic en ***Personalizar.*** Aparece el siguiente cuadro de diálogo:



En ***Título***, ingrese ***Dendograma de datos antropométricos***.

En ***Etiquetar el eje de Y con*** elija ***Distancia.***



Haga clic en ***Aceptar*** en cada cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts., Peso (Kg.), ...

```
Variables estandarizadas, Distancia euclediana, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	**Nivel de distancia**	**Conglomerados incorporados**		Nuevo conglomerado
1	4	76.3492	**0.87219**	**1**	**4**	1
2	3	71.8157	**1.03937**	**1**	**3**	1
3	2	66.1312	**1.24900**	**2**	**5**	2
4	1	39.6809	**2.22443**	**1**	**2**	1

Interpretación de los resultados

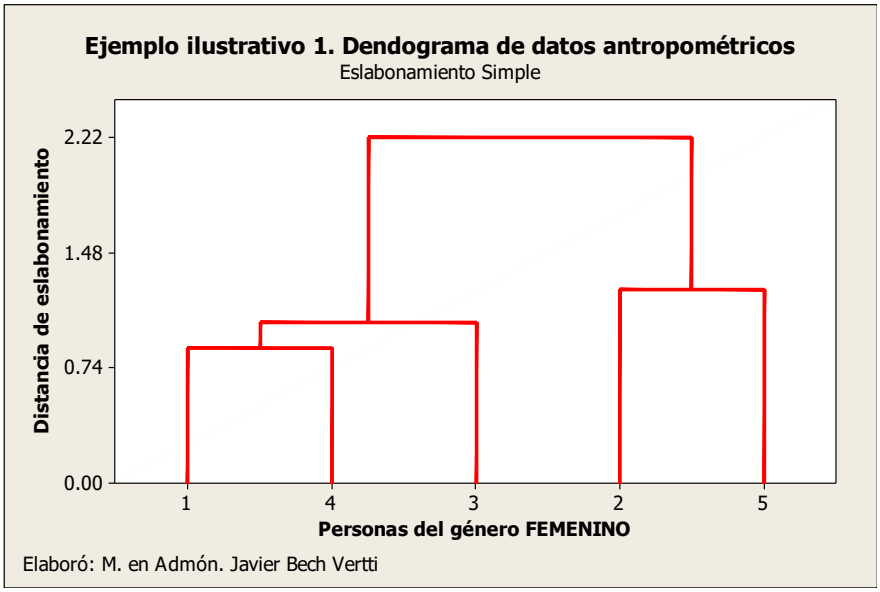
Minitab muestra los pasos de eslabonamiento en la ventana Sesión. En cada paso, se unen dos conglomerados (personas). La tabla muestra los conglomerados que se unieron, la distancia entre ellos, el nivel de similitud correspondiente, el número de identificación del conglomerado nuevo (este número siempre corresponde al número más pequeño de los dos conglomerados que se unieron), el número de observaciones en el conglomerado nuevo y el número de conglomerados. El eslabonamiento continúa hasta que sólo queda un conglomerado.

Los pasos de eslabonamiento muestran que en principio, en el paso 1, las personas más parecidas son la 1 y la 4 y forman el primer conglomerado a una distancia de eslabonamiento de 0.87219, posteriormente en el paso 2 se incorpora a este conglomerado la persona 3 a una distancia de eslabonamiento de 1.03937. En el paso 3 se observa que se forma un segundo conglomerado con las personas 2 y 5 a una distancia de eslabonamiento de 1.24900 para finalmente unirse en el paso 4 ambos conglomerados a una distancia de eslabonamiento de 2.22443

Esto indica que tal vez dos conglomerados son razonablemente suficientes para la partición o solución final. Si esta agrupación tiene sentido intuitivo para los datos, entonces es probable que sea una elección apropiada.

Dendrograma

La agrupación final de conglomerados (también denominada partición final) es la agrupación de conglomerados que, si los resultados son favorables, identificará grupos cuyas observaciones o variables compartan características comunes. La decisión acerca de la agrupación final también se denomina Dendograma. El dendograma completo (diagrama de árbol) es una representación gráfica de la conformación en amalgama de los elementos en un solo conglomerado.

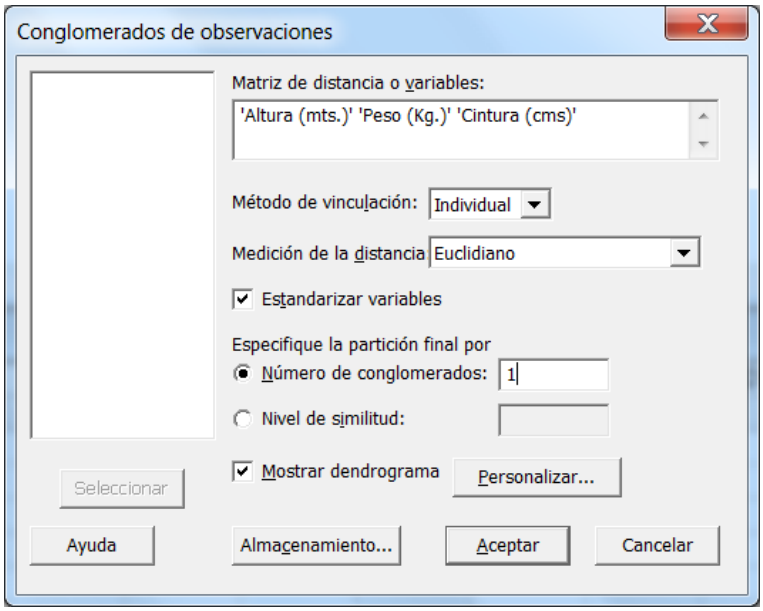


Por opción predeterminada, el dendograma presenta el nivel de similitud en el eje vertical (de forma alterna, como en este caso, puede mostrar el nivel de distancia) y las diferentes observaciones se enumeran con el eje horizontal. La gráfica muestra la manera en la cual se formaron los conglomerados, bien sea uniendo dos observaciones individuales o pareando una observación individual con un conglomerado existente (como en este caso la persona 3). Puede determinar en cuáles distancias de eslabonamiento se forman los conglomerados así como la composición de los conglomerados de la solución final.

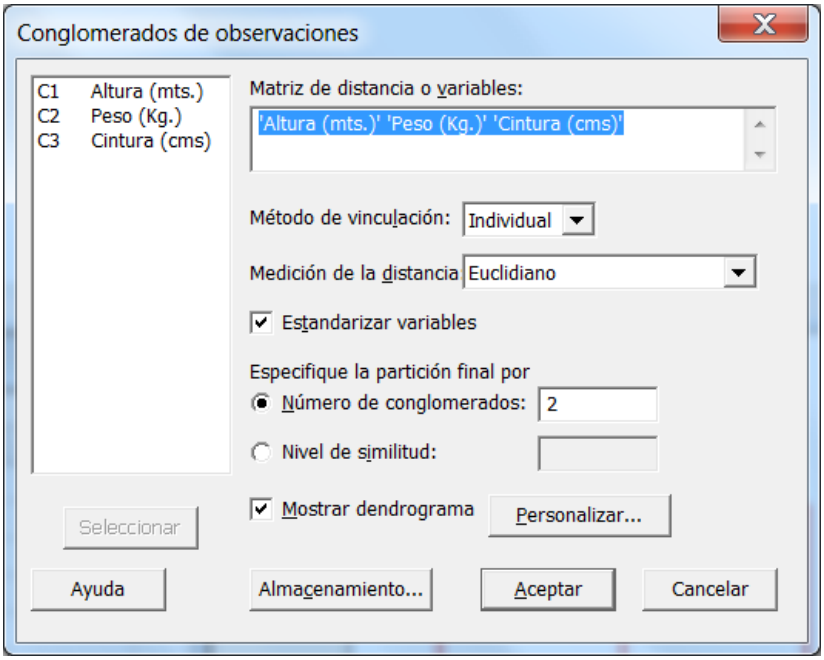
¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados, aunque en la práctica puede ejecutar primero un análisis de conglomerados **sin especificar una partición final con el objeto de decidir dónde colocarla**. Examine los niveles de distancia en los resultados de la ventana Sesión y en el dendograma. Puede ver las distancias de eslabonamiento en el dendograma colocando el puntero de su ratón sobre una línea horizontal. La distancia de eslabonamiento en cualquier paso es el valor de la distancia mínima en el paso, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Después de elegir dónde desea realizar su partición, vuelva a ejecutar el procedimiento del conglomerado, utilizando, el Número de conglomerados para obtener un número establecido de conjuntos de la siguiente manera:

En la parte inicial del análisis apareció el siguiente cuadro de Diálogo:



Ahora en ***Especifique la partición final por***, elija ***Número de conglomerados*** e ingrese ***2***.



Haga clic en ***Aceptar*** en el cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts., Peso (Kg.), ...

```
Variables estandarizadas, Distancia euclediana, Enlace simple
Pasos de amalgamación
```

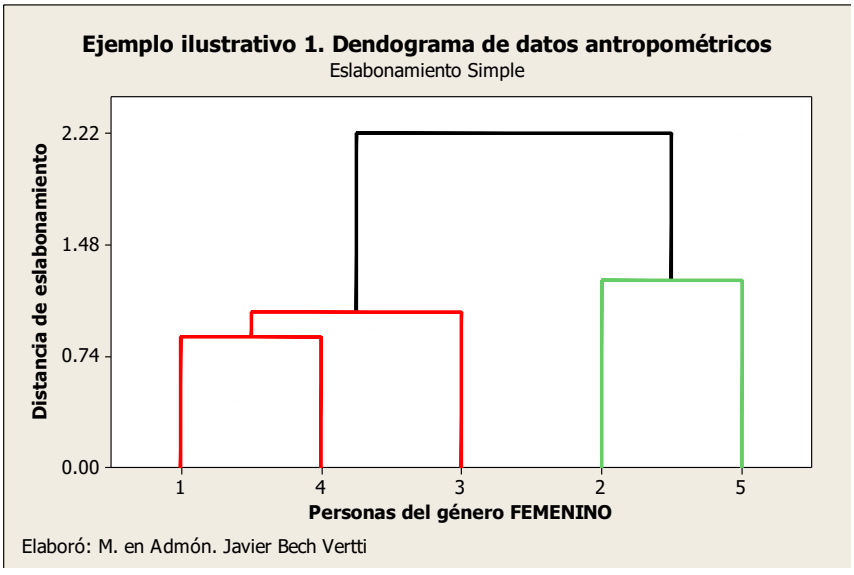
Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado
1	4	76.3492	0.87219	1	4	1
2	3	71.8157	1.03937	1	3	1
3	2	66.1312	1.24900	2	5	2
4	1	39.6809	2.22443	1	2	1

Partición final
Número de conglomerados: 2

	Número de observaciones	Dentro de la suma de cuadrados del conglomerado	Distancia promedio desde el centroide	Distancia máxima desde centroide
Conglomerado1	3	1.48147	0.665049	0.856782
Conglomerado2	2	0.78000	0.624502	0.624502

Cuando usted especifica la partición o solución final, Minitab muestra entre otras una tabla adicional que resume cada conglomerado según el número de observaciones, en este caso el primer conglomerado está formado por tres mujeres (1,4 y 3) y el segundo conglomerado por las mujeres (2 y 5)

Dendrograma



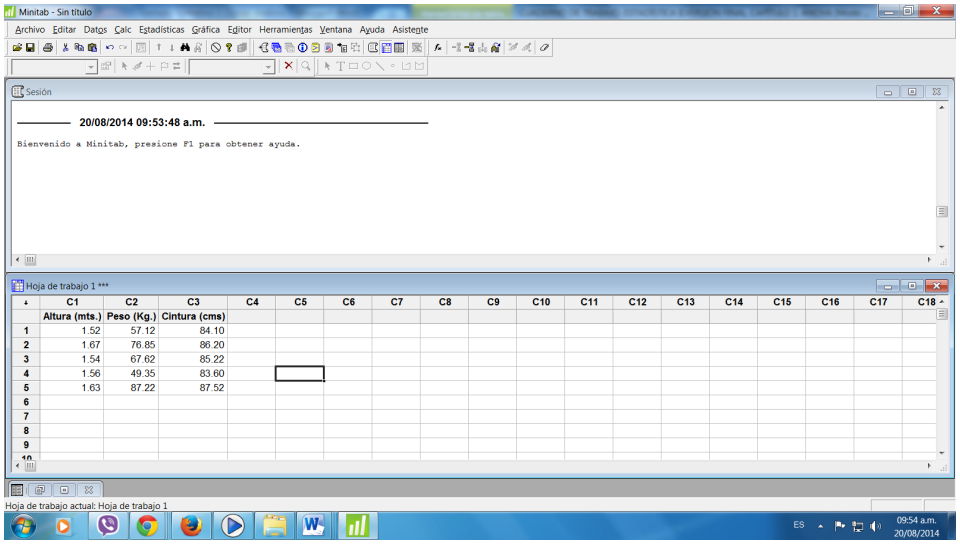
· Con el eslabonamiento completo, o "del vecino más lejano", la distancia entre dos conglomerados es la distancia máxima entre una observación en un conglomerado y una observación en el otro conglomerado. Este método garantiza que todas las observaciones en un conglomerado se encuentren dentro de una distancia máxima y tiende a producir conglomerados con diámetros similares. Los resultados pueden ser sensibles a valores atípicos.

Examine los conglomerados resultantes en la partición o solución final para observar si la agrupación parece lógica (Las personas que forman cada conglomerado aparecen con un color distinto en cada conglomerado). La observación de dendogramas para determinar diferentes agrupaciones finales puede ayudarle a decidir cuáles de ellas tienen más sentido para sus datos.

Solución. Eslabonamiento Completo:

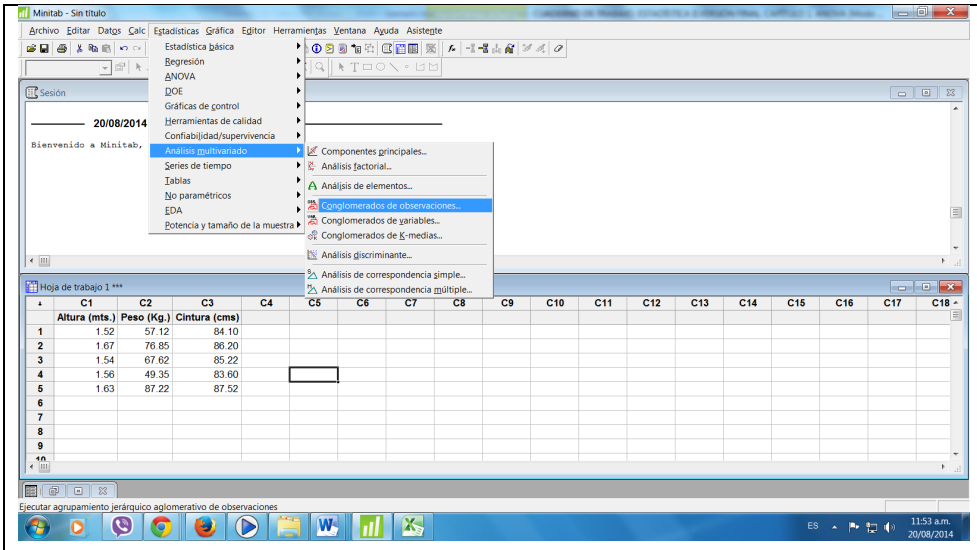
Método	La distancia entre dos conglomerados es...	Razones para utilizar este método
Completo	La distancia máxima entre un elemento en un conglomerado y un elemento en el otro conglomerado. Se conoce también como el método del "vecino más lejano".	Asegura que todos los elementos en un conglomerado se encuentren dentro de una distancia máxima y tiende a producir conglomerados con diámetros similares. Los resultados pueden ser sensibles a valores atípicos.

Comenzamos introduciendo los datos en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:

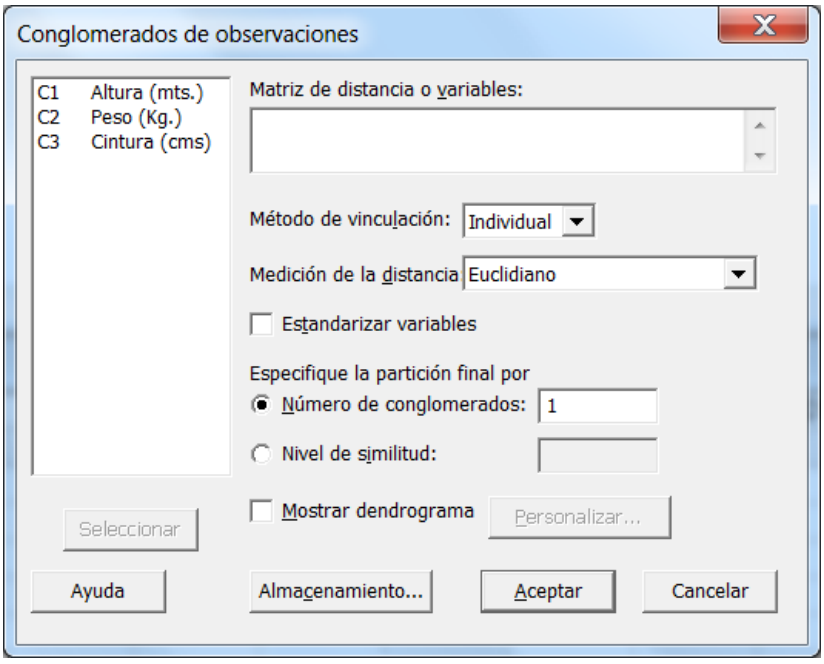


	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	Altura (mts.)	Peso (Kg.)	Cintura (cms)															
1	1.52	57.12	84.10															
2	1.67	76.85	86.20															
3	1.54	67.62	85.22															
4	1.56	49.35	83.60															
5	1.63	87.22	87.52															
6																		
7																		
8																		
9																		
10																		

Elija de la barra menú ***Estadísticas> Análisis Multivariado>Conglomerado de observaciones.***



Aparece el siguiente cuadro de Diálogo:

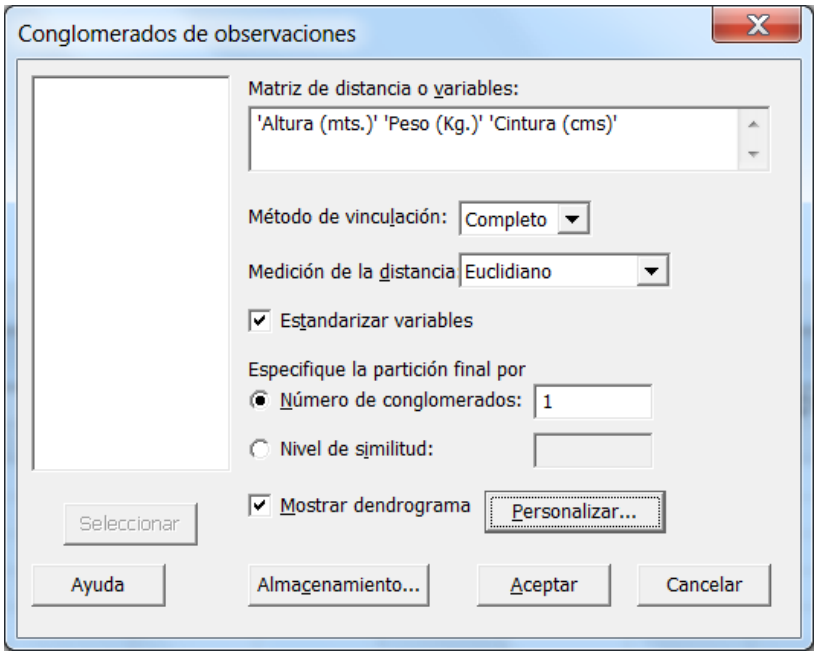


En ***Matriz de distancia o variables***, ingrese las tres variables dando doble click sobre cada una de ellas o seleccione las tres y oprima el botón Seleccionar.

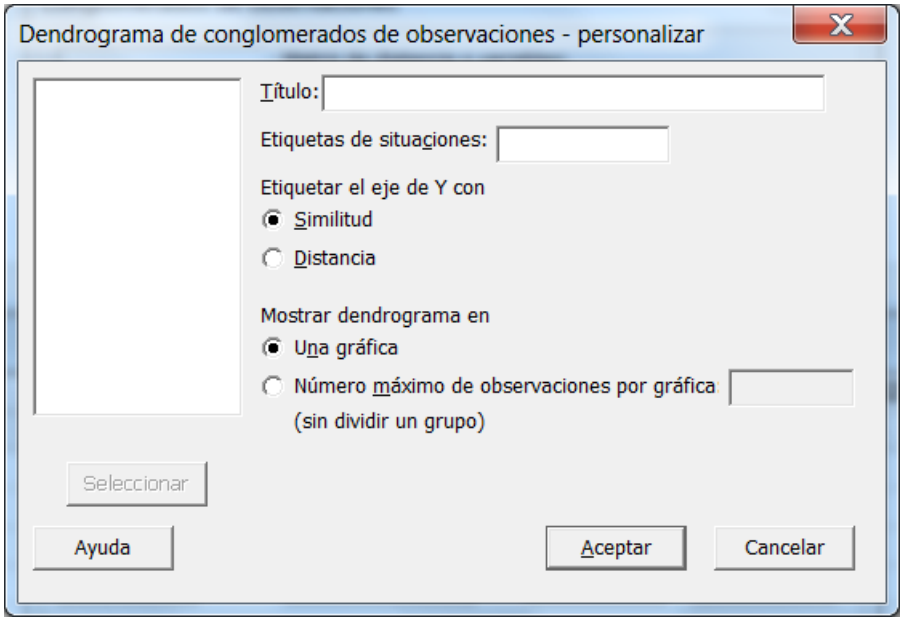
En ***Método de vinculación***, elija ***Completo*** en ***Medición de la distancia*** elija ***Euclidiano***.

Marque la opción ***Estandarizar variables***.

Marque la opción ***Mostrar dendograma***.

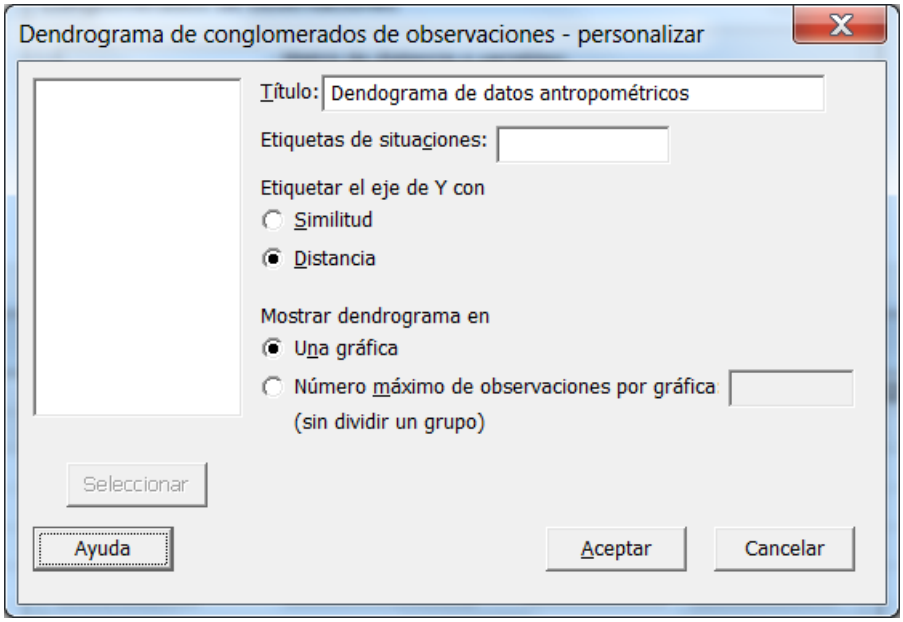


Haga clic en ***Personalizar.*** Aparece el siguiente cuadro de diálogo:



En ***Título***, ingrese ***Dendograma de datos antropométricos***.

En ***Etiquetar el eje de Y con*** elija ***Distancia.***



Haga clic en ***Aceptar*** en cada cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts., Peso (Kg.), ...

```
Variables estandarizadas, Distancia euclediana, Enlace completo
Pasos de amalgamación
```

Paso	Número de grupos	Nivel de semejanzal	Nivel de distancia	Grupos incorporados	Nuevo grupo	Número de obs. en el grupo nuevo
1	4	76.3492	0.87219	1 4	1	2
2	3	66.1312	1.24900	2 5	2	2
3	2	56.2472	1.61350	1 3	1	3
4	1	0.0000	3.68777	1 2	1	5

Interpretación de los resultados

Minitab muestra los pasos de eslabonamiento en la ventana Sesión. En cada paso, se unen dos conglomerados (personas). La tabla muestra los conglomerados que se unieron, la distancia entre ellos, el nivel de similitud correspondiente, el número de identificación del conglomerado nuevo (este número siempre corresponde al número más pequeño de los dos conglomerados que se unieron), el número de observaciones en el conglomerado nuevo y el número de conglomerados. El eslabonamiento continúa hasta que sólo queda un conglomerado.

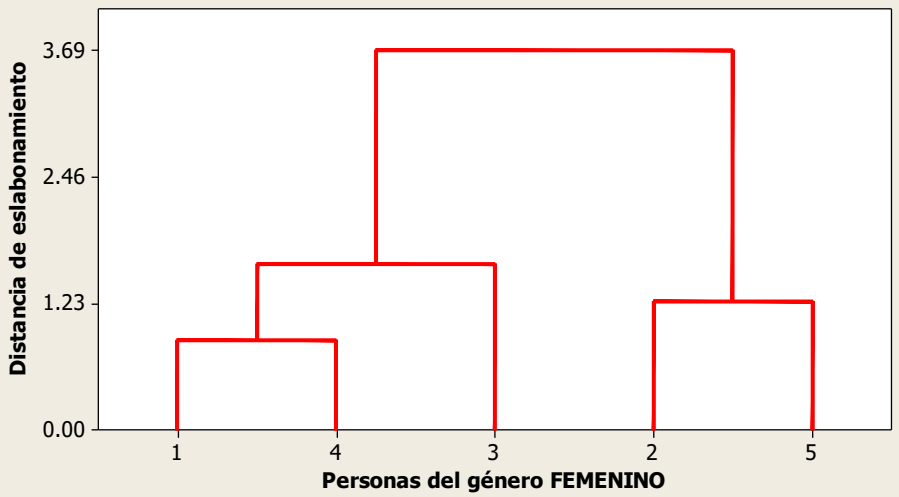
Los pasos de eslabonamiento muestran que en principio, en el paso 1, las personas más parecidas son la 1 y la 4 y forman el primer conglomerado a una distancia de eslabonamiento de 0.87219, posteriormente en el paso 2 se forma un segundo conglomerado con las personas 2 y 5 a una distancia de eslabonamiento de 1.24900. En el paso 3 se incorpora al primer conglomerado la persona 3 a una distancia de eslabonamiento de 1.61350 para finalmente unirse en el paso 4 ambos conglomerados a una distancia de eslabonamiento de 3.68777.

Esto indica que tal vez dos conglomerados son razonablemente suficientes para la partición o solución final. Si esta agrupación tiene sentido intuitivo para los datos, entonces es probable que sea una elección apropiada.

Dendrograma

Ilustra la información en la tabla de eslabonamiento en la forma de un diagrama de árbol.

Ejemplo ilustrativo 1. Dendograma de datos antropométricos



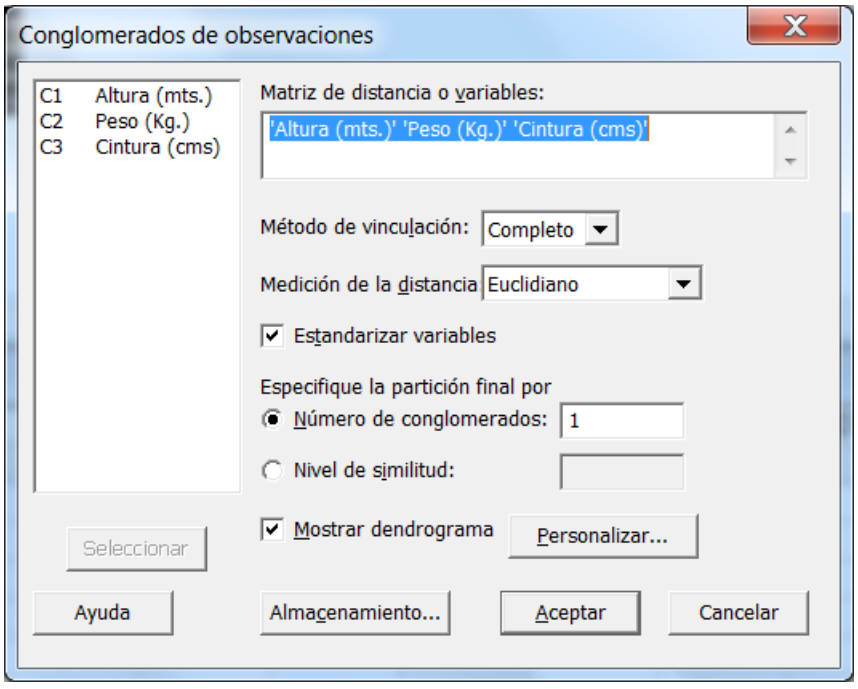
Elaboró: M. en Admón. Javier Bech Vertti

Por opción predeterminada, el dendograma presenta el nivel de similitud en el eje vertical (de forma alterna, como en este caso, puede mostrar el nivel de distancia) y las diferentes observaciones se enumeran con el eje horizontal. La gráfica muestra la manera en la cual se formaron los conglomerados, bien sea uniendo dos observaciones individuales o pareando una observación individual con un conglomerado existente (como en este caso la persona 3). Puede determinar en cuáles distancias de eslabonamiento se forman los conglomerados así como la composición de los conglomerados de la solución final.

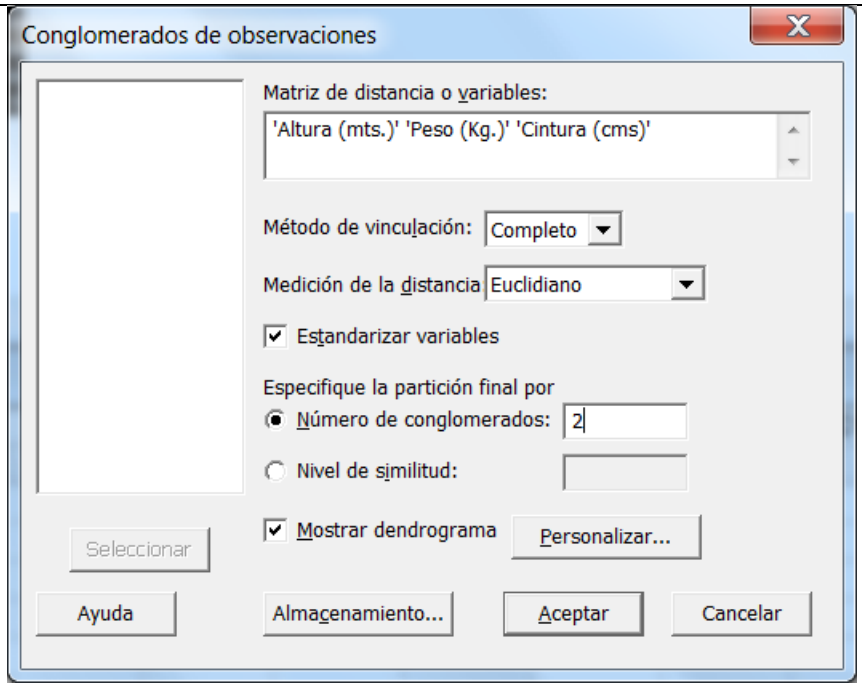
¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados, aunque en la práctica puede ejecutar primero un análisis de conglomerados **sin especificar una partición final con el objeto de decidir dónde colocarla**. Examine los niveles de distancia en los resultados de la ventana Sesión y en el dendograma. Puede ver las distancias de eslabonamiento en el dendograma colocando el puntero de su ratón sobre una línea horizontal. La distancia de eslabonamiento en cualquier paso es el valor de la distancia mínima en el paso, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Después de elegir dónde desea realizar su partición, vuelva a ejecutar el procedimiento del conglomerado, utilizando, el Número de conglomerados para obtener un número establecido de conjuntos de la siguiente manera:

En la parte inicial del análisis del eslabonamiento completo apareció el siguiente cuadro de Diálogo:



Ahora en ***Especifique la partición final por***, elija ***Número de conglomerados*** e ingrese ***2***.



Haga clic en ***Aceptar*** en el cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts., Peso (Kg.), ...

```
Variables estandarizadas, Distancia euclediana, Enlace completo
Pasos de amalgamación
```

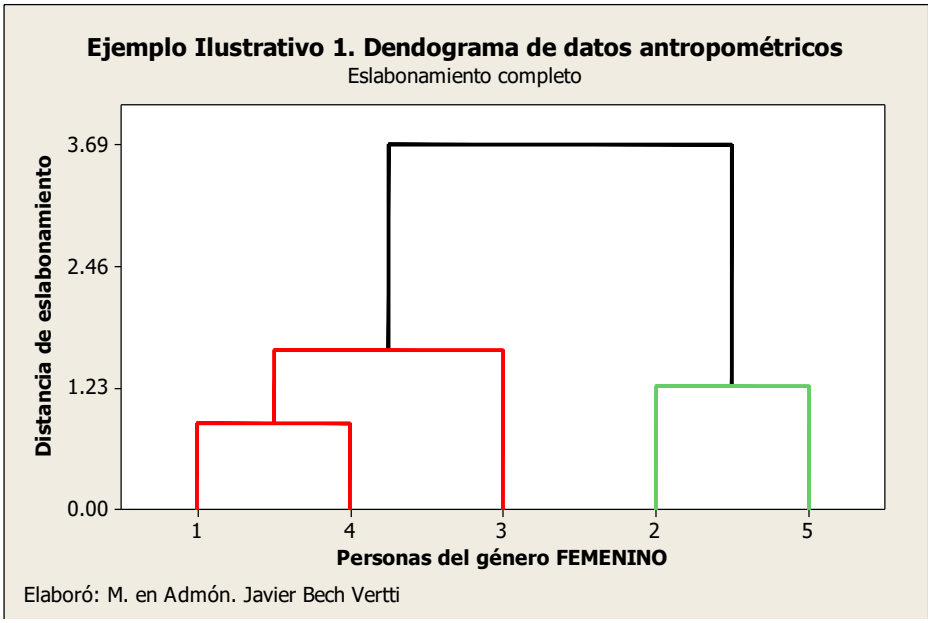
Paso	Número de grupos	Nivel de semejanzal	Nivel de distancia	Grupos incorporados		Nuevo grupo	Número de obs. en el grupo nuevo
1	4	76.3492	0.87219	1	4	1	2
2	3	66.1312	1.24900	2	5	2	2
3	2	56.2472	1.61350	1	3	1	3
4	1	0.0000	3.68777	1	2	1	5

Partición final
Número de grupos: 2

	Número de observaciones	Dentro de la suma de cuadrados del grupo	Distancia promedio desde el centroide	Distancia máxima desde centroide
Grupo1	3	1.48147	0.665049	0.856782
Grupo2	2	0.78000	0.624502	0.624502

Cuando usted especifica la partición o solución final, Minitab muestra entre otras una tabla adicional que resume cada conglomerado según el número de observaciones, en este caso el primer conglomerado está formado por tres mujeres (1,4 y 3) y el segundo conglomerado por las mujeres (2 y 5)

Dendrograma



Elaboró: M. en Admón. Javier Bech Vertti

Examine los conglomerados resultantes en la partición o solución final para observar si la agrupación parece lógica (Las personas que forman cada conglomerado aparecen con un color distinto en cada conglomerado). La observación de dendogramas para determinar diferentes agrupaciones finales puede ayudarle a decidir cuáles de ellas tienen más sentido para sus datos.

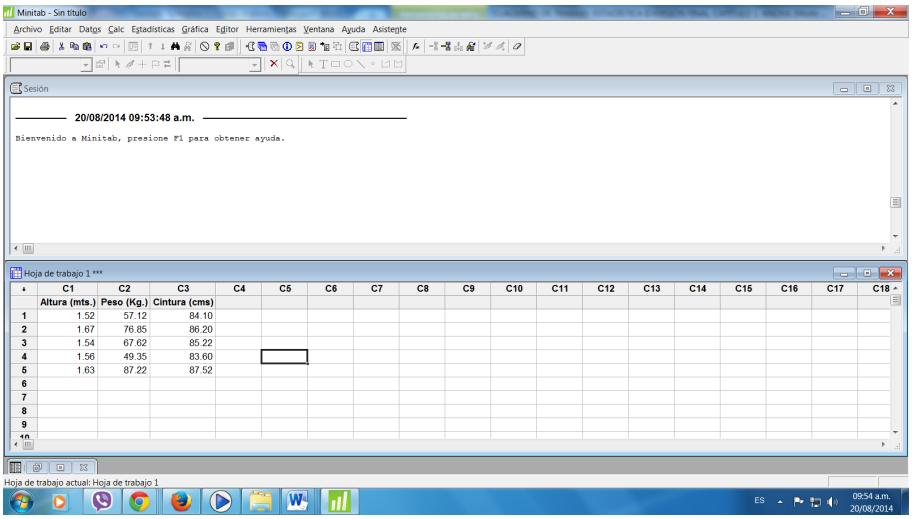
Solución. Eslabonamiento Promedio:

Método	La distancia entre dos conglomerados es...	Razones para utilizar este método
Promedio	La distancia media entre un elemento en un conglomerado y un elemento en el otro conglomerado	En vista de que los métodos de enlace simples o completos agrupan conglomerados sobre la base de distancias entre pares individuales, el enlace promedio utiliza una medida de ubicación más central.

Comenzamos introduciendo los datos en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:

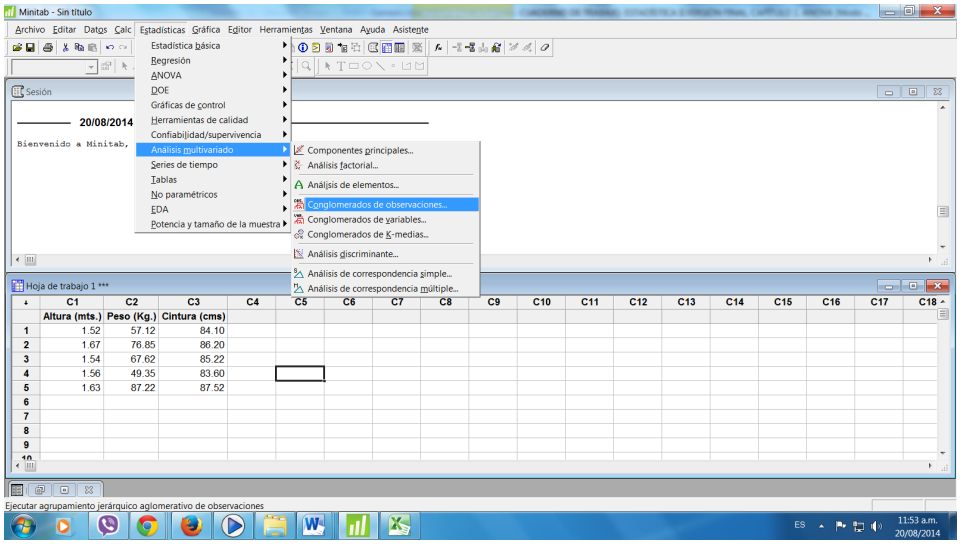
· Con el eslabonamiento promedio, la distancia entre dos conglomerados es la distancia media entre una observación en un conglomerado y una observación en el otro conglomerado. En vista de que los métodos de eslabonamiento simple o completos agrupan conglomerados sobre la base de distancias entre pares individuales, el eslabonamiento promedio

utiliza una medida de ubicación más central.



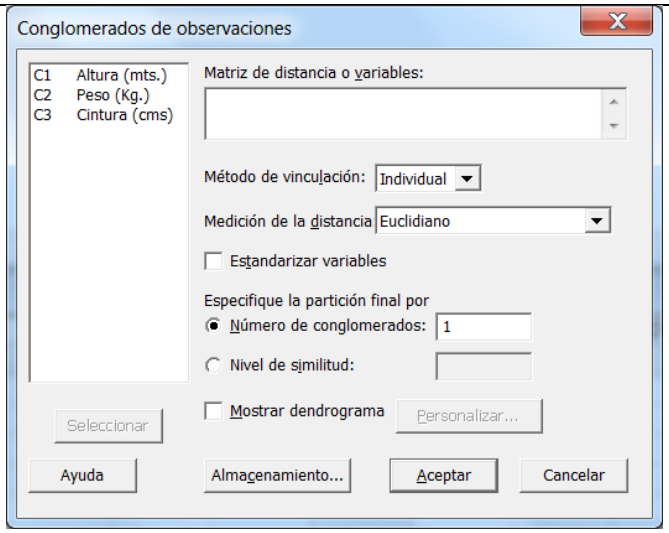
	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	Altura (mts.)	Peso (Kg.)	Cintura (cms)															
1	1.52	57.12	84.10															
2	1.67	76.85	86.20															
3	1.54	67.62	85.22															
4	1.56	49.35	83.60															
5	1.63	87.22	87.52															
6																		
7																		
8																		
9																		
10																		

Elija de la barra menú ***Estadísticas> Análisis Multivariado>Conglomerado de observaciones.***



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	Altura (mts.)	Peso (Kg.)	Cintura (cms)															
1	1.52	57.12	84.10															
2	1.67	76.85	86.20															
3	1.54	67.62	85.22															
4	1.56	49.35	83.60															
5	1.63	87.22	87.52															
6																		
7																		
8																		
9																		
10																		

Aparece el siguiente cuadro de Diálogo:

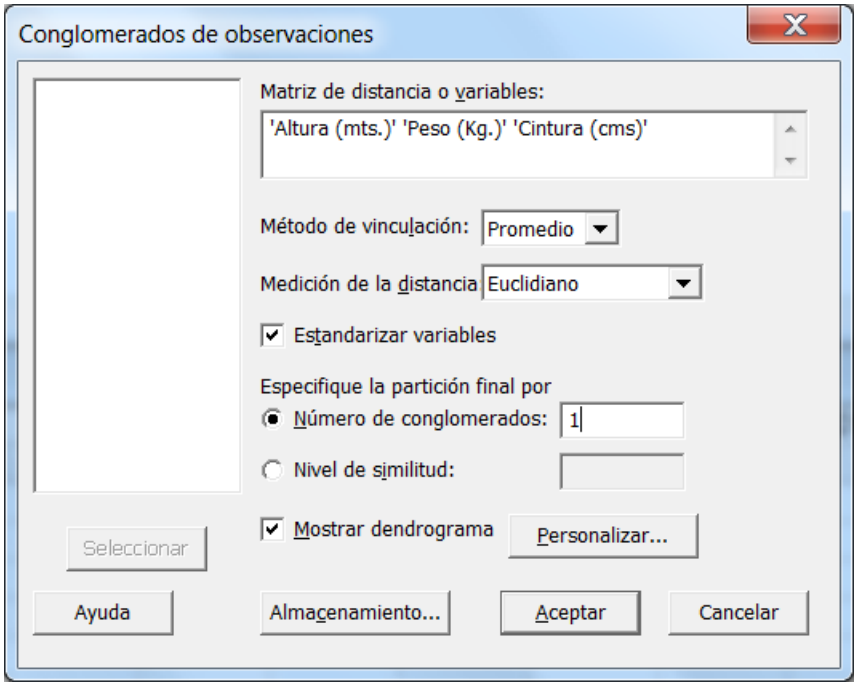


En ***Matriz de distancia o variables***, ingrese las tres variables dando doble click sobre cada una de ellas o seleccione las tres y oprima el botón Seleccionar.

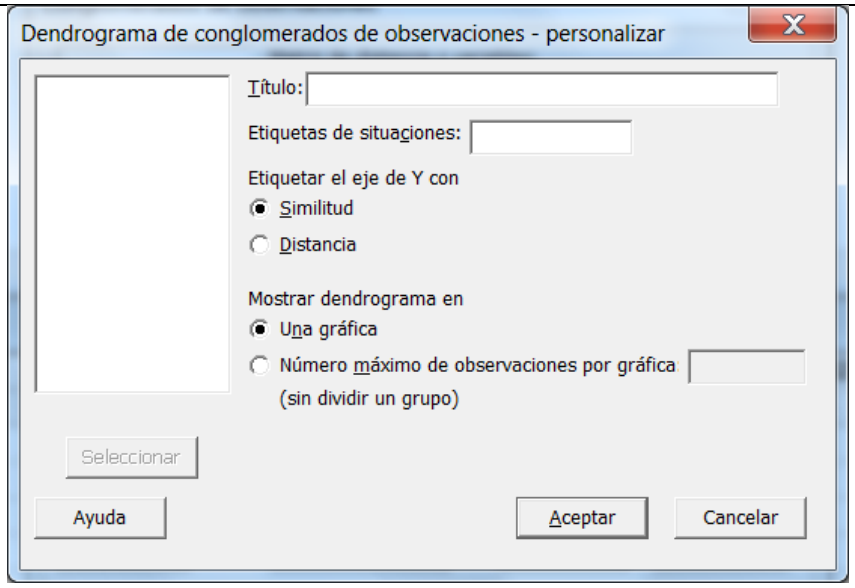
En ***Método de vinculación***, elija ***Promedio*** en ***Medición de la distancia*** elija ***Euclidiano***.

Marque la opción ***Estandarizar variables***.

Marque la opción ***Mostrar dendograma***.

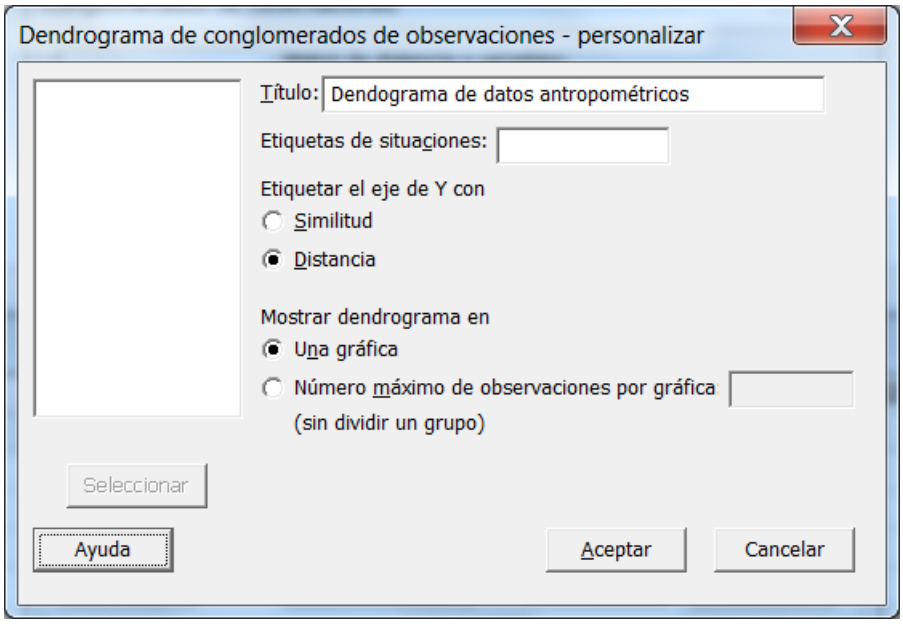


Haga clic en ***Personalizar.*** Aparece el siguiente cuadro de diálogo:



En ***Título***, ingrese ***Dendograma de datos antropométricos***.

En ***Etiquetar el eje de Y con*** elija ***Distancia.***



Haga clic en ***Aceptar*** en cada cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts., Peso (Kg.), ...

```
Variables estandarizadas, Distancia euclediana, Enlace de promedios
Pasos de amalgamación
```

Paso	Número de grupos	Nivel de semejanzal	Nivel de distancia	Grupos incorporados		Nuevo grupo	Número de obs. en el grupo nuevo
1	4	76.3492	0.87219	1	4	1	2
2	3	66.1312	1.24900	2	5	2	2
3	2	64.0314	1.32644	1	3	1	3
4	1	19.8532	2.95563	1	2	1	5

Interpretación de los resultados

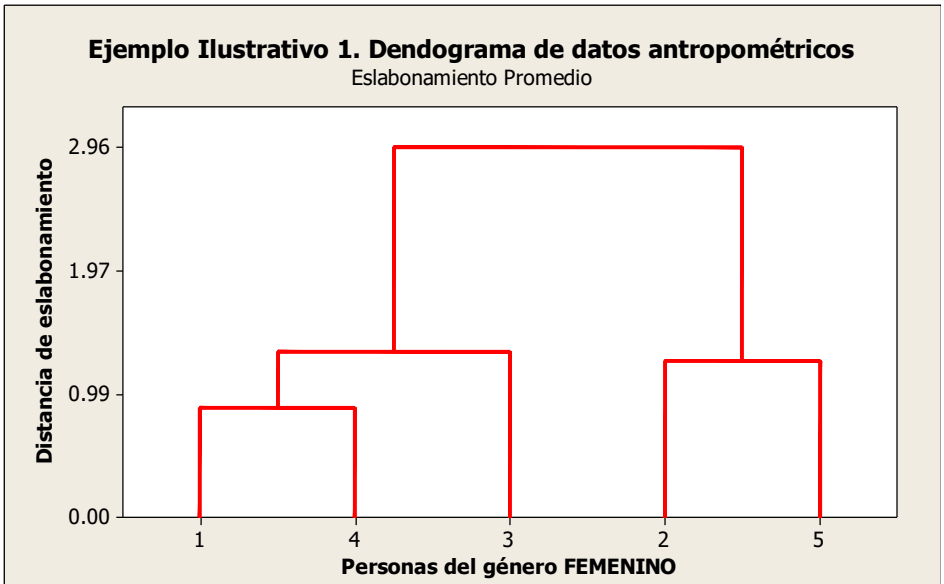
Minitab muestra los pasos de eslabonamiento en la ventana Sesión. En cada paso, se unen dos conglomerados (personas). La tabla muestra los conglomerados que se unieron, la distancia entre ellos, el nivel de similitud correspondiente, el número de identificación del conglomerado nuevo (este número siempre corresponde al número más pequeño de los dos conglomerados que se unieron), el número de observaciones en el conglomerado nuevo y el número de conglomerados. El eslabonamiento continúa hasta que sólo queda un conglomerado.

Los pasos de eslabonamiento muestran que en principio, en el paso 1, las personas más parecidas son la 1 y la 4 y forman el primer conglomerado a una distancia de eslabonamiento de 0.87219, posteriormente en el paso 2 se forma un segundo conglomerado con las personas 2 y 5 a una distancia de eslabonamiento de 1.24900. En el paso 3 se incorpora al primer conglomerado la persona 3 a una distancia de eslabonamiento de 1.32644 para finalmente unirse en el paso 4 ambos conglomerados a una distancia de eslabonamiento de 2.95563

Esto indica que tal vez dos conglomerados son razonablemente suficientes para la partición o solución final. Si esta agrupación tiene sentido intuitivo para los datos, entonces es probable que sea una elección apropiada.

Dendrograma

Ilustra la información en la tabla de eslabonamiento en la forma de un diagrama de árbol.



Elaboró: M. en Admón. Javier Bech Vertti

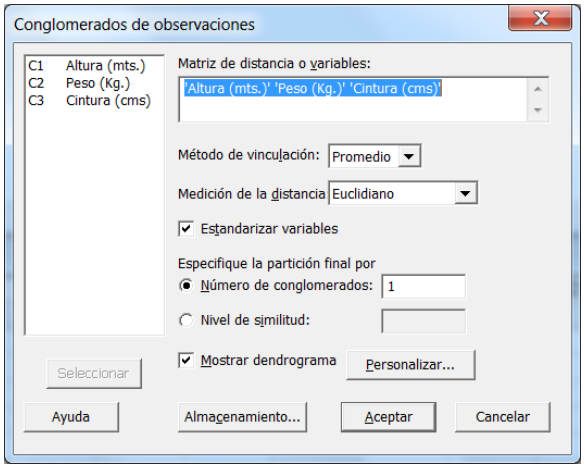
Por opción predeterminada, el dendograma presenta el nivel de similitud en el eje vertical (de forma alterna, como en este caso, puede mostrar el nivel de distancia) y las diferentes observaciones se enumeran con el eje horizontal. La gráfica muestra la manera en la cual se formaron los conglomerados, bien sea uniendo dos observaciones individuales o pareando una observación individual con un conglomerado existente (como en este caso la persona 3). Puede determinar en cuáles distancias de eslabonamiento se forman los conglomerados así como la composición de los conglomerados de la solución final.

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados, aunque en la práctica puede ejecutar primero un análisis de conglomerados **sin especificar una partición final con el objeto de decidir dónde colocarla**. Examine los niveles de distancia en los resultados de la ventana Sesión y en el dendograma. Puede ver las distancias de eslabonamiento en el dendograma colocando el puntero de su ratón sobre una línea horizontal. La distancia de eslabonamiento en cualquier paso es el valor de la

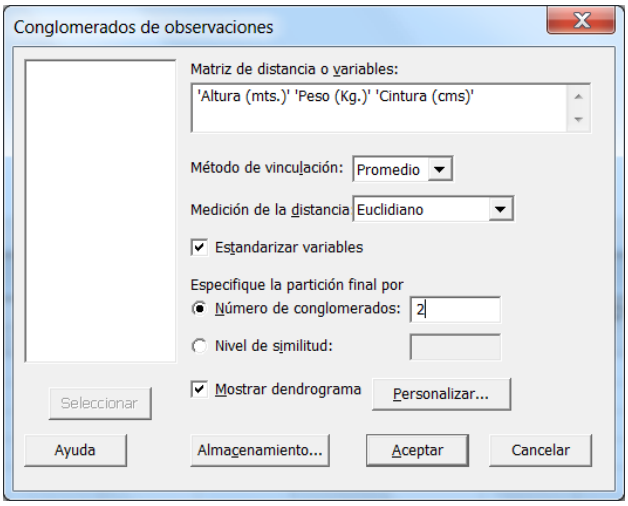
distancia mínima en el paso, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Después de elegir dónde desea realizar su partición, vuelva a ejecutar el procedimiento del conglomerado, utilizando, el Número de conglomerados para obtener un número establecido de conjuntos de la siguiente manera:

En la parte inicial del análisis del eslabonamiento completo apareció el siguiente cuadro de Diálogo:



Ahora en ***Especifique la partición final por***, elija ***Número de conglomerados*** e ingrese ***2***



Haga clic en ***Aceptar*** en el cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts., Peso (Kg.), ...

```
Variables estandarizadas, Distancia euclediana, Enlace de promedios
Pasos de amalgamación
```

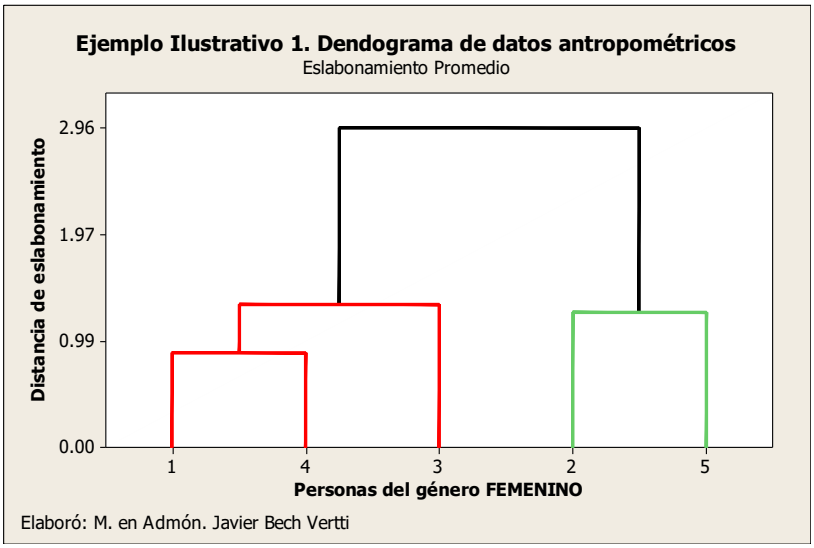
Paso	Número de grupos	Nivel de semejanzal	Nivel de distancia	Grupos incorporados		Nuevo grupo	Número de obs. en el grupo nuevo
1	4	76.3492	0.87219	1	4	1	2
2	3	66.1312	1.24900	2	5	2	2
3	2	64.0314	1.32644	1	3	1	3
4	1	19.8532	2.95563	1	2	1	5

Partición final
Número de grupos: 2

	Número de observaciones	Dentro de la suma de cuadrados del grupo	Distancia promedio desde el centroide	Distancia máxima desde centroide
Grupo1	3	1.48147	0.665049	0.856782
Grupo2	2	0.78000	0.624502	0.624502

Cuando usted especifica la partición o solución final, Minitab muestra entre otras una tabla adicional que resume cada conglomerado según el número de observaciones, en este caso el primer conglomerado está formado por tres mujeres (1,4 y 3) y el segundo conglomerado por las mujeres (2 y 5)

Dendrograma



Examine los conglomerados resultantes en la partición o solución final para observar si la agrupación parece lógica (Las personas que forman cada conglomerado aparecen con un color distinto en cada conglomerado). La observación de dendogramas para determinar diferentes agrupaciones finales puede ayudarle a decidir cuáles de ellas tienen más sentido para sus datos.

2.2.1.1 EJERCICIOS DE **REFUERZO**

EJERCICIOS DE REFUERZO 2.2.1.1 ANÁLISIS DE CONGLOMERADOS. MEDIDA DE DISTANCIA

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar

2.2.1.1.1. Los alumnos de mercadotecnia tienen que hacer una investigación acerca de si trae mejores resultados el trabajo entre dos personas juntas con un parecido en algún aspecto (en este caso calificación) que otras diferentes en tal aspecto.

Persona	Calificación No. 1	Calificación No. 2
1	5.5	4.9
2	9.6	6.3
3	8.1	9.8
4	3.8	7.3
5	8.6	7.0

a) Obtenga la pareja de personas más parecidas y la menos parecida
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos, simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima

2.2.1.1.2. Una compañía de investigación de mercados le solicitan realizar un estudio sobre algunas marcas de cierto producto del mercado para lo cual se analizan tres variables obtenidas de un análisis de factores. Las variables que se han identificado más con cada uno de los factores son las siguientes: Con el factor 1, la variable X_1; con el factor 2, la variable X_2 y con el factor 3, la variable X_3.

Suponga que dichos factores se utilizarán para realizar la segmentación de mercado, para lo cual se proporcionan los puntajes de 5 entrevistas:

ENTREVISTA	**VARIABLE**		
	X_1	**X_2**	**X_3**
1	1.35	2.70	1.35
2	6.75	9.45	6.75
3	2.70	1.35	1.35
4	9.45	9.45	6.75
5	4.05	2.70	2.70

que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

Promedio:	4.86	5.13	3.78
Desviación estándar	3.25	3.98	2.77

a) Obtenga la pareja de personas entrevistadas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.
c) Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

2.2.1.2 EJEMPLO ILUSTRATIVO

EJEMPLO ILUSTRATIVO 2.2.1.2 ANÁLISIS DE CONGLOMERADOS. MATRIZ DE SIMILITUD.

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de las mujeres mexicanas, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras . Para ser más precisos en ciertos resultados, el estudio antropométrico sustituyó el indicador de cintura por el de busto contemplando igualmente 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y el busto de cinco personas del género **FEMENINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Busto
1	1.52	57.12	Mediano
2	1.67	76.85	Grande
3	1.54	67.62	Mediano
4	1.56	49.35	Pequeño
5	1.63	87.22	Grande
6	1.65	82.54	Grande
7	1.68	53.12	Mediano

Criterio de similitud para recodificar los datos:

Altura: existe(1) si es ≥ a 1.60 mts.
no existe(0) si es < a 1.60 mts.

Peso: existe(1) si es ≥ a 60 Kgs.
No existe(0) si es < a 60 Kgs.

Busto: existe(1) si es Mediano
no existe (0) si es cualquier otro tamaño

a) Obtenga la pareja de personas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.
c) Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

Resolución del inciso a)
Los datos recodificados quedaron como sigue:

Persona	Altura (mts.)	Peso (Kg.)	Busto
1	0	0	1
2	1	1	0
3	0	1	1
4	0	0	0
5	1	1	0
6	1	1	0
7	1	0	1

Cálculo de las similitudes a encontrar:

El método de eslabonamiento que se elija determinará cómo se define la similitud entre dos conglomerados. En cada etapa del eslabonamiento, los dos conglomerados más semejantes se unen. Al comienzo, cuando cada observación constituye un conglomerado, la similitud entre los conglomerados es simplemente la similitud entre las observaciones. Posteriormente, después de que se unifican las observaciones, es necesaria una regla de eslabonamiento para calcular las similitudes entre conglomerados, cuando hay múltiples observaciones en un conglomerado.

Se recomienda probar con varios métodos de eslabonamiento y comparar los resultados. Dependiendo de las características de los datos, algunos métodos pueden proveer "mejores" resultados que otros.

El coeficiente de similitud que se usará para resolver este ejemplo es el de parejas simples. El coeficiente es:

$$S_{ij} = \frac{a + d}{p}$$

Se debe cuantificar la similitud entre las personas con base en este coeficiente, donde ***a*** es la presencia del atributo en ambas personas, ***d*** es la ausencia del atributo en ambas personas y ***p*** es el total de características (o variables) que serán evaluadas.
Se empieza a comparar por pareja de personas, en este caso la persona 1 y la 2, como en ninguna de las características hay presencia del atributo en ambas (***a***) ni ausente en ambas (***d***), entonces tanto ***a*** como ***d*** tienen valor de 0 en el coeficiente, en cuanto a ***p*** tenemos que son tres características para evaluar (Altura, Peso y Busto); la similitud entre las personas 1 y 2 es de 0 como a continuación se presenta:

$$S_{1.2} = \frac{0 + 0}{3} = 0$$

De igual forma se procede con las personas 1 y 3, donde hay un atributo presente (***a***) en ambos en la característica de Busto y un atributo ausente (***d***) en ambos en la característica de Altura; por lo tanto, la similitud entre las personas 1 y 3 es de 2/3 como a continuación se presenta:

$$S_{1.3} = \frac{1 + 1}{3} = 2/3$$

Y así sucesivamente, se obtienen las medidas de similitud para todas las parejas que se pretende comparar obteniendo los siguientes resultados:

$$S_{1.2}=\frac{0+0}{3}=0 \qquad S_{2.3}=\frac{1+0}{3}=1/3 \qquad S_{3.5}=\frac{1+0}{3}=1/3 \qquad S_{5.6}=\frac{2+1}{3}=\frac{3}{3}=1$$

$$S_{1.3}=\frac{1+1}{3}=2/3 \qquad S_{2.4}=\frac{0+1}{3}=1/3 \qquad S_{3.6}=\frac{1+0}{3}=1/3 \qquad S_{5.7}=\frac{1+0}{3}=1/3$$

$$S_{1.4}=\frac{0+2}{3}=2/3 \qquad S_{2.5}=\frac{2+1}{3}=3/3 \qquad S_{3.7}=\frac{1+0}{3}=1/3 \qquad S_{6.7}=\frac{1+0}{3}=1/3$$

$$S_{1.5}=\frac{0+0}{3}=0 \qquad S_{2.6}=\frac{2+1}{3}=3/3 \qquad S_{4.5}=\frac{0+1}{3}=1/3$$

$$S_{1.6}=\frac{0+0}{3}=0 \qquad S_{2.7}=\frac{1+0}{3}=1/3 \qquad S_{4.6}=\frac{0+1}{3}=1/3$$

$$S_{1.7}=\frac{1+1}{3}=2/3 \qquad S_{3.4}=\frac{0+1}{3}=1/3 \qquad S_{4.7}=\frac{0+1}{3}=1/3$$

Con la información anterior se construye una matriz cuadrada en este caso de 7 x 7 ya que tenemos 7 personas a la que le nombraremos ***matriz de similitud*** la cual representa que tan similares son las personas con respecto a la similitud entre ellas:

	1	2	3	4	5	6	7
1	1	0	2/3	2/3	0	0	2/3
2	0	1	1/3	1/3	1	1	1/3
3	2/3	1/3	1	1/3	1/3	1/3	1/3
4	2/3	1/3	1/3	1	1/3	1/3	1/3
5	0	1	1/3	1/3	1	1	1/3
6	0	1	1/3	1/3	1	1	1/3
7	2/3	1/3	1/3	1/3	1/3	1/3	1

Como esta matriz es una matriz espejo, es decir la matriz superior es igual a la matriz inferior y la diagonal contiene 1´s ya que la similitud máxima que hay al comparar una persona consigo es de uno, sólo será necesario utilizar la información de la matriz inferior como se muestra a continuación:

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

Al observar la matriz de similitud se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya similitud entre sí es la que se acerque más o sea igual a 1, en este caso tenemos tres parejas, la primera formada por las personas 2 y 5, con una medida de similitud de 1; la segunda formada por las personas 2 y 6, con una similitud también de 1 y la tercera formada por las personas 5 y 6, con una similitud de 1.

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

La pareja de personas menos parecida será aquella cuya similitud entre sí es la se acerque más o sea igual a 0. En este caso tenemos tres parejas, la primera formada por las personas 1 y 2, con una medida de similitud de 0; la segunda formada por las personas 1 y 5, con una similitud también de 0 y la tercera formada por las personas 1 y 6, con una similitud de 0.

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

Resolución al inciso b)

Vinculación o eslabonamiento:

1. **Eslabonamiento Simple:**

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:
Paso 1. Agrupar. Reducir la matriz con la similitud más grande, es decir cuya similitud entre sí se acerque más o sea igual a 1 , para formar el primer conglomerado.

· Con el eslabonamiento simple, o del "vecino más cercano", la similitud entre dos conglomerados es la similitud máxima entre una observación en un conglomerado y una observación en el otro conglomerado. El eslabonamiento simple es una opción apropiada

cuando los conglomerados están claramente separados. Cuando las observaciones están cerca entre sí, el eslabonamiento simple tiende a identificar conglomerados largos similares a cadenas que pueden tener una separación relativamente grande entre observaciones en cualquier extremo de la cadena. ·

Paso 2. Eslabonar. Llenar la matriz con la similitud más grande, o sea contrastar las similitudes de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan similitudes, los elementos más próximos se elegirán con la similitud más grande. Como existen tres parejas de personas con similitudes igual a 1 (2 con 5, 2 con 6 y 5 con 6) podemos elegir cualquiera de ellas para formar el primer conglomerado, en este caso tomaremos la pareja formada por las personas 5 y 6.

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

Una vez seleccionada la similitud de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 5 y 6, como se muestra en la siguiente matriz inferior:

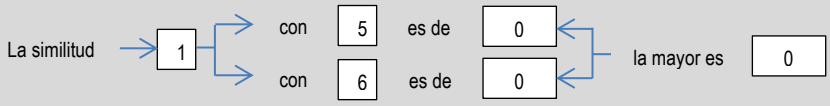
	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	~~0~~	~~1~~	~~1/3~~	~~1/3~~	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

El nuevo conglomerado (5,6) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 7 x 7 a 6 x 6, como se muestra a continuación:

	(5,6)	1	2	3	4	7
(5,6)	1					
1		1				
2			1			
3				1		
4					1	
7						1

Paso 2. Eslabonar

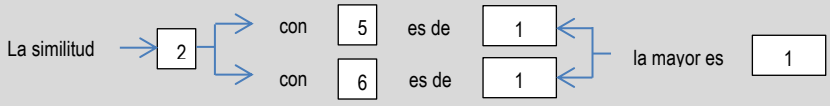
Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la similitud más grande, es decir se compara al nuevo conglomerado (5,6) con el resto de las personas; al comparar (5,6) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con la 5 que es 0 y la de la persona 1 con la 6 que es 0, eligiendo entonces la mayor similitud de las dos personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2			1			
3				1		
4					1	
7						1

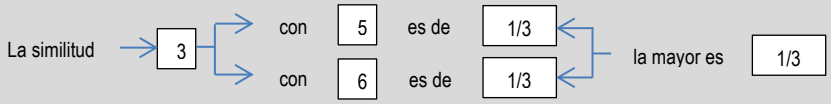
Siguiendo el mismo criterio se eslabona la persona 2 con la pareja de personas (5,6), eligiendo entonces la mayor similitud de las dos personas que en este caso es 1. Lo anterior se puede resumir con el siguiente diagrama de flujo:



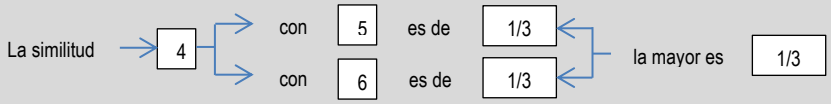
El valor eslabonado se coloca en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1		1			
3				1		
4					1	
7						1

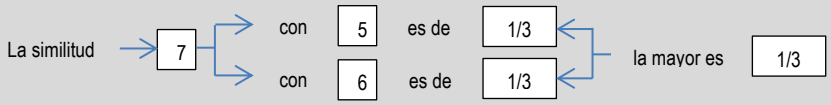
Sigue eslabonar a la persona 3 con la pareja de personas (5,6), eligiendo entonces la mayor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con la pareja de personas (5,6), eligiendo entonces la mayor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 7 con la pareja de personas (5,6), eligiendo entonces la mayor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1		1			
3	1/3			1		
4	1/3				1	
7	1/3					1

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1	0	1			
3	1/3	2/3	1/3	1		
4	1/3	2/3	1/3	1/3	1	
7	1/3	2/3	1/3	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 2 que se eslabonaría al conglomerado formado por las personas (5,6) con la similitud

mayor que en este caso es de 1 como se muestra a continuación:

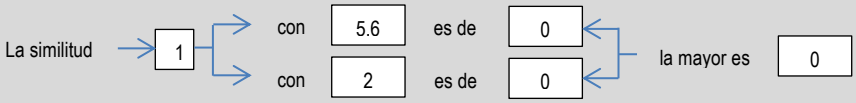
	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1	0	1			
3	1/3	2/3	1/3	1		
4	1/3	2/3	1/3	1/3	1	
7	1/3	2/3	1/3	1/3	1/3	1

La nueva matriz se reduce de 6 x 6 a una de 5 x 5 de la siguiente manera:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1		1			
3			1		
4				1	
7					1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la similitud mayor, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con las 5 y 6 que es 0 y la de la persona 1 con la 2 que es 0, eligiendo entonces la similitud mayor de las dos combinaciones o tres personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:

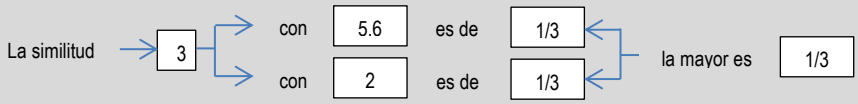


El valor eslabonado se coloca en la matriz reducida

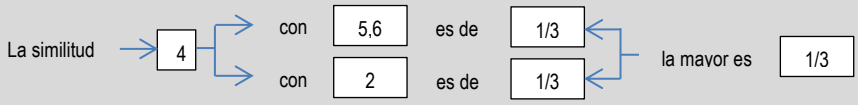
	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3					
4				1	
7					1

Sigue eslabonar a la persona 3 con el conglomerado de personas (5,6,2), eligiendo entonces la mayor similitud de las tres personas. Primero se compara a la persona 3 con las personas 5 y 6 con una similitud de 1/3 y segundo a la

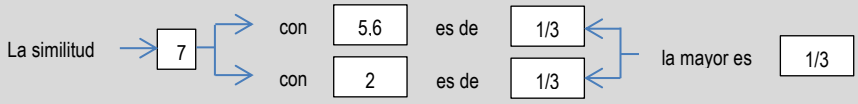
persona 3 con la persona 2 con una similitud de 1/3 eligiendo la similitud mayor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con el conglomerado de personas (5,6,2), eligiendo entonces la mayor similitud de las tres personas. Primero se compara a la persona 4 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 4 con la persona 2 con una similitud de 1/3 eligiendo la similitud mayor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 7 con el conglomerado de personas (2,5,6), eligiendo entonces la mayor similitud de las tres personas. Primero se compara a la persona 7 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 7 con la persona 2 con una similitud de 1/3 eligiendo la similitud mayor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3		1		
4	1/3			1	
7	1/3				1

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3	2/3	1		
4	1/3	2/3	1/3	1	
7	1/3	2/3	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 1 y 3 o con las personas 1 con 7. Como ambos grupos tienen la misma similitud de 2/3 se podría seleccionar cualquiera de los 2. Suponga que se eligió a la persona 1 y 7 como se muestra a continuación:

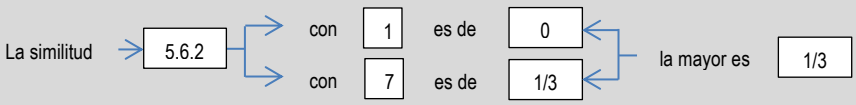
	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3	2/3	1		
4	1/3	2/3	1/3	1	
7	1/3	2/3	1/3	1/3	1

La nueva matriz se reduce de 5 x 5 a una de 4 x 4 de la siguiente manera:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)		1		
3			1	
4				1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la similitud mayor, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6,2) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con las personas (5,6,2) que es 0 y la persona 7 con las (5,6,2) que es 1/3, eligiendo entonces la similitud mayor de las dos combinaciones o cinco personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



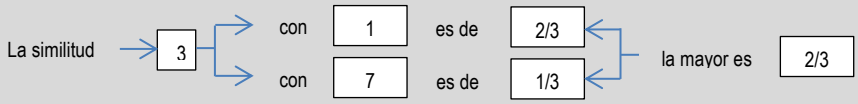
El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/3	1		
3			1	
4				1

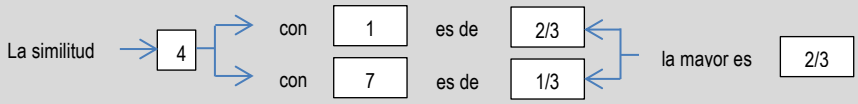
En la etapa anterior podemos observar que ya tenemos el eslabonamiento de la persona 3 con el conglomerado (5,6,2) con una similitud de 1/3 y a la persona 4 con el conglomerado (5,6,2) con una similitud de 1/3, estos valores se pasan directamente a la matriz reducida:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/3	1		
3	1/3		1	
4	1/3			1

Sigue eslabonar a la persona 3 con el nuevo conglomerado de personas (1,7) eligiendo entonces la mayor similitud de las tres personas. Primero se compara a la persona 3 con las personas 1 con una similitud de 2/3 y segundo a la persona 3 con la persona 7 con una similitud de 1/3 eligiendo la similitud mayor de ambas comparaciones que es 2/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con el nuevo conglomerado de personas (1,7) eligiendo entonces la mayor similitud de las tres personas. Primero se compara a la persona 4 con las personas 1 con una similitud de 1/3 y segundo a la persona 4 con la persona 7 con una similitud de 1/3 eligiendo la similitud mayor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/3	1		
3	1/3	2/3	1	
4	1/3	2/3		1

El valor faltante de la matriz se llena con el mismo dato de la matriz original:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/3	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, como es el caso de la persona 3 o 4 que se eslabonan al conglomerado formado por las personas (1,7) con la similitud más grande de 2/3. Se puede elegir cualquiera de las dos personas, se elegirá a la persona 4 como se muestra a continuación:

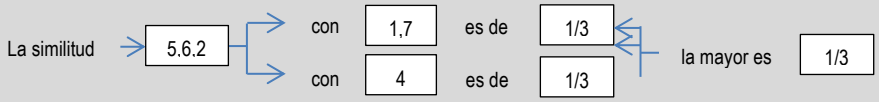
	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/3	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)		1	
3			1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la similitud mayor, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6,2) con la persona 1 y7 se debe verificar en la matriz anterior las similitudes entre las personas 1 y ,7 con las personas (5,6,2) que es 1/3 y la persona 4 con las personas (5,6,2) que es 1/3, eligiendo entonces la similitud mayor de las dos combinaciones o seis personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



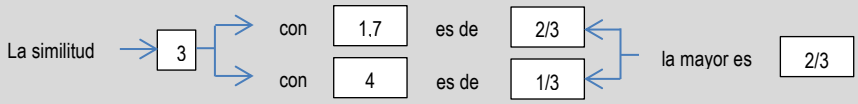
El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)	1/3	1	
3			1

En la etapa anterior podemos observar que ya tenemos el eslabonamiento de la persona 3 con el conglomerado (5,6,2) con una similitud de 1/3 este valor se pasa directamente a la matriz reducida:

	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)	1/3	1	
3	1/3		1

Sigue eslabonar a la persona 3 con el conglomerado de personas (1,7,4) eligiendo entonces la mayor similitud de las cuatro personas. Primero se compara a la persona 3 con las personas 1 y 7 con una similitud de 2/3 y segundo a la persona 3 con la persona 4 con una similitud de 1/3 eligiendo la similitud mayor de ambas comparaciones que es 2/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)	1/3	1	
3	1/3	2/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, como es el caso de la persona 3 que se puede eslabonar a cualquiera de los dos conglomerados (5,6,2) o (1,7,3) con la similitud más grande de 1/3. Se elige el conglomerado formado por las personas(1,7,4) como se muestra a continuación:

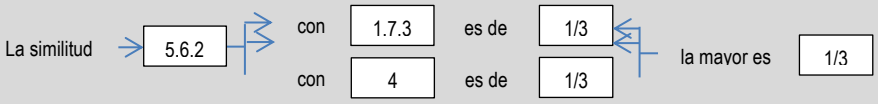
	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)	1/3	1	
3	1/3	1/3	1

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(5,6,2)	(1,7,4, 3)
(5,6,2)	1	
(1,7,4, 3)		1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la similitud mayor, es decir se compara al conglomerado (5,6,2) con el resto de las personas que forman el otro conglomerado (1,7,4,3); al comparar (5,6,2) con las personas (1,7,4,3) se debe verificar en la matriz anterior las similitudes entre las personas 1,7,3 con las personas (5,6,2) que es 1/3 y la persona 3 con las (5,6,2) que es 1/3, eligiendo entonces la similitud mayor de las dos combinaciones o siete personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor resultante se acomoda en la matriz reducida:

	(2,5,6)	(1,7,3, 4)
(2,5,6)	1	
(1,7,3, 4)	1/3	1

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	5	6	2
2	5	1	2	5	3
3	4	2/3	1	7	2
4	3	2/3	1	4	3
5	2	2/3	1	3	4
6	1	1/3	1	2	7

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos

uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 5 y la 6 y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 2 al primer conglomerado formado por las personas 5 y 6 como se muestra en la columna 3 a una similitud de 1 o 100%. En la etapa 3 se forma un nuevo conglomerado como se aprecia en la columna 5 formado por las personas 1 y 7 a una similitud de 2/3 ó 66.667%. En la etapa 4 se incorporó la persona 4 al segundo conglomerado formado por las personas 1 y 7 como se muestra en la columna 3 a una similitud de 2/3 ó 66.667%. En la etapa 5 se incorporó la persona 3 al sugundo conglomerado formado por las personas 1,7 y 3 como se muestra en la columna 3 a una similitud de 2/3 ó 66.667% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 1/3 ó 33.333%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Altura, Peso, Busto

Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación

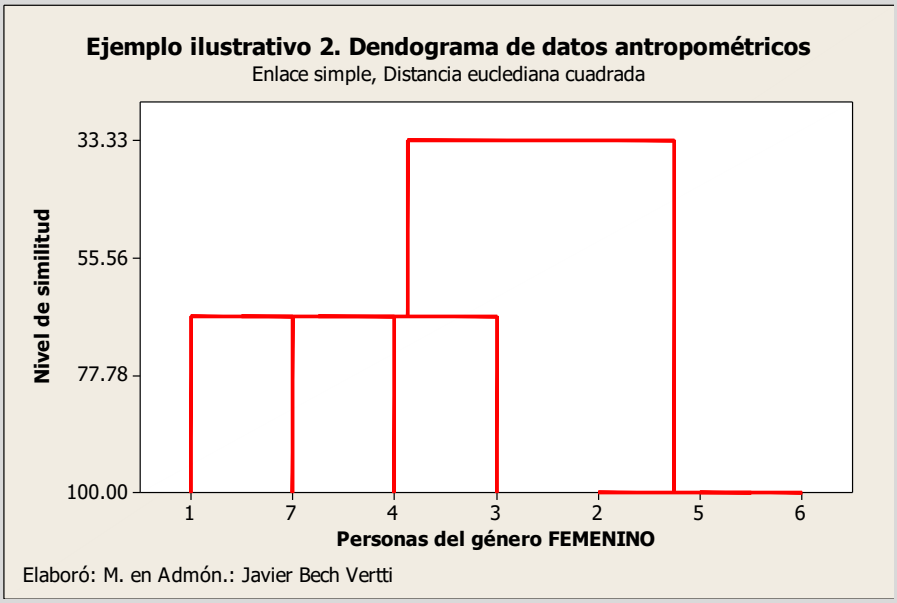
Paso	Número de grupos	Nivel de semejanzal	Nivel de distancia	Grupos incorporados		Nuevo grupo	Número de obs. En el grupo nuevo
1	6	100.000	0	5	6	5	2
2	5	100.000	0	2	5	2	3
3	4	66.667	1	1	7	1	2
4	3	66.667	1	1	4	1	3
5	2	66.667	1	1	3	1	4
6	1	33.333	2	1	2	1	7

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (5,6,2) y el conglomerado 2 por las personas (1,7,4,3).

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 5 y 6 los valores de similitud cambian de 2/3 a 1/3 cuando se unen ambos grupos o conglomerados por lo tanto, la solución más factible parece ser la de 2 conglomerados (5,6,2,) y (1,7,4,3).

· Con el eslabonamiento completo, o "del vecino más lejano", la similitud entre dos conglomerados es la similitud mínima entre una observación en un conglomerado y una observación en el otro conglomerado. Este método garantiza que todas las

2. Eslabonamiento Completo:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:
Paso 1. Agrupar. Reducir la matriz con la similitud más grande, es decir cuya similitud entre sí se acerque más o sea igual a 1, para formar el primer conglomerado.
Paso 2. Eslabonar. Llenar la matriz con la similitud más pequeña, es decir cuya similitud entre sí se acerque más o sea igual a 0 al contrastar las similitudes de

observaciones en un conglomerado se encuentren dentro de una similitud mínima y tiende a producir conglomerados con diámetros similares. Los resultados pueden ser sensibles a valores atípicos.

los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan similitudes, los elementos más próximos se elegirán con la similitud más grande. Como existen tres parejas de personas con similitudes igual a 1 (2 con 5, 2 con 6 y 5 con 6) podemos elegir cualquiera de ellas para formar el primer conglomerado, en este caso tomaremos la pareja formada por las personas 5 y 6.

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

Una vez seleccionada la similitud de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 5 y 6, como se muestra en la siguiente matriz inferior:

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	~~0~~	~~1~~	~~1/3~~	~~1/3~~	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

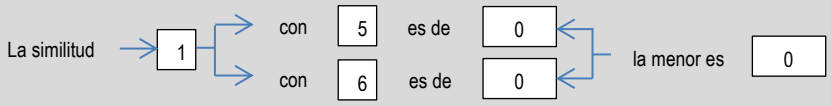
El nuevo conglomerado (5,6) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 7 x 7 a 6 x 6, como se muestra a continuación:

	(5,6)	1	2	3	4	7
(5,6)	1					
1		1				
2			1			
3				1		
4					1	
7						1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la

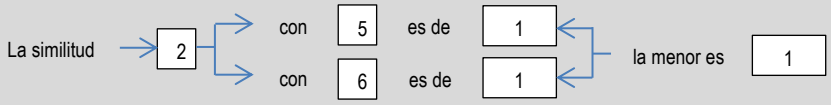
similitud más grande, es decir se compara al nuevo conglomerado (5,6) con el resto de las personas; al comparar (5,6) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con la 5 que es 0 y la de la persona 1 con la 6 que es 0, eligiendo entonces la menor similitud de las dos personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2			1			
3				1		
4					1	
7						1

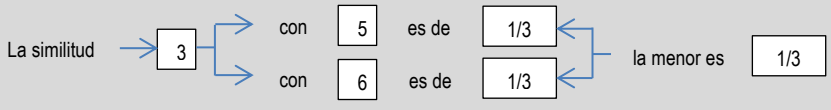
Siguiendo el mismo criterio se eslabona la persona 2 con la pareja de personas (5,6), eligiendo entonces la menor similitud de las dos personas que en este caso es 1. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

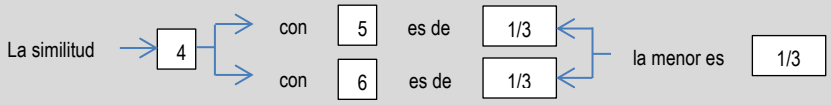
	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1		1			
3				1		
4					1	
7						1

Sigue eslabonar a la persona 3 con la pareja de personas (5,6), eligiendo entonces la menor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:

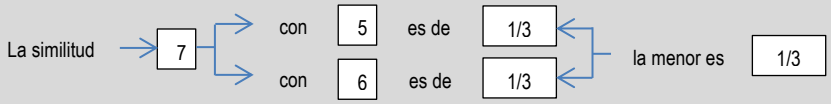


Sigue eslabonar a la persona 4 con la pareja de personas (5,6), eligiendo entonces la menor similitud de las dos personas que en este caso es 1/3. Lo

anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 7 con la pareja de personas (5,6), eligiendo entonces la menor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1		1			
3	1/3			1		
4	1/3				1	
7	1/3					1

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1	0	1			
3	1/3	2/3	1/3	1		
4	1/3	2/3	1/3	1/3	1	
7	1/3	2/3	1/3	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 2 que se eslabonaría al conglomerado formado por las personas (5,6) con la similitud mayor que en este caso es de 1 como se muestra a continuación:

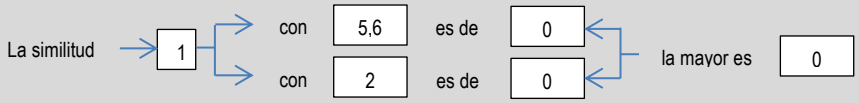
	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1	0	1			
3	1/3	2/3	1/3	1		
4	1/3	2/3	1/3	1/3	1	
7	1/3	2/3	1/3	1/3	1/3	1

La nueva matriz se reduce de 6 x 6 a una de 5 x 5 de la siguiente manera:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1		1			
3			1		
4				1	
7					1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la similitud menor, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con las 5 y 6 que es 0 y la de la persona 1 con la 2 que es 0, eligiendo entonces la similitud menor de las dos combinaciones o tres personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:

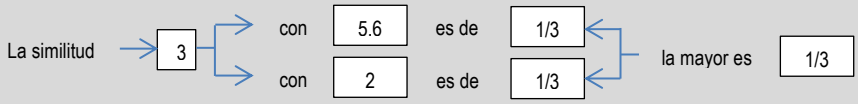


El valor eslabonado se coloca en la matriz reducida

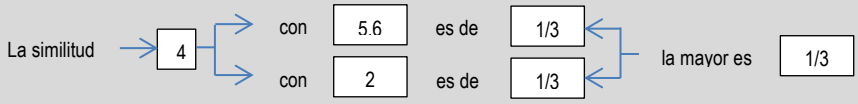
	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3					
4				1	
7					1

Sigue eslabonar a la persona 3 con el conglomerado de personas (5,6,2), eligiendo entonces la menor similitud de las tres personas. Primero se compara a la persona 3 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 3 con la persona 2 con una similitud de 1/3 eligiendo la similitud menor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el

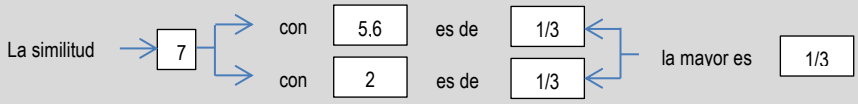
siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con el conglomerado de personas (5,6,2), eligiendo entonces la menor similitud de las tres personas. Primero se compara a la persona 4 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 4 con la persona 2 con una similitud de 1/3 eligiendo la similitud menor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 7 con el conglomerado de personas (2,5,6), eligiendo entonces la menor similitud de las tres personas. Primero se compara a la persona 7 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 7 con la persona 2 con una similitud de 1/3 eligiendo la similitud menor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3		1		
4	1/3			1	
7	1/3				1

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3	2/3	1		
4	1/3	2/3	1/3	1	
7	1/3	2/3	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 1 y 3 o con las personas 1 con 7. Como ambos grupos tienen la misma similitud de 2/3 se podría seleccionar cualquiera de los 2. Suponga que se eligió a la persona 1 y 7 como se muestra a continuación:

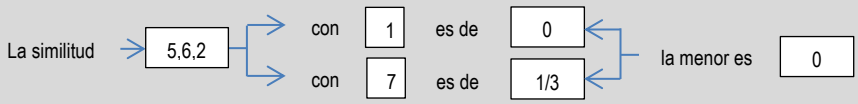
	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3	2/3	1		
4	1/3	2/3	1/3	1	
7	1/3	2/3	1/3	1/3	1

La nueva matriz se reduce de 5 x 5 a una de 4 x 4 de la siguiente manera:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)		1		
3			1	
4				1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la similitud menor, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6,2) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con las personas (5,6,2) que es 0 y la persona 7 con las (5,6,2) que es 1/3, eligiendo entonces la similitud menor de las dos combinaciones o cinco personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:



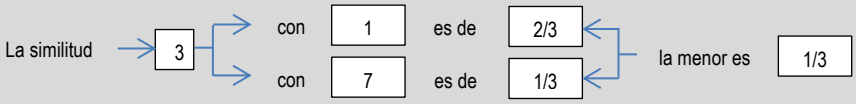
El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	0	1		
3			1	
4				1

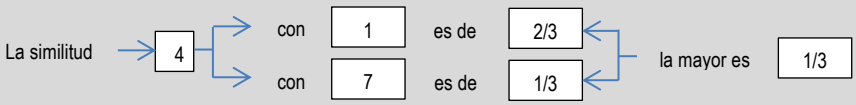
En la etapa anterior podemos observar que ya tenemos el eslabonamiento de la persona 3 con el conglomerado (5,6,2) con una similitud de 1/3 y a la persona 4 con el conglomerado (5,6,2) con una similitud de 1/3, estos valores se pasan directamente a la matriz reducida:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	0	1		
3	1/3		1	
4	1/3			1

Sigue eslabonar a la persona 3 con el nuevo conglomerado de personas (1,7) eligiendo entonces la menor similitud de las tres personas. Primero se compara a la persona 3 con las personas 1 con una similitud de 2/3 y segundo a la persona 3 con la persona 7 con una similitud de 1/3 eligiendo la similitud menor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con el nuevo conglomerado de personas (1,7) eligiendo entonces la menor similitud de las tres personas. Primero se compara a la persona 4 con las personas 1 con una similitud de 2/3 y segundo a la persona 4 con la persona 7 con una similitud de 1/3 eligiendo la similitud menor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	0	1		
3	1/3	1/3	1	
4	1/3	1/3		1

El valor faltante de la matriz se llena con el mismo dato de la matriz original:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	0	1		
3	1/3	1/3	1	
4	1/3	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, como es el caso de las personas 3 o 4 que pueden formar un nuevo conglomerado con una similitud más grande de 1/3 o se pueden eslabonar al conglomerado formado por las personas (1,7) con la similitud más grande de 1/3 o al conglomerado formado por las personas (5,6,2) con la similitud más grande de 1/3. Como se puede elegir cualquiera de las tres opciones, se elegirá en este caso formar un nuevo conglomerado como se muestra a continuación:

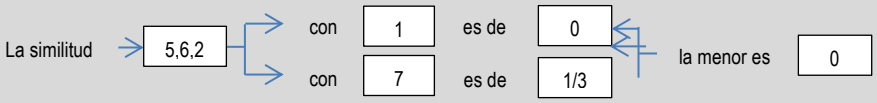
	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	0	1		
3	1/3	1/3	1	
4	1/3	1/3	1/3	1

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(5,6,2)	(1,7)	(3,4)
(5,6,2)	1		
(1,7)		1	
(3,4)			1

Paso 2. Eslabonar

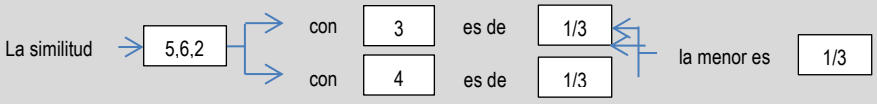
Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la similitud menor, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6,2) con la persona 1 y7 se debe verificar en la matriz anterior las similitudes entre la persona 1 con las personas (5,6,2) que es 0 y la persona 7 con las personas (5,6,2) que es 1/3, eligiendo entonces la similitud menor de las dos combinaciones o cinco personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7)	(3,4)
(5,6,2)	1		
(1,7)	0	1	
(3,4)			1

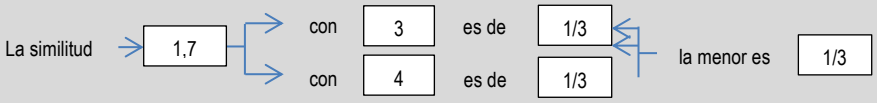
Sigue eslabonar a las personas 3 y 4 con el conglomerado de personas (5,6,2) eligiendo entonces la menor similitud de las cinco personas. Primero se compara a la persona 3 con las personas (5,6,2) con una similitud de 1/3 y segundo a la persona 4 con las personas (5,6,2) con una similitud de 1/3 eligiendo la similitud menor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7)	(3,4)
(5,6,2)	1		
(1,7)	0	1	
(3,4)	1/3		1

Sigue eslabonar a las personas 3 y 4 con el conglomerado de personas (1,7) eligiendo entonces la menor similitud de las cuatro personas. Primero se compara a la persona 3 con las personas (1,7) con una similitud de 1/3 y segundo a la persona 4 con las personas (1,7) con una similitud de 1/3 eligiendo la similitud menor de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7)	(3,4)
(5,6,2)	1		
(1,7)	0	1	
(3,4)	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, como en este caso el conglomerado (3,4) se puede eslabonar a cualquiera de los dos conglomerados (5,6,2) o (1,7) con la similitud más grande de 1/3, se elige el conglomerado formado por las personas(5,6,2) como se muestra a continuación:

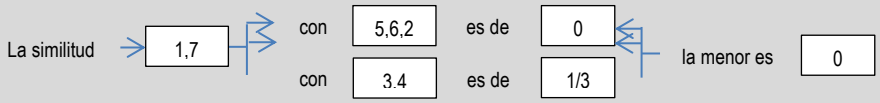
	(5,6,2)	(1,7)	(3,4)
(5,6,2)	1		
(1,7)	0	1	
(3,4)	1/3	1/3	1

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(5,6,2, 3,4)	(1,7)
(5,6,2, 3,4)	1	
(1,7)		1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la similitud menor, es decir se compara al conglomerado (1,7) con el resto de las personas que forman los otros conglomerados (5,6,2) y (3,4); al comparar (1,7) con las personas (5,6,2) se debe verificar en la matriz anterior las similitudes entre las personas 1,7 con las personas (5,6,2) que es 0 y las personas 1 y 7 con las (3,4) que es 1/3, eligiendo entonces la similitud menor de las dos combinaciones o siete personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor resultante se acomoda en la matriz reducida:

	(5,6,2, 3,4)	(1,7)
(5,6,2, 3,4)	1	
(1,7)	0	1

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	5	6	2
2	5	1	2	5	3
3	4	2/3	1	7	2
4	3	1/3	3	4	2
5	2	1/3	2	3	5
6	1	0	1	2	7

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 5 y la 6 y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 2 al primer conglomerado formado por las personas 5 y 6 como se muestra en la columna 3 a una similitud de 1 o 100%. En la etapa 3 se forma un segundo o nuevo conglomerado como se aprecia en la columna 5 formado por las personas 1 y 7 a una similitud de 2/3 ó 66.667%. En la etapa 4 se forma un nuevo conglomerado como se aprecia en la columna 5 formado por las personas 3 y 4 a una similitud de 1/3 ó 33.333%. En la etapa 5 se incorporó el último conglomerado formado por las personas 3 y 4 al segundo y primer

conglomerado formado por las personas 5,6,2 como se muestra en la columna 3 a una similitud de 1/3 ó 33.333% y finalmente en la etapa 6 se eslabonaron los dos conglomerados (5,6,2,3,4) y el (1,7) a una similitud final de 0 ó 0.000%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

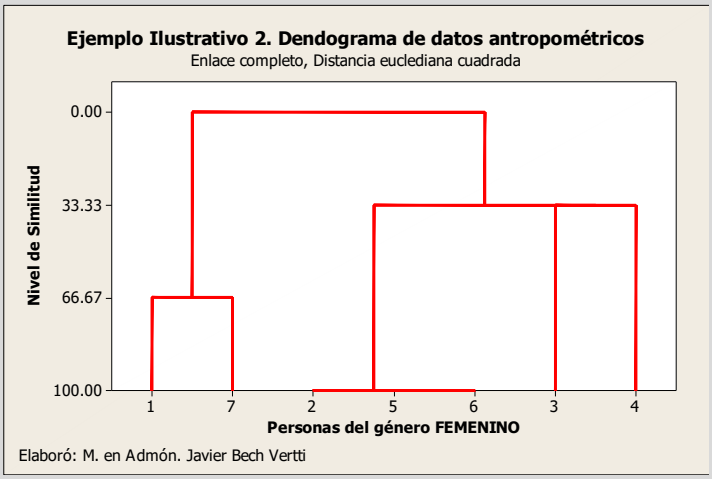
Análisis de observaciones de conglomerado: Altura, Peso, Busto

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación
```

Paso	Número de grupos	Nivel de semejanzal	Nivel de distancia	Grupos incorporados		Nuevo grupo	Número de obs. en el grupo nuevo
1	6	100.000	0	5	6	5	2
2	5	100.000	0	2	5	2	3
3	4	66.667	1	1	7	1	2
4	3	33.333	2	3	4	3	2
5	2	33.333	2	2	3	2	5
6	1	0.000	3	1	2	1	7

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en este caso el conglomerado 1 está formado por las personas (5,6,2,3,4) y el conglomerado 2 por las personas (1,7). Minitab puede dibujar los conglomerados en diferente orden, primero el 1 o el 2 o bien primero el 2 y luego el 1, lo cual no cambia la solución.
Cabe mencionar que mientras a menor similitud se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 5 y 6 los valores de similitud cambian de 0.00 a 1/3 cuando se unen ambos grupos o conglomerados por lo tanto, la solución más factible parece ser la de 2 conglomerados (5,6,2,3,4) y (2,7).

· Con el eslabonamiento promedio, la similitud entre dos conglomerados es la similitud media entre una observación en un conglomerado y una observación en el otro conglomerado. En vista de que los métodos de eslabonamiento simple o completos agrupan conglomerados sobre la base de similitudes entre pares individuales, el eslabonamiento promedio utiliza una medida de ubicación más central.

3. Eslabonamiento Promedio:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la similitud más grande, es decir cuya similitud entre sí se acerque más o sea igual a 1 , para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la similitud promedio, o sea contrastar las similitudes de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan similitudes, los elementos más próximos se elegirán con la similitud más grande. Como existen tres parejas de personas con similitudes igual a 1 (2 con 5, 2 con 6 y 5 con 6) podemos elegir cualquiera de ellas para formar el primer conglomerado, en este caso tomaremos la pareja formada por las personas 5 y 6.

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

Una vez seleccionada la similitud de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 5 y 6, como se muestra en la siguiente matriz inferior:

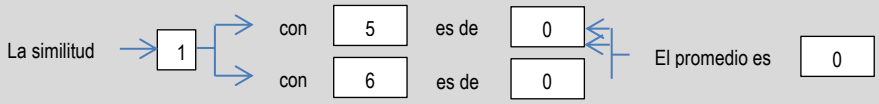
	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

El nuevo conglomerado (5,6) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 7 x 7 a 6 x 6, como se muestra a continuación:

	(5,6)	1	2	3	4	7
(5,6)	1					
1		1				
2			1			
3				1		
4					1	
7						1

Paso 2. Eslabonar

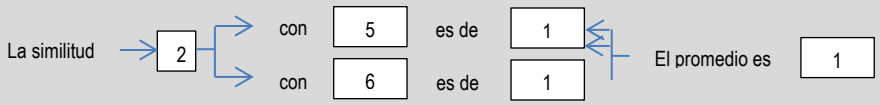
Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la similitud promedio, es decir se compara al nuevo conglomerado (5,6) con el resto de las personas; al comparar (5,6) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con la 5 que es 0 y la de la persona 1 con la 6 que es 0, eligiendo entonces la mayor similitud de las dos personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2			1			
3				1		
4					1	
7						1

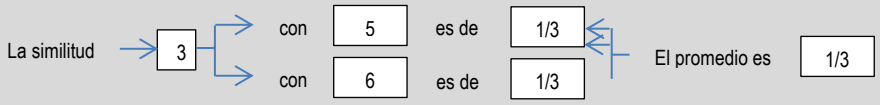
Siguiendo el mismo criterio se eslabona la persona 2 con la pareja de personas (5,6), eligiendo entonces la similitud promedio de las dos personas que en este caso es 1. Lo anterior se puede resumir con el siguiente diagrama de flujo:



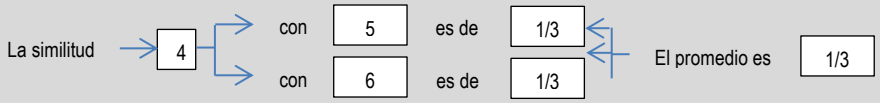
El valor eslabonado se coloca en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1		1			
3				1		
4					1	
7						1

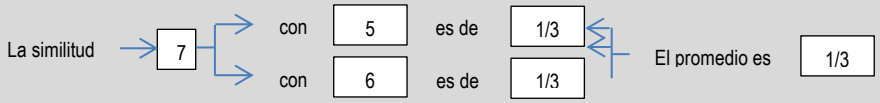
Sigue eslabonar a la persona 3 con la pareja de personas (5,6), eligiendo entonces la mayor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con la pareja de personas (5,6), eligiendo entonces la mayor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 7 con la pareja de personas (5,6), eligiendo entonces la mayor similitud de las dos personas que en este caso es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1		1			
3	1/3			1		
4	1/3				1	
7	1/3					1

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original:

	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1	0	1			
3	1/3	2/3	1/3	1		
4	1/3	2/3	1/3	1/3	1	
7	1/3	2/3	1/3	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 2 que se eslabonaría al conglomerado formado por las personas (5,6) con la similitud mayor que en este caso es de 1 como se muestra a continuación:

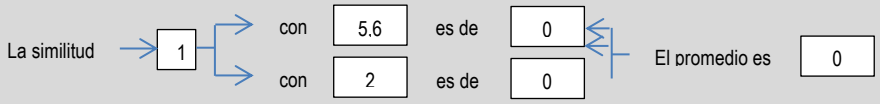
	(5,6)	1	2	3	4	7
(5,6)	1					
1	0	1				
2	1	0	1			
3	1/3	2/3	1/3	1		
4	1/3	2/3	1/3	1/3	1	
7	1/3	2/3	1/3	1/3	1/3	1

La nueva matriz se reduce de 6 x 6 a una de 5 x 5 de la siguiente manera:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1		1			
3			1		
4				1	
7					1

Paso 2. Eslabonar

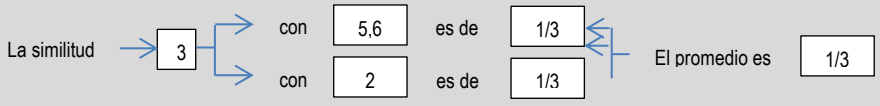
Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la similitud promedio, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con las 5 y 6 que es 0 y la de la persona 1 con la 2 que es 0, eligiendo entonces la similitud mayor de las dos combinaciones o cuatro personas que en este caso es 0. Lo anterior se puede resumir con el siguiente diagrama de flujo:



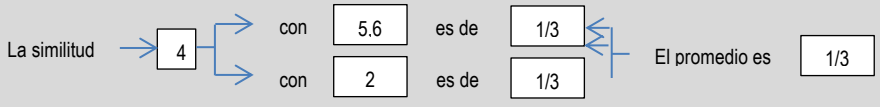
El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3					
4				1	
7					1

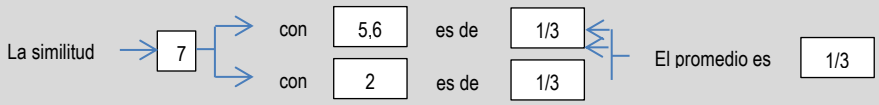
Sigue eslabonar a la persona 3 con el conglomerado de personas (5,6) eligiendo entonces la similitud promedio de las tres personas. Primero se compara a la persona 3 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 3 con la persona 2 con una similitud de 1/3 eligiendo la similitud promedio de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con el conglomerado de personas (5,6), eligiendo entonces la similitud promedio de las tres personas. Primero se compara a la persona 4 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 4 con la persona 2 con una similitud de 1/3 eligiendo la similitud promedio de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 7 con el conglomerado de personas (2,5), eligiendo entonces la similitud promedio de las tres personas. Primero se compara a la persona 7 con las personas 5 y 6 con una similitud de 1/3 y segundo a la persona 7 con la persona 2 con una similitud de 1/3 eligiendo la similitud promedio de ambas comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3		1		
4	1/3			1	
7	1/3				1

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original:

	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3	2/3	1		
4	1/3	2/3	1/3	1	
7	1/3	2/3	1/3	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 1 y 3 o con las personas 1 con 4 o con las personas 1 y 7. Como los tres grupos tienen la misma similitud de 2/3 se podría seleccionar cualquiera de los 3. Suponga que se eligió a la persona 1 y 7 como se muestra a continuación:

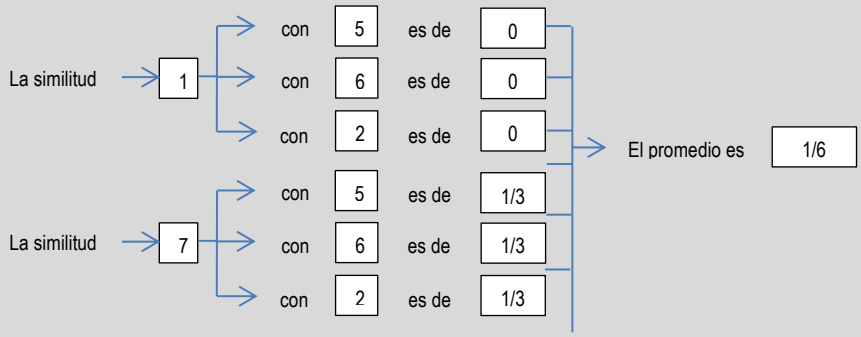
	(5,6,2)	1	3	4	7
(5,6,2)	1				
1	0	1			
3	1/3	2/3	1		
4	1/3	2/3	1/3	1	
7	1/3	2/3	1/3	1/3	1

La nueva matriz se reduce de 5 x 5 a una de 4 x 4 de la siguiente manera:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)		1		
3			1	
4				1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la similitud promedio, es decir se compara al conglomerado (5,6,2) con el resto de las personas; al comparar (5,6,2) con la persona 1 se debe verificar en la matriz anterior las similitudes entre la persona 1 con las personas (5,6,2) que son 0 y la persona 7 con las (5,6,2) que son 1/3, eligiendo entonces la similitud promedio de las seis combinaciones que en este caso es 1/6. Lo anterior se puede resumir con el siguiente diagrama de flujo:



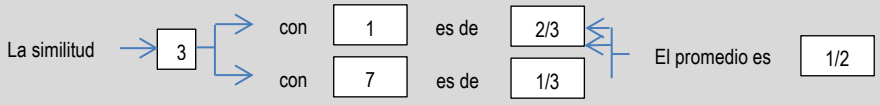
El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/6	1		
3			1	
4				1

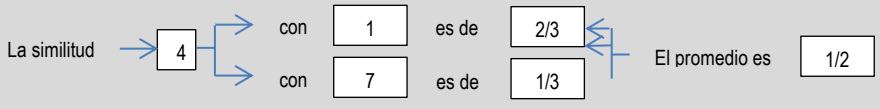
Sigue eslabonar a la persona 3 con el conglomerado de personas (5,6,2), eligiendo entonces la similitud promedio de las cuatro personas. Primero se compara a la persona 3 con las personas 5 con una similitud de 1/3, segundo a la persona 3 con la persona 6 con una similitud de 1/3 y tercero a la persona 3 con la persona 2 con una similitud de 1/3, eligiendo la similitud promedio de las tres comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:

Sigue eslabonar a la persona 4 con el conglomerado de personas (5,6,2), eligiendo entonces la similitud promedio de las cuatropersonas. Primero se compara a la persona 4 con las personas 5 con una similitud de 1/3, segundo a la persona 4 con la persona 6 con una similitud de 1/3 y tercero a la persona 4 con la persona 2 con una similitud de 1/3, eligiendo la similitud promedio de las tres comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:

Sigue eslabonar a la persona 3 con el conglomerado de personas (1,7), eligiendo entonces la similitud promedio de las tres personas. Primero se compara a la persona 3 con las personas 1 con una similitud de 2/3 y segundo a la persona 3 con la persona 7 con una similitud de 1/3 eligiendo la similitud promedio de ambas comparaciones que es 1/2. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 4 con el conglomerado de personas (1,7), eligiendo entonces la similitud promedio de las tres personas. Primero se compara a la persona 4 con las personas 1 con una similitud de 2/3 y segundo a la persona 4 con la persona 7 con una similitud de 1/3 eligiendo la similitud promedio de ambas comparaciones que es 1/2. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/6	1		
3	1/3	1/2	1	
4	1/3	1/2		1

El valor faltante de la matriz se llena con el mismo dato de la matriz original:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/6	1		
3	1/3	1/2	1	
4	1/3	1/2	1/3	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, como es el caso de la persona 3 o 4 que se eslabonan al conglomerado formado por las personas (1,7) con la similitud más grande de 1/2. Se puede elegir cualquiera de las dos personas, se elegirá a la persona 4 como se muestra a continuación:

	(5,6,2)	(1,7)	3	4
(5,6,2)	1			
(1,7)	1/6	1		
3	1/3	1/2	1	
4	1/3	1/2	1/3	1

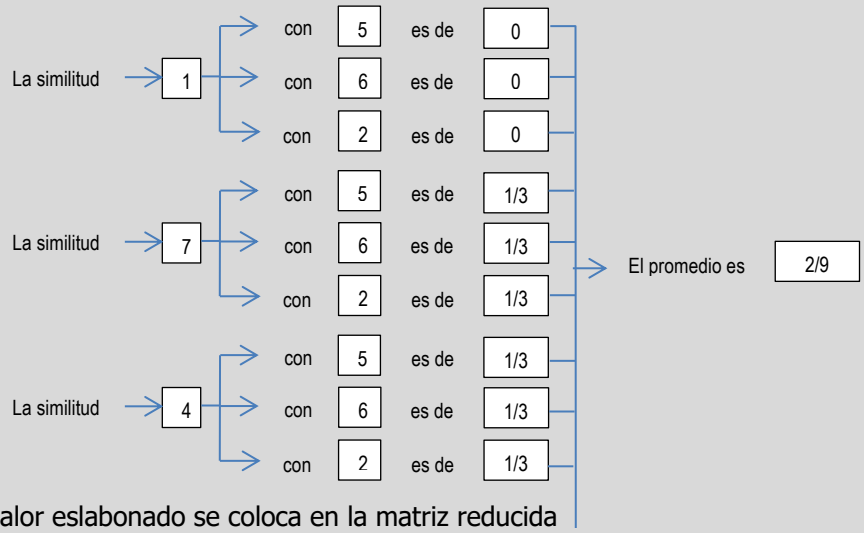
La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)		1	
3			1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la similitud promedio, es decir se compara al conglomerado (5,6,2) con el resto de las personas del conglomerado (1,7,4); al comparar (5,6,2) con la persona 1 se debe verificar en la matriz original las similitudes entre las personas 1 y 5 que es 0, a la persona 1 y 6 que es 0 y a la persona 1 y 2 que es 0, al comparar (5,6,2) con la persona 7 se debe verificar en la matriz original las similitudes entre las personas 7 y 5 que es 1/3, a la persona 7 y 6 que es 1/3 y a la persona 7 y 2 que es 1/3, asimismo al comparar (5,6,2) con la persona 4 se debe verificar en la matriz

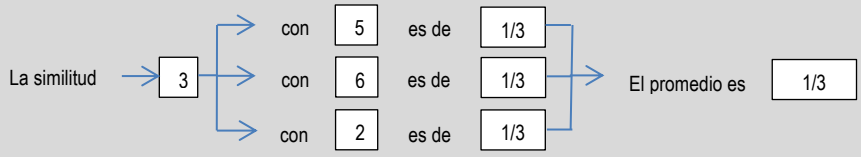
original las similitudes entre las personas 4 y 5 que es 1/3, a la persona 4 y 6 que es 1/3 y a la persona 4 y 2 que es 1/3, eligiendo la similitud promedio de las nueve combinaciones o seis personas que es 2/9. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)	2/9	1	
3			1

Sigue eslabonar a la persona 3 con el conglomerado de personas (5,6,2) eligiendo entonces la similitud promedio de las cuatro personas en la matriz original. Se compara a la persona 3 con la personas 5 con una similitud de 1/3, a la persona 3 con la persona 6 con una similitud de 1/3 y a la persona 3 con la persona 2 con una similitud de 1/3 eligiendo la similitud promedio de las tres comparaciones que es 1/3. Lo anterior se puede resumir con el siguiente diagrama de flujo:



Sigue eslabonar a la persona 3 con el conglomerado de personas (1,7,4) eligiendo entonces la similitud promedio de las cuatro personas en la matriz original. Se compara a la persona 3 con las personas 1 con una similitud de 2/3, a la persona 3 con la persona 7 con una similitud de 1/3 y a la persona 3 con la persona 4 con una similitud de 1/3 eligiendo la similitud promedio de las tres comparaciones que es 4/9. Lo anterior se puede resumir con el siguiente diagrama de flujo:

Los valores resultantes se acomodan en la matriz reducida:

	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)	2/9	1	
3	1/3	4/9	1

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, como es el caso de la persona 3 que se puede eslabonar a cualquiera de los dos conglomerados (5,6,2) o (1,7,3) con la similitud más grande de 1/3. Se elige el conglomerado formado por las personas(1,7,4) como se muestra a continuación:

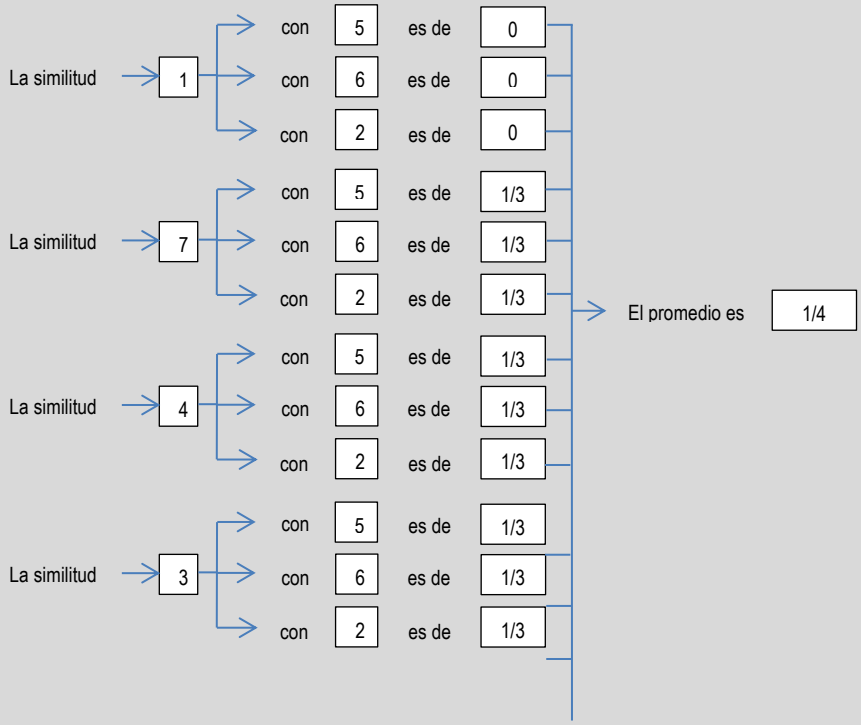
	(5,6,2)	(1,7,4)	3
(5,6,2)	1		
(1,7,4)	2/9	1	
3	1/3	4/9	1

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(5,6,2)	(1,7,4, 3)
(5,6,2)	1	
(1,7,4, 3)		1

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la similitud promedio, es decir se compara al conglomerado (5,6,2) con el resto de las personas que forman el otro conglomerado (1,7,4,3); al comparar (5,6,2) con las personas (1,7,4,3) se debe verificar en la matriz original las similitudes entre todas las combinaciones, es decir a las personas 1 y 5 que es 0, a la persona 1 y 6 que es 0, a la persona 1 y 2 que es 0 y así sucesivamente hasta la persona 3 y 2 que es 1/3, eligiendo entonces la similitud promedio de las doce combinaciones o siete personas que en este caso es 1/4 o 0.25. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor resultante se acomoda en la matriz reducida:

	(5,6,2)	(1,7,3, 4)
(5,6,2)	1	
(1,7,3, 4)	1/4	1

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	5	6	2
2	5	1	2	5	3
3	4	2/3	1	7	2
4	3	1/2	1	4	3
5	2	4/9	1	3	4
6	1	1/4	1	2	7

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 5 y la 6 y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 2 al primer conglomerado formado por las personas 5 y 6 como se muestra en la columna 3 a una similitud de 1 o 100%. En la etapa 3 se forma un nuevo conglomerado como se aprecia en la columna 5 formado por las personas 1 y 7 a una similitud de 2/3 ó 66.667%. En la etapa 4 se incorporó la persona 4 al segundo conglomerado formado por las personas 1 y 7 como se muestra en la columna 3 a una similitud de 1/2 ó 50.00%. En la etapa 5 se incorporó la persona 3 al sugundo conglomerado formado por las personas 1,7 y 3 como se muestra en la columna 3 a una similitud de 4/9 ó 44.444% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 1/4 ó 25.00%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Altura, Peso, Busto

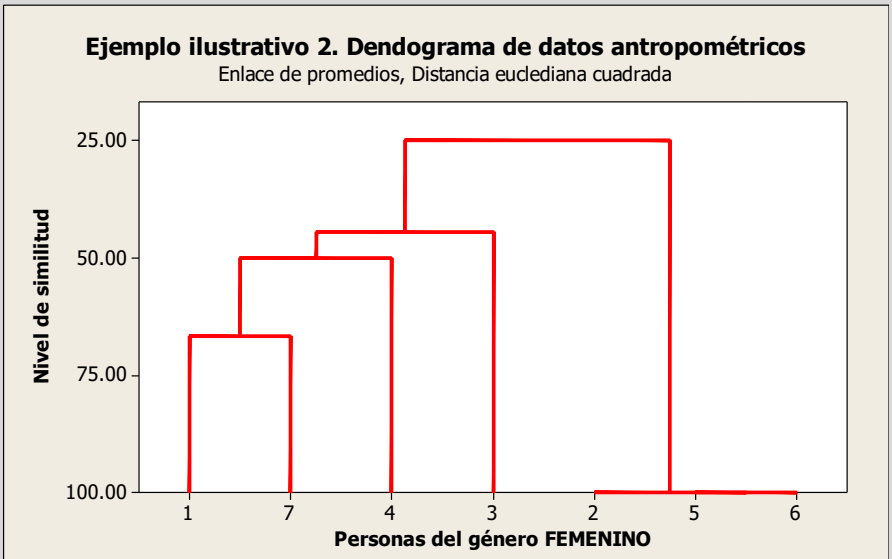
```
Distancia euclediana cuadrada, Enlace de promedios
Pasos de amalgamación


      Número                                                  Número de
          de    Nivel de   Nivel de     Grupos     Nuevo   obs. en el
Paso  grupos  semejanzal  distancia  incorporados  grupo  grupo nuevo
   1       6     100.000    0.00000  5           6      5            2
   2       5     100.000    0.00000  2           5      2            3
   3       4      66.667    1.00000  1           7      1            2
   4       3      50.000    1.50000  1           4      1            3
   5       2      44.444    1.66667  1           3      1            4
   6       1      25.000    2.25000  1           2      1            7
```

Dendrograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (5,6,2) y el conglomerado 2 por las personas (1,7,4,3).

Cabe mencionar que mientras a menor similitud se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Elaboró: M. en Admón: Javier Bech Vertti

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso

a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 5 y 6 los valores de distancia cambian abruptamente de 4/9 a 1/4 cuando se unen ambos grupos, por lo tanto, la solución más factible parece ser la de 2 conglmerados (5,6,2) y (1,7,4,3)

2.2.1.2 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE APRENDIZAJE 2.2.1.2 ANÁLISIS DE CONGLOMERADOS. MATRIZ DE SIMILITUD.

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de los hombres mexicanos, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras . Para ser más preciso en ciertos resultados, el estudio antropométrico sustituyó el indicador de cintura por el de la cadera contemplando igualmente 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y la cadera de cinco personas del género **MASCULINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Cadera
1	1.73	63.56	Estrecha
2	1.85	83.99	Ancha
3	1.70	74.91	Ancha
4	1.63	54.48	Estrecha
5	1.93	95.34	Ancha

6	1.69	69.00	Estrecha
7	1.67	70.00	Ancha

Criterio de similitud para recodificar los datos:

Altura: existe(1) si es ≥ a 1.71 mts.
no existe(0) si es < a 1.71 mts.

Peso: existe(1) si es ≥ a 71 Kgs.
no existe(0) si es < a 71 Kgs.

Cadera: existe(1) si es Ancha
no existe (0) si es cualquier otro tipo

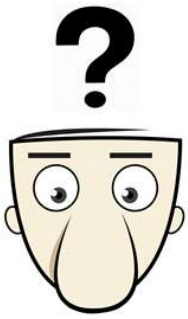
a) Obtenga la pareja de personas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

2.2.1.2 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra al final del cuaderno de trabajo en el anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 2.2.1.2 ANÁLISIS DE CONGLOMERADOS. MATRIZ DE SIMILITUD

Suponga que se desea segmentar a los clientes de un supermercado con base a la actitud que tienen cuando salen de compras. De acuerdo a una investigación previa, se identificaron cuatro variables de actitud y se pidió a los siete entrevistados que expresaran su grado de acuerdo con las dos primeras de las cuatro afirmaciones en una escala de siete puntos en la cual 1 significa en desacuerdo y 7 significa de acuerdo. Para las dos restantes se quiso ser más preciso en los resultados y se modificó la escala de medición cuantitativa por una cualitativa
Las afirmaciones podrían ser las siguientes:
V1 Salir de compras es divertido



V2 Combino la salida de compras con la comida fuera de casa
V3 Puedo ahorrar mucho dinero si comparo precios
V4 Gusto por salir de compras con mi familia

Si los resultados para cada una de las afirmaciones se considera que son variables, se tienen cuatro, las cuales se representarán con V1, V2, V3 y V4; por otra parte, las respuestas de cada uno de los entrevistados constituyen los siete casos, que se representarán con C1, C2,…,C7. Los valores de las variables para los siete casos, se presentan a continuación:

ENTREVISTADO	**VARIABLE**			
	V1	**V2**	**V3**	**V4**
C1	1	2	2	NADA
C2	5	7	7	MUCHO
C3	2	2	1	NADA
C4	7	5	7	MUCHO
C5	3	3	2	NADA
C6	6	4	6	MUCHO
C7	4	6	7	POCO

Criterio de similitud para recodificar los datos:

V1: existe(1) si es ≥ a 4.
no existe(0) si es < a 4.

V2: existe(1) si es ≥ a 3.
no existe(0) si es < a 3.

V3: existe(1) si es ≥ a 3.
no existe(0) si es < a 3.

V4: existe(1) si es Mucho
no existe (0) si es cualquier otro gusto

a) Obtenga la pareja de personas entrevistadas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

C.A.

EJEMPLO ILUSTRATIVO EN MINITAB 17

EJEMPLO ILUSTRATIVO INTEGRAL EN MINITAB 17. ANÁLISIS DE CONGLOMERADOS. MEDIDA DE SIMILITUD

MINITAB 17

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de las mujeres mexicanas, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras . Para ser más precisos en ciertos resultados, el estudio antropométrico sustituyó el indicador de cintura por el de busto contemplando igualmente 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y el busto de cinco personas del género **FEMENINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Busto
1	1.52	57.12	Mediano
2	1.67	76.85	Grande
3	1.54	67.62	Mediano
4	1.56	49.35	Pequeño
5	1.63	87.22	Grande
6	1.65	82.54	Grande
7	1.68	53.12	Mediano

Criterio de similitud para recodificar los datos:

Altura: existe(1) si es ≥ a 1.60 mts.
no existe(0) si es < a 1.60 mts.

Peso: existe(1) si es ≥ a 60 Kgs.
no existe(0) si es < a 60 Kgs.

Busto: existe(1) si es Mediano
no existe (0) si es cualquier otro tamaño

a) Obtenga la pareja de personas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

Resolución del inciso a)

Los datos recodificados quedaron como sigue:

Persona	Altura (mts.)	Peso (Kg.)	Busto
1	0	0	1
2	1	1	0
3	0	1	1
4	0	0	0
5	1	1	0
6	1	1	0
7	1	0	1

Como esta matriz es una matriz espejo, es decir la matriz superior es igual a la matriz inferior y la diagonal contiene 1´s ya que la similitud máxima que hay al comparar una persona consigo es de uno, sólo será necesario utilizar la información de la matriz inferior como se muestra a continuación:

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

Al observar la matriz de similitud se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya similitud entre sí es la que se acerque más o sea igual a 1, en este caso tenemos tres parejas, la primera formada por las personas 2 y 5, con una medida de similitud de 1; la segunda formada por las personas 2 y 6, con una similitud también de 1 y la tercera formada por las personas 5 y 6, con una similitud de 1.

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

La pareja de personas menos parecida será aquella cuya similitud entre sí es la se acerque más o sea igual a 0. En este caso tenemos tres parejas, la primera formada por las personas 1 y 2, con una medida de similitud de 0; la segunda formada por las personas 1 y 5, con una similitud también de 0 y la tercera formada por las personas 1 y 6, con una similitud de 0.

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	2/3	1/3	1				
4	2/3	1/3	1/3	1			
5	0	1	1/3	1/3	1		
6	0	1	1/3	1/3	1	1	
7	2/3	1/3	1/3	1/3	1/3	1/3	1

· Con el eslabonamiento simple, o del "vecino más cercano", la similitud entre dos conglomerados es la similitud máxima entre una observación en un conglomerado y una observación en el otro conglomerado. El eslabonamiento simple es una opción apropiada cuando los conglomerados están claramente separados. Cuando las observaciones están cerca entre sí, el eslabonamiento simple tiende a identificar conglomerados largos similares a cadenas que pueden tener una separación relativamente grande entre observaciones en cualquier extremo de la cadena. ·

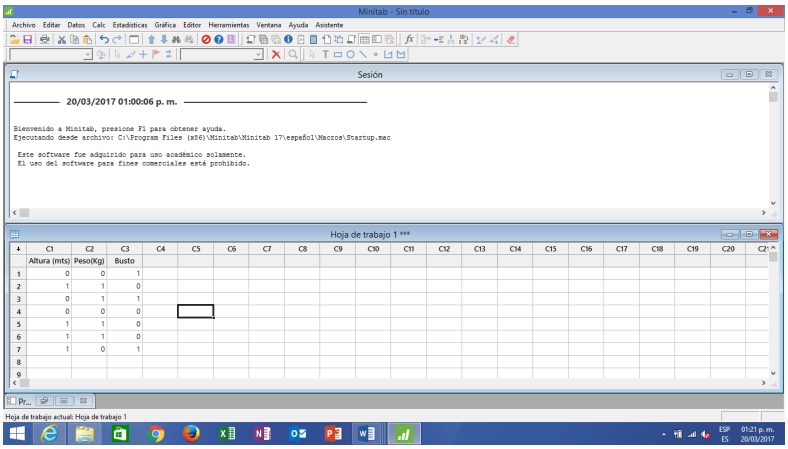
Resolución del inciso b)

Solución. Eslabonamiento Simple o Individual:

Método	La distancia entre dos conglomerados es...	Razones para utilizar este método
Simple o Individual	La distancia mínima entre un elemento en un conglomerado y un elemento en el otro conglomerado. Se conoce también como el método del "vecino más cercano"	Más apropiado para observaciones que están claramente separadas. Cuando están cerca, el enlace simple o individual tiende a identificar conglomerados largos similares a cadenas que pueden tener una separación relativamente grande entre elementos en cualquier extremo de la cadena.

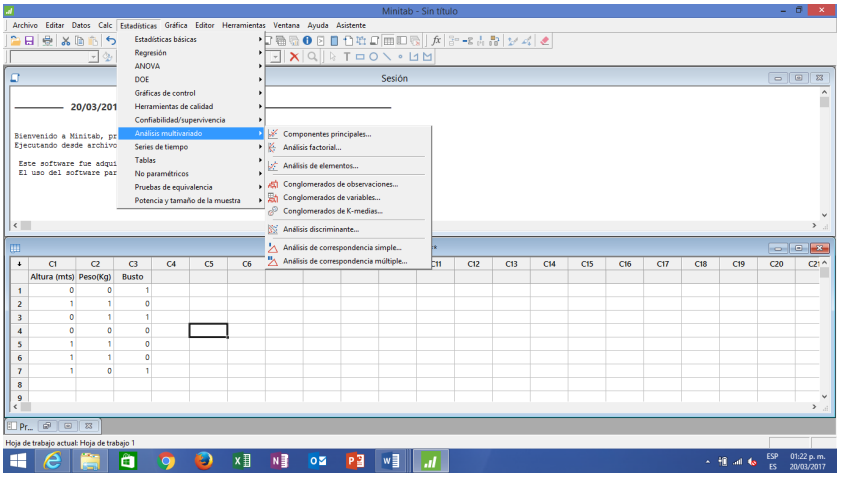
Cuando el número de observaciones **es extenso y/o existen muchas variables,** los cálculos manuales son tediosos. Existen muchos paquetes de software que pueden mostrar los resultados entre ellos **Minitab.**

Comenzamos introduciendo los **datos recodificados** en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:

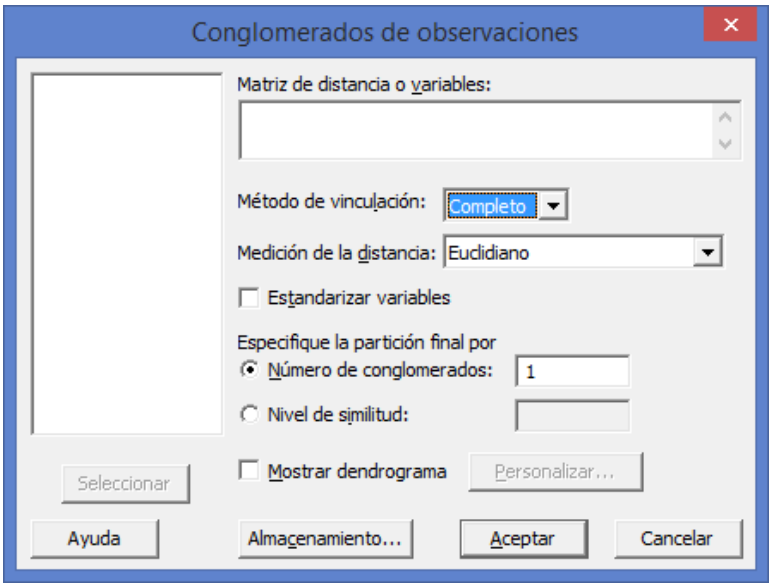


Elija de la barra menú ***Estadísticas> Análisis Multivariado>Conglomerado de observaciones.***

La opción de ***conglomerados de observaciones*** se utiliza para clasificar los elementos en grupos cuando éstos se desconocen inicialmente. En **Minitab** este procedimiento utiliza **el método jerárquico aglomerativo** que comienza con todos los elementos separados, cada uno de ellos formando su propio conglomerado. En el primer paso, los dos elementos más cercanos se unen entre sí. En el siguiente paso, un tercer elemento se une a los dos primeros u otros dos elementos se unen para formar un conglomerado diferente. Este proceso continúa hasta que todos los conglomerados se unan en un solo conglomerado; sin embargo, este conglomerado individual no es útil para propósitos de clasificación, por lo tanto, usted debe decidir cuántos grupos son lógicos para sus datos y clasificarlos adecuadamente.



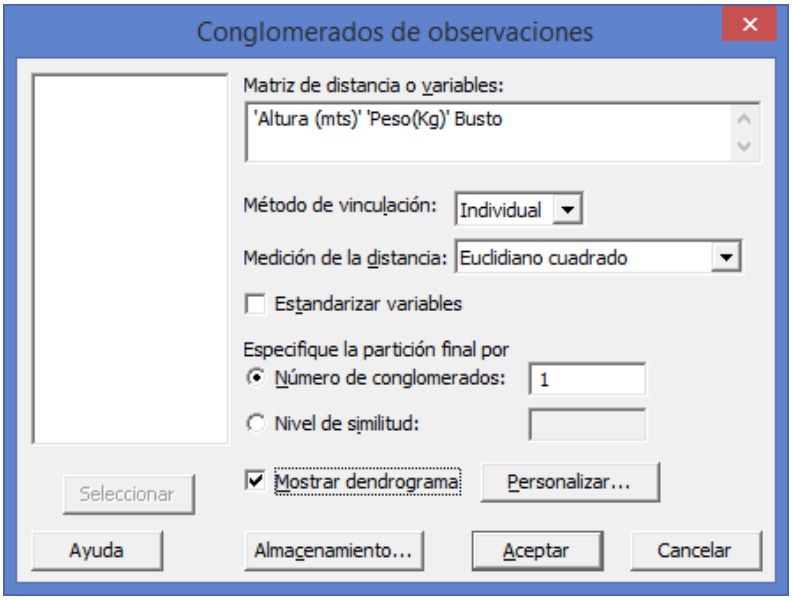
Aparece el siguiente cuadro de Diálogo:



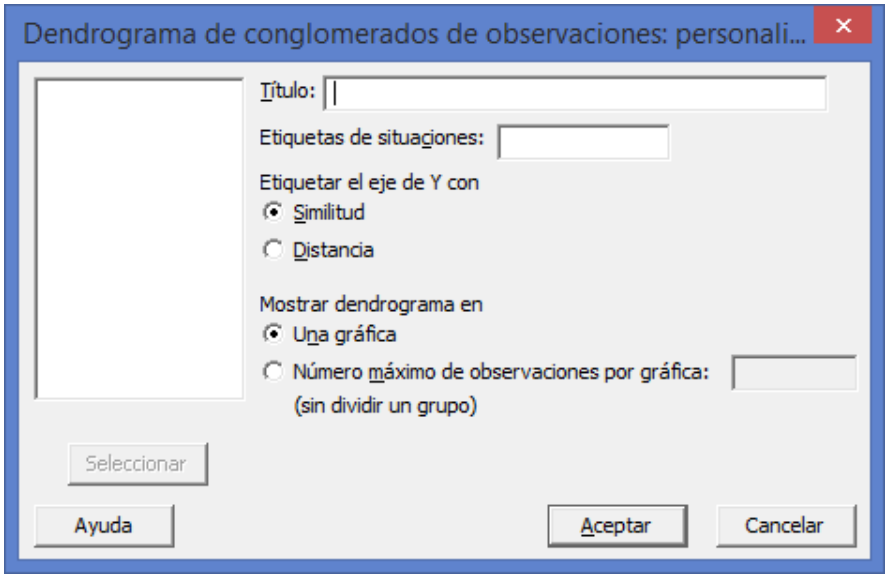
En ***Matriz de distancia o variables***, de un click para que aparezcan las tres variables e ingrese las tres variables dando doble click sobre cada una de ellas o seleccione las tres y oprima el botón Seleccionar.

En ***Método de vinculación***, elija ***Individual*** en ***Medición de la distancia*** elija ***Euclidiano cuadrado***.

Marque la opción ***Mostrar dendograma***.

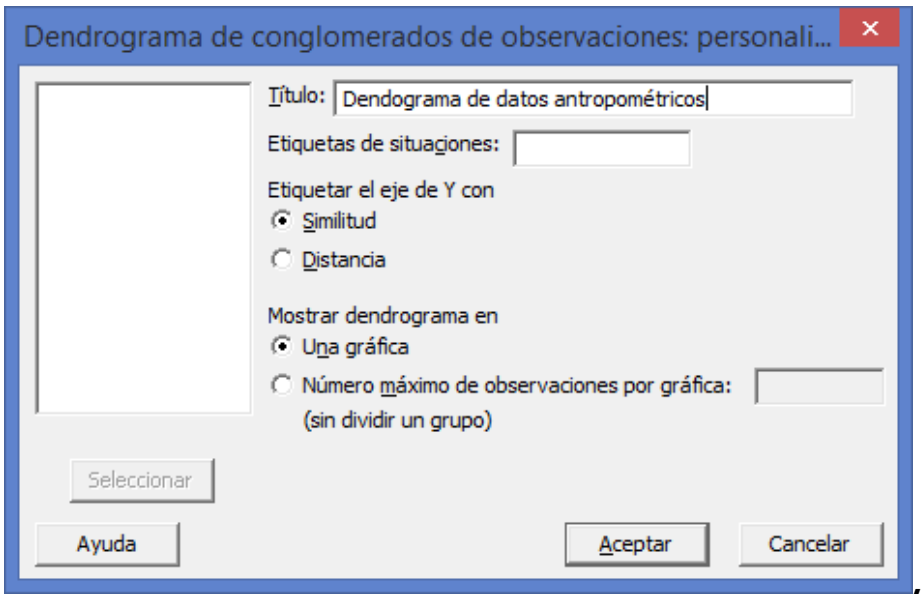


Haga clic en ***Personalizar.*** Aparece el siguiente cuadro de diálogo:



En ***Título***, ingrese ***Dendograma de datos antropométricos***.

En ***Etiquetar el eje de Y con*** elija ***Similitud*** aunque por default ya aparece esta opción



Haga clic en ***Aceptar*** en cada cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts), Peso(Kg), Busto

```
Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación


Número de obs. en
          Número de    Nivel de   Nivel de  Conglomerados         Nuevo
el conglomerado
Paso  conglomerados  semejanzal  distancia   incorporados  conglomerado
nuevo
   1              6     100.000          0  5             6             5
2
   2              5     100.000          0  2             5             2
3
   3              4      66.667          1  1             7             1
2
   4              3      66.667          1  1             4             1
3
   5              2      66.667          1  1             3             1
4
   6              1      33.333          2  1             2             1
7
```

Interpretación de los resultados

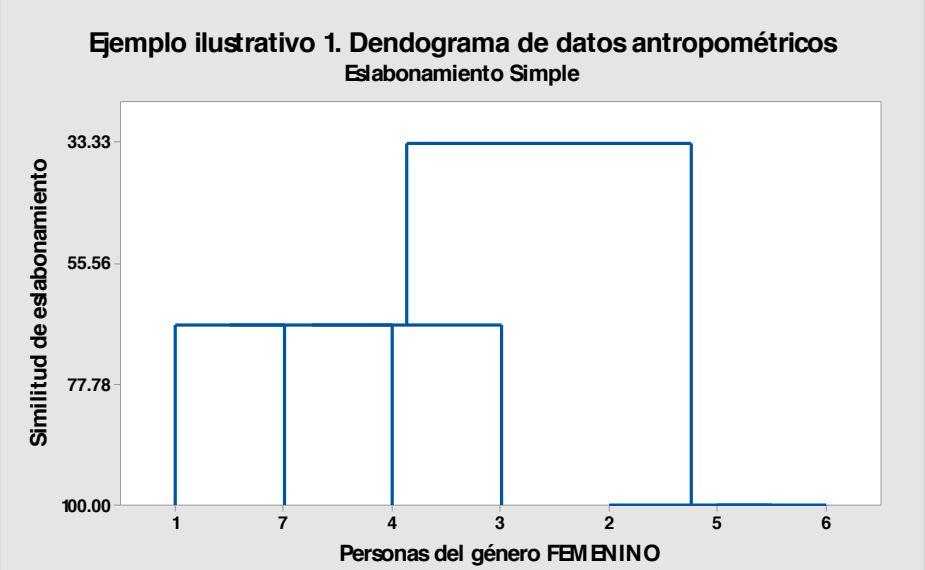
Minitab muestra los pasos de eslabonamiento en la ventana Sesión. En cada paso, se unen dos conglomerados (personas). La tabla muestra los conglomerados que se unieron, la distancia entre ellos, el nivel de similitud correspondiente, el número de identificación del conglomerado nuevo (este número siempre corresponde al número más pequeño de los dos conglomerados que se unieron), el número de observaciones en el conglomerado nuevo y el número de conglomerados. El eslabonamiento continúa hasta que sólo queda un conglomerado.

Los pasos de eslabonamiento muestran que en principio, en el paso 1, las personas más parecidas son la 6 y la 5 y forman el primer conglomerado a una similitud de eslabonamiento de 100.00, posteriormente en el paso 2 se incorpora a este conglomerado la persona 2 a una similitud de eslabonamiento también de 100.00. En el paso 3 se observa que se forma un segundo conglomerado con las personas 7 y 1 a una similitud de eslabonamiento de 66.667. En el paso 4 se observa que la persona 4 se une al segundo conglomerado con una similitud de eslabonamiento igual de 66.667. En el paso 5 se observa que la persona 3 se une al segundo conglomerado con una similitud de eslabonamiento igual de 66.667 para finalmente unirse en el paso 6 ambos conglomerados a una similitud de eslabonamiento de 33.333

Esto indica que tal vez dos conglomerados son razonablemente suficientes para la partición o solución final. Si esta agrupación tiene sentido intuitivo para los datos, entonces es probable que sea una elección apropiada.

Dendrograma

La agrupación final de conglomerados (también denominada partición final) es la agrupación de conglomerados que, si los resultados son favorables, identificará grupos cuyas observaciones o variables compartan características comunes. La decisión acerca de la agrupación final también se denomina Dendograma. El dendograma completo (diagrama de árbol) es una representación gráfica de la conformación en amalgama de los elementos en un solo conglomerado.



Elaboró: M. en Admón. Javier Bech Vertti

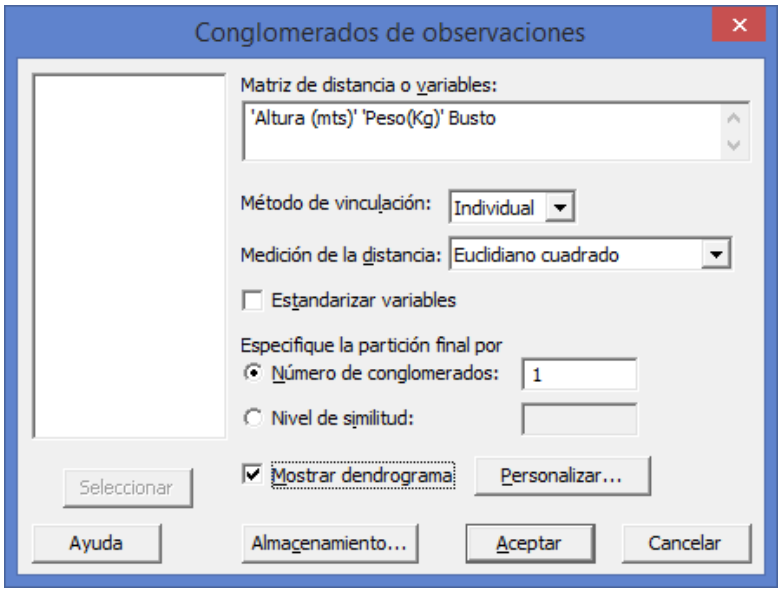
Por opción predeterminada, el dendograma presenta el nivel de similitud en el eje y las diferentes observaciones se enumeran con el eje horizontal. La gráfica muestra la manera en la cual se formaron los conglomerados, bien sea uniendo dos observaciones individuales o pareando una observación individual con un conglomerado existente. Puede determinar en cuáles similitudes de eslabonamiento se forman los conglomerados así como la composición de los conglomerados de la solución final.

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados, aunque en la práctica puede ejecutar primero un análisis de conglomerados **sin especificar una partición final con el objeto de decidir dónde colocarla**. Examine los niveles de similitud en los resultados de la ventana Sesión y en el dendograma. Puede ver las similitudes de eslabonamiento en el dendograma colocando el puntero de su ratón sobre una línea horizontal. La similitud de eslabonamiento en cualquier paso es el valor de la similitud mínima en el paso, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde

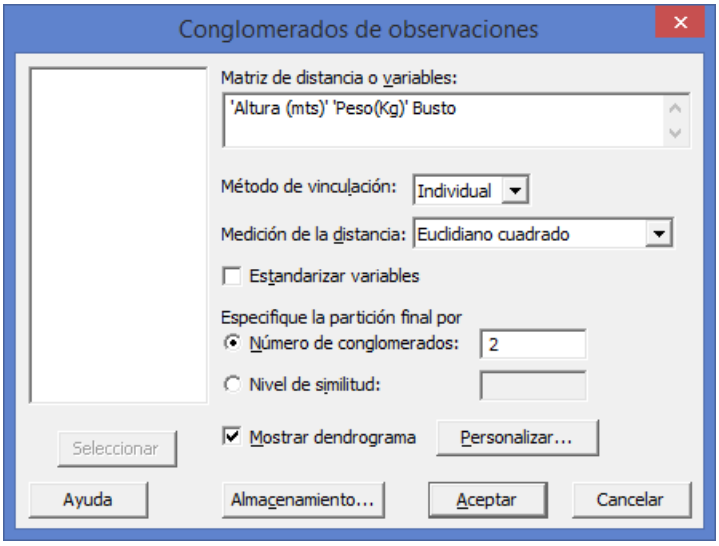
los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Después de elegir dónde desea realizar su partición, vuelva a ejecutar el procedimiento del conglomerado, utilizando, el número de conglomerados para obtener un número establecido de conjuntos de la siguiente manera:

En la parte inicial del análisis apareció el siguiente cuadro de Diálogo:



Ahora en ***Especifique la partición final por***, elija ***Número de conglomerados*** e ingrese ***2***.



Haga clic en ***Aceptar*** en el cuadro de diálogo.

Salida de la ventana Sesión

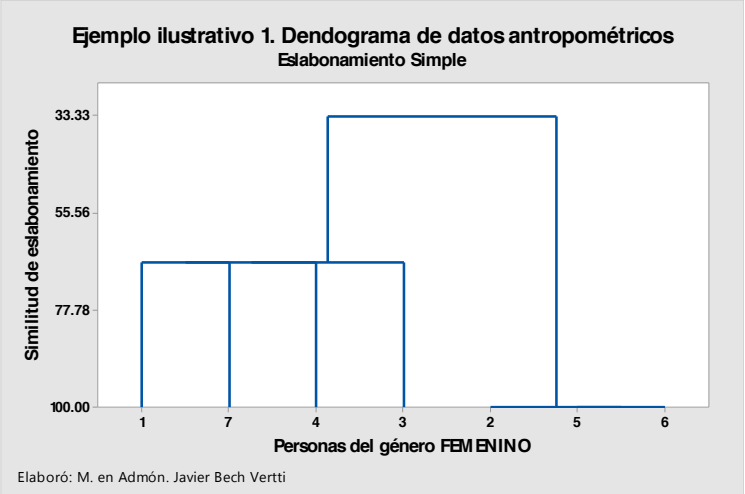
Análisis de observaciones de conglomerado: Altura (mts), Peso(Kg), Busto

```
Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100.000	0	5	6	5	2
2	5	100.000	0	2	5	2	3
3	4	66.667	1	1	7	1	2
4	3	66.667	1	1	4	1	3
5	2	66.667	1	1	3	1	4
6	1	33.333	2	1	2	1	7

Cuando usted especifica la partición o solución final, Minitab muestra entre otras una tabla adicional que resume cada conglomerado según el número de observaciones, en este caso el primer conglomerado está formado por cuatro mujeres (1,7,4 y 3) y el segundo conglomerado por las mujeres (2,5 y 6)

Dendrograma



Examine los conglomerados resultantes en la partición o solución final para observar si la agrupación parece lógica (Las personas que forman cada conglomerado aparecen con un color distinto en cada conglomerado). La observación de dendogramas para determinar diferentes agrupaciones finales

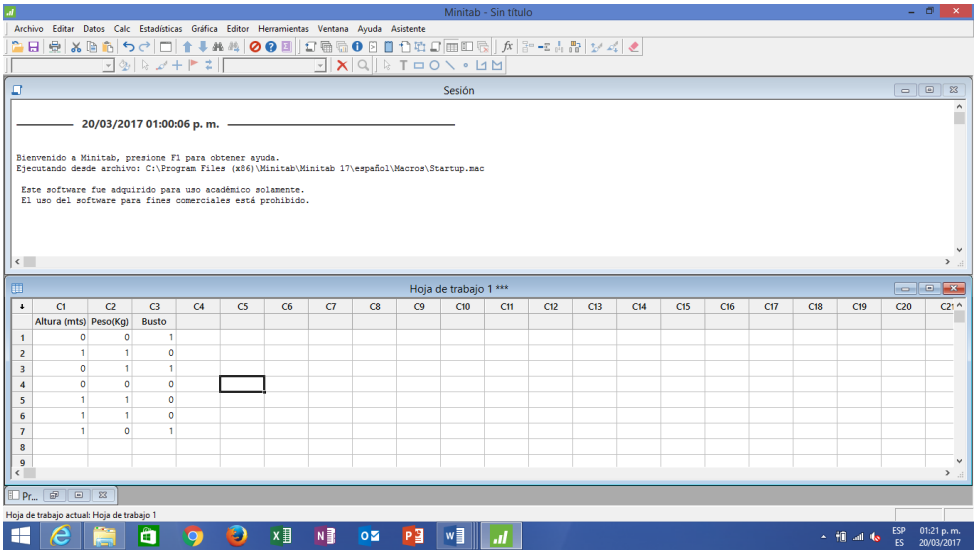
puede ayudarle a decidir cuáles de ellas tienen más sentido para sus datos.

Solución. Eslabonamiento Completo:

Método	La distancia entre dos conglomerados es...	Razones para utilizar este método
Completo	La distancia máxima entre un elemento en un conglomerado y un elemento en el otro conglomerado. Se conoce también como el método del "vecino más lejano".	Asegura que todos los elementos en un conglomerado se encuentren dentro de una distancia máxima y tiende a producir conglomerados con diámetros similares. Los resultados pueden ser sensibles a valores atípicos.

· Con el eslabonamiento completo, o "del vecino más lejano", la similitud entre dos conglomerados es la similitud mínima entre una observación en un conglomerado y una observación en el otro conglomerado. Este método garantiza que todas las observaciones en un conglomerado se encuentren dentro de una similitud mínima y tiende a producir conglomerados con diámetros similares. Los resultados pueden ser sensibles a valores atípicos.

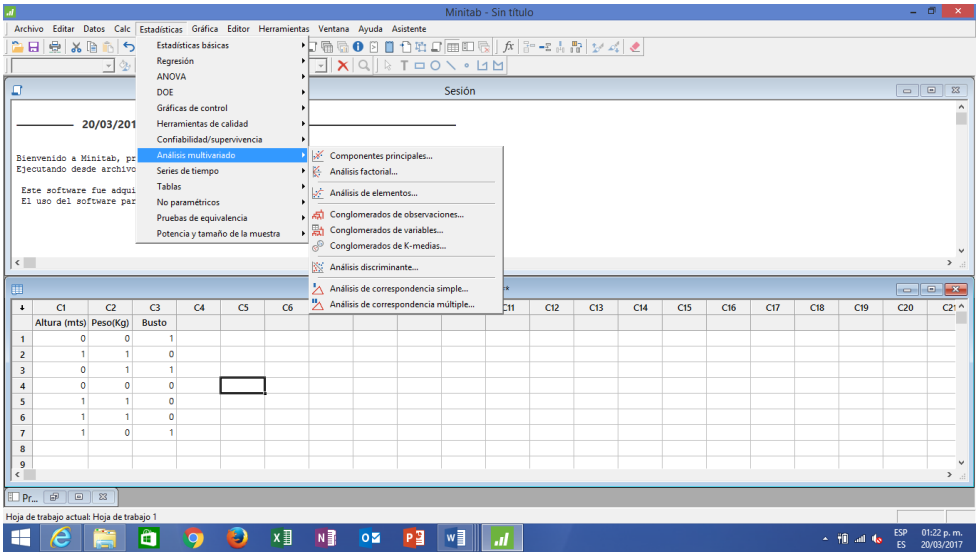
Comenzamos introduciendo los **datos recodificados** en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:



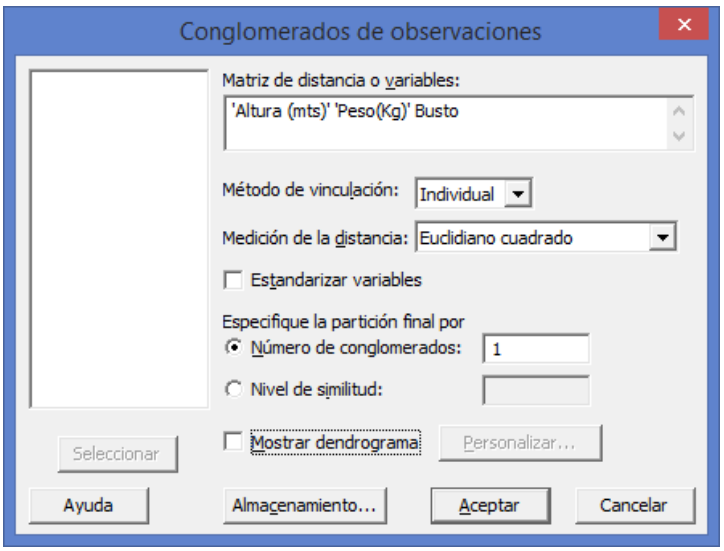
Elija de la barra menú ***Estadísticas> Análisis Multivariado>Conglomerado de observaciones.***

La opción de ***conglomerados de observaciones*** se utiliza para clasificar los elementos en grupos cuando éstos se desconocen inicialmente. En **Minitab** este procedimiento utiliza **el método jerárquico aglomerativo** que comienza con

todos los elementos separados, cada uno de ellos formando su propio conglomerado. En el primer paso, los dos elementos más cercanos se unen entre sí. En el siguiente paso, un tercer elemento se une a los dos primeros u otros dos elementos se unen para formar un conglomerado diferente. Este proceso continúa hasta que todos los conglomerados se unan en un solo conglomerado; sin embargo, este conglomerado individual no es útil para propósitos de clasificación, por lo tanto, usted debe decidir cuántos grupos son lógicos para sus datos y clasificarlos adecuadamente.



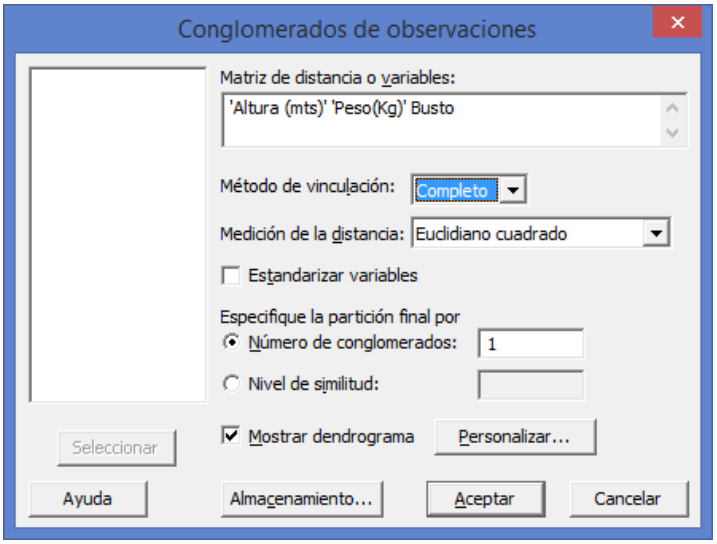
Aparece el siguiente cuadro de Diálogo:



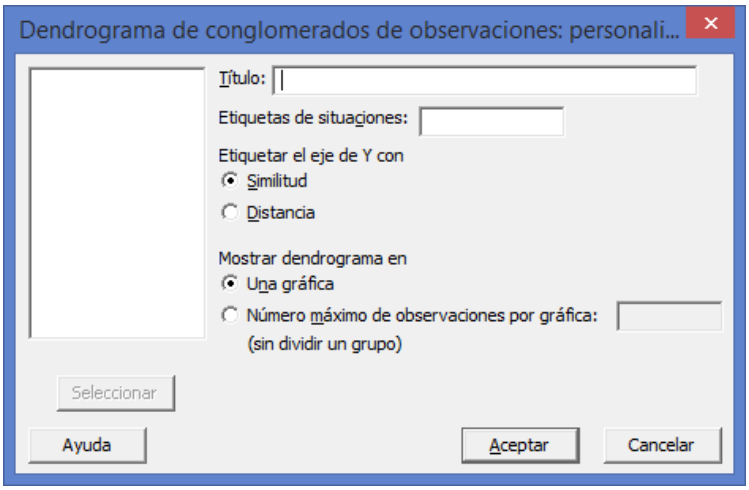
En ***Matriz de distancia o variables***, de un click para que aparezcan las tres variables e ingrese las tres variables dando doble click sobre cada una de ellas o seleccione las tres y oprima el botón Seleccionar.

En ***Método de vinculación***, elija ***Completo*** en ***Medición de la distancia*** elija ***Euclidiano Cuadrado***

Marque la opción ***Mostrar dendograma***.



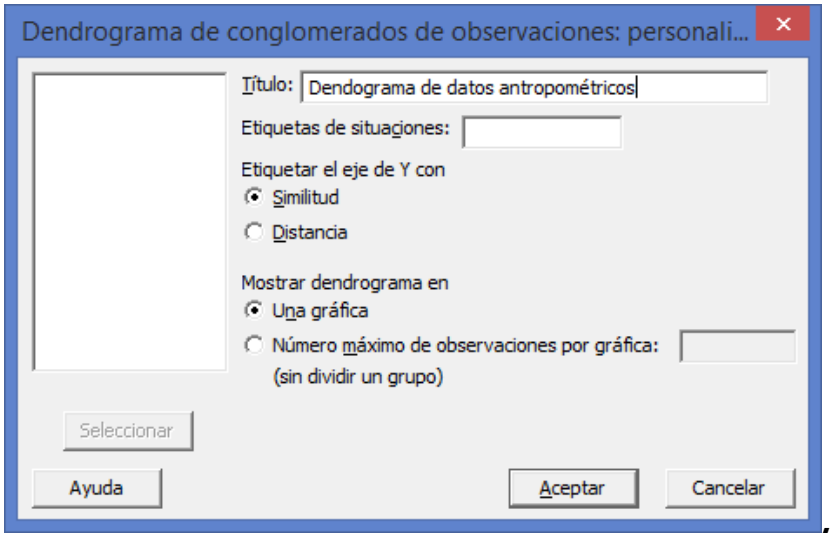
Haga clic en ***Personalizar.*** Aparece el siguiente cuadro de diálogo:



En ***Título***, ingrese ***Dendograma de datos antropométricos***.

En ***Etiquetar el eje de Y con*** elija ***Similitud*** aunque por default ya aparece esta

opción



Haga clic en ***Aceptar*** en cada cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts), Peso(Kg), Busto

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación


Número de obs. en
          Número de    Nivel de   Nivel de  Conglomerados         Nuevo
el conglomerado
Paso  conglomerados  semejanzal  distancia   incorporados  conglomerado
nuevo
   1              6     100.000          0  5             6             5
2
   2              5     100.000          0  2             5             2
3
   3              4      66.667          1  1             7             1
2
   4              3      33.333          2  3             4             3
2
   5              2      33.333          2  2             3             2
5
   6              1       0.000          3  1             2             1
7
```

Interpretación de los resultados

Minitab muestra los pasos de eslabonamiento en la ventana Sesión. En cada paso, se unen dos conglomerados (personas). La tabla muestra los conglomerados que se unieron, la distancia entre ellos, el nivel de similitud correspondiente, el

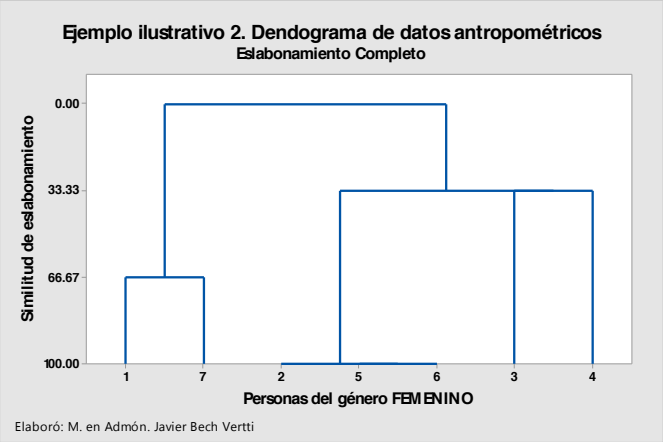
número de identificación del conglomerado nuevo (este número siempre corresponde al número más pequeño de los dos conglomerados que se unieron), el número de observaciones en el conglomerado nuevo y el número de conglomerados. El eslabonamiento continúa hasta que sólo queda un conglomerado.

Los pasos de eslabonamiento muestran que en principio, en el paso 1, las personas más parecidas son la 6 y la 5 y forman el primer conglomerado a una similitud de eslabonamiento de 100.00, posteriormente en el paso 2 se incorpora a este conglomerado la persona 2 a una similitud de eslabonamiento también de 100.00. En el paso 3 se observa que se forma un segundo conglomerado con las personas 7 y 1 a una similitud de eslabonamiento de 66.667. En el paso 4 se observa que se forma un tercer conglomerado con las persona 4 y 3 a una similitud de eslabonamiento igual de 33.333. En el paso 5 se observa que el primer conglomerado formado por las personas 6,5 y 2 se une con el tercer conglomerado formado con las personas 4 y 3 a una similitud de eslabonamiento de 33.333 para finalmente unirse en el paso 6 ambos conglomerados al segundo conglomerado formado por las personas 7 y 1 a una similitud de eslabonamiento de 0.000.

Esto indica que tal vez tres conglomerados son razonablemente suficientes para la partición o solución final. Si esta agrupación tiene sentido intuitivo para los datos, entonces es probable que sea una elección apropiada.

Dendrograma

Ilustra la información en la tabla de eslabonamiento en la forma de un diagrama de árbol.



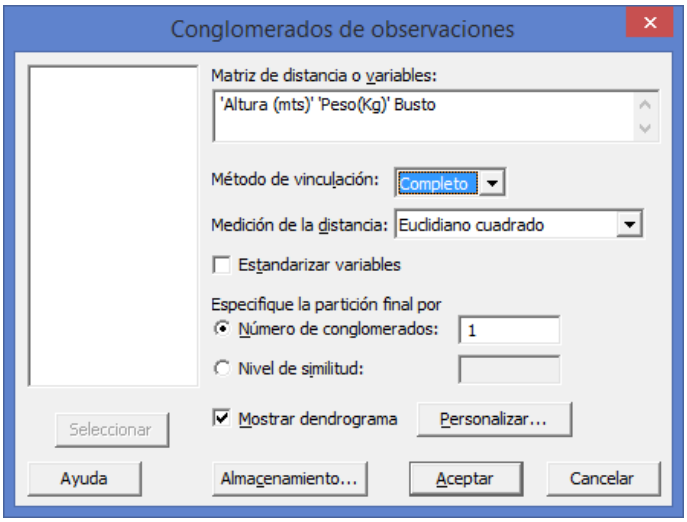
Por opción predeterminada, el dendograma presenta el nivel de similitud en el eje vertical (de forma alterna, como en este caso, puede mostrar el nivel de distancia) y las diferentes observaciones se enumeran con el eje horizontal. La gráfica muestra la manera en la cual se formaron los conglomerados, bien sea uniendo dos observaciones individuales o pareando una observación individual con

un conglomerado existente. Puede determinar en cuáles similitudes de eslabonamiento se forman los conglomerados así como la composición de los conglomerados de la solución final.

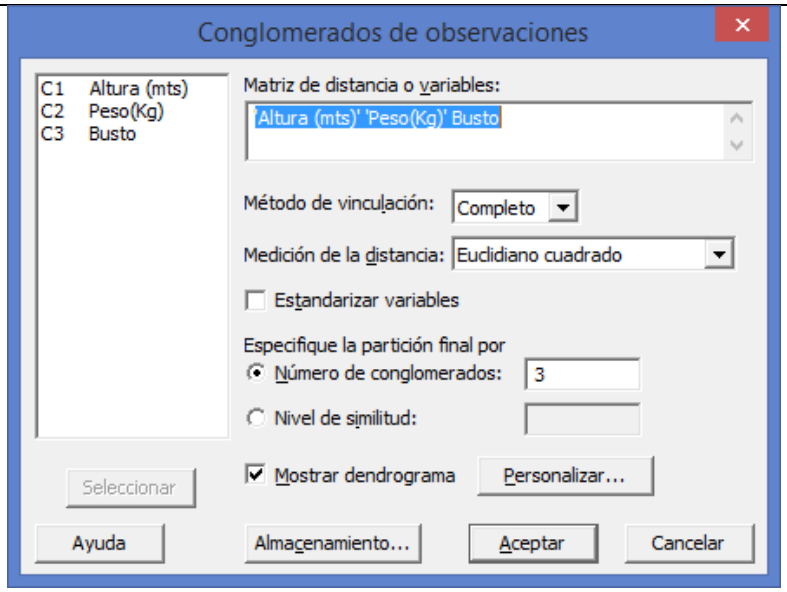
¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados, aunque en la práctica puede ejecutar primero un análisis de conglomerados **sin especificar una partición final con el objeto de decidir dónde colocarla**. Examine los niveles de similutud en los resultados de la ventana Sesión y en el dendograma. Puede ver las similitudes de eslabonamiento en el dendograma colocando el puntero de su ratón sobre una línea horizontal. La similitud de eslabonamiento en cualquier paso es el valor de la similitud mínima en el paso, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Después de elegir dónde desea realizar su partición, vuelva a ejecutar el procedimiento del conglomerado, utilizando, el número de conglomerados para obtener un número establecido de conjuntos de la siguiente manera:

En la parte inicial del análisis del eslabonamiento completo apareció el siguiente cuadro de Diálogo:



Ahora en ***Especifique la partición final por***, elija ***Número de conglomerados*** e ingrese 3.



Haga clic en ***Aceptar*** en el cuadro de diálogo.

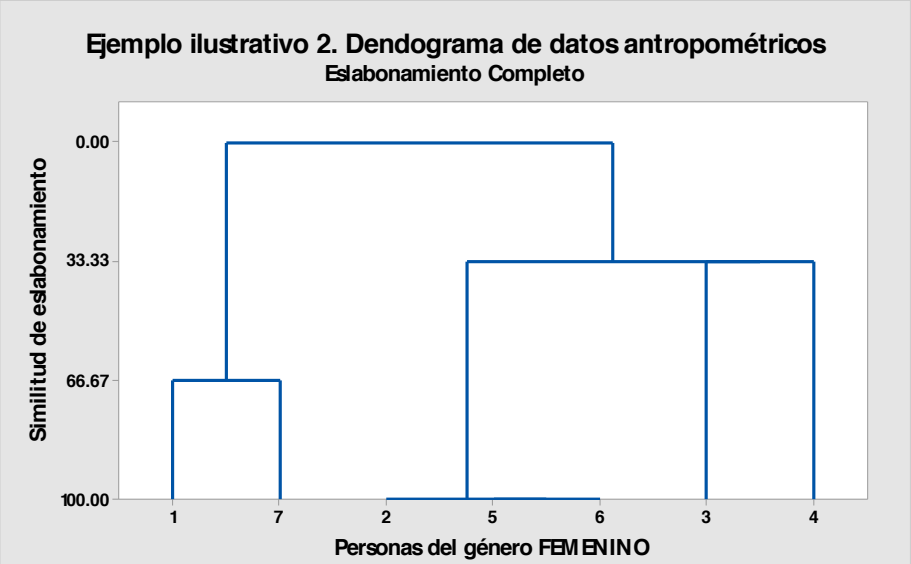
Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts), Peso(Kg), Busto

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación


Número de obs. en
          Número de    Nivel de   Nivel de  Conglomerados          Nuevo
el conglomerado
Paso  conglomerados  semejanzal  distancia   incorporados  conglomerado
nuevo
   1               6     100.000          0  5             6             5
2
   2               5     100.000          0  2             5             2
3
   3               4      66.667          1  1             7             1
2
   4               3      33.333          2  3             4             3
2
   5               2      33.333          2  2             3             2
5
   6               1       0.000          3  1             2             1
7
```

Cuando usted especifica la partición o solución final, Minitab muestra entre otras una tabla adicional que resume cada conglomerado según el número de observaciones, en este caso el primer conglomerado está formado por dos mujeres (7 y 1) el segundo conglomerado por tres mujeres (6,5 y 2) y el tercer conglomerado por las mujeres (4 y 3)

Dendrograma



Elaboró: M. en Admón. Javier Bech Vertti

Examine los conglomerados resultantes en la partición o solución final para observar si la agrupación parece lógica (Las personas que forman cada conglomerado aparecen con un color distinto en cada conglomerado). La observación de dendogramas para determinar diferentes agrupaciones finales puede ayudarle a decidir cuáles de ellas tienen más sentido para sus datos.

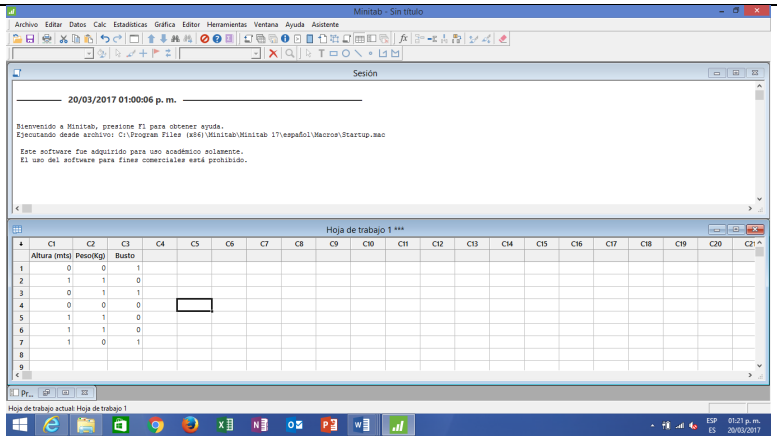
Solución. Eslabonamiento Promedio:

Método	La distancia entre dos conglomerados es...	Razones para utilizar este método
Promedio	La distancia media entre un elemento en un conglomerado y un elemento en el otro conglomerado	En vista de que los métodos de enlace simples o completos agrupan conglomerados sobre la base de distancias entre pares individuales, el enlace promedio utiliza una medida de ubicación más central.

Comenzamos introduciendo los **datos recodificados** en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:

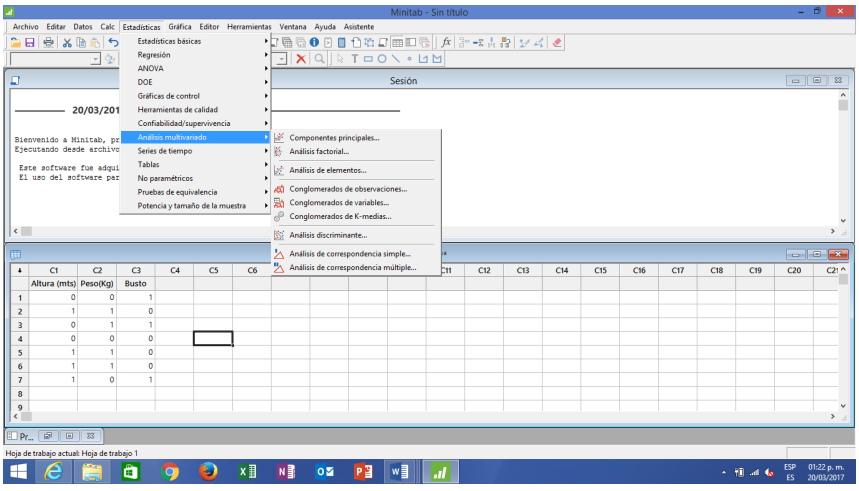
· Con el eslabonamiento promedio, la similitud entre dos conglomerados es la similitud media entre una observación en un conglomerado y una observación en el otro conglomerado. En vista de que los métodos de eslabonamiento simple o completos agrupan conglomerados sobre la base de similitudes entre pares

individuales, el eslabonamiento promedio utiliza una medida de ubicación más central.

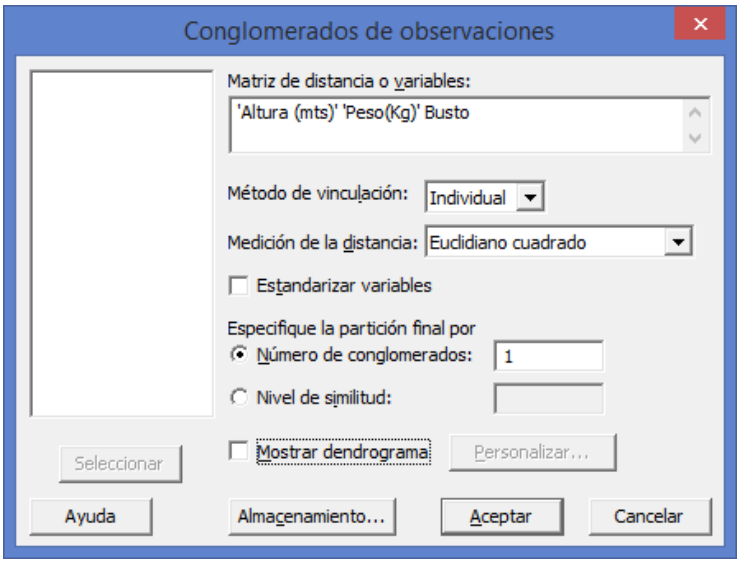


Elija de la barra menú ***Estadísticas> Análisis Multivariado>Conglomerado de observaciones.***

La opción de ***conglomerados de observaciones*** se utiliza para clasificar los elementos en grupos cuando éstos se desconocen inicialmente. En **Minitab** este procedimiento utiliza **el método jerárquico aglomerativo** que comienza con todos los elementos separados, cada uno de ellos formando su propio conglomerado. En el primer paso, los dos elementos más cercanos se unen entre sí. En el siguiente paso, un tercer elemento se une a los dos primeros u otros dos elementos se unen para formar un conglomerado diferente. Este proceso continúa hasta que todos los conglomerados se unan en un solo conglomerado; sin embargo, este conglomerado individual no es útil para propósitos de clasificación, por lo tanto, usted debe decidir cuántos grupos son lógicos para sus datos y clasificarlos adecuadamente.

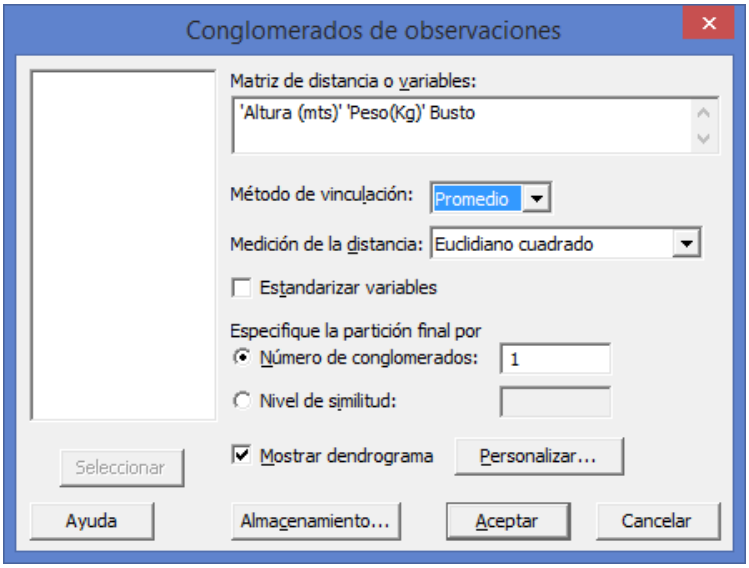


Aparece el siguiente cuadro de Diálogo:

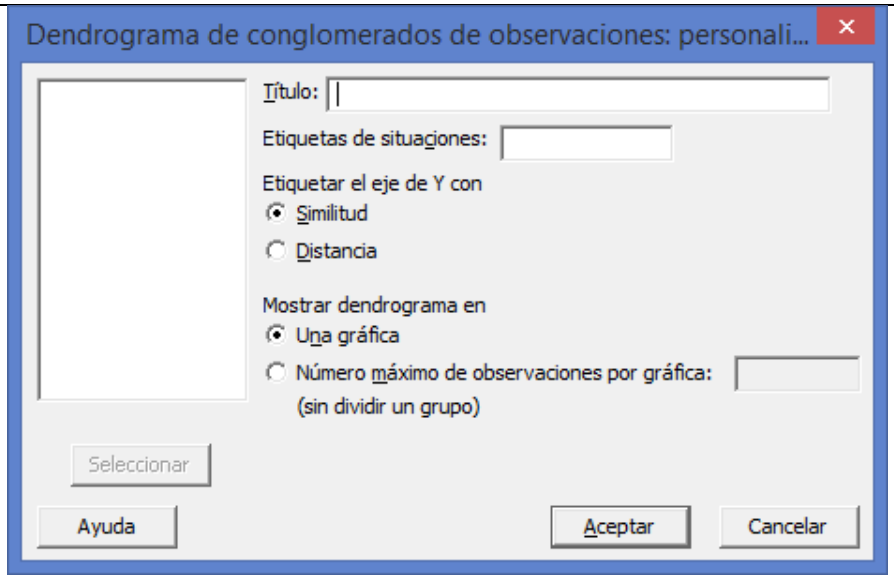


En ***Método de vinculación***, elija ***Promedio*** en ***Medición de la distancia*** elija ***Euclidiano Cuadrado***

Marque la opción ***Mostrar dendograma***.

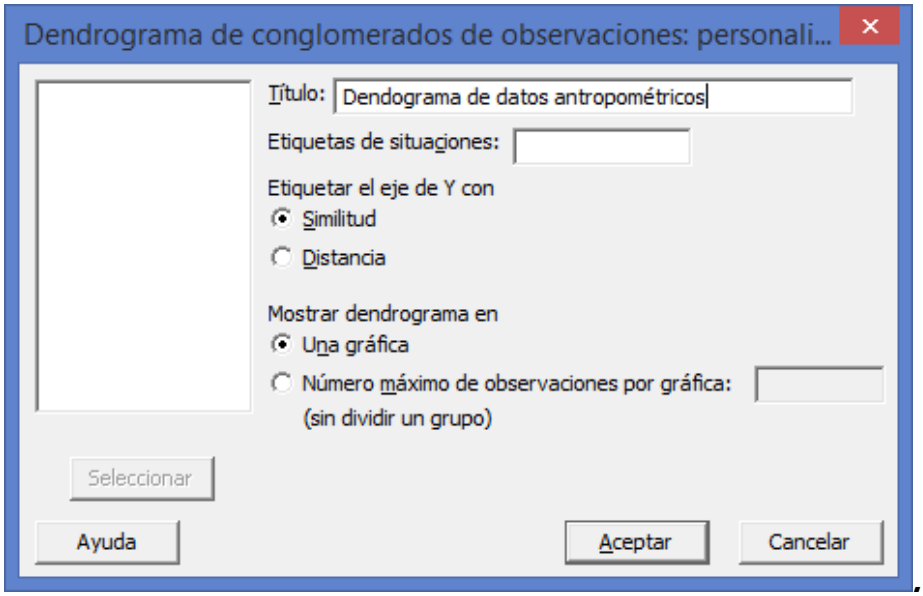


Haga clic en ***Personalizar.*** Aparece el siguiente cuadro de diálogo:



En ***Título***, ingrese ***Dendograma de datos antropométricos***.

En ***Etiquetar el eje de Y con*** elija ***Similitud*** aunque por default ya aparece esta opción



Haga clic en ***Aceptar*** en cada cuadro de diálogo.

Salida de la ventana Sesión

Análisis de observaciones de conglomerado: Altura (mts), Peso(Kg), Busto

```
Distancia eucledíana cuadrada, Enlace de promedios
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100.000	0.00000	5	6	5	2
2	5	100.000	0.00000	2	5	2	3
3	4	66.667	1.00000	1	7	1	2
4	3	50.000	1.50000	1	4	1	3
5	2	44.444	1.66667	1	3	1	4
6	1	25.000	2.25000	1	2	1	7

Interpretación de los resultados

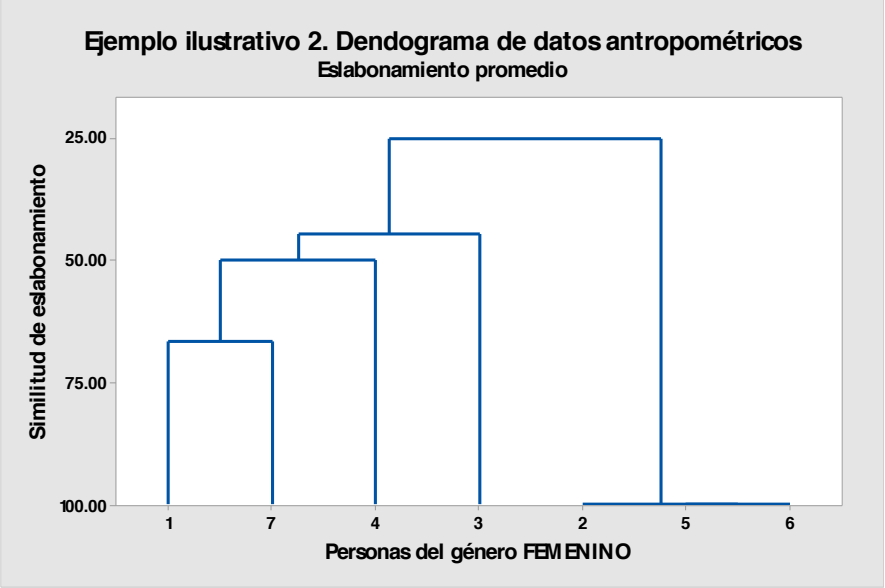
Minitab muestra los pasos de eslabonamiento en la ventana Sesión. En cada paso, se unen dos conglomerados (personas). La tabla muestra los conglomerados que se unieron, la distancia entre ellos, el nivel de similitud correspondiente, el número de identificación del conglomerado nuevo (este número siempre corresponde al número más pequeño de los dos conglomerados que se unieron), el número de observaciones en el conglomerado nuevo y el número de conglomerados. El eslabonamiento continúa hasta que sólo queda un conglomerado.

Los pasos de eslabonamiento muestran que en principio, en el paso 1, las personas más parecidas son la 6 y la 5 y forman el primer conglomerado a una similitud de eslabonamiento de 100.00, posteriormente en el paso 2 se incorpora a este conglomerado la persona 2 a una similitud de eslabonamiento también de 100.00. En el paso 3 se observa que se forma un segundo conglomerado con las personas 7 y 1 a una similitud de eslabonamiento de 66.667. En el paso 4 se observa que se incorpora a este conglomerado la persona 4 a una similitud de eslabonamiento de 50.000. En el paso 5 se incorpora a este mismo conglomerado la persona 3 a una similitud de eslabonamiento de 44.444 para finalmente unirse en el paso 6 ambos conglomerados a una similitud de eslabonamiento de 25.000

Esto indica que tal vez dos conglomerados son razonablemente suficientes para la partición o solución final. Si esta agrupación tiene sentido intuitivo para los datos, entonces es probable que sea una elección apropiada.

Dendrograma

Ilustra la información en la tabla de eslabonamiento en la forma de un diagrama de árbol.

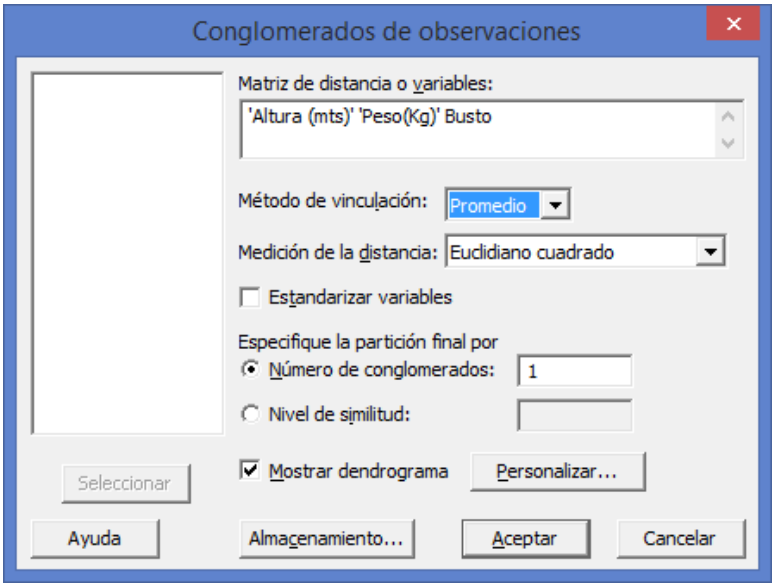


Por opción predeterminada, el dendograma presenta el nivel de similitud en el eje vertical (de forma alterna, como en este caso, puede mostrar el nivel de distancia) y las diferentes observaciones se enumeran con el eje horizontal. La gráfica muestra la manera en la cual se formaron los conglomerados, bien sea uniendo dos observaciones individuales o pareando una observación individual con un conglomerado existente (como en este caso la persona 3). Puede determinar en cuáles distancias de eslabonamiento se forman los conglomerados así como la composición de los conglomerados de la solución final.

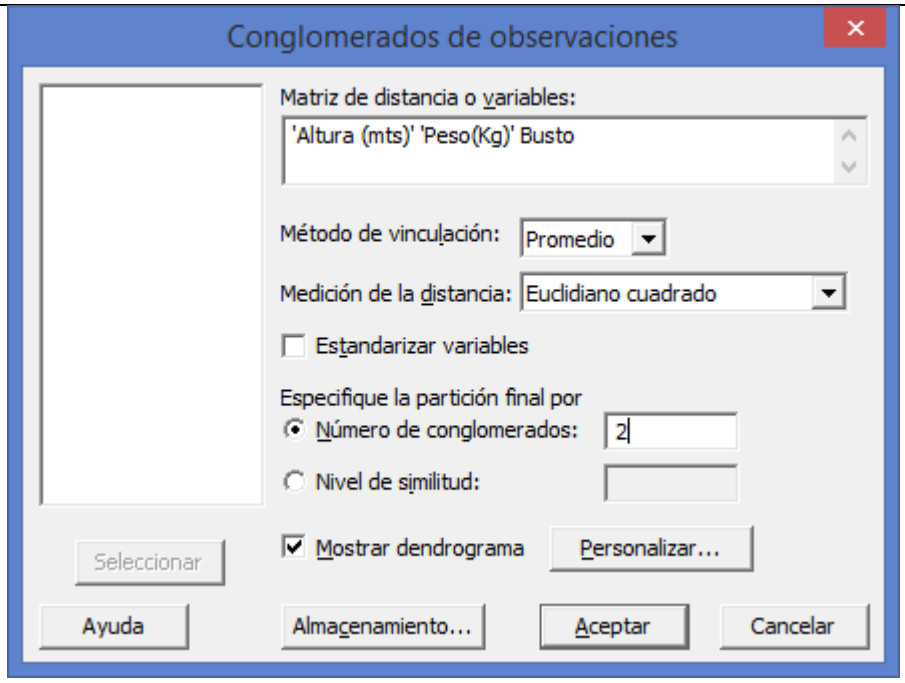
¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados, aunque en la práctica puede ejecutar primero un análisis de conglomerados **sin especificar una partición final con el objeto de decidir dónde colocarla**. Examine los niveles de similitud en los resultados de la ventana Sesión y en el dendograma. Puede ver las similitudes de eslabonamiento en el dendograma colocando el puntero de su ratón sobre una línea horizontal. La similitud de eslabonamiento en cualquier paso es el valor de la similitud mínima en el paso, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Después de elegir dónde desea realizar su partición, vuelva a ejecutar el procedimiento del conglomerado, utilizando, el número de conglomerados para obtener un número establecido de conjuntos de la siguiente manera:

En la parte inicial del análisis del eslabonamiento promedio apareció el siguiente cuadro de Diálogo:



Ahora en ***Especifique la partición final por***, elija ***Número de conglomerados*** e ingrese ***2***.



Haga clic en ***Aceptar*** en el cuadro de diálogo.

Salida de la ventana Sesión

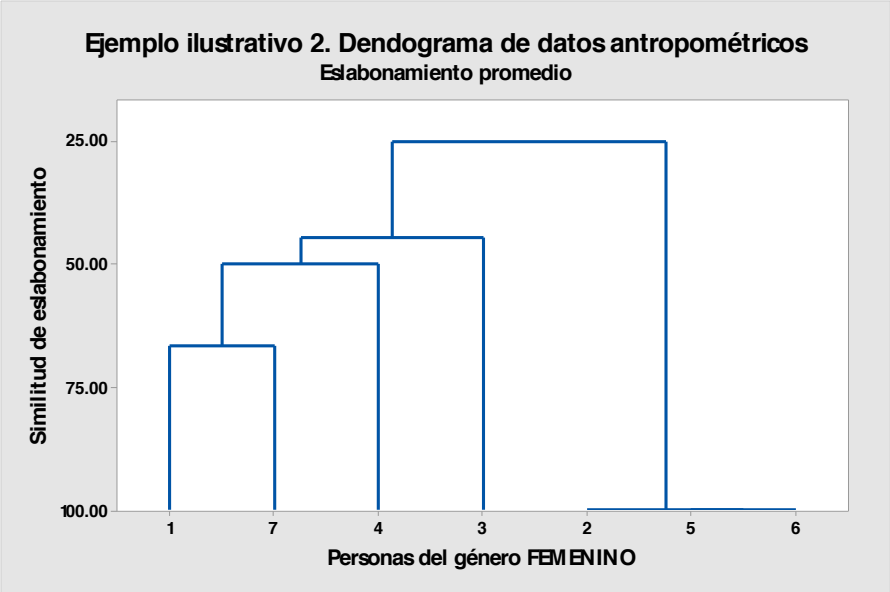
Análisis de observaciones de conglomerado: Altura (mts), Peso(Kg), Busto

```
Distancia eucledіana cuadrada, Enlace de promedios
Pasos de amalgamación


Número de obs. en
          Número de    Nivel de   Nivel de  Conglomerados         Nuevo
el conglomerado
Paso  conglomerados  semejanzal  distancia   incorporados  conglomerado
nuevo
   1              6     100.000    0.00000  5             6             5
2
   2              5     100.000    0.00000  2             5             2
3
   3              4      66.667    1.00000  1             7             1
2
   4              3      50.000    1.50000  1             4             1
3
   5              2      44.444    1.66667  1             3             1
4
   6              1      25.000    2.25000  1             2             1
7
```

Cuando usted especifica la partición o solución final, Minitab muestra entre otras una tabla adicional que resume cada conglomerado según el número de observaciones, en este caso el primer conglomerado está formado por cuatro

mujeres (7,1,4 y 3) y el segundo conglomerado por tres mujeres (6,5 y 2)

Dendrograma



Examine los conglomerados resultantes en la partición o solución final para observar si la agrupación parece lógica (Las personas que forman cada conglomerado aparecen con un color distinto en cada conglomerado). La observación de dendogramas para determinar diferentes agrupaciones finales puede ayudarle a decidir cuáles de ellas tienen más sentido para sus datos.

2.2.1.2 EJERCICIOS DE **REFUERZO**

EJERCICIOS DE REFUERZO 2.2.1.2 ANÁLISIS DE CONGLOMERADOS. MATRIZ DE SIMILITUD

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

2.2.1.2.1 Con el fin de ser más objetivos en el resultado, los alumnos de mercadotecnia agregaron otra forma de evaluar las calificaciones como *acreditado* o no *acreditado*. Recodifique usted los datos y con base en esto, obtenga nuevamente cuál es la pareja mas parecida a través de una medida de similitud.

Persona	Calificación No. 1	Calificación No. 2	Calificación No. 3
1	5.5	4.9	7.3
2	9.6	6.3	6.8
3	8.1	9.8	9.1
4	3.8	7.3	8.2
5	8.6	7.0	8.3

Criterio de similitud para recodificar los datos:

Calificación No. 1: Acreditado (1) si es ≥ a 7.
No Acreditado (0) si es < a 7.

Calificación No. 2: Acreditado (1) si es ≥ a 7.
No acreditado (0) si es < a 7.

Calificación No. 3: Acreditado (1) si es ≥ a 7.
No Acreditado (0) si es < a 7.

a) Obtenga la(s) pareja(s) de personas más parecidas y la(s) menos parecida(s).
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

2.2.1.2.2. Se recabó información de la altura, el peso y el color de ojos de 5 personas. Con estas tres características se desea realizar un estudio de mercado

que permita a los industriales de la confección desarrollar un sistema de patronaje unisex y colores adecuados. Los resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Color ojos
1	1.70	70	Azul
2	1.82	93	Café
3	1.67	83	Negro
4	1.60	60	Café
5	1.90	105	Café

Criterio de similitud para recodificar los datos:

Altura: existe(1) si es ≥ a 1.80 mts.
no existe(0) si es < a 1.80 mts.

Peso: existe(1) si es ≥ a 71 Kgs.
no existe(0) si es < a 71 Kgs.

Cadera: existe(1) si es Café
no existe (0) si es cualquier otro color

a) Obtenga la pareja de personas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

2.2.1.2.3. Se recabó información de cinco familias que solicitaron un préstamo bancario. La Institución les aplicó un cuestionario para conocer si contaban con

casa propia, auto propio y cuanto eran sus ingresos mensuales. Los resultados se presentan a continuación:

Familia	Casa	Auto	Ingreso Mes ($)
1	No	No	9,000
2	Si	Si	15,000
3	No	Si	5,000
4	No	No	8,000
5	Si	Si	17,000

Criterio de similitud para recodificar los datos:

Casa: Presencia (1) si la respuesta es (Sí)
Ausencia (0) Si la respuesta es (No)

Auto: Presencia (1) si La respuesta es (Sí).
Ausencia (0) si la respuesta es (No).

Ingreso: Existe (1) si es ≥ a $ 10,000.
No Existe (0) si es < a $ 10,000.

a) Obtenga la pareja de familias más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

AUTOEVALUACIÓN CON REACTIVOS DE FALSO Ó VERDADERO

EN CADA UNO DE LOS REACTIVOS, CONTESTE CON UNA F SI CONSIDERA QUE LA AFIRMACIÓN ES FALSA Y CON UNA V SI CONSIDERA QUE LA AFIRMACIÓN ES VERDADERA.

1. Uno de los objetivos principales del Factor Análisis es identificar nuevas variables significativas subyacentes. **( )**

2. En el análisis de factores se analiza las relaciones entre un gran número de variables y trata de explicarlas por medio de un concepto (atributos, características o dimensiones) que las agrupe. **( )**

3. Un eigenvalor es la variación explicada por la solución de factores para cada variable. **( )**

4. Cada carga del factor indica la variación que cada uno de los factores explica del conjunto de variables. **( )**

5. Uno de los criterios para obtener factores es el ortogonal, el cual extrae factores con cierto grado de relación entre ellos. **( )**

6. Si existe alguna variable que no tenga carga significativa en algún factor se puede eliminar esa variable de la base de datos y volver a correr de nuevo la solución. **( )**

7. En análisis de factores cuando las escalas en que se miden las variables son distintas se utiliza la matriz de correlaciones. **( )**

8. Una medida de distancia es una medida no negativa **( )**

9. Si la similaridad entre dos objetos es CERO, esto significa que son diferentes **( )**

10. El enlace simple calcula la distancia entre los objetos más próximos **( )**

11. El centroide de un conglomerado es el vector de las medias de las variables **( )**

12. El dendrograma es la representación gráfica de una clasificación jerárquica **( )**

13. Entre más similares son dos objetos, su distancia euclideana es más pequeña **( )**

14. La matriz de correlaciones es la inversa de la matriz de varianzas-covarianzas **( )**

15. Para un Análisis de Factores es necesario que las variables estén muy correlacionadas **( )**

16. La finalidad del análisis de factores es reducir un conjunto de objetos en grupos homogéneos. **( )**

17. Los eigenvalores son las raíces latentes de la matriz de varianzas-covarianzas **( )**

18. Un objetivo del análisis de componentes principales es reducir la dimensionalidad de los datos originales **( )**

19. El análisis de factores Q. describe al mercado no a las variables **( )**

20. El análisis de factores Q. describe a las variables no al mercado **( )**

21. El análisis de factores R. describe a las variables no al mercado **( )**

22. La carga de un factor es la relación que llegan a tener cada una de las personas u

objetos con cada uno de los factores. **( )**

23. Para calcular la comunalidad, cada carga se eleva al cuadrado y se suma para cada factor **( )**

24. Para calcular el Eigenvalor , cada carga se eleva al cuadrado y se suma para cada factor **( )**

25. Para calcular la comunalidad, cada carga se eleva al cuadrado y se suma para cada variable **( )**

26. Para calcular el Eigenvalor , cada carga se eleva al cuadrado y se suma para cada variable **( )**

27. El modelo de Componentes principales asume que la variación específica es tan pequeña que pudiéramos considerarla como cero y por lo tanto la variación total se considera igual a la variación común **( )**

28. El modelo de Factor común asume que la variación específica es tan pequeña que pudiéramos considerarla como cero y por lo tanto la variación total se considera igual a la variación común **( )**

29. El modelo de Factor común asume que la variación específica es importante y que por lo tanto se deberá eliminar del modelo de variación y con el resto intentar agrupar las variables **( )**

30. El modelo de Componentes principales asume que la variación específica es importante y que por lo tanto se deberá eliminar del modelo de variación y con el resto intentar agrupar las variables **( )**

31. En el criterio Ortogonal se va a extraer características que sean independientes, es decir, que tienen una relación igual a cero. **( )**

32. En el criterio Oblicuo se va a extraer características que sean independientes, es decir, que tienen una relación igual a cero **( )**

33. En el criterio Oblicuo se extraen factores con cierto grado de relación entre ellos. **( )**

34. En el criterio Ortogonal se extraen factores con cierto grado de relación entre ellos. **( )**

35. En el criterio denominado de la raíz latente, se considera que un factor debe ser retenido en la solución si su Eigenvalor es mayor a 1. **( )**

36. En el criterio denominado de porcentaje de variación explicada acumulada , se considera que un factor debe ser retenido en la solución si su Eigenvalor es mayor a 1. **( )**

37. El criterio del porcentaje de variación explicada acumulada establece que se deben considerar n factores como solución inicial, si el porcentaje de variación explicada acumulada es mayor al 60 % **( )**

38. El criterio del porcentaje de variación explicada acumulada establece que se deben considerar n factores como solución inicial, si el porcentaje de variación explicada acumulada es mayor al 95 % **( )**

39. El criterio del porcentaje de variación explicada acumulada establece que se deben considerar n factores como solución inicial, si el porcentaje de variación explicada acumulada es mayor al 95 % **( )**

40. El criterio del porcentaje de variación explicada acumulada establece que se deben considerar n factores como solución inicial, si el porcentaje de variación explicada acumulada es mayor al 50 % **( )**

41. En la matriz rotada casi todas las cargas resultan significativas por eso no se agrupan a las variables en este paso, solamente se determina el *número de factores* en los que quedarán agrupadas las variables de manera tentativa **( )**

42. En la matriz no rotada casi todas las cargas resultan significativas por eso no se agrupan a las variables en este paso, solamente se determina el *número de factores* en los que quedarán agrupadas las variables de manera tentativa. **( )**

43. La matriz rotada permite definir no solo el número de factores sino la agrupación de las variables en dichos factores, procurando obtener así una solución final lógica. **( )**

44. La matriz no rotada permite definir no solo el número de factores sino la agrupación de las variables en dichos factores, procurando obtener así una solución final lógica. **( )**

45. En la matriz rotada se ajustan los ejes coordenados o ejes factor en la dirección en que giran las manecillas del reloj con el fin de obtener una solución mas sencilla y con mayor significado teórico **( )**

46. En la matriz no rotada se ajustan los ejes coordenados o ejes factor en la dirección en que giran las manecillas del reloj con el fin de obtener una solución mas sencilla y con mayor significado teórico **( )**

47. La prioridad 2 se usa para una carga significativa de $\pm$ 0.5 para arriba. **( )**

48. La prioridad 1 se usa para una carga significativa de $\pm$ 0.5 para arriba. **( )**

49. La prioridad 3 se usa para una carga significativa de $\pm$ 0.5 para arriba. **( )**

50. La prioridad 4 se usa para una carga significativa de $\pm$ 0.5 para arriba. **( )**

51. La prioridad 2 se usa para una carga significativa entre $\pm$ 0.4 a $\pm$ 0.5 **( )**

52. La prioridad 1 se usa para una carga significativa entre $\pm$ 0.4 a $\pm$ 0.5 **( )**

53. La prioridad 3 se usa para una carga significativa entre $\pm$ 0.4 a $\pm$ 0.5 **( )**

54. La prioridad 4 se usa para una carga significativa entre $\pm$ 0.4 a $\pm$ 0.5 **( )**

55. La prioridad 3 se usa para una carga significativa entre $\pm$ 0.3 a $\pm$ 0.4 **( )**

56. La prioridad 1 se usa para una carga significativa entre $\pm$ 0.3 a $\pm$ 0.4 **( )**

57. La prioridad 2 se usa para una carga significativa entre $\pm$ 0.3 a $\pm$ 0.4 **( )**

58. La prioridad 4 se usa para una carga significativa entre $\pm$ 0.3 a $\pm$ 0.4 **( )**

59. La prioridad 4 se usa para una carga significativa entre $\pm$ 0.2 a $\pm$ 0.3 **( )**

60. La prioridad 1 se usa para una carga significativa entre $\pm$ 0.2 a $\pm$ 0.3 **( )**

61. La prioridad 2 se usa para una carga significativa entre $\pm$ 0.2 a $\pm$ 0.3 **( )**

62. La prioridad 3 se usa para una carga significativa entre $\pm$ 0.2 a $\pm$ 0.3 **( )**

63. En la medida de distancia se tendría como atención central agrupar aquellas personas u objetos con menores distancias entre si (lo mas parecido) **( )**

64. En la medida de similitud se tendría como atención central agrupar aquellas personas u objetos con menores distancias entre si (lo mas parecido) **( )**

65. La medida de similitud se utiliza generalmente cuando se manejan características o variables cualitativas, indicando con un 1 la presencia del atributo y con 0 la ausencia del atributo **( )**

66. La medida de distancia se utiliza generalmente cuando se manejan características o variables cualitativas, indicando con un 1 la presencia del atributo y con 0 la ausencia del atributo **( )**

67. En el método jerárquico aglomerativo se supone que cada uno de los elementos

que se pretende agrupar constituye un grupo o conglomerado por lo tanto existen tantos grupos como personas u objetos a ser agrupados. **( )**

68. En el método jerárquico divisional se supone que cada uno de los elementos que se pretende agrupar constituye un grupo o conglomerado por lo tanto existen tantos grupos como personas u objetos a ser agrupados. **( )**

69. En el método jerárquico divisional supone que existe un solo grupo o conglomerado conteniendo a todas las personas u objetos y con base en diferencias los va dividiendo hasta terminar formando tantos grupos como objetos manejados en la técnica. **( )**

70. En el método jerárquico aglomerativo se supone que existe un solo grupo o conglomerado conteniendo a todas las personas u objetos y con base en diferencias los va dividiendo hasta terminar formando tantos grupos como objetos manejados en la técnica. **( )**

71. En el método jerárquico aglomerativo, el procedimiento consiste en ir agrupando los elementos parecidos por parejas y evaluar lo parecido que llega a ser una pareja o grupo con los restantes elementos que no se han fusionado **( )**

72. En el método jerárquico divisional, el procedimiento consiste en ir agrupando los elementos parecidos por parejas y evaluar lo parecido que llega a ser una pareja o grupo con los restantes elementos que no se han fusionado **( )**

73. El eslabonamiento simple asume que al momento de fusionarse dos personas u objetos, prevalecen las características de aquel elemento cuya distancia sea la mínima o, bien cuya similitud sea la máxima (mas parecido). **( )**

74. El eslabonamiento completo asume que al momento de fusionarse dos personas u objetos, prevalecen las características de aquel elemento cuya distancia sea la mínima o, bien cuya similitud sea la máxima (mas parecido). **( )**

75. El eslabonamiento promedio asume que al momento de fusionarse dos personas u objetos, prevalecen las características de aquel elemento cuya distancia sea la mínima o, bien cuya similitud sea la máxima (mas parecido). **( )**

76. El eslabonamiento completo considera que cuando dos elementos se unen, el grado de parecido de ese grupo a un tercer elemento va a estar dado por la máxima distancia o la mínima de las similitudes (menos parecido). **( )**

77. El eslabonamiento simple considera que cuando dos elementos se unen, el grado de parecido de ese grupo a un tercer elemento va a estar dado por la máxima distancia o la mínima de las similitudes (menos parecido). **( )**

78. El eslabonamiento promedio considera que cuando dos elementos se unen, el grado de parecido de ese grupo a un tercer elemento va a estar dado por la máxima distancia o la mínima de las similitudes (menos parecido). **( )**

79. El eslabonamiento promedio asume que al unirse dos elementos y formar un grupo, las características de ellos con respecto a un tercero estarán dadas por el promedio de las que originalmente poseían, lo cual puede representarse de la siguiente manera. **( )**

80. El eslabonamiento simple asume que al unirse dos elementos y formar un grupo, las características de ellos con respecto a un tercero estarán dadas por el promedio de las que originalmente poseían, lo cual puede representarse de la siguiente **( )**

manera.

81. El eslabonamiento completo asume que al unirse dos elementos y formar un grupo, las características de ellos con respecto a un tercero estarán dadas por el promedio de las que originalmente poseían, lo cual puede representarse de la siguiente manera. **( )**

AUTOEVALUACIÓN CON REACTIVOS PARA COMPLETAR EL ENUNCIADO

EN CADA UNO DE LOS REACTIVOS, ESCRIBA LA O LAS PALABRAS NECESARIAS PARA COMPLETAR EL ENUNCIADO CORRESPONDIENTE:

1. El análisis de_____se considere una técnica estructural cuyo objetivo básico es el de resumir información donde sus variables son cuantitativas ò métricas y la mayoría de ellas deben estar medidas en escala de intervalo ò de razón.
2. El análisis de _____analiza las relaciones entre un gran número de variables y trata de explicarlas por medio de un concepto (atributos, características o dimensiones) que las agrupe.
3. El análisis de _____condensa la información de las variables originales mediante características comunes a dichas variables y que las representan lógicamente.
4. Encontrar características que describan a núcleos poblacionales (personas) es uno de los objetivos del Factor análisis. Si se persigue este objetivo se llama_____.
5. Identificar un conjunto de dimensiones o características que se encuentran latentes (las cuales no se detectan fácilmente) dentro de un conjunto grande de variables es uno de los objetivos de Factor análisis. Cuando se está interesado en este objetivo se le llama _____.
6. Desde el punto de vista matemático un _____es una combinación lineal de variables originales de la investigación.
7. La _____ de un factor es la relación que llegan a tener cada una de las variables con cada uno de los factores. Es una medida de correlación entre una variable y un factor.
8. El _____es la variación que cada uno de los factores logra explicar del conjunto de variables.
9. Para calcular el _____, cada carga se eleva al cuadrado y se suma para cada factor.
10. Para calcular la _____, cada carga se eleva al cuadrado y se suma para cada variable.
11. El porcentaje de _____ _____representa la relación entre la variación explicada por el factor, con respecto a la variación máxima, la cual es igual al numero de variables, ya que cada una de ellas tiene una variación de uno por estar estandarizadas; por lo tanto, la mayor variación que se puede llegar a explicar sería la suma de las variaciones de cada variable.
12. La _____es la variación explicada por la solución de factores para cada variable y por estar estandarizadas, el valor máximo posible de la comunalidad es de 1.

13. El modelo de _____ _____asume que la variación específica es tan pequeña que pudiéramos considerarla como cero y por lo tanto la variación total se considera igual a la variación común. Los paquetes computacionales si no se les especifica algún otro manejan este modelo.
14. El modelo de _____ _____ asume que la variación específica es importante y que por lo tanto se deberá eliminar del modelo de variación y con el resto intentar agrupar las variables. En el criterio _____se va a extraer características que sean independientes, es decir, que tienen una relación igual a cero.
15. En el criterio _____ se extraen factores con cierto grado de relación entre ellos.
16. En el criterio denominado de la _____ _____, se considera que un factor debe ser retenido en la solución si su Eigenvalor es mayor a 1.
17. El criterio del porcentaje de _____ _____ _____ establece que se deben considerar n factores como solución inicial, si el porcentaje de variación explicada acumulada es mayor al 60 %.
18. En la matriz no _____ casi todas las cargas resultan significativas por eso no se agrupan a las variables en este paso, solamente se determina el *número de factores* en los que quedarán agrupadas las variables de manera tentativa.
19. En la matriz _____ se ajustan los ejes coordenados o ejes factor en la dirección en que giran las manecillas del reloj con el fin de obtener una solución mas sencilla y con mayor significado teórico.
20. En una carga significativa de $\pm$ 0.5 para arriba se usa la prioridad _____.
21. En una carga sólo entre $\pm$ 0.4 a $\pm$ 0.5 se usa la prioridad _____.
22. En una carga sólo entre $\pm$ 0.3 a $\pm$ 0.4 se usa la prioridad _____.
23. En una carga sólo entre $\pm$ 0.2 a $\pm$ 0.4 se usa la prioridad _____.
24. El análisis de _____ se utiliza básicamente para la segmentación mercados en donde lo que se busca es formar grupos de objetos o personas lo mas homogéneos posible, al comparar los elementos que conforman un grupo, y lo mas heterogéneo, al comparar un elemento de un grupo con otro.
25. La medida de _____ se utiliza generalmente cuando se manejan características o variables cualitativas, indicando con un 1 la presencia del atributo y con 0 la ausencia del atributo.
26. En la medida de _____se tendría como atención central agrupar aquellas personas u objetos con menores distancias entre si (lo mas parecido).
27. Para realizar el análisis de agrupamiento los datos de las variables deben estar _____, si son métricas, es decir, eliminar sus unidades de manera que se encuentren todas en términos de *z.*
28. El algoritmo que se habrá de seguir para formar grupos de personas u objetos similares entre si se le conoce como método de _____ _____**.**
29. En el método _____ _____ se supone que cada uno de los elementos que se pretende agrupar constituye un grupo o conglomerado por lo tanto existen tantos grupos como personas u objetos a ser agrupados.
30. En el método _____ _____ se supone que existe un solo grupo o conglomerado conteniendo a todas las personas u objetos y con base en diferencias los va dividiendo hasta terminar formando tantos grupos como objetos manejados en la técnica.
31. En el método _____ _____, el procedimiento consiste en ir agrupando los elementos parecidos por parejas y evaluar lo parecido que llega a ser una pareja o grupo con los restantes elementos que no se han fusionado.

32. Un _____ consiste en volver a establecer el grado de parecido de un nuevo grupo con el resto de los elementos que no habìan sido agrupados y existen básicamente tres formas o criterios para medir ese grado de parecido.
33. El eslabonamiento _____asume que al momento de fusionarse dos personas u objetos, prevalecen las características de aquel elemento cuya distancia sea la mínima o, bien cuya similitud sea la máxima (mas parecido).
34. El eslabonamiento _____ considera que cuando dos elementos se unen, el grado de parecido de ese grupo a un tercer elemento va a estar dado por la máxima distancia o la mínima de las similitudes (menos parecido).
35. El eslabonamiento _____ asume que al unirse dos elementos y formar un grupo, las características de ellos con respecto a un tercero estarán dadas por el promedio de las que originalmente poseían, lo cual puede representarse de la siguiente manera.
36. El _____es la representación grafica de los grupos o conglomerados formados donde en el eje horizontal se establecen los elementos que se están agrupando, de preferencia en el orden en que se fueron agrupando y en el eje vertical se manejan las medidas de distancia (o similitud) a las que se fueron uniendo los grupos.

CAPÍTULO 3. MÉTODOS DE CLASIFICACIÓN

OBJET.33IVO 3.1 El alumno validará la función discriminante para 2 ó más grupos

ANTECEDENTES

CONCEPTOS DE:

Ecuación lineal, sistema de ecuaciones lineales, Matriz aumentada, representación matricial, operaciones elementales, ecuación algebraica, tamaño de una matriz, tipo de matrices, igualdad de matrices, operaciones con matrices, matriz de cofactores, matriz inversa, determinantes.

3.1.1 EL ANÁLISIS DISCRIMINANTE

CONCEPTOS **BÁSICOS**

ANÁLISIS **DISCRIMINANTE**

A una compañía que se especializa en tarjetas de crédito evidentemente le gustaría poder clasificar a los solicitantes de estas tarjetas en dos grupos: 1) los que son buenos riesgos de crédito, y 2) los que son malos riesgos de crédito. Entonces, a quienes considere como buenos riesgos de crédito se les ofrecerían las tarjetas, mientras que a los que se consideran como malos riesgos no se les haría el ofrecimiento. Para obtener ayuda para hacer esta determinación, la compañía podrá considerar varias características demográficas medibles de cada individuo. Por ejemplo, puede considerar el nivel educativo, el salario, los adeudos y la historia crediticia como posibles predicciones de los méritos crediticios. Entonces, la compañía intentaría usar esta información para decidir si debe aprobarse un solicitante de una tarjeta de crédito. El método multivariado que ayudaría a clasificar a los solicitantes en uno de los dos grupos de riesgo de crédito se llama análisis discriminante.

El análisis discriminante (DA: *discriminante análisis*) se usa principalmente para clasificar individuos o unidades experimentales en dos o más poblaciones definidas de manera única. Para desarrollar una regla de discriminante que clasifique las unidades experimentales en una de varias categorías posibles, el investigador debe tener una muestra aleatoria de unidades experimentales de cada grupo posible de clasificación. Entonces, el DA proporciona los métodos que permitirán a los investigadores establecer reglas que se puedan emplear para clasificar otras unidades experimentales en uno de los grupos de clasificación.

En el ejemplo de las tarjetas de crédito, se elabora una regla discriminante al usar datos demográficos proveniente de individuos de los que se sabe constituyen buenos riesgos de crédito, como también datos similares de aquellos que se sabe son malos riesgos. Entonces, los nuevos solicitantes de las tarjetas se clasifican en uno de los dos grupos de riesgo usando la regla resultante.

El análisis discriminante pertenece al grupo de las técnicas funcionales, es decir, su principal objetivo es realizar un pronóstico.

El análisis discriminante busca distinguir y clasificar las observaciones de una investigación. Explica el porquè de esas diferencias.

Otro objetivo que se persigue con el análisis discriminante es pronosticar a cual grupo pertenecerán personas de acuerdo con sus características.

Esta técnica maneja *n* variables independientes, todas ellas métricas y una sola variable dependiente no métrica nominal (categórica), que representara a los grupos en que se dividió el mercado.

El discriminante genera una ecuación que permitirá realizar pronósticos y que determinarà si realmente se ha segmentado el mercado. Dicha ecuación està representada por la "Función Discriminante Lineal de Fisher"

Una observación se clasifica en un grupo si la distancia cuadrada (también llamada distancia de Mahalanobis) de observación hasta el centro del grupo (media) es el mínimo. Se parte del supuesto de que las matrices de covarianzas son iguales para todos los grupos. Existe una parte única de la fórmula de la distancia cuadrada para cada grupo y se denomina función discriminante lineal para ese grupo. Para cualquier observación, el grupo con la distancia cuadrada más pequeña tiene la función discriminante lineal más grande y la observación se clasifica entonces en este grupo.

El análisis discriminante lineal tiene la propiedad de la distancia cuadrada simétrica: la función discriminante lineal del grupo i evaluado con la media del grupo j es igual a la función discriminante lineal del grupo j evaluado con la media del grupo i.

Hemos descrito los casos más simples, no las matrices de las covarianzas iguales y a priori. Si se considera que la distancia de Mahalanobis

es una manera razonable de medir la distancia de una observación a un grupo, entonces, no se necesita crear suposiciones de la distribución subyacente de sus datos.

$$Z= K_1 X_1 + K_2 X_2 + K_3 X_3 +...+ K_n X_n .$$

Donde:

Z:representa el puntaje discriminante o puntaje discriminatorio (número que servirá para efectuar la clasificación).

K: representa el factor de ponderación o de importancia que tiene cada una de las variables para discriminar.

X: cada una de las variables independientes.

El análisis discriminante va a manejar variables significativas en la diferenciación de los grupos; esto es , va a tratar de encontrar los pesos de las variables que cumplan con lo anterior.

En el análisis discriminante se buscará como criterio separar lo mas posible (diferenciar) a los promedios de cada grupo, pero con respecto a la dispersión de los datos, de manera que se aumente esta relación.

SUPUESTOS DEL ANALISIS DISCRIMINANTE:

1.- Los datos deberán estar distribuidos normalmente, es decir, que la mayoría de ellos se encuentren concentrados alrededor de su promedio.

2.- La dispersión de los datos (las varianzas y covarianzas) deberán ser iguales entre los grupos, lo cual significa que el experimento se realice en las mismas circunstancias y no debe haber diferentes fuentes de variación que haga que los grupos sean diferentes.

3.- Al llevar a cabo una clasificación, se supone a priori, que el resultado obtenido es tan confiable como lo indica el porcentaje de clasificaciones correctas obtenidas en la etapa de validación del modelo.

El análisis discriminante implica las siguientes etapas:

1.- Derivación de la ecuación para obtener el valor de las constantes que formaràn al modelo discriminante.

2.- La validación del modelo (se verifica el grado de error que tendría el modelo para efectuar pronósticos.

3.- La interpretación de los resultados del modelo (de aquí se obtendrán las estrategias concretas para el problema que se haya planteado).

Se utiliza Análisis discriminante para clasificar las observaciones en dos o más grupos si se tiene una muestra con grupos conocidos. El Análisis discriminante también se puede utilizar para investigar en que manera las variables contribuyen a la separación de grupos.

CONSIDERACIONES ANTES DE DERIVAR LA ECUACION DISCRIMINANTE

1.- Las categorías o grupos de la variable independiente deben ser 2, 3, máximo 4, pero no más de 4.

2.- Asegurarse de haber hecho una lista exhaustiva de todos los posibles grupos de la variable dependiente.

3.- Los grupos que forman a la variable dependiente deben ser mutuamente excluyentes.

4.- En cuanto a la muestra, esta deberá partirse en dos y utilizar para la etapa de derivación de 60 a 70 % de los datos, y el restante 40 o 30% se debe dejar para la etapa de validación del modelo, pues no es nada confiable el grado de error al pronosticar con los mismos datos que se derivaron de la ecuación que si se usaron datos diferentes que en la derivación.

Al dividir la muestra deben conservarse las proporciones de la muestra original en las partes, tanto para derivación como para validación.

TIPOS DE ANALISIS DISCRIMINANTE

De acuerdo con el número de grupos que conforman a la variable dependiente los modelos discriminantes pueden clasificarse como simples o múltiples.

La técnica de discriminante será simple, si la variable dependiente está formada por dos grupos y se considerará múltiple cuando la variable dependiente tiene más de dos grupos.

MODELO DISCRIMINANTE SIMPLE

La derivación de la ecuación del discriminante simple con **dos variables** independientes se ve representada con la siguiente expresión:

$$Z = K_1 X_1 + K_2 X_2$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las K. Hay que recordar que se tiene como criterio el aumentar las distancias de los promedios de cada grupo, considerando la dispersión de los datos.

Para llevar a cabo la derivación del modelo de discriminante simple y encontrar las *K*, se deben resolver las siguientes ecuaciones normales:

$$\sum X_1^{'2} \text{ K}_1 + \sum X_1' X_2' \text{ K}_2 = X_1 \text{ | Gpo. 1 - } X_1 \text{ | Gpo. 2}$$

$$\sum X_1' X_2' \text{ K}_1 + \sum X_2^{'2} \text{ K}_2 = X_2 \text{ | Gpo. 1 - } X_2 \text{ | Gpo. 2}$$

Donde:

$$\sum X_1^{'2} = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2^{'2} = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1 \bar{X}_2)$$

En donde $\sum X_i^{'2}$ representa la varianza de los datos de la variable en un grupo y otro, y la $\sum X_i' X_j'$ representa la covarianza de los datos de las variables en uno y otro grupo. Debe entenderse por varianza la dispersión de los datos con respecto a su media, mientras la covarianza es la dispersión de los datos, tomando en cuenta las dos variables y como si esta fuera una.

Para la derivación de la ecuación del discriminante simple con **tres variables** independientes se ve representada con la siguiente expresión:

$$Z = K_1 X_1 + K_2 X_2 + K_3 X_3$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las K. Hay que recordar que se tiene como criterio el aumentar las distancias de los promedios de cada grupo, considerando la dispersión de los datos.

Para llevar a cabo la derivación del modelo de discriminante simple y encontrar las *K*, se deben resolver las siguientes ecuaciones normales:

$$\sum X_1'^2 K_1 + \sum X_1' X_2' K_2 + \sum X_1' X_3' K_3 = \bar{X}_1 \,|Grupo\ 1 - \bar{X}_1|Grupo\ 2$$

$$\sum X_1' X_2' K_1 + \sum X_2'^2 K_2 + \sum X_2' X_3' K_3 = \bar{X}_2 \,|Grupo\ 1 - \bar{X}_2|Grupo\ 2$$

$$\sum X_1' X_3' K_1 + \sum X_2' X_3' K_2 + \sum X_3'^2 K_3 = \bar{X}_3 \,|Grupo\ 1 - \bar{X}_3|Grupo\ 2$$

Donde:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\,\bar{X}_2)$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\,\bar{X}_3)$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\,\bar{X}_3)$$

En donde $\sum X_i'^2$ representa la varianza de los datos de la variable en un grupo y otro, y la $\sum X_i' X_j'$ representa la covarianza de los datos de las variables en uno y otro grupo. Debe entenderse por varianza la dispersión de los datos con respecto a su media, mientras la covarianza es la dispersión de los datos, tomando en cuenta las dos variables y como si esta fuera una.

Este mismo modelo puede generalizarse para cuando se manejan más variables independientes en el modelo, como se presenta a continuación.

$$Z=K_1X_1+K_2X_2+K_3X_3+...+K_nX_n$$

$\sum X'^2_1 K_1 + \sum X'_1 X'_2 K_2 + \sum X'_1 X'_3 K_3 + \dots \sum X'_1 X'_n K_n = \overline{X}_1$ | Gpo. 1 - $\overline{X}_1$ | Gpo. 2

$\sum X'_1 X'_2 K_1 + \sum X'^2_2 K_2 + \sum X'_2 X'_3 K_3 + \dots \sum X'_2 X'_n K_n = \overline{X}_2$ | Gpo. 1 - $\overline{X}_2$ | Gpo. 2

$\sum X'_1 X'_3 K_1 + \sum X'_2 X'_3 K_2 + \sum X'^2_3 K_3 + \dots \sum X'_3 X'_n K_n = \overline{X}_3$ | Gpo. 1 - $\overline{X}_3$ | Gpo. 2

. . . . . .

. . . . . .

$\sum X'_1 X'_n K_1 + \sum X'_2 X'_n K_2 + \sum X'_3 X'_n K_3 + \dots \sum X'^2_n K_n = \overline{X}_n$ | Gpo. 1 - $\overline{X}_n$ | Gpo. 2

La diagonal principal de este conjunto de ecuaciones es la varianza y el resto es la covarianza.

Los promedios de cada grupo de las variables independientes se conoce con el nombre de centroides.

Mediante la resolución de estos sistemas de ecuaciones se obtendrá como resultado la ecuación discriminante. Una vez hecho esto, se pasa a una etapa previa de validación del modelo, por medio de la cual se prueba al modelo.

PRUEBA DE HIPÓTESIS PARA EL MODELO DISCRIMINANTE

Esta prueba de hipótesis tiene como objetivo probar desde el punto de vista estadístico, si la diferencia en los centroides o la diferencia que presentan los promedios de las variables involucradas en el modelo son significativas. Si los son, quiere decir que el modelo es bueno para discriminar.

Para llevar a cabo esta prueba, se usa la prueba Mahalanobis o D^2 , mediante la cual se podrá obtener el valor de la F-calculada y compararla contra la F- critica de tablas y realizar la prueba de hipótesis.

En esta prueba, la hipótesis nula ò H_0 establece que el modelo discriminante **no** es bueno para discriminar o clasificar observaciones, mientras que la hipótesis alternativa ò H_1 establece que el modelo discriminante **es** bueno.

La distancia D^2 se obtiene por medio del siguiente cálculo:

$$D^2 = (n_1 + n_2 - 2) \left[\sum K_i (\overline{X}_i \big|_1 - \overline{X}_i \big|_2 \right]$$

La D^2 es un valor absoluto (positivo), ya que evalúa la distancia entre los promedios. La D^2 se compara con la F- crítica para tomar la decisión de rechazar o no la H_0 , para lo cual se debe transformar la D^2 en F-calculada mediante la siguiente formula:

$$F_{calculada} = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2) D^2}{\left[(n_1 + n_2 - 2) p\right] * (n_1 + n_2)}$$

Esta F-calculada se compara contra la F-crìtica de tablas con los siguientes grados de libertad:

$$\text{F-crìtica} [\alpha , p , (n_1 + n_2 - p - 1)]$$

Donde:

n_1 : número de datos manejados en el grupo 1.

n_2 : número de datos manejados en el grupo 2.

p: número de grupos manejados en la variable dependiente.

Cuando se ha concluido que el modelo es bueno para hacer clasificaciones, se procede a realizar la validación del mismo. Esta etapa consiste en determinar la magnitud del error o la confiabilidad que se tendrá al realizar pronósticos con el modelo. Si en la prueba anterior se concluyera que el modelo no es bueno para discriminar, entonces no tiene caso hacer la validación.

ETAPA DE VALIDACIÓN DEL MODELO DISCRIMINANTE

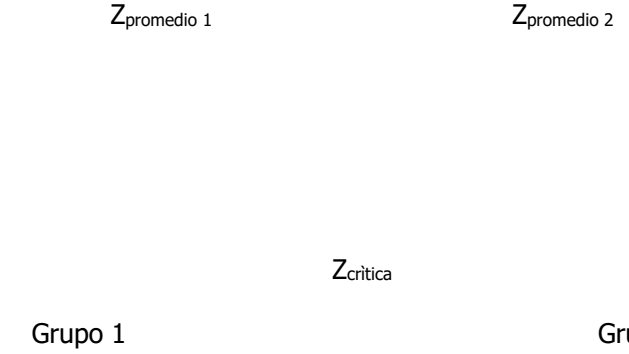
La etapa de validación tiene como objetivo obtener un indicador que evalúe lo acertado que es el modelo para efectuar clasificaciones, es decir, como medio para obtener una medida de la confiabilidad al pronosticar con el mismo. A dicha medida se le conoce como ***porcentaje de clasificaciones correctas*** .

En el análisis discriminante lo que se busca es determinar las variables que aumenten el valor relativo de distancia de los promedios, con respecto a la dispersión de los datos, o mejor dicho, aquellas variables que diferencien lo mejor posible los promedios de los grupos.

Las proyecciones de los datos se pueden reflejar en el eje de las Z y tienen una forma o distribución normal. Aquí se observan las opiniones de los grupos que forman la variable dependiente. Si esta curva normal se viera como sombra, se esperaría una mayor concentración de datos alrededor del centroide o promedio.

En la curva normal suelen traslaparse los resultados de los grupos, por lo que es necesario establecer una línea divisoria entre dichos resultados de los grupos. Al valor de corte que marca el punto de la línea divisoria se le llama : **Z crítica**. Lo deseado es que las colas de las curvas de cada grupo se crucen lo menos posible y así lograr una separación entre los promedios de los grupos.

El área de traslape entre ambas curvas representa la probabilidad de error al pronosticar, es decir, representan los datos clasificados incorrectamente.



Si el área de error es el traslape entre ambas curvas normales, lo que se va a tratar de hacer es tener un menor error, es decir distanciar más los promedios de los grupos para que se traslapen menos las curvas.

PASOS PARA LA ETAPA DE VALIDACIÓN

1. Se obtiene la **Z *promedio*** para cada grupo, lo cual se logra sustituyendo en la ecuación obtenida los promedios de las variables independientes para cada grupo. Una vez obtenidas las Z *promedio*, se acomodan de izquierda a derecha, del más negativo al más positivo.

2. Se determina la **Z *critica*** con la siguiente expresión:

Si $n_1 = n_2$ fórmula es :

$$Z_{critica} = (\overline{Z}_1 + \overline{Z}_2)/2$$

Si n_1 **es diferente a** n_2 se debe obtener un promedio ponderado y la fòrmula es :

$$Z_{critica} = (n_2\overline{Z}_1 + n_1\overline{Z}_2)/(n_1 + n_2)$$

3. Se obtiene la Z para cada observación y se pronostica en qué grupo cae cada una de las observaciones de acuerdo con la Z *critica.*

4. Se verifica si esta clasificación es correcta o incorrecta.

5. De acuerdo con el punto anterior, se obtiene el porcentaje de clasificaciones correctas, la cual será una medida de la confiabilidad o seguridad al pronosticar en el futuro.

6. Se procede a obtener el porcentaje mínimo de clasificaciones correctas (C), mediante la siguiente fórmula:

$$C = \sum p_i^2 \text{ ó bien } p_i^2 + (1 - p_i)^2$$

Donde:

p_i : porcentaje de datos de un grupo cualquiera (de la derivación, del total de la muestra o de la validación).

7. Por último se compara el porcentaje de clasificaciones correctas contra el porcentaje mínimo aceptable de clasificaciones correctas. Si el primero es menor al segundo, entonces se puede concluir que

la ecuación o el modelo discriminante no es útil para pronosticar observaciones. Si sucede lo contrario, se puede concluir que el modelo es bueno para clasificar observaciones.

ETAPA DE INTERPRETACIÓN DEL ANÁLISIS DISCRIMINANTE

La última etapa de esta técnica es la interpretación de los resultados obtenidos. En esta etapa se evalúa la importancia relativa que posee cada una de las variables independientes para discriminar.

Método de los pesos discriminantes estandarizados

Este método consiste en estandarizar las constantes de la ecuación discriminante, de manera que puedan ser comparadas.

Para jerarquizar las variables con base en su importancia para discriminar, se considerará únicamente la **magnitud del peso, no su signo**. Este método requiere que las variables independientes no presenten problemas de multicolinealidad y suponer que las variables están interactuando.

Para estandarizar las constantes se utilizará la siguiente fórmula:

$$K'_n = (K_n) [\sum X'^2_n / (n_1 + n_2 - 2)]^{1/2}$$

Una vez hecho esto, se puede concluir que mientras mayor sea el peso discriminante, más importante será la variable para discriminar.

3.1.1.1 EJEMPLO ILUSTRATIVO

EJEMPLO ILUSTRATIVO

3.1.1.1

ANÁLISIS DISCRIMINANTE

El **análisis discriminante** es una técnica estadística multivariante cuya finalidad es describir (si existen) las diferencias significativas entre *g* grupos de objetos ($g > 1$) sobre los que se observan *p* variables (variables discriminantes). Más concretamente, se comparan y describen las medias de las p variables clasificadoras a través de los *g* grupos.

En caso de que estas diferencias existan, intentará explicar en qué sentido se dan y proporcionar procedimientos de asignación sistemática de nuevas observaciones con grupo desconocido a uno de los grupos analizados, utilizando para ello sus valores en las *p* variables clasificadoras (estos sí, conocidos).

Podemos ver este procedimiento como un modelo de predicción de una variable respuesta categórica (variable grupo) a partir de *p* variables explicativas

Se está elaborando una investigación para analizar dos de las marcas de laptops de mayor penetración en el mercado. Las pruebas aplicadas abarcan aspectos como: decoloración, resistencia a los golpes y calentamiento interno.

El cuestionario en su parte principal incluía entre otras preguntas las siguientes:

1. Indique el grado de importancia que tiene para usted los siguientes aspectos:

	Nada Imp.(1)				Muy imp.(5)
a) Decoloración al lavado	____	____	____	____	____
b) Que no se agriete tan fácil	____	____	____	____	____
c) Que el color se mantenga firme	____	____	____	____	____
d) Que la cubierta no sufra desgaste	____	____	____	____	____

2. La experiencia personal al usar la laptop ACME en los siguientes aspectos ha sido:

	Nunca.(1)				Siempre.(5)
a) La cubierta mantiene su color	____	____	____	____	____

generalmente continuas (variables clasificatorias). Por poner algún ejemplo, podemos hablar del reconocimiento de formas, de texto, o del diagnóstico automático.

b) La cubierta no se raya con el uso ____ ____ ____ ____ ____

c) Se calienta con el uso prolongado ____ ____ ____ ____ ____

La segmentación de mercados realizada encontró que una parte de la población estaba de acuerdo con los aspectos de decoloración (X_1), resistencia a los golpes (X_2) y calentamiento interno (X_3) y la otra en desacuerdo. Para comprobar lo anterior se tomaron respuestas de diez personas entrevistadas con el fin de validar la segmentación.

Los datos son:

No. De sujeto	**Grupos**	**Decoloración (X_1)**	**Resistencia a los golpes (X_2)**	**Calentamiento interno (X_3)**
1	Acuerdo	4	4	3
2	Acuerdo	3	5	2
3	Acuerdo	5	3	4
4	Acuerdo	4	2	2
5	Acuerdo	2	4	2
	Media Gpo. 1	3.60	3.60	2.60
6	Desacuerdo	3	2	5
7	Desacuerdo	2	4	1
8	Desacuerdo	2	1	3
9	Desacuerdo	1	2	2
10	Desacuerdo	1	1	1
	Media Gpo. 2	1.80	2.00	2.40

a) Obtenga el modelo discriminante que nos ayude a comprobar la verdadera existencia de dos mercados.

b) Pruebe estadísticamente la existencia de los dos segmentos de mercado, es decir si el modelo es bueno para clasificar a las personas que están de acuerdo de los que no lo están.

c) Valide el modelo con los mismos datos de derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados indicando en orden de importancia las variables que logran segmentar y que estrategia se debería seguir.

e) ¿En qué grupo caería una persona que hubiera calificado el nuevo producto de la siguiente manera? : Decoloración: 1; Resistencia a los golpes: 3 y Calentamiento interno: 2

El Análisis discriminante es una técnica inferencial. Es una técnica típicamente multivariante porque suele usarse en contextos donde tenemos varias variables, pero evidentemente puede aplicarse con pocas variables, incluso con una sola variable, pero no es lo habitual.

Una característica esencial de esta técnica es que tenemos previamente definidas dos o más poblaciones; o sea, tenemos dos o más muestras de esas poblaciones con una serie de individuos de cada una de ellas de los que tenemos medidas una serie de variables.

Su finalidad básica es preparar esa información, seleccionarla, trabajarla, con una finalidad clasificadora. Futuros individuos, a los que les podremos medir esas variables, deberemos

Resolución del inciso a)

La ecuación para derivar el modelo discriminante tendrá la siguiente forma:

$$Z = K_1X_1 + K_2X_2 + K_3X_3$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las *K,* para ello se deben resolver las siguientes ecuaciones normales:

$$\sum X_1'^2 K_1 + \sum X_1'X_2'K_2 + \sum X_1' X_3'K_3 = \bar{X}_1\,|Grupo\ 1 - \bar{X}_1|Grupo\ 2$$

$$\sum X_1'X_2'K_1 + \sum X_2'^2 K_2 + \sum X_2' X_3'K_3 = \bar{X}_2\,|Grupo\ 1 - \bar{X}_2|Grupo\ 2$$

$$\sum X_1'X_3'K_1 + \sum X_2'X_3'K_2 + \sum X_3'^2\, K_3 = \bar{X}_3\,|Grupo\ 1 - \bar{X}_3|Grupo\ 2$$

clasificarlos como miembros de alguna de esas poblaciones.

Evidentemente partimos del supuesto de que esos nuevos individuos a clasificar pertenecen a una de esas poblaciones.

El Análisis discriminante tiene un nombre muy apropiado para lo que es su procedimiento. Porque lo que hace es iniciar, a partir de toda la información de que se dispone sobre las poblaciones y las variables, un proceso de discriminación, un proceso de separación lo mayor posible de esas poblaciones.

Por lo tanto, a partir de un conjunto de individuos que sabemos ciertamente a qué población pertenecen cada uno de ellos y a partir de los valores de todas las variables que disponemos mediante el Análisis discriminante tratamos de buscar qué combinaciones de esas variables nos permitirán discriminar lo más posible entre los grupos que tenemos.

Pongamos un ejemplo previo sencillo que puede ayudarnos a clarificar el camino de la explicación de esta técnica: Supongamos que queremos encontrar variables que nos permitan clasificar a una persona entre hombre o mujer teniendo únicamente la información de la medida de esa variable.

La variable edad no discrimina entre hombres y mujeres, no separa bien esos dos grupos. La variable

Donde:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\,\bar{X}_2)$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\,\bar{X}_3)$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\,\bar{X}_3)$$

Se sustituyen los valores correspondientes, obteniendo lo siguiente:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_1'^2 = [(4^2 + 3^2 + 5^2 + 4^2 + 2^2) - 5(3.60)^2] + [(3^2 + 2^2 + 2^2 + 1^2 + 1^2) - 5(1.80)^2] = 8$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_2'^2 = [(4^2 + 5^2 + 3^2 + 2^2 + 4^2) - 5(3.60)^2] + [(2^2 + 4^2 + 1^2 + 2^2 + 1^2) - 5(2.0)^2] = 11.20$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

altura ya discrimina más. La variable pie que calza discrimina más aún.

Elegir la edad para pronosticar el sexo nos llevaría a mucho error. La mitad de las veces nos equivocaríamos. Elegiendo la altura ya cometeríamos menos errores porque las poblaciones están más discriminadas, más separadas, respecto a esta variable. Eligiendo el pie que calza cometeríamos menos errores aún porque de las tres variable es la que discrimina mejor, es la que separa mejor a los dos grupos. Esto es así en la realidad: en mujeres y hombres con la misma altura los hombres tienen el pie más grande que las mujeres.

Si se entiende bien esta idea simple se entenderá perfectamente lo que persigue el Análisis discriminante. Porque esto es lo que hace la técnica: entre las variables de que disponemos y con las muestras que tenemos de las poblaciones en estudio debemos buscar qué variables y qué combinación de ellas es la que consigue separar más, discriminar más, esos grupos. La finalidad es usarlo como mecanismo para clasificar a un individuo futuro del que tendremos los valores que tiene de esas variables pero del que no sabremos a qué población pertenece.

A veces si tenemos más de una variable con una de esas variables nos bastará para conseguir una buena discriminación, pero a veces ninguna de ellas individualmente nos irá bien

$$\sum X_3'^2 = [(3^2 + 2^2 + 4^2 + 2^2 + 2^2) - 5(2.60)^2] + [(5^2 + 1^2 + 3^2 + 2^2 + 1^2) - 5(2.40)^2] = 14.40$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\bar{X}_2)$$

$$=[(4x4 + 3x5 + 5x3 + 4x2 + 2x4) - 5(3.60x3.60)] + [(3x2 + 2x4 + 2x1 + 1x2 + 1x1) - 5(1.80x2)] = -1.80$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\bar{X}_3)$$

$$=[(4x3 + 3x2 + 5x4 + 4x2 + 2x2) - 5(3.60x2.60)] + [(3x5 + 2x1 + 2x3 + 1x2 + 1x1) - 5(1.80x2.40)] = 7.60$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\bar{X}_3)$$

$$=[(4x3 + 5x2 + 3x4 + 2x2 + 4x2) - 5(3.60x2.60)] + [(2x5 + 4x1 + 1x3 + 2x2 + 1x1) - 5(2x2.4)] = -2.80$$

De lo anterior se obtienen tres ecuaciones, normales:

$$8K_1 - 1.80K_2 + 7.60K_3 = 1.80$$

$$-1.80K_1 + 11.2K_2 - 2.80K_3 = 1.60$$

$$7.6K_1 - 2.8K_2 + 14.4K_3 = 0.20$$

A partir de estas ecuaciones se obtienen los valores de *K* de cada variable. En este caso se utilizará la matriz inversa para la resolución del sistema lineal siguiendo los siguientes pasos:

y sí, en cambio, una combinación de ellas.

Paso 1. Expresar el sistema como A *K=D

Donde:

A= Matriz de coeficientes.

K= Coeficientes discriminantes.

D= Términos independientes.

$$\begin{bmatrix} 8 & -1.80 & 7.60 \\ -1.80 & 11.20 & -2.80 \\ 7.60 & -2.80 & 14.40 \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 1.80 \\ 1.60 \\ 0.20 \end{bmatrix}$$

Paso 2. Obtener el determinante de A aplicando la regla de Sarrus para comprobar si el sistema tiene solución, es decir si |A|≠0:

$$|A| = \begin{bmatrix} 8 & -1.80 & 7.60 \\ -1.80 & 11.20 & -2.80 \\ 7.60 & -2.80 & 14.40 \end{bmatrix} \begin{matrix} 8 & -1.80 & 7.60 \\ -1.80 & 11.20 & -2.80 \end{matrix}$$

$$(8x11.20x14.40) + (-1.80x - 2.80x7.60) + (7.60x - 1.80x - 2.80) - (7.60x11.20x7.60) - (-2.80x2.80x8) - (14.40x - 1.80x - 1.80) = 610.56$$

Como 610.56 es ≠ de 0 el sistema tiene solución.

Paso 3. Obtener la matriz transpuesta de A, que es:

$$A' = \begin{bmatrix} 8 & -1.80 & 7.60 \\ -1.80 & 11.20 & -2.80 \\ 7.60 & -2.80 & 14.40 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A', es decir formamos $\alpha A'$

$$\propto A^1 = \begin{bmatrix} + - + \\ - + - \\ + - + \end{bmatrix}$$

$$\begin{bmatrix} +\begin{vmatrix} 11.20 & -2.80 \\ -2.80 & 14.40 \end{vmatrix} \\ (11.20x - 14.40) - \\ (-2.80x - 2.80) \\ = 153.4400 \end{bmatrix} \quad \begin{bmatrix} -\begin{vmatrix} -1.80 & -2.80 \\ 7.60 & 14.40 \end{vmatrix} \\ (-1.80x14.40) - \\ (-2.80x7.60) \\ = -4.64 = 4.64 \end{bmatrix} \quad \begin{bmatrix} +\begin{vmatrix} -1.80 & 11.20 \\ 7.60 & -2.80 \end{vmatrix} \\ (-1.80x - 2.80) - \\ (11.20x7.60) \\ = -80.08 \end{bmatrix}$$

$$\begin{bmatrix} -\begin{vmatrix} -1.80 & 7.60 \\ -2.80 & 14.40 \end{vmatrix} \\ (-1.80x - 14.40) - \\ (7.60x - 2.80) \\ = -4.64 = 4.64 \end{bmatrix} \quad \begin{bmatrix} +\begin{vmatrix} 8 & 7.60 \\ 7.60 & 14.40 \end{vmatrix} \\ (8x14.40) - \\ (7.60x7.60) \\ = 57.44 \end{bmatrix} \quad \begin{bmatrix} -\begin{vmatrix} 8 & -1.80 \\ 7.60 & -2.80 \end{vmatrix} \\ (8x - 2.80) - \\ (-1.80x7.6) \\ = -8.72 = 8.72 \end{bmatrix}$$

$$\begin{bmatrix} +\begin{vmatrix} -1.80 & 7.60 \\ 11.20 & -2.80 \end{vmatrix} \\ (-1.80x - 2.80) - \\ (7.60x11.20) \\ = -80.080 \end{bmatrix} \quad \begin{bmatrix} -\begin{vmatrix} 8 & -1.80 \\ 7.60 & -2.80 \end{vmatrix} \\ (8x - 2.80) - \\ (-1.80x7.60) \\ = -8.72 = 8.72 \end{bmatrix} \quad \begin{bmatrix} +\begin{vmatrix} 8 & -1.80 \\ -1.80 & 11.20 \end{vmatrix} \\ (8x11.20) - \\ (-1.80x - 1.80) \\ = 86.36 \end{bmatrix}$$

Paso 5. Calculamos la matriz inversa de A:

$$A^{-1} = \frac{1}{|A|} \propto A^1 = \frac{1}{610.56}\begin{bmatrix} 153.44 & 4.64 & -80.08 \\ 4.64 & 57.44 & 8.72 \\ -80.08 & 8.72 & 86.36 \end{bmatrix} = \begin{bmatrix} \frac{153.44}{610.56} & \frac{4.64}{610.56} & \frac{-80.08}{610.56} \\ \frac{4.64}{610.56} & \frac{57.44}{610.56} & \frac{8.76}{610.56} \\ \frac{-80.08}{610.56} & \frac{8.72}{610.56} & \frac{86.36}{610.56} \end{bmatrix}$$

$$= \begin{bmatrix} .2513 & .0076 & -0.1312 \\ .0076 & .0941 & .0143 \\ -0.1312 & .0143 & .1414 \end{bmatrix}$$

Paso 6. Llevamos el valor anterior a la expresión $K = A^{-1} * D$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} .2513 & .0076 & -0.1312 \\ .0076 & .0941 & .0143 \\ -0.1312 & .0143 & .1414 \end{bmatrix}\begin{bmatrix} 1.80 \\ 1.60 \\ .20 \end{bmatrix} = \begin{bmatrix} .4383 \\ .1671 \\ -.1850 \end{bmatrix}$$

Paso 7. Obtenemos el modelo discriminante:

$$\boldsymbol{Z = .4383K_1 + .1671K_2 - .1850K_3}$$

Resolución del inciso b)

A continuación se prueba si el modelo es bueno para discriminar o clasificar a las personas que están de acuerdo con las que no lo están mediante los siguientes pasos:

Paso 1. Plantear el juego de hipótesis:

Paso 1. Plantear el juego de hipótesis:

H_0. El modelo discriminante no es bueno para discriminar o clasificar a las personas que están o no de acuerdo con las características de las laptops analizadas.

H_1. El modelo discriminante es bueno.

Paso 2. Establecer el estadístico de prueba:

Paso 2. Establecer el estadístico de prueba:

$$F\ calc = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2)D^2}{[(n_1 + n_2 - 2)p * (n_1 + n_2)]}$$

$$D^2 = (n_1 + n_2 - 2)\left[\sum K_1(\bar{X}_1|Grupo\ 1 - \bar{X}_1|Grupo2\right]$$

$$D^2 = (5 + 5 - 2)[(.4383)(1.80) + (.1671)(1.60) + (-.1850)(.20)] = 8.1544$$

$$F\ calc = \frac{[(5 + 5 - 2 - 1)(5 * 5)(8.1544)]}{[(5 + 5 - 2)2](5 + 5)} = \frac{7x25x8.1544}{16x10} = \frac{1427.0200}{160} = 8.9189$$

Paso 3. Establecer la región de Rechazo:

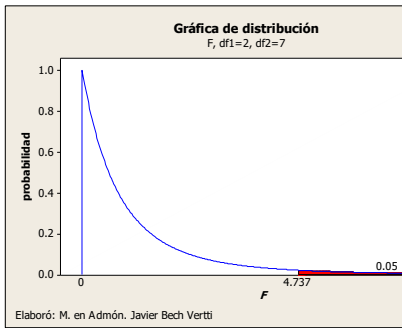
Paso 3. Establecer la región de Rechazo:

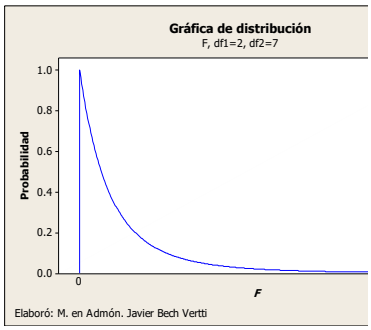
$Fcrit = [\propto_1 p_1(n_1 + n_2 - p - 1)]$

$Fcrit = 0.05, 2, 7$

$f_{0.05} = 4.74$

$f_{0.01} = 9.55$





Paso 4. Definir la regla de decisión:

Paso 4. Definir la regla de decisión:

Se rechaza H_0 si $F_{calc.} \geq 4.74$

Paso 5. Conclusiones:

Paso 5. Conclusiones:

-Estadística: Como 4.74 < 8.9189 < 9.55 ∴ la prueba es (S) y se rechaza H_0 con poca fuerza.

-Administrativa: Existe evidencia suficiente para decir que estadísticamente el modelo es bueno para clasificar a las personas que están de acuerdo con la que está en desacuerdo con las dos marcas de laptops.

Resolución del inciso c)

Cuando se ha concluido que el modelo discriminante es bueno para hacer clasificaciones, se debe proceder a validar el mismo. Esta validación consiste en determinar el grado de confiabilidad que se tendrá al hacer pronósticos con el modelo.

Paso 1. Se obtiene la Z promedio para cada grupo:

$$Z = .4383K_1 + .1671K_2 - .1850K_3$$

$$\bar{Z}_1 = .4383(3.60) + .1671(3.60) - .1850(2.60) = 1.6989$$

$$\bar{Z}_2 = .4383(1.80) + .1671(2.00) - .1850(2.40) = .6791$$

Paso 2. Se determina la Z crítica o de corte:

$$Z^*_{critica} = \frac{n_2\bar{Z}_1 + n_1\bar{Z}_2}{n_1 + n_2} = \frac{5(16989) + 5(06791)}{5 + 5} = 1.1890$$

*Esta fórmula puede usarse ya sea si $n_1=n_2$ o si $n_1 \neq n_2$

Paso 3. Se obtiene la Z para cada observación:

$Z_1 = .4383(4) + .1671(4) - .1850(3) = 1.866 > 1.1890$

$Z_2 = .4383(3) + .1671(5) - .1850(2) = 1.7804 > 1.1890$

$Z_3 = .4383(5) + .1671(3) - .1850(4) = 1.9528 > 1.1890$

$Z_4 = .4383(4) + .1671(2) - .1850(2) = 1.7174 > 1.1890$

$Z_5 = .4383(2) + .1671(4) - .1850(2) = 1.1750 > 1.1890$

$Z_6 = .4383(3) + .1671(2) - .1850(5) = .7241 > 1.1890$

$Z_7 = .4383(2) + .1671(4) - .1850(1) = 1.3600 > 1.1890$

$Z_8 = .4383(2) + .1671(1) - .1850(3) = .4887 > 1.1890$

$Z_9 = .4383(1) + .1671(2) - .1850(2) = .4025 > 1.1890$

$Z_{10} = .4383(1) + .1671(1) - .1850(1) = .4204 > 1.1890$

Paso 4. Se verifica si en las Z observadas la clasificación es correcta o incorrecta

Sujeto No.	**Grupos**	**Decoloración (X_1)**	**Resistencia a los golpes (X_2)**	**Calentamiento interno (X_3)**	**Z obs.**	**Clas. ✓ o X**
1	Acuerdo	4	4	3	1.8660>1.1890	✓
2	Acuerdo	3	5	2	1.7804>1.1890	✓
3	Acuerdo	5	3	4	1.9528>1.1890	✓
4	Acuerdo	4	2	2	1.7174>1.1890	✓

5	Acuerdo	2	4	2	1.1774<1.1890	**X**
	Media Gpo. 1	3.60	3.60	2.60		
6	Desacuerdo	3	2	5	0.7241<1.1890	✓
7	Desacuerdo	2	4	1	1.3600>1.1890	**X**
8	Desacuerdo	2	1	3	0.4887<1.1890	✓
9	Desacuerdo	1	2	2	0.4025<1.1890	✓
10	Desacuerdo	1	1	1	0.4204<1.1890	✓
	Media Gpo. 2	1.80	2.00	2.40		

Paso 5. Del paso anterior se obtiene el porcentaje de clasificaciones correctas (PCC):

$$PCC = \frac{\#\checkmark Grupo\ 1 + \#\checkmark Gupo\ 2}{\#\ de\ observaciones} * 100$$

$$PCC = \frac{4+4}{10} * 100 = 80\%$$

Paso 6. Se obtiene el porcentaje mínimo de clasificaciones correctas:

$$C = (p_i)^2 + (1 - p_1)^2$$

$$C = (\frac{5}{10})^2 + (1 - \frac{5}{10})^2$$

$$C = 0.25 + 0.25 = 0.50x100 = 50\%$$

Paso 7. Se compara el porcentaje de clasificaciones correctas contra el porcentaje mínimo de clasificaciones correctas:

Como 80% > 50% se puede validar el modelo para clasificar a los que están de acuerdo con los que están en desacuerdo al analizar ambas laptops.

Resolución del inciso d)

Para interpretar los resultados obtenidos anteriormente, se evalúa la importancia relativa que posee cada una de las variables independientes para discriminar. El método de los pesos discriminantes estandarizados consiste en estandarizar los coeficientes del modelo discriminante con el fin de que puedan ser comparados. Para jerarquizar las variables se considera solamente la magnitud del peso, no su signo, es decir el valor absoluto.

La fórmula para estandarizar los coeficientes discriminantes es:

$$K_n' = (K_n)\left[\frac{\sum X_n'^2}{(n_1 + n_2 - 2)}\right]^{0.05}$$

En el problema presentado, los coeficientes o constantes estandarizadas son:

1. Decoloración:

$$K_1 = (0.4383)\left[\frac{8}{(5+5-2)}\right]^{0.5} = 0.4383$$

2. Resistencia a golpes:

$$K_2 = (0.1671)\left[\frac{11.20}{(5+5-2)}\right]^{0.5} = 0.1977$$

3. Calentamiento interno:

$$K_3 = (-0.1850)\left[\frac{14.40}{(5+5-2)}\right]^{0.5} = -0.2482$$

-Interpretación: De acuerdo a los pesos discriminantes estandarizados podemos concluir que la variable que más discrimina o diferencia a un grupo de otro para los consumidores de laptops es la variable X_1 (decoloración) con un peso de .4383 por su mayor *K* estandarizada y la menos importante es la variable X_2 (resistencia a los golpes) con un peso de 0.1977.

-Estrategias a seguir: Lo que se pretendería como estrategia es tratar de actuar sobre la variable dependiente (grupos), manipulando las variables independientes (Decoloración, Resistencia a los golpes y calentamiento interno). Lo que se quisiera lograr es aumentar los valores de Z, para hacer que las personas del grupo ***Desacuerdo*** se pasen al grupo ***Acuerdo.*** Para nuestro problema, tenemos que si X_1 (Decoloración) aumenta, subirá a su vez el valor de Z, pues su signo es positivo. Asimismo tenemos que si X_2 (Resistencia a los golpes) aumenta, subirá el valor de Z pues su signo es también positivo, en cambio para X_3 (Calentamiento Interno), tenemos que va a afectar en forma inversa al valor de Z, por su signo negativo.

De las tres variables, la primera que se debe cuidar es X_1 (Decoloración), ya que tiene una mayor importancia para discriminar entre los grupos, mejorando tal vez la calidad de la pigmentación junto con un protector de color que ayude a combatir la decoloración, después podría pensarse en una estrategia en cuanto a la variable X_3 (Calentamiento Interno), en un segundo plano, para evitar un sobrecalentamiento con el uso prolongado y por último una estrategia para la variable X_2 (Resistencia a los golpes), en un tercer plano, para mejorar la resistencia a los golpes.

Resolución del inciso e)

$X_1 = 1$

$X_2 = 3$

$X_3 = 2$

Sustituimos los valores anteriores en el modelo discriminante:

$Z = .4383(1) + .1671(3) - ,1850(2) = .5696$

Comparamos el valor anterior con la Z crítica:

Como .5696 < 1.1890 ∴ se le clasificaría en el grupo de **Desacuerdo**.

3.1.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE
APRENDIZAJE

3.1.1.1

ANÁLISIS
DISCRIMINANTE

Una empresa del segmento de alimentos está desarrollando una nueva galleta fabricada con cereal del cual se considera que la cantidad de proteínas, vitaminas y minerales que llega a tener la mezcla son tres de los factores clave para que el consumidor le llegue a agradar o desagradar el sabor del producto. Para llevar a cabo la investigación, la empresa le da a degustar a cada persona la mezcla y le pide su opinión (le gusta o no le gusta). La empresa quiere saber qué características de su nuevo producto y que cantidades son útiles para diferenciar a las personas que les gusta o no la nueva galleta de cereal, es decir cual separa mejor a cada grupo.

La tabla siguiente refleja la valoración sobre estas tres características con diferentes cantidades de Proteínas y diferentes porcentajes de Vitaminas y Minerales efectuada a un grupo de 10 personas:

No. de persona	Grupos basados en la opinión sobre el producto	Evaluación de la nueva galleta		
		Proteínas en gramos (X_1)	Vitaminas en % (X_2)	Minerales en %(X_3)
1	Me gusta	8	6	7
2	Me gusta	9	7	9
3	Me gusta	7	8	7
4	Me gusta	8	3	9
5	Me gusta	8	6	8
	Media del grupo 1	**8**	**6**	**8**

6	No me gusta	2	2	4
7	No me gusta	3	2	2
8	No me gusta	6	2	2
9	No me gusta	4	2	3
10	No me gusta	5	2	4
	Media del grupo 2	**4**	**2**	**3**

a) Obtenga el modelo discriminante que nos ayude a comprobar la existencia de dos grupos.

b) Pruebe si el modelo es bueno para pronosticar o clasificar a las personas que les gusta o no la nueva galleta de cereal.

c) Valide el modelo discriminante con los mismos datos de la derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados indicando en orden de importancia las variables que logran segmentar y que estrategia se debería seguir.

e) ¿En qué grupo caería una persona que hubiera opinado sobre el nuevo producto de la siguiente manera? : Proteínas: 1; Vitaminas: 3 y Minerales: 2

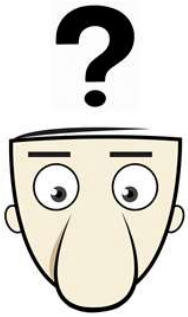
3.1.1.1 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra al final del cuaderno de trabajo en el anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN

3.1.1.1

ANÁLISIS DISCRIMINANTE



1. Supongamos que cierta empresa desea averiguar si uno de sus nuevos productos, una nueva y mejorada batidora tendrá éxito comercial

Para llevar a cabo la investigación, la empresa está interesada en identificar si es posible, a aquellos consumidores que comprarían el nuevo producto y aquellos que no lo comprarían.

Para a ayudar a identificar a los compradores potenciales, la empresa ha ideado escalas de valoración para tres características:- duración, funcionamiento y diseño- que son utilizados por los consumidores para evaluar el nuevo producto.

La empresa querría saber que características de su nuevo producto son útiles para diferenciar a los compradores de los NO compradores, es decir, cual separa mejor a cada grupo.

La tabla siguiente refleja la valoración sobre estas tres características con un precio especificado con un grupo de 10 compradores potenciales donde el 1 refleja la valoración más baja y 10 excelente.

No. De sujeto	Grupos basados en la intención de compra	Evaluación del nuevo producto		
		X_1 Duración	X_2 Funcionamiento	X_3 Diseño
1	Compraría	8	9	6
2	Compraría	6	7	5
3	Compraría	10	6	3
4	Compraría	9	4	4
5	Compraría	4	8	2

	Media del grupo 1	7.4	6.8	4
6	No compraría	5	4	7
7	No compraría	3	7	2
8	No compraría	4	5	5
9	No compraría	2	4	3
10	No compraría	2	2	2
	Media del grupo 2	3.2	4.4	3.8

a) Obtenga el modelo discriminante que nos ayude a comprobar la existencia de dos grupos.

b) Pruebe si el modelo es bueno para pronosticar o clasificar a las personas que comprarían de las que no comprarían la nueva y mejorada batidora.

c) Valide el modelo discriminante con los mismos datos de derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados indicando en orden de importancia las variables que logran segmentar y que estrategia se debería seguir.

e) ¿En qué grupo caería una persona que hubiera calificado el nuevo producto de la siguiente manera? : Duración: 7; Funcionamiento: 5 y Diseño: 6

3.1.1.1 EJERCICIOS DE **REFUERZO**

EJERCICIOS DE
REFUERZO

3.1.1.1

ANÁLISIS
DISCRIMINANTE

3.1.1.1.1 Supongamos que una compañía de Piscinas ha reunido datos de ingreso anual y tamaño de terreno de grupos de propietarios de piscina y no propietarios, quienes viven en el sureste de Mèxico. Además, hay datos disponibles para cada grupo sobre actitudes hacia los baños de sol, en escala de 0= *detestar los baños de sol* hasta 10= *extremadamente aficionado a tomar baños de sol.* Los datos se resumen a continuación:

No. De Sujeto	Grupo basado en Propiedad de Piscina	Ingreso (000'000'pesos) (X_1)	Tamaño del lote (Miles de m^2) ($X_{2)}$)	Medida de Actitud ($X_{3)}$)
1	Propietario	0.7	4.7	9
2	Propietario	0.9	5.6	10
3	Propietario	1.4	9.3	8
4	Propietario	0.7	2.3	5
5	Propietario	0.5	3.8	7
	Media Grupo 1	**0.84**	**5.14**	**7.8**
6	No Propietario	0.5	3.3	4
7	No Propietario	0.7	4.0	3
8	No Propietario	0.6	3.6	2
9	No Propietario	0.4	2.8	4
10	No Propietario	0.5	2.4	5
	Media Grupo 2	**0.54**	**3.22**	**3.6**

a) Obtener el modelo discriminante.

b) Probar si el modelo es bueno para pronosticar.

c) Validar el modelo con los mismos datos de derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados.

e) En que grupo caería una persona con las siguientes características:

Ingreso anual = $ 600,000

Tamaño del terreno= 4,180 metros cuadrados.

Actitud=8

3.1.1.1.2. Un departamento de créditos bancarios está interesado en el posible uso del análisis discriminante en la selección preliminar de solicitudes de crédito. De registros anteriores la Institución ha reunido información sobre dos grupos de adjudicatarios de crédito: a) Riesgos buenos y b) Riesgos dudosos; con base en tres características: Ingresos anuales (en millones de pesos) (X_1), Número de tarjetas de Crédito (X_2) y Edad del solicitante (X_3). Reúne información sobre una muestra aleatoria de 10 adjudicatarios de crédito que a continuación se presenta:

No. De sujeto	Grupos basados en la magnitud del riesgo	Evaluación del riesgo		
		X_1 Ingreso anual (en millones de pesos)	X_2 Tarjetas de crédito (número)	X_3 Edad (años)
1	Riesgo bueno	0.376	7	42
2	Riesgo bueno	0.348	6	47
3	Riesgo bueno	0.452	4	41
4	Riesgo bueno	0.486	5	37
5	Riesgo bueno	0.388	1	43

	Media del grupo 1	**0.410**	**5**	**42**
6	Riesgo dudoso	0.286	4	33
7	Riesgo dudoso	0.292	7	32
8	Riesgo dudoso	0.270	3	40
9	Riesgo dudoso	0.204	1	36
10	Riesgo dudoso	0.298	6	39
	Media del grupo 2	**0.270**	**4**	**36**

a) Obtenga el modelo discriminante que nos ayude a comprobar la existencia de dos grupos.

b) Pruebe si el modelo es bueno para pronosticar o clasificar el riesgo en la selección de solicitudes de crédito.

c) Valide el modelo discriminante con los mismos datos de derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados indicando en orden de importancia las variables que logran segmentar y que estrategia se debería seguir.

e) ¿En qué grupo caería una persona con un ingreso de $ 340,000 pesos, 3 tarjetas y 50 años?

3.1.1.1.3. Se está interesado en el posible uso del análisis discriminante para conocer si en una cierta Universidad no existen profesores intermedios, es decir , o son buenos o son malos; por lo que se analizaron 10 profesores con base en su capacidad docente (X_1) (metodología de enseñanza, exámenes representativos, calificaciones justas), experiencia en el área de la materia (X_2) y conocimiento en la misma (X_3); todo esto en una escala de 1-5 donde el uno significa malo y cinco bueno, para ver si realmente se segmentaban de esta forma. Los resultados obtenidos se muestran a continuación:

Profesor	Opinión	Cap.docente	Experiencia	Conocimiento
1	Bueno	3	4	3
2	Bueno	4	5	5

3	Bueno	4	5	2
4	Bueno	5	4	5
5	Bueno	5	4	2
	Media del grupo 1	**4.2**	**4.4**	**3.4**
6	Malo	1	2	1
7	Malo	1	1	1
8	Malo	2	1	2
9	Malo	1	1	1
10	Malo	3	2	1
	Media del grupo 2	**1.6**	**1.4**	**1.2**

a) Obtener el modelo discriminante

b) Probar si el modelo es bueno para segmentar

c) Valide el modelo con los mismos datos de derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados.

e) ¿En qué grupo caería un profesor que hubiera sido evaluado con con una Capacidad docente de 4, una experiencia de 3 y un conocimiento de 3?

OBJETIVO 3.2. El alumno comprenderá la estructura para realizar pruebas de hipótesis con datos multivariados

ANTECEDENTES

CONCEPTOS DE:

Parámetros, Estimadores, Estadísticos, Distribución muestral de la media, Distribución muestral de la varianza, Distribución muestral de la razón de varianza, Teorema del límite central, Prueba de hipótesis, Hipótesis nula, Hipótesis alternativa, Nivel de confianza, nivel de significancia, Valor p, Error tipoI, Error tipo II. Experimento, Factores controlados, Factores no controlados, Niveles de un factor, Variable de respuesta, Ensayos ó Réplicas, Unidad experimental, error experimental, Confusión, Aleatorización, Agrupamiento, Bloqueo, Balanceo, Análisis de varianza (ANOVA).

3.2.1 PRUEBAS SOBRE VARIAS VARIABLES (VECTOR DE MEDIAS)

CONCEPTOS BÁSICOS

INFERENCIA PARA UN VECTOR DE MEDIAS

PROCEDIMIENTOS DE INFERENCIA PARA UN VECTOR DE MEDIAS

En esta sección se consideran métodos para hacer inferencias acerca de un vector de parámetros de medias poblacionales.

Suponga que $x_1, x_2, \ldots, x_n$ es una muestra aleatoria proveniente de una distribución normal univariada, $N(\mu, \delta^2)$. Sea μ_0 una constante positiva especificada por el usuario. En su primer curso de estadística, se suele aprender como probar $\mathbf{H_0{:}\mu = \mu_0}$ **vs** $\mathbf{H_1{:}\ \mu \neq \mu_0}$, aplicando la prueba t de Student, que requiere que se calcule $t = \left(\bar{\mu} - \mu_0\right)/s/\sqrt{n}$, en donde $\bar{x}$ es la media de la muestra y s es la desviación estándar de esta última. Entonces se rechaza H_0 si $|t| > t_{\alpha/2.N-1}$, en donde $t_{\alpha.\nu}$ es el punto crítico superior de la distribución *t* de Student con ν grados de libertad.

ESTADÍSTICA t^2 DE HOTELLING.

Con el fin de generalizar la prueba t de Student al caso multivariado, suponga que $x_1, x_2, \ldots, x_n$ es una muestra aleatoria proveniente de una distribución normal multivariada, $N_p(\mu, \delta^2)$, en donde se conocen tanto μ como δ^2. Considere probar $\mathbf{H_0{:}\mu = \mu_0}$ **vs** $\mathbf{H_1{:}\ \mu \neq \mu_0}$, en donde $\boldsymbol{\mu_0}$ es un vector conocido (es decir, especificado por el usuario) de p x 1.

El estadístico de prueba, surge de manera natural del estadístico *t*.

Elevando al cuadrado este estadístico, se tiene:

$$t^2 = \frac{(\bar{\mu}_1 - \mu_0)^2}{\frac{s}{\sqrt{n}}} = (\bar{\mu}_1 - \mu_0)\left[s^2\left(\frac{1}{n}\right)\right]^{-1}(\bar{\mu}_1 - \mu_0)$$

luego,

$$t^2 = n(\bar{\mu}_1 - \mu_0)\left[s^2\right]^{-1}(\bar{\mu}_1 - \mu_0)$$

Si se reemplazan las medias de cada variable por los vectores de medias, y reemplazando la varianza s^2 por la matriz de varianzas y covarianzas muestral **S**, se tiene el estadístico $\boldsymbol{T^2}$, de Hotelling:

$$T^2 = n(\bar{\mu}_1 - \mu_0)'\left[S\right]^{-1}(\bar{\mu}_1 - \mu_0)$$

Entonces T^2 se conoce como ***estadística T^2 de Hotelling***. Cuando p=1, la estadística T^2 de Hotelling es igual al cuadrado de la estadística *t* de Student, es decir, $T^2 = t^2$.

Hotelling demostró que la siguiente transformación

$$\frac{n-p}{p(n-1)}T^2$$

está distribuida como una distribución *F*, con los grados de libertad del numerador iguales a ***p* variables** y los del denominador iguales a $n - p$. Se pueden usar estos resultados para probar hipótesis acerca de μ.

Para probar **H_0:μ= μ_0 vs H_1: μ ≠ μ_0**, en donde μ_0 es un vector conocido, calcule

$$T^2 = n(\hat{\mu} - \mu_0)'[S^2]^{-1}(\hat{\mu} - \mu_0)$$

y:

$$S^2 \frac{1}{n-1} SCPC = \frac{1}{n-1} D'D \qquad \text{(matriz de varianzas y covarianzas)}$$

y:

$$[S^2]^{-1} = \frac{1}{|S^2|}\alpha S' \qquad \text{(inversa de la matriz de varianzas y covarianzas)}$$

y rechace H_0 si

$$F_{calculada} = \frac{n-p}{p(n-1)}T^2 > F_{a.p.n-p}$$

en donde $F_{a.v1.v2}$ es el valor crítico superior α.100% de la distribución *F* con v_1 y v_2 grados de libertad.

3.2.1.1 EJEMPLO ILUSTRATIVO

EJEMPLO ILUSTRATIVO

3.2.1.1

VECTOR DE MEDIAS

El Director de Mercadotecnia de una cadena de cines desea medir el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines. Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Para que los potenciales consumidores recuerden la cadena de cines cuando se utiliza la PRENSA ESCRITA como medio masivo para la campaña de publicidad en los consumidores los valores medios de las variables deben ser por lo menos de (7,9,7). Para comprobar si se logró este objetivo, una vez lanzada la campaña en la PRENSA ESCRITA, se tomó una muestra de 7 potenciales clientes y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados son los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo:

$$X = \begin{bmatrix} 4 & 8 & 3 \\ 6 & 7 & 3 \\ 5 & 9 & 4 \\ 5 & 9 & 6 \\ 5 & 7 & 4 \\ 6 & 6 & 7 \\ 7 & 7 & 6 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 9 \\ 7 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de hipótesis

$$H_0 = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 9 \\ 7 \end{bmatrix}$$

$$H_1 = Al\ menos\ una\ M\ es\ \neq.$$

2. Estadístico de prueba

$$Fcalc = \frac{n-p}{p(n-1)} T^2$$

Donde: $T^2 = n(M - M_0)'[S^2]^{-1}(\widehat{M} - \widehat{M}_0)$

Donde: $S^2 = \frac{1}{n-1} SCPC = \frac{1}{n-1} D'D$ (Matriz de varianzas y covarianzas)=

$$[s^2]^2 = \frac{1}{|S|} \propto [S^2]'$$

$$D = \begin{bmatrix} 4 & 8 & 3 \\ 6 & 7 & 3 \\ 5 & 9 & 4 \\ 5 & 9 & 6 \\ 5 & 7 & 4 \\ 6 & 6 & 7 \\ 7 & 7 & 6 \end{bmatrix} - \begin{bmatrix} 5.43 & 7.57 & 4.71 \\ 5.43 & 7.57 & 4.71 \\ 5.43 & 7.57 & 4.71 \\ 5.43 & 7.57 & 4.71 \\ 5.43 & 7.57 & 4.71 \\ 5.43 & 7.57 & 4.71 \\ 5.43 & 7.57 & 4.71 \end{bmatrix} = \begin{bmatrix} -1.43 & .43 & -1.71 \\ .57 & -.57 & -1.71 \\ -0.43 & 1.43 & -.71 \\ -0.43 & 1.43 & 1.29 \\ -0.43 & -.57 & -.71 \\ .57 & -1.53 & 2.29 \\ 1.57 & -.57 & 1.29 \end{bmatrix}$$

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

Estadística de prueba

Un valor estandarizado calculado de los datos de muestra durante una prueba de hipótesis. Se utiliza para determinar si se desea rechazar la hipótesis nula. Cuando las pruebas de hipótesis comparan los datos de la muestra observados con lo que se esperaba en la hipótesis nula, la comparación se basará en la estadística de prueba. Los valores de una estadística de prueba corresponden a los valores p para la prueba de hipótesis. Por lo tanto, cuando los datos presenten una fuerte evidencia versus las suposiciones en la hipótesis nula, la magnitud de la estadística de prueba será grande y el valor p de la prueba pudiera ser suficientemente pequeño para

rechazar la hipótesis nula.

$$D'D$$

$$= \begin{bmatrix} -1.43 & .57 & -0.43 & -0.43 & -0.43 & .57 & 1.57 \\ .43 & -.57 & 1.43 & 1.43 & -.57 & -1.57 & -.57 \\ -1.71 & -1.71 & -.71 & 1.29 & -.71 & 2.29 & 1.29 \end{bmatrix} \begin{bmatrix} -1.43 & .43 & -1.71 \\ .57 & -.57 & -1.71 \\ -0.43 & 1.43 & -.71 \\ -0.43 & 1.43 & 1.29 \\ -0.43 & -.57 & -.71 \\ .57 & -1.53 & 2.29 \\ 1.57 & -.57 & 1.29 \end{bmatrix}$$

$$D'D = \begin{bmatrix} 5.7143 & -3.7143 & 4.8571 \\ -3.7143 & 7.7143 & -2.8571 \\ 4.8571 & -2.8571 & 15.4287 \end{bmatrix}$$

$$S^2 = \frac{1}{n-1} D'D = \frac{1}{7-1} D'D = \frac{1}{6} \quad \begin{bmatrix} 5.7143 & -3.7143 & 4.8571 \\ -3.7143 & 7.7143 & -2.8571 \\ 4.8571 & -2.8571 & 15.4287 \end{bmatrix}$$

$$S^2 = \frac{1}{6} \begin{bmatrix} 5.7143 & -3.7143 & 4.8571 \\ -3.7143 & 7.7143 & -2.8571 \\ 4.8571 & -2.8571 & 15.4287 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} \frac{5.7143}{6} & \frac{-3.7143}{6} & \frac{4.8571}{6} \\ \frac{-3.7143}{6} & \frac{7.7143}{6} & \frac{-2.8571}{6} \\ \frac{4.8571}{6} & \frac{-2.8571}{6} & \frac{15.4287}{6} \end{bmatrix}$$

Utilizando la regla de Sarrus

$$|S^2| = \begin{bmatrix} \frac{5.7143}{6} & \frac{-3.7143}{6} & \frac{4.8571}{6} \\ \frac{-3.7143}{6} & \frac{7.7143}{6} & \frac{-2.8571}{6} \\ \frac{4.8571}{6} & \frac{-2.8571}{6} & \frac{15.4287}{6} \\ \frac{5.7143}{6} & \frac{-3.7143}{6} & \frac{4.8571}{6} \\ \frac{-3.7143}{6} & \frac{7.7143}{6} & \frac{-2.8571}{6} \end{bmatrix} =$$

$$(5.7143/6 \times 7.7143/6 \times 15.4287/6) +$$
$$(-3.7143/6 \times -2.8571/6 \times 4.8571/6) +$$
$$(4.8571/6 \times -3.0766/6 \times -2.8571/6) -$$
$$(4.8571/6 \times 7.7143/6 \times 4.8571/6) -$$
$$(-2.8571/6 \times -2.8571/6 \times 5.7143/6) -$$
$$(15.4287/6 \times -3.7143/6 \times -3.7143/6) -$$

$= 1.5822$

Matriz de cofactores

$$\propto [S^2]'$$
$$+ - +$$
$$- + -$$
$$+ - +$$

$$= \begin{bmatrix} + \begin{vmatrix} 7.7143/6 & -2.8571/6 \\ -2.8571/6 & 15.4287/6 \end{vmatrix} & - \begin{vmatrix} -3.7143/6 & -2.8571/6 \\ 4.8571 & 15.4287/6 \end{vmatrix} & + \begin{vmatrix} -3.7143/6 & 7.7143/6 \\ 4.8571/6 & -2.8571/6 \end{vmatrix} \\ (7.7143/6 \times 15.4287/6) & (-3.0766 \times 15.4287/6) & (-3.7143 \times -2.8571/6) \\ -(-2.8571/6 \times -2.8571/6) & -(-2.8571/6 \times 4.8571/6) & (7.7143/6 \times 4.8571/6) \\ \underline{\mathbf{= 3.0794}} & = -1.2094 * -1 = \underline{\mathbf{1.2094}} & = \underline{\mathbf{-.7460}} \\ - \begin{vmatrix} -3.7143/6 & 4.8571/6 \\ -2.8571/6 & 15.4287/6 \end{vmatrix} & + \begin{vmatrix} 5.7143/6 & 4.8571/6 \\ 4.8571/6 & 15.4287/6 \end{vmatrix} & - \begin{vmatrix} 5.7143/6 & -3.7143/6 \\ 4.8571/6 & -2.8571/6 \end{vmatrix} \\ (-3.7143 \times 15.4287/6) & (5.7143 \times 15.4287/6) & (5.7143 \times -2.8571/6) \\ -(4.8571/6 \times -2.8571/6) & -(4.8571/6 \times 4.8571/6) & -(-3.7143/6 \times 4.8571/6) \\ = -1.2094 * -1 = \underline{\mathbf{1.2094}} & \underline{\mathbf{= 1.7937}} & = 2.0370 = -2.0370 \\ & & = 0.0476 * -1 = \underline{\mathbf{-0.0476}} \\ + \begin{vmatrix} -3.7143/6 & 4.8571/6 \\ 7.7143/6 & -2.8571/6 \end{vmatrix} & - \begin{vmatrix} 5.7143/6 & -3.7143/6 \\ 4.8571/6 & -2.8571/6 \end{vmatrix} & + \begin{vmatrix} 5.7143/6 & -3.7143/6 \\ -3.7143/6 & 7.7143/6 \end{vmatrix} \\ (-3.7143/6 \times -2.8571/6) & (5.7143 \times -2.8571/6) & (5.7143 \times 7.7143/6) \\ -(4.8571/6 \times 7.7143/6) & -(-3.0766/6 \times 4.8571/6) & (-3.7143/6 \times -3.7143/6) \\ \underline{\mathbf{= -.7460}} & = 0.0476 * -1 = \underline{\mathbf{-0.0476}} & \underline{\mathbf{= 0.8413}} \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{|S^2|} \propto [S^2]' = \frac{1}{1.5822} = \begin{bmatrix} \frac{3.0794}{1.5822} & \frac{1.2094}{1.5822} & \frac{-.7460}{1.5822} \\ \frac{1.2094}{1.5822} & \frac{1.7937}{1.5822} & \frac{-0.0476}{1.5822} \\ \frac{-.7460}{1.5822} & \frac{-0.0476}{1.5822} & \frac{0.8413}{1.5822} \end{bmatrix}$$

$$[S^2]^{-1} = \begin{bmatrix} 1.9463 & .7644 & -0.4715 \\ 0.7644 & 1.1337 & -0.0301 \\ -0.4715 & -0.0301 & 0.5317 \end{bmatrix}$$

Obtener T^2

$$T^2 = n[\widehat{M} - M_{H_0}]'[S^2]^{-1}[\widehat{M} - M_{H_0}]$$

$$T^2 = 7[5.43 - 7; 7.57 - 9; 4.71 - 7][S^2]^{-1} \begin{bmatrix} 5.43 - 7 \\ 7.57 - 9 \\ 4.71 - 7 \end{bmatrix}$$

$$T^2 = 7[-1.57; -1.43; -2.29][S^2]^{-1} \begin{bmatrix} -1.57 \\ -1.43 \\ -2.29 \end{bmatrix}$$

$$T^2 = [-10.99; -10.01; -16.03] \begin{bmatrix} 1.9463 & .7644 & -0.4715 \\ 0.7644 & 1.1337 & -0.0301 \\ -0.4715 & -0.0301 & 0.5317 \end{bmatrix} \begin{bmatrix} -1.57 \\ -1.43 \\ -2.29 \end{bmatrix}$$

$$T^2 = [-21.4833 - 19.2666 - 3.0401] \begin{bmatrix} -1.57 \\ -1.43 \\ -2.29 \end{bmatrix}$$

$$T^2 = \underline{68.2410}$$

$$Fcalc = \frac{n - p}{p(n - 1)} T^2$$

$$Fcalc = \frac{7 - 3}{3(7 - 1)} 68.2410 = .2222 * 68.2410 = \underline{15.1647}$$

3. Región de rechazo

$F_{\propto.p.n-p}$

$F_{0.05.3.4} = 6.59$

$F_{0.01.3.4} = 16.69$

4. Regla de decisión

Se rechaza H_0 si Fcalc >6.59

5. Conclusión

Estadística: Como 6.59<15.1647<16.69

∴ la prueba es (S) y se rechaza H_0 con poca potencia.

Adva: Existe evidencia suficiente (E.E.S). para decir que estadísticamente al menos uno de los parámetros de las tres variables estudiadas (X_1= Evocación de la marca, X_2= Evocación del Slogan y X_3= Evocación del logotipo) es diferente.

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su

Resolución del inciso b)

Intervalo de confianza para marca

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}} M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}} M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$M = 5.43 \mp 2.447 \frac{0.9759}{\sqrt{7}}$

muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

$$M = 5.43 \mp 2.447\frac{0.9759}{2.6458}$$

$$M = 5.43 \mp 2.447(.3688)$$

$$M = 5.43 \mp .9025\begin{cases} LIC = 4.5275 \\ LSC = 6.3325 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio de la evocación de la marca oscila entre 4.5275 y 6.3325. Como 7 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

Intervalo de confianza para slogan

$$M = \bar{x} \mp t_{\propto/2.n-1}\ \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1}\ \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6}\ \frac{S}{\sqrt{n}}$$

$$M = 7.57 \mp 2.447\frac{1.1339}{\sqrt{7}}$$

$$M = 7.57 \mp 2.447\frac{1.1339}{2.6458}$$

$$M = 7.57 \mp 2.447(.4286)$$

$$M = 7.57 \mp 1.0488\begin{cases} LIC = 6.5212 \\ LSC = 8.6188 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio de la evocación del slogan oscila entre 6.5212 y 8.6188. Como 9 se encuentra fuera del

intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

Intervalo de confianza para logotipo

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 4.71 \mp 2.447 \frac{1.6036}{\sqrt{7}}$$

$$M = 4.71 \mp 2.447 \frac{1.6036}{2.6458}$$

$$M = 4.71 \mp 2.447(.6061)$$

$$M = 4.71 \mp 1.4831 \begin{cases} LIC = 3.2269 \\ LSC = 6.1931 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio de la evocación del logotipo oscila entre 3.2269 y 6.1931. Como 7 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

3.2.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE APRENDIZAJE

3.2.1.1

VECTOR DE MEDIAS

El Director de Mercadotecnia de una cadena de cines desea medir el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines. Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Para que los potenciales consumidores recuerden la cadena de cines cuando se utiliza la TELEVISIÓN como medio masivo para la campaña de publicidad en los consumidores los valores medios de las variables deben ser por lo menos de (10,10,10). Para comprobar si se logró este objetivo, una vez lanzada la campaña en la TELEVISIÓN, se tomó una muestra de 6 potenciales clientes y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados son los siguientes:

Considere la siguiente muestra de tamaño n=6 extraída de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo:

$$X = \begin{bmatrix} 9 & 10 & 8 \\ 8 & 9 & 9 \\ 7 & 8 & 9 \\ 9 & 9 & 9 \\ 8 & 8 & 6 \\ 10 & 7 & 10 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

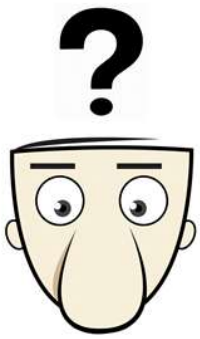
3.2.1.1 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra al final del cuaderno de trabajo en el anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN

3.2.1.1

VECTOR DE MEDIAS



El Director de Mercadotecnia de una cadena de cines desea medir el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines. Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Para que los potenciales consumidores recuerden la cadena de cines cuando se utiliza la REVISTA como medio masivo para la campaña de publicidad en los consumidores los valores medios de las variables deben ser por lo menos de (6,9,8). Para comprobar si se logró este objetivo, una vez lanzada la campaña en la REVISTA, se tomó una muestra de 7 potenciales clientes y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados son los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo:

$$X = \begin{bmatrix} 4 & 6 & 3 \\ 5 & 4 & 6 \\ 7 & 9 & 8 \\ 3 & 8 & 4 \\ 6 & 7 & 5 \\ 8 & 6 & 8 \\ 7 & 7 & 6 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las

medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 9 \\ 8 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.1.1 EJERCICIOS DE **REFUERZO**

EJERCICIO DE REFUERZO

3.2.1.1

VECTOR DE MEDIAS

NOTA:

3.2.1.1.1 El Director de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia de la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares), el tamaño del lote (en pies cuadrados) y el número de recámaras. Para este experimento se manejo uno de los niveles de la ubicación geográfica: **Norte**. Se tomaron lecturas en 7 casas unifamiliares y los resultados fueron los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada denominada **"BIENES1"** donde X_1= Valor de avalúo en la zona **Norte**, X_2= Tamaño del lote en la zona **Norte** y X_3= Número de recámaras en la zona **Norte**:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

$$X = \begin{bmatrix} 174.5 & 6.00 & 3 \\ 275.0 & 6.00 & 3 \\ 185.0 & 6.00 & 4 \\ 220.0 & 6.00 & 3 \\ 218.0 & 7.40 & 3 \\ 190.0 & 3.55 & 4 \\ 218.0 & 6.00 & 3 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula:

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 190.8 \\ 7.50 \\ 3.60 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.1.1.2 El Director de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia de la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares) , el tamaño del lote (en pies cuadrados) y el número de recámaras. Para este experimento se manejo uno de los niveles de la ubicación geográfica: **Sur**. Se tomaron lecturas en 7 casas unifamiliares y los resultados fueron los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada denominada **"BIENES1"** donde X_1= Valor de avalúo en la zona **Sur** , X_2= Tamaño del lote en la zona **Sur** y X_3= Número de recámaras en la zona **Sur**:

$$X = \begin{bmatrix} 189.9 & 7.20 & 3 \\ 305.0 & 12.10 & 3 \\ 189.9 & 7.00 & 5 \\ 115.0 & 18.92 & 3 \\ 210.0 & 10.00 & 3 \\ 181.0 & 7.00 & 4 \\ 180.0 & 7.00 & 4 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 155 \\ 7 \\ 3 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.1.1.3 El Director de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia de la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares) , el tamaño del lote (en pies cuadrados) y el número de recámaras. Para este experimento se manejo uno de los niveles de la ubicación geográfica: **Este**. Se tomaron lecturas en 7 casas unifamiliares y los resultados fueron los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada denominada **"BIENES1"** donde X_1= Valor de avalúo en la zona **Este** , X_2= Tamaño del lote en la zona **Este** y X_3= Número de recámaras en la zona **Este**:

$$X = \begin{bmatrix} 180.0 & 6.00 & 4 \\ 152.0 & 6.00 & 4 \\ 167.8 & 6.00 & 3 \\ 180.0 & 11.00 & 4 \\ 163.0 & 6.00 & 3 \\ 223.5 & 6.00 & 4 \\ 160.0 & 6.00 & 4 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 195 \\ 7.5 \\ 3 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.2 PRUEBAS DE HIPÓTESIS PARA UNA COMPARACIÓN DE MEDIAS PARA VARIAS MUESTRAS

CONCEPTOS BÁSICOS

DIFERENCIA DE MEDIAS MULTIVARIADAS

Estadística de prueba

Un valor estandarizado calculado de los datos de muestra durante una prueba de hipótesis. Se utiliza para determinar si desea rechazar la hipótesis nula. Cuando las pruebas de hipótesis comparan los datos de la muestra observados con lo que se esperaba en la hipótesis nula, la comparación se basará en la estadística de prueba. Los valores de una estadística de prueba corresponden a los valores p para la prueba de hipótesis. Por lo tanto, cuando los datos presenten una fuerte evidencia versus las suposiciones en la

PROCEDIMIENTO DE DOS MUESTRAS

Suponga que x_{i1}, x_{i2}, …,x_{ik}, son muestras aleatorias de dos poblaciones normales multivariadas, $n_p(\mu_{iq}, \delta_i)$, para i= 1,2. Además, suponga que las dos muestras aleatorias son independientes (no pareadas) entre sí. Considere probar Considere probar **H_0:$\mu_1 = \mu_2$ vs H_1: $\mu_1 \neq \mu_2$**. Si $\delta^2_1=\delta^2_2$, existe una prueba exacta de H_0 . Si las dos matrices de vartianzas-covarianzas son distintas, sólo se dispone de una prueba aproximada de H_0. Anteriormente, se dio una prueba respecto a $\delta^2_1=\delta^2_2$. Cuando $\delta^2_1=\delta^2_2$, denote su valor comun por δ. En este caso la mejor estimación de δ es:

Estadísticos de prueba:

El estadístico de prueba para el caso univariado, bajo el supuesto de que las varianzas poblacionales son iguales, es el estadístico ***t*** de Student, con $n_1 + n_2 -2$ grados de libertad:

$$t = \frac{(\bar{\mu}_1 - \bar{\mu}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \text{ en donde } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Bajo la hipótesis nula, de que las medias poblacionales son iguales, el estadístico ***t,*** se convierte en:

hipótesis nula, la magnitud de la estadística de prueba será grande y el valor p de la prueba pudiera ser suficientemente pequeño para rechazar la hipótesis nula.

$$t = \frac{(\bar{\mu}_1 - \bar{\mu}_2)}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Para el caso Multivariado, el estadístico de prueba, surge de manera natural del estadístico *t*.

Elevando al cuadrado este estadístico, se tiene:

$$t^2 = \frac{(\bar{\mu}_1 - \bar{\mu}_2)^2}{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = (\bar{\mu}_1 - \bar{\mu}_2)\left[s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right]^{-1}(\bar{\mu}_1 - \bar{\mu}_2)$$

$$= (\bar{\mu}_1 - \bar{\mu}_2)\left[s^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)\right]^{-1}(\bar{\mu}_1 - \bar{\mu}_2)$$

luego, $$t^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\bar{\mu}_1 - \bar{\mu}_2)\left[s^2\right]^{-1}(\bar{\mu}_1 - \bar{\mu}_2)$$

Si se reemplazan las medias de cada variable por los vectores de medias de cada grupo, y reemplazando la varianza s^2 por la matriz de varianzas y covarianzas muestral **S**, se tiene el estadístico $\boldsymbol{T^2}$, de Hotelling:

$$T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\bar{\mu}_1 - \bar{\mu}_2)'\left[\mathbf{S}\right]^{-1}(\bar{\mu}_1 - \bar{\mu}_2)$$

donde:

$$S = \frac{W_1 + W_2}{n_1 + n_2 - 2}$$ (matriz de varianzas y covarianzas ponderadas ó combinadas)

si:

$$W_1 = SCPC_1 = D_1^{'} D_1 \qquad \text{y} \qquad W_2 = SCPC_2 = D_2^{'} D_2$$

y:

$$S^{-1} = \frac{1}{|S|}\alpha S' \quad \text{(inversa de la matriz de varianzas y covarianzas)}$$

Un valor grande de T^2 de Hotelling es una evidencia de que los vectores de medias son diferentes para las dos poblaciones muestreadas.

Hotelling demostró en 1931 que la siguiente transformación de T^2 tiene distribución F:

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} T^2$$

con p y $(N-p-1)$ grados de libertad, donde p es el número de variables independientes y $N = n_1 + n_2$ es el número total de sujetos.

Si

$$F_{Calculada} = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2 > F_{\alpha.p.n_1+n_2-p-1}$$

entonces se rechaza H_0.

Si se rechaza la hipótesis nula la prueba multivariado no determina por cual de las variables se está rechazando por lo que se recomienda usar el método del intervalo de confianza univariada para la diferencia de medias de cada variable y conocer, de ser el caso, la magnitud de las diferencias.

Existen tres casos para determinar el intervalo de confianza de los cuales los más usados son el caso 2 y el caso 3 ya que por regla generel no se conoce la varianza poblacional:

t de 2 muestras calcula un intervalo de confianza y realiza una prueba de hipótesis de la diferencia entre las medias de dos poblaciones cuando las s son desconocidas y las muestras han sido extraídas independientemente. Este procedimiento se basa en la distribución t y, en el caso de

Caso 1. Variancias poblacionales conocidas:

$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm z\ \alpha / 2\ \sqrt{\ } \sigma^2_1 / n_1 + \sigma^2_2 / n_2 \therefore$

$LIC = (\bar{x}_1 - \bar{x}_2) - z\ \alpha / 2\ \sqrt{\ } \sigma^2_1 / n_1 + \sigma^2_2 / n_2$

$LSC = (\bar{x}_1 - \bar{x}_2) + z\ \alpha / 2\ \sqrt{\ } \sigma^2_1 / n_1 + \sigma^2_2 / n_2$

muestras pequeñas, funciona mejor si sus datos se extraen de distribuciones que son normales o cercanas a normales. A medida que el tamaño de la muestra aumenta, usted puede tener mayor confianza en los resultados

Caso 2. Variancias poblacionales desconocidas pero iguales:

$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t\alpha / 2.n_1 + n_2 - 2\ S_p \sqrt{1/ n_1 + 1/ n_2} \therefore$

$LIC=\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) - t\alpha / 2.\ n_1 + n_2 - 2\ \ S_p \sqrt{1/ n_1 + 1/ n_2}$

$LSC=\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) + t\alpha / 2.\ n_1 + n_2 - 2\ \ S_p \sqrt{1/ n_1 + 1/ n_2}$

Donde S_p (Desviación Estandar Agrupada)= $\sqrt{ (n_1 - 1)\ s^2_1 + (n_2 - 1)\ s^2_2 / n_1 + n_2 - 2}$

Caso 3. Variancias desconocidas pero desiguales:

$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm t\alpha / 2.v \sqrt{ S^2_1 / n_1 + S^2_2 / n_2} \therefore$

$LIC=(\bar{x}_1 - \bar{x}_2) - t\alpha / 2.v \sqrt{ S^2_1 / n_1 + S^2_2 / n_2}$

$LSC=(\bar{x}_1 - \bar{x}_2) + t\alpha / 2.v \sqrt{ S^2_1 / n_1 + S^2_2 / n_2}$

Donde los grados de libertad se determinan de la siguiente manera:

$$v = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}}$$

Para saber si las varianzas poblacionales desconocidas son iguales o no y determinar que caso usar si el 2 o el 3 se debe realizar una prueba bilateral de igualdad de varianzas de la siguiente manera:

Juego de Hipótesis

$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad\ \ "Caso\ 2"$

$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos\ \ "Caso\ 3"$

Estadístico de prueba

$$Fcalc = \frac{{S_1}^2}{{S_2}^2}$$

Regla de decisión

Se rechaza Ho si F calculada es $\leq$ a $F_{1-\alpha/2}$ crítica ó si F calculada es $\geq$ a $F_{\alpha/2}$ crítica si la prueba es bilateral.

3.2.2.1 EJEMPLO ILUSTRATIVO

EJEMPLO ILUSTRATIVO

3.2.2.1

DIFERENCIA DE MEDIAS MULTIVARIADA

El Director de Mercadotecnia de una cadena de cines desea medir el grado de influencia que llega a tener el medio masivo utilizado para su campaña de publicidad en los consumidores. Esta influencia se maneja por medio del "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Los tipos de medios manejados son dos: Prensa escrita y Televisión.

Una vez lanzada la campaña en la PRENSA ESCRITA y en TELEVISIÓN, se tomó una muestra aleatoria de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo de 5 potenciales clientes que leyeron la campaña en la prensa escrita y 6 que la vieron en Televisión y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados fueron los siguientes:

Medio utilizado					
Prensa escrita			Televisión		
Marca	Slogan	Logotipo	Marca	Slogan	Logotipo
4	8	3	9	10	8

6	7	3	8	9	9
5	9	4	7	8	9
5	9	6	9	9	9
5	7	4	8	8	6
			10	7	10

e) Existe evidencia suficiente que indique que el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines es el mismo en ambos medios utilizados?. Pruébelo.
f) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de Hipótesis

$$H_0: \begin{bmatrix} M_{11} \\ M_{21} \\ M_{31} \end{bmatrix} = \begin{bmatrix} M_{12} \\ M_{22} \\ M_{32} \end{bmatrix}$$

$$H_1: Al\ menos\ 1\ combinación\ es\ \neq$$

2. Estadistico de prueba

$$Fcalc = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

Donde $T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)^1 [S^2]^{-1} (\widehat{M}_1 - \widehat{M}_2)$

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia

entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

$$S^2 = \frac{W_1+W_2}{n_1+n_2-2}$$

$$W_1 = SCPC_1 = D'D_1$$

$$W_2 = SCPC_2 = D_2'D_2$$

$$D_1 = \begin{bmatrix} 4 & 8 & 3 \\ 6 & 7 & 3 \\ 5 & 9 & 4 \\ 5 & 9 & 6 \\ 5 & 7 & 4 \end{bmatrix} - \begin{bmatrix} 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 0 \end{bmatrix}$$

$$D_1'D_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & -1 \\ -1 & -1 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix}$$

$$W_1 = D_1'D_1 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 9 & 10 & 8 \\ 8 & 9 & 9 \\ 7 & 8 & 9 \\ 9 & 9 & 9 \\ 8 & 8 & 6 \\ 10 & 7 & 10 \end{bmatrix} - \begin{bmatrix} 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \end{bmatrix} = \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \\ -1.5 & -0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & -2.5 \\ 1.5 & -1.5 & 1.5 \end{bmatrix}$$

$$D_2'D_2 = \begin{bmatrix} 0.5 & -0.5 & -1.5 & 0.5 & -0.5 & 1.5 \\ 1.5 & 0.5 & -0.5 & 0.5 & -0.5 & -1.5 \\ -0.5 & 0.5 & 0.5 & 0.5 & -2.5 & 1.5 \end{bmatrix} \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \\ -1.5 & -0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & -2.5 \\ 1.5 & -1.5 & 1.5 \end{bmatrix}$$

$$= \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix}$$

$$W_2 = D_2'D_2 = \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix}$$

$$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2}$$

$$= \frac{\begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix} + \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix}}{5 + 6 - 2} = \frac{\begin{bmatrix} 7.5/9 & -1.5/9 & 2.5/9 \\ -1.5/9 & 9.5/9 & 1.5/9 \\ 2.5/9 & 1.5/9 & 15.5/9 \end{bmatrix}}{9}$$

$$|S^2| = \frac{\begin{bmatrix} \frac{7.5}{9} & \frac{-1.5}{9} & \frac{2.5}{9} \\ \frac{-1.5}{9} & \frac{9.5}{9} & \frac{1.5}{9} \\ \frac{2.5}{9} & \frac{1.5}{9} & \frac{15.5}{9} \end{bmatrix}}{\begin{bmatrix} \frac{7.5}{9} & \frac{-1.05}{9} & \frac{2.5}{9} \\ \frac{-1.5}{9} & \frac{9.5}{9} & \frac{1.5}{9} \end{bmatrix}} =$$

$$(7.5/9 \times 9.5/9 \times 15.5/9) + (-1.5/9 \times 1.5/9 \times$$

$$2.5/9) + (2.5/9 \times -1.5/9 \times 1.5/9) - (2.5/9 \times$$

$$9.5/9 \times 2.5/9) - (1.5/9 \times 1.5/9 \times 7.5/9) -$$

$$(15.5/9 \times -1.5/9 \times -1.5/9)$$

$$= 1.3470$$

Como es ≠ de 0 tiene solución

$$[S^2]^1 = \begin{bmatrix} \frac{7.5}{9} & \frac{-1.5}{9} & \frac{2.5}{9} \\ \frac{-1.5}{9} & \frac{9.5}{9} & \frac{1.5}{9} \\ \frac{2.5}{9} & \frac{1.5}{9} & \frac{15.5}{9} \end{bmatrix}$$

Matriz de cofactores

$$\propto= [S^2]^2 \begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix} \begin{bmatrix} +\begin{vmatrix} \frac{9.5}{9} & \frac{1.5}{9} \\ \frac{1.5}{9} & \frac{15.5}{9} \end{vmatrix} & -\begin{vmatrix} \frac{-1.5}{9} & \frac{1.5}{9} \\ \frac{2.5}{9} & \frac{15.5}{9} \end{vmatrix} & +\begin{vmatrix} \frac{-1.5}{9} & \frac{9.5}{9} \\ \frac{2.5}{9} & \frac{1.5}{9} \end{vmatrix} \\ (9.5/9 \times 15.5/9) - (1.5/9 \times 1.5/9) & (-1.5/9 \times 15.5/9) - (1.5/9 \times 2.5/9) & (-1.5/9 \times 1.5/9) - (9.5/9 \times 2.5/9) \\ = \underline{1.7901} & = -0.3333 = \underline{0.3333} & = \underline{-0.3210} \\ -\begin{vmatrix} \frac{-1.5}{9} & \frac{2.5}{9} \\ \frac{1.5}{9} & \frac{15.5}{9} \end{vmatrix} & +\begin{vmatrix} \frac{7.5}{9} & \frac{2.5}{9} \\ \frac{2.5}{9} & \frac{15.5}{9} \end{vmatrix} & -\begin{vmatrix} \frac{7.5}{9} & \frac{-1.5}{9} \\ \frac{2.5}{9} & \frac{1.5}{9} \end{vmatrix} \\ (-1.5/9 \times 15.5/9) - (2.5/9 \times 1.5/9) & (7.5/9 \times 15.5/9) - (2.5/9 \times 2.5/9) & (7.5/9 \times 1.5/9) - (-1.5/9 \times 2.5/9) \\ = -0.3333 = \underline{0.3333} & = \underline{1.3580} & = 0.1852 = \underline{-0.1852} \\ +\begin{vmatrix} \frac{-1.5}{9} & \frac{2.5}{9} \\ \frac{9.5}{9} & \frac{1.5}{9} \end{vmatrix} & -\begin{vmatrix} \frac{7.5}{9} & \frac{-1.5}{9} \\ \frac{2.5}{9} & \frac{1.5}{9} \end{vmatrix} & +\begin{vmatrix} \frac{7.5}{9} & \frac{-1.5}{9} \\ \frac{-1.5}{9} & \frac{9.5}{9} \end{vmatrix} \\ (-1.5/9 \times 1.5/9) - (2.5/9 \times 9.5/9) & (7.5/9 \times 1.5/9) - (-1.5/9 \times 2.5/9) & (7.5/9 \times 9.5/9) - (-1.5/9 \times -1.5/9) \\ = \underline{-0.3210} & = 0.1852 = \underline{-0.1852} & = \underline{0.8519} \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{1.3470} = \begin{bmatrix} 1.7901 & 0.3333 & -0.3210 \\ 0.3333 & 1.3580 & -0.1852 \\ -0.3210 & -0.1852 & 0.8519 \end{bmatrix}$$

$$[S^2]^{-1} \begin{bmatrix} \frac{1.7901}{1.3470} & \frac{0.3333}{1.3470} & \frac{-0.3210}{1.3470} \\ \frac{0.3333}{1.3470} & \frac{1.3580}{1.3470} & \frac{-0.1852}{1.3470} \\ \frac{-0.3210}{1.3470} & \frac{-0.1852}{1.3470} & \frac{0.8519}{1.3470} \end{bmatrix} = \begin{bmatrix} 1.3290 & 0.2474 & -0.2383 \\ 0.2474 & 1.0082 & -0.1375 \\ -0.2383 & -0.1375 & 0.6324 \end{bmatrix}$$

$$T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)\left(\widehat{M}_1 - \widehat{M}_2\right)'[S^2]^{-1}\left(\widehat{M}_1 - \widehat{M}_2\right)$$

$$T^2 = \left(\frac{5*6}{5+6}\right)(5-8.5; 8-8.5; 4-8.5)[S^2]^{-1}\begin{bmatrix}5-8.5\\8-8.5\\4-8.5\end{bmatrix}$$

$$T^2 = \left(\frac{30}{11}\right)(-3.5; -0.5; -4.5)[S^2]^{-1}\begin{bmatrix}-3.5\\-0.5\\-4.5\end{bmatrix}$$

$$T^2 = (2.7273)(-3.5; -0.5; -4.5)[S^2]^{-1}\begin{bmatrix}-3.5\\-0.5\\-4.5\end{bmatrix}$$

$$T^2 = [-9.5456; -1.3637; -12.2729]\begin{bmatrix}1.3290 & 0.2474 & -0.2383\\0.2474 & 1.0082 & -0.1375\\-0.2383 & -0.1375 & 0.6324\end{bmatrix}\begin{bmatrix}-3.5\\-0.5\\-4.5\end{bmatrix}$$

$$T^2 = [-10.098 - 2.049 - 5.2992]\begin{bmatrix}-3.5\\-0.5\\-4.5\end{bmatrix}$$

$$T^2 = \underline{60.2139}$$

$$Fcalc = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2$$

$$Fcalc = \frac{5+6-3-1}{3(5+6-2)}60.2139$$

$$Fcalc = \frac{7}{27}60.2139$$

$$Fcalc = 0.2593 \times 60.2139$$

$$\boldsymbol{Fcalc = 15.6110}$$

3. Región de rechazo

$F_{\alpha.p.n1+n2-p-1} = F_{0.05.3.5+6-3-1} = F_{0.05.3.7} = 4.35$

$F_{\alpha.p.n1+n2-p-1} = F_{0.01.3.5+6-3-1} = F_{0.01.3.7} = 8.45$

4. Regla de decisión

Se rechaza H_0 si Fcalc > 4.35

5. Conclusión:

- Estadistica: Como 15.6110 > 4.35 y > 8.45 ∴

La prueba es A.S y se rechaza con mucha potencia

- Adva: Existe evidencia suficiente para decir que el grado de evocación o recuerdo de la Marca, el Slogan y el Logotipo que tiene el consumidor sobre esta cadena de cines no es el mismo en ambos medios utilizados en al menos una de las evocaciones de la Marca, el Slogan o el Logotipo.

Resolución del inciso b)

Nota importante: Para construir el intervalo de confianza se debe probar primero si las varianzas de las muestras son iguales o no para saber que caso de debe usar para construir el intervalo de confianza (Caso 2 para varianzas iguales o Caso 3 para varianzas desiguales) por lo que primero hay que realizar la prueba de homocedasticidad de varianzas.

Prueba de 2 varianzas

Una prueba de hipótesis para determinar si dos varianzas de la población son significativamente diferentes.

Prueba de Homocedasticidad para Marca

1. Juego de Hipótesis

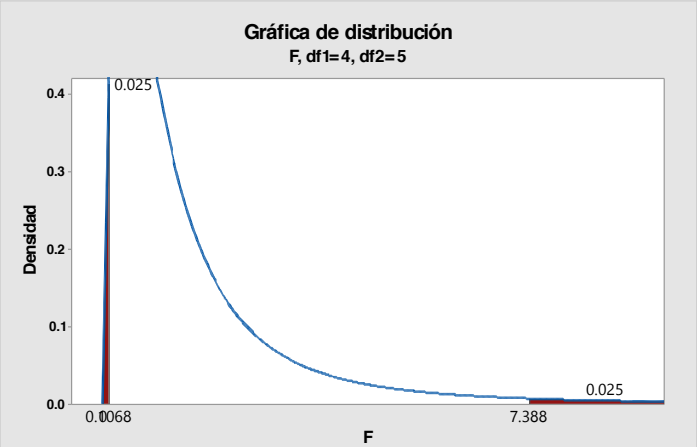
$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$

$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{11}}^2}{{S_{12}}^2} = \frac{0.500}{1.100} = \underline{0.4545}$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.0254.6} = 7.39$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.6-1.5-1}} = \frac{1}{F_{0.025-5.4}} = \frac{1}{9.36} = 0.1068$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc $\geq$ 7.39 ó si es $\leq$ 0.1068

5. Conclusión:

- Estadística: Como 0.1068 < 0.4545 < 7.39 ∴ no se rechaza H_0 y se utiliza el caso 2 para realizar la prueba de hipótesis de diferencia de medias.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

Intervalo de Confianza para Marca

$$M_{11} - M_{12} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp t\,{}^{\propto}/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)0.500 + (6-1)1.100}{5+6-2}}$$

$$Sp = \sqrt{\frac{4 * 0.500 + 5 * 1.100}{9}}$$

$$Sp = \sqrt{\frac{2 + 5.5}{9}}$$

$$Sp = \sqrt{\frac{7.5}{9}}$$

$$Sp = \sqrt{0.8333}$$

$$Sp = \underline{0.9129}$$

$$M_{11} - M_{12} = (5 - 8.5) \mp 2.262 * 0.9129 * \sqrt{\frac{1}{5} + \frac{1}{6}}$$

$$M_{11} - M_{12} = (-3.5) \mp 2.262 * 0.9129 * \sqrt{0.20 + 0.1667}$$

$$M_{11} - M_{12} = (-3.5) \mp 2.262 * 0.9129 * \sqrt{0.3667}$$

$$M_{11} - M_{12} = (-3.5) \mp 2.262 * 0.9129 * 0.6055$$

$$M_{11} - M_{12} = (-3.5) \mp 1.2503 \begin{cases} LIC = -4.7503 \\ LSC = -2.2497 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos, podemos decir que estadísticamente la evocación de la marca en la televisión es mayor que la evocación de la marca en la prensa escrita por un mínimo de 2.2497 y un máximo de 4.7503 puntos.

Nota: Como la Evocación de la Marca es diferente en ambos medios de publicidad es razonable que la prueba de hipótesis de diferencia de medias multivariada se haya rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Marca vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                          IC de 95%
                                            para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      0.707     0.500  (0.282, 2.916)
2        6      1.049     1.100  (0.567, 2.880)

Relación de desviaciones estándar = 0.674
Relación de varianzas = 0.455
```

ANOVA unidireccional: Marca vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1     33.409    33.4091    40.09    0.000
Error     9      7.500     0.8333
Total    10     40.909


Resumen del modelo

                      R-cuad.  R-cuad.
       S  R-cuad.  (ajustado)   (pred)
0.912871   81.67%      79.63%   73.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  5.000      0.707  (4.076, 5.924)
2        6  8.500      1.049  (7.657, 9.343)

Desv.Est. agrupada = 0.912871
```

Comparaciones en parejas de Tukey

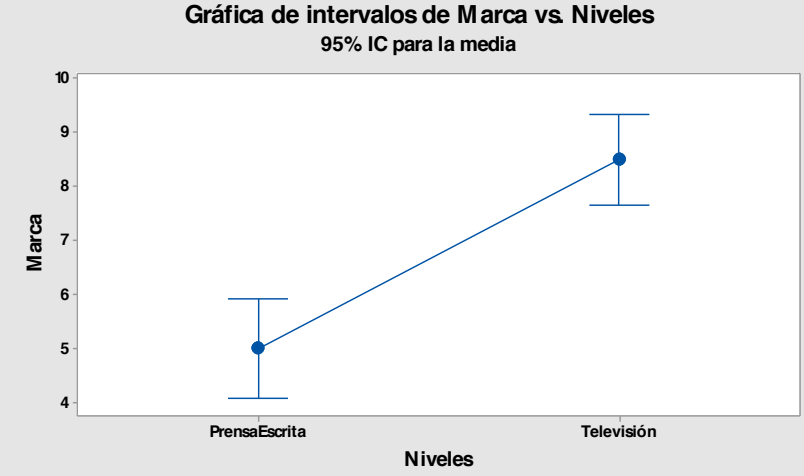
```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de
Valor p
de niveles      medias  diferencia     IC de 95%    Valor T
ajustado
2 - 1            3.500       0.553  (2.250, 4.750)     6.33
0.000

Nivel de confianza individual = 95.00%
```

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un

error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.

Gráfica de intervalos de Marca vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba de Homocedasticidad para Slogan

1. Juego de Hipótesis

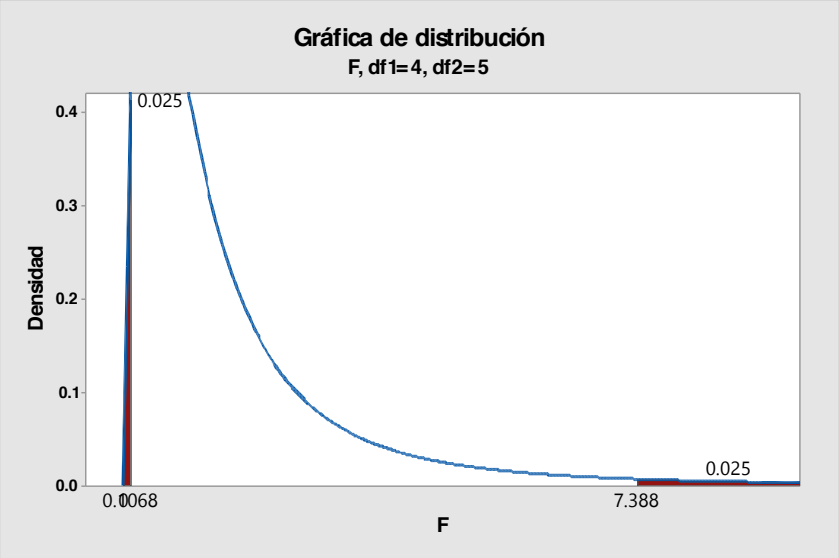
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{21}}^2}{{S_{22}}^2} = \frac{1}{1.100} = 0.9091$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.0254.6} = 7.39$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.6-1.5-1}} = \frac{1}{F_{0.025-5.4}} = \frac{1}{9.36} = 0.1068$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:

- Estadistica: Como 0.01068< 0.9091< 7.39 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adua: Existe evidencia suficiente para decir que estadísticamente las varianzas de slogan de ambas medias son iguales.

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

Intervalo de Confianza para Slogan

$$M_{21} - M_{22} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp t\,{}^{\propto}\!/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)1 + (6-1)1.100}{5+6-2}}$$

$$Sp = \sqrt{\frac{4*1 + 5*1.100}{9}}$$

$$Sp = \sqrt{\frac{4 + 5.5}{9}}$$

$$Sp = \sqrt{\frac{9.5}{9}}$$

$$Sp = \sqrt{1.0556}$$

$$Sp = \underline{1.0274}$$

$$M_{21} - M_{22} = (8 - 8.5) \mp 2.262\ (.9129)\sqrt{\frac{1}{5} + \frac{1}{6}}$$

$$M_{21} - M_{22} = -0.5 \mp 2.262\ (1.0274)\sqrt{0.2 + 0.1667}$$

$$M_{21} - M_{22} = -0.5 \mp 2.262\ (1.0274)\sqrt{0.3667}$$

$$M_{21} - M_{22} = -0.5 \mp 2.262\ (1.0274)(0.6056)$$

$$M_{21} - M_{22} = -0.5 \mp 1.4074 \begin{cases} -1.9074 \\ 0.9074 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la evocación de la marca en la prensa escrita es la misma que la evocación de la marca en la televisión

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Slogan vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                      IC de 95%
                                        para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      1.000     1.000  (0.610, 2.696)
2        6      1.049     1.100  (0.567, 2.880)

Relación de desviaciones estándar = 0.953
Relación de varianzas = 0.909
```

ANOVA unidireccional: Slogan vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
```

```
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1     0.6818     0.6818     0.65    0.442
Error     9     9.5000     1.0556
Total    10    10.1818


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
1.02740    6.70%       0.00%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  8.000      1.000  (6.961, 9.039)
2        6  8.500      1.049  (7.551, 9.449)

Desv.Est. agrupada = 1.02740
```

Comparaciones en parejas de Tukey

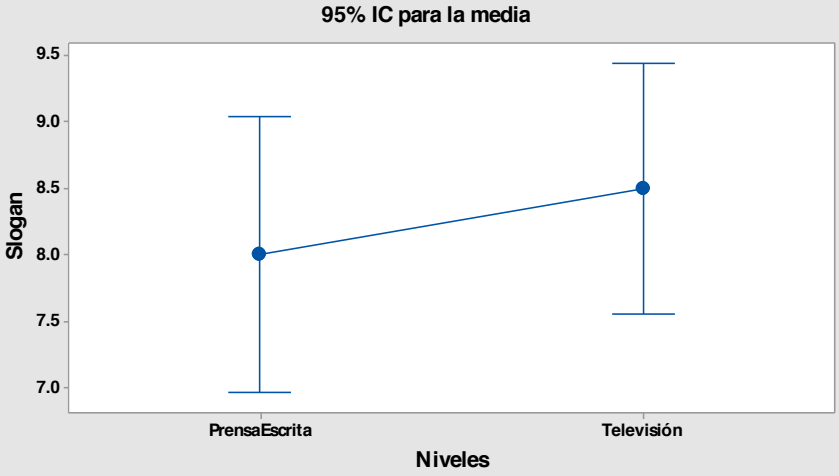
```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de
Valor p
de niveles      medias  diferencia     IC de 95%     Valor T
ajustado
2 - 1            0.500       0.622  (-0.907, 1.907)     0.80
0.442

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Slogan vs. Niveles

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el

intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba de Homocedasticidad para Logotipo

1. Juego de Hipótesis

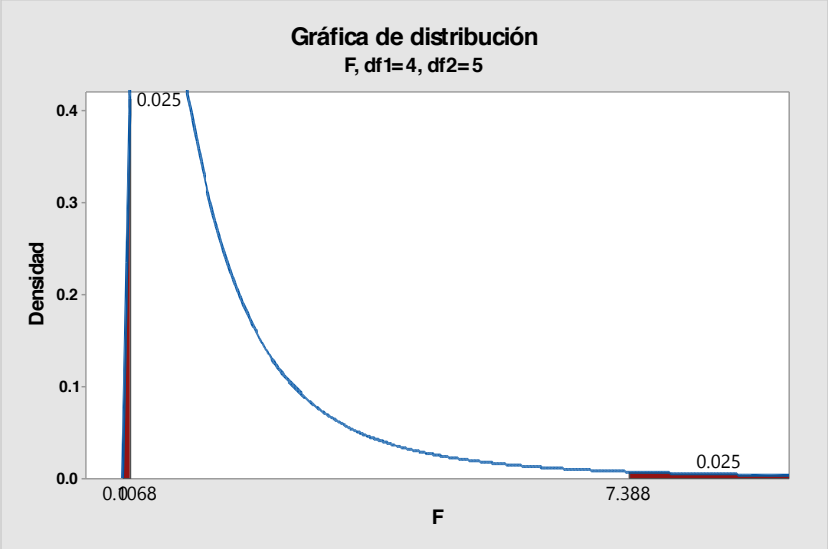
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{31}}^2}{{S_{32}}^2} = \frac{1.500}{1.900} = 0.7895$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.0254.6} = 7.39$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.6-1.5-1}} = \frac{1}{F_{0.025-5.4}} = \frac{1}{9.36} = 0.1068$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:

- Estadística: Como 0.1068< 0.7895<7.39 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de logotipo de ambas medias son iguales.

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

Intervalo de Confianza para Logotipo

$$M_{31} - M_{32} = \left(\overline{X}_{31} - \overline{X}_{32}\right) \mp t\,{}^{\propto}\!/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)1.500 + (6-1)1.900}{5+6-2}}$$

$$Sp = \sqrt{\frac{4 * 1.500 + 5 * 1.900}{9}}$$

$$Sp = \sqrt{\frac{6 + 9.5}{9}}$$

$$Sp = \sqrt{\frac{15.5}{9}}$$

$$Sp = \sqrt{1.7222}$$

$$Sp = \underline{1.3123}$$

$$M_{31} - M_{32} = (4 - 8.5) \mp 2.262\ (1.3123)\sqrt{\frac{1}{5} + \frac{1}{6}}$$

$$M_{31} - M_{32} = -4.5 \mp 2.262\ (1.3123)\sqrt{0.2 + 0.1667}$$

$$M_{31} - M_{32} = -4.5 \mp 2.262\ (1.3123)\sqrt{0.3667}$$

$$M_{31} - M_{32} = -4.5 \mp 2.262\ (1.3123)(0.6056)$$

$$M_{31} - M_{32} = -4.5 \mp 1.7977 \begin{cases} -6.2977 \\ -2.7023 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos, podemos decir que estadísticamente la evocación de la marca en la televisión es mayor que la evocación de la marca en la prensa escrita por un mínimo de 2.7023 y un máximo de 6.2977 puntos.

Nota: Como la Evocación del Logotipo es diferente en ambos medios de publicidad es razonable que la prueba de hipótesis de diferencia de medias multivariada se haya rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Logotipo vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                        IC de 95%
                                          para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      1.225     1.500  (0.489, 5.050)
2        6      1.378     1.900  (0.391, 7.219)

Relación de desviaciones estándar = 0.889
Relación de varianzas = 0.789
```

ANOVA unidireccional: Logotipo vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      55.23     55.227    32.07    0.000
Error     9      15.50      1.722
Total    10      70.73


Resumen del modelo

                    R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
1.31233   78.08%      75.65%   67.40%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  4.000      1.225  (2.672, 5.328)
2        6  8.500      1.378  (7.288, 9.712)

Desv.Est. agrupada = 1.31233
```

Comparaciones en parejas de Tukey

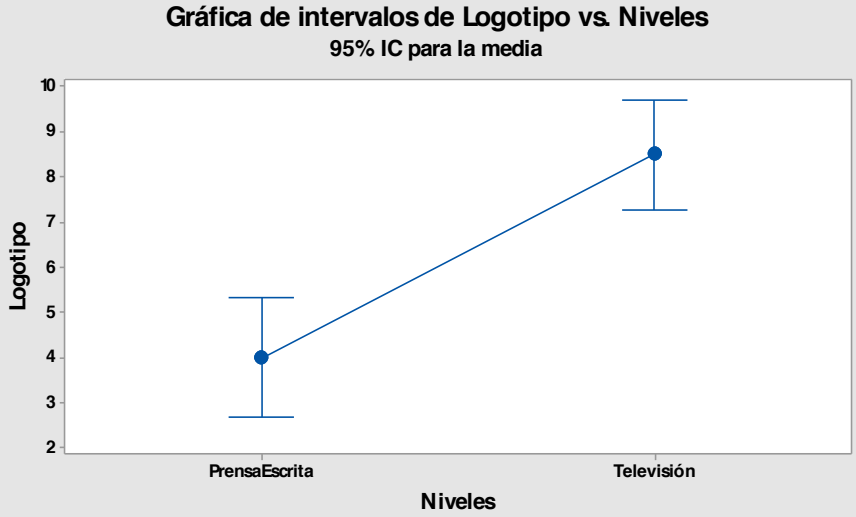
```
Pruebas simultáneas de Tukey para diferencias de las medias

           Diferencia
Diferencia      de las       EE de
Valor p
de niveles      medias  diferencia     IC de 95%    Valor T
ajustado
2 - 1            4.500       0.795  (2.702, 6.298)     5.66
0.000

Nivel de confianza individual = 95.00%
```

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un

error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.

Gráfica de intervalos de Logotipo vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.

3.2.2.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE APRENDIZAJE

3.2.2.1

DIFERENCIA DE MEDIAS MULTIVARIADA

El Director de Mercadotecnia de una cadena de cines desea medir el grado de influencia que llega a tener el medio masivo utilizado para su campaña de publicidad en los consumidores. Esta influencia se maneja por medio del "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Los tipos de medios manejados son dos: Prensa escrita y Revistas.

Una vez lanzada la campaña en la PRENSA ESCRITA y en REVISTAS, se tomó una muestra aleatoria de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo de 5 potenciales clientes que leyeron la campaña en la prensa escrita y 6 que la vieron en Revistas y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados fueron los siguientes:

Medio utilizado					
Prensa escrita			Revistas		
Marca	Slogan	Logotipo	Marca	Slogan	Logotipo
4	8	3	4	6	3
6	7	3	5	4	6
5	9	4	7	9	8
5	9	6	3	8	4
5	7	4	6	7	5
			8	6	8
			7	7	6

a) Existe evidencia suficiente que indique que el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines es el mismo en ambos medios utilizados?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

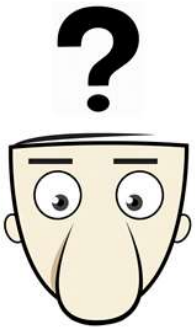
3.2.2.1 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra al final del cuaderno de trabajo en el anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTOEVALUACIÓN 3.2.2.1 DIFERENCIA DE MEDIAS MULTIVARIADA

El Director de Mercadotecnia de una cadena de cines desea medir el grado de influencia que llega a tener el medio masivo utilizado para su campaña de publicidad en los consumidores. Esta influencia se maneja por medio del "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Los tipos de medios manejados son dos: Televisión y Revistas.

Una vez lanzada la campaña en la TELEVISIÓN y en REVISTAS, se tomó una muestra aleatoria de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo de 6 potenciales clientes que leyeron la campaña en la Televisión y 7 que la vieron en



Revistas y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados fueron los siguientes:

Medio utilizado					
Televisión			Revistas		
Marca	Slogan	Logotipo	Marca	Slogan	Logotipo
9	10	8	4	6	3
8	9	9	5	4	6
7	8	9	7	9	8
9	9	9	3	8	4
8	8	6	6	7	5
			8	6	8
			7	7	6

a) Existe evidencia suficiente que indique que el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines es el mismo en ambos medios utilizados?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.2.1 EJERCICIOS DE **REFUERZO**

EJERCICIO DE **REFUERZO**

3.2.2.1

DIFERENCIA **DE MEDIAS MULTIVARIADA**

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**.

3.2.2.1.1. El Director de Mercadotecnia de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia que llega a tener la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares), el tamaño del lote (en pies cuadrados) y número de recámaras. Para este experimento se manejaron dos de los niveles de la ubicación geográfica: **Norte y Sur**. Se tomaron lecturas en 5 casas unifamiliares ubicadas en la zona Norte y 7 ubicadas en la zona Sur y los resultados fueron los siguientes:

Considere la siguiente muestra extraída de una población normal bivariada denominada **"BIENES1"** donde X_{11}= Valor de avalúo en la zona Norte, X_{12}= Tamaño del lote en la zona norte, X_{31}= Número de recámaras en la zona Norte, X_{12}= Valor del avalúo en la zona Sur, X_{22}= Tamaño del Lote en la zona Sur y X_{32}= Número de recámaras en la zona Sur:

Nota: El subíndice indica número de variable y zona. Ejemplo: X_{11} indica Variable 1 Valor en zona 1 Norte

ZONA NORTE			**ZONA SUR**		
X_{11}	**X_{21}**	**X_{31}**	**X_{12}**	**X_{22}**	**X_{32}**
Valor	**Tamaño**	**Número**	**Valor**	**Tamaño**	**Número**
265	26	2	181	7	4
245	6	5	220	9	4
210	16	2	190	7	4
239	9	3	160	6	2
299	11	4	176	6	4
			181	7	3
			185	6	3

Es importante mencionar que **pueden existir diferencias** en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

a) Existe evidencia suficiente que indique que la influencia que llega a tener la ubicación geográfica de las casas unifamiliares de las zonas Norte y Sur sobre el valor de las mismas, el tamaño del lote y el número de recámaras es el mismo en ambas ubicaciones?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.2.1.2. El Director de Mercadotecnia de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia que llega a tener la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares), el tamaño del lote (en pies cuadrados) y número de recámaras. Para este experimento se manejaron dos de los niveles de la ubicación geográfica: **Norte y Este**. Se tomaron lecturas en 5 casas unifamiliares ubicadas en la zona Norte y 7 ubicadas en la zona Este y los resultados fueron los siguientes:

Considere la siguiente muestra extraída de una población normal bivariada denominada **"BIENES1"** donde X_{11}= Valor de avalúo en la zona Norte, X_{12}= Tamaño del lote en la zona Norte, X_{31}= Número de recámaras en la zona Norte, X_{12}= Valor del avalúo en la zona Este, X_{22}= Tamaño del Lote en la zona Este y X_{32}= Número de recámaras en la zona Este:

Nota: El subíndice indica número de variable y zona. Ejemplo: X_{11} indica Variable 1 Valor en zona 1 Norte

ZONA NORTE			**ZONA ESTE**		
X_{11}	**X_{21}**	**X_{31}**	**X_{12}**	**X_{22}**	**X_{32}**
Valor	**Tamaño**	**Número**	**Valor**	**Tamaño**	**Número**
265	26	2	180	6	4

245	6	5	152	6	4
210	16	2	168	6	3
239	9	3	180	11	4
299	11	4	163	6	3
			224	6	4
			160	6	4

a) Existe evidencia suficiente que indique que la influencia que llega a tener la ubicación geográfica de las casas unifamiliares de las zonas Norte y Este sobre el valor de las mismas, el tamaño del lote y el número de recámaras es el mismo en ambas ubicaciones?. Pruébelo.

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.2.1.3. El Director de Mercadotecnia de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia que llega a tener la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares), el tamaño del lote (en pies cuadrados) y número de recámaras. Para este experimento se manejaron dos de los niveles de la ubicación geográfica: **Sur y Este**. Se tomaron lecturas en 5 casas unifamiliares ubicadas en la zona Sur y 7 ubicadas en la zona Este y los resultados fueron los siguientes:

Considere la siguiente muestra extraída de una población normal bivariada denominada **"BIENES1"** donde X_{11}= Valor de avalúo en la zona Sur, X_{12}= Tamaño del lote en la zona Sur, X_{31}= Número de recámaras en la zona Sur, X_{12}= Valor del avalúo en la zona Este, X_{22}= Tamaño del Lote en la zona Este y X_{32}= Número de recámaras en la zona Este:

Nota: El subíndice indica número de variable y zona. Ejemplo: X_{11} indica Variable 1 Valor en zona 1 Sur

ZONA SUR			**ZONA ESTE**		
X_{11}	X_{21}	X_{31}	X_{12}	X_{22}	X_{32}
Valor	**Tamaño**	**Número**	**Valor**	**Tamaño**	**Número**
190	7	4	180	6	4
160	6	2	152	6	4
176	6	4	168	6	3
181	7	3	180	11	4
185	6	3	163	6	3
			224	6	4
			160	6	4

a) Existe evidencia suficiente que indique que la influencia que llega a tener la ubicación geográfica de las casas unifamiliares de las zonas Sur y Este sobre el valor de las mismas, el tamaño del lote y el número de recámaras es el mismo en ambas ubicaciones?. Pruébelo.

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

3.2.3 ANÁLISIS DE VARIANZA MULTIVARIADA (MANOVA)

CONCEPTOS BÁSICOS MANOVA

ANALISIS DE VARIANZA MULTIVARIADO (MANOVA).

Es considerado como el caso general del análisis de varianza y permite pronosticar simultáneamente a dos o más variables dependientes, las cuales representan un factor. Esta técnica se conoce como MANOVA y maneja variables con las mismas características que el ANOVA.

MANOVA DE UN FACTOR (Caso balanceado)

Las variables que maneja esta técnica son una independiente y dos o mas dependientes.

En esta técnica, *un caso balanceado* significa que en un nivel de clasificación, las variables dependientes tienen el mismo número de observaciones.

Para que los procedimientos de los contrastes multivariantes del análisis multivariado sean válidos, se deberán cumplir los siguientes supuestos:

1. Las observaciones deben ser independientes.
2. Las matrices de varianzas-covarianzas deben ser iguales para todos los grupos del tratamiento.
3. El conjunto de las *p* variables dependientes debe seguir una distribución normal multivariante, es decir, cualquier combinación lineal de las variables dependientes debe seguir una distribución normal.
4. Debe existir linealidad y muy poca o nula multicolinealidad entre variables dependientes.

El MANOVA proporciona varios criterios con los que valorar las diferencias multivariantes entre los grupos, uno de ellos es el *Lambda de Wilks*

(conocido también como estadístico U).

Este estadístico compara si los grupos son de algún modo diferentes sin estar afectados por el hecho de que los grupos difieran en al menos una combinación lineal de las variables dependientes.

Lambda toma valores entre 0 y 1 de forma que cuanto más se acerca al cero, mayor es la dispersión entre los grupos y por lo tanto es mayor la significación.

El estadístico lambda de Wilks (£) se obtiene de la siguiente manera:

$$£ = \frac{|Error|}{|TR + Error|}$$

Donde |Error| es el determinante de la matriz de dispersión multivariante entre los grupos, y |TR+Error| es el determinante de la suma de TR + Error, donde TR es la matriz de dispersión multivariante entre los grupos o tratamientos.

Para obtener los valores de las matrices de varianzas–covarianzas del Error y del Total se deberán utilizar las siguientes fórmulas matriciales:

- Para la matriz Total:
  TOTAL = SCPC = D'D

- Para la matriz Error:
  ERROR = $\Sigma SCPC_i$

donde:

SCPC = Matriz de sumas de cuadrados y productos cruzados

D = Matriz de desviaciones del total de parejas de datos de las VD.

D' = Transpuesta de la Matriz de desviaciones del total de parejas de las

VD.

$\Sigma SCPC_i$ = Sumatoria de las matrices SCPC de cada nivel i del factor.

El valor de lambda se debe convertir a un valor *F,* el cual dependerá del número de variables dependientes y del número de niveles en el factor mediante la siguiente tabla de equivalencias:

Caso	**No. Var. Dep.**	**No. Niv. Factor**	**Equivalencia**
1	p=1	g≥2	$F_{calculada}=(\Sigma n_i - g/g - 1)(1 - £/£)$; $F_{crìtica}=[\alpha,(g-1),(\sum ni - g)]$
2	p=2	g≥2	$F_{calculada}=(\Sigma n_i - g - 1/g - 1)(1 - \sqrt{£}/\sqrt{£})$; $F_{crìtica}=[\alpha,2(g-1),2(\sum ni - g - 1)]$
3	p≥1	g=2	$F_{calculada}=(\Sigma n_i - p - 1/p)(1 - £/£)$; $F_{crìtica}=[\alpha,p,(\sum ni - p - 1)]$
4	p≥1	g=3	$F_{calculada}=(\Sigma n_i - p - 2/p)(1 - \sqrt{£}/\sqrt{£})$; $F_{crìtica}=[\alpha,2p,2(\sum ni - p - 2)]$

Donde:

n_i: número de pares de datos para cada uno de los niveles del factor manejado en el experimento.

p: número de variables dependientes.

g: número de niveles del factor.

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre

COMPARACIONES MÚLTIPLES: EL MÉTODO T DE TUKEY EN EL DISEÑO MANOVA

medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.

Con la finalidad de determinar **cuáles de las *k* medias son significativamente diferentes** de las otras podemos utilizar el procedimiento de **Tukey**. Este método es un ejemplo de un procedimiento de comparación ***post hoc*** (o ***a posteriori),*** pues las hipótesis de interés son formuladas ***después*** de que los datos han sido inspeccionados.

Para usar el procedimiento de **Tukey,** simplemente se ordenan en forma descendente las medias de los tratamientos y se comparan las diferencias observadas entre cada par de promedios con el valor correspondiente al **rango ó alcance crítico**. Si $\left|\overline{X}_{i.}-\overline{X}_{i.'}\right| \geq$ **rango ó alcance crítico**, se concluye que las medias poblacionales $\mu_{i.}$ y $\mu_{j.}$ son diferentes. El **rango ó alcance crítico** se obtiene entonces de la cantidad dada en la ecuación siguiente general:

$$\boldsymbol{rango}\ \text{ó}\ \boldsymbol{alcance\ crítico} = \boldsymbol{q} \propto, (k.N-k)\sqrt{\frac{CME}{nh}}$$

Si en uno ó más grupos hay tamaños de muestra desiguales, se reemplaza ***n*** de la ecuación anterior por la llamada media armónica:

$$\boldsymbol{n_h(media\ armónica)} = \frac{\boldsymbol{k}}{\sum_{i=1}^{k}\frac{\boldsymbol{1}}{\boldsymbol{n_i}}}$$

Con el **método de Tukey** se puede establecer también un conjunto de intervalos de confianza estimados simultáneamente para las **verdaderas diferencias entre cada par de medias**. Lo anterior se logra sumando y restando el alcance o rango crítico a las diferencias en cada par de medias muestrales de la siguiente manera general

$$(\overline{\boldsymbol{X}}_{\boldsymbol{i.}}-\overline{\boldsymbol{X}}_{\boldsymbol{i.}}^{'})-\boldsymbol{q} \propto, (k.N-k)\sqrt{\frac{CME}{nh}} \leq (\boldsymbol{\mu_{i.}}-\boldsymbol{\mu_{i.}^{'}}) \leq (\overline{\boldsymbol{X}}_{\boldsymbol{i.}}-\overline{\boldsymbol{X}}_{\boldsymbol{i.}}^{'})+\boldsymbol{q} \propto, (k.N-k)\sqrt{\frac{CME}{nh}}$$

Si en uno ó más grupos hay tamaños de muestra desiguales, se reemplaza ***n*** de la ecuación anterior por la llamada media armónica para cada comparación por parejas de las medias de muestra:

$$\boldsymbol{n_h(media\ armónica)} = \frac{\boldsymbol{k(grupos\ comparados)}}{\sum_{i=1}^{k}\frac{\boldsymbol{1}}{\boldsymbol{n_{i(grupos\ comparados)}}}}$$

Ahora bien, de acuerdo a la tabla de equivalencias anteriores el rango

crítico quedaría de la siguiente manera:

Caso 1

$$\boldsymbol{rango}\ \text{ó}\ \boldsymbol{alcance\ crítico} = \boldsymbol{q_{\alpha,k,(\sum n_i - g)}}\sqrt{\frac{\boldsymbol{SCE/(\sum n_i - g)}}{\boldsymbol{nh}}}$$

Caso 2

$$\boldsymbol{rango}\ \text{ó}\ \boldsymbol{alcance\ crítico} = \boldsymbol{q_{\alpha,k,2(\sum n_i - g - 1)}}\sqrt{\frac{\boldsymbol{SCE/2(\sum n_i - g - 1)}}{\boldsymbol{nh}}}$$

Caso 3

$$\boldsymbol{rango}\ \text{ó}\ \boldsymbol{alcance\ crítico} = \boldsymbol{q_{\alpha,k,(\sum n_i - p - 1)}}\sqrt{\frac{\boldsymbol{SCE/(\sum n_i - p - 1)}}{\boldsymbol{nh}}}$$

Caso 4

$$\boldsymbol{rango}\ \text{ó}\ \boldsymbol{alcance\ crítico} = \boldsymbol{q_{\alpha,k,2(\sum n_i - p - 2)}}\sqrt{\frac{\boldsymbol{SCE/2(\sum n_i - p - 2)}}{\boldsymbol{nh}}}$$

Si en uno ó más grupos hay tamaños de muestra desiguales, se reemplaza ***n*** de la ecuación anterior por la llamada media armónica:

$$\boldsymbol{n_h(media\ armónica)} = \frac{\boldsymbol{k}}{\boldsymbol{\sum_{i=1}^{k}\frac{1}{n_i}}}$$

Con el **método de Tukey** se puede establecer también un conjunto de intervalos de confianza estimados simultáneamente para las **verdaderas diferencias entre cada par de medias**. Lo anterior se logra sumando y restando el alcance o rango crítico a las diferencias en cada par de medias muestrales.

Asimismo de acuerdo a la tabla de equivalencias anteriores los intervalos de confianza quedarían de la siguiente manera:

Caso 1

$$(\boldsymbol{\bar{X}_{i.} - \bar{X}'_{i.}}) - \boldsymbol{q_{\alpha,k,(\sum n_i - g)}}\sqrt{\frac{\boldsymbol{SCE/(\sum n_i - g)}}{\boldsymbol{nh}}} \leq (\boldsymbol{\mu_{i.} - \mu'_{i.}})$$

$$\leq (\boldsymbol{\bar{X}_{i.} - \bar{X}'_{i.}}) + \boldsymbol{q_{\alpha,k,(\sum n_i - g)}}\sqrt{\frac{\boldsymbol{SCE/(\sum n_i - g)}}{\boldsymbol{nh}}}$$

Caso 2

$$(\bar{X}_{i.} - \bar{X}'_{i.}) - q_{\alpha,k,2(\sum n_i - g - 1)} \sqrt{\frac{SCE/2(\sum n_i - g - 1)}{nh}} \leq (\mu_{i.} - \mu'_{i.}) \leq (\bar{X}_{i.} - \bar{X}'_{i.}) + q_{\alpha,k,2(\sum n_i - g - 1)} \sqrt{\frac{SCE/2(\sum n_i - g - 1)}{nh}}$$

Caso 3

$$(\bar{X}_{i.} - \bar{X}'_{i.}) - q_{\alpha,k,(\sum n_i - p - 1)} \sqrt{\frac{SCE/(\sum n_i - p - 1)}{nh}} \leq (\mu_{i.} - \mu'_{i.}) \leq (\bar{X}_{i.} - \bar{X}'_{i.}) + q_{\alpha,k,(\sum n_i - p - 1)} \sqrt{\frac{SCE/(\sum n_i - p - 1)}{nh}}$$

Caso 4

$$(\bar{X}_{i.} - \bar{X}'_{i.}) - q_{\alpha,k,2(\sum n_i - p - 2)} \sqrt{\frac{SCE/2(\sum n_i - p - 2)}{nh}} \leq (\mu_{i.} - \mu'_{i.}) \leq (\bar{X}_{i.} - \bar{X}'_{i.}) + q_{\alpha,k,2(\sum n_i - p - 2)} \sqrt{\frac{SCE/2(\sum n_i - p - 2)}{nh}}$$

Si en uno ó más grupos hay tamaños de muestra desiguales, se reemplaza $\boldsymbol{n}$ de la ecuación anterior por la llamada media armónica para cada comparación por parejas de las medias de muestra:

$$n_h(\boldsymbol{media\ armónica}) = \frac{k(\boldsymbol{grupos\ comparados})}{\sum_{i=1}^{k} \frac{1}{n_{i(\boldsymbol{grupos\ comparados})}}}$$

3.2.3.1 EJEMPLO ILUSTRATIVO

EJEMPLO ILUSTRATIVO

3.2.3.1

MANOVA

El Director de Mercadotecnia de una cadena de cines desea medir el grado de influencia que llega a tener el medio masivo utilizado para su campaña de publicidad en los consumidores. Esta influencia se maneja por medio del "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Los tipos de medios manejados son tres: Prensa escrita, Televisión y Revistas.

Una vez lanzada la campaña en los tres medios, se tomó una muestra aleatoria de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del slogan y X_3= Evocación del logotipo de 5 potenciales clientes que leyeron la campaña en la prensa escrita y 6 que la vieron en Televisión y 7 que la leyeron en revistas. Se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados fueron los siguientes:

Medio utilizado								
Prensa escrita			Televisión			Revistas		
Marca	Slogan	Logotipo	Marca	Slogan	Logotipo	Marca	Slogan	Logotipo
4	8	3	9	10	8	4	6	3
6	7	3	8	9	9	5	4	6
5	9	4	7	8	9	7	9	8
5	9	6	9	9	9	3	8	4
5	7	4	8	8	6	6	7	5
			10	7	10	8	6	8
						7	7	6

a) ¿Existe evidencia suficiente que indique que el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines es el mismo en los tres medios utilizados?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable independiente para ver si el medio utilizado influye sobre el "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines".
c) Realice las pruebas Post-hoc de Tukey para cada variable independiente y construya los intervalos de confianza que indique, en su caso, cual medio utilizado logro mayor "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines" y por cuanto es la diferencia.

Resolución del inciso a)

Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

Prueba de hipótesis:

1. Juego de Hipótesis

$$H_0 = \begin{bmatrix}\mu_{11}\\ \mu_{21}\\ \mu_{31}\end{bmatrix} = \begin{bmatrix}\mu_{12}\\ \mu_{22}\\ \mu_{32}\end{bmatrix} = \begin{bmatrix}\mu_{13}\\ \mu_{23}\\ \mu_{33}\end{bmatrix}$$

2. Estadístico de prueba

Como No. de Variables Dependientes : p≥1 y

No. de Niveles del Factor: g=3

Usamos la siguiente fórmula (Caso 4):

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Donde\ £ = \frac{|error|}{|tratamiento|} = \frac{|SCE|}{|SCT|}$$

$$SCE = \sum_{i=1}^{3} SCEi = \sum_{i=1}^{3} D_i' D_i$$

$$SCT = D'D$$

Para SCE:

$$D_1 = \begin{bmatrix} 4 & 8 & 3 \\ 6 & 7 & 3 \\ 5 & 9 & 4 \\ 5 & 9 & 6 \\ 5 & 7 & 4 \end{bmatrix} - \begin{bmatrix} 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 0 \end{bmatrix}$$

$$D'_1 D_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & -1 \\ -1 & -1 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix}$$

$$D^2 = \begin{bmatrix} 9 & 10 & 8 \\ 8 & 9 & 9 \\ 7 & 8 & 9 \\ 9 & 9 & 9 \\ 8 & 8 & 6 \\ 10 & 7 & 10 \end{bmatrix} - \begin{bmatrix} 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \end{bmatrix} = \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \\ -1.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & -2.5 \\ 1.5 & -1.5 & 1.5 \end{bmatrix}$$

$$D'_2 D_2 = \begin{bmatrix} 0.5 & -0.5 & -1.5 & 0.5 & -0.5 & 1.5 \\ 1.5 & 0.5 & -0.5 & 0.5 & -0.5 & -1.5 \\ -0.5 & 0.5 & 0.5 & 0.5 & -2.5 & 1.5 \end{bmatrix} \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \\ -1.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & -2.5 \\ 1.5 & -1.5 & 1.5 \end{bmatrix}$$

$$= \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix}$$

$$D_3 = \begin{bmatrix} 4 & 6 & 3 \\ 5 & 4 & 6 \\ 7 & 9 & 8 \\ 3 & 8 & 4 \\ 6 & 7 & 5 \\ 8 & 6 & 8 \\ 7 & 7 & 6 \end{bmatrix} \begin{bmatrix} 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \end{bmatrix} = \begin{bmatrix} -1.71 & -0.71 & -2.71 \\ -0.71 & -2.71 & 0.29 \\ 1.29 & 2.29 & 2.29 \\ -2.71 & 1.29 & -1.71 \\ 0.29 & 0.29 & -0.71 \\ 2.29 & -0.71 & 2.29 \\ 1.29 & 0.29 & 0.29 \end{bmatrix}$$

$D'_3 D_3$

$$= \begin{bmatrix} -1.71 & -0.71 & 1.29 & -2.71 & 0.29 & 2.29 & 1.29 \\ 0.71 & -2.71 & 2.29 & 1.29 & 0.29 & -0.71 & 0.29 \\ -2.71 & 0.29 & 2.29 & -1.71 & 0.71 & 2.29 & 0.29 \end{bmatrix} \begin{bmatrix} -1.71 & -0.71 & -2.71 \\ -0.71 & -2.71 & 0.29 \\ 1.29 & 2.29 & 2.29 \\ -2.71 & 1.29 & -1.71 \\ 0.29 & 0.29 & -0.71 \\ 2.29 & -0.71 & 2.29 \\ 1.29 & 0.29 & 0.29 \end{bmatrix}$$

$$= \begin{bmatrix} 19.4287 & 1.4287 & 17.4287 \\ 1.4287 & 15.4287 & 2.4287 \\ 17.4287 & 2.4287 & 21.4287 \end{bmatrix}$$

$$SCE = \sum_{i=1}^{3} D'_i D_i$$

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix} + \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix} + \begin{bmatrix} 19.4287 & 1.4287 & 17.4287 \\ 1.4287 & 15.4287 & 2.4287 \\ 17.4287 & 2.4287 & 21.4287 \end{bmatrix}$$

$$SCE = \begin{bmatrix} 26.93 & -0.0713 & 19.9287 \\ -0.0713 & 24.9287 & 3.9287 \\ 19.9287 & 3.9287 & 36.9287 \end{bmatrix}$$

Utilizando la regla de Sarrus:

$$|SCE| = \begin{vmatrix} 26.93 & -0.0713 & 19.9287 \\ -0.0713 & 24.9287 & 3.9287 \\ 19.9287 & 3.9287 & 36.9287 \\ 26.93 & -0.0713 & 19.9287 \\ -0.0713 & 24.9287 & 3.9287 \end{vmatrix} =$$

$$\begin{gathered} (26.93 \times 24.9287 \times 36.9287) + (-0.0713 \times 3.9287 \times \\ 19.9287) + (19.9287 \times -0.0713 \times 3.9287) - \\ (19.9287 \times 24.9287 \times 19.9287) - (3.9287 \times 3.9287 \times \\ 26.93) - (36.9287 \times .0.0713 \times -0.0713) \end{gathered} = 14{,}463.82$$

Para SCT:

$$D\begin{bmatrix} 4 & 8 & 3 \\ 6 & 7 & 3 \\ 5 & 9 & 4 \\ 5 & 9 & 6 \\ 5 & 7 & 4 \\ 9 & 10 & 8 \\ 8 & 9 & 9 \\ 7 & 8 & 9 \\ 9 & 9 & 9 \\ 8 & 8 & 6 \\ 10 & 7 & 10 \\ 4 & 6 & 3 \\ 5 & 4 & 6 \\ 7 & 9 & 8 \\ 3 & 8 & 4 \\ 6 & 7 & 5 \\ 8 & 6 & 8 \\ 7 & 7 & 6 \end{bmatrix} - \begin{bmatrix} 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \\ 6.44 & 7.67 & 6.17 \end{bmatrix} = \begin{bmatrix} -2.44 & 0.33 & -3.17 \\ -0.44 & -0.67 & -3.17 \\ -1.44 & 1.33 & -2.17 \\ -1.44 & 1.33 & -0.17 \\ -1.44 & -0.67 & -2.17 \\ 2.56 & 2.33 & 1.83 \\ 1.56 & 1.33 & 2.83 \\ 0.56 & 0.33 & 2.83 \\ 2.56 & 1.33 & 2.83 \\ 1.56 & 0.33 & -0.17 \\ 3.56 & -0.67 & 3.83 \\ -2.44 & -1.67 & -3.17 \\ -1.44 & -3.67 & -0.17 \\ 0.56 & 1.33 & 1.83 \\ -3.44 & 0.33 & -2.17 \\ -0.44 & -0.67 & -1.17 \\ 1.56 & -1.67 & 1.83 \\ 0.56 & -0.67 & -0.17 \end{bmatrix}$$

$$SCT = D'D = \begin{bmatrix} 2.44 & -0.44 & -1.44 & \dots & 0.56 \\ 0.33 & -0.67 & 1.33 & \dots & -0.67 \\ -3.17 & -3.17 & -2.17 & \dots & -0.17 \end{bmatrix}\begin{bmatrix} 2.44 & 0.33 & -3.17 \\ -0.44 & -0.67 & -3.17 \\ -1.44 & 1.33 & -2.17 \\ \vdots & \vdots & \vdots \\ 0.56 & -0.67 & -0.17 \end{bmatrix}$$

$$= \begin{bmatrix} 66.46 & 12.66 & 66.6 \\ 12.66 & 36 & 15 \\ 66.67 & 15 & 94.48 \end{bmatrix}$$

Utilizando la regla de Sarrus:

$$|SCT| = \begin{vmatrix} 66.46 & 12.66 & 66.6 \\ 12.66 & 36 & 15 \\ 66.67 & 15 & 94.48 \\ 66.46 & 12.66 & 66.6 \\ 12.66 & 36 & 15 \end{vmatrix}$$

$$= \begin{matrix} (66.46 \times 36 \times 94.48) + (12.66 \times 15 \times 66.67) + \\ (66.67 \times 12.66 \times 15) - (66.67 \times 36 \times 66.67) - \\ (15 \times 15 \times 66.46) - (94.48 \times 12.66 \times 12.66) \end{matrix} = 61{,}258$$

$$£ = \frac{|SCE|}{|SCT|} = \frac{14{,}463.82}{61{,}258} = \underline{0.2361}$$

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1 - \sqrt{£})}{\sqrt{£}}$$

$$Fcalc = \frac{(18 - 3 - 2)}{3}\frac{(1 - \sqrt{0.2361})}{\sqrt{0.2361}}$$

$$Fcalc = \frac{(13)}{3}\frac{(1 - 0.4859}{0.4859}$$

$$Fcalc = (4.333)(1.0580)$$

$$Fcalc = 4.5848$$

3. Región de rechazo

$$Fcrit = \left[\propto, 2, p, 2\left(\sum n_1 - p - 2\right)\right]$$

$$Fcrit = \left[0.05{,}2(3){,}2\left(\sum 18 - 3 - 2\right)\right]$$

$$Fcrit = [0.05{,}6{,}2(13)]$$

$$Fcrit = 0.05{,}6{,}26$$

$$F0.05 = 2.47$$

$$F0.01 = 3.59$$

4. Regla de Decisión

Se rechaza H_0 sí $fcal \geq 2.47$

5. Conclusión

Estadista: Como 4.5848 > 2.47 > 3.59 ∴ la prueba es (A.S) y se rechaza con mucha potencia.

Administrativa: Existe evidencia suficiente para decir que estadísticamente que al menos una de las medias de los 3 grupos NO es igual.

Resolución del inciso b)

Prueba de Hipótesis para Marca:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

La hipótesis nula es que el grado de influencia que llega a tener el medio masivo utilizado para la campaña de publicidad en los consumidores es el mismo para recordar la **Marca** sobre esta cadena de cines.

$$H_0: \mu_{11.} = \mu_{12.} = \mu_{13.}$$

La hipótesis alternativa es No todos los medios utilizados para la campaña de publicidad son iguales para recordar la **Marca** sobre esta cadena de cines.

H_1: No todos los medios utilizados para la campaña de publicidad son iguales.

También se puede pensar que la hipótesis alternativa sea "al menos uno de los medios utilizados para la campaña de publicidad para recordar la **Marca** sobre esta cadena de cines es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de la puntuación asignada a la **Marca** con base los medios utilizados para la campaña de publicidad y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los medios utilizado para la campaña de publicidad tiene una diferencia significativa para recordar la **Marca** de esta cadena de cines sobre los otros medios utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Medioss	Puntuación							Total fila	Medias filas
	1	2	3	4	5	6	7		
Prensa	4	6	5	5	5			$X_{1.} = 25$	$\bar{X}_{1.} = 5$
Televisión	9	8	7	9	8	10		$X_{2.} = 51$	$\bar{X}_{2.} = 8.5$
Revistas	4	5	7	3	6	8	7	$X_{3.} = 40$	$\bar{X}_{3.} = 5.71429$
					Totales de filas:			$X_{..} = 116$	$\bar{X}.. = 6.44444$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (4^2 + 6^2 + \cdots + 7^2) - \frac{116^2}{18} = 814 - 747.555 = 66.44444$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{25^2}{5} + \frac{51^2}{6} + \frac{40^2}{7}\right) - \frac{116^2}{18} = 39.51588$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 4^2 + 6^2 + \cdots + 7^2 - (\frac{25^2}{5} + \frac{51^2}{6} + \frac{40^2}{7})$$
$$= 814 - 707.07143 = 26.92857$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamien-tos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 39.51588$	$CMt = {SCt}/{g.l.}$ $= {39.51588}/{2}$ $= 19.75794$	$F_{Calc.} = {CMt}/{CM}$ $= {19.75794}/{1.}$ $= 11.01$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 26.92857$	$CME = {SCE}/{g.l.}$ $= {26.92857}/{15}$ $= 1.79524$	
Total	18-1=17	$SCT = 66.44445$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **11.01**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el medio masivo utilizado para la campaña de publicidad de los consumidores no es el mismo en al menos uno de los tres medios utilizados.

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

Prueba de Hipótesis para Slogan:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el medio masivo utilizado para la campaña de publicidad en los consumidores es el mismo para recordar el **Slogan** sobre esta cadena de cines.

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los medios utilizados para la campaña de publicidad son iguales para recordar el **Slogan** sobre esta cadena de cines.

H_1: No todos los medios utilizados para la campaña de publicidad son iguales.

También se puede pensar que la hipótesis alternativa sea "al menos uno de los medios utilizados para la campaña de publicidad para recordar el **Slogan** sobre esta cadena de cines es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de la puntuación asignada al **Slogan** con base los medios utilizados para la campaña de publicidad y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los medios utilizado para la campaña de publicidad tiene una diferencia significativa para recordar el **Slogan** de esta cadena de cines sobre los otros medios utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Medios	Puntuación							Total fila	Medias filas
	1	2	3	4	5	6	7		
Prensa	8	7	9	9	7			$X_{1.} = 40$	$\bar{X}_{1.} = 8$
Televisión	10	9	8	9	8	7		$X_{2.} = 51$	$\bar{X}_{2.} = 8.5$
Revistas	6	4	9	8	7	6	7	$X_{3.} = 47$	$\bar{X}_{3.} = 6.71429$
					Totales de filas:			$X_{..} = 138$	$\bar{X}.. = 7.66667$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (8^2 + 7^2 + \cdots + 7^2) - \frac{138^2}{18} = 1094 - 1058 = 36$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{40^2}{5} + \frac{51^2}{6} + \frac{47^2}{7}\right) - \frac{138^2}{18} = 11.07$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 8^2 + 7^2 + \cdots + 7^2 - (\frac{40^2}{5} + \frac{51^2}{6} + \frac{47^2}{7})$$
$$= 1094 - 1069.07143 = 24.93$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamien-tos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 11.07$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{11.07}/_{2} = 5.535$	$F_{Calc.} = {}^{CMt}/_{CM}$ $= {}^{5.535}/_{1.662} =$ **3.33030**
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 24.93$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{24.93}/_{15} = 1.662$	
Total	18-1=17	$SCT = 36$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador

y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **3.33**, que es menor al valor crítico de 3.68; por lo tanto la hipótesis nula **NO** se rechaza y llegamos a la conclusión de que todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el medio masivo utilizado para la campaña de publicidad de los consumidores es el mismo en los tres medios utilizados.

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

Prueba de Hipótesis para Logotipo:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el medio masivo utilizado para la campaña de publicidad en los consumidores es el mismo para recordar el **Logotipo** sobre esta cadena de cines.

$$H_0\colon\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los medios utilizados para la campaña de publicidad son iguales para recordar el **Logotipo** sobre esta cadena de cines.

$$H_1\colon \text{No todos los medios utilizados para la campaña de publicidad son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los medios utilizados para la campaña de publicidad para recordar el **Logotipo** sobre esta cadena de cines es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de la puntuación asignada al **Logotipo** con base los medios utilizados para la campaña de publicidad y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los medios utilizado para la campaña de publicidad tiene una diferencia significativa para recordar el **Logotipo** de esta cadena de cines sobre los otros medios utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Medios	Puntuación							Total fila	Medias filas
	1	2	3	4	5	6	7		
Prensa	3	3	4	6	4			$X_{1.} = 20$	$\bar{X}_{1.}$ =4
Televisión	8	9	9	9	6	10		$X_{2.} = 51$	$\bar{X}_{2.} = 8.5$
Revistas	3	6	8	4	5	8	6	$X_{3.} = 40$	$\bar{X}_{3.} = 5.71429$
					Totales de filas:			$X_{..} = 111$	$\bar{X}.. = 6.16667$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (3^2 + 3^2 + \cdots + 6^2) - \frac{111^2}{18} = 779 - 684.5 = 94.5$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{20^2}{5} + \frac{51^2}{6} + \frac{40^2}{7}\right) - \frac{111^2}{18} = 57.57$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = (3^2 + 3^2 + \cdots + 6^2) - (\frac{25^2}{5} + \frac{51^2}{6} + \frac{40^2}{7})$$
$$= 779 - 742.07143 = 36.93$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamientos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 57.57$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{57.57}/_{2}$ $= 28.785$	$F_{Calc.} = {}^{CMt}/_{CM}$ $= {}^{28.785}/_{2.46}$ $= 11.69$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 36.93$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{36.93}/_{15}$ $= 2.462$	
Total	18-1=17	$SCT = 94.50$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **11.69**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el medio masivo utilizado para la campaña de publicidad de los consumidores no es el mismo en al menos uno de los tres medios utilizados.

Resolución del inciso c)

Prueba de Tukey

Intervalo de confianza para Marca (Prensa escrita-Televisión) .Modelo desbalanceado.

$$M_{11} - M_{12} = (\bar{X}_{11} - \bar{X}_{12}) \mp q \propto,(k.N-k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = (5 - 8.5) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = -3.5 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{26.93}{15} = \frac{26.93}{15} = \underline{1.795333}$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{11} - M_{12} = -3.5 \mp 3.67\sqrt{\frac{1.795333}{5.4545}}$$

resultante sea igual al valor que se especifique.

$$M_{11} - M_{12} = -3.5 \mp 3.67\sqrt{0.329147126}$$

$$M_{11} - M_{12} = -3.5 \mp 3.67(0.573713453)$$

$$M_{11} - M_{12} = -3.5 \mp 2.105 \begin{cases} LIC = -5.605 \\ LSC = -1.395 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos podemos decir que estadísticamente el grado de evocación del **Medio televisivo sobre la Marca** es mayor que en el de la **Prensa escrita sobre la Marca** por un mínimo de 1.395 y un máximo de 5.605 puntos.

Intervalo de confianza para Marca (Prensa escrita-Revistas)

$$M_{11} - M_{13} = (\bar{X}_{11} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = (5 - 5.714) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = -0.714 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = 5.8333$$

$$M_{11} - M_{13} = -0.714 \mp 3.67\sqrt{\frac{1.795333}{5.8333}}$$

$$M_{11} - M_{13} = -0.714 \mp 3.67(0.5547)$$

$$M_{11} - M_{13} = -0.714 \mp 2.036 \begin{cases} LIC = -2.750 \\ LSC = 1.322 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de evocación de la **Prensa escrita** sobre la **Marca** es el mismo que el de **Revistas** sobre la **Marca.**

Intervalo de confianza para Marca (Televisión-Revista)

$$M_{12} - M_{13} = (\bar{X}_{12} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique

$$M_{12} - M_{13} = (8.5 - 5.714) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = 2.786 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{6}+\frac{1}{7}} = 6.4615$$

$$M_{12} - M_{13} = 2.786 \mp 3.67\sqrt{\frac{1.7953}{6.4615}}$$

$$M_{12} - M_{13} = 2.786 \mp 3.67(0.527110711)$$

$$M_{12} - M_{13} = 2.786 \mp 1.934 \begin{cases} LIC = 4.720 \\ LSC = 0.851 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de evocación de las **Revistas** sobre la **Marca** es mayor que el de la **Televisión** sobre la **Marca** por un mínimo de 0.851 y un máximo de 4.720 puntos.

Intervalo de confianza para Slogan (Prensa escrita-Televisión)

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es

$$M_{21} - M_{22} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = (8 - 8.5) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = -0.5 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{24.9287}{15)} = \frac{24.9287}{15} = \underline{1.661913}$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

el nivel de confianza del intervalo.

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.

$$M_{21} - M_{22} = 0.5 \mp 3.67\sqrt{\frac{1.661913}{5.4545}}$$

$$M_{21} - M_{22} = 0.5 \mp 3.67(0.551984229)$$

$$M_{21} - M_{22} = -0.5 \mp 2.026\begin{cases}-2.526\\ 1.526\end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de evocación de la **Prensa escrita** sobre **Slogan** es el mismo que el de la **Televisión** sobre **Slogan.**

Intervalo de confianza para Slogan (Prensa escrita – Revista)

$$M_{21} - M_{23} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = (8 - 6.714) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = 1.286 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{21} - M_{23} = 1.286 \mp 3.67\sqrt{\frac{1.661913}{5.8326}}$$

$$M_{21} - M_{23} = 1.286 \mp 3.67(0.533795941)$$

$$M_{21} - M_{23} = 1.286 \mp 1.959\begin{cases}-0.673\\ 3.245\end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente el grado de evocación de la **Prensa escrita** sobre **Slogan** es el mismo que el de las **Revistas** sobre **Slogan**.

Intervalo de confianza para Slogan (Televisión-Revistas)

$$M_{22} - M_{23} = (\bar{X}_{22} - \bar{X}_{23}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = (8.5 - 6.714) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = 1.786 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{22} - M_{23} = 1.786 \mp 3.67\sqrt{\frac{1.661913}{6.4620}}$$

$$M_{22} - M_{23} = 1.786 \mp 3.67(0.507131592)$$

$$M_{22} - M_{23} = 1.786 \mp 1.861\begin{cases} LIC = -0.076 \\ LSC = 3.647 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente el grado de evocación de la **Televisión** sobre **Slogan** es el mismo que el de las **Revistas** sobre **Slogan.**

Intervalo de confianza para Logotipo (Prensa- Televisión)

$$M_{31} - M_{32} = (\bar{X}_{31} - \bar{X}_{32}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = (4 - 8.5) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = -4.5 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N - k} = \frac{SCE}{18 - 3} = \frac{36.9287}{15} = 2.461913$$

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{31} - M_{32} = -4.5 \mp 3.67\sqrt{\frac{2.461913}{5.4545}}$$

$$M_{31} - M_{32} = -4.5 \mp 3.67(0.671829203)$$

$$M_{31} - M_{32} = -4.5 \mp 2.466\begin{cases} LIC = -6.966 \\ LSC = -2.034 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos, podemos decir que estadísticamente el grado de evocación de la **Televisión** sobre el **Logotipo** es mayor que el de la de la **Prensa escrita** sobre el **Logotipo** por un mínimo de 2.034 y un máximo de 6.966 puntos.

Intervalo de confianza para Logotipo (Prensa escrita – Revista)

$$M_{31} - M_{33} = (\bar{X}_{31} - \bar{X}_{33}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = (4 - 5.714) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = 1.714 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{31} - M_{33} = 1.714 \mp 3.67\sqrt{\frac{2.461913}{5.8326}}$$

$$M_{31} - M_{33} = 1.714 \mp 3.67(0.649688612)$$

$$M_{31} - M_{33} = 1.714 \mp 2.384\begin{cases} -0.670 \\ 4.098 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente el grado de evocación de la **Prensa escrita** sobre el **Logotipo** es el mismo que el de la **Televisión** sobre el **Logotipo.**

Intervalo de confianza para Logotipo (Televisión-Revista)

$$M_{32} - M_{33} = (\bar{X}_{32} - \bar{X}_{33}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = (8.5 - 5.714) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = 2.786 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{32} - M_{33} = 2.786 \mp 3.67\sqrt{\frac{2.461913}{6.4620}}$$

$$M_{32} - M_{33} = 2.786 \mp 3.67(0.617238311)$$

$$M_{32} - M_{33} = 2.786 \mp 2.265\begin{cases} LIC = 0.521 \\ LSC = 5.051 \end{cases}$$

Interpretación:

Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de evocación de la **Televisión** sobre el **Logotipo** es mayor que el de las **Revistas** sobre el **Logotipo** por un mínimo de 0.521 y un máximo de 5.051 puntos.

3.2.3.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE
APRENDIZAJE

3.2.3.1

MANOVA

El Director de mercadotecnia de una compañía de Bienes Raíces desea medir el grado de influencia que llega a tener la zona donde se localizan las casas unifamiliares de cierta ciudad en las magnitudes de los atributos de las mismas. Esta influencia es medida a través de tres variables dependientes: Valor comercial de la casa, Tamaño del lote y el Número de habitaciones. Las zonas manejadas son tres: Norte, Sur y Este. El efecto se medirá con base en el grado de influencia que la Zona de la Ciudad ejerce sobre cada una de las tres variables dependientes mediante los siguientes criterios:

1.- Que tanto la Zona de la Ciudad llega a influir sobre el Valor de las casas.

2.- Que tanto la Zona de la Ciudad llega a influir sobre el Tamaño del lote donde están construidas las casas.

3.- Que tanto la Zona de la Ciudad llega a influir sobre el Número de habitaciones que tienen las casas.

Nota importante: Donde se realizó la encuesta el número de habitaciones incluye todos los espacios no solo las recámaras, es decir, sala, comedor, cocina, etc.)

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Valor comercial de la casa (en cientos de dólares) X_2= Tamaño del lote (en cientos de pies cuadrados) y X_3= Número de habitaciones, de 5 propietarios de casas en la zona Norte, 6 en la zona Sur y 7 en la zona Este. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Zona de la ciudad								
Norte			Sur			Este		
Valor	Tamaño	Habita-ciones	Valor	Tamaño	Habita-ciones	Valor	Tamaño	Habita-ciones
1081	367	6	1093	186	4	595	111	4
1214	96	4	931	186	4	905	111	5
1204	226	5	890	130	3	607	111	4
1255	136	4	931	135	3	866	102	5
1053	120	3	911	126	4	758	105	4
			718	130	4	850	102	4
						870	102	4

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de las casas es el mismo en las tres zonas?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si la zona de la ciudad influye sobre las magnitudes promedio de los atributos de las casas.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que zona o zonas los atributos son mayores y en que magnitud.

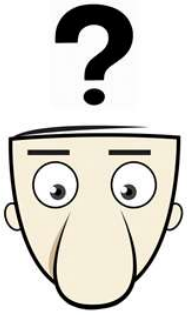
3.2.3.1 EJERCICIO DE AUTOEVALUACIÓN

A continuación se presenta un ejercicio de autoevaluación el cual pone a prueba su comprensión del material anterior. La respuesta a este ejercicio de autoevaluación se encuentra al final del cuaderno de trabajo en el anexo de respuestas. Le recomendamos enfáticamente resolverlo y posteriormente revisar su respuesta como retroalimentación de su aprendizaje

AUTO**EVALUACIÓN**

3.2.3.1

MANOVA



El Director de Industrias México desea medir el grado de influencia que llega a tener el área de trabajo de los empleados de Industrias México sobre los ingresos que reciben "antes de impuestos" (en millones de pesos), los años de escolaridad concluidos y la edad de los empleados. Las áreas de trabajo que se manejaron son tres: Gerencial, Servicios y Obreros.

El efecto se evalúa mediante tres criterios:

1.- Que tanto el Área de Trabajo llega a influir sobre el Ingreso de los empleados de Industrias México.

2.- Que tanto el Área de Trabajo llega a influir sobre la Escolaridad de los empleados de Industrias México.

3.- Que tanto el Área de Trabajo llega a influir sobre la Edad de los empleados de Industrias México.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Ingresos anuales "antes de impuestos" (en millones de pesos), X_2= Años de escolaridad concluidos y X_3= Edad de los empleados (En años cumplidos), de 5 empleados del área Gerencial, 6 del área de Servicios y 7 del área de Obreros. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Área de trabajo								
Gerencial			Servicios			Obreros		
Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad
1.234	16	43	0.542	12	40	0.210	11	20
0.894	15	62	0.326	14	21	0.230	12	18
0.950	16	43	0.380	14	34	0.204	9	53
0.702	16	28	0.436	12	65	0.292	13	27
0.718	14	36	0.404	10	63	0.240	10	22
			0.442	10	40	0.238	12	17
						0.214	11	45

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de los empleados es el mismo en las tres áreas de trabajo?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el área de trabajo influye sobre las magnitudes promedio de los empleados.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que área o áreas de trabajo los atributos son mayores y en que magnitud.

3.2.3.1 MANOVA

EJEMPLO ILUSTRATIVO EN MINITAB 17

EJEMPLO **ILUSTRATIVO I**NTEGRAL EN **MINITAB 17. 3.2.3.1. MANOVA**

El Director de Mercadotecnia de una cadena de cines desea medir el grado de influencia que llega a tener el medio masivo utilizado para su campaña de publicidad en los consumidores. Esta influencia se maneja por medio del "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Los tipos de medios manejados son tres: Prensa escrita, televisión y revistas.

Una vez lanzada la campaña en los tres medios, se tomó una muestra aleatoria de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del slogan y X_3= Evocación del logotipo de 5 potenciales clientes que leyeron la campaña en la prensa escrita y 6 que la vieron en Televisión y 7 que la leyeron en revistas. Se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados fueron los siguientes:

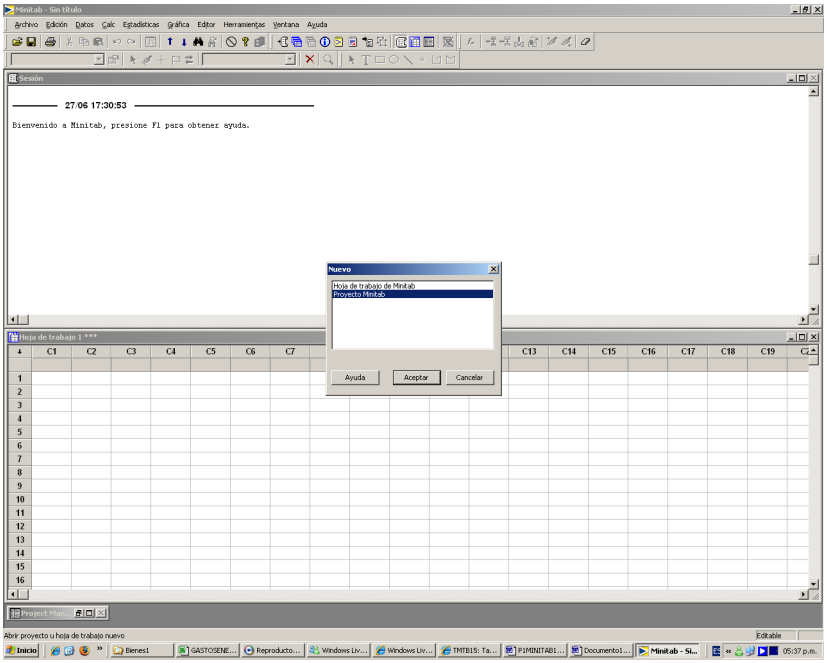
Medio utilizado								
Prensa escrita			Televisión			Revistas		
Marca	Slogan	Logotipo	Marca	Slogan	Logotipo	Marca	Slogan	Logotipo
4	8	3	9	10	8	4	6	3
6	7	3	8	9	9	5	4	6
5	9	4	7	8	9	7	9	8
5	9	6	9	9	9	3	8	4
5	7	4	8	8	6	6	7	5
			10	7	10	8	6	8
						7	7	6

a) ¿Existe evidencia suficiente que indique que el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines es el mismo en los tres medios utilizados?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable independiente para ver si el medio utilizado influye sobre el "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Asimismo realice las pruebas Post-hoc de Tukey para cada variable independiente y construya los intervalos de confianza que indique, en su caso, cual medio utilizado logro mayor "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines" y por cuanto es la diferencia.

Resolución del inciso a)

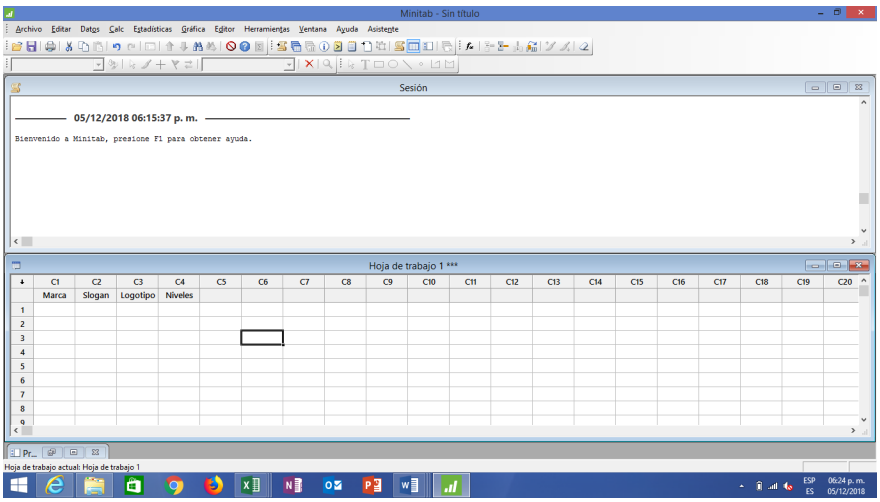
Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

Para crear la base de datos del ejemplo ilustrativo en **MINITAB17**, accede a **MINITAB VERSIÓN 17.** Cuando aparezca la pantalla da un clic en la opción **Archivo** de la barra menú y selecciona **Nuevo**. Aparecerá la ventana **Nuevo. Selecciona la opción Proyecto Minitab y Aceptar**

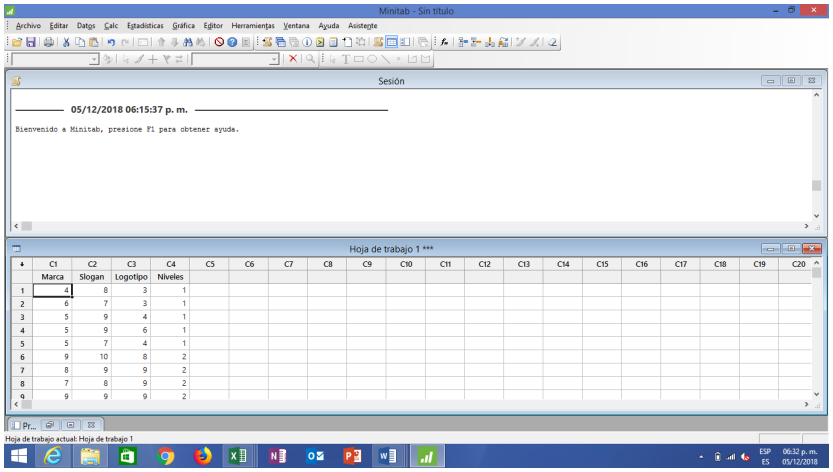


Procede a editar la base de datos de la siguiente manera:

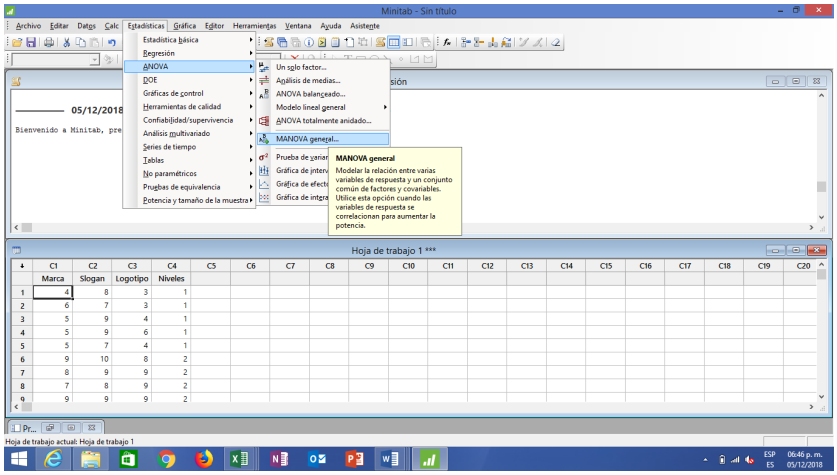
Coloca los nombres de cada **variable (En este caso Marca, Slogan, Logotipo y Niveles**), debajo de cada número de columna, en este caso **C1, C2, C3 y C4**:

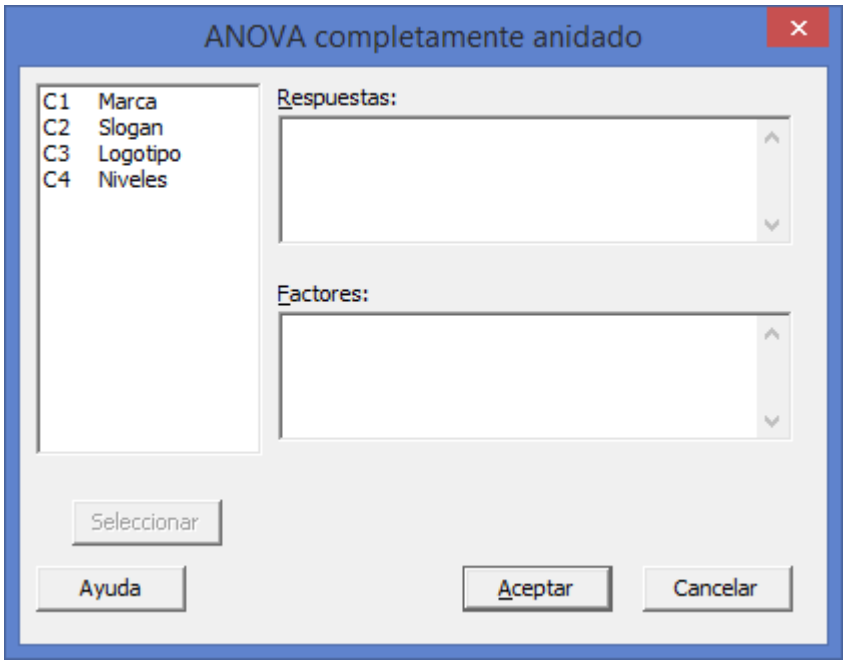


Una vez termines de colocar los nombres de cada **variable (En este caso Marca, Slogan, Logotipo y Niveles**, coloca el cursor en el primer nombre de la variable **(Marca)** y llena los datos de Marca en forma vertical de los tres medios utilizados (Prensa escrita, Televisión y Revistas). Cuando se terminen los 18 datos coloca el cursor en el segundo nombre de la variable (**Slogan**) y llena los 18 datos. Cuando se terminen los 18 datos de Slogan coloca el cursor en el tercer nombre de la variable **(Logotipo**) y llena los 18 datos. Cuando se terminen los 18 datos de Logotipo coloca el cursor en la cuarta columna (**Niveles**) y coloca en forma vertical cinco 1, seis 2 y siete 3. El número 1 corresponde al medio **Prensa escrita,** el 2 corresponde al medio **Televisión** y el 3 corresponde al Medio **Revistas**.



1 Elija **Estadísticas > ANOVA > MANOVA general**.





2 En **Respuestas**, ingrese las 3 variables o columnas numéricas que contengan las variables de respuesta (**Marca, Slogan y Logotipo**).

3 En **Factores**, ingresa la cuarta columna (**Niveles**)

4 Haga clic en **Aceptar**.

Salida de la ventana Sesión

Modelo lineal general: Marca, Slogan, Logotipo vs. Niveles

```
MANOVA para Niveles
s = 2    m = 0.0    n = 5.5

                  Estadística              GL
Criterio            de prueba      F  Núm  Denom      P
De Wilks              0.23606  4.585    6     26  0.003
Lawley-Hotelling      2.36823  4.736    6     24  0.003
De Pillai             0.96882  4.384    6     28  0.003
De Roy                1.91502
```

Interpretación de los resultados

Valor P

Determina si es adecuado rechazar la hipótesis nula en una prueba de hipótesis. Los valores **p** abarcan de 0 a 1. El valor **p** es la probabilidad de obtener una estadística de prueba que sea por lo menos tan extrema como el valor calculado, si la hipótesis nula es verdadera. Antes de realizar cualquier análisis, determine su nivel de significancia (alfa). Un valor comúnmente utilizado es 0.05. Si el valor **p** de una estadística de prueba es menor que su nivel de significancia, rechace la hipótesis nula.

Debido a lo indispensable de su función en las pruebas de hipótesis, los valores **p** se utilizan en muchas áreas de la estadística, entre las que figuran la estadística básica, los modelos lineales, la confiabilidad y los análisis multivariados, entre muchos otros. La clave es comprender lo que representan las hipótesis

La salida de la ventana sesión presenta los resultados de cuatro criterios o pruebas entre ellas el de **Lambda de Wilks** que arroja una Estadística de prueba de **0.23606** ≈ 02361 con una Fcalculada de **4.585** ≈ 4.5848 . El valor de p de 0.003 (nivel de significancia observado) indica que hay suficiente evidencia de que no todas las medias en cada variable son iguales en los tres medios utilizados cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

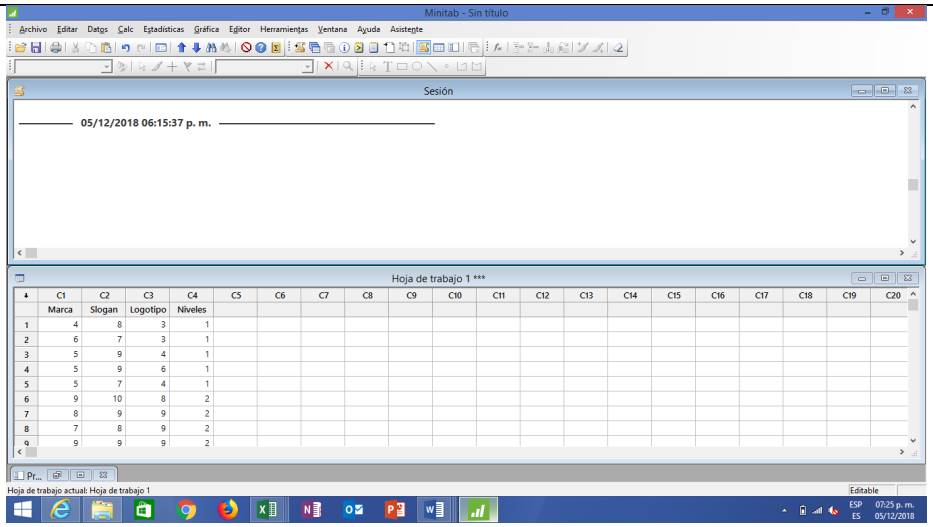
Resolución del inciso b)

1. Prueba ANOVA de Hipótesis para Marca y pruebas Post-hoc de Tukey para Marca:

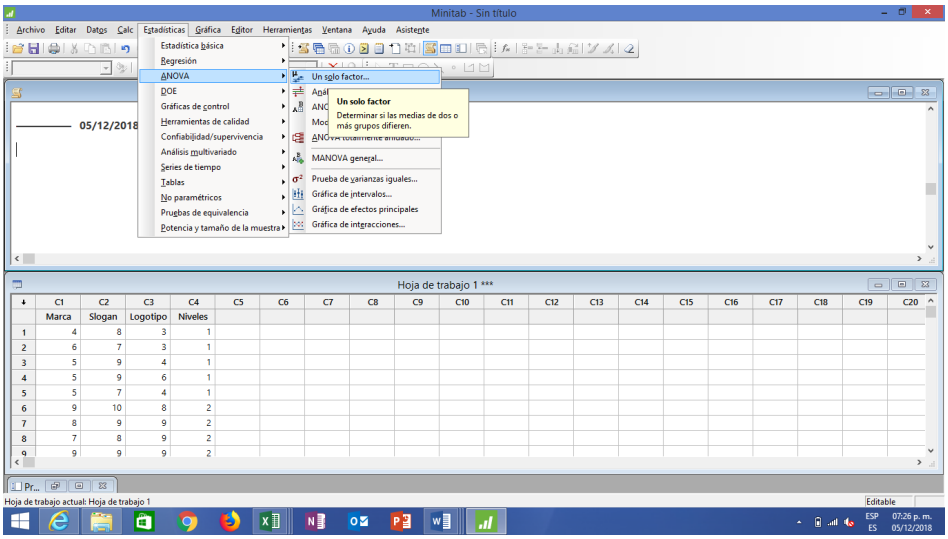
Cuando el número de observaciones en cada tratamiento es extenso y/o existen muchos tratamientos, los cálculos manuales son tediosos. Existen muchos paquetes de software que pueden mostrar los resultados entre ellos Minitab (Versión 17).

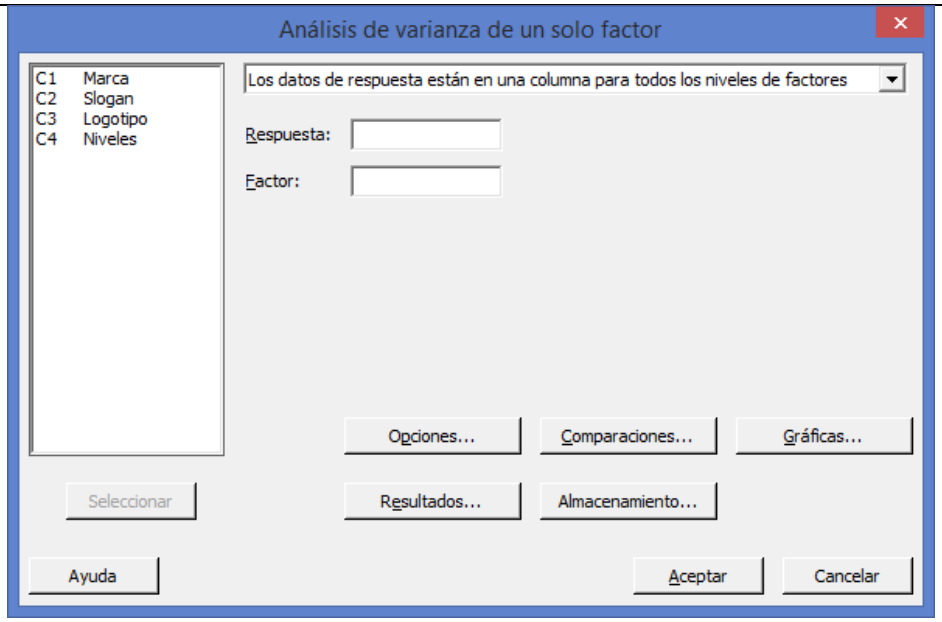
Comenzamos introduciendo los datos en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:

nulas y alternativas en cada prueba y luego utilizar el valor **p** para orientarse en la decisión de rechazar la nulidad.



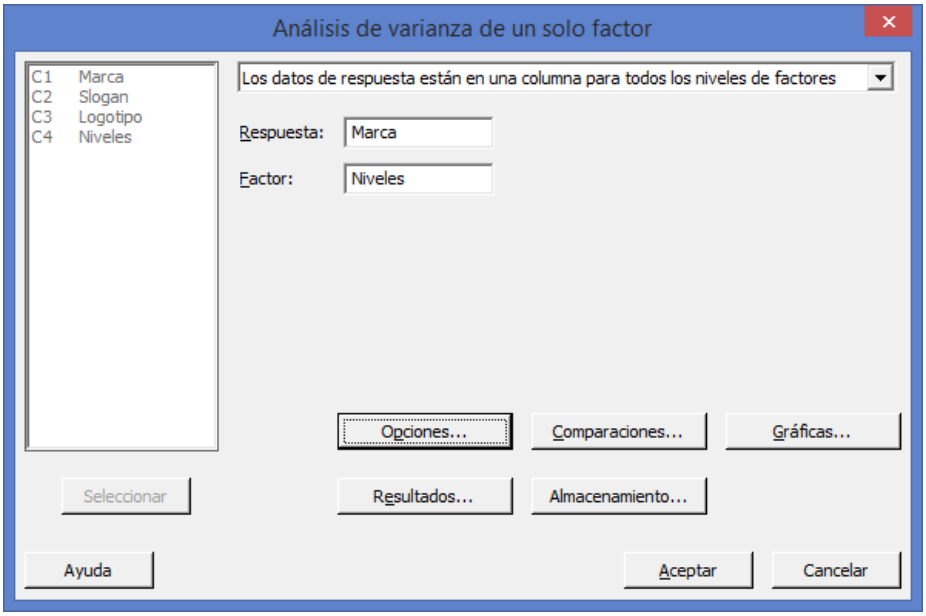
Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**





El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

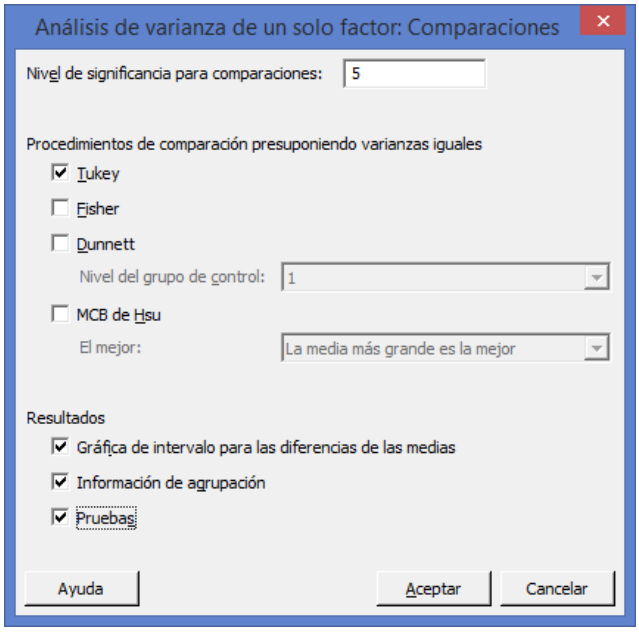
En Respuesta, ingrese **Marca**. En Factor, ingrese **Niveles**.



Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.

Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Marca vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2      39.52     19.758    11.01    0.001
Error    15      26.93      1.795
Total    17      66.44
```

```
Resumen del modelo

                      R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
1.33986   59.47%      54.07%   43.58%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  5.000      0.707  (3.723, 6.277)
2        6  8.500      1.049  (7.334, 9.666)
3        7  5.714      1.799  (4.635, 6.794)

Desv.Est. agrupada = 1.33986
```

Interpretación de los resultados

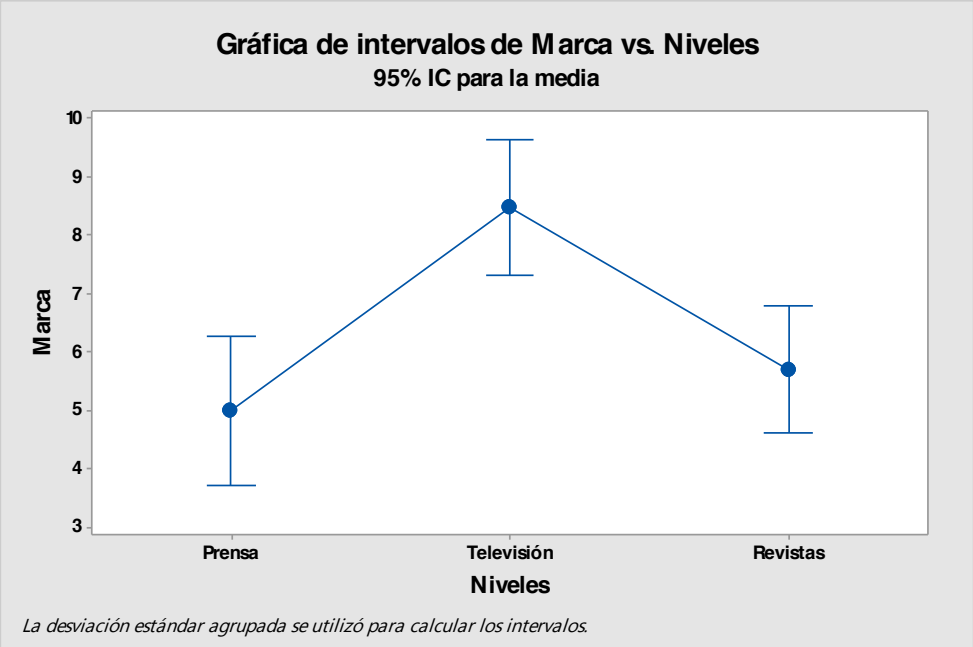
En la Tabla de ANOVA, para **Marca** en **los tres Medios** utilizados el valor de F calculada de 11.01 con un valor p (0.001) indica que hay suficiente evidencia de que no todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Comparaciones en parejas de Tukey

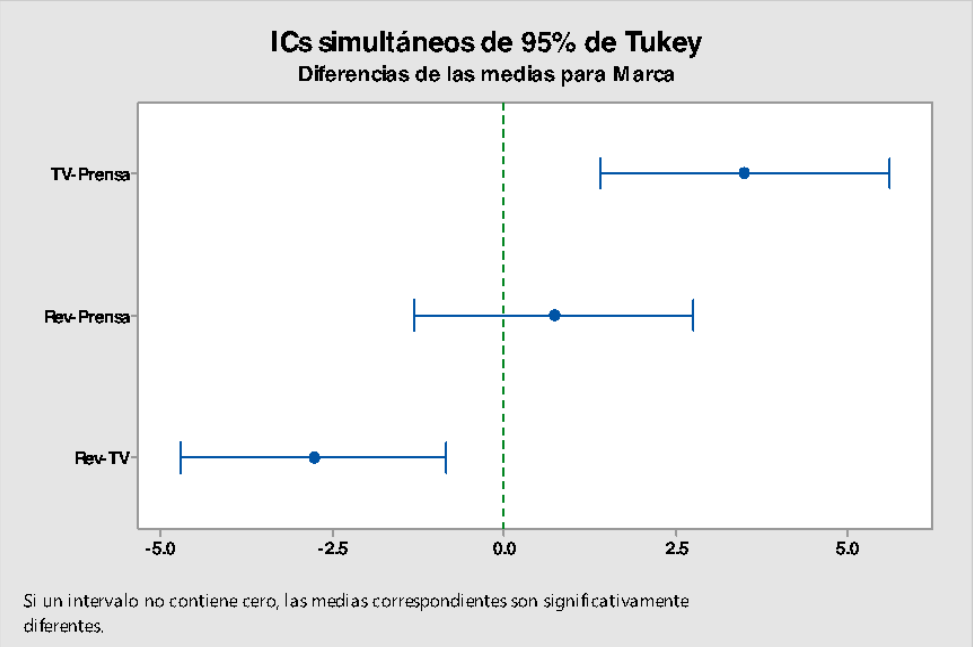
```
Agrupar información utilizando el método de Tukey y una confianza
de 95%

Niveles  N  Media  Agrupación
2        6  8.500  A
3        7  5.714    B
1        5  5.000    B

Las medias que no comparten una letra son significativamente
diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de
Valor p
de niveles      medias  diferencia      IC de 95%     Valor T
ajustado
2 - 1            3.500       0.811  ( 1.395,  5.605)     4.31
0.002
3 - 1            0.714       0.785  (-1.322,  2.750)     0.91
0.642
3 - 2           -2.786       0.745  (-4.720, -0.851)    -3.74
0.005

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Marca vs. Niveles



La desviación estándar agrupada se utilizó para calcular los intervalos.



Si un intervalo no contiene cero, las medias correspondientes son significativamente diferentes.

Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de la **Marca** en la **Televisión** (Nivel 2) con la **Marca** en la **Prensa escrita** (Nivel 1) oscila entre 1.395 y 5.605. Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de evocación de la Marca en la Televisión es mayor que el de la Marca en la Prensa escrita por un mínimo de 1.395 y un máximo de 5.605 puntos.

En cuanto al intervalo de confianza de la **Marca** en las **Revistas** (Nivel 3) con la **Marca** en la **Prensa escrita** (Nivel 1) oscila entre -1.322 y 2.750. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de evocación de la Marca en las Revistas es igual que el de la Marca en la Prensa escrita.

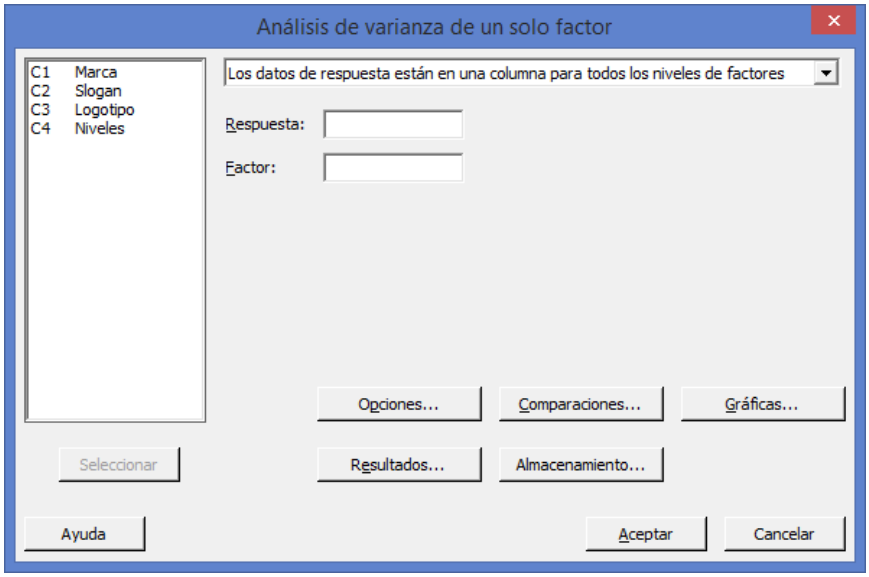
Por último podemos observar que el intervalo de confianza de la **Marca** en las **Revistas** (Nivel 3) con la **Marca** en la **Televisión** (Nivel 2) oscila entre -4.720 y -0.851. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de evocación de la Marca en la Televisión es mayor que el de la Marca en las Revistas por un mínimo de 0.851 y un máximo de 4.720 puntos (en valor absoluto).

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

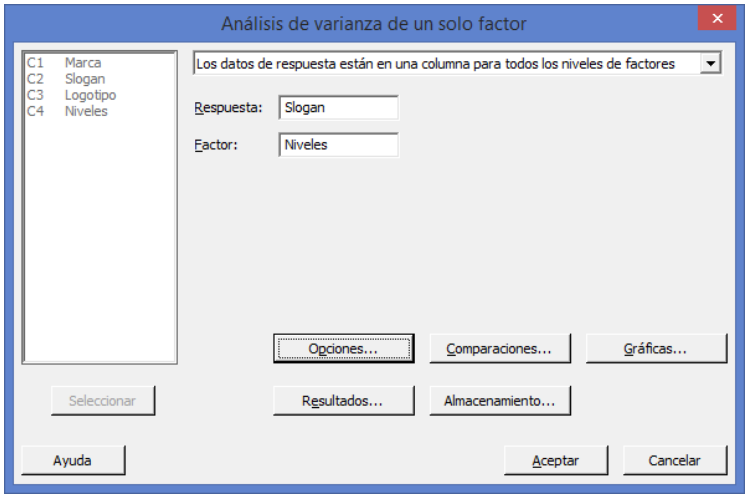
2. Prueba ANOVA de Hipótesis para Slogan y pruebas Post-hoc de Tukey para Slogan:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

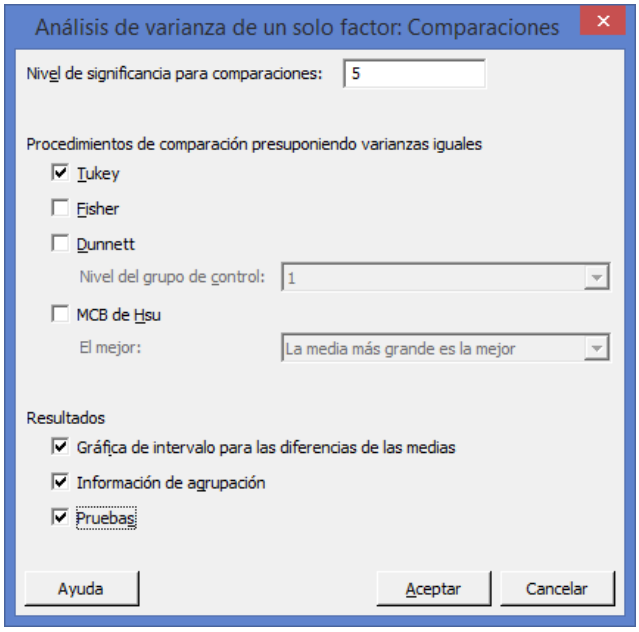


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Slogan**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Slogan vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2      11.07      5.536     3.33    0.064
Error    15      24.93      1.662
Total    17      36.00


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
1.28915   30.75%      21.52%    2.31%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  8.000      1.000  (6.771, 9.229)
2        6  8.500      1.049  (7.378, 9.622)
3        7  6.714      1.604  (5.676, 7.753)

Desv.Est. agrupada = 1.28915
```

Interpretación de los resultados

En la Tabla de ANOVA, para **Slogan** en **los tres Medios** utilizados el valor de F calculada de 3.33 con un valor p (0.064) indica que hay suficiente evidencia de que todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Método de Tukey

Utilizado en el análisis ANOVA para crear intervalos de confianza para todas las diferencias en parejas entre medias de los niveles de factor mientras controla la tasa de error por familia en un nivel que se especifique. Es importante considerar la tasa de error por familia al realizar múltiples comparaciones, porque las posibilidades de cometer un error de tipo I para una serie de comparaciones son mayores que el nivel de significancia para cualquier comparación individual. Para contrarrestar este nivel de significancia más alto, el método de Tukey ajusta el intervalo de confianza para cada intervalo individual, de manera que el nivel de confianza simultáneo resultante sea igual al valor que se especifique.

Comparaciones en parejas de Tukey

Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles	N	Media	Agrupación
2	6	8.500	A
1	5	8.000	A
3	7	6.714	A

Las medias que no comparten una letra son significativamente diferentes.

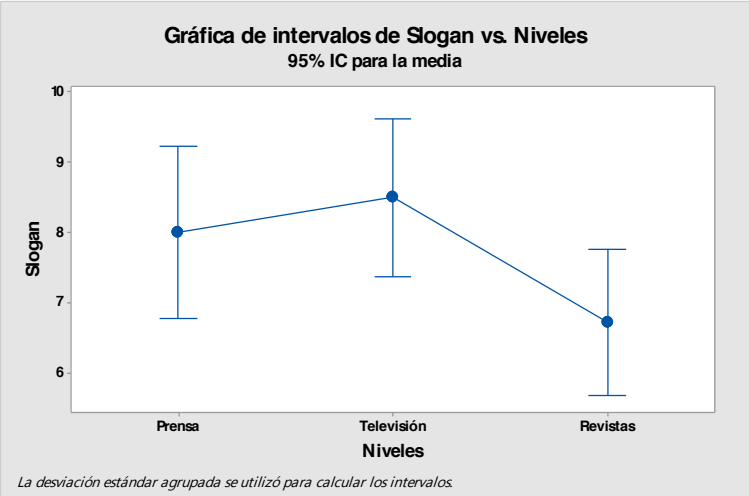
Pruebas simultáneas de Tukey para diferencias de las medias

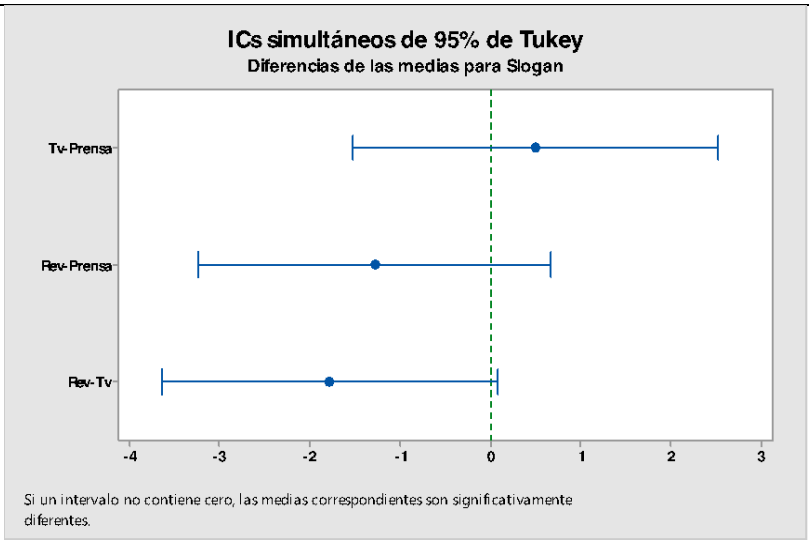
Diferencia de niveles	Diferencia de las medias	EE de diferencia	IC de 95%	Valor T	Valor p ajustado
2 - 1	0.500	0.781	(-1.526, 2.526)	0.64	0.800
3 - 1	-1.286	0.755	(-3.245, 0.673)	-1.70	0.236
3 - 2	-1.786	0.717	(-3.647, 0.076)	-2.49	0.061

Nivel de confianza individual = 97.97%

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Slogan vs. Niveles





Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Slogan** en la **Televisión** (Nivel 2) con la **Slogan** en la **Prensa escrita** (Nivel 1) oscila entre -1.526 y 2.526. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de evocación del Slogan en Televisión es igual que el Slogan en la Prensa escrita.

En cuanto al intervalo de confianza del **Slogan** en las **Revistas** (Nivel 3) con el **Slogan** en la **Prensa escrita** (Nivel 1) oscila entre -3.245 y 0.677. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de evocación de la Slogan en las Revistas es igual que el de la Slogan en la Prensa escrita.

Por último podemos observar que el intervalo de confianza del **Slogan** en las **Revistas** (Nivel 3) con el **Slogan** en la **Televisión** (Nivel 2) oscila entre -3.647 y 0.076. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de evocación de la Slogan en las Revistas es igual que el de la Slogan en la Televisión.

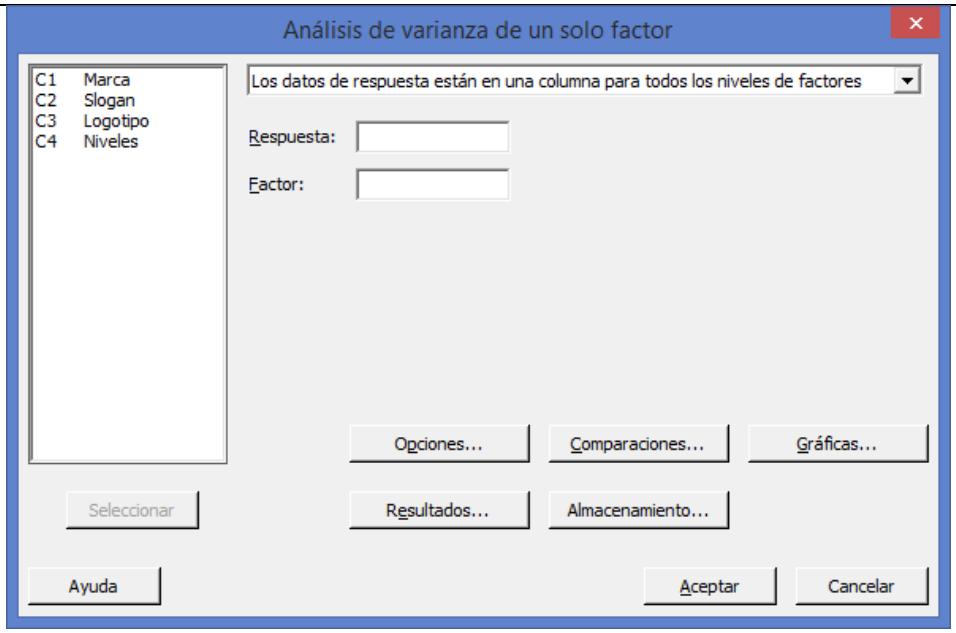
3. Prueba ANOVA de Hipótesis para Slogan y pruebas Post-hoc de Tukey para Logotipo:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

Prueba de hipótesis

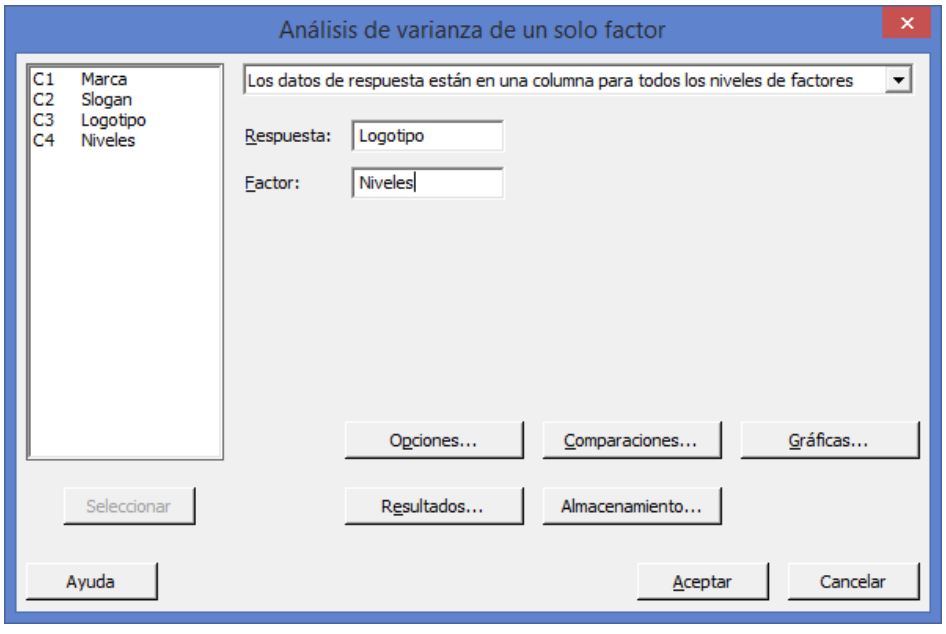
Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan

hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

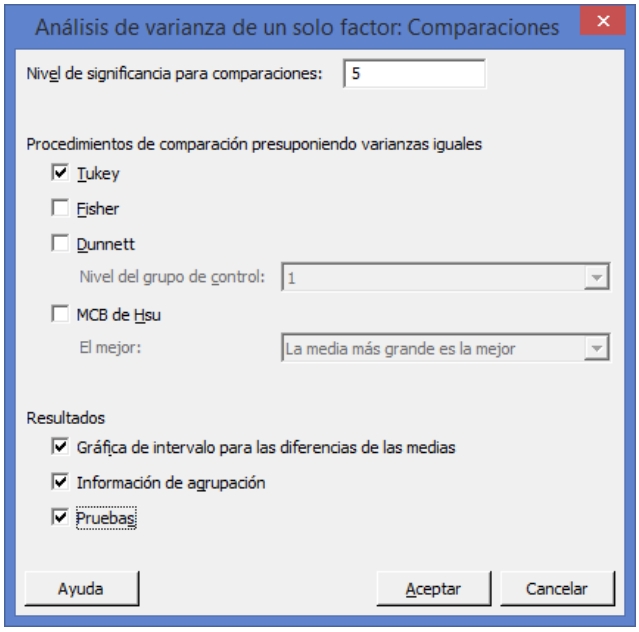


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Logotipo**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Logotipo vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2      57.57     28.786    11.69    0.001
Error    15      36.93      2.462
Total    17      94.50
```

Prueba de hipótesis

Procedimiento que evalúa dos enunciados mutuamente excluyentes sobre una población. Una prueba de hipótesis utiliza datos de muestra para determinar a cuál enunciado respaldan mejor los datos. Estos dos enunciados se denominan hipótesis nula e hipótesis alternativa. Siempre son enunciados sobre los atributos de las poblaciones, tales como el valor de un parámetro, la diferencia entre parámetros correspondientes de múltiples poblaciones o el tipo de distribución que describe mejor a la población

```
Resumen del modelo

                      R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
1.56905   60.92%      55.71%   44.74%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  4.000      1.225  (2.504, 5.496)
2        6  8.500      1.378  (7.135, 9.865)
3        7  5.714      1.890  (4.450, 6.978)

Desv.Est. agrupada = 1.56905
```

Valor P

Determina si es adecuado rechazar la hipótesis nula en una prueba de hipótesis. Los valores **p** abarcan de 0 a 1. El valor **p** es la probabilidad de obtener una estadística de prueba que sea por lo menos tan extrema como el valor calculado, si la hipótesis nula es verdadera. Antes de realizar cualquier análisis, determine su nivel de significancia (alfa). Un valor comúnmente utilizado es 0.05. Si el valor **p** de una estadística de prueba es menor que su nivel de significancia, rechace la hipótesis nula.

Interpretación de los resultados

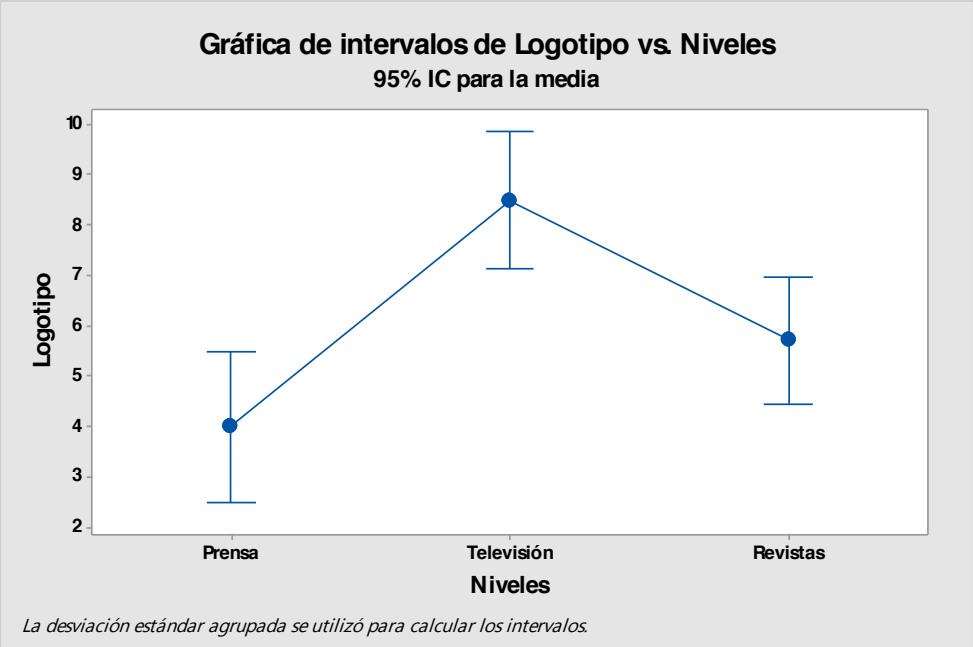
En la Tabla de ANOVA, para **Logotipo** en **los tres Medios** utilizados el valor de F calculada de 11.69 con un valor p (0.001) indica que hay suficiente evidencia de que NO todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Comparaciones en parejas de Tukey

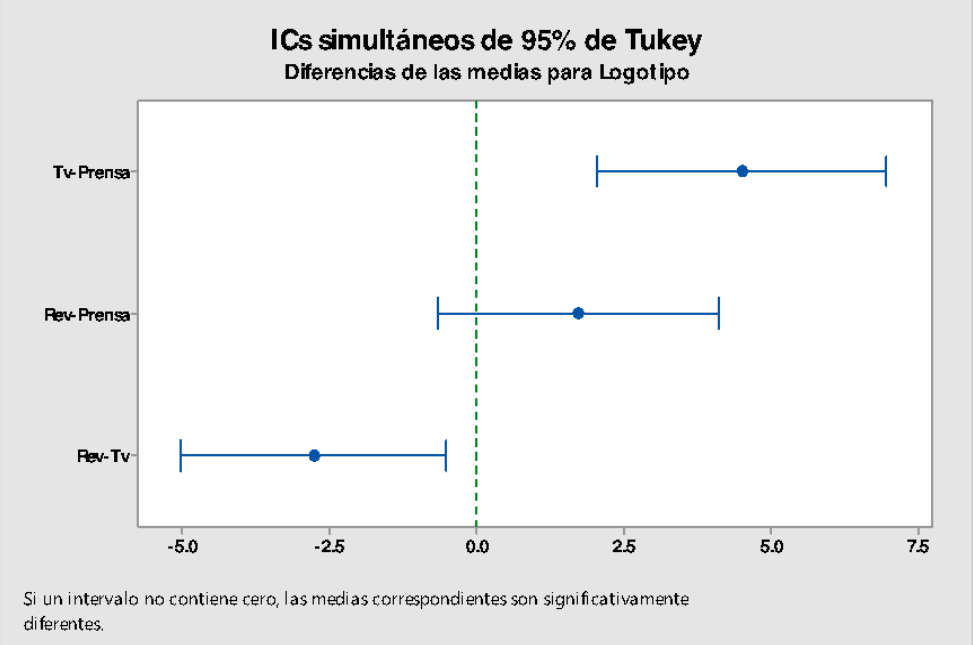
```
Agrupar información utilizando el método de Tukey y una confianza
de 95%

Niveles  N  Media  Agrupación
2        6  8.500  A
3        7  5.714    B
1        5  4.000    B

Las medias que no comparten una letra son significativamente
diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia       de las       EE de
Valor p
de niveles       medias  diferencia      IC de 95%     Valor T
ajustado
2 - 1             4.500       0.950  ( 2.034,  6.966)     4.74
0.001
3 - 1             1.714       0.919  (-0.670,  4.098)     1.87
0.183
3 - 2            -2.786       0.873  (-5.051, -0.520)    -3.19
0.016

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Logotipo vs. Niveles



La desviación estándar agrupada se utilizó para calcular los intervalos.



Si un intervalo no contiene cero, las medias correspondientes son significativamente diferentes.

Intervalo de confianza

Un intervalo de confianza es un rango de valores, derivado de estadísticas de muestra, que posiblemente incluya el valor de un parámetro desconocido de la población. Debido a su naturaleza aleatoria, es poco probable que dos muestras de una población dada generen intervalos de confianza idénticos. Sin embargo, si repitió muchas veces su muestra, un determinado porcentaje de los intervalos de confianza resultantes incluiría el parámetro desconocido de la población. El porcentaje de estos intervalos de confianza que incluyen el parámetro es el nivel de confianza del intervalo.

Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza del **Logotipo** en la **Televisión** (Nivel 2) con el **Logotipo** en la **Prensa escrita** (Nivel 1) oscila entre 2.024 y 6.966. Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de evocación del Logotipo en la Televisión es mayor que el del Logotipo en la Prensa escrita por un mínimo de 2.024 y un máximo de 6.966 puntos.

En cuanto al intervalo de confianza del **Logotipo** en las **Revistas** (Nivel 3) con el **Logotipo** en la **Prensa escrita** (Nivel 1) oscila entre -0.670 y 4.098. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de evocación de Logotipo en las Revistas es igual que el del Logotipo en la Prensa escrita.

Por último podemos observar que el intervalo de confianza del **Logotipo** en las **Revistas** (Nivel 3) con el **Logotipo** en la **Televisión** (Nivel 2) oscila entre -5.051 y -0.520. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de evocación del Logotipo en la Televisión es mayor que el del Logotipo en las Revistas por un mínimo de 0.520 y un máximo de 5.051 puntos (en valor absoluto).

3.2.3.1 EJERCICIOS DE **REFUERZO**

EJERCICIOS DE REFUERZO

3.2.3.1

MANOVA

NOTA:

El uso de un software estadístico como **Excel o Minitab**, entre otros, reduce de gran manera el tiempo de cálculo y la probabilidad de cometer errores en los cálculos aritméticos, sin embargo se deben **comprender primero los pasos del proceso**. Por lo mismo es **muy importante que primero resuelva el ejercicio en forma manual** y **posteriormente utilice un software para comparar sus resultados**. Es importante mencionar que

3.2.3.1.1. En cierta empresa, el departamento de recursos humanos está midiendo la efectividad de 3 distintos programas de inducción para su personal, los cuales consisten en Dinámicas, Casos o Seminarios. La efectividad será medida basándose en el grado de identificación que el personal logra con su empresa. Este último se evalúa a través de 3 criterios:

1. ¿Qué tanto el empleado llega a conocer los objetivos de la empresa? Y,
2. ¿Qué tanta motivación se logró dar a la persona durante el programa de capacitación?
3. ¿Qué tanta habilidad obtuvo el empleado para desempeñar el nuevo puesto?

Los datos obtenidos son los siguientes:

Método de Inducción								
Dinámica			Casos			Seminarios		
Conocimiento	Motivación	Habilidad	Conocimiento	Motivación	Habilidad	Conocimiento	Motivación	Habilidad
9	3	6	1	3	0	4	8	1
6	2	9	1	1	2	1	9	0
8	5	7	2	2	1	2	7	3
9	4	9	0	3	1	2	6	2
8	6	4	2	2	0	1	8	3
			0	1	2	1	9	3
						3	9	2

pueden existir diferencias en las respuestas debido a la cantidad de dígitos que se **utilizan en los cálculos manuales. Se sugiere utilizar aproximaciones de 5 dígitos.**

a) ¿Existe evidencia suficiente que indique que el grado de identificación que el personal logra con su empresa es el mismo en los tres Métodos de Inducción utilizados?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable independiente para ver si el Método utilizado influye sobre el "grado de identificación que el personal logra con su empresa".
c) Realice las pruebas Post-hoc de Tukey para cada variable independiente y construya los intervalos de confianza que indique, en su caso, cual Método de Inducción utilizado logró mayor "grado de identificación del personal con su empresa" y por cuanto es la diferencia.

3.2.3.1.2. El Director de mercadotecnia de una compañía de Bienes Raíces desea medir el grado de influencia que llega a tener el estilo arquitectónico donde se localizan las casas unifamiliares de cierta ciudad en las magnitudes de los atributos de las mismas. Esta influencia es medida a través de tres variables dependientes: Valor comercial de la casa, Tamaño del lote y el Número de Recamaras. Los estilos arquitectónicos manejados son tres: Moderno, Hacienda y Colonial

El efecto se evalúa mediante tres criterios:

1.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Valor de las Casas?.

2.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Tamaño del Lote donde se ubican las casas?.

3.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Número de recámaras de las casas?.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Valor comercial de la casa(en millones de pesos) X_2= Tamaño del lote (en metros cuadrados) y X_3= Número de Recámaras, de 5 propietarios de casas Con estilo arquitectónico Moderno, 6 con estilo Hacienda y 7 Con estilo Colonial. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Estilo Arquitectónico de las Casas								
Moderno			Hacienda			Colonial		
Valor	Tama-ño	Recámaras	Valor	Tama-ño	Recámaras	Valor	Tama-ño	Recámaras
3.74	600	4	6.10	600	5	4.58	1700	4
3.80	1100	4	5.50	750	3	3.50	1518	4
3.60	618	3	5.58	600	5	4.78	1218	4

4.36	600	4	4.98	840	4	5.34	1625	4
3.36	800	4	4.36	600	4	3.80	1000	3
			4.20	700	4	3.60	1125	5
						3.96	1600	4

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de las casas es el mismo en los tres estilos arquitectónicos?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el estilo arquitectónico influye sobre las magnitudes promedio de los atributos de las casas.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que estilo o estilos los atributos son mayores y en que magnitud.

3.2.3.1.3. El Director de Industrias México desea medir el grado de influencia que llega a tener el Nivel de Satisfacción con la empresa de los empleados de Industrias México sobre los ingresos que reciben "antes de impuestos" (en millones de pesos), los años de escolaridad concluidos y la edad de los empleados. Los niveles de satisfacción que se manejaron son tres: Muy Satisfecho, Moderadamente Satisfecho y Muy Insatisfecho

El efecto se evalúa mediante tres criterios:

1.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre el Ingreso de los empleados de Industrias México.

2.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre la Escolaridad de los empleados de Industrias México.

3.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre la Edad de los empleados de Industrias México.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Ingresos anuales "antes de impuestos" (en millones de pesos), X_2= Años de escolaridad concluidos y X_3= Edad de los empleados (En años cumplidos), de 5 empleados muy satisfechos, 6 empleados moderadamente satisfechos y 7 empleados muy insatisfechos. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Nivel de Satisfacción								
Muy Satisfecho			Moderadamente Satisfecho			Muy Insatisfecho		
Ingre-sos	Escolari-dad	Edad	Ingre-sos	Escolari-dad	Edad	Ingre-sos	Escolari-dad	Edad
0.854	15	35	0.550	14	49	0.320	18	36
0.974	19	38	0.656	17	48	0.540	19	47
0.894	16	26	0.472	16	29	0.332	15	43
1.284	16	39	0.646	13	30	0.570	12	36
0.626	12	34	0.358	14	27	0.272	13	30
			0.744	16	36	0.310	11	27
						0.462	16	42

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de los empleados es el mismo en los tres niveles de satisfacción de los empleados con la empresa?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el nivel de satisfacción de los empleados con la empresa influye sobre las magnitudes promedio de los empleados.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que nivel o niveles de satisfacción de los empleados con la empresa, los atributos son mayores y en que magnitud.

AUTOEVALUACIÓN CON REACTIVOS DE

FALSO Ó VERDADERO

EN CADA UNO DE LOS REACTIVOS, CONTESTE CON UNA F SI CONSIDERA QUE LA AFIRMACIÓN ES FALSA Y CON UNA V SI CONSIDERA QUE LA AFIRMACIÓN ES VERDADERA.

1. La forma específica de una distribución F depende del número de grados de libertad en el numerador.

( )

2. En el análisis de varianza, debemos suponer que las muestras se seleccionan de una población normal sin importar si cada una de estas poblaciones tiene la misma varianza.

( )

3. En el diseño aleatorizado en bloques además de las suposiciones del análisis de varianza en un sentido se necesita suponer que no existe efecto de interacción entre los tratamientos y los bloques.

( )

4. La distribución F es discreta, esto significa que puede tocar una cantidad infinita de valores entre cero y más infinito.

( )

5. En una distribución F, conforme el número de grados de libertad aumenta, tanto en el numerador como en el denominador, la distribución se aproxima a una distribución normal.

( )

6. Para utilizar ANOVA se debe suponer que la mayoría de las poblaciones deben seguir la distribución normal.

( )

7. Para utilizar ANOVA se debe suponer que las poblaciones deben tener varianzas iguales (homoscedasticidad).

( )

8. Para probar la homoscedasticidad en el análisis de varianza se utiliza la prueba T de Tukey.

( )

9. En el diseño aleatorizado en bloques puede haber mas de una réplica para cada combinación de tratamientos y bloques.

( )

10. El método de Tukey está diseñado específicamente para comparar dos ó más medias cualesquiera.

( )

11. Para probar la igualdad de varianzas de dos ó más muestras se utiliza la prueba de Levene. **( )**

12. La forma específica de una distribución F depende del numero de grados de libertad en el numerador y denominador de la razón F. **( )**

13. Una prueba de hipótesis donde se utiliza el estadístico F debe ser siempre de extremo derecho. **( )**

14. Los tamaños de las muestras ó tratamientos en el análisis de varianza siempre deben ser iguales. **( )**

15. La forma específica de una distribución F depende del número de grados de libertad en el denominador. **( )**

16. El análisis de varianza se aplica para probar si las medias de dos o más poblaciones pueden considerarse iguales. **( )**

17. En el análisis de varianza si la razón de las varianzas entre tratamientos y la varianza dentro de los tratamientos se acerca a 1 se rechaza la hipótesis nula de igualdad de medias. **( )**

18. Cuando hay una gran interacción entre los niveles de dos factores, el diagrama de interacción la exhibe por medio de líneas no paralelas. **( )**

19. La distribución F es sesgada positivamente, es decir la cola larga de la distribución se encuentra a la derecha. **( )**

20. Usamos el análisis de varianza para decidir si las muestras fueron extraídas de poblaciones que tiene la misma media. **( )**

21. Una segunda variable de tratamiento ó factor que cuando se incluye en el análisis de varianza tiene el efecto de reducir el error experimental se llama variable de bloqueo. **( )**

22. A la suma de las diferencias elevadas al cuadrado entre las observaciones y sus medias de tratamiento se le denomina variación de tratamiento. **( )**

23. A la suma de las diferencias elevadas al cuadrado entre cada observación y la media total se le denomina variación total. **( )**

24. Si el valor del estadístico calculado de F en un análisis de varianza es pequeño, la tendencia a pensar que exista una diferencia entre las medias de los tratamientos es mayor. **( )**

25. Si el valor del estadístico calculado de F en un análisis de varianza es pequeño, la tendencia a pensar que no exista una diferencia entre las medias de los tratamientos es mayor. **( )**

26. Una técnica estadística con la cual se puede probar la igualdad de dos medias se denomina análisis de varianza. **( )**

27. A la suma de las diferencias elevadas al cuadrado entre la media de cada tratamiento y la media total o global se le denomina variación de tratamiento. **( )**

( )

28. Una familia de distribuciones diferenciadas por dos parámetros y utilizada frecuentemente en el análisis de varianza se denomina distribución t de Student.

( )

29. Una familia de distribuciones diferenciadas por un solo parámetro y utilizada frecuentemente para probar la igualdad de dos medias se denomina distribución t de Student.

( )

30. El análisis de varianza se basa en comparar la varianza dentro de los tratamientos con la varianza entre tratamientos. F

( )

31. En el análisis de varianza si la razón de las varianzas entre tratamientos y la varianza dentro de los tratamientos se aleja de 1 se rechaza la hipótesis nula de igualdad de medias.

( )

32. En el análisis de varianza, una vez calculada la razón F, el numerador y el denominador deberán ser aproximadamente iguales si la hipótesis nula de igualdad de medias no es verdadera. F

( )

33. En un diseño de bloques aleatorizados debe haber por lo menos dos réplicas por celda.

( )

34. Para utilizar ANOVA se debe suponer que las muestras se deben seleccionar en forma aleatoria.

( )

35. A la suma de las diferencias elevadas al cuadrado entre la media de cada tratamiento y la media total o global se le denomina variación total.

( )

36. El número máximo de réplicas que puede tener un diseño de dos factores es tres.

( )

37. En el diseño aleatorizado en bloques además de las suposiciones del análisis de varianza en un sentido se necesita suponer que existe efecto de interacción entre los tratamientos y los bloques.

( )

38. El análisis de varianza se basa en comparar la varianza entre tratamientos con la varianza dentro de los tratamientos.

( )

39. A la suma de las diferencias elevadas al cuadrado entre las observaciones y sus medias de tratamiento se le denomina variación aleatoria (error).

( )

40. Cuando no hay interacción, las diferencias entre las medias para los niveles de un factor son iguales para todos los niveles del otro factor; esta es la razón por la que los factores que no interactúan muestran líneas cruzadas en un diagrama de interacción.

( )

41. En el diseño aleatorizado en bloques no todos los tratamientos ni todos los bloques deben tener el mismo tamaño

( )

42. En el análisis de varianza en dos sentidos o dos factores debe haber por lo menos dos replicas para cada combinación de factores.

( )

43. A la suma de las diferencias elevadas al cuadrado entre cada observación y la media total se le denomina variación de tratamiento.

( )

44. El numero máximo de réplicas que puede tener el diseño aleatorizado en bloques es una. **( )**

45. En el diseño de dos factores además de las suposiciones del análisis de varianza en un sentido se necesita suponer que existe efecto de interacción entre los niveles de ambos factores. **( )**

AUTOEVALUACIÓN CON REACTIVOS DE OPCIÓN MÚLTIPLE

EN CADA UNO DE LOS REACTIVOS SIGUIENTES, SELECCIONE LA OPCIÓN QUE CONSIDERE CORRECTA.

1. La distribución *F*:
- a) No puede ser negativa.
- b) No puede ser positiva.
- c) Es la misma que la distribución t.
- d) Es la misma que la distribución X^2.

2. Los valores de *t*:
- a) Todos son negativos.
- b) Todos son positivos.
- c) Dependen de los grados de libertad.
- d) Todos los anteriores.

3. Conforme se aumenta el tamaño de la muestra, la distribución *t* se aproxima a:
- a) La distribución binomial.
- b) La distribución normal estándar o distribución z.
- c) La distribución de Poisson.
- d) La distribución *F.*

4. El estadístico de prueba para probar una hipótesis con muestras pequeñas, cuando no se conoce la desviación estándar poblacional es:

a)z b)t c)*F* d)X^2 .

5. Una muestra tiene 20 observaciones. Si se realiza una prueba de extremo derecho en la que se usa la distribución t como estadístico de prueba y el nivel de significación es 0.01, el valor crítico es:
- a) 2.5280.

b) -2.8609.
c) -2.0930.
d) 2.8609.

6. Una muestra tiene 31 observaciones. Si se realiza una prueba bilateral en la que se usa la distribución t como estadístico de prueba y el nivel de significación es 0.05, el valor crítico es:
a) $\mp$2.4573.
b) -2.4573.
c) $\mp$2.0423.
d) +2.7500.

7. Cometemos un error de tipo I cuando:
a) Se rechaza una hipótesis nula que es verdadera.
b) No se rechaza una hipótesis alternativa que es verdadera.
c) Se rechaza una hipótesis alternativa que es verdadera.
d) No se rechaza una hipótesis nula que es falsa.

8. Cometemos un error de tipo II cuando:
a) Se rechaza una hipótesis nula que es verdadera.
b) No se rechaza una hipótesis alternativa que es verdadera.
c) Se rechaza una hipótesis alternativa que es verdadera.
d) No se rechaza una hipótesis nula que es falsa.

9. Una muestra tiene 21 observaciones y otra 30. Si se realiza una prueba de ANOVA en la que se usa la distribución F como estadístico de prueba y el nivel de significación es 0.05, el valor crítico es:
a) 2.21.
b) 1.94.
c) 2.57.
d) 2.86.

10. Si planteamos el siguiente juego de hipótesis $H_o: \mu \geq 240$ y $H_{1:}\mu < 240$
a) Se trata de una prueba de extremo derecho.
b) Se trata de una prueba bilateral o de dos colas.
c) Se trata de una prueba de extremo izquierdo.
d) Todas las anteriores.

11. Si planteamos el siguiente juego de hipótesis $H_o: \mu = 240$ y $H_{1:}\mu \neq 240$
a) Se trata de una prueba de extremo derecho.
b) Se trata de una prueba bilateral o de dos colas.
c) Se trata de una prueba de extremo izquierdo.
d) Todas las anteriores.

12. Si planteamos el siguiente juego de hipótesis $H_o: \mu \leq 240$ y $H_{1:}\mu > 240$
a) Se trata de una prueba de extremo derecho.
b) Se trata de una prueba bilateral o de dos colas.
c) Se trata de una prueba de extremo izquierdo.
d) Todas las anteriores.

13. Se obtiene una muestra aleatoria de 21 observaciones. Se realiza una prueba de hipótesis de extremo izquierdo con un nivel de significancia de 0.05. El valor del estadístico de t calculada es -2.8453, esto indica que: -2.086

a) No se debe rechazar la H_0.
b) Se debe rechazar H_0.
c) No se debe rechazar H_1.
d) b y c solamente.

14. Se quiere probar la hipótesis de que μ_2 es mayor que μ_1

a) Se debe emplear una prueba de extremo izquierdo.
b) Se debe emplear una prueba bilateral.
c) Se debe emplear una prueba de extremo derecho.
d) Con la información que se tiene no se puede determinar si se debe utilizar una prueba de extremo izquierdo, de extremo derecho, o una prueba bilateral.

15. Se quiere probar la hipótesis nula de que μ_2 es igual que μ_1

a) Se debe emplear una prueba de extremo izquierdo.
b) Se debe emplear una prueba bilateral.
c) Se debe emplear una prueba de extremo derecho.
d) Con la información que se tiene no se puede determinar si se debe utilizar una prueba de extremo izquierdo, de extremo derecho, o una prueba bilateral.

16. Se quiere probar la hipótesis nula de que μ_1 es mayor que μ_2

a) Se debe emplear una prueba de extremo izquierdo.
b) Se debe emplear una prueba bilateral.
c) Se debe emplear una prueba de extremo derecho.
d) Con la información que se tiene no se puede determinar si se debe utilizar una prueba de extremo izquierdo, de extremo derecho, o una prueba bilateral.

17. Se quiere probar la hipótesis nula de que μ_1 es igual que μ_2 y μ_3

a) Se debe emplear una prueba de extremo izquierdo.
b) Se debe emplear una prueba bilateral.
c) Se debe emplear una prueba de extremo derecho
d) Con la información que se tiene no se puede determinar si se debe utilizar una prueba de extremo izquierdo, de extremo derecho, o una prueba bilateral.

18. Se realizó una prueba ANOVA de dos factores y se rechazó la hipótesis nula para el Factor 1. Esto indica que:

a) Hay demasiados grados de libertad.
b) No hay diferencia entre las medias de los tratamientos del Factor 1.
c) Hay diferencia, en por lo menos uno de los tratamientos del Factor 1.
d) a y c solamente.

19. Se realizó una prueba ANOVA de dos factores y no se rechazó la hipótesis nula para el Factor 1. Esto indica que:

a) Hay demasiados grados de libertad.
b) No hay diferencia entre las medias de los tratamientos del Factor 1.
c) Hay diferencia, en por lo menos uno de los tratamientos del Factor 1.

d) a y b solamente.

20. ¿Cúal de los siguientes enunciados es un paso en la realización del análisis de variancia?

a) Determinar una estimación de la variancia de la población dentro de los tratamientos.
b) Determinar una estimación de la variancia de la población entre los tratamientos.
c) Determinar la diferencia entre la frecuencia observada y la esperada en cada clase.
d) a y b pero no c.

21. Una prueba bilateral con dos variancias va a realizarse en las muestra 1 y 2 con n_1 = 15 y n_2 = 12. Si α = .10, ¿Cuál de los siguientes enunciados representa el valor superior con que debería compararse s_1^2 / s_2^2 ?

a) $\frac{1}{F_{0.05,11,14}}$
b) $\frac{1}{F_{0.05,14,11}}$
$c)$ $F_{0.05,11,14}$
$d)$ $F_{0.05,14,11}$

22. Se realizó una prueba ANOVA de dos factores y se rechazó la hipótesis nula para el Factor 1. Esto indica que:

a) Hay demasiados grados de libertad.
b) No hay diferencia entre las medias de los tratamientos del Factor 1.
c) Hay diferencia, en por lo menos uno de los tratamientos del Factor 1
d) a y c solamente.

23. Una prueba bilateral con dos variancias va a realizarse en las muestra 1 y 2 con n_1 = 12 y n_2 = 15. Si α = .10, ¿Cuál de los siguientes enunciados representa el valor inferior con que debería compararse s_1^2 / s_2^2 ?

a) $\frac{1}{F_{0.05,11,14}}$
b) $\frac{1}{F_{0.05,14,11}}$
c) $F_{0.05,11,14}$
d) $F_{0.05,14,11}$

24. Suponga que está comparando 5 grupos sometidos a diferentes métodos de tratamiento y que ha seleccionado una muestra de tamaño 10 en cada uno y se ha calculado $\bar{X}$ en cada tratamiento ¿Cómo debe calcular ahora la gran media?

a) Multiplicando cada media muestral por $1/5$ y sumando estos valores. Dividiendo luego la suma entre 50.
b) Sumando las 5 medias muestrales y dividiendo entre 50 el resultado.
c) Sumando las 5 medias muestrales y multiplicando por $1/5$ el resultado.
d) Sumando las 5 medias muestrales.

25. ¿Cúal de las siguientes distribuciones tiene un par de grados de libertad?

a) De Poisson.
b) Normal.

c) *F.*
d) Binominal.

GLOSARIO **GENERAL**

AGLOMERATIVO. Significa que tiene muchos grupos, éstos se unen hasta formar un todo.

AGRUPAR. Selección de los elementos más parecidos entre si.

ALEATORIA, ASIGNACIÓN. Uso de métodos aleatorios para designar pacientes a tratamientos distintos o viceversa**.**

ALFA La probabilidad de cometer un error de tipo I. Se representa por la letra griega α.

ANÁLISIS CONJUNTO. Técnica multivariante que se utiliza para conocer las preferencias hacia productos o servicios por parte de los consumidores.

ANÁLISIS DE BLOQUES AL AZAR. Sirve para aislar el efecto después de realizado el experimento, pero su variable incontrolable es de naturaleza no métrica e influye en la variable dependiente.

ANÁLISIS DE CLASES LATENTES O DE SEGMENTOS LATENTES. Método para encontrar la variable no observada o latente, compuesta por subgrupos o segmentos, por lo que sus conglomerados son K distintas categorías de una variable latente nominal.

ANÁLISIS DE CONGLOMERADOS. Técnica del grupo estructural que tiene por objeto resumir información y que sirve para clasificar en grupos o categorías. Técnica empleada en la segmentación de mercados, que se basa en formar grupos que contengan elementos lo más similares entre sí y diferentes al compararse entre grupos.

ANÁLISIS DE CONJUNTO. Técnica que estudia el efecto que tienen diversas variables independientes no métricas sobre una variable dependiente no métrica (ordinal).

ANÁLISIS DE CORRELACIÓN CANÓNICA. Consiste en determinar el grado de relación entre dos conjuntos de múltiples variables.

ANÁLISIS DE CORRELACIÓN. Técnica de con que se determina el grado de relación lineal que hay entre variables.

ANÁLISIS DE CORRESPONDENCIA. Establece relaciones entre variables independientes cualitativas, enriqueciendo así la información de las tablas de contingencia.

ANÁLISIS DE COVARIANZA. Técnica que pretende aislar o controlar los efectos de una variable que durante el desarrollo del experimento no pudo controlarse y afecta las mediciones de la variable dependiente. Busca como afecta la variable incontrolable (de naturaleza métrica) a la variable dependiente para aislar y controlar su efecto.

ANÁLISIS DE FACTORES O ANÁLISIS FACTORIAL. Método mediante el cual se encuentran variables no observadas o latentes, las cuales son métricas. Técnica estructural o de interdependencia en la que todas las variables tienen la misma importancia. Método mediante el cual se encuentran variables no observadas o latentes. Reduce un sistema múltiple de variables métricas a un menor número de perfiles latentes.

ANÁLISIS DE PERFILES LATENTES. Método que sirve para encontrar variables no observadas o latente no métricas. Reduce un sistema múltiple de variables métricas a un menor número de perfiles latentes.

ANÁLISIS DE RASGOS LATENTES. Método por el cual se analizan datos categóricos de forma binaria para encontrar variables no observadas o latentes.

ANÁLISIS DE REGRESIÓN. Busca saber la cantidad de variación que la covariable pudo incluir en la variable dependiente. Es una técnica multivariante, cuyo objetivo es desarrollar un modelo matemático lineal. Permite conocer la relación causa-efecto que hay entre varas variables independientes y una dependiente.

ANÁLISIS DE RESIDUALES. Respecto a regresión, análisis de las diferencias entre Y y $\hat{Y}$ para valorar las premisas y proporciona guías sobre qué tan bien se ajusta la ecuación a los datos.

ANÁLISIS DE RESIDUALES. Respecto a regresión, análisis de las diferencias entre Y y $\hat{Y}$ para

ANÁLISIS DE SUPERVIVENCIA. Estudia el efecto de una variable independiente (tiempo, variable métrica) cuando la variable dependiente se expresa en tiempo que tarda en ocurrir un suceso.

ANALISIS DE VARIANZA (ANOVA). Técnica estadística con que se prueba la igualdad de 3 o más medias muéstrales y que, por tanto, permite hacer inferencias sobre si las muestras provienen de poblaciones que tienen la misma media.

ANÁLISIS DE VARIANZA DE UN FACTOR. Analiza una variable dependiente medida bajo cierta condición experimental identificada por una variable independiente.

ANÁLISIS DE VARIANZA MULTIVARIADO (MANOVA). Es un análisis de varianza cuyo objetivo es explicar "n" variables dependientes a partir de "m" independientes. Técnica que permite pronosticar simultáneamente 2 o más variables dependientes, que representan un concepto o factor.

ANÁLISIS DE VARIANZA UNIVARIADO. Es un análisis de varianza cuyo objetivo es explicar una variable dependiente a partir de "m" variables independientes.

ANÁLISIS DE VARIANZA. Técnica funcional que utiliza una o más variables independientes, todas ellas no métricas, y trata de explicar el comportamiento de una o mas variables dependientes métricas. Examina el efecto que la variable no métrica tiene en la variable dependiente.

ANÁLISIS DISCRIMINANTE. Es una técnica de clasificación o asignación de un elemento a un grupo del cual se conocen sus atributos. Técnica funcional que permite explicar las diferencias entre grupos de personas u objetos mediante variables independientes con el objetivo de pronosticar o clasificar nuevas observaciones entre los grupos.

ANÁLISIS MULTIVARIANTE. Métodos o técnicas de análisis e interpretación de un conjunto de variables que conservan su grado de interacción.

ANCOVA. Análisis de la covarianza, del inglés Analysis of Covariance. Es una combinación del análisis de la varianza y de la regresión. Corrige el ANOVA del efecto que pueda tener una variable no controlada, covariable, sobre la variable respuesta mediante la función de regresión de la variable respuesta sobre la covariable.

ANOVA. Análisis de la varianza, del inglés Analysis of Variance. Se utiliza también el término ADEVA, de su traducción al castellano. Es un método para contrastar la homogeneidad de más de dos variables poblacionales mediante la descomposición de la suma total de los cuadrados de las diferencias entre valores observados de esas variables poblacionales y su media muestral, como medida de la heterogeneidad de esas poblaciones. Las variables poblacionales para cada tratamiento son concreciones de una variable general llamada variable respuesta o variable dependiente.

BETA Probabilidad de cometer un error de tipo II. Se representa por la letra griega β.

BLOQUE, DISEÑO EN. En el análisis de varianza, un protocolo donde los sujetos en cada bloque o grupo están asignados a tratamiento diferente.

BLOQUE. Unidades de prueba incontrolables que representan un grupo experimental y son equivalentes a los niveles de los factores. Es una segunda fuente de variación, además de los tratamientos.

BLOQUES AL AZAR. Técnica que elimina posibles fuentes de variación que no pudieron controlarse durante el experimento.

CARGA DE FACTOR. Término que indica el grado de relación que llegan a tener cada una de las variables con cada uno de los factores.

CASO BALANCEADO. Manejo del mismo número de datos en cada uno de los niveles del factor.

CENTROIDE. Es el promedio o centro de los grupos.

CLASIFICACIÓN. Es uno de los procesos fundamentales de la ciencia que ordena los fenómenos para poderlos entender.

COEFICIENTE (r) DE CORRELACIÓN DE PEARSON . Una medida de la relación lineal entre dos mediciones numéricas hechas en el mismo conjunto de sujetos. Oscila de -1 a +1, con el cero indicando la ausencia de relación**.** Raíz cuadrada del coeficiente de determinación. Su signo indica la dirección de la relación entre dos variables, directa o inversa. Índice estadístico que mide la relación lineal entre dos variables cuantitativas.

COEFICIENTE (R^2) DE DETERMINACIÓN . Medida de la proporción de variación de *Y*, la variable dependiente, que se explica con la línea de regresión; esto es, por la relación de las *Y* con la variable independiente. Se interpreta como la cantidad de variación en una variable que puede definirse por el conocimiento de una segunda variable. Se utiliza para saber cuánto se logró captar o explicar del comportamiento de la variable dependiente.

COEFICIENTE DE KENDALL'S TAU. Se utiliza cuando se mide la relación entre variables ordinales. Toma valores entre -1 y 1.

COLINEALIDAD. Surge cuando dos variables independientes se relacionan entre sí, ya sea linealmente o no.

COMPONENTES PRINCIPALES Y FACTOR COMÚN. Análisis que se usan para encontrar la interdependencia (dependencia interna) entre variables.

COMUNALIDAD. Variación explicada por el conjunto de factores para cada variable.

CONFIANZA, INTERVALO DE (IC). Espacio calculado a partir de los datos de una muestra, que tiene una probabilidad dada de comprender el parámetro desconocido.

CONFIANZA, LIMITES DE. Delimitan a un intervalo de confianza. Se calculan de los datos de la muestra y tienen una probabilidad dada de que el parámetro desconocido se ubique entre éstos.

CORRELACIÓN CANÓNICA. Coeficiente que mide, para cada función discriminante, el grado en que se diferencian las medias por cada una de las funciones.

COVARIANZA. Es el promedio de la suma de los productos, es decir, es la suma de los productos cruzados dividido entre los grados de libertad.

CRITERIO COMPLETO O DISTANCIAS MÁXIMAS. Supone que cuando dos elementos se unen, el parecido de ese grupo a un tercer elemento estará dado por las características del elemento con el cual tenga más diferencias, es decir, se eslabona con la máxima de las distancias o la mínima de las similitudes.

CRITERIO SIMPLE O DISTANCIAS MÍNIMAS. Asume que al fusionarse dos objetos se tomarán como características las de aquel elemento que sea más parecido a aquel con el que se compara.

CUADRADO DE YOUDEN. Técnica que permite dos fuentes experimentales de variabilidad y que es menos restrictiva que los cuadrados latinos.

CUADRADO GRECOLATINO. Aumento del número de factores bloque en la extensión del cuadrado latino.

CUADRADO MEDIO DENTRO DEGRUPO. Estimación de la variación en el análisis de varianza. Se usa en el denominador de la prueba estadística ***F***.

CUADRADO MEDIO ENTRE GRUPOS. Estimación de la variación en el análisis de varianza. Se usa en el numerador de la estadística ***F***.

CUADRADOS LATINO. Modelo que aprovecha la simetría del experimento factorial al seleccionar un conjunto de condiciones experimentales con la condición de que cada nivel de un factor aparezca una vez con cada uno de los niveles de los otros factores. Técnica que supone, en su modelo general, que solamente existen los efectos principales de los factores.

CUADRADOS MEDIOS. Cociente entre la suma de cuadrados y los grados de libertad.

CUADRADOS SUMAS DE. Cantidad calculada en el análisis de varianza y usada para obtener cuadrados medios para la prueba ***F***.

DATO. Colección de observaciones numéricas o no numéricas con una característica común.

DENDOGRAMA. Diagrama en cuyo eje horizontal aparecen los elementos agrupados, de preferencia en el orden en que se unieron.

DETECTOR AUTOMÁTICO DE INTERACCIÓN. Técnica del análisis multivariante que permite conocer la relación entre una variable dependiente en escala métrica (de intervalo o de razón) y varias variables independientes en escala métrica o no (nominales, de intervalo, de razón).

DIAGRAMA DE DISPERSIÓN. Grafica de puntos sobre una rejilla rectangular; las coordenadas *X* y *Y* de cada punto de corresponden a las dos mediciones hachas en algún elemento particular de la muestra, y el patrón de puntos indica la relación existente entre las dos variables.

DIAGRAMA DE TÉMPANOS O DE CARÁMBANOS. Simplifica los elementos que se agruparon paso a paso. DIMENSIÓN. Variables simultáneas que hay en un conjunto.

DISEÑO DE TRES FACTORES. Diseño en el que se tienen tres factores con un número diferente de niveles y las interacciones entre ellos no son nulas.

DISEÑO ESTADÍSTICO DE EXPERIMENTOS. Es el proceso de planeación de un experimento que puede analizarse mediante métodos estadísticos con el objetivo de encontrar soluciones apropiadas.

DISEÑO FACTORIAL 2^2. Técnica de análisis de varianza que sólo utiliza dos factores con dos niveles cada uno.

DISEÑO FACTORIAL 2^K EN DOS BLOQUES. Técnica que se emplea cuando el diseño factorial 2^2 necesita más de un bloque para su realización.

DISEÑO FACTORIAL 3^K. Variación del diseño 2^2 que consta de *K* factores con tres niveles cada uno.

DISEÑO FACTORIAL FRACCIONARIO DE DOS NIVELES. Diseño con k factores con dos niveles que mantienen la propiedad de ortogonalidad entre los factores y donde se suponen nulas las interacciones de orden alto.

DISEÑO FACTORIAL O DE DOS FACTORES. Modelo de análisis que se utiliza cuando se tienen dos factores con un número diferente de niveles y las interacciones entre ambas no son nulas.

DISEÑO JERÁRQUICO O ANIDADO. Diseño en el que existen dos factores bloque anidados.

DISEÑO POR BLOQUES INCOMPLETOS. Técnica que se emplea cuando algunos experimentos que usan diseño por bloques no pueden realizar los ensayos o mediciones de todas las combinaciones de tratamientos dentro de cada bloque.

DISEÑO RETICULAR. Técnica que se realiza del mismo modo que un diseño de bloque incompleto y que se utiliza cuando hay muchas combinaciones de tratamientos.

DISIMILARIDAD. Las proximidades se interpretan en términos de lo diferente.

DISTANCIA ENTRE DOS GRUPOS. Distancia que existe entre las medianas de las características de los individuos que componen los grupos.

DISTRIBUCIÓN *F*. Se emplea como el estadístico de prueba en el ANOVA, así como en otras pruebas.

DISTRIBUCIÓN *t*. Fue investigada y dada a conocer por William S. Gosset, en 1908, bajo el seudónimo de *Student.* Es similar a la distribución normal.

ECUACIÓN DE ESTIMACIÓN. Fórmula matemática que relaciona la variable desconocida con las variables conocidas en el análisis de regresión.

ECUACIÓN DE REGRESIÓN MÚLTIPLE. Es una ecuación que define la relación lineal entre dos ó más variables.

ECUACIÓN DE REGRESIÓN. Es una ecuación que define la relación lineal entre dos variables.

EFECTO DE UN FACTOR. Es el cambio en la variable dependiente producido por un cambio en el nivel del factor.

EFECTOS ALEATORIOS. Los tratamientos son aleatorios cuando los tratamientos de la experimentación son una parte aleatoria del colectivo total de tratamientos sobre el que tenemos que inducir las conclusiones del contraste.

EFECTOS FIJOS. Los tratamientos son fijos cuando los tratamientos de la experimentación coinciden con los tratamientos totales sobre los que tenemos que sacar conclusiones

EFECTOS PRINCIPALES. Efectos que los factores pueden tener sobre la variable dependiente.

EFECTOS. Medida de la influencia de los tratamientos en la heterogeneidad de las poblaciones.

EIGENVALOR. Variación o cantidad de información que cada uno de los factores logra captar del conjunto de variables.

ENFOQUE PROMEDIO. Consiste en utilizar como puntajes los promedios de las variables que se agruparon con cada factor.

ERROR DE TIPO I Se presenta cuando se rechaza una H_0 verdadera.

ERROR DE TIPO II Se presenta cuando no se rechaza H_0 falsa.

ERROR ESTÁNDAR DE ESTIMACIÓN MÚLTIPLE. Medida de confiabilidad de la ecuación de estimación, que indica la variabilidad de los puntos observados alrededor de un plano de regresión (en el caso de dos variables independientes); es decir, hasta qué punto los valores observados difieren de los predichos en el plano de regresión.

ERROR ESTÁNDAR DE ESTIMACIÓN. Medida de la confiabilidad de la ecuación de estimación, que indica la variabilidad de los puntos observados alrededor de la línea de regresión; es decir, hasta qué punto los valores observados difieren de los predichos en la línea de regresión.

ERROR ESTÁNDAR DE UN COEFICIENTE DE REGRESIÓN. Medida de la incertidumbre respecto al valor exacto de un coeficiente de regresión. Medida de la variabilidad de los coeficientes de regresión de la muestra alrededor del verdadero coeficiente de regresión de la población.

ERROR ESTÁNDAR. Valor de la distancia para un valor particular.

ERROR MUESTRAL. Error debido a la aleatoriedad muestral, a que sea un elemento y no otro el observado. Es el error debido a la aleatoridad en la selección de los elementos muestrales

ERRORES. Son la diferencia entre el dato real y el dato pronosticado.

ESCALAS MULTIDIMENSIONALES. Técnica que traduce los juicios del consumidor en similaridad o preferencia mediante distancias representadas en una gráfica.

ESLABONAMIENTO PROMEDIO. Grado de similitud que tiene un nuevo grupo respecto a otros elementos no agrupados.

ESLABONAMIENTO. Consiste en volver a establecer el grado de parecido de un nuevo grupo con el resto de los elementos.

ESTADÍSTICA. Conjunto de teorías y métodos desarrollados para la recolección, manejo y análisis de datos.

ESTADÍSTICO F. Función de datos muestrales que sigue la ley F de Snedecor al definirse como cociente de dos sumas de cuadrados divididas por sus grados de libertad, es decir como cociente de dos cuadrados medios.

ESTIMACIONES O ESTÍMULOS. Combinación de todos los atributos con todos los niveles.

EXPERIMENTACIÓN. Es un proceso para descubrir algo acerca de un sistema en particular.

EXPERIMENTAL, ESTUDIO. Estudio comparativo en donde hay una intervención o manipulación. Se designa como una prueba o serie clínica cuando se aplica a seres humanos.

EXPERIMENTO. Proceso planeado de obtención de datos. Es una prueba en la cual se inducen cambios en los datos de entrada para observar las causas de los cambios en los datos de salida.

EXPERIMENTOS CONTROLADOS. El investigador puede manipular el medio en el que se lleva a cabo el experimento, así como a los participantes.

EXPERIMENTOS NATURALES. El investigador actúa como observador, sin poder seleccionar las unidades de prueba.

F CALCULADA. Estadístico que se usa como prueba de la significancia de una variable explicatoria individual.

FACTOR DE BLOQUEO. Factor que puede afectar a la variable dependiente, o dicho de otra forma a la homogeneidad de las poblaciones, y, por ello, lo introducimos como factor secundario para eliminar la influencia que pudiera tener sobre las conclusiones del contraste. Véase explicación detallada en el modelo III del ANOVA.

FACTOR DE INFLACIÓN DE VARIANZAS. Es un indicador de cuánto las varianzas de los β.s crecen y se vuelven imprecisas.

FACTOR. Característica que es el centro de indagación en un estudio**.** Cada una de las causas que influyen en la heterogeneidad de las poblaciones en estudio. Si hay un solo factor lo representamos por A, si hay dos o más lo representamos por A, B, C. Desde el punto matemático, es una combinación lineal (aditiva) de las variables originales de la investigación.

FACTORES O TRATAMIENTOS. Son las variables independientes del experimento.

GRADO DE LIBERTAD Es el número de elementos en una muestra que pueden variar libremente. Un parámetro en algunas distribuciones de probabilidad de uso común, por ejemplo, distribución ***t***, ji cuadrada y ***F.*** Número de términos independientes de una suma.

GRAN MEDIA. Media del grupo completo de sujetos de todas las muestras del experimento.

HETEROSCEDASTICIDAD. Es cuando las varianzas no son constantes.

HIPÓTESIS ALTERNATIVA La conclusión que se acepta cuando se demuestra que la hipótesis nula es falsa. También se conoce como hipótesis de investigación.

HIPÓTESIS NULA Hipótesis, o suposición, acerca de un parámetro de la población que deseamos probar, generalmente una suposición del status quo (situación actual).

HIPÓTESIS Suposición, o conjetura, que hacemos sobre un parámetro de la población.

HOMOCEDASTICIDAD. Supuesto de igualdad de as varianzas poblacionales.

HOMOSCEDASTICIDAD U HOMOGENEIDAD DE VARIANZAS. Es cuando las fuentes de variación de un experimento son constantes o iguales. Indica homogeneidad de varianzas o que las varianzas son constantes.

IGUALDAD FUNDAMENTAL DEL ANÁLISIS DE LA VARIANZA. Descomposición de la suma total de cuadrados de los elementos muestrales respecto de su media global en dos términos, el que mide la influencia del factor en la heterogeneidad y el que mide la influencia del error muestral o aleatoriedad muestral en dicha heterogeneidad.

INFORMACIÓN ALEATORIZADA EN BLOQUES. Esto sucede cuando la información reparte por bloques de manera aleatoria entre los tratamientos.

INFORMACIÓN COMPLETAMENTE ALEATORIZADA. Esto ocurre cuando la información de la experimentación se reparte de forma totalmente aleatoria entre los tratamientos.

INTERSECCIÓN EN *Y*. Constante de cualquier recta, cuyo valor representa el valor de la variable *Y* cuando la variable *X* tiene un valor de 0.

INTERSECCIÓN EN *Y*. Constante de cualquier recta, cuyo valor representa el valor de la variable *Y* cuando la variable *X* tiene un valor de 0.

LAMBDA SE WILKS. Medida que evalúa la capacidad discriminante de la función y proporción de la varianza total explicada por los puntajes discriminantes.

LÍNEA DE REGRESIÓN. Línea ajustada a un conjunto de puntos de datos para estimar la relación entre dos variables.

LINEALIDAD O ADITIVIDAD. Suma de los efectos de cada una de las variables independientes.

MATRIZ DE CLASIFICACIÓN O CONFUSIÓN. Presentan los datos reales contra los pronosticados y se basa en el conteo de lo correcta e incorrectamente clasificado.

MATRIZ IDENTIDAD. Es aquella que en su diagonal principal tiene 1 y fuera de ella, ceros; esto indicaría que no existen correlaciones entre las variables.

MATRIZ NO ROTADA. Resultado de la descomposición espectral que muestra las eigenvectores correspondientes.

METAANÁLISIS. Conjunto de técnicas que permite revisar y combinar diversos resultados experimentales de distintos estudios previos para contestar una misma pregunta científica o emitir una misma conclusión.

MÉTODO BONFERRONI. Técnica estadística que ajusta el nivel de significancia con relación al número de pruebas estadísticas realizadas simultáneamente sobre un conjunto de datos.

MÉTODO C DE DUNNET. Es idéntico al de Games-Howell excepto en la forma de corregir los grados de libertad de la distribución del rango estudentizado.

MÉTODO DE MÍNIMOS CUADRADOS. Técnica con que se ajusta una recta mediante un conjunto de puntos, de manera que se minimice la suma de los cuadrados de las distancias verticales entre *n* puntos y la línea.

MÉTODO DE REASIGNACIÓN. Permite que un individuo asignado a un grupo sea reasignado a otro en un paso posterior, ya que no estaba en el grupo adecuado; al reasignarse se optimiza el criterio de selección.

MÉTODO DE REDUCCIÓN DE DIMENSIONES. Busca dimensiones latentes en un conjunto según sus características y con base en ellas determina un calificativo que describa al conglomerado.

MÉTODO DE R-E-G-W-F. Se basa en la distribución F; es un método que no resulta apropiado cuando los grupos tienen datos diferentes o en un caso no balanceado.

MÉTODO DE TUKEY. Método que se basa en la construcción de intervalos de confianza de las diferencias por pares.

MÉTODO DE WARD. Busca a los dos grupos o conglomerados cuya unión conlleve el menor incremento de la varianza.

MÉTODO DEL CENTROIDE. Criterio que considera que, cuando dos elementos se unen y forman un grupo, las características que prevalecen respecto a un tercer elemento están dadas por el promedio de las características originales.
MÉTODO DUNNET. Método que sirve para comparar cada grupo con un grupo de control pues controla la tasa de error para $k - 1$ comparaciones.
MÉTODO GABRIEL. Método que se basa en la distribución del módulo máximo estudentizado.
MÉTODO GAMES-HOWELL. Se basa en la distribución del rango estudentizado y en un estadístico T en el que se corrigen los grados de libertad mediante la ecuación de Welch.
MÉTODO GT2 DE HOCHBERG. Método similar a la prueba de Tukey que se basa en la distribución del módulo máximo estudentizado.
MÉTODO JERÁRQUICO DIVISIONAL. Supone que existe un solo grupo que contiene a todos los objetos y basándose en las diferencias los separa hasta terminar con tantos grupos como objetos se tienen.
MÉTODO JERÁRQUICO. Algoritmo que se sigue para formar grupos de personas u objetos similares entre sí.
MÉTODO R-E-G-W-Q. Método que utiliza el mismo estadístico que la prueba de Newman-Keuls o la de Tukey pero controla el nivel de significancia de cada comparación.
MÉTODO SIDAK. Método basado en la distribución t de *Student* que controla la tasa de error al evaluar cada comparación con un nivel de significancia $\alpha = 1 - (1 - \alpha)1/k$.
MÉTODO T2 DE TAMHANE. Método conservador basado en la distribución t que utiliza comparaciones por parejas.
MÉTODO T3 DE DUNNETT. Modificación del estadístico T2 de Tamhane que utiliza la distribución del módulo máximo estudentizado.
MÉTODO WALLET-DUNCAN. Utiliza la distribución t-*student* y una aproximación bayesiana.
METODOLOGÍA ESTADÍSTICA. Es el único enfoque objetivo que analiza problemas que involucren datos sujetos a errores experimentales.
MÍNIMAS DIFERENCIAS SIGNIFICATIVAS. Determina el valor mínimo necesario para que dos tratamientos se consideren diferentes y lo utiliza para comparar los pares de medias que se deseen evaluar.
MODELO ADITIVO. Es el que explica mejor el comportamiento de la variable dependiente.
MUESTRA ALEATORIA. Muestra de n sujetos (u objetos) seleccionada de una población de modo que cada uno tenga una probabilidad conocida de estar en la muestra**.**
MUESTRAS DEPENDIENTES. Las muestras dependientes se caracterizan porque se hace una medición, después una intervención, y de nuevo se realiza una medición. Las muestras en pares también son dependientes, ya que un mismo individuo u objeto es miembro de ambas muestras. Ejemplo: diez participantes en un maratón se pesaron antes y después de la carrera. Se quiere estudiar la cantidad media de peso corporal que pierden los participantes.
MUESTRAS INDEPENDIENTES. Las muestras tomadas aleatoriamente no están relacionadas una con otra. Se quiere estudiar la edad promedio de los internos en las prisiones A y B. Se toma una muestra aleatoria de 28 internos en la prisión A, y una muestra de 19 internos en la prisión B. Una persona no puede ser interno de ambas prisiones. Las muestras son independientes, no están relacionadas.
MULTICOLINEALIDAD. Problema estadístico que en ocasiones se presenta en el análisis de regresión múltiple; en él se reduce la confiabilidad de los coeficientes de regresión, a causa de un alto nivel de correlación entre las variables independientes. Es cuando dos o más de dos variables independientes están correlacionadas, es decir, no son independientes entre sí.
NIVEL DE SIGNIFICANCIA Valor que indica el porcentaje de los valores muestrales que se halla fuera de ciertos límites, suponiendo que la hipótesis nula sea correcta, esto es, la probabilidad de rechazarla cuando es verdadera.
NIVELES DE FACTOR. Son los valores que toman las variables independientes.
NIVELES DEL FACTOR O TRATAMIENTOS. Cada uno de los valores posibles del factor. Los representamos por A, si el factor se representa por A.
NIVELES. Valores que poseen las variables.

PENDIENTE. Constante de cualquier recta, cuyo valor representa en qué medida el cambio de cada unidad de la variable independiente modifica la variable dependiente.

PORCENTAJE DE CLASIFICACIONES CORRECTAS. Indicador que evalúa lo acertado que es el modelo para clasificar como medio para obtener una medida de la confiabilidad al pronosticar. Medida de la confiabilidad o seguridad que existe al pronosticar y que se obtiene con los datos empleados en la etapa de validación.

PORCENTAJE DE VARIACIÓN EXPLICADA. Relación entre la variación explicada por el factor con respecto a la variación total manejada por el conjunto de variables.

PRINCIPIO DE MÍNIMOS CUADRADOS. Técnica empleada para obtener la ecuación de regresión, minimizando la suma de los cuadrados de las distancias verticales entre los valores verdaderos de ***Y*** y los valores pronosticados de ***Y.***

PRODUCTOS CRUZADOS. Son la covarianza de los datos de las variables, sin considerar los grupos.

PROMEDIO ENTRE GRUPOS. Define la distancia entre 2 grupos como el promedio de las distancias entre todas las combinaciones posible por parejas, al combinar todos los grupos, en los cuales un miembro del par pertenece a cada uno de los grupos formados anteriormente.

PROMEDIO INTRAGRUPOS. Combinan grupos de manera que la media de las distancias entre todos los casos en el grupo resultante sea lo más pequeña posible.

PROXIMIDAD. Grado de similitud o diferencia que existe entre parejas de individuos, objetos y variables.

PRUEBA DE DOS COLAS O BILATERAL. Método donde la hipótesis alternativa especifica una desviación a partir de la hipótesis nula en las dos direcciones. La región crítica o de rechazo se localiza en ambos extremos de la distribución de la estadística de prueba.

PRUEBA DE HIPÓTESIS. Un enfoque para la inferencia estadística que conduce a una decisión para rechazar o no la hipótesis nula.

PRUEBA DE INTERVALOS DE DUNCAN. Se ordenan las "r" medias muestrales en forma creciente de magnitud, y después se calculan las deferencias entre las medias muestrales adyacentes.

PRUEBA DE NEWMAN-KEULS. Método que ordena las "r" medias muestrales de forma decreciente según su magnitud y su nivel de significancia para cada conjunto de medias separadas "r" pasos es siempre α.

PRUEBA DE SCHEFFÉ. Se usa en la distribución F y permite controlar la tasa de error para el conjunto total de comparaciones que se plantean con j medias. Busca determinar si las diferencias entre 2 promedios es lo suficientemente grande para considerarse diferentes.

PRUEBA DE UNA COLA O DE UN EXTREMO. Prueba donde la hipótesis alternativa especifica una desviación de la hipótesis nula sólo en una dirección. La región crítica o de rechazo se localiza en un extremo de la distribución de la prueba estadística.

PRUEBA ESTADÍSTICA. Procedimiento empleado para probar una hipótesis nula (por ejemplo, prueba *t*, prueba ji cuadrada, prueba *F*).

PRUEBA T DE TUKEY. Una prueba *a posteriori* para hacer comparaciones apareadas múltiples entre medias y después de obtener una prueba ***F*** significativa en el análisis de varianza. Es el método más recomendado por los estadígrafos.

PRUEBAS POST-HOC. Tipo particular de contraste que se emplea para saber qué media difiere de otra. Métodos para comparar medias después del Análisis de varianza.

PUNTAJES O PUNTUACIONES. Nueva dimensión o variable que surge a partir del análisis de factores. Datos con los que se manejaría cada factor como variable.

RAÍZ CARACTERÍSTICA. Variación explicada por cada factor.

RAZÓN *F* . Aquella que se utiliza en el análisis de varianza, entre otras pruebas, para comparar la magnitud de dos estimaciones de la varianza de la población y determina si ambas estimaciones son aproximadamente iguales; en el análisis de varianza, se emplea la razón de la varianza entre columnas con la varianza dentro de columnas.

RAZÓN F CALCULADA. Estadístico que se usa para probar la significancia de la regresión como un todo.

REGIÓN CRÍTICA. Región (o conjunto de valores) donde debe ocurrir una prueba estadística para rechazar la hipótesis nula.

REGRESIÓN LOGÍSTICA. Estudia si una variable binomial (o dicotómica) depende o no de una o unas variables métricas.
REGRESIÓN MÚLTIPLE. Proceso estadístico por medio del cual se utilizan, algunas variables para predecir otra variable.
REGRESIÓN. (de **Y** en **X**) proceso por el cual se determina una ecuación para predecir **Y** a partir de **X**. Proceso general de predecir una variable a partir de otra con medios estadísticos, usando datos anteriores.
RELACIÓN CURVILÍNEA. Nexo de dos variables que es descrito por una línea curva.
RELACIÓN DIRECTA. Dependencia total entre 2 variables cuando una de ellas aumenta, la otra también lo hace en idéntica proporción.
RELACIÓN INVERSA: Relación entre dos variables en la cual, al aumentar la variable independiente, disminuye la variable dependiente.
RELACIÓN LINEAL. Tipo particular de asociación entre dos variables, que puede ser descrita matemáticamente con una recta.
RELACIÓN O ASOCIACIÓN INVERSA. Dependencia total entre 2 variables cuando una de ellas aumenta, la otra disminuye en idéntica proporción.
RESIDUAL. Diferencia entre el valor probable (predicción) y el valor real de la variable dependiente (resultado o respuesta) en regresión.
SEMILLA INICIAL. Promedios de los grupos con los cuales se inicia un análisis.
SIMILARIDAD. Las proximidades se interpretan en términos de lo parecido.
SUMA DE CUADRADOS DENTRO DE LAS MUESTRAS. Suma de cuadrados debida al error muestral.
SUMA DE CUADRADOS ENTRE LAS MUESTRAS. Suma de cuadrados debida al factor.
SUMA DE CUADRADOS. Representa la varianza de los datos de la variable sin considerar diferencia de grupos.
TABLA DE CONTINGENCIA. La que tiene renglones R y columnas C. cada renglón corresponde a un nivel de una variable cada columna, a un nivel de otra variable. Las partes del cuerpo de las tablas son las frecuencias con que ocurre cada combinación de variables.
TABLA DEL ANÁLISIS DE LA VARIANZA. Tabla donde se recogen todos los datos necesarios para realizar el contraste con el ANOVA. En el mismo sentido hablamos de la tabla del ANCOVA o del MANOVA.
TÉCNICA DE ANÁLISIS DE VARIANZA. En esta técnica se manejan los efectos de 2 factores, cada uno de ellos con un número diferente de niveles.
TÉCNICA DE K-PROMEDIOS. Agrupa los elementos para obtener el promedio de cada característica usada al segmentar en cada grupo y verificar que la asignación de los elementos en los conglomerados sea correcta.
TÉCNICAS DE MODELADO. Métodos con que se decide que variables incluir en un modelo de regresión y las diferentes maneras de incluirlas.
TÉCNICAS ESTRUCTURALES O INTERDEPENDIENTES. A diferencia de las técnicas funcionales, que buscan relaciones causa-efecto entre las variables, éstas buscan algo en común entre las variables o personas para unirlas y así poder resumir información.
TÉCNICAS FUNCIONALES O DEPENDIENTES. Son técnicas que implican un modelo, ecuación o función formada por las variables involucradas. En ellos intervienen variables dependientes e independientes.
TRANSFORMACIONES. Manipulaciones matemáticas para darle a una variable una forma diferente, de modo que se ajuste a las curvas y también a las líneas por regresión.
UTILIDAD. Valor asignado a cada uno de los niveles de los atributos escogidos para el producto.
VALOR CRÍTICO Un valor que es el punto divisorio entre la región en la que se rechaza la hipótesis nula, y la región en la que no se rechaza. En una prueba de una cola hay sólo un valor crítico. En una prueba de dos colas hay dos valores críticos, uno en cada cola. Cantidad que una prueba estadística debe exceder (en un sentido de valor absoluto) para poder rechazar la hipótesis nula.
VALOR *p* Probabilidad de encontrar, para el estadístico de prueba, un valor tan extremo o más que el obtenido con los datos muestrales, dado que la hipótesis nula sea verdadera.
VALORACIÓN PREDICTIVA DEL MODELO. Verificación del grado de error para efectuar pronósticos.

VALORES ATÍPICOS. Son valores o muy grandes o muy chicos con respecto a los demás o a la mayoría de los residuales o errores; son valores muy alejados de todas las observaciones.

VARIABLE DEPENDIENTE. Variable en estudio o variable respuesta, es decir a variable que nos interesa medir al aplicar los tratamientos como niveles del factor. Dicho factor sería independiente y el objeto del ANOVA es medir la influencia del factor (independiente) sobre la variable dependiente. Las variables que se asocian a los tratamientos son concreciones, para esos tratamientos, de la variable dependiente general. De tal manera que decir que esas variables son homogéneas equivale a decir que no existe influencia del factor sobre la variable dependiente.

VARIABLE FICTICIA. Aquella que toma el valor 0 o 1, permitiéndonos incluir en un modelo de regresión factores cualitativos tales como sexo, estado civil y nivel de escolaridad.

VARIABLE INDEPENDIENTE. La variable, o variables, conocidas en el análisis de regresión. . Variable explicatoria o "predictora" en un estudio. Se conoce en ocasiones como un factor en **ANOVA**.

VARIABLE LATENTE. Variable que no puede ser medida u observada directamente, pero sí representada o medida por varias variables llamadas *indicadores*. Sirve para reducir a un número menor de variables.

VARIABLE. Símbolo que puede tomar un valor cualquiera y representar a un conjunto determinado de datos. En un estudio, característica de interés que tiene valores diferentes para distintos sujetos u objetos.

VARIABLES BLOQUE. Variables que aparecen de forma natural y obligada sin interactuar con el factor que se estudia.

VARIABLES CANTITATIVAS O MÉTRICAS. Identifican a un objeto o persona en forma cuantitativa o numérica, pues es diferente de otro en cantidad o grado.

VARIABLES CUALITATIVAS O NO MÉTRICAS. Miden cualidades que no se pueden medir porque las observaciones sobre ellas sólo se pueden clasificar o enumerar.

VARIABLES POBLACIONALES. Las representamos por ξ_i y están asociadas a los conjuntos o grupos de valores que se obtienen aplicando los tratamientos A_i. Por ello podemos hablar de variables de tratamiento o grupos de tratamiento.

VARIACIÓN EXPLICADA (SS_B) ENTRE GRUPOS. Esta variación muestra lo distinto que son los datos y resulta de comparar cada promedio de cada nivel con el promedio general de los datos.

VARIACIÓN NO EXPLICADA (SS_W) DENTRO DE LOS GRUPOS. Es la diferencia de los datos dentro del mismo nivel de factor y es el resultado de la comparación de cada dato de cada nivel de factor con respecto a su correspondiente promedio de nivel de factor.

VARIACIÓN TOTAL (SS_T). Se refiere a lo distinto que son los datos sin clasificarlos.

VARIANCIA DENTRO DE TRATAMIENTOS. Estimación de la variancia de la población que se basa en las variancias dentro de las muestras o tratamientos *k*, que usa un promedio ponderado de k variancias muéstrales.

VARIANCIA ENTRE TRATAMIENTOS. Estimación de la variancia de la población derivada de la variancia entre las medias muéstrales.

VARIANZA. Es el promedio de una suma de cuadrados, es decir, es la suma de cuadrados dividido entre los grados de libert

APÉNDICE **TABLAS.**

SECCIÓN **1**

TABLA A1 PUNTOS PORCENTUALES DEL RANGO STUDENTIZADO. PARTE 1

g.l. del error	α	t= número de niveles del tratamiento									
		2	3	4	5	6	7	8	9	10	11
2	.05	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.40
	.01	13.90	19.02	22.56	25.37	27.76	29.86	31.73	33.41	34.93	36.29
3	.05	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72
	.01	8.26	10.62	12.17	13.32	14.24	15.00	15.65	16.21	16.71	17.16
4	.05	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03
	.01	6.51	8.12	9.17	9.96	10.58	11.10	11.54	11.92	12.26	12.57
5	.05	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17
	.01	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48
6	.05	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65
	.01	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30

7	.05	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30
	.01	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55
8	.05	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05
	.01	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03
9	.05	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87
	.01	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65
10	.05	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72
	.01	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36
11	.05	3.11	3.82	4.26	4.57	4.82	5.03	5.30	5.35	5.49	5.61
	.01	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13
12	.05	3.08	3.77	4.20	4.52	4.75	4.95	5.12	5.27	5.39	5.51
	.01	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94
13	.05	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43
	.01	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79
14	.05	3,03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36
	.01	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66

TABLA A1 PUNTOS PORCENTUALES DEL RANGO STUDENTIZADO. PARTE 2

g.l. del error	α	t= número de niveles del tratamiento									
		2	3	4	5	6	7	8	9	10	11
15	.05	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31
	.01	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55
16	.05	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26
	.01	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46
17	.05	2.97	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21
	.01	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38
18	.05	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17
	.01	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31
19	.05	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14
	.01	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25
20	.05	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	4.01	5.11
	.01	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19
21	.05	2.94	3.56	4.21	4.42	4.60	4.74	4.87	4.98	5.08	5.17
	.01	4.00	4.61	4.99	5.26	5.47	5.65	5.79	5.92	6.04	6.14
22	.05	2.93	3.55	3.93	4.20	4.41	4.58	4.72	4.85	4.96	5.06
	.01	3.99	4.59	4.96	5.22	5.43	5.61	5.75	5.88	5.99	6.10
23	.05	2.93	3.54	3.91	4.18	4.39	4.56	4.70	4.63	4.94	5.03
	.01	3.97	4.57	4.93	5.20	5.40	5.57	5.72	5.84	5.95	6.05
24	.05	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01

	.01	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02
25	.05	2.91	3.52	3.89	4.15	4.36	4.53	4.67	4.79	4.90	4.99
	.01	3.94	4.53	4.89	5.14	5.35	5.51	5.65	5.78	5.89	5.98
26	.05	2.91	3.51	3.88	4.14	4.35	4.51	4.65	4.77	4.88	4.98
	.01	3.94	4.53	4.89	5.14	5.35	5.51	5.65	5.78	5.89	5.98
27	.05	2.90	3.51	3.87	4.13	4.33	4.50	4.64	4.76	4.86	4.96
	.01	3.92	4.49	4.85	5.10	5.30	5.46	5.60	5.72	5.83	5.92
28	.05	2.90	3.50	3.86	4.12	4.32	4.49	4.62	4.74	4.85	4.94
	.01	3.90	4.48	4.83	5.08	5.28	5.44	5.58	5.70	5.80	5.90
29	.05	2.89	3.49	3.85	4.11	4.31	4.47	4.61	4.73	4.84	4.93
	.01	3.90	4.47	4.81	5.06	5.26	5.42	5.56	5.67	5.78	5.87
30	.05	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92
	.01	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85
31	.05	2.88	3.48	3.84	4.09	4.29	4.45	4.59	4.71	4.81	4.90
	.01	3.88	4.44	4.79	5.03	5.23	5.38	5.52	5.63	5.74	5.83
32	.05	2.88	3.48	3.83	4.09	4.28	4.45	4.58	4.70	4.80	4.89
	.01	3.87	4.43	4.77	5.02	5.21	5.37	5.50	5.61	5.72	5.81
33	.05	2.88	3.47	3.83	4.08	4.28	4.44	4.57	4.69	4.79	4.88
	.01	3.87	4.42	4.76	5.00	5.20	5.35	5.48	5.60	5.70	5.79

TABLA A1 PUNTOS PORCENTUALES DEL RANGO STUDENTIZADO. PARTE 3

g.l. del error	α	t= número de niveles del tratamiento									
		2	3	4	5	6	7	8	9	10	11
34	.05	2.87	3.47	3.82	4.07	4.27	4.43	4.56	4.68	4.78	4.87
	.01	3.86	4.41	4.75	4.99	5.18	5.34	5.47	5.58	5.68	5.77
35	.05	2.87	3.46	3.81	4.07	4.26	4.42	4.56	4.67	4.77	4.86
	.01	3.85	4.40	4.74	4.98	5.17	5.32	5.45	5.57	5.67	5.75
36	.05	2.87	3.46	3.81	4.06	4.25	4.41	4.55	4.66	4.76	4.85
	.01	3.85	4.40	4.73	4.97	5.16	5.31	5.44	5.55	5.65	5.74
37	.05	2.87	3.45	3.80	4.05	4.25	4.41	4.54	4.66	4.76	4.85
	.01	3.84	4.39	4.72	4.96	5.15	5.30	5.43	5.54	5.64	5.72
38	.05	2.86	3.45	3.80	4.05	4.24	4.40	4.53	4.65	4.75	4.84
	.01	3.83	4.38	4.71	4.95	5.13	5.29	5.41	5.53	5.62	5.71
39	.05	2.86	3.45	3.79	4.04	4.24	4.39	4.53	4.64	4.74	4.83
	.01	3.83	4.37	4.70	4.94	5.12	5.28	5.40	5.51	5.61	5.70
40	.05	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82
	.01	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69
41	.05	2.86	3.44	3.79	4.03	4.23	4.38	4.51	4.63	4.73	4.82
	.01	3.82	4.36	4.69	4.92	5.11	5.26	5.38	5.49	5.59	5.67
42	.05	2.85	3.44	3.78	4.03	4.22	4.38	4.51	4.62	4.72	4.81
	.01	3.82	4.35	4.68	4.91	5.10	5.25	5.37	5.48	5.58	5.66
43	.05	2.85	3.43	3.78	4.03	4.22	4.37	4.50	4.62	4.72	4.80

	.01	3.81	4.35	4.67	4.91	5.09	5.24	5.36	5.47	5.57	5.65
44	.05	2.85	3.43	3.78	4.02	4.21	4.37	4.50	4.61	4.71	4.80
	.01	3.81	4.34	4.67	4.90	5.08	5.23	5.35	5.46	5.56	5.64
45	.05	2.85	3.43	3.77	4.02	4.21	4.36	4.49	4.61	4.70	4.79
	.01	3.80	4.34	4.66	4.89	5.07	5.22	5.34	5.45	5.55	5.63
50	.05	2.84	3.42	3.76	4.00	4.19	4.34	4.47	4.58	4.68	4.77
	.01	3.79	4.32	4.63	4.86	5.04	5.19	5.31	5.41	5.51	5.59
60	.05	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73
	.01	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53
120	.05	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64
	.01	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.37
∞	.05	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55
	.01	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23

APÉNDICE TABLAS.

SECCIÓN 2

TABLA A2 VALORES CRÍTICOS DE t. PARTE 1

Grados de libertad	Áreas de la cola superior					
	0.25	0.10	0.05	0.025	0.01	0.005
1	1.00000	3.07768	6.31375	12.70620	31.82052	63.65674
2	0.81650	1.88562	2.91999	4.30265	6.96456	9.92484
3	0.76489	1.63774	2.35336	3.18245	4.54070	5.84091
4	0.74070	1.53321	2.13185	2.77645	3.74695	4.60409
5	0.72669	1.47588	2.01505	2.57058	3.36493	4.03214
6	0.71756	1.43976	1.94318	2.44691	3.14267	3.70743
7	0.71114	1.41492	1.89458	2.36462	2.99795	3.49948
8	0.70639	1.39682	1.85955	2.30600	2.89646	3.35539
9	0.70272	1.38303	1.83311	2.26216	2.82144	3.24984
10	0.69981	1.37218	1.81246	2.22814	2.76377	3.16927
11	0.69745	1.36343	1.79588	2.20099	2.71808	3.10581
12	0.69548	1.35622	1.78229	2.17881	2.68100	3.05454
13	0.69383	1.35017	1.77093	2.16037	2.65031	3.01228

14	0.69242	1.34503	1.76131	2.14479	2.62449	2.97684
15	0.69120	1.34061	1.75305	2.13145	2.60248	2.94671
16	0.69013	1.33676	1.74588	2.11991	2.58349	2.92078
17	0.68920	1.33338	1.73961	2.10982	2.56693	2.89823
18	0.68836	1.33039	1.73406	2.10092	2.55238	2.87844
19	0.68762	1.32773	1.72913	2.09302	2.53948	2.86093
20	0.68695	1.32534	1.72472	2.08596	2.52798	2.84534
21	0.68635	1.32319	1.72074	2.07961	2.51765	2.83136
22	0.68581	1.32124	1.71714	2.07387	2.50832	2.81876
23	0.68531	1.31946	1.71387	2.06866	2.49987	2.80734
24	0.68485	1.31784	1.71088	2.06390	2.49216	2.79694
25	0.68443	1.31635	1.70814	2.05954	2.48511	2.78744
26	0.68404	1.31497	1.70562	2.05553	2.47863	2.77871
27	0.68368	1.31370	1.70329	2.05183	2.47266	2.77068
28	0.68335	1.31253	1.70113	2.04841	2.46714	2.76326
29	0.68304	1.31143	1.69913	2.04523	2.46202	2.75639
30	0.68276	1.31042	1.69726	2.04227	2.45726	2.75000
31	0.68249	1.30946	1.69552	2.03951	2.45282	2.74404
32	0.68223	1.30857	1.69389	2.03693	2.44868	2.73848
33	0.68200	1.30774	1.69236	2.03452	2.44479	2.73328
34	0.68177	1.30695	1.69092	2.03224	2.44115	2.72839
35	0.68156	1.30621	1.68957	2.03011	2.43772	2.72381

Fuente: Valores generados con Microsoft Excel.

TABLA A2 VALORES CRÍTICOS DE t. PARTE 2

Grados de libertad	Áreas de la cola superior					
	0.25	0.10	0.05	0.025	0.01	0.005
36	0.68137	1.30551	1.68830	2.02809	2.43449	2.71948
37	0.68118	1.30485	1.68709	2.02619	2.43145	2.71541
38	0.68100	1.30423	1.68595	2.02439	2.42857	2.71156
39	0.68083	1.30364	1.68488	2.02269	2.42584	2.70791
40	0.68067	1.30308	1.68385	2.02108	2.42326	2.70446
41	0.68052	1.30254	1.68288	2.01954	2.42080	2.70118
42	0.68038	1.30204	1.68195	2.01808	2.41847	2.69807
43	0.68024	1.30155	1.68107	2.01669	2.41625	2.69510
44	0.68011	1.30109	1.68023	2.01537	2.41413	2.69228
45	0.67998	1.30065	1.67943	2.01410	2.41212	2.68959
46	0.67986	1.30023	1.67866	2.01290	2.41019	2.68701
47	0.67975	1.29982	1.67793	2.01174	2.40835	2.68456
48	0.67964	1.29944	1.67722	2.01063	2.40658	2.68220
49	0.67953	1.29907	1.67655	2.00958	2.40489	2.67995
50	0.67943	1.29871	1.67591	2.00856	2.40327	2.67779
51	0.67933	1.29837	1.67528	2.00758	2.40172	2.67572
52	0.67924	1.29805	1.67469	2.00665	2.40022	2.67373
53	0.67915	1.29773	1.67412	2.00575	2.39879	2.67182
54	0.67906	1.29743	1.67356	2.00488	2.39741	2.66998
55	0.67898	1.29713	1.67303	2.00404	2.39608	2.66822
56	0.67890	1.29685	1.67252	2.00324	2.39480	2.66651
57	0.67882	1.29658	1.67203	2.00247	2.39357	2.66487
58	0.67874	1.29632	1.67155	2.00172	2.39238	2.66329
59	0.67867	1.29607	1.67109	2.00100	2.39123	2.66176

60	0.67860	1.29582	1.67065	2.00030	2.39012	2.66028
61	0.67853	1.29558	1.67022	1.99962	2.38905	2.65886
62	0.67847	1.29536	1.66980	1.99897	2.38801	2.65748
63	0.67840	1.29513	1.66940	1.99834	2.38701	2.65615
64	0.67834	1.29492	1.66901	1.99773	2.38604	2.65485
65	0.67828	1.29471	1.66864	1.99714	2.38510	2.65360
66	0.67823	1.29451	1.66827	1.99656	2.38419	2.65239
67	0.67817	1.29432	1.66792	1.99601	2.38330	2.65122
68	0.67811	1.29413	1.66757	1.99547	2.38245	2.65008
69	0.67806	1.29394	1.66724	1.99495	2.38161	2.64898
70	0.67801	1.29376	1.66691	1.99444	2.38081	2.64790
71	0.67796	1.29359	1.66660	1.99394	2.38002	2.64686
72	0.67791	1.29342	1.66629	1.99346	2.37926	2.64585
73	0.67787	1.29326	1.66600	1.99300	2.37852	2.64487
74	0.67782	1.29310	1.66571	1.99254	2.37780	2.64391
75	0.67778	1.29294	1.66543	1.99210	2.37710	2.64298
76	0.67773	1.29279	1.66515	1.99167	2.37642	2.64208
77	0.67769	1.29264	1.66488	1.99125	2.37576	2.64120
78	0.67765	1.29250	1.66462	1.99085	2.37511	2.64034
79	0.67761	1.29236	1.66437	1.99045	2.37448	2.63950
80	0.67757	1.29222	1.66412	1.99006	2.37387	2.63869

Fuente: Valores generados con Microsoft Excel.

TABLA A2 VALORES CRÍTICOS DE t. PARTE 3

Grados de libertad	Áreas de la cola superior					
	0.25	0.10	0.05	0.025	0.01	0.005
81	0.67753	1.29209	1.66388	1.98969	2.37327	2.63790
82	0.67749	1.29196	1.66365	1.98932	2.37269	2.63712
83	0.67746	1.29183	1.66342	1.98896	2.37212	2.63637
84	0.67742	1.29171	1.66320	1.98861	2.37156	2.63563
85	0.67739	1.29159	1.66298	1.98827	2.37102	2.63491
86	0.67735	1.29147	1.66277	1.98793	2.37049	2.63421
87	0.67732	1.29136	1.66256	1.98761	2.36998	2.63353
88	0.67729	1.29125	1.66235	1.98729	2.36947	2.63286
89	0.67726	1.29114	1.66216	1.98698	2.36898	2.63220
90	0.67723	1.29103	1.66196	1.98667	2.36850	2.63157
91	0.67720	1.29092	1.66177	1.98638	2.36803	2.63094
92	0.67717	1.29082	1.66159	1.98609	2.36757	2.63033
93	0.67714	1.29072	1.66140	1.98580	2.36712	2.62973
94	0.67711	1.29062	1.66123	1.98552	2.36667	2.62915
95	0.67708	1.29053	1.66105	1.98525	2.36624	2.62858
96	0.67705	1.29043	1.66088	1.98498	2.36582	2.62802
97	0.67703	1.29034	1.66071	1.98472	2.36541	2.62747
98	0.67700	1.29025	1.66055	1.98447	2.36500	2.62693
99	0.67698	1.29016	1.66039	1.98422	2.36461	2.62641
100	0.67695	1.29007	1.66023	1.98397	2.36422	2.62589
101	0.67693	1.28999	1.66008	1.98373	2.36384	2.62539
102	0.67690	1.28991	1.65993	1.98350	2.36346	2.62489
103	0.67688	1.28982	1.65978	1.98326	2.36310	2.62441
104	0.67686	1.28974	1.65964	1.98304	2.36274	2.62393

105	0.67683	1.28967	1.65950	1.98282	2.36239	2.62347
106	0.67681	1.28959	1.65936	1.98260	2.36204	2.62301
107	0.67679	1.28951	1.65922	1.98238	2.36170	2.62256
108	0.67677	1.28944	1.65909	1.98217	2.36137	2.62212
109	0.67675	1.28937	1.65895	1.98197	2.36105	2.62169
110	0.67673	1.28930	1.65882	1.98177	2.36073	2.62126
115	0.67663	1.28896	1.65821	1.98081	2.35921	2.61926
120	0.67654	1.28865	1.65765	1.97993	2.35782	2.61742
125	0.67646	1.28836	1.65714	1.97912	2.35655	2.61573
130	0.67638	1.28810	1.65666	1.97838	2.35537	2.61418
135	0.67631	1.28785	1.65622	1.97769	2.35429	2.61274
140	0.67625	1.28763	1.65581	1.97705	2.35328	2.61140
145	0.67619	1.28742	1.65543	1.97646	2.35234	2.61016
150	0.67613	1.28722	1.65508	1.97591	2.35146	2.60900
170	0.67594	1.28655	1.65387	1.97402	2.34848	2.60506
200	0.67572	1.28580	1.65251	1.97190	2.34514	2.60063
∞	0.67449	1.28155	1.64486	1.95997	2.32635	2.57583

Fuente: Valores generados con Microsoft Excel.

TABLA A3 VALORES CRÍTICOS DE F $\alpha = 0.05$. PARTE 1

Denominador gl_2	Grados de libertad en el numerador gl_1									
	1	2	3	4	5	6	7	8	9	10
1	161.4476	199.5000	215.7073	224.5832	230.1618	233.9860	236.7684	238.8826	240.5432	241.8817
2	18.51282	19.00000	19.16429	19.24679	19.29641	19.32953	19.35322	19.37099	19.38483	19.39590
3	10.12796	9.55209	9.27663	9.11718	9.01346	8.94065	8.88674	8.84524	8.81230	8.78552
4	7.70865	6.94427	6.59138	6.38823	6.25606	6.16313	6.09421	6.04104	5.99878	5.96437
5	6.60789	5.78614	5.40945	5.19217	5.05033	4.95029	4.87587	4.81832	4.77247	4.73506
6	5.98738	5.14325	4.75706	4.53368	4.38737	4.28387	4.20666	4.14680	4.09902	4.05996
7	5.59145	4.73741	4.34683	4.12031	3.97152	3.86597	3.78704	3.72573	3.67667	3.63652
8	5.31766	4.45897	4.06618	3.83785	3.68750	3.58058	3.50046	3.43810	3.38813	3.34716
9	5.11736	4.25649	3.86255	3.63309	3.48166	3.37375	3.29275	3.22958	3.17889	3.13728
10	4.96460	4.10282	3.70826	3.47805	3.32583	3.21717	3.13546	3.07166	3.02038	2.97824
11	4.84434	3.98230	3.58743	3.35669	3.20387	3.09461	3.01233	2.94799	2.89622	2.85362
12	4.74723	3.88529	3.49029	3.25917	3.10588	2.99612	2.91336	2.84857	2.79638	2.75339
13	4.66719	3.80557	3.41053	3.17912	3.02544	2.91527	2.83210	2.76691	2.71436	2.67102
14	4.60011	3.73889	3.34389	3.11225	2.95825	2.84773	2.76420	2.69867	2.64579	2.60216
15	4.54308	3.68232	3.28738	3.05557	2.90129	2.79046	2.70663	2.64080	2.58763	2.54372
16	4.49400	3.63372	3.23887	3.00692	2.85241	2.74131	2.65720	2.59110	2.53767	2.49351
17	4.45132	3.59153	3.19678	2.96471	2.81000	2.69866	2.61430	2.54796	2.49429	2.44992
18	4.41387	3.55456	3.15991	2.92774	2.77285	2.66130	2.57672	2.51016	2.45628	2.41170
19	4.38075	3.52189	3.12735	2.89511	2.74006	2.62832	2.54353	2.47677	2.42270	2.37793
20	4.35124	3.49283	3.09839	2.86608	2.71089	2.59898	2.51401	2.44706	2.39281	2.34788
21	4.32479	3.46680	3.07247	2.84010	2.68478	2.57271	2.48758	2.42046	2.36605	2.32095

22	4.30095	3.44336	3.04913	2.81671	2.66127	2.54906	2.46377	2.39650	2.34194	2.29670
23	4.27934	3.42213	3.02800	2.79554	2.64000	2.52766	2.44223	2.37481	2.32011	2.27473
24	4.25968	3.40283	3.00879	2.77629	2.62065	2.50819	2.42263	2.35508	2.30024	2.25474
25	4.24170	3.38519	2.99124	2.75871	2.60299	2.49041	2.40473	2.33706	2.28210	2.23647
26	4.22520	3.36902	2.97515	2.74259	2.58679	2.47411	2.38831	2.32053	2.26545	2.21972
27	4.21001	3.35413	2.96035	2.72777	2.57189	2.45911	2.37321	2.30531	2.25013	2.20429
28	4.19597	3.34039	2.94669	2.71408	2.55813	2.44526	2.35926	2.29126	2.23598	2.19004
29	4.18296	3.32765	2.93403	2.70140	2.54539	2.43243	2.34634	2.27825	2.22287	2.17684
30	4.17088	3.31583	2.92228	2.68963	2.53355	2.42052	2.33434	2.26616	2.21070	2.16458
31	4.15962	3.30482	2.91133	2.67867	2.52254	2.40943	2.32317	2.25491	2.19936	2.15316
32	4.14910	3.29454	2.90112	2.66844	2.51225	2.39908	2.31274	2.24440	2.18877	2.14249
33	4.13925	3.28492	2.89156	2.65887	2.50264	2.38939	2.30298	2.23456	2.17886	2.13250
34	4.13002	3.27590	2.88260	2.64989	2.49362	2.38031	2.29383	2.22534	2.16956	2.12314
35	4.12134	3.26742	2.87419	2.64147	2.48514	2.37178	2.28524	2.21668	2.16083	2.11434
36	4.11317	3.25945	2.86627	2.63353	2.47717	2.36375	2.27714	2.20852	2.15261	2.10605
37	4.10546	3.25192	2.85880	2.62605	2.46965	2.35618	2.26951	2.20083	2.14485	2.09824
38	4.09817	3.24482	2.85174	2.61899	2.46255	2.34903	2.26230	2.19356	2.13753	2.09086
39	4.09128	3.23810	2.84507	2.61231	2.45583	2.34226	2.25549	2.18668	2.13060	2.08387
40	4.08475	3.23173	2.83875	2.60597	2.44947	2.33585	2.24902	2.18017	2.12403	2.07725
50	4.03431	3.18261	2.79001	2.55718	2.40041	2.28644	2.19920	2.12992	2.07335	2.02614
60	4.00119	3.15041	2.75808	2.52522	2.36827	2.25405	2.16654	2.09697	2.04010	1.99259
120	3.92012	3.07178	2.68017	2.44724	2.28985	2.17501	2.08677	2.01643	1.95876	1.91046
∞	3.84155	2.99582	2.60500	2.37202	2.21419	2.09869	2.00968	1.93851	1.87998	1.83080

Fuente: Valores generados con Microsoft Excel.

TABLA A3 VALORES CRÍTICOS DE F $\alpha = 0.05$. PARTE 2

Denominador gl_2	Grados de libertad en el numerador gl_1									
	11	12	13	14	15	16	17	18	19	20
1	242.9834	243.9060	244.6898	245.3639	245.9499	246.4639	246.9184	247.3232	247.6860	248.0130
2	19.40496	19.41251	19.41890	19.42438	19.42914	19.43329	19.43696	19.44022	19.44314	19.44577
3	8.76333	8.74464	8.72868	8.71490	8.70287	8.69229	8.68290	8.67452	8.66699	8.66019
4	5.93581	5.91173	5.89114	5.87335	5.85781	5.84412	5.83197	5.82112	5.81136	5.80254
5	4.70397	4.67770	4.65523	4.63577	4.61876	4.60376	4.59044	4.57853	4.56782	4.55813
6	4.02744	3.99994	3.97636	3.95593	3.93806	3.92228	3.90826	3.89571	3.88441	3.87419
7	3.60304	3.57468	3.55034	3.52923	3.51074	3.49441	3.47988	3.46686	3.45514	3.44452
8	3.31295	3.28394	3.25902	3.23738	3.21841	3.20163	3.18670	3.17332	3.16125	3.15032
9	3.10249	3.07295	3.04755	3.02547	3.00610	2.98897	2.97370	2.96000	2.94765	2.93646
10	2.94296	2.91298	2.88717	2.86473	2.84502	2.82757	2.81201	2.79805	2.78545	2.77402
11	2.81793	2.78757	2.76142	2.73865	2.71864	2.70091	2.68510	2.67090	2.65808	2.64645
12	2.71733	2.68664	2.66018	2.63712	2.61685	2.59888	2.58284	2.56843	2.55541	2.54359
13	2.63465	2.60366	2.57693	2.55362	2.53311	2.51492	2.49867	2.48407	2.47087	2.45888
14	2.56550	2.53424	2.50726	2.48373	2.46300	2.44461	2.42818	2.41340	2.40004	2.38790
15	2.50681	2.47531	2.44811	2.42436	2.40345	2.38488	2.36827	2.35333	2.33982	2.32754
16	2.45637	2.42466	2.39725	2.37332	2.35222	2.33348	2.31672	2.30164	2.28798	2.27557
17	2.41256	2.38065	2.35306	2.32895	2.30769	2.28880	2.27189	2.25667	2.24289	2.23035
18	2.37416	2.34207	2.31430	2.29003	2.26862	2.24959	2.23255	2.21720	2.20330	2.19065
19	2.34021	2.30795	2.28003	2.25561	2.23406	2.21490	2.19773	2.18226	2.16825	2.15550
20	2.30999	2.27758	2.24951	2.22496	2.20327	2.18398	2.16670	2.15112	2.13701	2.12416
21	2.28292	2.25036	2.22216	2.19747	2.17567	2.15626	2.13887	2.12319	2.10898	2.09603
22	2.25852	2.22583	2.19750	2.17269	2.15078	2.13126	2.11377	2.09799	2.08369	2.07066
23	2.23642	2.20361	2.17516	2.15024	2.12822	2.10860	2.09101	2.07515	2.06075	2.04764

24	2.21631	2.18338	2.15482	2.12980	2.10767	2.08796	2.07028	2.05433	2.03986	2.02666
25	2.19793	2.16489	2.13623	2.11111	2.08889	2.06909	2.05132	2.03529	2.02074	2.00747
26	2.18107	2.14793	2.11917	2.09395	2.07164	2.05176	2.03391	2.01780	2.00318	1.98984
27	2.16554	2.13230	2.10345	2.07815	2.05575	2.03579	2.01787	2.00169	1.98699	1.97359
28	2.15120	2.11787	2.08893	2.06354	2.04107	2.02103	2.00304	1.98678	1.97203	1.95856
29	2.13791	2.10449	2.07547	2.05000	2.02746	2.00735	1.98928	1.97297	1.95815	1.94462
30	2.12556	2.09206	2.06296	2.03742	2.01480	1.99462	1.97650	1.96012	1.94524	1.93165
31	2.11405	2.08048	2.05131	2.02569	2.00301	1.98276	1.96457	1.94813	1.93320	1.91956
32	2.10331	2.06966	2.04042	2.01474	1.99199	1.97168	1.95343	1.93694	1.92195	1.90826
33	2.09325	2.05954	2.03023	2.00448	1.98167	1.96130	1.94300	1.92645	1.91141	1.89767
34	2.08382	2.05004	2.02066	1.99486	1.97199	1.95157	1.93321	1.91660	1.90151	1.88773
35	2.07496	2.04111	2.01167	1.98581	1.96288	1.94241	1.92400	1.90735	1.89221	1.87838
36	2.06661	2.03270	2.00321	1.97729	1.95431	1.93378	1.91532	1.89862	1.88344	1.86956
37	2.05873	2.02477	1.99522	1.96925	1.94622	1.92564	1.90713	1.89039	1.87516	1.86124
38	2.05129	2.01728	1.98767	1.96165	1.93857	1.91794	1.89939	1.88260	1.86733	1.85338
39	2.04425	2.01018	1.98053	1.95445	1.93133	1.91066	1.89206	1.87523	1.85992	1.84593
40	2.03758	2.00346	1.97376	1.94764	1.92446	1.90375	1.88511	1.86824	1.85289	1.83886
50	1.98606	1.95153	1.92143	1.89493	1.87138	1.85031	1.83133	1.81413	1.79846	1.78412
60	1.95221	1.91740	1.88702	1.86024	1.83644	1.81511	1.79589	1.77845	1.76255	1.74798
120	1.86929	1.83370	1.80255	1.77503	1.75050	1.72846	1.70854	1.69043	1.67388	1.65868
∞	1.78874	1.75227	1.72025	1.69187	1.66649	1.64362	1.62287	1.60395	1.58661	1.57063

Fuente: Valores generados con Microsoft Excel.

TABLA A3 VALORES CRÍTICOS DE F $\alpha = 0.05$. PARTE 3

Denomi-	Grados de libertad en el numerador gl_1									
nador gl_2	21	22	23	24	25	30	40	60	120	∞
1	248.3093	248.5790	248.8255	249.0517	249.2600	250.0951	251.1431	252.1957	253.2528	254.3143
2	19.44815	19.45031	19.45228	19.45409	19.45575	19.46241	19.47074	19.47906	19.48739	19.49572
3	8.65402	8.64839	8.64324	8.63850	8.63414	8.61658	8.59441	8.57200	8.54935	8.52645
4	5.79453	5.78723	5.78054	5.77439	5.76872	5.74588	5.71700	5.68774	5.65811	5.62808
5	4.54933	4.54129	4.53393	4.52715	4.52090	4.49571	4.46379	4.43138	4.39845	4.36500
6	3.86489	3.85640	3.84862	3.84146	3.83484	3.80816	3.77429	3.73980	3.70467	3.66887
7	3.43487	3.42604	3.41795	3.41049	3.40361	3.37581	3.34043	3.30432	3.26745	3.22976
8	3.14037	3.13128	3.12293	3.11524	3.10813	3.07941	3.04278	3.00530	2.96692	2.92758
9	2.92626	2.91693	2.90837	2.90047	2.89318	2.86365	2.82593	2.78725	2.74752	2.70668
10	2.76360	2.75407	2.74532	2.73725	2.72978	2.69955	2.66086	2.62108	2.58012	2.53788
11	2.63584	2.62613	2.61720	2.60897	2.60136	2.57049	2.53091	2.49012	2.44802	2.40448
12	2.53281	2.52293	2.51386	2.50548	2.49773	2.46628	2.42588	2.38417	2.34099	2.29620
13	2.44794	2.43792	2.42870	2.42020	2.41232	2.38033	2.33918	2.29660	2.25241	2.20644
14	2.37681	2.36665	2.35731	2.34868	2.34069	2.30821	2.26635	2.22295	2.17781	2.13070
15	2.31632	2.30603	2.29657	2.28783	2.27973	2.24679	2.20428	2.16011	2.11406	2.06585
16	2.26423	2.25383	2.24425	2.23541	2.22721	2.19384	2.15071	2.10581	2.05890	2.00964
17	2.21890	2.20839	2.19871	2.18977	2.18148	2.14771	2.10400	2.05841	2.01066	1.96039
18	2.17909	2.16847	2.15870	2.14966	2.14129	2.10714	2.06289	2.01664	1.96810	1.91685
19	2.14383	2.13313	2.12326	2.11414	2.10569	2.07119	2.02641	1.97954	1.93024	1.87803
20	2.11240	2.10160	2.09165	2.08245	2.07392	2.03909	1.99382	1.94636	1.89632	1.84319
21	2.08419	2.07331	2.06328	2.05400	2.04540	2.01025	1.96452	1.91649	1.86574	1.81171
22	2.05873	2.04777	2.03767	2.02832	2.01964	1.98420	1.93802	1.88945	1.83802	1.78311

23	2.03563	2.02460	2.01442	2.00501	1.99627	1.96054	1.91394	1.86484	1.81276	1.75700
24	2.01458	2.00348	1.99324	1.98376	1.97496	1.93896	1.89195	1.84236	1.78964	1.73306
25	1.99532	1.98415	1.97385	1.96431	1.95545	1.91919	1.87180	1.82173	1.76840	1.71100
26	1.97763	1.96639	1.95603	1.94643	1.93751	1.90101	1.85325	1.80272	1.74879	1.69061
27	1.96131	1.95002	1.93959	1.92994	1.92097	1.88424	1.83613	1.78515	1.73065	1.67169
28	1.94622	1.93487	1.92439	1.91469	1.90567	1.86871	1.82026	1.76886	1.71380	1.65408
29	1.93222	1.92082	1.91029	1.90053	1.89147	1.85429	1.80552	1.75370	1.69811	1.63765
30	1.91920	1.90775	1.89716	1.88736	1.87825	1.84087	1.79179	1.73957	1.68345	1.62227
31	1.90706	1.89555	1.88492	1.87507	1.86592	1.82834	1.77896	1.72636	1.66973	1.60784
32	1.89571	1.88415	1.87348	1.86358	1.85439	1.81662	1.76696	1.71398	1.65686	1.59428
33	1.88507	1.87347	1.86275	1.85281	1.84358	1.80564	1.75569	1.70236	1.64475	1.58149
34	1.87508	1.86344	1.85268	1.84270	1.83342	1.79531	1.74510	1.69142	1.63334	1.56941
35	1.86569	1.85400	1.84320	1.83318	1.82387	1.78559	1.73512	1.68111	1.62258	1.55799
36	1.85683	1.84510	1.83427	1.82421	1.81486	1.77642	1.72570	1.67136	1.61239	1.54716
37	1.84847	1.83671	1.82583	1.81574	1.80636	1.76776	1.71680	1.66215	1.60274	1.53688
38	1.84057	1.82876	1.81785	1.80773	1.79831	1.75957	1.70838	1.65342	1.59359	1.52710
39	1.83308	1.82124	1.81029	1.80014	1.79069	1.75180	1.70039	1.64513	1.58489	1.51779
40	1.82598	1.81410	1.80312	1.79294	1.78346	1.74443	1.69280	1.63725	1.57661	1.50891
50	1.77095	1.75879	1.74753	1.73708	1.72734	1.68716	1.63368	1.57565	1.51147	1.43827
60	1.73459	1.72222	1.71077	1.70012	1.69019	1.64914	1.59427	1.53431	1.46727	1.38929
120	1.64467	1.63170	1.61966	1.60844	1.59796	1.55434	1.49520	1.42901	1.35189	1.25387
∞	1.55585	1.54213	1.52935	1.51740	1.50621	1.45921	1.39409	1.31817	1.22157	1.00776

Fuente: Valores generados con Microsoft Excel.

TABLA A4 VALORES CRÍTICOS DE F $\alpha = 0.01$. PARTE 1

Denominador gl_2	Grados de libertad en el numerador gl_1									
	1	2	3	4	5	6	7	8	9	10
1	4052.180	4999.500	5403.352	5624.583	5763.649	5858.986	5928.355	5981.070	6022.473	6055.846
2	98.50251	99.00000	99.16620	99.24937	99.29930	99.33259	99.35637	99.37421	99.38809	99.39920
3	34.11622	30.81652	29.45670	28.70990	28.23708	27.91066	27.67170	27.48918	27.34521	27.22873
4	21.19769	18.00000	16.69437	15.97702	15.52186	15.20686	14.97576	14.79889	14.65913	14.54590
5	16.25818	13.27393	12.05995	11.39193	10.96702	10.67225	10.45551	10.28931	10.15776	10.05102
6	13.74502	10.92477	9.77954	9.14830	8.74590	8.46613	8.26000	8.10165	7.97612	7.87412
7	12.24638	9.54658	8.45129	7.84665	7.46044	7.19140	6.99283	6.84005	6.71875	6.62006
8	11.25862	8.64911	7.59099	7.00608	6.63183	6.37068	6.17762	6.02887	5.91062	5.81429
9	10.56143	8.02152	6.99192	6.42209	6.05694	5.80177	5.61287	5.46712	5.35113	5.25654
10	10.04429	7.55943	6.55231	5.99434	5.63633	5.38581	5.20012	5.05669	4.94242	4.84915
11	9.64603	7.20571	6.21673	5.66830	5.31601	5.06921	4.88607	4.74447	4.63154	4.53928
12	9.33021	6.92661	5.95254	5.41195	5.06434	4.82057	4.63950	4.49937	4.38751	4.29605
13	9.07381	6.70096	5.73938	5.20533	4.86162	4.62036	4.44100	4.30206	4.19108	4.10027
14	8.86159	6.51488	5.56389	5.03538	4.69496	4.45582	4.27788	4.13995	4.02968	3.93940
15	8.68312	6.35887	5.41696	4.89321	4.55561	4.31827	4.14155	4.00445	3.89479	3.80494
16	8.53097	6.22624	5.29221	4.77258	4.43742	4.20163	4.02595	3.88957	3.78042	3.69093
17	8.39974	6.11211	5.18500	4.66897	4.33594	4.10151	3.92672	3.79096	3.68224	3.59307
18	8.28542	6.01290	5.09189	4.57904	4.24788	4.01464	3.84064	3.70542	3.59707	3.50816
19	8.18495	5.92588	5.01029	4.50026	4.17077	3.93857	3.76527	3.63052	3.52250	3.43382
20	8.09596	5.84893	4.93819	4.43069	4.10268	3.87143	3.69874	3.56441	3.45668	3.36819
21	8.01660	5.78042	4.87405	4.36882	4.04214	3.81173	3.63959	3.50563	3.39815	3.30983
22	7.94539	5.71902	4.81661	4.31343	3.98796	3.75830	3.58666	3.45303	3.34577	3.25761
23	7.88113	5.66370	4.76488	4.26357	3.93919	3.71022	3.53902	3.40569	3.29863	3.21060

24	7.82287	5.61359	4.71805	4.21845	3.89507	3.66672	3.49593	3.36287	3.25599	3.16807
25	7.76980	5.56800	4.67546	4.17742	3.85496	3.62717	3.45675	3.32394	3.21722	3.12941
26	7.72125	5.52633	4.63657	4.13996	3.81834	3.59108	3.42099	3.28840	3.18182	3.09411
27	7.67668	5.48812	4.60091	4.10562	3.78477	3.55799	3.38822	3.25583	3.14939	3.06175
28	7.63562	5.45294	4.56809	4.07403	3.75389	3.52756	3.35807	3.22587	3.11955	3.03199
29	7.59766	5.42045	4.53779	4.04487	3.72540	3.49947	3.33025	3.19822	3.09201	3.00452
30	7.56248	5.39035	4.50974	4.01788	3.69902	3.47348	3.30450	3.17262	3.06652	2.97909
31	7.52977	5.36239	4.48369	3.99281	3.67453	3.44934	3.28059	3.14886	3.04285	2.95548
32	7.49928	5.33634	4.45943	3.96948	3.65173	3.42688	3.25834	3.12675	3.02082	2.93351
33	7.47080	5.31203	4.43679	3.94770	3.63046	3.40591	3.23757	3.10611	3.00026	2.91300
34	7.44414	5.28928	4.41561	3.92733	3.61056	3.38631	3.21815	3.08681	2.98103	2.89381
35	7.41912	5.26794	4.39575	3.90824	3.59191	3.36793	3.19995	3.06872	2.96301	2.87583
36	7.39560	5.24789	4.37710	3.89031	3.57440	3.35068	3.18286	3.05173	2.94609	2.85895
37	7.37344	5.22902	4.35954	3.87343	3.55792	3.33444	3.16677	3.03574	2.93016	2.84305
38	7.35254	5.21122	4.34299	3.85752	3.54238	3.31913	3.15161	3.02067	2.91515	2.82807
39	7.33279	5.19441	4.32736	3.84250	3.52771	3.30468	3.13730	3.00644	2.90097	2.81392
40	7.31410	5.17851	4.31257	3.82829	3.51384	3.29101	3.12376	2.99298	2.88756	2.80055
50	7.17058	5.05661	4.19934	3.71955	3.40768	3.18643	3.02017	2.89001	2.78496	2.69814
60	7.07711	4.97743	4.12589	3.64905	3.33888	3.11867	2.95305	2.82328	2.71845	2.63175
120	6.85089	4.78651	3.94910	3.47953	3.17355	2.95585	2.79176	2.66291	2.55857	2.47208
∞	6.63515	4.60538	3.78182	3.31936	3.01744	2.80216	2.63951	2.51146	2.40751	2.32110

Fuente: Valores generados con Microsoft Excel.

TABLA A4 VALORES CRÍTICOS DE F $\alpha = 0.01$. PARTE 2

Denominador gl_2	Grados de libertad en el numerador gl_1									
	11	12	13	14	15	16	17	18	19	20
1	6083.316	6106.320	6125.864	6142.673	6157.284	6170.101	6181.434	6191.528	6200.575	6208.730
2	99.40828	99.41585	99.42226	99.42775	99.43251	99.43668	99.44035	99.44362	99.44654	99.44917
3	27.13257	27.05182	26.98306	26.92380	26.87219	26.82686	26.78671	26.75091	26.71878	26.68979
4	14.45228	14.37359	14.30650	14.24863	14.19820	14.15386	14.11457	14.07951	14.04803	14.01961
5	9.96265	9.88828	9.82481	9.77001	9.72222	9.68016	9.64287	9.60957	9.57966	9.55265
6	7.78957	7.71833	7.65748	7.60490	7.55899	7.51857	7.48271	7.45066	7.42186	7.39583
7	6.53817	6.46909	6.41003	6.35895	6.31433	6.27501	6.24010	6.20889	6.18082	6.15544
8	5.73427	5.66672	5.60891	5.55887	5.51512	5.47655	5.44228	5.41163	5.38405	5.35909
9	5.17789	5.11143	5.05451	5.00521	4.96208	4.92402	4.89019	4.85992	4.83266	4.80800
10	4.77152	4.70587	4.64961	4.60083	4.55814	4.52045	4.48692	4.45691	4.42987	4.40539
11	4.46244	4.39740	4.34162	4.29324	4.25087	4.21344	4.18013	4.15029	4.12340	4.09905
12	4.21982	4.15526	4.09985	4.05176	4.00962	3.97237	3.93921	3.90950	3.88271	3.85843
13	4.02452	3.96033	3.90520	3.85734	3.81537	3.77825	3.74520	3.71556	3.68884	3.66461
14	3.86404	3.80014	3.74524	3.69754	3.65570	3.61868	3.58570	3.55611	3.52942	3.50522
15	3.72990	3.66624	3.61151	3.56394	3.52219	3.48525	3.45231	3.42275	3.39608	3.37189
16	3.61616	3.55269	3.49810	3.45063	3.40895	3.37205	3.33914	3.30960	3.28293	3.25874
17	3.51851	3.45520	3.40072	3.35333	3.31169	3.27482	3.24193	3.21240	3.18573	3.16152
18	3.43379	3.37061	3.31622	3.26888	3.22729	3.19043	3.15754	3.12801	3.10132	3.07710
19	3.35960	3.29653	3.24221	3.19491	3.15334	3.11650	3.08361	3.05406	3.02736	3.00311
20	3.29411	3.23112	3.17686	3.12960	3.08804	3.05120	3.01830	2.98873	2.96201	2.93774
21	3.23587	3.17295	3.11874	3.07150	3.02995	2.99311	2.96019	2.93061	2.90386	2.87956
22	3.18374	3.12089	3.06671	3.01949	2.97795	2.94109	2.90816	2.87855	2.85178	2.82745
23	3.13682	3.07402	3.01987	2.97267	2.93112	2.89425	2.86130	2.83167	2.80487	2.78050

24	3.09437	3.03161	2.97749	2.93029	2.88873	2.85185	2.81888	2.78923	2.76239	2.73800
25	3.05577	2.99306	2.93895	2.89175	2.85019	2.81329	2.78030	2.75061	2.72375	2.69932
26	3.02053	2.95785	2.90375	2.85655	2.81498	2.77807	2.74505	2.71534	2.68845	2.66399
27	2.98823	2.92557	2.87149	2.82429	2.78270	2.74577	2.71273	2.68299	2.65607	2.63158
28	2.95851	2.89588	2.84180	2.79460	2.75300	2.71605	2.68299	2.65322	2.62627	2.60174
29	2.93108	2.86847	2.81440	2.76719	2.72558	2.68860	2.65552	2.62573	2.59874	2.57419
30	2.90569	2.84310	2.78902	2.74181	2.70018	2.66319	2.63008	2.60026	2.57325	2.54866
31	2.88211	2.81953	2.76546	2.71823	2.67659	2.63958	2.60645	2.57660	2.54956	2.52494
32	2.86016	2.79759	2.74353	2.69629	2.65463	2.61760	2.58444	2.55457	2.52750	2.50285
33	2.83968	2.77712	2.72305	2.67580	2.63413	2.59708	2.56390	2.53399	2.50690	2.48222
34	2.82052	2.75797	2.70390	2.65664	2.61495	2.57788	2.54467	2.51474	2.48762	2.46292
35	2.80256	2.74002	2.68594	2.63867	2.59697	2.55987	2.52665	2.49669	2.46954	2.44481
36	2.78569	2.72315	2.66907	2.62179	2.58007	2.54296	2.50971	2.47973	2.45255	2.42779
37	2.76982	2.70728	2.65320	2.60591	2.56417	2.52704	2.49376	2.46376	2.43656	2.41177
38	2.75485	2.69232	2.63823	2.59093	2.54918	2.51202	2.47873	2.44870	2.42147	2.39666
39	2.74072	2.67819	2.62410	2.57678	2.53501	2.49784	2.46452	2.43447	2.40722	2.38239
40	2.72735	2.66483	2.61073	2.56340	2.52162	2.48442	2.45108	2.42101	2.39374	2.36888
50	2.62503	2.56250	2.50833	2.46089	2.41896	2.38160	2.34807	2.31780	2.29032	2.26524
60	2.55867	2.49612	2.44188	2.39435	2.35230	2.31480	2.28113	2.25070	2.22305	2.19781
120	2.39900	2.33630	2.28181	2.23395	2.19150	2.15357	2.11943	2.08851	2.06035	2.03459
∞	2.24790	2.18492	2.13004	2.08170	2.03871	2.00018	1.96540	1.93381	1.90497	1.87850

Fuente: Valores generados con Microsoft Excel.

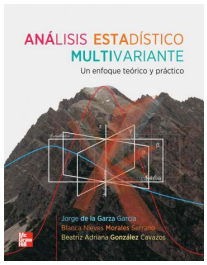
TABLA A4 VALORES CRÍTICOS DE F $\alpha = 0.01$. PARTE 3

Denominador gl_2	Grados de libertad en el numerador gl_1									
	21	22	23	24	25	30	40	60	120	∞
1	6216.118	6222.843	6228.990	6234.630	6239.825	6260.648	6286.782	6313.030	6339.391	6365.861
2	99.45155	99.45371	99.45569	99.45750	99.45917	99.46583	99.47416	99.48250	99.49083	99.49916
3	26.66350	26.63955	26.61764	26.59752	26.57898	26.50453	26.41081	26.31635	26.22114	26.12518
4	13.99383	13.97033	13.94882	13.92906	13.91085	13.83766	13.74538	13.65220	13.55810	13.46306
5	9.52812	9.50576	9.48529	9.46647	9.44912	9.37933	9.29119	9.20201	9.11177	9.02043
6	7.37219	7.35063	7.33088	7.31272	7.29597	7.22853	7.14322	7.05674	6.96902	6.88003
7	6.13238	6.11134	6.09205	6.07432	6.05795	5.99201	5.90845	5.82357	5.73729	5.64954
8	5.33641	5.31571	5.29673	5.27926	5.26314	5.19813	5.11561	5.03162	4.94605	4.85881
9	4.78556	4.76508	4.74629	4.72900	4.71303	4.64858	4.56665	4.48309	4.39777	4.31056
10	4.38313	4.36278	4.34411	4.32693	4.31106	4.24693	4.16529	4.08186	3.99648	3.90899
11	4.07688	4.05662	4.03803	4.02091	4.00509	3.94113	3.85957	3.77607	3.69044	3.60245
12	3.83633	3.81612	3.79757	3.78049	3.76469	3.70079	3.61918	3.53547	3.44944	3.36082
13	3.64254	3.62236	3.60383	3.58675	3.57096	3.50704	3.42529	3.34129	3.25476	3.16540
14	3.48317	3.46300	3.44447	3.42739	3.41159	3.34760	3.26564	3.18127	3.09419	3.00403
15	3.34984	3.32966	3.31112	3.29403	3.27822	3.21411	3.13191	3.04713	2.95945	2.86844
16	3.23668	3.21649	3.19793	3.18081	3.16497	3.10073	3.01825	2.93305	2.84474	2.75284
17	3.13944	3.11923	3.10064	3.08350	3.06764	3.00324	2.92046	2.83481	2.74585	2.65304
18	3.05500	3.03476	3.01615	2.99897	2.98308	2.91852	2.83542	2.74931	2.65970	2.56597
19	2.98098	2.96071	2.94207	2.92487	2.90894	2.84420	2.76079	2.67421	2.58394	2.48929
20	2.91558	2.89528	2.87660	2.85936	2.84340	2.77848	2.69475	2.60771	2.51678	2.42120
21	2.85737	2.83704	2.81832	2.80105	2.78505	2.71995	2.63590	2.54839	2.45681	2.36031
22	2.80523	2.78486	2.76611	2.74880	2.73276	2.66749	2.58311	2.49515	2.40292	2.30549

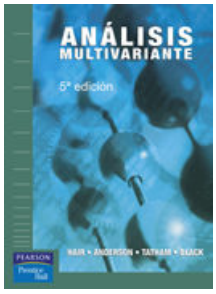
23	2.75825	2.73785	2.71907	2.70172	2.68565	2.62019	2.53550	2.44708	2.35421	2.25586
24	2.71571	2.69527	2.67646	2.65907	2.64296	2.57733	2.49232	2.40346	2.30996	2.21070
25	2.67701	2.65653	2.63768	2.62026	2.60411	2.53831	2.45299	2.36369	2.26956	2.16940
26	2.64164	2.62113	2.60224	2.58479	2.56860	2.50262	2.41701	2.32728	2.23254	2.13148
27	2.60919	2.58865	2.56973	2.55224	2.53602	2.46987	2.38396	2.29381	2.19846	2.09653
28	2.57933	2.55875	2.53979	2.52227	2.50602	2.43970	2.35350	2.26294	2.16700	2.06419
29	2.55174	2.53113	2.51214	2.49458	2.47830	2.41182	2.32534	2.23437	2.13785	2.03418
30	2.52618	2.50553	2.48651	2.46892	2.45260	2.38597	2.29921	2.20785	2.11076	2.00624
31	2.50243	2.48175	2.46270	2.44508	2.42873	2.36194	2.27491	2.18317	2.08552	1.98015
32	2.48031	2.45960	2.44052	2.42286	2.40648	2.33954	2.25225	2.16014	2.06194	1.95574
33	2.45965	2.43891	2.41980	2.40211	2.38570	2.31861	2.23107	2.13859	2.03985	1.93283
34	2.44031	2.41955	2.40040	2.38269	2.36625	2.29902	2.21123	2.11838	2.01912	1.91129
35	2.42218	2.40138	2.38221	2.36447	2.34800	2.28063	2.19260	2.09940	1.99962	1.89099
36	2.40513	2.38431	2.36511	2.34734	2.33084	2.26333	2.17507	2.08153	1.98124	1.87183
37	2.38909	2.36824	2.34901	2.33121	2.31468	2.24704	2.15855	2.06468	1.96389	1.85370
38	2.37395	2.35307	2.33381	2.31599	2.29944	2.23167	2.14295	2.04876	1.94748	1.83652
39	2.35965	2.33874	2.31946	2.30161	2.28503	2.21714	2.12820	2.03369	1.93194	1.82022
40	2.34611	2.32518	2.30588	2.28800	2.27140	2.20338	2.11423	2.01941	1.91719	1.80472
50	2.24226	2.22111	2.20158	2.18349	2.16666	2.09759	2.00659	1.90903	1.80260	1.68314
60	2.17465	2.15334	2.13363	2.11536	2.09837	2.02848	1.93602	1.83626	1.72632	1.60066
120	2.01091	1.98906	1.96882	1.95002	1.93249	1.86001	1.76285	1.65569	1.53299	1.38055
∞	1.85410	1.83152	1.81055	1.79101	1.77275	1.69660	1.59247	1.47321	1.32486	1.01099

Fuente: Valores generados con Microsoft Excel.

BIBLIOGRAFÍA



1. Jorge de la Garza García; Blanca Nieves Morales Serrano; Beatriz Adriana González Cavazos. **ANÁLISIS ESTADÍSTICO MULTIVARIANTE.** Editorial McGracHill. Primera Edición, 2013. Formato: Rústico. Idioma: español. País: México. ISBN: 978-607-15-0817-1. No. de páginas: 648.



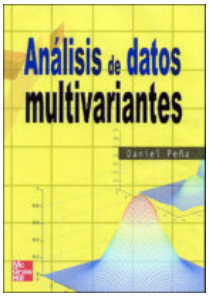
2. J.F. Hair; R.E. Anderson; R.L. Tatham. **ANÁLISIS MULTIVARIANTE** Editorial Prentice-Hall. Edición, 1999. Formato: Rústico Tapa Blanda. Idioma: Español. País: México. ISBN: 97884832203513. No. de páginas: 832.

3. Cesar Perez Lopez. **TÉCNICAS DE ANÁLISIS MULTIVARIANTE DE DATOS. APLICACIONES CON SPSS**. Editorial Pearson Educación. Edición, 2004. Formato: Rústico Tapa blanda. Idioma: español. País: España. ISBN: 9788420541044. No. de páginas: 672.

4. Rosario Martínez Arias. **EL ANÁLISIS MULTIVARIANTE EN LA INVESTIGACIÓN CIENTÍFICA**. Editorial La Muralla. Edición, 2008. Formato: Rústico. Idioma: español. País: España. ISBN: 9788471337788. No. de páginas: 144.

5. Joaquín Aldas Manzano. **ANÁLISIS MULTIVARIANTE APLICADO CON R**. Editorial S.A. Ediciones Edición, 2017. Formato: Rústico Tapa Blanda. Idioma: español. País: México. ISBN: 9788428329699. No. de páginas: 678.

6. Daniel Peña. **ANÁLISIS DE DATOS MULTIVARIANTES**. Editorial S.A. McGrawHill/Interamericana de España. Edicione, 2002. Formato: Rústico Tapa Blanda. Idioma: español. País: España. ISBN: 9788448136109. No. de páginas: 529.



7. Cesar Perez. **ANÁLISIS MULTIVARIANTE DE DATOS. APLICACIONES CON IBM, SPSS, SAS Y STATGRAPHICS**. Editorial Garceta Grupo Editorial Edición, 2016. Formato: Rústico Tapa Blanda. Idioma: Castellano. País: México. ISBN: 978849281204. No. de páginas:

8. Cesar Perez. **TÉCNICAS ESTADÍSTICAS MULTIVARIANTES CON SPSS**. Editorial Garceta Grupo Editorial Edición, 2016. Formato: Rústico Tapa Blanda. Idioma: Castellano. País: México. ISBN: 9788415452737. No. de páginas: 564

9. **MANUAL DE MINITAB 17**. Versión en español para Windows. Edición, 2014-2016. Formato: electrónico. Idioma: español. País: México.

10. Lind, Douglas A., Marchal, William G. y Mason, Robert D. **ESTADÍSTICA PARA ADMINISTRACIÓN Y ECONOMÍA**. Editorial Alfaomega. Onceava edición, 2004. Formato: Rústico. Idioma: español. País: México. ISBN: 9701509749. No. de páginas: 830.

Minitab® 18

11. Hildebrand, David K y Ott R., Lyman. **ESTADISTICA APLICADA** a la administración y a la economía. Editorial Addison–Wesley Iberoamericana, 1997. Formato: Rústico. Idioma: español. País: México. ISBN: 0201625520. No. de páginas: 943.

12. Cesar Pérez. **ESTADÍSTICA APLICADA A TRAVÉS DE EXCEL.** Editorial Prentce Hall. Edición 2002. Formato: Rústico. Idioma: Español. Pais: México. ISBN: 8420535362

"CUADERNO DE TRABAJO: **ANÁLISIS MULTIVARIADO"**

ANEXO **DE RESPUESTAS**

PARA LA CARRERA DE **MERCADOTECNIA** Y CARRERAS AFINES ADSCRITAS AL
CENTRO DE **CIENCIAS ECONÓMIAS Y ADMINISTRATIVAS**

Elaboró: Arq. Y M. en Admón**. Javier Bech Vertti**

Aguascalientes, Ags. **Agosto de 2017**

RESPUESTAS A LAS ACTIVIDADES DE APRENDIZAJE DEL CAPÍTULO 1

1.3.1.1 ACTIVIDAD DE APRENDIZAJE

ACTIVIDAD DE APRENDIZAJE 1.3.1.1

DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **NORTE** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

215.0	267.0	247.0	235.0	170.0	155.0
195.0	199.9	218.0	190.0	215.0	179.0
160.0	169.0	299.9	180.0	229.5	195.0
189.0	179.0	172.0	213.0	180.0	179.9
249.0	218.0	230.0	175.0	225.0	225.0

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas sin recortar los datos o valores atípicos generado por Minitab:

Tallo y hoja de VALOR N = 30

Unidad de hoja = 1.0

1 15 5

3 16 09

9 17 025999

12 18 009

(4) 19 0559

14 20

14 21 35588

9 22 559

6 23 05

4 24 79

2 25

2 26 7

1 27

1 28

1 29 9

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas recortando los datos o valores atípicos generado por Minitab:

Tallo y hoja de VALOR N = 30

Unidad de hoja = 1.0

1 15 5

3 16 09

9 17 025999

12 18 009

(4) 19 0559

14 20

14 21 35588

9 22 559

6 23 05

4 24 79

2 25

2 26 7

HI 299

Otra forma de ordenarlos es colocarlos en forma ascendente:

155.0	175.0	180.0	199.9	218.0	235.0
160.0	179.0	189.0	213.0	225.0	247.0
169.0	179.0	190.0	215.0	225.0	249.0
170.0	179.9	195.0	215.0	229.5	267.0
172.0	180.0	195.0	218.0	230.0	299.9

Resumen de 5 Números: Resultados aproximados a las décimas.

1. Valor Mínimo: **155.0**
2. Cuartil 1 Q_1=Dato 7 + 0.75 (Dato 8-Dato 7)= 179.0 + 0.75 (179.0-179.0)=**179.0**

   P(Q_1)= 1 $\frac{30+1}{4}$= 7.75
3. Mediana $\tilde{X}=\frac{Dato\ 16-Dato15}{2}=\frac{199.9+195.00}{2}$=**197.4**

   P($\tilde{X}$)=$\frac{n+1}{2}=\frac{30+1}{2}$=15.5
4. Cuartil 3 Q_3=Dato 23+ 0.25(Dato 24- Dato 23)=225.0+0.25(229.5-225.0)=**226.1**

   P(Q_3)= 3 $\frac{30+1}{4}$= 23.25
5. Valor Máximo= **299.9**
6.

Esquema de vallas: Resultados aproximados a las décimas

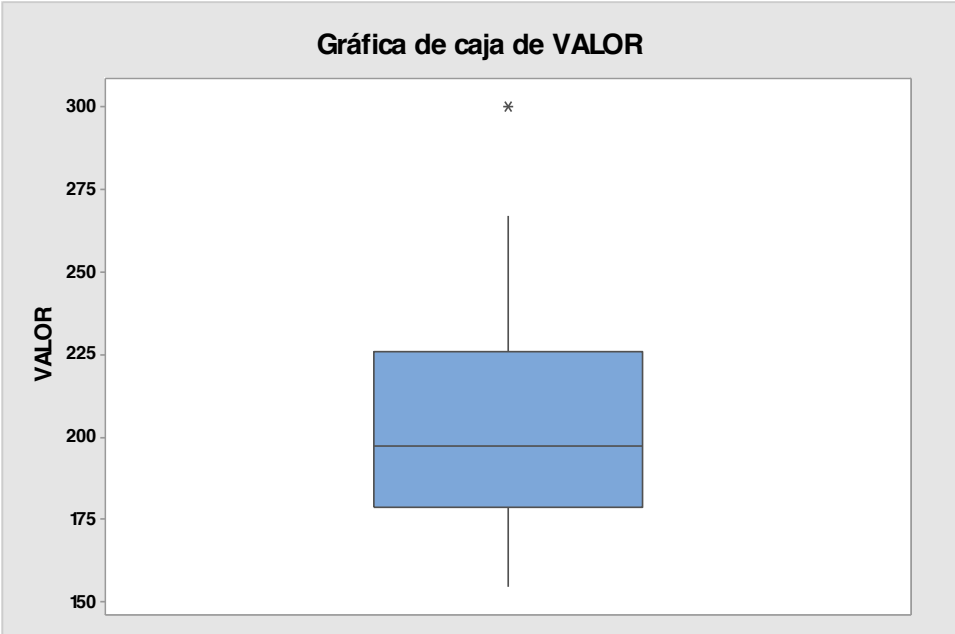
VEI (Valla Externa Inferior)= Q_1-3(RIQ)=179.0-3(226.1-179.0)= 179.0-3(47.1)=**37.7**

VII (Valla Interna Inferior)= Q_1-1.5(RIQ)=179.0-1.5(226.1-179.0)= 179.0-1.5(47.1)=**108.35**

VIS (Valla Interna Superior)= Q_{3+}1.5(RIQ)=226.1+1.5(226.1-179.0)= 226.1+1.5(47.1)=**296.7**

VES (Valla Externa Superior)= Q_3-3(RIQ)=226.1+3(226.1-179.0)= 226.1+3(47.1)=**367.4**

Diagrama de Caja y Brazos



Interpretación: En el diagrama anterior se puede observar que existe 1 valor atípico: **299.9** entre las **VIS** y la **VES (**al que le denominamos comúnmente valor aberrante) que debe ser estudiado con detalle para ver el tratamiento que se le va a dar en estudios subsecuentes ya que es posible que provoquen un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la izquierda de la caja, sugiere un sesgo hacia la derecha producto tal vez de valor o caso atípico).

1.3.1.2 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 1.3.1.2**

DATOS **ATÍPICOS**

Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota

que el valor más pequeño o más grande de los datos de la variable **TAMAÑO DEL LOTE** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.**

1 Abra la base de datos que creó **"BIENES1".** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***TAMAÑO***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

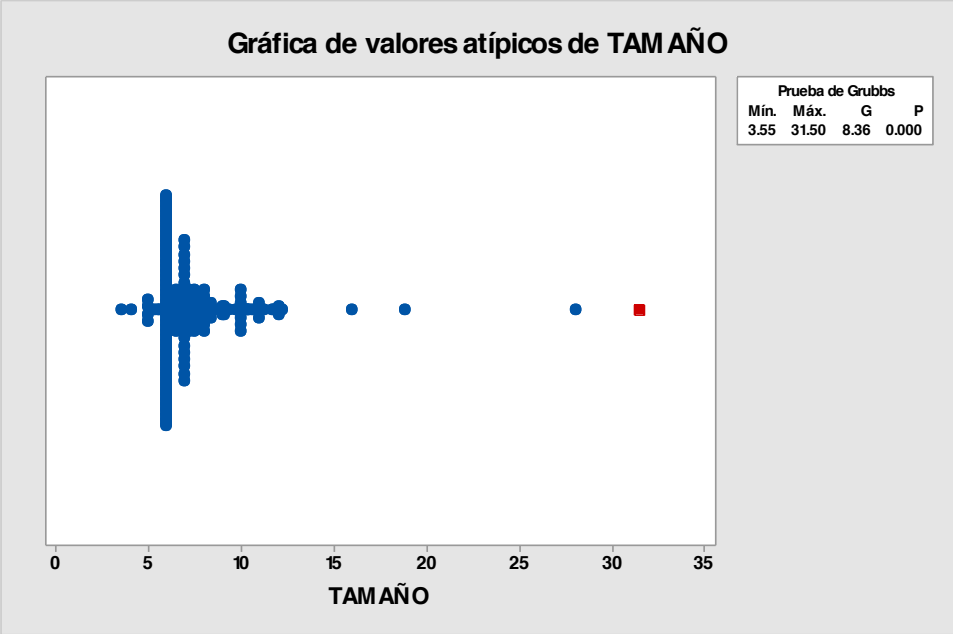
Salida de la ventana Sesión

Prueba de valores atípicos: TAMAÑO

```
Método

Hipótesis nula          Todos los valores de los datos provienen de la misma
población normal
Hipótesis alterna       El valor más pequeño o más grande de los datos es un
valor atípico
Nivel de significancia  α = 0.05


Prueba de Grubbs

Variable    N  Media  Desv.Est.   Mín.    Máx.     G      P
TAMAÑO    233  7.471      2.876  3.550  31.500  8.36  0.000


Valor atípico

                 Valor
Variable  Fila  atípico
TAMAÑO      76     31.5
```

Gráfica de valores atípicos de TAMAÑO

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **TAMAÑO DEL LOTE** muestran el valor más pequeño es de 3.55 y el más grande de 31.50. La estadística de G de 8.36 indica que el valor de datos más pequeño ó más grande de 3.55 ó 31.50, es 8.36 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.000. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico y corresponde al caso o la casa 76 con un valor máximo de 31.5 pies cuadrados.

Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

1.3.1.3 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 1.3.1.3**

DATOS ATÍPICOS PARA UN NIVEL DE UN **FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO**

Se reciben datos sobre una encuesta realizada a las casas ubicadas en la zona **SUR** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en esta zona **SUR** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

c) Crear un subconjunto de hoja de trabajo.
d) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

Respuesta al inciso a):

Crear un subconjunto de hoja de trabajo.

Nota importante:

Al **crear un subconjunto de hoja de trabajo** se divide, o desapila, la hoja de trabajo activa en hojas de trabajo nuevas basándose en una o más "Por" variables. Crear subconjunto de hoja de trabajo siempre copian los datos en hojas de trabajo **nuevas**. Como la base de datos **"BIENES 1"** contiene todas las casas ubicadas en la zona **NORTE, SUR y ESTE** debemos crear **un subconjunto de hoja de trabajo** que solo incluya las casas ubicadas en el **SUR**. Esto nos permite analizar los diferentes niveles de un factor por separado en caso de que la prueba de **ANOVA** haya sido significativa y exista la posibilidad de que las medias de la variable **VALOR** no sea la misma en los tres niveles del Factor **UBICACIÓN**

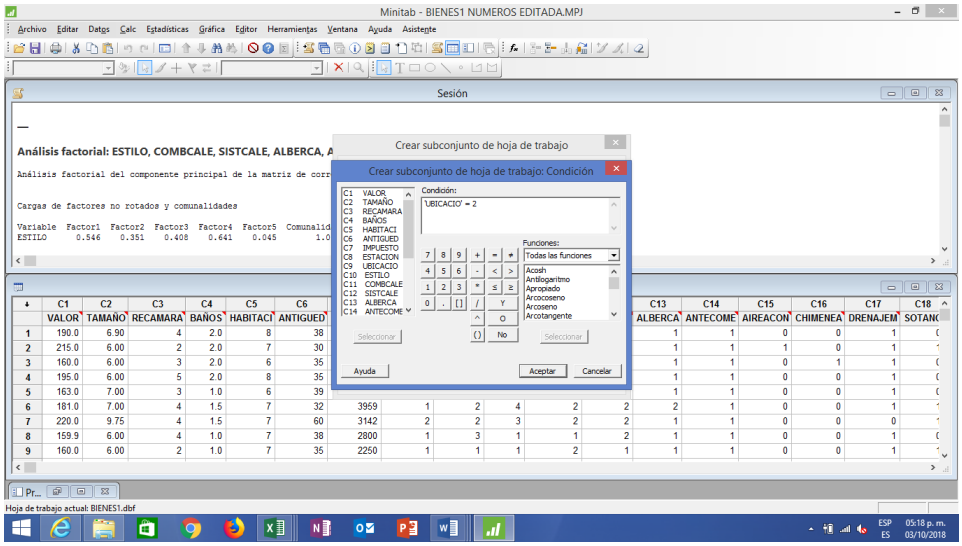
1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Datos > Crear subconjunto de hoja de trabajo**.

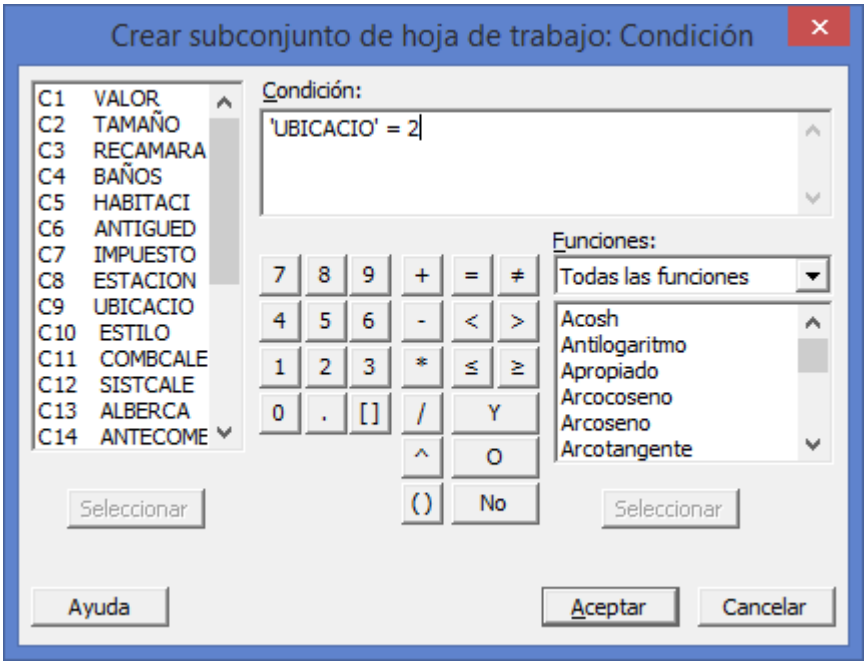
3 En **Incluir o excluir**, elija **Especificar las filas que se incluirán.**

4 En **Especificar las filas que se incluirán**, elija **Filas que coincidan**. Haga clic en **Condición**.

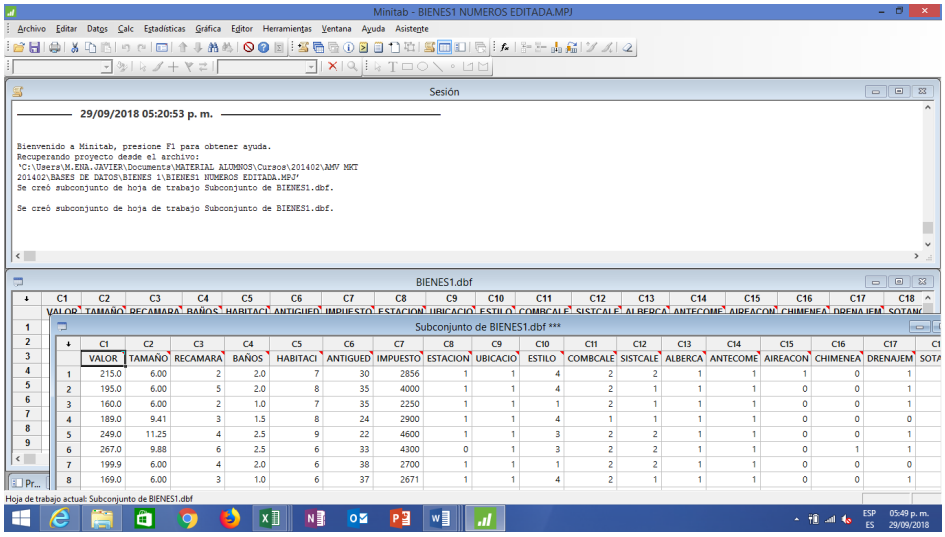
5 En **Condición**, ingrese ***UBICACIÓN=2.*** Haga clic en **Aceptar** en cada cuadro de diálogo.



Nota importante: Minitab es sensible al nombre que se le asignó a la variable por lo cual es conveniente no escribir el nombre de la variable sino seleccionarla de la parte izquierda de la ventana y solo añadir **=2.**



Minitab Crea un subconjunto de hoja nueva que incluye las casas ubicadas en la zona **SUR**



Respuesta al inciso b):

Realizar la Prueba de valores atípicos de Minitab

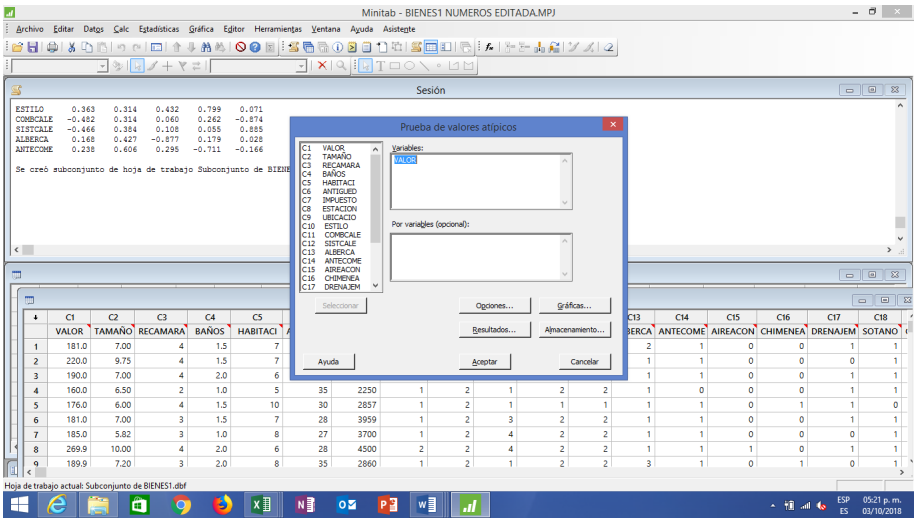
1 En la nueva hoja de trabajo que Minitab generó de la base de datos **"BIENES1"** que incluye solo las casas ubicadas en la zona **SUR** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***Valor***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.



Salida de la ventana Sesión

Prueba de valores atípicos: VALOR

Método

Hipótesis nula	Todos los valores de los datos provienen de la misma población normal
Hipótesis alterna	El valor más pequeño o más grande de los datos es un valor atípico
Nivel de significancia	α = 0.05

Prueba de Grubbs

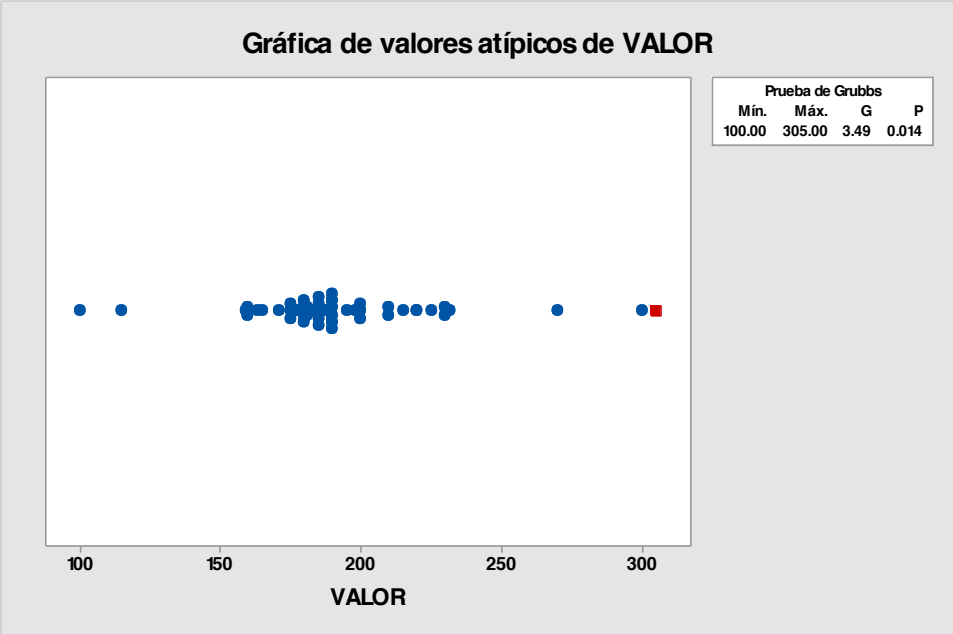
Variable	N	Media	Desv.Est.	Mín.	Máx.	G	P
VALOR	60	191.32	32.61	100.00	305.00	3.49	0.014

Valor atípico

Variable	Fila	Valor atípico
VALOR	35	305

Gráfica de valores atípicos de VALOR

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **VALOR** muestran el valor más pequeño es de 100 y el más grande de 305. La estadística de G de 3.49 indica que el valor de datos más pequeño ó más grande de 100 ó 305, es 3.49 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.014. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico. En este caso corresponde al valor 305 que se observa en rojo en la gráfica y que pertenece al caso o **fila 35 del nuevo subconjunto de hoja de trabajo**.

Es importante aclarar que hay que ubicar dicho caso en la base de datos "BIENES1" original por si es el caso en que haya que eliminirlo para depurar dicha base de datos.

1.3.1.4 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 1.3.1.4**

DATOS **ATÍPICOS . DIAGRAMA DE CAJA Y BRAZOS SIMPLE**

Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **NORTE** que forman

su cartera de ventas. De la base de datos **"BIENES1"** se eligieron aleatoriamente a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

215.0	267.0	247.0	235.0	170.0	155.0
195.0	199.9	218.0	190.0	215.0	179.0
160.0	169.0	299.9	180.0	229.5	195.0
189.0	179.0	172.0	213.0	180.0	179.9
249.0	218.0	230.0	175.0	225.0	225.0

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

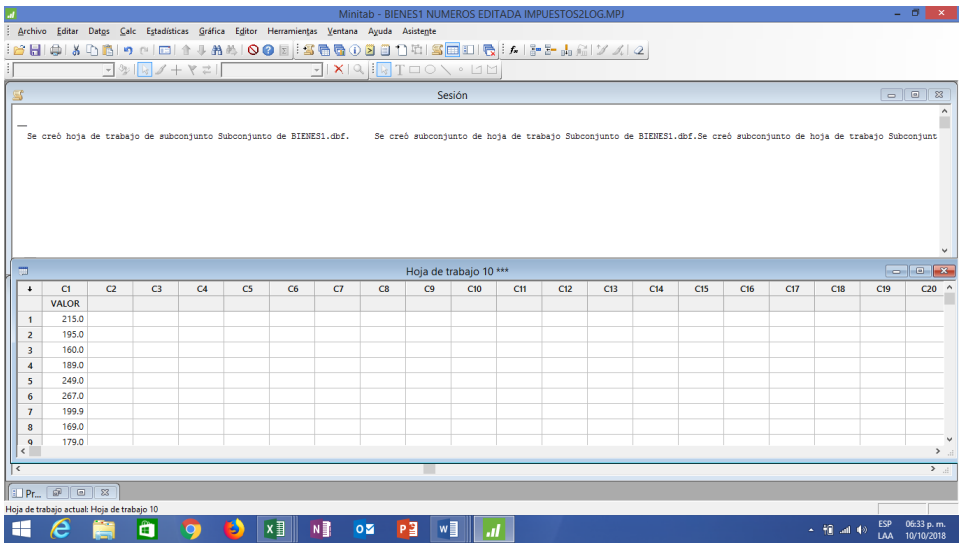
Para introducir los datos agrega una hoja de trabajo vacía al proyecto actual:

1 Elija **Archivo > Nuevo**.

2 Seleccione **Hoja de trabajo de Minitab** y haga clic en **Aceptar**.

3 En **C1** coloca el nombre de la variable **VALOR**.

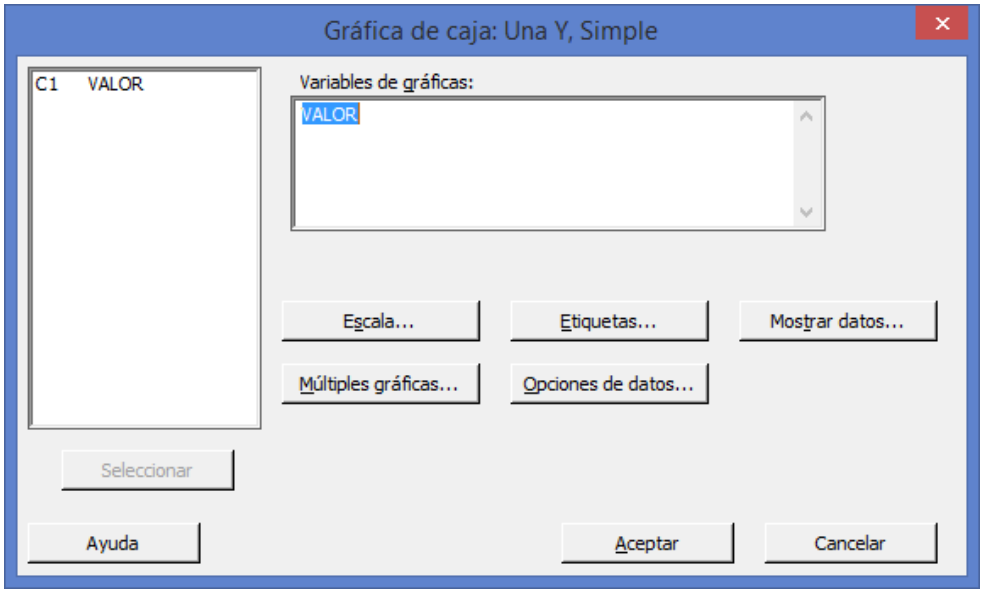
4 Teclea los datos de la tabla en forma vertical.



5 Elija **Gráfica > Gráfica de caja**.

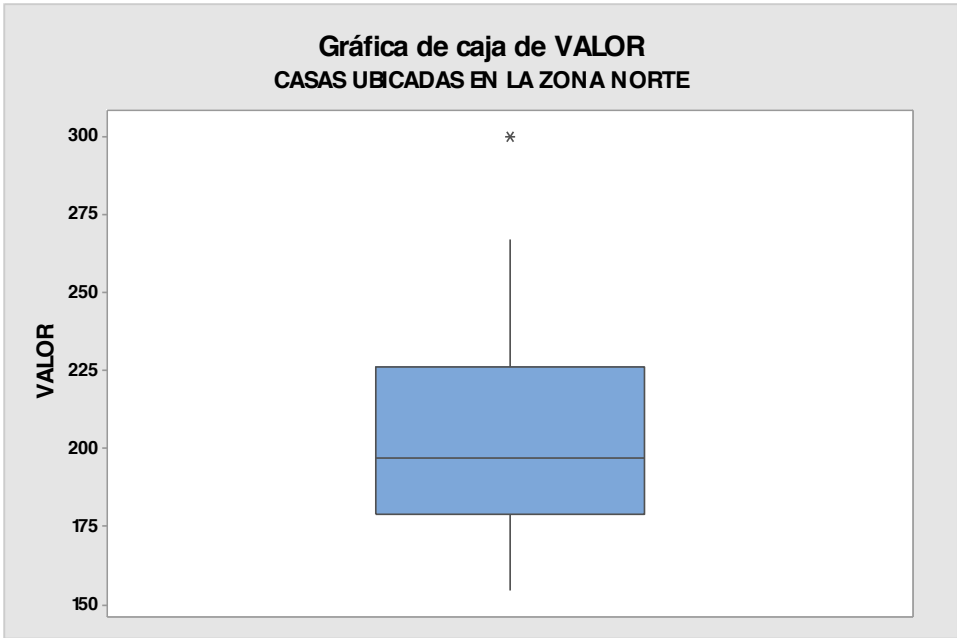
6 En **Una Y**, elija **Simple**. Haga clic en **Aceptar**.

7 En **Variables de graficas**, ingrese ***VALOR***.

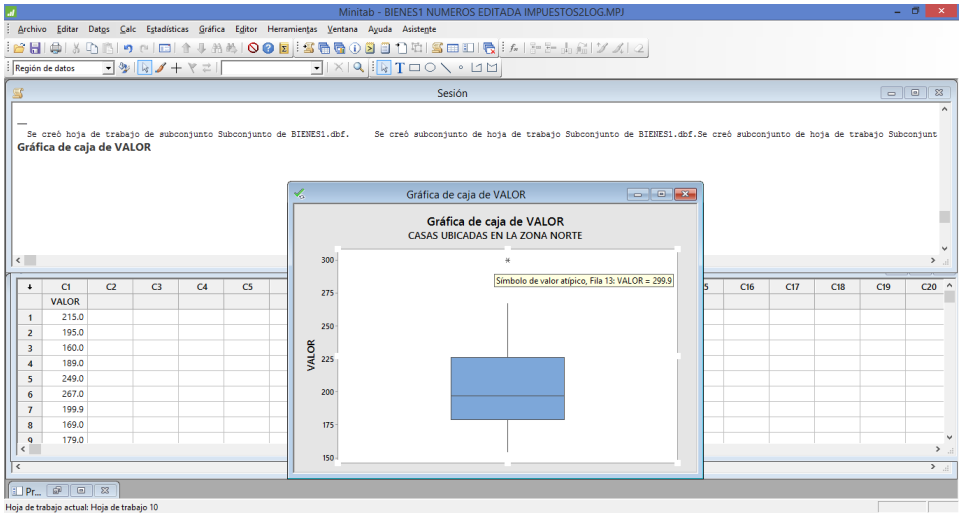


8 Haga clic en **Aceptar**.

Salida de la ventana Gráfica



Interpretación: En el diagrama anterior se puede observar que existe 1 valor atípico que pueden ser identificados colocando el puntero del mouse en cada uno de ellos proporcionandonos la siguiente información:



Fila 13 Valor 299.9

Este valor ó dato ó caso atípico debe ser estudiado con detalle para ver el tratamiento que se le va a dar en estudios subsecuentes ya que es posible que provoque un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la izquierda de la caja sugiere un sesgo hacia la derecha producto tal vez de los valores o casos atípicos.

1.3.1.5 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 1.3.1.5**

DATOS **ATÍPICOS . DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE**

Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas con estilo **MODERNO** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

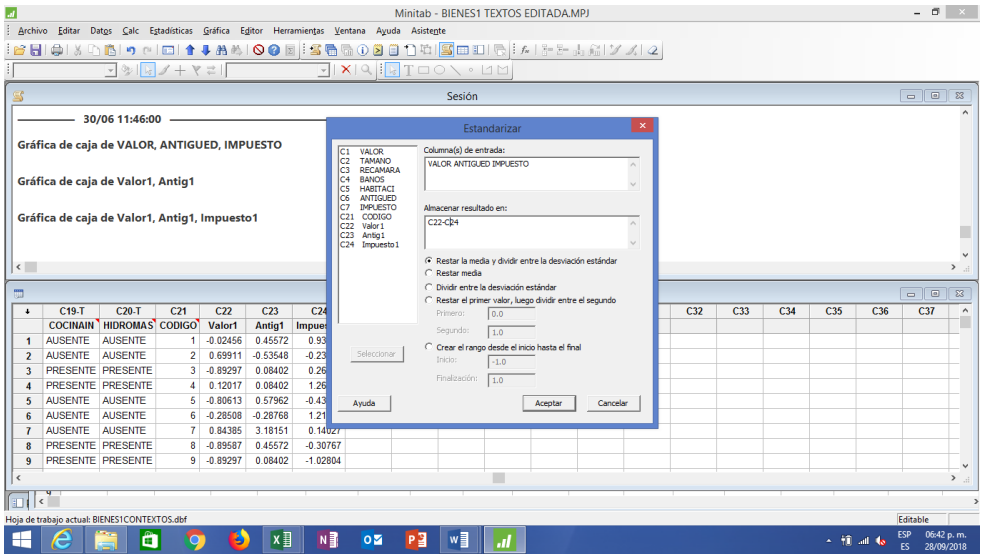
Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con el estilo MODERNO** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables de la siguiente manera:

1 Con la hoja de trabajo con el estilo **MODERNO** elija **Calc > Estandarizar**.

2 En **Columnas de entrada**, ingrese las columnas que desea estandarizar. En este caso ingrese **VALOR, ANTIGUEDAD E IMPUESTOS.**

3 En **Almacenar resultados en**, ingrese las columnas en las que desea almacenar los datos estandarizados. En este caso escriba **C22-C24**, haga clic en **Aceptar**



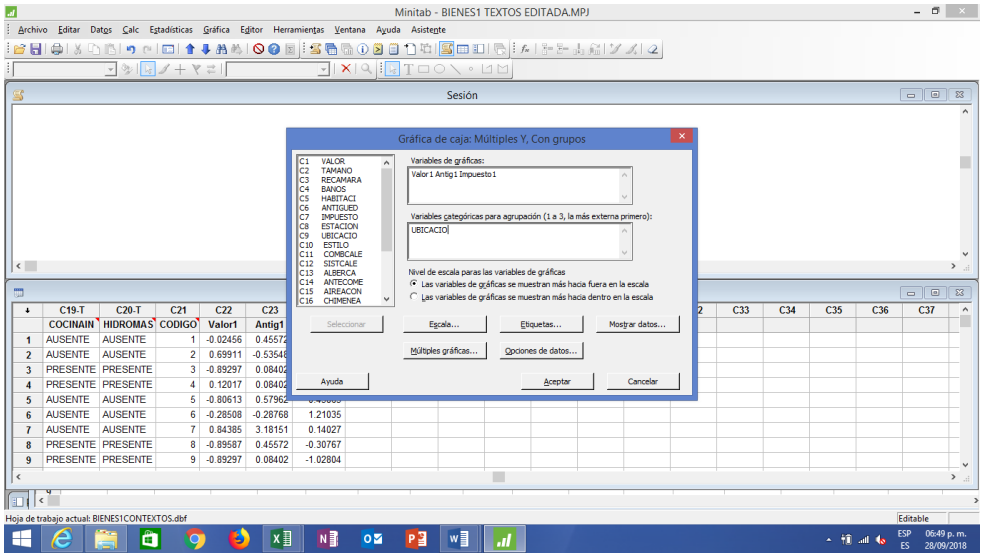
4 Los datos estandarizados aparecerán en las columas 22,23 y 24 de la base original **"BIENES1"**. Colocar el nombre de las variables estantarazadas como **Valor 1, Antig1 e Impuesto 1** para identificarlas.

1 Elija **Gráfica > Gráfica de caja**.

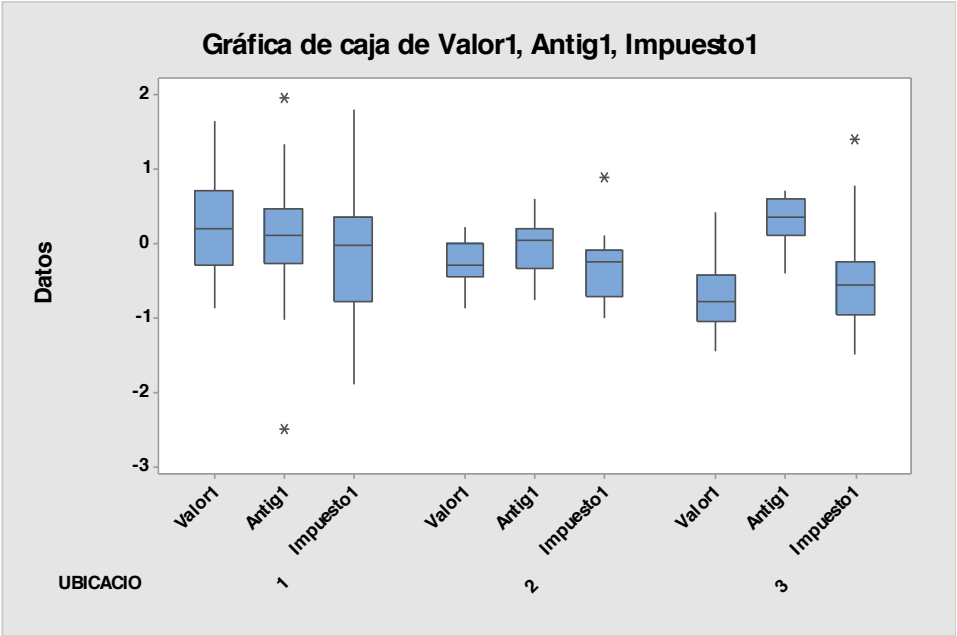
2 En **Múltiples Y**, elija **Con grupos**. Haga clic en **Aceptar**.

3 En **Variables de gráficas**, ingrese ***Valor1, Antig1 e Impuesto1***

4 En **Variables categóricas para agrupación (1 a 3, la más externa primero)**, ingrese ***UBICACIÓN.***



5 En **Nivel de escala para variables de gráficas**, elija **Las variables de gráficas se muestran más hacia dentro en la escala**. Haga clic en **Aceptar**.



Interpretación de los resultados

La gráfica de caja muestra:

· Para las casas con estilo **MODERNO**, en general en todos los niveles de **UBICACIÓN** de las casas las tres variables muestran valores atípicos que hay que estudiar con excepción de **Valor1** e **Impuestos1** en la zona **NORTE**, **Valor1 y Antig1** de la zona **SUR** y **Valor1 y Antig1** de la zona **ESTE**

· Para las casas ubicadas en la zona **NORTE** en la variable **Antig1** se detectan dos valores atípicos, en **Impuestos1** de la zona **SUR** se detecta un valor atípico superior y en **Impuestos1** de la zona **ESTE** un valor atípico superior.

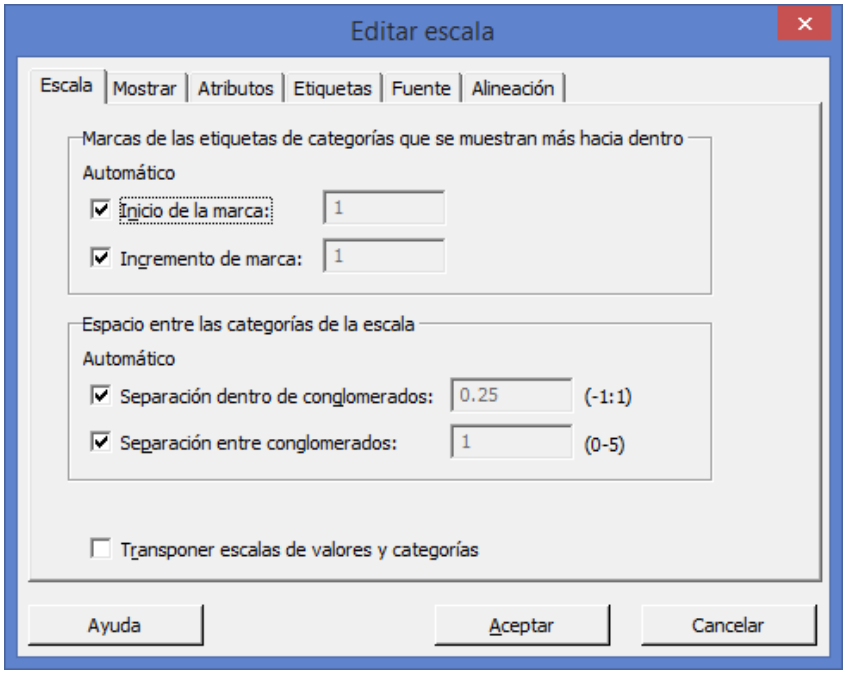
Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

Asimismo se recomienda realizar análisis estadísticos, tales como **MANOVA balanceado**, para examinar más a fondo la relación entre los factores.

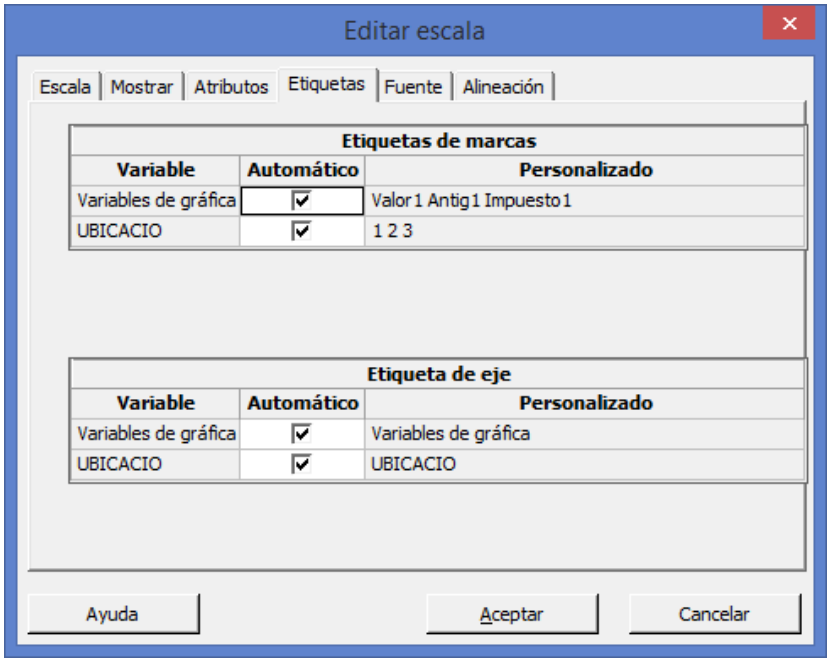
Para obtener información precisa de Q1, mediana, Q3, rango intercuartil, bigotes o brazos y N, coloque el cursor sobre cualquier parte de la gráfica de caja. Para ver el valor del eje Y y el número de fila de un valor atípico, coloque el cursor sobre el valor atípico.

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN Y Etiquetado del subtitulo del nivel MODERNO del "Factor" ESTILO:

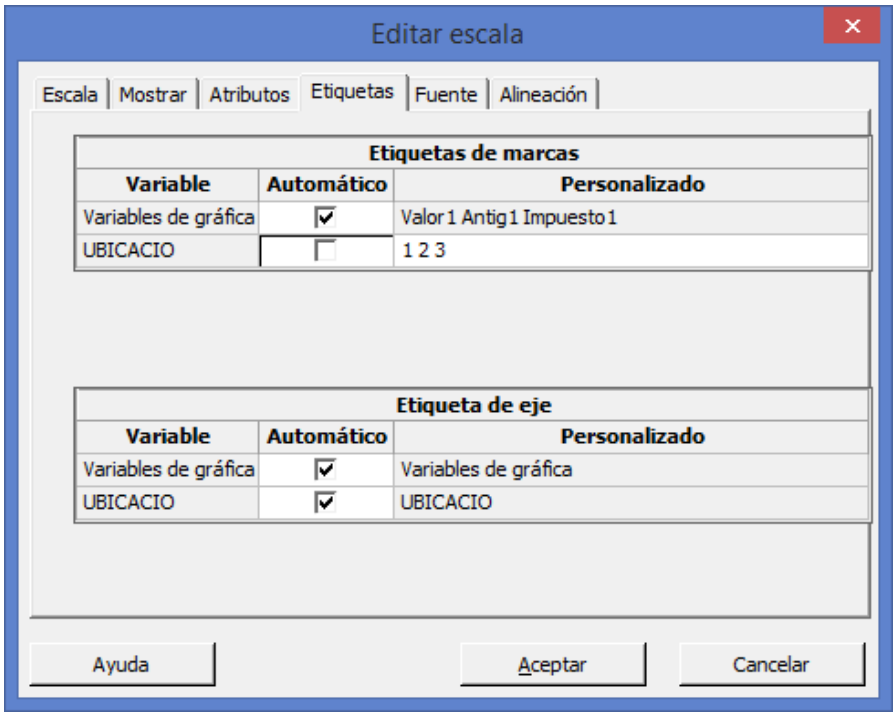
1 Haga doble clic en el número 1 en el renglón de UBICACIÓN para que Minitab despliegue la ventana ***Editar Escala.***



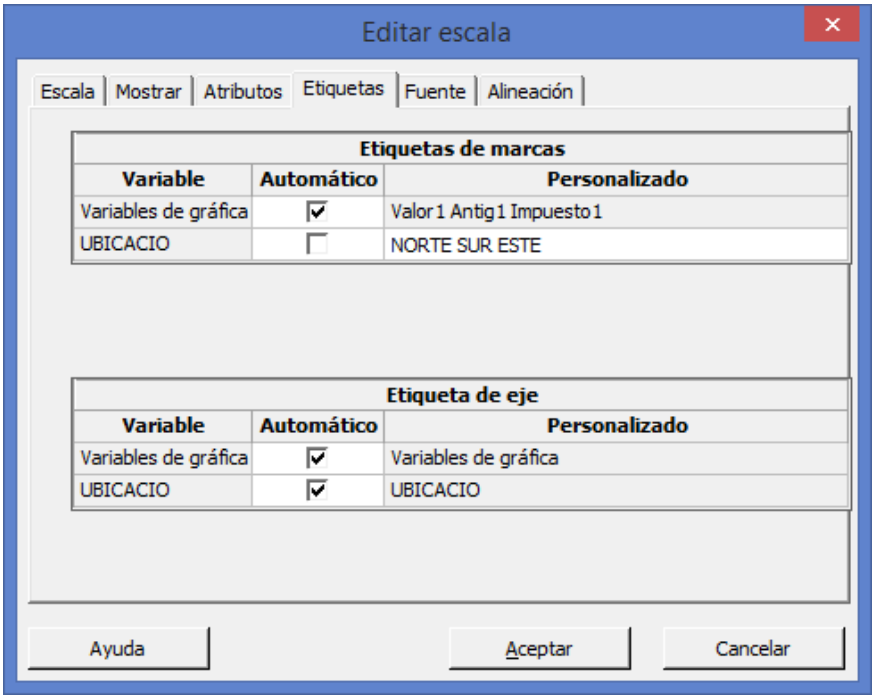
2 Seleccionar la pestaña **Etiquetas**



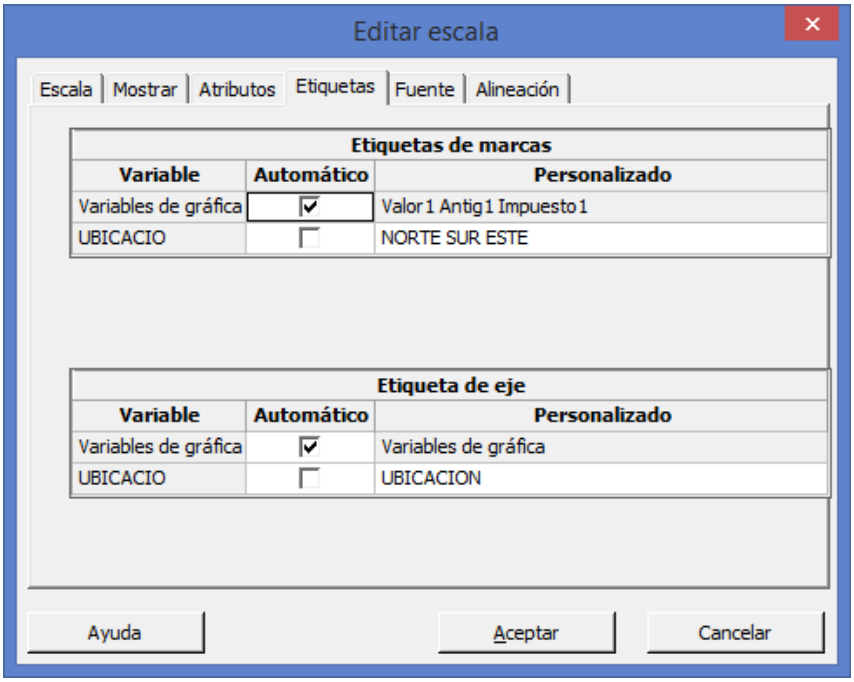
3 Desactivar la casilla **UBICACIÓN**



4 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE.**

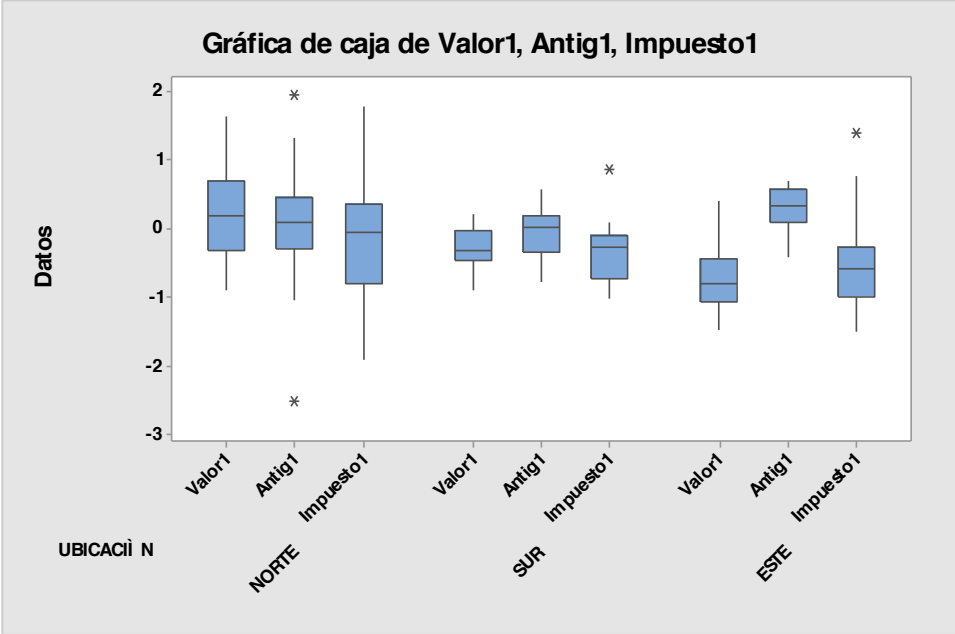


5 En **Etiqueta de eje** , si es el caso, desactivar la casilla de **UBICACIO** y escribir **UBICACIÓN.**

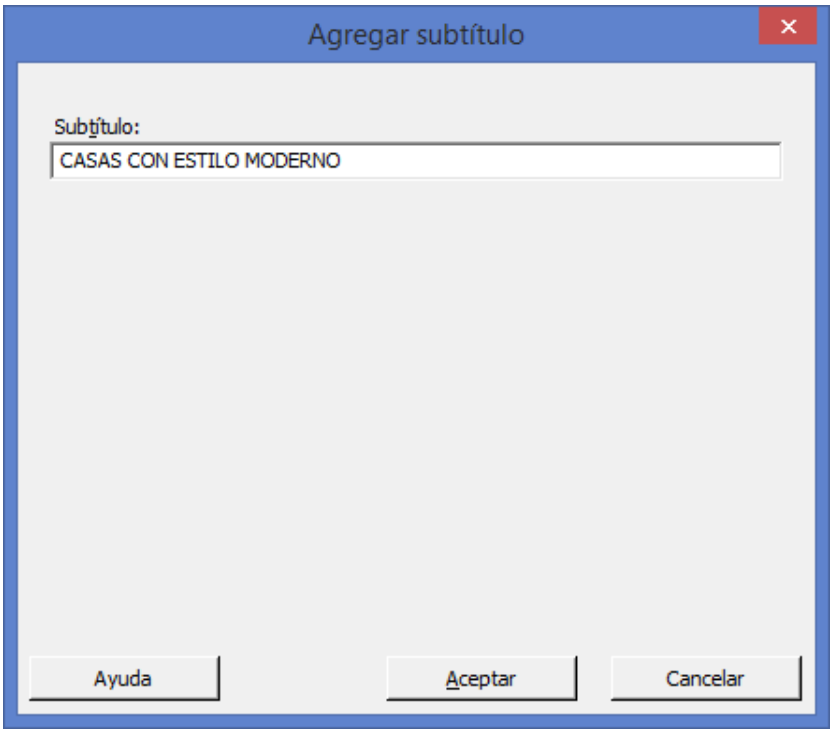


6 Haga clic en **Aceptar**.

7 Aparece la gráfica debidamente etiquetada.

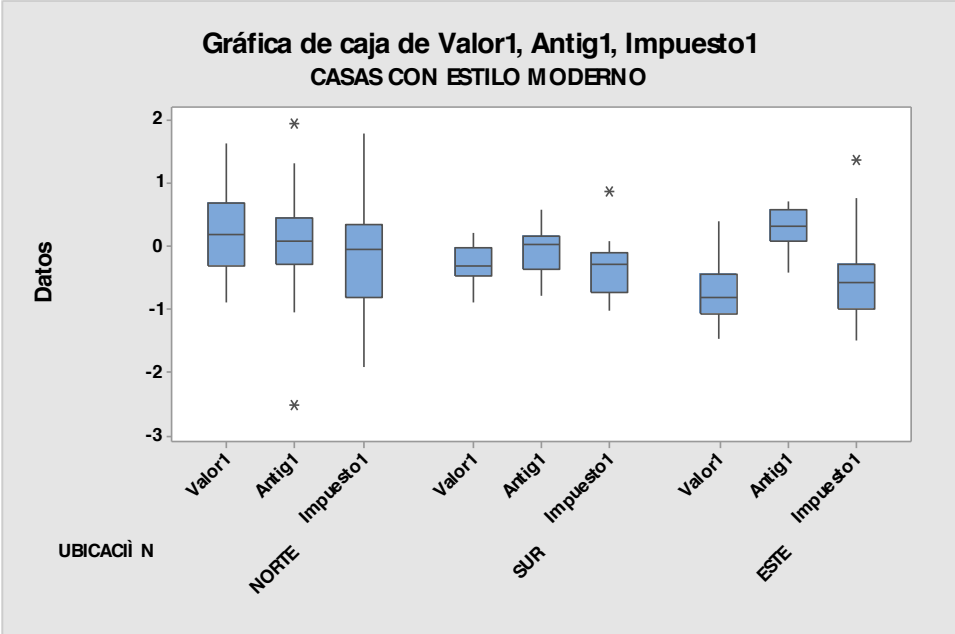


8 Coloque el mouse en el título de la grafica y presione el botón derecho del mouse. Seleccione la opción Agregar y la opción Subtitulo y se escribe **CASAS CON ESTILO MODERNO**.



9 Haga clic en Aceptar.

10 Aparece la gráfica debidamente etiquetada.



1.3.2.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 1.3.2.1**

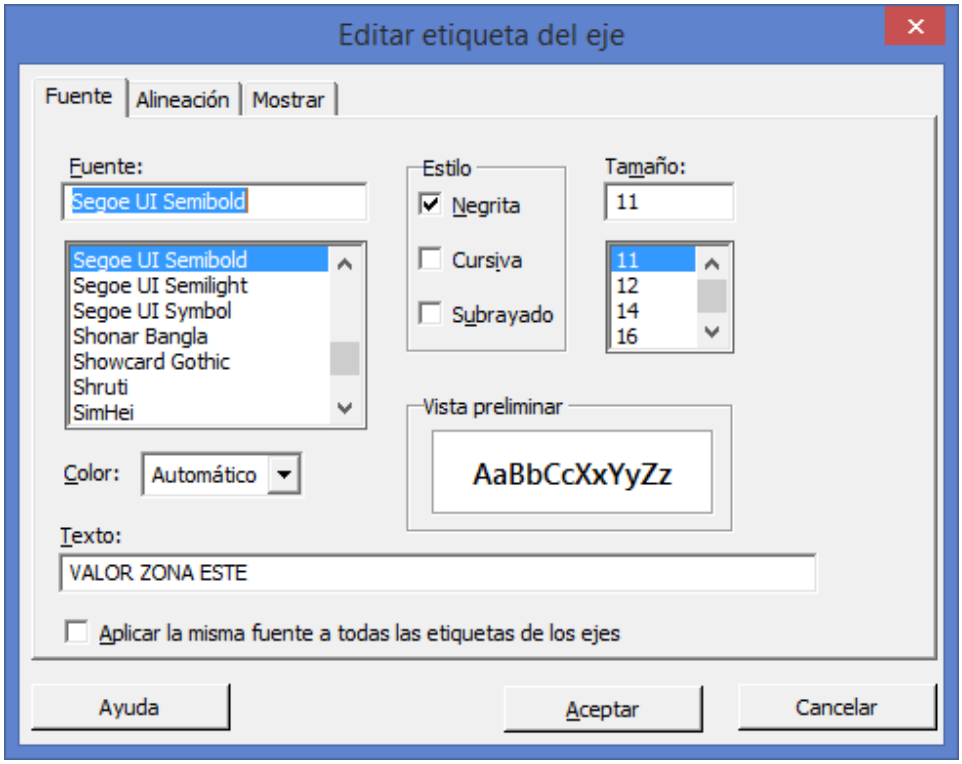
PRUEBA **DE NORMALIDAD. PRUEBA DE ANDERSON DARLING**

Se midió el **VALOR** de las casas en la zona **ESTE** en la base de datos "**BIENES1"** que se construyó anteriormente.

Se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas ubicadas en el **ESTE** e interprete los resultados-

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje.

Nota: Etiquete la gráfica dando doble clic en **VALOR** para que aparezca la ventana ***Editar etiqueta de eje*** y en la parte inferior donde dice ***Texto*** escriba **VALOR ZONA ESTE**. Haga clic en **Aceptar.**



1 Abra la base de datos que creó **"BIENES1"** y cree un subconjunto de hoja de trabajo para las casas ubicadas en el **ESTE**

2 Elija **Estadísticas > Estadísticas básicas > Prueba de normalidad**.

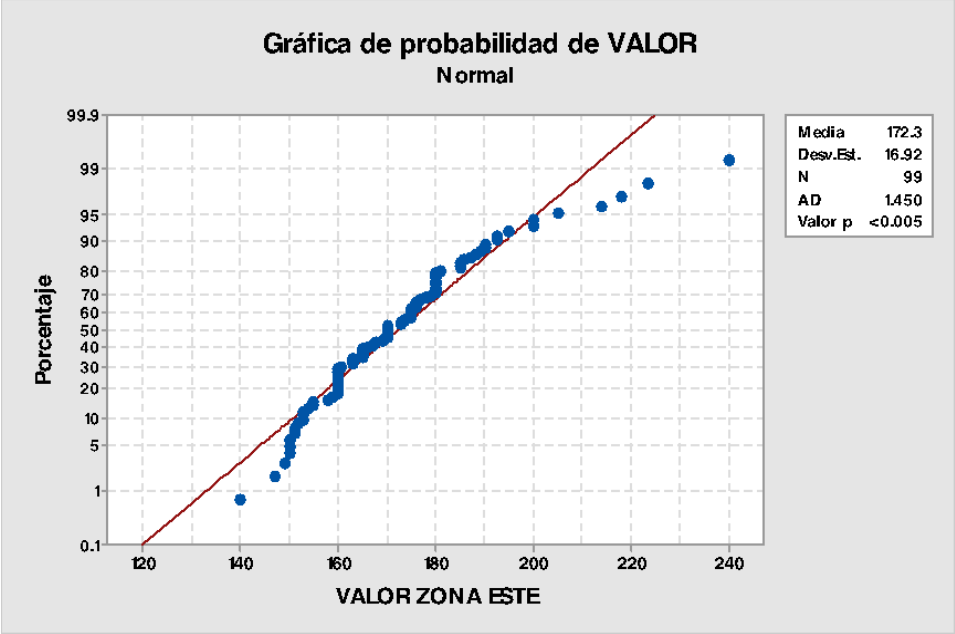
3 En **Variable**, ingrese ***Valor.*** Haga clic en **Aceptar**.

Salida de la ventana Gráfica

Genera una **gráfica de probabilidad** normal y realiza una **prueba de hipótesis** para examinar si las observaciones siguen o no una **distribución normal** . Para la **prueba de normalidad** , las hipótesis son,

H_0: los datos siguen una distribución normal vs. H_1: los datos no siguen una distribución normal

La escala vertical de la gráfica se asemeja a la escala vertical del papel de probabilidad normal. El eje horizontal es una escala lineal. La línea forma un estimado de la función de distribución acumulada para la **población** de la cual se extrajeron los datos. Con la gráfica se muestran **estimaciones numéricas de los parámetros de la población** , Media y Desviación Estándar, el tamaño de la muestra ó población, el valor de la prueba de normalidad (Estadístico de Anderson Darling) y el **valor p** asociado.



Interpretación de los resultados

La salida gráfica es una gráfica de probabilidades normales versus los datos. Los datos se alejan de la línea ajustada de una manera más evidente en los extremos, o colas de la distribución. El valor p de la prueba de Anderson-Darling indica que, en niveles p menores a 0.005, hay evidencia de que los datos no siguen una distribución normal.

Los problemas de no normalidad puede solucionarse por medio de **transformaciones de datos** y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una o varias de las variables por lo que también se debe verificar el supuesto de homocedasticidad y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

1.3.2.2 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 1.3.2.2**

PRUEBA **DE HOMOCEDASTICIDAD. PRUEBA DE LEVENE. TRANSFORMACIONES. ETIQUETADO DE GRÁFICOS**

Usted pretende diseñar un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad en cada una de las casas con **UBICACIÓN** diferente. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos. **Sin embargo, aquí se seleccionan dos métodos para demostrar las capacidades de Minitab para probar primeramente que se cumpla el supuesto de igualdad de varianzas (Homocedasticidad).**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**)**:**

Utilice el paquete Minitab para resolver esta Actividad de Aprendizaje de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

PRUEBA DE LEVENE EN MINITAB 17

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***Valor***

4 En **Factores**, ingrese ***UBICACIÓN***. Haga clic en **Aceptar**.

Salida de la ventana Sesión

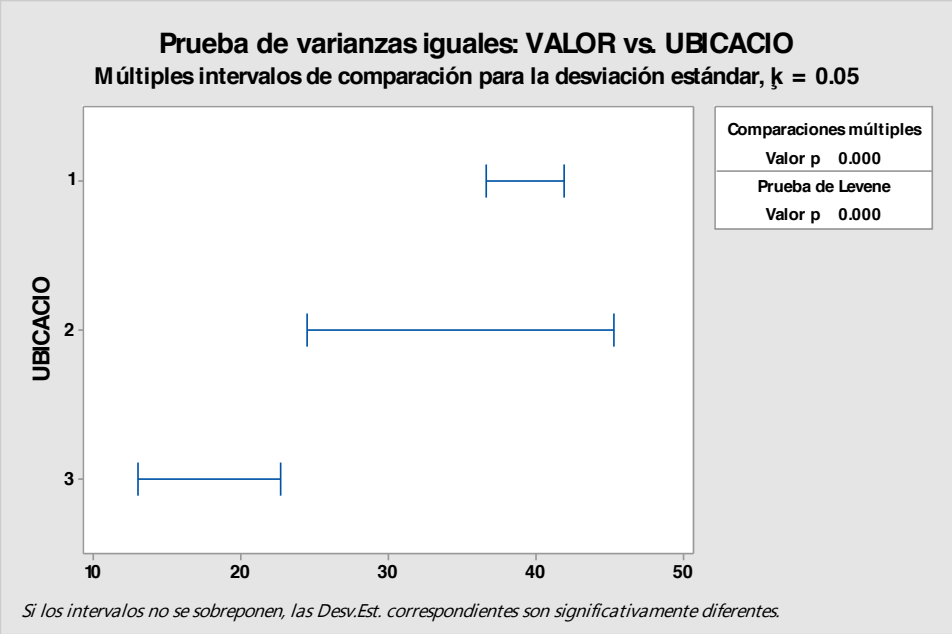
Prueba de varianzas iguales: VALOR vs. UBICACIO

Método

```
Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05


Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.          IC
       1  74    38.4944  (32.0204, 47.8245)
       2  60    32.6097  (22.0593, 50.2093)
       3  99    16.9209  (13.0610, 22.4648)

Nivel de confianza individual = 98.3333%


Pruebas

                         Estadística
Método                     de prueba  Valor p
Comparaciones múltiples            —    0.000
Levene                         15.97    0.000
```

Prueba de varianzas iguales: VALOR vs. UBICACIO

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de las pruebas **Levene** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **VALOR** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada,

logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva

TRANSFORMACIONES EN MINITAB 17 Y PRUEBA DE LEVENE

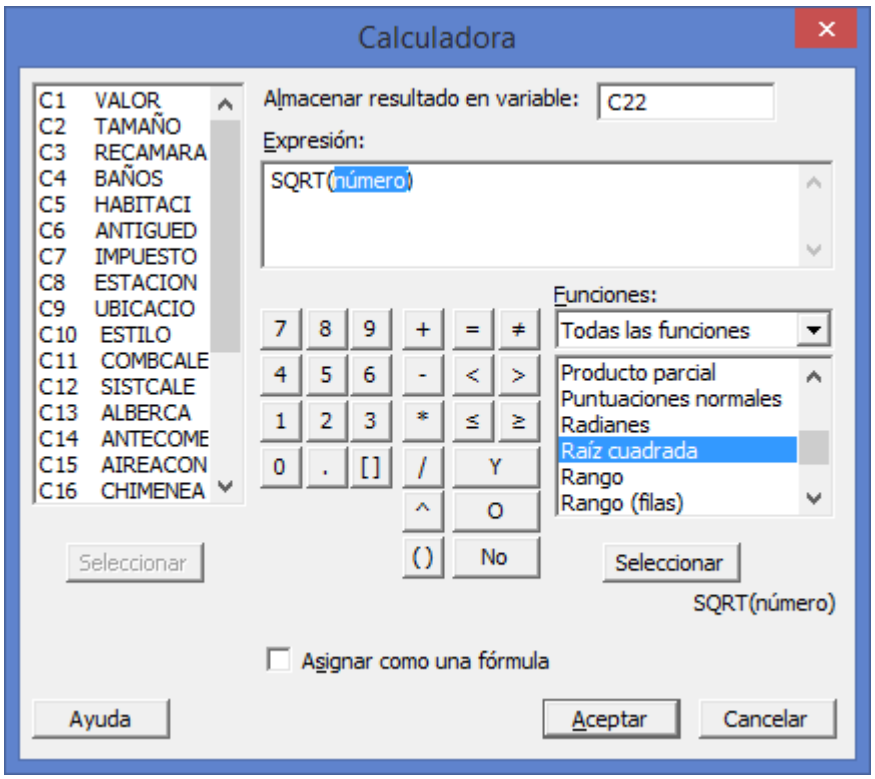
Transformaciónes Raiz Cuadrada, 1/X y Log de la variable VALOR con Minitab:

1 Abra la base de datos que creó **"BIENES1"**

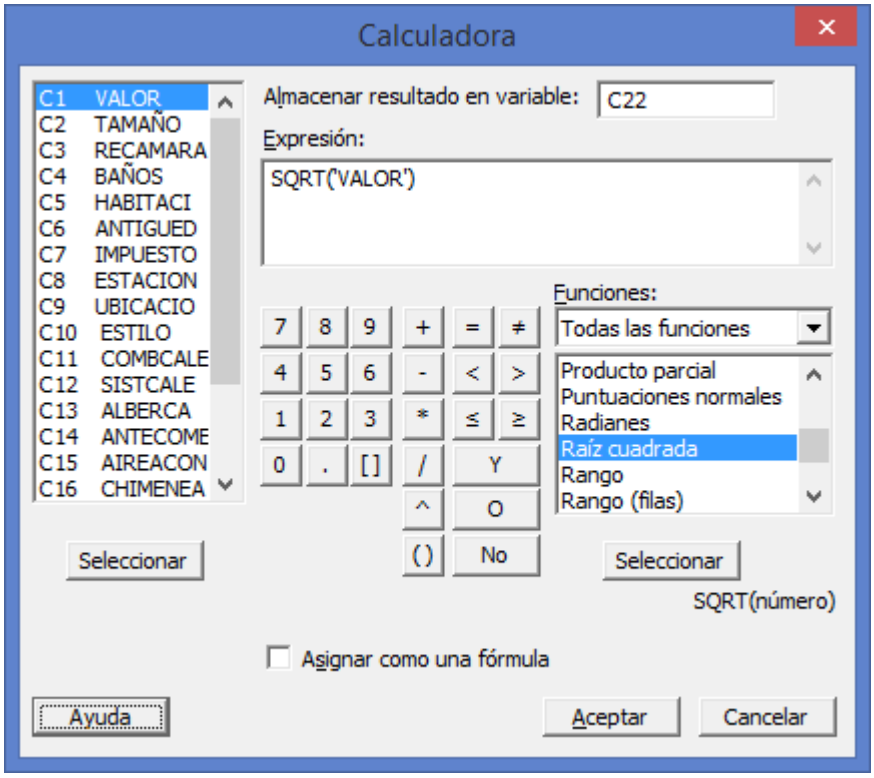
2 Elija **Calc > Calculadora**.

3 En **Almacenar resultado en variable**, ingrese *C22*

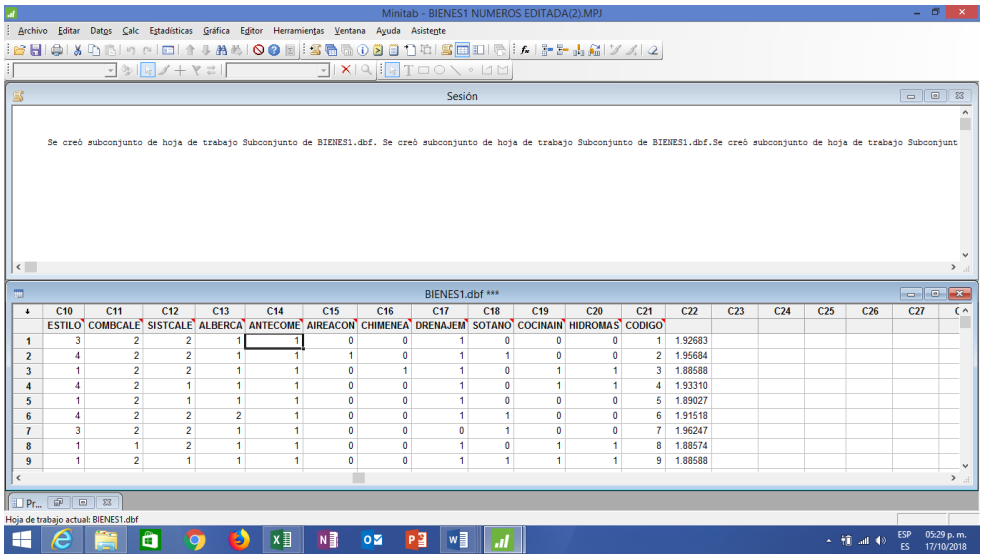
4 En **Expresión**, busque en **Funciones**: Raíz Cuadrada y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



5 En **Expresión** aparece **SQRT(número).** Seleccione la variable **VALOR** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.

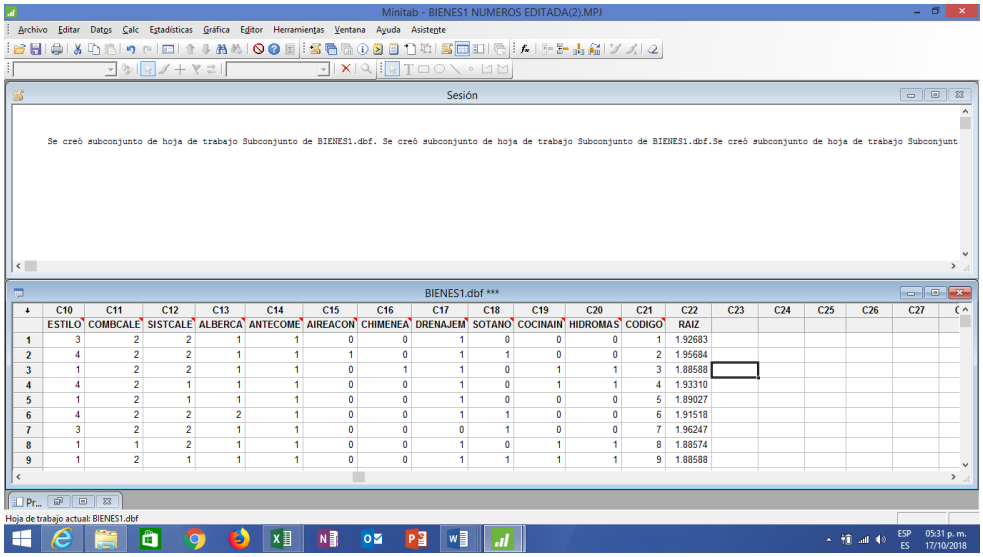


6 Haga clic en **Aceptar**.



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO						
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683					
2	4	2	2	1	1	1	0	1	1	0	0	2	1.95684					
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588					
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310					
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027					
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518					
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247					
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574					
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588					

7 En C22 teclee el nombre **RAÍZ**

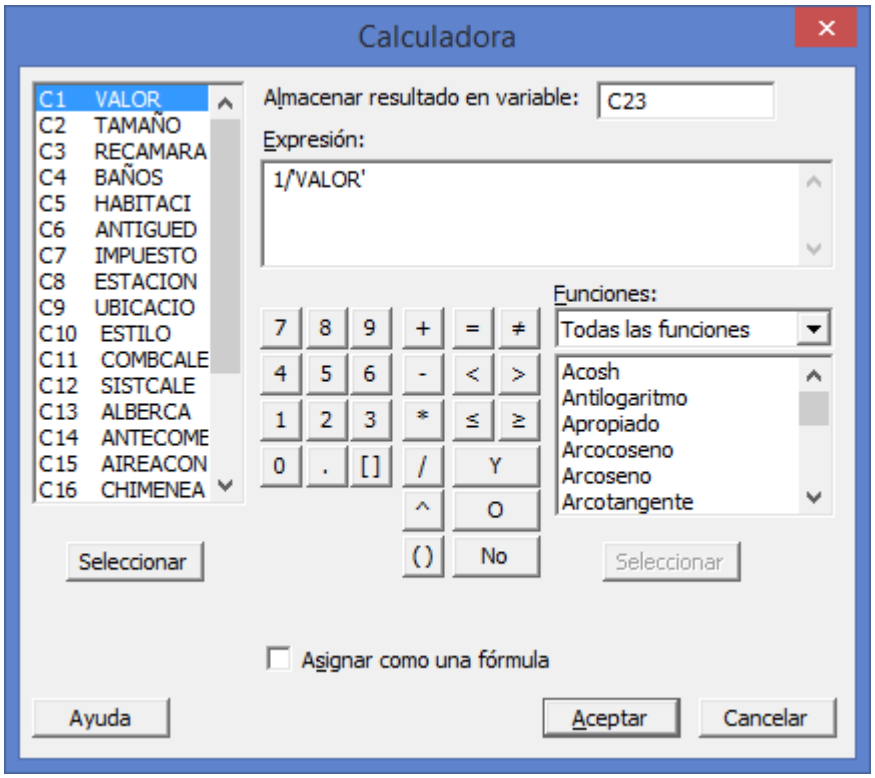


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAIZ					
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683					
2	4	2	2	1	1	0	0	1	1	0	0	2	1.95684					
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588					
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310					
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027					
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518					
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247					
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574					
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588					

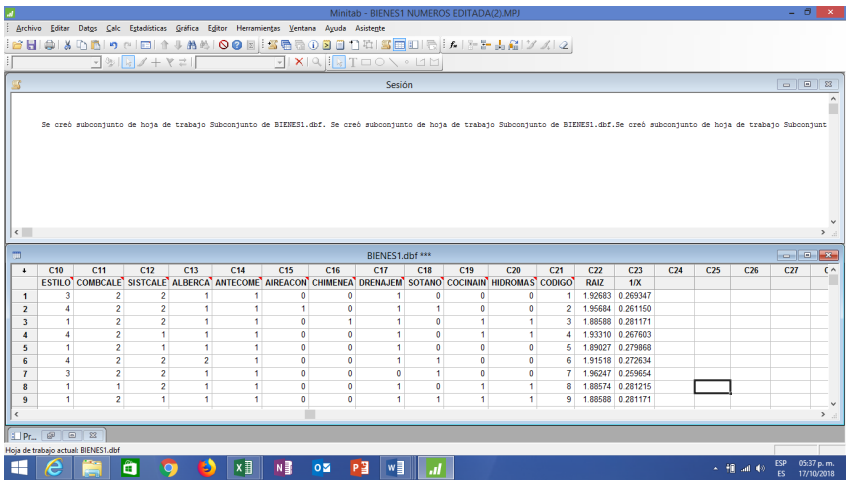
8 Elija **Calc > Calculadora**.

9 En **Almacenar resultado en variable**, ingrese *C23*

10 En **Expresión**, para obtener la Inversa, **Teclee 1/** y seleccione la variable **VALOR** del lado izquierdo o doble clic.



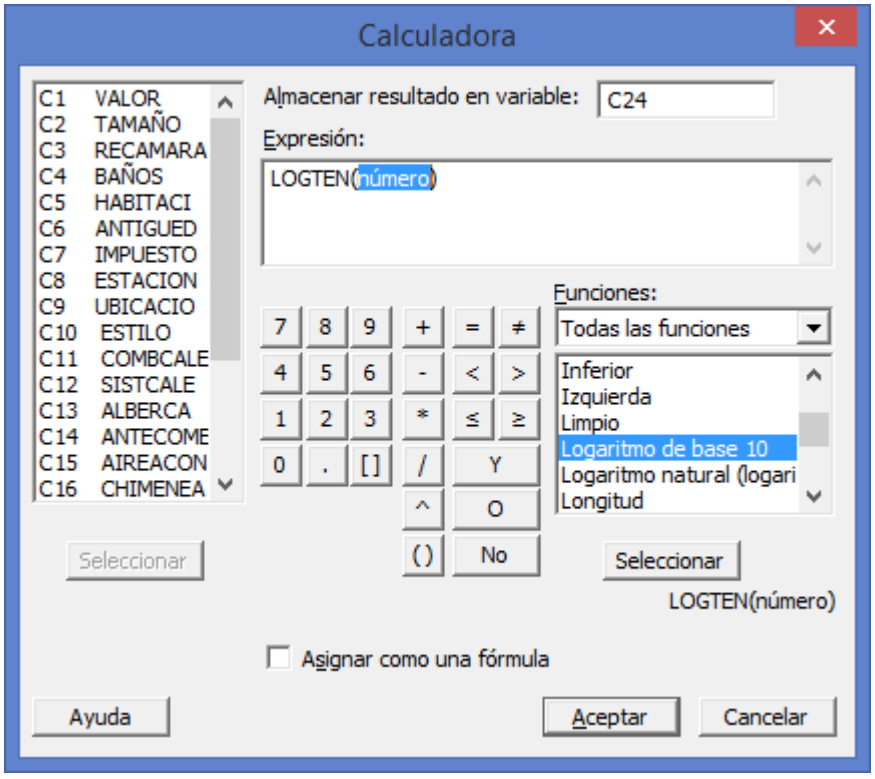
2 Haga clic en **Aceptar**. En **C23** escriba el nombre **1/X**



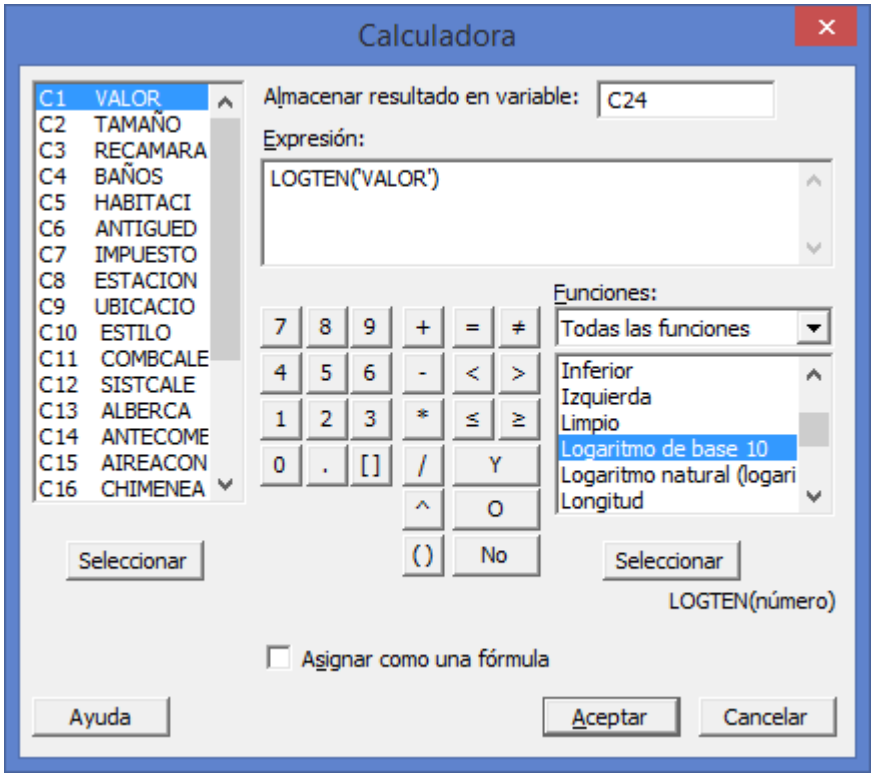
3 Elija **Calc > Calculadora**.

13 En **Almacenar resultado en variable**, ingrese *C24*

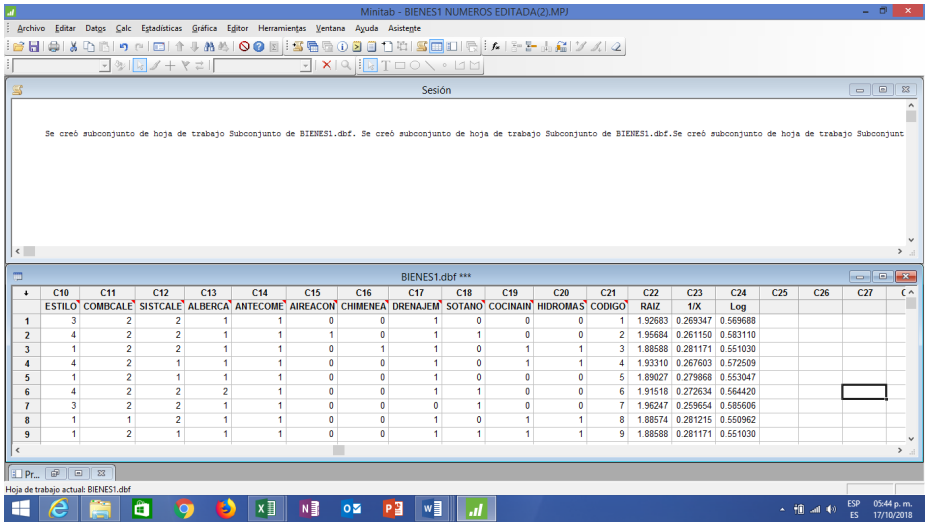
14 En **Expresión**, busque en **Funciones**: **Logaritmo base 10** y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



15 En **Expresión** aparece **LOGTEN(número).** Seleccione la variable **VALOR** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.



16 Haga clic en **Aceptar.** En **C24** escriba el nombre **Log**



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAIZ	1/X	Log			
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683	0.269347	0.569688			
2	4	2	2	1	1	1	0	1	1	0	0	2	1.95684	0.261150	0.583110			
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588	0.281171	0.551030			
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310	0.267603	0.572509			
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027	0.279868	0.553047			
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518	0.272634	0.564420			
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247	0.259654	0.585606			
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574	0.281215	0.550962			
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588	0.281171	0.551030			

Repita la prueba de Levene para las TRES transformaciones para ver si se corrigió el problema de homocedasticidad y si fué así cual transformación resulto mejor.

A. CON LA RAÍZ CUADRADA DE VALOR

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***RAÍZ***

4 En **Factores**, ingrese ***UBICACIÓN***. Haga clic en **Aceptar**.

Salida de la ventana Sesión

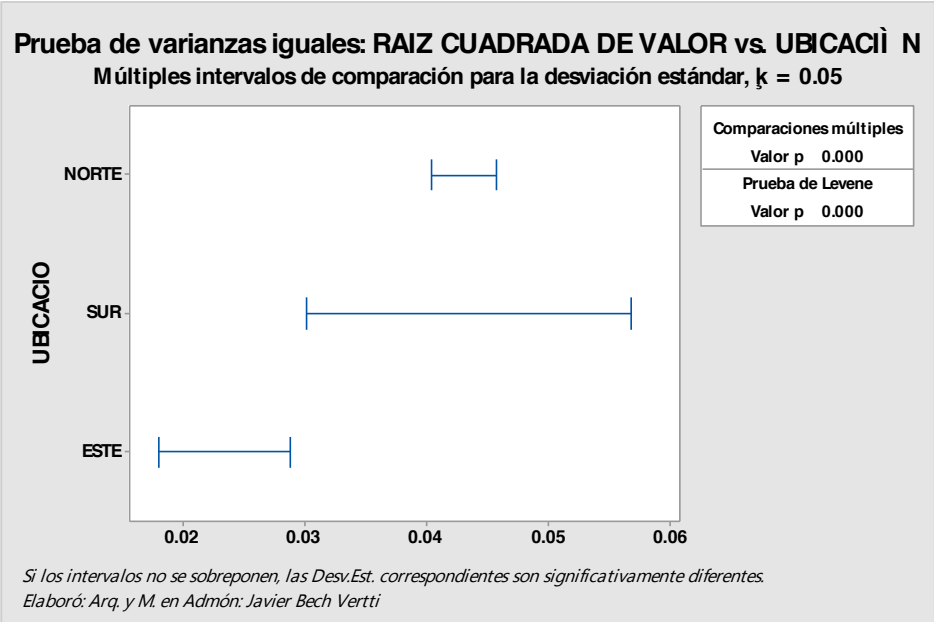
Prueba de varianzas iguales: RAIZ vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05


Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.            IC
       1  74  0.0422377  (0.0359390, 0.0512998)
       2  60  0.0405782  (0.0276796, 0.0619597)
       3  99  0.0225194  (0.0180347, 0.0288162)

Nivel de confianza individual = 98.3333%


Pruebas

                          Estadística
Método                      de prueba  Valor p
Comparaciones múltiples           —    0.000
```

```
Levene                          11.14     0.000
```

Prueba de varianzas iguales: RAIZ vs. UBICACIO

Salida de la ventana Gráfica



Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **RAIZ CUADRADA** del **VALOR** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

B. CON LA LA INVERSA DE VALOR (1/X)

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***1/X***

4 En **Factores**, ingrese ***UBICACIÓN***. Haga clic en **Aceptar**.

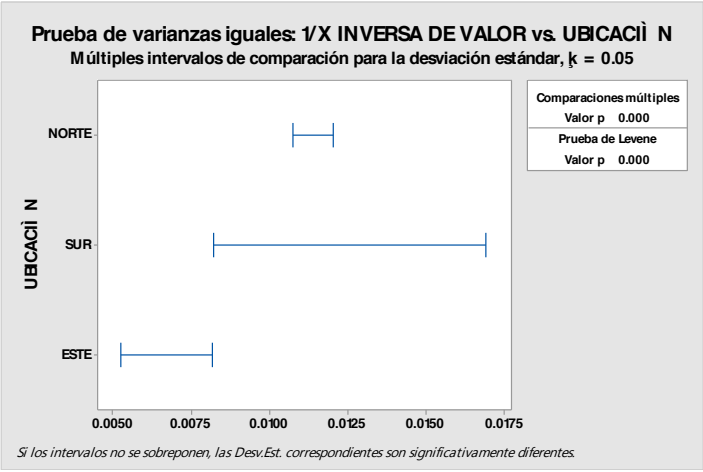
Salida de la ventana Sesión

Prueba de varianzas iguales: 1/X vs. UBICACIO

```
Método


Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.            IC

       1  74  0.0111782  (0.0095794, 0.0134799)

       2  60  0.0115403  (0.0076270, 0.0181869)

       3  99  0.0064543  (0.0052380, 0.0081502)


Nivel de confianza individual = 98.3333%

Pruebas

                          Estadística

Método                      de prueba  Valor p

Comparaciones múltiples            —    0.000

Levene                          8.78    0.000
```

Prueba de varianzas iguales: 1/X vs. UBICACIO

Salida de la ventana Gráfica



Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **INVERSA** del **VALOR** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

C. CON EL LOGARITMO BASE 10 DE VALOR

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***Log***

4 En **Factores**, ingrese ***UBICACIÓN***. Haga clic en **Aceptar**.

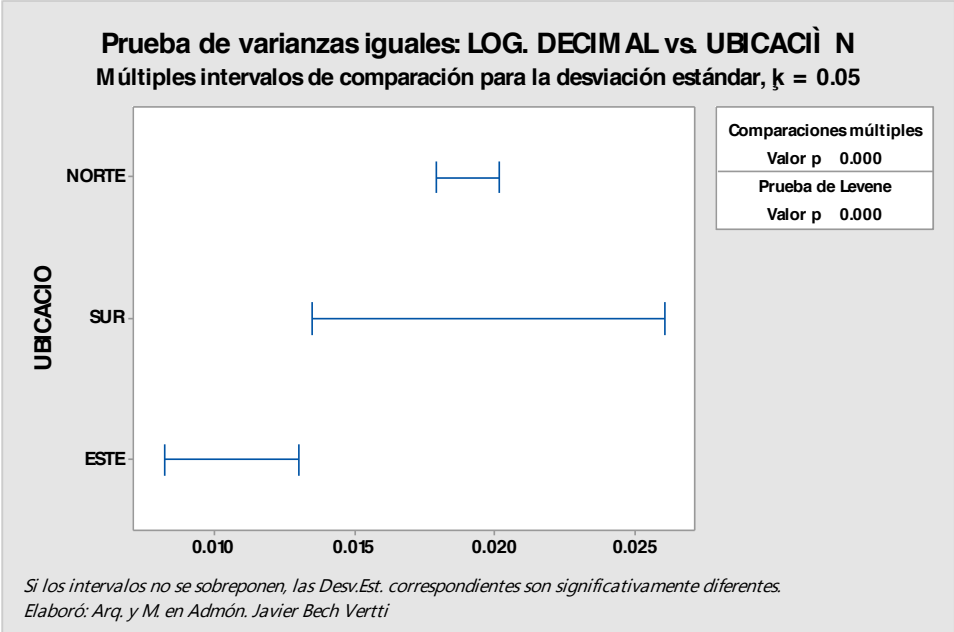
Salida de la ventana Sesión

Prueba de varianzas iguales: Log vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.            IC

       1  74  0.0186864  (0.0159409, 0.0226370)

       2  60  0.0183643  (0.0124248, 0.0282711)

       3  99  0.0102332  (0.0082331, 0.0130345)


Nivel de confianza individual = 98.3333%

Pruebas


                         Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.000

Levene                         10.36    0.000
```

Prueba de varianzas iguales: Log vs. UBICACIO

Salida de la ventana Gráfica



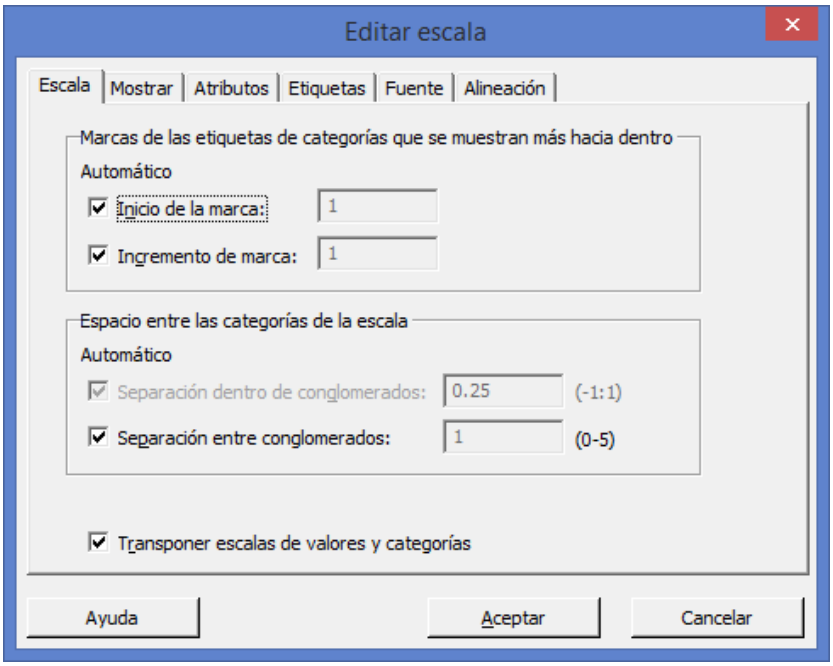
Interpretación de los resultados

Para el ejemplo para evaluar la transformación del **LOGARITMO DECIMAL** del **VALOR** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

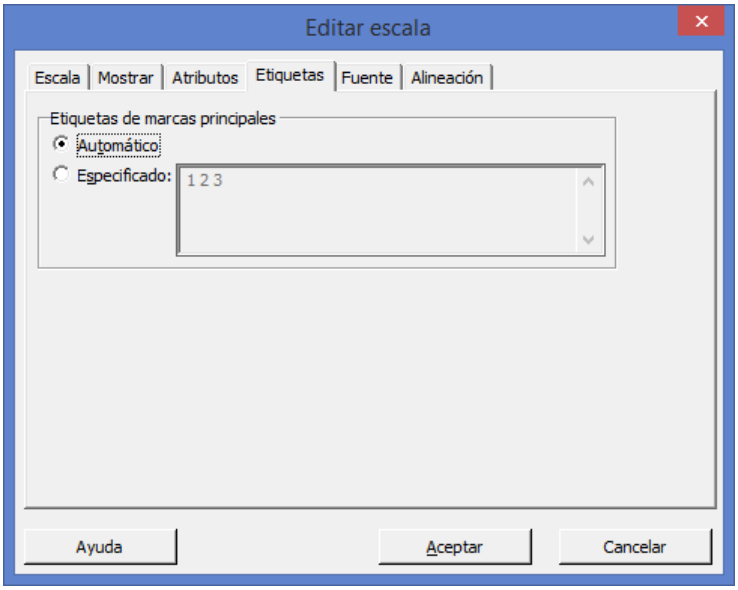
Nota Importante: No es del alcance de este trabajo tomar la decisión de si se debe o no trabajar con los datos transformados o sin transformar para usar alguna técnica multivariada sino solamente mostrar la forma de hacer dichas transformaciones.

ETIQUETADO DE LOS NIVELES DE LA VARIABLE O "FACTOR" UBICACIÓN EN MINITAB 17:

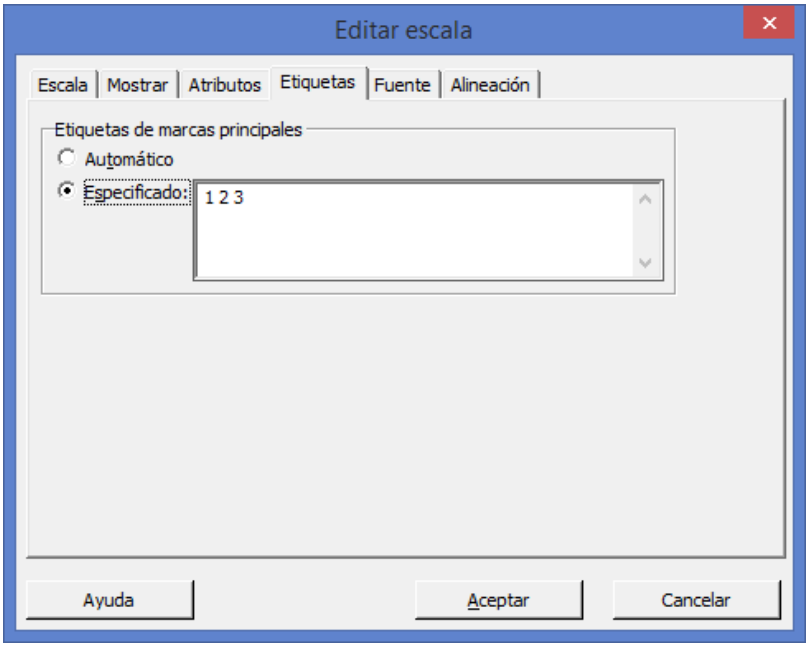
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **UBICACIÓN** para que Minitab despliegue la ventana ***Editar Escala.***



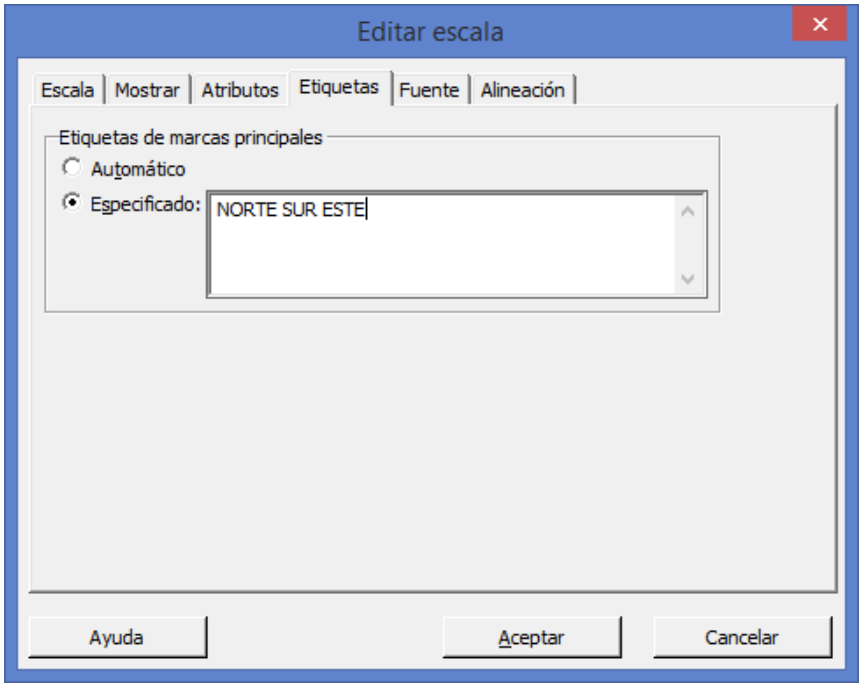
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

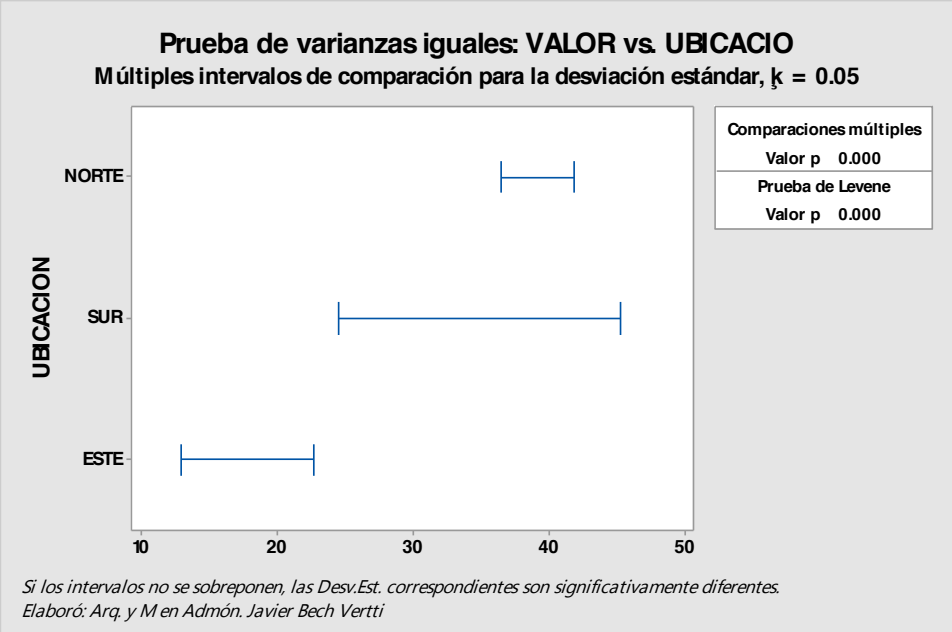


4 Borrar los números 1 2 3 4 5 y escribir **NORTE SUR ESTE**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



1.3.2.3 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 1.3.2.3**

PRUEBA **DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT**

Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con base en la **UBICACIÓN** de las casas. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas.

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***VALOR***

4 En **Factores**, ingrese ***UBICACIÓN***

5 En el botón **Opciones** active la casilla que dice: ***Utilice la prueba basándose en la distribución normal.*** Haga clic en **Aceptar** y nuevamente **Aceptar**

Salida de la ventana Sesión

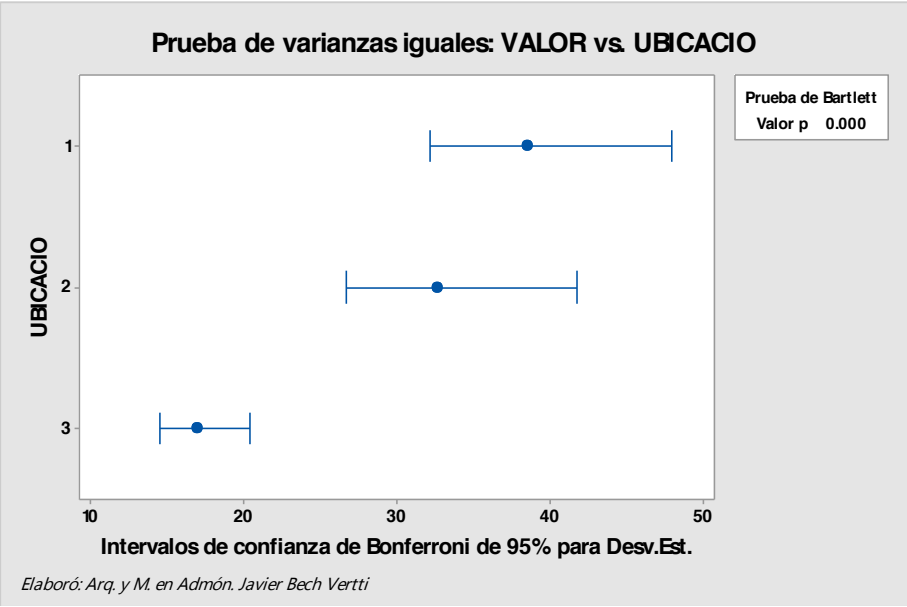
Prueba de varianzas iguales: VALOR vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales
Hipótesis alterna       Por lo menos una varianza es diferente
Nivel de significancia  α = 0.05

Se utiliza el método de Bartlett. Este método es exacto sólo para datos
     normales.


Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.          IC
       1  74    38.4944  (32.0901, 47.8492)
       2  60    32.6097  (26.6832, 41.6492)
       3  99    16.9209  (14.4361, 20.3650)

Nivel de confianza individual = 98.3333%
```

```
Pruebas

          Estadística
Método      de prueba  Valor p
Bartlett        56.55    0.000
```

Prueba de varianzas iguales: VALOR vs. UBICACIO

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de la prueba de **Barlett** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los**

datos provengan de distribuciones continuas, pero no necesariamente normales). Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **VALOR** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

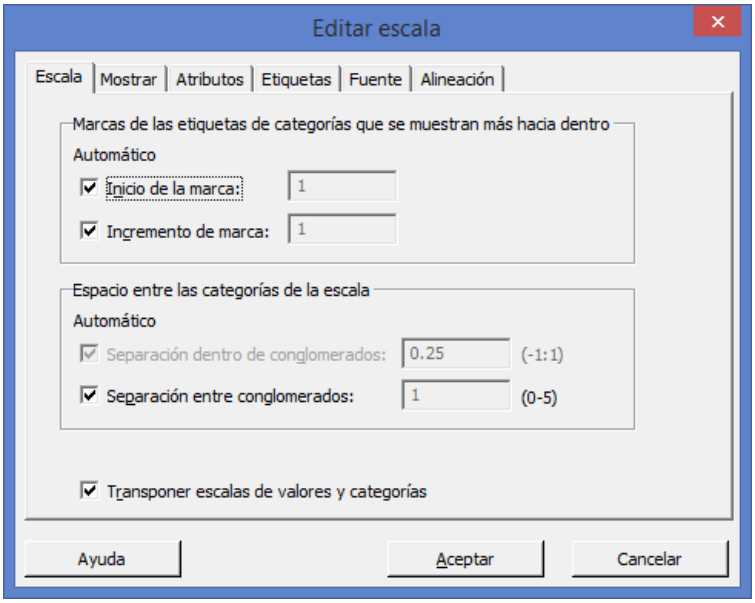
Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

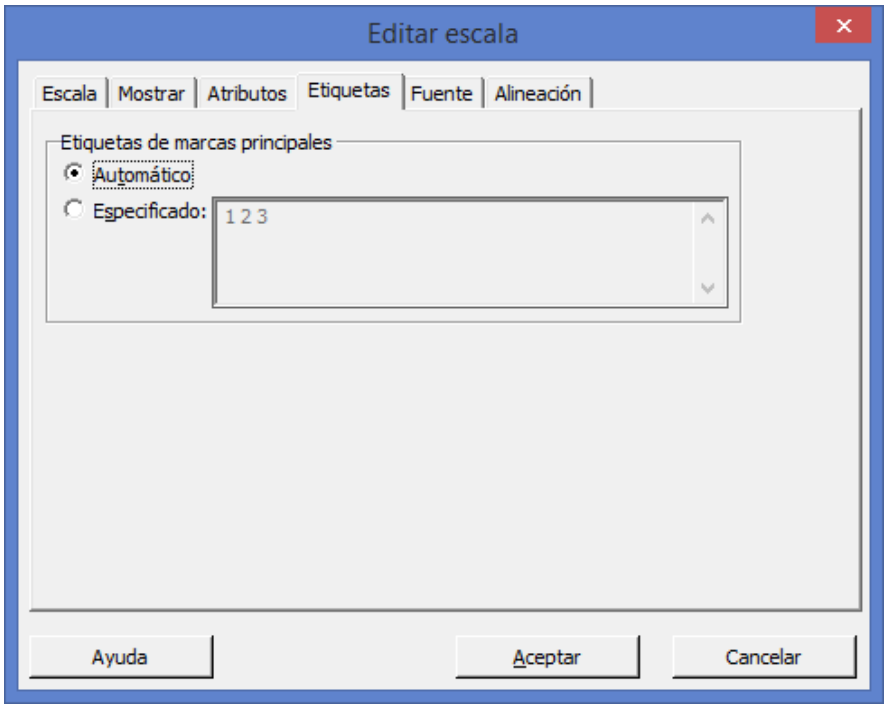
Asímismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución y en la falta de homocedasticidad.

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN:

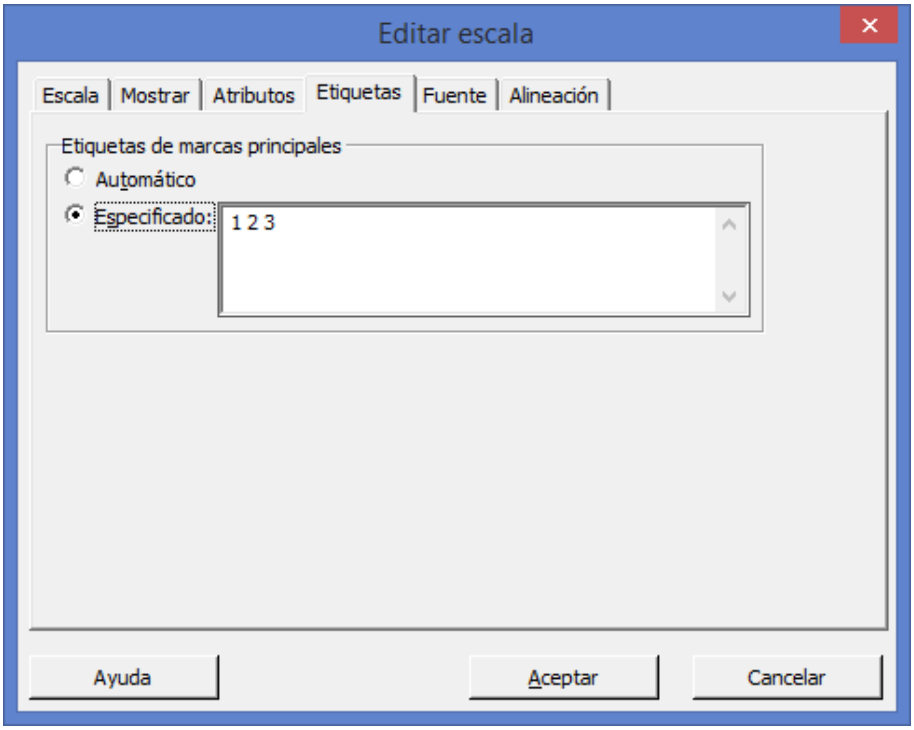
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **UBICACIÓN** para que Minitab despliegue la ventana ***Editar Escala.***



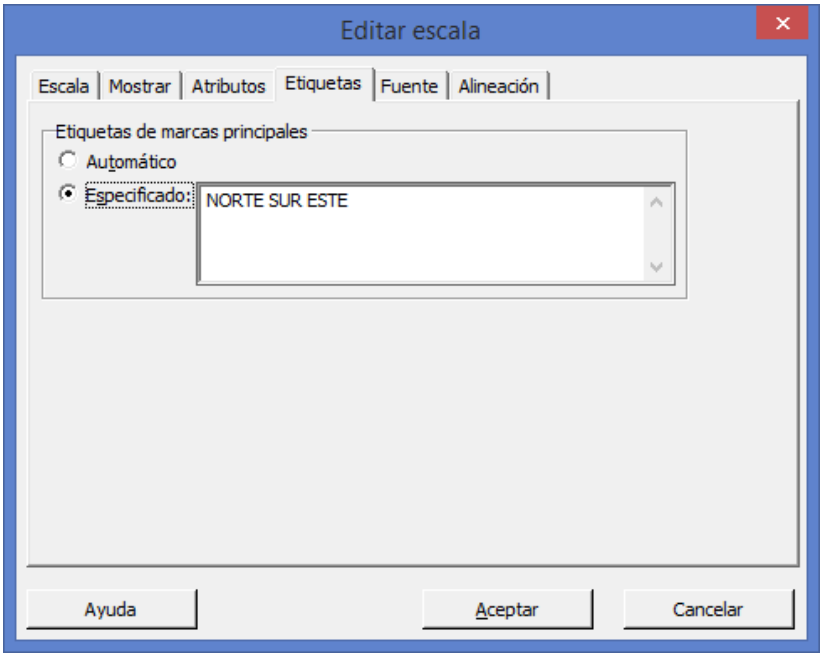
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

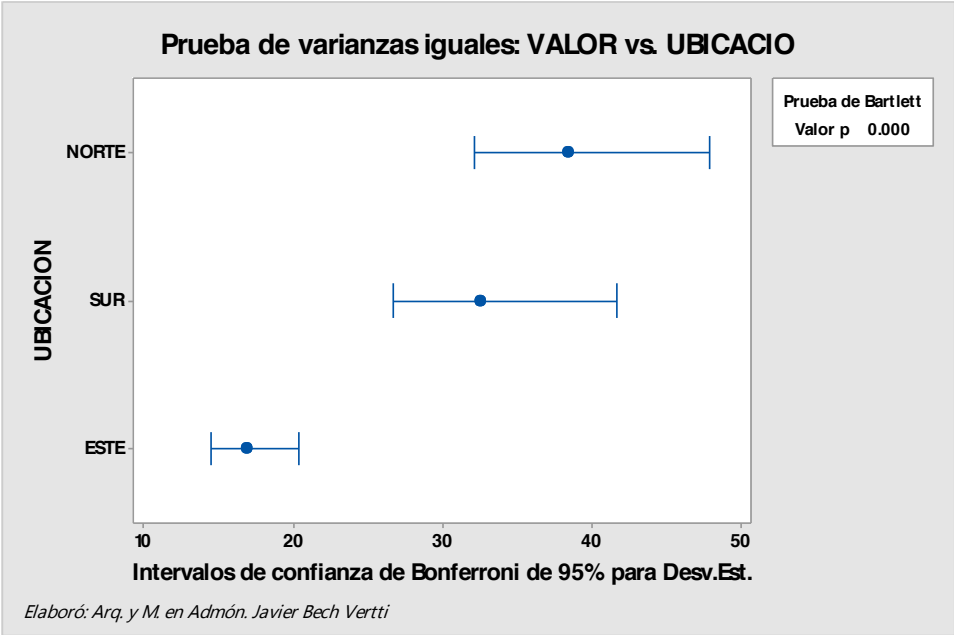


4 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

1.4.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE APRENDIZAJE 1.4.1.1

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ INVERSA

Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

x + y + z = 11

2x – y +z = 5

3x + 2y +z= 24

Paso 1. Expresar el sistema como A x X = K

$$\underbrace{\begin{bmatrix}1 & 1 & 1\\ 2 & -1 & 1\\ 3 & 2 & 1\end{bmatrix}}_{A} \text{ x } \underbrace{\begin{bmatrix}x\\ y\\ z\end{bmatrix}}_{X} = \underbrace{\begin{bmatrix}11\\ 5\\ 24\end{bmatrix}}_{K}$$

Paso 2. Buscar el valor del determinante $|A|$=

Aplicando la regla de Sarrus:

$$|A| = \begin{vmatrix}1 & 1 & 1\\ 2 & -1 & 1\\ 3 & 2 & 1\\ 1 & 1 & 1\\ 2 & -1 & 1\end{vmatrix} = (1)(-1)(1)+(2)(2)(1)+(3)(1)(1)-(3)(-1)(1)-(1)(2)(1)-(2)(1)(1)=5\neq 0 \text{ (Tiene Inversa)}$$

Paso 3. Obtener la matriz transpuesta de la matriz A, que es:

$$A^T = A' = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A′, es decir formar α A′

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} -1 & 2 \\ 1 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} & +\begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} & +\begin{vmatrix} 1 & 3 \\ 1 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 1 & 1 \end{vmatrix} \\ +\begin{vmatrix} 2 & 3 \\ -1 & 2 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 1 & 2 \end{vmatrix} & +\begin{vmatrix} 1 & 2 \\ 1 & -1 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(-1-2) & -(1-2) & +(1+1) \\ -(2-3) & +(1-3) & -(1-2) \\ +((4+3) & -(2-3) & +(-1-2) \end{bmatrix} = \begin{bmatrix} -3 & +1 & +2 \\ +1 & -2 & +1 \\ +7 & +1 & -3 \end{bmatrix}$$

Paso 5. Calcular la inversa de la matriz A′= $\frac{1}{|A|}$ α A′

$$\frac{1}{5}\begin{bmatrix} -3 & +1 & +2 \\ +1 & -2 & +1 \\ +7 & +1 & -3 \end{bmatrix} = \begin{bmatrix} -3/5 & +1/5 & +2/5 \\ +1/5 & -2/5 & +1/5 \\ +7/5 & +1/5 & -3/5 \end{bmatrix}$$

Paso 6. Llevar este valor a la expresión X=A^{-1} x K y tenemos:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -3/5 & +1/5 & +2/5 \\ +1/5 & -2/5 & +1/5 \\ +7/5 & +1/5 & -3/5 \end{bmatrix} x \begin{bmatrix} 11 \\ 5 \\ 24 \end{bmatrix} =$$

$$\begin{bmatrix} \left(-\frac{3}{5} + \frac{1}{5} + \frac{2}{5}\right) * \begin{bmatrix}11\\5\\24\end{bmatrix} \\ \left(+\frac{1}{5} - \frac{2}{5} + \frac{1}{5}\right) * \begin{bmatrix}11\\5\\24\end{bmatrix} \\ \left(+\frac{7}{5} + \frac{1}{5} - \frac{3}{5}\right) * \begin{bmatrix}11\\5\\24\end{bmatrix} \end{bmatrix} = \begin{bmatrix} -33/5 & +5/5 & +48/5 \\ 11/5 & -10/5 & +24/5 \\ 77/5 & +5/5 & -72/5 \end{bmatrix}$$

$$= \begin{bmatrix}20/5\\25/5\\10/5\end{bmatrix} = \begin{bmatrix}4\\5\\2\end{bmatrix} = \begin{bmatrix}x\\y\\z\end{bmatrix}$$

Paso 7. La solución es x=4, y=5, z=2

1.4.1.2 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE EN EXCEL 1.4.1.2**

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**

Resolver el siguiente sistema, empleando el método de la MATRIZ INVERSA con Excel:

x + y + z = 11

2x – y +z = 5

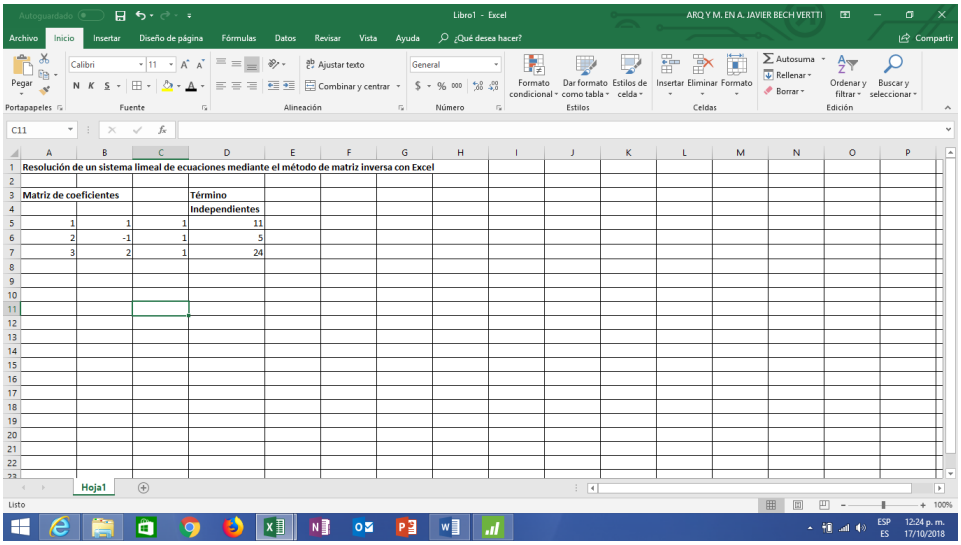
3x + 2y +z= 24

Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

5. Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).

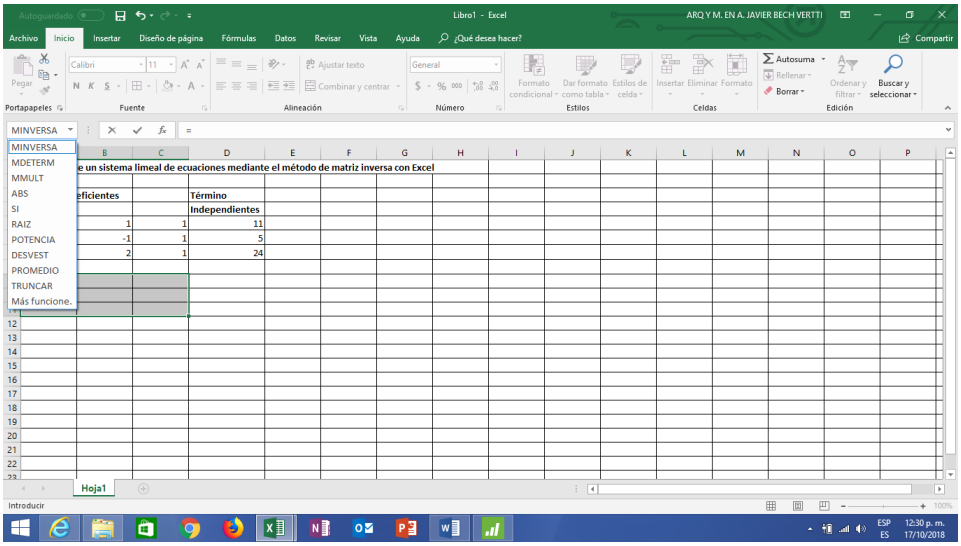
6. Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.

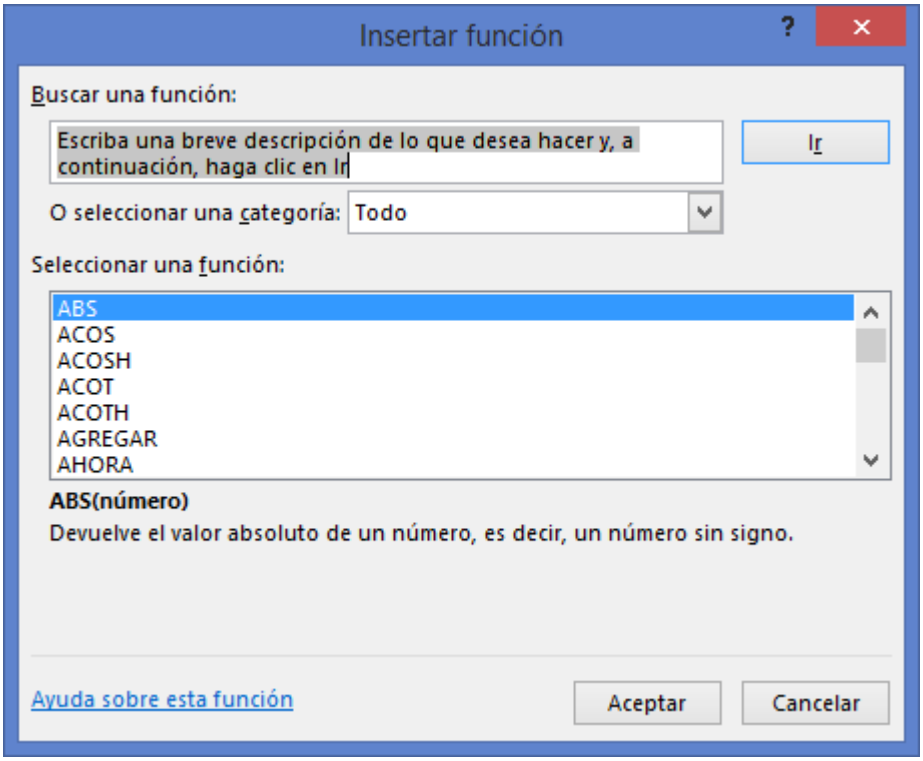
7. Introduce los valores independientes en otro rango al lado.



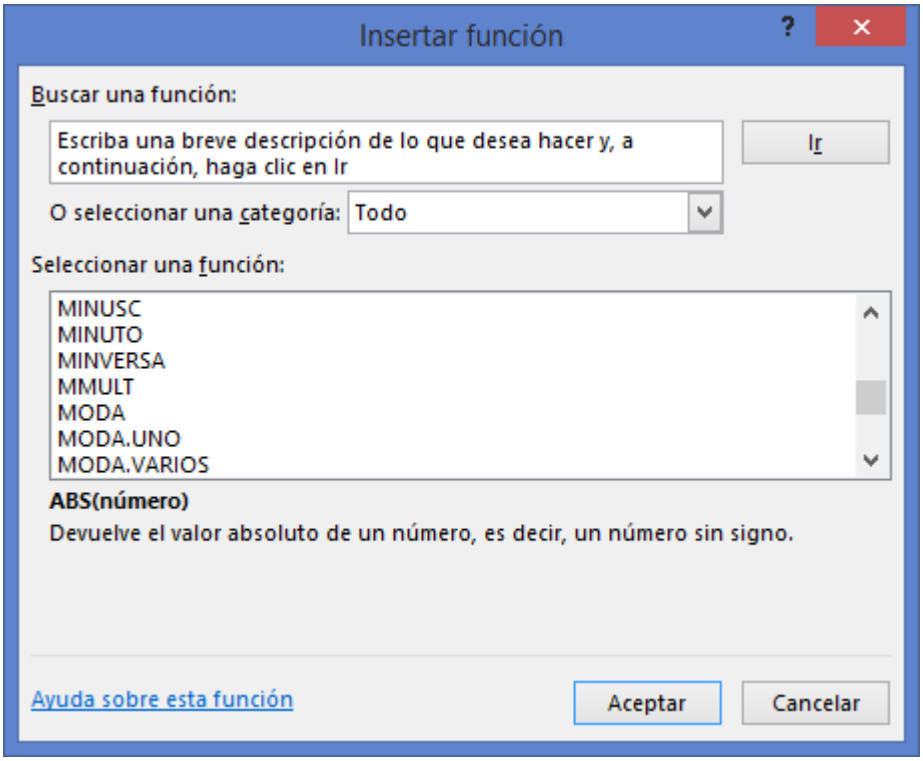
8. Utiliza la función matricial **MINVERSA ()** para calcular la matriz inversa de los coeficientes introducidos. (Recuerda que para introducir la función debes pulsar Ctrl + Mayús + Intro).

Como la matriz de coeficientes es de 3 x 3 se necesitas 9 celdas para mostrar los resultados de la matris Inversa para lo cual selecciona con el mouse 9 celdas debajo de la matriz de coeficientes dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MINVERSA** () seleccionando la opción **Más funciones** en la parte inferior

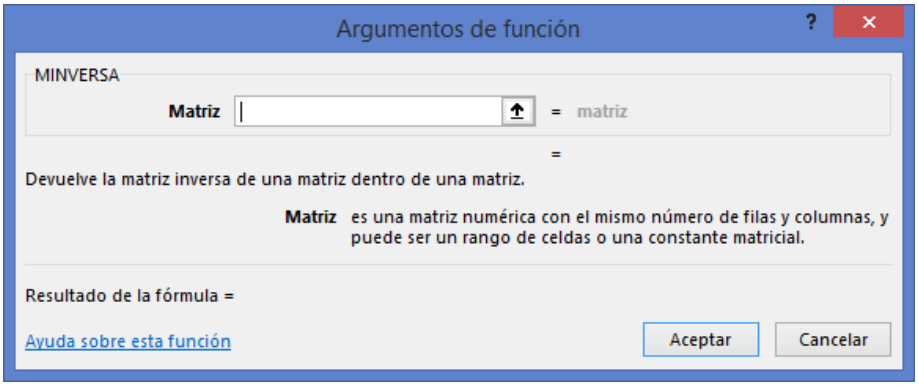




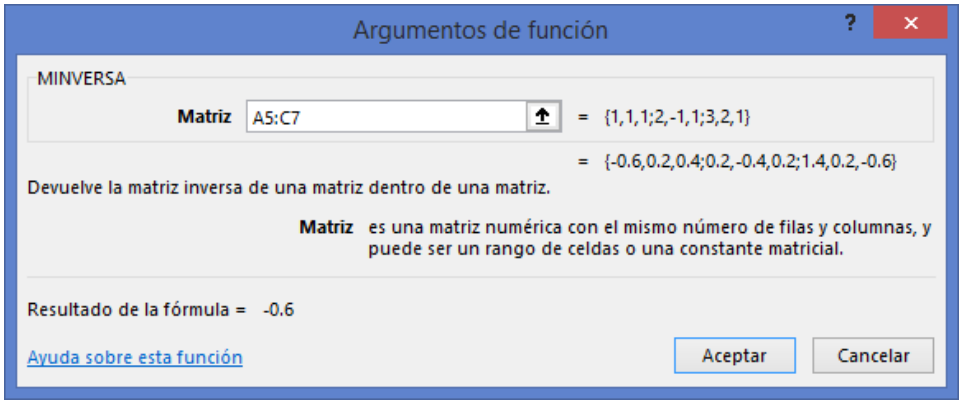
En Seleccionar una función busca **MINVERSA** y da un clic.



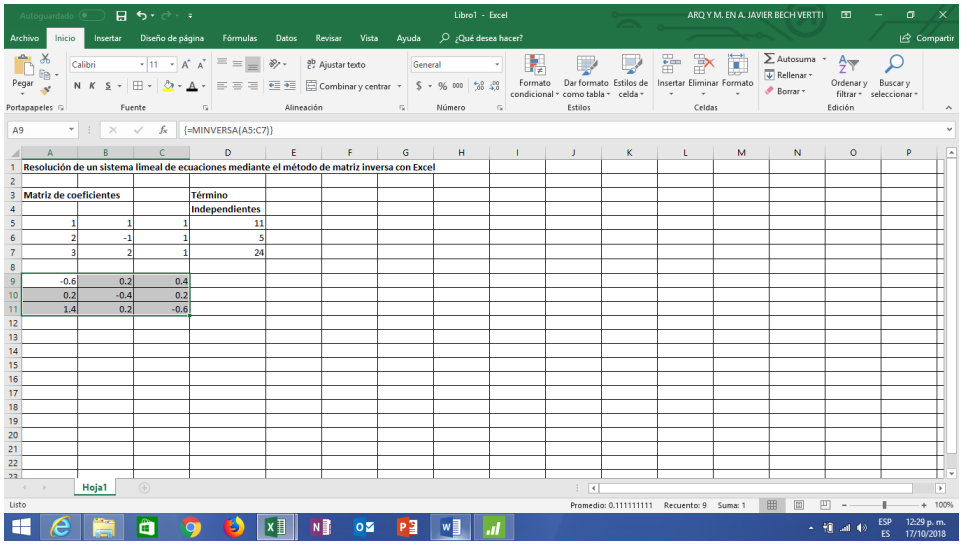
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



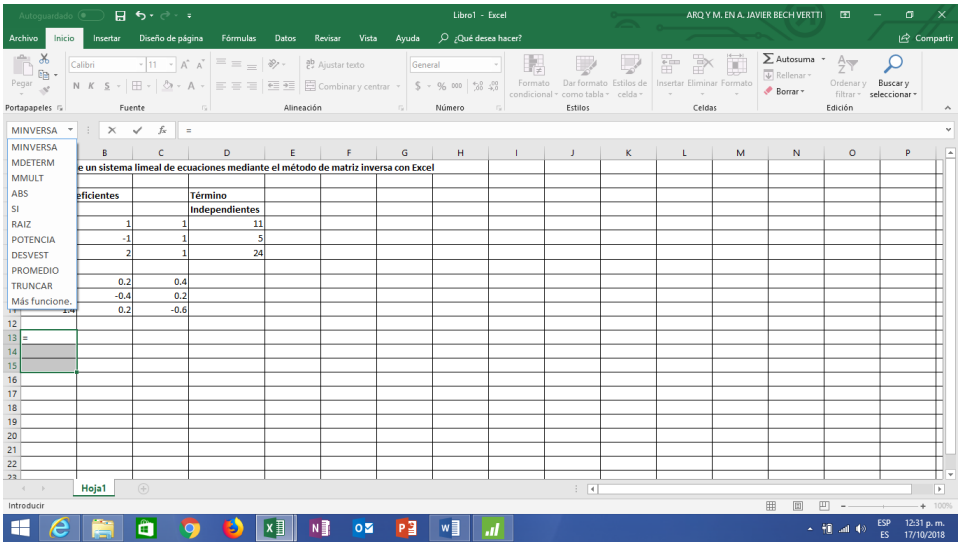
Selecciona las celdas de la matriz de coeficientes, en este caso A5:C7

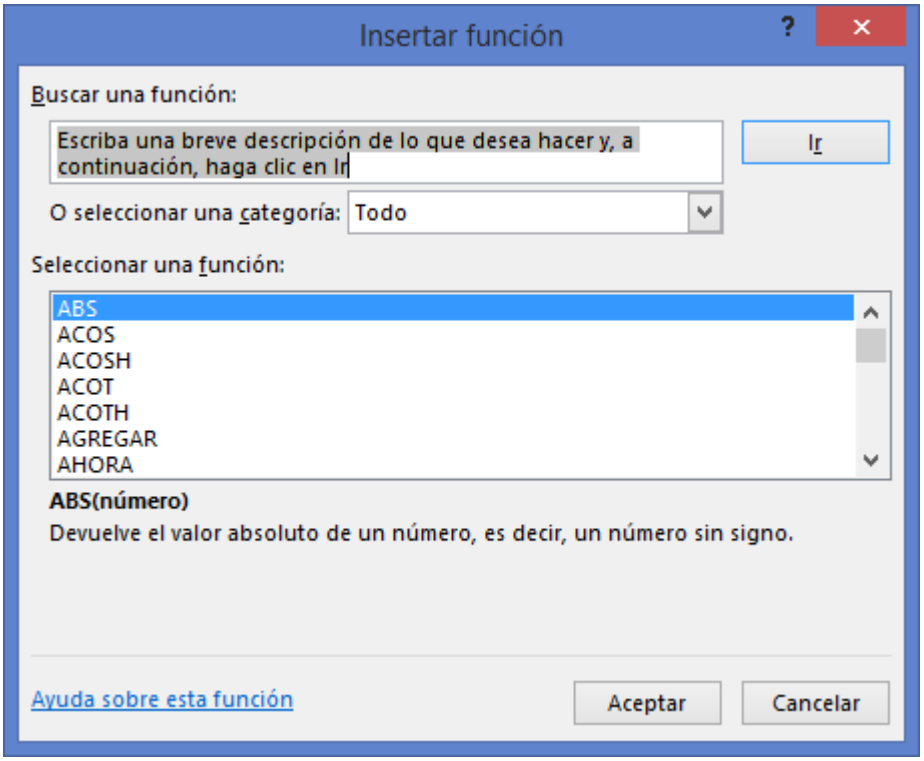


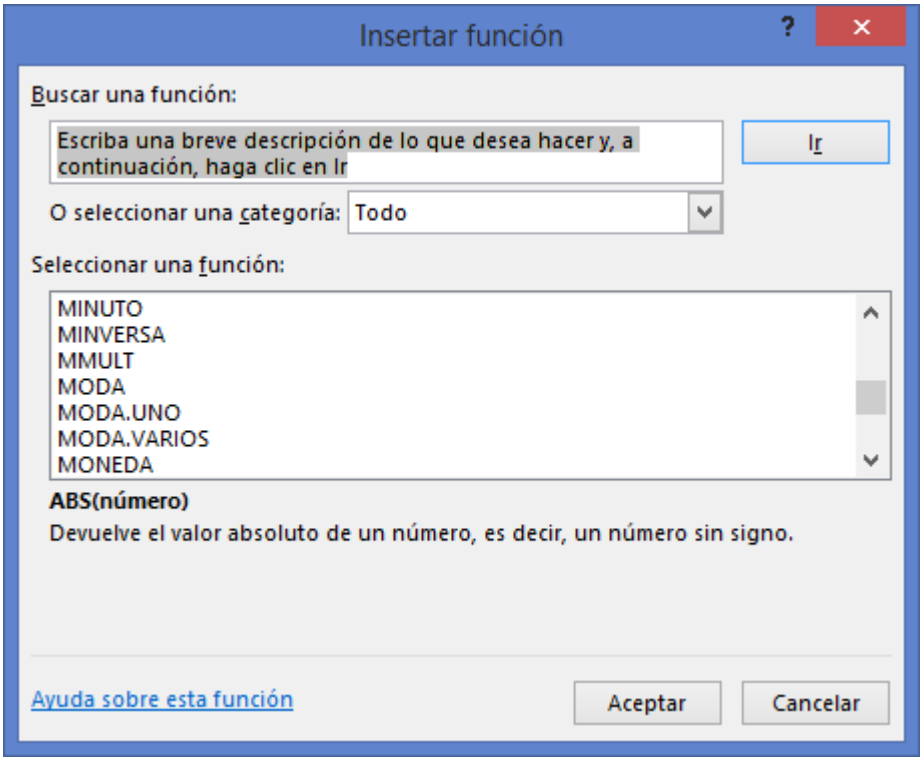
Ojo, no pulses Aceptar. La matriz inversa se determinará al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter)**.



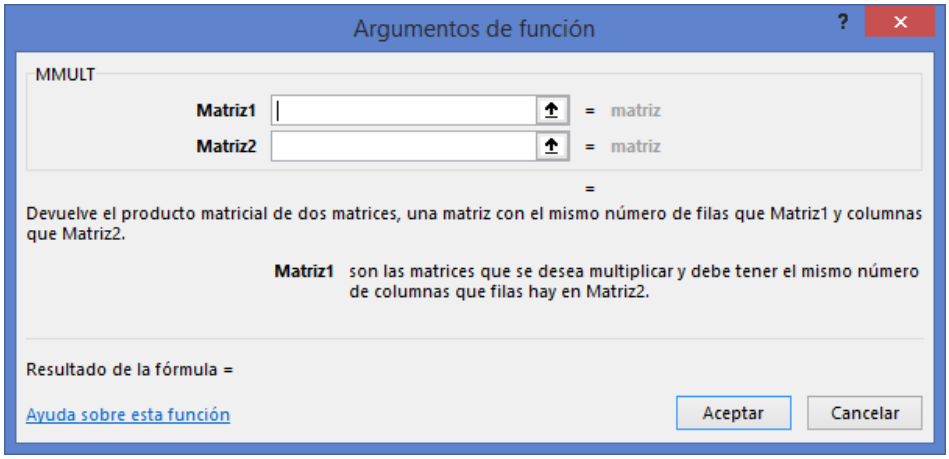
Como la matriz de los términos constantes es un vector de 1 x 3 se necesitas 3 celdas para mostrar los resultados de las Incógnitas **X, Y y Z** para lo cual selecciona con el mouse 3 celdas debajo de la **matriz inversa** dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MMULT** () seleccionando la opción **Más funciones** en la parte inferior


Libro1 - Excel
ARQ Y M. EN A. JAVIER BECH VERTTI
MINVERSA
MDETERM
MMULT
ABS
SI
RAIZ
POTENCIA
DESVEST
PROMEDIO
TRUNCAR
Más funcione...
e un sistema lineal de ecuaciones mediante el método de matriz inversa con Excel
eficientes
Término
Independientes
Hoja1
Introducir

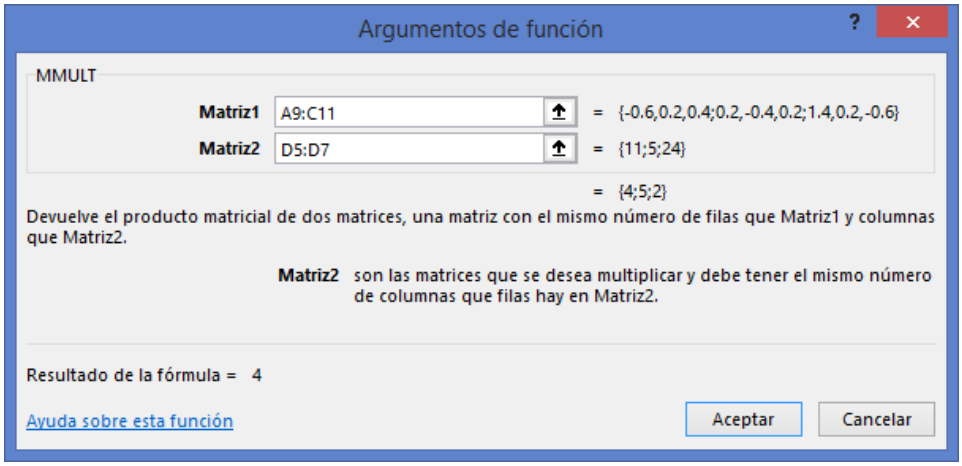

Insertar función
Buscar una función:
Escriba una breve descripción de lo que desea hacer y, a continuación, haga clic en Ir
Ir
O seleccionar una categoría: Todo
Seleccionar una función:
ABS
ACOS
ACOSH
ACOT
ACOTH
AGREGAR
AHORA
ABS(número)
Devuelve el valor absoluto de un número, es decir, un número sin signo.
Ayuda sobre esta función
Aceptar
Cancelar



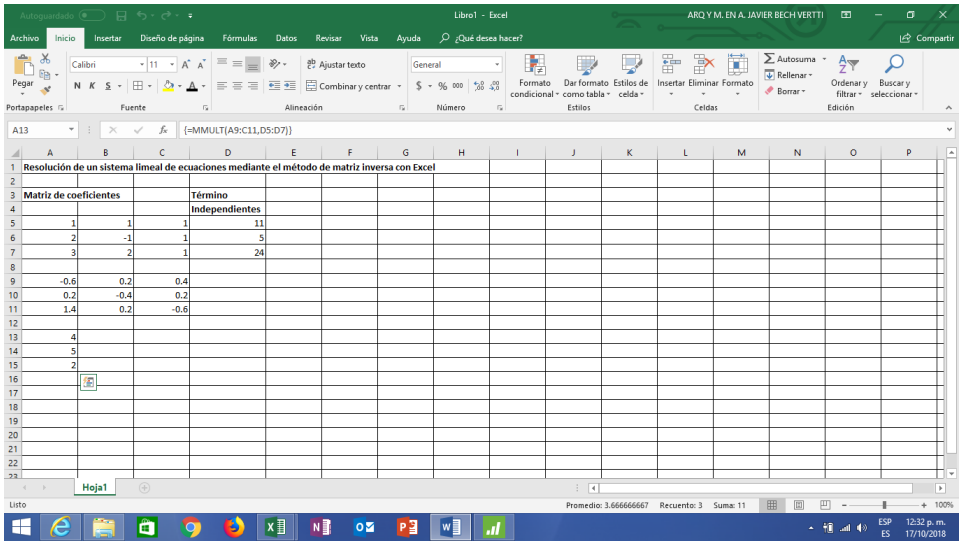
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



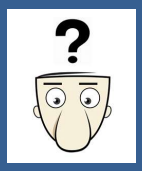
En el campo que dice Matriz 1 selecciona las celdas de la matriz de inversa, en este caso A9:C11 y en el campo que dice Matriz 2 selecciona las celdas de los términos constantes, en este caso D5:D7



Ojo, no pulses Aceptar. Los resultados de las incógnitas **X, Y, Z** se determinarán al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter), en este caso 4,5,2.** Por lo tanto las incognitas quedan **X=4, Y=5 y Z=2.**



Con lo anterior queda terminado el ejemplo ilustrativo de **resolución de un sistema lineal de ecuaciones mediante el método de matriz inversa** en el paquete **EXCEL**. **Aplica el procedimiento para salir de Window**



RESPUESTAS A LOS EJERCICIOS DE AUTOEVALUACIÓN DEL CAPÍTULO 1

1.3.1.1 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 1.3.1.1

DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE

Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

181.0	181.0	179.9	200.0	230.0	174.9
220.0	185.0	210.0	115.0	183.0	100.0
190.0	269.9	159.9	199.9	215.0	174.9
160.0	189.9	165.0	189.9	159.0	189.9
176.0	299.9	185.0	189.9	180.0	210.0

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas sin recortar los datos o valores atípicos generado por Minitab:

Tallo y hoja de VALOR N = 30

Unidad de hoja = 10

2 1 01

2 1

4 1 55

10 1 667777

(12) 1 888888888899

8 2 0111

4 2 23

2 2

2 2 6

1 2 9

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas recortando los datos o valores atípicos generado por Minitab:

Tallo y hoja de VALOR N = 30

Unidad de hoja = 1.0

LO 100, 115

4 15 99

6 16 05

10 17 4469

(10) 18 0113559999

10 19 09

8 20 0

7 21 005

4 22 0

3 23 0

HI 269, 299

Otra forma de ordenarlos es colocarlos en forma ascendente:

100.0	165.0	180.0	185.0	190.0	215.0
115.0	174.9	181.0	189.9	199.9	220.0
159.0	174.9	181.0	189.9	200.0	230.0
159.9	176.0	183.0	189.9	210.0	269.9
160.0	179.9	185.0	189.9	210.0	299.9

Resumen de 5 Números: Resultados aproximados a las décimas.

1. Valor Mínimo: **100.0**
2. Cuartil 1 Q_1=Dato 7 + 0.75 (Dato 8-Dato 7)= 174.9 + 0.75 (174.9-174.9)=**174.9**

   $P(Q_1)= 1\ \frac{30+1}{4}= 7.75$

3. Mediana $\tilde{X}=\frac{Dato\ 16-Dato15}{2}=\frac{185.0+185.00}{2}$=**185.0**

P($\tilde{X}$)=$\frac{n+1}{2}=\frac{30+1}{2}$=15.5

4. Cuartil 3 Q_3=Dato 23+ 0.25(Dato 24- Dato 23)=200.0+0.25(210.0-200.0)=**202.5**

P(Q_3)= 3 $\frac{30+1}{4}$= 23.25

5. Valor Máximo= **299.9**

Esquema de vallas: Resultados aproximados a las décimas

VEI (Valla Externa Inferior)= Q_1-3(RIQ)=174.9-3(202.5-174.9)= 174.9-3(27.6)=**92.1**

VII (Valla Interna Inferior)= Q_1-1.5(RIQ)=174.9-1.5(202.5-174.9)= 174.9-1.5(27.6)=**133.5**

VIS (Valla Interna Superior)= Q_{3+}1.5(RIQ)=202.5+1.5(202.5-174.9)= 202.5+1.5(27.6)=**243.9**

VES (Valla Externa Superior)= Q_3-3(RIQ)=202.5+3(202.5-174.9)= 202.5+3(27.6)=**285.3**

Diagrama de Caja y Brazos

Interpretación: En el diagrama anterior se puede observar que existen 4 valores atípicos: **100.0, 115.0** entre las **VEI y la VII (a los que denominamos aberrantes),** el **269.9** entre las **VIS y la VES (al que denominamos aberrante)** y el **299.9 despues de la VES (al que denominamos aberrante lejano)** los que deben ser estudiados con detalle para ver el tratamiento que se les va a dar en estudios subsecuentes ya que es posible que provoquen un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la izquierda de la caja, sugiere un sesgo hacia la derecha producto tal vez de los valores o casos atípicos).

1.3.1.2 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 1.3.1.2

DATOS ATÍPICOS

Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **ANTIGÜEDAD DE LAS CASAS** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver esta Autoevaluación.**

1 Abra la base de datos que creó **"BIENES1".** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***ANTIGUED***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

Salida de la ventana Sesión

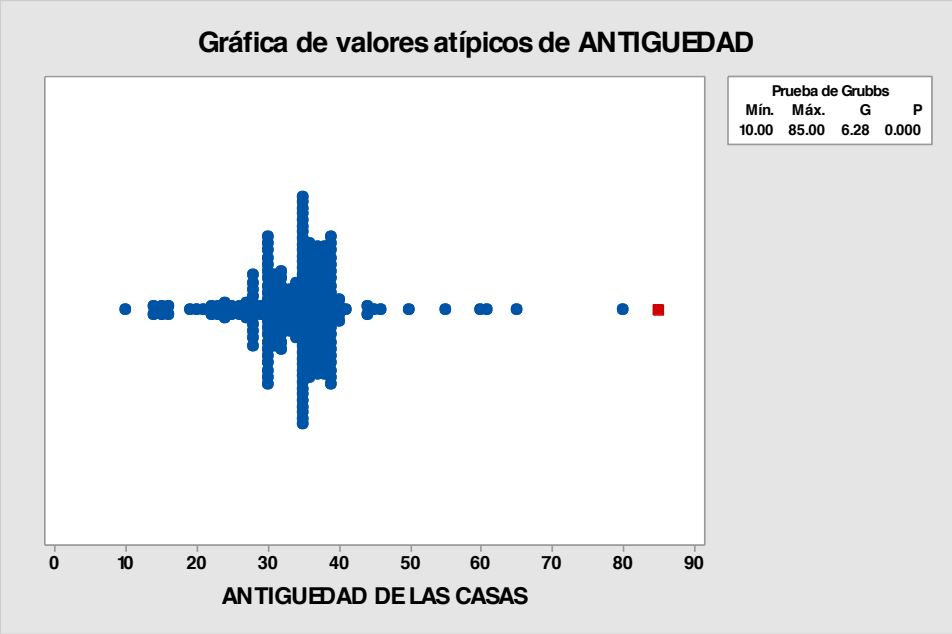
Prueba de valores atípicos: ANTIGUED

```
Método

Hipótesis nula          Todos los valores de los datos provienen de la misma
población normal
```

```
Hipótesis alterna       El valor más pequeño o más grande de los datos es un
valor atípico
Nivel de significancia  α = 0.05


Prueba de Grubbs

Variable    N   Media  Desv.Est.    Mín.    Máx.     G      P
ANTIGUED  233  34.322      8.071  10.000  85.000  6.28  0.000


Valor atípico

                 Valor
Variable  Fila  atípico
ANTIGUED   124       85
```

Gráfica de valores atípicos de ANTIGUED

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **ANTIGÜEDAD DE LAS CASAS** muestran el valor más pequeño es de 10.00 y el más grande de 85.00. La estadística de G de 6.28 indica que el valor de datos más pequeño ó más grande de 10.00 ó 85.00, es 6.28 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida,

entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.000. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico y corresponde al caso o la casa 124 con un valor máximo de 85.00 años.

Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

1.3.1.3 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 1.3.1.3

DATOS ATÍPICOS PARA UN NIVEL DE UN FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO

Se reciben datos sobre una encuesta realizada a las casas ubicadas en la zona **NORTE** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en esta zona **NORTE** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

c) Crear un subconjunto de hoja de trabajo.
d) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver esta Autoevaluación.

Respuesta al inciso a):

Crear un subconjunto de hoja de trabajo.

Nota importante:

Al **crear un subconjunto de hoja de trabajo** se divide, o desapila, la hoja de trabajo activa en hojas de trabajo nuevas basándose en una o más "Por" variables. Crear subconjunto de hoja de trabajo siempre copian los datos en hojas de trabajo **nuevas**. Como la base de datos **"BIENES 1"** contiene todas las casas ubicadas en la zona **NORTE, SUR y ESTE** debemos crear **un subconjunto de hoja de trabajo** que solo incluya las casas ubicadas en el **NORTE**. Esto nos permite analizar los diferentes niveles de un factor por separado en caso de que la prueba de **ANOVA** haya sido significativa y exista la posibilidad de que las medias de la variable **VALOR** no sea la misma en los tres niveles del Factor **UBICACIÓN**

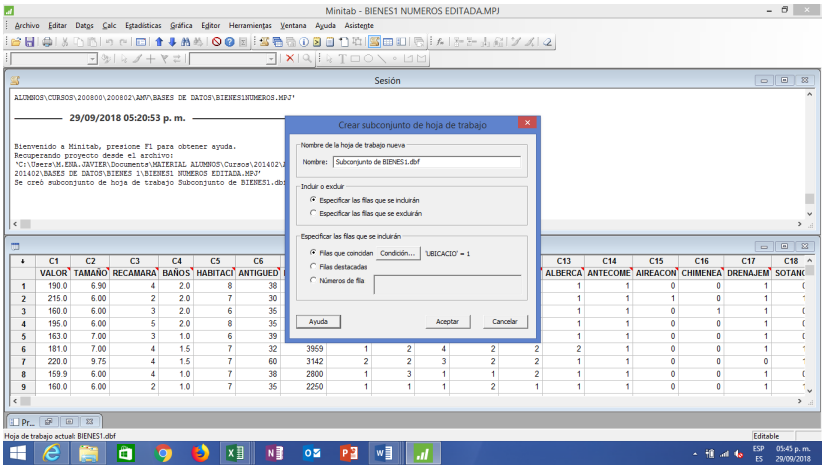
1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Datos > Crear subconjunto de hoja de trabajo**.

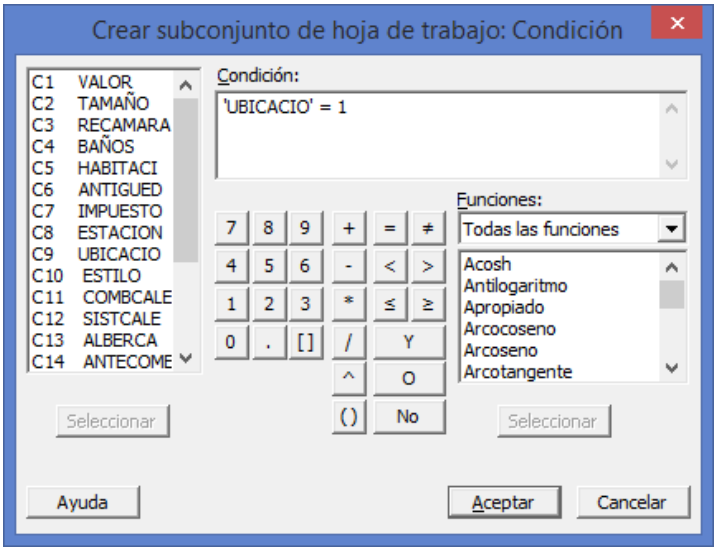
3 En **Incluir o excluir**, elija **Especificar las filas que se incluirán.**

4 En **Especificar las filas que se incluirán**, elija **Filas que coincidan**. Haga clic en **Condición**.

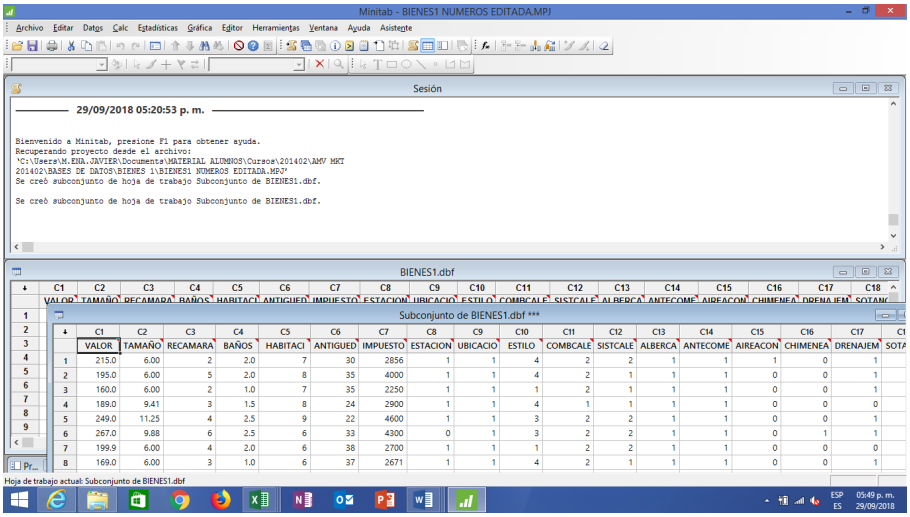
5 En **Condición**, ingrese ***UBICACIÓN=1.*** Haga clic en **Aceptar** en cada cuadro de diálogo.



Nota importante: Minitab es sensible al nombre que se le asignó a la variable por lo cual es conveniente no escribir el nombre de la variable sino seleccionarla de la parte izquierda de la ventana y solo añadir **=1.**



Minitab Crea un subconjunto de hoja nueva que incluye las casas ubicadas en la zona **NORTE**



Respuesta al inciso b):

Realizar la Prueba de valores atípicos de Minitab

1 En la nueva hoja de trabajo que Minitab generó de la base de datos **"BIENES1"** que incluye solo las casas ubicadas en la zona **NORTE** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***Valor***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

Salida de la ventana Sesión

Prueba de valores atípicos: VALOR

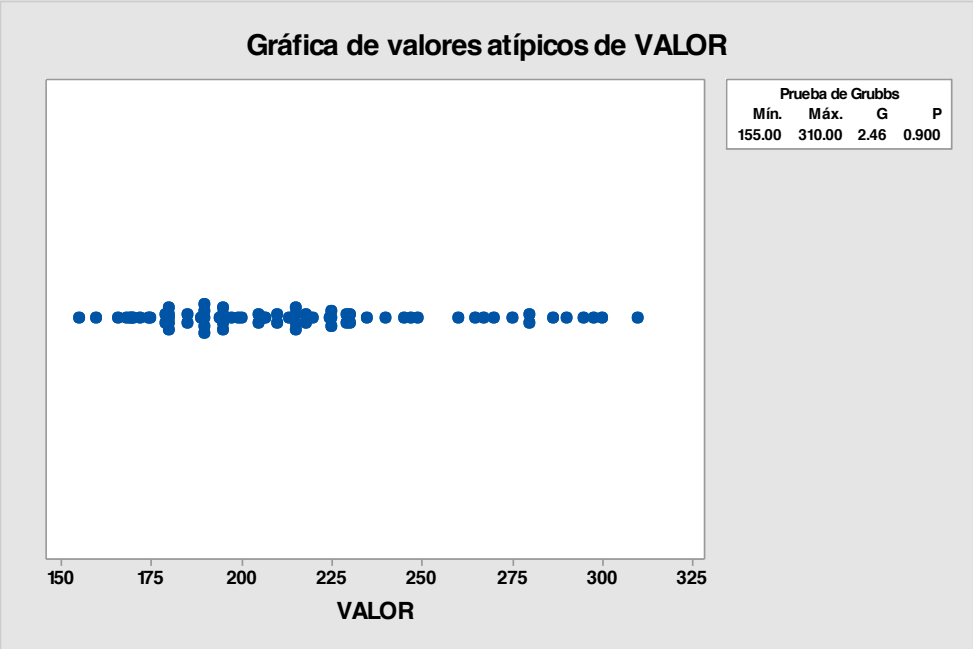
Método

Hipótesis nula	Todos los valores de los datos provienen de la misma población normal
Hipótesis alterna	El valor más pequeño o más grande de los datos es un valor atípico
Nivel de significancia	α = 0.05

```
Prueba de Grubbs


Variable   N   Media  Desv.Est.    Mín.    Máx.     G      P
VALOR     74  215.23      38.49  155.00  310.00  2.46  0.900


* NOTA * No hay valor atípico en el nivel de significancia de 5%
```

Gráfica de valores atípicos de VALOR

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **VALOR** muestran el valor más pequeño es de 155 y el más grande de 310. La estadística de G de 2.46 indica que el valor de datos más pequeño ó más grande de 155 ó 310, es 3.45 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.900. Este valor p es mayor que el nivel de significancia que se eligió, de modo que no se puede rechazar la hipótesis nula y concluir que el valor más pequeño ó el más grande no es un valor atípico.

1.3.1.4 EJERCICIO DE AUTOEVALUACIÓN

AUTOEVALUACIÓN 1.3.1.4

DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE

Una agencia de Bienes Raíces desea examinar el **VALOR** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **VALOR** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

181.0	181.0	179.9	200.0	230.0	174.9
220.0	185.0	210.0	115.0	183.0	100.0
190.0	269.9	159.9	199.9	215.0	174.9
160.0	189.9	165.0	189.9	159.0	189.9
176.0	299.9	185.0	189.9	180.0	210.0

Utilice el paquete Minitab para resolver esta Autoevaluación.

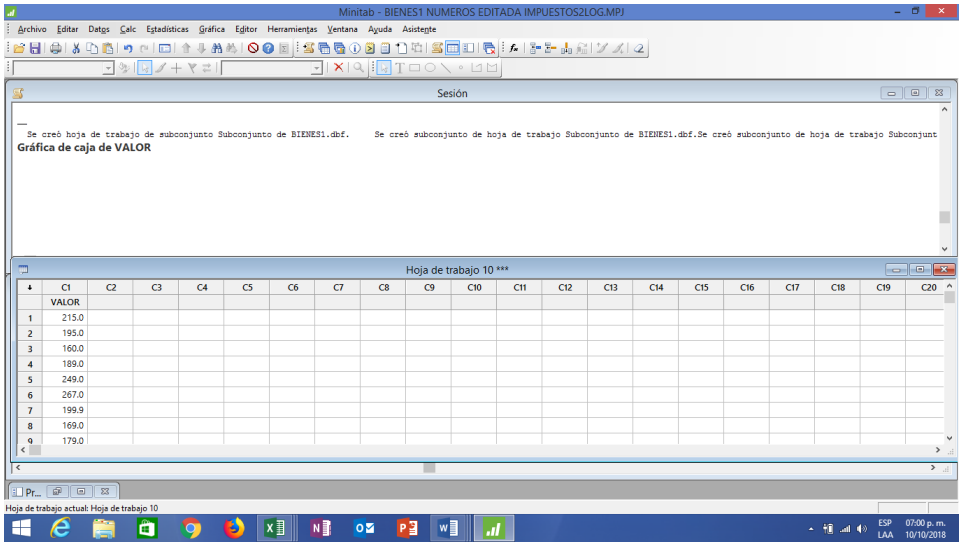
Para introducir los datos agrega una hoja de trabajo vacía al proyecto actual:

1 Elija **Archivo > Nuevo**.

2 Seleccione **Hoja de trabajo de Minitab** y haga clic en **Aceptar**.

3 En **C1** coloca el nombre de la variable **VALOR**.

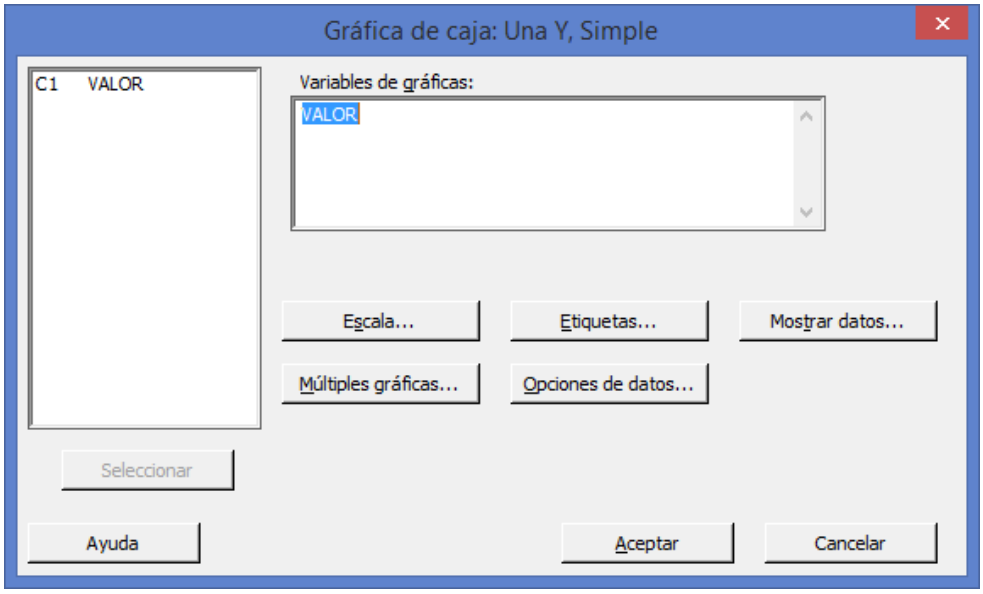
4 Teclea los datos de la tabla en forma vertical.



5 Elija **Gráfica > Gráfica de caja**.

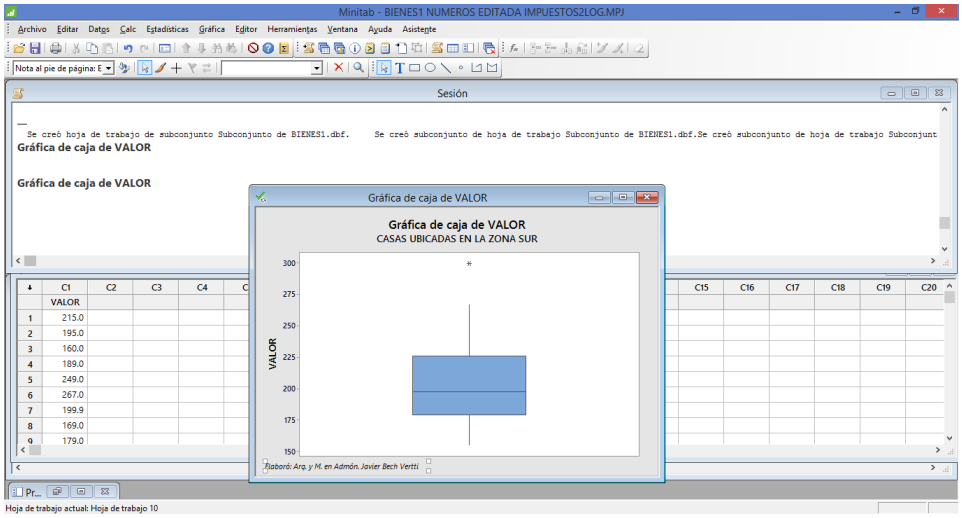
6 En **Una Y**, elija **Simple**. Haga clic en **Aceptar**.

7 En **Variables de graficas**, ingrese ***VALOR***.

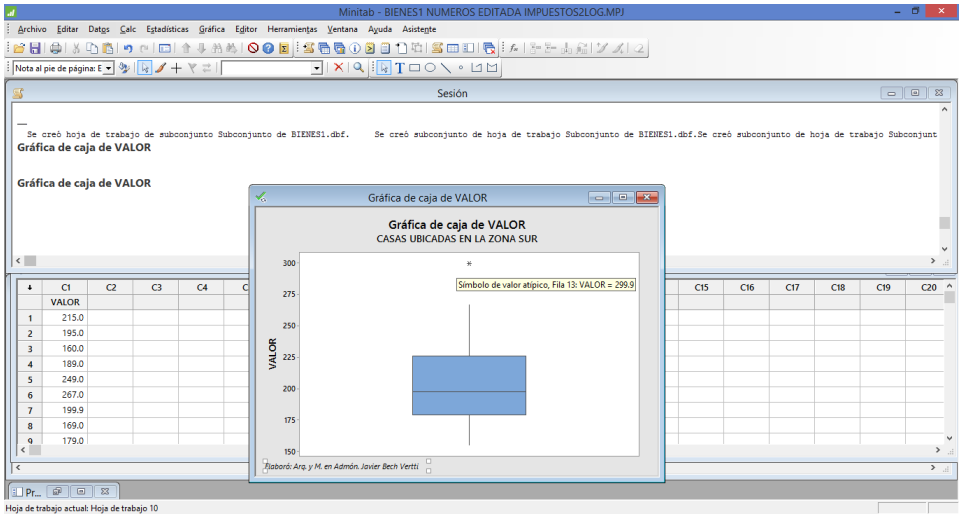


8 Haga clic en **Aceptar**.

Salida de la ventana Gráfica



Interpretación: En el diagrama anterior se puede observar que existe 1 valor atípico que pueden ser identificados colocando el puntero del mouse en cada uno de ellos proporcionandonos la siguiente información:



Fila 13 Valor 299.9

Este valor ó dato ó caso atípico debe ser estudiado con detalle para ver el tratamiento que se le va a dar en estudios subsecuentes ya que es posible que provoque un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la izquierda de la caja sugiere un sesgo hacia la derecha producto tal vez de los valores o casos atípicos.

1.3.1.5 EJERCICIO DE AUTOEVALUACIÓN

AUTOEVALUACIÓN 1.3.1.5

DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE

Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **CON AIRE ACONDICIONADO** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

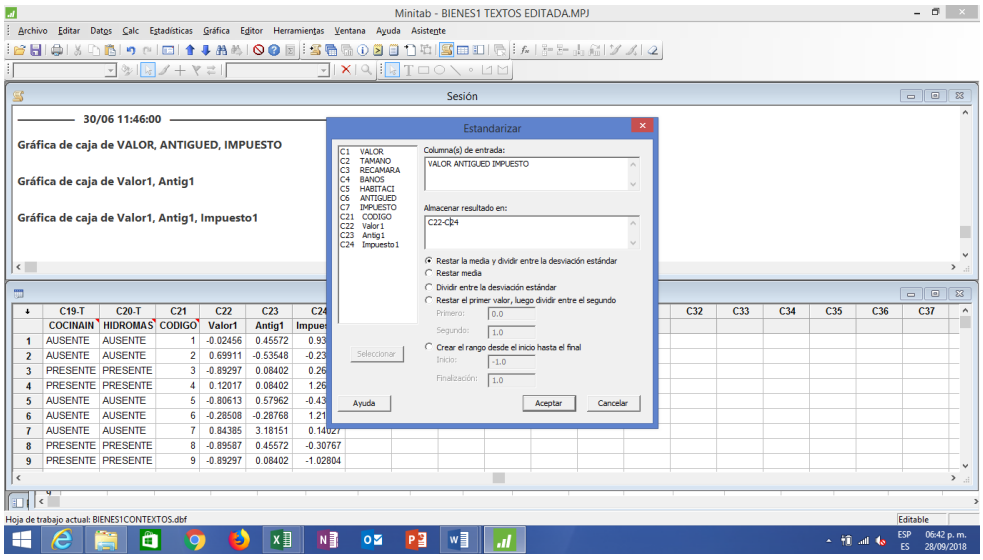
Utilice el paquete Minitab para resolver esta Autoevaluación.

Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las caas que tienen AIRE ACONDICIONADO** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables de la siguiente manera:

1 Con la hoja de trabajo con las casas que tienen **AIRE ACONDICIONADO** seleccionar **Calc > Estandarizar**.

2 En **Columnas de entrada**, ingrese las columnas que desea estandarizar. En este caso ingrese **VALOR, ANTIGUEDAD E IMPUESTOS.**

3 En **Almacenar resultados en**, ingrese las columnas en las que desea almacenar los datos estandarizados. En este caso escriba **C22-C24**, haga clic en **Aceptar**



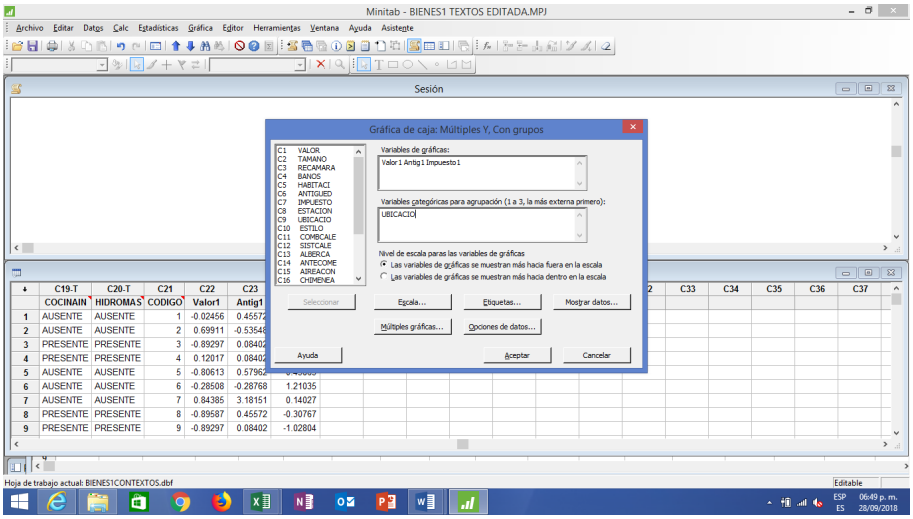
4 Los datos estandarizados aparecerán en las columas **22,23 y 24** de la base original **"BIENES1"**. Colocar el nombre de las variables estantarazadas como **Valor 1, Antig1 e Impuesto 1** para identificarlas.

5 Elija **Gráfica > Gráfica de caja**.

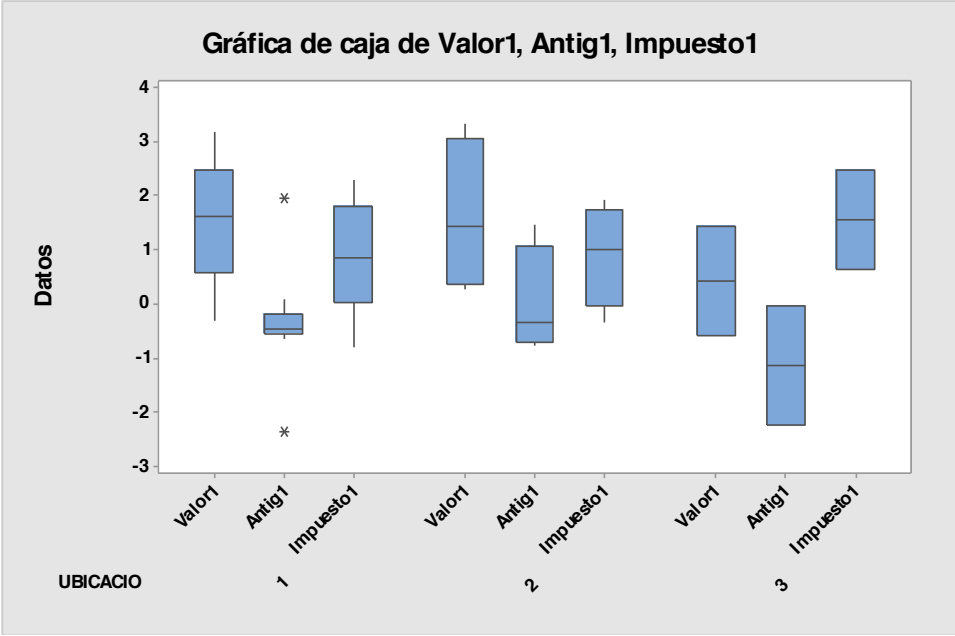
6 En **Múltiples Y**, elija **Con grupos**. Haga clic en **Aceptar**.

7 En **Variables de gráficas**, ingrese ***Valor1, Antig1 e Impuesto1***

8 En **Variables categóricas para agrupación (1 a 3, la más externa primero)**, ingrese ***UBICACIÓN.***



9 En **Nivel de escala para variables de gráficas**, elija **Las variables de gráficas se muestran más hacia dentro en la escala**. Haga clic en **Aceptar**.



Interpretación de los resultados

La gráfica de caja muestra:

· Para las casas con **AIRE ACONDICIONADO**, en general en todos los niveles de **UBICACIÓN** de las casas las tres variables no muestran valores atípicos con excepción de **Antig1** en la zona **NORTE.**

· Para las casas ubicadas en la zona **NORTE** en la variable **Antig1** se detectan dos valores atípicos en ambos extremos.

· Llama la atención la ausencia de bigotes o brazos en las csas ubicadas en la zona **ESTE** donde habría que estudiar el efecto de la Curtósis que pudiera influir en la Normalidad de los datos.

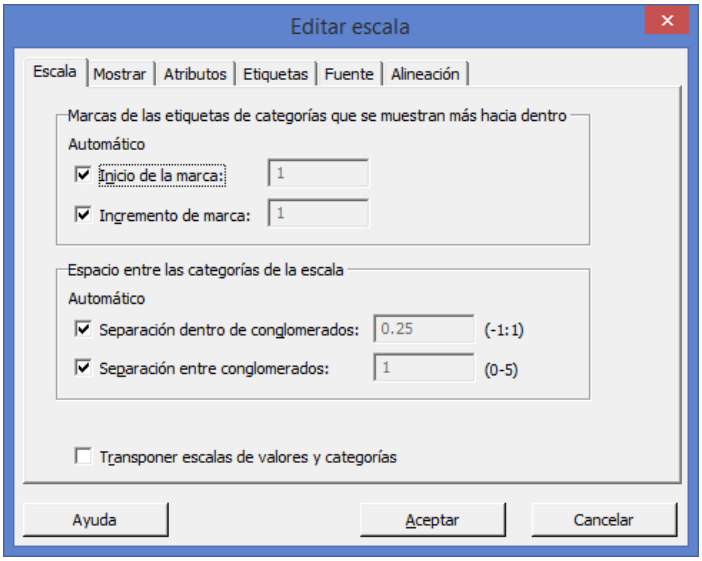
Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

Asimismo se recomienda realizar análisis estadísticos, tales como **MANOVA balanceado**, para examinar más a fondo la relación entre los factores.

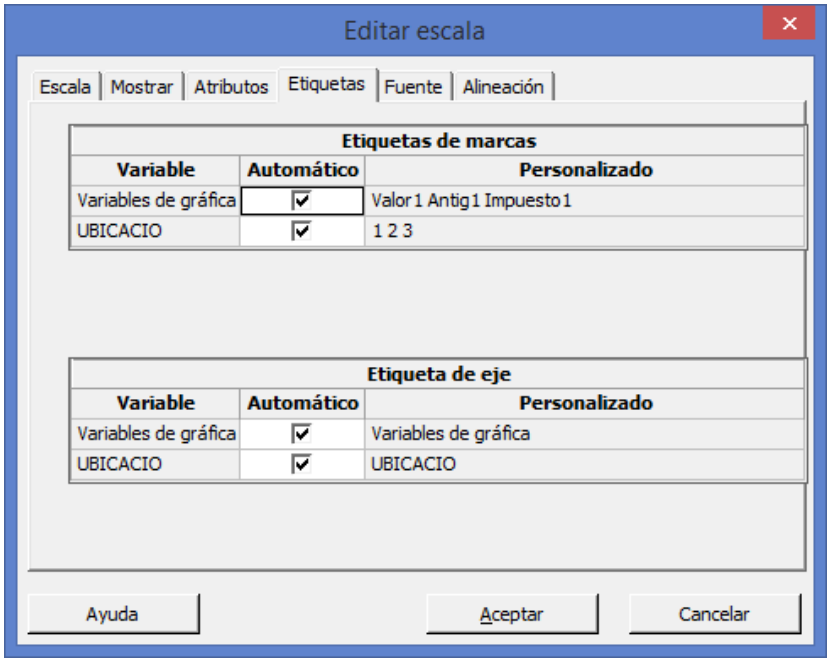
Para obtener información precisa de Q1, mediana, Q3, rango intercuartil, bigotes o brazos y N, coloque el cursor sobre cualquier parte de la gráfica de caja. Para ver el valor del eje Y y el número de fila de un valor atípico, coloque el cursor sobre el valor atípico.

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN Y Etiquetado del subtitulo del nivel AIRE ACONDICIONADO:

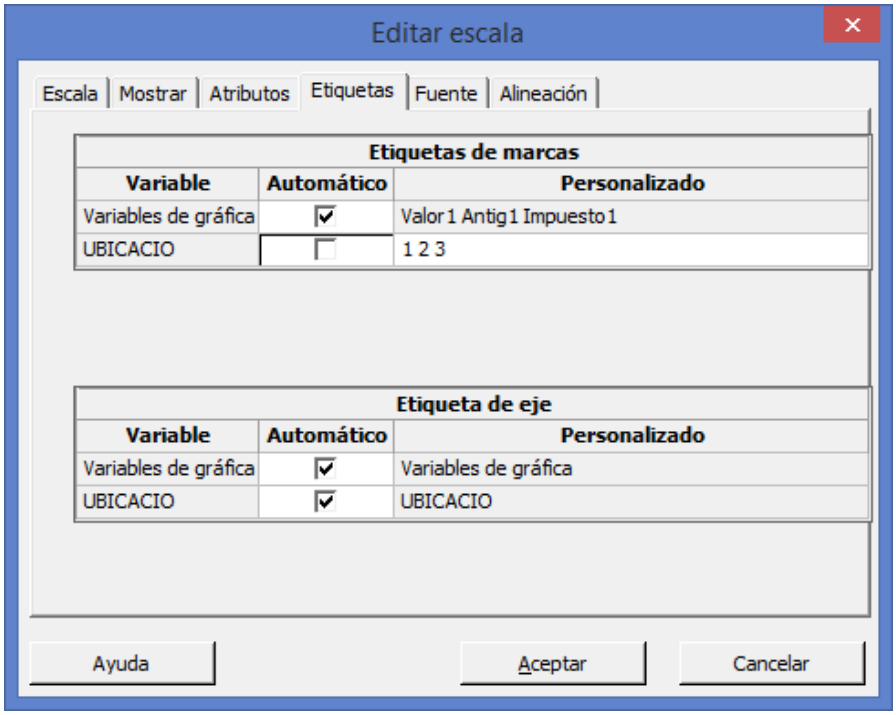
1 Haga doble clic en el número 1 en el renglón de UBICACIÓN para que Minitab despliegue la ventana ***Editar Escala.***



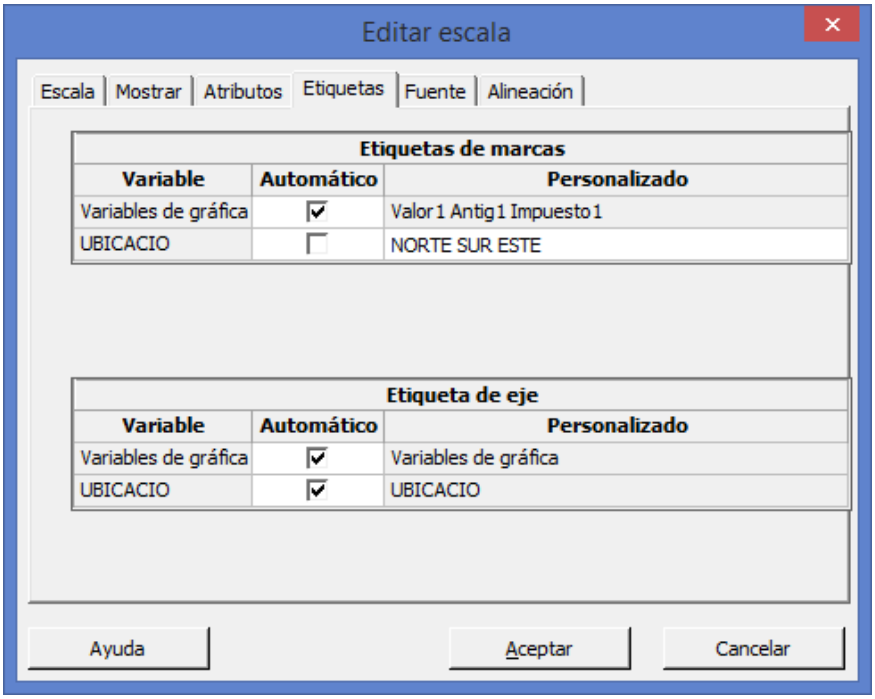
4 Seleccionar la pestaña **Etiquetas**



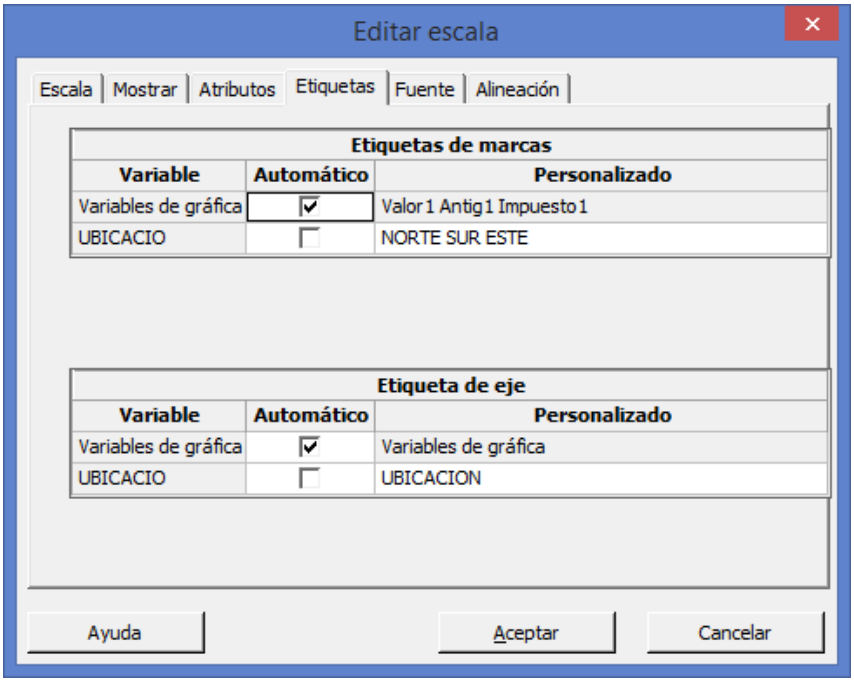
5 Desactivar la casilla **UBICACIÓN**



6 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE.**

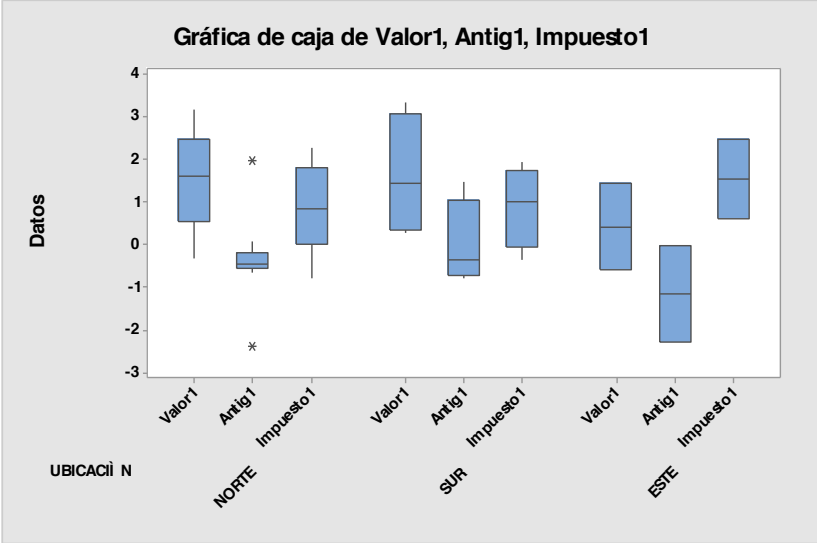


7 En **Etiqueta de eje** , si es el caso, desactivar la casilla de **UBICACIO** y escribir **UBICACIÓN.**

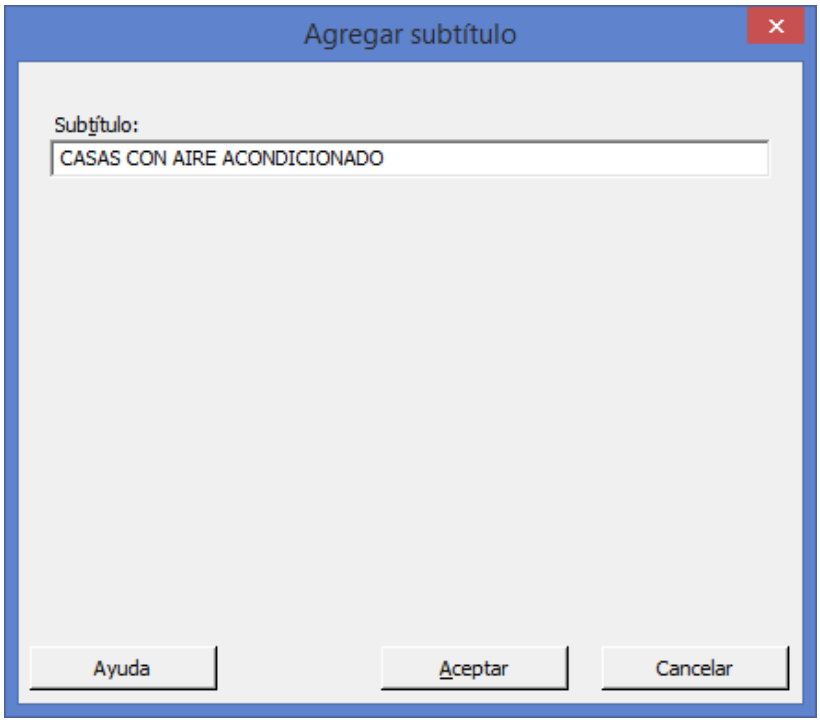


8 Haga clic en **Aceptar**.

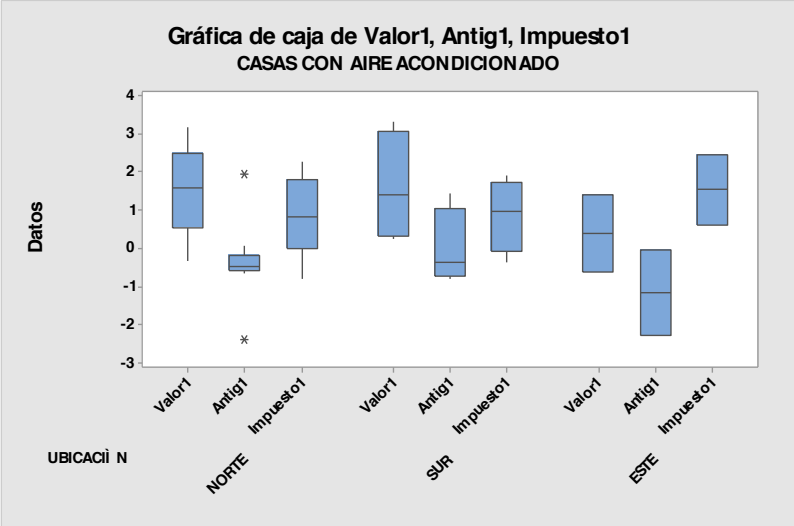
9 Aparece la gráfica debidamente etiquetada.



10 Coloque el mouse en el título de la grafica y presione el botón derecho del mouse. Seleccione la opción Agregar y la opción Subtitulo y se escribe **CASAS CON AIRE ACONDICIONADO**.



11 Haga clic en Aceptar.
12 Aparece la gráfica debidamente etiquetada.



1.3.2.1 EJERCICIO DE AUTOEVALUACIÓN

AUTOEVALUACIÓN 1.3.2.1

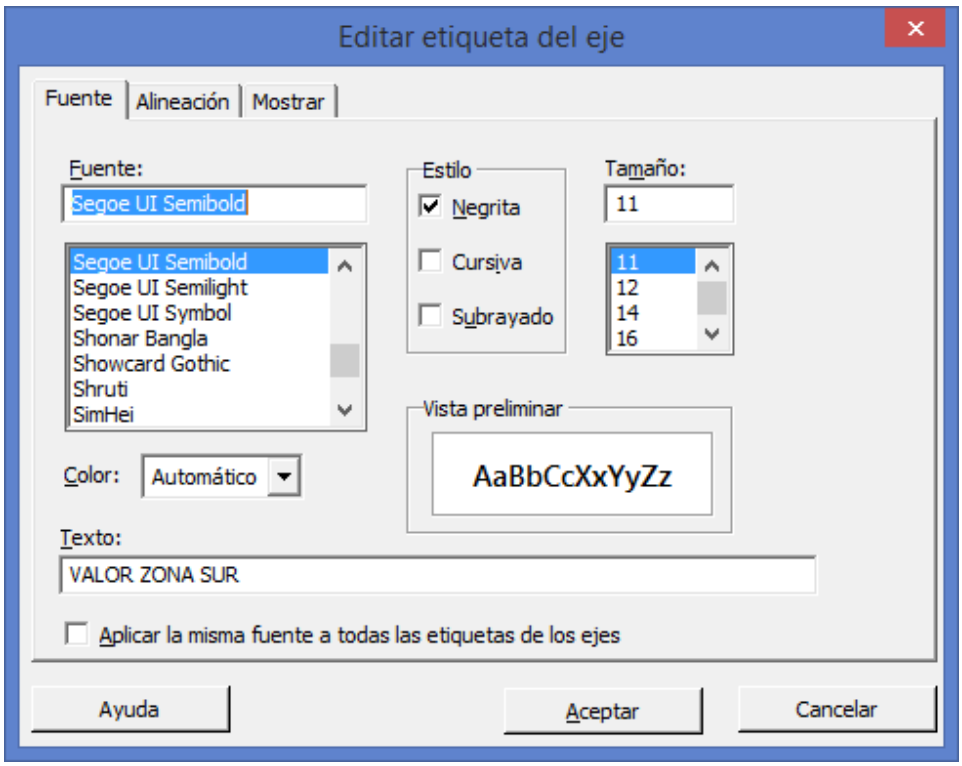
PRUEBA DE NORMALIDAD. PRUEBA DE ANDERSON DARLING

Se midió el **VALOR** de las casas en la zona **SUR** en la base de datos "**BIENES1"** que se construyó anteriormente.

Se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas ubicadas en el **SUR** e interprete los resultados.

Utilice el paquete Minitab para resolver esta Autoevaluación.

Nota: Etiquete la gráfica dando doble clic en **VALOR** para que aparezca la ventana ***Editar etiqueta de eje*** y en la parte inferior donde dice ***Texto*** escriba **VALOR ZONA SUR**. Haga clic en **Aceptar.**



1 Abra la base de datos que creó **"BIENES1"** y cree un subconjunto de hoja de trabajo para las casas ubicadas en el **SUR**

2 Elija **Estadísticas > Estadísticas básicas > Prueba de normalidad**.

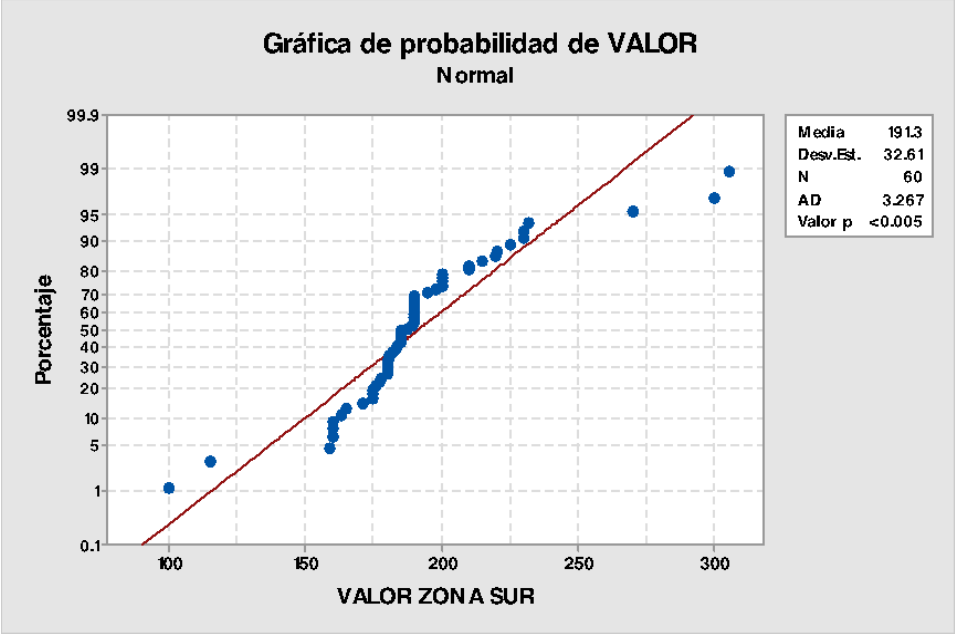
3 En **Variable**, ingrese ***Valor.*** Haga clic en **Aceptar**.

Salida de la ventana Gráfica

Genera una gráfica de probabilidad normal y realiza una prueba de hipótesis para examinar si las observaciones siguen o no una distribución normal . Para la prueba de normalidad , las hipótesis son,

H_0: los datos siguen una distribución normal vs. H_1: los datos no siguen una distribución normal

La escala vertical de la gráfica se asemeja a la escala vertical del papel de probabilidad normal. El eje horizontal es una escala lineal. La línea forma un estimado de la función de distribución acumulada para la población de la cual se extrajeron los datos. Con la gráfica se muestran estimaciones numéricas de los parámetros de la población , Media y Desviación Estándar, el tamaño de la muestra ó población, el valor de la prueba de normalidad (Estadístico de Anderson Darling) y el valor p asociado.



Interpretación de los resultados

La salida gráfica es una gráfica de probabilidades normales versus los datos. Los datos se alejan de la línea ajustada de una manera más evidente en los extremos, o colas de la distribución. El valor p de la prueba de Anderson-Darling indica que, en niveles p menores a 0.005, hay evidencia de que los datos no siguen una distribución normal.

Los problemas de no normalidad puede solucionarse por medio de **transformaciones de datos** y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una o varias de las variables por lo que también se debe verificar el supuesto de homocedasticidad y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

1.3.2.2 EJERCICIO DE AUTOEVALUACIÓN

AUTOEVALUACIÓN 1.3.2.2

PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE LEVENE. TRANSFORMACIONES. ETIQUETADO DE GRÁFICOS.

Se pretende diseñar un experimento para evaluar el valor de 233 casas de diferentes estilos. Se mide el **VALOR** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad con base en **si tienen o no CHIMENEA**. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de comparación múltiple apropiado para sus datos.

Antes de realizar el **análisis de varianza,** se verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**)**:**

Utilice el paquete Minitab para resolver este Ejercicio de Autoevaluación de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

PRUEBA DE LEVENE EN MINITAB 17

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***Valor***

4 En **Factores**, ingrese ***CHIMENEA.*** Haga clic en **Aceptar**.

Salida de la ventana Sesión

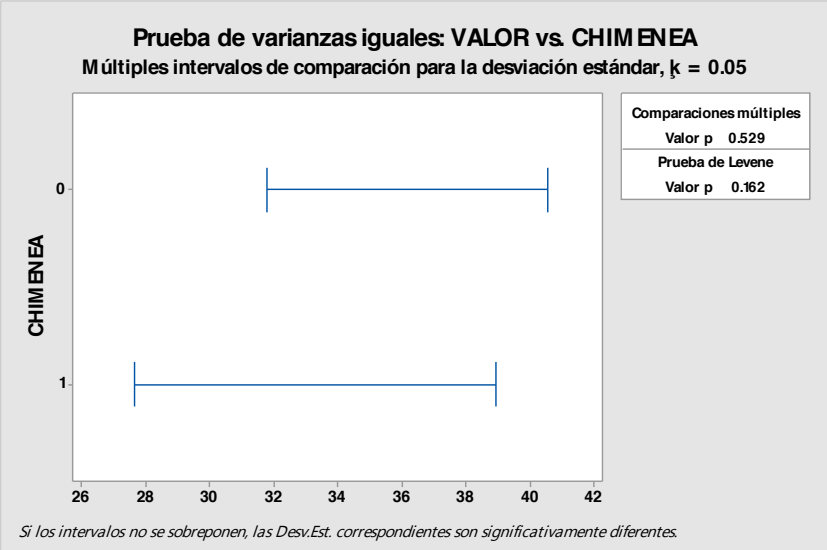
Prueba de varianzas iguales: VALOR vs. CHIMENEA

Método

```
Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA    N  Desv.Est.          IC

       0  154    35.6644  (29.9574, 43.0856)

       1   79    32.3843  (23.4647, 45.9994)

Nivel de confianza individual = 97.5%

Pruebas

                         Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.529

Levene                          1.96    0.162
```

Prueba de varianzas iguales: VALOR vs. CHIMENEA

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de**

Bonferroni de 95% para la desviación estándar de la respuesta en cada nivel. Los resultados de las pruebas **Levene** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **VALOR** de las casas con base en si tienen o no **CHIMENEA**, los **valores de p** de 0.162 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

TRANSFORMACIONES EN MINITAB 17

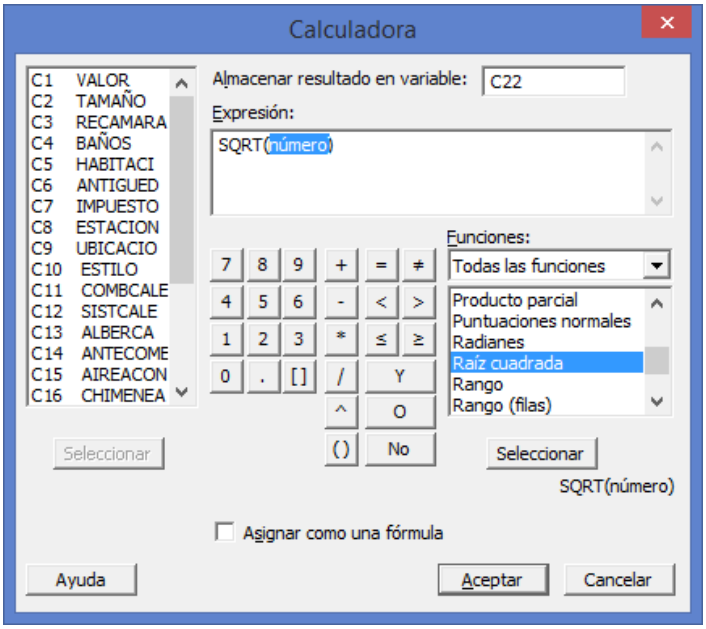
Transformaciónes Raiz Cuadrada, 1/X y Log de la variable VALOR con Minitab:

2 Abra la base de datos que creó **"BIENES1"**

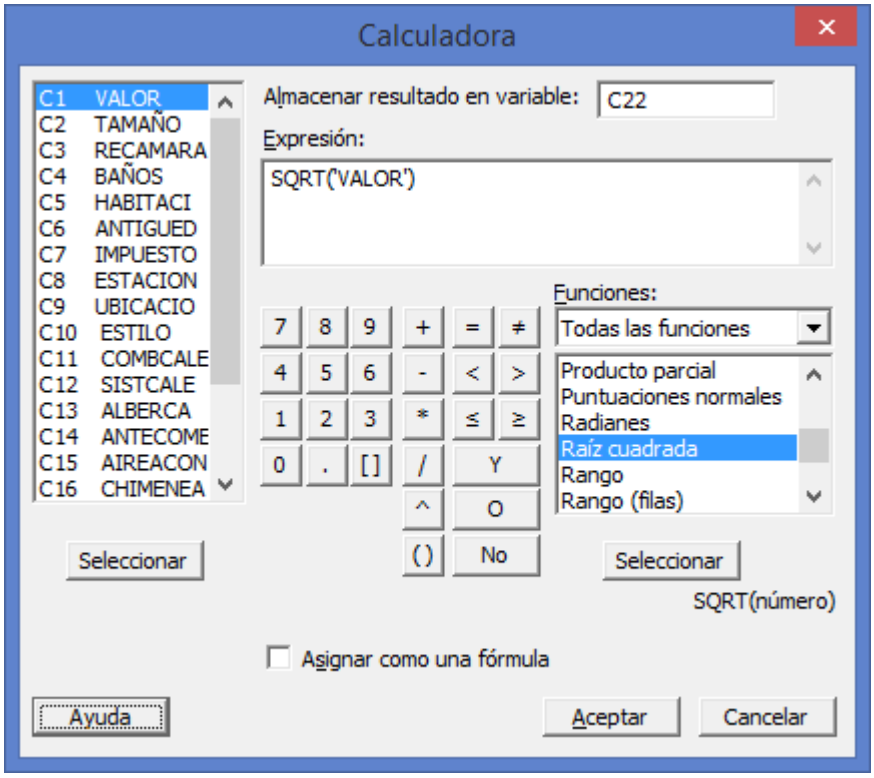
2 Elija **Calc > Calculadora**.

4 En **Almacenar resultado en variable**, ingrese *C22*

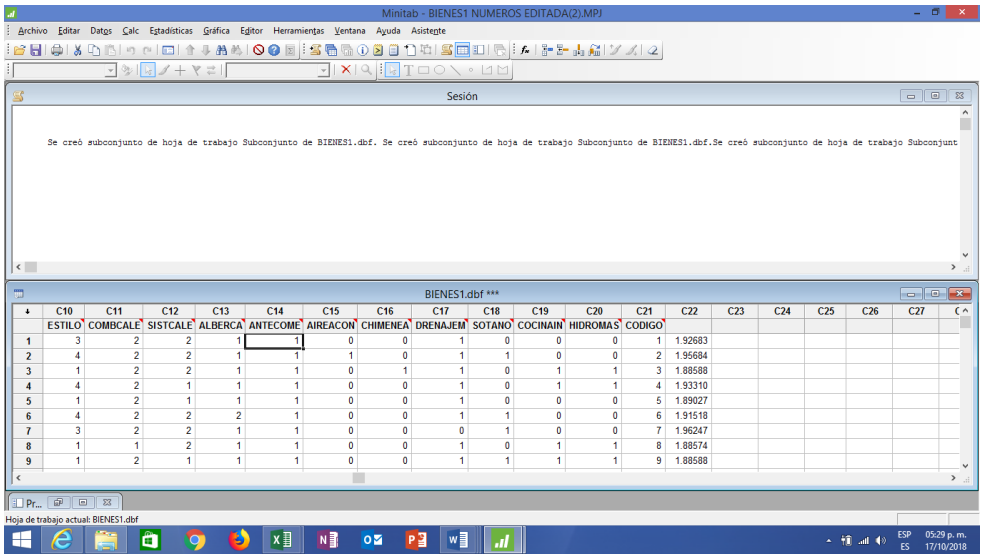
4 En **Expresión**, busque en **Funciones**: Raíz Cuadrada y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



5 En **Expresión** aparece **SQRT(número).** Seleccione la variable **VALOR** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.

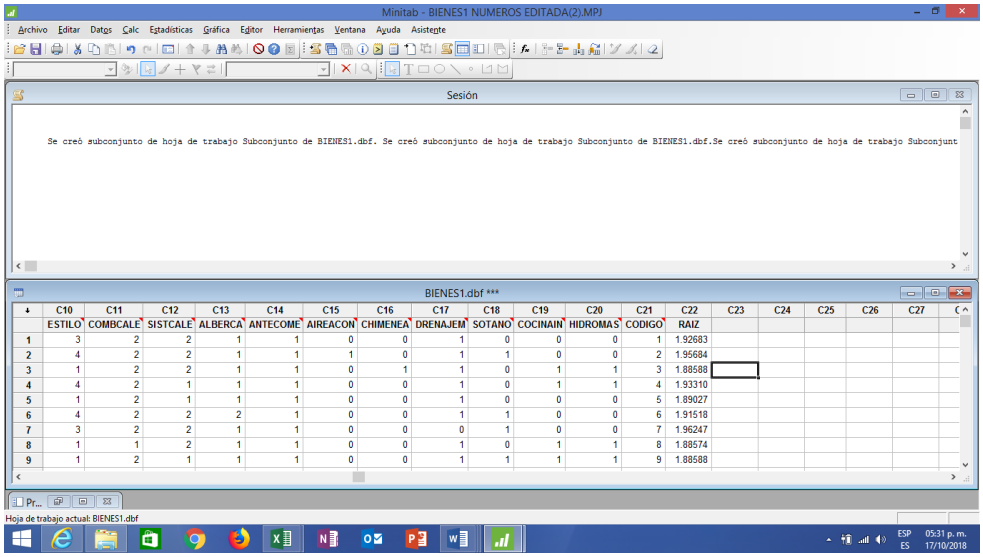


6 Haga clic en **Aceptar**.



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO						
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683					
2	4	2	2	1	1	1	0	1	1	0	0	2	1.95684					
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588					
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310					
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027					
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518					
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247					
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574					
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588					

7 En C22 teclee el nombre **RAÍZ**

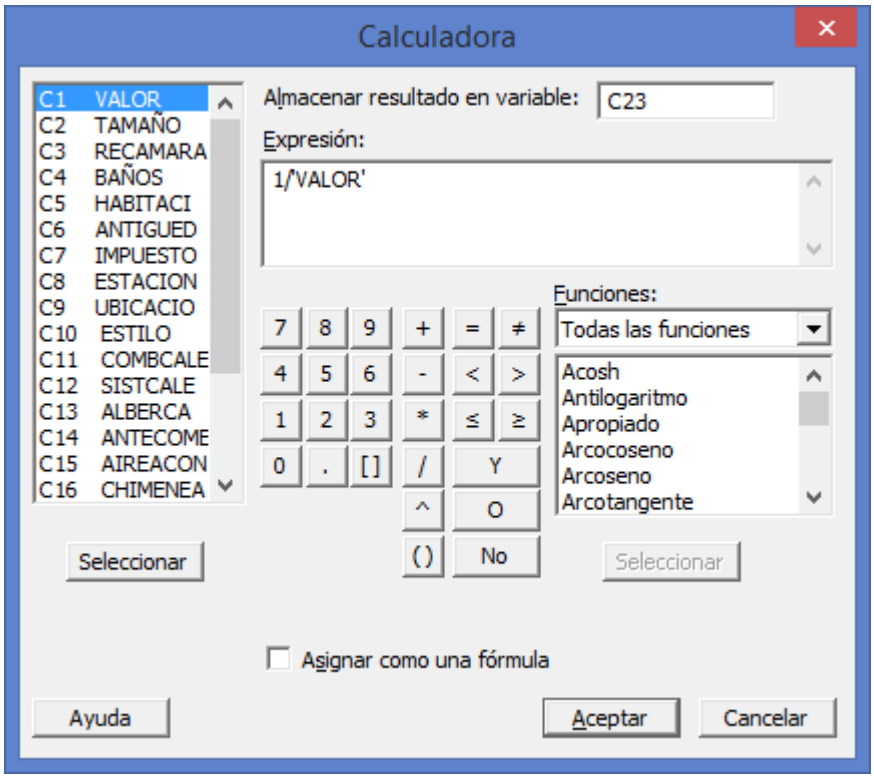


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAIZ					
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683					
2	4	2	2	1	1	1	0	1	1	0	0	2	1.95684					
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588					
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310					
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027					
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518					
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247					
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574					
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588					

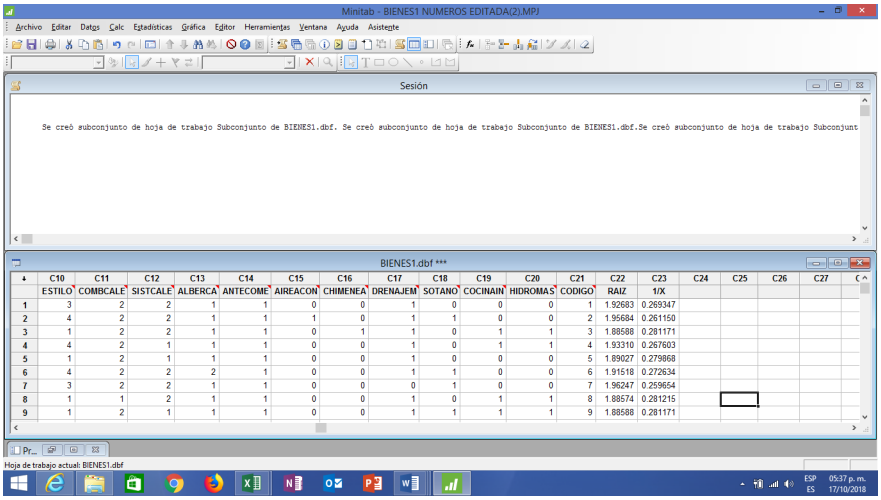
9 Elija **Calc > Calculadora**.

9 En **Almacenar resultado en variable**, ingrese *C23*

10 En **Expresión**, para obtener la Inversa, **Teclee 1/** y seleccione la variable **VALOR** del lado izquierdo o doble clic.



2 Haga clic en **Aceptar**. En **C23** escriba el nombre **1/X**

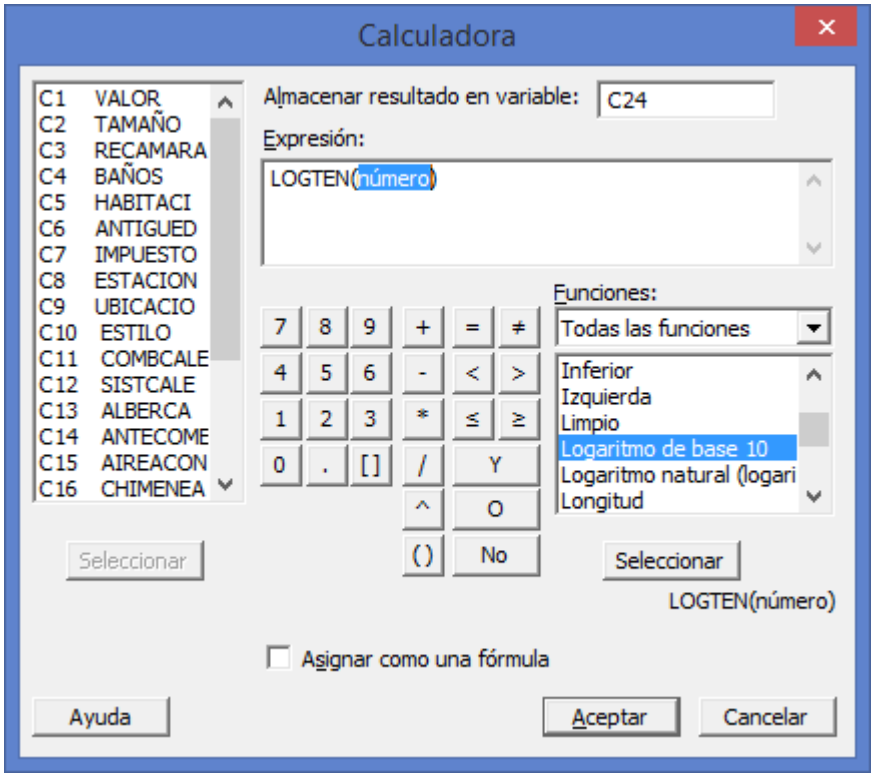


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAIZ	1/X				
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683	0.269347				
2	4	2	2	1	1	1	0	1	1	0	0	2	1.95684	0.261150				
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588	0.281171				
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310	0.267603				
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027	0.279868				
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518	0.272634				
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247	0.259654				
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574	0.281215				
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588	0.281171				

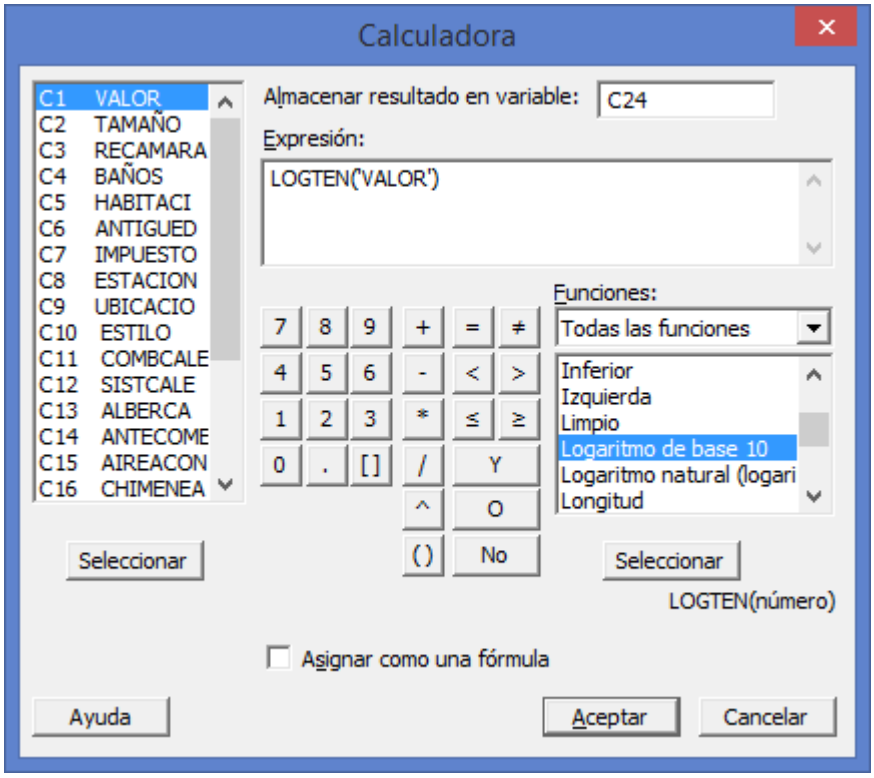
3 Elija **Calc > Calculadora**.

13 En **Almacenar resultado en variable**, ingrese *C24*

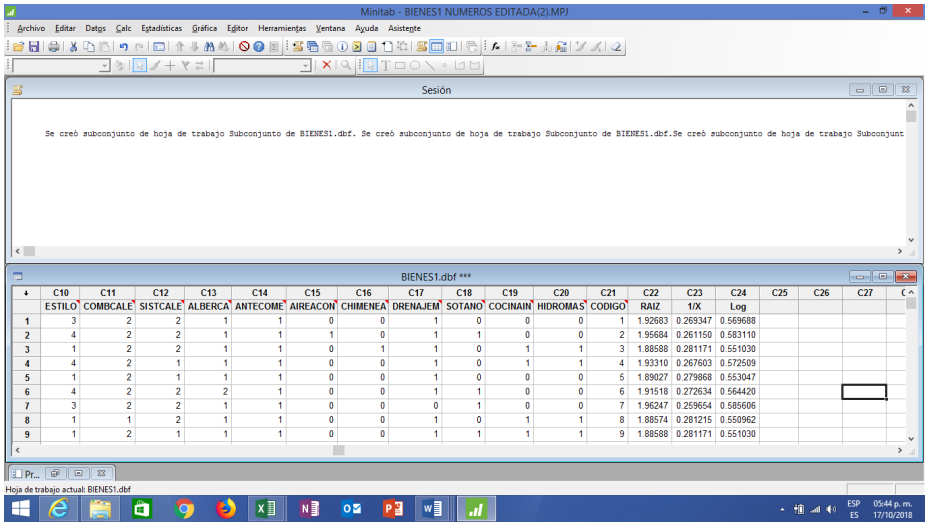
14 En **Expresión**, busque en **Funciones**: **Logaritmo base 10** y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



15 En **Expresión** aparece **LOGTEN(número).** Seleccione la variable **VALOR** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.



16 Haga clic en **Aceptar.** En **C24** escriba el nombre **Log**



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAIZ	1/X	Log			
1	3	2	2	1	1	0	0	1	0	0	0	1	1.92683	0.269347	0.569688			
2	4	2	2	1	1	1	0	1	1	0	0	2	1.95684	0.261150	0.583110			
3	1	2	2	1	1	0	1	1	0	1	1	3	1.88588	0.281171	0.551030			
4	4	2	1	1	1	0	0	1	0	1	1	4	1.93310	0.267603	0.572509			
5	1	2	1	1	1	0	0	1	0	0	0	5	1.89027	0.279868	0.553047			
6	4	2	2	2	1	0	0	1	1	0	0	6	1.91518	0.272634	0.564420			
7	3	2	2	1	1	0	0	0	1	0	0	7	1.96247	0.259654	0.585606			
8	1	1	2	1	1	0	0	1	0	1	1	8	1.88574	0.281215	0.550962			
9	1	2	1	1	1	0	0	1	1	1	1	9	1.88588	0.281171	0.551030			

Repita la prueba de Levene para las TRES transformaciones para ver si se corrigió el problema de homocedasticidad y si fué así cual transformación resulto mejor.

A. CON LA RAÍZ CUADRADA DE VALOR

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***RAÍZ***

4 En **Factores**, ingrese ***CHIMENEA.*** Haga clic en **Aceptar**.

Salida de la ventana Sesión

Prueba de varianzas iguales: RAIZ vs. CHIMENEA

Método

Hipótesis nula	Todas las varianzas son iguales
Hipótesis alterna	Por lo menos una varianza es diferente
Nivel de significancia	α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA	N	Desv.Est.	IC
0	154	0.0428682	(0.0365610, 0.0510059)
1	79	0.0370621	(0.0280509, 0.0503980)

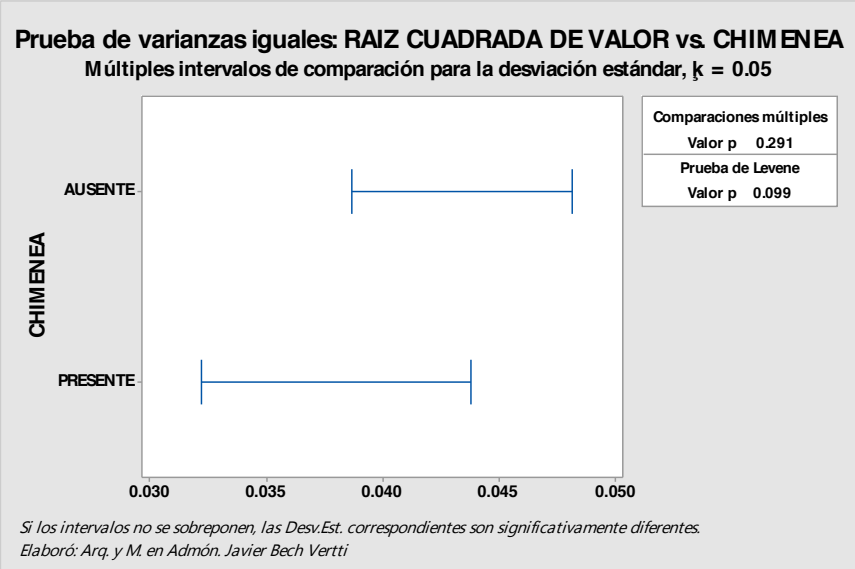
Nivel de confianza individual = 97.5%

Pruebas

Método	Estadística de prueba	Valor p
Comparaciones múltiples	—	0.291
Levene	2.74	0.099

Prueba de varianzas iguales: RAIZ vs. CHIMENEA

Salida de la ventana Gráfica



Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **RAIZ CUADRADA** del **VALOR** de las casas con o sin **CHIMENEA** de las mismas, los **valores de p** de 0.099 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

B. CON LA LA INVERSA DE VALOR (1/X)

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***1/X***

4 En **Factores**, ingrese ***CHIMENEA***. Haga clic en **Aceptar**.

Salida de la ventana Sesión

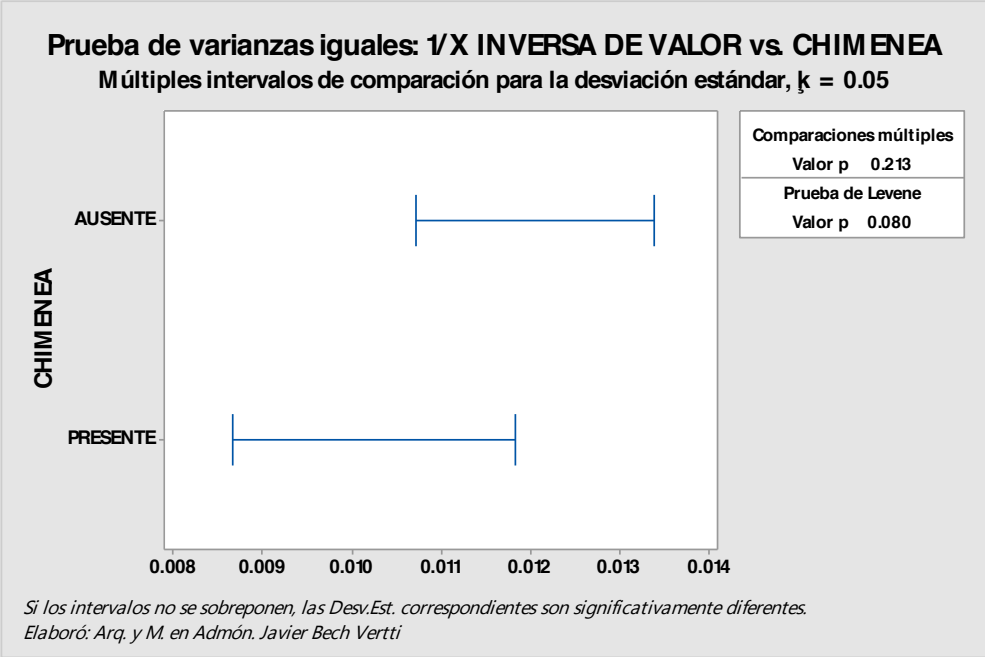
Prueba de varianzas iguales: 1/X vs. CHIMENEA

```
Método

Hipótesis nula          Todas las varianzas son iguales
```

```
Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA    N  Desv.Est.            IC

       0  154  0.0118928  (0.0100825, 0.0142353)

       1   79  0.0099990  (0.0077039, 0.0133569)

Nivel de confianza individual = 97.5%

Pruebas

                          Estadística

Método                      de prueba  Valor p

Comparaciones múltiples            —    0.213

Levene                          3.09    0.080
```

Prueba de varianzas iguales: 1/X vs. CHIMENEA

Salida de la ventana Gráfica



Si los intervalos no se sobreponen, las Desv.Est. correspondientes son significativamente diferentes.
Elaboró: Arq. y M. en Admón. Javier Bech Vertti

Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **INVERSA** del **VALOR** de las casas con o sin

CHIMENEA de las mismas, los **valores de p** de 0.080 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

C. CON EL LOGARITMO BASE 10 DE VALOR

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***Log***

4 En **Factores**, ingrese ***CHIMENEA***. Haga clic en **Aceptar**.

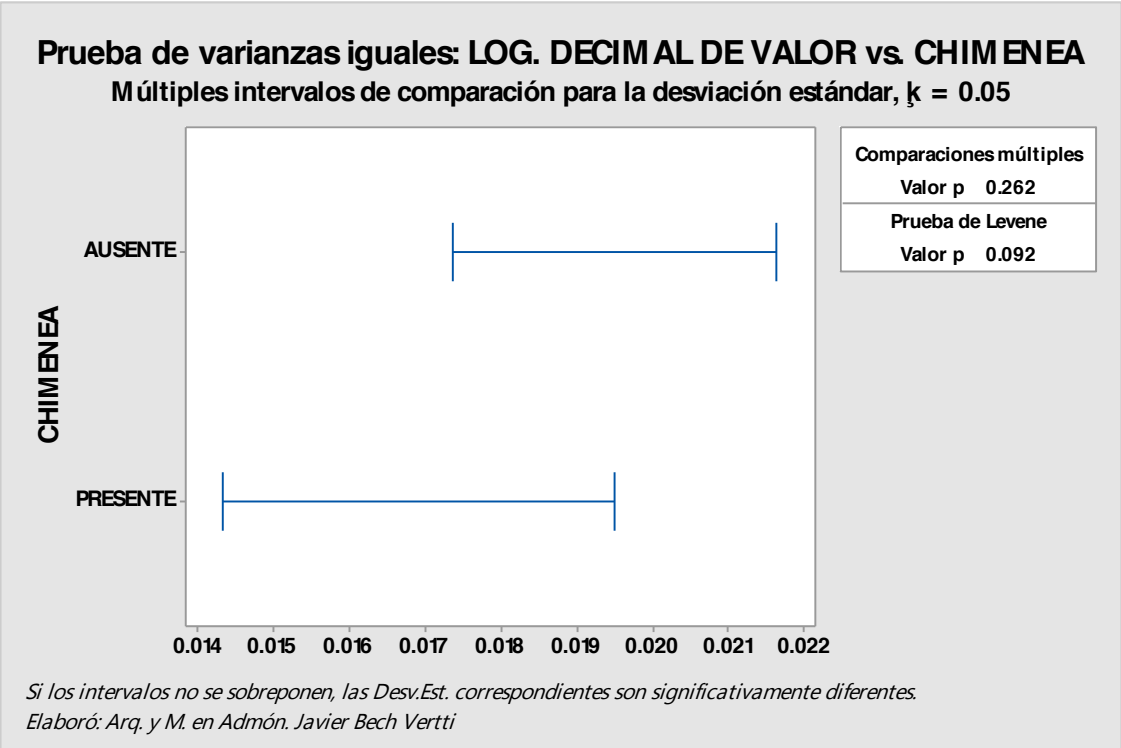
Salida de la ventana Sesión

Prueba de varianzas iguales: Log vs. CHIMENEA

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA    N  Desv.Est.            IC

       0  154  0.0192542  (0.0164062, 0.0229303)

       1   79  0.0165010  (0.0125643, 0.0223039)

Nivel de confianza individual = 97.5%


Pruebas

                          Estadística

Método                      de prueba  Valor p

Comparaciones múltiples            —    0.262

Levene                          2.86    0.092
```

Prueba de varianzas iguales: Log vs. CHIMENEA

Salida de la ventana Gráfica



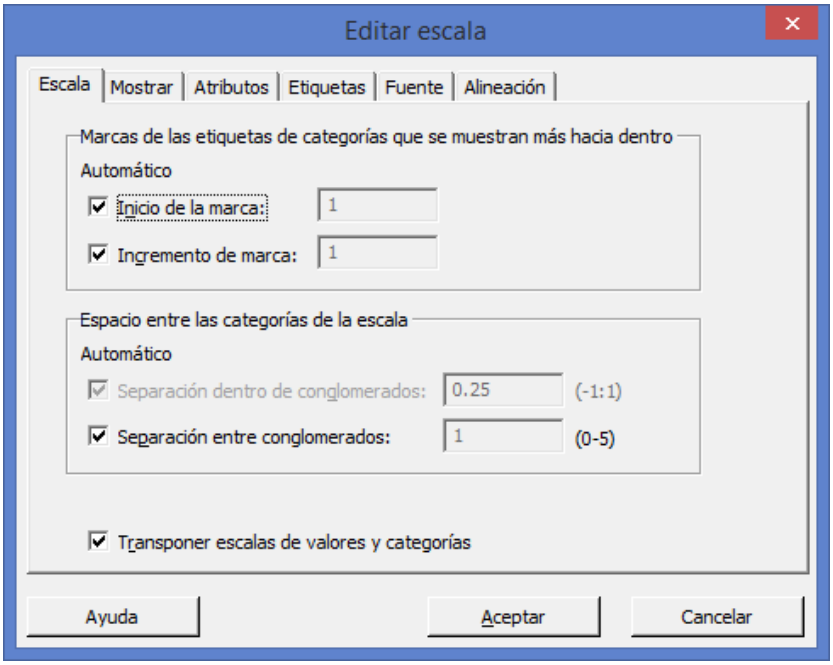
Interpretación de los resultados

Para el ejemplo para evaluar la transformación del **LOGARITMO DECIMAL** del **VALOR** de las casas con o sin **CHIMENEA** de las mismas, los **valores de p** de 0.092 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

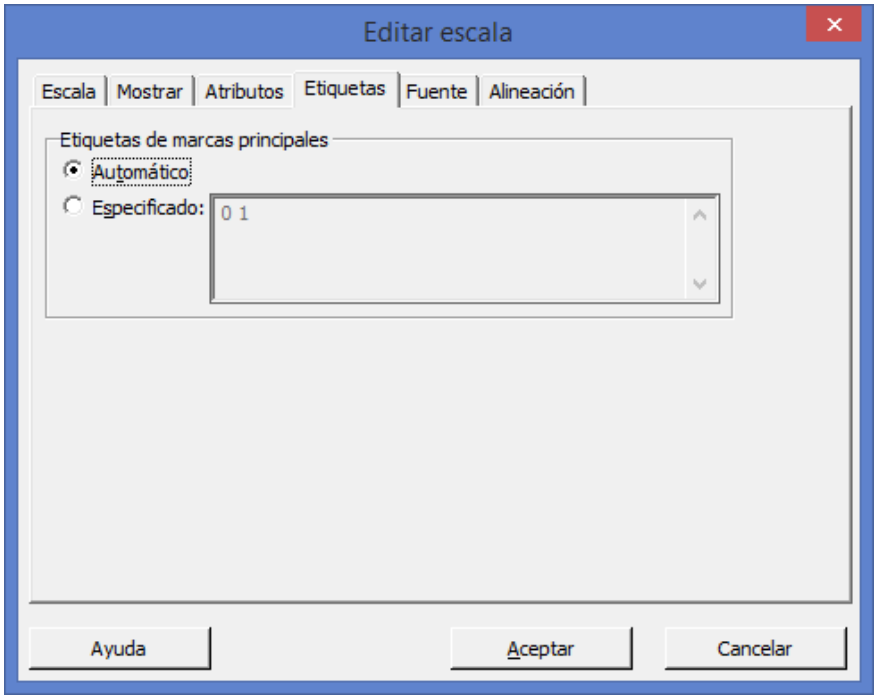
Nota Importante: No es del alcance de este trabajo tomar la decisión de si se debe o no trabajar con los datos transformados o sin transformar para usar alguna técnica multivariada sino solamente mostrar la forma de hacer dichas transformaciones.

ETIQUETADO DE LOS NIVELES DE LA VARIABLE O "FACTOR" CHIMENEA EN MINITAB 17:

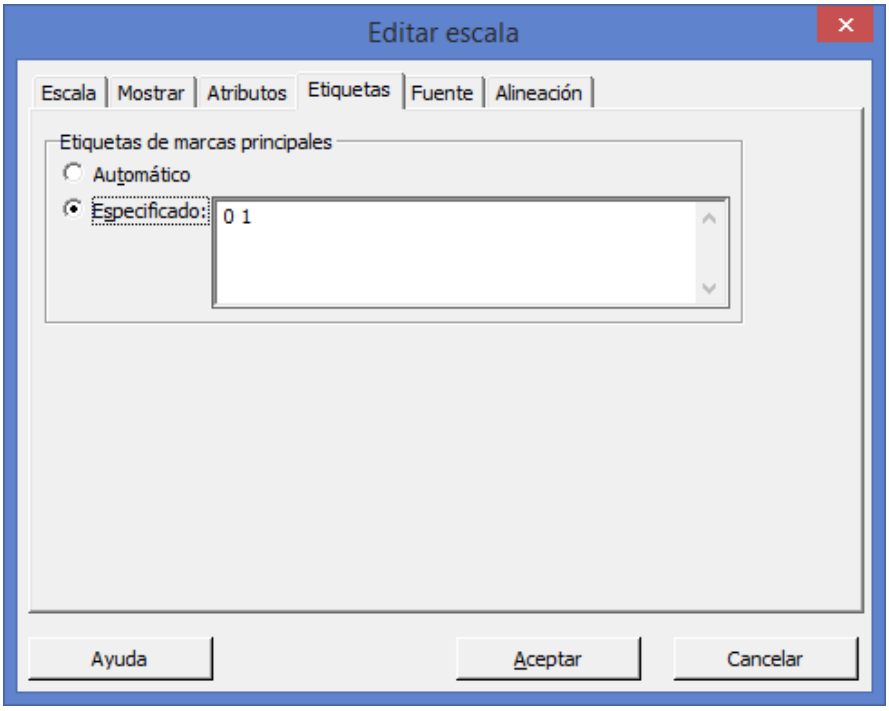
1 Haga doble clic en el número 0 del eje vertical de la gráfica que dice **CHIMENEA** para que Minitab despliegue la ventana ***Editar Escala.***



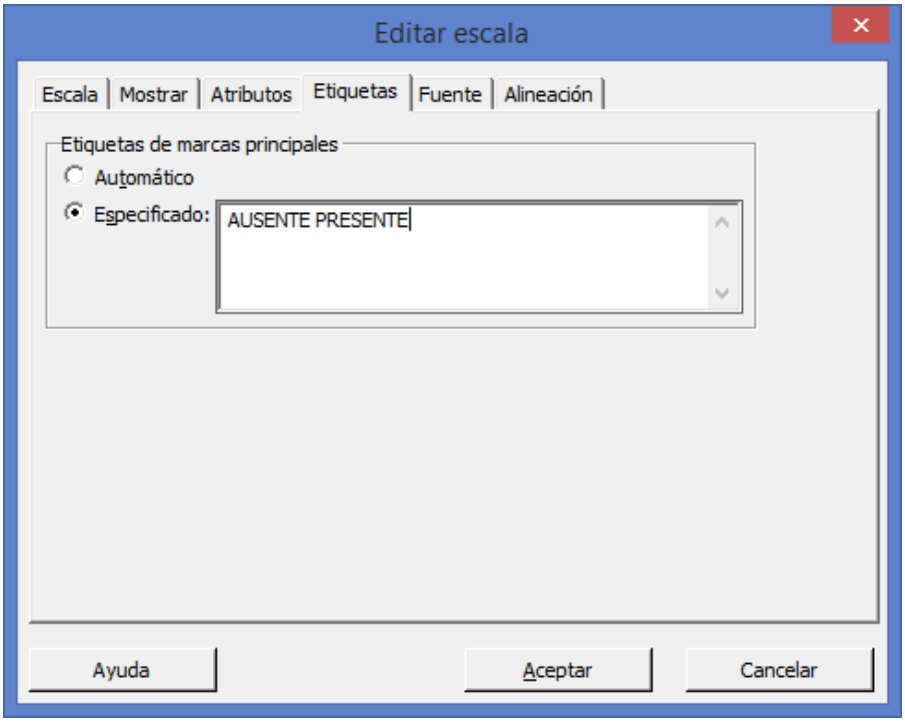
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

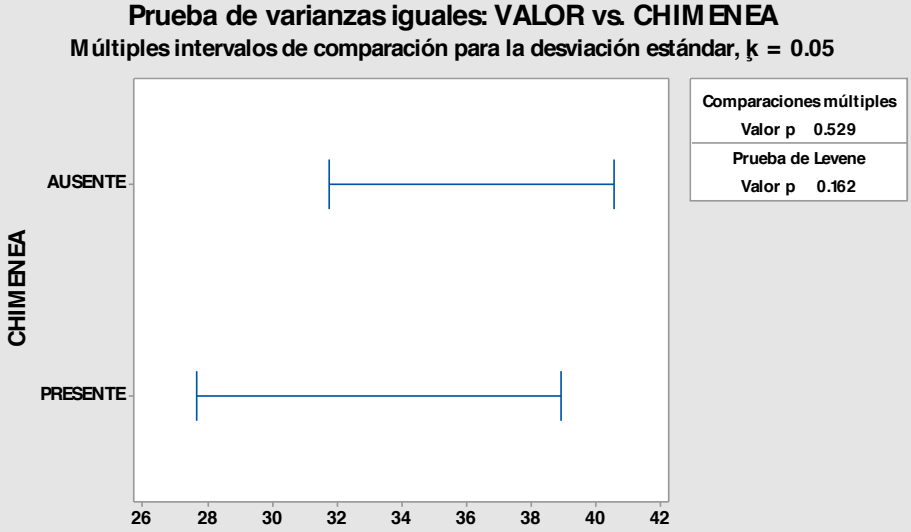


4 Borrar los números 0 1 y escribir **AUSENTE PRESENTE**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



Si los intervalos no se sobreponen, las Desv.Est. correspondientes son significativamente diferentes.
Elaboró: Arq. y M. en Admón. Javier Bech Vertti

1.3.2.3 EJERCICIO DE AUTOEVALUACIÓN

AUTOEVALUACIÓN 1.3.2.3

PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT

Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **VALOR** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con respecto al tipo de **ALBERCA** ubicada en la propiedad. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas.

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***VALOR***

4 En **Factores**, ingrese ***ALBERCA***

5 En el botón **Opciones** active la casilla que dice: ***Utilice la prueba basándose en la distribución normal.*** Haga clic en **Aceptar** y nuevamente **Aceptar**

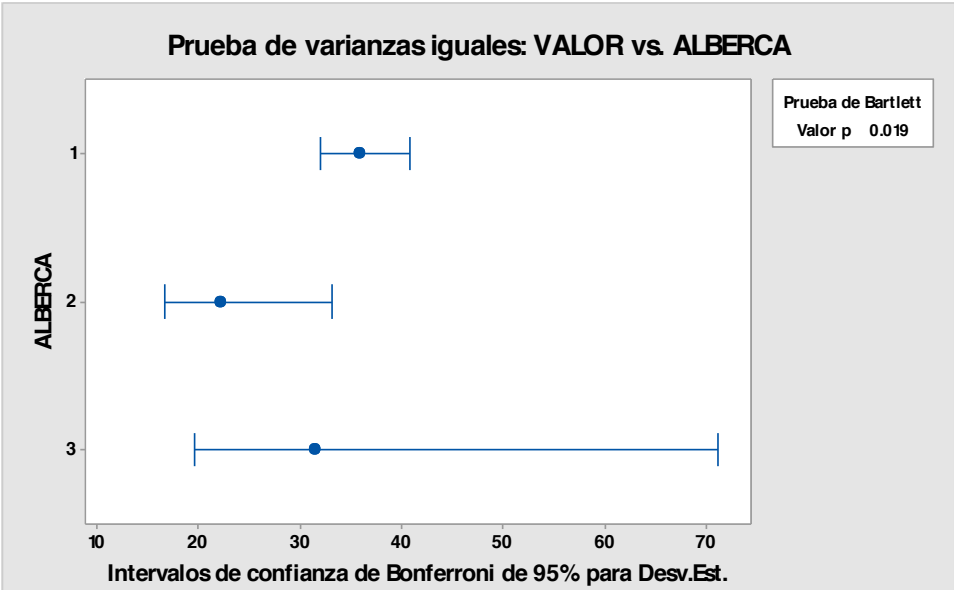
Salida de la ventana Sesión

Prueba de varianzas iguales: VALOR vs. ALBERCA

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Se utiliza el método de Bartlett. Este método es exacto sólo para datos

     normales.

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

ALBERCA    N  Desv.Est.          IC

      1  198    35.8031  (31.9324, 40.6731)

      2   26    22.1891  (16.5363, 33.0845)

      3    9    31.3925  (19.5699, 71.0908)

Nivel de confianza individual = 98.3333%

Pruebas


          Estadística

Método      de prueba  Valor p

Bartlett         7.88    0.019
```

Prueba de varianzas iguales: VALOR vs. ALBERCA

Salida de la ventana Gráfica



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

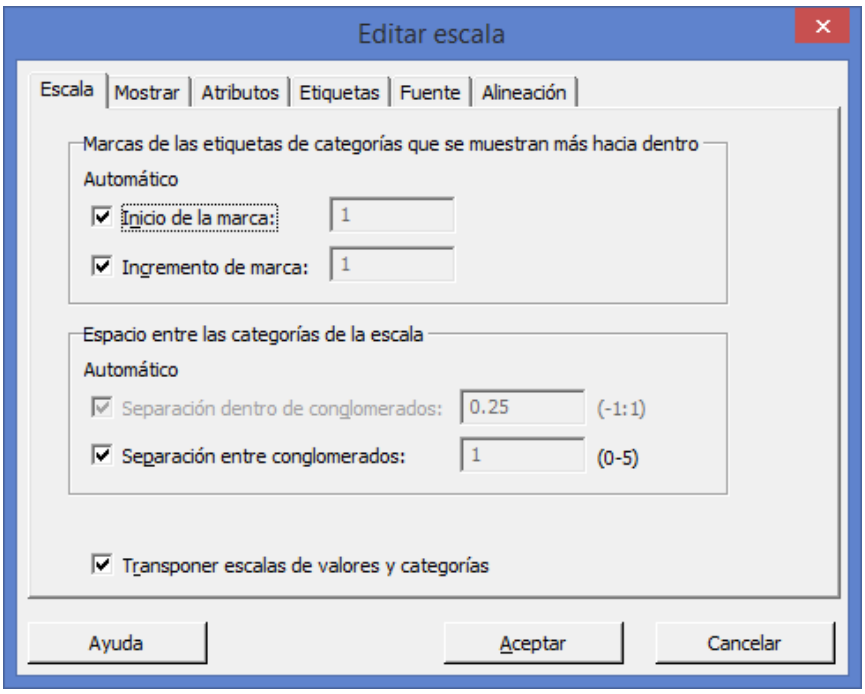
Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de la prueba de **Barlett** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

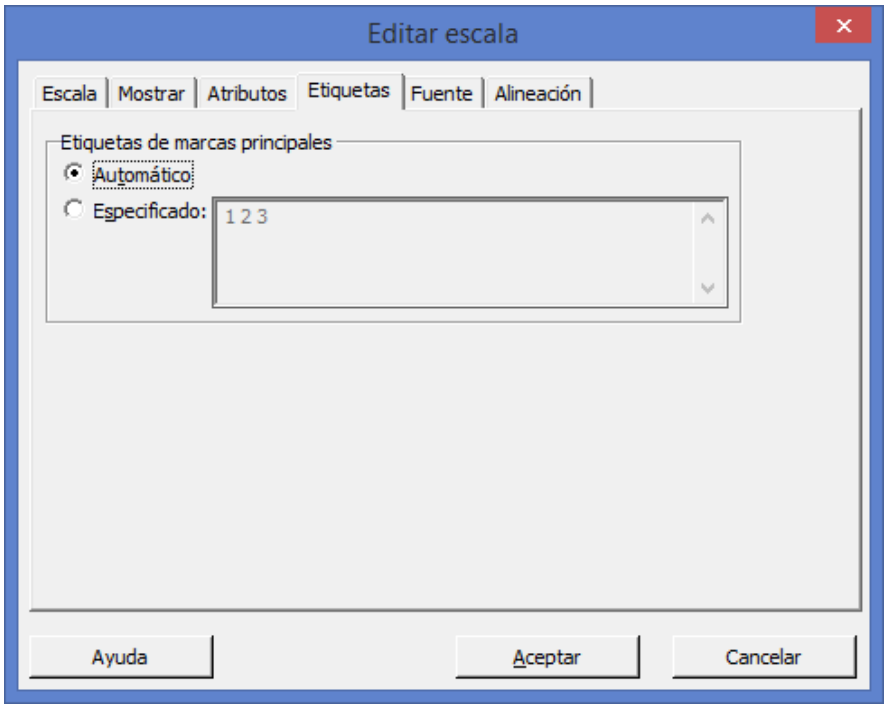
Para el ejemplo para evaluar el **VALOR** de las casas con base en el tipo de **ALBERCA** ubicada en la propiedad de las mismas, los **valores de p** de 0.019 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

Etiquetado de los niveles de la variable o "Factor" ALBERCA:

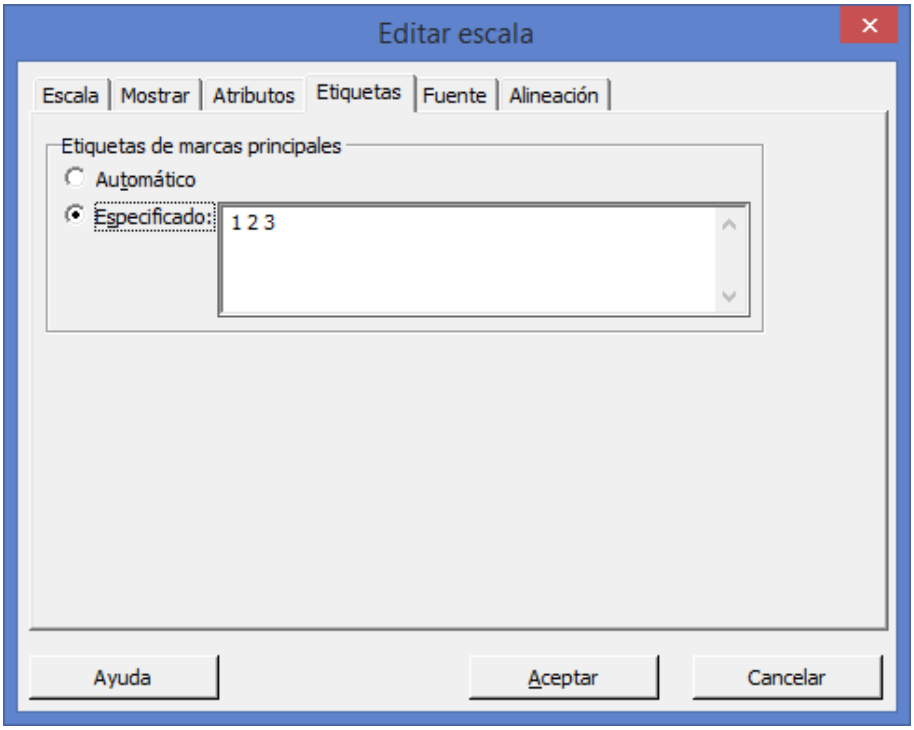
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **ALBERCA** para que Minitab despliegue la ventana ***Editar Escala.***



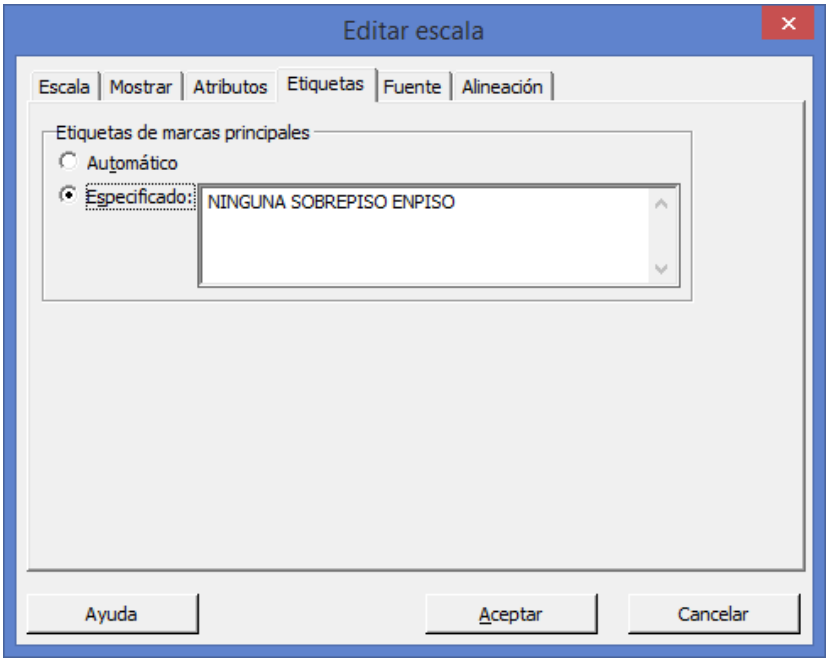
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

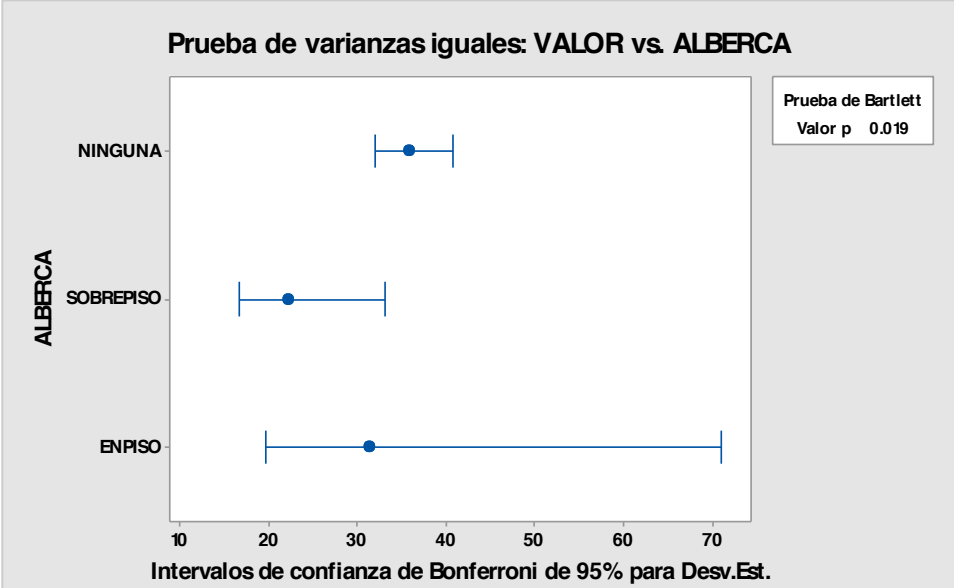


4 Borrar los números 1 2 3 y escribir **NINGUNA SOBREPISO ENPISO**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

1.4.1.1 EJERCICIO DE AUTOEVALUACIÓN

AUTOEVALUACIÓN 1.4.1.1

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**

Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$4x + 2y + 3z = 4$$

$$-3x - y - 2z = -6$$

$$-2x - y - z = -5$$

Paso 1. Expresar el sistema como A x X = K

$$\underbrace{\begin{bmatrix}-4 & 2 & 3\\ -3 & -1 & -2\\ -2 & -1 & -1\end{bmatrix}}_{A} \quad x \quad \underbrace{\begin{bmatrix}x\\ y\\ z\end{bmatrix}}_{X} = \quad \underbrace{\begin{bmatrix}4\\ -6\\ -5\end{bmatrix}}_{K}$$

Paso 2. Buscar el valor del determinante $|A|$=

Aplicando la regla de Sarrus:

$$|A| = \begin{vmatrix}4 & 2 & 3\\ -3 & -1 & -2\\ -2 & -1 & -1\\ 4 & 2 & 3\\ -3 & -1 & -2\end{vmatrix} =$$ (4)(-1)(-1)+(-3)(-1)(3)+(-2)(2)(-2)-(2)(-1)(3)-(4)(-1)(-2)-(-3)(2)(-1)=1≠0 (Tiene Inversa)

Paso 3. Obtener la matriz transpuesta de la matriz A, que es:

$$A^{T}=A'=\begin{bmatrix}4 & -3 & -2\\ 2 & -1 & -1\\ 3 & -2 & -1\end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A′, es decir formar α A′

$$\alpha A'=\begin{bmatrix}+\begin{vmatrix}-1 & -1\\ -2 & -1\end{vmatrix} & -\begin{vmatrix}2 & -1\\ 3 & -1\end{vmatrix} & +\begin{vmatrix}2 & -1\\ 3 & -2\end{vmatrix}\\ -\begin{vmatrix}-3 & -2\\ -2 & -1\end{vmatrix} & +\begin{vmatrix}4 & -2\\ 3 & -1\end{vmatrix} & -\begin{vmatrix}4 & -3\\ 3 & -2\end{vmatrix}\\ +\begin{vmatrix}-3 & -2\\ -1 & -1\end{vmatrix} & -\begin{vmatrix}4 & -2\\ 2 & -1\end{vmatrix} & +\begin{vmatrix}4 & -3\\ 2 & -1\end{vmatrix}\end{bmatrix}$$

$$=\begin{bmatrix}+(1-2) & -(-2+3) & +(-4+3)\\ -(3-4) & +(-4+6) & -(-8+9)\\ +((3-2) & -(-4+4) & +(-4+6)\end{bmatrix}=\begin{bmatrix}-1 & -1 & -1\\ +1 & +2 & -1\\ +1 & 0 & +2\end{bmatrix}$$

Paso 5. Calcular la inversa de la matriz A´= $\frac{1}{|A|}$ α A´

$$\frac{1}{1}\begin{bmatrix}-1 & -1 & -1\\ +1 & +2 & -1\\ +1 & 0 & +2\end{bmatrix}=\begin{bmatrix}-1 & -1 & -1\\ +1 & +2 & -1\\ +1 & 0 & +2\end{bmatrix}$$

Paso 6. Llevar este valor a la expresión X=A^{-1} x K y tenemos:

$$\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\begin{bmatrix}-1 & -1 & -1\\ +1 & +2 & -1\\ +1 & 0 & +2\end{bmatrix}x\begin{bmatrix}4\\ -6\\ -5\end{bmatrix}=$$

$$\begin{bmatrix}(-1 & -1 & -1) & * & \begin{bmatrix}4\\ -6\\ -5\end{bmatrix}\\ (+1 & +2 & -1) & * & \begin{bmatrix}4\\ -6\\ -5\end{bmatrix}\\ (+1 & 0 & +2) & * & \begin{bmatrix}4\\ -6\\ -5\end{bmatrix}\end{bmatrix}=\begin{bmatrix}-4 & +6 & +5\\ +4 & -12 & +5\\ +4 & 0 & -10\end{bmatrix}$$

$$=\begin{bmatrix}7\\ -3\\ -6\end{bmatrix}=\begin{bmatrix}x\\ y\\ z\end{bmatrix}$$

Paso 7. La solución es x=7, y=-3, z=-6

1.4.1.2 EJERCICIO DE AUTOEVALUACIÓN

AUTOEVALUACIÓN EN EXCEL1.4.1.2

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**

Resolver en **Excel** el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$x + 2y - 3z = 5$$

$$3x - 22y + 6z = 4$$

$$7x - 6y - 3z = 15$$

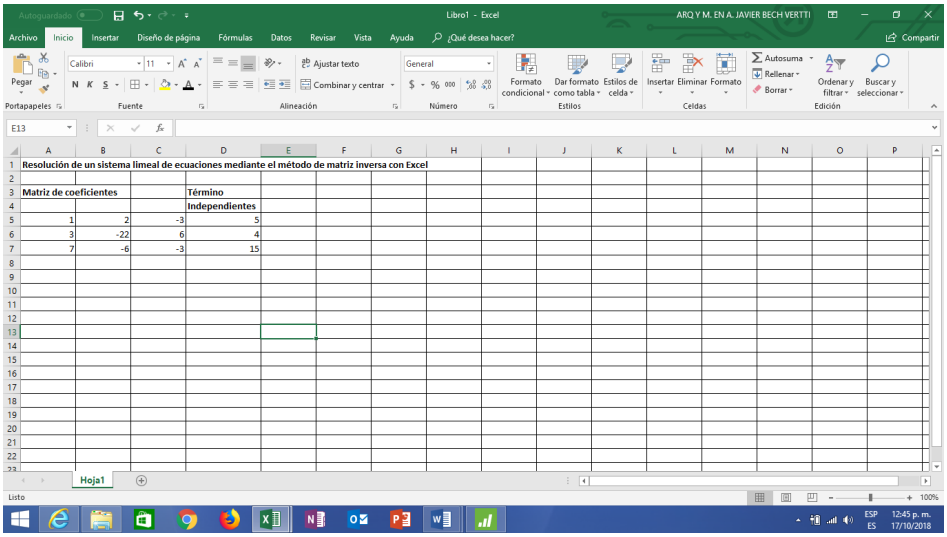
Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).

- Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.
- Introduce los valores independientes en otro rango al lado.

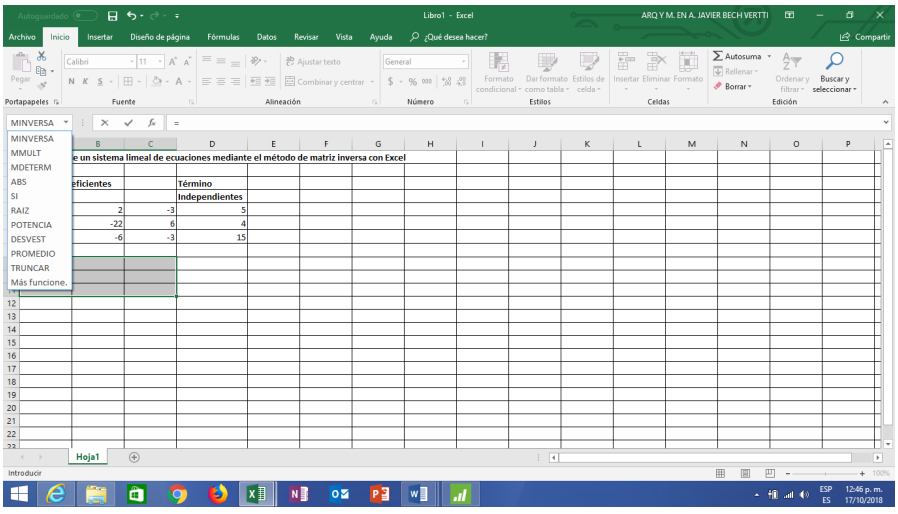
Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

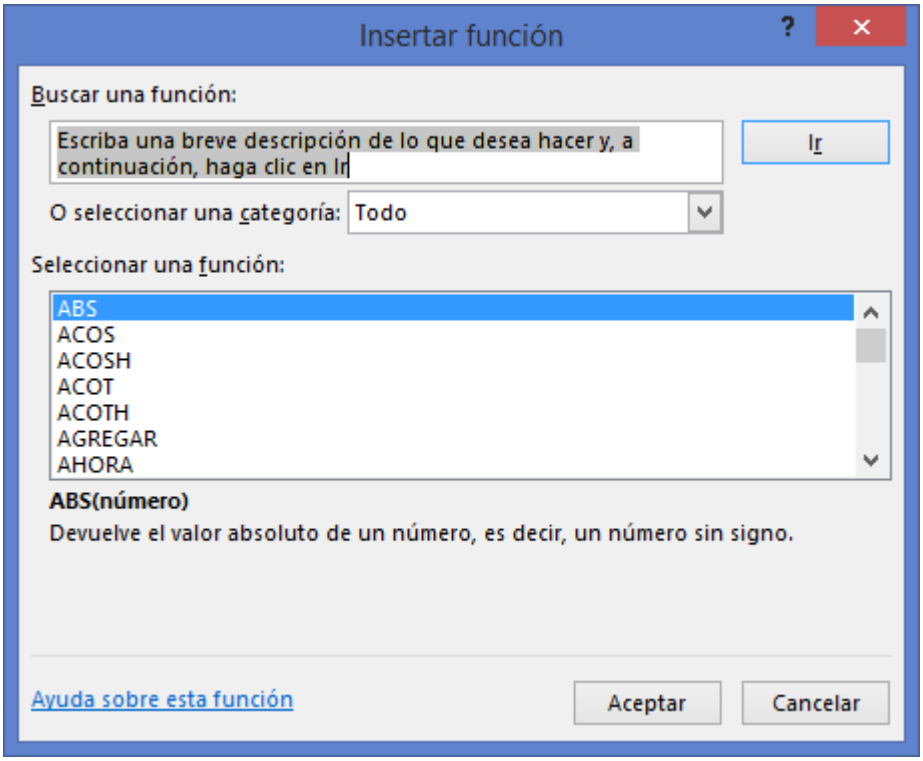
- Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).
- Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.
- Introduce los valores independientes en otro rango al lado.



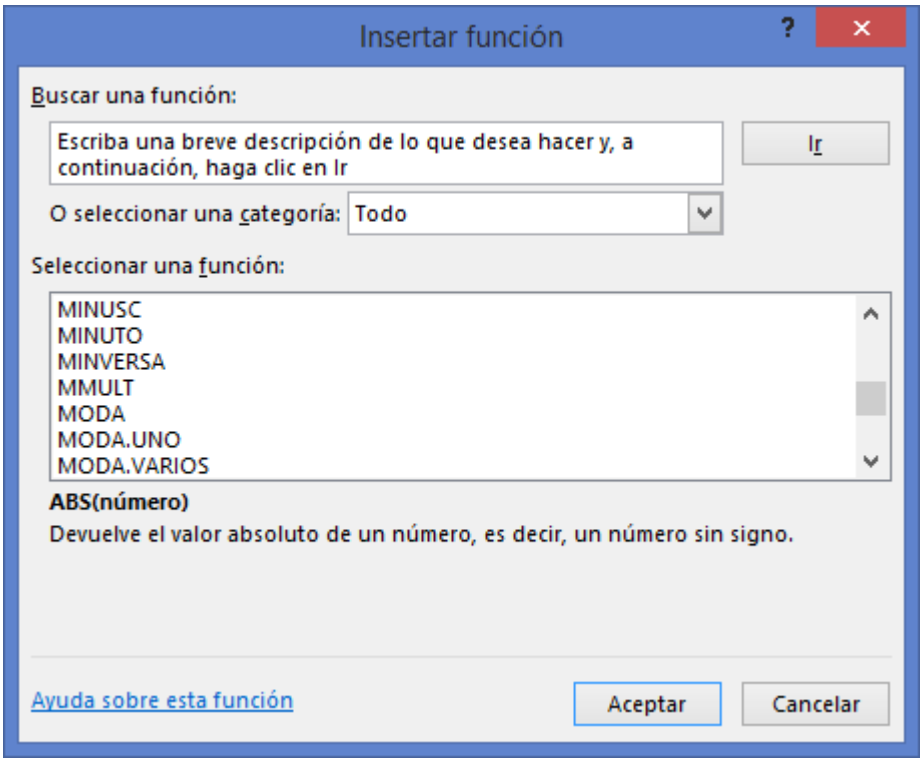
. Utiliza la función matricial **MINVERSA ()** para calcular la matriz inversa de los coeficientes introducidos. (Recuerda que para introducir la función debes pulsar Ctrl + Mayús + Intro).

Como la matriz de coeficientes es de 3 x 3 se necesitas 9 celdas para mostrar los resultados de la matris Inversa para lo cual selecciona con el mouse 9 celdas debajo de la matriz de coeficientes dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MINVERSA** () seleccionando la opción **Más funciones** en la parte inferior

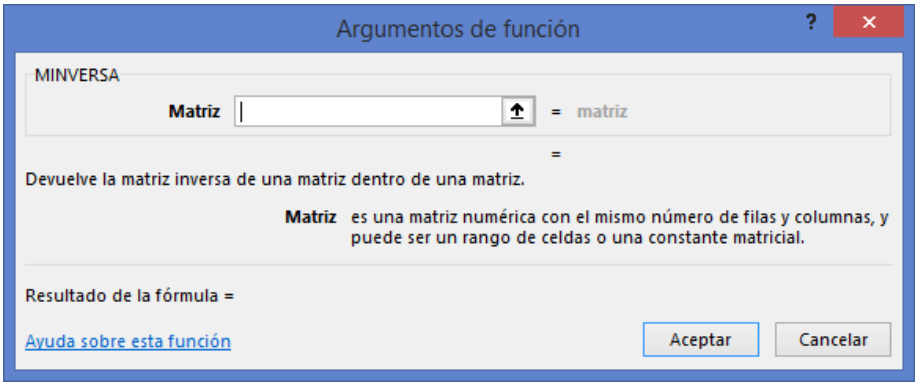




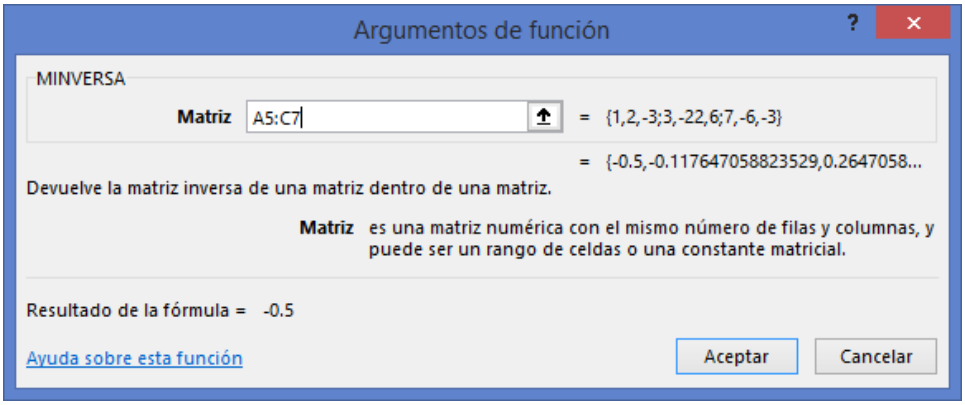
En Seleccionar una función busca **MINVERSA** y da un clic.



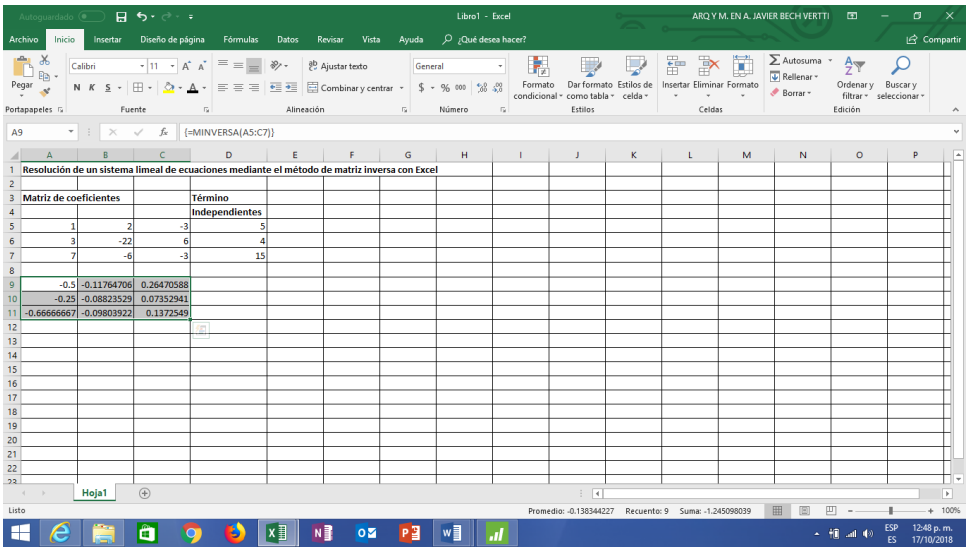
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



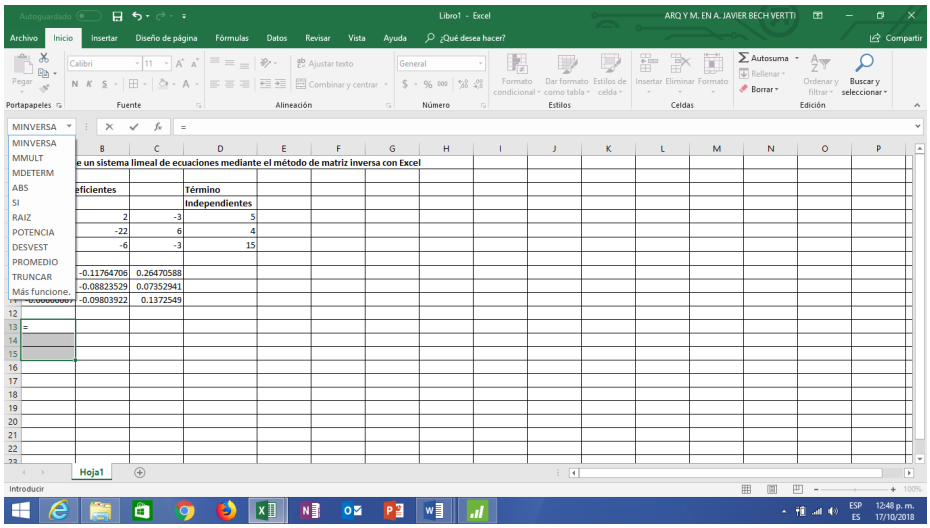
Selecciona las celdas de la matriz de coeficientes, en este caso A5:C7

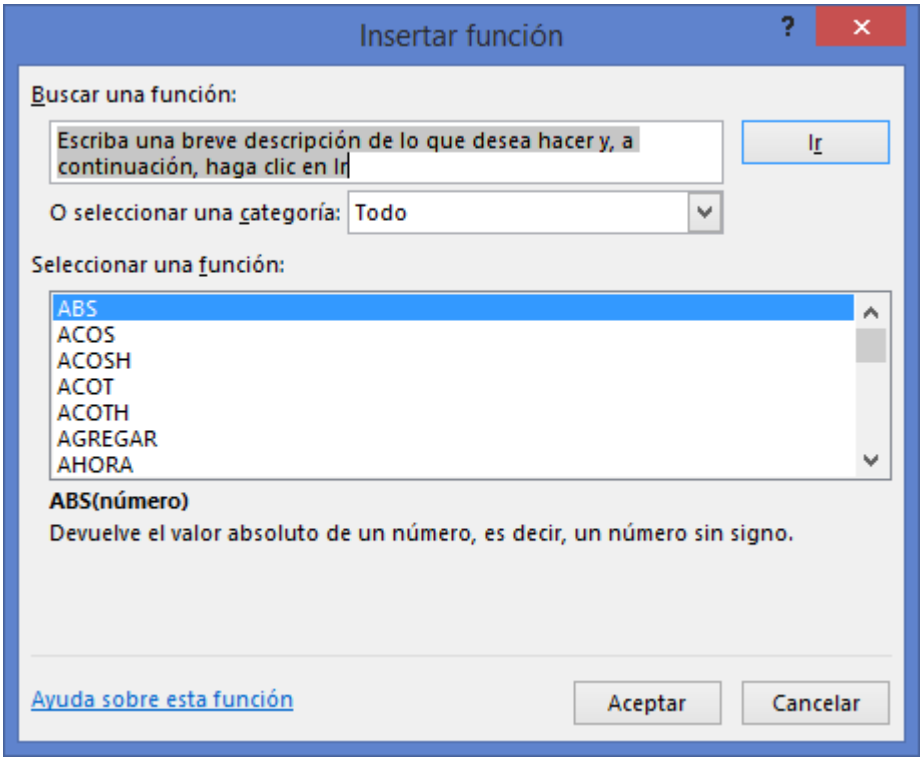


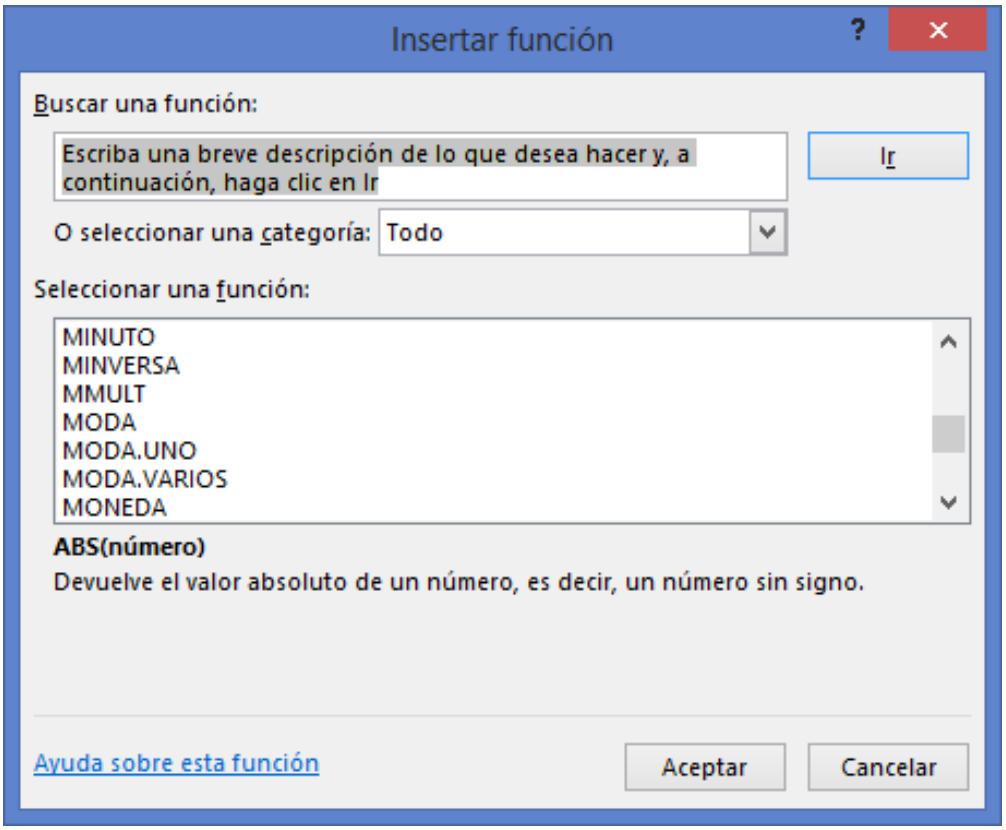
Ojo, no pulses Aceptar. La matriz inversa se determinará al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter)**.



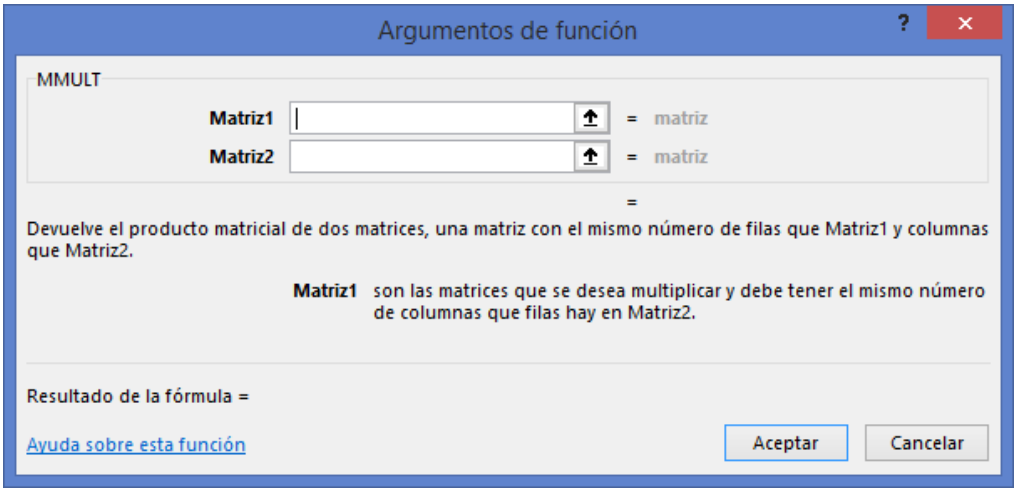
Como la matriz de los términos constantes es un vector de 1 x 3 se necesitas 3 celdas para mostrar los resultados de las Incógnitas **X, Y y Z** para lo cual selecciona con el mouse 3 celdas debajo de la **matriz inversa** dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MMULT** () seleccionando la opción **Más funciones** en la parte inferior



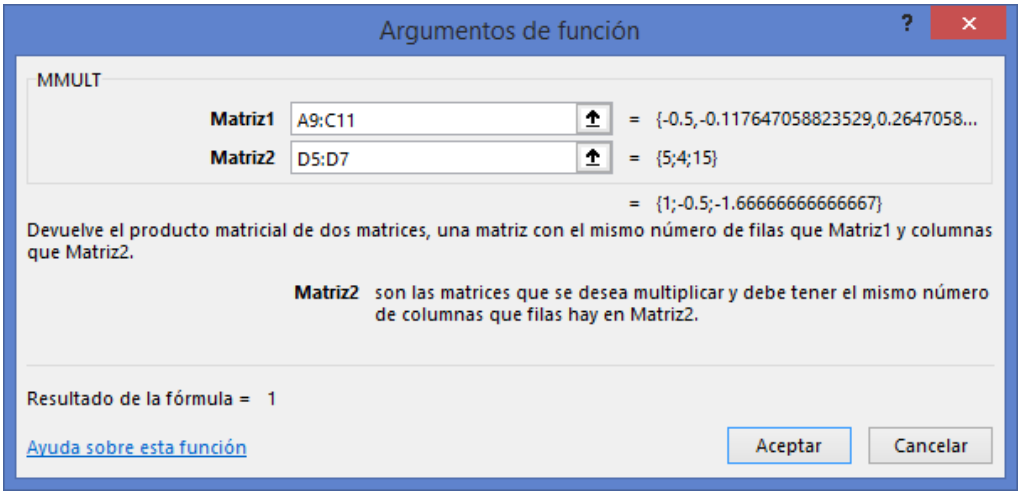




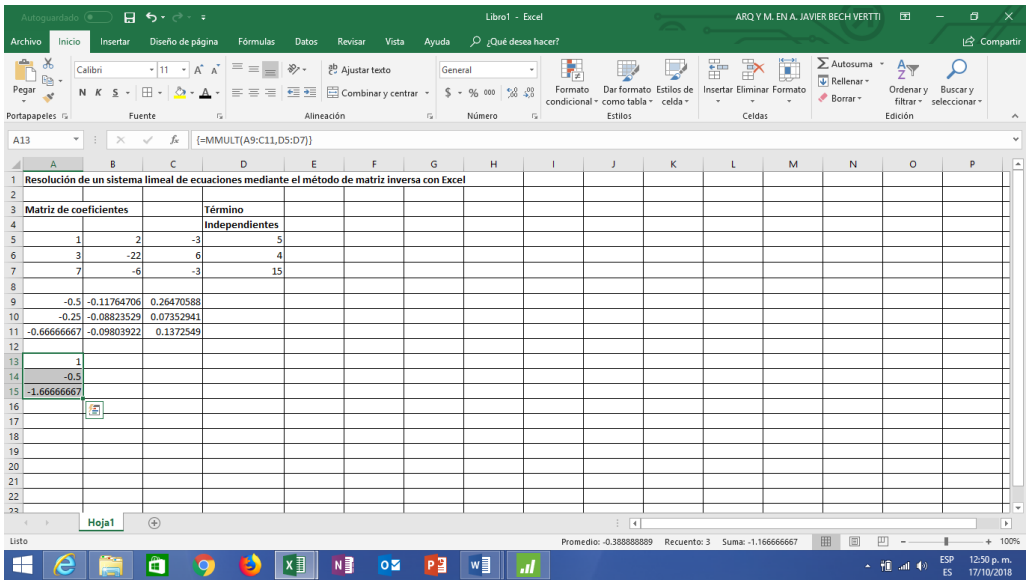
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



En el campo que dice Matriz 1 selecciona las celdas de la matriz de inversa, en este caso A9:C11 y en el campo que dice Matriz 2 selecciona las celdas de los términos constantes, en este caso D5:D7



Ojo, no pulses Aceptar. Los resultados de las incógnitas **X, Y, Z** se determinarán al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter), en este caso 1,-0.5,-1.66666667.** Por lo tanto las incognitas quedan **X=1.00, Y=-0.50 y Z=-1.67 aproximando a dos cifras.**



Con lo anterior queda terminado el ejemplo ilustrativo de **resolución de un sistema lineal de ecuaciones mediante el método de matriz inversa** en el paquete **EXCEL**. **Aplica el procedimiento para salir de Windows**

RESPUESTAS A LOS EJERCICIOS DE REFUERZO DEL CAPÍTULO 1

1.3.1.1.1 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 1.3.1.1.1

DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

1.3.1.1.1 Una agencia de Bienes Raíces desea examinar el **IMPUESTO** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **IMPUESTO** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

3959	3959	2911	3871	3638	2500
3142	3700	3930	2400	2708	1775
2600	4500	2378	2540	2989	3000
2250	2860	1800	4200	1814	2386
2857	4343	2700	3600	2800	3500

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas sin recortar los datos o valores atípicos generado por Minitab:

Tallo y hoja de IMPUESTO N = 30

Unidad de hoja = 100

1 1 7

3 1 88

3 2

6 2 233

9 2 455

12 2 677

(5) 2 88899

13 3 01

11 3

11 3 5

10 3 667

7 3 8999

3 4

3 4 23

1 4 5

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas recortando los datos o valores atípicos generado por Minitab:

Tallo y hoja de IMPUESTO N = 30

Unidad de hoja = 100

1 1 7

3 1 88

3 2

6 2 233

9 2 455

12 2 677

(5) 2 88899

13 3 01

11 3

11 3 5

10 3 667

7 3 8999

3 4

3 4 23

1 4 5

Otra forma de ordenarlos es colocarlos en forma ascendente:

1775	2386	2700	2911	3600	3959
1800	2400	2708	2989	3638	3959
1814	2500	2800	3000	3700	4200
2250	2540	2857	3142	3871	4343
2378	2600	2860	3500	3930	4500

Resumen de 5 Números: Resultados aproximados a las décimas.

1. Valor Mínimo: **1775**
2. Cuartil 1 Q_1=Dato 7 + 0.75 (Dato 8-Dato 7)= 2400 + 0.75 (2500-2400)=**2475**

   P(Q_1)= 1 $\frac{30+1}{4}$= 7.75
3. Mediana $\tilde{X}$= $\frac{Dato\ 16-Dato15}{2}$=$\frac{2911+2860}{2}$=**2886**

   P($\tilde{X}$)=$\frac{n+1}{2}$=$\frac{30+1}{2}$=15.5
4. Cuartil 3 Q_3=Dato 23+ 0.25(Dato 24- Dato 23)=3700+0.25(3871-3700)=**3743**

   P(Q_3)= 3 $\frac{30+1}{4}$= 23.25
5. Valor Máximo= **4500**

Esquema de vallas: Resultados aproximados a las décimas

VEI (Valla Externa Inferior)= Q_1-3(RIQ)=2475-3(3743-2475)= 2475-3(1268)=

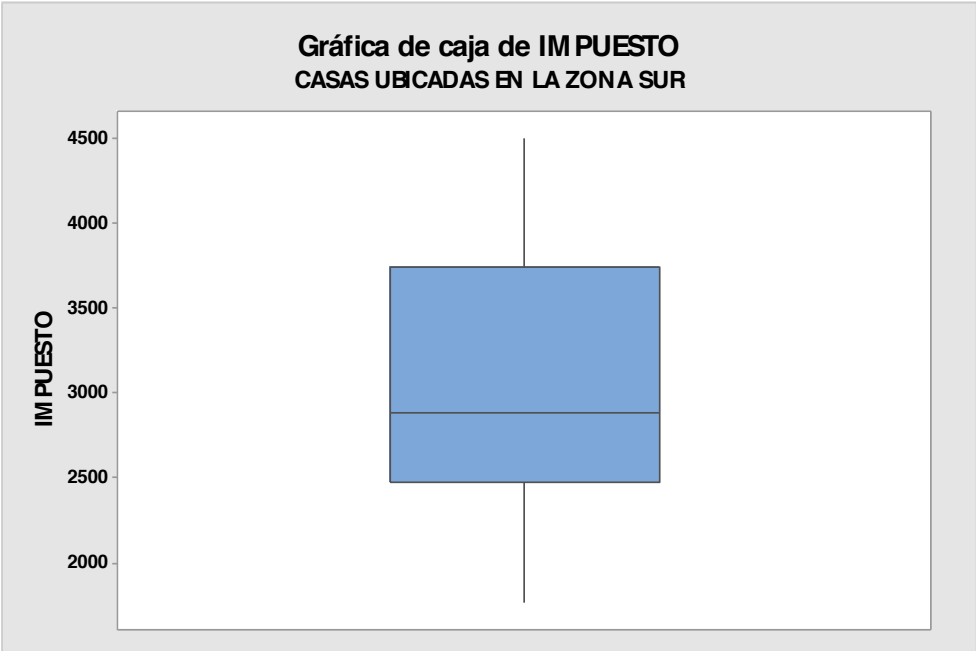
-**1329**

VII (Valla Interna Inferior)= Q_1-1.5(RIQ)=475-1.5(3743-2475)= 2475-1.5(1268)=**573**

VIS (Valla Interna Superior)= Q_{3+}1.5(RIQ)=3743+1.5(3743-2475)= 3743+1.5(1268)=**5645**

VES (Valla Externa Superior)= Q_3-3(RIQ)=3743+3(3743-2475)= 3743+3(1268)=**7547**

Diagrama de Caja y Brazos



Interpretación: En el diagrama anterior se puede observar que no existen casos o valores atípicos. La posición de la mediana dentro de la caja cargada al lado izquierdo nos puede sugerir un ligero sego a la derecha que debe ser tomado en cuenta para analizar la Normalidad de la distribución.

1.3.1.1.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.1.2**

DATOS ATÍPICOS**. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.**

1.3.1.1.2 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **NORTE** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas el **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

30	33	30	34	41	35
35	38	30	44	35	37
35	37	15	40	28	34
24	35	35	35	14	25
22	31	31	38	30	33

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas sin recortar los datos o valores atípicos generado por Minitab:

Tallo y hoja de ANTIGUED N = 30

Unidad de hoja = 1.0

1 1 4

2 1 5

4 2 24

6 2 58

(10) 3 0000113344

14 3 55555557788

3 4 014

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas recortando los

datos o valores atípicos generado por Minitab:

Tallo y hoja de ANTIGUED N = 30

Unidad de hoja = 1.0

LO 14, 15

5 2 45

5 2

6 2 8

12 3 000011

14 3 33

(9) 3 445555555

7 3 77

5 3 88

3 4 01

HI 44

Otra forma de ordenarlos es colocarlos en forma ascendente:

14	28	31	34	35	38
15	30	31	35	35	38
22	30	33	35	35	40
24	30	33	35	37	41
25	30	34	35	37	44

Resumen de 5 Números: Resultados aproximados a las décimas.

1. Valor Mínimo: **14**
2. Cuartil 1 Q_1=Dato 7 + 0.75 (Dato 8-Dato 7)= 30 + 0.75 (30-30)=**30**

   P(Q_1)= 1 $\frac{30+1}{4}$= 7.75
3. Mediana $\tilde{X}$= $\frac{Dato\ 16-Dato15}{2}$=$\frac{34+34}{2}$=**34**

   P($\tilde{X}$)=$\frac{n+1}{2}$=$\frac{30+1}{2}$=15.5

4. Cuartil 3 Q_3=Dato 23+ 0.25(Dato 24- Dato 23)=35+0.25(37-35)=**35.5**

$P(Q_3)= 3\frac{30+1}{4}= 23.25$

5. Valor Máximo= **44**

Esquema de vallas: Resultados aproximados a las décimas

VEI (Valla Externa Inferior)= Q_1-3(RIQ)=30-3(35.5-30)= 30-3(5.5)=

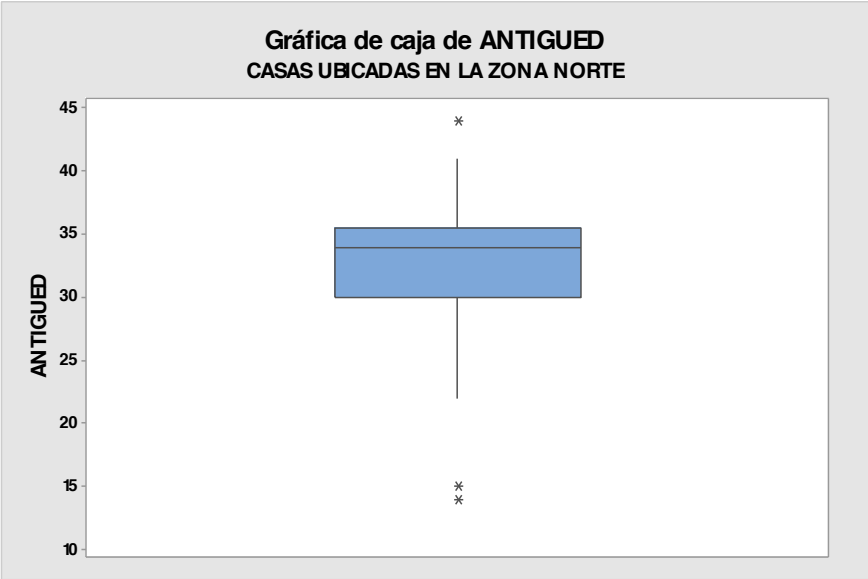
13.5

VII (Valla Interna Inferior)= Q_1-1.5(RIQ)=30-1.5(35.5-30)= 30-1.5(5.5)=**21.75**

VIS (Valla Interna Superior)= Q_{3+}1.5(RIQ)=35.5+1.5(35.5-30)= 35.5+1.5(5.5)=**43.75**

VES (Valla Externa Superior)= Q_3-3(RIQ)=35.5+3(35.5-30)= 35.5+3(5.5)=**52**

Diagrama de Caja y Brazos



Interpretación: En el diagrama anterior se puede observar que existen 3 valores atípicos: **14, 15** entre las **VEI y la VII (a los que denominamos aberrantes)** y el **44** entre las **VIS y la VES (al que denominamos aberrante)** los que deben ser estudiados con detalle para ver el tratamiento que se les va a dar en estudios subsecuentes ya que es posible que provoquen un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la derecha de la caja, sugiere un sesgo hacia la izquierda producto tal vez de los valores o casos atípicos).

1.3.1.1.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.1.3**

DATOS **A**TÍPICOS**. DIAGRAMA DE CAJA Y BRAZOS SIMPLE.**

1.3.1.1.3 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. Se levanta en principio una encuesta que más adelante se llamará **"BIENES1"** a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

32	28	35	30	61	34
60	27	30	16	39	21
27	28	39	28	31	34
35	35	30	31	55	38
30	14	28	31	32	24

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas sin recortar los datos o valores atípicos generado por Minitab:

Tallo y hoja de ANTIGUED N = 30

Unidad de hoja = 1.0

1 1 4

2 1 6

4 2 14

10 2 778888

(11) 3 00001112244

9 3 555899

3 4

3 4

3 5

3 5 5

2 6 01

Para ordenar los datos en forma ascendente usamos un diagrama de tallo y hojas recortando los datos o valores atípicos generado por Minitab:

Tallo y hoja de ANTIGUED N = 30

Unidad de hoja = 1.0

LO 14, 16

3 2 1

3 2

4 2 4

6 2 77

10 2 8888

(7) 3 0000111

13 3 22

11 3 44555

6 3

6 3 899

HI 55, 60, 61

Otra forma de ordenarlos es colocarlos en forma ascendente:

14	27	30	31	34	39
16	28	30	31	35	39
21	28	30	32	35	55
24	28	30	32	35	60
27	28	31	34	38	61

Resumen de 5 Números: Resultados aproximados a las décimas.

1. Valor Mínimo: **14**
2. Cuartil 1 Q_1=Dato 7 + 0.75 (Dato 8-Dato 7)= 28 + 0.75 (28-28)=**28**

   P(Q_1)= 1 $\frac{30+1}{4}$= 7.75
3. Mediana $\tilde{X}$= $\frac{Dato\ 16-Dato15}{2}$=$\frac{31+31}{2}$=**31**

   P($\tilde{X}$)=$\frac{n+1}{2}$=$\frac{30+1}{2}$=15.5
4. Cuartil 3 Q_3=Dato 23+ 0.25(Dato 24- Dato 23)=35+0.25(35-35)=**35**

   P(Q_3)= 3 $\frac{30+1}{4}$= 23.25
5. Valor Máximo= **61**

Esquema de vallas: Resultados aproximados a las décimas

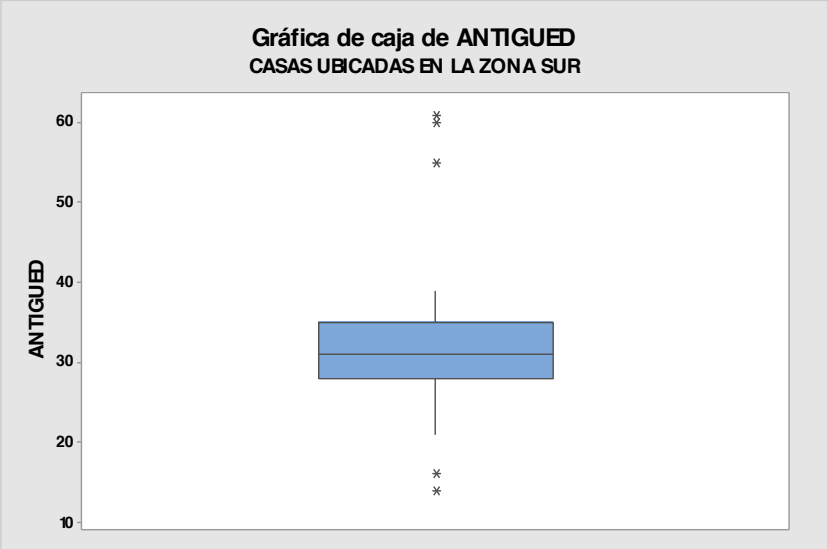
VEI (Valla Externa Inferior)= Q_1-3(RIQ)=28-3(35-28)= 28-3(7)=**7**

VII (Valla Interna Inferior)= Q_1-1.5(RIQ)=28-1.5(35-28)= 28-1.5(7)=**17.5**

VIS (Valla Interna Superior)= Q_{3+}1.5(RIQ)=35+1.5(35-28)= 35+1.5(7)=**45.5**

VES (Valla Externa Superior)= Q_3-3(RIQ)=35+3(35-28)= 35+3(7)=**56**

Diagrama de Caja y Brazos



Interpretación: En el diagrama anterior se puede observar que existen 3 valores atípicos: **14, 16** entre las **VEI y la VII (a los que denominamos aberrantes)** el **55** entre las **VIS y la VES (al que denominamos aberrante)** y **60 y 61** después de la **VES (a los que denominamos aberrantes lejanos)** los que deben ser estudiados con detalle para ver el tratamiento que se les va a dar en estudios subsecuentes ya que es posible que provoquen un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, ligeramente a la izquierda de la caja, sugiere un sesgo hacia la derecha producto tal vez de los valores o casos atípicos).

1.3.1.2.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.2.1**

DATOS **ATÍPICOS**

1.3.1.2.1. Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **NÚMERO DE RECÁMARAS** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.**

1 Abra la base de datos que creó **"BIENES1".** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***RECAMARA***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

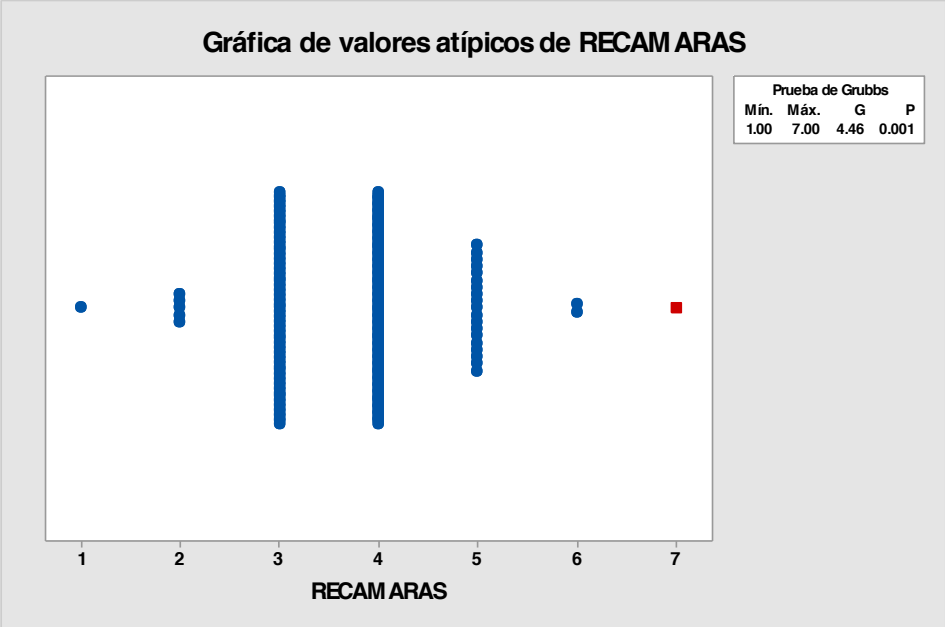
Salida de la ventana Sesión

Prueba de valores atípicos: RECAMARA

```
Método

Hipótesis nula          Todos los valores de los datos provienen de la misma
población normal
Hipótesis alterna       El valor más pequeño o más grande de los datos es un
valor atípico
Nivel de significancia  α = 0.05



Prueba de Grubbs

Variable    N   Media  Desv.Est.    Mín.    Máx.     G      P
RECAMARA  233  3.6395     0.7535  1.0000  7.0000  4.46  0.001


Valor atípico

                Valor
Variable  Fila  atípico
RECAMARA   103        7
```

Gráfica de valores atípicos de RECAMARA

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **NÚMERO DE RECÁMARAS** muestran el valor más pequeño es de 1.00 y el más grande de 7.00. La estadística de G de 4.46 indica que el valor de datos más pequeño ó más grande de 1.00 ó 7.00, es 4.46 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.000. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico y corresponde al caso o la casa 103 con un valor máximo de 7 recamaras.

Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante

1.3.1.2.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.2.2**

DATOS **ATÍPICOS**

1.3.1.2.2. Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **NÚMERO DE BAÑOS** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño

o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.**

1 Abra la base de datos que creó **"BIENES1".** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***BAÑOS***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

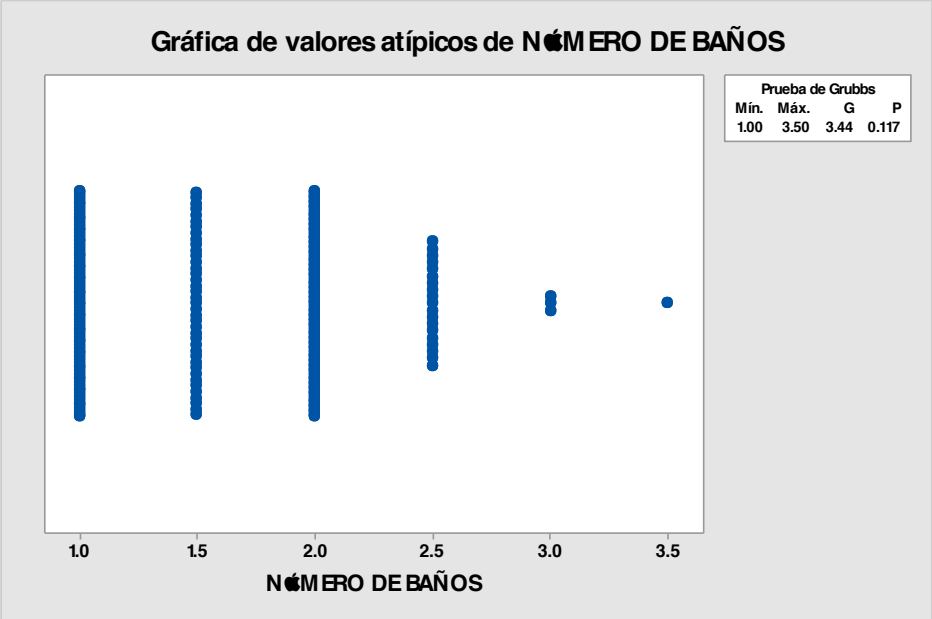
Salida de la ventana Sesión

Prueba de valores atípicos: BAÑOS

```
Método

Hipótesis nula          Todos los valores de los datos provienen de la misma
población normal
Hipótesis alterna       El valor más pequeño o más grande de los datos es un
valor atípico
Nivel de significancia  α = 0.05


Prueba de Grubbs

Variable    N   Media  Desv.Est.    Mín.    Máx.     G      P
BAÑOS     233  1.6330     0.5429  1.0000  3.5000  3.44  0.117

* NOTA * No hay valor atípico en el nivel de significancia de 5%
```

Gráfica de valores atípicos de BAÑOS

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **NÚMERO DE BAÑOS** muestran el valor más pequeño es de 1.00 y el más grande de 3.50. La estadística de G de 3.34 indica que el valor de datos más pequeño ó más grande de 1.00 ó 3.50, es 3.34 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.117. Este valor p es mayor que el nivel de significancia que se eligió, de modo que no se puede rechazar la hipótesis nula y concluir que el valor más pequeño o mas grande no es un valor atípico

Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

1.3.1.2.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.2.3**

DATOS **ATÍPICOS**

1.3.1.2.3. Se reciben datos sobre una encuesta realizada a 233 casas para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **IMPUESTOS ANUALES** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la Prueba de valores atípicos de Minitab para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico. **Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.**

1 Abra la base de datos que creó **"BIENES1".** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***IMPUESTO***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

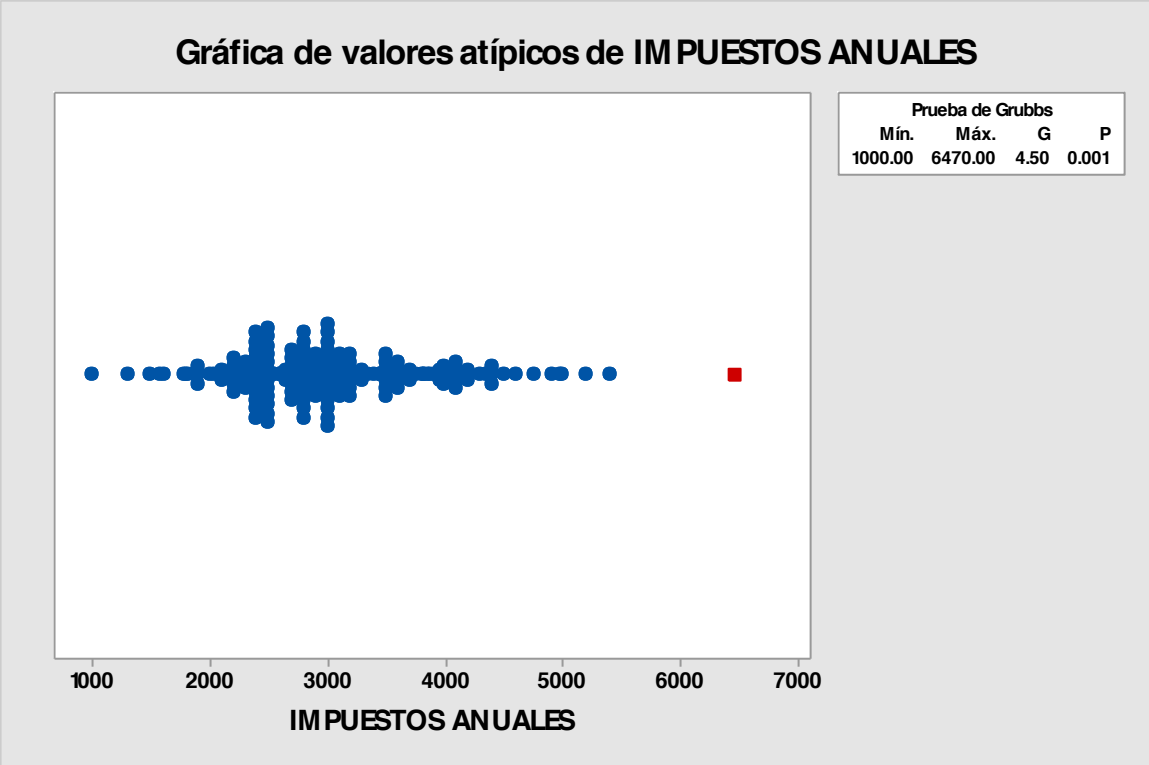
Salida de la ventana Sesión

Prueba de valores atípicos: IMPUESTO

```
Método

Hipótesis nula          Todos los valores de los datos provienen de la misma
población normal
Hipótesis alterna       El valor más pequeño o más grande de los datos es un
valor atípico
Nivel de significancia  α = 0.05


Prueba de Grubbs

Variable    N   Media  Desv.Est.    Mín.    Máx.     G      P
IMPUESTO  233  3034.9      763.5  1000.0  6470.0  4.50  0.001


Valor atípico

                Valor
Variable  Fila  atípico
IMPUESTO   226     6470
```

Gráfica de valores atípicos de IMPUESTO

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **IMPUESTOS ANUALES** muestran el valor más pequeño es de 1000.00 y el más grande de 6470.00. La estadística de G de 4.50 indica que el valor de datos más pequeño ó más grande de 1000.00 ó 6470.00, es 4.50 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.001. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico y corresponde al caso o la casa 226 con un valor máximo de 6470.00 dólares.

Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

1.3.1.3.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.3.1**

DATOS ATÍPICOS PARA UN NIVEL DE UN **FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO**

1.3.1.3.1 Se reciben datos sobre una encuesta realizada a las casas con estilo **MODERNO** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en las casas con estilo **MODERNO** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

c) Crear un subconjunto de hoja de trabajo.
d) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

Respuesta al inciso a):

Crear un subconjunto de hoja de trabajo.

Nota importante:

Al **crear un subconjunto de hoja de trabajo** se divide, o desapila, la hoja de trabajo activa en hojas de trabajo nuevas basándose en una o más "Por" variables. Crear subconjunto de hoja de trabajo siempre copian los datos en hojas de trabajo **nuevas**. Como la base de datos **"BIENES 1"** contiene todas las casas con estilo **MODERNO, HACIENDA, COLONIAL, GRANJA Y CON DESNIVELES** debemos crear **un subconjunto de hoja de trabajo** que solo incluya las casas con estilo **MODERNO**. Esto nos permite analizar los diferentes niveles de un factor por separado en caso de que la prueba de **ANOVA** haya sido significativa y exista la posibilidad de que las medias de la variable **VALOR** no sea la misma en los cinco niveles del Factor **ESTILO**

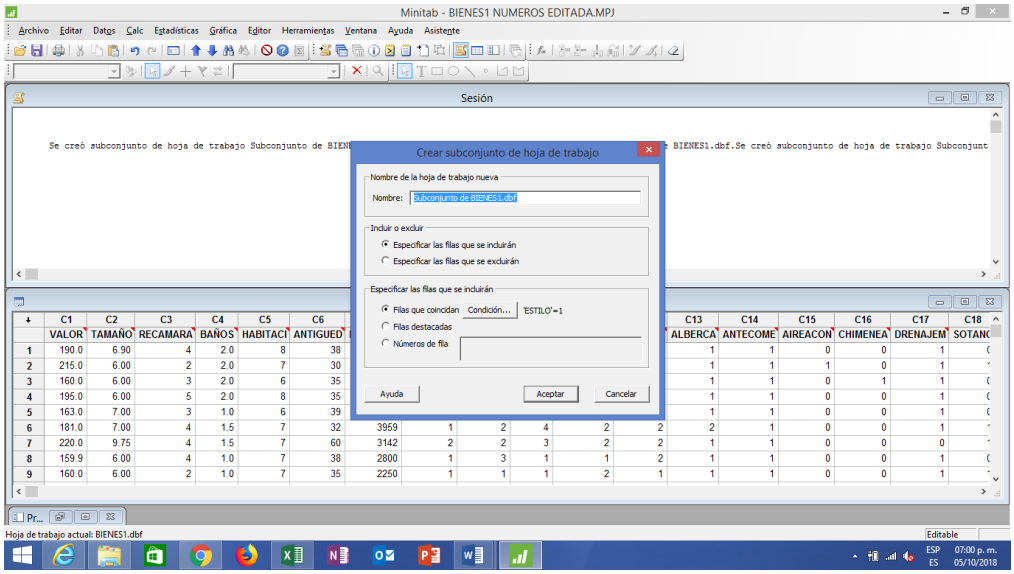
1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Datos > Crear subconjunto de hoja de trabajo**.

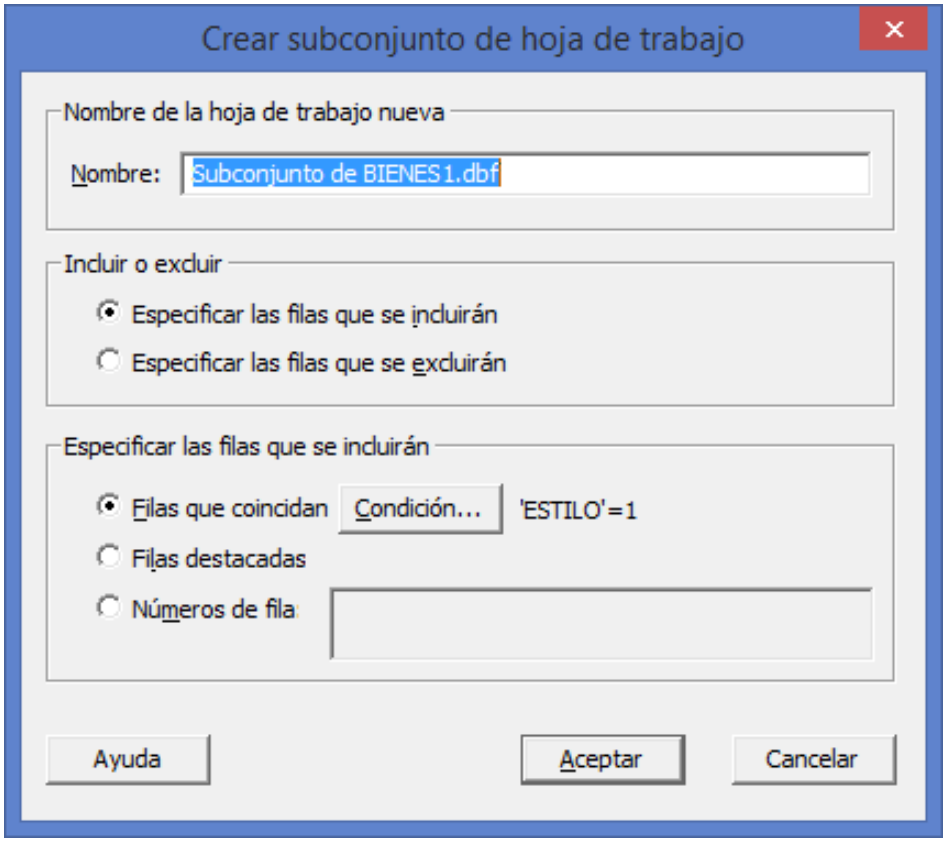
3 En **Incluir o excluir**, elija **Especificar las filas que se incluirán.**

4 En **Especificar las filas que se incluirán**, elija **Filas que coincidan**. Haga clic en **Condición**.

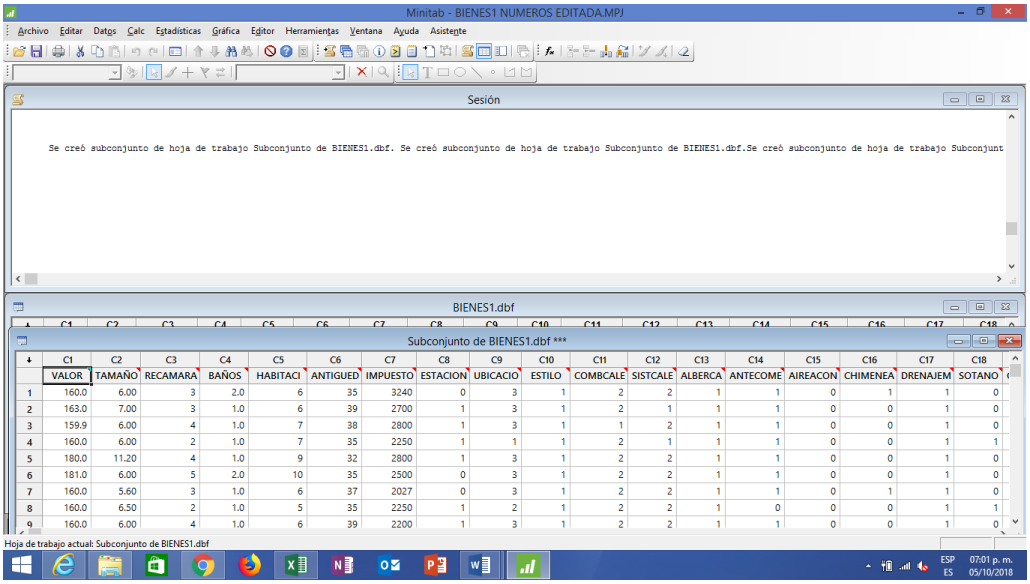
5 En **Condición**, ingrese ***ESTILO=1.*** Haga clic en **Aceptar** en cada cuadro de diálogo.



Nota importante: Minitab es sensible al nombre que se le asignó a la variable por lo cual es conveniente no escribir el nombre de la variable sino seleccionarla de la parte izquierda de la ventana y solo añadir **=1.**



Minitab Crea un subconjunto de hoja nueva que incluye las casas con estilo **MODERNO**



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	VALOR	TAMAÑO	RECAMARA	BAÑOS	HABITACI	ANTIGUED	IMPUESTO	ESTACION	UBICACIO	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO
1	160.0	6.00	3	2.0	6	35	3240	0	3	1	2	2	1	1	0	1	1	0
2	163.0	7.00	3	1.0	6	39	2700	1	3	1	2	1	1	1	0	0	1	0
3	159.9	6.00	4	1.0	7	38	2800	1	3	1	1	2	1	1	0	0	1	0
4	160.0	6.00	2	1.0	7	35	2250	1	1	1	2	1	1	1	0	0	1	1
5	180.0	11.20	4	1.0	9	32	2800	1	3	1	2	2	1	1	0	0	1	0
6	181.0	6.00	5	2.0	10	35	2500	0	3	1	2	2	1	1	0	0	1	0
7	160.0	5.60	3	1.0	6	37	2027	0	3	1	2	2	1	1	0	1	1	0
8	160.0	6.50	2	1.0	5	35	2250	1	2	1	2	2	1	0	0	0	1	1
9	160.0	6.00	4	1.0	6	39	2200	1	3	1	2	2	1	1	0	0	1	0

Respuesta al inciso b):

Realizar la Prueba de valores atípicos de Minitab

1 En la nueva hoja de trabajo que Minitab generó de la base de datos **"BIENES1"** que incluye solo las casas con estilo **MODERNO** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***Valor***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

Salida de la ventana Sesión

Prueba de valores atípicos: VALOR

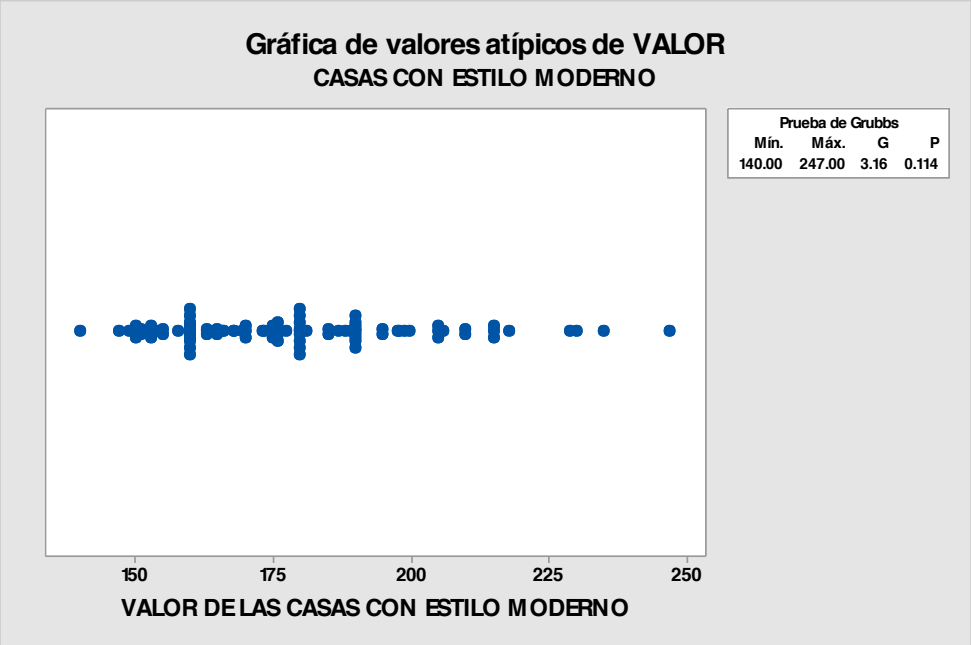
Método

Hipótesis nula	Todos los valores de los datos provienen de la misma población normal
Hipótesis alterna	El valor más pequeño o más grande de los datos es un valor atípico
Nivel de significancia	α = 0.05

```
Prueba de Grubbs


Variable   N   Media  Desv.Est.    Mín.    Máx.     G      P

VALOR     97  178.23      21.73  140.00  247.00  3.16  0.114


* NOTA * No hay valor atípico en el nivel de significancia de 5%
```

Gráfica de valores atípicos de VALOR

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **VALOR en las casas con estilo MODERNO** muestran el valor más pequeño es de 140 y el más grande de 247. La estadística de G de 3.16 indica que el valor de datos más pequeño ó más grande de 140 ó 247, es 3.16 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.114. Este valor p es mayor que el nivel de significancia que se eligió, de modo que no se puede rechazar la hipótesis nula y concluir que el valor más pequeño ó el más grande no es un valor atípico.

1.3.1.3.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.3.2**

DATOS ATÍPICOS PARA UN NIVEL DE UN **FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO**

1.3.1.3.2 Se reciben datos sobre una encuesta realizada a las casas con estilo **HACIENDA** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en las casas con estilo **HACIENDA** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

c) Crear un subconjunto de hoja de trabajo.
d) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

Respuesta al inciso a):

Crear un subconjunto de hoja de trabajo.

Nota importante:

Al **crear un subconjunto de hoja de trabajo** se divide, o desapila, la hoja de trabajo activa en hojas de trabajo nuevas basándose en una o más "Por" variables. Crear subconjunto de hoja de trabajo siempre copian los datos en hojas de trabajo **nuevas**. Como la base de datos **"BIENES 1"** contiene todas las casas con estilo **MODERNO, HACIENDA, COLONIAL, GRANJA Y CON DESNIVELES** debemos crear **un subconjunto de hoja de trabajo** que solo incluya las casas con estilo **HACIENDA**. Esto nos permite analizar los diferentes niveles de un factor por separado en caso de que la prueba de **ANOVA** haya sido significativa y exista la posibilidad de que las medias de la variable **VALOR** no sea la misma en los cinco niveles del Factor **ESTILO**

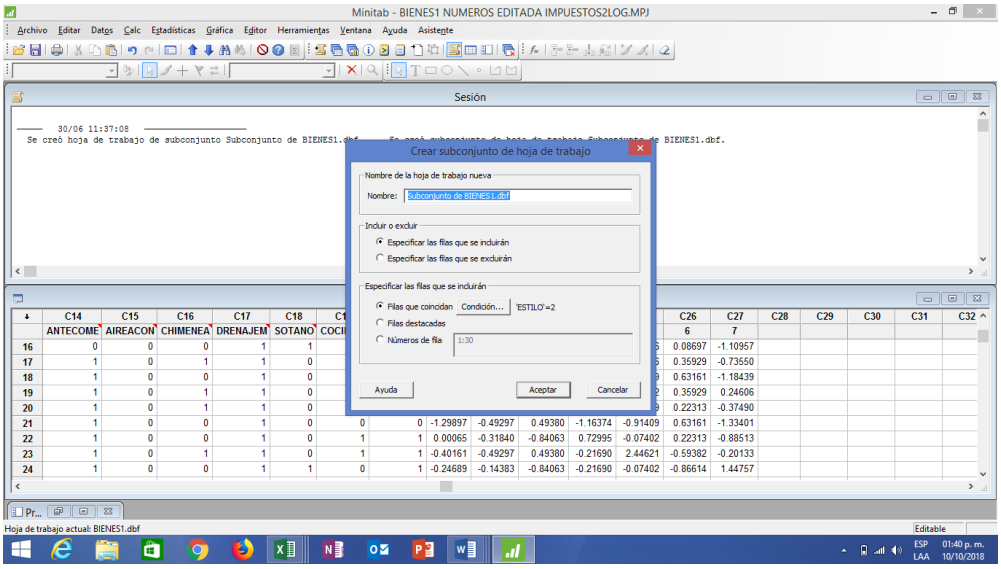
1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Datos > Crear subconjunto de hoja de trabajo**.

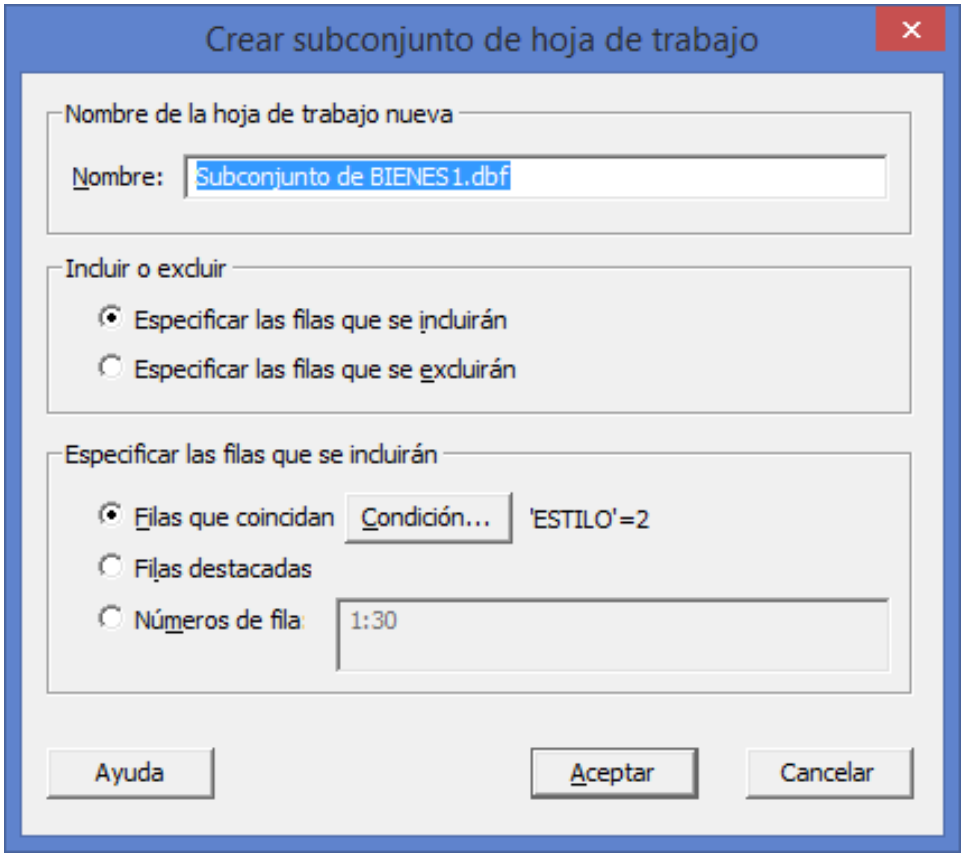
3 En **Incluir o excluir**, elija **Especificar las filas que se incluirán.**

4 En **Especificar las filas que se incluirán**, elija **Filas que coincidan**. Haga clic en **Condición**.

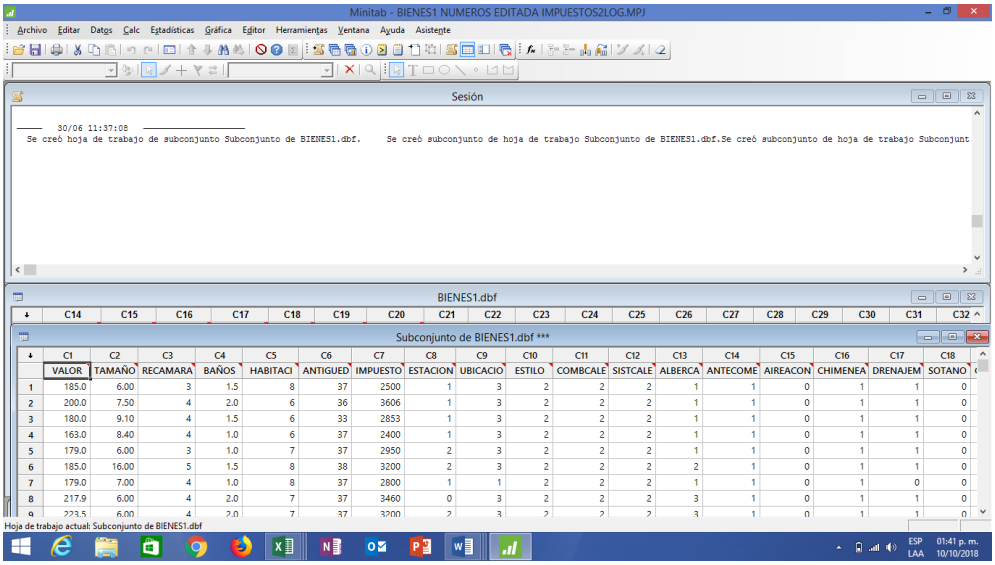
5 En **Condición**, ingrese ***ESTILO=2.*** Haga clic en **Aceptar** en cada cuadro de diálogo.



Nota importante: Minitab es sensible al nombre que se le asignó a la variable por lo cual es conveniente no escribir el nombre de la variable sino seleccionarla de la parte izquierda de la ventana y solo añadir **=2.**



Minitab Crea un subconjunto de hoja nueva que incluye las casas con estilo **MODERNO**



Respuesta al inciso b):

Realizar la Prueba de valores atípicos de Minitab

1 En la nueva hoja de trabajo que Minitab generó de la base de datos **"BIENES1"** que incluye solo las casas con estilo **HACIENDA** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***Valor***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

Salida de la ventana Sesión

Prueba de valores atípicos: VALOR

Método

Hipótesis nula Todos los valores de los datos provienen de la misma población normal
Hipótesis alterna El valor más pequeño o más grande de los datos es un valor atípico
Nivel de significancia $\alpha = 0.05$

Prueba de Grubbs

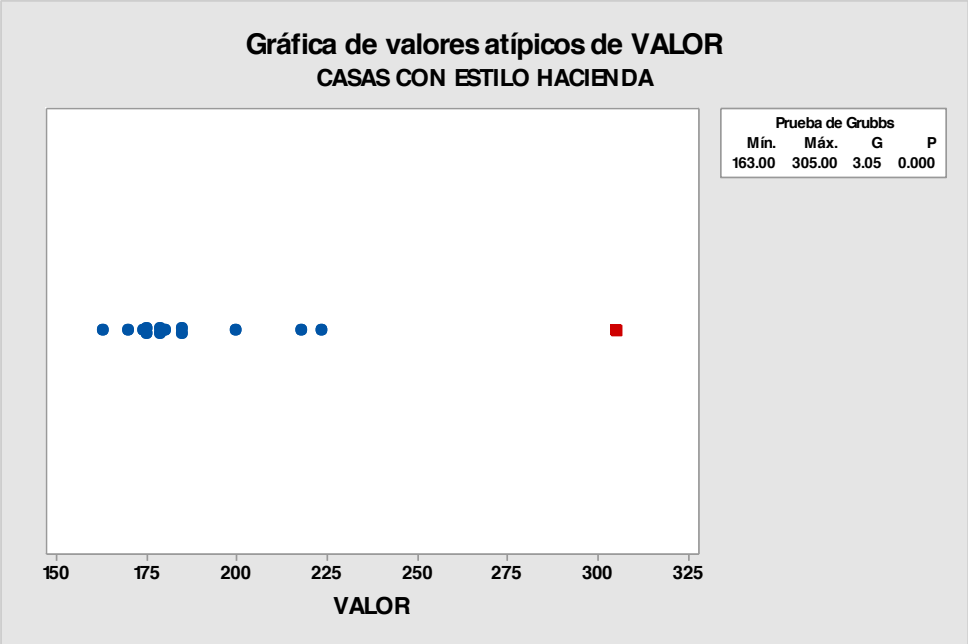
Variable	N	Media	Desv.Est.	Mín.	Máx.	G	P
VALOR	14	193.67	36.44	163.00	305.00	3.05	0.000

Valor atípico

Variable	Fila	Valor atípico
VALOR	11	305

Gráfica de valores atípicos de VALOR

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **VALOR en las casas con estilo HACIENDA** muestran el valor más pequeño es de 163 y el más grande de 305. La estadística de G de 3.05 indica que el valor de datos más pequeño ó más grande de 163 ó 305, es 3.05 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.000. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico. En este caso corresponde al valor 305 que se observa en rojo en la gráfica y que pertenece al caso o **fila 11 del nuevo subconjunto de hoja de trabajo**.

Es importante aclarar que hay que ubicar dicho caso en la base de datos "BIENES1" original por si es el caso en que haya que eliminirlo para depurar dicha base de datos.

1.3.1.3.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.3.3**

DATOS ATÍPICOS PARA UN NIVEL DE UN **FACTOR. SUBCONJUNTO DE HOJA DE TRABAJO**

1.3.1.3.3 Se reciben datos sobre una encuesta realizada a las casas con estilo **GRANJA** de la base de datos **"BIENES 1"** para realizar diferentes estudios sobre ella. Nota que el valor más pequeño o más grande de los datos de la variable **VALOR** en las casas con estilo **GRANJA** parecen inusualmente pequeño o grande. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea utilizar la **Prueba de valores atípicos de Minitab** para determinar si el valor más pequeño o el más grande se puede considerar como valor atípico.

c) Crear un subconjunto de hoja de trabajo.
d) Realizar la Prueba de valores atípicos de Minitab

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

Respuesta al inciso a):

Crear un subconjunto de hoja de trabajo.

Nota importante:

Al **crear un subconjunto de hoja de trabajo** se divide, o desapila, la hoja de trabajo activa en hojas de trabajo nuevas basándose en una o más "Por" variables. Crear subconjunto de hoja de trabajo siempre copian los datos en hojas de trabajo **nuevas**. Como la base de datos **"BIENES 1"** contiene todas las casas con estilo **MODERNO, HACIENDA, COLONIAL, GRANJA Y CON DESNIVELES** debemos crear **un subconjunto de hoja de trabajo** que solo incluya las casas con estilo **GRANJA**. Esto nos permite analizar los diferentes niveles de un factor por separado en caso de que la prueba de **ANOVA** haya sido significativa y exista la posibilidad de que las medias de la variable **VALOR** no sea la misma en los cinco niveles del Factor **ESTILO**

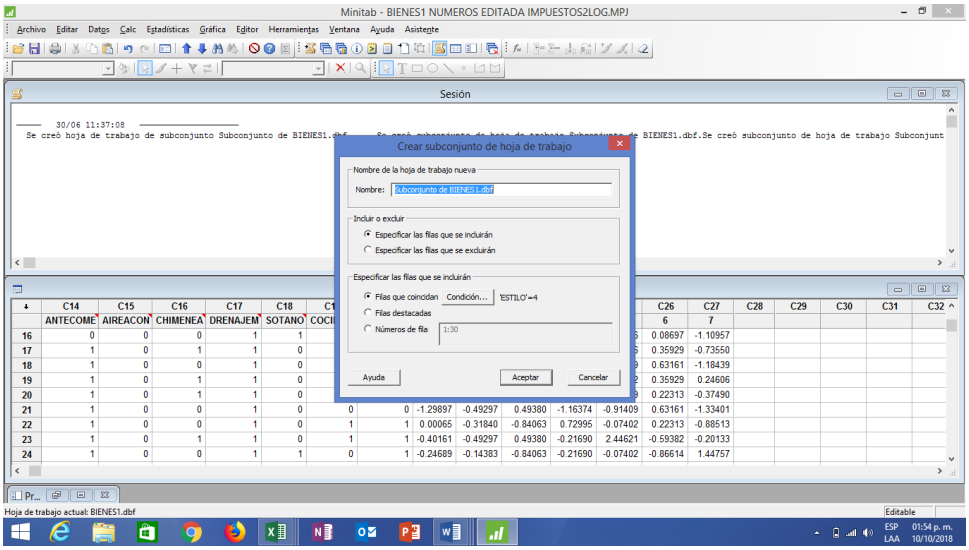
1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Datos > Crear subconjunto de hoja de trabajo**.

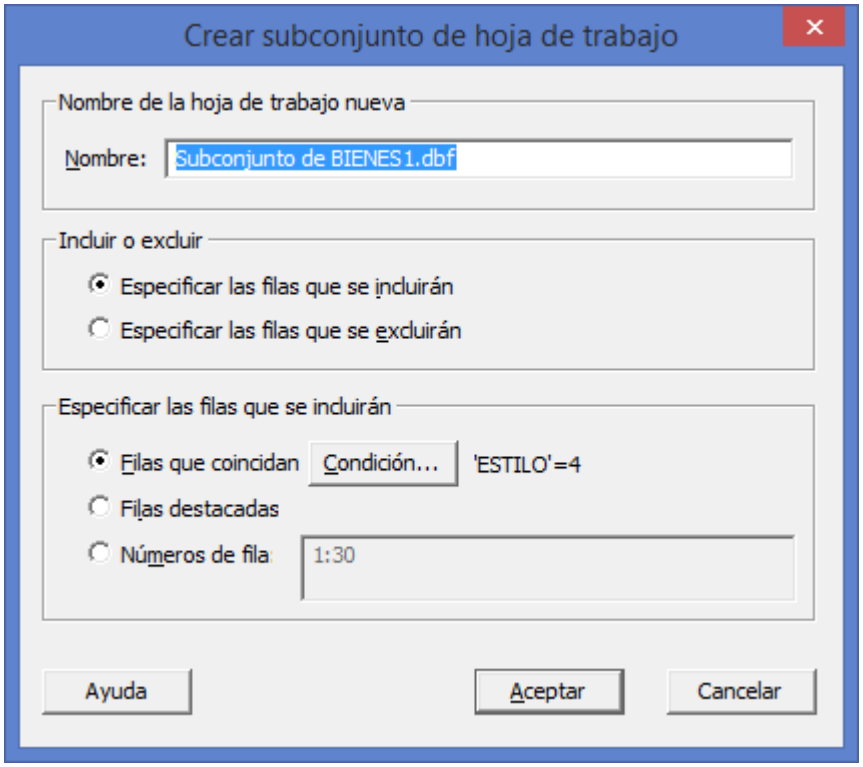
3 En **Incluir o excluir**, elija **Especificar las filas que se incluirán.**

4 En **Especificar las filas que se incluirán**, elija **Filas que coincidan**. Haga clic en **Condición**.

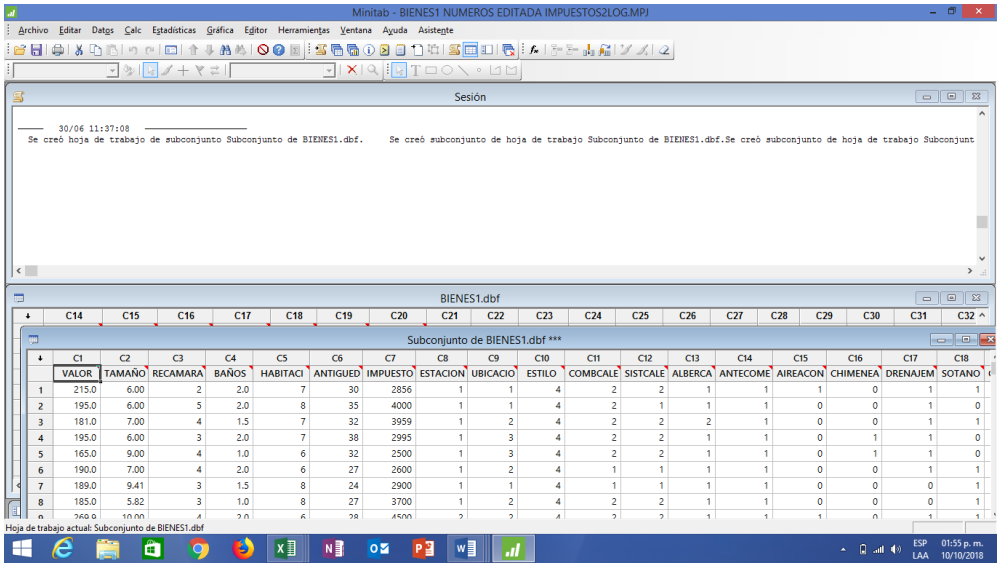
5 En **Condición**, ingrese ***ESTILO=4.*** Haga clic en **Aceptar** en cada cuadro de diálogo.



Nota importante: Minitab es sensible al nombre que se le asignó a la variable por lo cual es conveniente no escribir el nombre de la variable sino seleccionarla de la parte izquierda de la ventana y solo añadir **=4.**



Minitab Crea un subconjunto de hoja nueva que incluye las casas con estilo **GRANJA**



	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18
	VALOR	TAMAÑO	RECAMARA	BAÑOS	HABITACI	ANTIGUED	IMPUESTO	ESTACION	UBICACIO	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO
1	215.0	6.00	2	2.0	7	30	2856	1	1	4	2	2	1	1	1	0	1	1
2	195.0	6.00	5	2.0	8	35	4000	1	1	4	2	1	1	1	0	0	1	0
3	181.0	7.00	4	1.5	7	32	3959	1	2	4	2	2	2	1	0	0	1	1
4	195.0	6.00	3	2.0	7	38	2995	1	3	4	2	2	1	1	0	1	1	0
5	165.0	9.00	4	1.0	6	32	2500	1	3	4	2	2	1	1	0	1	1	0
6	190.0	7.00	4	2.0	6	27	2600	1	2	4	1	1	1	1	0	0	1	1
7	189.0	9.41	3	1.5	8	24	2900	1	1	4	1	1	1	1	0	0	0	1
8	185.0	5.82	3	1.0	8	27	3700	1	2	4	2	2	1	1	0	0	0	1
9	269.9	10.00	4	2.0	6	28	4500	2	2	4	2	2	1	1	0	0	1	1

Respuesta al inciso b):

Realizar la Prueba de valores atípicos de Minitab

1 En la nueva hoja de trabajo que Minitab generó de la base de datos **"BIENES1"** que incluye solo las casas con estilo **GRANJA** Elija **Estadísticas > Estadísticas básicas > Prueba de valor atípico**.

3 En **Variable**, ingrese ***Valor***

4 Haga clic en **Opciones**.

5 En **¿Qué desea determinar?**, elija **El valor más pequeño o más grande de los datos es un valor atípico**.

6 Haga clic en **Aceptar** en cada cuadro de diálogo.

Salida de la ventana Sesión

Prueba de valores atípicos: VALOR

Método

Hipótesis nula Todos los valores de los datos provienen de la misma población normal

Hipótesis alterna El valor más pequeño o más grande de los datos es un valor atípico

Nivel de significancia $\alpha = 0.05$

Prueba de Grubbs

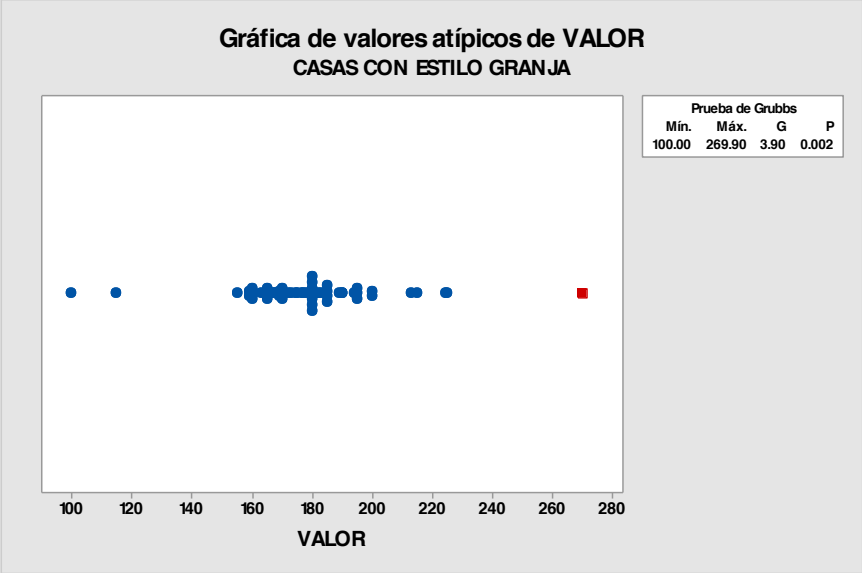
Variable	N	Media	Desv.Est.	Mín.	Máx.	G	P
VALOR	60	178.41	23.49	100.00	269.90	3.90	0.002

Valor atípico

Variable	Fila	Valor atípico
VALOR	9	269.9

Gráfica de valores atípicos de VALOR

Salida de la ventana Gráfica



Interpretación de los resultados

Los resultados correspondientes a los datos de **VALOR en las casas con estilo GRANJA** muestran el valor más pequeño es de 100 y el más grande de 269.9. La estadística de G de 3.90 indica que el valor de datos más pequeño ó más grande de 100 ó 269.9, es 3.90 desviaciones estándar menos más que la media. El valor p indica que, si todos los valores pertenecen en realidad a la misma población normalmente distribuida, entonces la probabilidad de obtener un valor mínimo así de pequeño o un valor máximo así de grande es 0.002. Este valor p es menor que el nivel de significancia que se eligió, de modo que se puede rechazar la hipótesis nula y concluir que el valor más grande es un valor atípico. En este caso corresponde al valor 269.9 que se observa en rojo en la gráfica y que pertenece al caso o **fila 9 del nuevo subconjunto de hoja de trabajo**.

Es importante aclarar que hay que ubicar dicho caso en la base de datos "BIENES1" original por si es el caso en que haya que eliminirlo para depurar dicha base de datos

1.3.1.4.1 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 1.3.1.4.1

DATOS ATÍPICOS . DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

1.3.1.4.1 Una agencia de Bienes Raíces desea examinar el **IMPUESTO** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. De la base de datos **"BIENES1"** se seleccionan aleatoriamente a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas el **IMPUESTO** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

3959	3959	2911	3871	3638	2500
3142	3700	3930	2400	2708	1775
2600	4500	2378	2540	2989	3000
2250	2860	1800	4200	1814	2386
2857	4343	2700	3600	2800	3500

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo

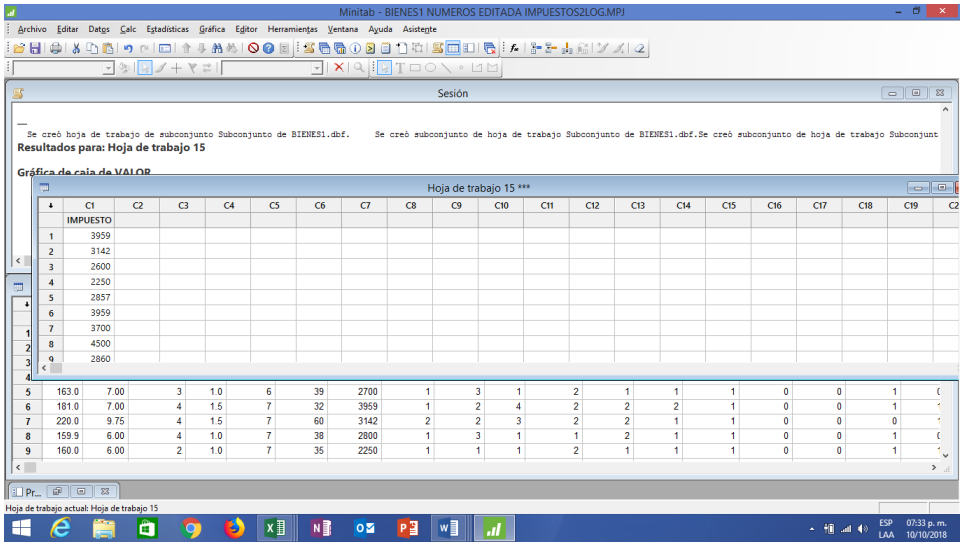
Para introducir los datos agrega una hoja de trabajo vacía al proyecto actual:

1 Elija **Archivo > Nuevo**.

2 Seleccione **Hoja de trabajo de Minitab** y haga clic en **Aceptar**.

3 En **C1** coloca el nombre de la variable **IMPUESTO**.

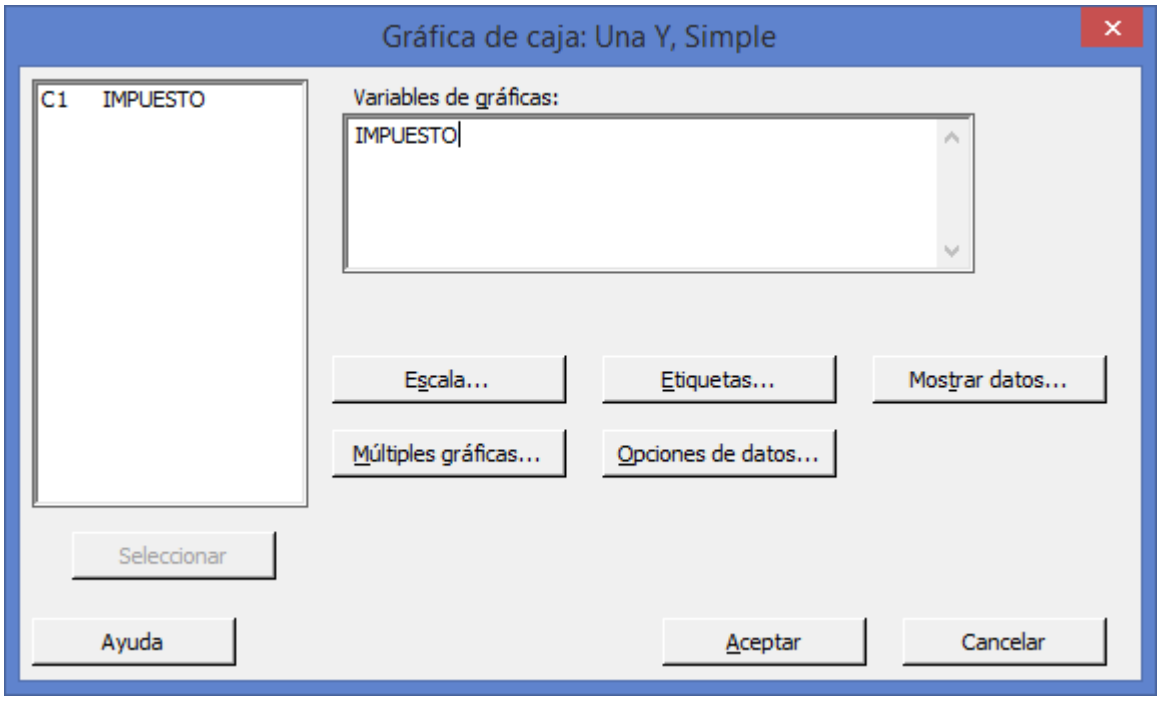
4 Teclea los datos de la tabla en forma vertical.



5 Elija **Gráfica > Gráfica de caja**.

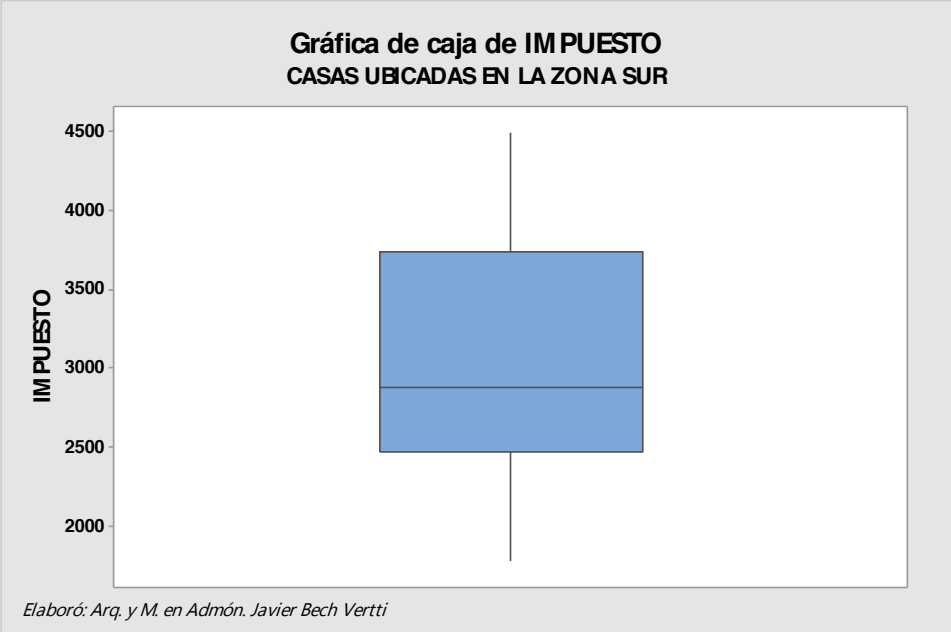
6 En **Una Y**, elija **Simple**. Haga clic en **Aceptar**.

7 En **Variables de graficas**, ingrese ***IMPUESTO***.



8 Haga clic en **Aceptar**.

Salida de la ventana Gráfica



Interpretación: En el diagrama anterior se puede observar que no existen casos o valores atípicos. La posición de la mediana dentro de la caja cargada al lado izquierdo nos puede sugerir un ligero sesgo a la derecha que debe ser tomado en cuenta para analizar la Normalidad de la distribución.

1.3.1.4.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.4.2**

DATOS **ATÍPICOS . DIAGRAMA DE CAJA Y BRAZOS SIMPLE.**

1.3.1.4.2 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **NORTE** que forman su cartera de ventas. De la base de datos **"BIENES1"** se seleccionan aleatoriamente a 30 de ellas ubicadas en la zona **NORTE** que incluyen 21 preguntas entre ellas la **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

30	33	30	34	41	35
35	38	30	44	35	37
35	37	15	40	28	34
24	35	35	35	14	25
22	31	31	38	30	33

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo

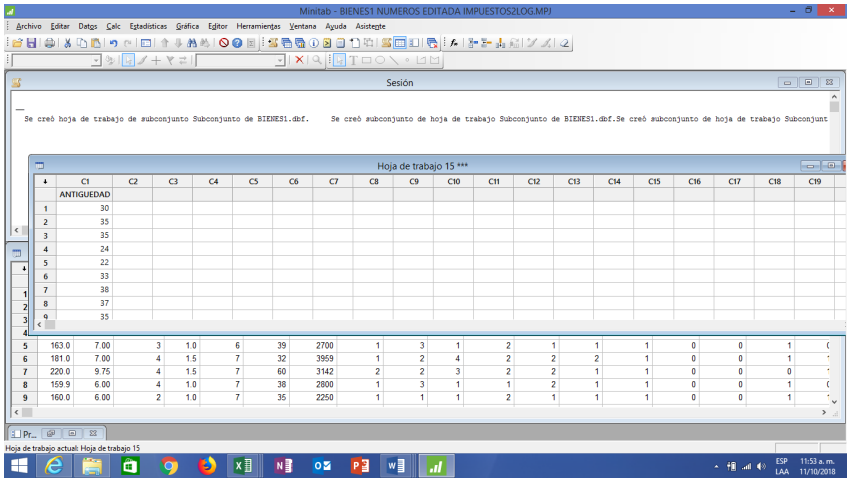
Para introducir los datos agrega una hoja de trabajo vacía al proyecto actual:

1 Elija **Archivo > Nuevo**.

2 Seleccione **Hoja de trabajo de Minitab** y haga clic en **Aceptar**.

3 En **C1** coloca el nombre de la variable **ANTIGUEDAD**.

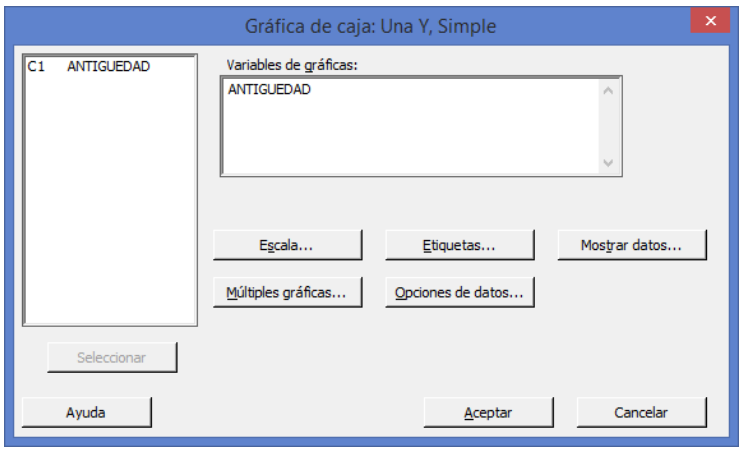
4 Teclea los datos de la tabla en forma vertical.



5 Elija **Gráfica > Gráfica de caja**.

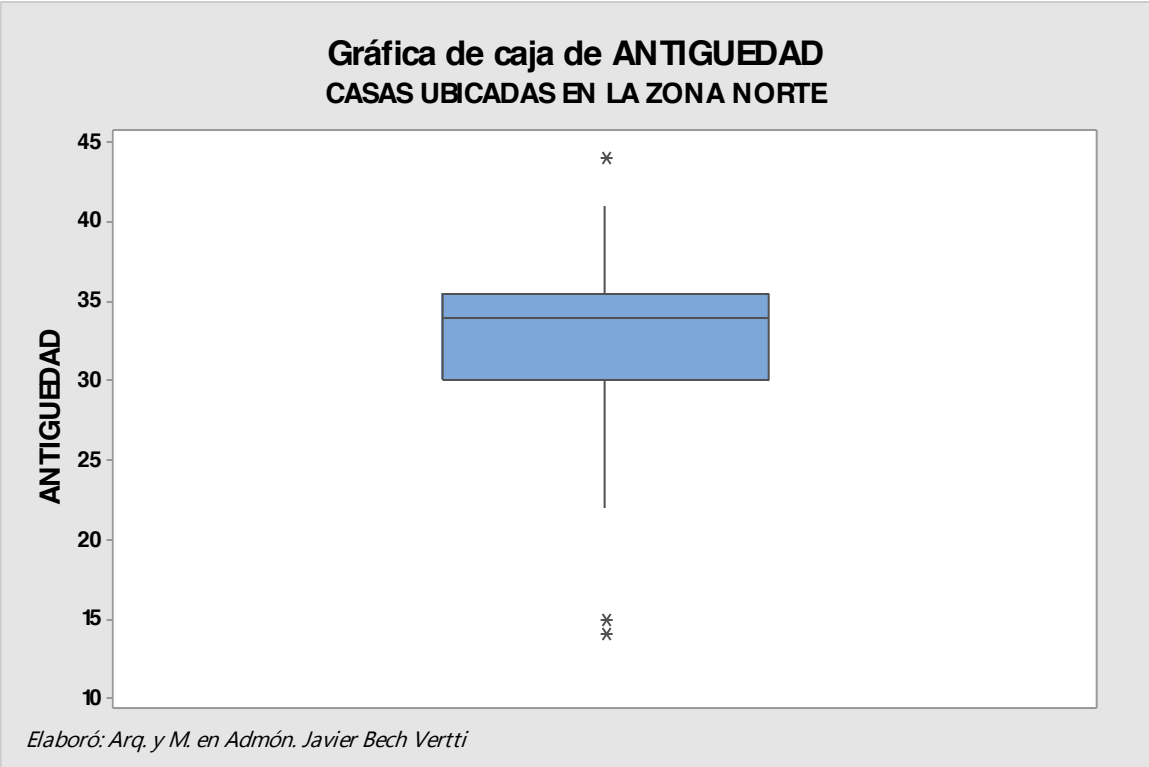
6 En **Una Y**, elija **Simple**. Haga clic en **Aceptar**.

7 En **Variables de graficas**, ingrese ***ANTIGUEDAD***.



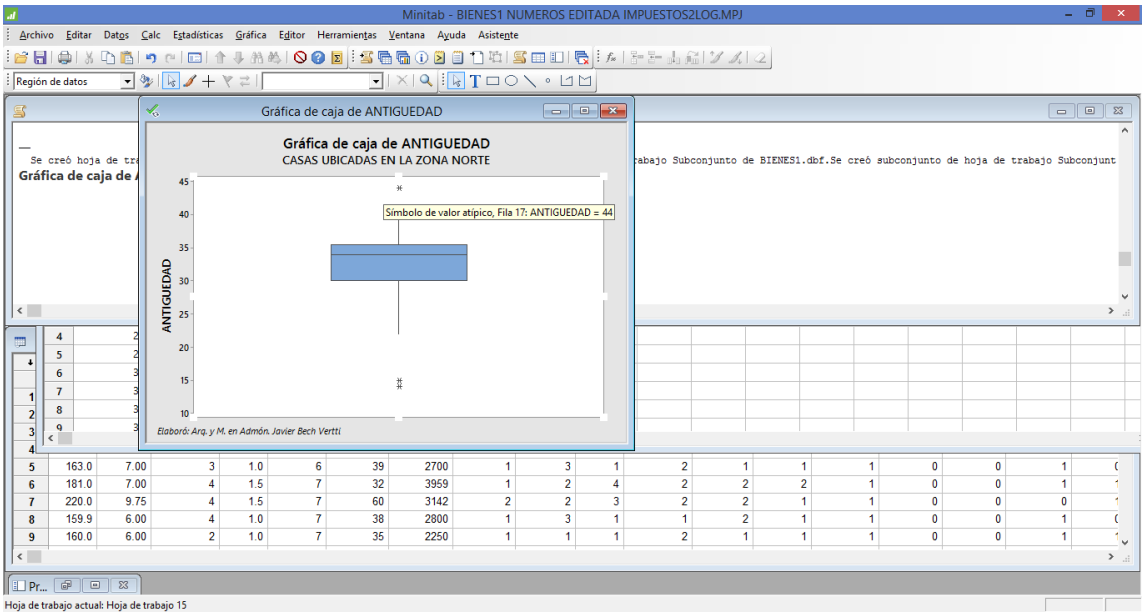
8 Haga clic en **Aceptar**.

Salida de la ventana Gráfica



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

Interpretación: En el diagrama anterior se puede observar que existen 3 valores atípicos que pueden ser identificados colocando el puntero del mouse en cada uno de ellos proporcionandonos la siguiente información:



Fila 24 Antiguedad 14

Fila 13 Antigüedad 15

Fila 17 Antigüedad 44

Ests valores ó datos ó casos atípicos deben ser estudiados con detalle para ver el tratamiento que se les van a dar en estudios subsecuentes ya que es posible que provoque un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la derecha de la caja sugiere un sesgo hacia la izquierda producto tal vez de los valores o casos atípicos.

1.3.1.4.3 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 1.3.1.4.3

DATOS ATÍPICOS . DIAGRAMA DE CAJA Y BRAZOS SIMPLE.

1.3.1.4.3 Una agencia de Bienes Raíces desea examinar la **ANTIGUEDAD** de las casas ubicadas en la zona **SUR** que forman su cartera de ventas. De la base de datos **"BIENES1"** se seleccionan aleatoriamente a 30 de ellas ubicadas en la zona **SUR** que incluyen 21 preguntas entre ellas la **ANTIGUEDAD** de las mismas. Cree una gráfica de caja y brazos para examinar la **distribución de los valores** de las misma y revisar la posibilidad de encontrar **valores atípicos** que pudieran influir negativamente en el estudio que se pretende realizar más adelante.

Los datos se muestran a continuación:

32	28	35	30	61	34
60	27	30	16	39	21
27	28	39	28	31	34
35	35	30	31	55	38
30	14	28	31	32	24

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo

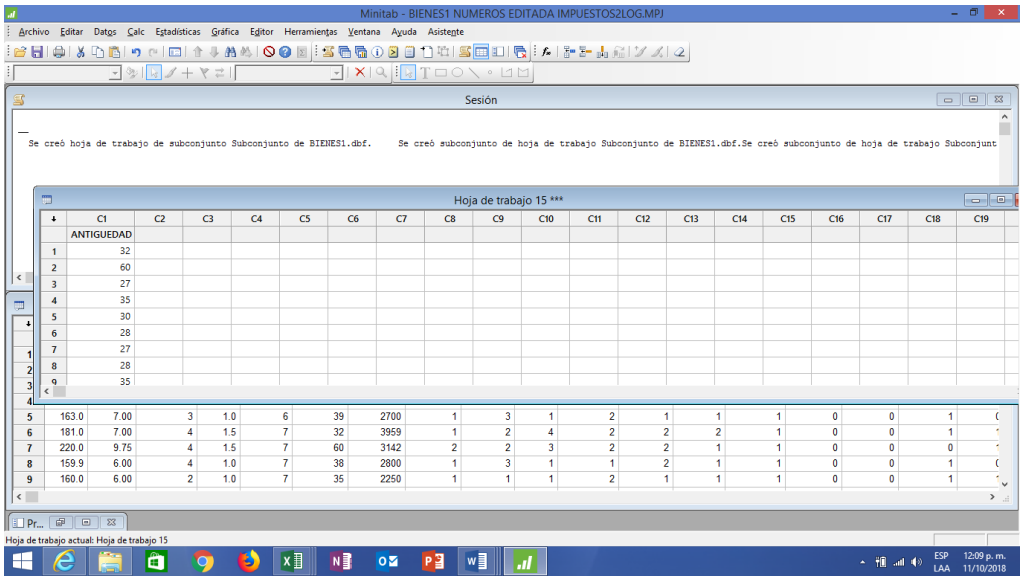
Para introducir los datos agrega una hoja de trabajo vacía al proyecto actual:

1 Elija **Archivo > Nuevo**.

2 Seleccione **Hoja de trabajo de Minitab** y haga clic en **Aceptar**.

3 En **C1** coloca el nombre de la variable **ANTIGUEDAD**.

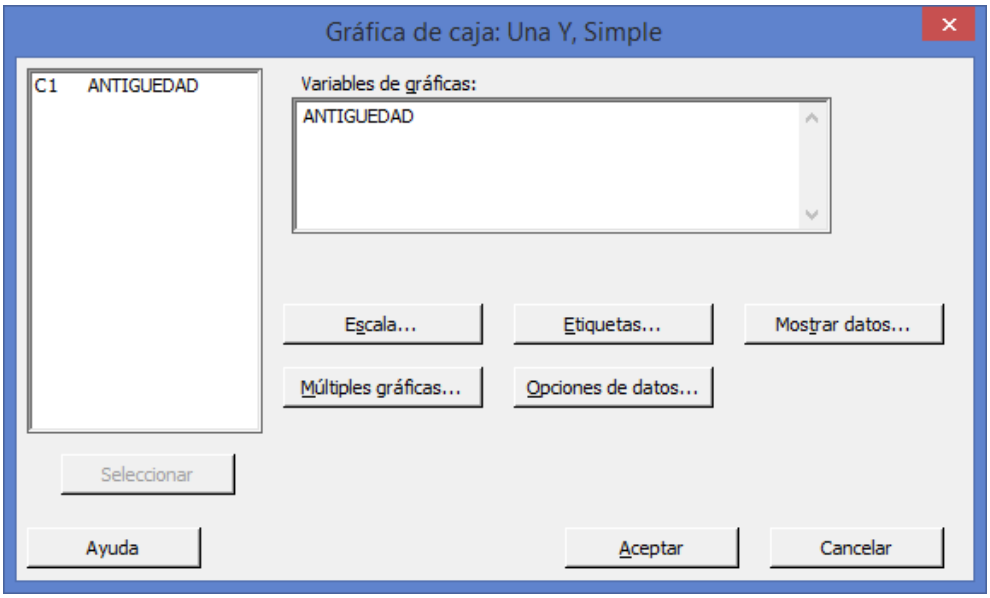
4 Teclea los datos de la tabla en forma vertical.



5 Elija **Gráfica > Gráfica de caja**.

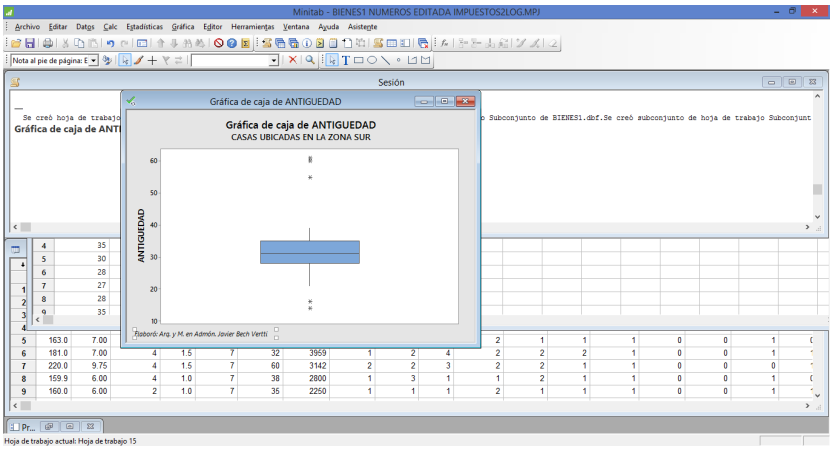
6 En **Una Y**, elija **Simple**. Haga clic en **Aceptar**.

7 En **Variables de graficas**, ingrese ***ANTIGUEDAD***.

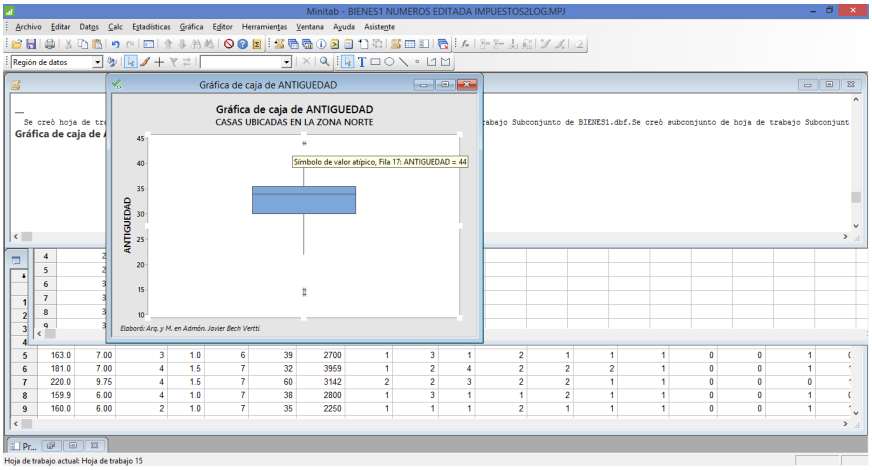


8 Haga clic en **Aceptar**.

Salida de la ventana Gráfica



Interpretación: En el diagrama anterior se puede observar que existen 3 valores atípicos que pueden ser identificados colocando el puntero del mouse en cada uno de ellos proporcionandonos la siguiente información:



Fila 10 Antiguedad 14

Fila 17 Antigüedad 16

Fila 24 Antigüedad 55

Fila 2 Antigüedad 60

Fila 21 Antigüedad 61

Estos valores ó datos ó casos atípicos deben ser estudiados con detalle para ver el tratamiento que se les van a dar en estudios subsecuentes ya que es posible que provoque un sesgo rompiendo algunos supuestos entre ellos el de Normalidad (la posición de la mediana, a la izquierda de la caja sugiere un sesgo hacia la derecha producto tal vez de los valores o casos atípicos.

1.3.1.5.1 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 1.3.1.5.1

DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE

1.3.1.5.1 Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **SIN SÓTANO** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

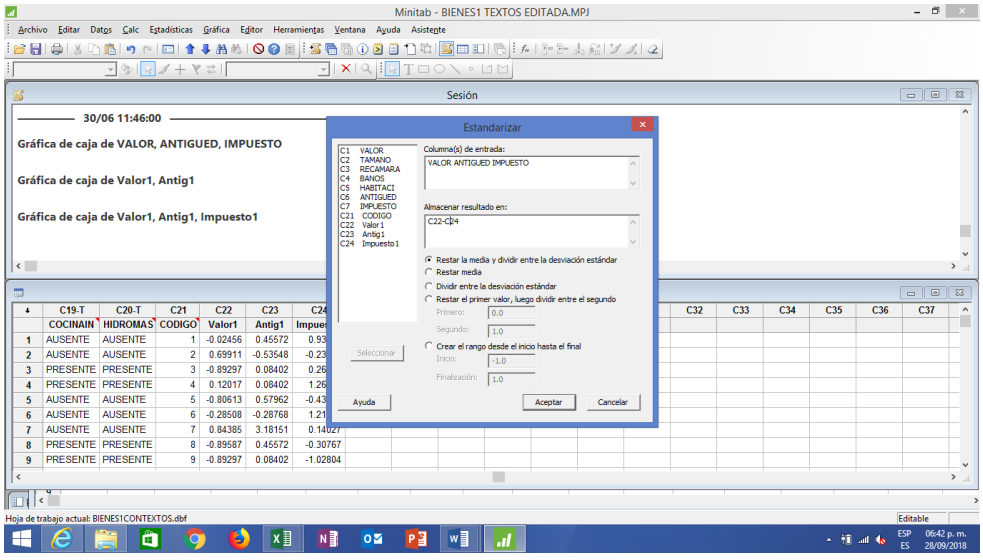
Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las caas que no tienen SOTANO** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables de la siguiente manera:

1 Con la hoja de trabajo con las casas que **no** tienen **SÓTANO** seleccionar **Calc > Estandarizar**.

2 En **Columnas de entrada**, ingrese las columnas que desea estandarizar. En este caso ingrese **VALOR, ANTIGUEDAD E IMPUESTOS.**

3 En **Almacenar resultados en**, ingrese las columnas en las que desea almacenar los datos estandarizados. En este caso escriba **C22-C24**, haga clic en **Aceptar**



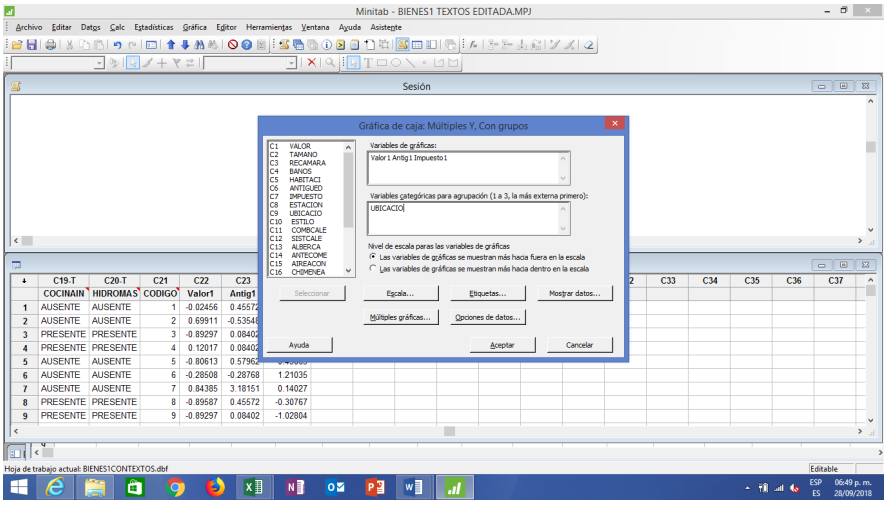
4 Los datos estandarizados aparecerán en las columas **22,23 y 24** de la base original **"BIENES1"**. Colocar el nombre de las variables estantarazadas como **Valor 1, Antig1 e Impuesto 1** para identificarlas.

5 Elija **Gráfica > Gráfica de caja**.

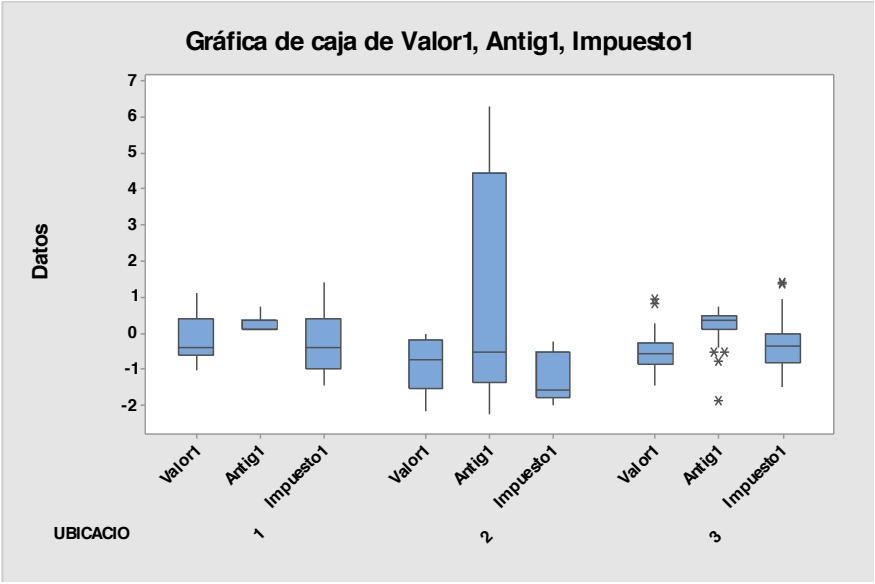
6 En **Múltiples Y**, elija **Con grupos**. Haga clic en **Aceptar**.

7 En **Variables de gráficas**, ingrese ***Valor1, Antig1 e Impuesto1***

8 En **Variables categóricas para agrupación (1 a 3, la más externa primero)**, ingrese ***UBICACIÓN.***



9 En **Nivel de escala para variables de gráficas**, elija **Las variables de gráficas se muestran más hacia dentro en la escala**. Haga clic en **Aceptar**.



Interpretación de los resultados

La gráfica de caja muestra:

· Para las casas **SIN SÓTANO**, en general en todos los niveles de **UBICACIÓN** de las casas las tres variables no muestran valores atípicos con excepción de **Valor1**, **Antig1 e Impuesto1** en la zona **ESTE.**

· Para las casas ubicadas en la zona **ESTE** en la variable V**alor1** se detectan dos valores atípicos en la parte superior, en **Antig1** se detectan varios valores atípicos en la parte o extremo inferior y en **Impuestos1** se detecta un valor atípico en el extremo superior. Como los valores atípicos detectados en la zona **ESTE** se encuentran en un extremo habría que estudiar el sesgo producido que pudiera influir en la Normalidad de los datos.

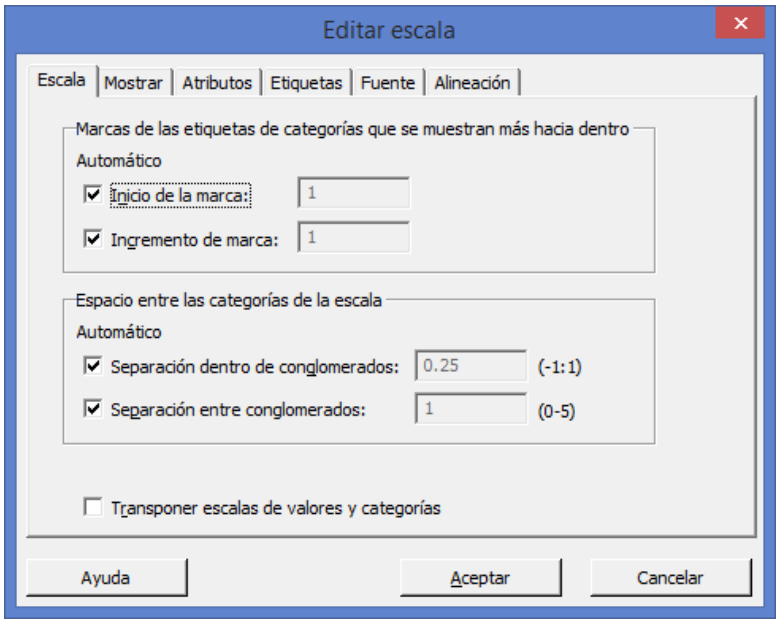
Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

Asimismo se recomienda realizar análisis estadísticos, tales como **MANOVA balanceado**, para examinar más a fondo la relación entre los factores.

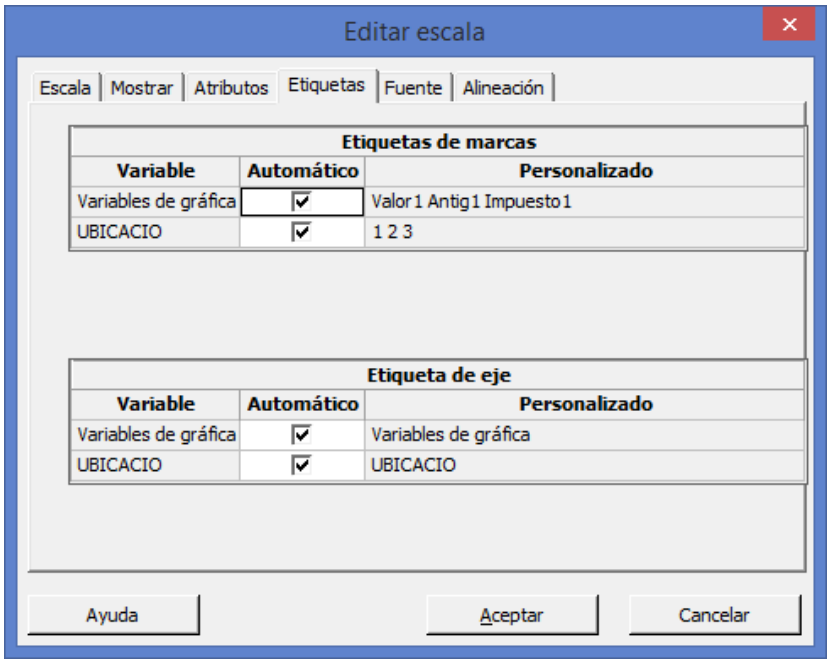
Para obtener información precisa de Q1, mediana, Q3, rango intercuartil, bigotes o brazos y N, coloque el cursor sobre cualquier parte de la gráfica de caja. Para ver el valor del eje Y y el número de fila de un valor atípico, coloque el cursor sobre el valor atípico.

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN Y Etiquetado del subtitulo del nivel SIN SÓTANO (Ausente):

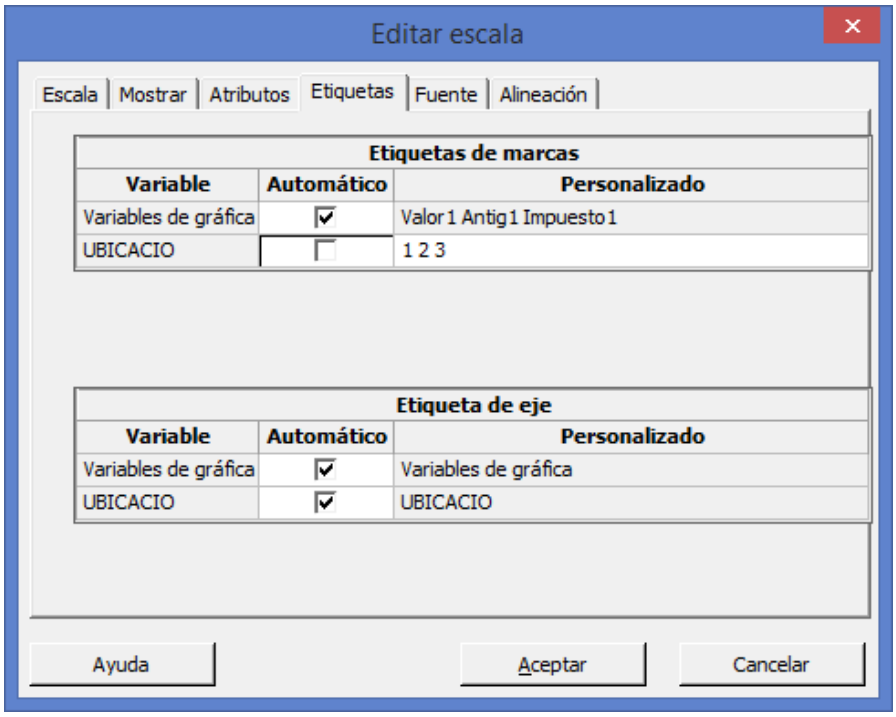
1 Haga doble clic en el número 1 en el renglón de UBICACIÓN para que Minitab despliegue la ventana ***Editar Escala.***



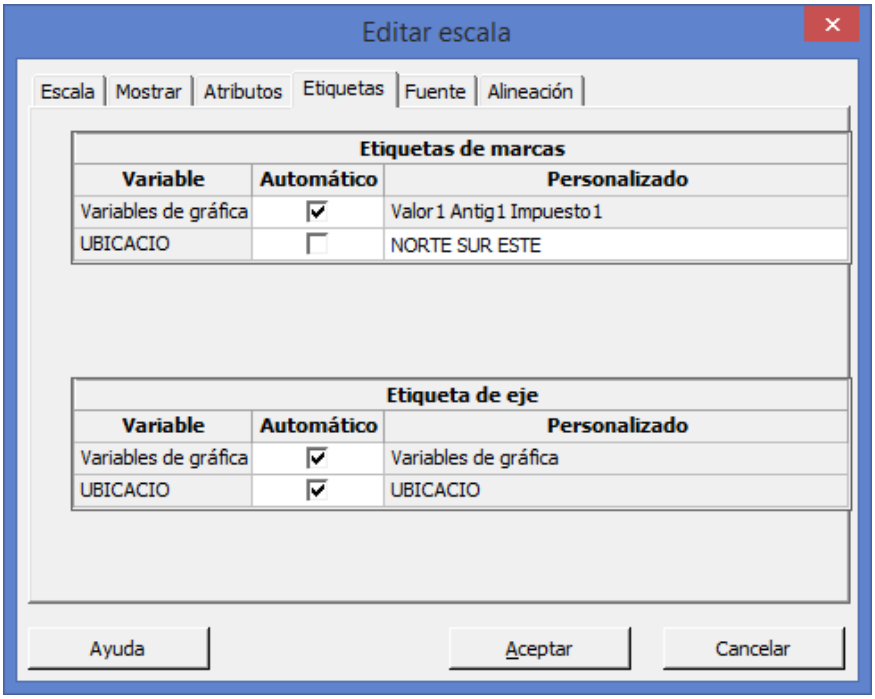
2 Seleccionar la pestaña **Etiquetas**



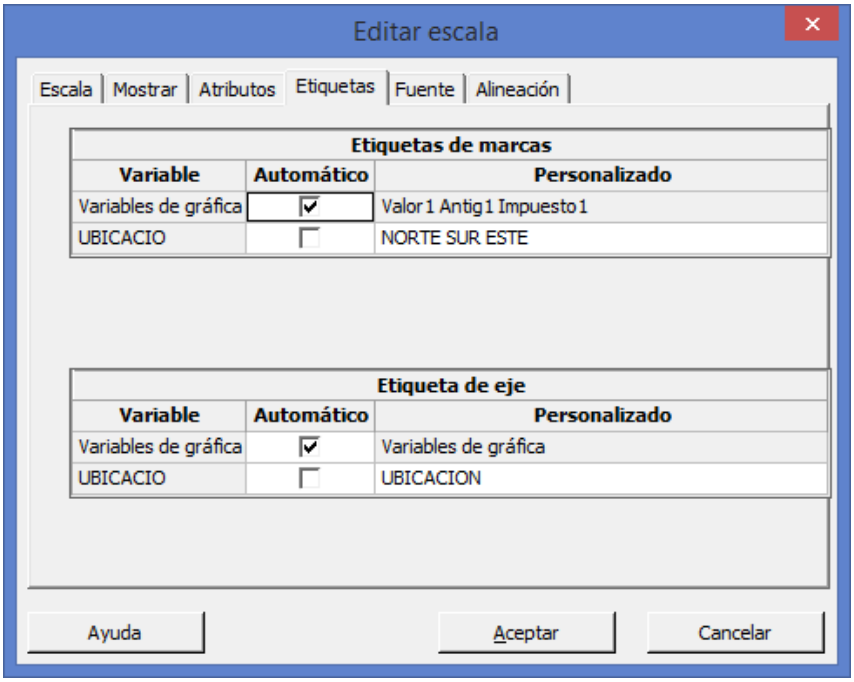
3 Desactivar la casilla **UBICACIÓN**



4 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE.**

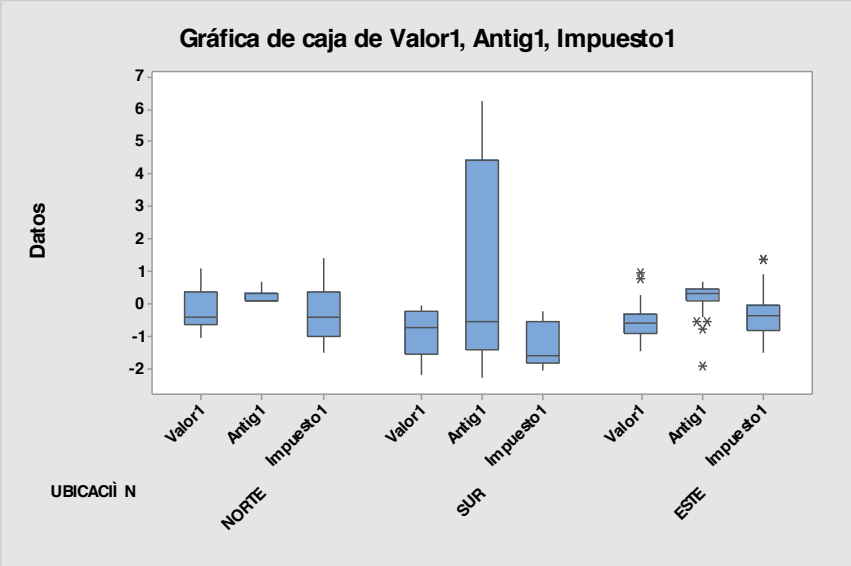


5 En **Etiqueta de eje** , si es el caso, desactivar la casilla de **UBICACIO** y escribir **UBICACIÓN.**



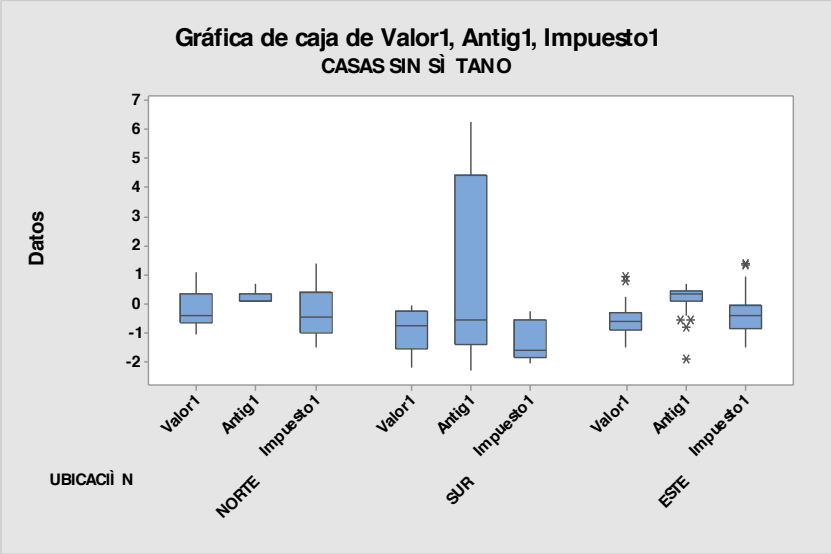
6 Haga clic en **Aceptar**.

7 Aparece la gráfica debidamente etiquetada.



8 Coloque el mouse en el título de la grafica y presione el botón derecho del mouse. Seleccione la opción Agregar y la opción Subtitulo y se escribe **CASAS SIN SÓTANO**.

9 Haga clic en Aceptar.
10 Aparece la gráfica debidamente etiquetada.



1.3.1.5.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.1.5.2**

DATOS **ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE**

1.3.1.5.2 Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **SIN CHIMENEA** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

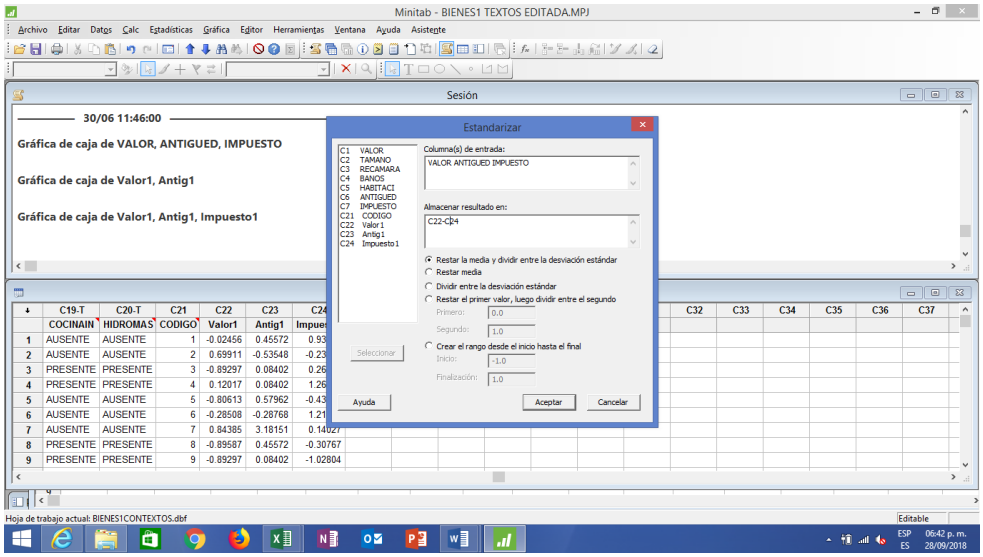
Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las casas que no tienen CHIMENEA** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables de la siguiente manera:

1 Con la hoja de trabajo con las casas que **no** tienen **CHIMENEA** seleccionar **Calc > Estandarizar**.

2 En **Columnas de entrada**, ingrese las columnas que desea estandarizar. En este caso ingrese **VALOR, ANTIGUEDAD E IMPUESTOS.**

3 En **Almacenar resultados en**, ingrese las columnas en las que desea almacenar los datos estandarizados. En este caso escriba **C22-C24**, haga clic en **Aceptar**



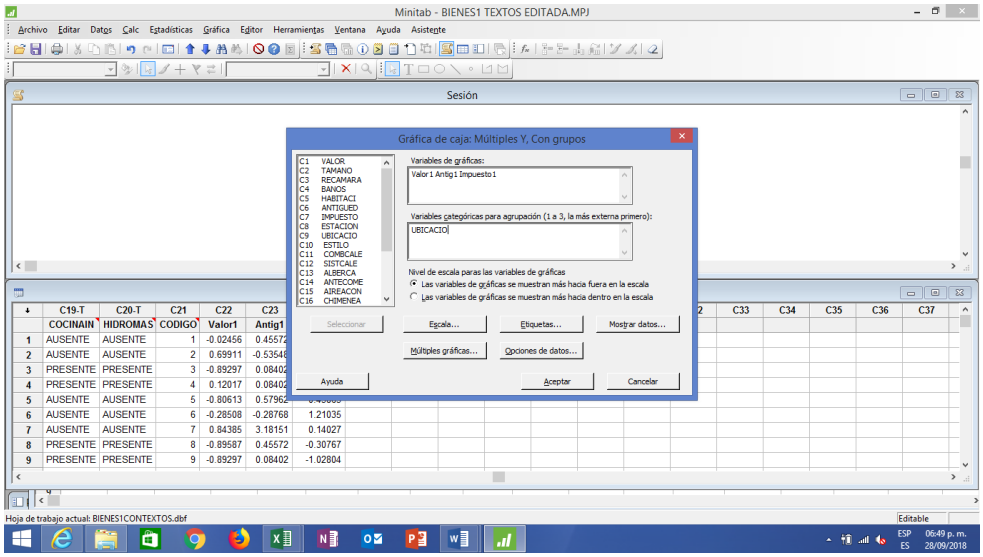
4 Los datos estandarizados aparecerán en las columas **22,23 y 24** de la base original **"BIENES1"**. Colocar el nombre de las variables estantarazadas como **Valor 1, Antig1 e Impuesto 1** para identificarlas.

5 Elija **Gráfica > Gráfica de caja**.

6 En **Múltiples Y**, elija **Con grupos**. Haga clic en **Aceptar**.

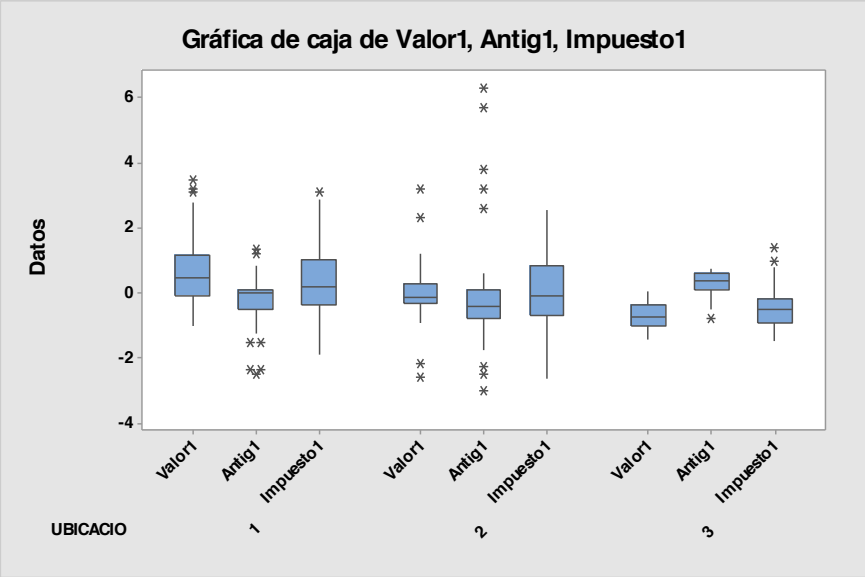
7 En **Variables de gráficas**, ingrese ***Valor1, Antig1 e Impuesto1***

8 En **Variables categóricas para agrupación (1 a 3, la más externa primero)**, ingrese ***UBICACIÓN.***



9 En **Nivel de escala para variables de gráficas**, elija **Las variables de gráficas se muestran**

más hacia dentro en la escala. Haga clic en **Aceptar**.



Interpretación de los resultados

La gráfica de caja muestra:

· Para las casas **SIN CHIMENEA**, en general en todos los niveles de **UBICACIÓN** de las casas las tres variables muestran valores atípicos con excepción de **Impuesto1** en la zona **SUR** y **Valor 1** en la zona **ESTE.**

· Para las casas ubicadas en la zona **NORTE** en la variable V**alor1** se detectan valores atípicos en la parte superior donde habría que estudiar el sesgo producido que pudiera influir en la Normalidad de los datos así como en **Impuestos1** que se detecta un valor atípico en el extremo superior. En **Antig1** se detectan varios valores atípicos en ambos extremos. En la zona **SUR** se detectan valores atípicos en ambos extremos de **Valor 1 y de Antig 1** siendo estos últimos más numerosos y extremos sobre todo en la parte superior. Como los valores atípicos detectados en la zona **ESTE** se encuentran en un solo extremo tanto en **Antig 1** como en **Impuestos1** habría que estudiar el sesgo producido que pudiera influir en la Normalidad de los datos.

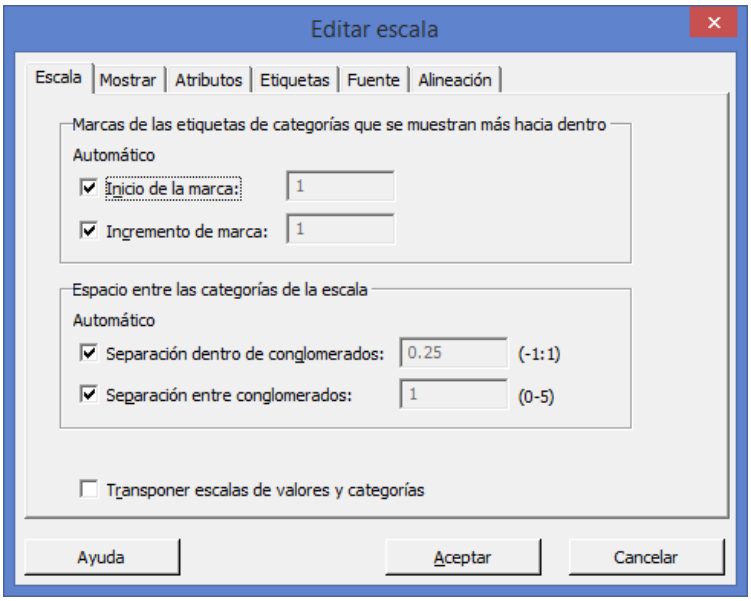
Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

Asimismo se recomienda realizar análisis estadísticos, tales como **MANOVA balanceado**, para examinar más a fondo la relación entre los factores.

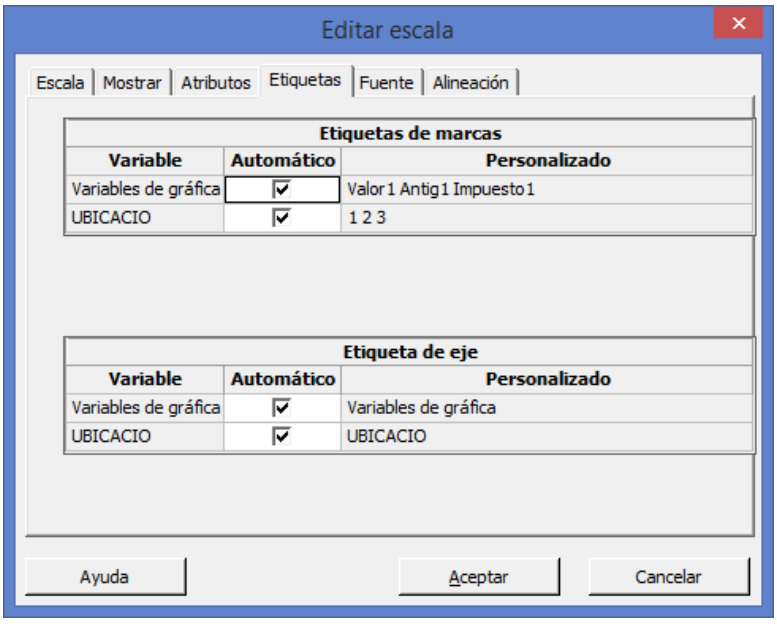
Para obtener información precisa de Q1, mediana, Q3, rango intercuartil, bigotes o brazos y N, coloque el cursor sobre cualquier parte de la gráfica de caja. Para ver el valor del eje Y y el número de fila de un valor atípico, coloque el cursor sobre el valor atípico.

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN Y Etiquetado del subtitulo del nivel SIN CHIMENEA (Ausente):

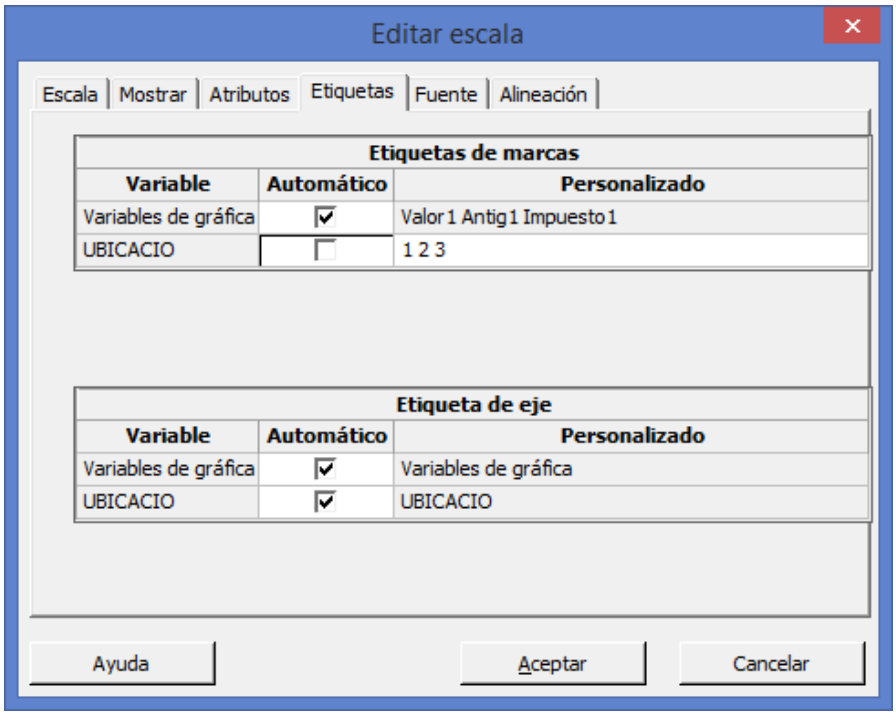
1 Haga doble clic en el número 1 en el renglón de UBICACIÓN para que Minitab despliegue la ventana ***Editar Escala.***



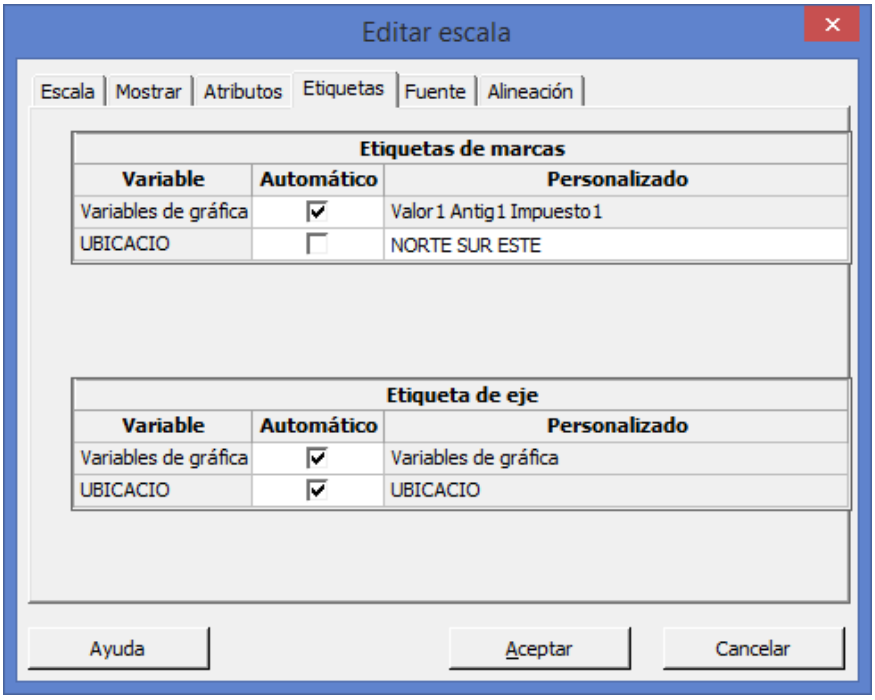
2 Seleccionar la pestaña **Etiquetas**



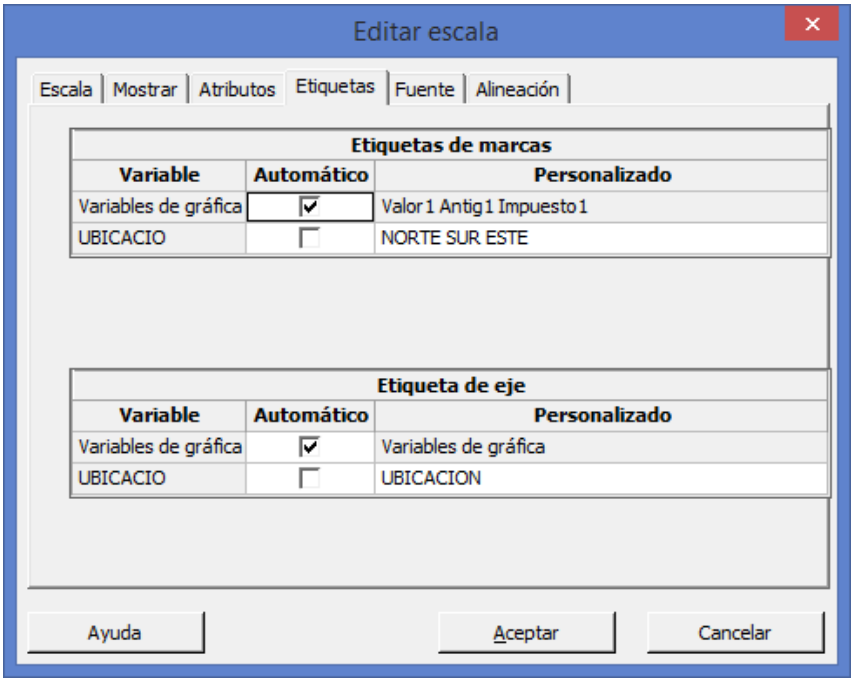
3 Desactivar la casilla **UBICACIÓN**



4 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE.**

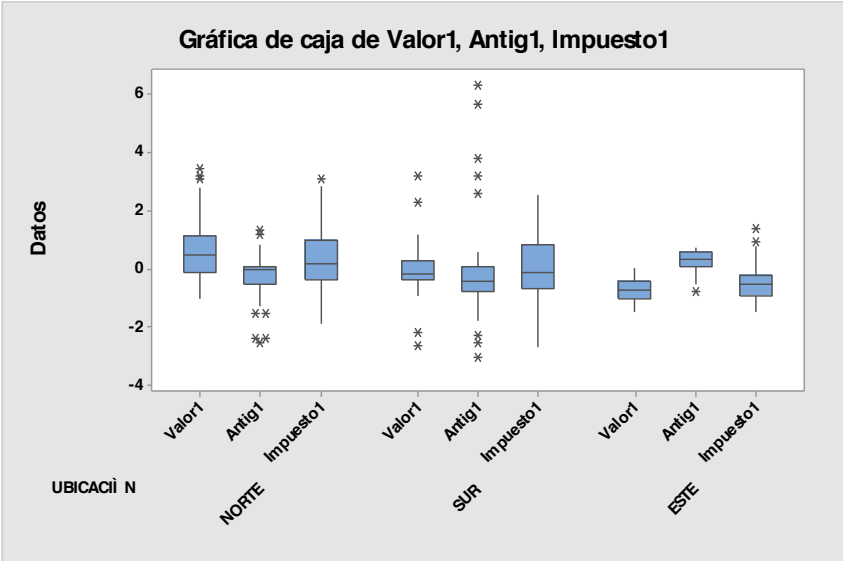


5 En **Etiqueta de eje** , si es el caso, desactivar la casilla de **UBICACIO** y escribir **UBICACIÓN.**

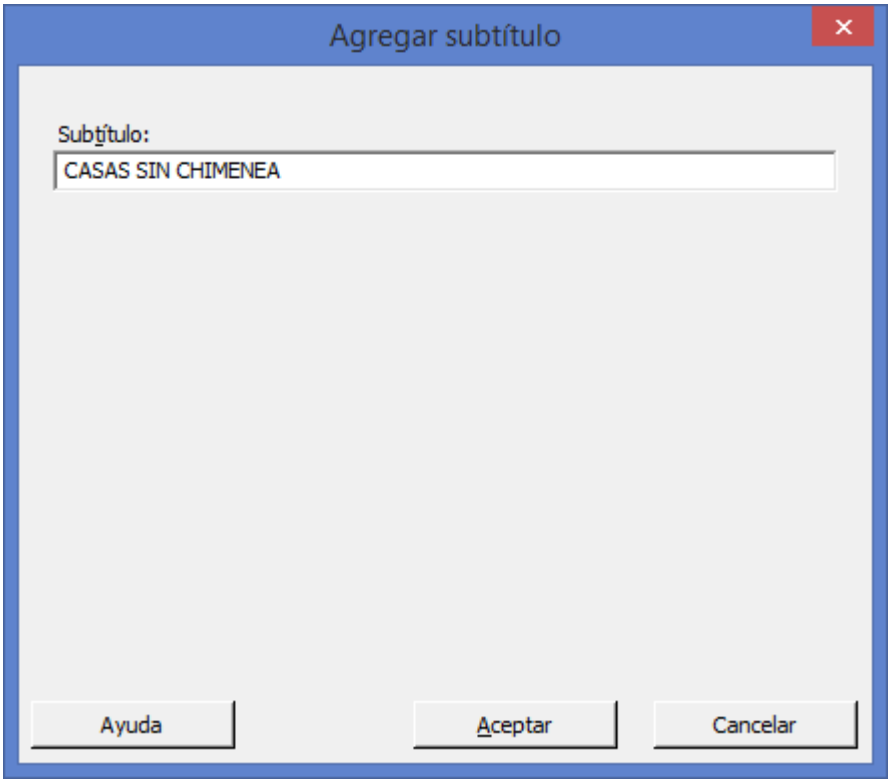


6 Haga clic en **Aceptar**.

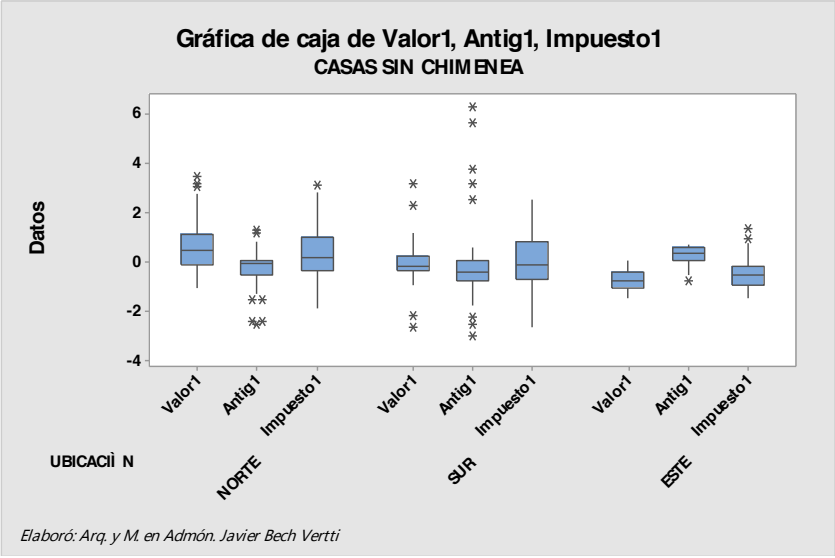
7 Aparece la gráfica debidamente etiquetada.



8 Coloque el mouse en el título de la grafica y presione el botón derecho del mouse. Seleccione la opción Agregar y la opción Subtitulo y se escribe **CASAS SIN CHIMENEA**



9 Haga clic en Aceptar.
10 Aparece la gráfica debidamente etiquetada.



1.3.1.5.3 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 1.3.1.5.3

DATOS ATÍPICOS. DIAGRAMA DE CAJA Y BRAZOS MÚLTIPLE

1.3.1.5.3 Se reciben datos sobre una encuesta **"BIENES1"** realizada a 233 casas para realizar diferentes estudios sobre ella. Antes de analizar los datos con mayor profundidad y aplicar alguna técnica multivariada, se desea identificar la posible existencia de valores atípicos que pudieran influenciar negativamente el estudio. Para su identificación se desea utilizar diagramas de caja múltiples que incluyan las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** de las casas **SIN TINA DE HIDROMASAJE** en los 3 niveles de la **UBICACIÓN DE LAS CASAS**.

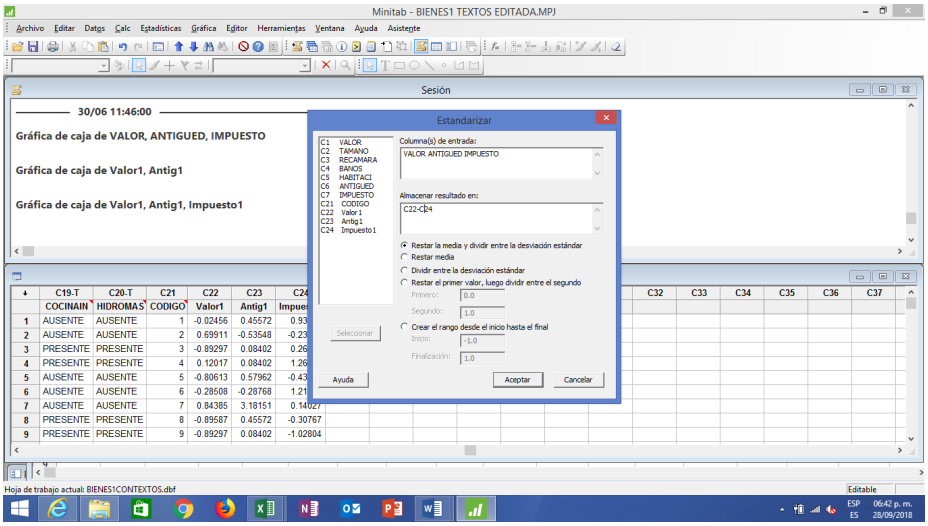
Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo.

Nota importante: Primero hay que crear un **subconjunto de hoja de trabajo con las casas que no tienen TINA DE HIDROMASAJE** y después, como las variables **VALOR, ANTIGÜEDAD E IMPUESTOS** están medidas en diferentes escalas se deben **estandariza**r las variables de la siguiente manera:

1 Con la hoja de trabajo con las casas que **no** tienen **TINA DE HIDROMASAJE** seleccionar **Calc > Estandarizar**.

2 En **Columnas de entrada**, ingrese las columnas que desea estandarizar. En este caso ingrese **VALOR, ANTIGUEDAD E IMPUESTOS.**

3 En **Almacenar resultados en**, ingrese las columnas en las que desea almacenar los datos estandarizados. En este caso escriba **C22-C24**, haga clic en **Aceptar**



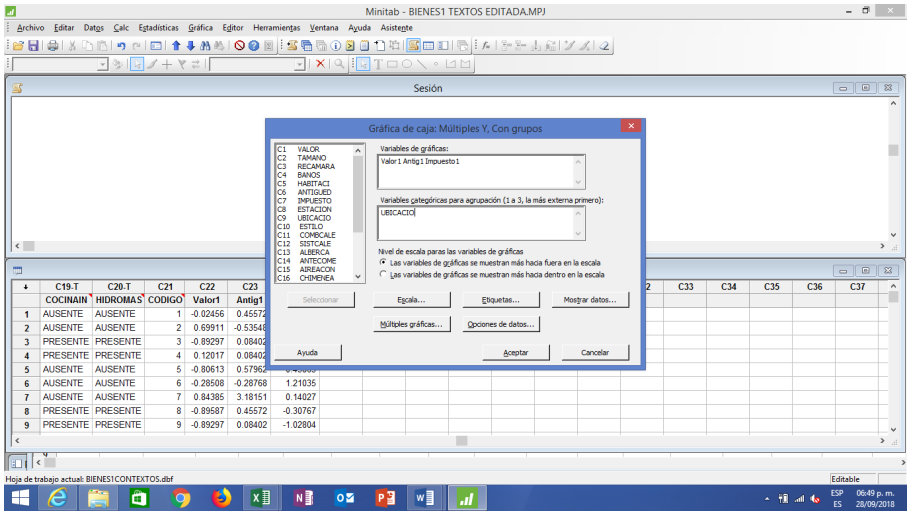
4 Los datos estandarizados aparecerán en las columas **22,23 y 24** de la base original **"BIENES1"**. Colocar el nombre de las variables estantarazadas como **Valor 1, Antig1 e Impuesto 1** para identificarlas.

5 Elija **Gráfica > Gráfica de caja**.

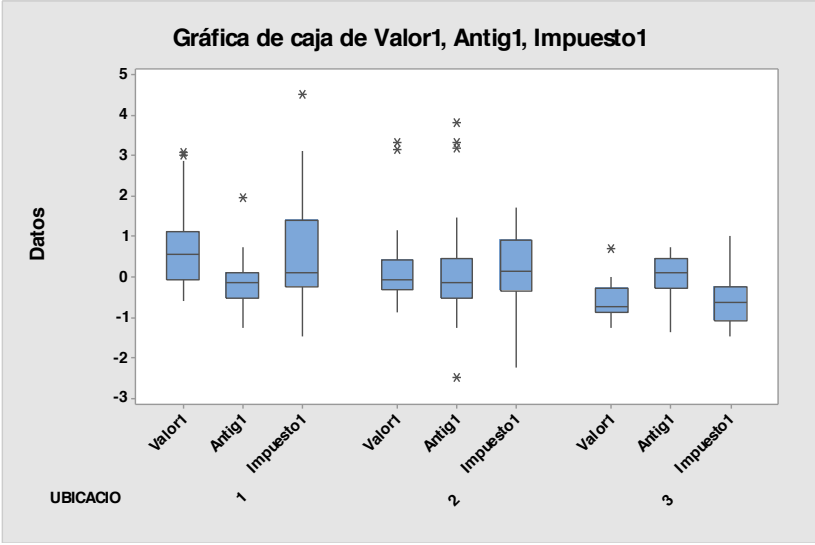
6 En **Múltiples Y**, elija **Con grupos**. Haga clic en **Aceptar**.

7 En **Variables de gráficas**, ingrese ***Valor1, Antig1 e Impuesto1***

8 En **Variables categóricas para agrupación (1 a 3, la más externa primero)**, ingrese ***UBICACIÓN.***



9 En **Nivel de escala para variables de gráficas**, elija **Las variables de gráficas se muestran más hacia dentro en la escala**. Haga clic en **Aceptar**.



Interpretación de los resultados

La gráfica de caja muestra:

· Para las casas **SIN TINA DE HIDROMASAJE**, en general en todos los niveles de **UBICACIÓN** de las casas las tres variables muestran valores atípicos con excepción de **Impuesto1** en la zona **SUR** y **Antig1 e Impuestos1** en la zona **ESTE.**

· Para las casas ubicadas en la zona **NORTE** en las tres se detectan valores atípicos en la parte superior donde habría que estudiar el sesgo producido que pudiera influir en la Normalidad de los datos. En la zona **SUR** se detectan valores atípicos en **Valor1** donde habría que estudiar el sesgo producido que pudiera influir en la Normalidad de los datos y en ambos extremos de **Antig1**. Como los valores atípicos detectados en la zona **ESTE** se encuentran en un solo extremo de **Valor1** habría que estudiar el sesgo producido que pudiera influir en la Normalidad de los datos.

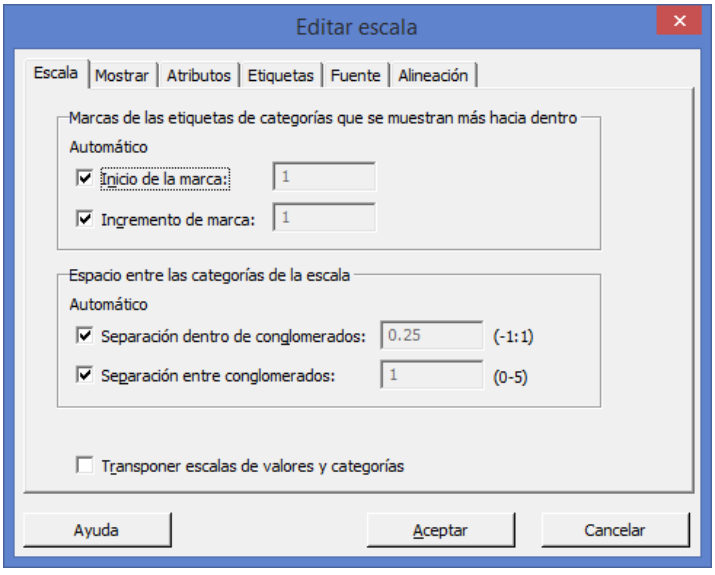
Se recomienda **estudiar minuciosamente cada uno de los valores atípicos** para ver que tipo de tratamiento se les puede dar para disminuir el riesgo potencial de influencia en el estudio multivariado que se pudiera realizar más adelante.

Asimismo se recomienda realizar análisis estadísticos, tales como **MANOVA balanceado**, para examinar más a fondo la relación entre los factores.

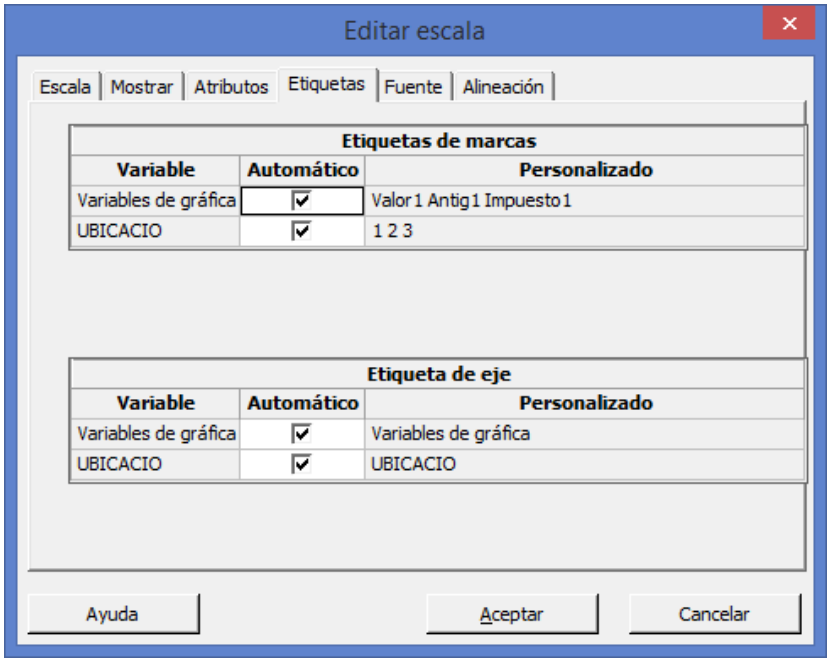
Para obtener información precisa de Q1, mediana, Q3, rango intercuartil, bigotes o brazos y N, coloque el cursor sobre cualquier parte de la gráfica de caja. Para ver el valor del eje Y y el número de fila de un valor atípico, coloque el cursor sobre el valor atípico.

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN Y Etiquetado del subtitulo del nivel SIN TINA DE HIDROMASAJE (Ausente):

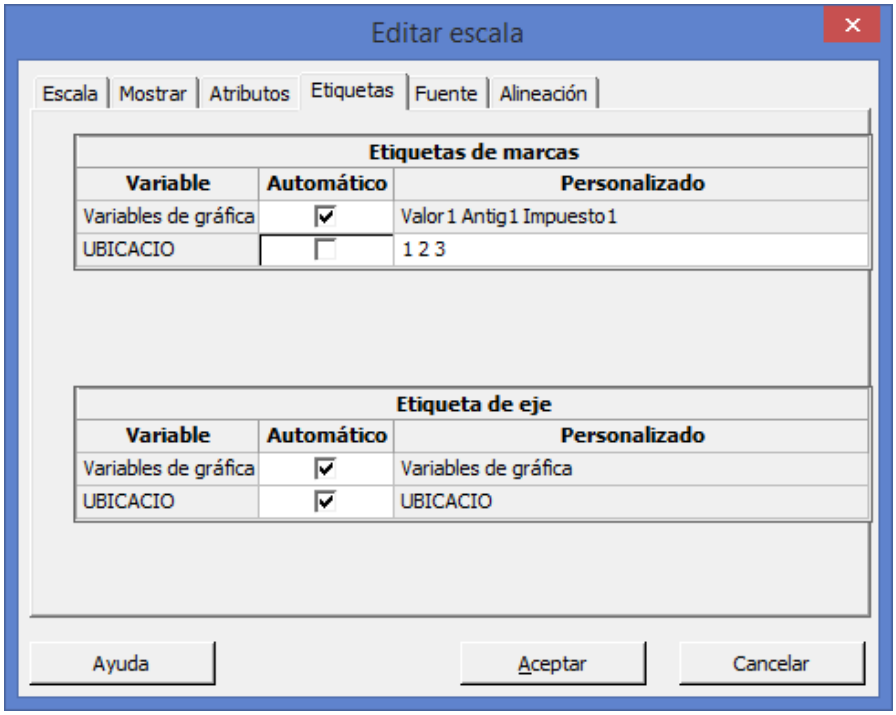
1 Haga doble clic en el número 1 en el renglón de UBICACIÓN para que Minitab despliegue la ventana ***Editar Escala.***



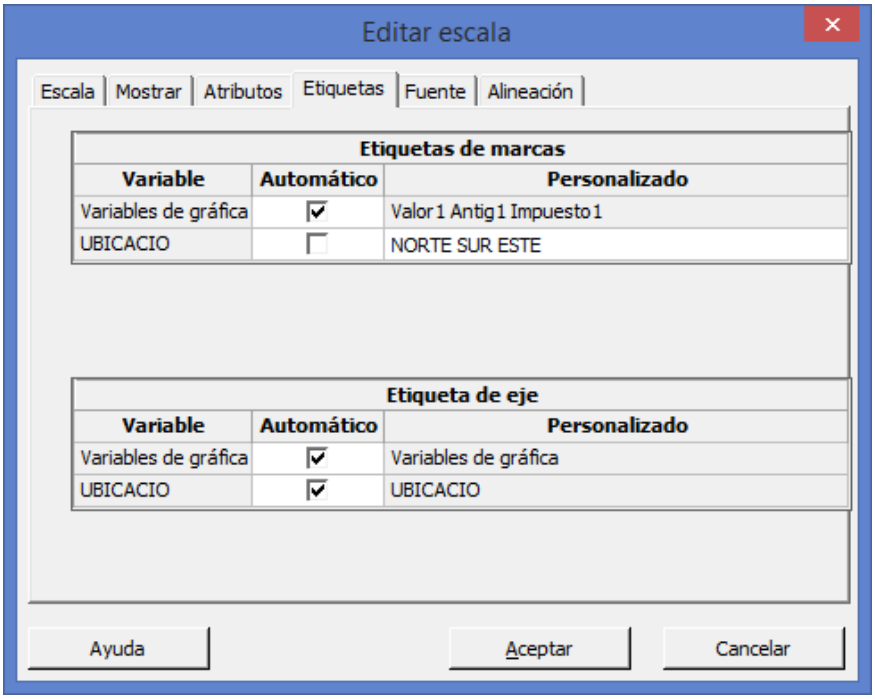
4 Seleccionar la pestaña **Etiquetas**



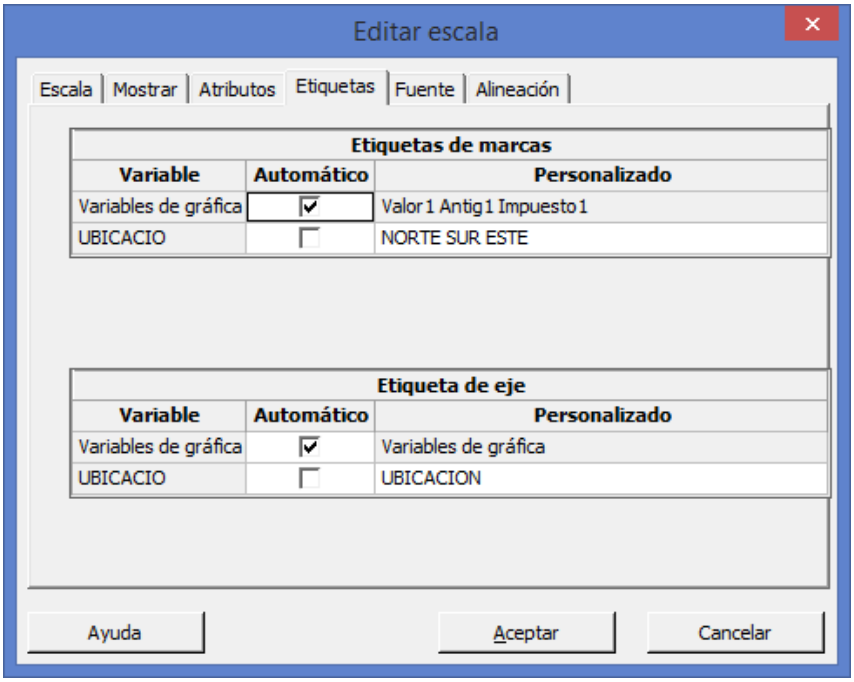
5 Desactivar la casilla **UBICACIÓN**



6 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE.**

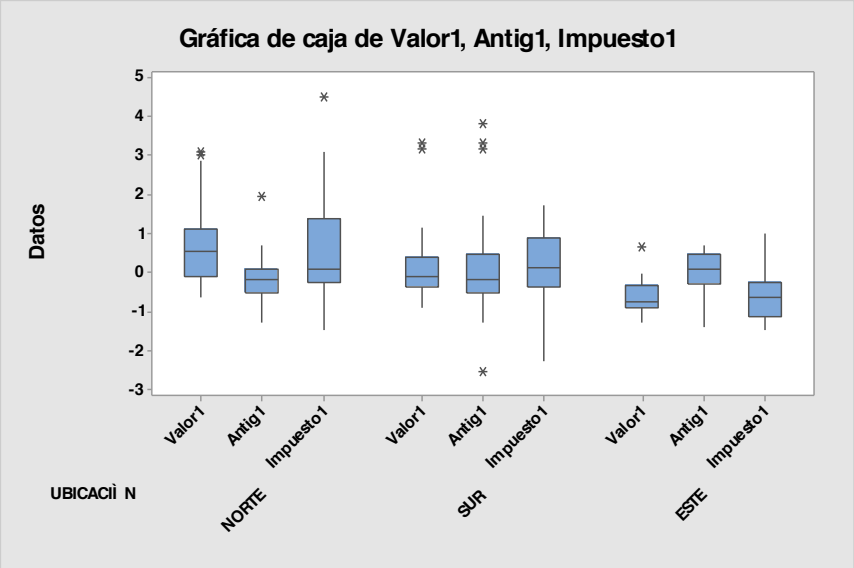


7 En **Etiqueta de eje** , si es el caso, desactivar la casilla de **UBICACIO** y escribir **UBICACIÓN.**

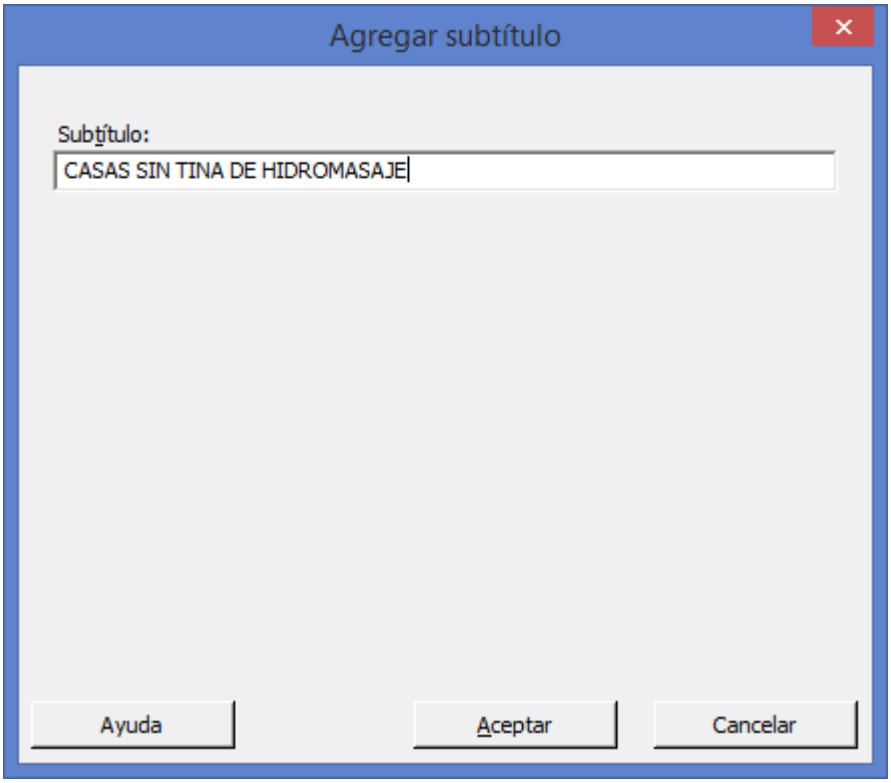


8 Haga clic en **Aceptar**.

9 Aparece la gráfica debidamente etiquetada.

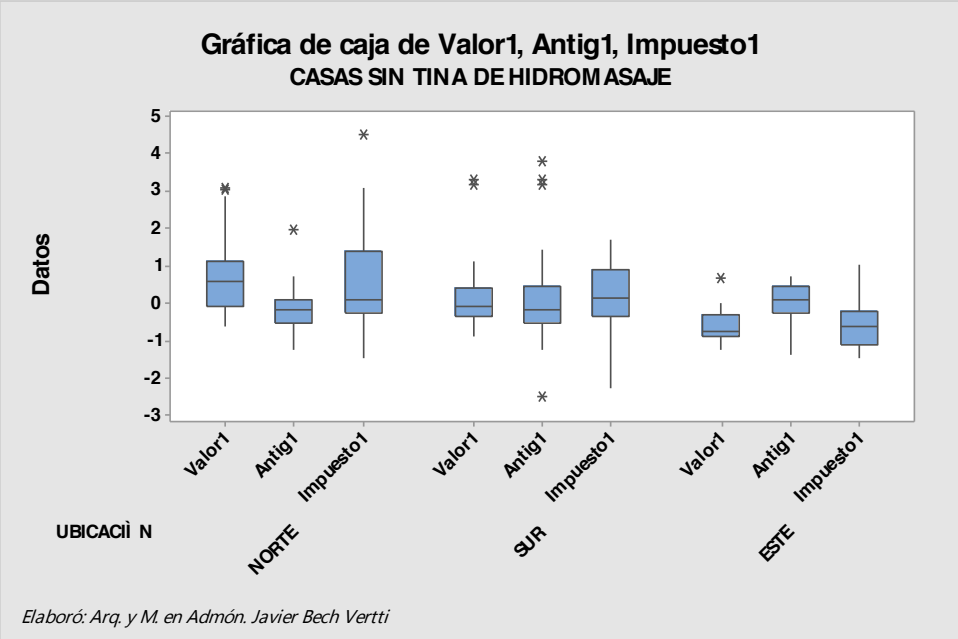


10 Coloque el mouse en el título de la grafica y presione el botón derecho del mouse. Seleccione la opción Agregar y la opción Subtitulo y se escribe **CASAS SIN TINA DE HIDROMASAJE**



11 Haga clic en Aceptar.

12 Aparece la gráfica debidamente etiquetada.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

1.3.2.1.1 EJERCICIO DE **REFUERZO**

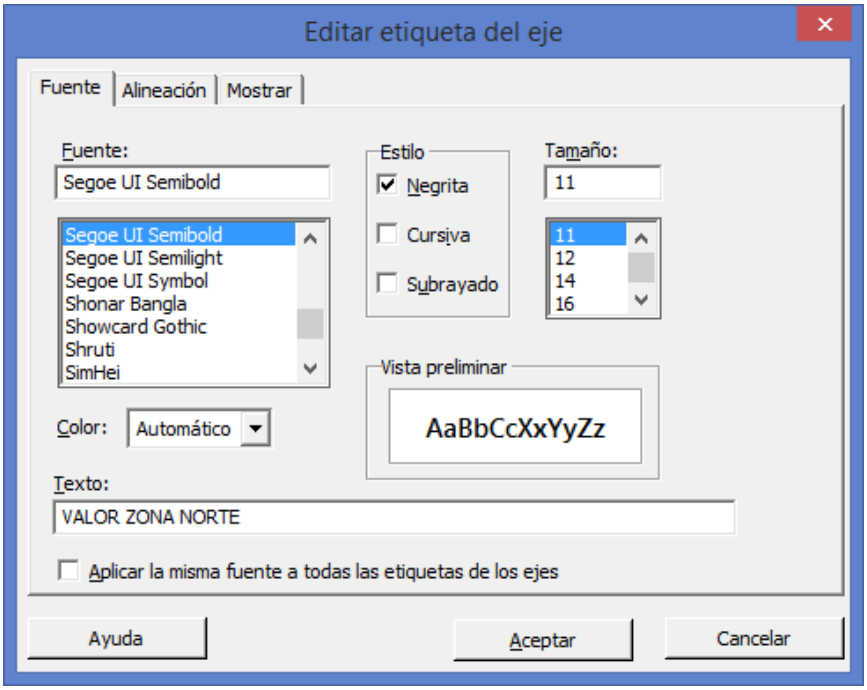
EJERCICIO DE **REFUERZO 1.3.2.1.1**

PRUEBA **DE NORMALIDAD. PRUEBA DE ANDERSON DARLING.**

1.3.2.1.1 Se midió el **VALOR** de las casas en la zona **NORTE** en la base de datos "**BIENES1"** que se construyó con anterioridad.

Utilizando el paquete **Minitab** se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas en la zona **NORTE** e interprete los resultados.

Nota: Etiquete la gráfica dando doble clic en **VALOR** para que aparezca la ventana ***Editar etiqueta de eje*** y en la parte inferior donde dice ***Texto*** escriba **VALOR ZONA NORTE**. Haga clic en **Aceptar.**



1 Abra la base de datos que creó **"BIENES1"** y cree un subconjunto de hoja de trabajo para las casas ubicadas en el **NORTE**

2 Elija **Estadísticas > Estadísticas básicas > Prueba de normalidad**.

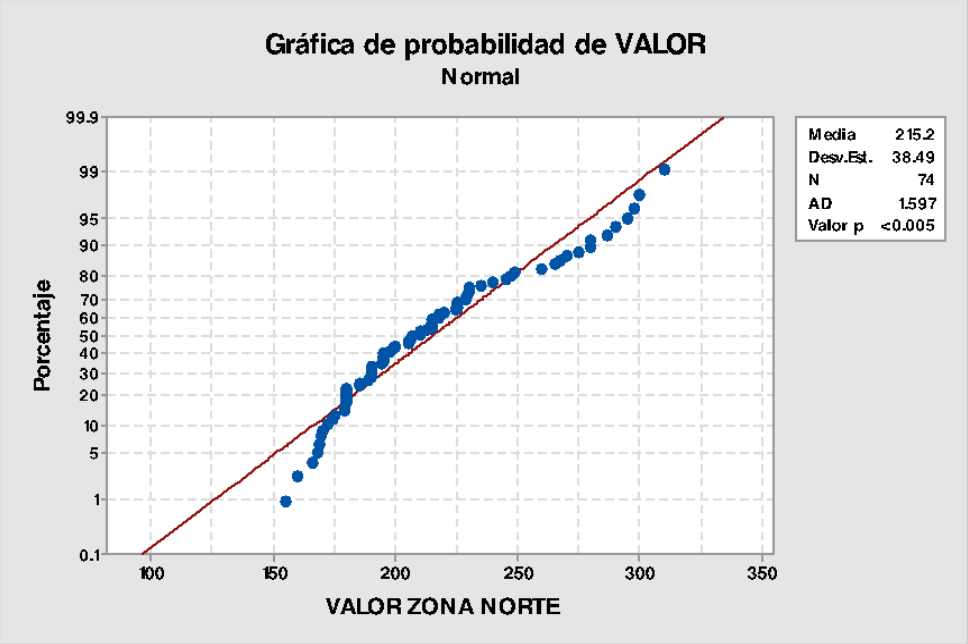
3 En **Variable**, ingrese ***Valor***. Haga clic en **Aceptar**.

Salida de la ventana Gráfica

Genera una gráfica de probabilidad normal y realiza una prueba de hipótesis para examinar si las observaciones siguen o no una distribución normal . Para la prueba de normalidad , las hipótesis son,

H_0: los datos siguen una distribución normal vs. H_1: los datos no siguen una distribución normal

La escala vertical de la gráfica se asemeja a la escala vertical del papel de probabilidad normal. El eje horizontal es una escala lineal. La línea forma un estimado de la función de distribución acumulada para la población de la cual se extrajeron los datos. Con la gráfica se muestran estimaciones numéricas de los parámetros de la población , Media y Desviación Estándar, el tamaño de la muestra ó población, el valor de la prueba de normalidad (Estadístico de Anderson Darling) y el valor p asociado.



Interpretación de los resultados

La salida gráfica es una gráfica de probabilidades normales versus los datos. Los datos se alejan de la línea ajustada de una manera más evidente en los extremos, o colas de la distribución. El valor p de la prueba de Anderson-Darling indica que, en niveles p menores a 0.005, hay evidencia de que los datos no siguen una distribución normal.

Los problemas de no normalidad puede solucionarse por medio de **transformaciones de datos** y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una o varias de las variables por lo que también se debe verificar el supuesto de homocedasticidad y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

1.3.2.1.2 EJERCICIO DE **REFUERZO**

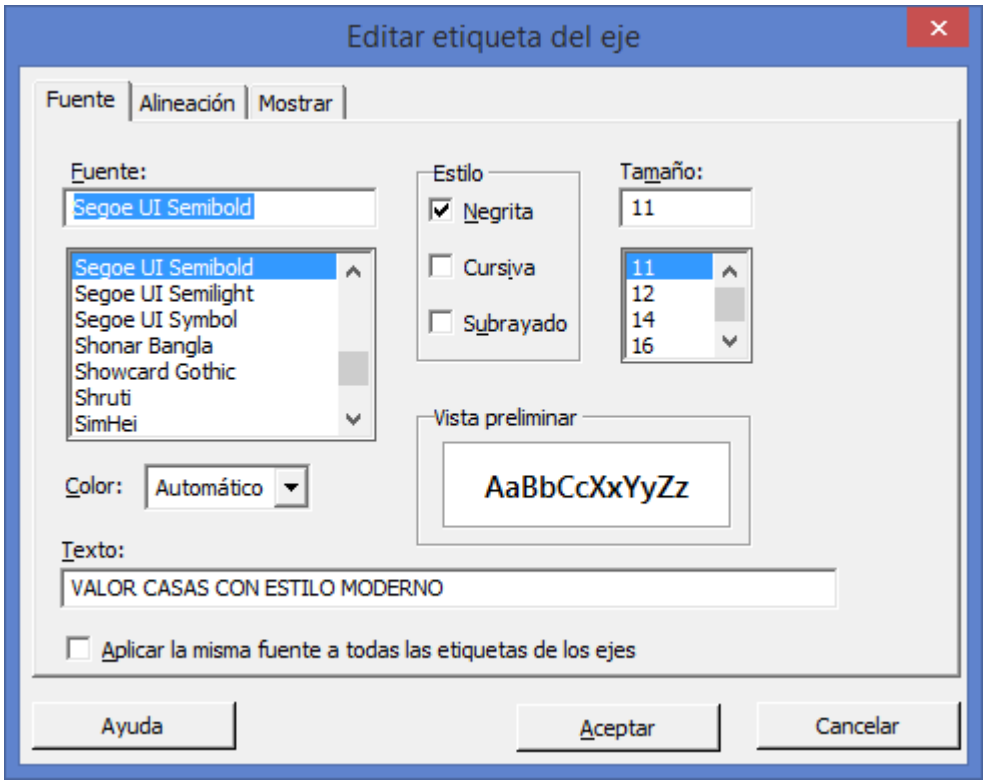
EJERCICIO DE **REFUERZO 1.3.2.1.2**

PRUEBA **DE NORMALIDAD. PRUEBA DE ANDERSON DARLING.**

1.3.2.1.2 Se midió el **VALOR** de las casas con estilo **MODERNO** en la base de datos "**BIENES1"** que se construyó con anterioridad.

Utilizando el paquete **Minitab** se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas con estilo **MODERNO** e interprete los resultados.

Nota: Etiquete la gráfica dando doble clic en **VALOR** para que aparezca la ventana ***Editar etiqueta de eje*** y en la parte inferior donde dice ***Texto*** escriba **VALOR CASAS CON ESTILO MODERNO**. Haga clic en **Aceptar.**



1 Abra la base de datos que creó **"BIENES1"** y cree un subconjunto de hoja de trabajo para las casas con estilo **MODERNO**

2 Elija **Estadísticas > Estadísticas básicas > Prueba de normalidad**.

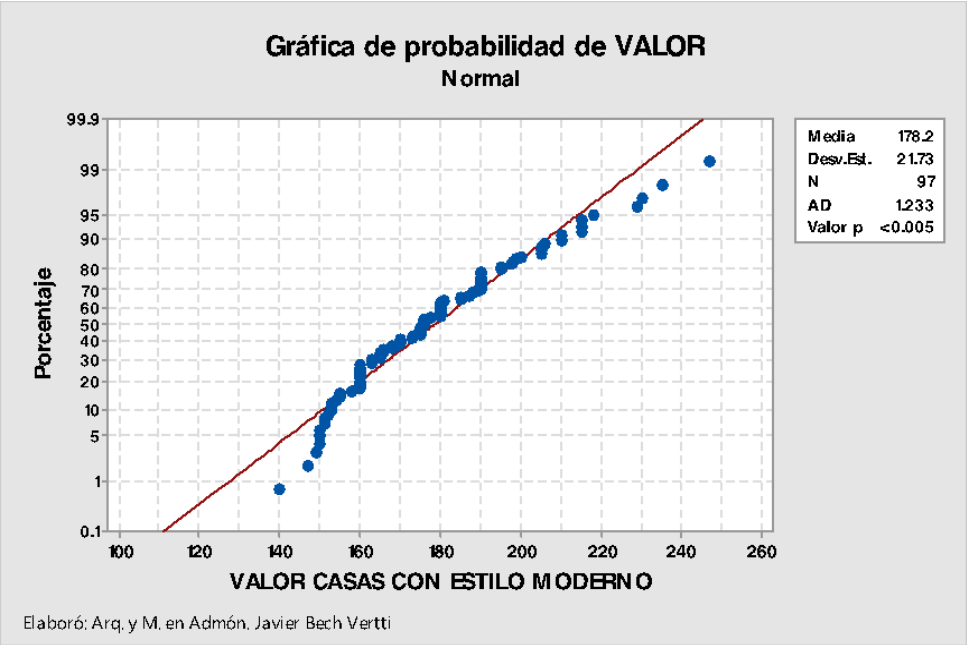
3 En **Variable**, ingrese ***Valor.*** Haga clic en **Aceptar**.

Salida de la ventana Gráfica

Genera una **gráfica de probabilidad** normal y realiza una **prueba de hipótesis** para examinar si las observaciones siguen o no una **distribución normal** . Para la **prueba de normalidad** , las hipótesis son,

H_0: los datos siguen una distribución normal vs. H_1: los datos no siguen una distribución normal

La escala vertical de la gráfica se asemeja a la escala vertical del papel de probabilidad normal. El eje horizontal es una escala lineal. La línea forma un estimado de la función de distribución acumulada para la **población** de la cual se extrajeron los datos. Con la gráfica se muestran **estimaciones numéricas de los parámetros de la población** , Media y Desviación Estándar, el tamaño de la muestra ó población, el valor de la prueba de normalidad (Estadístico de Anderson Darling) y el **valor p** asociado.



Interpretación de los resultados

La salida gráfica es una gráfica de probabilidades normales versus los datos. Los datos se alejan de la línea ajustada de una manera más evidente en los extremos, o colas de la distribución. El valor p de la prueba de Anderson-Darling indica que, en niveles p menores a 0.005, hay evidencia de que los datos no siguen una distribución normal.

Los problemas de no normalidad puede solucionarse por medio de **transformaciones de datos** y como se

mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una o varias de las variables por lo que también se debe verificar el supuesto de homocedasticidad y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

1.3.2.1.3 EJERCICIO DE REFUERZO

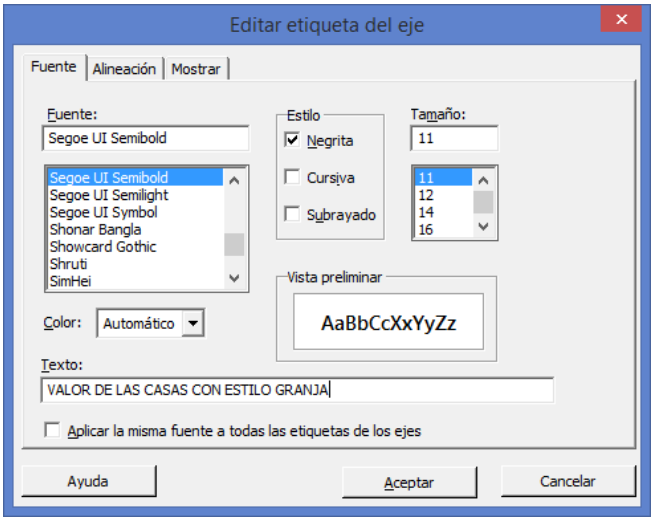
EJERCICIO DE REFUERZO 1.3.2.1.3

PRUEBA DE NORMALIDAD. PRUEBA DE ANDERSON DARLING.

1.3.2.1.3 Se midió el **VALOR** de las casas con estilo **GRANJA** en la base de datos "**BIENES1"** que se construyó con anterioridad.

Utilizando el paquete **Minitab** se desea determinar si estos datos siguen una distribución normal, de modo que utiliza una Prueba de normalidad de Anderson Darling, creando primero un subconjunto de hoja de trabajo para las casas con estilo **GRANJA** e interprete los resultados.

Nota: Etiquete la gráfica dando doble clic en **VALOR** para que aparezca la ventana ***Editar etiqueta de eje*** y en la parte inferior donde dice ***Texto*** escriba **VALOR DE LAS CASAS CON ESTILO GRANJA**. Haga clic en **Aceptar.**



1 1 Abra la base de datos que creó **"BIENES1"** y cree un subconjunto de hoja de trabajo para las casas con estilo **GRANJA**

2 Elija **Estadísticas > Estadísticas básicas > Prueba de normalidad**.

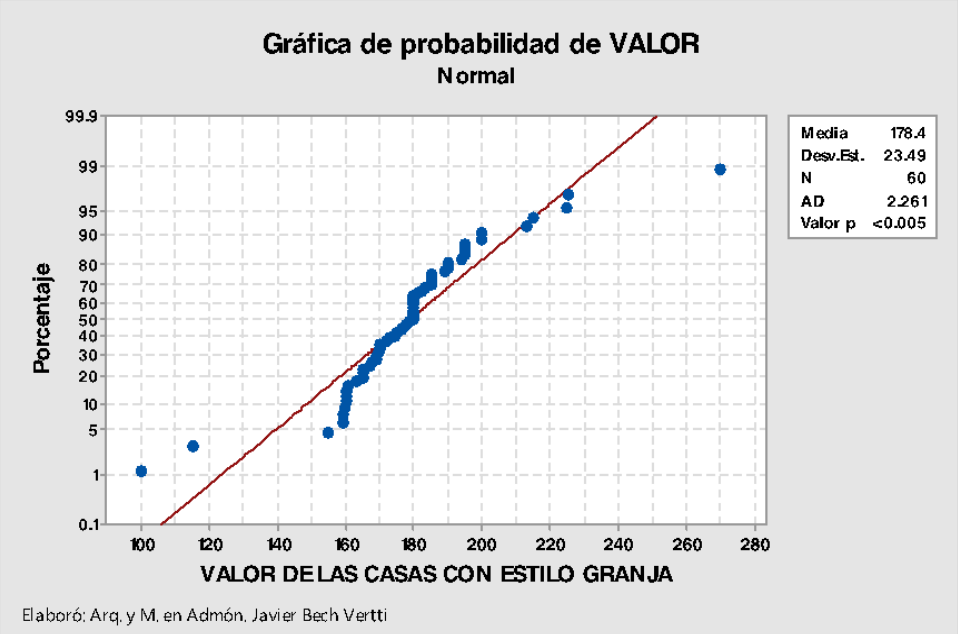
3 En **Variable**, ingrese ***Valor***. Haga clic en **Aceptar**.

Salida de la ventana Gráfica

Genera una **gráfica de probabilidad** normal y realiza una **prueba de hipótesis** para examinar si las observaciones siguen o no una **distribución normal** . Para la **prueba de normalidad** , las hipótesis son,

H_0: los datos siguen una distribución normal vs. H_1: los datos no siguen una distribución normal

La escala vertical de la gráfica se asemeja a la escala vertical del papel de probabilidad normal. El eje horizontal es una escala lineal. La línea forma un estimado de la función de distribución acumulada para la **población** de la cual se extrajeron los datos. Con la gráfica se muestran **estimaciones numéricas de los parámetros de la población** , Media y Desviación Estándar, el tamaño de la muestra ó población, el valor de la prueba de normalidad (Estadístico de Anderson Darling) y el **valor p** asociado.



Interpretación de los resultados

La salida gráfica es una gráfica de probabilidades normales versus los datos. Los datos se alejan de la línea ajustada de una manera más evidente en los extremos, o colas de la distribución. El valor p de la prueba de Anderson-Darling indica que, en niveles p menores a 0.005, hay evidencia de que los datos no siguen una distribución normal. **Recomendación: Verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.**

Los problemas de no normalidad puede solucionarse por medio de **transformaciones de datos** y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una o varias de

las variables por lo que también se debe verificar el supuesto de homocedasticidad y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

1.3.2.2.1 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 1.3.2.2.1

PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE LEVENE. TRANSFORMACIONES. ETIQUETADO DE GRÁFICOS.

1.3.2.2.1 Se pretende diseñar un experimento para evaluar el **IMPUESTO** que pagan 233 casas en diferentes zonas de la ciudad. Se mide el **IMPUESTO** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad en cada una de las casas con **ESTILO** diferente. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de comparación múltiple apropiado para los datos.

Antes de realizar el **análisis de varianza,** verifica la suposición de que la varianza es igual en todos los niveles del Factor **ESTILO** utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**) utilizando el paquete de cómputo **Minitab:**

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

PRUEBA DE LEVENE EN MINITAB 17

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

Nota importante: Verifica que en la ventana de Prueba de varianzas iguales el primer campo

diga "Los datos de respuesta están en una columna para todos los niveles de los factores"

3 En **Respuesta**, ingrese ***IMPUESTO***

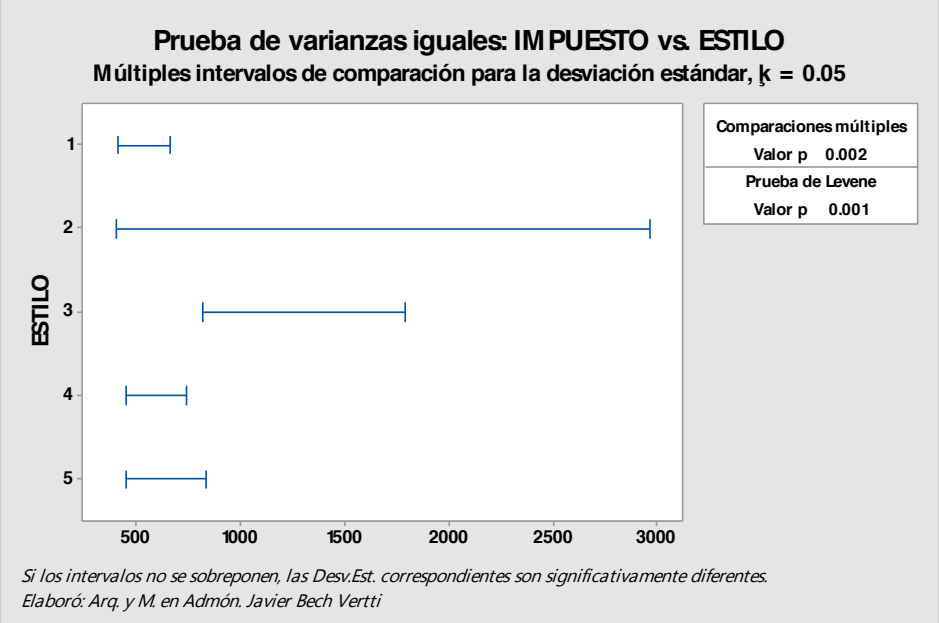
4 En **Factores**, ingrese ***ESTILO.*** Haga clic en **Aceptar**.

Salida de la ventana Sesión

Prueba de varianzas iguales: IMPUESTO vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

ESTILO   N  Desv.Est.          IC

     1  97     509.20  (412.612,  645.54)

     2  15     979.79  (269.539, 4299.96)

     3  23    1130.71  (702.530, 2049.38)

     4  60     558.92  (431.820,  755.87)

     5  38     582.90  (416.535,  875.03)


Nivel de confianza individual = 99%

Pruebas


                         Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.002

Levene                          4.56    0.001
```

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de las pruebas **Levene** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **IMPUESTO** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.001 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales rompiéndose el supuesto de homocedasticidad necesario para estas pruebas.

Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de

forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

TRANSFORMACIONES EN MINITAB 17

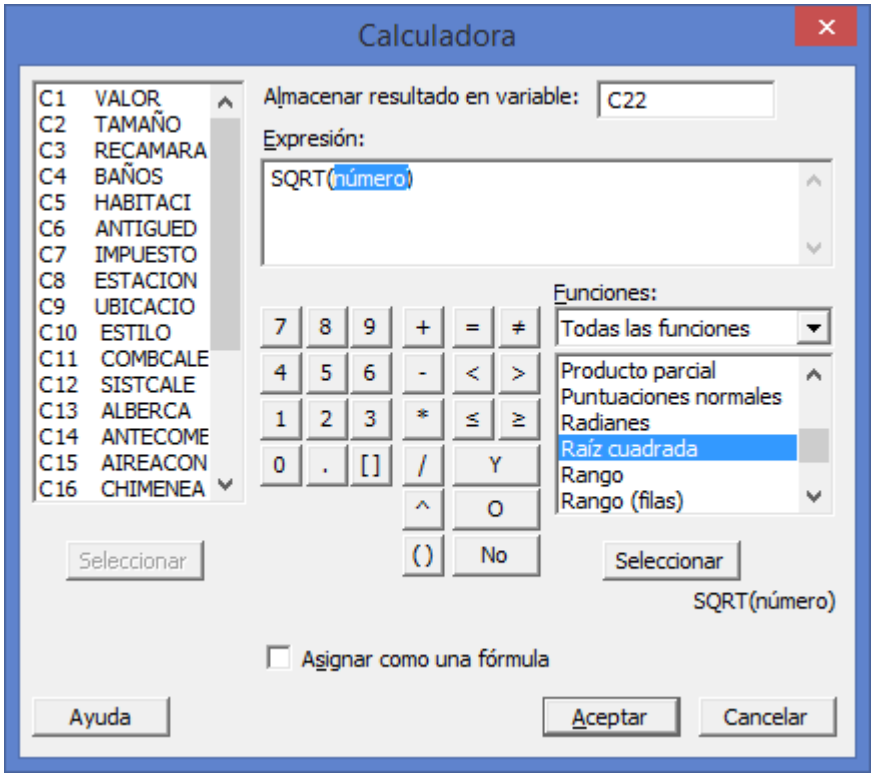
Transformaciónes Raiz Cuadrada, 1/X y Log de la variable IMPUESTO con Minitab:

11 Abra la base de datos que creó **"BIENES1"**

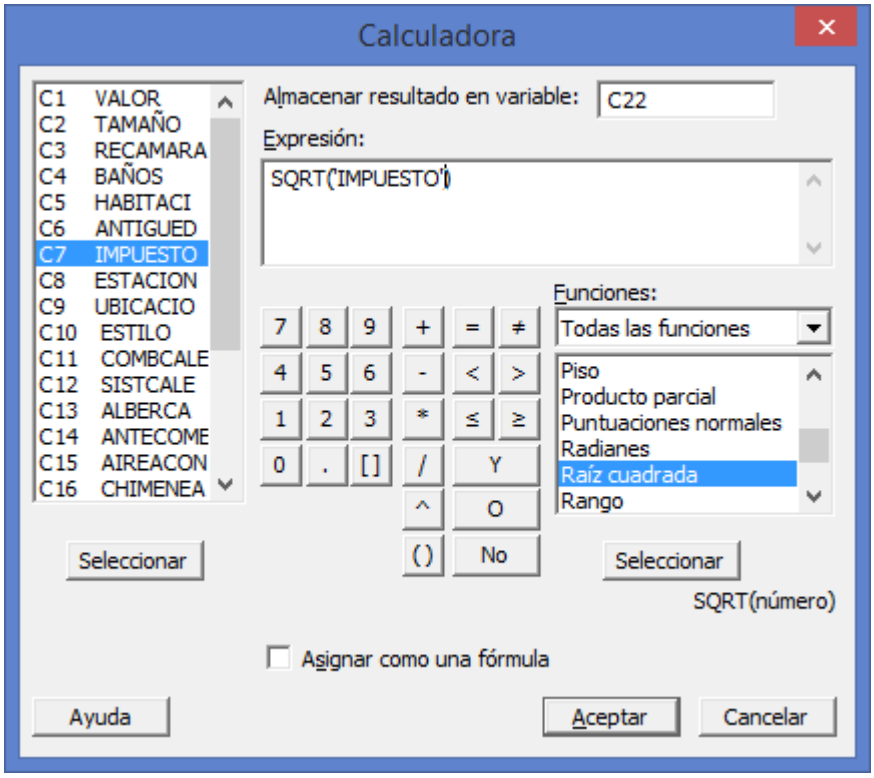
2 Elija **Calc > Calculadora**.

5 En **Almacenar resultado en variable**, ingrese *C22*

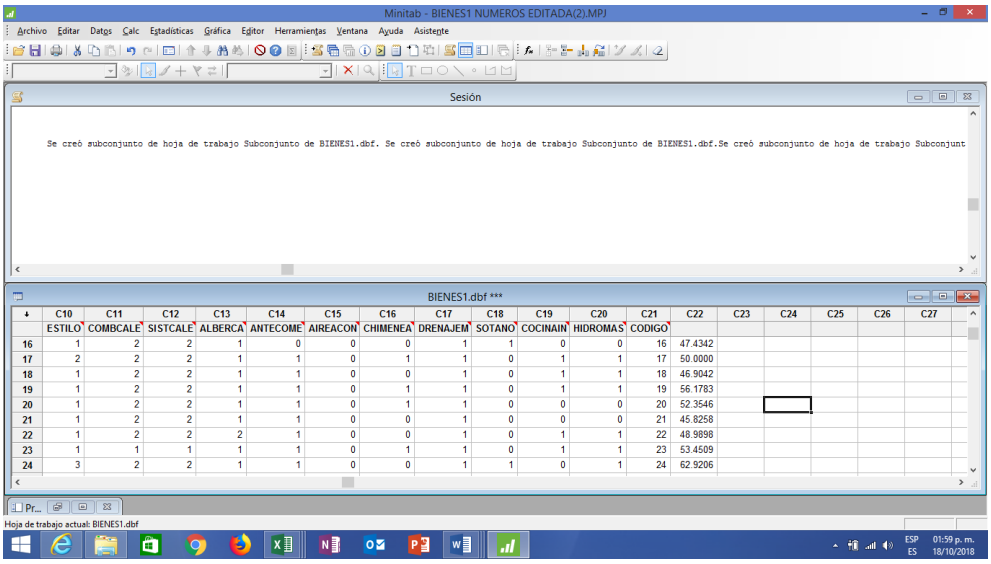
4 En **Expresión**, busque en **Funciones**: Raíz Cuadrada y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



5 En **Expresión** aparece **SQRT(número).** Seleccione la variable **IMPUESTO** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.

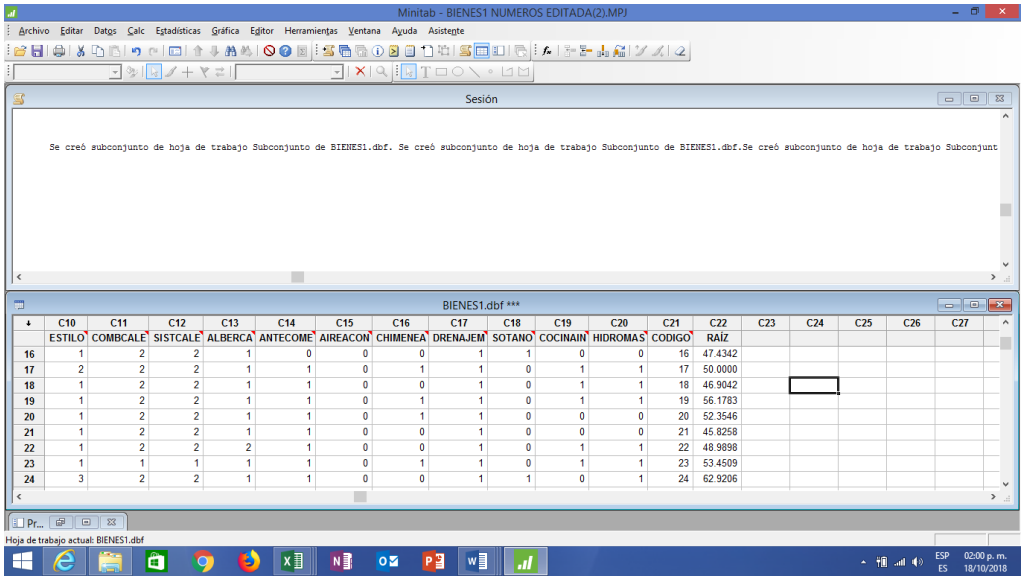


6 Haga clic en **Aceptar**.



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206

7 En C22 teclee el nombre **RAÍZ**

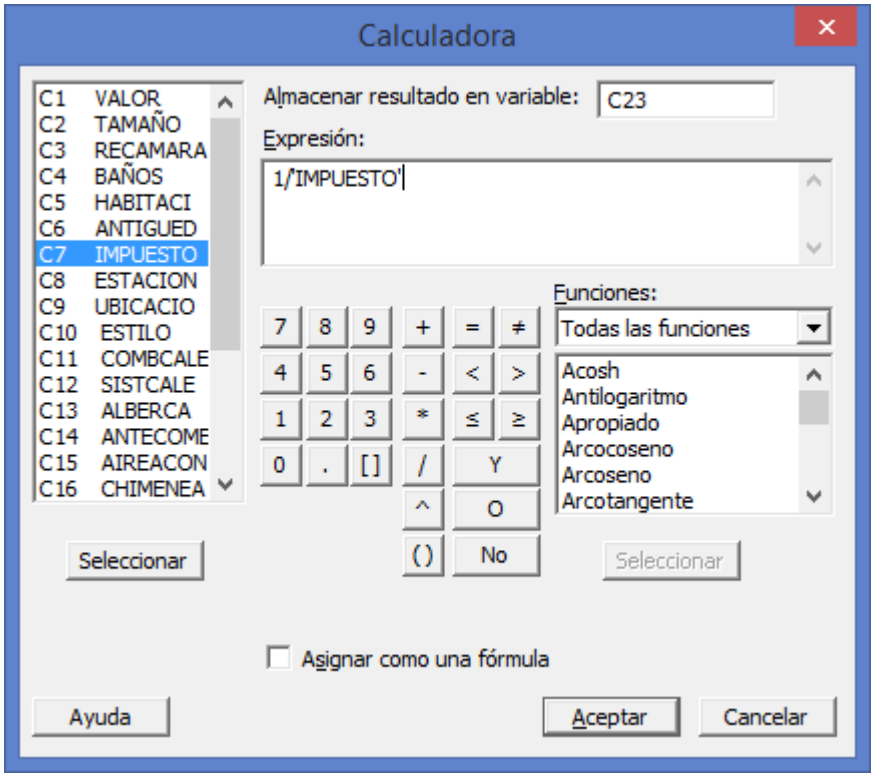


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ					
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342					
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000					
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042					
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783					
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546					
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258					
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898					
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509					
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206					

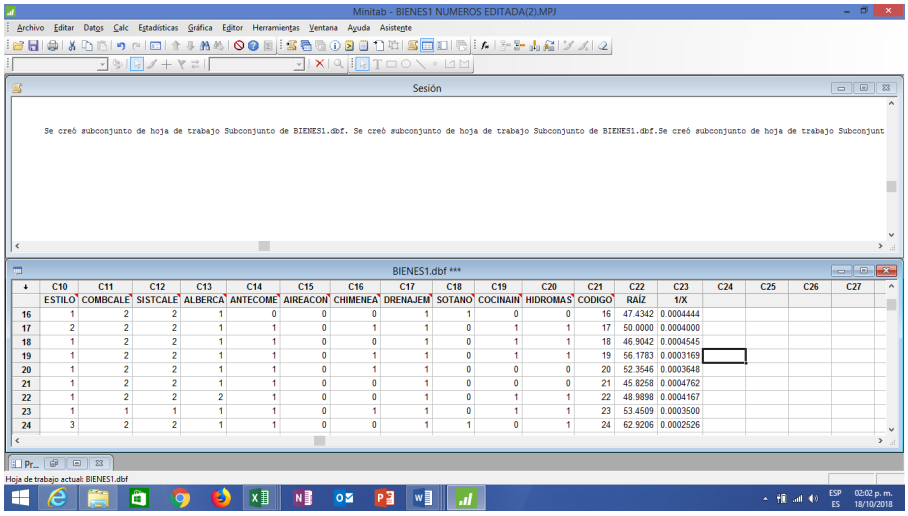
10 Elija **Calc > Calculadora**.

9 En **Almacenar resultado en variable**, ingrese ***C23***

10 En **Expresión**, para obtener la Inversa, **Teclee 1/** y seleccione la variable **IMPUESTO** del lado izquierdo o doble clic.



13 Haga clic en **Aceptar**. En **C23** escriba el nombre **1/X**

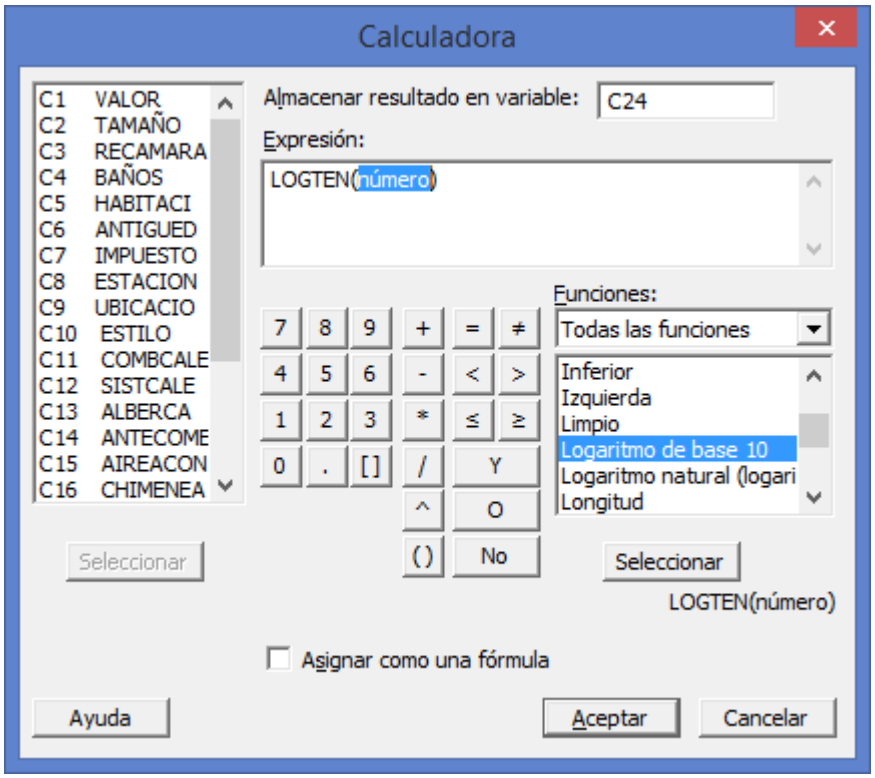


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ	1/X				
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342	0.0004444				
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000	0.0004000				
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042	0.0004545				
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783	0.0003169				
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546	0.0003648				
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258	0.0004762				
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898	0.0004167				
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509	0.0003500				
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206	0.0002526				

14 Elija **Calc > Calculadora**.

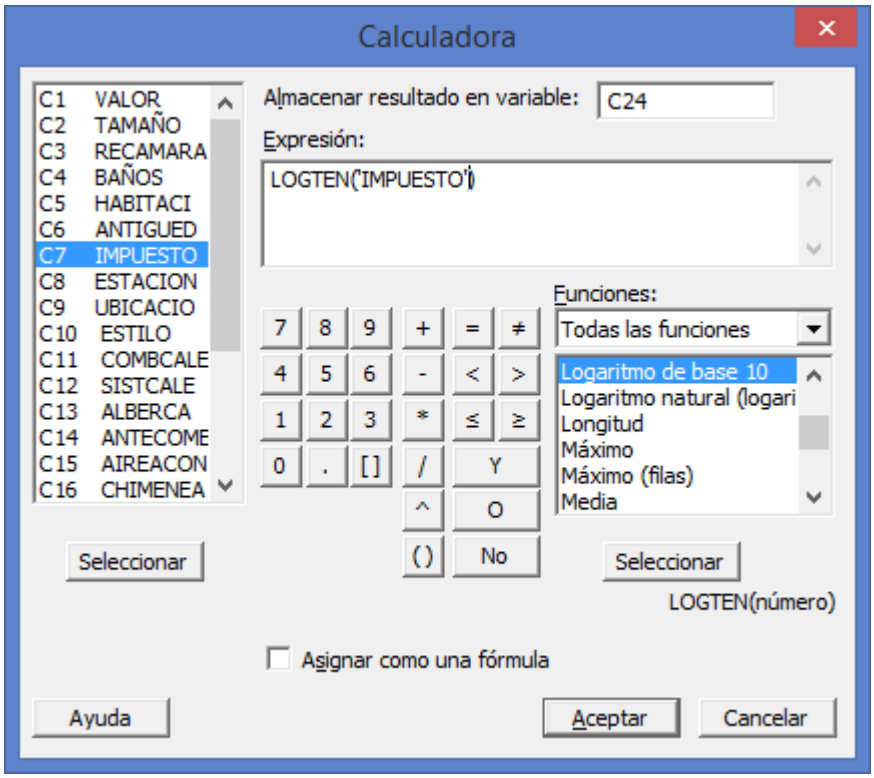
13 En **Almacenar resultado en variable**, ingrese ***C24***

14 En **Expresión**, busque en **Funciones**: **Logaritmo base 10** y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.

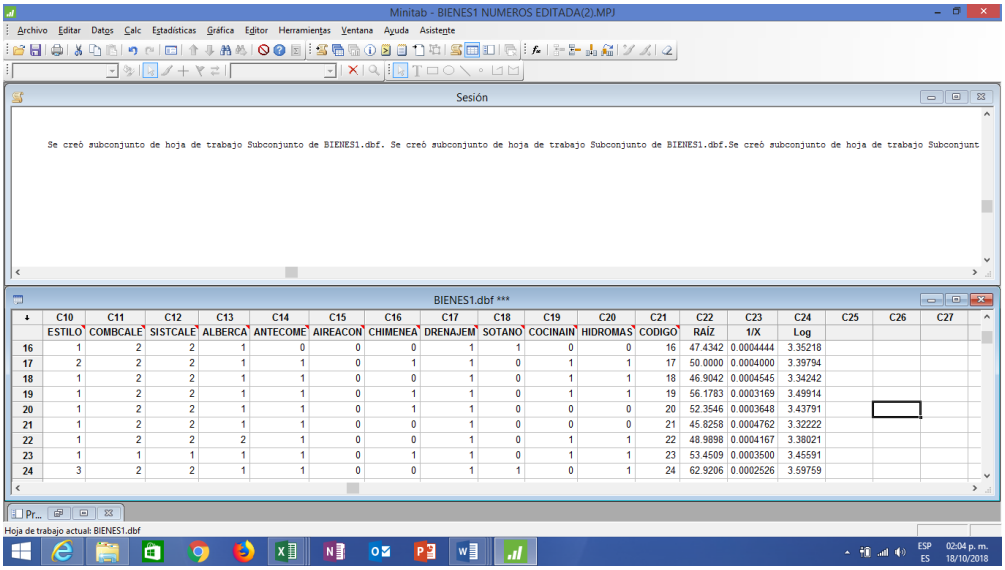


15 En **Expresión** aparece **LOGTEN(número).** Seleccione la variable **IMPUESTO** del lado izquierdo

y haga clic en el botón **Seleccionar** o doble clic.



16 Haga clic en **Aceptar.** En **C24** escriba el nombre **Log**



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ	1/X	Log			
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342	0.0004444	3.35218			
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000	0.0004000	3.39794			
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042	0.0004545	3.34242			
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783	0.0003169	3.49914			
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546	0.0003648	3.43791			
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258	0.0004762	3.32222			
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898	0.0004167	3.38021			
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509	0.0003500	3.45591			
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206	0.0002526	3.59759			

Repita la prueba de Levene para las TRES transformaciones para ver si se corrigió el problema de homocedasticidad y si fué así cual transformación resulto mejor.

A. CON LA RAÍZ CUADRADA DE IMPUESTO

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***RAÍZ***

4 En **Factores**, ingrese ***ESTILO.*** Haga clic en **Aceptar**.

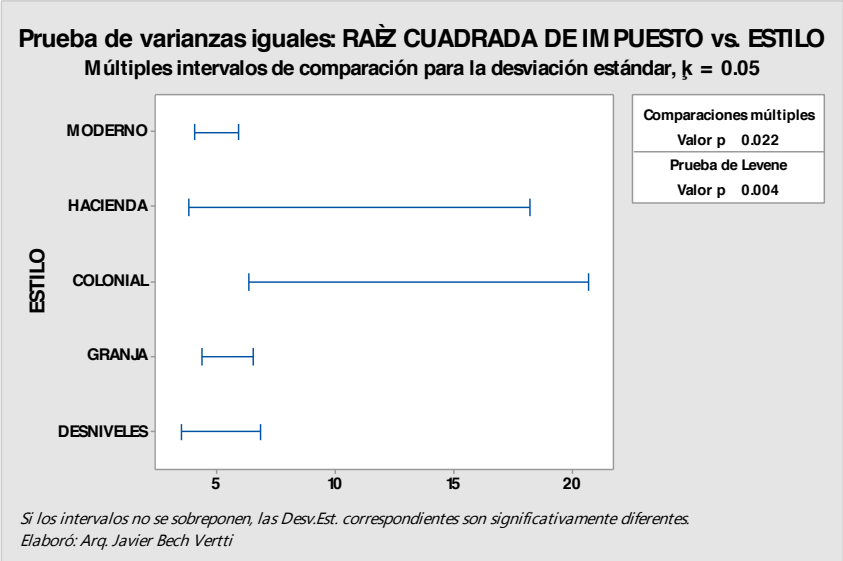
Salida de la ventana Sesión

Prueba de varianzas iguales: RAÍZ vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar


ESTILO   N  Desv.Est.          IC

     1  97     4.8390  (3.98045,  6.0432)

     2  15     7.5642  (2.33311, 29.6085)

     3  23    10.7479  (6.01235, 21.6363)

     4  60     5.2251  (4.11020,  6.9404)

     5  38     4.7212  (3.51113,  6.8098)


Nivel de confianza individual = 99%


Pruebas

                        Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.022

Levene                          3.89    0.004
```

Prueba de varianzas iguales: RAÍZ vs. ESTILO

Salida de la ventana Gráfica



Para el ejemplo para evaluar la transformación de la **RAIZ CUADRADA** del **IMPUESTO** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.004 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

B. CON LA LA INVERSA DE IMPUESTO (1/X)

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***1/X***

4 En **Factores**, ingrese ***ESTILO.*** Haga clic en **Aceptar**.

Salida de la ventana Sesión

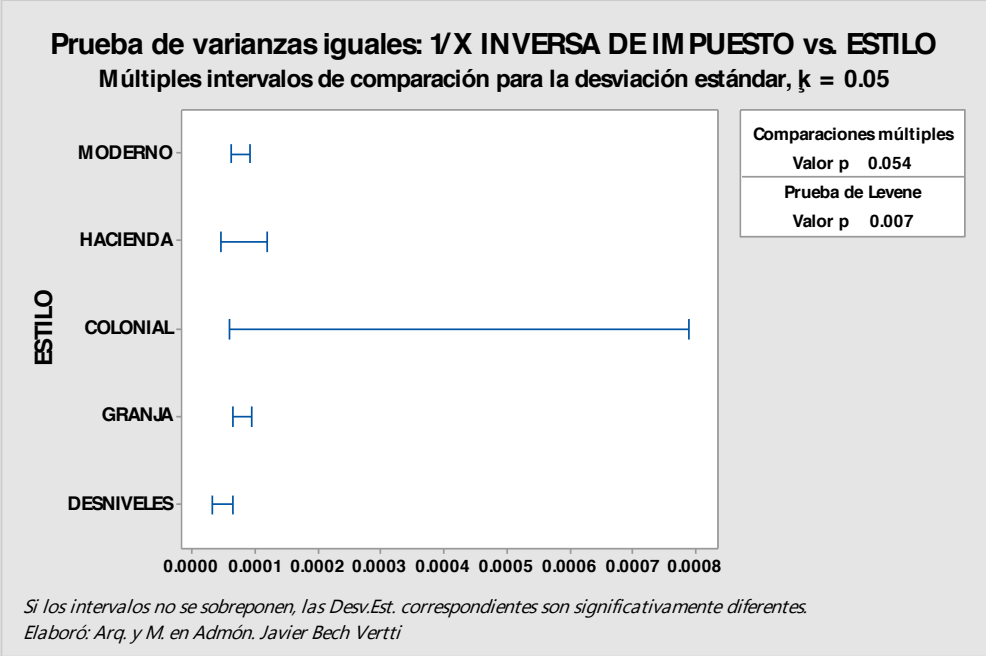
Prueba de varianzas iguales: 1/X vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05
```

```
Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

ESTILO   N  Desv.Est.            IC

     1  97  0.0000725  (0.0000568, 0.0000951)

     2  15  0.0000649  (0.0000303, 0.0001678)

     3  23  0.0002023  (0.0000836, 0.0005511)

     4  60  0.0000748  (0.0000566, 0.0001032)

     5  38  0.0000420  (0.0000334, 0.0000566)

Nivel de confianza individual = 99%

Pruebas

                        Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.054

Levene                          3.62    0.007
```

Prueba de varianzas iguales: 1/X vs. ESTILO

Salida de la ventana Gráfica



Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **INVERSA** del **IMPUESTO** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.007 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

C. CON EL LOGARITMO BASE 10 DE IMPUESTO

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***Log***

4 En **Factores**, ingrese ***ESTILO***. Haga clic en **Aceptar**.

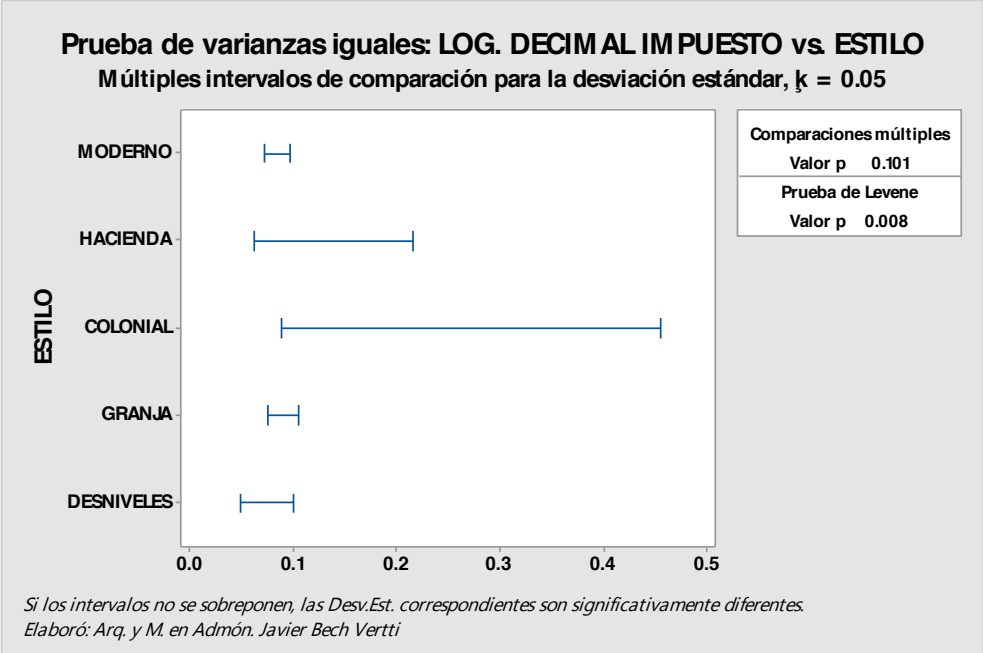
Salida de la ventana Sesión

Prueba de varianzas iguales: Log vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

ESTILO   N  Desv.Est.            IC

     1  97   0.081038  (0.0666267, 0.101255)

     2  15   0.104035  (0.0366471, 0.356571)

     3  23   0.187016  (0.0939630, 0.419164)

     4  60   0.086150  (0.0678713, 0.114256)

     5  38   0.066950  (0.0513849, 0.093572)

Nivel de confianza individual = 99%

Pruebas

                         Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.101

Levene                          3.57    0.008
```

Prueba de varianzas iguales: Log vs. ESTILO

Salida de la ventana Gráfica



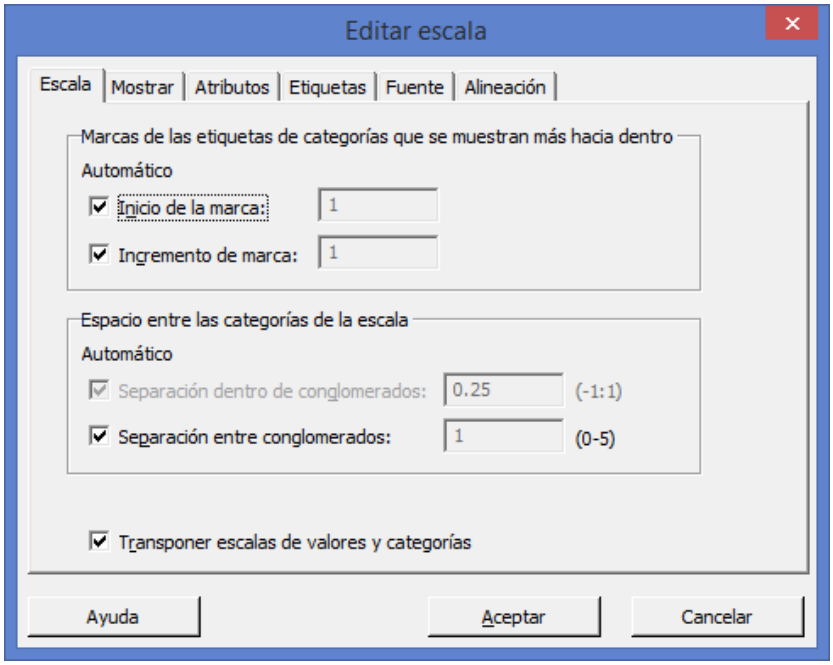
Interpretación de los resultados

Para el ejemplo para evaluar la transformación del **LOGARITMO DECIMAL** del **IMPUESTO** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.008 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

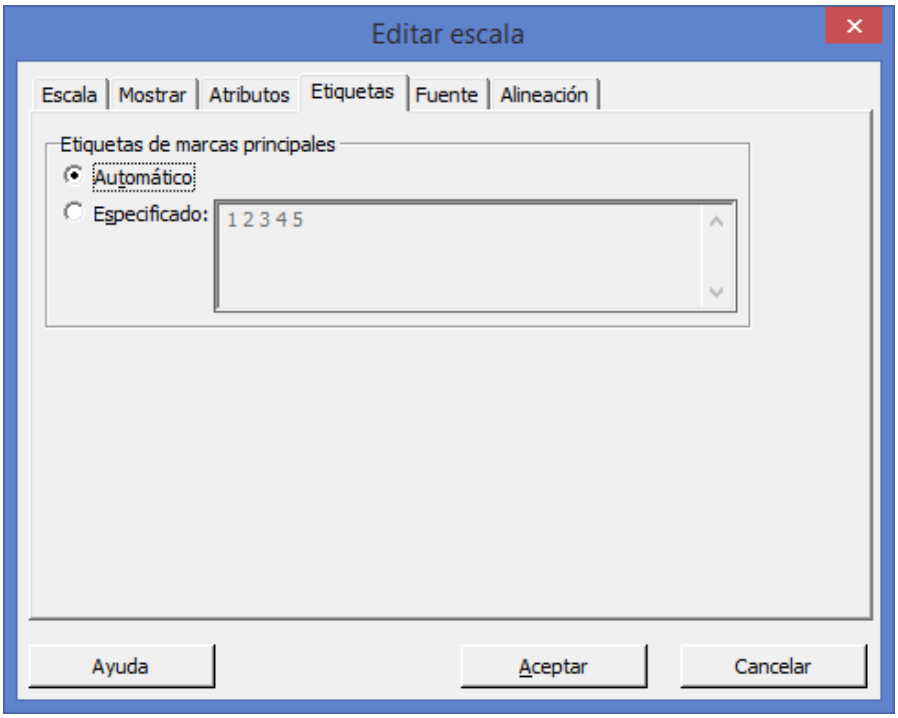
Nota Importante: No es del alcance de este trabajo tomar la decisión de si se debe o no trabajar con los datos transformados o sin transformar para usar alguna técnica multivariada sino solamente mostrar la forma de hacer dichas transformaciones.

ETIQUETADO DE LOS NIVELES DE LA VARIABLE O "FACTOR" ESTILO EN MINITAB 17:

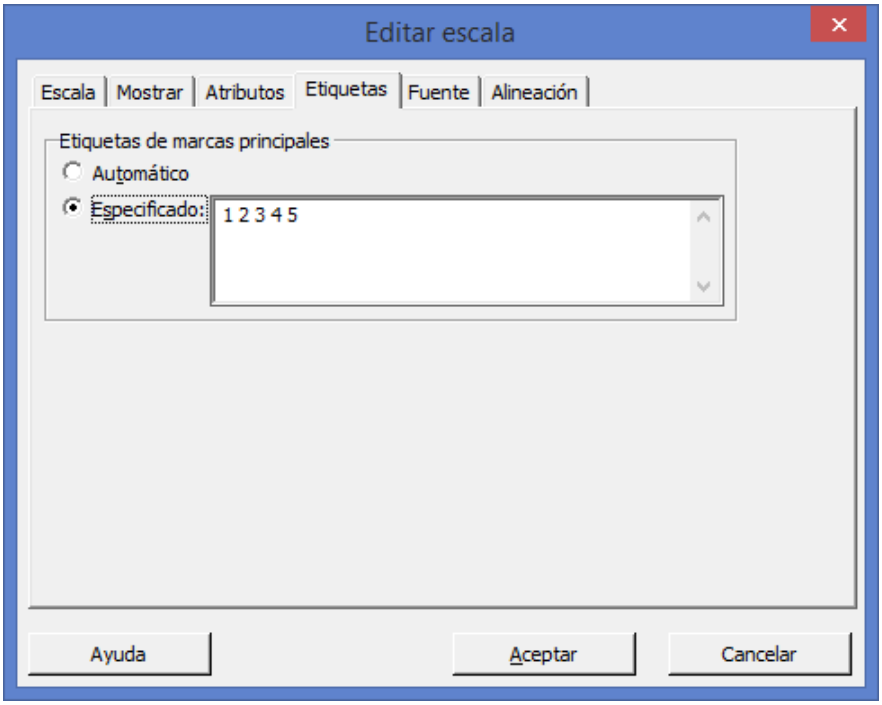
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **ESTILO** para que Minitab despliegue la ventana ***Editar Escala.***



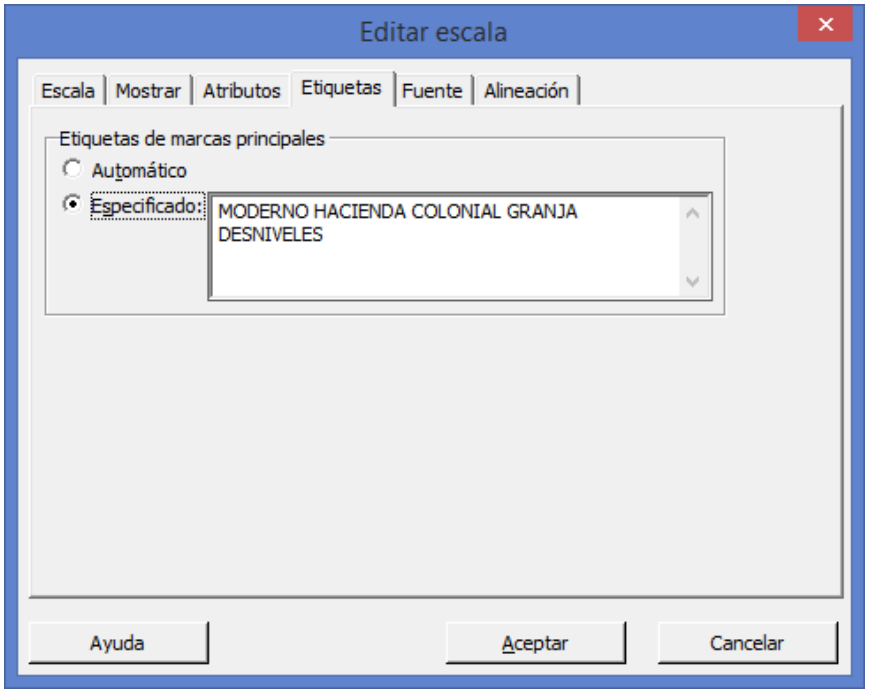
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

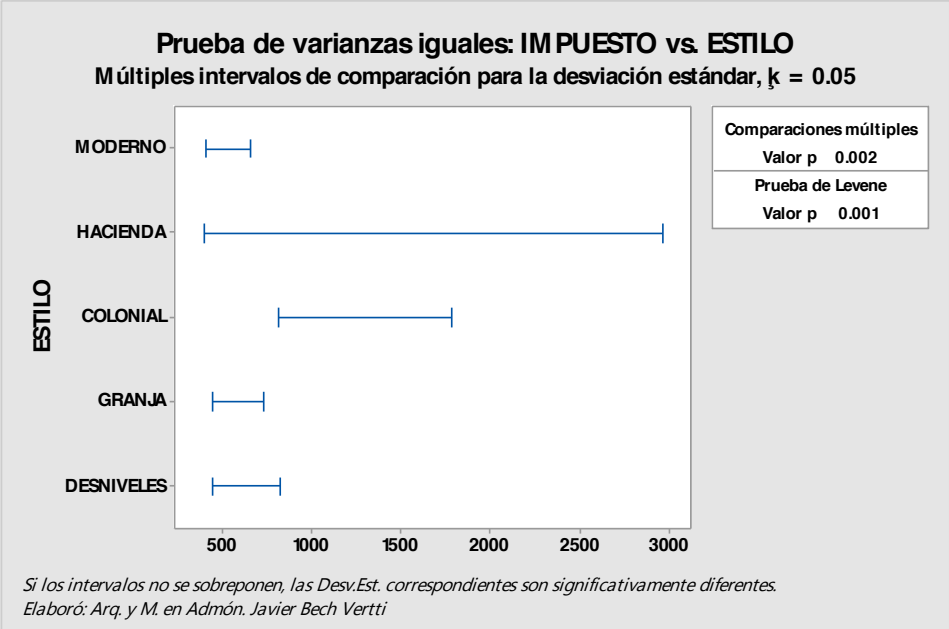


4 Borrar los números 1 2 3 4 5 y escribir **MODERNO HACIENDA COLONIAL GRANJA DESNIVELES**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



1.3.2.2.2 EJERCICIO DE **REFUERZO.O**

EJERCICIO DE **REFUERZO 1.3.2.2.2**

PRUEBA **DE HOMOCEDASTICIDAD. PRUEBA DE LEVENE**. TRANSFORMACIONES. ETIQUETADO DE GRÁFICOS.

1.3.2.2.2 Se pretende diseñar un experimento para evaluar el **IMPUESTO** que pagan 233 casas en diferentes zonas de la ciudad. Se mide el **IMPUESTO** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad con respecto a su **UBICACIÓN** diferente. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de comparación múltiple apropiado para los datos.

Antes de realizar el **análisis de varianza,** verifica la suposición de que la varianza es igual en todos los niveles del Factor **UBICACIÓN** utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**) utilizando el paquete de cómputo **Minitab:**

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y

etiquete debidamente todos los gráficos de presentación.

PRUEBA DE LEVENE EN MINITAB 17

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

Nota importante: Verifica que en la ventana de Prueba de varianzas iguales el primer campo diga "Los datos de respuesta están en una columna para todos los niveles de los factores"

3 En **Respuesta**, ingrese ***IMPUESTO***

4 En **Factores**, ingrese ***UBICACIÓN***. Haga clic en **Aceptar**.

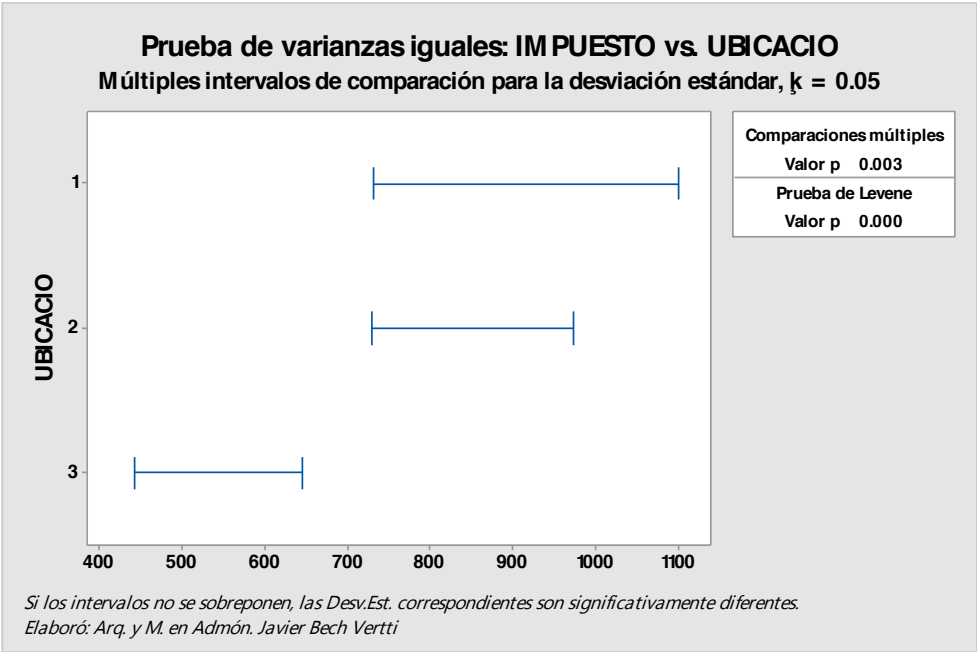
Salida de la ventana Sesión

Prueba de varianzas iguales: IMPUESTO vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.          IC

       1  74    882.083  (684.062, 1175.46)

       2  60    825.970  (675.654, 1051.69)

       3  99    527.133  (416.917,  683.00)

Nivel de confianza individual = 98.3333%

Pruebas


                          Estadística

Método                      de prueba  Valor p

Comparaciones múltiples            —    0.003

Levene                          8.46    0.000
```

Prueba de varianzas iguales: IMPUESTO vs. UBICACIO

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de las pruebas **Levene** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **IMPUESTO** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales rompiéndose el supuesto de homocedasticidad necesario para estas pruebas.

Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no

normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva

Asimismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución.

TRANSFORMACIONES EN MINITAB 17

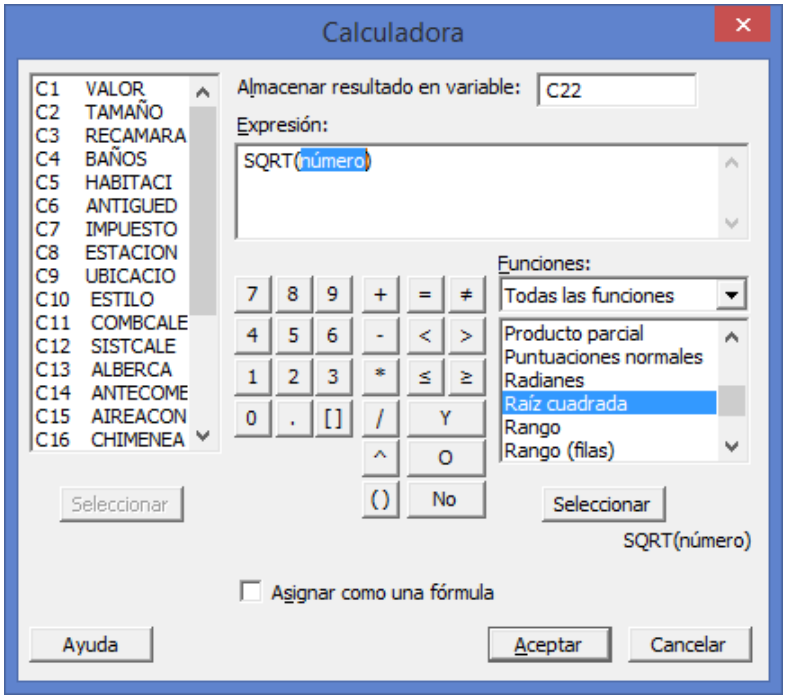
Transformaciónes Raiz Cuadrada, 1/X y Log de la variable IMPUESTO con Minitab:

12 Abra la base de datos que creó **"BIENES1"**

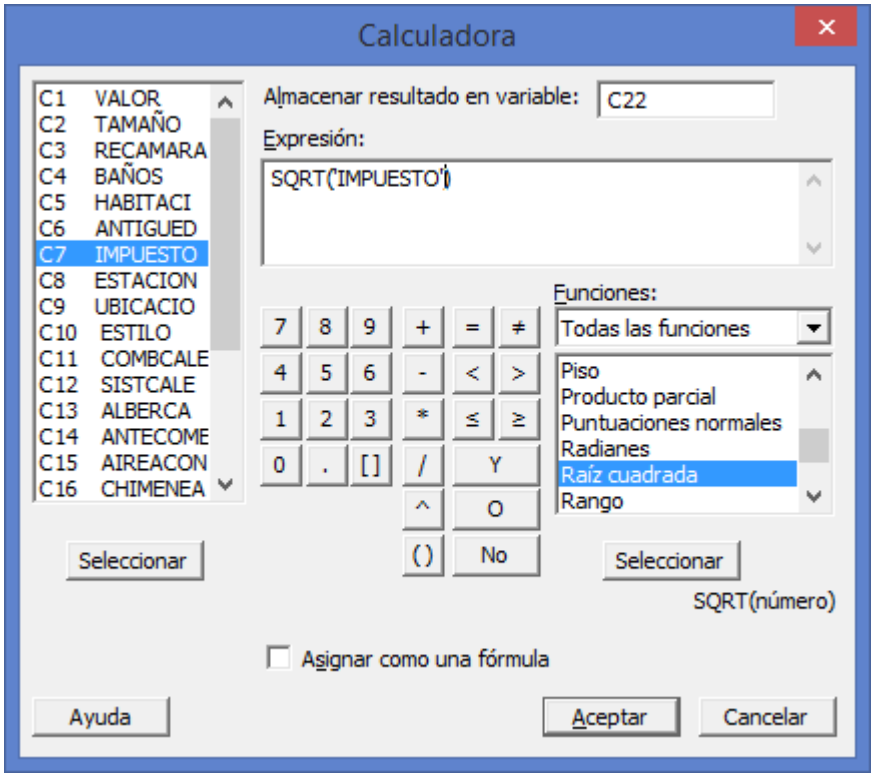
2 Elija **Calc > Calculadora**.

6 En **Almacenar resultado en variable**, ingrese *C22*

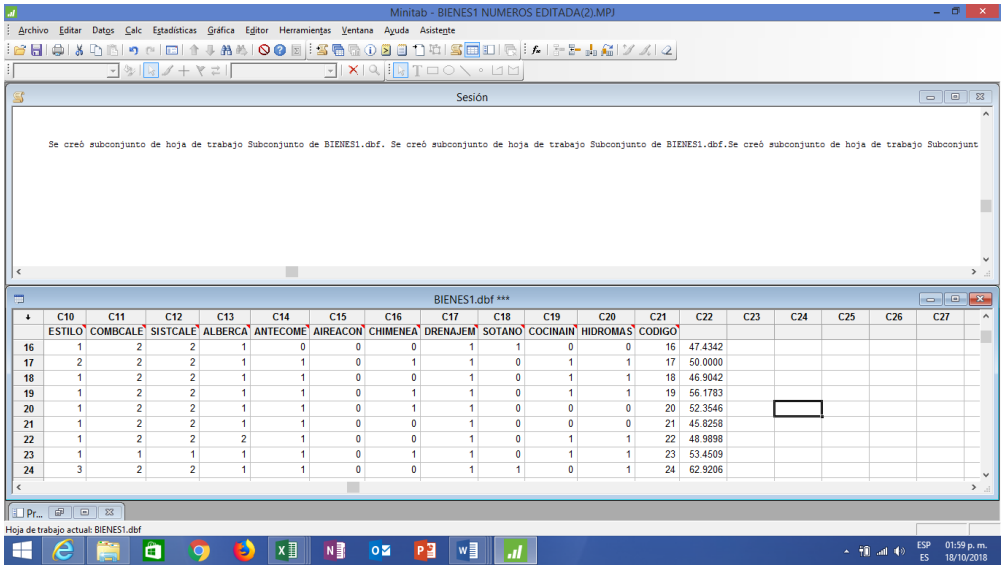
4 En **Expresión**, busque en **Funciones**: Raíz Cuadrada y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



5 En **Expresión** aparece **SQRT(número).** Seleccione la variable **IMPUESTO** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.

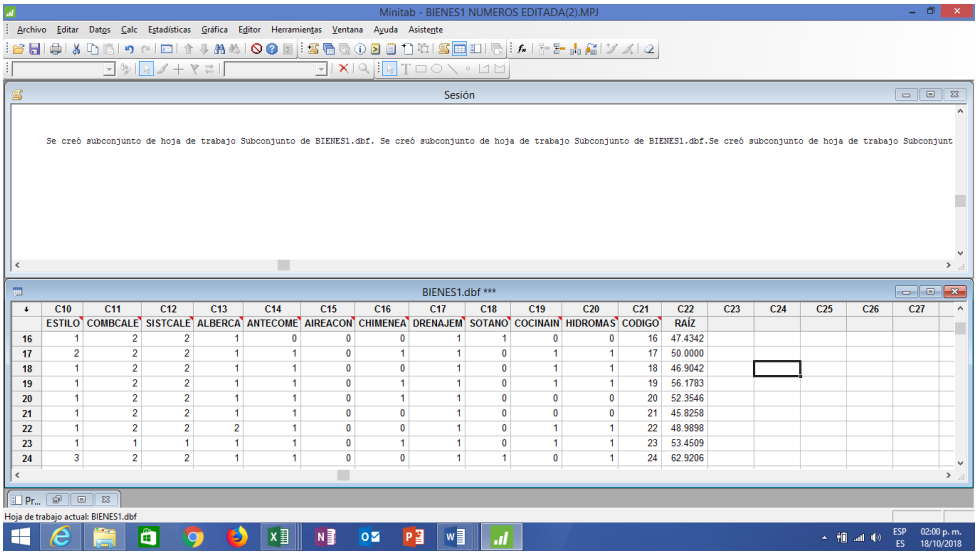


6 Haga clic en **Aceptar**.



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO						
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342					
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000					
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042					
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783					
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546					
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258					
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898					
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509					
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206					

7 En C22 teclee el nombre **RAÍZ**

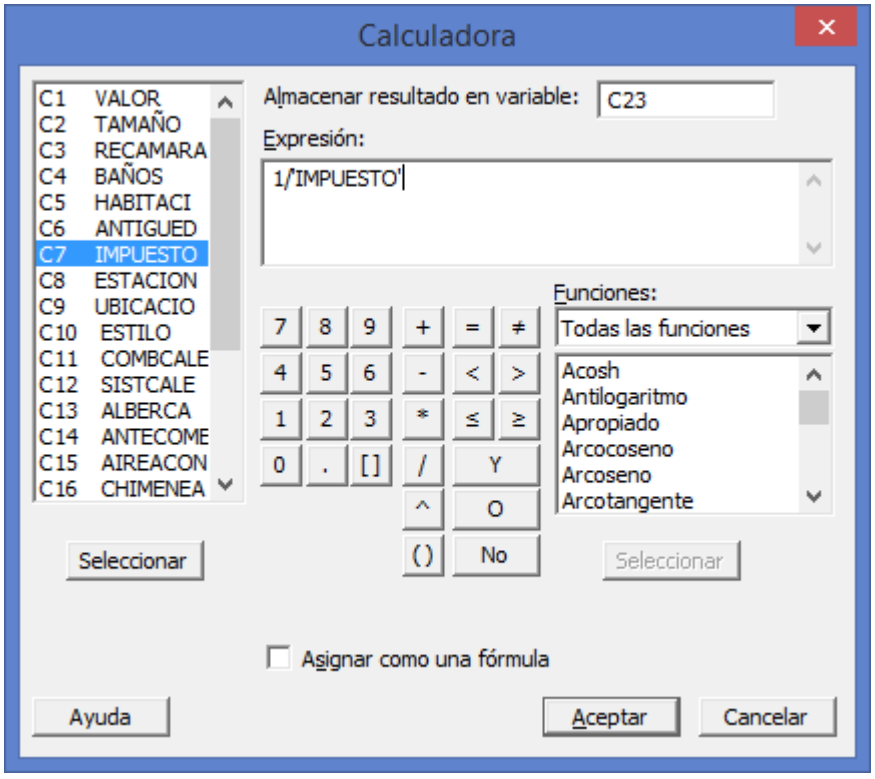


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ					
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342					
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000					
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042					
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783					
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546					
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258					
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898					
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509					
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206					

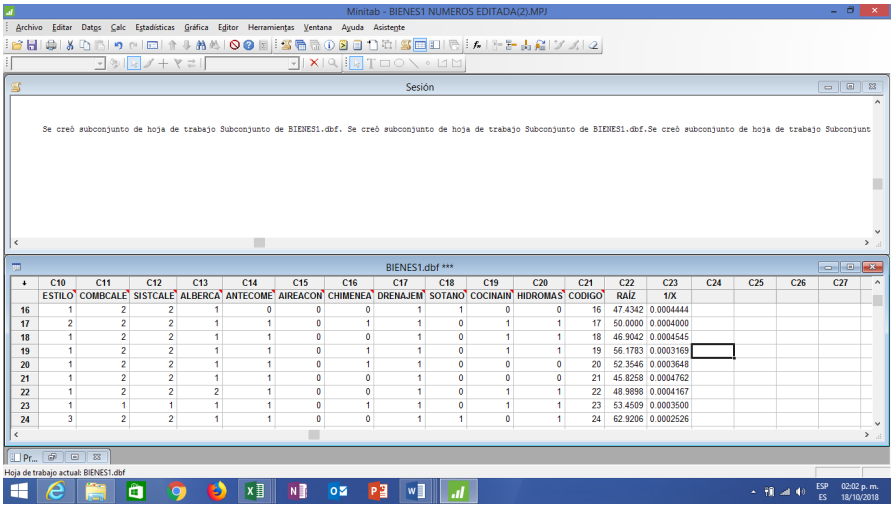
11 Elija **Calc > Calculadora**.

9 En **Almacenar resultado en variable**, ingrese ***C23***

10 En **Expresión**, para obtener la Inversa, **Teclee 1/** y seleccione la variable **IMPUESTO** del lado izquierdo o doble clic.



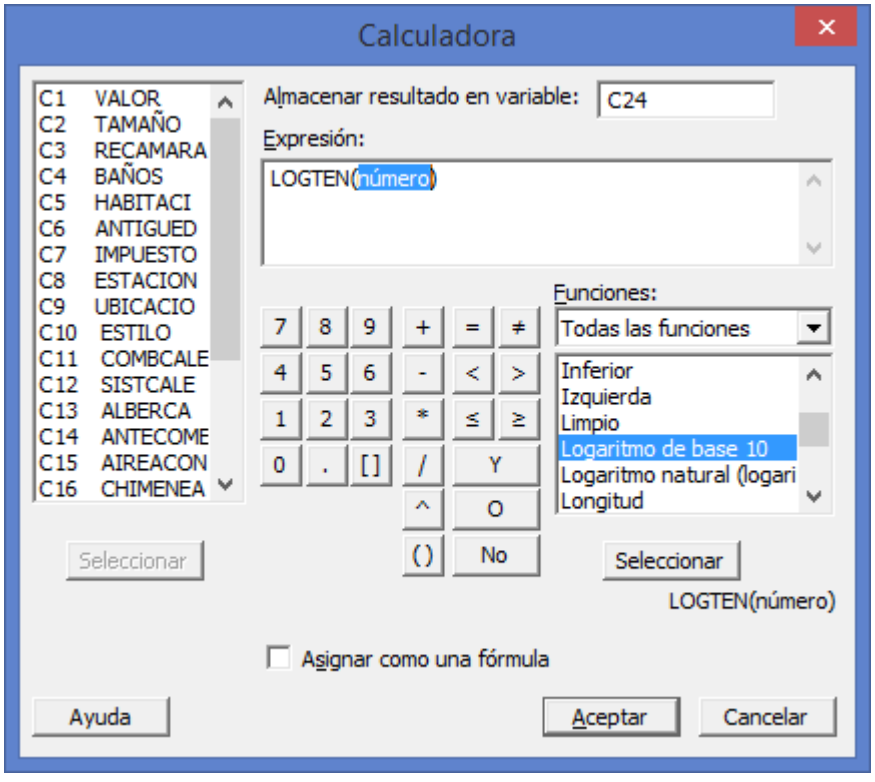
15 Haga clic en **Aceptar**. En **C23** escriba el nombre **1/X**



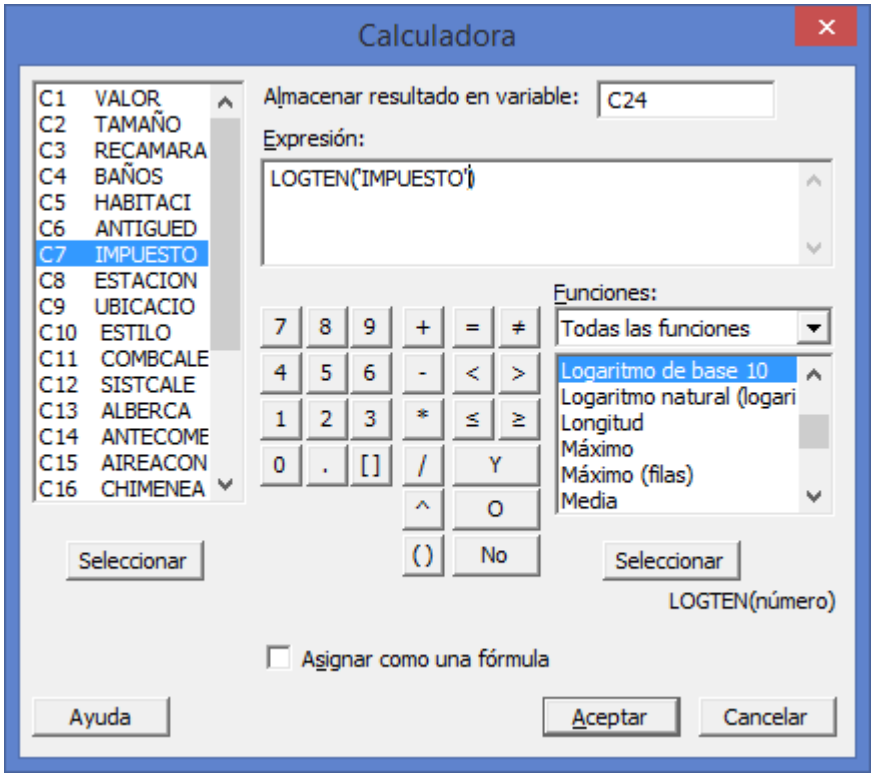
16 Elija **Calc > Calculadora**.

13 En **Almacenar resultado en variable**, ingrese ***C24***

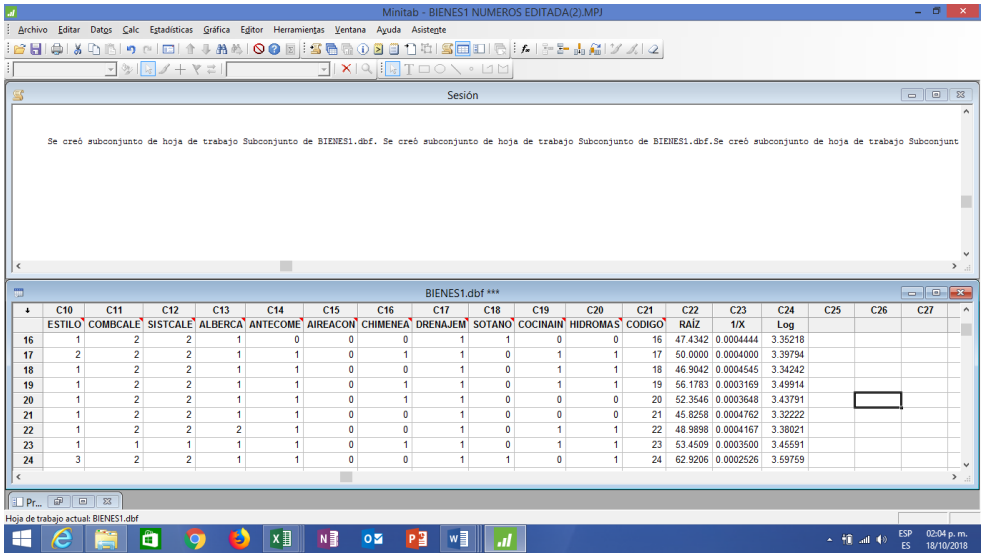
14 En **Expresión**, busque en **Funciones**: **Logaritmo base 10** y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



15 En **Expresión** aparece **LOGTEN(número).** Seleccione la variable **IMPUESTO** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.



16 Haga clic en **Aceptar.** En **C24** escriba el nombre **Log**



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ	1/X	Log			
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342	0.0004444	3.35218			
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000	0.0004000	3.39794			
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042	0.0004545	3.34242			
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783	0.0003169	3.49914			
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546	0.0003648	3.43791			
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258	0.0004762	3.32222			
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898	0.0004167	3.38021			
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509	0.0003500	3.45591			
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206	0.0002526	3.59759			

Repita la prueba de Levene para las TRES transformaciones para ver si se corrigió el problema de homocedasticidad y si fué así cual transformación resulto mejor.

A. CON LA RAÍZ CUADRADA DE IMPUESTO

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***RAÍZ***

4 En **Factores**, ingrese ***UBICACIÓN.*** Haga clic en **Aceptar**.

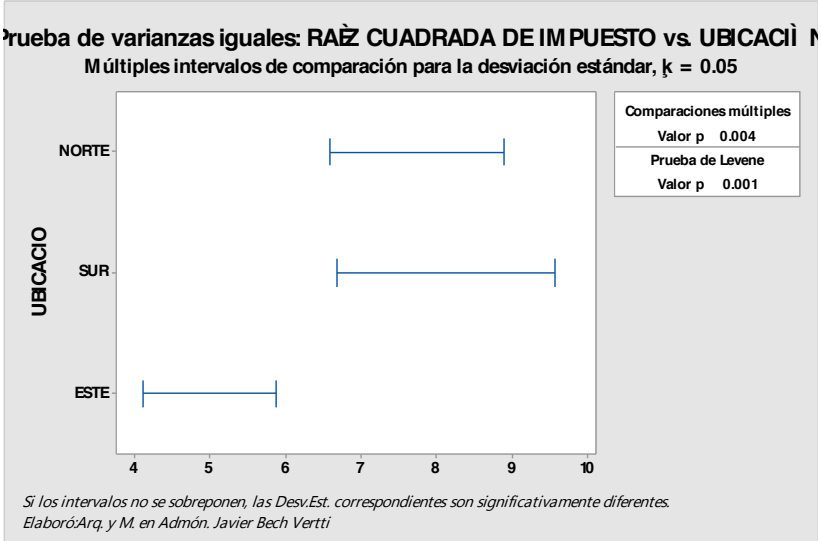
Salida de la ventana Sesión

Prueba de varianzas iguales: RAÍZ vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.          IC

       1  74    7.51925  (6.05261,  9.6536)

       2  60    7.83410  (6.15809, 10.3804)

       3  99    4.84238  (3.96630,  6.0585)

Nivel de confianza individual = 98.3333%


Pruebas

                          Estadística

Método                      de prueba  Valor p

Comparaciones múltiples           —    0.004

Levene                         7.53    0.001
```

Prueba de varianzas iguales: RAÍZ vs. UBICACIO

Salida de la ventana Gráfica



Para el ejemplo para evaluar la transformación de la **RAIZ CUADRADA** del **IMPUESTO** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.001 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

B. CON LA LA INVERSA DE IMPUESTO (1/X)

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***1/X***

4 En **Factores**, ingrese ***UBICACIÓN.*** Haga clic en **Aceptar**.

Salida de la ventana Sesión

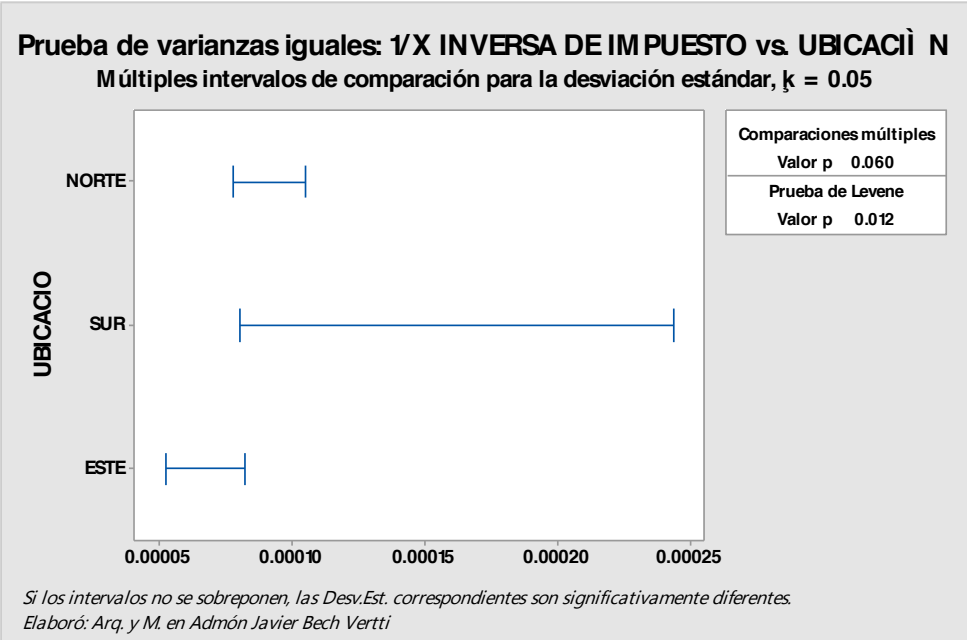
Prueba de varianzas iguales: 1/X vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.           IC

       1  74  0.0000890  (0.0000664, 0.0001232)
```

```
2   60   0.0001371   (0.0000815, 0.0002402)

3   99   0.0000649   (0.0000554, 0.0000779)


Nivel de confianza individual = 98.3333%

Pruebas

                          Estadística
Método                      de prueba  Valor p
Comparaciones múltiples             —    0.060
Levene                           4.55    0.012
```

Prueba de varianzas iguales: 1/X vs. UBICACIO

Salida de la ventana Gráfica



Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **INVERSA** del **IMPUESTO** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.012 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es

diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

C. CON EL LOGARITMO BASE 10 DE IMPUESTO

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***Log***

4 En **Factores**, ingrese ***UBICACIÓN.*** Haga clic en **Aceptar**.

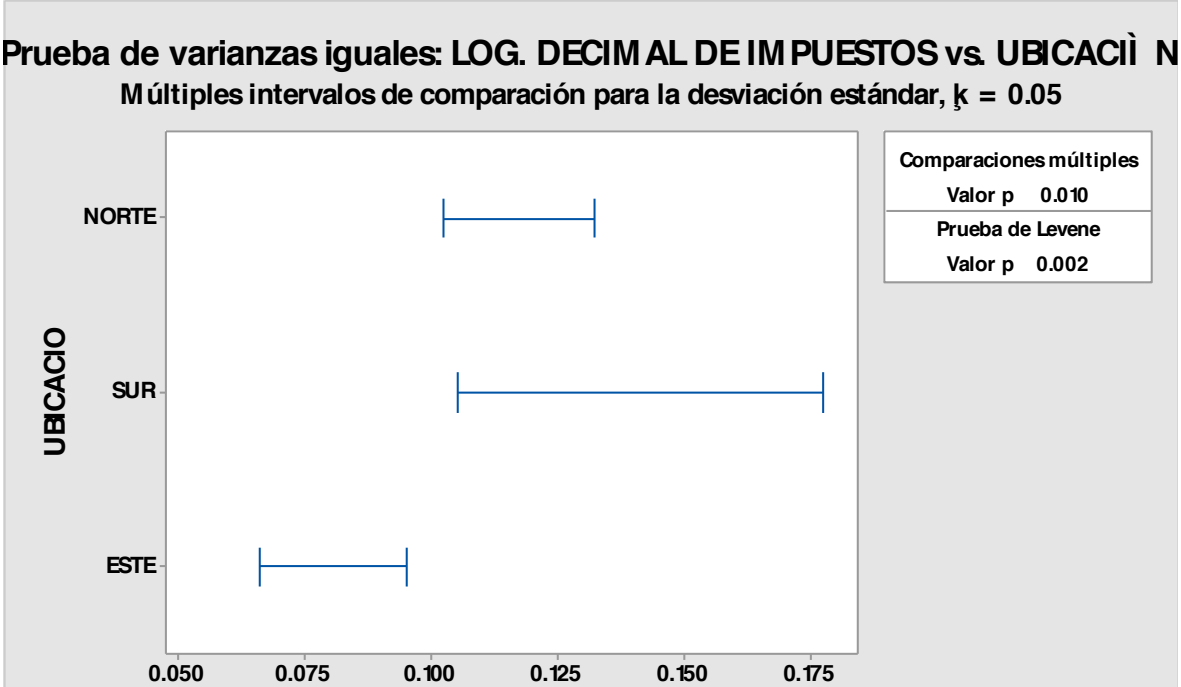
Salida de la ventana Sesión

Prueba de varianzas iguales: Log vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.            IC

       1  74   0.114628  (0.0928061, 0.146315)

       2  60   0.134033  (0.0981559, 0.190631)

       3  99   0.078315  (0.0657447, 0.095601)


Nivel de confianza individual = 98.3333%

Pruebas

                          Estadística

Método                      de prueba  Valor p

Comparaciones múltiples            —    0.010

Levene                          6.38    0.002
```

Prueba de varianzas iguales: Log vs. UBICACIO

Salida de la ventana Gráfica



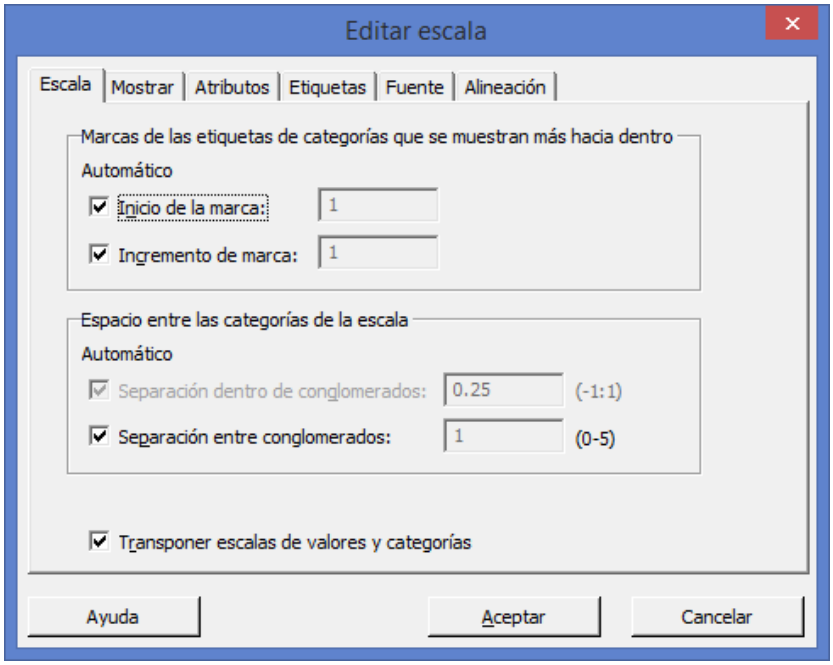
Interpretación de los resultados

Para el ejemplo para evaluar la transformación del **LOGARITMO DECIMAL** del **IMPUESTO** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.002 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

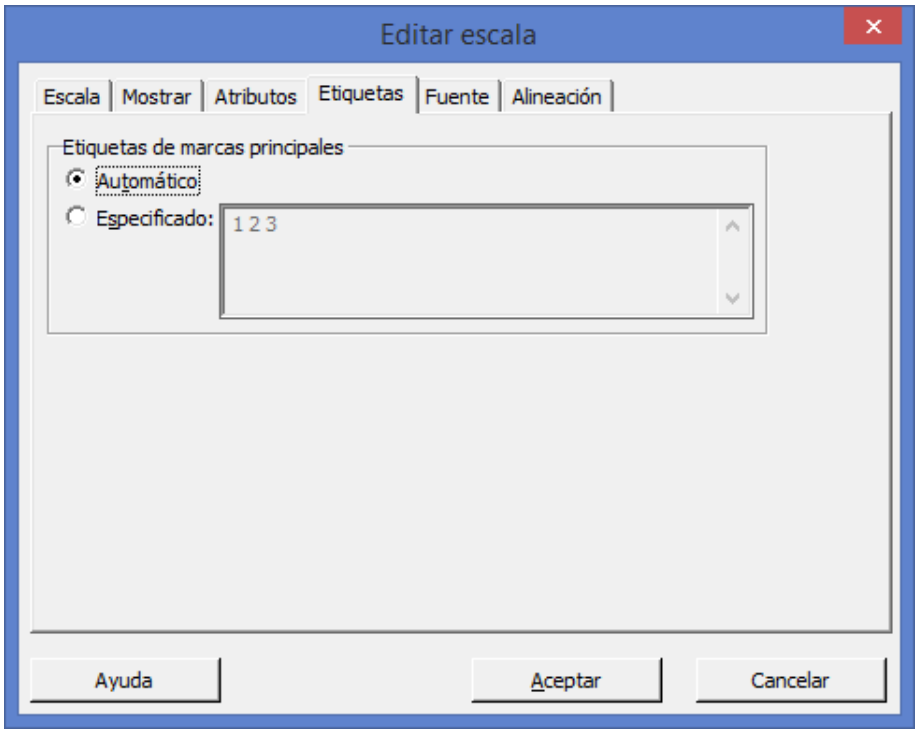
Nota Importante: No es del alcance de este trabajo tomar la decisión de si se debe o no trabajar con los datos transformados o sin transformar para usar alguna técnica multivariada sino solamente mostrar la forma de hacer dichas transformaciones.

ETIQUETADO DE LOS NIVELES DE LA VARIABLE O "FACTOR" UBICACIÓN EN MINITAB 17:

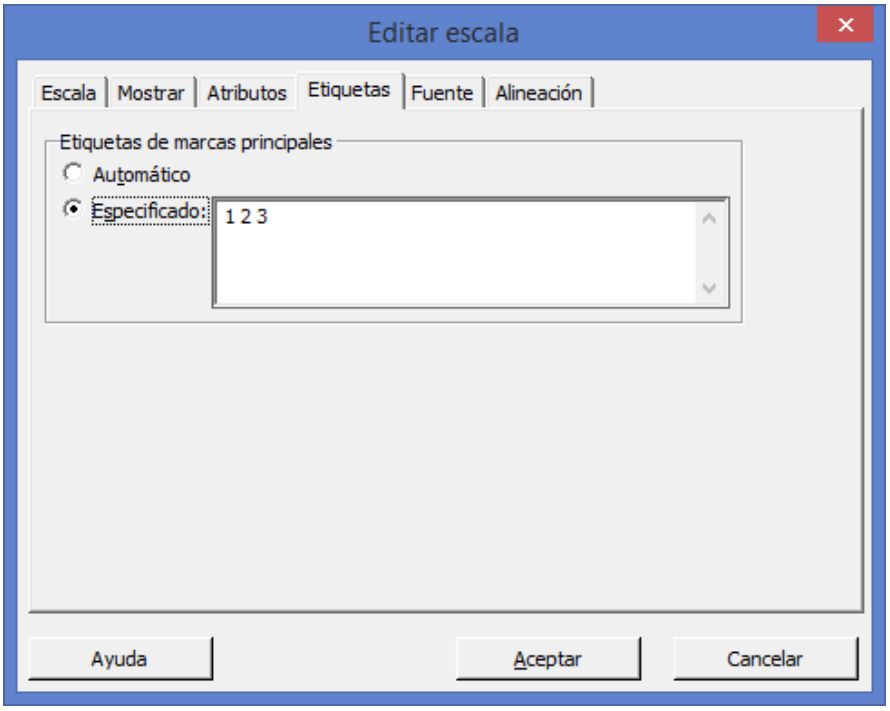
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **UBICACIÓN** para que Minitab despliegue la ventana ***Editar Escala.***



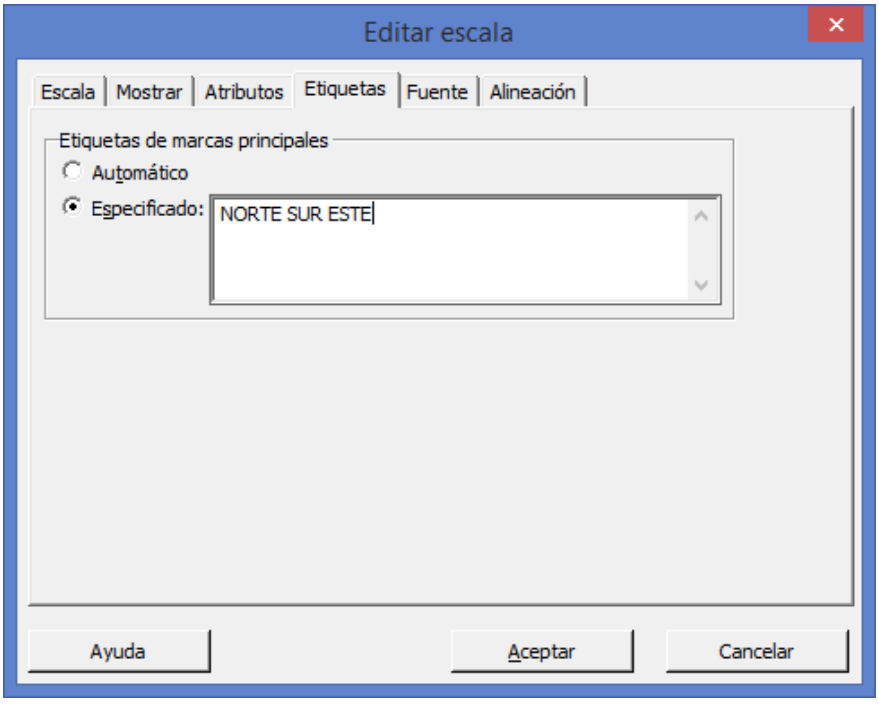
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

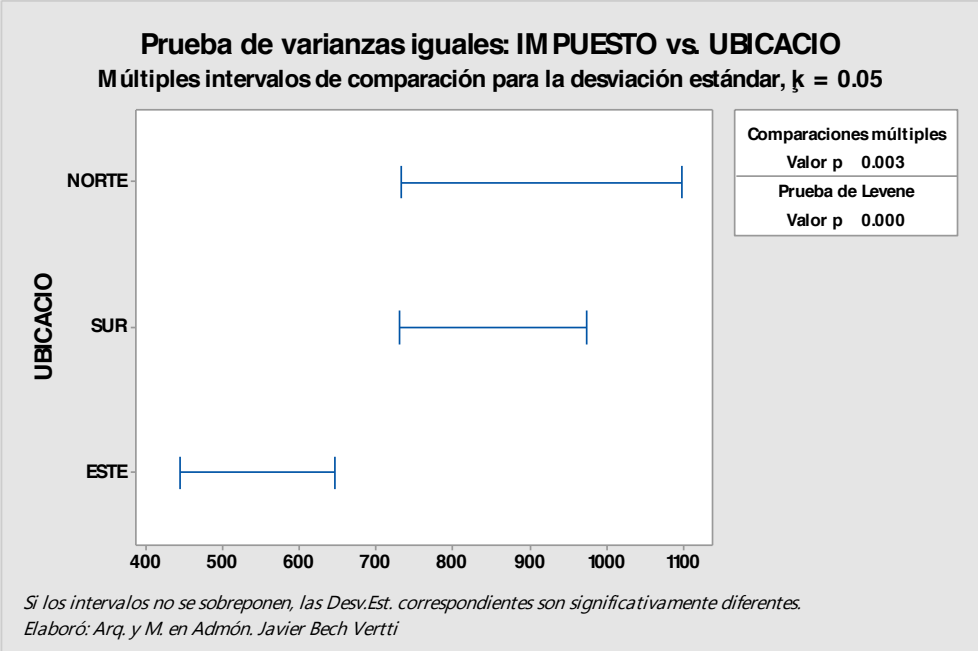


4 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



1.3.2.2.3 EJERCICIO DE **REFUERZO.O**

EJERCICIO DE **REFUERZO 1.3.2.2.3**

PRUEBA **DE HOMOCEDASTICIDAD. PRUEBA DE LEVENE. TRANSFORMACIONES. ETIQUETADO DE GRÁFICOS.**

1.3.2.2.3 Se pretende diseñar un experimento para evaluar el **IMPUESTO** que pagan 233 casas en diferentes zonas de la ciudad. Se mide el **IMPUESTO** de las casas de la base de datos **"BIENES 1"** que se construyó con anterioridad con respecto a si tienen o no **CHIMENEA**. Como se desea probar la igualdad de las medias y evaluar las diferencias de las medias, se utiliza el procedimiento **ANOVA** de un factor (datos apilados) con comparaciones múltiples. Por lo general, se escogería un método de comparación múltiple apropiado para los datos.

Antes de realizar el **análisis de varianza,** verifica la suposición de que la varianza es igual en todos los niveles del Factor **CHIMENEA** utilizando la Prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones continuas, **pero no necesariamente normales** (Prueba de **Levene**) utilizando el paquete de cómputo **Minitab:**

Utilice el paquete Minitab para resolver este Ejercicio de Refuerzo de Igualdad de Varianzas e interprete su resultado, realice las Transformaciones de Raíz Cuadrada, Inversa y Logaritmo

decimal repitiendo las pruebas Levene para cada transformación e interprete los resultados y etiquete debidamente todos los gráficos de presentación.

PRUEBA DE LEVENE EN MINITAB 17

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

Nota importante: Verifica que en la ventana de Prueba de varianzas iguales el primer campo diga "Los datos de respuesta están en una columna para todos los niveles de los factores"

3 En **Respuesta**, ingrese ***IMPUESTO***

4 En **Factores**, ingrese ***CHIMENEA***. Haga clic en **Aceptar**.

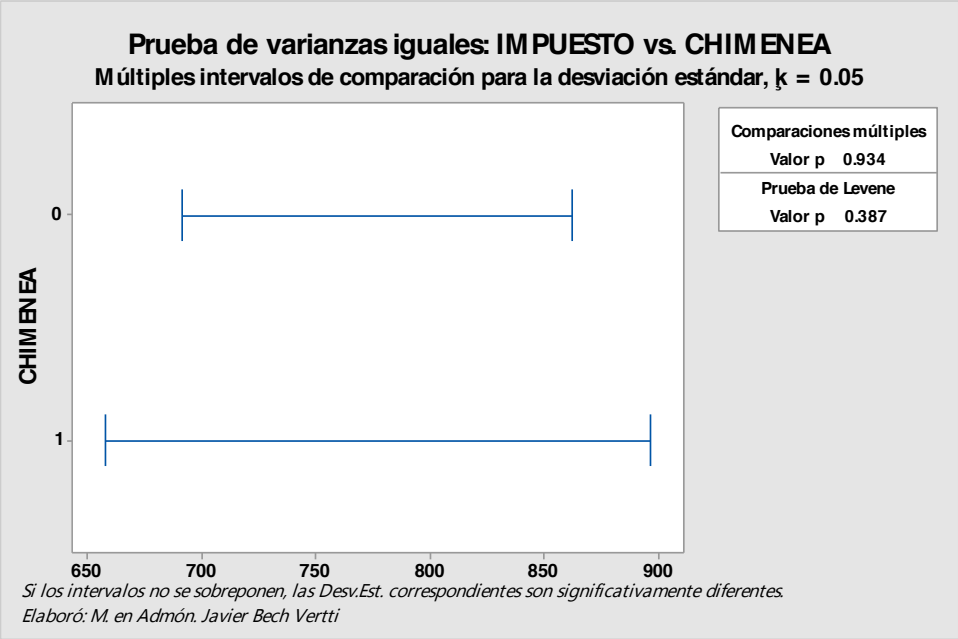
Salida de la ventana Sesión

Prueba de varianzas iguales: IMPUESTO vs. CHIMENEA

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA    N  Desv.Est.          IC

       0  154    767.169  (669.211,  892.46)

       1   79    758.403  (550.899, 1074.55)

Nivel de confianza individual = 97.5%

Pruebas


                         Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.934

Levene                          0.75    0.387
```

Prueba de varianzas iguales: IMPUESTO vs. CHIMENEA

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de las pruebas **Levene** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **IMPUESTO** de las casas con base en si tienen o no **CHIMENEA**, los **valores de p** de 0.387 es mayor que la opción razonable de **alfa**, así que no se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

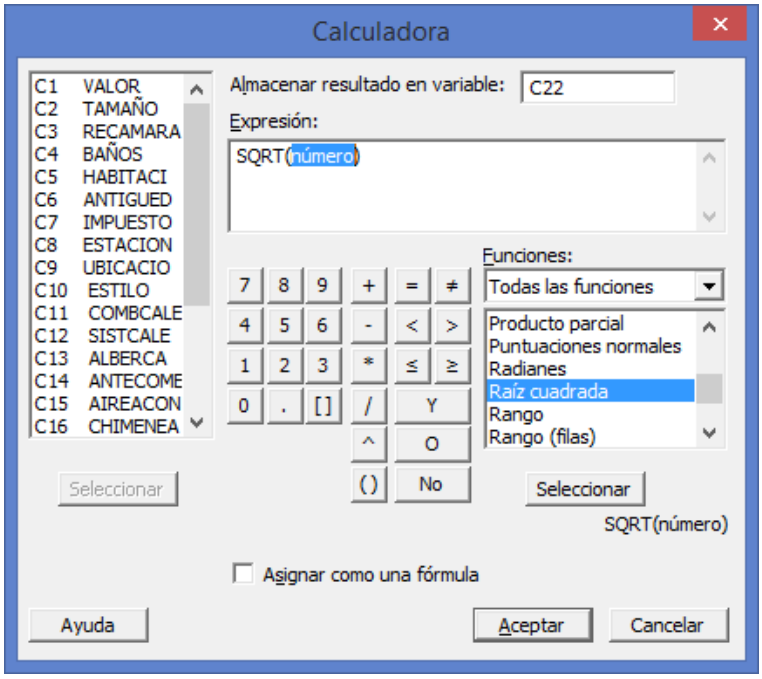
TRANSFORMACIONES EN MINITAB 17

Transformaciónes Raiz Cuadrada, 1/X y Log de la variable IMPUESTO con Minitab:

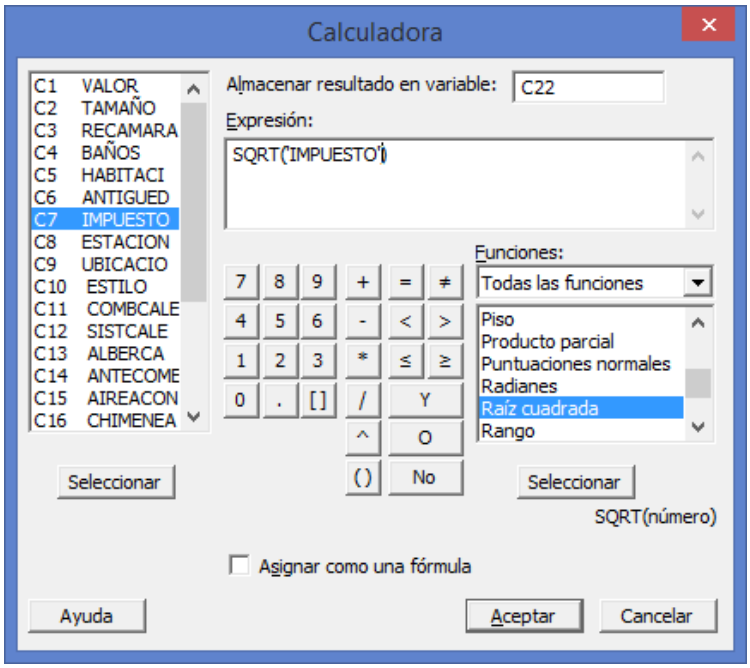
1 Abra la base de datos que creó **"BIENES1"**
2 Elija **Calc > Calculadora**.

3 En **Almacenar resultado en variable**, ingrese *C22*

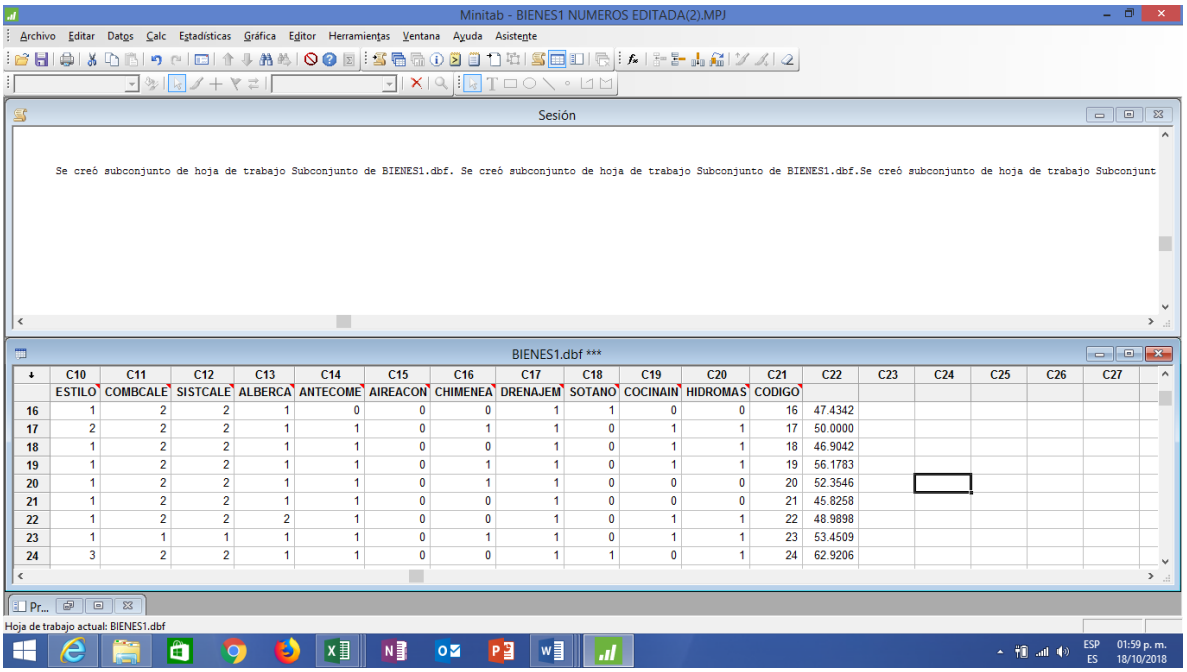
4 En **Expresión**, busque en **Funciones**: Raíz Cuadrada y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



5 En **Expresión** aparece **SQRT(número).** Seleccione la variable **IMPUESTO** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.

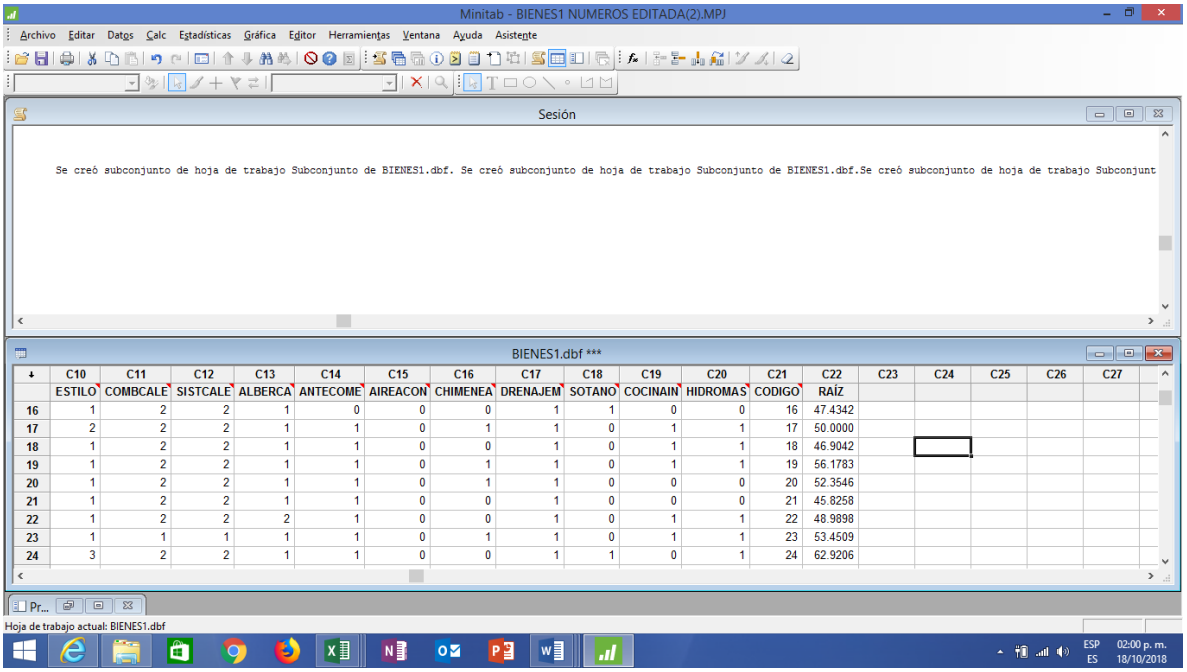


6 Haga clic en **Aceptar**.



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO						
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342					
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000					
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042					
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783					
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546					
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258					
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898					
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509					
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206					

7 En C22 teclee el nombre **RAÍZ**

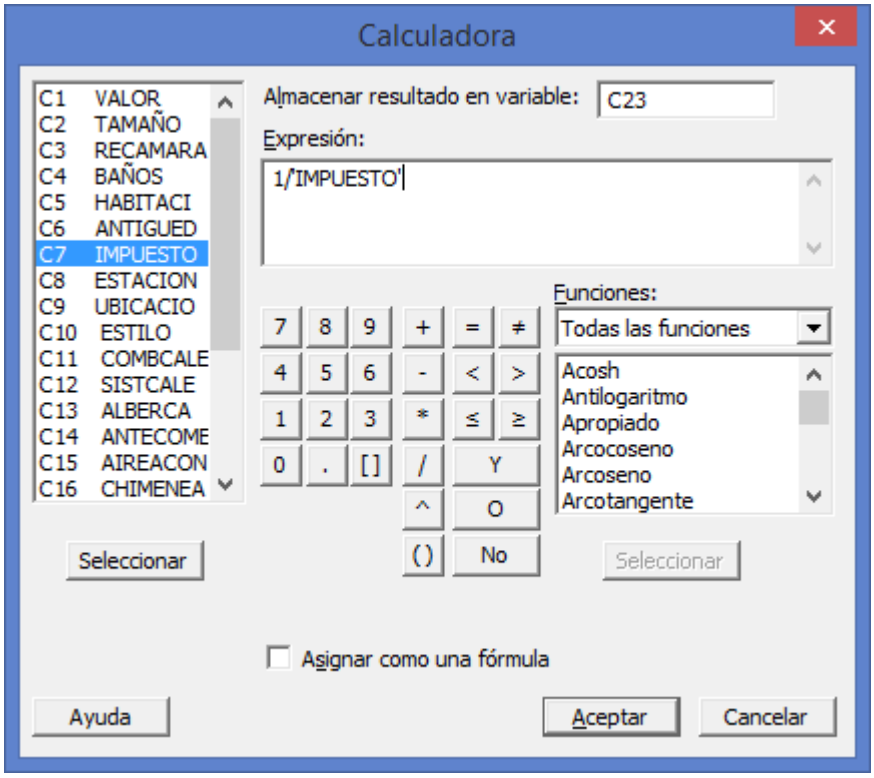


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ					
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342					
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000					
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042					
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783					
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546					
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258					
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898					
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509					
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206					

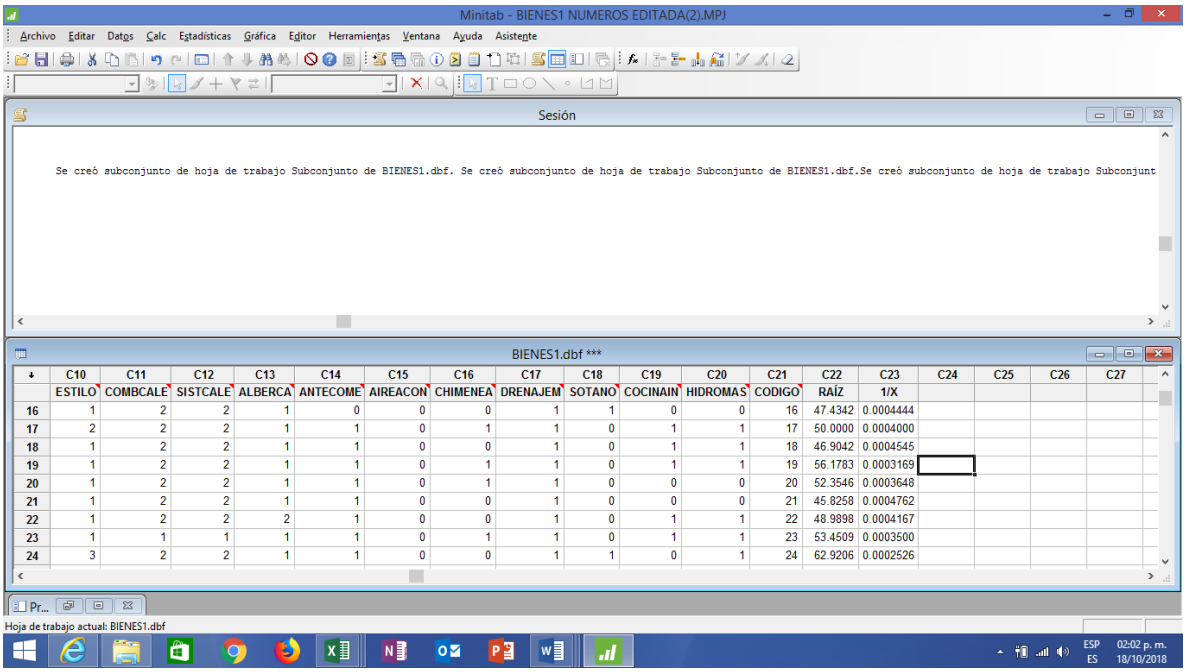
8 Elija **Calc > Calculadora**.

9 En **Almacenar resultado en variable**, ingrese ***C23***

10 En **Expresión**, para obtener la Inversa, **Teclee 1/** y seleccione la variable **IMPUESTO** del lado izquierdo o doble clic.



11 Haga clic en **Aceptar**. En **C23** escriba el nombre **1/X**

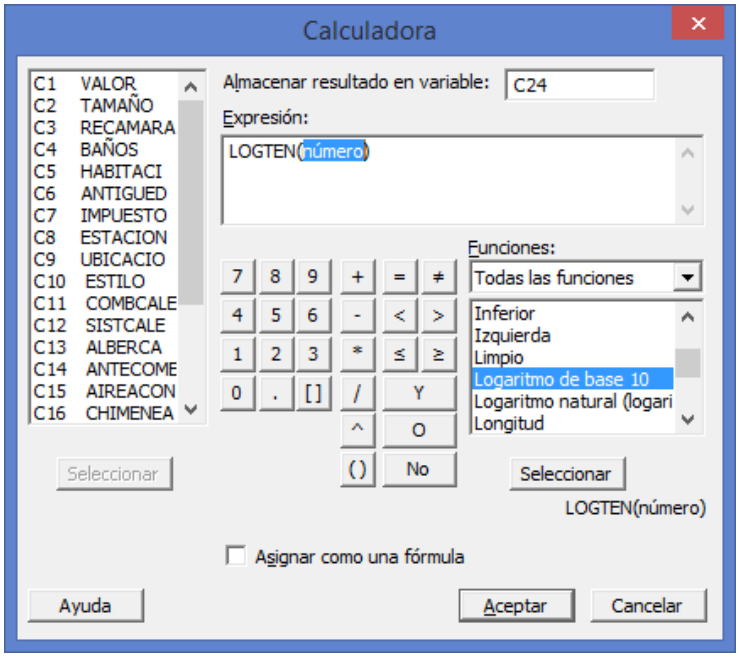


	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ	1/X				
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342	0.0004444				
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000	0.0004000				
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042	0.0004545				
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783	0.0003169				
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546	0.0003648				
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258	0.0004762				
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898	0.0004167				
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509	0.0003500				
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206	0.0002526				

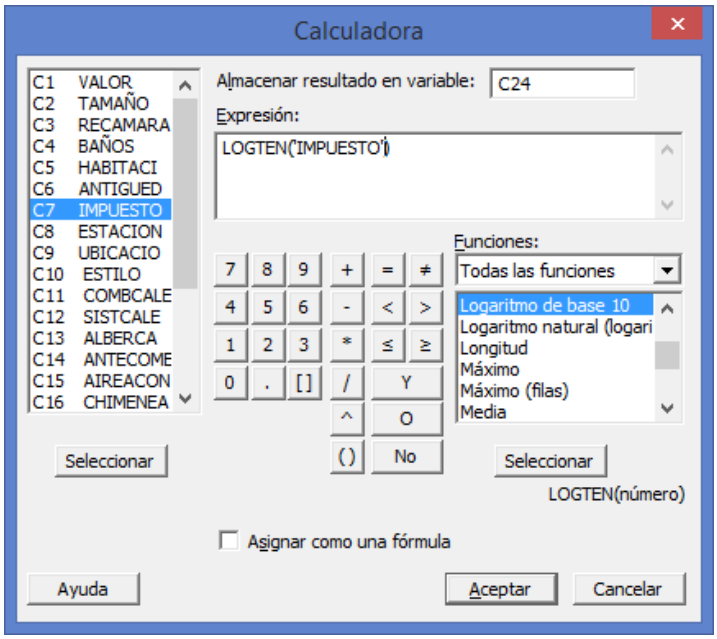
12 Elija **Calc > Calculadora**.

13 En **Almacenar resultado en variable**, ingrese ***C24***

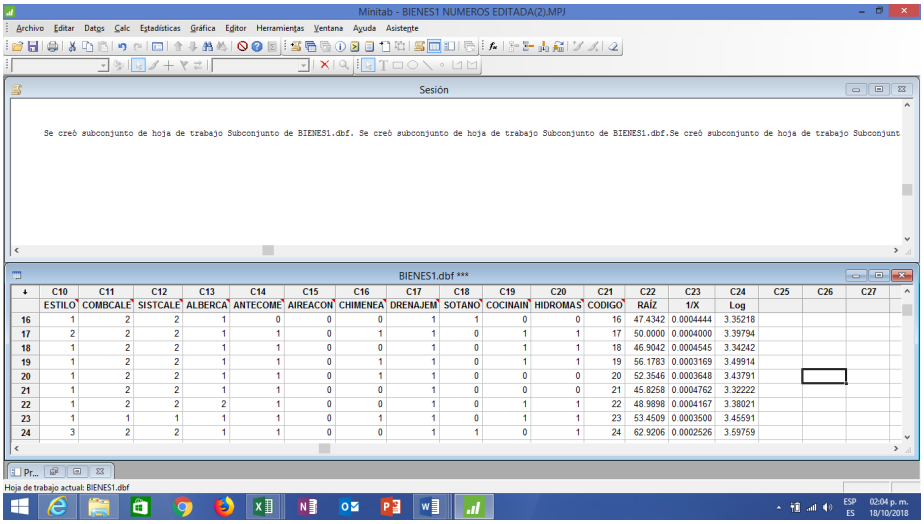
14 En **Expresión**, busque en **Funciones**: **Logaritmo base 10** y selecciónela. Haga clic en el botón S**eleccionar** o simplemente **doble clic**.



15 En **Expresión** aparece **LOGTEN(número).** Seleccione la variable **IMPUESTO** del lado izquierdo y haga clic en el botón **Seleccionar** o doble clic.



16 Haga clic en **Aceptar.** En **C24** escriba el nombre **Log**



	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24	C25	C26	C27
	ESTILO	COMBCALE	SISTCALE	ALBERCA	ANTECOME	AIREACON	CHIMENEA	DRENAJEM	SOTANO	COCINAIN	HIDROMAS	CODIGO	RAÍZ	1/X	Log			
16	1	2	2	1	0	0	0	1	1	0	0	16	47.4342	0.0004444	3.35218			
17	2	2	2	1	1	0	1	1	0	1	1	17	50.0000	0.0004000	3.39794			
18	1	2	2	1	1	0	0	1	0	1	1	18	46.9042	0.0004545	3.34242			
19	1	2	2	1	1	0	1	1	0	1	1	19	56.1783	0.0003169	3.49914			
20	1	2	2	1	1	0	1	1	0	0	0	20	52.3546	0.0003648	3.43791			
21	1	2	2	1	1	0	0	1	0	0	0	21	45.8258	0.0004762	3.32222			
22	1	2	2	2	1	0	0	1	0	1	1	22	48.9898	0.0004167	3.38021			
23	1	1	1	1	1	0	1	1	0	1	1	23	53.4509	0.0003500	3.45591			
24	3	2	2	1	1	0	0	1	1	0	1	24	62.9206	0.0002526	3.59759			

Repita la prueba de Levene para las TRES transformaciones para ver si se corrigió el problema de homocedasticidad y si fué así cual transformación resulto mejor.

A. CON LA RAÍZ CUADRADA DE IMPUESTO

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***RAÍZ***

4 En **Factores**, ingrese ***CHIMENEA.*** Haga clic en **Aceptar**.

Salida de la ventana Sesión

Prueba de varianzas iguales: RAÍZ vs. CHIMENEA

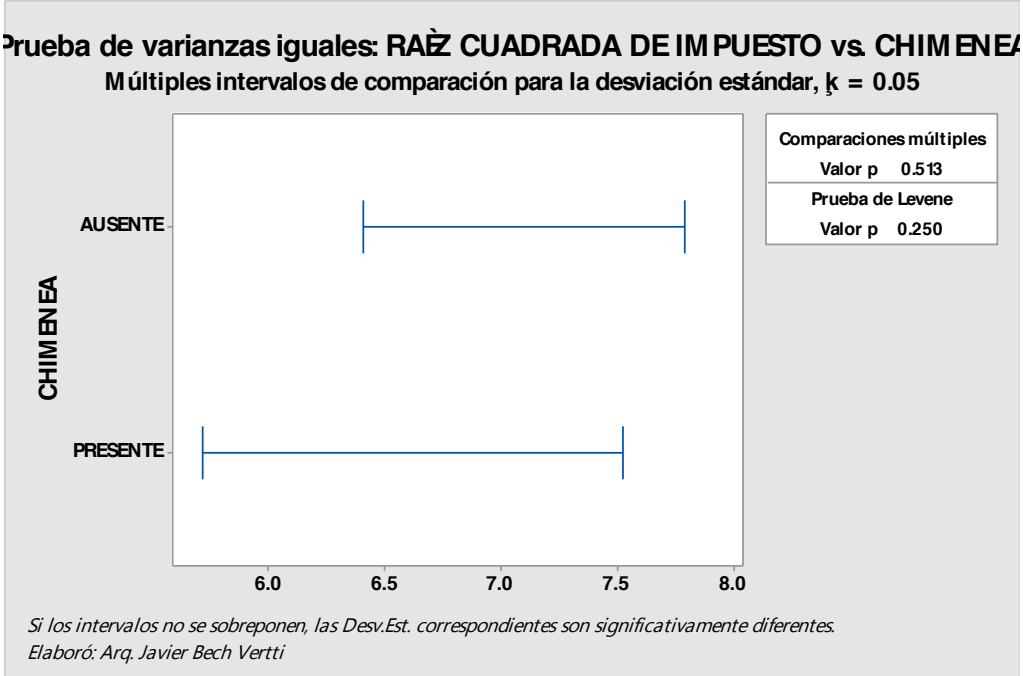
```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA    N  Desv.Est.          IC

       0  154    7.01817  (6.11823, 8.16940)

       1   79    6.47441  (5.01941, 8.59505)

Nivel de confianza individual = 97.5%
```

Pruebas

Método	Estadística de prueba	Valor p
Comparaciones múltiples	—	0.513
Levene	1.33	0.250

Prueba de varianzas iguales: RAÍZ vs. CHIMENEA

Salida de la ventana Gráfica



Para el ejemplo para evaluar la transformación de la **RAIZ CUADRADA** del **IMPUESTO** de las casas con base en si tienen o no **CHIMENEA**, los **valores de p** de 0.250 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

B. CON LA LA INVERSA DE IMPUESTO (1/X)

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***1/X***

4 En **Factores**, ingrese ***CHIMENEA.*** Haga clic en **Aceptar**.

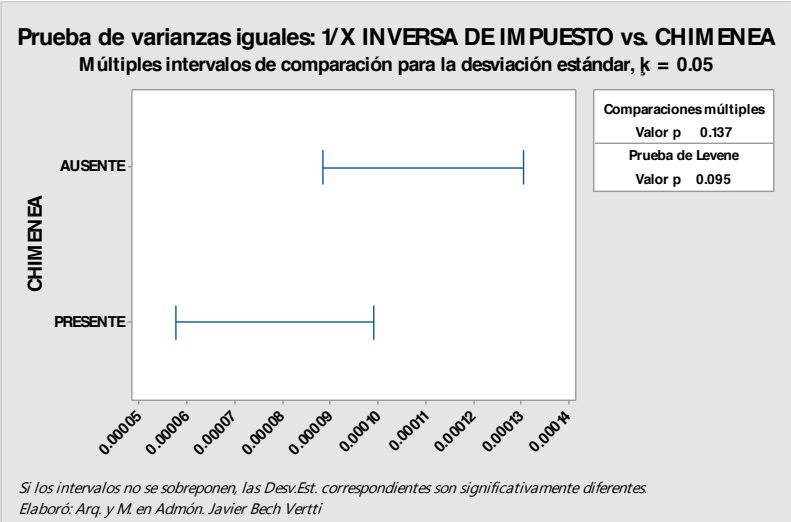
Salida de la ventana Sesión

Prueba de varianzas iguales: 1/X vs. CHIMENEA

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA    N  Desv.Est.            IC

       0  154  0.0001067  (0.0000785, 0.0001473)

       1   79  0.0000745  (0.0000617, 0.0000925)

Nivel de confianza individual = 97.5%

Pruebas

                         Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.137

Levene                          2.81    0.095
```

Prueba de varianzas iguales: 1/X vs. CHIMENEA

Salida de la ventana Gráfica



Interpretación de los resultados

Para el ejemplo para evaluar la transformación de la **INVERSA** del **IMPUESTO** de las casas con base en si tienen o no **CHIMENEA**, los **valores de p** de 0.095 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

C. CON EL LOGARITMO BASE 10 DE IMPUESTO

1 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

2 En **Respuesta**, ingrese ***Log***

3 En **Factores**, ingrese ***CHIMENEA***. Haga clic en **Aceptar**.

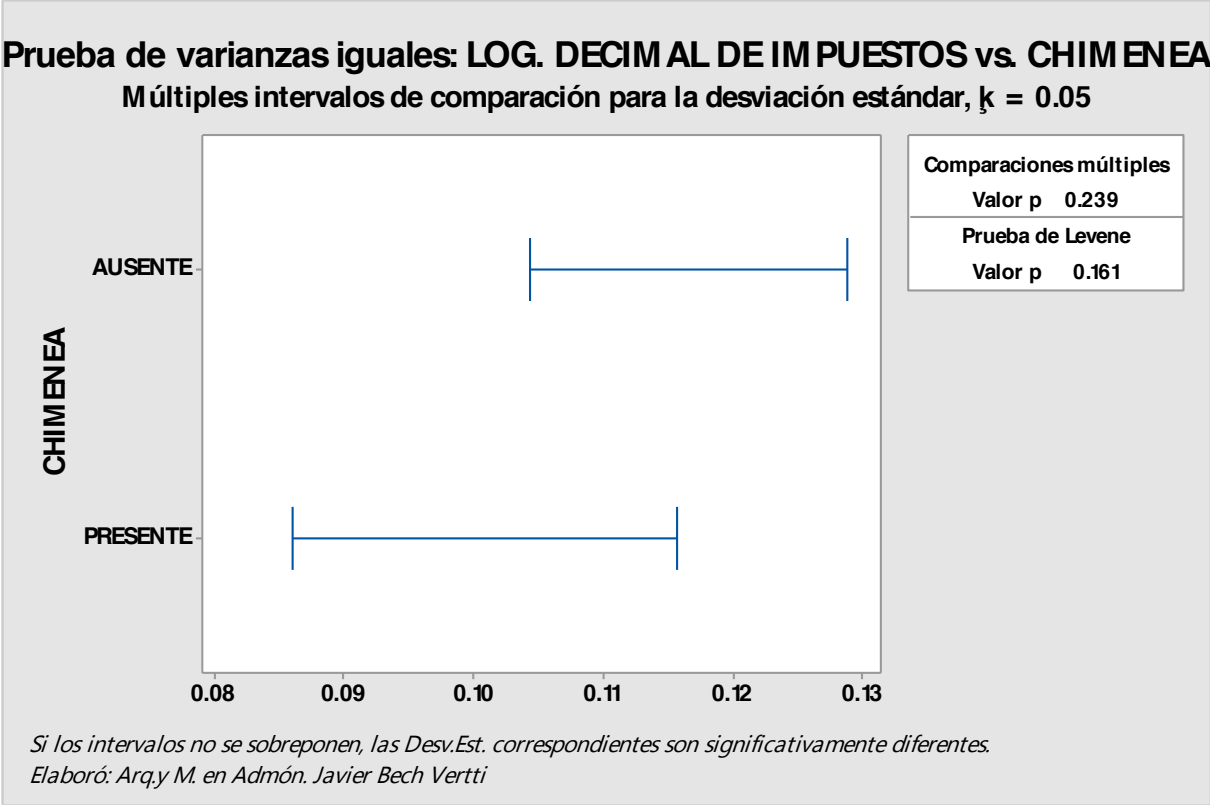
Salida de la ventana Sesión

Prueba de varianzas iguales: Log vs. CHIMENEA

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

CHIMENEA    N  Desv.Est.            IC

       0  154   0.115161  (0.0976676, 0.137794)

       1   79   0.098484  (0.0798916, 0.124948)

Nivel de confianza individual = 97.5%

Pruebas

                         Estadística

Método                     de prueba  Valor p

Comparaciones múltiples            —    0.239

Levene                          1.98    0.161
```

Prueba de varianzas iguales: Log vs. CHIMENEA

Salida de la ventana Gráfica



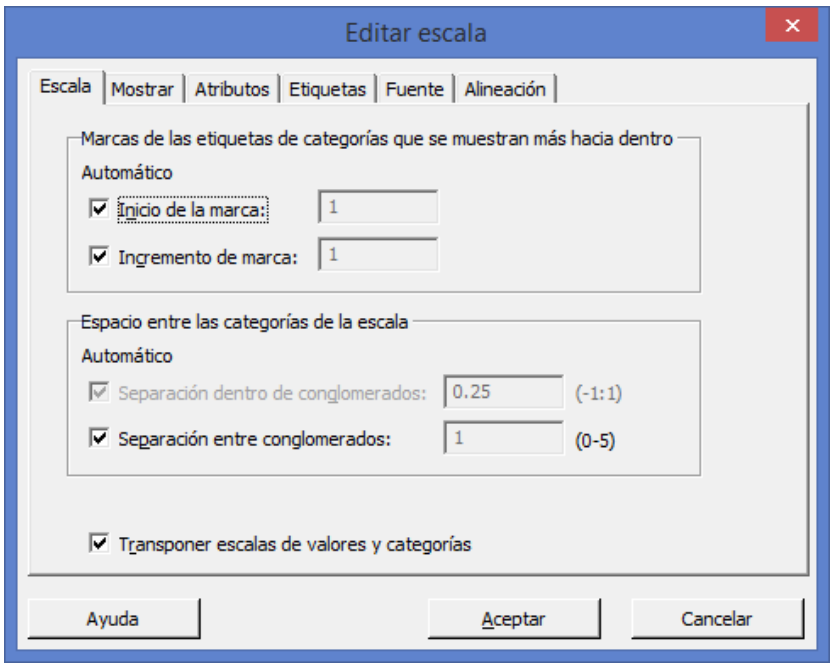
Interpretación de los resultados

Para el ejemplo para evaluar la transformación del **LOGARITMO DECIMAL** del **IMPUESTO** de las casas con base en si tienen o no **CHIMENEA**, los **valores de p** de 0.161 es mayor que la opción razonable de **alfa**, así que **NO** se puede rechazar la **hipótesis nula** de que las varianzas son iguales. Es decir, estos datos proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

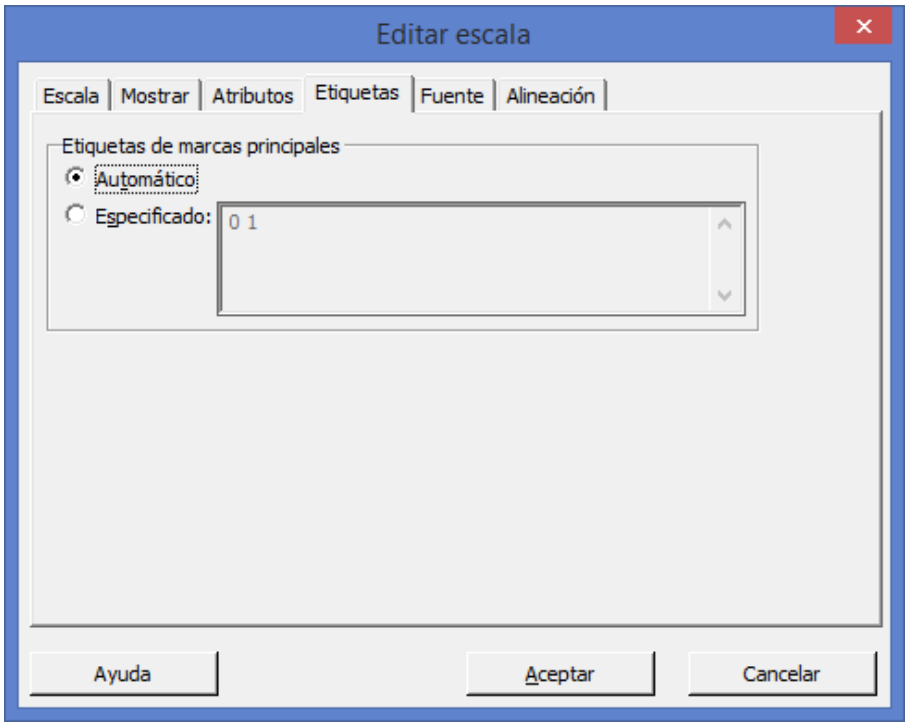
Nota Importante: No es del alcance de este trabajo tomar la decisión de si se debe o no trabajar con los datos transformados o sin transformar para usar alguna técnica multivariada sino solamente mostrar la forma de hacer dichas transformaciones.

ETIQUETADO DE LOS NIVELES DE LA VARIABLE O "FACTOR" CHIMENEA EN MINITAB 17:

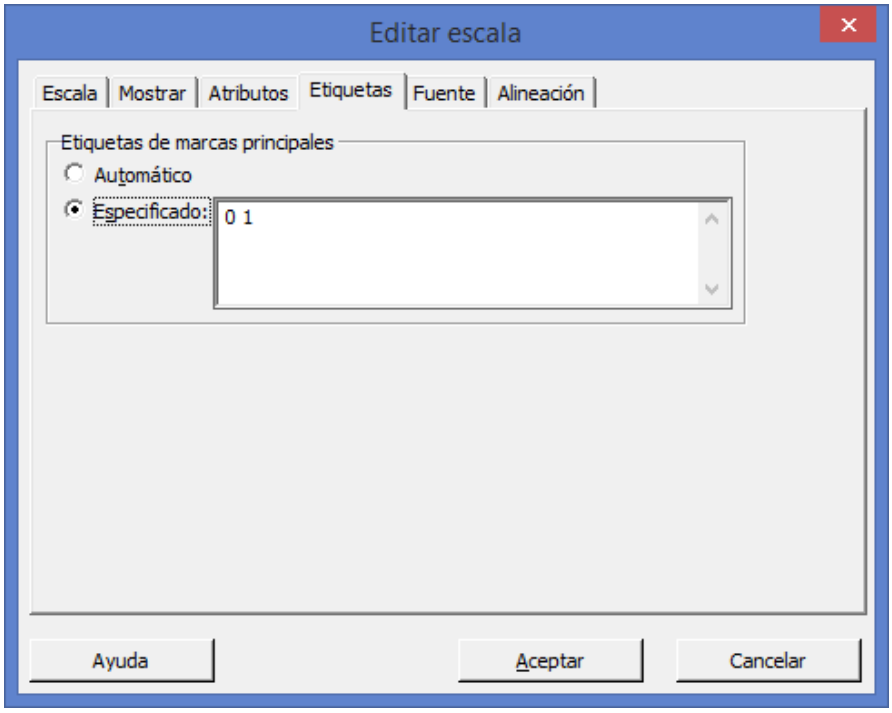
1 Haga doble clic en el número 0 del eje vertical de la gráfica que dice **CHIMENEA** para que Minitab despliegue la ventana ***Editar Escala.***



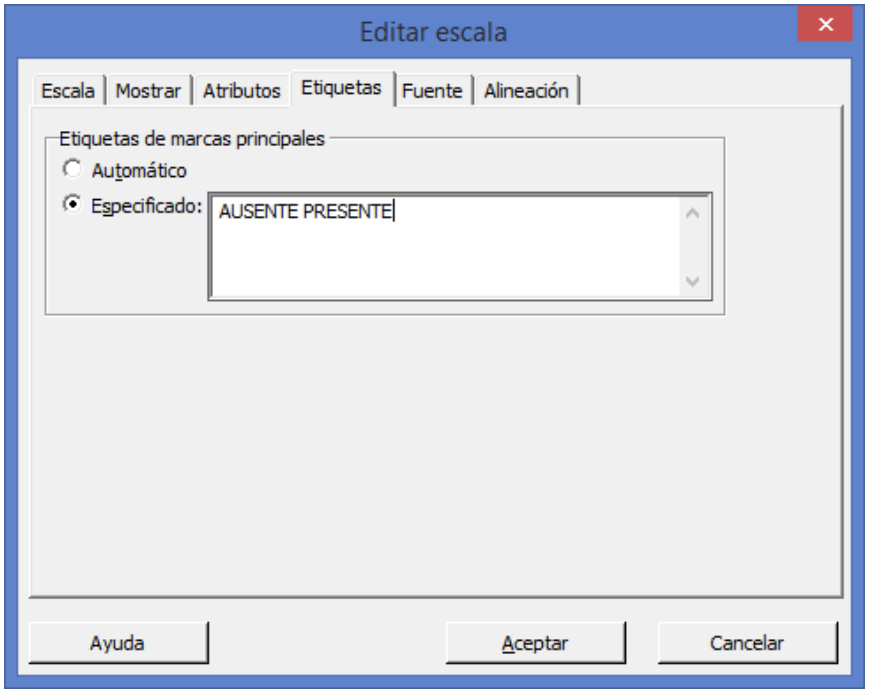
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

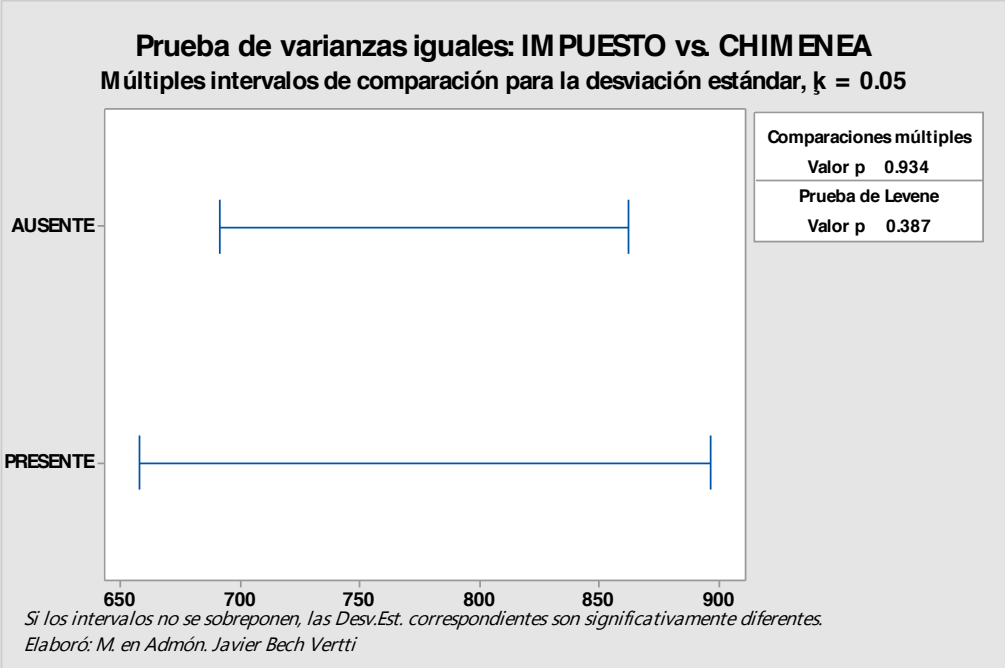


4 Borrar los números 0 1 y escribir **AUSENTE PRESENTE**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



1.3.2.3.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.2.3.1**

PRUEBA **DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT**

1.3.2.3.1 Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **IMPUESTO** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con **ESTILO** diferente. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas.

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***IMPUESTO***

4 En **Factores**, ingrese ***ESTILO***

5 En el botón **Opciones** active la casilla que dice: ***Utilice la prueba basándose en la distribución normal.*** Haga clic en **Aceptar** y nuevamente **Aceptar**

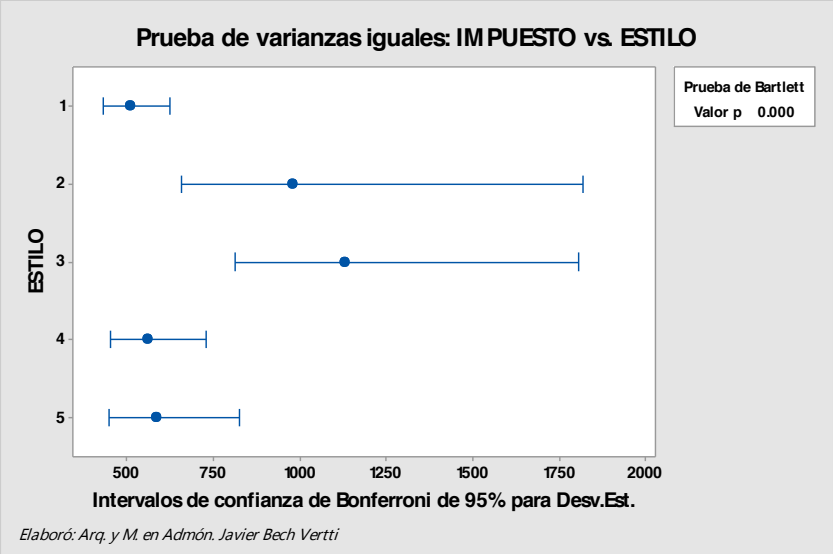
Salida de la ventana Sesión

Prueba de varianzas iguales: IMPUESTO vs. ESTILO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Se utiliza el método de Bartlett. Este método es exacto sólo para datos

     normales.

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

ESTILO   N  Desv.Est.          IC

     1  97     509.20  (428.708,  623.33)

     2  15     979.79  (655.071, 1816.14)

     3  23    1130.71  (810.707, 1804.01)

     4  60     558.92  (450.748,  728.06)

     5  38     582.90  (447.123,  822.44)

Nivel de confianza individual = 99%

Pruebas


          Estadística

Método      de prueba  Valor p

Bartlett        39.02    0.000
```

Prueba de varianzas iguales: IMPUESTO vs. ESTILO

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de la prueba de **Barlett** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **IMPUESTO** de las casas con base en el **ESTILO** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

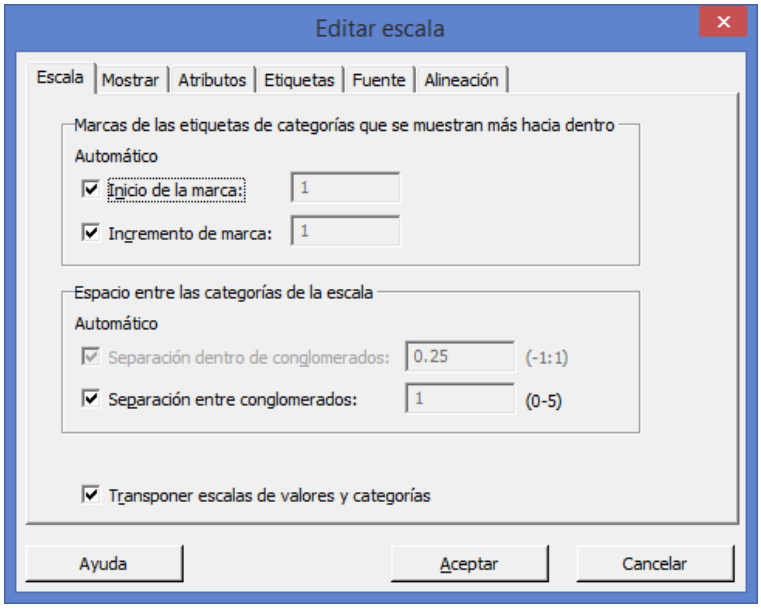
Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

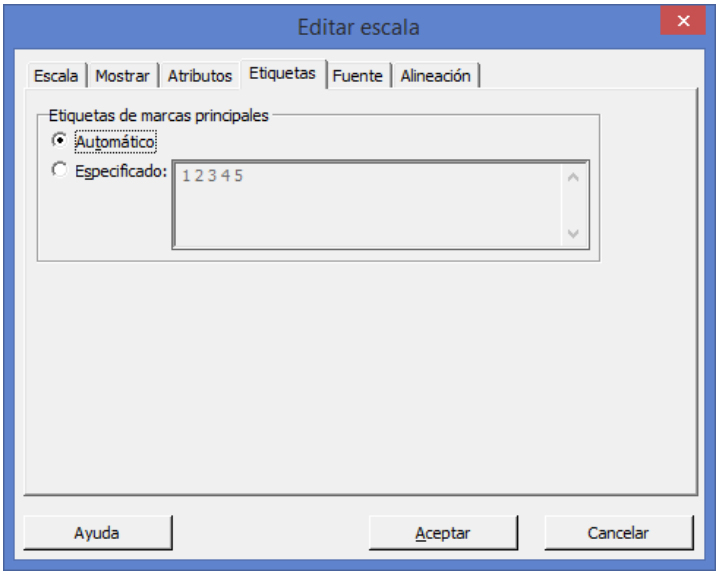
Asímismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución y en la falta de homocedasticidad.

Etiquetado de los niveles de la variable o "Factor" ESTILO:

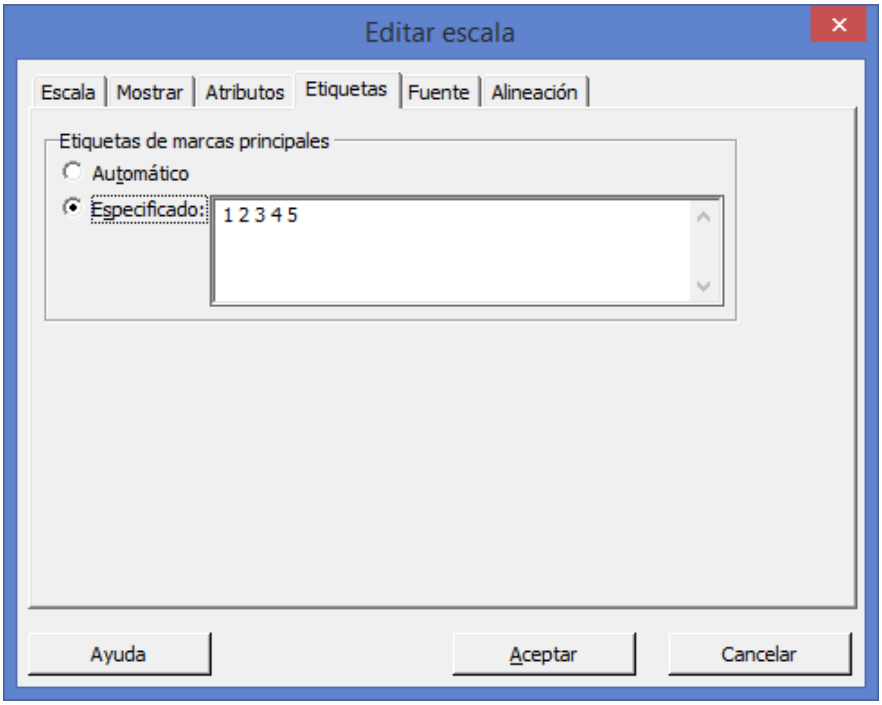
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **ESTILO** para que Minitab despliegue la ventana ***Editar Escala.***



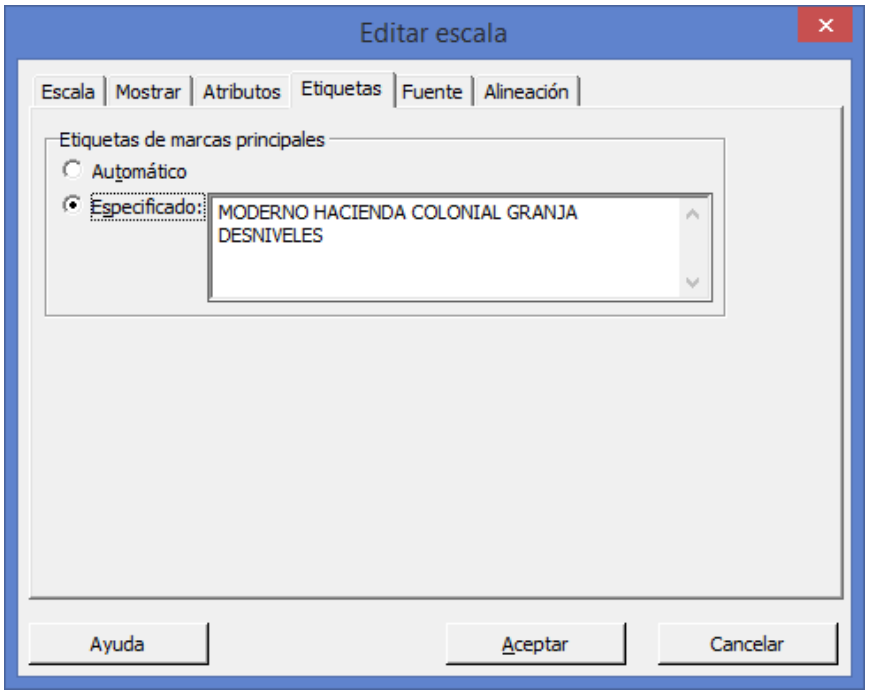
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

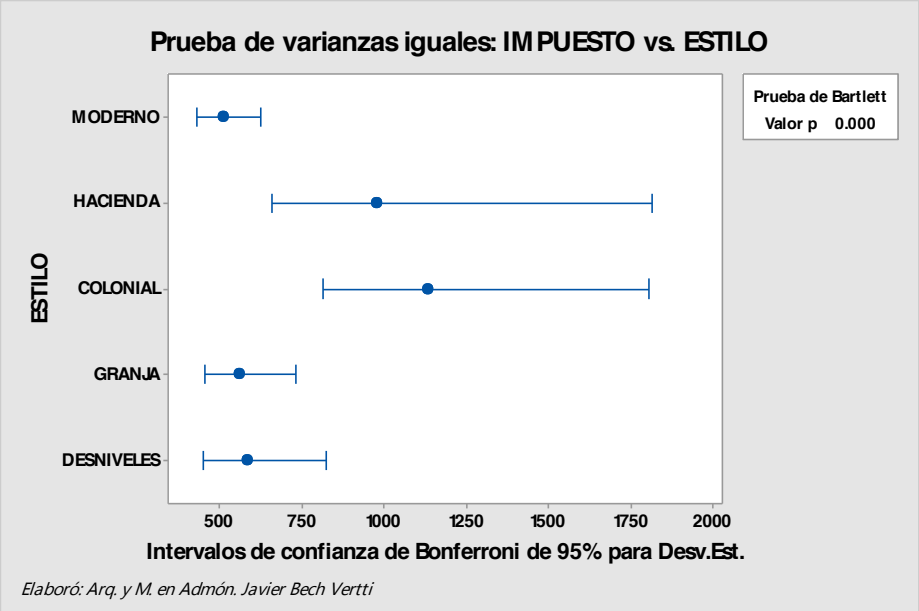


4 Borrar los números 1 2 3 4 5 y escribir **MODERNO HACIENDA COLONIAL GRANJA DESNIVELES**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



1.3.2.3.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.3.2.3.2**

PRUEBA **DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT**

1.3.2.3.2 Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **IMPUESTO** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con base en el **IMPUESTO** que pagan las casas. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas.

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***IMPUESTO***

4 En **Factores**, ingrese ***UBICACIÓN***

5 En el botón **Opciones** active la casilla que dice: ***Utilice la prueba basándose en la distribución normal.*** Haga clic en **Aceptar** y nuevamente **Aceptar**

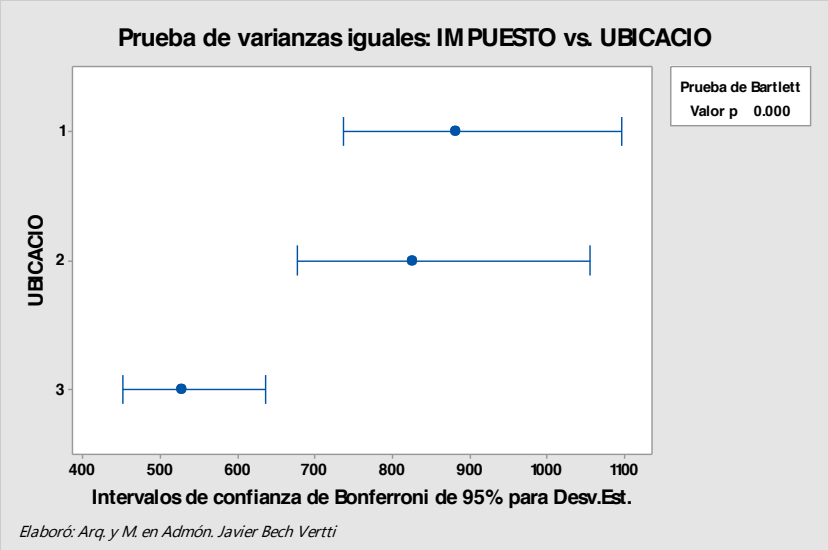
Salida de la ventana Sesión

Prueba de varianzas iguales: IMPUESTO vs. UBICACIO

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05


Se utiliza el método de Bartlett. Este método es exacto sólo para datos

     normales.

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

UBICACIO   N  Desv.Est.          IC

       1  74    882.083  (735.332, 1096.45)

       2  60    825.970  (675.858, 1054.93)

       3  99    527.133  (449.724,  634.43)


Nivel de confianza individual = 98.3333%

Pruebas


          Estadística

Método      de prueba  Valor p

Bartlett        24.66    0.000
```

Prueba de varianzas iguales: IMPUESTO vs. UBICACIO

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de la prueba de **Barlett** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **IMPUESTO** de las casas con base en la **UBICACIÓN** de las mismas, los **valores de p** de 0.000 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

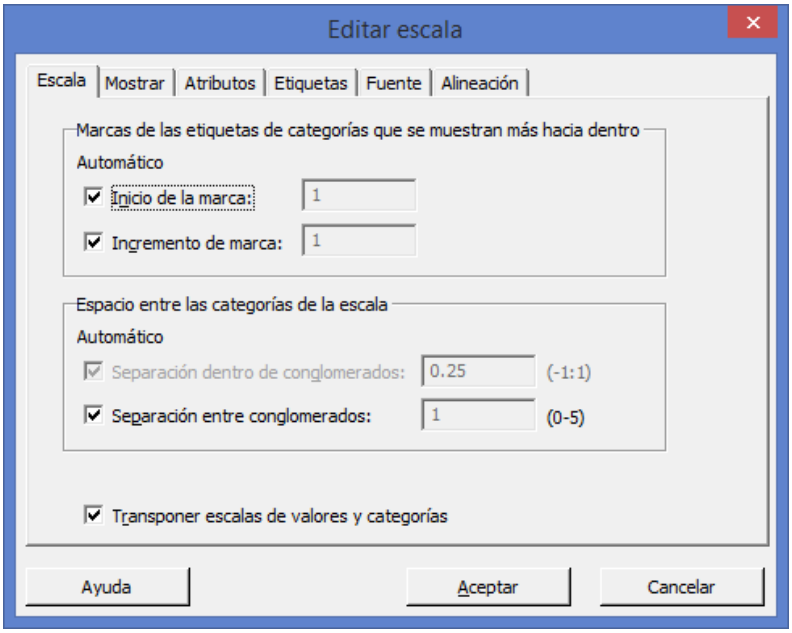
Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

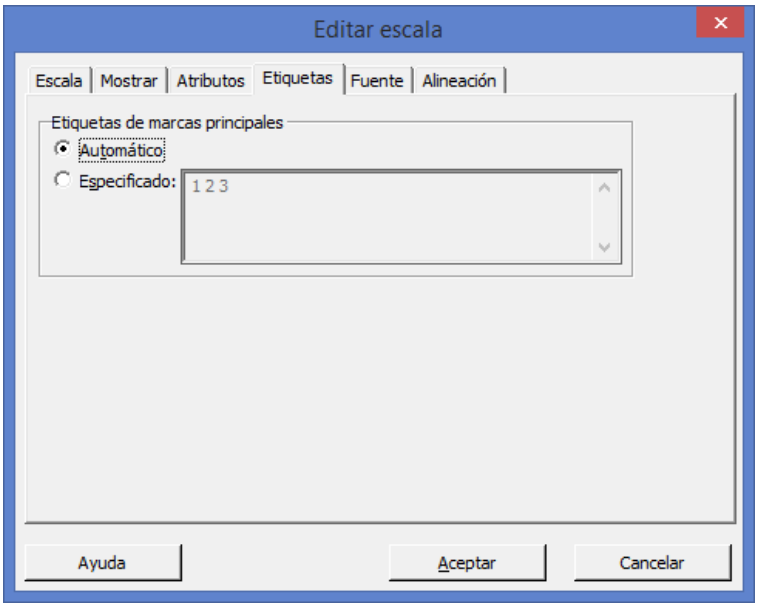
Asímismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución y en la falta de homocedasticidad.

Etiquetado de los niveles de la variable o "Factor" UBICACIÓN:

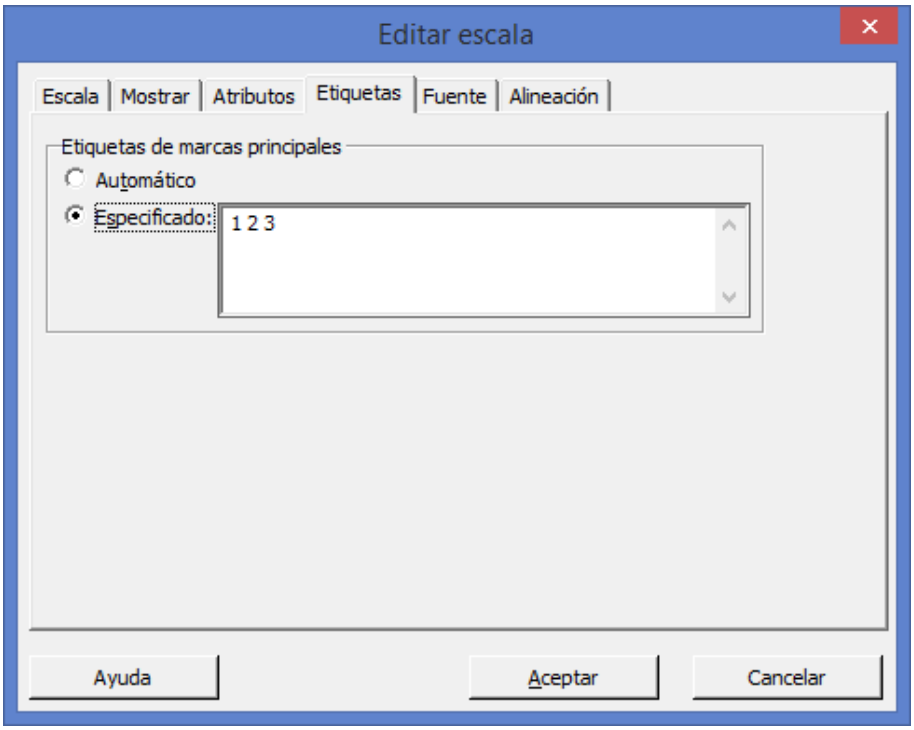
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **UBICACIÓN** para que Minitab despliegue la ventana ***Editar Escala.***



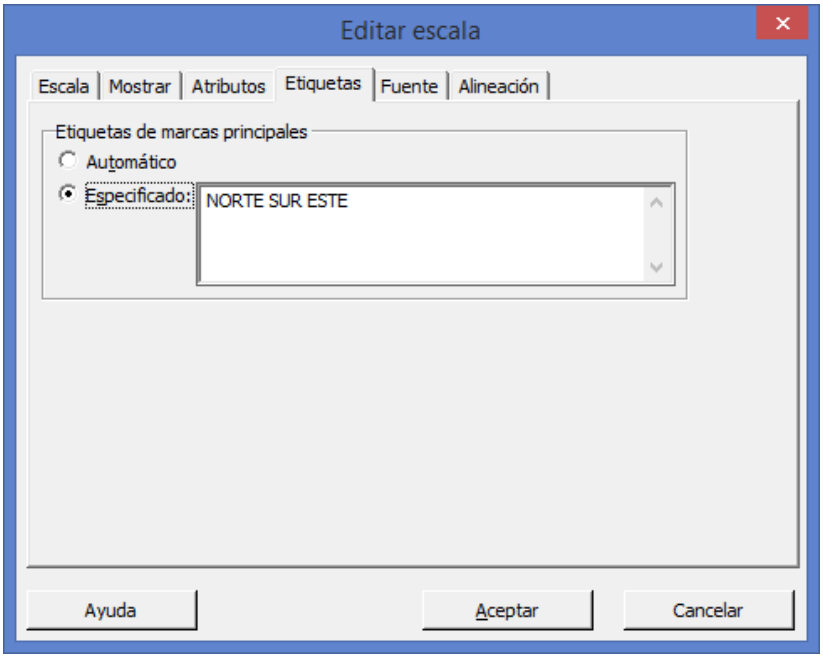
2 Seleccionar la pestaña **Etiquetas**



3 Activar la casilla **Especificado**

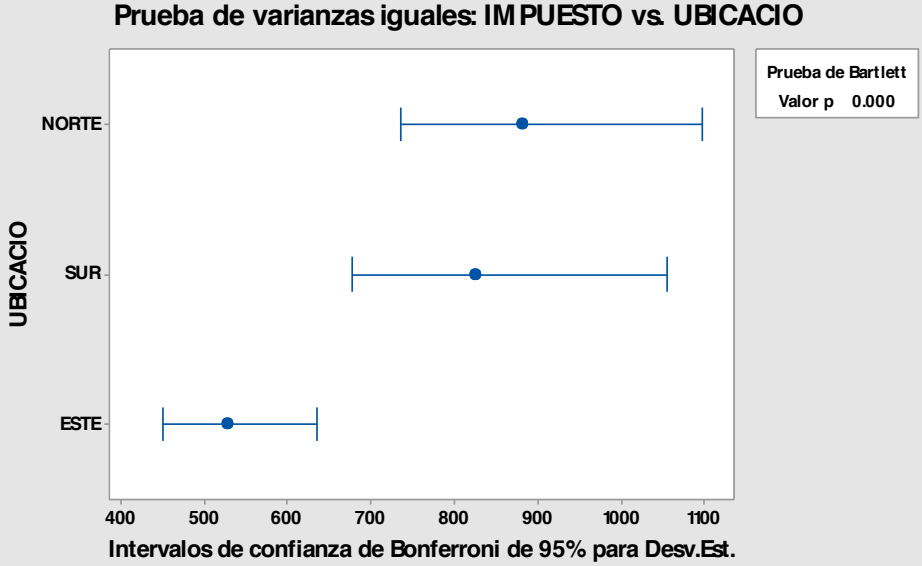


4 Borrar los números 1 2 3 y escribir **NORTE SUR ESTE**



5 Haga clic en **Aceptar**.

6 Aparece la gráfica debidamente etiquetada.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

1.3.2.3.3 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 1.3.2.3.3

PRUEBA DE HOMOCEDASTICIDAD. PRUEBA DE BARLETT

1.3.2.3.3 Usted diseña un experimento para evaluar el valor de 233 casas de diferentes estilos. Usted mide el **IMPUESTO** de las casas de la base de datos **"BIENES1"** que se construyó en el ejemplo anterior en cada una de las casas con respecto al tipo de **ALBERCA** ubicada en la propiedad. Como usted desea probar la igualdad de las medias y evaluar las diferencias de las medias, utiliza el procedimiento ANOVA de un factor (datos apilados) con comparaciones múltiples. Por lo general, usted escogería un método de comparación múltiple apropiado para sus datos.

Realice la prueba de igualdad de varianzas suponiendo que los datos provienen de distribuciones normales (Prueba de **Barlett**) utilizando el paquete de cómputo Minitab**:**

Antes de realizar el **análisis de varianza,** usted verifica la suposición de que la varianza es igual utilizando la Prueba de igualdad de varianzas.

1 Abra la base de datos que creó **"BIENES1"**

2 Elija **Estadísticas > ANOVA > Prueba de varianzas iguales**.

3 En **Respuesta**, ingrese ***IMPUESTO***

4 En **Factores**, ingrese ***ALBERCA***

5 En el botón **Opciones** active la casilla que dice: ***Utilice la prueba basándose en la distribución normal.*** Haga clic en **Aceptar** y nuevamente **Aceptar**

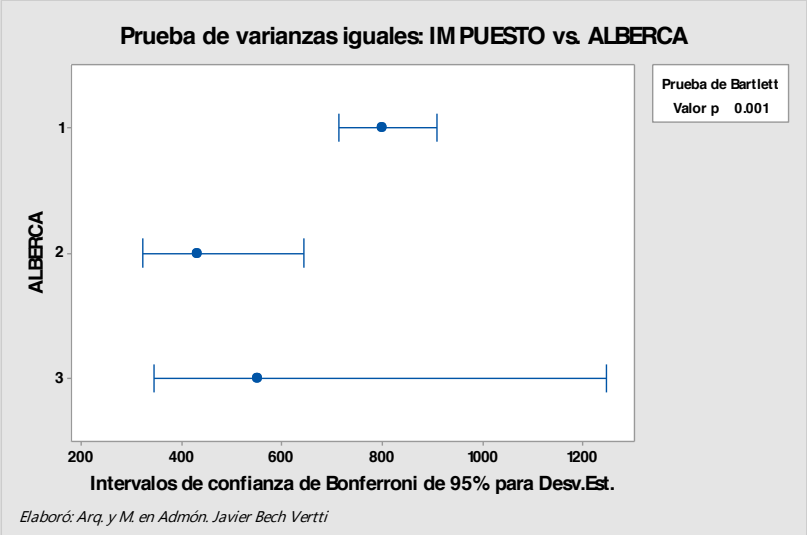
Salida de la ventana Sesión

Prueba de varianzas iguales: IMPUESTO vs. ALBERCA

```
Método

Hipótesis nula          Todas las varianzas son iguales

Hipótesis alterna       Por lo menos una varianza es diferente

Nivel de significancia  α = 0.05

Se utiliza el método de Bartlett. Este método es exacto sólo para datos

     normales.

Intervalos de confianza de Bonferroni de 95% para desviaciones estándar

ALBERCA    N  Desv.Est.          IC

      1  198    799.574  (713.132,  908.33)

      2   26    431.117  (321.288,  642.81)

      3    9    550.403  (343.117, 1246.43)


Nivel de confianza individual = 98.3333%

Pruebas

          Estadística

Método      de prueba  Valor p

Bartlett        13.43    0.001
```

Prueba de varianzas iguales: IMPUESTO vs. ALBERCA

Salida de la ventana Gráfica



Interpretación de los resultados

La **Prueba de igualdad de varianzas** genera una gráfica que muestra **Intervalos de confianza de Bonferroni de 95%** para la desviación estándar de la respuesta en cada nivel. Los resultados de la prueba de **Barlett** se muestran tanto en la ventana Sesión, como en la gráfica. **(Interprete la prueba de Bartlett cuando los datos provengan de distribuciones normales; utilice la prueba de Levene cuando los datos provengan de distribuciones continuas, pero no necesariamente normales)**. Note que el nivel de confianza de 95% se aplica a la **familia de intervalos** y la asimetría de los intervalos se debe al sesgo de la distribución chi-cuadrada.

Para el ejemplo para evaluar el **IMPUESTO** de las casas con base en el tipo de **ALBERCA** de las mismas, los **valores de p** de 0.001 es menor que la opción razonable de **alfa**, así que se puede rechazar la **hipótesis nula** de que las varianzas no son iguales ó de que al menos una es diferente. Es decir, estos datos no proveen evidencia suficiente para afirmar que las poblaciones tienen varianzas iguales.

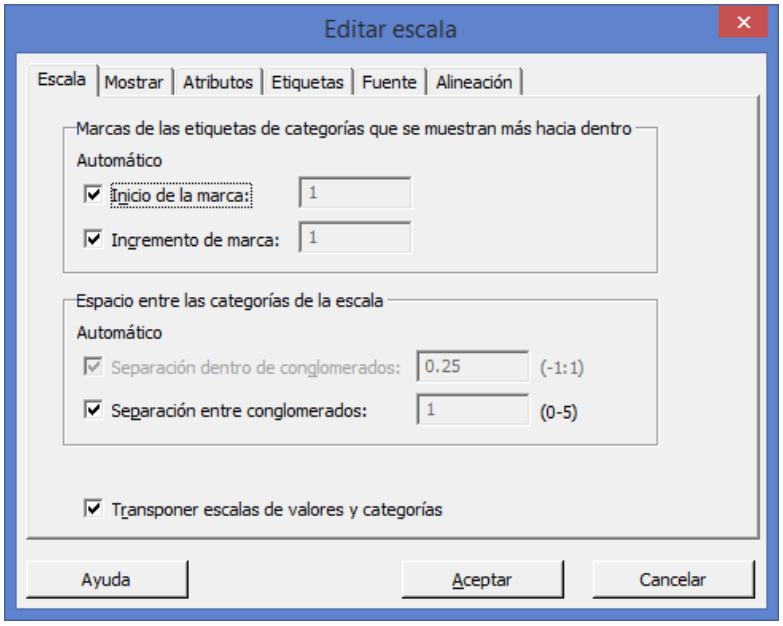
Los problemas de heterocedasticidad pueden solucionarse por medio de transformaciones de datos, similares a las usadas para conseguir normalidad y como se mencionó antes, en muchas ocasiones la heterocedasticidad es reflejo de la no normalidad de una de las variables y la correlación de la no normalidad resuelve igualmente la dispersión de la varianza.

Para las distribuciones no normales, las dos formas más comunes son las distribuciones "planas" y las distribuciones asimétricas. Para la distribución plana, las transformaciones más comunes son la inversa (es decir, 1/Y o 1/X). Las distribuciones asimétricas, pueden ser transformadas empleando la raíz cuadrada, logaritmos o incluso la inversa de la variable. Las distribuciones asimétricas negativas se transforman de forma más efectiva usando la raíz cuadrada, mientras que por lo regular, el logaritmo funciona mejor para la asimetría positiva.

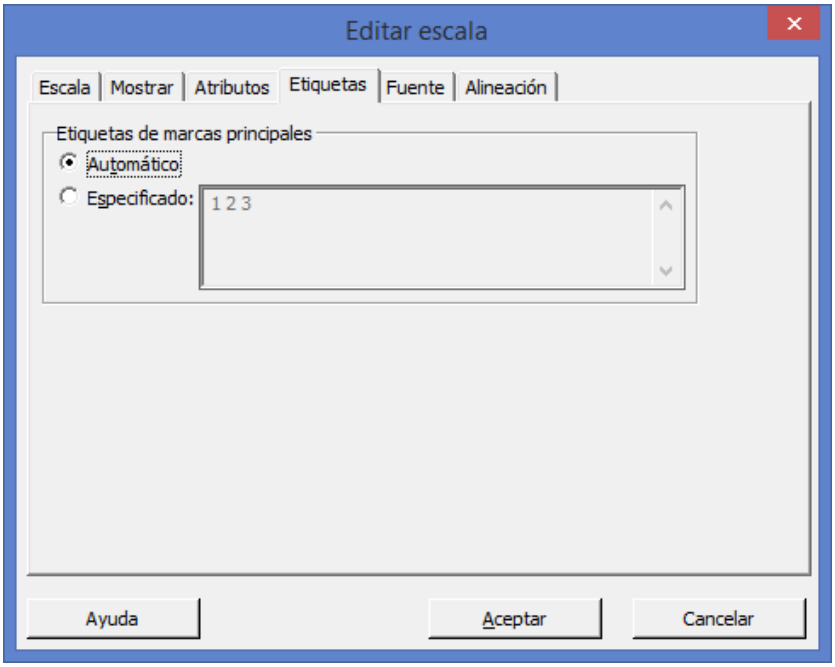
Asímismo se debe verificar la presencia de valores o casos atípicos que pudieran influir en la Normalidad de la distribución y en la falta de homocedasticidad.

Etiquetado de los niveles de la variable o "Factor" ALBERCA:

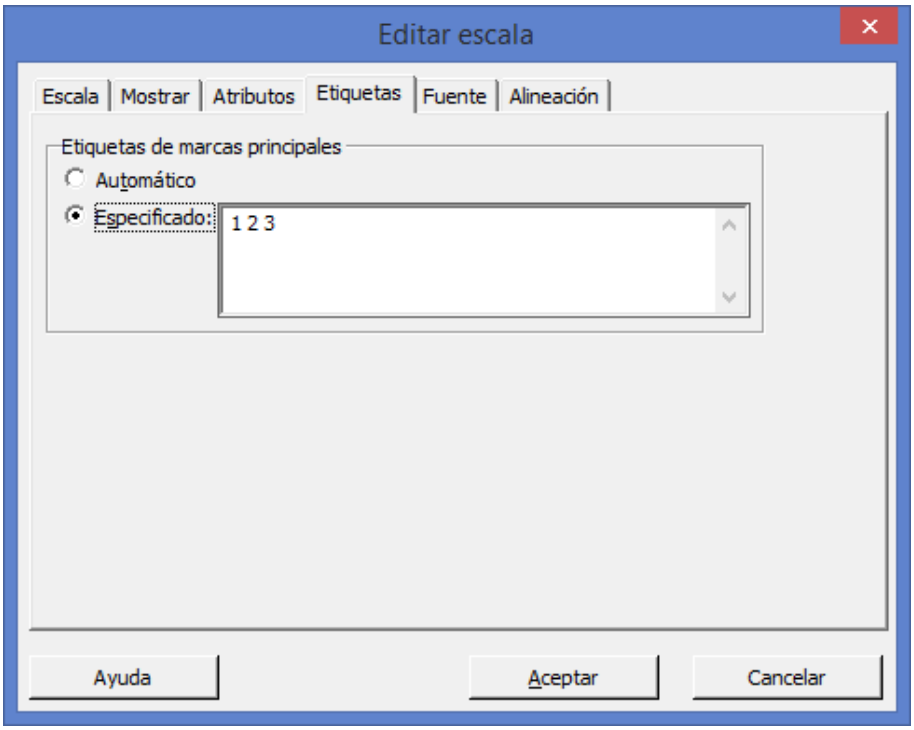
1 Haga doble clic en el número 1 del eje vertical de la gráfica que dice **ALBERCA** para que Minitab despliegue la ventana ***Editar Escala.***



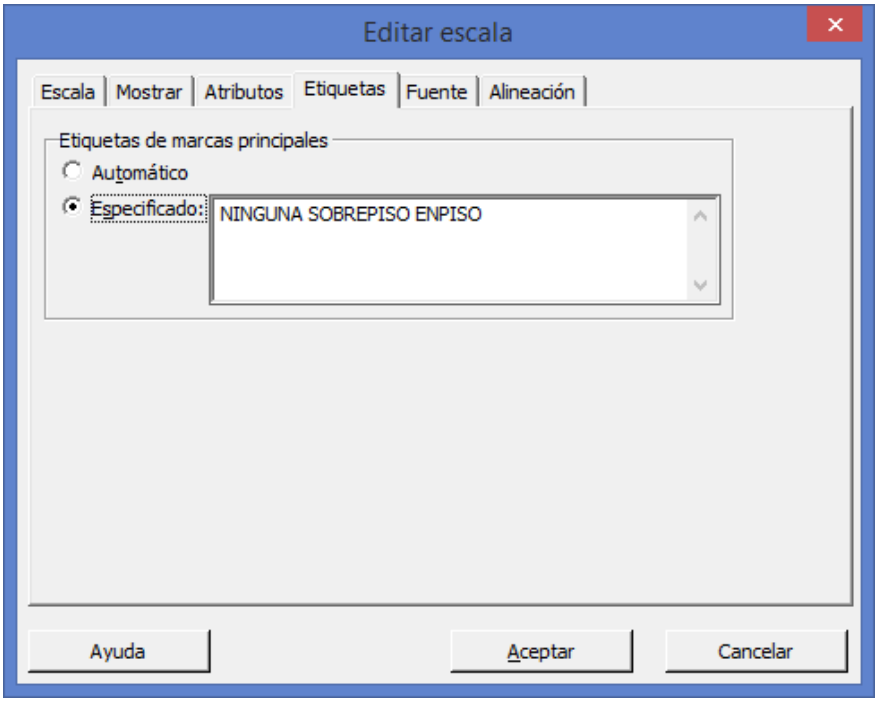
2 Seleccionar la pestaña **Etiquetas**



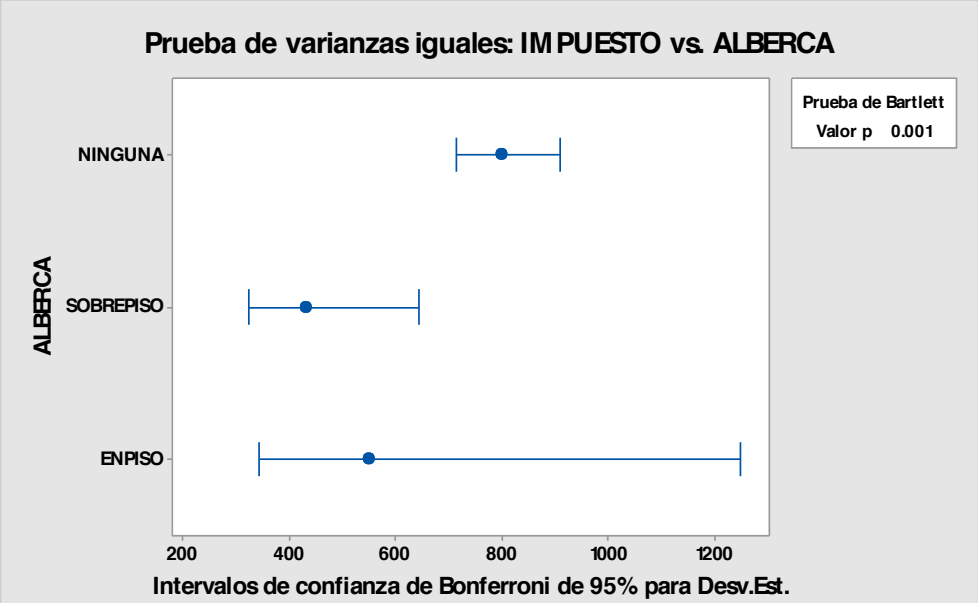
3 Activar la casilla **Especificado**



4 Borrar los números 1 2 3 y escribir **NINGUNA SOBREPISO ENPISO**



5 Haga clic en **Aceptar**.
6 Aparece la gráfica debidamente etiquetada.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

1.4.1.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.4.1.1.1**

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**

1.4.1.1.1 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$3x - 5y + 7z = 28$$

$$2x + 6y - 9z = -23$$

$$4x - 2y - 5z = 9$$

Paso 1. Expresar el sistema como A x X = K

$$\underbrace{\begin{bmatrix} 3 & -5 & +7 \\ +2 & +6 & -9 \\ 4 & -2 & -5 \end{bmatrix}}_{A} \quad x \quad \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{X} = \quad \underbrace{\begin{bmatrix} 28 \\ -23 \\ 9 \end{bmatrix}}_{K}$$

Paso 2. Buscar el valor del determinante $|A|$=

Aplicando la regla de Sarrus:

$$|A| = \begin{vmatrix} 3 & -5 & +7 \\ +2 & +6 & -9 \\ 4 & -2 & -5 \\ 3 & -5 & +7 \\ 2 & +6 & -9 \end{vmatrix} =$$ (3)(6)(-5)+(2)(-2)(7)+(4)(-5)(-9)-(4)(6)(7)-(3)(-2)(-9)-(2)(-5)(-5)=-210≠0 (Tiene Inversa)

Paso 3. Obtener la matriz transpuesta de la matriz A, que es:

$$A^T = A' = \begin{bmatrix} 3 & 2 & 4 \\ -5 & 6 & -2 \\ 7 & -9 & -5 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A´, es decir formar α A´

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 6 & -2 \\ -9 & -5 \end{vmatrix} & -\begin{vmatrix} -5 & -2 \\ 7 & -5 \end{vmatrix} & +\begin{vmatrix} -5 & 6 \\ 7 & -9 \end{vmatrix} \\ -\begin{vmatrix} 2 & 4 \\ -9 & -5 \end{vmatrix} & +\begin{vmatrix} 3 & 4 \\ 7 & -5 \end{vmatrix} & -\begin{vmatrix} 3 & 2 \\ 7 & -9 \end{vmatrix} \\ +\begin{vmatrix} 2 & 4 \\ 6 & -2 \end{vmatrix} & -\begin{vmatrix} 3 & 4 \\ -5 & -2 \end{vmatrix} & +\begin{vmatrix} 3 & 2 \\ -5 & 6 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(-30-18) & -(25+14) & +(45-42) \\ -(-10+36) & +(-15-28) & -(-27-14) \\ +((-4-24) & -(-6+20) & +(18+10) \end{bmatrix} = \begin{bmatrix} -48 & -39 & +3 \\ -26 & -43 & +41 \\ -28 & -14 & +28 \end{bmatrix}$$

Paso 5. Calcular la inversa de la matriz $A' = \frac{1}{|A|}$ α A'

$$\frac{1}{-210}\begin{bmatrix}-48 & -39 & +3\\ -26 & -43 & +41\\ -28 & -14 & +28\end{bmatrix} = \begin{bmatrix}48/210 & +39/210 & -3/210\\ +26/210 & +43/210 & -41/210\\ +28/210 & +14/210 & -28/210\end{bmatrix}$$

Paso 6. Llevar este valor a la expresión $X=A^{-1}$ x K y tenemos:

$$\begin{bmatrix}x\\ y\\ z\end{bmatrix} = \begin{bmatrix}+48/210 & +39/210 & -3/210\\ +26/210 & +43/210 & -41/210\\ +28/210 & +14/210 & -28/210\end{bmatrix} x \begin{bmatrix}28\\ -23\\ 9\end{bmatrix} =$$

$$\begin{bmatrix}\left(-\frac{48}{210} + \frac{39}{210} + \frac{-3}{210}\right) * \begin{bmatrix}28\\ -23\\ 9\end{bmatrix}\\ \left(+\frac{26}{210} + \frac{43}{210} - \frac{41}{210}\right) * \begin{bmatrix}28\\ -23\\ 9\end{bmatrix}\\ \left(+\frac{28}{210} + \frac{14}{210} - \frac{28}{210}\right) * \begin{bmatrix}28\\ -23\\ 9\end{bmatrix}\end{bmatrix}$$

$$= \begin{bmatrix}-1344/210 & -897/210 & -27/210\\ 728/210 & -989/210 & -369/210\\ 784/210 & -322/210 & -252/210\end{bmatrix}$$

$$= \begin{bmatrix}420/210\\ -630/210\\ 210/210\end{bmatrix} = \begin{bmatrix}2\\ -3\\ 1\end{bmatrix} = \begin{bmatrix}x\\ y\\ z\end{bmatrix}$$

Paso 7. La solución es x=2, y=-3, z=1

1.4.1.1.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.4.1.1.2**

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**

1.4.1.1.2 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$2x + 3y + z = 4$$

$$x + 2y + 2z = 6$$

$$5x + y + 4x = 21$$

Paso 1. Expresar el sistema como A x X = K

$$\underbrace{\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 2 \\ 5 & 1 & 4 \end{bmatrix}}_{A} \quad x \quad \underbrace{\begin{bmatrix} x \\ y \\ z \end{bmatrix}}_{X} = \quad \underbrace{\begin{bmatrix} 4 \\ 6 \\ 21 \end{bmatrix}}_{K}$$

Paso 2. Buscar el valor del determinante $|A|$=

Aplicando la regla de Sarrus:

$$|A| = \begin{vmatrix} 2 & 3 & 1 \\ 1 & 2 & 2 \\ 5 & 1 & 4 \\ 2 & 3 & 1 \\ 1 & 2 & 2 \end{vmatrix} = (2)(2)(4)+(1)(1)(1)+(5)(3)(2)-(5)(2)(1)-(2)(1)(2)-(1)(3)(4)=21\neq 0 \text{ (Tiene Inversa)}$$

Paso 3. Obtener la matriz transpuesta de la matriz A, que es:

$$A^T=A'=\begin{bmatrix}2 & 1 & 5\\ 3 & 2 & 1\\ 1 & 2 & 4\end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A′, es decir formar α A′

$$\alpha A'=\begin{bmatrix}+\begin{vmatrix}2 & 1\\ 2 & 4\end{vmatrix} & -\begin{vmatrix}3 & 1\\ 1 & 4\end{vmatrix} & +\begin{vmatrix}3 & 2\\ 1 & 2\end{vmatrix}\\ -\begin{vmatrix}1 & 5\\ 2 & 4\end{vmatrix} & +\begin{vmatrix}2 & 5\\ 1 & 4\end{vmatrix} & -\begin{vmatrix}2 & 1\\ 1 & 2\end{vmatrix}\\ +\begin{vmatrix}1 & 5\\ 2 & 1\end{vmatrix} & -\begin{vmatrix}2 & 5\\ 3 & 1\end{vmatrix} & +\begin{vmatrix}2 & 1\\ 3 & 2\end{vmatrix}\end{bmatrix}$$

$$=\begin{bmatrix}+(8-2) & -(12-1) & +(6-2)\\ -(4-10) & +(8-5) & -(4-1)\\ +((1-10) & -(2-15) & +(4-3)\end{bmatrix}=\begin{bmatrix}+6 & -11 & +4\\ +6 & +3 & -3\\ -9 & +13 & +1\end{bmatrix}$$

Paso 5. Calcular la inversa de la matriz A′= $\frac{1}{|A|}$ α A′

$$\frac{1}{21}\begin{bmatrix}+6 & -11 & +4\\ +6 & +3 & -3\\ -9 & +13 & +1\end{bmatrix}=\begin{bmatrix}+6/21 & -11/21 & +4/21\\ +6/21 & +3/21 & -3/21\\ -9/21 & +13/21 & +1/21\end{bmatrix}$$

Paso 6. Llevar este valor a la expresión $X=A^{-1}$ x K y tenemos:

$$\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\left[\begin{bmatrix}+6/21 & -11/21 & +4/21\\ +6/21 & +3/21 & -3/21\\ -9/21 & +13/21 & +1/21\end{bmatrix}\right]x\begin{bmatrix}4\\ 6\\ 21\end{bmatrix}=$$

$$\begin{bmatrix} \left(+\frac{6}{21} \quad -\frac{11}{21} \quad +\frac{4}{21}\right) & * & \begin{bmatrix} 4 \\ 6 \\ 21 \end{bmatrix} \\ \left(+\frac{6}{21} \quad +\frac{3}{21} \quad -\frac{3}{21}\right) & * & \begin{bmatrix} 4 \\ 6 \\ 21 \end{bmatrix} \\ \left(-\frac{9}{21} \quad +\frac{13}{21} \quad +\frac{1}{21}\right) & * & \begin{bmatrix} 4 \\ 6 \\ 21 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} +24/21 & -66/21 & +84/21 \\ +24/21 & +18/21 & -63/21 \\ -36/21 & +78/21 & +21/21 \end{bmatrix}$$

$$= \begin{bmatrix} 42/21 \\ -21/21 \\ 63/21 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Paso 7. La solución es x=2, y=-1, z=3

1.4.1.1.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.4.1.1.3**

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA**

1.4.1.1.3 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$2x + 5y - 3z = 17$$

$$6x - 2y \quad -5z = -3$$

$$3x + 7y + 4z = -18$$

Paso 1. Expresar el sistema como A x X = K

$$\underbrace{\begin{bmatrix}2 & 5 & -3\\ 6 & -2 & -5\\ 3 & 7 & 4\end{bmatrix}}_{A} \text{ x } \underbrace{\begin{bmatrix}x\\ y\\ z\end{bmatrix}}_{X} = \underbrace{\begin{bmatrix}17\\ -3\\ -18\end{bmatrix}}_{K}$$

Paso 2. Buscar el valor del determinante $|A|$=

Aplicando la regla de Sarrus:

$$|A| = \begin{vmatrix}2 & 5 & -3\\ 6 & -2 & -5\\ 3 & 7 & 4\\ 2 & 5 & -3\\ 6 & -2 & -5\end{vmatrix} =$$ (2)(-2)(4)+(6)(7)(-3)+(3)(5)(-5)-(3)(-2)(-3)-(2)(7)(-5)-(6)(5)(4)=-285≠0 (Tiene Inversa)

Paso 3. Obtener la matriz transpuesta de la matriz A, que es:

$$A^T=A'=\begin{bmatrix}2 & 6 & 3\\ 5 & -2 & 7\\ -3 & -5 & 4\end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A′, es decir formar α A′

$$\alpha A'=\begin{bmatrix}+\begin{vmatrix}-2 & 7\\ -5 & 4\end{vmatrix} & -\begin{vmatrix}5 & 7\\ -3 & 4\end{vmatrix} & +\begin{vmatrix}5 & 2\\ -3 & -5\end{vmatrix}\\ -\begin{vmatrix}6 & 3\\ -5 & 4\end{vmatrix} & +\begin{vmatrix}2 & 3\\ -3 & 4\end{vmatrix} & -\begin{vmatrix}2 & 6\\ -3 & -5\end{vmatrix}\\ +\begin{vmatrix}6 & 3\\ -2 & 7\end{vmatrix} & -\begin{vmatrix}2 & 3\\ 5 & 7\end{vmatrix} & +\begin{vmatrix}2 & 6\\ 5 & -2\end{vmatrix}\end{bmatrix}$$

$$=\begin{bmatrix}+(8+35) & -(20+21) & +(-25-6)\\ -(24+15) & +(8+9) & -(-10+18)\\ +((42+6) & -(14-15) & +(-4-30)\end{bmatrix}=\begin{bmatrix}+27 & -41 & -31\\ -39 & +17 & -8\\ +48 & +1 & -34\end{bmatrix}$$

Paso 5. Calcular la inversa de la matriz A′= $\frac{1}{|A|}$ α A′

$$\frac{1}{-285}\begin{bmatrix}+27 & -41 & -31\\ -39 & +17 & -8\\ +48 & +1 & -34\end{bmatrix}=\begin{bmatrix}-27/285 & +41/285 & +31/285\\ +39/285 & -17/285 & +8/285\\ -48/285 & -1/285 & +34/285\end{bmatrix}$$

Paso 6. Llevar este valor a la expresión $X=A^{-1}$ x K y tenemos:

$$\begin{bmatrix}x\\ y\\ z\end{bmatrix}=\left[\begin{bmatrix}-27/285 & +41/285 & +31/285\\ +39/285 & -17/285 & +8/285\\ -48/285 & -1/285 & +34/285\end{bmatrix}\right]x\begin{bmatrix}17\\ -3\\ -18\end{bmatrix}=$$

$$\begin{bmatrix}\left(-\frac{27}{285}+\frac{41}{285}+\frac{31}{285}\right) * \begin{bmatrix}17\\ -3\\ -18\end{bmatrix}\\ \left(+\frac{39}{285}-\frac{17}{285}+\frac{8}{285}\right) * \begin{bmatrix}17\\ -3\\ -18\end{bmatrix}\\ \left(-\frac{48}{285}-\frac{1}{285}+\frac{34}{285}\right) * \begin{bmatrix}17\\ -3\\ -18\end{bmatrix}\end{bmatrix}=\begin{bmatrix}-459/285 & -123/285 & -558/285\\ +663/285 & +51/285 & -144/285\\ -816/285 & +3/285 & -612/285\end{bmatrix}$$

$$=\begin{bmatrix}-1140/285\\ 570/285\\ -1425/285\end{bmatrix}=\begin{bmatrix}-4\\ 2\\ -5\end{bmatrix}=\begin{bmatrix}x\\ y\\ z\end{bmatrix}$$

Paso 7. La solución es x=-4, y=2, z=-5

1.4.1.2.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.4.1.2.1**

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA EN EXCEL**

1.4.1.2.1 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$x + y + 2z = 31$$

$$x + z + 2y = 32$$

$$y + z + 2x = 33$$

Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

. Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).

. Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.

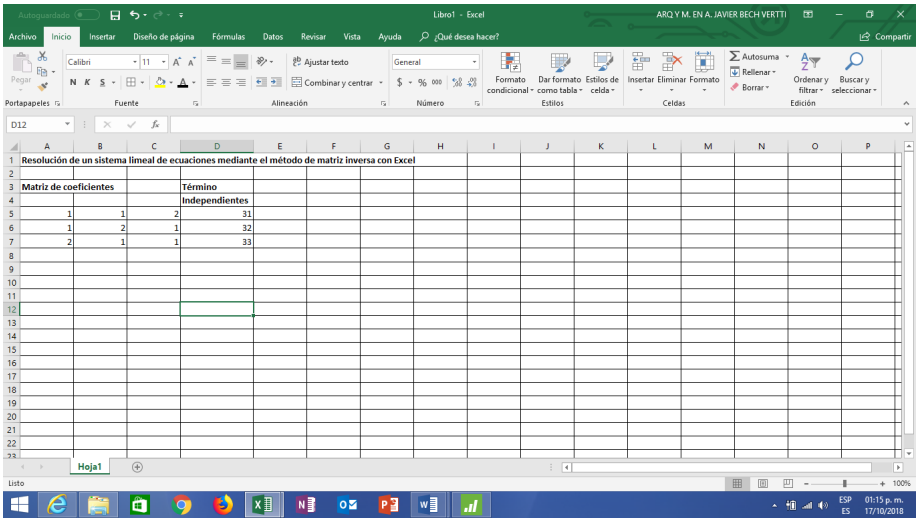
. Introduce los valores independientes en otro rango al lado.

Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

. Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).

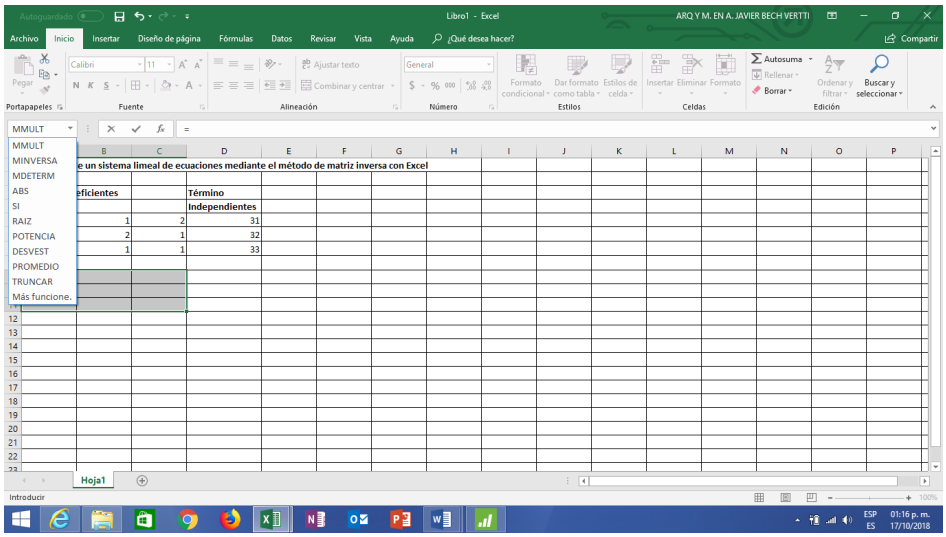
. Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.

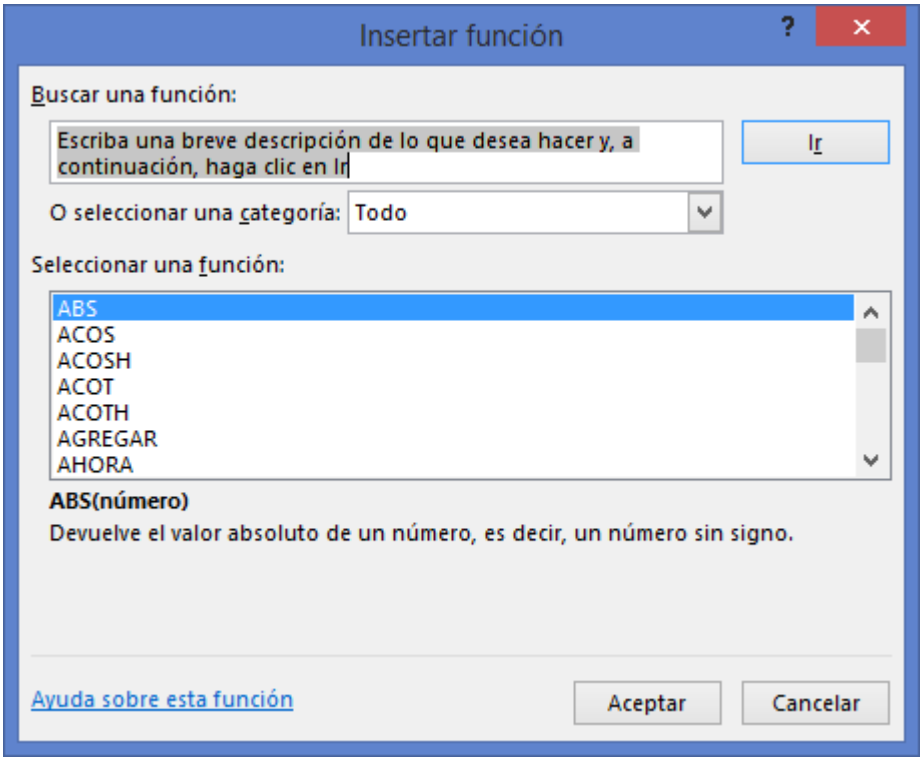
. Introduce los valores independientes en otro rango al lado.



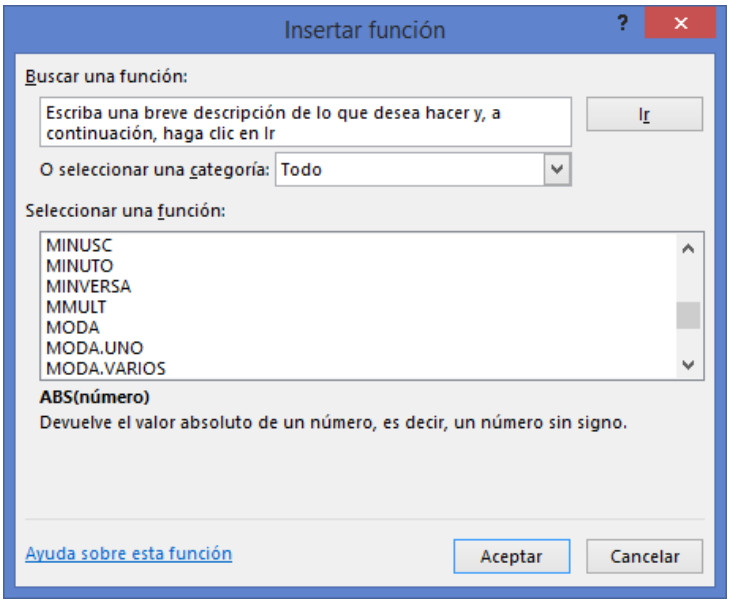
. Utiliza la función matricial **MINVERSA ()** para calcular la matriz inversa de los coeficientes introducidos. (Recuerda que para introducir la función debes pulsar Ctrl + Mayús + Intro).

Como la matriz de coeficientes es de 3 x 3 se necesitas 9 celdas para mostrar los resultados de la matris Inversa para lo cual selecciona con el mouse 9 celdas debajo de la matriz de coeficientes dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MINVERSA** () seleccionando la opción **Más funciones** en la parte inferior

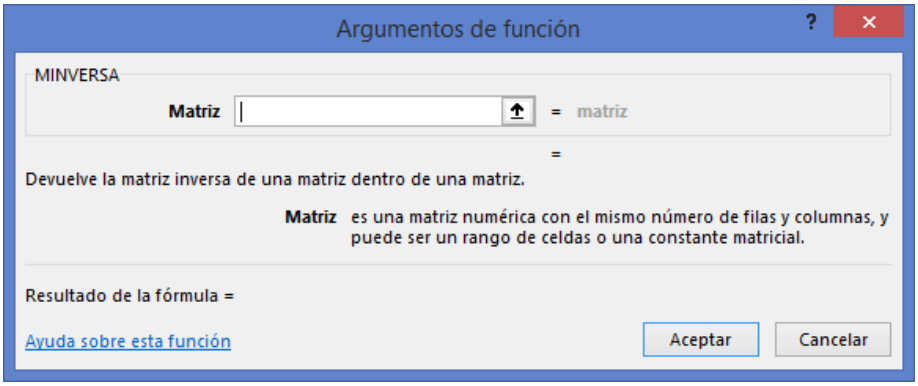




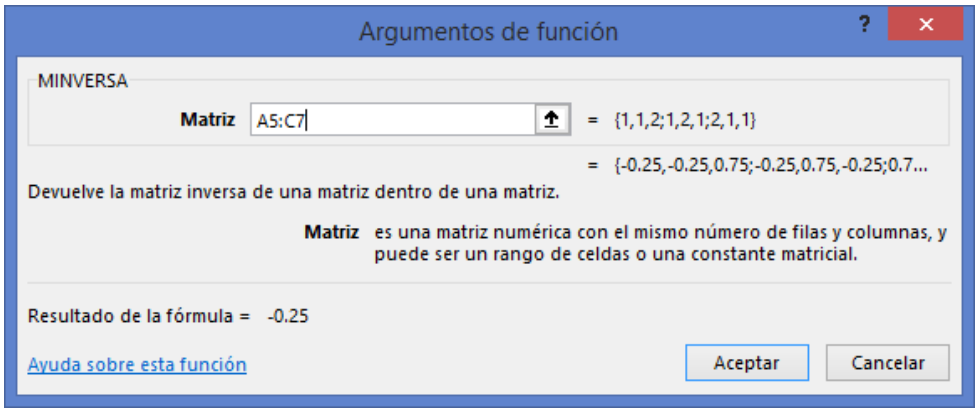
En Seleccionar una función busca **MINVERSA** y da un clic.



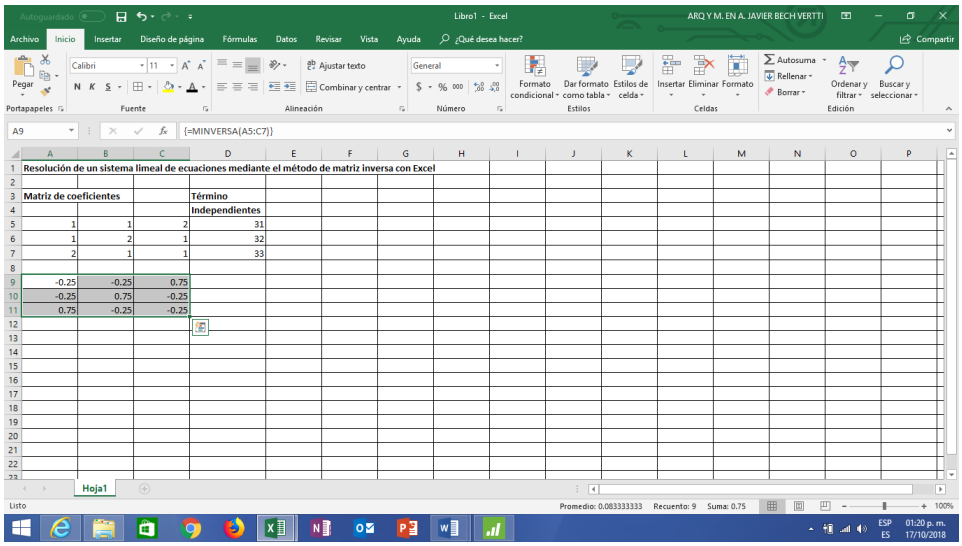
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



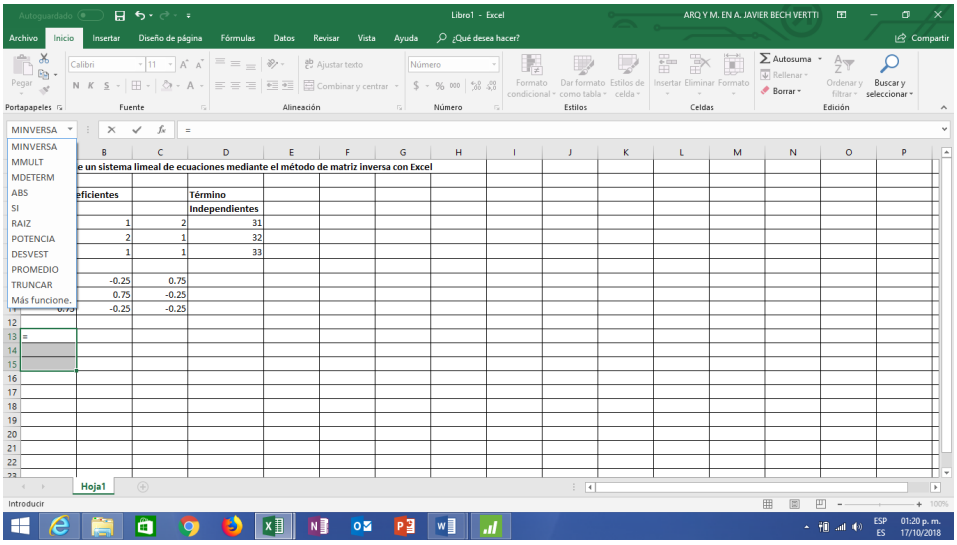
Selecciona las celdas de la matriz de coeficientes, en este caso A5:C7

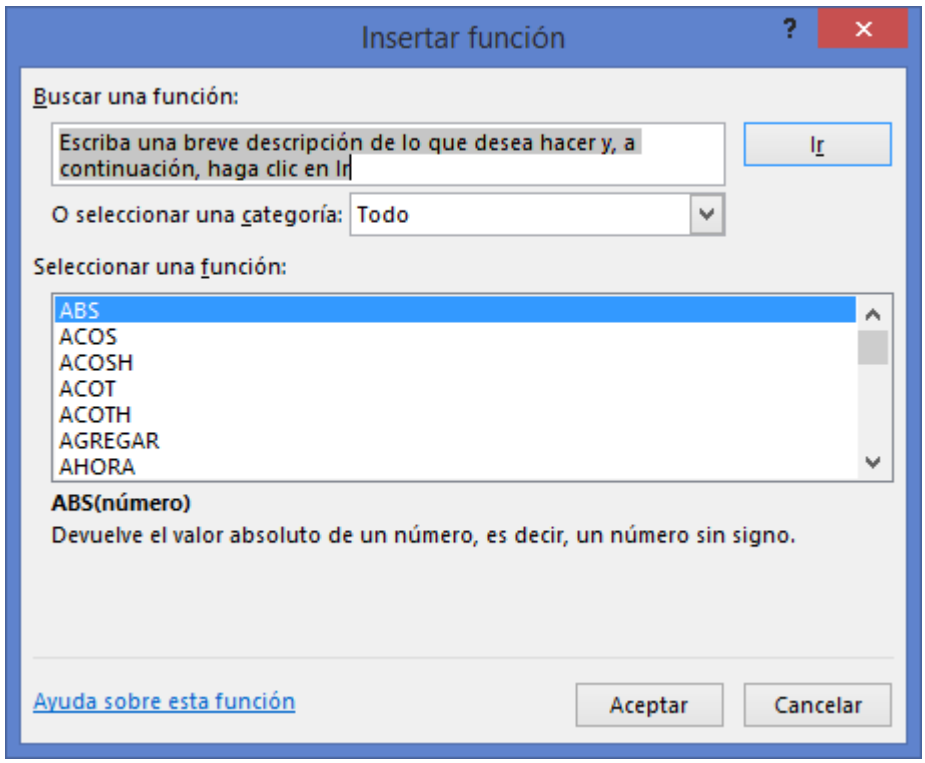


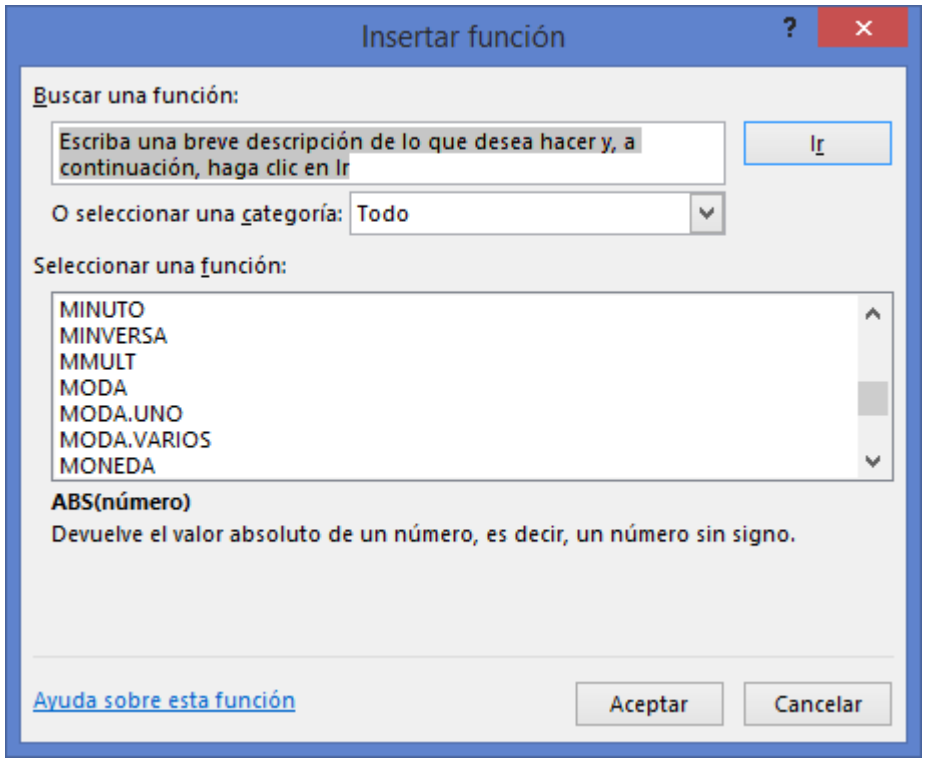
Ojo, no pulses Aceptar. La matriz inversa se determinará al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter)**.



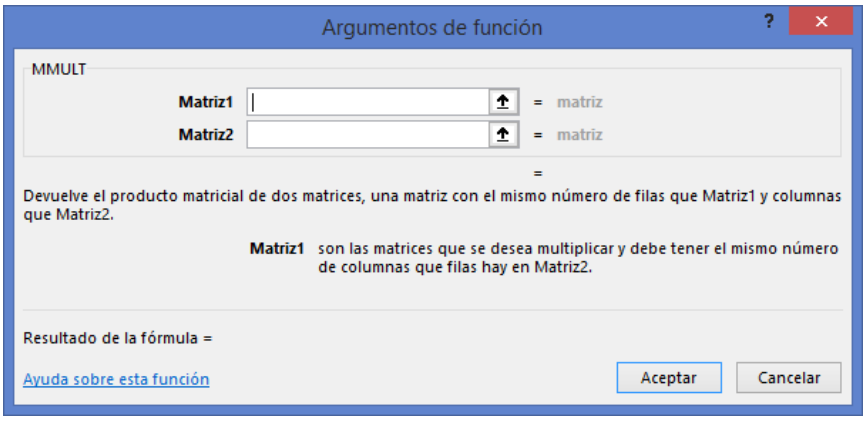
Como la matriz de los términos constantes es un vector de 1 x 3 se necesitas 3 celdas para mostrar los resultados de las Incógnitas **X, Y y Z** para lo cual selecciona con el mouse 3 celdas debajo de la **matriz inversa** dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MMULT** () seleccionando la opción **Más funciones** en la parte inferior



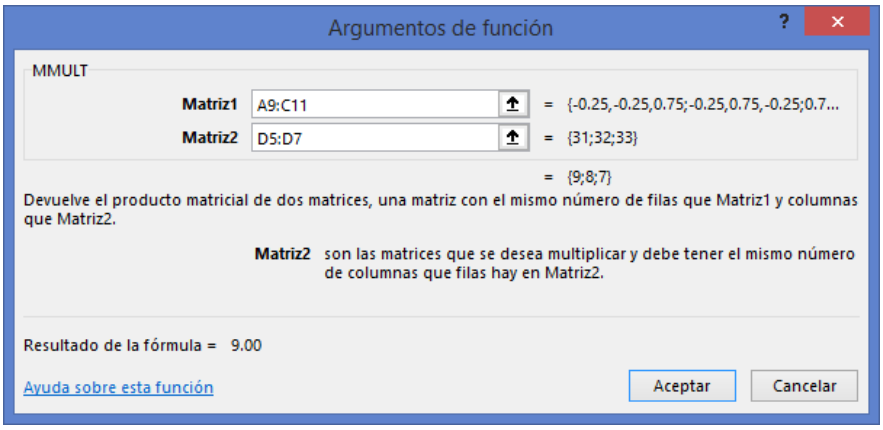




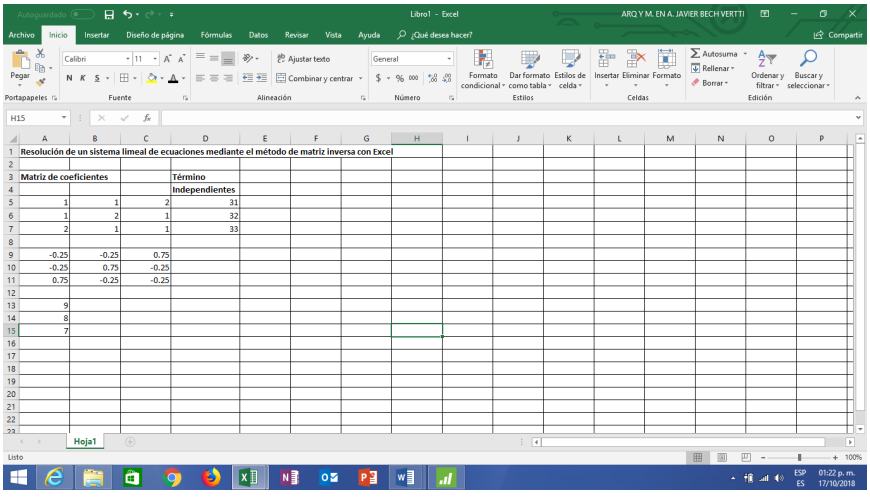
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



En el campo que dice Matriz 1 selecciona las celdas de la matriz de inversa, en este caso A9:C11 y en el campo que dice Matriz 2 selecciona las celdas de los términos constantes, en este caso D5:D7



Ojo, no pulses Aceptar. Los resultados de las incógnitas **X, Y, Z** se determinarán al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter), en este caso 9, 8, 7.** Por lo tanto las incognitas quedan **X=9, Y= 8 y Z=** 7



Con lo anterior queda terminado el ejemplo ilustrativo de **resolución de un sistema lineal de ecuaciones mediante el método de matriz inversa** en el paquete **EXCEL**. **Aplica el procedimiento para salir de Windows**

1.4.1.2.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.4.1.2.2**

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA EN EXCEL**

1.4.1.2.2 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$9x - 2y + 12z = 20$$

$$6x + 4y - 3z = 3$$

$$3x + 2y - 3z = -1$$

Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

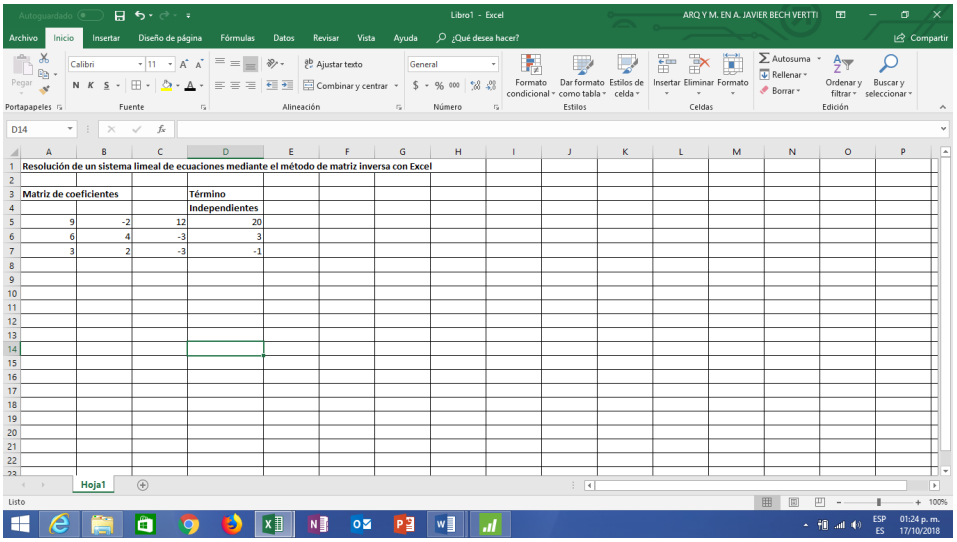
- Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).
- Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.
- Introduce los valores independientes en otro rango al lado.

Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

- Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).

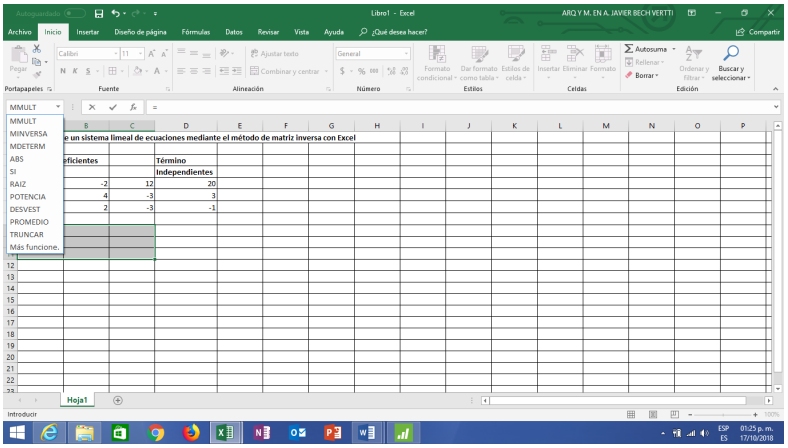
. Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.

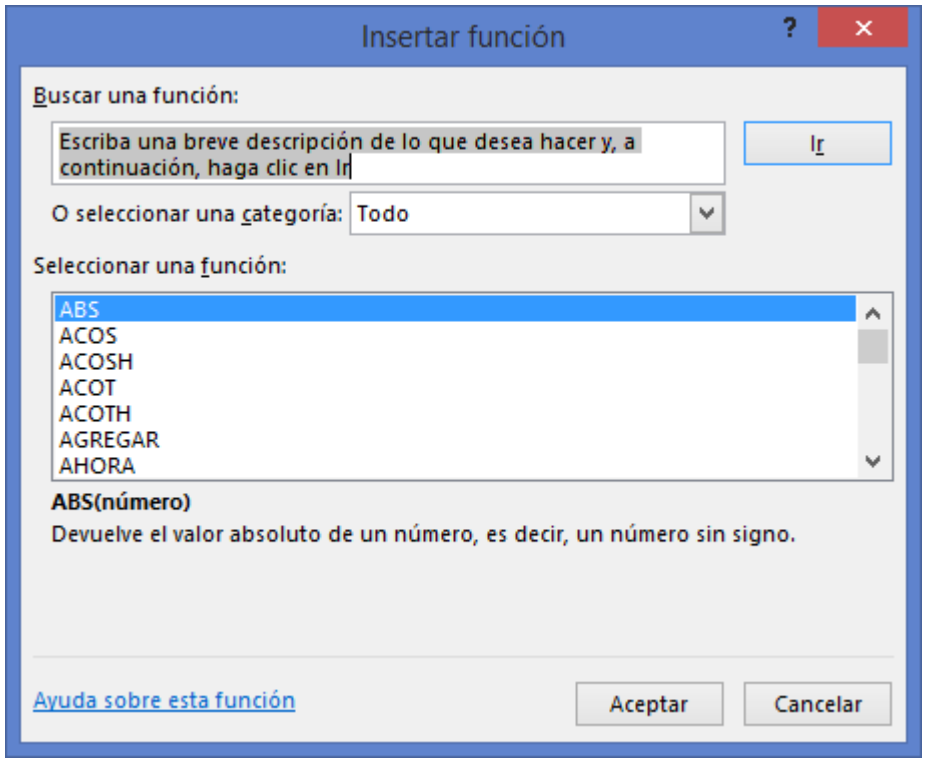
. Introduce los valores independientes en otro rango al lado.



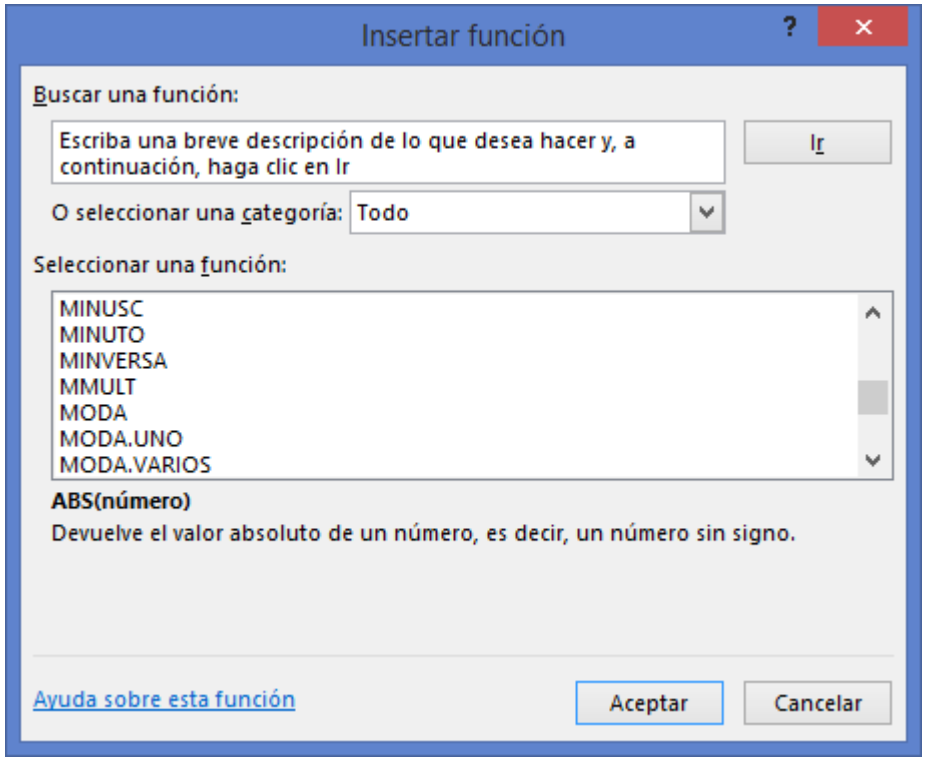
. Utiliza la función matricial **MINVERSA ()** para calcular la matriz inversa de los coeficientes introducidos. (Recuerda que para introducir la función debes pulsar Ctrl + Mayús + Intro).

Como la matriz de coeficientes es de 3 x 3 se necesitas 9 celdas para mostrar los resultados de la matris Inversa para lo cual selecciona con el mouse 9 celdas debajo de la matriz de coeficientes dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MINVERSA** () seleccionando la opción **Más funciones** en la parte inferior

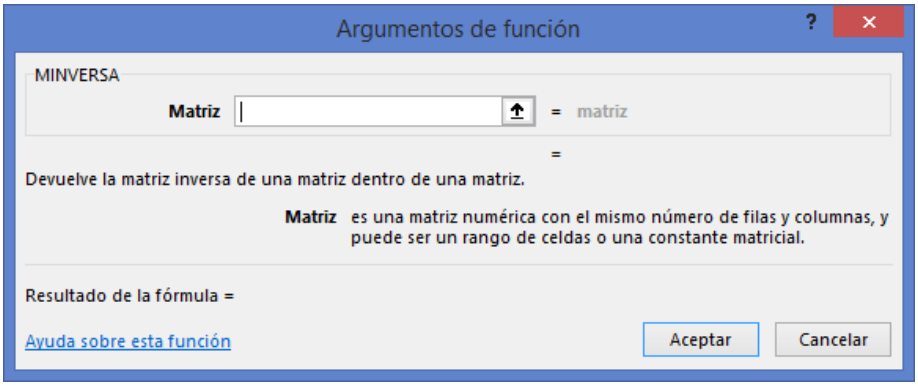




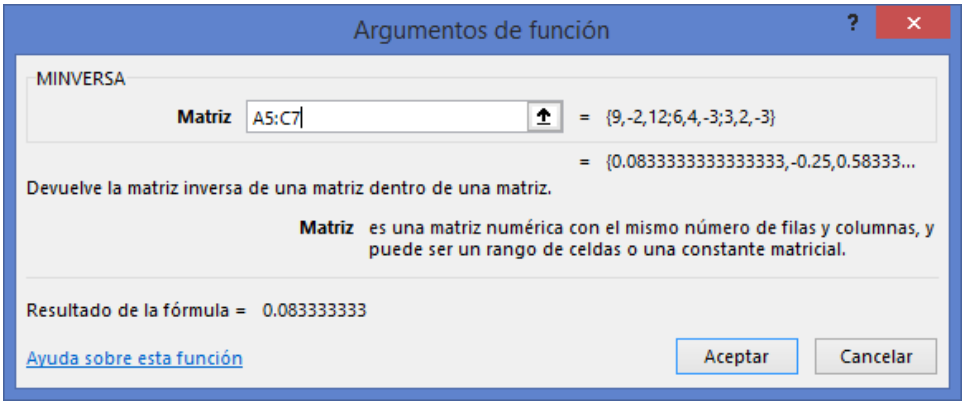
En Seleccionar una función busca **MINVERSA** y da un clic.



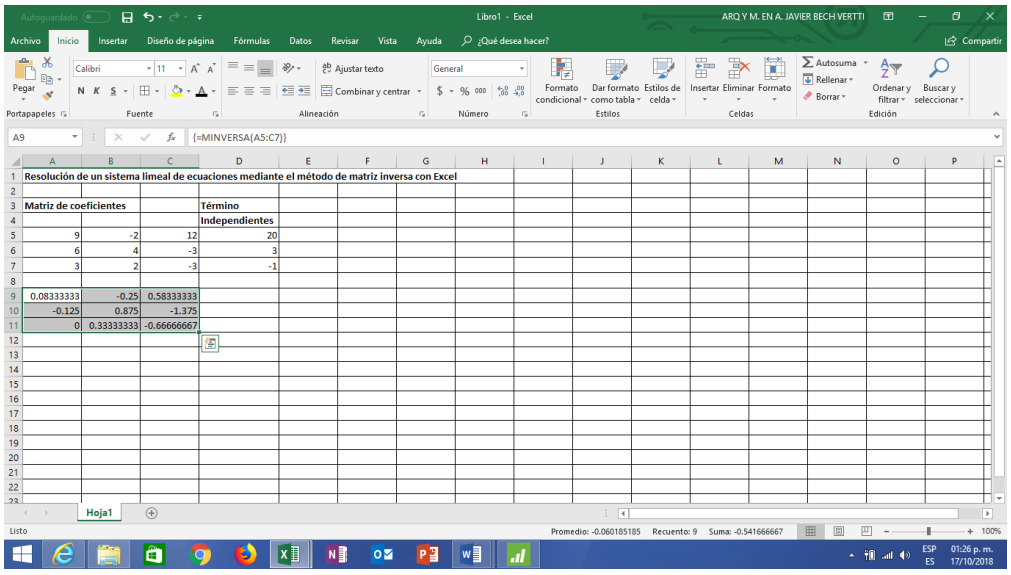
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



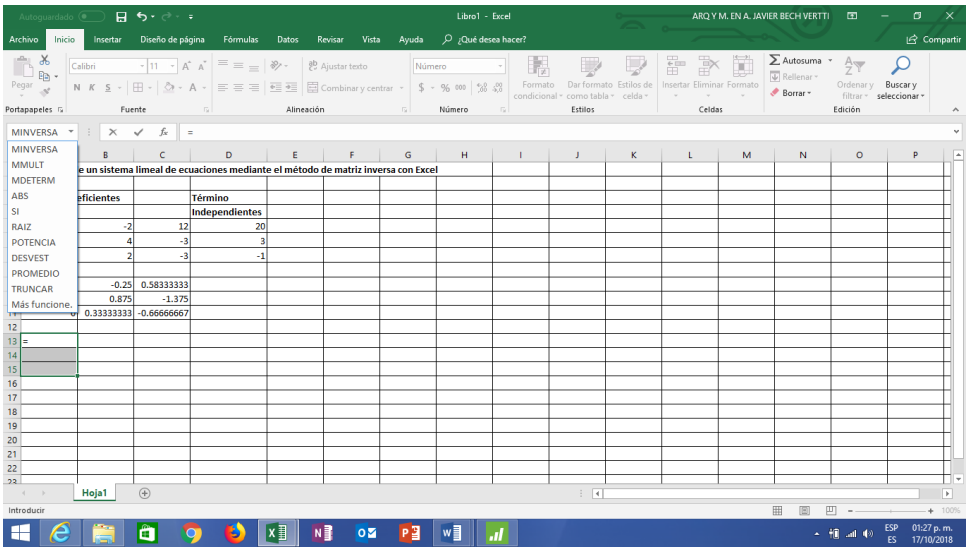
Selecciona las celdas de la matriz de coeficientes, en este caso A5:C7

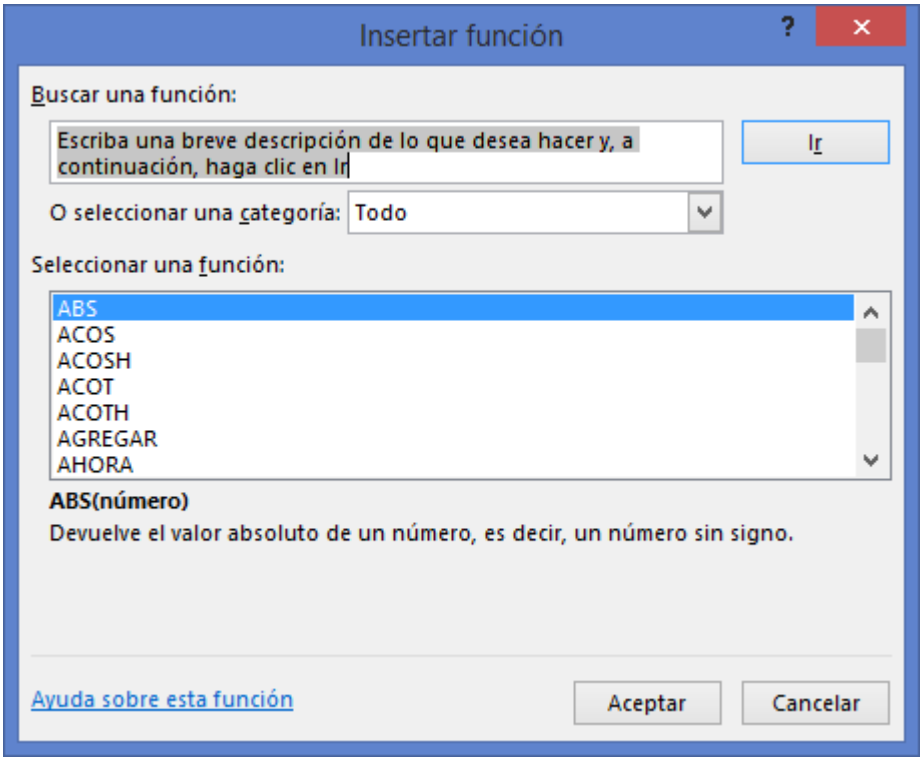


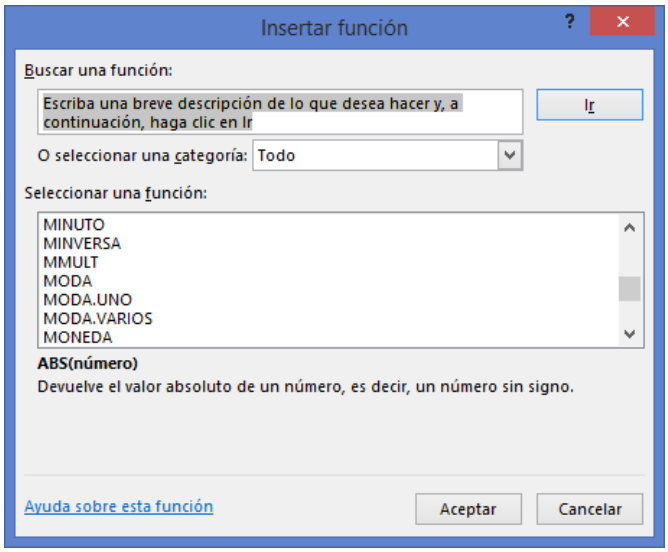
Ojo, no pulses Aceptar. La matriz inversa se determinará al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter)**.



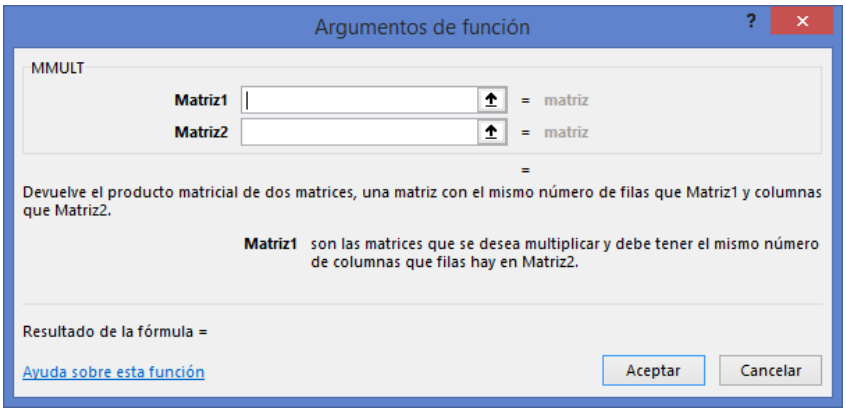
Como la matriz de los términos constantes es un vector de 1 x 3 se necesitas 3 celdas para mostrar los resultados de las Incógnitas **X, Y y Z** para lo cual selecciona con el mouse 3 celdas debajo de la **matriz inversa** dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MMULT** () seleccionando la opción **Más funciones** en la parte inferior



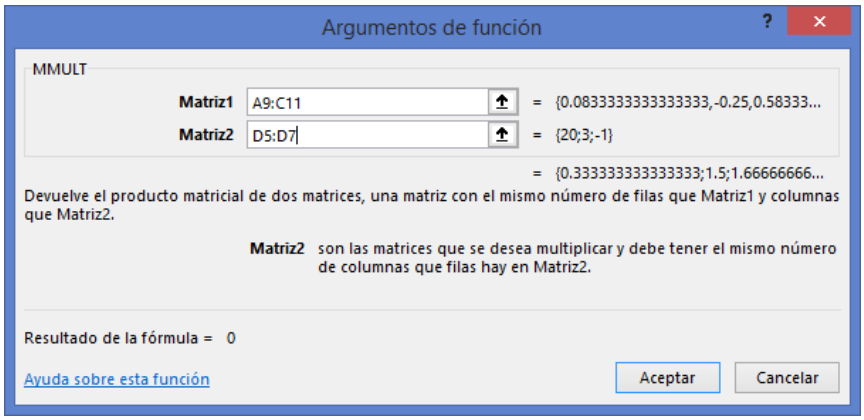




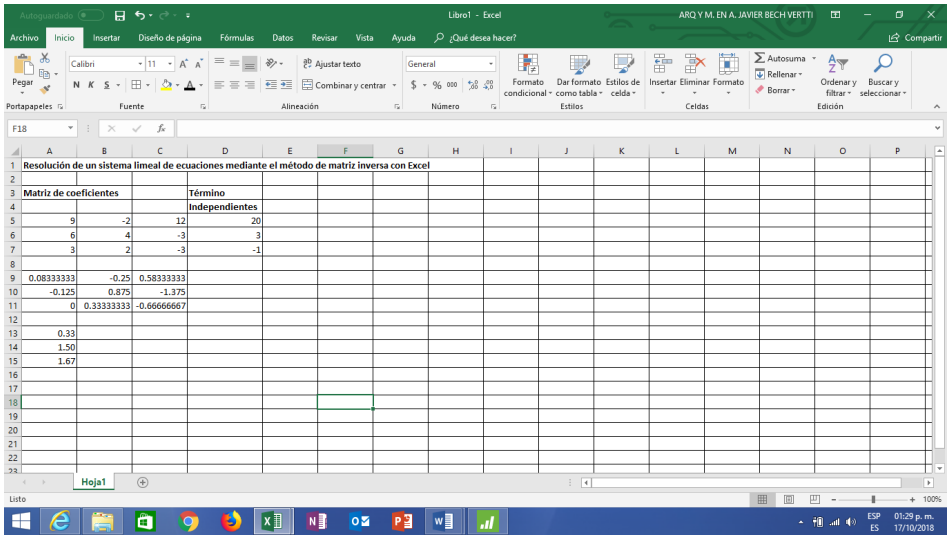
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



En el campo que dice Matriz 1 selecciona las celdas de la matriz de inversa, en este caso A9:C11 y en el campo que dice Matriz 2 selecciona las celdas de los términos constantes, en este caso D5:D7



Ojo, no pulses Aceptar. Los resultados de las incógnitas **X, Y, Z** se determinarán al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter), en este caso 0.33, 1.50, 1.67.** Por lo tanto las incognitas quedan **X=0.33, Y=1.50 y Z=1.67 aproximando a dos cifras.**



Con lo anterior queda terminado el ejemplo ilustrativo de **resolución de un sistema lineal de ecuaciones mediante el método de matriz inversa** en el paquete **EXCEL**. **Aplica el procedimiento para salir de Windows**

1.4.1.2.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 1.4.1.2.3**

RESOLUCIÓN DE UN SISTEMA LINEAL DE ECUACIONES MEDIANTE EL MÉTODO DE MATRIZ **INVERSA EN EXCEL**

1.4.4.2.3 Resolver el siguiente sistema, empleando el método de la **MATRIZ INVERSA** y aplicando la regla de Sarrus para obtener el determinante de **A** o sea $|A|$

$$2x - y + 3z = 14$$

$$2y - z + 3x = 24$$

$$2z - x + 3y = 10$$

Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

. Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).

. Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.

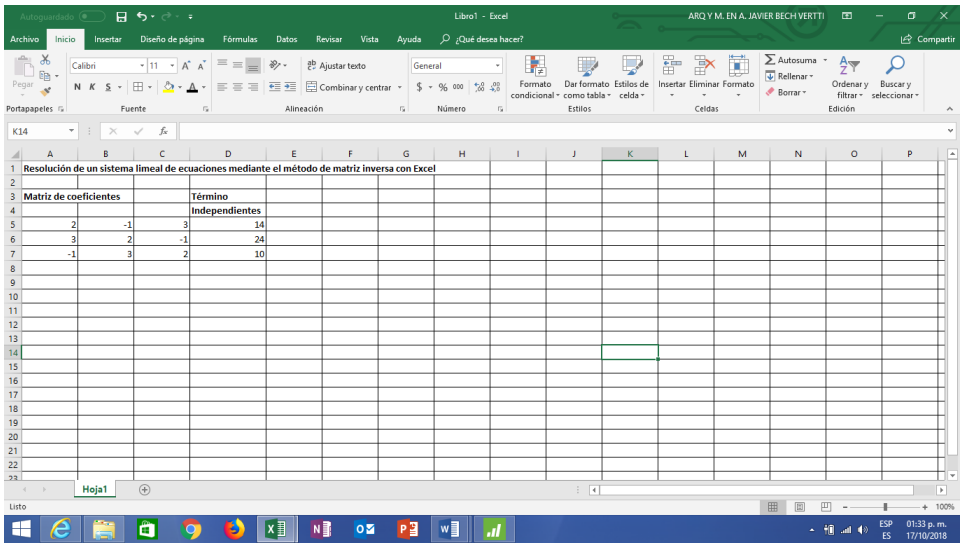
. Introduce los valores independientes en otro rango al lado.

Los pasos que debes seguir para resolver sistemas de ecuaciones con **Excel** es el siguiente:

. Introduce la ecuación de forma que los valores como figura en la imagen de debajo, es decir, los valores deben situarse en celdas independientes para luego poder tomarlas de referencia para los cálculos. Si las ecuaciones originales están en el formato 3x -8 = -2y, deben cambiarse para despejar el valor independiente (3x + 2y = 8).

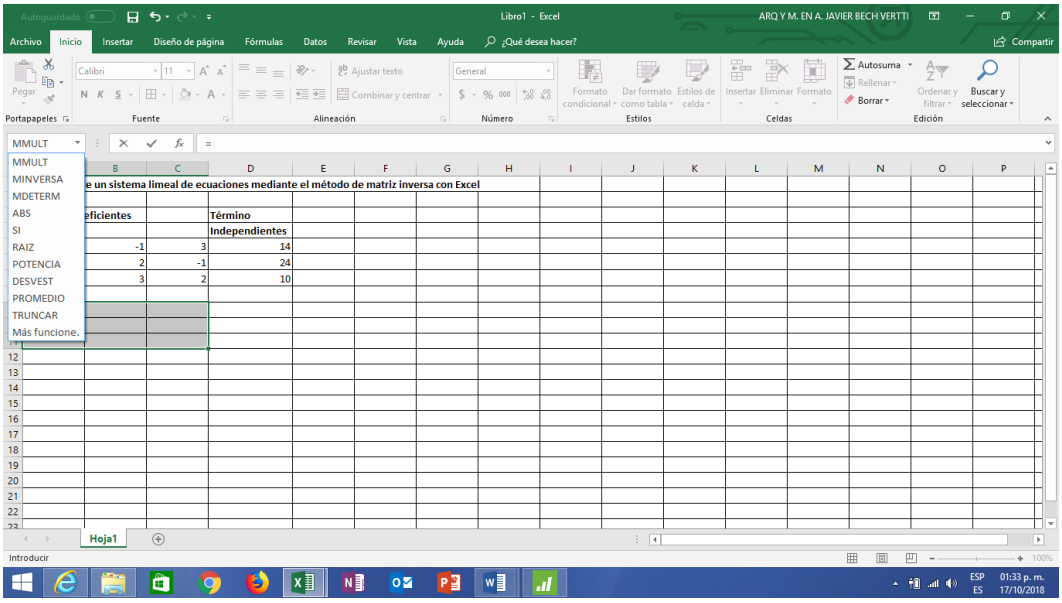
. Introduce los coeficientes del sistema en una matriz de n x n celdas, donde n representa el número de ecuaciones.

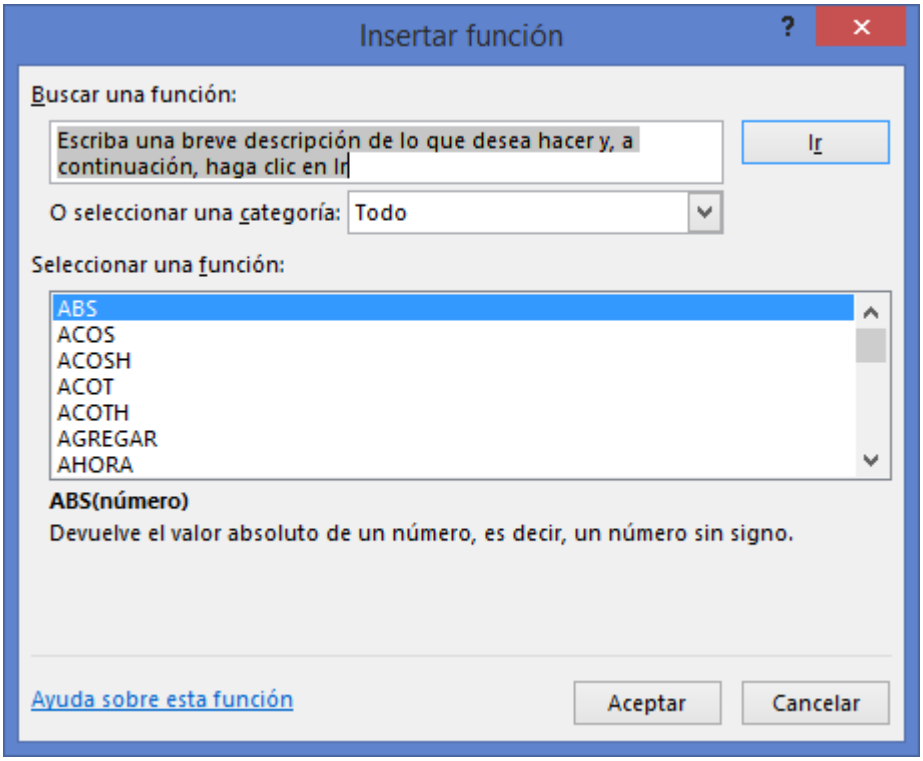
. Introduce los valores independientes en otro rango al lado.



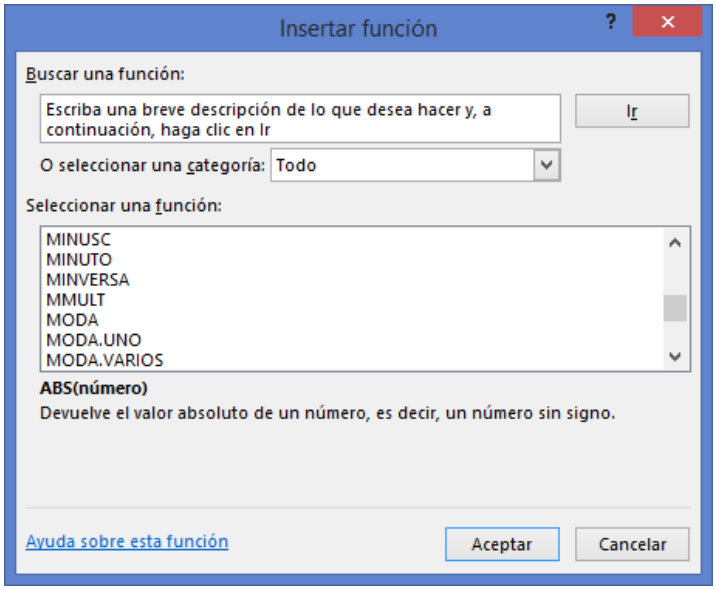
. Utiliza la función matricial **MINVERSA ()** para calcular la matriz inversa de los coeficientes introducidos. (Recuerda que para introducir la función debes pulsar Ctrl + Mayús + Intro).

Como la matriz de coeficientes es de 3 x 3 se necesitas 9 celdas para mostrar los resultados de la matris Inversa para lo cual selecciona con el mouse 9 celdas debajo de la matriz de coeficientes dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MINVERSA** () seleccionando la opción **Más funciones** en la parte inferior

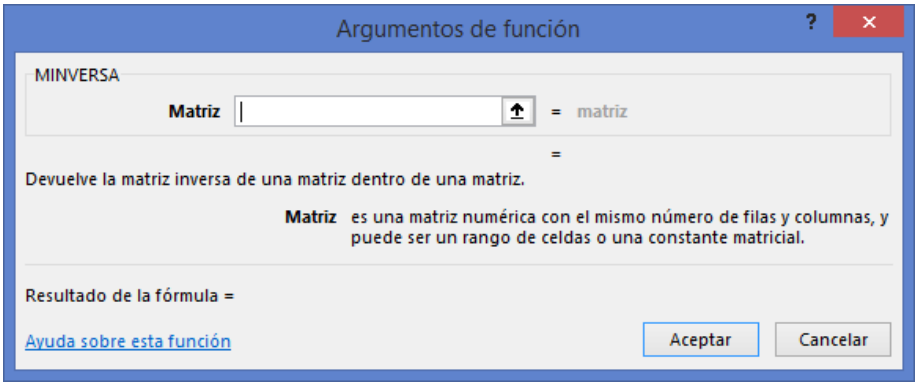




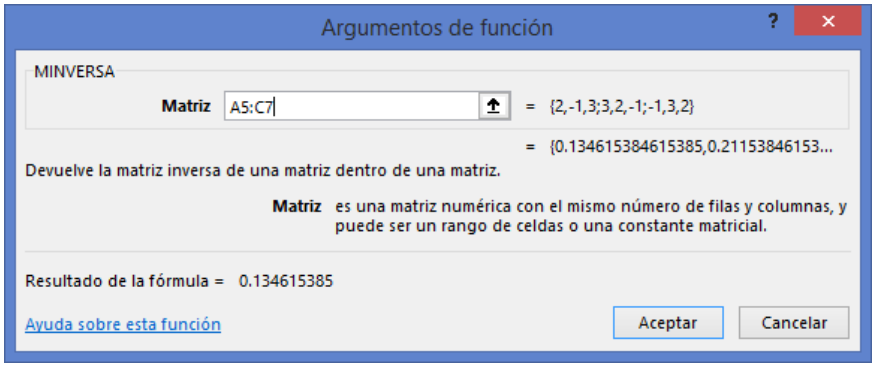
En Seleccionar una función busca **MINVERSA** y da un clic.



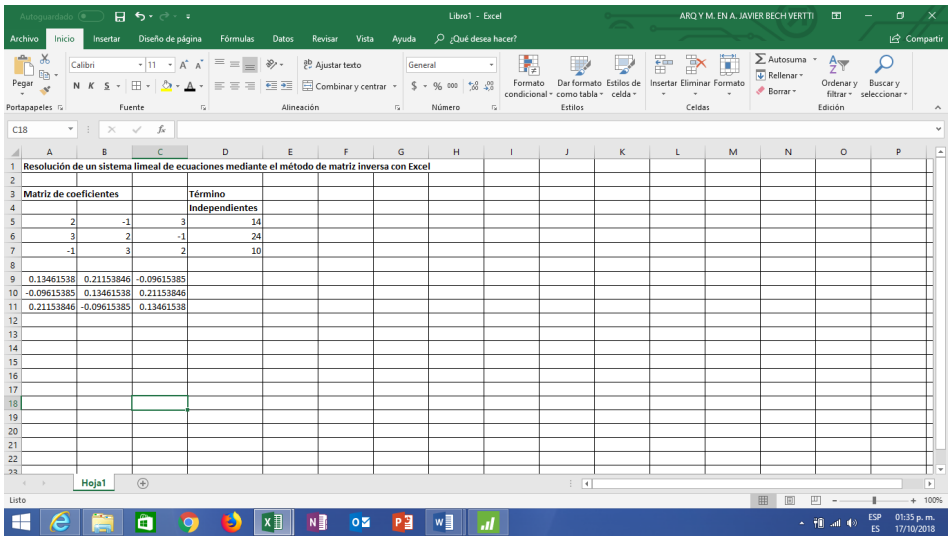
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



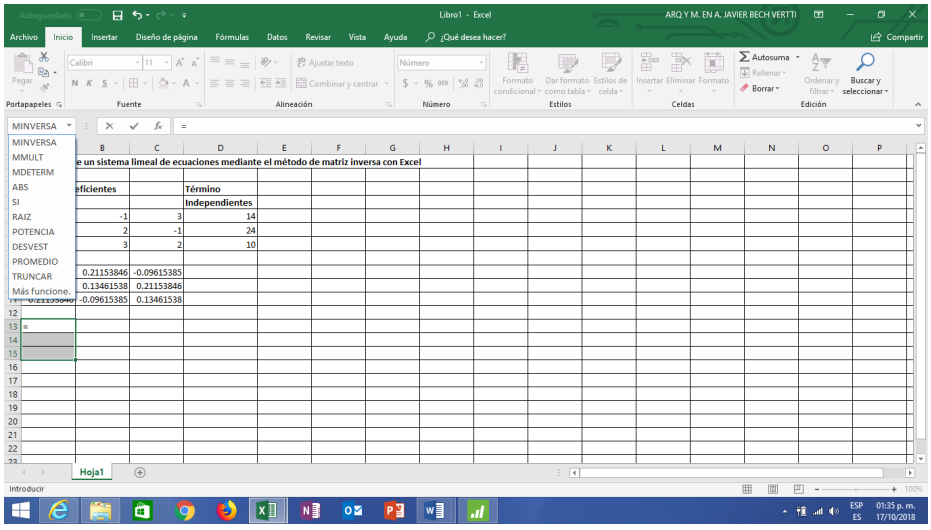
Selecciona las celdas de la matriz de coeficientes, en este caso A5:C7

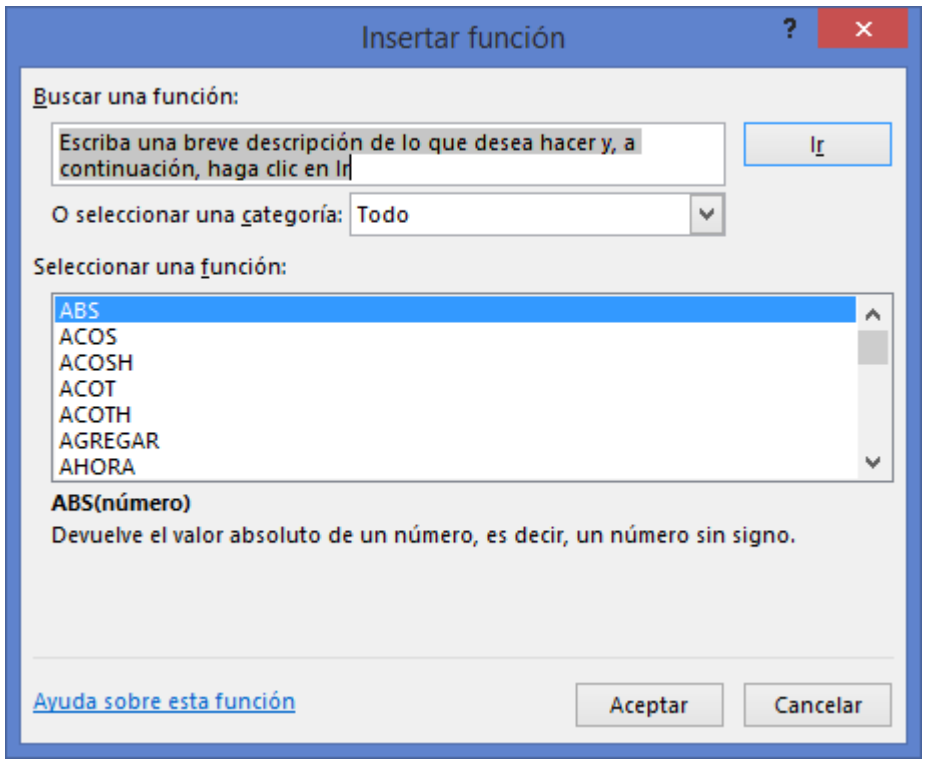


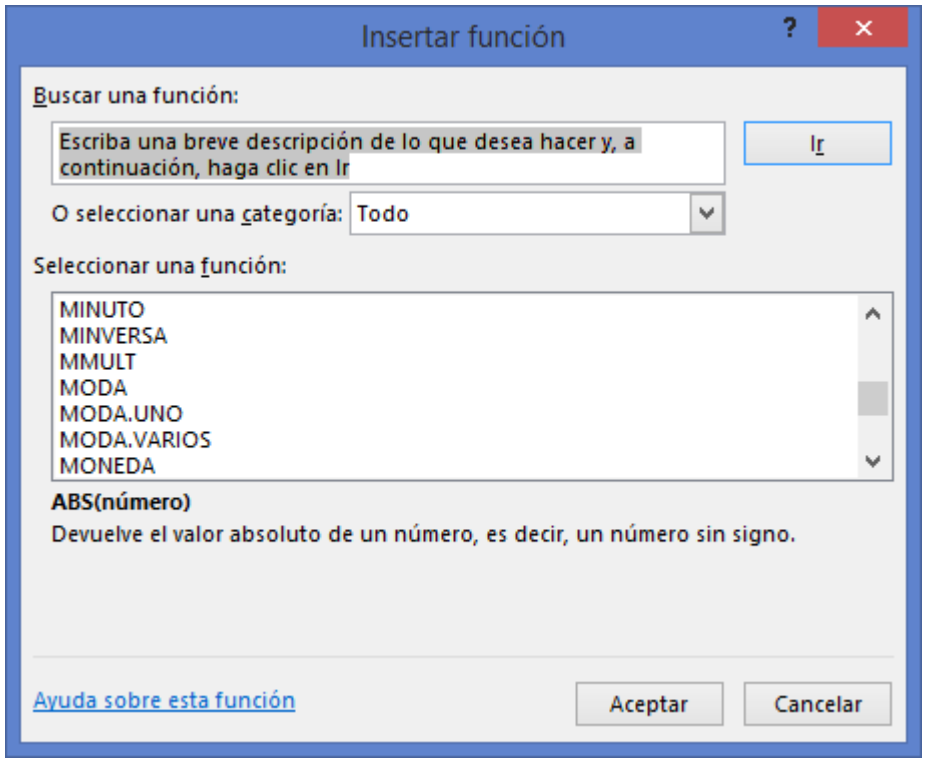
Ojo, no pulses Aceptar. La matriz inversa se determinará al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter)**.



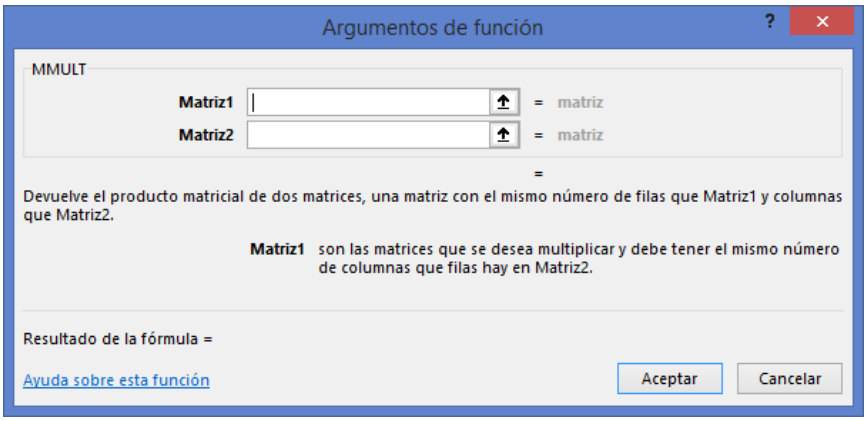
Como la matriz de los términos constantes es un vector de 1 x 3 se necesitas 3 celdas para mostrar los resultados de las Incógnitas **X, Y y Z** para lo cual selecciona con el mouse 3 celdas debajo de la **matriz inversa** dejando un renglón libre. Teclea el signo = para que se activen las funciones de ecuaciones en el margen izquierdo superior y busca la función matricial **MMULT** () seleccionando la opción **Más funciones** en la parte inferior



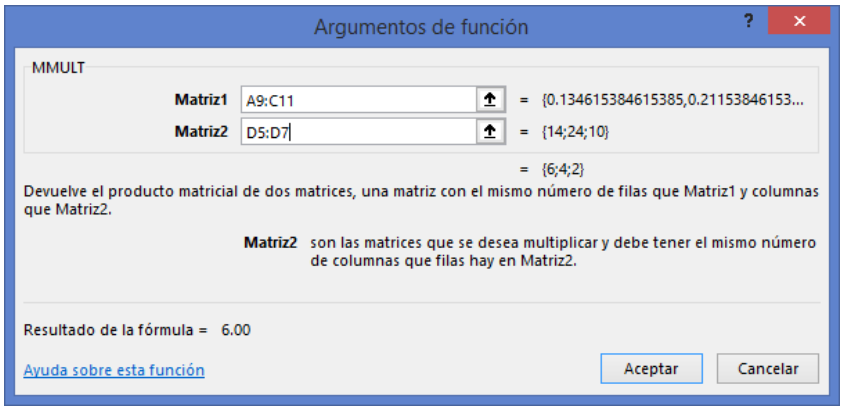




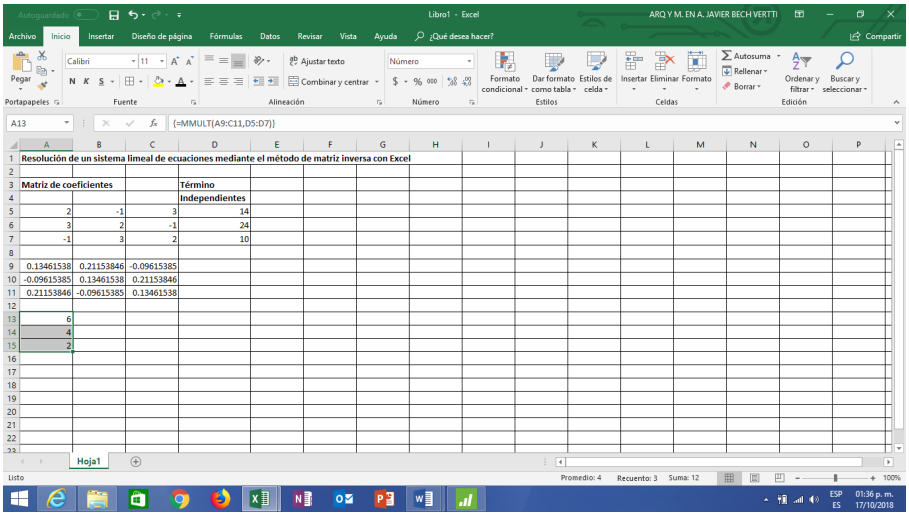
Da un clic en **Aceptar** para que aparezca el cuadro de dialogo Argumentos de función



En el campo que dice Matriz 1 selecciona las celdas de la matriz de inversa, en este caso A9:C11 y en el campo que dice Matriz 2 selecciona las celdas de los términos constantes, en este caso D5:D7



Ojo, no pulses Aceptar. Los resultados de las incógnitas **X, Y, Z** se determinarán al pulsar **simultáneamente** las teclas **Ctrl + Mayús (La flecha hacia arriba) + Intro (Enter), en este caso 6, 4, 2.** Por lo tanto las incognitas quedan **X=6, Y=4 y Z=2 .**



Con lo anterior queda terminado el ejemplo ilustrativo de **resolución de un sistema lineal de ecuaciones mediante el método de matriz inversa** en el paquete **EXCEL**. **Aplica el procedimiento para salir de Windows**

RESPUESTAS A LA AUTOEVALUACIÓN CON REACTIVOS DE **FALSO Ó VERDADERO** DEL CAPÍTULO **1**

1.V	2.V	3.V	4.F	5.F	6.V	7.F	8.V	9.V
10.F	11.F	12.F	13.V	14.F	15.F	16.V	17.F	18.V
19.V	20.V	21.V	22.F	23.V	24.F	25.V	26.V	

RESPUESTAS A LA AUTOEVALUACIÓN CON REACTIVOS DE **OPCIÓN MÚLTIPLE** DEL CAPÍTULO **1**

1.d	2.b	3.c	4.d	5.b	6.d	7.d	8.d	9.a
10.b	11.a	12.b	13.c	14.d				

RESPUESTAS A LAS ACTIVIDADES DE APRENDIZAJE DEL CAPÍTULO 2

2.1.1.1 ACTIVIDAD DE APRENDIZAJE

ACTIVIDAD DE APRENDIZAJE 2.1.1.1 Factor Análisis

El Director de mercadotecnia de una compañía de Bienes Raíces desea conocer cuales dimensiones o indicadores básicos influyen al momento de pronosticar el valor de una casa o vivienda unifamiliar al momento en que un potencial cliente desea ya sea comprar o vender su casa con el objeto de brindar un servicio confiable y rápido. Para llevar a cabo la investigación dividió la Ciudad en tres grandes zonas: NORTE, SUR Y ESTE. En cada zona se aplicó una encuesta a 20 propietarios de casas unifamiliares. El cuestionario estaba conformado por 20 variables que se consideraba podían formar parte de dichos indicadores X_1= Tamaño del lote y X_2= Número de recámaras, , X_3= Número de baños, X_4= Número de habitaciones y X_5= Antigüedad de la casa, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico, X_9= Tipo de combustible para la calefacción, X_{10}= Tipo de sistema para la calefaccción, X_{11}= Tipo de alberca, X_{12}= Cocina con antecomedor y X_{13}= Aire acondicionado, X_{14}= Chimenea, X_{15}= Drenaje municipal X_{16}= Sótano, X_{17}= Cocina integral, X_{18}= Tina de hidromasaje.
Los resultados en la zona SUR fueron los siguientes:

Casa No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	31.5	5	2.0	6	31	4200	0	5	2
2	6.8	4	2.5	8	10	4980	1	3	2
3	6.0	4	1.5	10	30	2857	1	1	1
4	9.0	3	1.5	6	25	2400	1	4	2
5	7.0	3	1.5	7	28	3959	1	3	2
6	6.5	2	1.0	5	35	2250	1	1	2
7	9.1	3	1.5	5	30	4134	1	5	2
8	9.8	4	1.5	7	60	3142	2	3	2
9	6.0	3	1.0	8	30	2500	1	4	2
10	6.0	4	1.0	7	34	2500	1	1	2
11	10.0	4	2.0	7	34	3000	1	1	2
12	10.0	3	1.5	6	30	3930	1	5	2
13	7.0	4	2.0	7	35	3100	1	1	2
14	12.1	3	2.0	7	46	3640	0	2	2
15	8.5	3	1.0	7	35	2725	1	4	2
16	7.0	5	1.5	6	28	4100	1	5	1
17	6.0	3	1.0	6	39	2378	0	1	1
18	11.0	3	2.0	7	33	2767	0	4	2
19	10.0	4	2.0	6	28	4500	2	4	2
20	9.9	3	1.5	8	30	3871	1	5	2

Casa No.	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
1	2	1	1	0	0	1	1	1	1
2	2	1	1	0	0	0	1	1	1
3	1	1	1	0	1	1	0	1	1
4	2	1	1	0	1	0	1	1	1
5	2	1	1	0	0	1	1	0	1
6	2	1	0	0	0	1	1	0	0
7	2	1	1	0	1	1	1	0	0
8	2	1	1	0	0	0	1	0	0
9	2	1	1	0	0	1	1	1	1
10	2	1	1	0	0	1	1	1	1
11	2	1	1	0	0	1	1	1	1
12	2	1	1	1	0	1	1	1	1
13	2	1	0	0	1	1	1	0	0
14	2	1	1	1	1	1	1	1	0
15	2	2	1	0	0	1	1	0	0
16	1	1	1	0	1	1	1	0	0
17	1	2	1	0	0	0	1	0	0
18	1	2	1	1	1	1	1	0	0
19	2	1	1	1	0	1	1	1	1
20	2	1	1	0	0	1	1	1	1

a) ¿Cuántas dimensiones o indicadores esperaría tener como solución inicial en esta investigación?. Explique su respuesta tomando en cuenta los criterios para analizar el número de factores en la solución inicial.
b) Teniendo como supuesto que la solución inicial es correcta, ¿Qué variables se identifican más con cada factor?. Explique su respuesta tomando en cuenta los criterios para identificar cargas significativas y sus prioridades.
c) Según la información de la matriz rotada, ¿con cuál de los factores se identificó mejor el conjunto de 18 variables?. Justifique.
d) ¿Cuál de las variables ha sido mejor explicada en esta investigación por la solución de factores?. Justifique.

Respuesta al inciso a.

Se supone que ya se realizaron todos los pasos necesarios hasta antes de generar la matriz de correlación, la cual es la siguiente:

Matriz de correlación

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1.000								
X_2	0.394	1.000							
X_3	0.368	0.045	1.000						
X_4	-0.217	0.180	0.136	1.000					
X_5	0.068	-0.085	-0.281	-0.102	1.000				
X_6	0.320	0.421	0.647	-0.054	-0.404	1.000			
X_7	-0.378	0.125	-0.076	0.050	0.042	0.158	1.000		
X_8	0.386	0.085	0.100	-0.243	-0.265	0.512	0.071	1.000	
X_9	0.244	-0.283	0.238	-0.200	0.010	0.124	0.182	0.204	1.000
X_{10}	0.182	-0.169	0.088	-0.200	0.003	0.202	0.371	0.111	0.840
X_{11}	-0.075	-0.283	-0.238	-0.050	0.144	-0.382	-0.442	-0.27	-0.216
X_{12}	0.168	0.225	0.059	0.238	-0.090	0.282	-0.062	0.444	-0.140
X_{13}	0.122	-0.169	0.352	-0.134	0.093	0.228	-0.139	0.206	0.210
X_{14}	-0.097	0.071	0.240	0.037	-0.010	-0.056	-0.253	0.020	-0.279
X_{15}	0.146	**0.000**	-0.059	0.022	-0.052	0.076	-0.093	0.111	0.140
X_{16}	0.147	-0.155	0.040	-0.654	0.064	0.141	-0.043	0.305	0.546
X_{17}	0.248	0.203	0.277	0.376	-0.331	0.198	0.019	0.057	0.183
X_{18}	0.153	0.203	0.159	0.376	-0.530	0.238	0.205	0.121	0.183

	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}
X_1									
X_2									
X_3									
X_4									
X_5									
X_6									
X_7									
X_8									
X_9									
X_{10}	1.000								
X_{11}	.0.490	1.000							
X_{12}	-0.167	0.140	1.000						
X_{13}	-0.062	0.140	0.167	1.000					
X_{14}	-0.149	-0.015	-0.105	0.157	1.000				
X_{15}	0.062	-0.140	-0.167	0.250	0.105	1.000			
X_{16}	0.459	0.096	-0.076	0.115	-0.313	-0.115	1.000		
X_{17}	0.302	-0.464	0.369	0.201	-0.179	0.050	-0.208	1.000	
X_{18}	0.302	-0.464	0.369	-0.050	-0.390	0.050	-0.208	**0.798**	1.000

A partir de la matriz de correlación, se puede identificar que las variables con mayor relación son la 17 con la 18 es decir X_{17}= Cocina integral y X_{18}= Tina de hidromasaje, con una correlación de **0.798**, mientras las que tienen menor relación son la 2 con la 15 o sea X_2= Número de recámaras con X_{15}= Drenaje municipal, con una correlación de **0.000**. Sin embargo, a partir de la matriz de correlación, es muy difícil encontrar a las variables afines que se puedan agrupar en factores y resumir así la información. Se necesita verificar entonces la matriz de factores no rotada que proporcione mayor información por lo que se corrió un análisis de factores y se obtuvo la siguiente matriz:

Cargas de factores no rotados :

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
X_1	0.460	0.115	-0.505	-0.052	-0.045	-0.614	-0.241	0.007	0.077
X_2	0.383	-0.450	-0.218	0.280	0.457	-0.382	-0.095	-0.129	-0.154
X_3	0.561	-0.155	-0.426	0.409	-0.040	-0.065	0.467	0.182	-0.144
X_4	0.099	-0.709	0.331	-0.057	-0.227	-0.047	0.191	-0.103	-0.176
X_5	-0.449	0.274	0.006	0.075	-0.034	-0.483	0.170	-0.649	0.084
X_6	0.709	-0.044	-0.381	0.198	0.295	0.234	0.041	-0.020	-0.219
X_7	0.243	0.118	0.559	0.224	0.397	0.331	0.092	-0.455	-0.086
X_8	0.451	0.183	-0.478	-0.278	0.225	0.357	-0.302	-0.189	0.200
X_9	0.477	0.708	0.134	0.029	-0.300	-0.030	0.163	0.016	0.013
X_{10}	0.556	0.645	0.406	0.088	-0.079	-0.100	0.039	-0.003	0.091
X_{11}	-0.602	0.034	-0.386	-0.508	-0.068	0.006	0.135	0.097	-0.303
X_{12}	0.338	-0.354	-0.255	-0.653	0.145	0.063	0.102	-0.324	0.137
X_{13}	0.176	0.117	-0.490	-0.028	-0.539	0.255	0.329	-0.337	-0.144
X_{14}	-0.273	-0.323	-0.423	0.505	-0.169	0.232	0.041	0.016	0.493
X_{15}	0.121	0.034	-0.107	0.262	-0.540	0.114	-0.648	-0.221	-0.300
X_{16}	0.137	0.819	-0.218	-0.153	0.202	-0.003	0.096	0.159	-0.047
X_{17}	0.708	-0.325	0.227	-0.202	-0.356	-0.139	0.089	0.017	0.206
X_{18}	0.734	-0.317	0.397	-0.293	-0.131	0.002	-0.094	0.124	0.014
Eigenvalue	3.8406	2.9077	2.3624	1.5907	1.4786	1.2199	1.0642	1.0435	0.6938
% Var	0.213	0.162	0.131	0.088	0.082	0.068	0.059	0.058	0.039

Variable	Factor 10	Factor 11	Factor 12	Factor 13	Factor 14	Factor 15	Factor 16	Factor 17	Factor 18
X_1	-0.063	-0.220	-0.137	0.028	-0.009	0.007	-0.003	-0.001	-0.042
X_2	0.063	0.288	-0.147	-0.020	0.061	-0.084	0.026	0.006	0.018
X_3	-0.098	-0.005	-0.039	0.046	-0.77	0.121	-0.026	0.029	0.017
X_4	-0.441	0.031	0.081	-0.198	-0.011	-0.035	-0.006	0.007	-0.025
X_5	-0.014	-0.034	0.115	-0.049	-0.061	0.022	-0.051	-0.023	0.023
X_6	0.003	-0.180	0.281	0.023	0.061	-0.041	-0.018	-0.050	-0.002
X_7	0.007	-0.023	-0.257	0.053	0.024	0.060	-0.014	-0.010	-0.028
X_8	-0.227	-0.089	-0.116	-0.197	-0.009	0.018	-0.007	0.015	0.027
X_9	-0.308	0.078	-0.107	0.076	-0.059	-0.038	0.097	-0.037	0.013

X_{10}	-0.151	-0.004	0.089	0.081	0.167	-0.058	-0.072	0.045	0.006
X_{11}	-0.184	0.034	-0.205	0.080	0.133	0.007	-0.065	-0.025	0.008
X_{12}	-0.070	0.161	0.173	0.215	-0.026	0.024	0.030	0.018	-0.011
X_{13}	0.317	-0.042	-0.100	-0.054	0.002	-0.089	0.01	0.022	-0.009
X_{14}	-0.154	0.156	-0.048	0.065	0.021	-0.033	-0.052	-0.023	-0.013
X_{15}	-0.043	0.163	0.070	0.060	-0.005	0.059	-0.014	0.002	0.000
X_{16}	0.085	0.340	0.089	-0.157	-0.057	0.024	-0.040	-0.008	-0.029
X_{17}	0.213	0.076	-0.018	-0.107	0.179	0.091	0.014	-0.031	0.006
X_{18}	0.115	0.014	-0.113	0.051	-0.185	-0.055	-0.097	-0.019	0.012
Eigenvalue	0.6124	0.3813	0.3505	0.2012	0.1378	0.0589	0.0387	0.0111	0067
% Var	0.034	0.021	0.019	0.011	0.008	0.003	0.002	0.001	0.000

Esta matriz no rotada muestra las cargas de factor que, como ya se mencionó, indican la relación de cada variable con cada factor .

Aquí la mejor solución en cuanto al número de factores sería un factor que agrupara a las diez y ocho variables. Todas tendrían algo en común entre sí, mientras que la peor solución serian diez y ocho factores, lo cual representaría que las variables no tuvieron nada en común y no se podría resumir nada la información.
Tanto por el criterio de la raíz latente como por el porcentaje de variación explicada acumulada, se tiene como solución inicial ocho factores, ya que el *Eigenvalue* de los ocho primeros es mayor a 1 y el porcentaje de variación explicada acumulada está entre 60 % y 95%, es de 86.20%.

En este problema de diez y ocho variables la varianza ó variación máxima que se podría obtener, se tiene que F1 explica 3.846 el F2 explica 2.9070, el F3 explica 2.3624, el F4 explica 1.5907, el F5 explica 1.4786, el F6 explica 1.2199, el F7 explica 1.0642 y el F8 explica 1.0435 de acuerdo con su *Eigenvalue*. El total de la variación es de 15.5075 (total de los *Eigenvalue* de los 8 factores y se pierde 2.4925 (18 – 15.5075) de información.

Respuesta al inciso b.

Una vez se ha determinado el número de factores para la solución inicial se trata de agrupar a las variables en esos ocho factores, lo cual se determina analizando la matriz de factores rotada, que es la siguiente:
Cargas de factores rotados y comunalidades

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Comunalidad
X_1	0.304	0.061	**-0.636**	-0.242	**-0.537**	0.030	-0.158	-0.214	0.920
X_2	-0.268	0.132	**-0.848**	-0.112	0.078	0.156	-0.053	0.042	0.856
X_3	0.105	0.135	**-0.658**	0.115	-0.066	**-0.595**	0.269	0.194	0.944
X_4	-0.424	**0.707**	0.022	0.141	0.119	-0.061	-0.050	0.079	0.726
X_5	0.032	-0.237	0.041	0.104	0.033	-0.043	**-0.946**	0.015	**0.968**
X_6	0.143	0.001	**-0.611**	-0.384	0.220	-0.284	0.402	-0.002	0.832
X_7	0.160	0.054	-0.021	-0.072	**0.928**	0.111	-0.073	0.061	0.918
X_8	0.166	-0.181	-0.179	**-0.797**	0.035	-0.100	0.260	-0.202	0.847
X_9	**0.889**	0.061	0.033	0.044	0.055	-0.238	0.006	-0.095	0.865
X_{10}	**0.892**	0.145	-0.072	0.092	0.271	0.062	0.011	**-0.87**	0.916
X_{11}	-0.261	-0.301	0.453	-0.210	**-0.535**	-0.068	-0.158	0.279	0.802
X_{12}	-0.176	0.389	-0.066	**-0.784**	-0.080	-0.082	-0.075	0.229	0.871
X_{13}	0.111	0.044	0.094	-0.253	-0.119	**-0.849**	-0.124	-0.158	0.863
X_{14}	**-0.526**	-0.268	-0.136	0.246	-0.038	-0.459	0.129	-0.203	0.698
X_{15}	0.010	0.042	0.019	-0.006	-0.041	-0.113	0.010	**-0.923**	0.870
X_{16}	**0.709**	-0.485	0.012	-0.207	-0.127	-0.013	0.062	0.189	0.836
X_{17}	0.207	**0.856**	-0.180	-0.098	-0.038	-0.111	0.134	-0.067	0.853
X_{18}	0.211	**0.823**	-0.116	-0.191	0.102	0.174	0.329	-0.051	0.924
Eigenvalue	2.9933	2.6178	2.2545	1.7847	1.6399	1.5469	1.4254	1.2450	15.5075
% Var	0.166	0.145	0.125	0.099	0.091	0.086	0.079	0.069	0.862

Ni el *Eigenvalue* ni la comunalidad total, ni el porcentaje de variación explicada total cambian al rotar los factores, es decir, no cambian de la matriz no rotada a la rotada, **solo se distribuyeron las cargas de**

factor y ahora sí algunas cargas tienen valores más grandes que en la no rotada como para poder determinar la forma como se agrupan las variables en cada factor.
Lo que si cambia es la importancia en que cada factor explica a las variables (*Eigenvalue* de cada factor) y en la matriz rotada es posible que cada factor hacia la derecha explique menos a las variables.

Para determinar en cual factor se agrupara cada variable debe analizarse la matriz rotada, buscando las cargas significativas. Al identificar las cargas significativas, no importa el signo que tengan, sino la magnitud de la relación, el signo negativo solo expresa una relación inversa. En principio se buscan cargas mayores a 0.50 utilizando la Prioridad 1, en caso de que una variable no tuviera una carga mayor a 0.5 se aplicarían las prioridades 2, 3 o 4, según sea el caso, de los criterios de cargas significativas en toda la matriz rotada. En este ejemplo se tiene:

X_1= Tamaño del lote se identifica con el F3 con una carga de 0.636
X_2= Número de recámaras se identifica con el F3 con una carga de 0.848
X_3= Número de baños se identifica con el F3 con una carga de 0.658
X_4= Número de habitaciones se identifica con el F2 con una carga de 0.707
X_5= Antigüedad de la casa se identifica con el F7 con una carga de 0.946
X_6= Impuestos anuales se identifica con el F3 con una carga de 0.611
X_7= Tipo de estacionamiento interior se identifica con el F5 con una carga de 0.928
X_8= Estilo arquitectónico se identifica con el F4 con una carga de 0.797
X_9= Tipo de combustible para la calefacción se identifica con el F1 con una carga de 0.889
X_{10}= Tipo de sistema para la calefacción se identifica con el F1 con una carga de 0.892
X_{11}= Tipo de alberca se identifica con el F5 con una carga de 0.535
X_{12}= Cocina con antecomedor se identifica con el F4 con una carga de 0.784
X_{13}= Aire acondicionado se identifica con el F6 con una carga de 0.849
X_{14}= Chimenea se identifica con el F1 con una carga de 0.526
X_{15}= Drenaje municipal se identifica con el F8 con una carga de 0.923
X_{16}= Sótano se identifica con el F1 con unas cargas de 0.709
X_{17}= Cocina integral se identifica con el F2 con una carga de 0.856
X_{18}= Tina de hidromasaje se identifica con el F2 con una carga de 0.823

Respuesta al inciso c.

El porcentaje de varianza, que es el porcentaje de variación explicada con respecto a la máxima variación que se podría tener, se obtiene, por ejemplo para F1 dividiendo 2.9933/18, lo que es igual a 0.166 o 16.6 %. El Factor que ha explicado más al conjunto de variables es el 1, y después el 2 y así sucesivamente lo cual es lógico pues cada factor adicional explica menor cantidad, ya que cada vez queda menos por explicar, en este caso, por ejemplo, queda por explicar 2.4925 de variación; si se agregan los 10 factores que faltan, ya cada uno explicaría muy poco y no convendría. El porcentaje de variación explicada acumulada total es de 0.862 o de 86.2% de un máximo por explicar de 100%.

Respuesta al inciso d.

Se puede observar que la variable mejor explicada por el conjunto de factores es la variable X_5=Antigüedad de la casa, ya que es la que tiene la mayor comunalidad de 0.968 o de 96.8%.

2.2.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 2.2.1.1 Cluster Análisis. Medida de Distancia.**

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de los hombres mexicanos, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras . El estudio antropométrico contempló 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y la cintura de cinco personas del género **MASCULINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Cintura (cms)
1	1.73	63.56	89.10
2	1.85	83.99	92.20
3	1.70	74.91	90.22
4	1.63	54.48	88.60
5	1.93	95.34	93.50
Promedio:	**1.77**	**74.46**	**90.72**
Desviación estándar	**0.12**	**16.16**	**2.08**

a. Obtenga la pareja de personas más parecidas y la menos parecida.
b. Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.
c. Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

Resolución del inciso a)

Estandarización de variables

Antes que nada hay que se deben estandarizar las variables, es decir, eliminar el efecto de las unidades y hacerlas comparables, de manera que todas estén en términos de "Z" mediante la siguiente expresión:

$$Z = \frac{x - \bar{x}}{s}$$

$Z_{11} = \frac{1.73 - 1.77}{.12} = -0.33$	$Z_{12} = \frac{63.56 - 74.46}{16.16} = -.67$	$Z_{13} = \frac{89.10 - 890.72}{2.08} = -.78$
$Z_{21} = \frac{1.85 - 1.77}{0.12} = .67$	$Z_{22} = \frac{83.99 - 74.46}{16.16} = .59$	$Z_{23} = \frac{92.20 - 90.72}{2.08} = .71$
$Z_{31} = \frac{1.70 - 1.77}{.12} = -0.58$	$Z_{32} = \frac{74.91 - 74.46}{16.16} = 0.03$	$Z_{33} = \frac{90.22 - 90.72}{2.08} = -.24$
$Z_{41} = \frac{1.63 - 1.77}{.12} = -1.17$	$Z_{42} = \frac{54.48 - 74.46}{16.16} = -1.24$	$Z_{43} = \frac{88.60 - 90.72}{2.08} = -1.02$
$Z_{51} = \frac{1.93 - 1.77}{.12} = 1.33$	$Z_{45} = \frac{95.34 - 74.46}{16.16} = 1.29$	$Z_{53} = \frac{93.50 - 90.72}{2.08} = 1.34$

Los datos estandarizados de las 5 personas, al considerar 3 variables, se presenta de la siguiente manera:

Cálculo de las distancias:

$$d_{ij} = \left[\sum_{K=1}^{r} (x_{ik} - x_{jk})^2\right]^{0.5}$$

$$d_{1,2} = [(-0.33 - .67)^2 + (-.67 - .59)^2 + (-.78 - .71)^2]^{0.5} = 2.19$$

$$d_{1,3} = [(-0.33 - (-0.58))^2 + (-.67 - .03)^2 + (-.78 - (-.24))^2]^{0.5} = .92$$

$$d_{1,4} = [(-0.33 - (-1.17))^2 + (-.67 - (-1.24))^2 + (-.78 - (-1.02))^2]^{0.5} = 1.04$$

$$d_{1,5} = [(-0.33 - 1.33)^2 + (-.67 - 1.29)^2 + (-.78 - 1.34)^2]^{0.5} = 3.33$$

$$d_{2,3} = [(.67 - (-0.58))^2 + (.59 - 0.03)^2 + (.71 - (-.24))^2]^{0.5} = 1.67$$

$$d_{2,4} = [(.67 - (-1.17))^2 + (.59 - (-1.24))^2 + (.71 - (-1.02))^2]^{0.5} = 3.12$$

$$d_{2,5} = [(.67 - 1.33)^2 + (.59 - 1.29)^2 + (.71 - 1.34)^2]^{0.5} = 1.15$$

$$d_{3,4} = [(-0.58 - (-1.17))^2 + (0.03 - (-1.24))^2 + (-.24 - (-1.02))^2]^{0.5} = 1.60$$

$$d_{3,5} = [(-0.58 - 1.33)^2 + (0.03 - 1.29)^2 + (-.24 - 1.34)^2]^{0.5} = 2.78$$

$$d_{4,5} = [(-1.17 - 1.33)^2 + (-1.24 - 1.29)^2 + (-1.02 - 1.34)^2]^{0.5} = 4.27$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5 personas a la que le nombraremos ***matriz de distancia*** la cual representa que tan parecidas son las personas con respecto a la distancia entre ellas.

Como esta matriz es una matriz espejo, es decir la matriz superior es igual a la matriz inferior y la diagonal contiene ceros ya que la distancia que hay al comparar una persona consigo misma es nula, sólo será necesario utilizar la información de la matriz inferior como se muestra a continuación:

	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	3.33	1.15	2.78	4.27	0

Al observar la matriz de distancia se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya distancia entre sí es la menor que está formada por las personas 1 y 3, con una medida de distancia de 0.92 y la menos parecida será aquella cuya distancia entre sí es la mayor que está formada por las personas 4 y 5, con una medida de distancia de 4.27.

	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	~~3.33~~	~~1.15~~	~~2.78~~	4.27	0

Resolución al inciso b)

Vinculación o eslabonamiento:

4. Eslabonamiento Simple:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer cluster.

Paso 2. Eslabonar. Llenar la matriz con la distancia menor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia más corta, que en este caso es 0.92

	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	3.33	1.15	2.78	4.27	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	3.33	1.15	2.78	4.27	0

El nuevo cluster (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

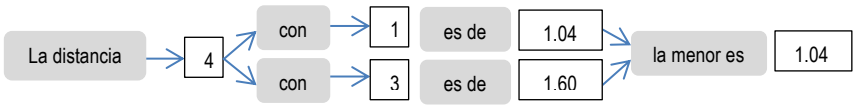
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al nuevo cluster (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 2.19 y la de la persona 2 con la 3 que es 1.67, eligiendo entonces la distancia menor de las dos personas que en este caso es 1.67. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	1.67	0		
4			0	
5				0

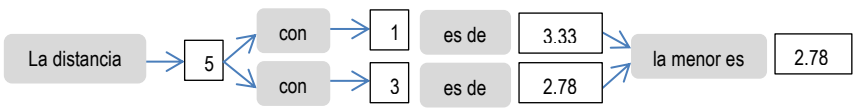
Repetimos el mismo criterio para eslabonar la persona 4 con la pareja de personas (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4 con 1 que es 1.04 y la de la personas 4 con 3 que es 1.60, eligiendo entonces la distancia menor de las dos personas que en este caso es 1.04. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	1.67	0		
4	1.04		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 3.33 y la de la personas 5 con 3 que es 2.78, eligiendo entonces la distancia menor de las dos personas que en este caso es 2.78. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	1.67	0		
4	1.04		0	
5	2.78			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	1.67	0		
4	1.04	3.12	0	
5	2.78	1.15	4.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo cluster o para eslabonar una persona al cluster ya formado, que en este caso sería la persona 3 que se eslabonaría al cluster formado por las personas (1,3) con la distancia menor de 1.04 como se muestra a continuación:

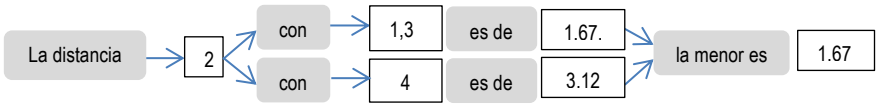
	(1,3)	2	4	5
(1,3)	0			
2	1.67	0		
4	1.04	3.12	0	
5	2.78	1.15	4.27	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3,4)	2	5
(1,3,4)	0		
2		0	
5			0

Paso 2. Eslabonar

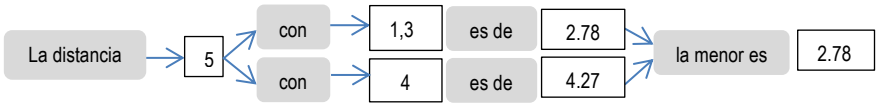
Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al cluster (1,3,4) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las 1 y 3 que es 1.67 y la de la persona 2 con la 4 que es 3.12, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 1.67. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3,4)	2	5
(1,3,4)	0		
2	1.67	0	
5			0

Repetimos el mismo criterio para eslabonar la persona 5 con el cluster (1,3,4); al comparar (1,3,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con las personas 1 y 3 que es 2.78 y la de la personas 5 con la 4 que es 4.27, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.78. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3,4)	2	5
(1,3,4)	0		
2	1.67	0	
5	2.78		0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3,4)	2	5
(1,3,4)	0		
2	1.67	0	
5	2.78	1.15	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo cluster o para eslabonar una persona al cluster ya formado, que en este caso serían las personas 2 y 5 para formar un nuevo conglomerado con la distancia menor de 1.15 como se muestra a continuación:

	(1,3,4)	2	5
(1,3,4)	0		
2	1.67	0	
5	2.78	1.15	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,3,4)	(2,5)
(1,3,4)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al cluster (1,3,4) con el resto de las personas; al comparar (1,3,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las personas (1,3,4) que es 1.67 y la de la persona 5 con las (1,3,4) que es

2.78, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 1.67. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida:

	(1,3,4)	(2,5)
(1,3,4)	0	
(2,5)	1.67	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 1.67.

Tabla de eslabonamiento

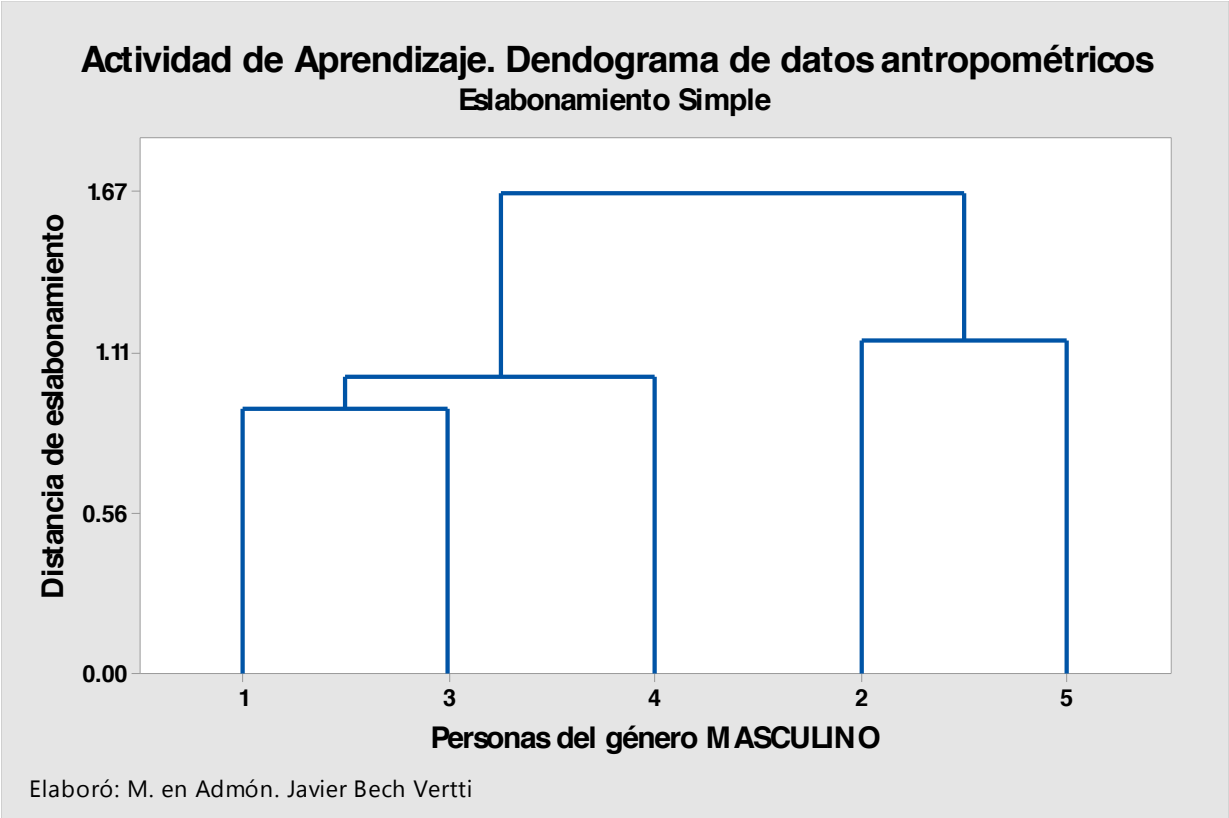
La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.92	1	3	1
2	3	1.04	1	4	1
3	2	1.15	2	5	2
4	1	1.67	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La segunda columna no muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 1 y la 3 y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 4 al primer conglomerado formado por las personas 1 y 3 como se muestra en la columna 4 a una distancia de 1.04. En la etapa 3 se forma un nuevo conglomerado como se aprecia en la columna 5 formado por las personas 2 y 5 a una distancia de 1.15 y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de 1.67 Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

```
Variables estandarizadas, Distancia euclediana, Enlace simple
Pasos de amalgamación


                                                                  Número de obs. en
          Número de    Nivel de   Nivel de  Conglomerados         Nuevo    el conglomerado
Paso  conglomerados  semejanzal  distancia   incorporados  conglomerado              nuevo
   1              4     78.4155    0.91942  1           3             1                  2
   2              3     75.8054    1.03060  1           4             1                  3
   3              2     72.9778    1.15104  2           5             2                  2
   4              1     60.9116    1.66501  1           2             1                  5
```

Dendograma



El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,4) y el conglomerado 2 por las personas (2,5).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

1. **Eslabonamiento Completo:**

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia mayor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, las personas más parecidas se elegirán con la distancia más corta, que en este caso es 0.92

	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	3.33	1.15	2.78	4.27	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	3.33	1.15	2.78	4.27	0

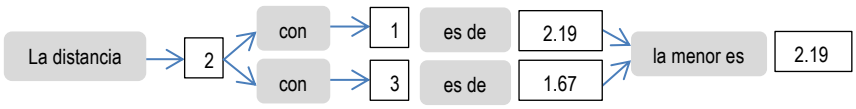
El nuevo conglomerado (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al nuevo conglomerado

(1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 2.19 y la de la persona 2 con la 3 que es 1.67, eligiendo entonces la distancia mayor de las dos personas que en este caso es 2.19. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

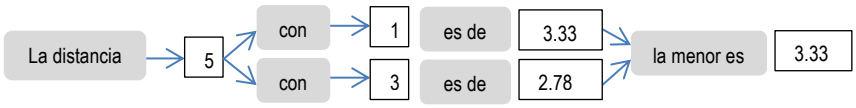
	(1,3)	2	4	5
(1,3)	0			
2	2.19	0		
4			0	
5				0

Repetimos el mismo criterio para eslabonar a la persona 4 con la pareja de personas (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre las personas 4 con 1 que es 1.04 y la de la personas 4 con 3 que es 1.60, eligiendo entonces la distancia mayor de las dos personas que en este caso es 1.60. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	2.19	0		
4	1.60		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 3.33 y la de la personas 5 con 3 que es 2.78, eligiendo entonces la distancia mayor de las dos personas que en este caso es 3.33. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	2.19	0		
4	1.60		0	
5	3.33			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	2.19	0		
4	1.60	3.12	0	
5	3.33	1.15	4.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo cluster o para eslabonar una persona al cluster ya formado. En este se forma un nuevo conglomerado con las personas 2 y 5 con una distancia de 1.15 como se muestra a continuación:

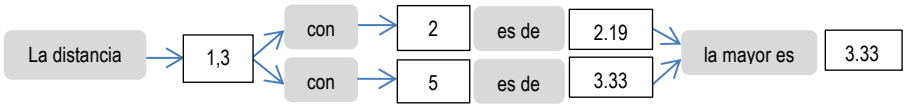
	(1,3)	2	4	5
(1,3)	0			
2	2.19	0		
4	1.60	3.12	0	
5	3.33	1.15	4.27	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3)	(2,5)	4
(1,3)	0		
(2,5)		0	
4			0

Paso 2. Eslabonar

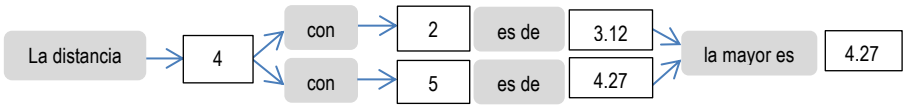
Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre el conglomerado (1,3) con la persona 2 que es 2.19 y al conglomerado (1,3) con la de la persona 5 que es 3.33, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 3.33. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,5)	4
(1,3)	0		
(2,5)	3.33	0	
4			0

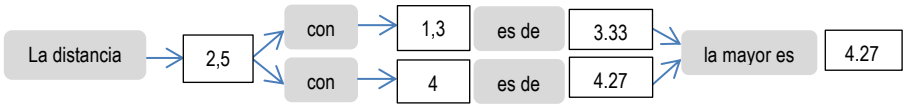
Repetimos el mismo criterio para eslabonar la persona 4 con el conglomerado (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre la persona 4 con la persona 2 que es 3.12 y la de la personas 4 con la 5 que es 4.27, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 4.27. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,5)	4
(1,3)	0		
(2,5)	3.33	0	
4	1.60		0

Repetimos el mismo criterio para eslabonar la persona (2,5) con el conglomerado (); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre la persona 3 con la persona 2 que es 2.34 y la de la personas 3 con la 5 que es 2.46, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 2.46. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,5)	4
(1,3)	0		
(2,5)	3.33	0	
4	1.60	4.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona a un conglomerado ya formado, que en este caso sería la persona 4 que se eslabonaría al conglomerado formado por las personas (1,3) con la distancia menor de 1.60 como se muestra a continuación:

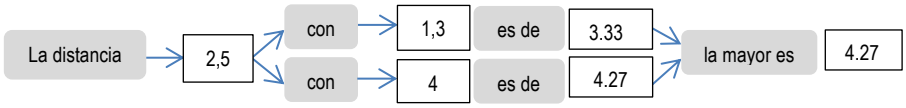
	(1,3)	(2,5)	4
(1,3)	0		
(2,5)	3.33	0	
4	1.60	4.27	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,3,4)	(2,5)
(1,3,4)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (1,3,4) con el conglomerado (2,5); al comparar (1,3,4) con el conglomerado (2,5) las personas 2 y 5 se debe verificar en la matriz anterior las distancias entre las personas 2 y 5 con las personas (1,3) que es 3.33 y las personas 2 y 5 con la persona 4 que es 4.27, eligiendo entonces la distancia mayor de las dos combinaciones o cinco personas que en este caso es 4.27. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,3,4)	(2,5)
(1,3,4)	0	
(2,5)	4.27	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 4.27.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	.92	1	3	1
2	3	1.15	2	5	2
3	2	1.60	1	4	1
4	1	4.27	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan **cuatro etapas**. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas

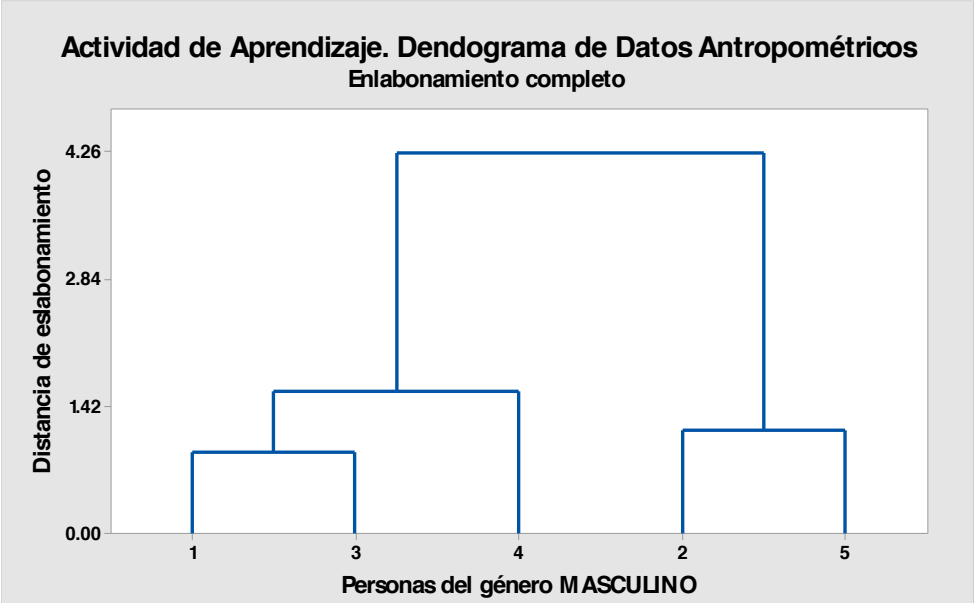
en la **columna 4** que en este caso fueron la 1 y la 3 a una distancia de **0.92** y en la **quinta columna** nos señala la formación del primer conglomerado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 5 a una distancia de **1.15** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 3 al primer conglomerado formado por las personas 1 y 4 a una distancia de **1.60** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **4.27**. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Altura, Peso, Cintura

```
Variables estandarizadas, Distancia euclediana, Enlace completo
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	78.4155	0.91942	1	3	1	2
2	3	72.9778	1.15104	2	5	2	2
3	2	62.5678	1.59447	1	4	1	3
4	1	0.0000	4.25962	1	2	1	5

Dendrograma



Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro

caso el conglomerado 1 está formado por las personas (1,3,4) y el conglomerado 2 por las personas (2,5).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 3 y 4 los valores de distancia cambian abruptamente de 1.60 a 4.27 cuando se unen ambos grupos, por lo tanto, la solución más factible parece ser la de 2 conglomerados (1,3,4) y (2,5).

5. Eslabonamiento Promedio:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia promedio, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.92

	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	3.33	1.15	2.78	4.27	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

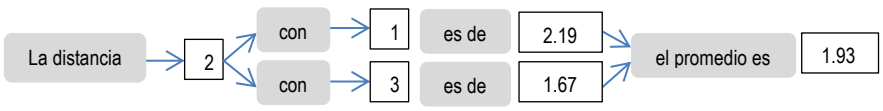
	1	2	3	4	5
1	0				
2	2.19	0			
3	.92	1.67	0		
4	1.04	3.12	1.60	0	
5	3.33	1.15	2.78	4.27	0

El nuevo cluster (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al nuevo conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 2.19 y la de la persona 2 con la 3 que es 1.67, eligiendo entonces la distancia promedio de las dos personas que en este caso es 1.93. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	1.93	0		
4			0	
5				0

Repetimos el mismo criterio para eslabonar la persona 4 con la pareja de personas (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4 con 1 que es 1.04 y la de la personas 4 con 3 que es 1.60, eligiendo entonces la distancia promedio de las dos personas que en este caso es 1.32. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	1.93	0		
4	1.32		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 3.33 y la de la personas 5 con 3 que es 2.78, eligiendo entonces la distancia promedio de las dos personas que en este caso es 3.06. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	1.93	0		
4	1.32		0	
5	3.06			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	1.93	0		
4	1.32	3.12	0	
5	3.06	1.15	4.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 2 y 5 a una distancia menor de 1.15 como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2	1.93	0		
4	1.32	3.12	0	
5	3.06	1.15	4.27	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3)	(2,5)	4
(1,3)	0		
(2,5)		0	
4			0

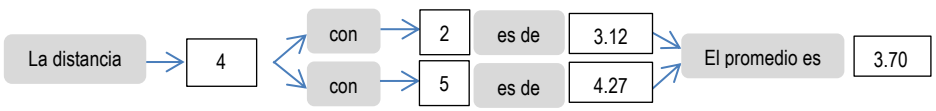
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado(1,3) con las personas del conglomerado (2,5); al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre las persona (1,3) y la persona 2 que es 1.93 y la de las personas (1,3) con la persona 5 que es 3.06, eligiendo entonces la distancia promedio de las dos combinaciones o cuatro personas que en este caso es 2.50. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,5)	4
(1,3)	0		
(2,5)	2.50	0	
4			0

Repetimos el mismo criterio para eslabonar a la persona 4 con el conglomerado (2,5); al comparar (2,5) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4 con la persona 2 que es 3.12 y la de la personas 4 con la 5 que es 4.27, eligiendo entonces la distancia promedio de las dos combinaciones o tres personas que en este caso es 3.70. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,5)	4
(1,3)	0		
(2,5)	2.50	0	
4	1.32	3.70	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 4 para eslabonarse al conglomerado formado por las personas (1,3) con la distancia menor de 1.32 como se muestra a continuación:

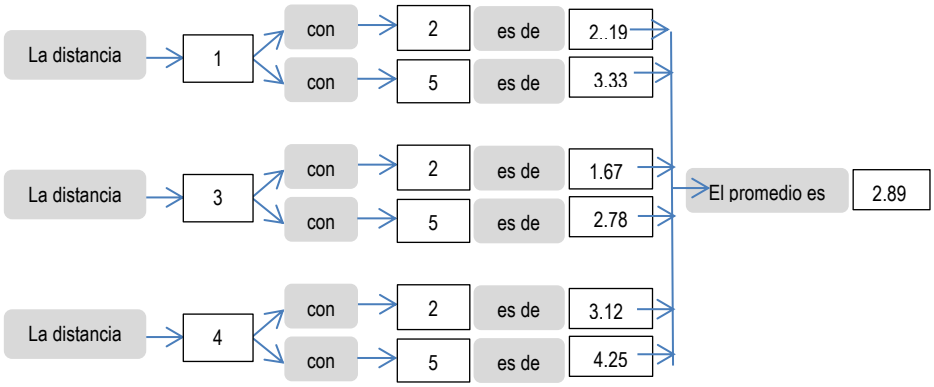
	(1,3)	(2,5)	4
(1,3)	0		
(2,5)	2.50	0	
4	1.32	3.70	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,3,4)	(2,5)
(1,3,4)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado (1,3,4) con el resto de las personas del conglomerado (2,5); al comparar (2,5) con las personas (1,3,4) se debe verificar **por única vez en la primera matriz de distancias** entre las personas 1,3 y 4 con las personas 2,5, eligiendo entonces la distancia promedio de las seis combinaciones o cinco personas que en este caso es 2.89. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,3,4)	(2,5)
(1,3,4)	0	
(2,5)	2.89	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 2.89.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.92	1	3	1
2	3	1.15	2	5	2
3	2	1.32	1	4	1
4	1	2.89	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **cuarta columna** que en este caso fueron la 1 y la 3 a una distancia de **0.92** y en la **quinta columna** nos señala el número de conglomerado formado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 5 a una distancia de **1.15** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 4 al primer conglomerado formado por las personas 1 y 3 a una distancia de **1.32** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **2.89**. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

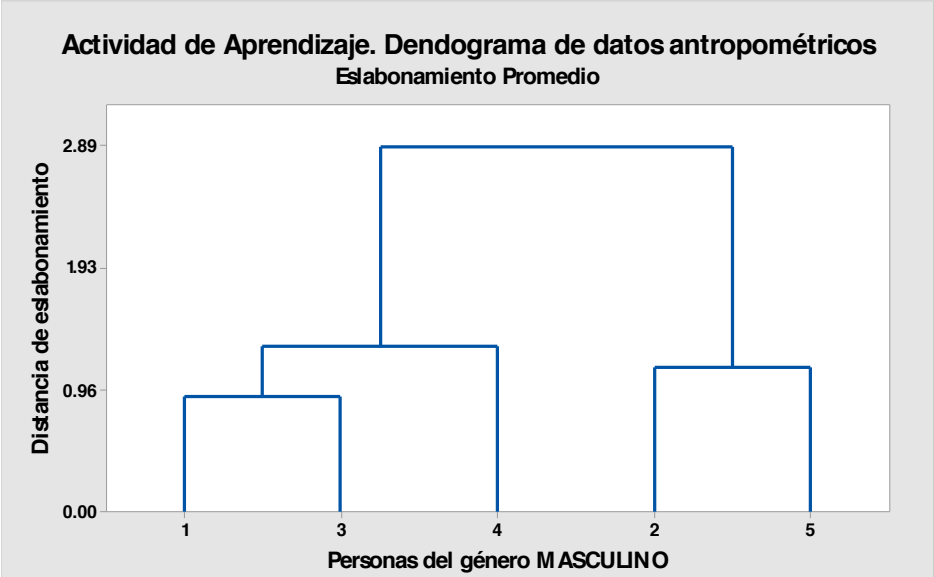
Variables estandarizadas, Distancia euclediana, Enlace de promedios
Pasos de amalgamación

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	78.4155	0.91942	1	3	1	2
2	3	72.9778	1.15104	2	5	2	2
3	2	69.1866	1.31253	1	4	1	3
4	1	32.1558	2.88990	1	2	1	5

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,4) y el conglomerado 2 por las personas (2,5).



Elaboró: M. en Admón. Javier Bech Vertti

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos. En este caso entre la etapa 3 y 4 los valores de distancia cambian

abruptamente de 1.32 a 2.89 cuando se unen ambos grupos, por lo tanto, la solución más factible parece ser la de 2 conglomerados (1,3,4) y (2,5).

Respuesta al inciso c)

Nombre de los conglomerados

Una vez elegidos el número de conglomerados definitivo, para finalizar el proceso es deseable colocar un nombre apropiado a los conglomerados. En este caso el conglomerado 1 está formado por las personas 1,4 y 3 y el conglomerado 2 está formado por las personas 2 y 5. Para conocer el tipo de personas que se encuentran en cada conglomerado se deben obtener la media de los datos originales de quienes la conforman variable por variable como se muestra a continuación:

Para el conglomerado 1:

- La media de la altura es de: (1.52+1.54+1.56)/3=1.54 Mts.
- La media del peso es de: (57.12+67.62+49.35)/3=58.03 Kg.
- La media de la cintura es de: (84.10+85.22+83.60)/3=84.30 Cms.

Para el conglomerado 2:

- La media de la altura es de: (1.67+1.63)/2=1.65 Mts.
- La media del peso es de: (76.85+87.22)/2=82.03 Kg.
- La media de la cintura es de: (86.20+87.52)/2=86.86 Cms.

Los resultados del conglomerado 1 nos indican que las mujeres de este grupo son más bajitas, más delgadas y de menor cintura que las mujeres del conglomerados 2, por lo que un posible nombre para el conglomerado 1 podría ser "talla pequeña" y para el conglomerado 2 "talla grande"

2.2.1.2 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 2.2.1.2 Cluster Análisis. Medida de Similitud**

Con el fin de incrementar el confort de las diferentes piezas para vestir, los industriales de la confección desean desarrollar un sistema de patronaje y tallas flexibles, pero con medidas reales del cuerpo de los hombres mexicanos, experimentado de mejor manera con la elongación de músculos y la elasticidad de alguna fibras. Para ser más preciso en ciertos resultados, el estudio antropométrico sustituyó el indicador de cintura por el de la cadera contemplando igualmente 3 indicadores útiles para diseñar y fabricar piezas de vestir, la estatura, el peso y la cadera de cinco personas del género **MASCULINO** cuyos resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Cadera
1	1.73	63.56	Estrecha

2	1.85	83.99	Ancha
3	1.70	74.91	Ancha
4	1.63	54.48	Estrecha
5	1.93	95.34	Ancha
6	1.69	69.00	Estrecha
7	1.67	70.00	Ancha

Criterio de similitud para recodificar los datos:

Altura: existe(1) si es ≥ a 1.71 mts.
no existe(0) si es < a 1.71 mts.

Peso: existe(1) si es ≥ a 71 Kgs.
no existe(0) si es < a 71 Kgs.

Cadera: existe(1) si es Ancha
no existe (0) si es cualquier otro tipo

c) Obtenga la pareja de personas más parecidas y la menos parecida.
d) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

Resolución del inciso a)

Los datos recodificados quedaron como sigue:

Persona	Altura (mts.)	Peso (Kg.)	Cadera
1	1	0	0
2	1	1	1
3	0	1	1
4	0	0	0
5	1	1	1
6	0	0	0
7	0	0	1

Cálculo de las similitudes a encontrar:

El coeficiente de similitud que se usará para resolver este ejemplo es el de parejas simples. El coeficiente es:

$$S_{ij} = \frac{a+d}{p}$$

Se debe cuantificar la similitud entre las personas con base en este coeficiente, donde ***a*** es la presencia del atributo en ambas personas, ***d*** es la ausencia del atributo en ambas personas y ***p*** es el total de características

(o variables) que serán evaluadas.

Se empieza a comparar por pareja de personas, en este caso la persona 1 y la 2, como en ninguna de las características hay presencia del atributo en ambas (***a***) ni ausente en ambas (***d***), entonces tanto ***a*** como ***d*** tienen valor de 0 en el coeficiente, en cuanto a ***p*** tenemos que son tres características para evaluar (Altura, Peso y Busto); la similitud entre las personas 1 y 2 es de 0 como a continuación se presenta:

$$S_{1.2} = \frac{0+0}{3} = 0$$

De igual forma se procede con las personas 1 y 3, donde hay un atributo presente (***a***) en ambos en la característica de Busto y un atributo ausente (***d***) en ambos en la característica de Altura; por lo tanto, la similitud entre las personas 1 y 3 es de 2/3 como a continuación se presenta:

$$S_{1.3} = \frac{1+1}{3} = 2/3$$

Y así sucesivamente, se obtienen las medidas de similitud para todas las parejas que se pretende comparar obteniendo los siguientes resultados:

$S_{1.2} = \frac{1+0}{3} = 1/3$ $S_{2.3} = \frac{2+0}{3} = 2/3$ $S_{3.5} = \frac{2+0}{3} = 2/3$ $S_{5.6} = \frac{0+0}{3} = 0$

$S_{1.3} = \frac{0+0}{3} = 0$ $S_{2.4} = \frac{0+0}{3} = 0$ $S_{3.6} = \frac{1+0}{3} = 1/3$ $S_{5.7} = \frac{1+0}{3} = 1/3$

$S_{1.4} = \frac{0+2}{3} = 2/3$ $S_{2.5} = \frac{3+0}{3} = 3/3$ $S_{3.7} = \frac{1+1}{3} = 2/3$ $S_{6.7} = \frac{0+2}{3} = 2/3$

$S_{1.5} = \frac{1+0}{3} = 1/3$ $S_{2.6} = \frac{0+0}{3} = 0$ $S_{4.5} = \frac{0+0}{3} = 0$

$S_{1.6} = \frac{0+2}{3} = 2/3$ $S_{2.7} = \frac{1+0}{3} = 1/3$ $S_{4.6} = \frac{0+3}{3} = 3/3$

$S_{1.7} = \frac{0+1}{3} = 1/3$ $S_{3.4} = \frac{0+1}{3} = 1/3$ $S_{4.7} = \frac{0+2}{3} = 2/3$

Con la información anterior se construye una matriz cuadrada en este caso de 7 x 7 ya que tenemos 7 personas a la que le nombraremos ***matriz de similitud*** la cual representa que tan similares son las personas con respecto a la similitud entre ellas:

	1	2	3	4	5	6	7
1	1						
2	1/3	1					
3	0	2/3	1				
4	2/3	0	1/3	1			
5	1/3	1	2/3	0	1		
6	2/3	0	1/3	**1**	0	1	
7	1/3	1/3	2/3	2/3	1/3	2/3	1

Al observar la matriz de similitud se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya similitud entre sí es la que se acerque más o sea igual a 1, en este caso tenemos dos parejas, la primera formada por las personas 2 y 5, con una medida de similitud de 1; y la segunda formada por las personas 4 y 6, con una similitud también de 1.

La pareja de personas menos parecida será aquella cuya similitud entre sí es la se acerque más o sea igual a 0. En este caso tenemos cinco parejas, la primera formada por las personas 1 y 3, con una medida de similitud de 0; la segunda formada por las personas 2 y 4, con una similitud también de 0, la tercera formada por las personas 2 y 6, con una similitud de 0, la cuarta formada por las personas 4 y 5, con una similitud de 0 y la quinta formada por las personas 5 y 6 con una similitud de 0.

Resolución al inciso b)

Vinculación o eslabonamiento:

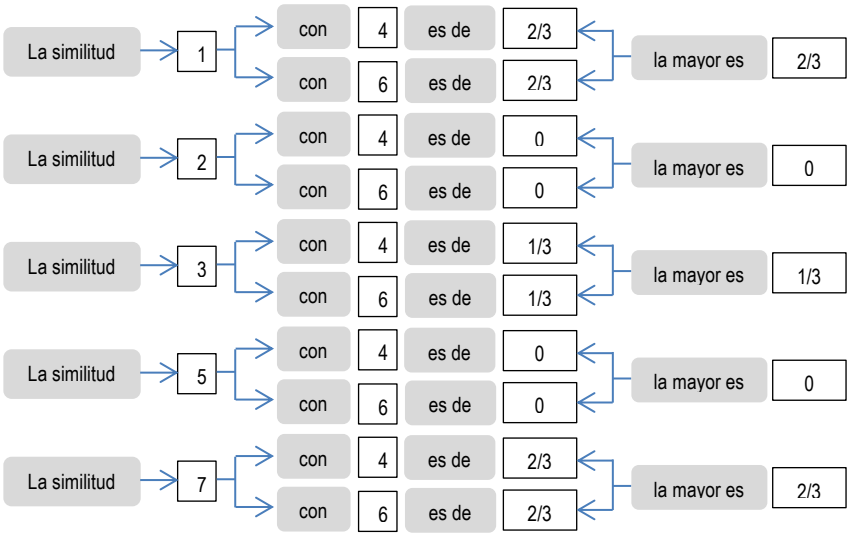
1. **Eslabonamiento Simple:**

Paso 1: Reducir la matriz con la similitud más grande, la más cercana o igual a 1

	1	2	3	4	5	6	7
1	1						
2	1/3	1					
3	0	2/3	1				
4	2/3	0	1/3	1			
5	1/3	1	2/3	0	1		
6	2/3	0	1/3	**1**	0	1	
7	1/3	1/3	2/3	2/3	1/3	2/3	1

Paso 2: Llenar la matriz con la similitud más grande, la más cercana o igual a 1

	(4,6)	1	2	3	5	7
(4,6)	1					
1	2/3	1				
2	0	1/3	1			
3	1/3	0	2/3	1		
5	0	1/3	1	2/3	1	
7	2/3	1/3	1/3	2/3	1/3	1

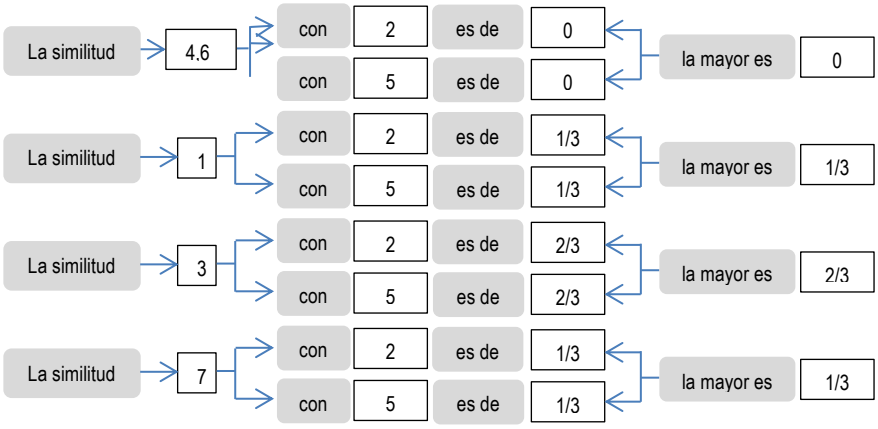


Paso 1: Reducir la matriz con la similitud más grande, la más cercana o igual a 1

	(4,6)	1	2	3	5	7
(4,6)	1					
1	2/3	1				
2	0	1/3	1			
3	1/3	0	2/3	1		
5	0	1/3	1	2/3	1	
7	2/3	1/3	1/3	2/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, la más cercana o igual a 1

	(4,6)	(2,5)	1	3	7
(4,6)	1				
(2,5)	0	1			
1	2/3	1/3	1		
3	1/3	2/3	0	1	
7	2/3	1/3	1/3	1/3	1

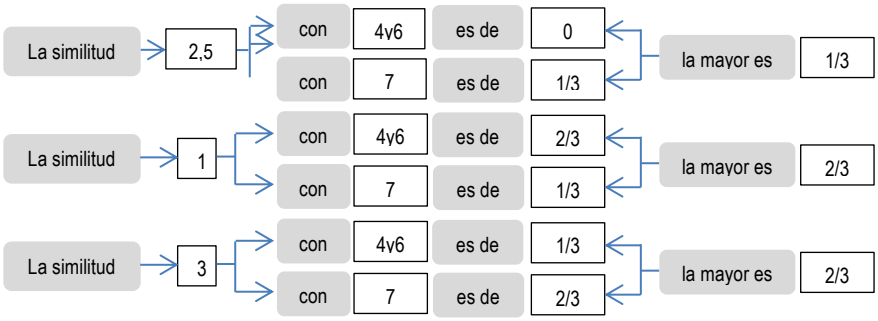


Paso 1: Reducir la matriz con la similitud más grande, lo más cercana o igual a 1

	(4,6)	(2,5)	1	3	7
(4,6)	1				
(2,5)	0	1			
1	2/3	1/3	1		
3	1/3	2/3	0	1	
7	2/3	1/3	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercana o igual a 1

	(4,6,7)	(2,5)	1	3
(4,6,7)	1			
(2,5)	1/3	1		
1	2/3	1/3	1	
3	2/3	2/3	0	1

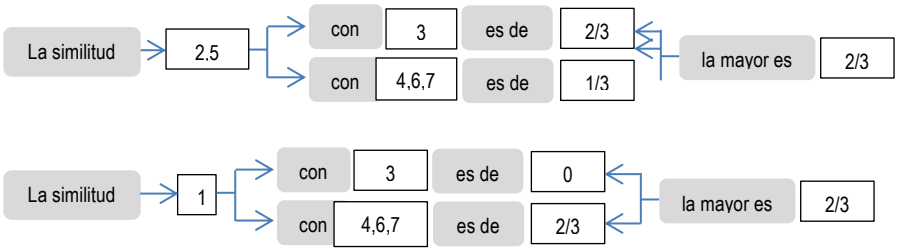


Paso 1: Reducir la matriz con la similitud más grande, lo mas cercana o igual a 1

	(4,6,7)	(2,5)	1	3
(4,6,7)	1			
(2,5)	1/3	1		
1	2/3	1/3	1	
3	2/3	2/3	0	1

Paso 2: Llenar la matriz con la similitud más grande, la más cercana o igual a 1

	(3,4,6,7)	(2,5)	1
(3,4,6,7)	1		
(2,5)	2/3	1	
1	2/3	1/3	1



Paso 1: Reducir la matriz con la similitud más grande, la más cercana o igual a 1

	(3,4,6,7)	(2,5)	1
(3,4,6,7)	1		
(2,5)	2/3	1	
1	2/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, la más cercana o igual a 1

	(1,3,4,6,7)	(2,5)
(1,3,4,6,7)	1	
(2,5)	2/3	1

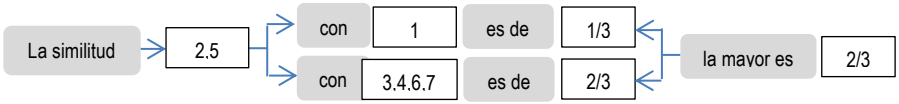


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	4	6	4
2	5	1	2	5	2
3	4	2/3	4	7	4
4	3	2/3	3	4	3
5	2	2/3	2	3	2
6	1	2/3	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 4 y la 6 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 formo otro conglomerado con las personas 2 y 5 con una similitud de 1 o 100. En la etapa 3 se incorporó la persona 7 al primer conglomerado con una similitud de 2/3 o 66.667%. En la etapa 4 se incorporó la persona 3 al primer conglomerado formado por las personas 4,6 y 7 como se muestra en la columna 3 a una similitud de 2/3 ó 66.667%. En la etapa 5 se incorporó la persona 1 al primer

conglomerado formado por las personas 3,4,6 y 7 como se muestra en la columna 3 a una similitud de 2/3 ó 66.667% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 2/3 ó 66.667%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

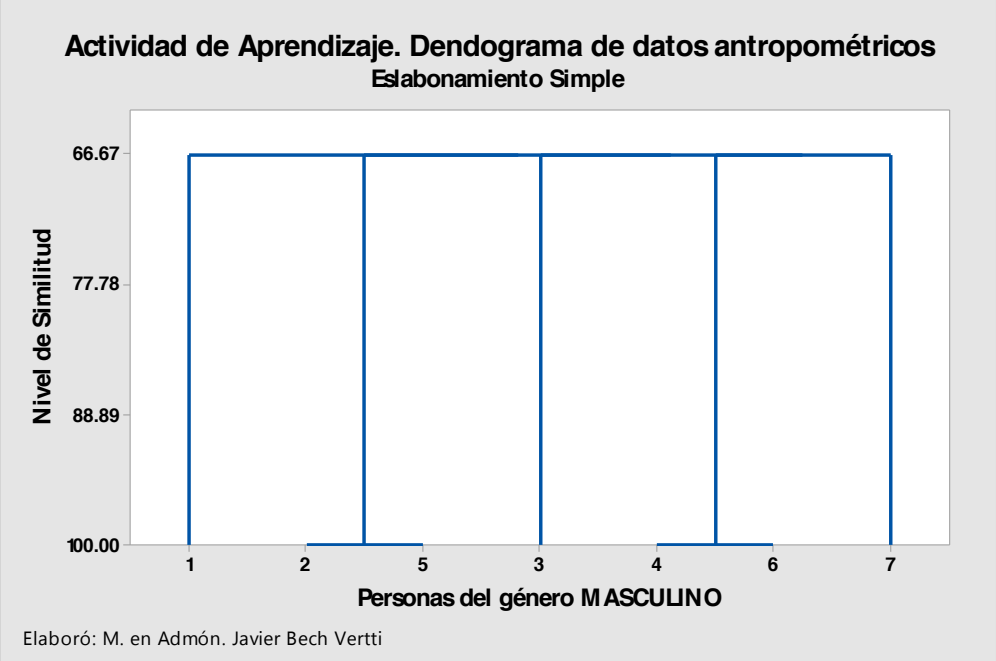
Análisis de observaciones de conglomerado: Altura, Peso, Cadera

```
Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100.000	0	4	6	4	2
2	5	100.000	0	2	5	2	2
3	4	66.667	1	4	7	4	3
4	3	66.667	1	3	4	3	4
5	2	66.667	1	2	3	2	6
6	1	66.667	1	1	2	1	7

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,4,6,7) y el conglomerado 2 por las personas (2,5).
Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Número de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

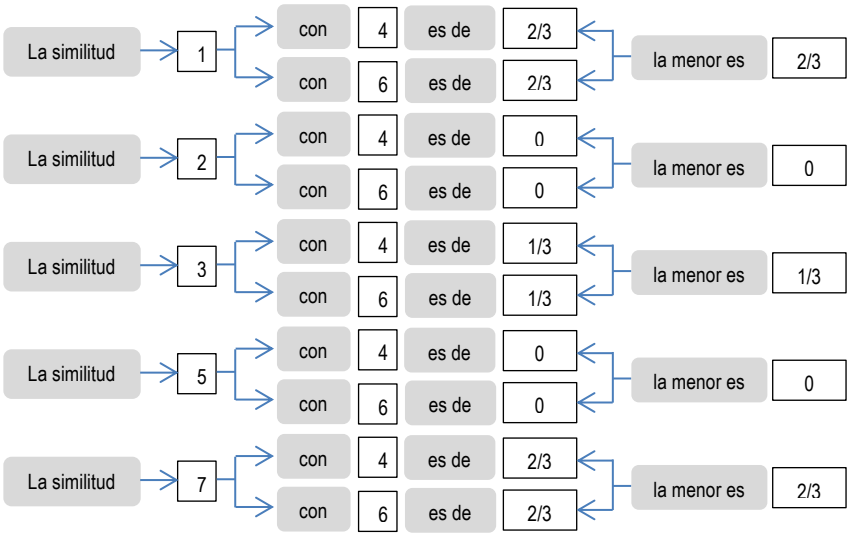
2. Eslabonamiento Completo:

Paso 1: Reducir la similitud (matriz) con la similitud más grande, la más cercana o igual a 1.

	1	2	3	4	5	6	7
1	1						
2	1/3	1					
3	0	2/3	1				
4	2/3	0	1/3	1			
5	1/3	1	2/3	0	1		
6	2/3	0	1/3	**1**	0	1	
7	1/3	1/3	2/3	2/3	1/3	2/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercana o igual a 0.

	(4,6)	1	2	3	5	7
(4,6)	1					
1	2/3	1				
2	0	1/3	1			
3	1/3	0	2/3	1		
5	0	1/3	1	2/3	1	
7	2/3	1/3	1/3	2/3	1/3	1

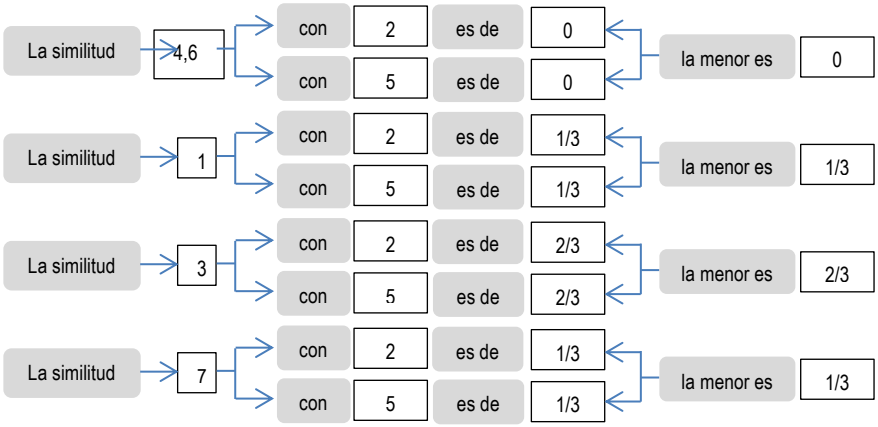


Paso 1: Reducir la similitud (matriz) con la similitud más grande, la más cercana o igual a 1.

	(4,6)	1	2	3	5	7
(4,6)	1					
1	2/3	1				
2	0	1/3	1			
3	1/3	0	2/3	1		
5	0	1/3	1	2/3	1	
7	2/3	1/3	1/3	2/3	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercana o igual a 0.

	(4,6)	(2,5)	1	3	7
(4,6)	1				
(2,5)	0	1			
1	2/3	1/3	1		
3	1/3	2/3	0	1	
7	2/3	2/3	1/3	2/3	1

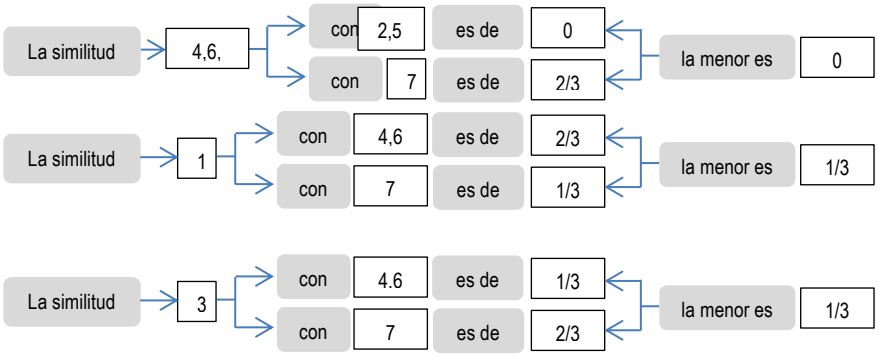


Paso 1: Reducir la similitud (matriz) con la similitud más grande, la más cercana o igual a 1.

	(4,6)	(2,5)	1	3	7
(4,6)	1				
(2,5)	0	1			
1	2/3	1/3	1		
3	1/3	2/3	0	1	
7	2/3	2/3	1/3	2/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercana o igual a 0.

	(4,6,7)	(2,5)	1	3
(4,6,7)	1			
(2,5)	0	1		
1	1/3	1/3	1	
3	1/3	2/3	0	1

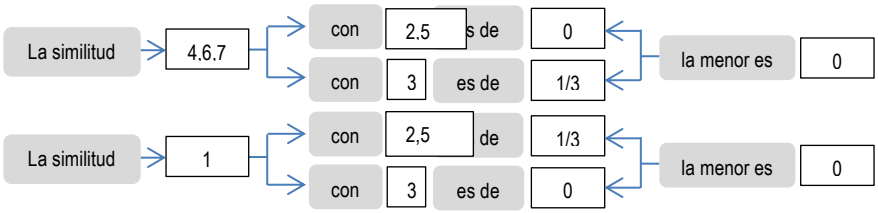


Paso 1: Reducir la similitud (matriz) con la similitud más grande, la más cercana o igual a 1

	(4,6,7)	(2,5)	1	3
(4,6,7)	1			
(2,5)	0	1		
1	1/3	1/3	1	
3	1/3	2/3	0	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercana o igual a 0.

	(4,6,7)	(2,3,5)	1
(4,6,7)	1		
(2,3,5)	0	1	
1	1/3	0	1



Paso 1: Reducir la similitud (matriz) con la similitud más grande, la más cercana o igual a 1

	(4,6,7)	(2,3,5)	1
(4,6,7)	1		
(2,3,5)	0	1	
1	1/3	0	1

Paso 2: Llenar la matriz con la similitud más pequeña, la mas cercana o igual a 0.

	(1,4,6,7)	(2,3,5)
(1,4,6,7)	1	
(2,3,5)	0	1

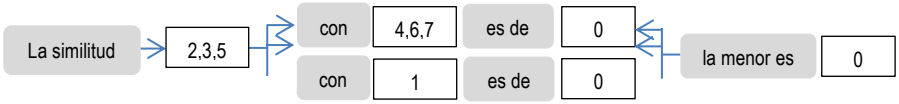


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	4	6	4
2	5	1	2	5	2
3	4	2/3	4	7	4
4	3	2/3	2	3	2
5	2	1/3	1	4	1
6	1	0	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 4 y la 6 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 formo otro conglomerado con las personas 2 y 5 con una similitud de

1 o 100%. En la etapa 3 se incorporó la persona 7 al primer conglomerado con una similitud de 2/3 o 66.667%. En la etapa 4 se incorporó la persona 3 al segundo conglomerado formado por las personas 2 y 5 como se muestra en la columna 3 a una similitud de 2/3 ó 66.667%. En la etapa 5 se incorporó la persona 1 al primer conglomerado formado por las personas 4,6 y 7 como se muestra en la columna 3 a una similitud de 1/3 ó 33.333% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 0 ó 0.000%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

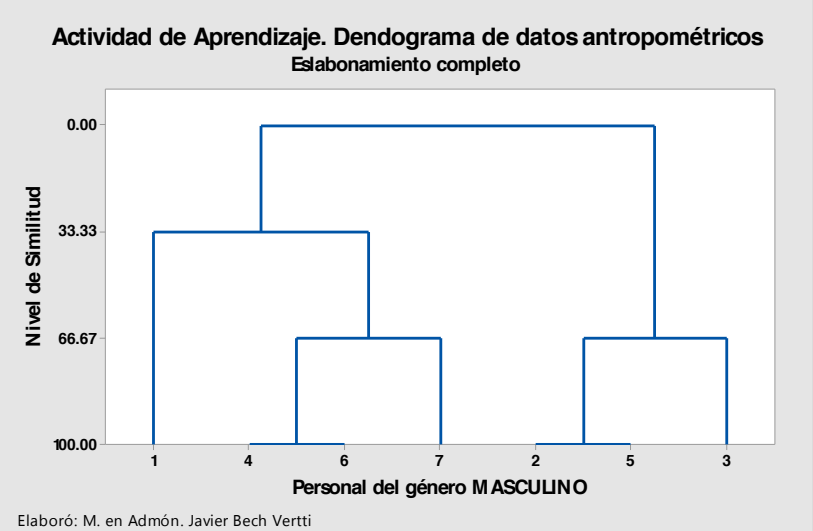
Análisis de observaciones de conglomerado: Altura, Peso, Cadera

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100.000	0	4	6	4	2
2	5	100.000	0	2	5	2	2
3	4	66.667	1	4	7	4	3
4	3	66.667	1	2	3	2	3
5	2	33.333	2	1	4	1	4
6	1	0.000	3	1	2	1	7

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en este caso el conglomerado 1 está formado por las personas (1,4,6,7) y el conglomerado 2 por las personas (2,3,5). Minitab puede dibujar los conglomerados en diferente orden, primero el 1 o el 2 o bien primero el 2 y luego el 1, lo cual no cambia la solución.
Cabe mencionar que mientras a menor similitud se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



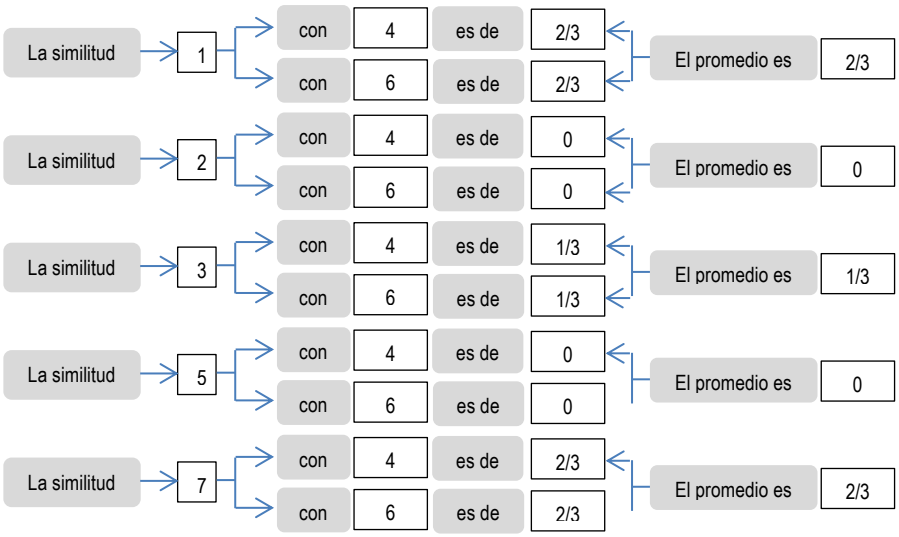
3. Eslabonamiento Promedio:

Paso 1. Reducir la matriz con la similitud más grande, la más cercana o igual a 1

	1	2	3	4	5	6	7
1	1						
2	1/3	1					
3	0	2/3	1				
4	2/3	0	1/3	1			
5	1/3	1	2/3	0	1		
6	2/3	0	1/3	**1**	0	1	
7	1/3	1/3	2/3	2/3	1/3	2/3	1

Paso 2. Llenar la matriz con la similitud promedio.

	(4,6)	1	2	3	5	7
(4,6)	1					
1	2/3	1				
2	0	1/3	1			
3	1/3	0	2/3	1		
5	0	1/3	1	2/3	1	
7	2/3	1/3	1/3	2/3	1/3	1

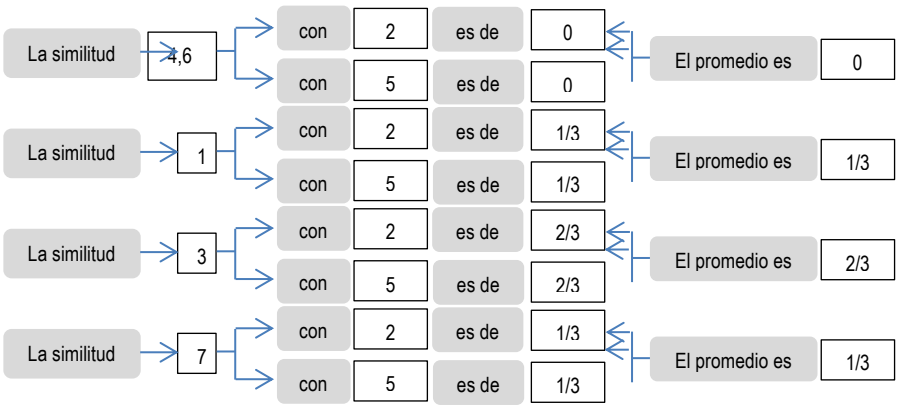


Paso 1. Reducir la matriz con la similitud más grande, la más cercano o igual a 1

	(4,6)	1	2	3	5	7
(4,6)	1					
1	2/3	1				
2	0	1/3	1			
3	1/3	0	2/3	1		
5	0	1/3	1	2/3	1	
7	2/3	1/3	1/3	2/3	1/3	1

Paso 2. Llenar la matriz con la similitud promedio.

	(4,6)	(2,5)	1	3	7
(4,6)	1				
(2,5)	0	1			
1	2/3	1/3	1		
3	1/3	2/3	0	1	
7	2/3	1/3	1/3	2/3	1

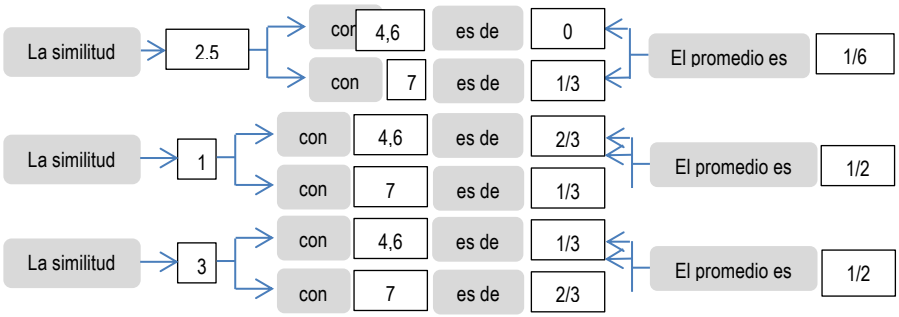


Paso 1. Reducir la matriz con la similitud más grande, la más cercana o igual a 1

	(4,6)	(2,5)	1	3	7
(4,6)	1				
(2,5)	0	1			
1	2/3	1/3	1		
3	1/3	2/3	0	1	
7	2/3	1/3	1/3	2/3	1

Paso 2. Llenar la matriz con la similitud promedio.

	(4,6,7)	(2,5)	1	3
(4,6,7)	1			
(2,5)	1/6	1		
1	1/2	1/3	1	
3	1/2	2/3	0	1

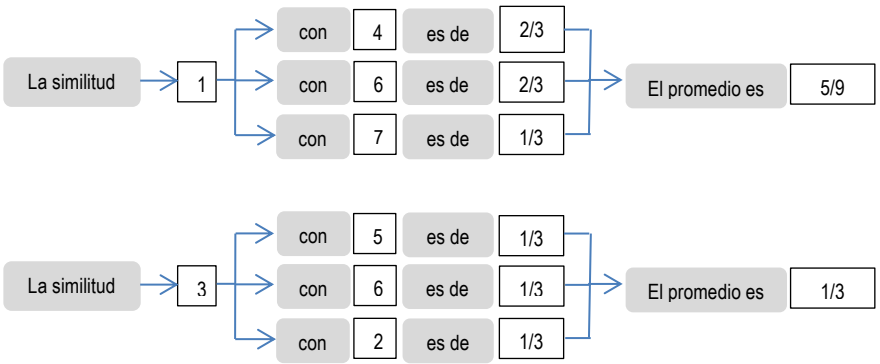


Paso 1. Reducir la matriz con la similitud más grande, lo mas cercano o igual a 1

	(4,6,7)	(2,5)	1	3
(4,6,7)	1			
(2,5)	1/6	1		
1	1/2	1/3	1	
3	1/2	2/3	0	1

Paso 2. Llenar la matriz con la similitud promedio.

	(4,6,7)	(2,3,5)	1
(4,6,7)	1		
(2,3,5)	1/2	1	
1	5/9	2/9	1



Paso 1. Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(4,6,7)	(2,3,5)	1
(4,6,7)	1		
(2,3,5)	1/2	1	
1	5/9	2/9	1

Paso 2. Llenar la matriz con la similitud promedio.

	(1,4,6,7)	(2,3,5)
(1,4,6,7)	1	
(2,3,5)	2/9	1

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la similitud promedio, es decir se compara al conglomerado (1,4,6,7) con el resto de las personas que forman el otro conglomerado (1,7,4,3); al comparar (1,4,6,7) con las personas (1,7,4,3) se debe verificar **en la matriz original las similitudes entre todas las combinaciones**, es decir a las personas 1 y 2 que es 1/3, a la persona 1 y 3 que es 0, a la persona 1 y 5 que es 1/3 y así sucesivamente hasta la persona 7 y 5 que es 1/3, eligiendo entonces la similitud promedio de las doce combinaciones o siete personas que en este caso es 2/9 o 22.222. Lo anterior se puede resumir con el siguiente diagrama de flujo:

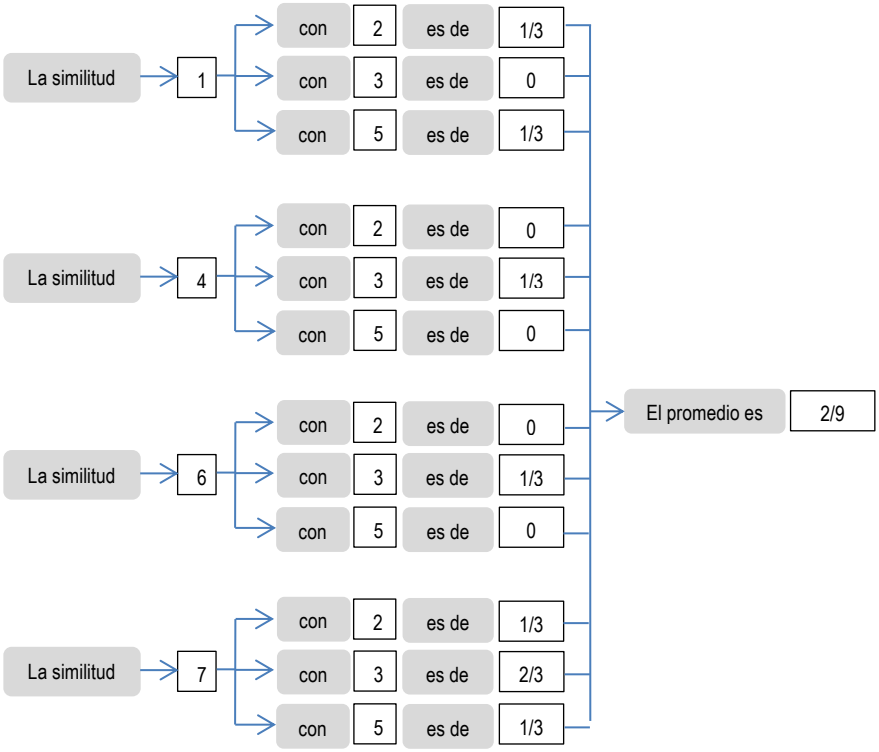


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	4	6	4
2	5	1	2	5	2
3	4	2/3	4	7	4
4	3	2/3	2	3	2
5	2	5/9	1	4	1
6	1	2/9	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos una nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 4 y la 6 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se formó otro conglomerado con las personas 2 y 5 con una similitud de 1 o 100%. En la etapa 3 se incorporó la persona 7 al primer conglomerado con una similitud de 2/3 o 66.667%. En la etapa 4 se incorporó la persona 3 al segundo conglomerado formado por las personas 2 y 5 como se muestra en la columna 3 a una similitud de 2/3 ó 66.667%. En la etapa 5 se incorporó la persona 1 al primer conglomerado formado por las personas 4,6 y 7 como se muestra en la columna 3 a una similitud de 5/9 ó 55.556% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 2/9 ó 22.222%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Altura, Peso, Cadera

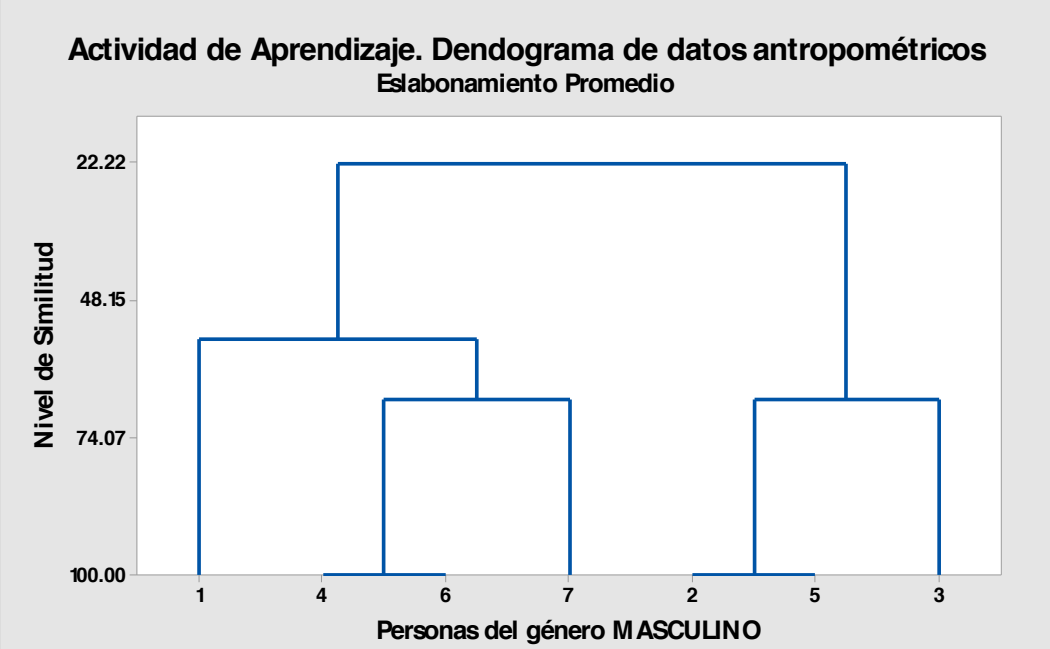
```
Distancia eucledíana cuadrada, Enlace de promedios
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados	Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100.000	0.00000	4	6	4	2
2	5	100.000	0.00000	2	5	2	2
3	4	66.667	1.00000	4	7	4	3
4	3	66.667	1.00000	2	3	2	3
5	2	55.556	1.33333	1	4	1	4
6	1	22.222	2.33333	1	2	1	7

Dendrograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,4,6,7) y el conglomerado 2 por las personas (2,3,5). Minitab puede dibujar los conglomerados en diferente orden, primero el 1 o el 2 o bien primero el 2 y luego el 1, lo cual no cambia la solución.

Cabe mencionar que mientras a menor similitud se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Elaboró: M. en Admón. Javier Bech Vertti

Respuesta al inciso c)

Nombre de los conglomerados

Una vez elegidos el número de conglomerados definitivo, para finalizar el proceso es deseable colocar un nombre apropiado a los conglomerados. En este caso el conglomerado 1 está formado por las personas 1,4,6 y 7 y el conglomerado 2 está formado por las personas 2,3 y 5. Para conocer el tipo de personas que se encuentran en cada conglomerado se deben obtener la media de los datos originales de quienes conforman variable por variable métrica y la moda de los datos originales de quienes conforman variable por variable no métrica como se muestra a continuación:

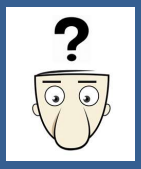
Para el conglomerado 1:

- La media de la altura es de: (1.73+1.63+1.69+1.67)/4=1.68 Mts.
- La media del peso es de: (63.56+54.48+69.00+70.00)/4=64.26 Kg.
- El tipo de cadera es de: (Estrecha;Estrecha;Estrecha;Ancha)=Estrecha

Para el conglomerado 2:

- La media de la altura es de: (1.85+1.70+1.93)/3=1.83 Mts.
- La media del peso es de: (83.99+74.91+95.34)/3=84.75 Kg.
- La tipo de cadera es de: (Ancha;Ancha;Ancha)=Ancha.

Los resultados del conglomerado 1 nos indican que los hombres de este grupo son menos altos, menos pesados y de menor cadera que los hombres del conglomerado 2, por lo que un posible nombre para el conglomerado 1 podría ser "talla XXS" y para el conglomerado 2 "talla S".



RESPUESTAS A LOS EJERCICIOS DE AUTOEVALUACIÓN DEL CAPÍTULO 2

2.1.1.1 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 2.1.1.1 Factor Análisis

El Director de mercadotecnia de una compañía de Bienes Raíces desea conocer cuales dimensiones o indicadores básicos influyen al momento de pronosticar el valor de una casa o vivienda unifamiliar al momento en que un potencial cliente desea ya sea comprar o vender su casa con el objeto de brindar un servicio confiable y rápido. Para llevar a cabo la investigación dividió la Ciudad en tres grandes zonas: NORTE, SUR Y ESTE. En cada zona se aplicó una encuesta a 20 propietarios de casas unifamiliares. El cuestionario estaba conformado por 20 variables que se consideraba podían formar parte de dichos indicadores X_1= Tamaño del lote y X_2= Número de recámaras, , X_3= Número de baños, X_4= Número de habitaciones y X_5= Antigüedad de la casa, X_6= Impuestos anuales, X_7= Tipo de estacionamiento interior, X_8= Estilo arquitectónico, , X_9= Tipo de sistema para la calefaccción, X_{10}= Tipo de alberca, X_{11}= Cocina con antecomedor, X_{12}= Chimenea, X_{13}= Cocina integral, X_{14}= Tina de hidromasaje.

Los resultados en la zona ESTE fueron los siguientes:

Casa No.	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	6.9	4	2.0	8	38	3750	1
2	6.0	3	2.0	6	35	3240	0
3	7.0	3	1.0	6	39	2700	1
4	6.0	4	1.0	7	38	2800	1
5	6.0	3	2.0	7	38	2995	1
6	9.0	4	1.0	6	32	2500	1
7	11.2	4	1.0	9	32	2800	1
8	6.0	5	2.0	10	35	2500	0
9	5.6	3	1.0	6	37	2027	0
10	6.0	3	1.5	8	37	2500	1
11	6.0	4	1.0	6	39	2200	1
12	6.0	4	2.0	7	37	3156	0
13	8.7	4	1.0	6	36	2741	1
14	6.0	4	1.0	6	39	2100	0
15	6.5	3	2.0	7	36	2400	0
16	6.5	3	1.0	6	35	2280	0
17	6.0	3	2.0	8	38	2700	1
18	6.0	3	1.5	7	35	2300	1
19	6.0	3	1.0	7	36	1889	0
20	6.0	4	1.0	6	35	2900	0

Casa No.	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	3	2	1	1	0	0	0
2	1	2	1	1	1	1	1
3	1	1	1	1	0	0	0
4	1	2	1	1	0	1	1
5	4	2	1	1	1	1	1
6	4	2	1	1	1	0	0
7	1	2	1	1	0	0	0
8	1	2	1	1	0	0	0
9	1	2	1	1	1	1	1
10	2	2	1	1	1	1	1
11	1	2	1	1	0	1	1
12	1	2	1	1	1	1	1
13	1	2	1	1	1	0	0
14	1	2	1	1	0	0	0
15	1	2	2	1	0	1	1
16	1	2	1	0	0	0	1
17	1	2	1	1	0	1	1
18	4	2	1	1	1	1	1
19	1	2	1	1	0	1	1
20	4	2	1	1	1	1	1

a) ¿Cuántas dimensiones o indicadores esperaría tener como solución inicial en esta investigación?. Explique su respuesta tomando en cuenta los criterios para analizar el número de factores en la solución inicial.
b) Teniendo como supuesto que la solución inicial es correcta, ¿Qué variables se identifican más con cada factor?. Explique su respuesta tomando en cuenta los criterios para identificar cargas significativas y sus prioridades.
c) Según la información de la matriz rotada, ¿con cuál de los factores se identificó mejor el conjunto de 14 variables?. Justifique.
d) ¿Cuál de las variables ha sido mejor explicada en esta investigación por la solución de factores?. Justifique.

Respuesta al inciso a.

Se supone que ya se realizaron todos los pasos necesarios hasta antes de generar la matriz de correlación, la cual es la siguiente:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_1	1.000						
X_2	0.277	1.000					
X_3	-0.315	-0.073	1.000				
X_4	0.171	0.346	0.473	1.000			
X_5	-0.617	-0.163	0.118	-0.193	1.000		
X_6	0.159	0.208	0.525	0.201	0.022	1.000	
X_7	0.402	**-0.009**	-0.087	0.139	0.651	0.256	1.000
X_8	0.023	-0.017	0.088	-0.083	-0.250	0.253	0.309
X_9	-0.056	0.214	0.198	0.195	-0.303	-0.040	-0.208
X_{10}	-0.029	-0.214	0.297	0.010	-0.040	-0.116	-0.254
X_{11}	0.029	0.214	0.198	0.195	0.154	0.179	0.254
X_{12}	-0.054	-0.162	0.087	-0.319	-0.258	0.169	0.010
X_{13}	-0.597	-0.450	0.264	-0.128	0.244	-0.088	-0.123
X_{14}	-0.627	-0.560	0.181	-0.021	0.180	-0.172	-0.242

	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
X_8	1.000						
X_9	0.141	1.000					
X_{10}	-0.141	0.053	1.000				
X_{11}	0.141	-0.053	0.053	1.000			
X_{12}	0.515	0.208	-0.208	0.208	1.000		

X_{13}	0.084	0.281	0.187	0.281	0.328	1.000	
X_{14}	0.021	0.310	0.168	-0.168	0.242	**0.899**	1.000

A partir de la matriz de correlación, se puede identificar que las variables con mayor relación son la 6 con la 13 es decir X_6= Impuestos anuales y X_{13}= Cocina integral, con una correlación de **0.899**, mientras las que tienen menor relación son la 2 con la 7 o sea X_2= Número de recámaras con X_7= Tipo de estacionamiento interior, con una correlación de **0.009**. Sin embargo, a partir de la matriz de correlación, es muy difícil encontrar a las variables afines que se puedan agrupar en factores y resumir así la información. Se necesita verificar entonces la matriz de factores no rotada que proporcione mayor información por lo que se corrió un análisis de factores y se obtuvo la siguiente matriz:

Cargas de factores no rotados :

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
X_1	-0.801	-0.065	-0.266	0.184	0.368	-0.157	-0.093
X_2	-0.616	-0.187	0.234	0.188	-0.552	-0.130	0.212
X_3	0.276	-0.609	0.591	0.029	0.178	0.300	0.073
X_4	-0.323	-0.384	0.627	0.211	-0.027	-0.078	-0.399
X_5	0.378	0.216	0.352	-0.715	-0.260	0.033	0.016
X_6	-0.213	-0.651	0.184	-0.256	0.080	0.478	0.146
X_7	-0.365	-0.311	-0.195	-0.556	0.206	-0.166	-0.521
X_8	0.001	-0.611	-0.522	-0.054	0.069	0.101	0.005
X_9	0.207	-0.367	0.005	0.690	-0.312	-0.104	-0.211
X_{10}	0.262	0.077	0.388	0.277	0.659	-0.235	0.295
X_{11}	-0.036	-0.509	0.162	-0.361	-0.066	-0.682	0.276
X_{12}	0.269	-0.537	-0.636	0.050	-0.55	0.019	0.278
X_{13}	0.877	-0.268	-0.011	-0.031	-0.020	-0.284	-0.133
X_{14}	0.918	-0.043	-0.085	0.133	0.010	0.019	-0.262
Eigenvalue	3.3207	2.2600	1.9040	1.7069	1.1344	1.019	0.8830
% Var	0.237	0.161	0.136	0.122	0.081	0.073	0.063

Variable	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
X_1	-0.0194	-0.204	-0.003	0.048	0.040	-0.130	0.000
X_2	0.157	-0.154	0.173	-0.216	-0.053	-0.033	0.000
X_3	-0.026	0.112	-0.185	-0.022	-0.156	-0.065	0.000
X_4	-0.083	0.331	0.039	-0.061	0.149	0.011	0.000
X_5	0.176	-0.160	-0.168	-0.005	0.129	-0.074	0.000
X_6	-0.151	-0.322	0.186	0.078	0.067	0.052	0.000
X_7	0.115	-0.176	-0.097	-0.143	-0.076	0.050	0.000
X_8	0.514	0.217	0.127	0.083	0.031	-0.041	0.000
X_9	0.150	-0.311	-0.235	0.133	0.022	0.025	0.000
X_{10}	0.280	-0.159	0.001	-0.118	0.060	0.030	0.000
X_{11}	-0.112	0.062	0.003	0.146	-0.024	0.018	0.000
X_{12}	-0.227	0.073	-0.212	-0.205	0.089	0.005	0.000
X_{13}	-0.126	-0.77	0.188	-0.009	-0.015	-0.028	0.000
X_{14}	-0.078	-0.107	0.192	-0.076	-0.004	-0.037	0.000
Eigenvalue	0.5868	0.5434	0.3260	0.1875	0.0915	0.0389	0.0000
% Var	0.042	0.039	0.023	0.013	0.007	0.003	0.000

Esta matriz no rotada muestra las cargas de factor que, como ya se mencionó, indican la relación de cada variable con cada factor .

Aquí la mejor solución en cuanto al número de factores sería un factor que agrupara a las catorce variables. Todas tendrían algo en común entre sí, mientras que la peor solución serian catorce factores, lo cual representaría que las variables no tuvieron nada en común y no se podría resumir nada la información.

Tanto por el criterio de la raíz latente como por el porcentaje de variación explicada acumulada, se tiene como solución inicial siete factores, ya que el *Eigenvalue* de los seis primeros es mayor a 1 y el porcentaje de variación explicada acumulada está entre 60 % y 95%, es de 81.0%

En este problema de catorce variables la varianza ó variación máxima que se podría obtener, se tiene que F1 explica 3.3207, el F2 explica 2.2600, el F3 explica 1.9040, el F4 explica 1.7069, el F5 explica 1.1344 y el F6

explica 1.019 de acuerdo con su *Eigenvalue*. El total de la variación es de 11.3429 (total de los *Eigenvalue* de los 6 factores y se pierde 2.6571 (14 – 11.3429) de información.

Respuesta al inciso b.

Una vez se ha determinado el número de factores para la solución inicial se trata de agrupar a las variables en esos seis factores, lo cual se determina analizando la matriz de factores rotada, que es la siguiente:
Cargas de factores rotados y comunalidades

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Comunalidad
X_1	**-0.916**	-0.141	-0.083	-0.013	0.129	-0.094	0.892
X_2	-0.439	0.244	0.143	0.412	**-0.553**	-0.276	0.826
X_3	0.279	-0.002	**0.873**	0.134	0.236	-0.068	0.918
X_4	-0.268	0.309	**0.565**	0.343	0.034	-0.301	0.696
X_5	**0.610**	0.382	0.083	**-0.520**	-0.255	-0.180	0.893
X_6	-0.135	-0.280	**0.801**	-0.182	-0.180	-0.024	0.804
X_7	-0.311	-0.300	0.101	**-0.501**	-0.086	-0.437	0.646
X_8	-0.066	-**0.788**	0.154	0.000	-0.082	-0.092	0.664
X_9	0.135	-0.187	0.077	**0.839**	-0.007	-0.011	0.762
X_{10}	0.047	0.239	0.095	0.088	**0.844**	-0.044	0.791
X_{11}	0.056	-0.095	0.106	-0.024	0.002	**-0.928**	0.885
X_{12}	0.182	**-0.841**	-0.040	0.142	-0.074	-0.056	0.771
X_{13}	**0.821**	-0.255	-0.034	0.164	0.295	-0.266	**0.924**
X_{14}	**0.817**	-0.218	-0.083	0.180	0.303	0.151	0.870
Eigenvalue	3.0716	2.0314	1.8280	1.6507	1.3886	1.3728	11.3429
% Var	0.219	0.145	0.131	0.188	0.099	0.098	0.810

Ni el *Eigenvalue* ni la comunalidad total, ni el porcentaje de variación explicada total cambian al rotar los factores, es decir, no cambian de la matriz no rotada a la rotada, **solo se distribuyeron las cargas de factor** y ahora sí algunas cargas tienen valores más grandes que en la no rotada como para poder determinar la forma como se agrupan las variables en cada factor.
Lo que si cambia es la importancia en que cada factor explica a las variables (*Eigenvalue* de cada factor) y en la matriz rotada es posible que cada factor hacia la derecha explique menos a las variables.
Para determinar en cual factor se agrupara cada variable debe analizarse la matriz rotada, buscando las cargas significativas. Al identificar las cargas significativas, no importa el signo que tengan, sino la magnitud de la relación, el signo negativo solo expresa una relación inversa. En principio se buscan cargas mayores a 0.50 utilizando la Prioridad 1, en caso de que una variable no tuviera una carga mayor a 0.5 se aplicarían las prioridades 2, 3 o 4, según sea el caso, de los criterios de cargas significativas en toda la matriz rotada. En este ejemplo se tiene:
X_1= Tamaño del lote se identifica con el F1 con una carga de 0.916
X_2= Número de recámaras se identifica con el F5 con una carga de 0.553
X_3= Número de baños se identifica con el F3 con una carga de 0.873
X_4= Número de habitaciones se identifica con el F3 con una carga de 0.565
X_5= Antigüedad de la casa se identifica con el **F1 y el F4** con unas cargas de **0.610 y 0.520**
X_6= Impuestos anuales se identifica con el F3 con una carga de 0.801
X_7= Tipo de estacionamiento interior se identifica con el F4 con una carga de 0.501
X_8= Estilo arquitectónico se identifica con el F2 con una carga de 0.788
X_9= Tipo de sistema para la calefacción se identifica con el F4 con una carga de 0.839
X_{10}= Tipo de alberca se identifica con el F5 con una carga de 0.844
X_{11}= Cocina con antecomedor se identifica con el F6 con una carga de 0.928
X_{12}= Chimenea se identifica con el F2 con una carga de 0.841
X_{13}= Cocina Integral se identifica con el F1 con una carga de 0.821
X_{14}= Tina de hidromasaje se identifica con el F1 con una carga de 0.817

En caso de que alguna variable tuviera carga significativa con dos o más factores, se debe analizar en la matriz de correlación con cuales variables tiene mayor relación. En este caso **X_5= Antigüedad de la casa** se identifica con el **F1 y el F4** con unas cargas de 0.610 y 0.520; entonces se debe checar la correlación con

todas las combinaciones que se pueden hacer con la variable X_5= Antigüedad con X_1= Tamaño del lote, X_{13}= Cocina Integral, X_{14}= Tina de Hidromasaje que forman el **Factor 1** y con X_7= Tipo de estacionamiento interior y X_9= Tipo de sistema para la calefacción que forman el **Factor 4**.

X_5= Antigüedad VS X_1= Tamaño del lote tienen una correlación de 0.617
X_5= Antigüedad VS X_7= Tipo de estacionamiento interior tienen una correlación de **0.651**
X_5= Antigüedad VS X_9= Tipo de combustible tienen una correlación de 0.303
X_5= Antigüedad VS X_{13}= Aire acondicionado tienen una correlación de 0.244
X_5= Antigüedad VS X_{14}= Chimenea tienen una correlación de 0.180

Como la máxima correlación que se obtuvo es con X_7= Tipo de estacionamiento interior, entonces la variable X_5= Antigüedad debe ir en el Factor 4.
La solución final óptima sería:

F1: X_1= Tamaño del lote, X_{13}= Cocina Integral y X_{14}= Tina de hidromasaje
F2: X_8= Estilo arquitectónico y X_{12}= Chimenea
F3: X_3= Número de baños, X_4= Número de habitaciones y X_6= Impuestos anuales
F4: X_5= Antigüedad, X_7= Tipo de estacionamiento interior y X_9= Tipo de sistema para la calefacción
F5: X_2= Número de recámaras y X_{10}= Tipo de alberca
F6: X_{11}= Cocina con antecomedor

Respuesta al inciso c.

El porcentaje de varianza, que es el porcentaje de variación explicada con respecto a la máxima variación que se podría tener, se obtiene, por ejemplo para F1 dividiendo 3.0716/14, lo que es igual a 0.219 o 21.9 %. El Factor que ha explicado más al conjunto de variables es el 1, y después el 2 y así sucesivamente lo cual es lógico pues cada factor adicional explica menor cantidad, ya que cada vez queda menos por explicar, en este caso, por ejemplo, queda por explicar 2.6571 de variación; si se agregan los 8 factores que faltan, ya cada uno explicaría muy poco y no convendría. El porcentaje de variación explicada acumulada total es de 0.810 o de 81.00 % de un máximo por explicar de 100%.

Respuesta al inciso d.

Se puede observar que la variable mejor explicada por el conjunto de factores es la variable X_{13}= Cocina Integral, ya que es la que tiene la mayor comunalidad de 0.924

2.2.1.1 EJERCICIO DE **AUTOEVALUACIÓN**

EJERCICIO DE **AUTOEVALUACIÓN 2.2.1.1 Cluster Análisis. Medida de Distancia**

Suponga que se desea segmentar a los clientes de un supermercado con base a la actitud que tienen cuando salen de compras. De acuerdo a una investigación previa, se identificaron tres variables de actitud y se pidió a los cinco entrevistados que expresaran su grado de acuerdo con tres afirmaciones en una escala de siete puntos en la cual 1 significa en desacuerdo y 7 significa de acuerdo.

Las afirmaciones podrían ser las siguientes:

V1 Salir de compras es divertido
V2 Puedo ahorrar mucho dinero si comparo precios
V3 Prefiero salir de compras con mi familia

Si los resultados para cada una de las afirmaciones se considera que son variables, se tienen tres, las cuales se representarán con V1, V2 y V3; por otra parte, las respuestas de cada uno de los entrevistados constituyen los cinco casos, que se representarán con C1, C2,…,C5. Los valores de las variables para los cinco casos, se presentan a continuación:

ENTREVISTADO	**VARIABLE**		
	V1	**V2**	**V3**
C1	1	2	1
C2	5	7	5
C3	2	1	1
C4	7	7	5
C5	3	2	2
Promedio:	3.6	3.8	2.8
Desviación estándar	2.41	2.95	2.049

d) Obtenga la pareja de personas entrevistadas más parecidas y la menos parecida.
e) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.
f) Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

Resolución del inciso a)

Estandarización de variables

Antes que nada hay que se deben estandarizar las variables, es decir, eliminar el efecto de las unidades y hacerlas comparables, de manera que todas estén en términos de "Z" mediante la siguiente expresión:

$$Z = \frac{x - \bar{x}}{s}$$

$$Z_{11} = \frac{1 - 3.6}{2.41} = -1.08 \qquad Z_{12} = \frac{2 - 3.8}{2.95} = -0.61 \qquad Z_{13} = \frac{1 - 2.8}{2.049} = -0.88$$

$$Z_{21} = \frac{5 - 3.6}{2.41} = 0.58 \qquad Z_{22} = \frac{7 - 3.8}{2.95} = 1.08 \qquad Z_{23} = \frac{5 - 2.8}{2.049} = 1.07$$

$$Z_{31} = \frac{2 - 3.6}{2.41} = -0.66 \qquad Z_{32} = \frac{1 - 3.8}{2.95} = -0.95 \qquad Z_{33} = \frac{1 - 2.8}{2.049} = -0.88$$

$$Z_{41} = \frac{7 - 3.6}{2.41} = 1.41 \qquad Z_{42} = \frac{7 - 3.8}{2.95} = 1.08 \qquad Z_{43} = \frac{5 - 2.8}{2.049} = 1.07$$

$$Z_{51} = \frac{3 - 3.6}{2.41} = -0.25 \qquad Z_{52} = \frac{2 - 3.8}{2.95} = -0.61 \qquad Z_{53} = \frac{2 - 2.8}{2.049} = -0.39$$

Los datos estandarizados de las 5 personas, al considerar 3 variables, se presenta de la siguiente manera:

Persona	V1	V2	V3
1	-1.08	-0.61	-0.88
2	0.58	1.08	1.07
3	-0.66	-0.95	-0.88
4	1.41	1.08	1.07
5	-0.25	-0.61	-0.39
Promedio:	0.00	0.00	0.00
Desviación estándar	1.00	1.00	1.00

Cálculo de las distancias:

$$d_{ij} = \left[\sum_{K=1}^{r}(x_{ik} - x_{jk})^2\right]^{0.5}$$

$$d_{1,2} = [(-1.08 - 0.58)^2 + (-.61 - 1.08)^2 + (-0.88 - 1.07)^2]^{0.5} = 3.07$$

$$d_{1,3} = 0.54$$

$$d_{1,4} = 3.59$$

$$d_{1,5} = 0.96$$

$$d_{2,3} = 3.08$$

$$d_{2,4} = 0.83$$

$$d_{2,5} = 2.39$$

$$d_{3,4} = 3.50$$

$$d_{3,5} = 0.72$$

$$d_{4,5} = 2.79$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5 personas a la que le nombraremos ***matriz de distancia*** la cual representa que tan parecidas son las personas con respecto a la distancia entre ellas.

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Como esta matriz es una matriz espejo, es decir la matriz superior es igual a la matriz inferior y la diagonal contiene ceros ya que la distancia que hay al comparar una persona consigo misma es nula, sólo será necesario utilizar la información de la matriz inferior como se muestra a continuación:

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Al observar la matriz de distancia se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya distancia entre sí es la menor que está formada por las personas 1 y 3, con una medida de distancia de 0.54 y la menos parecida será aquella cuya distancia entre sí es la mayor que está formada por las personas 1 y 4, con una medida de distancia de 3.59.

Resolución al inciso b)

Vinculación o eslabonamiento:

1. **Eslabonamiento Simple:**

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia menor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.54

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

El nuevo conglomerado (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

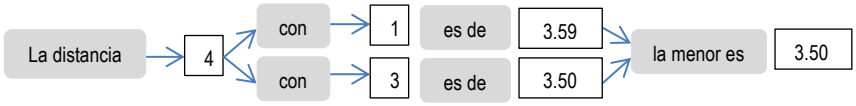
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al nuevo conglomerado (1,4) con el resto de las personas; al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.11 y la de la persona 2 con la 4 que es 3.06, eligiendo entonces la distancia menor de las dos personas que en este caso es 3.06. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4			0	
5				0

Repetimos el mismo criterio para eslabonar la persona 4 con la pareja de personas (1,3); al comparar (1,3 con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4con 1 que es 3.59 y la de la personas 4 con 3 que es 3.50, eligiendo entonces la distancia menor de las dos personas que en este caso es 3.50. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 0.96 y la de la personas 5 con 3 que es 0.72, eligiendo entonces la distancia menor de las dos personas que en este caso es 0.72. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50		0	
5	0.72			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50	0.83	0	
5	0.72	2.39	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 5 que se eslabonaría al conglomerado formado por las personas (1,3) con la distancia menor de 0.72 como se muestra a continuación:

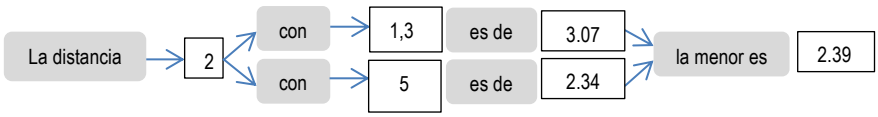
	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50	0.83	0	
5	0.72	2.39	2.79	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3,5)	2	4
(1,3,5)	0		
2		0	
4			0

Paso 2. Eslabonar

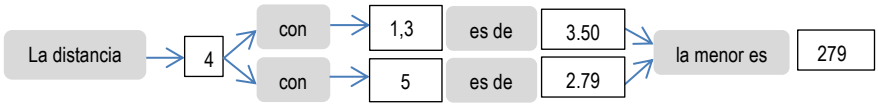
Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (1,3,5) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las 1 y 3 que es 3.07 y la de la persona 2 con la 5 que es 2.39, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.39. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4			0

Repetimos el mismo criterio para eslabonar la persona 4 con el conglomerado (1,3,5); al comparar (1,3,5) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4con las personas 1 y 3 que es 3.50 y la de la personas 4 con la 5 que es 2.79, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.79. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4	2.79		0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4	2.79	0.83	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso serían las personas 2 y 4 para formar un nuevo conglomerado con la distancia menor de 0.83 como se muestra a continuación:

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4	2.79	0.83	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (1,3,5) con el resto de las personas; al comparar (1,3,5) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las personas (1,3,5) que es 2.39 y la de la persona 4 con las (1,3,5) que es 2.79, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.39. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)	2.39	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 2.39.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.54	1	3	1
2	3	0.72	1	5	1
3	2	0.83	2	4	2
4	1	2.39	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La segunda columna no muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 1 y la 3 y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 5 al primer conglomerado formado por las personas 1 y 3 como se muestra en la columna 4 a una distancia de 0.72. En la etapa 3 se forma un nuevo conglomerado como se aprecia en la columna 5 formado por las personas 2 y 4 a una distancia de 0.83 y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de 2.39. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: V1, V2, V3

```
Variables estandarizadas, Distancia euclediana, Enlace simple
Pasos de amalgamación
```

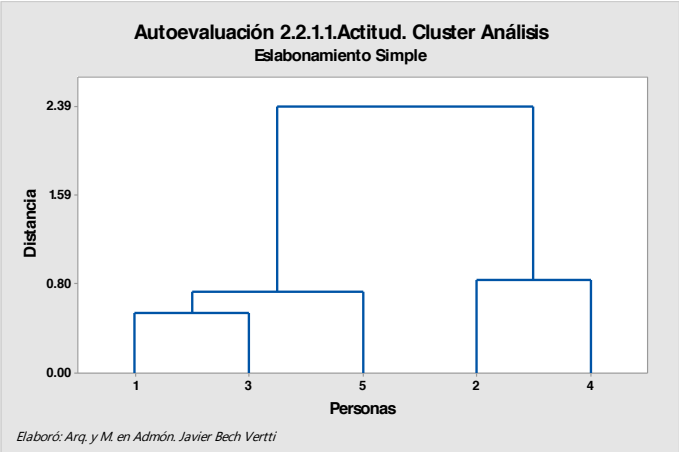
Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	85.0692	0.53606	1	3	1	2
2	3	79.8098	0.72488	1	5	1	3
3	2	76.8692	0.83045	2	4	2	2
4	1	33.4662	2.38874	1	2	1	5

Dendrograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,5) y el conglomerado 2 por las personas (2,4).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

2. Eslabonamiento Completo:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia mayor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.54

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

El nuevo conglomerado (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

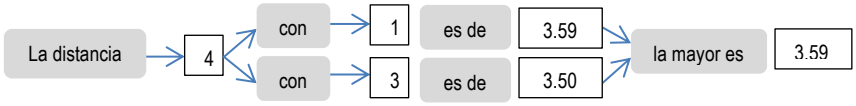
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al nuevo conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.07 y la de la persona 2 con la 3 que es 3.08, eligiendo entonces la distancia mayor de las dos personas que en este caso es 3.08. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4			0	
5				0

Repetimos el mismo criterio para eslabonar a la persona 4 con la pareja de personas (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre las personas 4 con 1 que es 3.59 y la de la personas 4 con 3 que es .50, eligiendo entonces la distancia mayor de las dos personas que en este caso es 3.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 0.96 y la de la personas 5 con 3 que es 0.72, eligiendo entonces la distancia mayor de las dos personas que en este caso es 0.96. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59		0	
5	0.96			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59	0.83	0	
5	0.96	2.39	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 2 y 5 con una distancia de 1.27 como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59	**0.83**	0	
5	0.96	2.39	2.79	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)		0	
5			0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre el conglomerado (1,3) con la persona 2 que es 3.08 y al conglomerado (1,3) con la de la persona 4 que es 3.59, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 3.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5			0

Repetimos el mismo criterio para eslabonar la persona 5 con el conglomerado (2,4); al comparar (2,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre la persona 5 con la persona 2 que es 2.39 y la de la persona 5 con la 4 que es 2.79, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 2.79. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5		2.79	0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz reducida anterior donde ya está el valor de 5 con 1 y 3 que es 0.96

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5	0.96	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona a un conglomerado ya formado, que en este caso sería la persona 3 que se eslabonaría al conglomerado formado por las personas (1,4) con la distancia menor de 1.62 como se muestra a continuación:

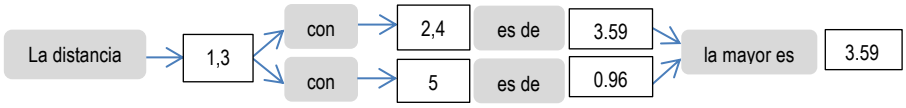
	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5	**0.96**	2.79	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (1,3,5) con el conglomerado (2,4); al comparar (1,3,5) con el conglomerado (2,4) las personas 2 y 4 se debe verificar en la matriz anterior las distancias entre las personas 2 y 4 con las personas (1,3) que es 3.59 y las personas 1 y 3 con la persona 5 que es 0.96, eligiendo entonces la distancia mayor de las dos combinaciones o cinco personas que en este caso es 3.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)	**3.59**	0

En este punto se detiene el procedimiento llegando a unir a las cinco personas en un solo grupo con una distancia de 3.59.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.53	1	3	1
2	3	0.83	2	4	2
3	2	0.96	1	5	1
4	1	3.59	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan **cuatro etapas**. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna**

nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **columna 4** que en este caso fueron la 1 y la 3 a una distancia de **0.53** y en la **quinta columna** nos señala la formación del primer conglomerado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 4 a una distancia de **0.83** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 5 al primer conglomerado formado por las personas 1 y 4 a una distancia de **0.96** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **3.59**. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: V1, V2, V3

```
Variables estandarizadas, Distancia euclediana, Enlace completo
Pasos de amalgamación
```

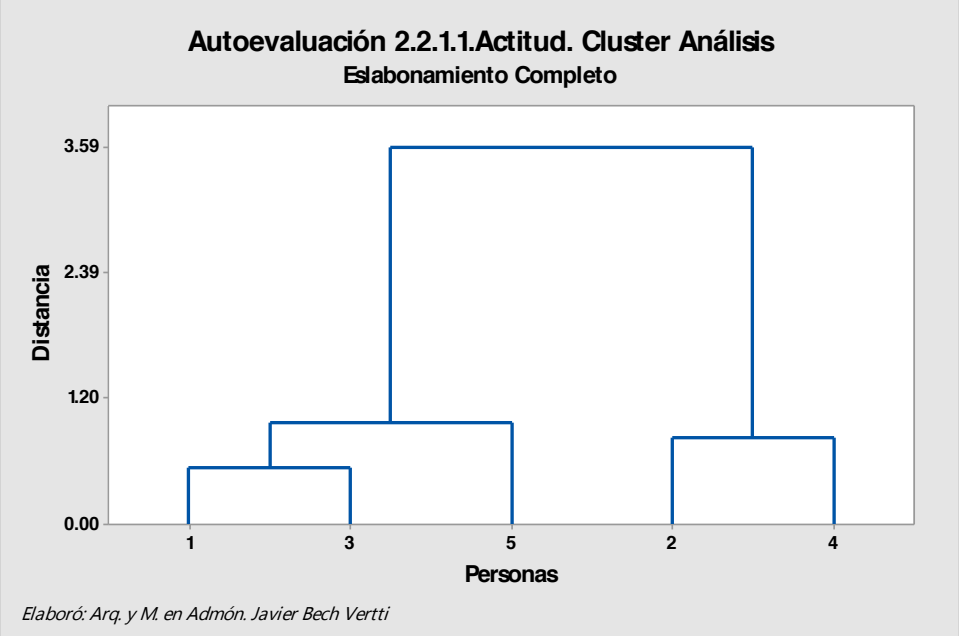
Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	85.0692	0.53606	1	3	1	2
2	3	76.8692	0.83045	2	4	2	2
3	2	73.1719	0.96320	1	5	1	3
4	1	0.0000	3.59026	1	2	1	5

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,5) y el conglomerado 2 por las personas (2,4).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

3. Eslabonamiento Promedio:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia promedio, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.54

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

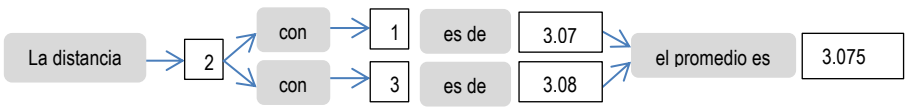
	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

El nuevo conglomerado (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

Paso 2. Eslabonar

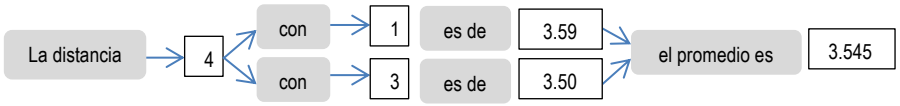
Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al nuevo conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.07 y la de la persona 2 con la 3 que es 3.08, eligiendo entonces la distancia promedio de las dos personas que en este caso es 3.075. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4			0	
5				0

Repetimos el mismo criterio para eslabonar la persona 4 con la pareja de personas (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4 con 1 que es 3.59 y la de la personas 4 con 3 que es 3.50, eligiendo entonces la distancia promedio de las dos personas que en este caso es 3.545. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre la persona 5 con 1 que es 0.96 y la de la persona 5 con 3 que es 0.72, eligiendo entonces la distancia promedio de las dos personas que en este caso es 0.84. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545		0	
5	0.84			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545	0.83	0	
5	0.84	2.39	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 2 y 4 a una distancia menor de 0.83 como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545	**0.83**	0	
5	0.84	2.39	2.79	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)		0	
5			0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado(1,3) con las personas del conglomerado (2,4); al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre las persona (1,3) y la persona 2 que es 3.075 y la de las personas (1,3) con la persona 4 que es 3.545, eligiendo entonces la distancia promedio de las dos combinaciones o cuatro personas que en este caso es 3.31. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5			0

Repetimos el mismo criterio para eslabonar a la persona 5 con el conglomerado (2,4); al comparar (2,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con las personas 2 que es 2.39 y la de la personas 5 con la 4 que es 2.79, eligiendo entonces la distancia promedio de las dos combinaciones o tres personas que en este caso es 2.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5		2.59	0

Como el valor del conglomerado (1,3) Con la persona 5 ya esta en la matriz inicial con un valor de 0.84 se pasa tal cual a la matriz reducida:

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5	0.84	2.59	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 3 para eslabonarse al conglomerado formado por las personas (1,4) con la distancia menor de 1.33 como se muestra a continuación:

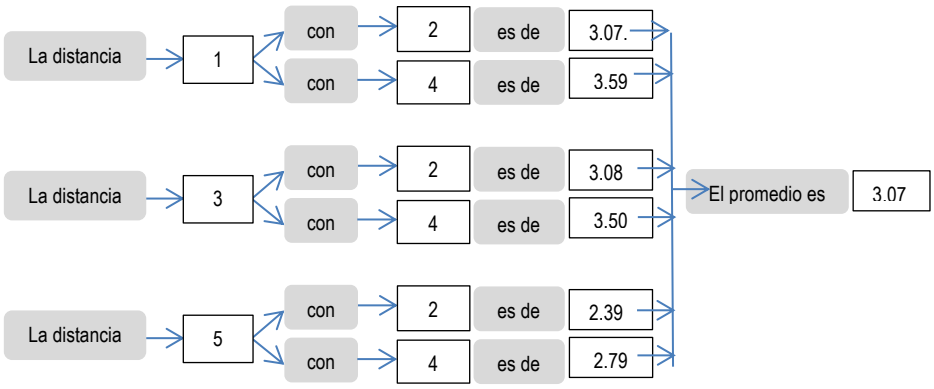
	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5	0.84	2.59	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado (1,3,5) con el resto de las personas del conglomerado (2,4); al comparar (2,4) con las personas (1,3,5) se debe verificar **por única vez en la primera matriz de distancias** entre las personas 1,3 y 5 con las personas 2,4, eligiendo entonces la distancia promedio de las seis combinaciones o cinco personas que en este caso es 3.07. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)	3.07	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 3.07.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.54	1	3	1
2	3	0.83	2	4	2
3	2	0.84	1	5	1
4	1	3.07	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **cuarta columna** que en este caso fueron la 1 y la 3 a una distancia de **0.54** y en la **quinta columna** nos señala el número de conglomerado formado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 4 a una distancia de **0.83** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 3 al primer conglomerado formado por las personas 1 y 3 a una distancia de **0.84** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **3.07** Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

```
Variables estandarizadas, Distancia euclediana, Enlace de promedios
Pasos de amalgamación
```

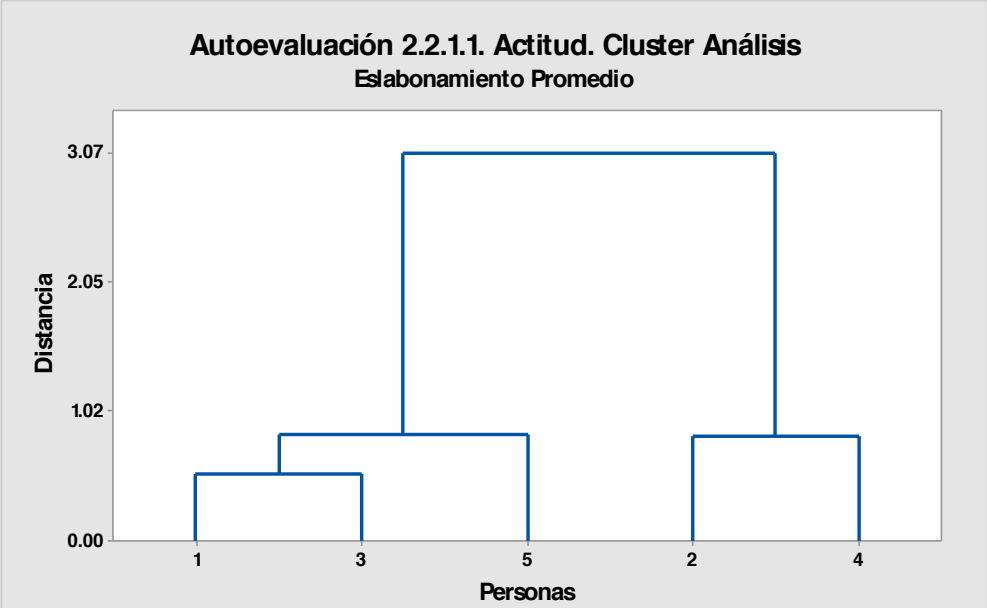
Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	85.0692	0.53606	1	3	1	2
2	3	76.8692	0.83045	2	4	2	2
3	2	76.4909	0.84404	1	5	1	3
4	1	14.4756	3.07055	1	2	1	5

Dendrograma

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,5) y el conglomerado 2 por las personas (2,4).



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Respuesta al inciso c)

Nombre de los conglomerados

Una vez elegidos el número de conglomerados definitivo, para finalizar el proceso es deseable colocar un nombre apropiado a los conglomerados. En este caso el conglomerado 1 está formado por las personas 1,3 y 5 y el conglomerado 2 está formado por las personas 2 y 4. Para conocer el tipo de personas que se encuentran en cada conglomerado se deben obtener la media de los datos originales de quienes la conforman variable por variable como se muestra a continuación:

Para el conglomerado 1:

- La media de la variable V1 es de: (1+2+3)/3=2 Ptos.
- La media de la variable V2 es de: (2+1+2)/3=1.67 Ptos.
- La media de la variable V3 es de: (1+1+2)/3=1.33 Ptos.

Para el conglomerado 2:

- La media de la variable V1 es de: (5+7)/2=6 Ptos.

- La media de la variable V2 es de: (7+7)/2=7 Ptos.
- La media de la variable V3 es de: (5+5)/2=5 Ptos.

Los resultados del conglomerado 1 nos indican que las personas de este grupo están menos de acuerdo con las afirmaciones que se proponen cuando salen de compras que las personas del conglomerado 2 por lo que un posible nombre para el conglomerado 1 podría ser "Poca Actitud" y para el conglomerado 2 " Mucha Actitud".

2.2.1.2 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 2.2.1.2 Cluster Análisis. Medida de Similitud

Suponga que se desea segmentar a los clientes de un supermercado con base a la actitud que tienen cuando salen de compras. De acuerdo a una investigación previa, se identificaron cuatro variables de actitud y se pidió a los siete entrevistados que expresaran su grado de acuerdo con las dos primeras de las cuatro afirmaciones en una escala de siete puntos en la cual 1 significa en desacuerdo y 7 significa de acuerdo. Para las dos restantes se quiso ser más preciso en los resultados y se modificó la escala de medición cuantitativa por una cualitativa

Las afirmaciones podrían ser las siguientes:

V1 Salir de compras es divertido
V2 Combino la salida de compras con la comida fuera de casa
V3 Puedo ahorrar mucho dinero si comparo precios
V4 Gusto por salir de compras con mi familia

Si los resultados para cada una de las afirmaciones se considera que son variables, se tienen cuatro, las cuales se representarán con V1, V2, V3 y V4; por otra parte, las respuestas de cada uno de los entrevistados constituyen los siete casos, que se representarán con C1, C2,…,C7. Los valores de las variables para los siete casos, se presentan a continuación:

ENTREVISTADO	VARIABLE			
	V1	V2	V3	V4
C1	1	2	2	NADA
C2	5	7	7	MUCHO
C3	2	2	1	NADA
C4	7	5	7	MUCHO
C5	3	3	2	NADA
C6	6	4	6	MUCHO
C7	4	6	7	POCO

Criterio de similitud para recodificar los datos:

V1: existe(1) si es ≥ a 4.
no existe(0) si es < a 4.

V2: existe(1) si es ≥ a 3.
no existe(0) si es < a 3.

V3: existe(1) si es ≥ a 3.
no existe(0) si es < a 3.

V4: existe(1) si es Mucho
no existe (0) si es cualquier otro gusto

a) Obtenga la pareja de personas entrevistadas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

Los datos recodificados quedaron como sigue:

ENTREVISTADO	V1	V2	V3	V4
1	0	0	0	0
2	1	1	1	1
3	0	0	0	0
4	1	1	1	1
5	0	1	0	0
6	1	1	1	1
7	1	1	1	0

Cálculo de las similitudes a encontrar:

El coeficiente de similitud que se usará para resolver este ejemplo es el de parejas simples. El coeficiente es:

$$S_{ij} = \frac{a+d}{p}$$

Se debe cuantificar la similitud entre las personas con base en este coeficiente, donde ***a*** es la presencia del atributo en ambas personas, ***d*** es la ausencia del atributo en ambas personas y ***p*** es el total de características (o variables) que serán evaluadas.
Se empieza a comparar por pareja de personas, en este caso la persona 1 y la 2, como en ninguna de las características hay presencia del atributo en ambas (***a***) ni ausente en ambas (***d***), entonces tanto ***a*** como ***d*** tienen valor de 0 en el coeficiente, en cuanto a ***p*** tenemos que son cuatro características para evaluar (V1,V2,V3 y V4); la similitud entre las personas 1 y 2 es de 0 como a continuación se presenta:

$$S_{1.2} = \frac{0+0}{4} = 0$$

De igual forma se procede con las personas 1 y 3, donde hay cero atributos presentes (***a***) en ambos y cuatro atributos ausentes (***d***) en ambos en todas las características; por lo tanto, la similitud entre las personas 1 y 3 es de 1 como a continuación se presenta:

$$S_{1.3} = \frac{0+1}{4} = 1$$

Y así sucesivamente, se obtienen las medidas de similitud para todas las parejas que se pretende comparar obteniendo los siguientes resultados:

$S_{1.2} = \frac{0+0}{4} = 0$ $S_{2.3} = \frac{0+0}{4} = 0$ $S_{3.5} = \frac{0+0}{4} = 0$ $S_{5.6} = \frac{1+0}{4} = 1/4$

$S_{1.3} = \frac{0+4}{4} = 1$ $S_{2.4} = \frac{4+0}{4} = 1$ $S_{3.6} = \frac{0+0}{4} = 0$ $S_{5.7} = \frac{1+1}{1} = 1/2$

$S_{1.4} = \frac{0+0}{4} = 0$ $S_{2.5} = \frac{1+0}{4} = 1/4$ $S_{3.7} = \frac{0+1}{4} = 1/4$ $S_{6.7} = \frac{3+0}{4} = 3/4$

$S_{1.5} = \frac{0+3}{4} = 3/4$ $S_{2.6} = \frac{4+0}{4} = 1$ $S_{4.5} = \frac{1+0}{4} = 1/4$

$S_{1.6} = \frac{0+0}{0} = 0$ $S_{2.7} = \frac{3+0}{4} = 3/4$ $S_{4.6} = \frac{4+0}{4} = 1$

$S_{1.7} = \frac{0+1}{4} = 1/4$ $S_{3.4} = \frac{0+0}{4} = 1/3$ $S_{4.7} = \frac{3+0}{4} = 3/4$

Con la información anterior se construye una matriz cuadrada en este caso de 7 x 7 ya que tenemos 7 personas a la que le nombraremos ***matriz de similitud*** la cual representa que tan similares son las personas con respecto a la similitud entre ellas:

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	**1**	0	1				
4	0	**1**	0	1			
5	3/4	1/4	3/4	1/4	1		
6	0	**1**	0	**1**	1/4	1	
7	1/4	3/4	1/4	3/4	1/2	3/4	1

Al observar la matriz de similitud se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya similitud entre sí es la que se acerque más o sea igual a 1, en este caso tenemos cuatro parejas, la primera formada por las personas 1 y 3; la segunda formada por las personas 2 y 4; la tercera formada por las personas 2,6 y la cuarta formada por las personas 4 y 6, todas con una medida de similitud de 1.
La pareja de personas menos parecida será aquella cuya similitud entre sí es la se acerque más o sea igual a 0. En este caso tenemos seis parejas, la primera formada por las personas 1 y 2; la segunda formada por las personas 1 y 4; la tercera formada por las personas 1 y 6; la cuarta formada por las personas 2 y 3; la quinta formada por las personas 3 y 4 y la sexta formada por las personas 3 y 6, todas con una medida de similitud de 0.

Resolución al inciso b)

Vinculación o eslabonamiento:

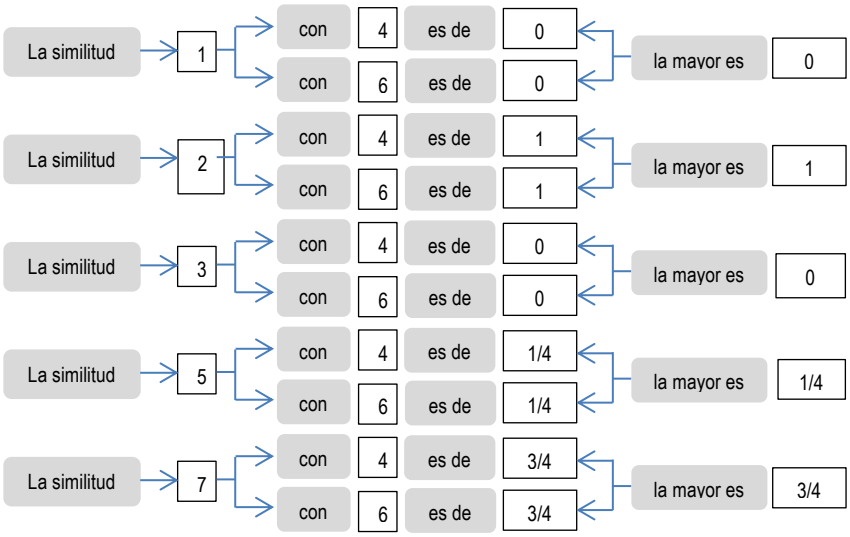
1. **Eslabonamiento Simple:**

Paso 1: Reducir la matriz con la similitud más grande. Cercana o igual a (1)

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	1	0	1				
4	0	1	0	1			
5	3/4	1/4	3/4	1/4	1		
6	0	1	0	**1**	1/4	1	
7	1/4	3/4	1/4	3/4	1/2	3/4	1

Paso 2: Llenar la matriz con la similitud más grande. Cercana o igual a (1

	(4,6)	1	2	3	5	7
(4,6)	1					
1	0	1				
2	1	0	1			
3	0	1	1	1		
5	1/4	3/4	1/4	3/4	1	
7	3/4	1/4	3/4	1/4	1/2	1

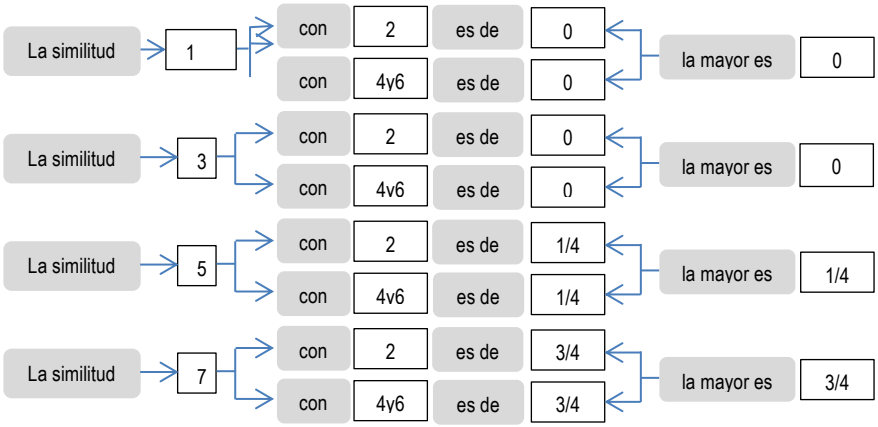


Paso 1: Reducir la matriz con la similitud más grande. (1)

	(4,6)	1	2	3	5	7
(4,6)	1					
1	0	1				
2	1	0	1			
3	0	1	1	1		
5	1/4	3/4	1/4	3/4	1	
7	3/4	1/4	3/4	1/4	1/2	1

Paso 2: Llenar la matriz con la similitud más grande. (1)

	(2,4,6)	1	3	5	7
(2,4,6)	1				
1	0	1			
3	0	1	1		
5	1/4	3/4	3/4	0	
7	3/4	1/4	1/4	1/2	1

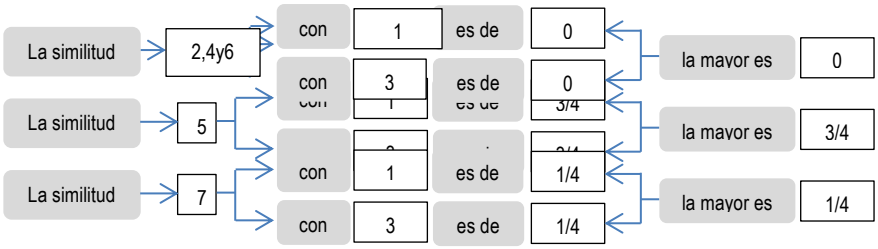


Paso 1: Reducir la matriz con la similitud más grande. Cercana o igual a (1)

	(2,4,6)	1	3	5	7
(2,4,6)	1				
1	0	1			
3	0	1	1		
5	1/4	3/4	3/4	0	
7	3/4	1/4	1/4	1/2	1

Paso 2: Llenar la matriz con la similitud más grande. Cercana o igual a (1)

	(2,4,6)	(1,3)	5	7
(2,4,6)	1			
(1,3)	0	1		
5	1/4	3/4	1	
7	3/4	1/4	1/2	1

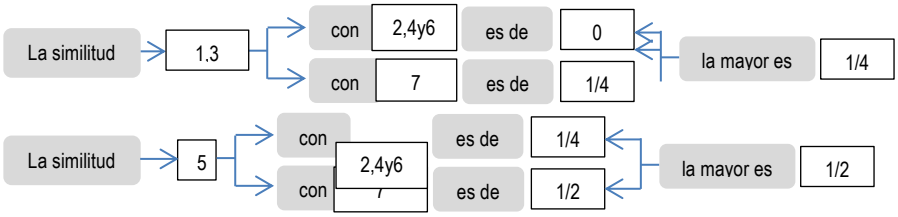


Paso 1: Reducir la matriz con la similitud más grande. Cercana o igual a (1)

	(2,4,6)	(1,3)	5	7
(2,4,6)	1			
(1,3)	0	1		
5	1/4	3/4	1	
7	3/4	1/4	1/2	1

Paso 2: Llenar la matriz con la similitud más grande. Cercana o igual a (1)

	(2,4,6,7)	(1,3)	5
(2,4,6,7)	1		
(1,3)	1/4	1	
5	1/2	3/4	1



Paso 1: Reducir la matriz con la similitud más grande. Cercana o igual a (1)

	(2,4,6,7)	(1,3)	5
(2,4,6,7)	1		
(1,3)	1/4	1	
5	1/2	3/4	1

Paso 2: Llenar la matriz con la similitud más grande. Cercana o igual (1)

	(2,4,6,7)	(1,2,5)
(2,4,6,7)	1	
(1,2,5)	1/2	1

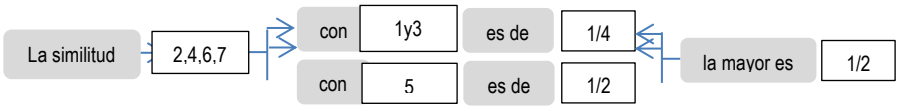


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	4	6	4
2	5	1	2	4	2
3	4	1	1	3	1

4	3	3/4	2	7	2
5	2	3/4	1	5	1
6	1	1/2	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos una nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 4 y la 6 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 2 al primer conglomerado con una similitud de 1 o 100%. En la etapa 3 se se formó un segundo conglomerado formado por las personas 1 y 3 con una similitud de 1 o 100%. En la etapa 4 se incorporó la persona 7 al primer conglomerado formado por las personas 2,4 y 6 como se muestra en la columna 3 a una similitud de 3/4 ó 75%. En la etapa 5 se incorporó la persona 5 al segundo conglomerado formado por las personas 1 y 3 como se muestra en la columna 3 a una similitud de 3/4 ó 75% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 1/2 ó 50%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

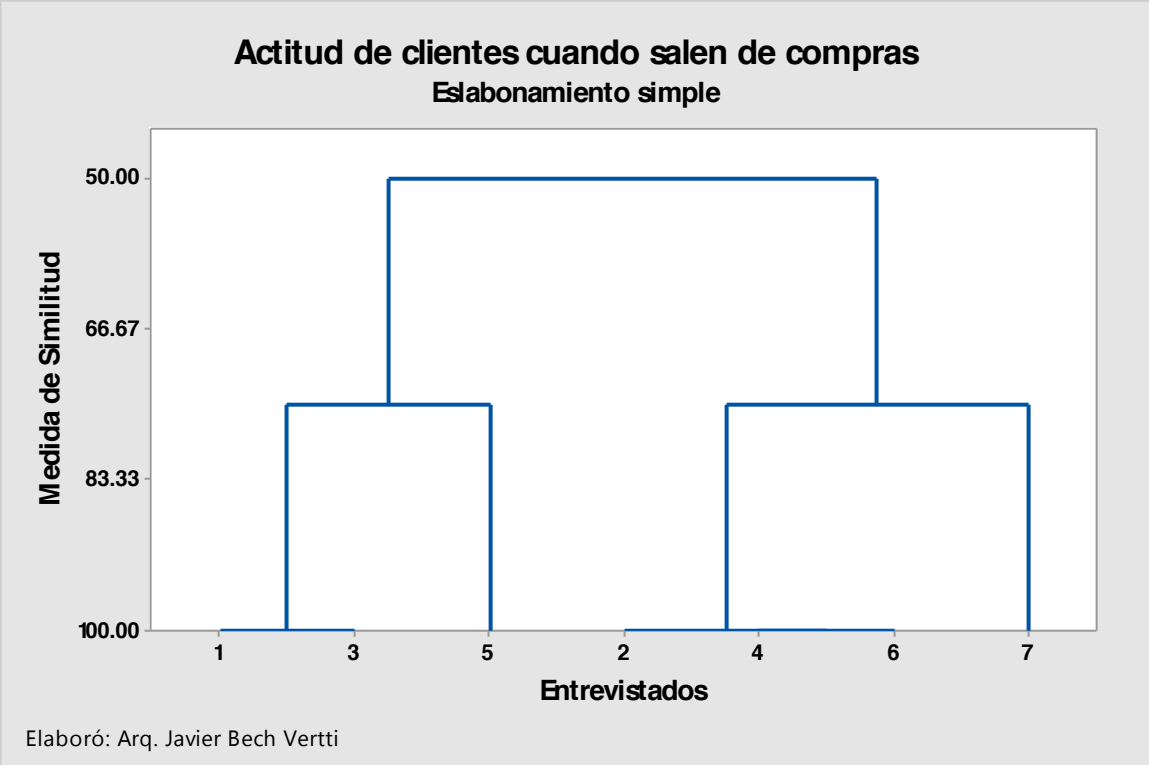
Análisis de observaciones de conglomerado: V1, V2, V3, V4

```
Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100	0	4	6	4	2
2	5	100	0	2	4	2	3
3	4	100	0	1	3	1	2
4	3	75	1	2	7	2	4
5	2	75	1	1	5	1	3
6	1	50	2	1	2	1	7

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (2,4,6,7) y el conglomerado 2 por las personas (1,3,5).
Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Elaboró: Arq. Javier Bech Vertti

Número de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

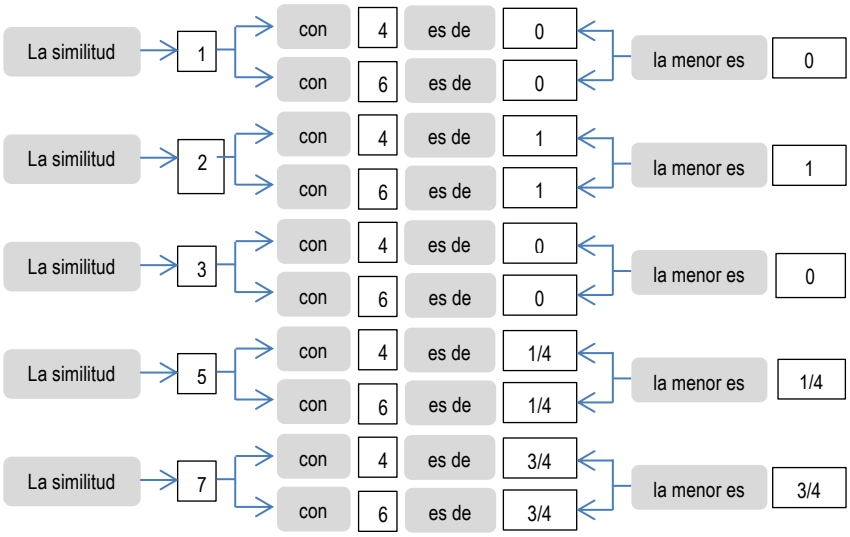
2. Eslabonamiento Completo:

Paso 1: Reducir la matriz con la similitud más grande, cercana o igual a 1

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	1	0	1				
4	0	1	0	1			
5	3/4	1/4	3/4	1/4	1		
6	0	1	0	**1**	1/4	1	
7	1/4	3/4	1/4	3/4	1/2	3/4	1

Paso 2: Llenar la matriz con la similitud más pequeña, cercana o igual a 0.

	(4,6)	1	2	3	5	7
(4,6)	1					
1	0	1				
2	1	0	1			
3	0	1	1	1		
5	1/4	3/4	1/4	3/4	1	
7	3/4	1/4	3/4	1/4	1/2	1

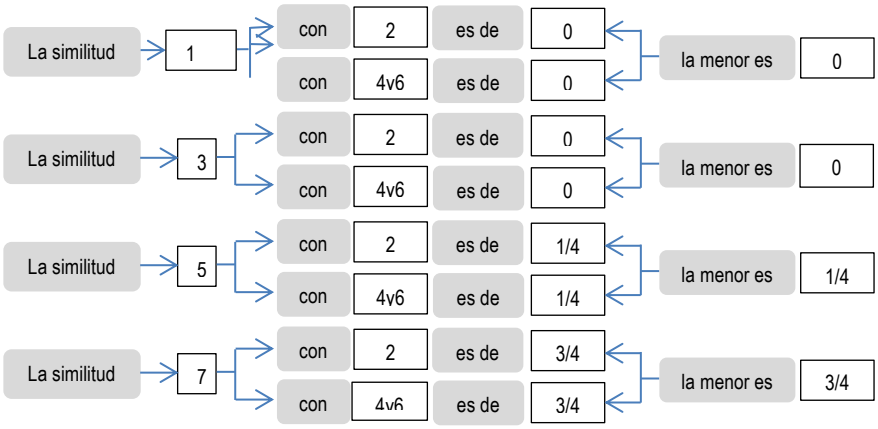


Paso 1: Reducir la matriz con la similitud más grande, cercana o igual a 1

	(4,6)	1	2	3	5	7
(4,6)	1					
1	0	1				
2	1	0	1			
3	0	1	1	1		
5	1/4	3/4	1/4	3/4	1	
7	3/4	1/4	3/4	1/4	1/2	1

Paso 2: Llenar la matriz con la similitud más pequeña, cercana o igual a 0.

	(2,4,6)	1	3	5	7
(2,4,6)	1				
1	0	1			
3	0	1	1		
5	1/4	3/4	3/4	0	
7	3/4	1/4	1/4	1/2	1



Paso 1: Reducir la matriz con la similitud más grande, cercana o igual a 1

	(2,4,6)	1	3	5	7
(2,4,6)	1				
1	0	1			
3	0	1	1		
5	1/4	3/4	3/4	0	
7	3/4	1/4	1/4	1/2	1

Paso 2: Llenar la matriz con la similitud más pequeña, cercana o igual a 0.

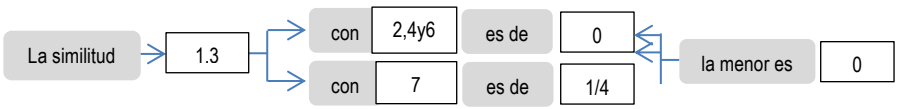
	(2,4,6)	(1,3)	5	7
(2,4,6)	1			
(1,3)	0	1		
5	1/4	3/4	1	
7	3/4	1/4	1/2	1

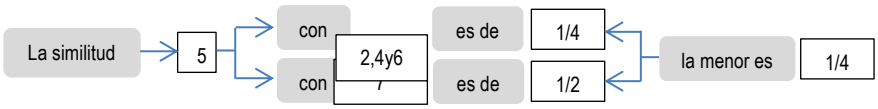
Paso 1: Reducir la matriz con la similitud más grande, cercana o igual a 1

	(2,4,6)	(1,3)	5	7
(2,4,6)	1			
(1,3)	0	1		
5	1/4	3/4	1	
7	3/4	1/4	1/2	1

Paso 2: Llenar la matriz con la similitud más pequeña, cercana o igual a 0.

	(2,4,6,7)	(1,3)	5
(2,4,6,7)	1		
(1,3)	0	1	
5	1/4	3/4	1





Paso 1: Reducir la matriz con la similitud más grande, cercana o igual a 1

	(2,4,6,7)	(1,3)	5
(2,4,6,7)	1		
(1,3)	0	1	
5	1/4	3/4	1

Paso 2: Llenar la matriz con la similitud más pequeña, cercana o igual a 0.

	(2,4,6,7)	(1,3,5)
(2,4,6,7)	1	
(1,3,5)	0	1

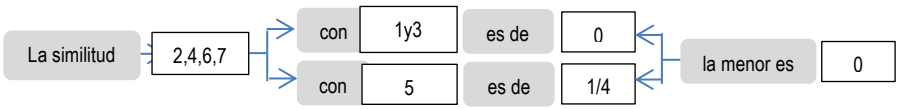


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	4	6	4
2	5	1	2	4	2
3	4	1	1	3	1
4	3	3/4	2	7	2
5	2	3/4	1	5	1
6	1	0	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan

seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 4 y la 6 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 2 al primer conglomerado con una similitud de 1 o 100%. En la etapa 3 se se formó un segundo conglomerado formado por las personas 1 y 3 con una similitud de 1 o 100%. En la etapa 4 se incorporó la persona 7 al primer conglomerado formado por las personas 2,4 y 6 como se muestra en la columna 3 a una similitud de 3/4 ó 75%. En la etapa 5 se incorporó la persona 5 al segundo conglomerado formado por las personas 1 y 3 como se muestra en la columna 3 a una similitud de 3/4 ó 75% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 0 ó 0%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

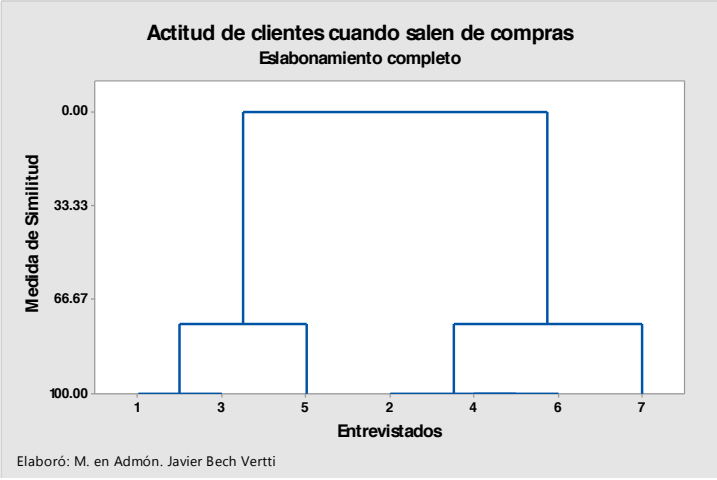
Análisis de observaciones de conglomerado: V1, V2, V3, V4

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100	0	4	6	4	2
2	5	100	0	2	4	2	3
3	4	100	0	1	3	1	2
4	3	75	1	2	7	2	4
5	2	75	1	1	5	1	3
6	1	0	4	1	2	1	7

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (2,4,6,7) y el conglomerado 2 por las personas (1,3,5).
Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Número de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

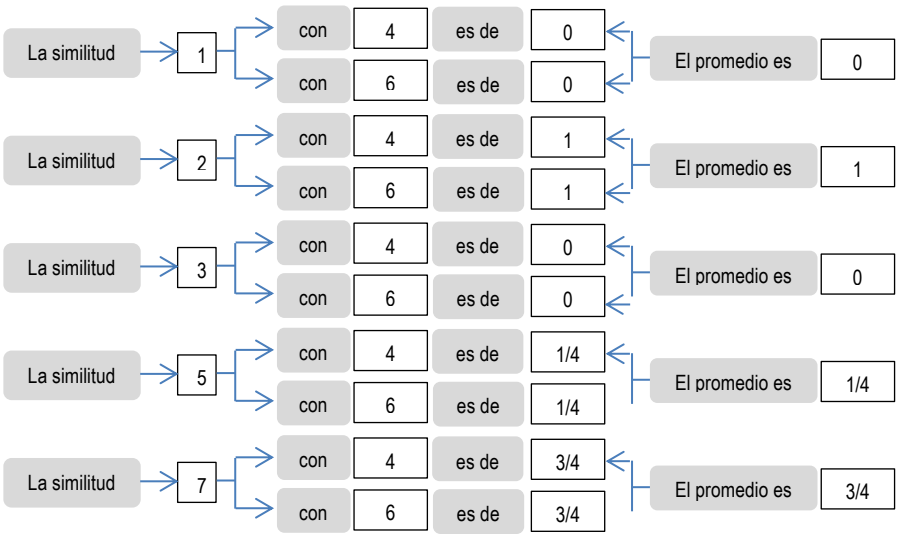
3. Eslabonamiento Promedio:

Paso 1. Reducir la matriz con la similitud más grande cercana o igual a 1

	1	2	3	4	5	6	7
1	1						
2	0	1					
3	1	0	1				
4	0	1	0	1			
5	3/4	1/4	3/4	1/4	1		
6	0	1	0	**1**	1/4	1	
7	1/4	3/4	1/4	3/4	1/2	3/4	1

Paso 2. Llenar la matriz con la similitud promedio.

	(4,6)	1	2	3	5	7
(4,6)	1					
1	0	1				
2	1	0	1			
3	0	1	1	1		
5	1/4	3/4	1/4	3/4	1	
7	3/4	1/4	3/4	1/4	1/2	1

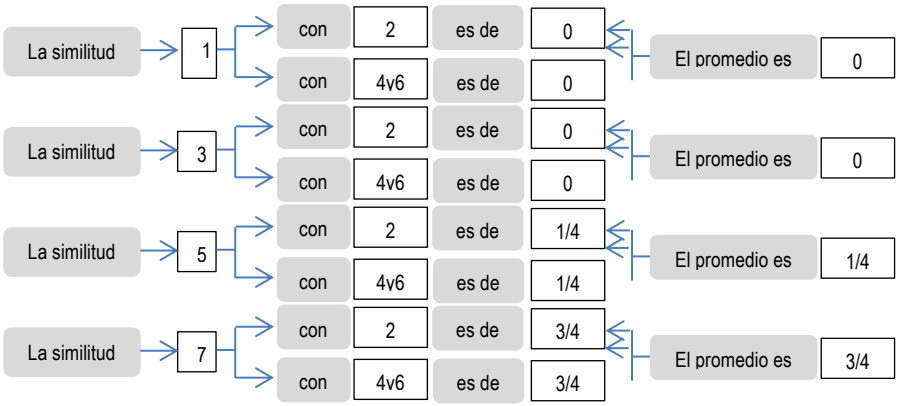


Paso 1. Reducir la matriz con la similitud más grande cercana o igual a 1

	(4,6)	1	2	3	5	7
(4,6)	1					
1	0	1				
2	1	0	1			
3	0	1	1	1		
5	1/4	3/4	1/4	3/4	1	
7	3/4	1/4	3/4	1/4	1/2	1

Paso 2. Llenar la matriz con la similitud promedio.

	(2,4,6)	1	3	5	7
(2,4,6)	1				
1	0	1			
3	0	1	1		
5	1/4	3/4	3/4	1	
7	3/4	1/4	1/4	1/2	1

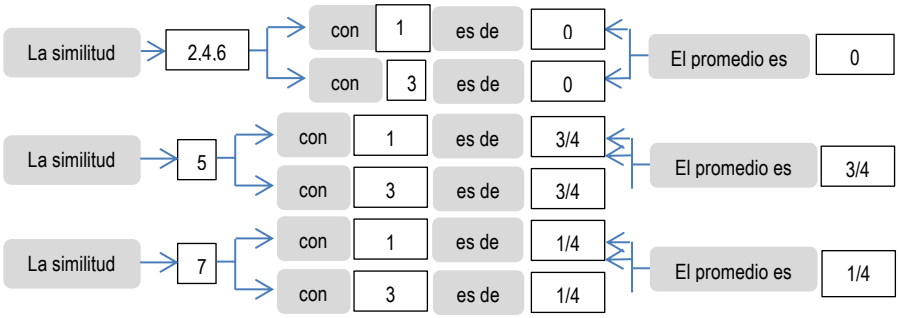


Paso 1. Reducir la matriz con la similitud más grande cercana o igual a 1

	(2,4,6)	1	3	5	7
(2,4,6)	1				
1	0	1			
3	0	1	1		
5	1/4	3/4	3/4	1	
7	3/4	1/4	1/4	1/2	1

Paso 2. Llenar la matriz con la similitud promedio.

	(2,4,6)	(1,3)	5	7
(2,4,6)	1			
(1,3)	0	1		
5	1/4	3/4	1	
7	3/4	1/4	1/2	1

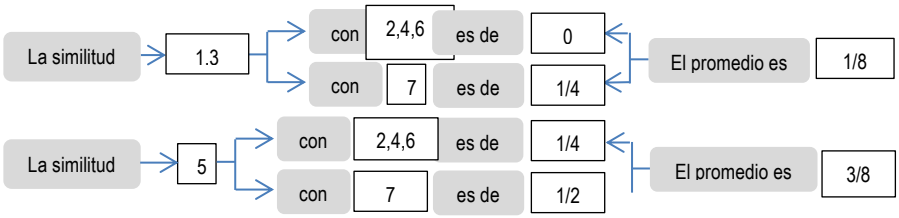


Paso 1. Reducir la matriz con la similitud más grande cercana o igual a 1

	(2,4,6)	(1,3)	5	7
(2,4,6)	1			
(1,3)	0	1		
5	1/4	3/4	1	
7	3/4	1/4	1/2	1

Paso 2. Llenar la matriz con la similitud promedio.

	(2,4,6,7)	(1,3)	5
(2,4,6,7)	1		
(1,3)	1/8	1	
5	3/8	3/4	1



Paso 1. Reducir la matriz con la similitud más grande cercana o igual a 1

	(2,4,6,7)	(1,3)	5
(2,4,6,7)	1		
(1,3)	1/8	1	
5	3/8	3/4	1

Paso 2. Llenar la matriz con la similitud promedio.

	(2,4,6,7)	(1,3,5)
(2,4,6,7)	1	
(1,3,5)	7/48	1

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener

cuidado de elegir la similitud promedio, es decir se compara al conglomerado (2,4,6,7) con el resto de las personas que forman el otro conglomerado (1,3,5); al comparar (2,4,6,7) con las personas (1,3,5) se debe verificar **en la matriz original las similitudes entre todas las combinaciones**, es decir a las personas 1 y 2 que es 0, a la persona 1 y 4 que es 0, a la persona 1 y 6 que es 0 y así sucesivamente hasta la persona 5 y 7 que es 1/2, eligiendo entonces la similitud promedio de las doce combinaciones o siete personas que en este caso es 7/48 o 14.58%. Lo anterior se puede resumir con el siguiente diagrama de flujo:

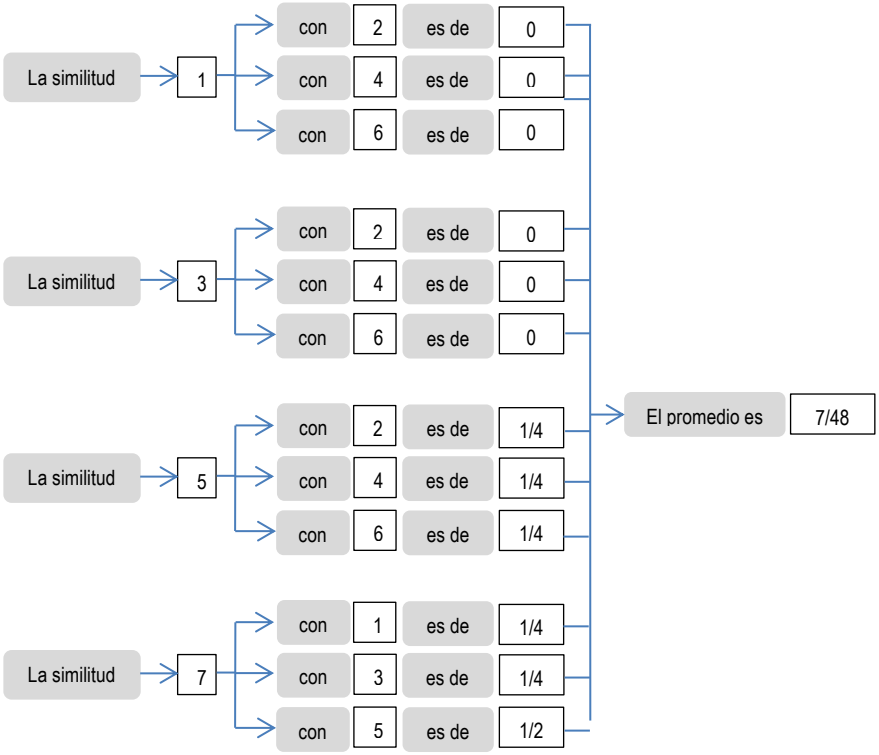


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	6	1	4	6	4
2	5	1	2	4	2
3	4	1	1	3	1
4	3	3/4	2	7	2
5	2	3/4	1	5	1
6	1	7/48	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos una nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 4 y la 6 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 2 al primer conglomerado con una similitud de 1 o

100%. En la etapa 3 se se formó un segundo conglomerado formado por las personas 1 y 3 con una similitud de 1 o 100%. En la etapa 4 se incorporó la persona 7 al primer conglomerado formado por las personas 2,4 y 6 como se muestra en la columna 3 a una similitud de 3/4 ó 75%. En la etapa 5 se incorporó la persona 5 al segundo conglomerado formado por las personas 1 y 3 como se muestra en la columna 3 a una similitud de 3/4 ó 75% y finalmente en la etapa 6 se eslabonaron ambos conglomerados a una similitud final de 7/48 ó 14.58%%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: V1, V2, V3, V4

```
Distancia euclediana cuadrada, Enlace de promedios
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	6	100.000	0.00000	4	6	4	2
2	5	100.000	0.00000	2	4	2	3
3	4	100.000	0.00000	1	3	1	2
4	3	75.000	1.00000	2	7	2	4
5	2	75.000	1.00000	1	5	1	3
6	1	14.583	3.41667	1	2	1	7

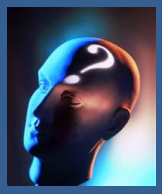
Dendrograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (2,4,6,7) y el conglomerado 2 por las personas (1,3,5). Minitab puede dibujar los conglomerados en diferente orden, primero el 1 o el 2 o bien primero el 2 y luego el 1, lo cual no cambia la solución.

Elaboró: M. en Admón. Javier Bech Vertti

Número de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.



RESPUESTAS A LOS EJERCICIOS DE REFUERZO DEL CAPÍTULO 2

2.1.1.1 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 2.1.1.1.1 Factor Análisis

Se tiene la siguiente matriz de factores no rotada:

Variable	Factor 1	Factor 2	Factor 3	Factor 4
A	0.67	0.49	-0.41	-0.35
B	0.55	0.58	0.45	-0.39
C	0.45	0.39	0.31	0.33
D	0.53	0.41	0.36	0.25
E	0.70	0.35	-0.29	0.28
F	0.65	0.48	0.45	0.32

a) Obtenga las Comunalidades, el Eigenvalue y el Porcentaje de Variación Explicada:
b) A partir de estos resultados, ¿Se puede saber cómo agrupar a las variables?. Explique.
c) ¿Cuál es la variable que más se identifica con las características?. Explique el porqué de su respuesta.

Respuesta al inciso a)

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Comunalidades
A	0.67	0.49	-0.41	-0.35	0.9796
B	0.55	0.58	0.45	-0.39	0.9935
C	0.45	0.39	0.31	0.33	0.5596
D	0.53	0.41	0.36	0.25	0.8902
E	0.70	0.35	-0.29	0.28	0.7750
F	0.65	0.48	0.45	0.32	0.9578
Eigenvalue	2.1473	1.2496	0.8829	0.6268	
V. Explicada	0.3579	0.2083	0.1472	0.1045	

Comunalidad

Se usa este término para nombrar a la variación explicada por la solución de factores para cada variable y por estar estandarizadas, el valor máximo posible de la comunalidad es de 1. Mientras más se aproxime el valor de la comunalidad a 1 quiere decir que se explico mejor la variable.

El valor de la comunalidad total se obtiene elevando al cuadrado cada carga y sumándolo por cada variable.

Eigenvalor

Se usa este término para denominar a la variación que cada uno de los factores logra explicar del conjunto de variables.

Para calcular el Eigenvalor, cada carga se eleva al cuadrado y se suma para cada factor.

Porcentaje de variación explicada

Representa la relación entre la variación explicada por el factor, con respecto a la variación máxima, la cual es igual al numero de variables, ya que cada una de ellas tiene una variación de uno por estar estandarizadas; por lo tanto, la mayor variación que se puede llegar a explicar sería la suma de las variaciones de cada variable.

A mayor porcentaje de varianza explicada quiere decir que el factor está explicando mejor a la variable.

Respuesta al inciso b)

Esta matriz nos proporciona información en relación a que tanto se identifica cada variable con cada factor; estas medidas se llaman *cargas de factores*. Como en esta matriz casi todas las cargas resultan significativas por eso no se agrupan a las variables en este paso, solamente se determina el *número de factores* en los que quedarán agrupadas las variables de manera tentativa.

Respuesta al inciso c)

Comunalidad

Se usa este término para nombrar a la variación explicada por la solución de factores para cada variable y por estar estandarizadas, el valor máximo posible de la comunalidad es de 1. Mientras más se aproxime el valor de la comunalidad a 1 quiere decir que se explico mejor la variable.

La variable B es la que mejor se identifica con las características por su comunalidad mas alta hasta el Factor 4 de 0.9935.

2.1.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 2.1.1.1.2 Factor Análisis**

Se tiene la siguiente matriz de factores no rotada.

	F1	F2
X_1	0.8386	-0.2196
X_2	-0.7387	-0.5203
X_3	0.5242	0.6748
X_4	0.5849	0.3674

a) Obtenga la variación explicada de cada factor y las comunalidades.
b) Se roto la matriz y generó la siguiente información:

	F1	F2
X_1	0.8050	0.3215
X_2	0.9033	-0.0202
X_3	-0.0208	0.8541
X_4	-0.2523	0.6430

Obtenga *Eigenvalue* y comunalidades.
c) ¿ Se ganó o perdió información con respecto a la matriz no rotada?
d) Agrupe o identifique a cada variable con el factor que le corresponda y concluya cómo quedaron agrupadas las variables.
e) ¿Cuál es la variable que más se identifica con las características?. Explique el porqué de su respuesta.

Respuesta al inciso a)

	F1	F2	Comunalidad
X_1	0.8386	-0.2196	0.7515
X_2	-0.7387	-0.5203	0.8116
X_3	0.5242	0.6748	0.7301
X_4	0.5849	0.3674	0.4771
Eigenvalue	1.8658	0.9093	
V. Explicada	0.4665	0.2273	

Respuesta al inciso b)

	F1	F2	Comunalidades
X_1	0.8050	0.3215	0.7514
X_2	0.9033	-0.0202	0.8164
X_3	-0.0208	0.8541	0.7299

X_4	-0.2523	0.6430	0.4771
Eigenvalue	1.5281	1.2467	
V. Explicada	0.3820	0.3117	

Respuesta al inciso c)

En el proceso de rotar los factores se redistribuyen las cargas cambiando el valor de los Eigenvalue para cada factor sin que la variación explicada total y las comunalidades cambien, a menos que se eliminen o agreguen factores por lo que al rotar no se explica mas o se pierde información solo se redistribuyen las cargas.

Respuesta al inciso d)

	F1	F2	Comunalidades
X_1	**0.8050**	0.3215	0.7514
X_2	**0.9033**	-0.0202	0.8164
X_3	-0.0208	**0.8541**	0.7299
X_4	-0.2523	**0.6430**	0.4771
Eigenvalue	1.5281	1.2467	
V. Explicada	0.3820	0.3117	

a) Carga altamente significativa es la de $\pm$ 0.5 para arriba (prioridad 1).
b) Carga sólo significativa es entre $\pm$ 0.4 a $\pm$ 0.5 (prioridad 2).
c) Carga sòlo significativa es entre $\pm$ 0.3 a $\pm$ 0.4 (prioridad 3).
d) Si es menor a 0.3 se considera que no existen cargas significativas (prioridad 4).

Utilizando la Prioridad 1 con cargas altamente significativas mayores a 0.5, en el Factor 1 quedarán las variables X_1 y X_2 con cargas significativas de 0.8050 y 0.9033 respectivamente y en el Factor 2 quedarán las variables X_3 y X_4 con cargas significativas de 0.8541 y 0.6430.

Respuesta al inciso e)

	F1	F2	Comunalidades
X_1	0.8050	0.3215	0.7514
X_2	0.9033	-0.0202	**0.8164**
X_3	-0.0208	0.8541	0.7299
X_4	-0.2523	0.6430	0.4771
Eigenvalue	1.5281	1.2467	
V. Explicada	0.3820	0.3117	

Comunalidad

Se usa este término para nombrar a la variación explicada por la solución de factores para cada variable y por estar estandarizadas, el valor máximo posible de la comunalidad es de 1. Mientras más se aproxime el valor de la comunalidad a 1 quiere decir que se explico mejor la variable.

La variable X_2 es la que mejor se identifica con las características por su comunalidad mas alta hasta el Factor 2 de 0.8164.

2.1.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 2.1.1.1.3 Factor Análisis**

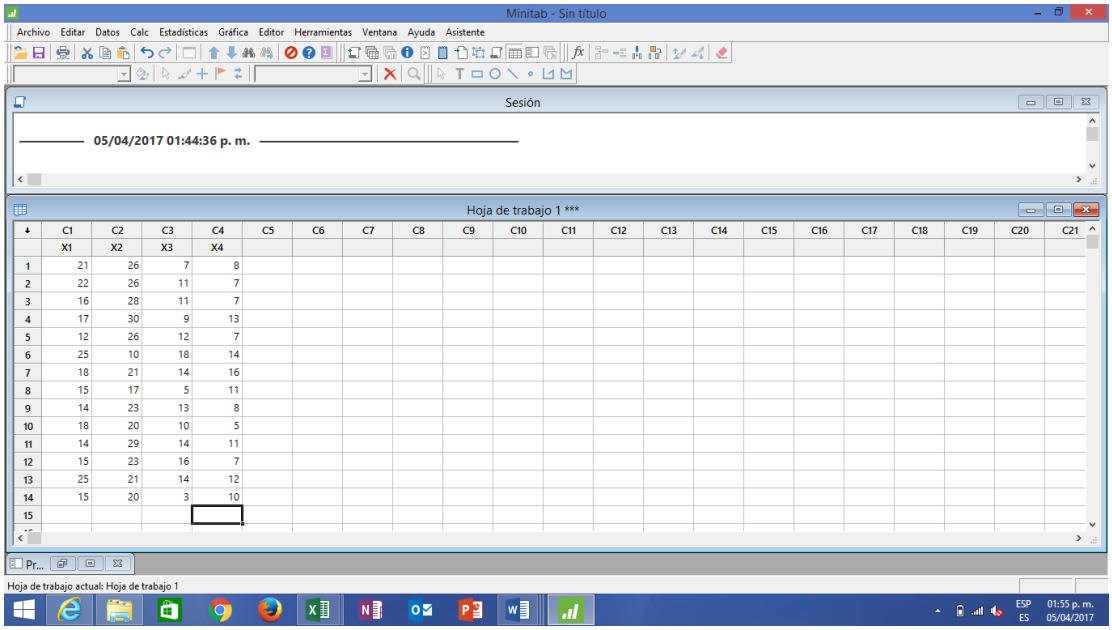
2.1.1.1.3 Una firma de investigación de mercados esta en el proceso de intentar relacionar pruebas de puntaje de actividad-interés a medidas de "éxito" de solicitantes de empleo como vendedores al detalle en un gran almacén de departamentos de cierta ciudad. Los puntajes de actividad-interés son: X_1=vivir en grupo; X_2afición por deportes al aire libre; X_3afición por la música y X_4=deseo de viajar.

Valores altos en X_i indican mayor interés o éxito respectivamente.

a) Utilice un programa de computo estadístico para correr un análisis de factores (Utilice el modelo de componentes principales y el criterio de rotación de factores Varimax)
b) Determine el número de factores óptimo y justifique su decisión.
c) Agrupe o identifique a cada variable con el factor que le corresponda y concluya cómo quedaron agrupadas las variables.
d) ¿Cuál es la variable que más se identifica con las características?. Explique el porqué de su respuesta.

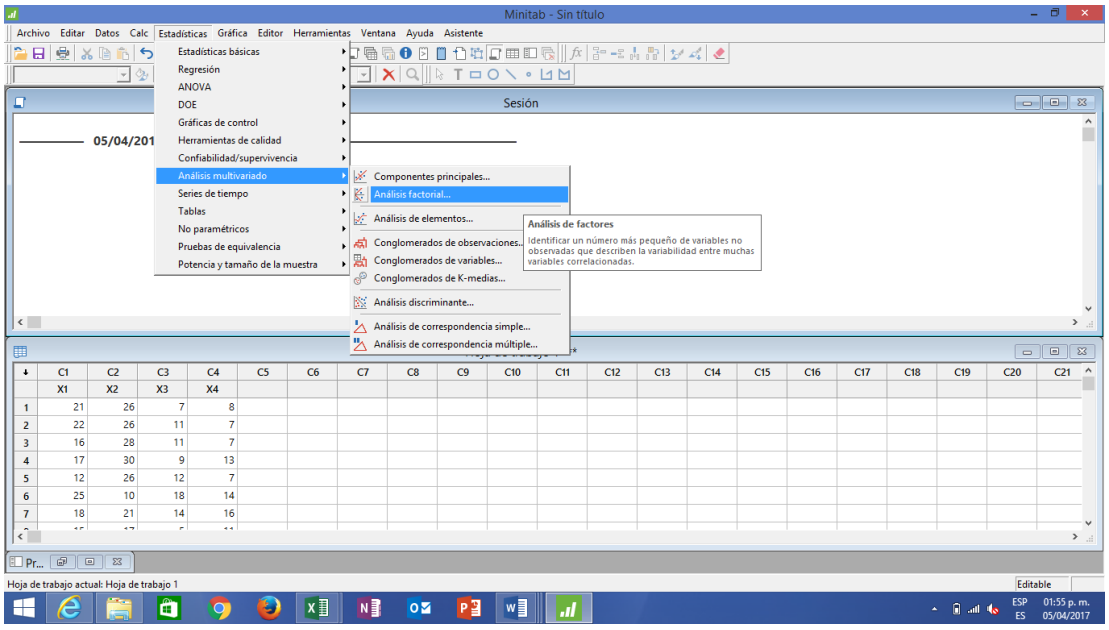
Persona	X_1	X_2	X_3	X_4
1	21	26	7	8
2	22	16	11	7
3	16	28	11	7
4	17	30	9	13
5	12	26	12	7
6	25	10	18	14
7	18	21	14	16
8	15	17	5	11
9	14	23	13	8
10	18	20	10	5
11	14	29	14	11
12	15	23	16	7
13	25	21	14	12
14	15	20	3	10

Respuesta al inciso a)

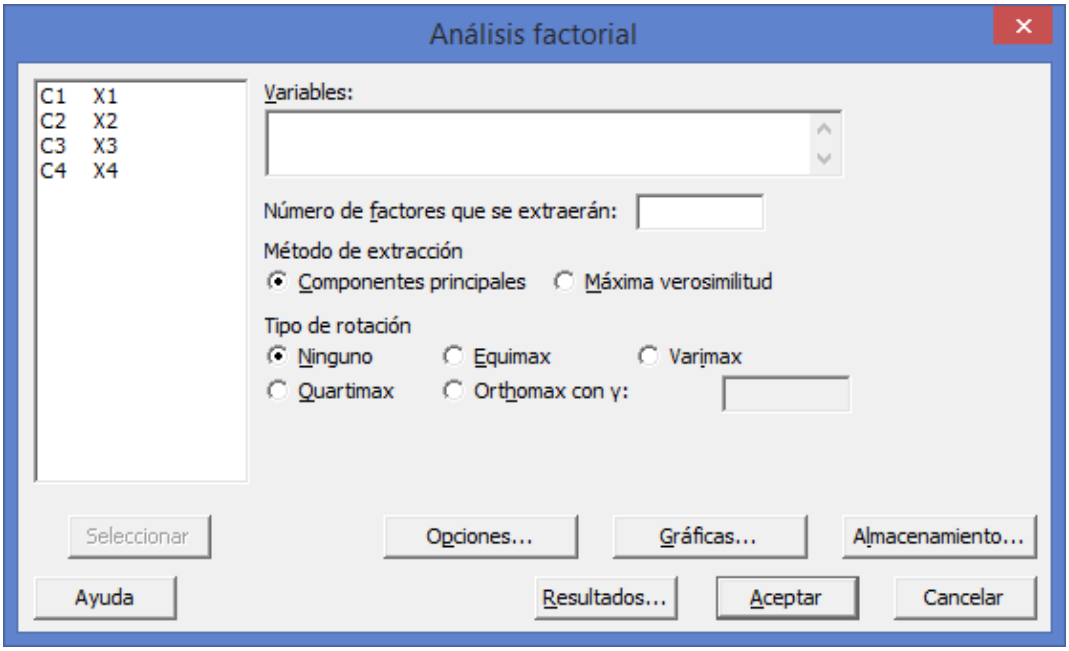


	C1	C2	C3	C4
	X1	X2	X3	X4
1	21	26	7	8
2	22	26	11	7
3	16	28	11	7
4	17	30	9	13
5	12	26	12	7
6	25	10	18	14
7	18	21	14	16
8	15	17	5	11
9	14	23	13	8
10	18	20	10	5
11	14	29	14	11
12	15	23	16	7
13	25	21	14	12
14	15	20	3	10
15				

Respuesta al inciso b)



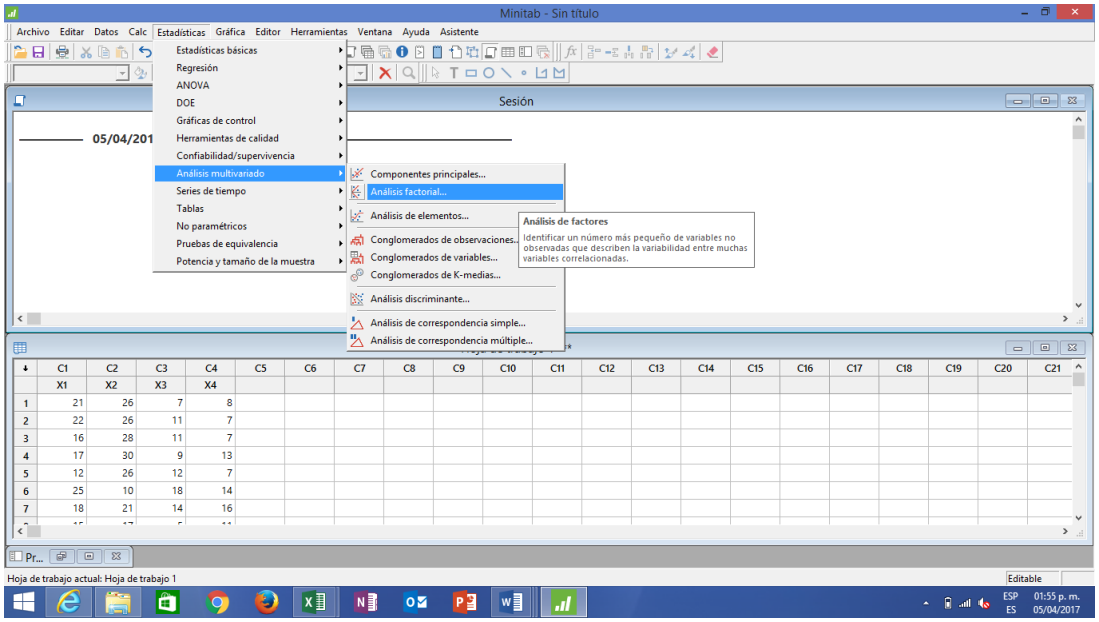
	C1	C2	C3	C4
	X1	X2	X3	X4
1	21	26	7	8
2	22	26	11	7
3	16	28	11	7
4	17	30	9	13
5	12	26	12	7
6	25	10	18	14
7	18	21	14	16

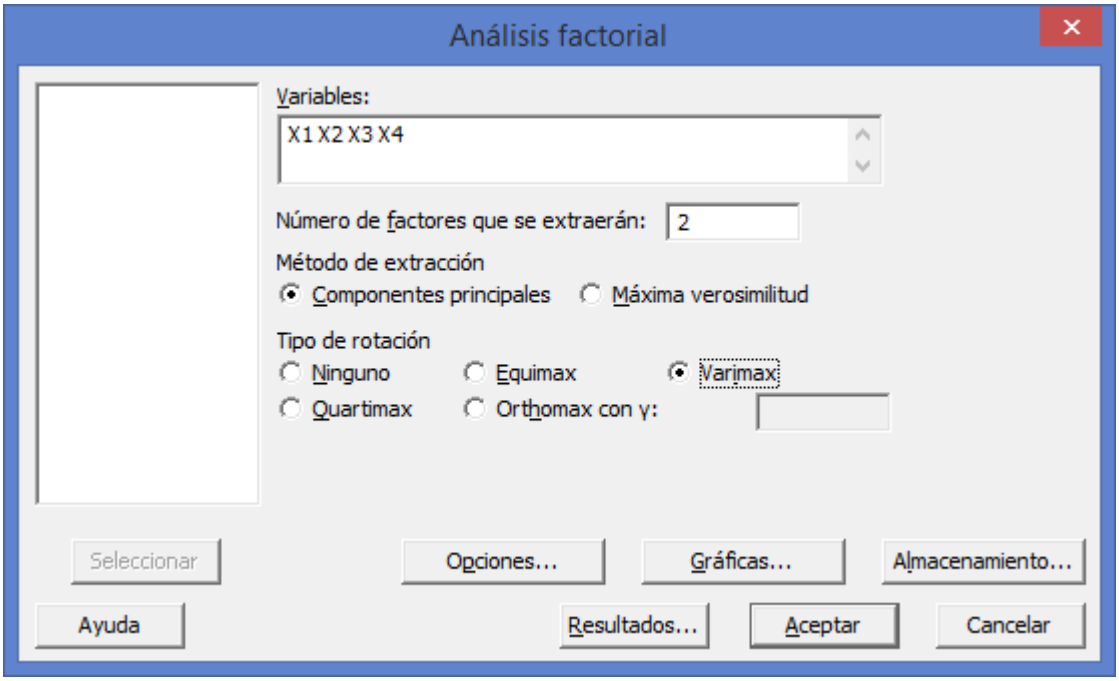


Análisis factorial: X1, X2, X3, X4

```
Análisis factorial del componente principal de la matriz de correlación


Cargas de factores no rotados y comunalidades

Variable  Factor1  Factor2  Factor3  Factor4  Comunalidad
X1          0.767    0.030   -0.353   -0.536        1.000
X2         -0.720   -0.401    0.295   -0.483        1.000
X3          0.538   -0.819    0.007    0.201        1.000
X4          0.676    0.190    0.709   -0.067        1.000

Varianza   1.8527   0.8678   0.7137   0.5657       4.0000
% Var       0.463    0.217    0.178    0.141        1.000

Por el criterio de raíz latente sólo se necesitaría un factor con un Eigenvalor mayor
a 1 pero por el criterio de Variación explicada que debe ser mayor al 60% se
necesitan dos factores para alcanzar un 0.680 ó 68% (0.463+0.217).
```

Respuesta al inciso c)





Análisis factorial: X1, X2, X3, X4

Análisis factorial del componente principal de la matriz de correlación

Cargas de factores rotados y comunalidades
Rotación Varimax

Variable	Factor1	Factor2	Comunalidad
X1	**0.700**	-0.314	0.588
X2	**-0.823**	-0.039	0.679
X3	0.118	**-0.972**	**0.959**
X4	**0.690**	-0.130	0.493
Varianza	1.6582	1.0623	2.7205
% Var	0.415	0.266	0.680

a) Carga altamente significativa es la de $\pm$ 0.5 para arriba (prioridad 1).
b) Carga sólo significativa es entre $\pm$ 0.4 a $\pm$ 0.5 (prioridad 2).
c) Carga sòlo significativa es entre $\pm$ 0.3 a $\pm$ 0.4 (prioridad 3).
d) Si es menor a 0.3 se considera que no existen cargas significativas (prioridad 4).

Utilizando la Prioridad 1 con cargas altamente significativas mayores a 0.5, en el Factor 1 quedarán las variables X_1, X_2 y X_4 con cargas significativas de 0.700, 0.823 y 0.690 respectivamente y en el Factor 2 quedará la variable X_3 con cargas significativa de 0.0.972.

Respuesta al inciso d)

Análisis factorial: X1, X2, X3, X4

Análisis factorial del componente principal de la matriz de correlación

Cargas de factores rotados y comunalidades
Rotación Varimax

Variable	Factor1	Factor2	Comunalidad
X1	**0.700**	-0.314	0.588
X2	**-0.823**	-0.039	0.679
X3	0.118	**-0.972**	**0.959**
X4	**0.690**	-0.130	0.493
Varianza	1.6582	1.0623	2.7205
% Var	0.415	0.266	0.680

Comunalidad

Se usa este término para nombrar a la variación explicada por la solución de factores para cada variable y por estar estandarizadas, el valor máximo posible de la comunalidad es de 1. Mientras más se aproxime el valor de la comunalidad a 1 quiere decir que se explico mejor la variable.

La variable X_3 es la que mejor se identifica con las características por su comunalidad mas alta hasta el Factor 2 de 0.0.959.

$$Z_{11} = \frac{5.50 - 7.12}{2.40} = -0.68 \qquad Z_{12} = \frac{4.90 - 7.06}{1.79} = -1.21$$

2.2.1.1 **EJERCICIO DE REFUERZO**

Persona	Calificación No. 1	Calificación No. 2
1	5.5	4.9
2	9.6	6.3
3	8.1	9.8
4	3.8	7.3
5	8.6	7.0

EJERCICIO DE **REFUERZO 2.2.1.1.1 Cluster Análisis. Medida de Distancia**

Los alumnos de mercadotecnia tienen que hacer una investigación acerca de si trae mejores resultados el trabajo entre dos personas juntas con un parecido en algún aspecto (en este caso calificación) que otras diferentes en tal aspecto.

a) Obtenga la pareja de personas más parecidas y la menos parecida

b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos, simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima

Resolución del inciso a)

Estandarización de variables

Antes que nada hay que se deben estandarizar las variables, es decir, eliminar el efecto de las unidades y hacerlas comparables, de manera que todas estén en términos de "Z" mediante la siguiente expresión:

Persona	Calificación No. 1	Calificación No. 2
1	5.5	4.9
2	9.6	6.3
3	8.1	9.8
4	3.8	7.3
5	8.6	7.0
Promedio	7.12	7.06
Desviación Estándar	2.4	1.79

$$Z = \frac{x - \bar{x}}{s}$$

$$Z_{21} = \frac{9.60 - 7.12}{2.40} = 1.03 \qquad Z_{22} = \frac{6.30 - 7.06}{1.79} = -0.42$$

$$Z_{31} = \frac{8.1 - 7.12}{2.40} = 0.41 \qquad Z_{32} = \frac{9.80 - 7.06}{1.79} = 1.53$$

$$Z_{41} = \frac{3.8 - 7.12}{2.40} = -1.38 \qquad Z_{42} = \frac{7.30 - 7.06}{1.79} = 0.13$$

$$Z_{51} = \frac{8.60 - 7.12}{2.40} = .83 \qquad Z_{52} = \frac{7.00 - 7.06}{1.79} = -0.03$$

Los datos estandarizados de las 5 personas, al considerar 2 variables, se presenta de la siguiente manera:

Persona	Altura estandarizada	Peso estandarizado
1	- 0.68	-1.21
2	1.03	-0.42
3	0.41	1.53
4	- 1.38	0.13
5	0.83	-0.03
Promedio:	0.00	0.00
Desviación estándar	1.00	1.00

Cálculo de las distancias:

$$d_{ij} = \left[\sum_{K=1}^{r} (x_{ik} - x_{jk})^2\right]^{0.5}$$

$$d_{1,2} = [(-0.68 - 1.03)^2 + (-1.21 - (-0.42))^2]^{0.5} = 1.88$$

$$d_{1,3} = [(-0.68 - 0.41))^2 + (-1.21 - 1.53)^2]^{0.5} = 2.94$$

$$d_{1,4} = [(-0.68 - (-1.38))^2 + (-1.21 - 0.13)^2]^{0.5} = 1.52$$

$$d_{1,5} = [(-0.68 - .83)^2 + (-1.21 - (-0.03))^2]^{0.5} = 1.74$$

$$d_{2,3} = [(1.03 - 0.41))^2 + (-0.42 - 1.53)^2]^{0.5} = 2.05$$

$$d_{2,4} = [(1.03 - (-1.38))^2 + (-0.42 - 0.13))^2]^{0.5} = 2.48$$

$$d_{2,5} = [(1.03 - .83)^2 + (-0.42 - (-0.03))^2]^{0.5} = 0.57$$

$$d_{3,4} = [(0.41 - (-1.38))^2 + (1.53 - 0.13))^2]^{0.5} = 2.27$$

$$d_{3,5} = [(-0.41 - 0.83)^2 + (1.53 - (0.03))^2]^{0.5} = 1.57$$

$$d_{4,5} = [(-1.38 - .83)^2 + (0.13 - (-0.03))^2]^{0.5} = 2.01$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5 personas a la que le nombraremos ***matriz de distancia*** la cual representa que tan parecidas son las personas con respecto a la distancia entre ellas.

	1	2	3	4	5
1	0	1.88	2.94	1.52	1.74
2	1.88	0	2.05	2.48	0.57
3	2.94	2.05	0	2.27	1.58
4	1.52	2.48	2.27	0	2.01
5	1.74	0.57	1.58	2.01	0

Como esta matriz es una matriz espejo, es decir la matriz superior es igual a la matriz inferior y la diagonal contiene ceros ya que la distancia que hay al comparar una persona consigo misma es nula, sólo será necesario utilizar la información de la matriz inferior como se muestra a continuación:

	1	2	3	4	5
1	0				
2	1.88	0			
3	2.94	2.05	0		
4	1.52	2.48	2.27	0	
5	1.74	0.57	1.57	2.01	0

Al observar la matriz de distancia se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya distancia entre sí es la menor que está formada por las personas 2 y 5, con una medida de distancia de 0.57 y la menos parecida será aquella cuya distancia entre sí es la mayor que está formada por las personas 1 y 3, con una medida de distancia de 2.94.

Resolución al inciso b)

Vinculación o eslabonamiento:

1. **Eslabonamiento Simple:**

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia menor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.57

	1	2	3	4	5
1	0				
2	1.88	0			
3	2.94	2.05	0		
4	1.52	2.48	2.27	0	
5	1.74	**0.57**	1.57	2.01	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 2 y 5, como se muestra en la siguiente matriz inferior:

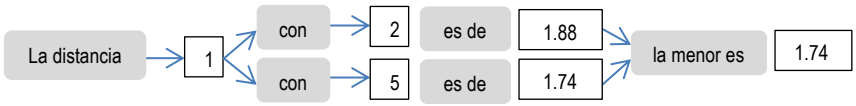
	1	2	3	4	5
1	0				
2	1.88	0			
3	2.94	2.05	0		
4	1.52	2.48	2.27	0	
5	1.74	**0.57**	1.57	2.01	0

El nuevo conglomerado (2,5) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(2,5)	1	3	4
(2,5)	0			
1		0		
3			0	
4				0

Paso 2. Eslabonar

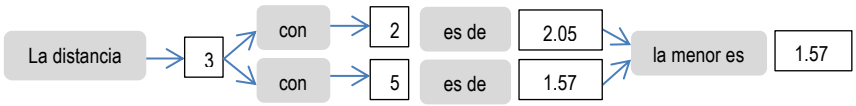
Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al nuevo conglomerado (2,5) con el resto de las personas; al comparar (2,5) con la persona 1 se debe verificar en la matriz anterior las distancias entre la persona 1 con la 2 que es 1.88 y la de la persona 1 con la 5 que es 1.57, eligiendo entonces la distancia menor de las dos personas que en este caso es 1.74. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.74	0		
3			0	
4				0

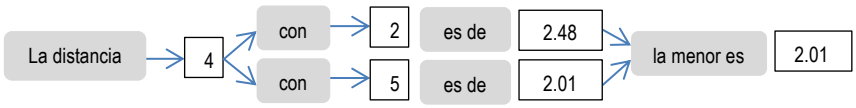
Repetimos el mismo criterio para eslabonar la persona 3 con la pareja de personas (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con 2 que es 2.05 y la de la personas 3 con 5 que es 1.57, eligiendo entonces la distancia menor de las dos personas que en este caso es 1.57. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.74	0		
3	1.57		0	
4				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 4 con la pareja de personas (2,5); al comparar (2,5) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4 con 2 que es 2.48 y la de la personas 4 con 5 que es 2.01, eligiendo entonces la distancia menor de las dos personas que en este caso es 2.01. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.74	0		
3	1.57		0	
4	2.01			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(2,5)	1	3	4
(2,5)	0			
1	1.74	0		
3	1.57	2.94	0	
4	2.01	1.52	2.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, en este caso se forma un segundo conglomerado formado por las personas (1,4) con la distancia menor de 1.52 como se muestra a continuación:

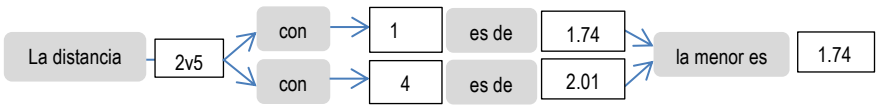
	(2,5)	1	3	4
(2,5)	0			
1	1.74	0		
3	1.57	2.94	0	
4	2.01	**1.52**	2.27	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)		0	
3			0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al primer conglomerado (2,5) con las personas del segundo conglomerado (1,4); al comparar (2,5) con (1,4) la persona 1 se debe verificar en la matriz anterior las distancias con las personas (2 y 5) que es 1.74 y la de la persona 4 con las 2 y 5 que es 2.01, eligiendo entonces la distancia menor de las dos combinaciones o cuatro personas que en este caso es 1.74. Lo anterior se puede resumir con el siguiente diagrama de flujo:



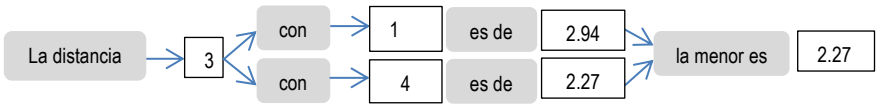
El valor eslabonado se coloca en la matriz reducida

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	1.74	0	
3			0

Repetimos el mismo criterio para eslabonar la persona 4 con el conglomerado (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con las personas 2 y 5 que es 1.57 y se pasa el dato directo en la matriz reducida de la siguiente forma:

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	1.74	0	
3	1.57		0

Repetimos el mismo criterio para eslabonar la persona 3 con el conglomerado (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con las personas 1 y 4. Al comparar a la persona 3 con 1 la distancia es de 2.94 y la persona 3 con 4 la distancia es de 2.27 eligiendo entonces la distancia menor de las dos combinaciones o cuatro personas que en este caso es 2.27 Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	1.74	0	
3	1.57	2.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería las persona 3 que se une al primer conglomerado (2,5) con la distancia menor de 1.74 como se muestra a continuación:

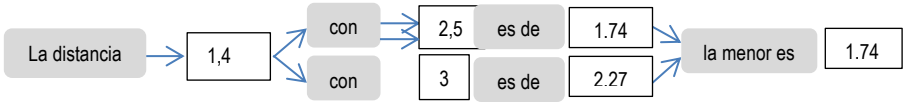
	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	**1.74**	0	
3	1.57	2.27	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(2,3,5)	(1,4)
(2,3,5)	0	
(1,4)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (2,3,5) con las personas del segundo conglomerado (1,4); al comparar (1,4) con las personas 2,3,5) se debe verificar en la matriz anterior las distancias entre la persona 1,4 con las personas (2,5) que es 1.74 y las personas (1,4) con la persona 3 que es 2.27, eligiendo entonces la distancia menor de las dos combinaciones o cinco personas que en este caso es 1.74. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)	**1.74**	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 1.74

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.57	2	5	2
2	3	1.52	1	4	1
3	2	1.57	2	3	2
4	1	1.74	1	2	1

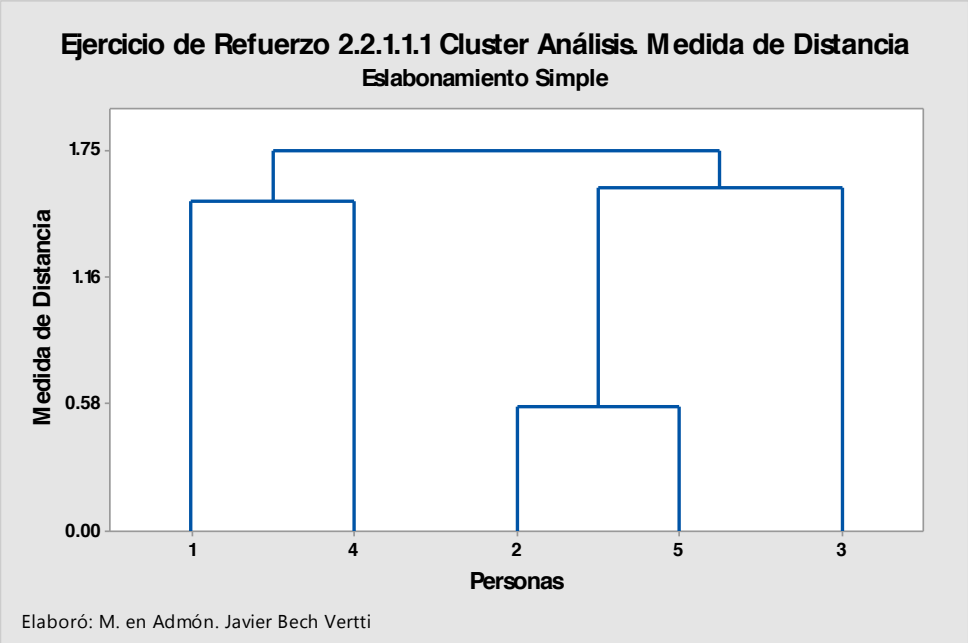
En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La segunda columna no muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 2 y la 5 y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se formó un segundo conglomerado con las personas 1 y 4 a una distancia de 1.52. En la etapa 3 se incorporó la persona 3 al primer conglomerado (2,5) a una distancia de 1.57 y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de 1.74. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Calif 1, Calif 2

```
Variables estandarizadas, Distancia euclediana, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	80.5757	0.57209	2	5	2	2
2	3	48.4844	1.51725	1	4	1	2
3	2	46.4091	1.57838	2	3	2	3
4	1	40.6847	1.74698	1	2	1	5

Dendrograma



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

2. **Eslabonamiento Completo:**

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia mayor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, las personas más parecidas se elegirán con la distancia menor, que en este caso es 0.57

	1	2	3	4	5
1	0				
2	1.88	0			
3	2.94	2.05	0		
4	1.52	2.48	2.27	0	
5	1.74	**0.57**	1.57	2.01	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 2 y 5, como se muestra en la siguiente matriz inferior:

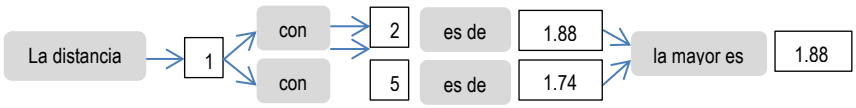
	1	2	3	4	5
1	0				
2	1.88	0			
3	2.94	2.05	0		
4	1.52	2.48	2.27	0	
5	1.74	**0.57**	1.57	2.01	0

El nuevo conglomerado (2,5) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(2,5)	1	3	4
(2,5)	0			
1		0		
3			0	
4				0

Paso 2. Eslabonar

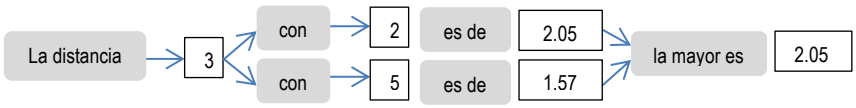
Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al nuevo conglomerado (2,5) con el resto de las personas; al comparar (2,5) con la persona 1 se debe verificar en la matriz anterior las distancias entre la persona 1 con la 2 que es 1.88 y la de la persona 1 con la 5 que es 1.74, eligiendo entonces la distancia mayor de las dos personas que en este caso es 1.88. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.88	0		
3			0	
4				0

Repetimos el mismo criterio para eslabonar a la persona 3 con la pareja de personas (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre las personas 3 con 2 que es 2.05 y la de la personas 3 con 5 que es 1.57, eligiendo entonces la distancia mayor de las dos personas que en este caso es 2.05. Lo anterior se puede resumir con el siguiente diagrama de flujo:

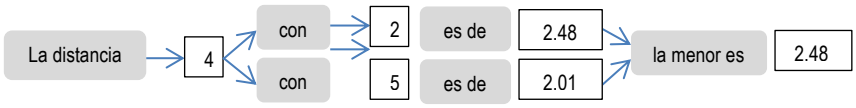


El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.88	0		
3	2.05		0	
4				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 4 con la pareja de personas (2,5); al comparar (2,5) con la persona 4 se debe verificar en la matriz anterior las distancias entre la persona 4 con 2

que es 2.48 y la de la persona 4 con 5 que es 2.01, eligiendo entonces la distancia mayor de las dos personas que en este caso es 2.48. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.88	0		
3	2.05		0	
4	2.48			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(2,5)	1	3	4
(2,5)	0			
1	1.88	0		
3	2.05	2.94	0	
4	2.48	1.52	2.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 1 y 4 con una distancia de 1.52 como se muestra a continuación:

	(2,5)	1	3	4
(2,5)	0			
1	1.88	0		
3	2.05	2.94	0	
4	2.48	**1.52**	2.27	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)		0	
3			0

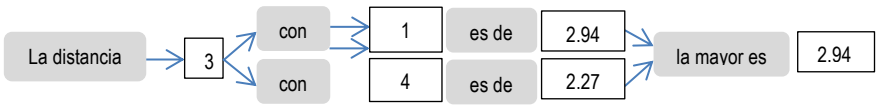
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (2,5) con el resto de las personas; al comparar (2,5) con la persona 1 se debe verificar en la matriz anterior las distancias entre el conglomerado (2,5) con la persona 1 que es 1.88 y al conglomerado (2,5) con la de la persona 4 que es 2.48, eligiendo entonces la distancia mayor de las dos combinaciones o cuatro personas que en este caso es 2.48. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.48	0	
3			0

Repetimos el mismo criterio para eslabonar la persona 3 con el conglomerado (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre la persona 3 con la persona 1 que es 2.94 y la de la persona 3 con la 4 que es 2.27, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 2.94. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.48	0	
3		2.94	0

Repetimos el mismo criterio para eslabonar la persona 3 con el conglomerado (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre la persona 3 con las personas (2 y 5) que ya están y que es 2.05. En este caso el valor se pasa tal cual a la matriz reducida.

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.48	0	
3	2.05	2.94	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona a un conglomerado ya formado, que en este caso sería la persona 3 que se eslabonaría al conglomerado formado por las personas (2,5) con la distancia menor de 2.05 como se muestra a continuación:

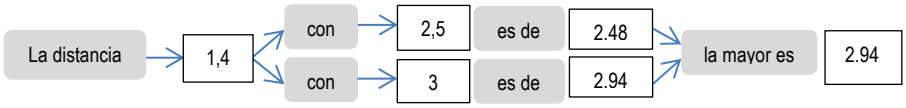
	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.48	0	
3	**2.05**	2.94	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(2,3,5)	(1,4)
(2,3,5)	0	
(1,4)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (2,3,5) con el conglomerado (1,4); al comparar (2,3,5) con el conglomerado (1,4) se debe verificar en la matriz anterior las distancias entre las personas 1 y 4 con las personas (2,5) que es 2.48 y las personas 1 y 4 con la persona 3 que es 2.94, eligiendo entonces la distancia mayor de las dos combinaciones o cinco personas que en este caso es 2.94. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(2,3,5)	(1,4)
(2,3,5)	0	
(1,4)	**2.94**	0

En este punto se detiene el procedimiento llegando a unir a las cinco personas en un solo grupo con una distancia de 2.94.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.57	2	5	2
2	3	1.52	1	4	1
3	2	2.05	2	3	2
4	1	2.94	1	2	1

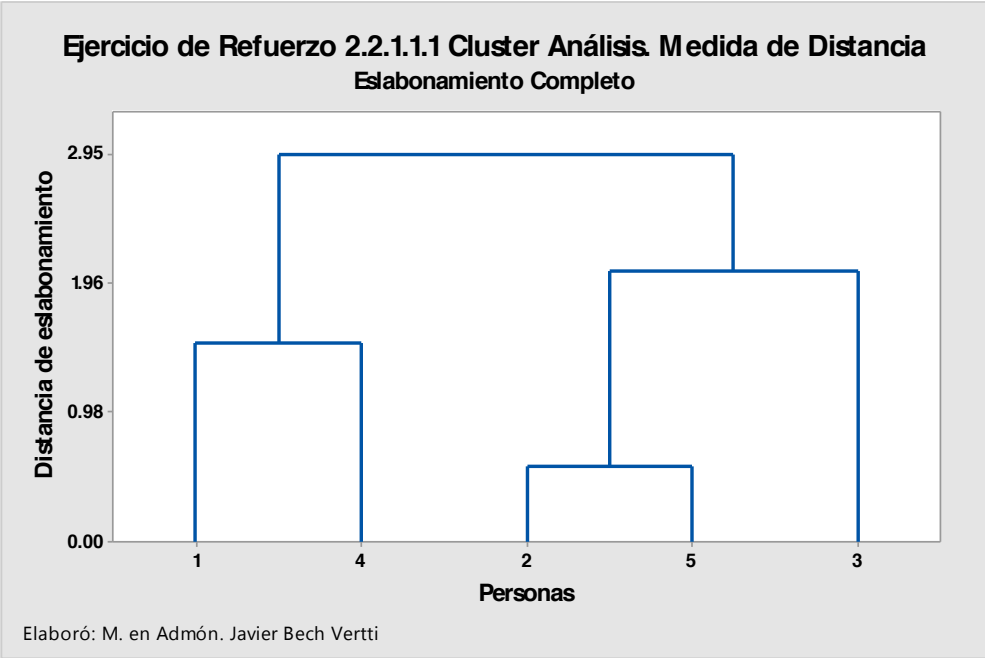
En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan **cuatro etapas**. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **columna 4** que en este caso fueron la 2 y la 5 a una distancia de **0.57** y en la **quinta columna** nos señala la formación del primer conglomerado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 1 y 4 a una distancia de **1.52** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 3 al primer conglomerado formado por las personas 2 y 5 a una distancia de **2.05** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **2.94**. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Calif 1, Calif 2

```
Variables estandarizadas, Distancia euclediana, Enlace completo
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	80.5757	0.57209	2	5	2	2
2	3	48.4844	1.51725	1	4	1	2
3	2	30.2783	2.05347	2	3	2	3
4	1	0.0000	2.94524	1	2	1	5

Dendograma



Elaboró: M. en Admón. Javier Bech Vertti

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (2,3,5) y el conglomerado 2 por las personas (1,4).
Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de

cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

3. Eslabonamiento Promedio:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia promedio, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.57

	1	2	3	4	5
1	0				
2	1.88	0			
3	2.94	2.05	0		
4	1.52	2.48	2.27	0	
5	1.74	**0.57**	1.57	2.01	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 2 y 5, como se muestra en la siguiente matriz inferior:

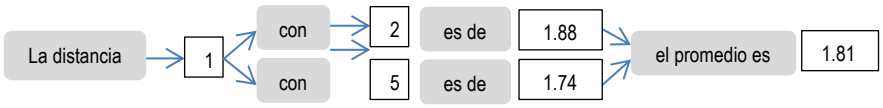
	1	2	3	4	5
1	0				
2	1.88	0			
3	2.94	2.05	0		
4	1.52	2.48	2.27	0	
5	1.74	**0.57**	1.57	2.01	0

El nuevo cluster (2,5) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(2,5)	1	3	4
(2,5)	0			
1		0		
3			0	
4				0

Paso 2. Eslabonar

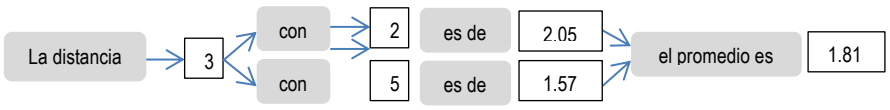
Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al nuevo conglomerado (2,5) con el resto de las personas; al comparar (2,5) con la persona 1 se debe verificar en la matriz anterior las distancias entre la persona 1 con la 2 que es 1.88 y la de la persona 1 con la 5 que es 1.74, eligiendo entonces la distancia promedio de las dos personas que en este caso es 1.81. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.81	0		
3			0	
4				0

Repetimos el mismo criterio para eslabonar la persona 3 con la pareja de personas (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con 2 que es 2.05 y la de la personas 3 con 5 que es 1.57, eligiendo entonces la distancia promedio de las dos personas que en este caso es 1.81. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.81	0		
3	1.81		0	
4				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 4 con la pareja de personas (2,5); al comparar (2,5) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4 con 2 que es 2.48 y la de la personas 4 con 5 que es 2.01, eligiendo entonces la distancia promedio de las dos personas que en este caso es 2.25. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(2,5)	1	3	4
(2,5)	0			
1	1.81	0		
3	1.81		0	
4	2.25			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(2,5)	1	3	4
(2,5)	0			
1	1.81	0		
3	1.81	2.49	0	
4	2.25	1.52	2.27	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 1 y 4 a una distancia menor de 1.52 como se muestra a continuación:

	(2,5)	1	3	4
(2,5)	0			
1	1.81	0		
3	1.81	2.49	0	
4	2.25	**1.52**	2.27	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)		0	
3			0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado(2,5) con las personas del conglomerado (1,4); al comparar (2,5) con la persona 1 se debe verificar en la matriz anterior las distancias entre las persona (2,5) y la persona 1 que es 1.81 y la de las personas (2,5) con la persona 4 que es 2.25, eligiendo entonces la distancia promedio de las dos combinaciones o cuatro personas que en este caso es 2.03. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.03	0	
3			0

Repetimos el mismo criterio para eslabonar a la persona 3 con el conglomerado (1,4); al comparar (1,4) con la persona 3 se debe verificar en la matriz anterior las distancias entre las persona 3 con la persona 1 que es 2.49 y la de la personas 3 con la 4 que es 2.27, eligiendo entonces la distancia promedio de las dos combinaciones o tres personas que en este caso es 2.38. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.03	0	
3		2.38	0

Repetimos el mismo criterio para eslabonar a la persona 3 con el conglomerado (2,5); al comparar (2,5) con la persona 3 se debe verificar en la matriz anterior la distancia entre las persona 3 con las persona 2 y 5 que es 1.81 pasando el dato directamente en la matriz reducida :

El valor eslabonado se coloca en la matriz reducida

	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.03	0	
3	1.81	2.38	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 3 para eslabonarse al conglomerado formado por las personas (1,4) con la distancia menor de 1.33 como se muestra a continuación:

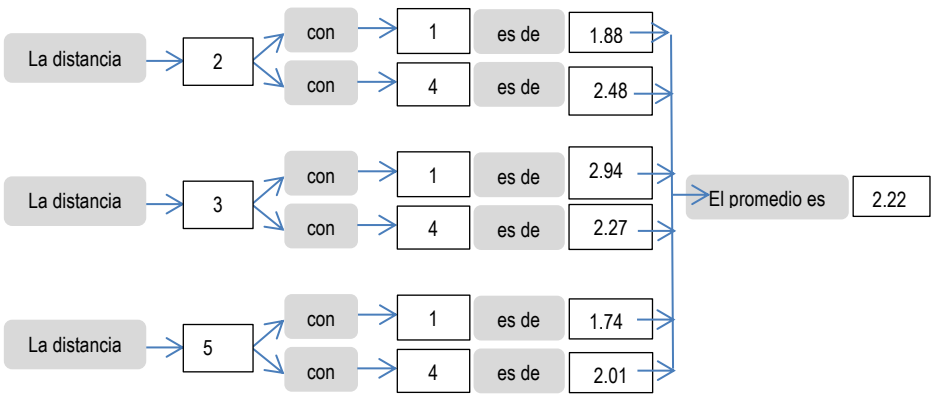
	(2,5)	(1,4)	3
(2,5)	0		
(1,4)	2.03	0	
3	**1.81**	2.38	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(2,3,5)	(1,4)
(2,3,5)	0	
(1,4)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado (2,3,5) con el resto de las personas del conglomerado (1,4); al comparar (2,3,5) con las personas (1,4) se debe verificar **por única vez en la primera matriz de distancias** entre las personas 1,4 y 3 con las personas 2,5, eligiendo entonces la distancia promedio de las seis combinaciones o cinco personas que en este caso es 3.02. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(2,3,5)	(1,4)
(2,3,5)	0	
(1,4)	**2.22**	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 2.22.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

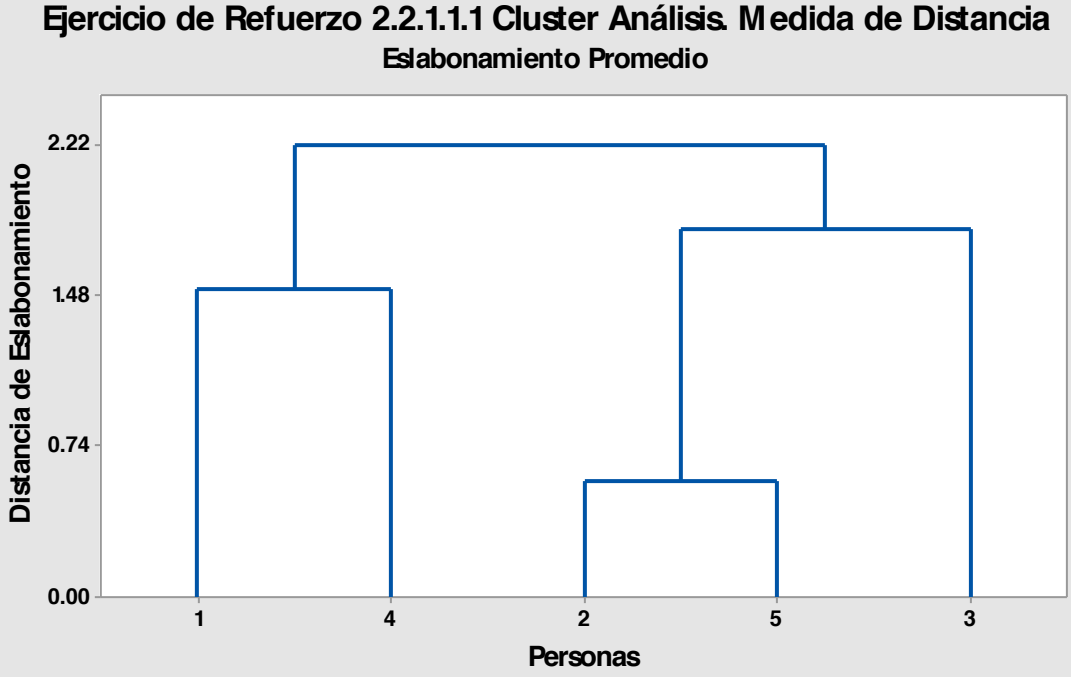
Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.57	1	4	1
2	3	1.52	2	5	2
3	2	1.81	1	3	1
4	1	2.22	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **cuarta columna** que en este caso fueron la 2 y la 5 a una distancia de **0.57** y en la **quinta columna** nos señala el número de conglomerado formado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 1 y 4 a una distancia de **1.52** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 3 al primer conglomerado formado por las personas 2 y 5 a una distancia de **1.81** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **2.22** Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Calif 1, Calif 2

```
Variables estandarizadas, Distancia euclediana, Enlace de promedios
Pasos de amalgamación


                                                                  Número de obs. en
          Número de    Nivel de   Nivel de  Conglomerados        Nuevo    el conglomerado
Paso  conglomerados  semejanzal  distancia   incorporados  conglomerado             nuevo
   1             4     80.5757    0.57209  2           5             2                 2
   2             3     48.4844    1.51725  1           4             1                 2
   3             2     38.3437    1.81592  2           3             2                 3
   4             1     24.4821    2.22418  1           2             1                 5
```

Dendrograma



Elaboró: M. en Admón. Javier Bech Vertti

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

2.2.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 2.2.1.1.2 Cluster Análisis. Medida de Distancia**

2.2.1.1.2. Una compañía de investigación de mercados le solicitan realizar un estudio sobre algunas marcas de cierto producto del mercado para lo cual se analizan tres variables obtenidas de un análisis de factores. Las variables que se han identificado más con cada uno de los factores son las siguientes: Con el factor 1, la variable X_1; con el factor 2, la variable X_2 y con el factor 3, la variable X_3.

Suponga que dichos factores se utilizarán para realizar la segmentación de mercado, para lo cual se proporcionan los puntajes de 5 entrevistas:

ENTREVISTA	**VARIABLE**		
	X_1	**X_2**	**X_3**
1	1.35	2.70	1.35
2	6.75	9.45	6.75
3	2.70	1.35	1.35
4	9.45	9.45	6.75
5	4.05	2.70	2.70
Promedio:	4.86	5.13	3.78
Desviación estándar	3.25	3.98	2.77

a) Obtenga la pareja de personas entrevistadas más parecidas y la menos parecida.
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.
c) Proponga un nombre tentativo a cada uno de los conglomerados conforme a la solución establecida.

Resolución del inciso a)

Estandarización de variables

Antes que nada hay que se deben estandarizar las variables, es decir, eliminar el efecto de las unidades y hacerlas comparables, de manera que todas estén en términos de "Z" mediante la siguiente expresión:

$$Z = \frac{x - \bar{x}}{s}$$

$$Z_{11} = \frac{1.35 - 4.86}{3.25} = -1.08 \qquad Z_{12} = \frac{2.70 - 5.13}{3.98} = -0.61 \qquad Z_{13} = \frac{1.35 - 3.78}{2.77} = -0.88$$

$$Z_{21} = \frac{6.75 - 4.86}{3.25} = 0.58 \qquad Z_{22} = \frac{9.45 - 5.13}{3.98} = 1.08 \qquad Z_{23} = \frac{6.75 - 3.78}{2.77} = 1.07$$

$$Z_{31} = \frac{2.70 - 4.86}{3.25} = -0.66 \qquad Z_{32} = \frac{1.35 - 5.13}{3.98} = -0.95 \qquad Z_{33} = \frac{1.35 - 3.78}{2.77} = -0.88$$

$$Z_{41} = \frac{9.45 - 4.86}{3.25} = 1.41 \qquad Z_{42} = \frac{9.45 - 5.13}{3.98} = 1.08 \qquad Z_{43} = \frac{6.75 - 3.78}{2.77} = 1.07$$

$$Z_{51} = \frac{4.05 - 4.86}{3.25} = -0.25 \qquad Z_{52} = \frac{2.70 - 5.13}{3.98} = -0.61 \qquad Z_{53} = \frac{2.70 - 3.78}{2.77} = -0.39$$

Los datos estandarizados de las 5 personas, al considerar 3 variables, se presenta de la siguiente manera:

Entrevista	V1	V2	V3
1	-1.08	-0.61	-0.88
2	0.58	1.08	1.07
3	-0.66	-0.95	-0.88
4	1.41	1.08	1.07
5	-0.25	-0.61	-0.39
Promedio:	0.00	0.00	0.00
Desviación estándar	1.00	1.00	1.00

Cálculo de las distancias:

$$d_{ij} = \left[\sum_{K=1}^{r} (x_{ik} - x_{jk})^2\right]^{0.5}$$

$$d_{1,2} = [(-1.08 - 0.58)^2 + (-.61 - 1.08)^2 + (-0.88 - 1.07)^2]^{0.5} = 3.07$$

$$d_{1,3} = 0.54$$

$$d_{1,4} = 3.59$$

$$d_{1,5} = 0.96$$

$$d_{2,3} = 3.08$$

$$d_{2,4} = 0.83$$

$$d_{2,5} = 2.39$$

$$d_{3,4} = 3.50$$

$$d_{3,5} = 0.72$$

$$d_{4,5} = 2.79$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5 entrevistas a la que le nombraremos ***matriz de distancia*** la cual representa que tan parecidas son las personas entrevistadas con respecto a la distancia entre ellas.

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Como esta matriz es una matriz espejo, es decir la matriz superior es igual a la matriz inferior y la diagonal contiene ceros ya que la distancia que hay al comparar una persona consigo misma es nula, sólo será necesario utilizar la información de la matriz inferior como se muestra a continuación:

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Al observar la matriz de distancia se puede entonces contestar lo planteado en el inciso a) con respecto a las personas entrevistadas más parecidas y las menos parecidas, en este caso la pareja de personas entrevistadas más parecidas es aquella cuya distancia entre sí es la menor que está formada por las personas 1 y 3, con una medida de distancia de 0.54 y la menos parecida será aquella cuya distancia entre sí es la mayor que está formada por las personas 1 y 4, con una medida de distancia de 3.59.

Resolución al inciso b)

Vinculación o eslabonamiento:

1. **Eslabonamiento Simple:**

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia menor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.54

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

El nuevo conglomerado (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

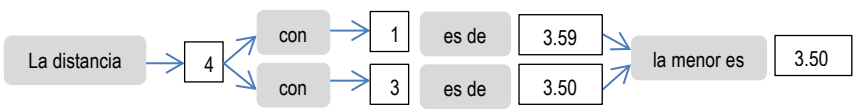
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al nuevo conglomerado (1,4) con el resto de las personas; al comparar (1,4) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.11 y la de la persona 2 con la 4 que es 3.06, eligiendo entonces la distancia menor de las dos personas que en este caso es 3.06. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4			0	
5				0

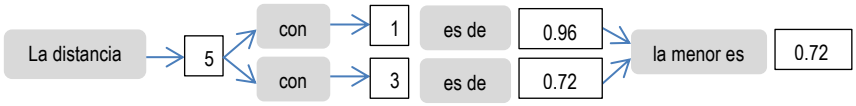
Repetimos el mismo criterio para eslabonar la persona 4 con la pareja de personas (1,3); al comparar (1,3 con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4con 1 que es 3.59 y la de la personas 4 con 3 que es 3.50, eligiendo entonces la distancia menor de las dos personas que en este caso es 3.50. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 0.96 y la de la personas 5 con 3 que es 0.72, eligiendo entonces la distancia menor de las dos personas que en este caso es 0.72. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50		0	
5	0.72			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50	0.83	0	
5	0.72	2.39	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 5 que se eslabonaría al conglomerado formado por las personas (1,3) con la distancia menor de 0.72 como se muestra a continuación:

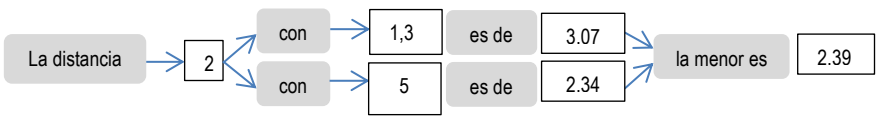
	(1,3)	2	4	5
(1,3)	0			
2	3.07	0		
4	3.50	0.83	0	
5	0.72	2.39	2.79	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3,5)	2	4
(1,3,5)	0		
2		0	
4			0

Paso 2. Eslabonar

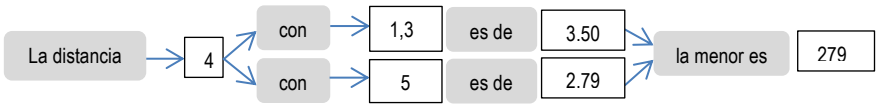
Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (1,3,5) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las 1 y 3 que es 3.07 y la de la persona 2 con la 5 que es 2.39, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.39. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4			0

Repetimos el mismo criterio para eslabonar la persona 4 con el conglomerado (1,3,5); al comparar (1,3,5) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4con las personas 1 y 3 que es 3.50 y la de la personas 4 con la 5 que es 2.79, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.79. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4	2.79		0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4	2.79	0.83	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso serían las personas 2 y 4 para formar un nuevo conglomerado con la distancia menor de 0.83 como se muestra a

continuación:

	(1,3,5)	2	4
(1,3,5)	0		
2	2.39	0	
4	2.79	0.83	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento simple o del vecino más cercano, al momento de eslabonar se debe tener cuidado de elegir la distancia menor, es decir se compara al conglomerado (1,3,5) con el resto de las personas; al comparar (1,3,5) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con las personas (1,3,5) que es 2.39 y la de la persona 4 con las (1,3,5) que es 2.79, eligiendo entonces la distancia menor de las dos combinaciones o tres personas que en este caso es 2.39. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)	2.39	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 2.39.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.54	1	3	1
2	3	0.72	1	5	1
3	2	0.83	2	4	2
4	1	2.39	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La segunda columna no muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 1 y la 3 y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 5 al primer conglomerado formado por las personas 1 y 3 como se muestra en la columna 4 a una distancia de 0.72. En la etapa 3 se forma un nuevo conglomerado como se aprecia en la columna 5 formado por las personas 2 y 4 a una distancia de 0.83 y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de 2.39. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: X1, X2, X3

```
Variables estandarizadas, Distancia euclediana, Enlace simple
Pasos de amalgamación
```

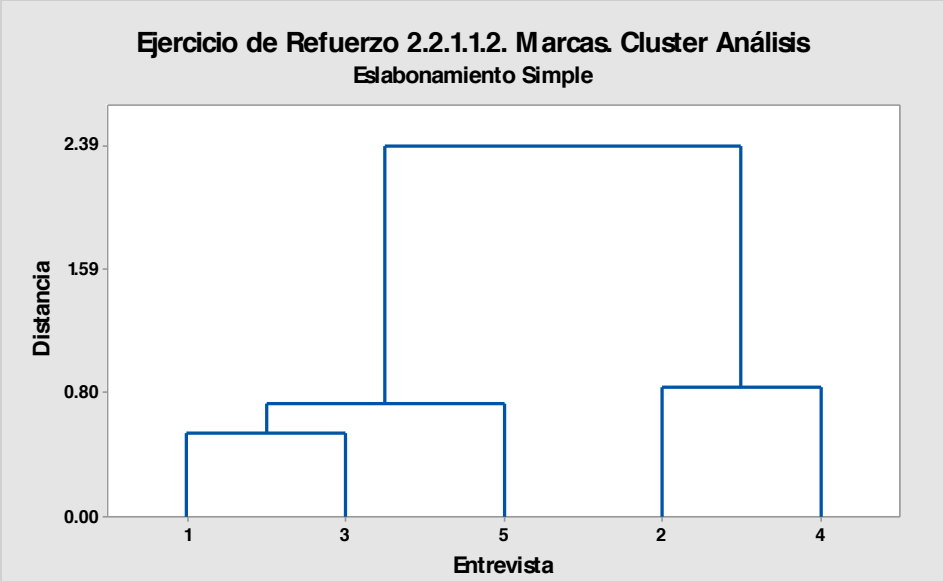
Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	85.0692	0.53606	1	3	1	2
2	3	79.8098	0.72488	1	5	1	3
3	2	76.8692	0.83045	2	4	2	2
4	1	33.4662	2.38874	1	2	1	5

Dendrograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,5) y el conglomerado 2 por las personas (2,4).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

2. Eslabonamiento Completo:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia mayor, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.54

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

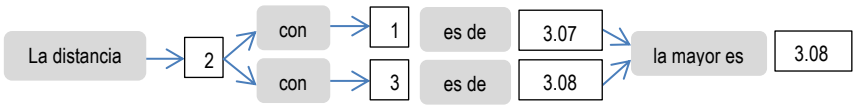
	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

El nuevo conglomerado (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al nuevo conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.07 y la de la persona 2 con la 3 que es 3.08, eligiendo entonces la distancia mayor de las dos personas que en este caso es 3.08. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4			0	
5				0

Repetimos el mismo criterio para eslabonar a la persona 4 con la pareja de personas (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre las personas 4 con 1 que es 3.59 y la de la personas 4 con 3 que es .50, eligiendo entonces la distancia mayor de las dos personas que en este caso es 3.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con 1 que es 0.96 y la de la personas 5 con 3 que es 0.72, eligiendo entonces la distancia mayor de las dos personas que en este caso es 0.96. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59		0	
5	0.96			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59	0.83	0	
5	0.96	2.39	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 2 y 5 con una distancia de 1.27 como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2	3.08	0		
4	3.59	**0.83**	0	
5	0.96	2.39	2.79	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)		0	
5			0

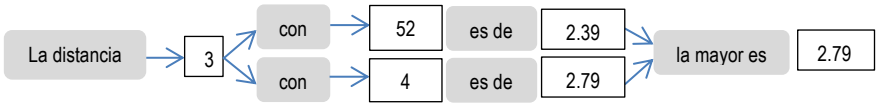
Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre el conglomerado (1,3) con la persona 2 que es 3.08 y al conglomerado (1,3) con la de la persona 4 que es 3.59, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 3.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5			0

Repetimos el mismo criterio para eslabonar la persona 5 con el conglomerado (2,4); al comparar (2,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre la persona 5 con la persona 2 que es 2.39 y la de la persona 5 con la 4 que es 2.79, eligiendo entonces la distancia mayor de las dos combinaciones o tres personas que en este caso es 2.79. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5		2.79	0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz reducida anterior donde ya está el valor de 5 con 1 y 3 que es 0.96

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5	0.96	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona a un conglomerado ya formado, que en este caso sería la persona 3 que se eslabonaría al conglomerado formado por las personas (1,4) con la distancia menor de 1.62 como se muestra a continuación:

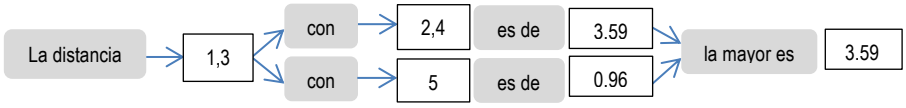
	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.59	0	
5	**0.96**	2.79	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento completo o del vecino más lejano, al momento de eslabonar se debe tener cuidado de elegir la distancia mayor, es decir se compara al conglomerado (1,3,5) con el conglomerado (2,4); al comparar (1,3,5) con el conglomerado (2,4) las personas 2 y 4 se debe verificar en la matriz anterior las distancias entre las personas 2 y 4 con las personas (1,3) que es 3.59 y las personas 1 y 3 con la persona 5 que es 0.96, eligiendo entonces la distancia mayor de las dos combinaciones o cinco personas que en este caso es 3.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,3,5)	(2,4)
(1,3,5)	0	
(2,4)	**3.59**	0

En este punto se detiene el procedimiento llegando a unir a las cinco personas en un solo grupo con una distancia de 3.59.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.53	1	3	1
2	3	0.83	2	4	2
3	2	0.96	1	5	1
4	1	3.59	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan **cuatro etapas**. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **columna 4** que en este caso fueron la 1 y la 3 a una distancia de **0.53** y en la **quinta columna** nos señala la formación del primer conglomerado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 4 a una distancia de **0.83** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 5 al primer conglomerado formado por las personas 1 y 4 a una distancia de **0.96** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **3.59**. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: X1, X2, X3

```
Variables estandarizadas, Distancia euclediana, Enlace completo
Pasos de amalgamación
```

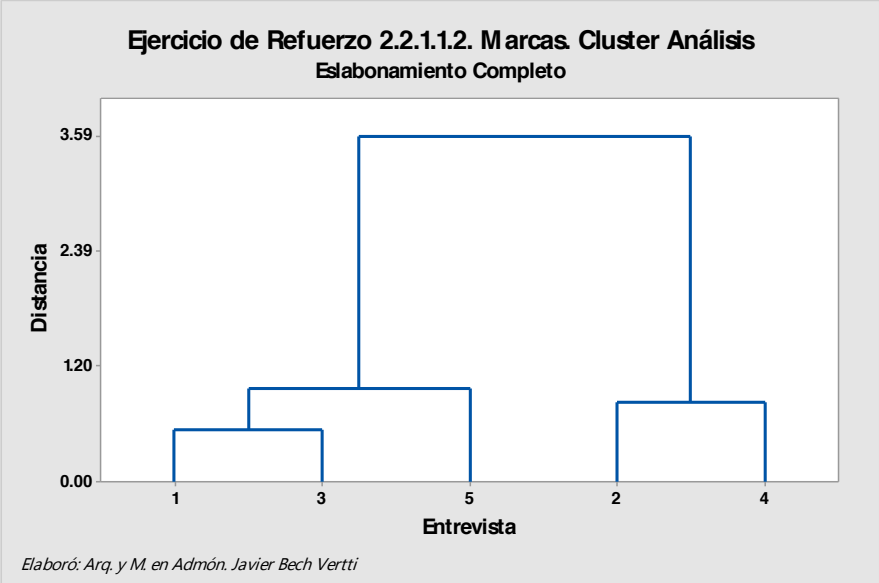
Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	85.0692	0.53606	1	3	1	2
2	3	76.8692	0.83045	2	4	2	2
3	2	73.1719	0.96320	1	5	1	3
4	1	0.0000	3.59026	1	2	1	5

Dendrograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,5) y el conglomerado 2 por las personas (2,4).

Cabe mencionar que mientras a mayor distancia se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

3. Eslabonamiento Promedio:

Una vez elegido el tipo de eslabonamiento que se va a aplicar, se realizarán los siguientes pasos:

Paso 1. Agrupar. Reducir la matriz con la distancia menor, es decir elegir los elementos más parecidos para formar el primer conglomerado.

Paso 2. Eslabonar. Llenar la matriz con la distancia promedio, o sea contrastar las distancias de los elementos agrupados con el resto de los elementos.

Paso 1. Agrupar

Como en este ejemplo se utilizan distancias, los elementos más parecidos se elegirán con la distancia menor, que en este caso es 0.54

	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

Una vez seleccionada la distancia de los elementos que se van a agrupar, se determina quiénes son las personas; al seguir por renglón y columna vemos que son las personas 1 y 3, como se muestra en la siguiente matriz inferior:

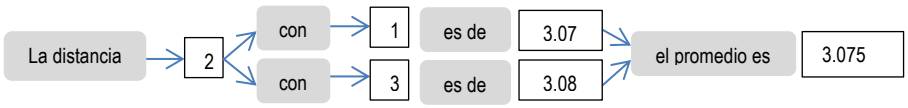
	1	2	3	4	5
1	0				
2	3.07	0			
3	0.54	3.08	0		
4	3.59	0.83	3.50	0	
5	0.96	2.39	0.72	2.79	0

El nuevo conglomerado (1,3) se pone en primer lugar dentro de la nueva matriz, cuyas dimensiones ahora son reducidas de 5 x 5 a 4 x 4, como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2		0		
4			0	
5				0

Paso 2. Eslabonar

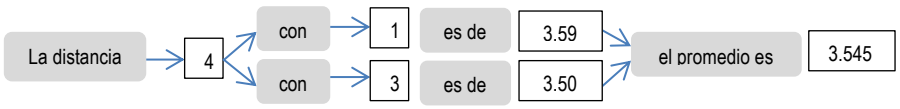
Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al nuevo conglomerado (1,3) con el resto de las personas; al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre la persona 2 con la 1 que es 3.07 y la de la persona 2 con la 3 que es 3.08, eligiendo entonces la distancia promedio de las dos personas que en este caso es 3.075. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4			0	
5				0

Repetimos el mismo criterio para eslabonar la persona 4 con la pareja de personas (1,3); al comparar (1,3) con la persona 4 se debe verificar en la matriz anterior las distancias entre las persona 4 con 1 que es 3.59 y la de la personas 4 con 3 que es 3.50, eligiendo entonces la distancia promedio de las dos personas que en este caso es 3.545. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545		0	
5				0

Utilizamos finalmente el mismo criterio para eslabonar la persona 5 con la pareja de personas (1,3); al comparar (1,3) con la persona 5 se debe verificar en la matriz anterior las distancias entre la persona 5 con 1 que es 0.96 y la de la persona 5 con 3 que es 0.72, eligiendo entonces la distancia promedio de las dos personas que en este caso es 0.84. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545		0	
5	0.84			0

Los valores faltantes de la matriz se llenan con los mismos datos de la matriz original.

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545	0.83	0	
5	0.84	2.39	2.79	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado. En este caso se forma un nuevo conglomerado con las personas 2 y 4 a una distancia menor de 0.83 como se muestra a continuación:

	(1,3)	2	4	5
(1,3)	0			
2	3.075	0		
4	3.545	**0.83**	0	
5	0.84	2.39	2.79	0

La nueva matriz se reduce de 4 x 4 a una de 3 x 3 de la siguiente manera:

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)		0	
5			0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado(1,3) con las personas del conglomerado (2,4); al comparar (1,3) con la persona 2 se debe verificar en la matriz anterior las distancias entre las persona (1,3) y la persona 2 que es 3.075 y la de las personas (1,3) con la persona 4 que es 3.545, eligiendo entonces la distancia promedio de las dos combinaciones o cuatro personas que en este caso es 3.31. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5			0

Repetimos el mismo criterio para eslabonar a la persona 5 con el conglomerado (2,4); al comparar (2,4) con la persona 5 se debe verificar en la matriz anterior las distancias entre las persona 5 con las personas 2 que es 2.39 y la de la personas 5 con la 4 que es 2.79, eligiendo entonces la distancia promedio de las dos combinaciones o tres personas que en este caso es 2.59. Lo anterior se puede resumir con el siguiente diagrama de flujo:

El valor eslabonado se coloca en la matriz reducida

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5		2.59	0

Como el valor del conglomerado (1,3) Con la persona 5 ya esta en la matriz inicial con un valor de 0.84 se pasa tal cual a la matriz reducida:

	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5	0.84	2.59	0

Paso 1. Agrupar

Una vez llena la matriz se repiten los dos pasos de **agrupar y eslabonar** ya sea para formar un nuevo conglomerado o para eslabonar una persona al conglomerado ya formado, que en este caso sería la persona 3 para eslabonarse al conglomerado formado por las personas (1,4) con la distancia menor de 1.33 como se muestra a continuación:

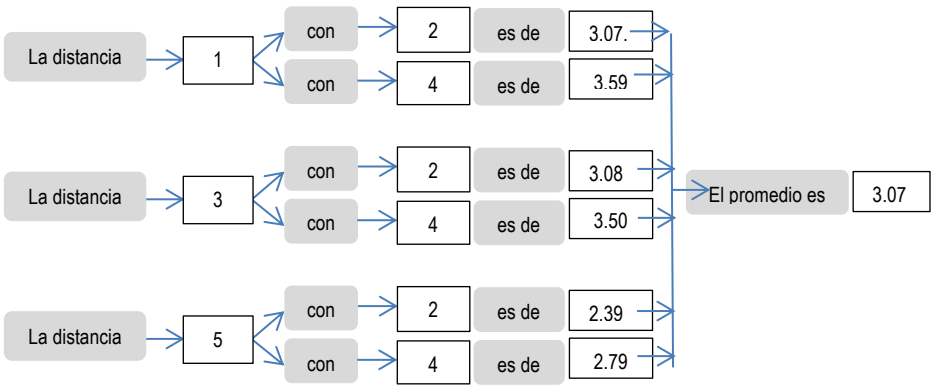
	(1,3)	(2,4)	5
(1,3)	0		
(2,4)	3.31	0	
5	0.84	2.59	0

La nueva matriz se reduce de 3 x 3 a una de 2 x 2 de la siguiente manera:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)		0

Paso 2. Eslabonar

Como se eligió el criterio de vinculación o eslabonamiento promedio, al momento de eslabonar se debe tener cuidado de elegir la distancia promedio, es decir se compara al conglomerado (1,3,5) con el resto de las personas del conglomerado (2,4); al comparar (2,4) con las personas (1,3,5) se debe verificar **por única vez en la primera matriz de distancias** entre las personas 1,3 y 5 con las personas 2,4, eligiendo entonces la distancia promedio de las seis combinaciones o cinco personas que en este caso es 3.07. Lo anterior se puede resumir con el siguiente diagrama de flujo:



El valor eslabonado se coloca en la matriz reducida:

	(1,4,3)	(2,5)
(1,4,3)	0	
(2,5)	3.07	0

En este punto se detiene el procedimiento llegando a unir las cinco personas en un solo grupo con una distancia de 3.07.

Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de distancia	Personas incorporadas		Nuevo conglomerado
1	4	0.54	1	3	1
2	3	0.83	2	4	2
3	2	0.84	1	5	1
4	1	3.07	1	2	1

En la tabla anterior se presenta en la **primera columna** el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La **segunda columna** nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La **tercera columna** nos muestra el valor de la distancia a la que se eslabonaron las dos primeras personas presentadas en la **cuarta columna** que en este caso fueron la 1 y la 3 a una distancia de **0.54** y en la **quinta columna** nos señala el número de conglomerado formado. En la etapa 2 se formó un segundo conglomerado eslabonando a las personas 2 y 4 a una distancia de **0.83** como se muestra en quinta columna. En la etapa 3 se eslabonó a la persona 3 al primer conglomerado formado por las personas 1 y 3 a una distancia de **0.84** y finalmente en la etapa 4 se eslabonaron ambos conglomerados a una distancia final de **3.07** Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que sólo se tomaron dos decimales para desarrollar el problema a mano, los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: X1, X2, X3

```
Variables estandarizadas, Distancia euclediana, Enlace de promedios
Pasos de amalgamación
```

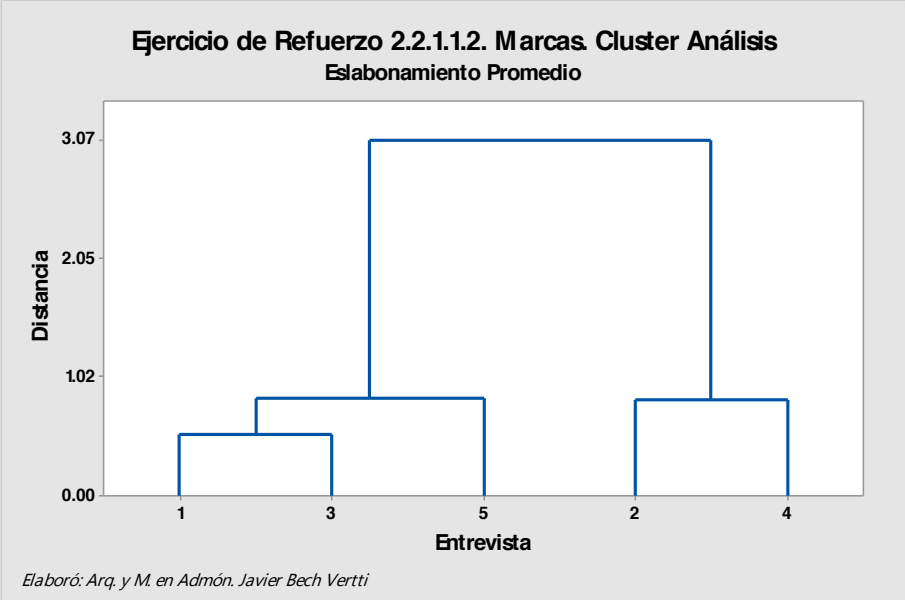
Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	85.0692	0.53606	1	3	1	2
2	3	76.8692	0.83045	2	4	2	2
3	2	76.4909	0.84404	1	5	1	3
4	1	14.4756	3.07055	1	2	1	5

Dendrograma

Dendograma

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el valor de distancia a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso el conglomerado 1 está formado por las personas (1,3,5) y el conglomerado 2 por las personas (2,4).



Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de distancia que requiere dentro de los conglomerados. Examine los niveles de distancia en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La distancia de eslabonamiento en cualquier etapa es el valor de la distancia mínima en la etapa, en relación con la distancia máxima entre observaciones dentro de los datos. El patrón de cómo los valores de distancia cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

Respuesta al inciso c)

Nombre de los conglomerados

Una vez elegidos el número de conglomerados definitivo, para finalizar el proceso es deseable colocar un nombre apropiado a los conglomerados. En este caso el conglomerado 1 está formado por las personas entrevistadas 1,3 y 5 y el conglomerado 2 está formado por las personas entrevistadas 2 y 4. Para conocer el tipo de personas entrevistadas que se encuentran en cada conglomerado se deben obtener la media de los datos originales de quienes la conforman variable por variable como se muestra a continuación:

Para el conglomerado 1:

- La media de la variable X_1 es de: (1.35+2.70+4.05)/3=2.70 Ptos.
- La media de la variable X_2 es de: (2.70+1.35+2.70)/3=2.25 Ptos.
- La media de la variable X_3 es de: (1.35+1.35+2.70)/3=1.80 Ptos.

Para el conglomerado 2:

- La media de la variable X_1 es de: (6.75+9.45)/2=8.1 Ptos.

- La media de la variable X_2 es de: (9.45+9.45)/2=9.45 Ptos.
- La media de la variable X_3 es de: (6.75+6.75)/2=6.75 Ptos.

Los resultados del conglomerado 1 nos indican que las personas entrevistadas de este grupo están nada de acuerdo con las características de las marcas que se proponen que las personas entrevistadas del conglomerado 2 por lo que un posible nombre para el conglomerado 1 podría ser "Nada de Acuerdo" y para el conglomerado 2 "Muy de Acuerdo".

2.2.1.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 2.2.1.2.1 Cluster Análisis. Medida de Similitud**

Con el fin de ser más objetivos en el resultado, los alumnos de mercadotecnia agregaron otra forma de evaluar las calificaciones como *acreditado* o no *acreditado*. Recodifique usted los datos y con base en esto, obtenga nuevamente cuál es la pareja mas parecida a través de una medida de similitud.

Persona	Calificación No. 1	Calificación No. 2	Calificación No. 3
1	5.5	4.9	7.3
2	9.6	6.3	6.8
3	8.1	9.8	9.1
4	3.8	7.3	8.2
5	8.6	7.0	8.3

Criterio de similitud para recodificar los datos:

Calificación No. 1: Acreditado (1) si es ≥ a 7.
No Acreditado (0) si es < a 7.

Calificación No. 2: Acreditado (1) si es ≥ a 7.
No acreditado (0) si es < a 7.

Calificación No. 3: Acreditado (1) si es ≥ a 7.
No Acreditado (0) si es < a 7.

a) Obtenga la(s) pareja(s) de personas más parecidas y la(s) menos parecida(s).
b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

Resolución al inciso a)

Los datos recodificados quedaron como sigue:

Persona	Calificación No. 1	Calificación No. 2	Calificación No. 3
1	0	0	1
2	1	0	0
3	1	1	0
4	0	1	1
5	1	1	1

Cálculo de las similitudes a encontrar:

El coeficiente de similitud que se usará para resolver este ejemplo es el de parejas simples. El coeficiente es:

$$S_{ij} = \frac{a+d}{p}$$

Se debe cuantificar la similitud entre las personas con base en este coeficiente, donde ***a*** es la presencia del atributo en ambas personas, ***d*** es la ausencia del atributo en ambas personas y ***p*** es el total de características (o variables) que serán evaluadas.

Se empieza a comparar por pareja de personas, en este caso la persona 1 y la 2, como en ninguna de las características hay presencia del atributo en ambas (***a***) ni ausente en ambas (***d***), entonces tanto ***a*** como ***d*** tienen valor de 0 en el coeficiente, en cuanto a ***p*** tenemos que son tres características para evaluar (Calificación No. 1, Calificación No. 2, Calificación No. 3); la similitud entre las personas 1 y 2 es de 1/3 como a continuación se presenta:

$$S_{1.2} = \frac{0+1}{3} = 1/3$$

De igual forma se procede con las personas 1 y 3, donde hay cero atributos presentes (***a***) en ambos y cuatro atributos ausentes (***d***) en ambos en todas las características; por lo tanto, la similitud entre las personas 1 y 3 es de 1 como a continuación se presenta:

$$S_{1.3} = \frac{1+0}{3} = 1/3$$

Y así sucesivamente, se obtienen las medidas de similitud para todas las parejas que se pretende comparar obteniendo los siguientes resultados:

$$S_{1.2} = \frac{0+1}{3} = 1/3 \qquad S_{2.3} = \frac{1+0}{3} = 1/3 \qquad S_{3.5} = \frac{3+0}{3} = 1$$

$$S_{1.3} = \frac{1+0}{3} = 1/3 \qquad S_{2.4} = \frac{0+0}{3} = 0 \qquad S_{4.5} = \frac{2+0}{3} = 2/3$$

$$S_{1.4} = \frac{1+1}{3} = 2/3 \qquad S_{2.5} = \frac{1+0}{3} = 1/34$$

$$S_{1.5} = \frac{1+0}{3} = 1/3 \qquad S_{3.4} = \frac{2+0}{3} = 2/3$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5

personas a la que le nombraremos ***matriz de similitud*** la cual representa que tan similares son las personas con respecto a la similitud entre ellas:

	1	2	3	4	5
1	1				
2	1/3	1			
3	1/3	1/3	1		
4	2/3	**0**	2/3	1	
5	1/3	1/3	**1**	2/3	1

Al observar la matriz de similitud se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya similitud entre sí es la que se acerque más o sea igual a 1, en este caso tenemos una pareja formada por las personas 3 y 5 con una medida de similitud de 1.

La pareja de personas menos parecida será aquella cuya similitud entre sí es la se acerque más o sea igual a 0. En este caso tenemos una pareja formada por las personas 2 y 4 con una medida de similitud de 0.

Resolución al inciso b)

Vinculación o eslabonamiento:

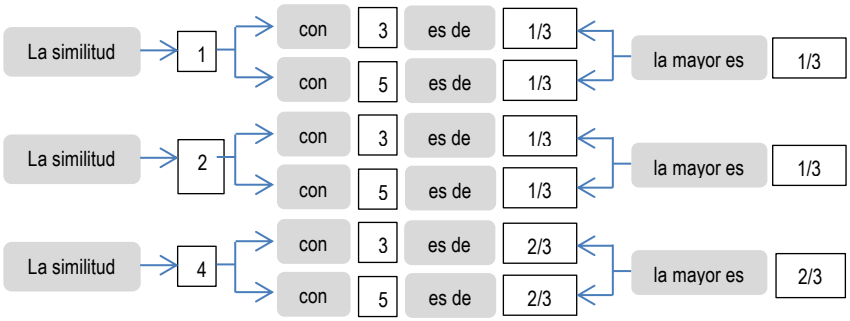
1. **Eslabonamiento Simple:**

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	1	2	3	4	5
1	1				
2	1/3	1			
3	1/3	1/3	1		
4	2/3	0	2/3	1	
5	1/3	1/3	**1**	2/3	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,5)	1	2	4
(3,5)	1			
1	1/3	1		
2	1/3	13	1	
4	2/3	2/3	0	1



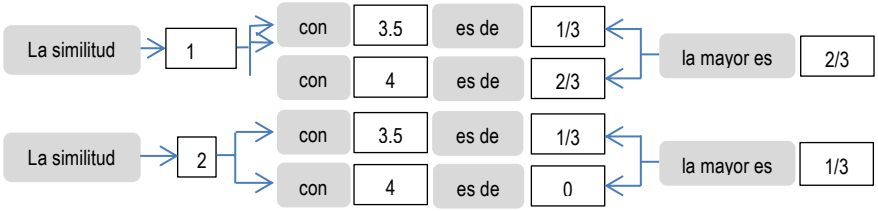
Los demás valores se pasan tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,5)	1	2	4
(3,5)	1			
1	1/3	1		
2	1/3	13	1	
4	**2/3**	2/3	0	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,4,5)	1	2
(3,4,5)	1		
1	2/3	1	
2	1/3	1/3	1



El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,4,5)	1	2
(3,4,5)	1		
1	**2/3**	1	
2	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(1,3,4,5)	2
(1,3,4,5)	1	
2	1/3	1

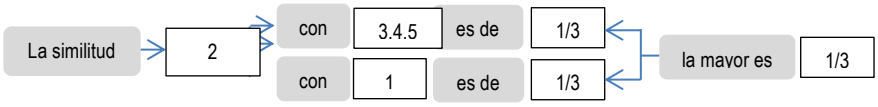


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	3	5	3
2	3	2/3	3	4	3
3	2	2/3	1	3	1
4	1	1/3	1	2	1

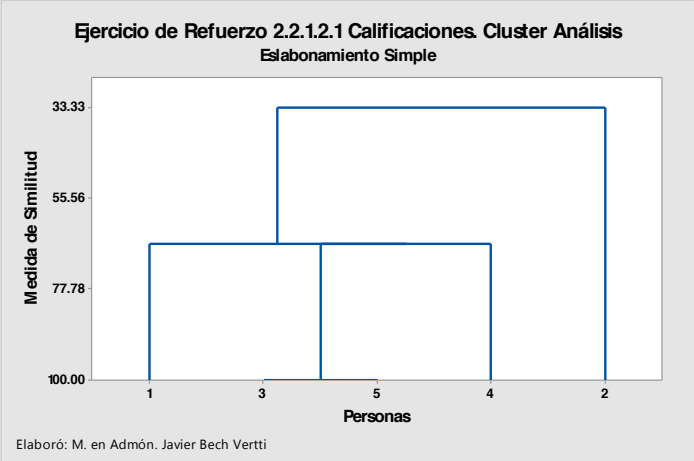
En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 3 y la 5 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 4 al primer conglomerado con una similitud de 2/3 o 66.667%. En la etapa 3 se se incorporó la persona 1 al primer conglomerado con una similitud de 2/3 o 66.667% y finalmente en la etapa 4 se incorporó la persona 2 al primer conglomerado a una similitud final de 1/3 ó 33.333%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Calif 1, Calif 2, Calif 3

```
Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0	3	5	3	2
2	3	66.667	1	3	4	3	3
3	2	66.667	1	1	3	1	4
4	1	33.333	2	1	2	1	5

Dendograma



El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.
Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay un solo conglomerado formado por todas las personas (1,2,3,4,5) a una similitud última de 1/3 o 33.333%.
Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Número de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

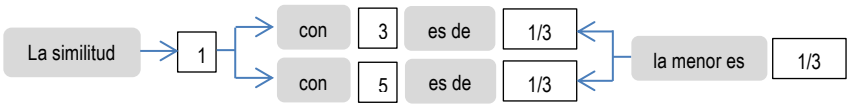
2. Eslabonamiento Completo:

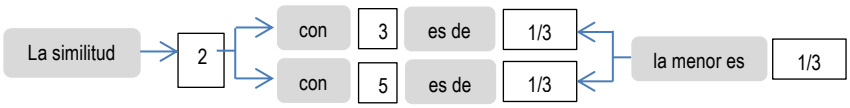
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

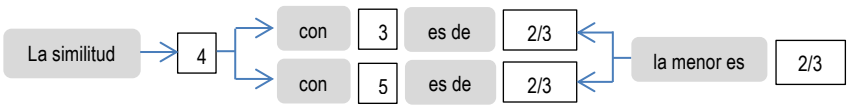
	1	2	3	4	5
1	1				
2	1/3	1			
3	1/3	1/3	1		
4	2/3	0	2/3	1	
5	1/3	1/3	**1**	2/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercano o igual a 0

	(3,5)	1	2	4
(3,5)	1			
1	1/3	1		
2	1/3	13	1	
4	2/3	2/3	0	1







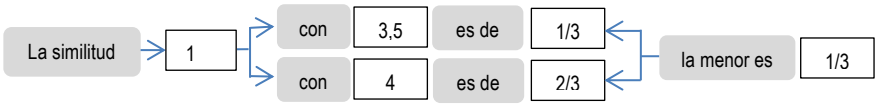
Los demás valores se pasan tal cual de la matriz anterior.

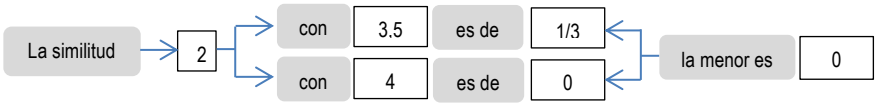
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,5)	1	2	4
(3,5)	1			
1	1/3	1		
2	1/3	13	1	
4	**2/3**	2/3	0	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercano o igual a 0

	(3,4,5)	1	2
(3,4,5)	1		
1	1/3	1	
2	0	1/3	1





El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,4,5)	1	2
(3,4,5)	1		
1	**1/3**	1	
2	0	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, lo más cercano o igual a 0

	(1,3,4,5)	2
(1,3,4,5)	1	
2	0	1

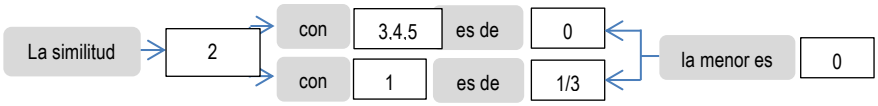


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	3	5	3
2	3	2/3	3	4	3
3	2	1/3	1	3	1
4	1	0	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el

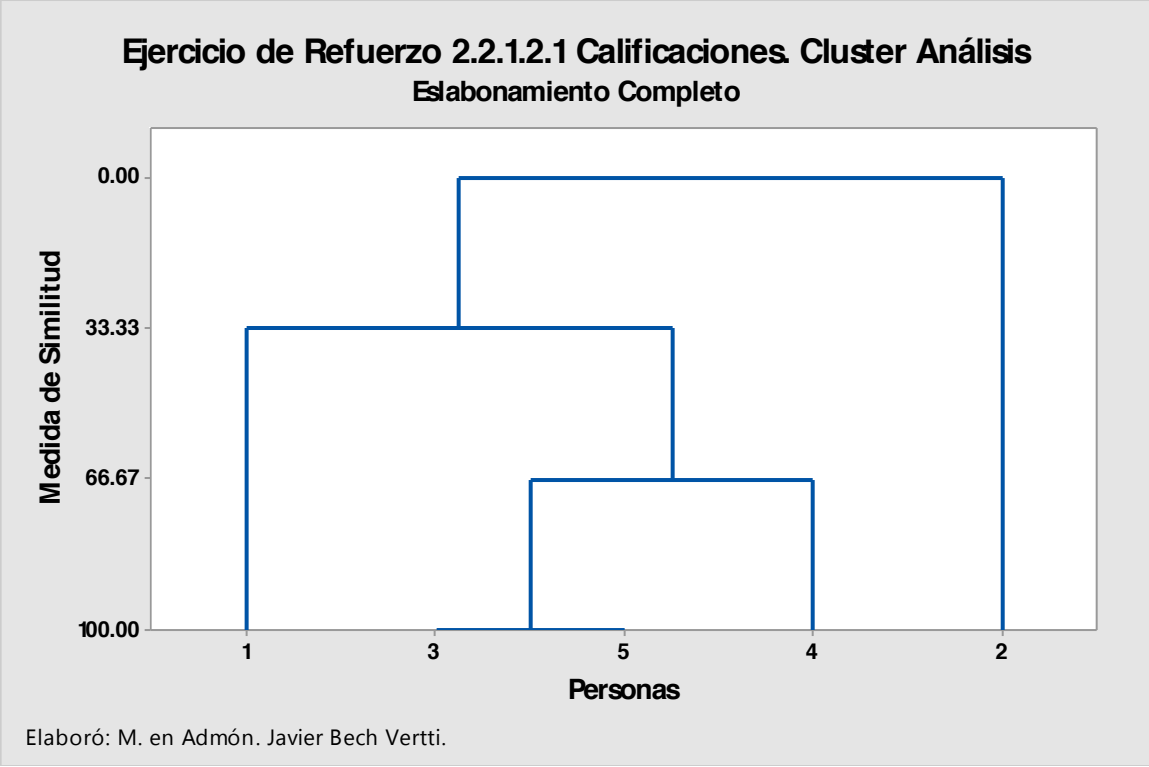
valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 3 y la 5 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 4 al primer conglomerado con una similitud de 2/3 o 66.667%. En la etapa 3 se se incorporó la persona 1 al primer conglomerado con una similitud de 1/3 o 33.333% y finalmente en la etapa 4 se incorporó la persona 2 al primer conglomerado a una similitud final de 0 ó 0.000%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Calif 1, Calif 2, Calif 3

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0	3	5	3	2
2	3	66.667	1	3	4	3	3
3	2	33.333	2	1	3	1	4
4	1	0.000	3	1	2	1	5

Dendograma



El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay un solo conglomerado formado por todas las personas (1,2,3,4,5) a una similitud última de 0 o 0.000%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Número de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

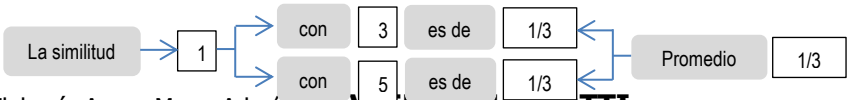
3. Eslabonamiento Promedio:

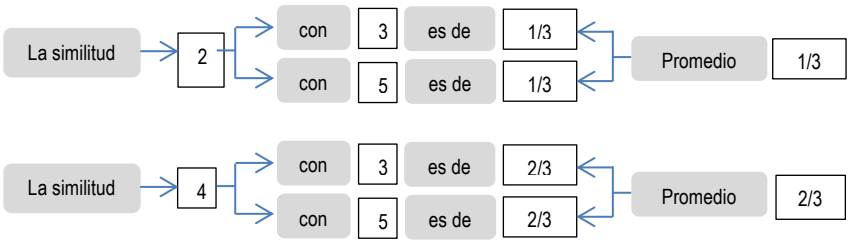
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	1	2	3	4	5
1	1				
2	1/3	1			
3	1/3	1/3	1		
4	2/3	0	2/3	1	
5	1/3	1/3	**1**	2/3	1

Paso 2: Llenar la matriz con la similitud promedio

	(3,5)	1	2	4
(3,5)	1			
1	1/3	1		
2	1/3	13	1	
4	2/3	2/3	0	1





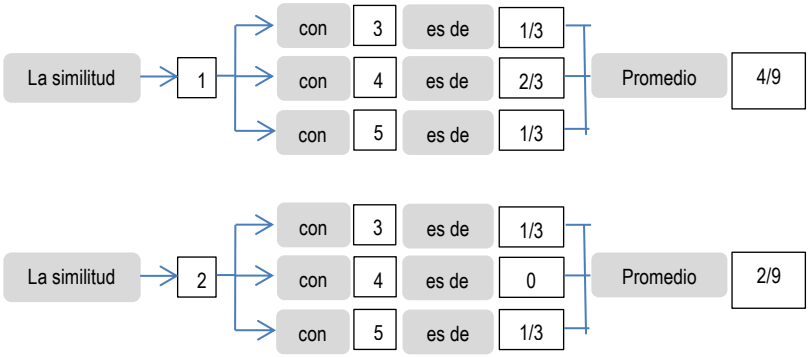
Los demás valores se pasan tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,5)	1	2	4
(3,5)	1			
1	1/3	1		
2	1/3	13	1	
4	**2/3**	2/3	0	1

Paso 2: Llenar la matriz con la similitud promedio considerando todas las combinaciones **con las similitudes de la matriz original inicial:**

	(3,4,5)	1	2
(3,4,5)	1		
1	4/9	1	
2	2/9	1/3	1



El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,4,5)	1	2
(3,4,5)	1		

1	**4/9**	1	
2	2/9	1/3	1

Paso 2: Llenar la matriz con la similitud promedio considerando todas las combinaciones **con las similitudes de la matriz original inicial:**

	(1,3,4,5)	2
(1,3,4,5)	1	
2	**1/4**	1

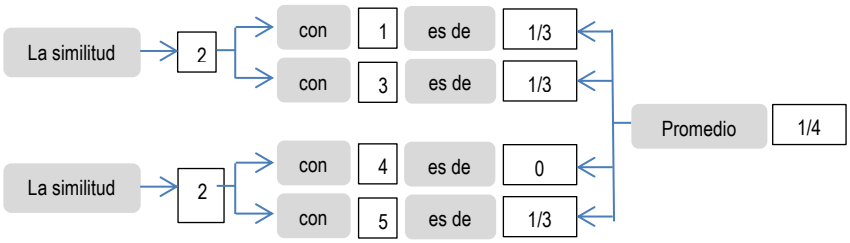


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	3	5	3
2	3	2/3	3	4	3
3	2	4/9	1	3	1
4	1	1/4	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos siete personas menos uno nos dan seis etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 7 x 7 y que en la primera etapa quedó reducida a una matriz de 6 x 6 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 5 que en este caso fueron la 3 y la 5 con una similitud de 1 o 100% y en la séptima columna nos señala la formación del primer conglomerado. En la etapa 2 se incorporó la persona 4 al primer conglomerado con una similitud de 2/3 o 66.667%. En la etapa 3 se se incorporó la persona 1 al primer conglomerado con una similitud de 4/9 o 44.444% y finalmente en la etapa 4 se incorporó la persona 2 al primer conglomerado a una similitud final de

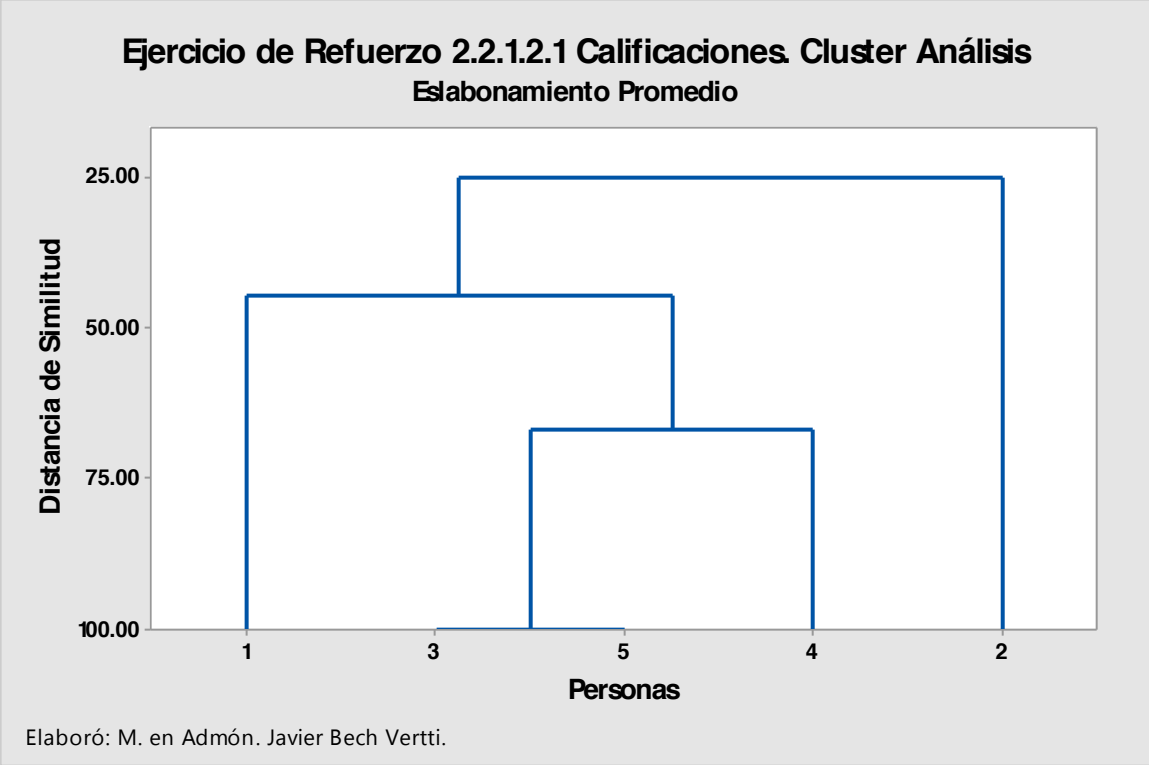
1/4 ó 25.000%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Calif 1, Calif 2, Calif 3

```
Distancia euclediana cuadrada, Enlace de promedios
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0.00000	3	5	3	2
2	3	66.667	1.00000	3	4	3	3
3	2	44.444	1.66667	1	3	1	4
4	1	25.000	2.25000	1	2	1	5

Dendograma



Elaboró: M. en Admón. Javier Bech Vertti.

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay un solo conglomerado formado por todas las personas (1,2,3,4,5) a una similitud última de 1/4 o 25.000%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Número de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

EJERCICIO DE **REFUERZO 2.2.1.2.2 Cluster Análisis. Medida de Similitud**

2.2.1.2.2. Se recabó información de la altura, el peso y el color de ojos de 5 personas. Con estas tres características se desea realizar un estudio de mercado que permita a los industriales de la confección desarrollar un sistema de patronaje unisex y colores adecuados. Los resultados se presentan a continuación:

Persona	Altura (mts.)	Peso (Kg.)	Color ojos
1	1.70	70	Azul
2	1.82	93	Café
3	1.67	83	Negro
4	1.60	60	Café
5	1.90	105	Café

Criterio de similitud para recodificar los datos:

Altura: existe(1) si es ≥ a 1.80 mts.

no existe(0) si es < a 1.80 mts.

Peso: existe(1) si es ≥ a 71 Kgs.

no existe(0) si es < a 71 Kgs.

Cadera: existe(1) si es Café

no existe (0) si es cualquier otro color

a) Obtenga la pareja de personas más parecidas y la menos parecida.

b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

Resolución al inciso a)

Los datos recodificados quedaron como sigue:

Persona	Altura (mts.)	Peso (Kg.)	Color ojos
1	0	0	0
2	1	1	1
3	0	1	0
4	0	0	1
5	1	1	1

Cálculo de las similitudes a encontrar:

El coeficiente de similitud que se usará para resolver este ejemplo es el de parejas simples. El coeficiente es:

$$S_{ij} = \frac{a + d}{p}$$

Se debe cuantificar la similitud entre las personas con base en este coeficiente, donde ***a*** es la presencia del atributo en ambas personas, ***d*** es la ausencia del atributo en ambas personas y ***p*** es el total de características (o variables) que serán evaluadas.

Se empieza a comparar por pareja de personas, en este caso la persona 1 y la 2, como en ninguna de las características hay presencia del atributo en ambas (***a***) ni ausente en ambas (***d***), entonces tanto ***a*** como ***d***

tienen valor de 0 en el coeficiente, en cuanto a ***p*** tenemos que son tres características para evaluar (Altura, Peso y Color de Ojos); la similitud entre las personas 1 y 2 es de 0 como a continuación se presenta:

$$S_{1.2} = \frac{0+0}{3} = 0$$

De igual forma se procede con las personas 1 y 3, donde hay cero atributos presentes (***a***) en ambos y dos atributos ausentes (***d***) en ambos en todas las características; por lo tanto, la similitud entre las personas 1 y 3 es de 2/3 como a continuación se presenta:

$$S_{1.3} = \frac{0+2}{3} = 2/3$$

Y así sucesivamente, se obtienen las medidas de similitud para todas las parejas que se pretende comparar obteniendo los siguientes resultados:

$$S_{1.2} = \frac{0+0}{3} = 0 \qquad S_{2.3} = \frac{1+0}{3} = 1/3 \qquad S_{3.5} = \frac{1+0}{3} = 1/3$$

$$S_{1.3} = \frac{0+2}{3} = 2/3 \qquad S_{2.4} = \frac{1+0}{3} = 1/3 \qquad S_{4.5} = \frac{1+0}{3} = 1/3$$

$$S_{1.4} = \frac{0+2}{3} = 2/3 \qquad S_{2.5} = \frac{3+0}{3} = \frac{3}{3} = 1$$

$$S_{1.5} = \frac{0+0}{3} = 0 \qquad S_{3.4} = \frac{1+0}{3} = 1/3$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5 personas a la que le nombraremos ***matriz de similitud*** la cual representa que tan similares son las personas con respecto a la similitud entre ellas:

	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	1	1/3	1/3	1

Al observar la matriz de similitud se puede entonces contestar lo planteado en el inciso a) con respecto a las personas más parecidas y las menos parecidas, en este caso la pareja de personas más parecidas es aquella cuya similitud entre sí es la que se acerque más o sea igual a 1, en este caso tenemos una pareja formada por las personas 2 y 5 con una medida de similitud de 1.

La pareja de personas menos parecida será aquella cuya similitud entre sí es la se acerque más o sea igual a 0. En este caso tenemos dos parejas formadas por las personas 1 con 2 y 1 con 5 con una medida de similitud de 0.

Resolución al inciso b)

Vinculación o eslabonamiento:

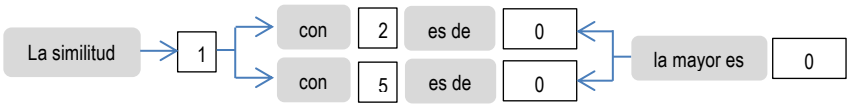
1. **Eslabonamiento Simple:**

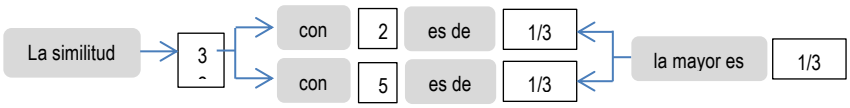
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

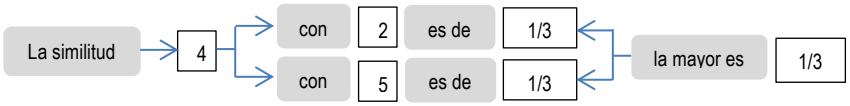
	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	1	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1







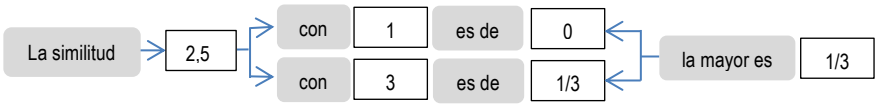
Los demás valores se pasan tal cual de la matriz anterior.

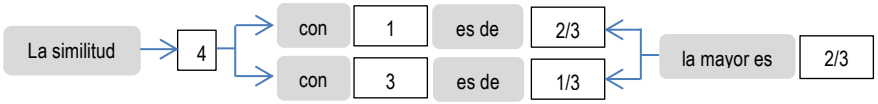
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,5)	1	2	4
(3,5)	1			
1	0	1		
2	1/3	**2/3**	1	
4	1/3	2/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/3	1	
4	1/3	2/3	1





El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/3	1	
4	1/3	**2/3**	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3,4)
(2,5)	1	
(1,3,4)	1/3	1

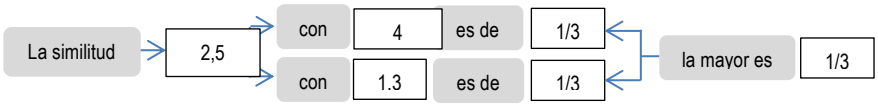


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	2	5	2
2	3	2/3	1	3	1
3	2	2/3	1	4	1
4	1	1/3	1	2	1

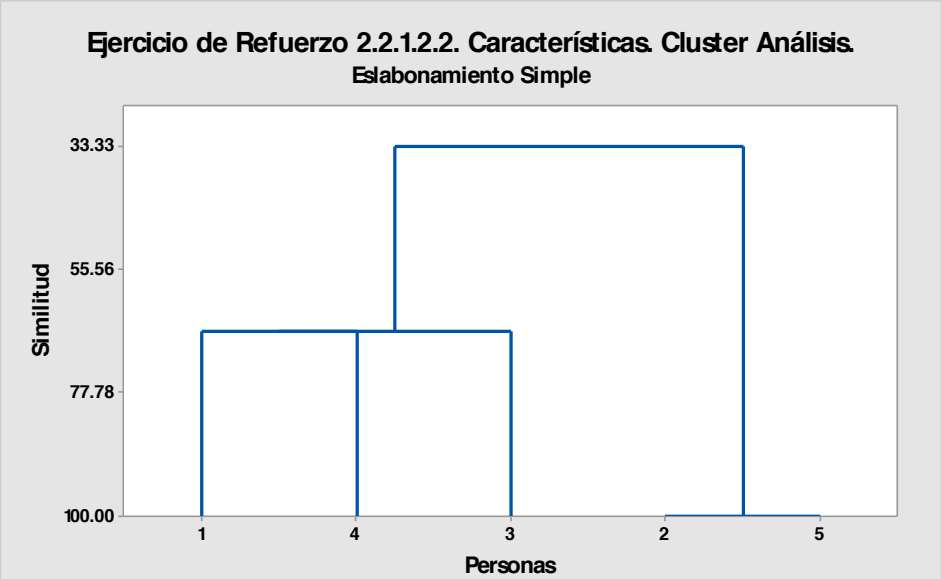
En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos uno nos dan cuatro etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 2 y la 5 con una similitud de 1 o 100% y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se formó un nuevo conglomerado con las personas 1 y 3 con una similitud de 2/3 o 66.67%. En la etapa 3 se incorporó la persona 4 al segundo conglomerado con una similitud de 2/3 o 66.667% % y finalmente en la etapa 4 se incorporó el segundo conglomerado al primer conglomerado a una similitud final de 1/3 ó 33.333%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Altura, Peso, Color

```
Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0	2	5	2	2
2	3	66.667	1	1	4	1	2
3	2	66.667	1	1	3	1	3
4	1	33.333	2	1	2	1	5

Dendrograma



El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay dos conglomerados uno formado por las personas 2 y 5 a una similitud de 1 o 100% y otro formado por las personas 1.3 y 4 una similitud de 2/3 o 66.67% uniéndose ambos conglomerados a una similitud final de 1/3 o 33.333%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

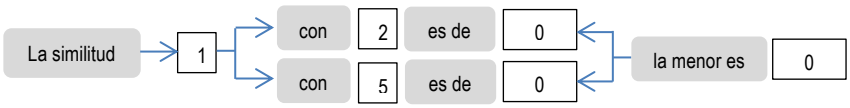
2. Eslabonamiento Completo:

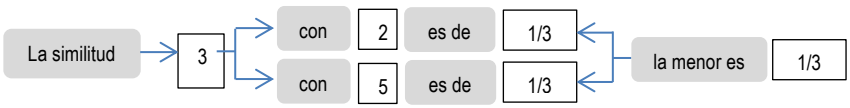
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

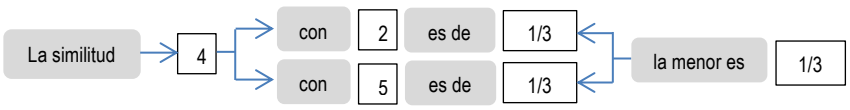
	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	**1**	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercano o igual a 0

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1







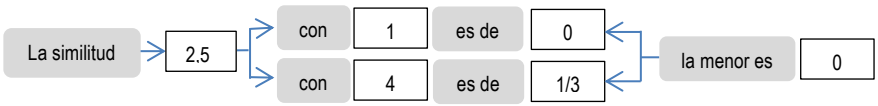
Los demás valores se pasan tal cual de la matriz anterior.

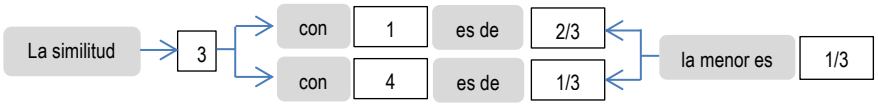
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercano o igual a 0

	(2,5)	(1,4)	3
(2,5)	1		
(1,4)	0	1	
3	1/3	1/3	1





El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,4)	3
(2,5)	1		
(1,4)	0	1	
3	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, lo más cercano o igual a 0

	(2,3,5)	(1,4)
(2,3,5)	1	
(1,4)	**0**	1

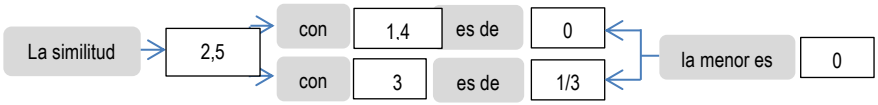


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	2	5	2
2	3	2/3	1	4	1
3	2	1/3	2	3	2
4	1	0	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos una nos dan

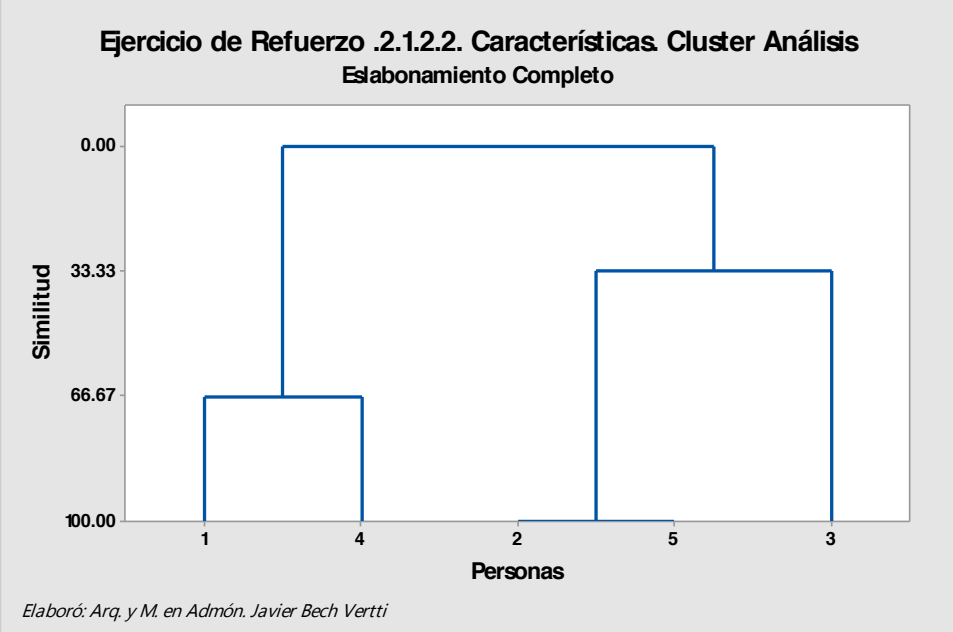
cuatro etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 2 y la 5 con una similitud de 1 o 100% y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se formó un nuevo conglomerado con las personas 1 y 4 con una similitud de 2/3 o 66.67%. En la etapa 3 se incorporó la persona 3 al primer conglomerado con una similitud de 1/3 o 33.333% % y finalmente en la etapa 4 se incorporó el segundo conglomerado al primer conglomerado a una similitud final de 0 ó 0.000%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Altura, Peso, Color

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0	2	5	2	2
2	3	66.667	1	1	4	1	2
3	2	33.333	2	2	3	2	3
4	1	0.000	3	1	2	1	5

Dendograma



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay dos conglomerados uno formado por las personas 1 y 4 a una similitud de 2/3 o 66.667% y otro formado por las personas 2,3 y 5 a una similitud de 1/3 o 33.333% uniéndose ambos conglomerados a una similitud final de 0 o 0.000%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

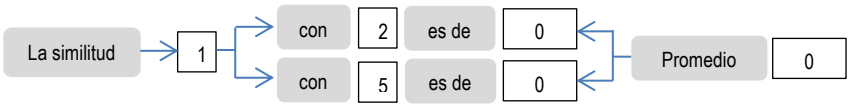
3. Eslabonamiento Promedio:

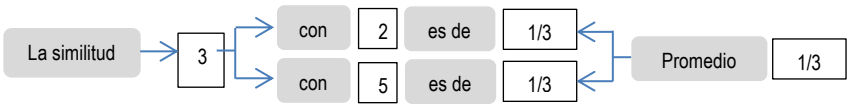
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

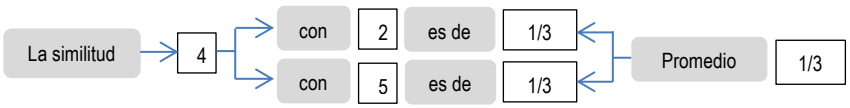
	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	**1**	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud promedio

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1







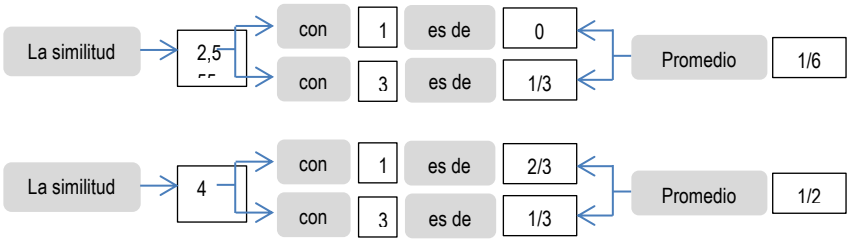
Los demás valores se pasan tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	**2/3**	1/3	1

Paso 2: Llenar la matriz con la similitud promedio considerando todas las combinaciones **con las similitudes de la matriz original inicial:**

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/6	1	
4	1/3	1/2	1



El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/6	1	
4	1/3	**1/2**	1

Paso 2: Llenar la matriz con la similitud promedio considerando todas las combinaciones **con las similitudes de la matriz original inicial:**

	(2,5)	(1,3,4)
(2,5)	1	
(1,3,4)	**4/18**	1

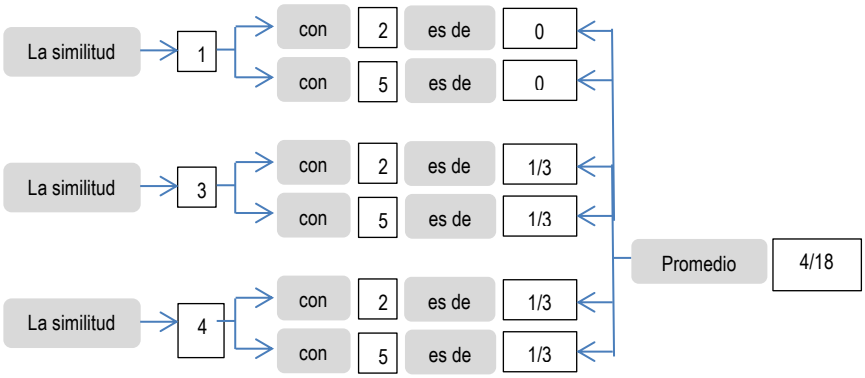


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	2	5	2
2	3	2/3	1	3	1
3	2	1/2	1	4	1
4	1	4/18	1	2	1

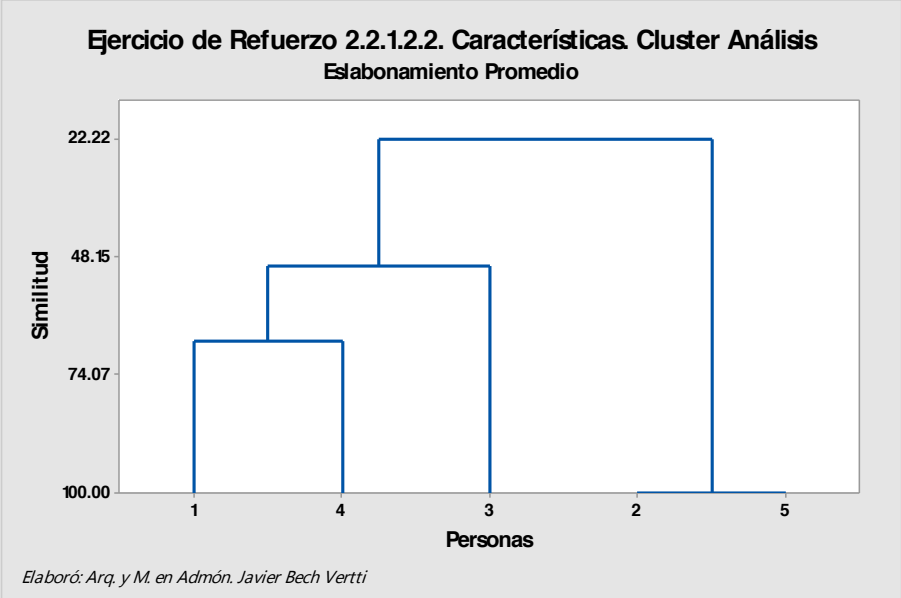
En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos una nos dan cuatro etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras personas presentadas en la columna 4 que en este caso fueron la 2 y la 5 con una similitud de 1 o 100% y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se formó un nuevo conglomerado con las personas 1 y 3 con una similitud de 2/3 o 66.67%. En la etapa 3 se incorporó la persona 4 al segundo conglomerado con una similitud de 1/2 o 50.000% % y finalmente en la etapa 4 se incorporó el segundo conglomerado al primer conglomerado a una similitud final de 4/18 ó 22.222%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Altura, Peso, Color

```
Distancia euclediana cuadrada, Enlace de promedios
Pasos de amalgamación


                                                                  Número de obs. en
          Número de    Nivel de   Nivel de  Conglomerados        Nuevo    el conglomerado
```

Paso	conglomerados	semejanzal	distancia	incorporados	conglomerado	nuevo
1	4	100.000	0.00000	2	5	2
2	3	66.667	1.00000	1	4	1
3	2	50.000	1.50000	1	3	1
4	1	22.222	2.33333	1	2	1

Dendograma



El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay un solo conglomerado formado por todas las personas (1,2,3,4,5) a una similitud última de 4/18 o 22.222%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

EJERCICIO DE **REFUERZO 2.2.1.2.3 Cluster Análisis. Medida de Similitud**

2.2.1.2.3. Se recabó información de cinco familias que solicitaron un préstamo bancario. La Institución les aplicó un cuestionario para conocer si contaban con casa propia, auto propio y cuanto eran sus ingresos mensuales. Los resultados se presentan a continuación:

Familia	Casa	Auto	Ingreso Mes ($)
1	No	No	9,000
2	Si	Si	15,000
3	No	Si	5,000
4	No	No	8,000
5	Si	Si	17,000

Criterio de similitud para recodificar los datos:

Casa: Presencia (1) si la respuesta es (Sí)

Ausencia (0) Si la respuesta es (No)

Auto: Presencia (1) si La respuesta es (Sí).

Ausencia (0) si la respuesta es (No).

Ingreso: Existe (1) si es ≥ a $ 10,000.

No Existe (0) si es < a $ 10,000.

a) Obtenga la pareja de familias más parecidas y la menos parecida.

b) Se le pide realice un agrupamiento jerárquico aglomerativo con los tres eslabonamientos simple, completo y promedio y represente cada procedimiento en un dendograma. Establezca la solución óptima.

Resolución al inciso a)

Los datos recodificados quedaron como sigue:

Familia	Casa	Auto	Ingreso Mes ($)
1	0	0	0
2	1	1	1
3	0	1	0
4	0	0	1
5	1	1	1

Cálculo de las similitudes a encontrar:

El coeficiente de similitud que se usará para resolver este ejemplo es el de parejas simples. El coeficiente es:

$$S_{ij} = \frac{a + d}{p}$$

Se debe cuantificar la similitud entre las personas con base en este coeficiente, donde ***a*** es la presencia del atributo en ambas personas, ***d*** es la ausencia del atributo en ambas personas y ***p*** es el total de características (o variables) que serán evaluadas.

Se empieza a comparar por pareja de personas, en este caso la persona 1 y la 2, como en ninguna de las características hay presencia del atributo en ambas (***a***) ni ausente en ambas (***d***), entonces tanto ***a*** como ***d*** tienen valor de 0 en el coeficiente, en cuanto a ***p*** tenemos que son tres características para evaluar (Casa, Auto e Ingreso); la similitud entre las personas 1 y 2 es de 0 como a continuación se presenta:

$$S_{1.2} = \frac{0 + 0}{3} = 0$$

De igual forma se procede con las personas 1 y 3, donde hay cero atributos presentes (***a***) en ambos y dos atributos ausentes (***d***) en ambos en todas las características; por lo tanto, la similitud entre las personas 1 y 3 es de 2/3 como a continuación se presenta:

$$S_{1.3} = \frac{0 + 2}{3} = 2/3$$

Y así sucesivamente, se obtienen las medidas de similitud para todas las parejas que se pretende comparar obteniendo los siguientes resultados:

$$S_{1.2} = \frac{0+0}{3} = 0 \qquad S_{2.3} = \frac{1+0}{3} = 1/3 \qquad S_{3.5} = \frac{1+0}{3} = 1/3$$

$$S_{1.3} = \frac{0+2}{3} = 2/3 \qquad S_{2.4} = \frac{1+0}{3} = 1/3 \qquad S_{4.5} = \frac{1+0}{3} = 1/3$$

$$S_{1.4} = \frac{0+2}{3} = 2/3 \qquad S_{2.5} = \frac{3+0}{3} = \frac{3}{3} = 1$$

$$S_{1.5} = \frac{0+0}{3} = 0 \qquad S_{3.4} = \frac{1+0}{3} = 1/3$$

Con la información anterior se construye una matriz cuadrada en este caso de 5 x 5 ya que tenemos 5 familias a la que le nombraremos ***matriz de similitud*** la cual representa que tan similares son las personas con respecto a la similitud entre ellas:

	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	1	1/3	1/3	1

Al observar la matriz de similitud se puede entonces contestar lo planteado en el inciso a) con respecto a las familias más parecidas y las menos parecidas, en este caso la pareja de familias más parecidas es aquella cuya similitud entre sí es la que se acerque más o sea igual a 1, en este caso tenemos una pareja formada por las familias 2 y 5 con una medida de similitud de 1.

La pareja de familias menos parecida será aquella cuya similitud entre sí es la se acerque más o sea igual a 0. En este caso tenemos dos parejas formadas por las familias 1 con 2 y 1 con 5 con una medida de similitud de 0.

Resolución al inciso b)

Vinculación o eslabonamiento:

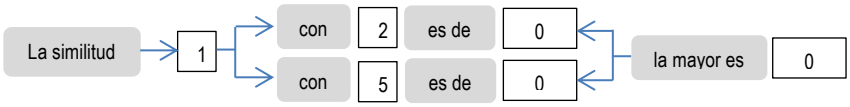
1. **Eslabonamiento Simple:**

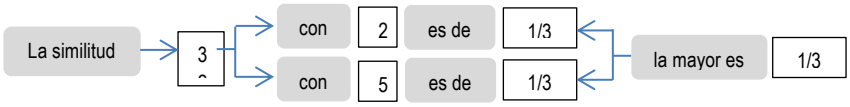
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

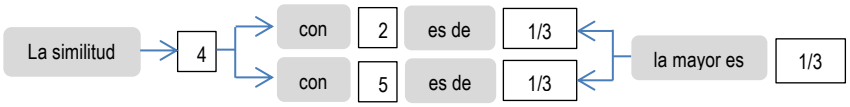
	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	1	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1







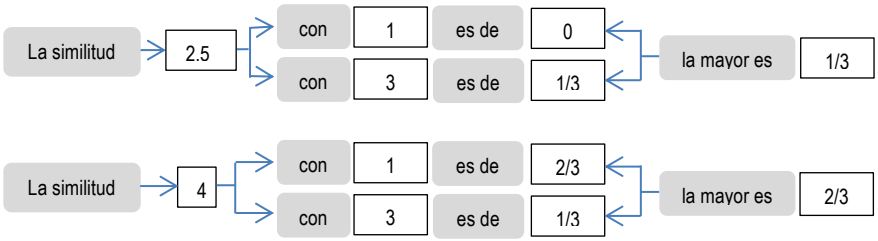
Los demás valores se pasan tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(3,5)	1	2	4
(3,5)	1			
1	0	1		
2	1/3	**2/3**	1	
4	1/3	2/3	1/3	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/3	1	
4	1/3	2/3	1



El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/3	1	
4	1/3	**2/3**	1

Paso 2: Llenar la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3,4)
(2,5)	1	
(1,3,4)	1/3	1

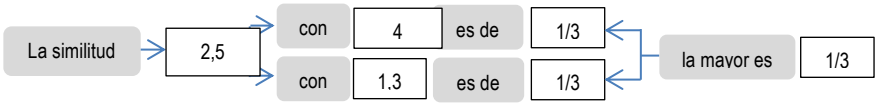


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	2	5	2
2	3	2/3	1	3	1
3	2	2/3	1	4	1
4	1	1/3	1	2	1

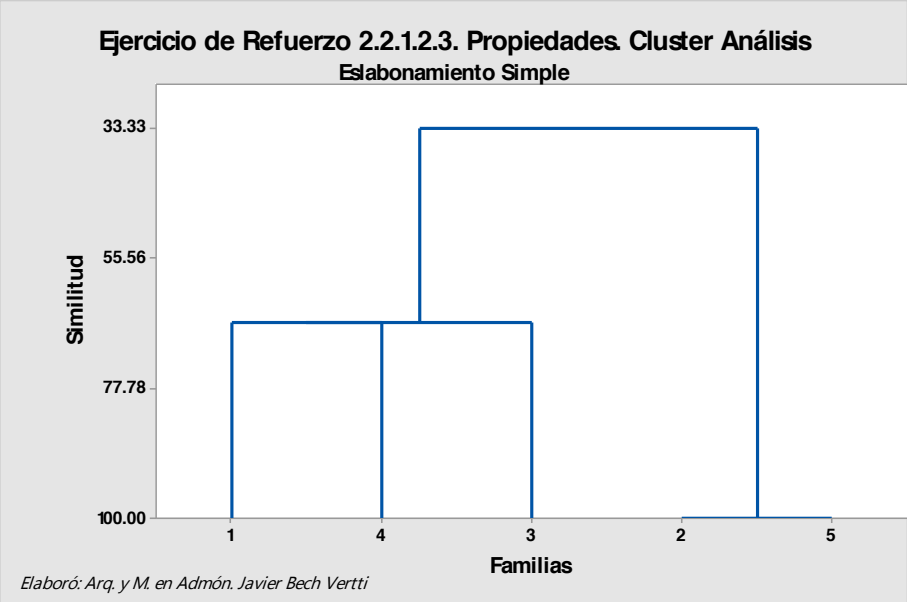
En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco familias menos uno nos dan cuatro etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras familias presentadas en la columna 4 que en este caso fueron la 2 y la 5 con una similitud de 1 o 100% y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se formó un nuevo conglomerado con las familias 1 y 3 con una similitud de 2/3 o 66.67%. En la etapa 3 se incorporó la familia 4 al segundo conglomerado con una similitud de 2/3 o 66.667% % y finalmente en la etapa 4 se incorporó el segundo conglomerado al primer conglomerado a una similitud final de 1/3 ó 33.333%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Casa, Auto, Ingreso Mes ($)

```
Distancia euclediana cuadrada, Enlace simple
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0	2	5	2	2
2	3	66.667	1	1	4	1	2
3	2	66.667	1	1	3	1	3
4	1	33.333	2	1	2	1	5

Dendrograma



El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay dos conglomerados uno formado por las familias 2 y 5 a una similitud de 1 o 100% y otro formado por las familias 1,3 y 4 una similitud de 2/3 o 66.67% uniéndose ambos conglomerados a una similitud final de 1/3 o 33.333%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

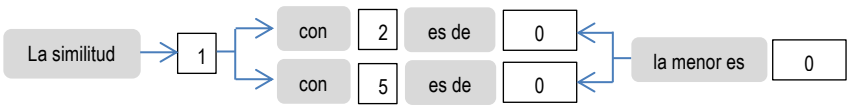
2. **Eslabonamiento Completo:**

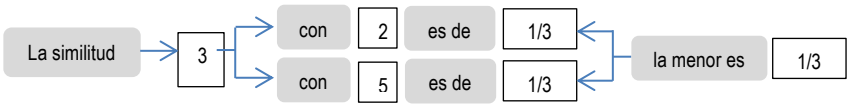
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

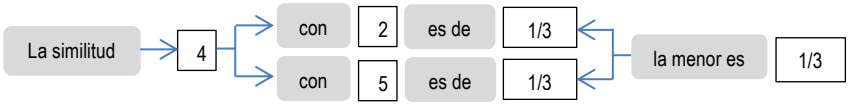
	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	**1**	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercano o igual a 0

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1







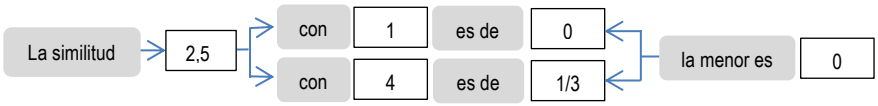
Los demás valores se pasan tal cual de la matriz anterior.

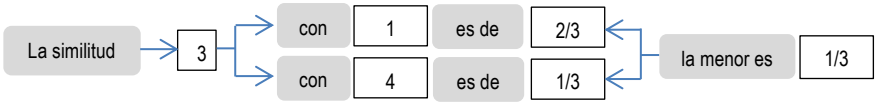
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, la más cercano o igual a 0

	(2,5)	(1,4)	3
(2,5)	1		
(1,4)	0	1	
3	1/3	1/3	1





El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,4)	3
(2,5)	1		
(1,4)	0	1	
3	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud más pequeña, lo más cercano o igual a 0

	(2,3,5)	(1,4)
(2,3,5)	1	
(1,4)	**0**	1

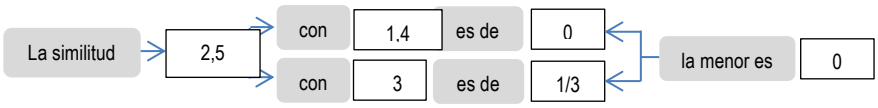


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	2	5	2
2	3	2/3	1	4	1
3	2	1/3	2	3	2
4	1	0	1	2	1

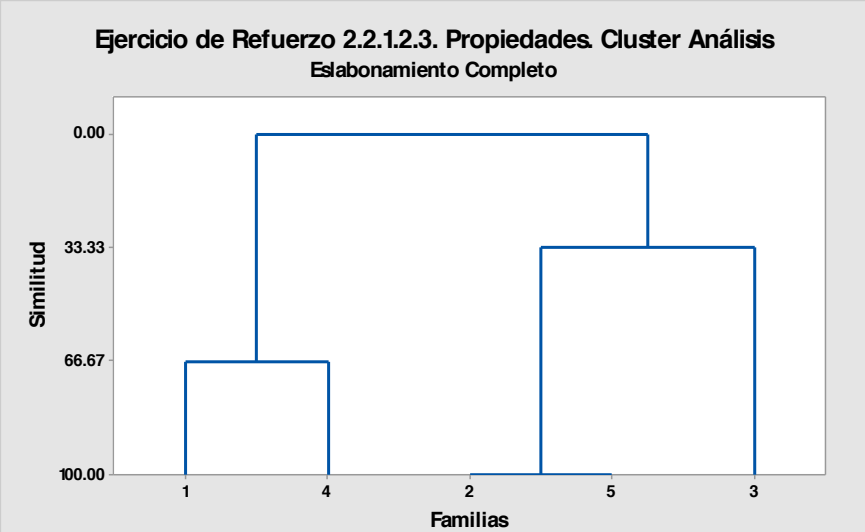
En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos una nos dan cuatro etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras familias presentadas en la columna 4 que en este caso fueron la 2 y la 5 con una similitud de 1 o 100% y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se formó un nuevo conglomerado con las familias 1 y 4 con una similitud de 2/3 o 66.67%. En la etapa 3 se incorporó la familia 3 al primer conglomerado con una similitud de 1/3 o 33.333% % y finalmente en la etapa 4 se incorporó el segundo conglomerado al primer conglomerado a una similitud final de 0 ó 0.000%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Casa, Auto, Ingreso Mes ($)

```
Distancia euclediana cuadrada, Enlace completo
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0	2	5	2	2
2	3	66.667	1	1	4	1	2
3	2	33.333	2	2	3	2	3
4	1	0.000	3	1	2	1	5

Dendrograma



El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay dos conglomerados uno formado por las familias 1 y 4 a una similitud de 2/3 o 66.667% y otro formado por las familias 2,3 y 5 a una similitud de 1/3 o 33.333% uniéndose ambos conglomerados a una similitud final de 0 o 0.000%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

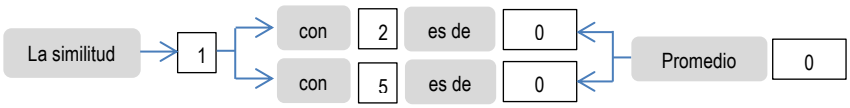
3. Eslabonamiento Promedio:

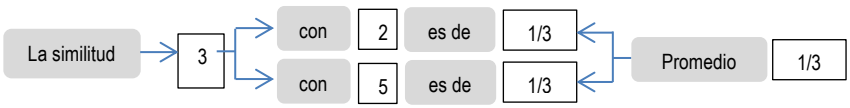
Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

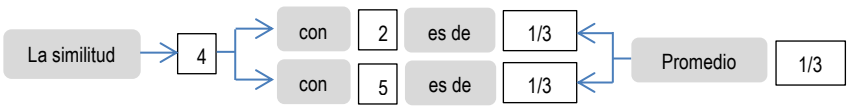
	1	2	3	4	5
1	1				
2	0	1			
3	2/3	1/3	1		
4	2/3	1/3	1/3	1	
5	0	**1**	1/3	1/3	1

Paso 2: Llenar la matriz con la similitud promedio

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	2/3	1/3	1







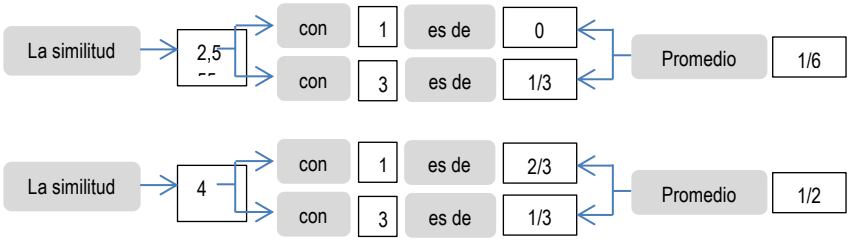
Los demás valores se pasan tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	1	3	4
(2,5)	1			
1	0	1		
3	1/3	2/3	1	
4	1/3	**2/3**	1/3	1

Paso 2: Llenar la matriz con la similitud promedio considerando todas las combinaciones **con las similitudes de la matriz original inicial:**

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/6	1	
4	1/3	1/2	1



El valor restante se pasa tal cual de la matriz anterior.

Paso 1: Reducir la matriz con la similitud más grande, lo más cercano o igual a 1

	(2,5)	(1,3)	4
(2,5)	1		
(1,3)	1/6	1	
4	1/3	**1/2**	1

Paso 2: Llenar la matriz con la similitud promedio considerando todas las combinaciones **con las similitudes de la matriz original inicial:**

	(2,5)	(1,3,4)
(2,5)	1	
(1,3,4)	**4/18**	1

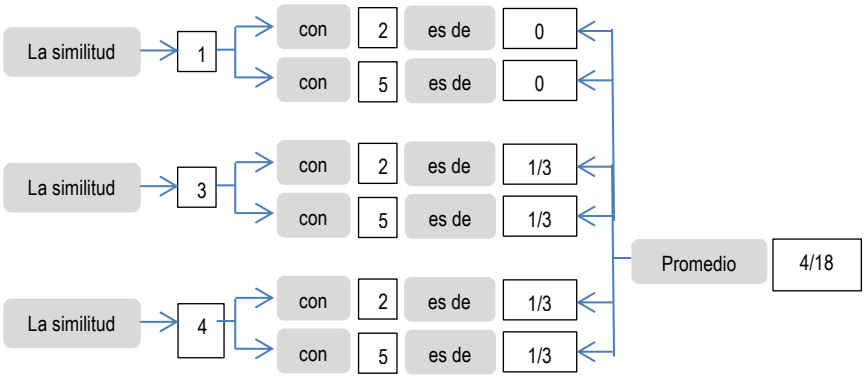


Tabla de eslabonamiento

La información de las etapas realizadas se puede resumir en la siguiente tabla:

Etapa	Número de conglomerados	Nivel de similitud	Personas incorporadas		Nuevo conglomerado
1	4	1	2	5	2
2	3	2/3	1	3	1
3	2	1/2	1	4	1
4	1	4/18	1	2	1

En la tabla anterior se presenta en la primera columna el número de etapas que se realizaron, la cual se puede calcular como el número de elementos menos uno, es decir como tenemos cinco personas menos una nos dan cuatro etapas. La segunda columna nos muestra la reducción de la matriz principal que era de 5 x 5 y que en la primera etapa quedó reducida a una matriz de 4 x 4 y así sucesivamente. La tercera columna nos muestra el valor de la similitud a la que se eslabonaron las dos primeras familias presentadas en la columna 4 que en este

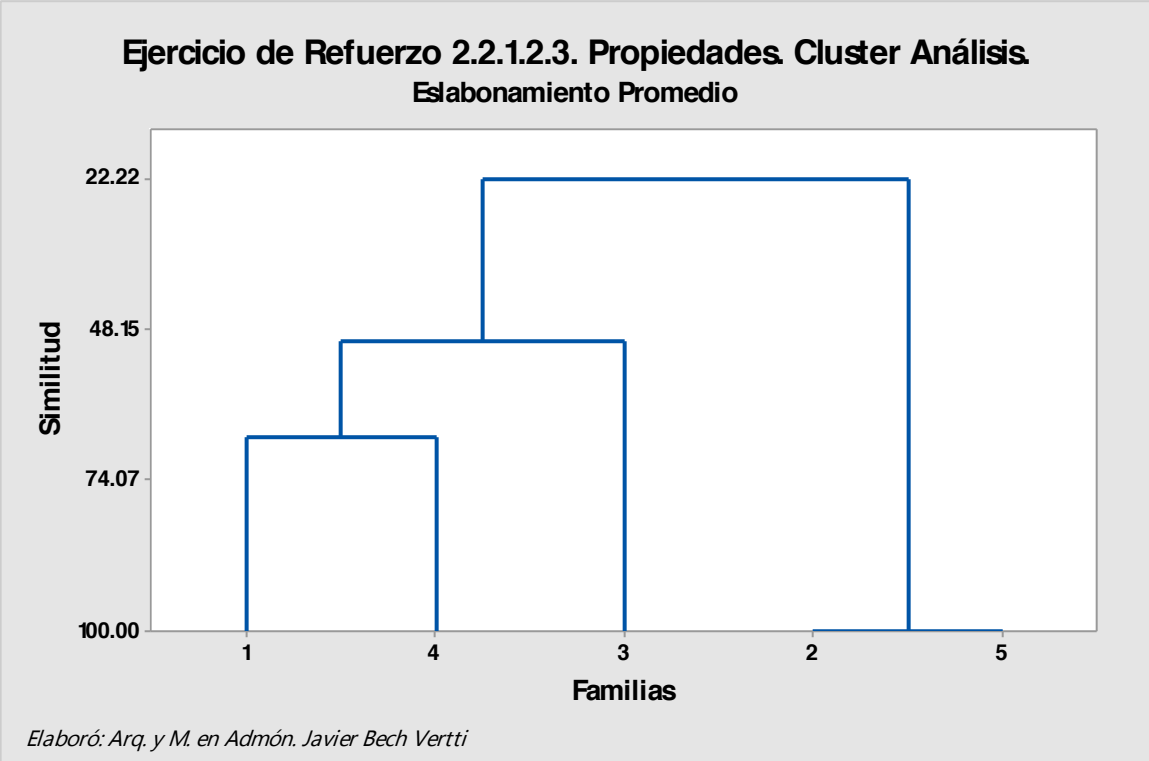
caso fueron la 2 y la 5 con una similitud de 1 o 100% y en la quinta columna nos señala la formación del primer conglomerado. En la etapa 2 se formó un nuevo conglomerado con las familias 1 y 3 con una similitud de 2/3 o 66.67%. En la etapa 3 se incorporó la familia 4 al segundo conglomerado con una similitud de 1/2 o 50.000% % y finalmente en la etapa 4 se incorporó el segundo conglomerado al primer conglomerado a una similitud final de 4/18 ó 22.222%. Minitab presenta la misma información en una tabla semejante haciendo hincapié que debido a que se tomaron fracciones para desarrollar el problema los resultados de Minitab pueden variar sensiblemente.

Análisis de observaciones de conglomerado: Casa, Auto, Ingreso Mes ($)

```
Distancia euclediana cuadrada, Enlace de promedios
Pasos de amalgamación
```

Paso	Número de conglomerados	Nivel de semejanzal	Nivel de distancia	Conglomerados incorporados		Nuevo conglomerado	Número de obs. en el conglomerado nuevo
1	4	100.000	0.00000	2	5	2	2
2	3	66.667	1.00000	1	4	1	2
3	2	50.000	1.50000	1	3	1	3
4	1	22.222	2.33333	1	2	1	5

Dendrograma



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

El dendograma es un gráfico formado por rectángulos que unen a los elementos agrupados en el eje horizontal y su altura en el eje vertical indica el nivel de similitud a la cual se unieron dichos elementos.

Es importante señalar que en el dendograma no deben cruzarse las líneas que indican los eslabonamientos para lo cual es conveniente construirlo conforme a los elementos que forman cada conglomerado en nuestro caso hay un solo conglomerado formado por todas las familias (1,2,3,4,5) a una similitud última de 4/18 o 22.222%.

Cabe mencionar que a menor nivel de similitud en que se realiza el eslabonamiento de los elementos para formar un conglomerado, es mayor la diferencia entre ellos.

Numero de conglomerados a dejar como solución

¿Cómo saber dónde cortar el dendograma?. Debido a que la definición de una agrupación útil depende por completo de su situación particular, debe especificar los criterios para colocar la partición final. Puede definir la colocación basándose en el número de grupos que desee obtener o por el nivel de similitud que requiere dentro de los conglomerados. Examine los niveles de similitud en los resultados de la tabla de eslabonamiento resumen y en el dendograma. La similitud de eslabonamiento en cualquier etapa es el valor de la similitud mínima en la etapa, en relación con la similitud máxima entre observaciones dentro de los datos. El patrón de cómo los valores de similitud cambian de paso a paso puede ayudar a elegir la agrupación final. El paso donde los valores cambian de manera abrupta podría identificar un buen punto para cortar el dendograma, si esto tiene sentido para sus datos.

RESPUESTAS A LA AUTOEVALUACIÓN CON REACTIVOS DE **FALSO Ó VERDADERO** DEL CAPÍTULO 2

1.V	2.V	3.V	4.V	5.F	6.V	7.V	8.V	9.V
10.V	11.V	12.V	13.V	14.F	15.V	16.F	17.V	18.V
19.V	20.F	21.V	22.F	23.F	24.V	25.V	26.F	27.V
28.F	29.V	30.F	31.V	32.F	33.V	34.F	35.V	36.F
37.V	38.F	39.F	40.F	41.F	42.V	43.V	44.F	45.V
46.F	47.F	48.V	49.F	50.F	51.V	52.F	53.F	54.F
55.V	56.F	57.F	58.F	59.V	60.F	61.F	62.F	63.V
64.F	65.V	66.F	67.V	68.F	69.V	70.F	71.V	72.F
73.V	74.F	75.F	76.V	77.F	78.F	79.V	80.F	81.F

RESPUESTAS A LA AUTOEVALUACIÓN CON REACTIVOS DE **PALABRA O PALABRAS PARA COMPLETAR EL ENUNCIADO** DEL CAPÍTULO 2

1.b	2.a	3.b	4.c	5.a	6.d	7.c	8.b	9.d	10.b
11.c	12.d	13.d	14.c	15.d	16.d	17.d	18.b	19.c	20.b
21.c	22.b	23.a	24.c	25.c	26.a	27.a	28.d	29.b	30.d

RESPUESTAS A LAS ACTIVIDADES DE APRENDIZAJE DEL CAPÍTULO 3

3.1.1.1 ACTIVIDAD DE APRENDIZAJE

ACTIVIDAD DE APRENDIZAJE 3.1.1.1 Análisis Discriminante

Una empresa del segmento de alimentos está desarrollando una nueva galleta fabricada con cereal del cual se considera que la cantidad de proteínas, vitaminas y minerales que llega a tener la mezcla son tres de los factores clave para que el consumidor le llegue a agradar o desagradar el sabor del producto. Para llevar a cabo la investigación, la empresa le da a degustar a cada persona la mezcla y le pide su opinión (le gusta o no le gusta). La empresa quiere saber qué características de su nuevo producto y que cantidades son útiles para diferenciar a las personas que les gusta o no la nueva galleta de cereal, es decir cual separa mejor a cada grupo.

La tabla siguiente refleja la valoración sobre estas tres características con diferentes cantidades de Proteínas y diferentes porcentajes de Vitaminas y Minerales efectuada a un grupo de 10 personas:

No. de persona	Grupos basados en la opinión sobre el producto	Evaluación de la nueva galleta		
		Proteínas en gramos (X_1)	Vitaminas en % (X_2)	Minerales en %(X_3)
1	Me gusta	8	6	7
2	Me gusta	9	7	9
3	Me gusta	7	8	7
4	Me gusta	8	3	9
5	Me gusta	8	6	8

	Media del grupo 1	8	6	8
6	No me gusta	2	2	4
7	No me gusta	3	2	2
8	No me gusta	6	2	2
9	No me gusta	4	2	3
10	No me gusta	5	2	4
	Media del grupo 2	4	2	3

a) Obtenga el modelo discriminante que nos ayude a comprobar la existencia de dos grupos.

b) Pruebe si el modelo es bueno para pronosticar o clasificar a las personas que les gusta o no la nueva galleta de cereal.

c) Valide el modelo discriminante con los mismos datos de la derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados indicando en orden de importancia las variables que logran segmentar y que estrategia se debería seguir.

e) ¿En qué grupo caería una persona que hubiera opinado sobre el nuevo producto de la siguiente manera? : Proteínas: 1; Vitaminas: 3 y Minerales: 2

Resolución del inciso a)

La ecuación para derivar el modelo discriminante tendrá la siguiente forma:

$$Z = K_1X_1 + K_2X_2 + K_3X_3$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las *K,* para ello se deben resolver las siguientes ecuaciones normales:

$$\sum X_1'^2 K_1 + \sum X_1'X_2'K_2 + \sum X_1'X_3'K_3 = \bar{X}_1\,|Grupo\ 1 - \bar{X}_1|Grupo\ 2$$

$$\sum X_1'X_2'K_1 + \sum X_2'^2 K_2 + \sum X_2'X_3'K_3 = \bar{X}_2\,|Grupo\ 1 - \bar{X}_2|Grupo\ 2$$

$$\sum X_1'X_3'K_1 + \sum X_2'X_3'K_2 + \sum X_3'^2 K_3 = \bar{X}_3\,|Grupo\ 1 - \bar{X}_3|Grupo\ 2$$

Donde:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\,\bar{X}_2)$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\,\bar{X}_3)$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\,\bar{X}_3)$$

Se sustituyen los valores correspondientes, obteniendo lo siguiente:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_1'^2 = [(8^2 + 9^2 + 7^2 + 8^2 + 8^2) - 5(8)^2] + [(2^2 + 3^2 + 6^2 + 4^2 + 5^2) - 5(4)^2] = 12$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_2'^2 = [(6^2 + 7^2 + 8^2 + 3^2 + 6^2) - 5(6)^2] + [(2^2 + 2^2 + 2^2 + 2^2 + 2^2) - 5(2)^2] = 14$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_3'^2 = [(7^2 + 9^2 + 7^2 + 9^2 + 8^2) - 5(8)^2] + [(4^2 + 2^2 + 2^2 + 3^2 + 4^2) - 5(3)^2] = 8$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\bar{X}_2)$$

$$=[(8x6 + 9x7 + 7x8 + 8x3 + 8x6) - 5(8x6)] + [(2x2 + 3x2 + 6x2 + 4x2 + 5x2) - 5(4x2)] = -1$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\bar{X}_3)$$

$$=[(8x7 + 9x9 + 7x7 + 8x9 + 8x8) - 5(8x8)] + [(2x4 + 3x2 + 6x2 + 4x3 + 5x4) - 5(4x3)] = 0$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\bar{X}_3)$$

$$=[(6x7 + 7x9 + 8x7 + 3x9 + 6x8) - 5(6x8)] + [(2x4 + 2x2 + 2x2 + 2x3 + 2x4) - 5(2x3)] = -4$$

De lo anterior se obtienen tres ecuaciones, normales:

$$12K_1 - 1K_2 + 0K_3 = 4$$

$$-1K_1 + 14K_2 - 4K_3 = 4$$

$$0K_1 - 4K_2 + 8K_3 = 5$$

A partir de estas ecuaciones se obtienen los valores de *K* de cada variable. En este caso se utilizará la matriz inversa para la resolución del sistema lineal siguiendo los siguientes pasos:

Paso 1. Expresar el sistema como A *K=D

Donde:

A= Matriz de coeficientes.

K= Coeficientes discriminantes.

D= Términos independientes.

$$\begin{bmatrix} 12 & -1 & 0 \\ -1 & 14 & -4 \\ 0 & -4 & 8 \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 5 \end{bmatrix}$$

Paso 2. Obtener el determinante de A aplicando la regla de Sarrus para comprobar si el sistema tiene solución, es decir si |A|≠0:

$$|A| = \begin{vmatrix} 12 & -1 & 0 \\ -1 & 14 & -4 \\ 0 & -4 & 8 \\ 12 & -1 & 0 \\ -1 & 14 & -4 \end{vmatrix} = (12*14*8)+ (-1*-4*0)+ (0*-1*-4)- (0*14*0)- (12*-4*8)- (-1*-1*8)=1144$$

Como 1144 es ≠ de 0 el sistema tiene solución.

Paso 3. Obtener la matriz transpuesta de A, que es:

$$A' = \begin{bmatrix} 12 & -1 & 0 \\ -1 & 14 & -4 \\ 0 & -4 & 8 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A', es decir formamos $\alpha A'$

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 14 & -4 \\ -4 & 8 \end{vmatrix} & -\begin{vmatrix} -1 & -4 \\ 0 & 8 \end{vmatrix} & +\begin{vmatrix} -1 & 14 \\ 0 & -4 \end{vmatrix} \\ -\begin{vmatrix} -1 & 0 \\ -4 & 8 \end{vmatrix} & +\begin{vmatrix} 12 & 0 \\ 0 & 8 \end{vmatrix} & -\begin{vmatrix} 12 & -1 \\ 0 & -4 \end{vmatrix} \\ +\begin{vmatrix} -1 & 0 \\ 14 & -4 \end{vmatrix} & -\begin{vmatrix} 12 & 0 \\ -1 & -4 \end{vmatrix} & +\begin{vmatrix} 12 & -1 \\ -1 & 14 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(112-16) & -(-8+0) & +(4-0) \\ -(-8-0) & +(96-0) & -(-48+0) \\ +((4-0) & -(-48-0) & +(168-1) \end{bmatrix} =$$

$$= \begin{bmatrix} 96 & 8 & 4 \\ 8 & 96 & 48 \\ 4 & 48 & 167 \end{bmatrix}$$

Paso 5. Calculamos la matriz inversa de A:

$$A^{-1} = \frac{1}{|A|} \propto A^1 = \frac{1}{1144}\begin{bmatrix} 96 & 8 & 4 \\ 8 & 96 & 48 \\ 4 & 48 & 167 \end{bmatrix} = \begin{bmatrix} \frac{96}{1144} & \frac{8}{1144} & \frac{4}{1144} \\ \frac{8}{1144} & \frac{96}{1144} & \frac{48}{1144} \\ \frac{4}{1144} & \frac{48}{1144} & \frac{167}{1144} \end{bmatrix}$$

Nota: Es conveniente dejar el resultado en forma de fracción para no perder precisión

Paso 6. Llevamos el valor anterior a la expresión $K = A^{-1} * D$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} \frac{96}{1144} & \frac{8}{1144} & \frac{4}{1144} \\ \frac{8}{1144} & \frac{96}{1144} & \frac{48}{1144} \\ \frac{4}{1144} & \frac{48}{1144} & \frac{167}{1144} \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 0.38112 \\ 0.57343 \\ 0.91171 \end{bmatrix}$$

Paso 7. Obtenemos el modelo discriminante:

$$\boldsymbol{Z = 0.38112K_1 + 0.57343K_2 + 0.91171K_3}$$

Resolución del inciso b)

A continuación se prueba si el modelo es bueno para discriminar o clasificar a las personas que están de acuerdo con las que no lo están mediante los siguientes pasos:

Paso 1. Plantear el juego de hipótesis:

H_0. El modelo discriminante no es bueno para discriminar o clasificar a las personas que están o no de acuerdo con las características de las laptops analizadas.

H_1. El modelo discriminante es bueno.

Paso 2. Establecer el estadístico de prueba:

$$F\ calc = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2)D^2}{[(n_1 + n_2 - 2)p * (n_1 + n_2)]}$$

$$D^2 = (n_1 + n_2 - 2)\left[\sum K_1(\bar{X}_1|Grupo\ 1 - \bar{X}_1|Grupo2\right]$$

$$D^2 = (5 + 5 - 2)[(0.38112)(4) + (0.57343)(4) + (0.91171)(5)] = 67.014$$

$$F\ calc = \frac{[(5 + 5 - 2 - 1)(5 * 5)(67.014)]}{[(5 + 5 - 2)2](5 + 5)} = \frac{7x25x67.014}{16x10} = \frac{11727.45}{160} = 73.30$$

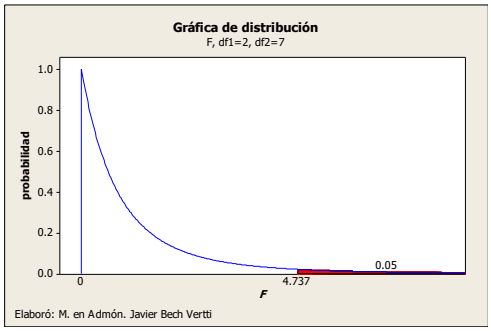
Paso 3. Establecer la región de Rechazo:

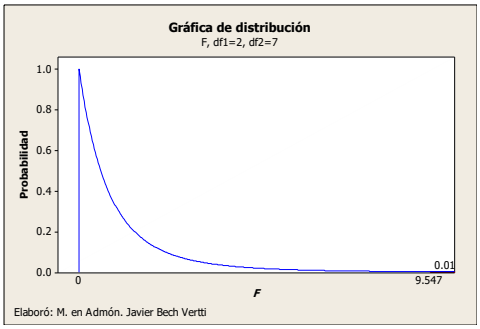
$Fcrit = [\propto_1 p_1(n_1 + n_2 - p - 1)]$

$Fcrit = 0.05, 2, 7$

$f_{0.05} = 4.74$

$f_{0.01} = 9.55$





Paso 4. Definir la regla de decisión:

Se rechaza H_0 si $F_{calc.} \geq 4.74$

Paso 5. Conclusiones:

-Estadística: Como 73.30>4.74 y > 9.55 ∴ la prueba es (AS) y se rechaza H_0 con mucha fuerza.

-Administrativa: Existe evidencia suficiente para decir que estadísticamente el modelo es bueno para clasificar o discriminar a los compradores de los no compradores.

Resolución del inciso c)

Cuando se ha concluido que el modelo discriminante es bueno para hacer clasificaciones, se debe proceder a validar el mismo. Esta validación consiste en determinar el grado de confiabilidad que se tendrá al hacer pronósticos con el modelo.

Paso 1. Se obtiene la Z promedio para cada grupo:

$Z = 0.38112K_1 + 0.57343K_2 + 0.91171K_3$

$\bar{Z}_1 = 0.38112(8) + 0.57343(6) + 0.91171(8) = 13.78$

$\bar{Z}_2 = 0.38112(4) + 0.57343(2) + 0.91171(3) = 5.41$

Paso 2. Se determina la Z crítica o de corte:

$$Z^*_{critica} = \frac{n_2\bar{Z}_1 + n_1\bar{Z}_2}{n_1 + n_2} = \frac{5(13.78) + 5(5.41)}{5 + 5} = 9.59$$

*Esta fórmula puede usarse ya sea si $n_1=n_2$ o si $n_1 \neq n_2$

Paso 3. Se obtiene la Z para cada observación:

$Z_1 = 0.38112(8) + 0.57343(6) + 0.91171(7) = 12.87 > 9.59$

$Z_2 = 0.38112(9) + 0.57343(7) + 0.91171(9) = 15.65 > 9.59$

$Z_3 = 0.38112(7) + 0.57343(8) + 0.91171(7) = 13.64 > 9.59$

$Z_4 = 0.38112(8) + 0.57343(3) + 0.91171(9) = 12.97 > 9.59$

$Z_5 = 0.38112(8) + 0.57343(6) + 0.91171(8) = 13.78 > 9.59$

$Z_6 = 0.38112(2) + 0.57343(2) + 0.91171(4) = 5.56 < 9.59$

$Z_7 = 0.38112(3) + 0.57343(2) + 0.91171(2) = 4.11 < 9.59$

$Z_8 = 0.38112(6) + 0.57343(2) + 0.91171(2) = 5.26 < 9.59$

$Z_9 = 0.38112(4) + 0.57343(2) + 0.91171(3) = 5.41 < 9.59$

$Z_{10} = 0.38112(5) + 0.57343(2) + 0.91171(4) = 6.70 < 9.59$

Paso 4. Se verifica si en las Z observadas la clasificación es correcta o incorrecta

Sujeto No.	**Grupos**	**Proteinas en gramos (X_1)**	**Vitaminas en % (X_2)**	**Minerales en % (X_3)**	**Z obs.**	**Clas. ✓o X**
1	Me gusta	8	6	7	12.87 > 9.59	✓
2	Me gusta	9	7	9	15.65 > 9.59	✓
3	Me gusta	7	8	7	13.64 > 9.59	✓
4	Me gusta	8	3	9	12.97 > 9.59	✓
5	Me gusta	8	6	8	13.78 > 9.59	✓

	Media Gpo. 1	8	6	8		
6	No me gusta	2	2	4	$5.56 < 9.59$	✓
7	No me gusta	3	2	2	$4.11 < 9.59$	✓
8	No me gusta	6	2	2	$5.26 < 9.59$	✓
9	No me gusta	4	2	3	$5.41 < 9.59$	✓
10	No me gusta	5	2	4	$6.70 < 9.59$	✓
	Media Gpo. 2	4	2	3		

Paso 5. Del paso anterior se obtiene el porcentaje de clasificaciones correctas (PCC):

$$PCC = \frac{\#\checkmark Grupo\ 1 + \#\checkmark Gupo\ 2}{\#\ de\ observaciones} * 100$$

$$PCC = \frac{5+5}{10} * 100 = 100\%$$

Paso 6. Se obtiene el porcentaje mínimo de clasificaciones correctas:

$$C = (p_i)^2 + (1 - p_1)^2$$

$$C = (\frac{5}{10})^2 + (1 - \frac{5}{10})^2$$

$$C = 0.25 + 0.25 = 0.50x100 = 50\%$$

Paso 7. Se compara el porcentaje de clasificaciones correctas contra el porcentaje mínimo de clasificaciones correctas:

Como 100% > 50% se puede validar el modelo para clasificar a los que opinan Me gusta de los que opinan No me gusta la nueva galleta fabricada con cereal al analizar ambos grupos.

Resolución del inciso c)

Para interpretar los resultados obtenidos anteriormente, se evalúa la importancia relativa que posee cada una de las variables independientes para discriminar. El método de los pesos discriminantes estandarizados consiste en estandarizar los coeficientes del modelo discriminante con el fin de que puedan ser comparados. Para jerarquizar las variables se considera solamente la magnitud del peso, no su signo, es decir el valor absoluto.

La fórmula para estandarizar los coeficientes discriminantes es:

$$K_n' = (K_n)\left[\frac{\sum X_n'^2}{(n_1 + n_2 - 2)}\right]^{0.05}$$

En el problema presentado, los coeficientes o constantes estandarizadas son:

1. Proteínas:
$$K_1 = (0.38112)\left[\frac{12}{(5+5-2)}\right]^{0.5} = 0.46677$$
2. Vitaminas:
$$K_2 = (0.57343)\left[\frac{14}{(5+5-2)}\right]^{0.5} = 0.75857$$
3. Minerales:
$$K_3 = (0.91171)\left[\frac{8}{(5+5-2)}\right]^{0.5} = 0.91171$$

-Interpretación: De acuerdo a los pesos discriminantes estandarizados **en valores absolutos** (sin considerar el signo), podemos concluir que la variable que más discrimina o diferencia a un grupo de otro para los posibles consumidores de la nueva galleta fabricada con cereal es la variable X_3 (Minerales) con un peso de .91171 por su mayor *K* estandarizada y la menos importante es la variable X_1 (Proteínas) con un peso de 0.46677 en valor absoluto.

-Estrategias a seguir: Lo que se pretendería como estrategia es tratar de actuar sobre la variable dependiente (grupos), manipulando las variables independientes (Proteínas, Vitaminas y Minerales). Lo que se quisiera lograr es aumentar los valores de Z, para hacer que las personas del grupo ***No me gusta*** se pasen al grupo ***Me gusta.*** Para nuestro problema, tenemos que si X_1 (Duración) aumenta, subirá a su vez el valor de Z, pues su signo es positivo. Tenemos que si X_3 (Minerales) aumenta, subirá el valor de Z pues su signo es positivo, Asimismo para X_2 (Vitaminas) y para X_1 (Proteínas, tenemos que va a afectar en forma directa al valor de Z, por su signo positivo.

De las tres variables, la primera que se debe cuidar es X_3 (Minerales), ya que tiene una mayor importancia para discriminar entre los grupos, aumentando más tal vez el porcentaje de Minerales en la nueva galleta con cereal, después podría pensarse en una estrategia en cuanto a la variable X_2 (Vitaminas), en un segundo plano, para incrementar el sabor de la galleta y por último una estrategia para la variable X_3 (Proteínas), en un tercer plano, para mejorar la nutrición de la nueva galleta.

Resolución del inciso d)

$X_1 = 1$

$X_2 = 3$

$X_3 = 2$

Sustituimos los valores anteriores en el modelo discriminante:

$Z = 0.38112(1) + 0.57343(3) + 0.91171(2) = 3.92$

Comparamos el valor anterior con la Z crítica:

Como 3.92< 9.59 ∴ se le clasificaría en el grupo 2 de **No me gusta**

3.2.1.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 3.2.1.1 Vector de Medias**

El Director de Mercadotecnia de una cadena de cines desea medir el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines. Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Para que los potenciales consumidores recuerden la cadena de cines cuando se utiliza la TELEVISIÓN como medio masivo para la campaña de publicidad en los consumidores los valores medios de las variables deben ser por lo menos de (10,10,10). Para comprobar si se logró este objetivo, una vez lanzada la campaña en la TELEVISIÓN, se tomó una muestra de 6 potenciales clientes y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados son los siguientes:

Considere la siguiente muestra de tamaño n=6 extraída de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo:

$$X = \begin{bmatrix} 9 & 10 & 8 \\ 8 & 9 & 9 \\ 7 & 8 & 9 \\ 9 & 9 & 9 \\ 8 & 8 & 6 \\ 10 & 7 & 10 \end{bmatrix}$$

c) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$$

d) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de Hipótesis

$$H_0: \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 10 \\ 10 \end{bmatrix}$$

$H_1: Al\ menos\ M\ es\ \neq$

2. Estadístico de prueba

$$Fcalc = \frac{n-p}{p(n-1)} \quad T^2$$

$$Donde\ T^2 = n(M - M_0)^1[S^2]^{-1}(\widehat{M} - \widehat{M}_0)$$

Donde: $S^2 = \frac{1}{n-1} SCPC = \frac{1}{n-1} D'D$ (Matriz de varianzas y covarianzas)

$$[S^2]^{-1} = \frac{1}{|S|} \quad \propto [S^2]$$

$$D = \begin{bmatrix} 9 & 10 & 8 \\ 8 & 9 & 9 \\ 7 & 8 & 9 \\ 9 & 9 & 9 \\ 8 & 8 & 6 \\ 10 & 7 & 10 \end{bmatrix} - \begin{bmatrix} 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \\ 8.5 & 8.5 & 8.5 \end{bmatrix} = \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \\ -1.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & -2.5 \\ 1.5 & -1.5 & 1.5 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 & -0.5 & -1.5 & 0.5 & -0.5 & 1.5 \\ 1.5 & 0.5 & -0.5 & 0.5 & -0.5 & -1.5 \\ -0.5 & 0.5 & 0.5 & 0.5 & -2.5 & 1.5 \end{bmatrix} \begin{bmatrix} 0.5 & 1.5 & -0.5 \\ -0.5 & 0.5 & 0.5 \\ -1.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 \\ -0.5 & -0.5 & -2.5 \\ 1.5 & -1.5 & 1.5 \end{bmatrix}$$

$$D'D = \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix}$$

$$S^2 = \frac{1}{n-1} = \frac{1}{6-1} - \frac{1}{5} \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix}$$

$$S^2 = \frac{1}{5} \begin{bmatrix} 5.5 & -0.5 & 2.5 \\ -0.5 & 5.5 & -1.5 \\ 2.5 & -1.5 & 9.5 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} \frac{5.5}{5} & \frac{-0.5}{5} & \frac{2.5}{5} \\ \frac{-0.5}{5} & \frac{5.5}{5} & \frac{-1.5}{5} \\ \frac{2.5}{5} & \frac{-1.5}{5} & \frac{9.5}{5} \end{bmatrix}$$

Utilizando la regla de Sarrus

$$|S^2| = \begin{bmatrix} \frac{5.5}{5} & \frac{-0.5}{5} & \frac{2.5}{5} \\ \frac{-0.5}{5} & \frac{5.5}{5} & \frac{-1.5}{5} \\ \frac{2.5}{5} & \frac{-1.5}{5} & \frac{9.5}{5} \\ \frac{5.5}{5} & \frac{-0.5}{5} & \frac{2.5}{5} \\ \frac{-0.5}{5} & \frac{5.5}{5} & \frac{-1.5}{5} \end{bmatrix} =$$

$$\begin{gathered} (5.5/5 \times 5.5/5 \times 9.5/5) + \\ (-0.5/5 \times -1.5/5 \times 2.5/5) + \\ (2.5/5 \times -0.5/5 \times -1.5/5) - \\ (2.5/5 \times 5.5/5 \times 2.5/5) - \\ (5.5/5 \times \ -1.5/5 \times -1.5/5) - \\ (-0.5/5 \times -0.5/5 \times -9.5/5) = \end{gathered}$$

$$= 1.936$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 1.1 & -0.3 \\ -0.3 & 1.9 \end{vmatrix} & -\begin{vmatrix} -0.1 & -0.3 \\ 0.5 & 1.9 \end{vmatrix} & +\begin{vmatrix} -0.1 & 1.1 \\ 0.5 & -0.3 \end{vmatrix} \\ -\begin{vmatrix} -0.1 & 0.5 \\ -0.3 & 1.9 \end{vmatrix} & +\begin{vmatrix} 1.1 & 0.5 \\ 0.5 & 1.9 \end{vmatrix} & -\begin{vmatrix} 1.1 & -0.1 \\ 0.5 & -0.3 \end{vmatrix} \\ +\begin{vmatrix} -0.1 & 0.5 \\ 1.1 & -0.3 \end{vmatrix} & -\begin{vmatrix} 1.1 & 0.5 \\ -0.1 & -0.3 \end{vmatrix} & +\begin{vmatrix} 1.1 & -0.1 \\ -0.1 & 1.1 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(2.09 - 0.09) & -(-0.19 + 0.15) & +(0.03 - 0.55) \\ -(-0.19 + 0.15) & +(2.09 - 0.25) & -(-0.33 + 0.05) \\ +((0.03 - 0.55) & -(-0.33 - 0.05) & +(1.21 - 0.01) \end{bmatrix} =$$

$$= \begin{bmatrix} 2 & 0.04 & -0.52 \\ 0.04 & 1.84 & 0.28 \\ -0.52 & 0.28 & 1.2 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{|S^2|} \propto [S^2]' = \frac{1}{1.936} = \begin{bmatrix} \frac{2}{1.936} & \frac{0.04}{1.936} & \frac{-0.52}{1.936} \\ \frac{0.04}{1.936} & \frac{1.84}{1.936} & \frac{0.28}{1.936} \\ \frac{-0.52}{1.936} & \frac{0.28}{1.936} & \frac{1.2}{1.936} \end{bmatrix}$$

$$[S^2]^{-1} = \begin{bmatrix} 1.03306 & 0.02066 & -0.2686 \\ 0.02066 & 0.95041 & 0.14463 \\ -0.2686 & 0.14463 & 0.61984 \end{bmatrix}$$

Obtener T^2

$$T^2 = n[\widehat{M} - M_{H_0}]'[S^2]^{-1}[\widehat{M} - M_{H_0}]$$

$$T^2 = 6[8.5 - 10; 8.5 - 10; 8.5 - 10][S^2]^{-1} \begin{bmatrix} 5.43 - 7 \\ 7.57 - 9 \\ 4.71 - 7 \end{bmatrix}$$

$$T^2 = 6[-1.57; -1.43; -2.29][S^2]^{-1} \begin{bmatrix} -1.5 \\ -1.5 \\ -1.5 \end{bmatrix}$$

$$T^2 = 6[-1.5; -1.5; -1.5] \begin{bmatrix} 1.03306 & 0.02066 & -0.2686 \\ 0.02066 & 0.95041 & 0.14463 \\ -0.2686 & 0.14463 & 0.61984 \end{bmatrix} \begin{bmatrix} -1.5 \\ -1.5 \\ -1.5 \end{bmatrix}$$

$$T^2 = [-9; -9; -9] \begin{bmatrix} 1.03306 & 0.02066 & -0.2686 \\ 0.02066 & 0.95041 & 0.14463 \\ -0.2686 & 0.14463 & 0.61984 \end{bmatrix} \begin{bmatrix} -1.5 \\ -1.5 \\ -1.5 \end{bmatrix}$$

$$T^2 = [-7.06612 - 10.04132 - 4.46281] \begin{bmatrix} -1.5 \\ -1.5 \\ -1.5 \end{bmatrix}$$

$$T^2 = \underline{32.35537}$$

$$Fcalc = \frac{n - p}{p(n - 1)} T^2$$

$$Fcalc = \frac{6 - 3}{3(6 - 1)} 32.35537 = .2 * 32.35537 = \underline{6.47107}$$

3. Región de rechazo

$F_{\propto.p.n-p}$
$F_{0.05.3.3} = 9.27$

$F_{0.01.3.3} = 29.45$

4. Regla de decisión

Se rechaza H_0 si Fcalc >9.27

5. Conclusión

Estadística: Como 6.47 <9.27

∴ la prueba es (NS) y no se rechaza H_0

Adva: Existe evidencia suficiente (E.E.S). para decir que estadísticamente los parámetros de las tres variables estudiadas (X_1= Evocación de la marca, X_2= Evocación del Slogan y X_3= Evocación del logotipo) es el mismo.

Resolución del inciso b)

Intervalo de confianza para marca

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.5} \frac{S}{\sqrt{n}}$$

$$M = 8.5 \mp 2.571 \frac{1.04881}{\sqrt{6}}$$

$$M = 8.5 \mp 2.571 \frac{1.04881}{2.44949}$$

$$M = 8.5 \mp 2.571(0.42817)$$

$$M = 8.5 \mp 1.10084 \begin{cases} LIC = 7.39916 \\ LSC = 9.60084 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 6, el verdadero valor promedio de la evocación de la marca oscila entre 7.39916 y 9.60084. Como 10 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

Intervalo de confianza para slogan

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.5} \frac{S}{\sqrt{n}}$$

$$M = 8.5 \mp 2.571 \frac{1.04881}{\sqrt{6}}$$

$$M = 8.5 \mp 2.571 \frac{1.04881}{2.44949}$$

$$M = 8.5 \mp 2.571(0.42817)$$

$$M = 8.5 \mp 1.10084 \begin{cases} LIC = 7.39916 \\ LSC = 9.60084 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 6, el verdadero valor promedio de la evocación del slogan oscila entre 7.39916 y 9.60084. Como 10 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

Intervalo de confianza para logotipo

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.5} \frac{S}{\sqrt{n}}$$

$$M = 8.5 \mp 2.571 \frac{1.3784}{\sqrt{6}}$$

$$M = 8.5 \mp 2.571 \frac{1.3784}{2.44949}$$

$$M = 8.5 \mp 2.571(0.56273)$$

$$M = 8.5 \mp 1.44678 \begin{cases} LIC = 7.05322 \\ LSC = 9.94678 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 6, el verdadero valor promedio de la evocación del logotipo oscila entre 7.05322 y 9.94678. Como 10 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

3.2.2.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 3.2.2.1 Diferencia de Medias Multivariada**

El Director de Mercadotecnia de una cadena de cines desea medir el grado de influencia que llega a tener el medio masivo utilizado para su campaña de publicidad en los consumidores. Esta influencia se maneja por medio del "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Los tipos de medios manejados son dos: Prensa escrita y Revistas.

Una vez lanzada la campaña en la PRENSA ESCRITA y en REVISTAS, se tomó una muestra aleatoria de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación

del logotipo de 5 potenciales clientes que leyeron la campaña en la prensa escrita y 6 que la vieron en Revistas y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados fueron los siguientes:

Medio utilizado					
Prensa escrita			Revistas		
Marca	Slogan	Logotipo	Marca	Slogan	Logotipo
4	8	3	4	6	3
6	7	3	5	4	6
5	9	4	7	9	8
5	9	6	3	8	4
5	7	4	6	7	5
			8	6	8
			7	7	6

a) Existe evidencia suficiente que indique que el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines es el mismo en ambos medios utilizados?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de Hipótesis

$$H_0: \begin{bmatrix} M_{11} \\ M_{21} \\ M_{31} \end{bmatrix} = \begin{bmatrix} M_{12} \\ M_{22} \\ M_{32} \end{bmatrix}$$

$H_1: Al\ menos\ 1\ combinación\ es\ \neq$

2. Estadístico de prueba

$$Fcalc = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

Donde $T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)^1 [S^2]^{-1} (\widehat{M}_1 - \widehat{M}_2)$

$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2}$

$W_1 = SCPC_1 = D'D_1$

$W_2 = SCPC_2 = D_2' D_2$

$$D_1 = \begin{bmatrix} 4 & 8 & 3 \\ 6 & 7 & 3 \\ 5 & 9 & 4 \\ 5 & 9 & 6 \\ 5 & 7 & 4 \end{bmatrix} - \begin{bmatrix} 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \\ 5 & 8 & 4 \end{bmatrix} = \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 0 \end{bmatrix}$$

$$D_1' D_1 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 1 & -1 \\ -1 & -1 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 & -1 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix}$$

$$W_1 = D_1' D_1 = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 4 & 6 & 3 \\ 5 & 4 & 6 \\ 7 & 9 & 8 \\ 3 & 8 & 4 \\ 6 & 7 & 5 \\ 8 & 6 & 8 \\ 7 & 7 & 6 \end{bmatrix} - \begin{bmatrix} 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \end{bmatrix} = \begin{bmatrix} -1.71 & -0.71 & -2.71 \\ -0.71 & -2.71 & 0.29 \\ 1.29 & 2.29 & 2.29 \\ -2.71 & 1.29 & -1.71 \\ 0.29 & 0.29 & -0.71 \\ 2.29 & -0.71 & 2.29 \\ 1.29 & 0.29 & 0.29 \end{bmatrix}$$

$$D_2' D_2 = \begin{bmatrix} -1.71 & -0.71 & 1.29 & -2.71 & 0.29 & 2.29 & 1.29 \\ -0.71 & -2.71 & 2.29 & 1.29 & 0.29 & -0.71 & 0.29 \\ -2.71 & 0.29 & 2.29 & -1.71 & -0.71 & 2.29 & 0.29 \end{bmatrix} \begin{bmatrix} -1.71 & -0.71 & -2.71 \\ -0.71 & -2.71 & 0.29 \\ 1.29 & 2.29 & 2.29 \\ -2.71 & 1.29 & -1.71 \\ 0.29 & 0.29 & -0.71 \\ 2.29 & -0.71 & 2.29 \\ 1.29 & 0.29 & 0.29 \end{bmatrix} = \begin{bmatrix} 19.42857 & 1.15857 & 17.42857 \\ 1.42857 & 15.42857 & 2.42857 \\ 17.42857 & 2.42857 & 21.42857 \end{bmatrix}$$

$$W_2 = D_2' D_2 = \begin{bmatrix} 19.42857 & 1.15857 & 17.42857 \\ 1.42857 & 15.42857 & 2.42857 \\ 17.42857 & 2.42857 & 21.42857 \end{bmatrix}$$

$$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2} = \frac{\begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 3 \\ 0 & 3 & 6 \end{bmatrix} + \begin{bmatrix} 19.42857 & 1.15857 & 17.42857 \\ 1.42857 & 15.42857 & 2.42857 \\ 17.42857 & 2.42857 & 21.42857 \end{bmatrix}}{5 + 7 - 2} = \begin{bmatrix} 2.14286 & 0.04286 & 1.74286 \\ 0.04286 & 1.94286 & 0.54286 \\ 1.74286 & 0.54286 & 2.74286 \end{bmatrix}$$

$$|S^2| = \begin{bmatrix} 2.14286 & 0.04286 & 1.74286 \\ 0.04286 & 1.94286 & 0.54286 \\ 1.74286 & 0.54286 & 2.74286 \\ 2.14286 & 0.04286 & 1.74286 \\ 0.04286 & 1.94286 & 0.54286 \end{bmatrix} = 4.96229$$

Como es $\neq$ de 0 tiene solución

Transpuesta de S^2

$$[S^2]^1 = \begin{bmatrix} 2.14286 & 0.04286 & 1.74286 \\ 0.04286 & 1.94286 & 0.54286 \\ 1.74286 & 0.54286 & 2.74286 \end{bmatrix}$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 1.94286 & 0.54286 \\ 0.54286 & 2.74286 \end{vmatrix} & -\begin{vmatrix} 0.04286 & 0.54286 \\ 1.74286 & 2.74286 \end{vmatrix} & +\begin{vmatrix} 0.04286 & 1.94286 \\ 1.74286 & 0.54286 \end{vmatrix} \\ -\begin{vmatrix} 0.04286 & 1.74286 \\ 0.54286 & 2.74286 \end{vmatrix} & +\begin{vmatrix} 2.14286 & 1.74286 \\ 1.74286 & 2.74286 \end{vmatrix} & -\begin{vmatrix} 2.14286 & 0.04286 \\ 1.74286 & 0.54286 \end{vmatrix} \\ +\begin{vmatrix} 0.04286 & 1.74286 \\ 1.94286 & 0.54286 \end{vmatrix} & -\begin{vmatrix} 2.14286 & 1.74286 \\ 0.04286 & 0.54286 \end{vmatrix} & +\begin{vmatrix} 2.14286 & 0.04286 \\ 0.04286 & 1.94286 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(5.32898 - 0.29469) & -(0.11755 - 0.94612) & +(0.02327 - 3.38612) \\ -(0.11755 - 0.94612) & +(5.87755 - 3.03755) & -(1.16327 - 0.07469) \\ +((0.02327 - 3.38612) & -(1.16327 - 0.07469) & +(4.16327 - 0.00184) \end{bmatrix} =$$

$$= \begin{bmatrix} 5.03429 & 0.082857 & -3.36286 \\ 0.082857 & 2.84 & -1.08857 \\ -3.36286 & .1.08857 & 4.16143 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{4.96229} = \begin{bmatrix} 5.03429 & 0.082857 & -3.36286 \\ 0.082857 & 2.84 & -1.08857 \\ -3.36286 & .1.08857 & 4.16143 \end{bmatrix} = \begin{bmatrix} 1.01451 & 0.16697 & -0.67768 \\ 0.16697 & 0.57232 & -0.21937 \\ -0.67768 & -0.21937 & 0.83861 \end{bmatrix}$$

$$T^2 = \left(\frac{n_1 n_2}{n_1+n_2}\right)\left(\hat{M}_1 - \hat{M}_2\right)[S^2]^{-1}\left(\hat{M}_1 - \hat{M}_2\right)$$

$$T^2 = \left(\frac{5*7}{5+7}\right)(5-5.71; 8-6.71; 4-5.71)[S^2]^{-1}\begin{bmatrix} 5-5.71 \\ 8-6.71 \\ 4-5.71 \end{bmatrix}$$

$$T^2 = \left(\frac{35}{12}\right)(-0.71; 1.29; -1.71)[S^2]^{-1}\begin{bmatrix} -0.71 \\ 1.29 \\ -1.71 \end{bmatrix}$$

$$T^2 = (2.91667)(-0.71; 1.29; -1.71)[S^2]^{-1}\begin{bmatrix} -0.71 \\ 1.29 \\ -1.71 \end{bmatrix}$$

$$T^2 = [-2.0833; 3.75; -5]\begin{bmatrix} 1.01451 & 0.16697 & -0.67768 \\ 0.16697 & 0.57232 & -0.21937 \\ -0.67768 & -0.21937 & 0.83861 \end{bmatrix}\begin{bmatrix} -0.71 \\ 1.29 \\ -1.71 \end{bmatrix}$$

$$T^2 = [1.9010; 2.8952; -3.60385]\begin{bmatrix} -0.71 \\ 1.29 \\ -1.71 \end{bmatrix}$$

$$T^2 = \underline{8.54253}$$

$$Fcalc = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2$$

$$Fcalc = \frac{5+7-3-1}{3(5+7-2)} 8.54253$$

$$Fcalc = \frac{8}{30} 8.54253$$

$$Fcalc = 0.26667 \times 8.54253$$

$$\boldsymbol{Fcalc = 2.278}$$

3. Región de rechazo

$$F_{\alpha.p.n1+n2-p-1} = F_{0.05.3.5+7-3-1} = F_{0.05.3.8} = 4.07$$

$$F_{\alpha.p.n1+n2-p-1} = F_{0.01.3.5+7-3-1} = F_{0.01.3.8} = 7.59$$

4. Regla de decisión

Se rechaza H_0 si Fcalc > 4.07

5. Conclusión:

- Estadistica: Como 2.278< 4.07 ∴

La prueba es N.S y no se rechaza H_0

- Adva: Existe evidencia suficiente para decir el grado de evocación o recuerdo de la Marca, el Slogan y el Logotipo que tiene el consumidor sobre esta cadena de cines es el mismo en ambos medios utilizados.

Resolución del inciso b)

Nota importante: Para construir el intervalo de confianza se debe probar primero si las varianzas de las muestras son iguales o no para saber que caso de debe usar para construir el intervalo de confianza (Caso 2 para varianzas iguales o Caso 3 para varianzas desiguales) por lo que primero hay que realizar la prueba de homocedasticidad de varianzas.

Prueba de Homocedasticidad para Marca

1. Juego de Hipótesis

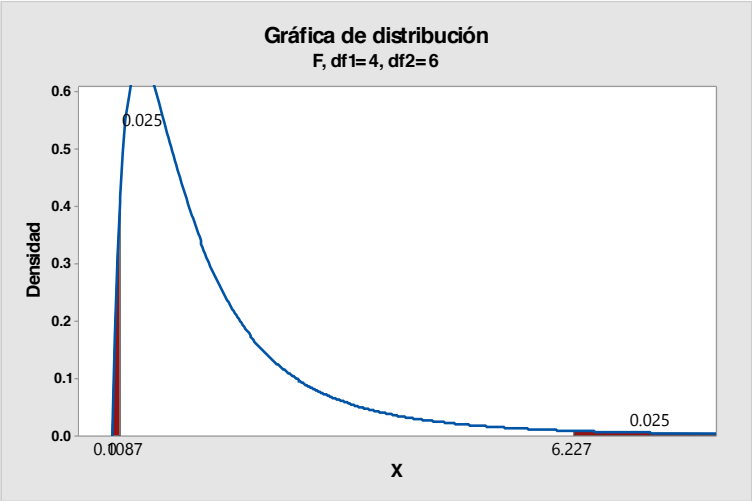
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{11}}^2}{{S_{12}}^2} = \frac{0.49985}{3.24} = 0.15427$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc ≥ 6.23 ó si es ≤ 0.1087

5. Conclusión:

- Estadistica: Como 0.1087< 0.15427< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de Confianza para Marca

$$M_{11} - M_{12} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp (t_{\alpha/2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)0.49985 + (7-1)3.24}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 0.49985 + 6 * 3.24}{10}}$$

$$Sp = \sqrt{\frac{1.9994 + 19.44}{10}}$$

$$Sp = \sqrt{\frac{21.4394}{10}}$$

$$Sp = \sqrt{2.14394}$$

$$Sp = 1.46422$$

$$M_{11} - M_{12} = (5 - 5.71) \mp 2.228 * 1.46422 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (-0.71) \mp 2.228 * 1.46422 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (-0.71) \mp 2.228 * 1.46422 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (-0.71) \mp 2.228 * 1.46422 * 0.58554$$

$$M_{11} - M_{12} = (-0.71) \mp 1.9102 \begin{cases} LIC = -2.62 \\ LSC = 1.20 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la evocación de la marca en la prensa escrita es la misma que la evocación de la marca en las revistas.

Nota: Como la Evocación de la Marca es el mismo en ambos medios de publicidad es razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se haya rechazado por esta variable.

Comprobación utilizando el Software Minitab 17

Prueba e IC para dos varianzas: Marca vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      0.707     0.500  (0.282, 2.916)
2        7      1.799     3.238  (1.055, 4.264)

Relación de desviaciones estándar = 0.393
Relación de varianzas = 0.154
```

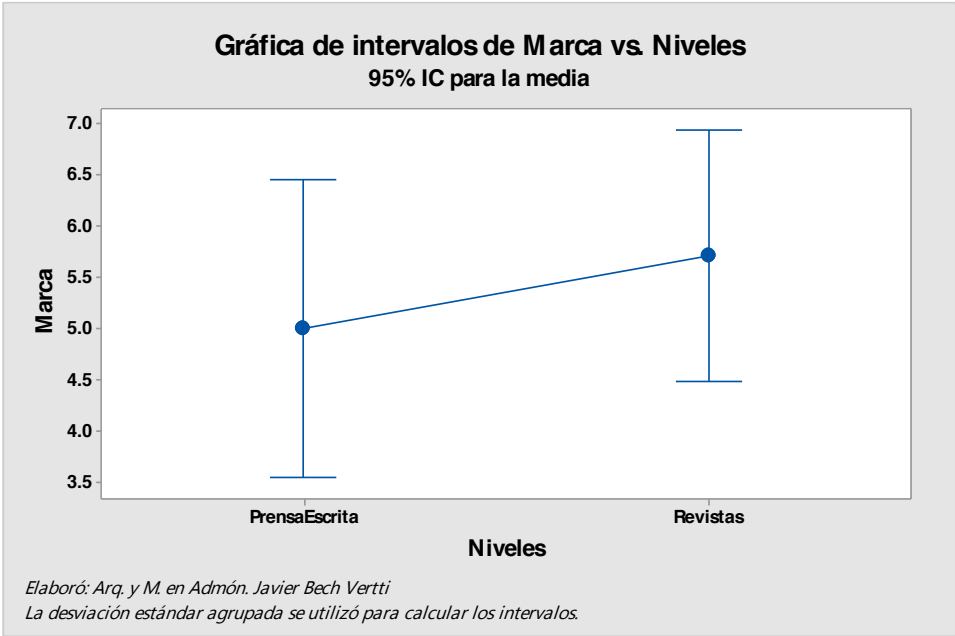
ANOVA unidireccional: Marca vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      1.488      1.488     0.69    0.424
Error    10     21.429      2.143
Total    11     22.917


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
1.46385    6.49%       0.00%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  5.000      0.707  (3.541, 6.459)
2        7  5.714      1.799  (4.481, 6.947)

Desv.Est. agrupada = 1.46385
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            0.714       0.857  (-1.196, 2.624)     0.83     0.424

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Marca vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba de Homocedasticidad para Slogan

1. Juego de Hipótesis

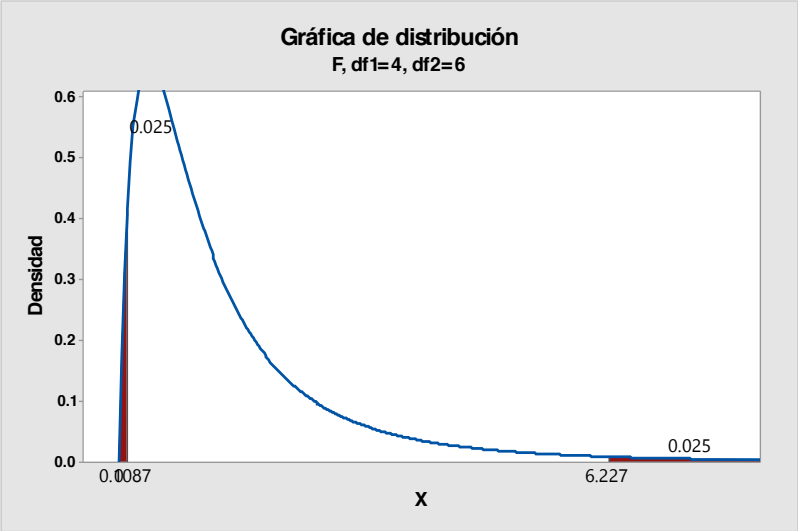
$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$

$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{21}}^2}{{S_{22}}^2} = \frac{1}{2.56} = 0.39063$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:
   - Estadística: Como 0.1087< 0.39063< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
   - Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de Confianza para Slogan

$$M_{21} - M_{22} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp t\,{}^{\propto}/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)1 + (7-1)\,2.56}{5+7-2}}$$

$$Sp = \sqrt{\frac{4*1 + 6*2.56}{10}}$$

$$Sp = \sqrt{\frac{4 + 15.36}{10}}$$

$$Sp = \sqrt{\frac{19.36}{10}}$$

$$Sp = \sqrt{1.936}$$

$$Sp = 1.39$$

$$M_{11} - M_{12} = (8 - 6.71) \mp 2.228 * 1.39 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (1.29) \mp 2.228 * 1.39 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (1.29) \mp 2.228 * 1.39 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (1.29) \mp 2.228 * 1.39 * 0.58554$$

$$M_{11} - M_{12} = (1.29) \mp 1.81337 \begin{cases} LIC = -0.523 \\ LSC = 3.103 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la evocación de la marca en la prensa escrita es la misma que la evocación de la marca en las revistas.

Nota: Como la Evocación del Slogan es el mismo en ambos medios de publicidad es razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se haya rechazado por esta variable.

Comprobación utilizando el Software Minitab 17

Prueba e IC para dos varianzas: Slogan vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      1.000     1.000  (0.610, 2.696)
2        7      1.604     2.571  (0.880, 4.060)

Relación de desviaciones estándar = 0.624
Relación de varianzas = 0.389
```

ANOVA unidireccional: Slogan vs. Niveles

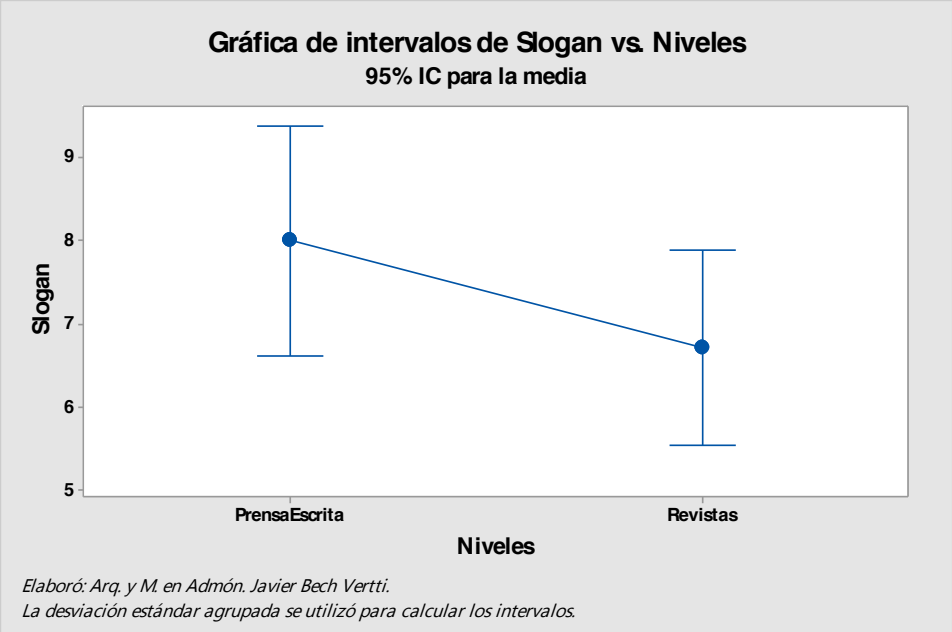
```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      4.821      4.821     2.48    0.146
Error    10     19.429      1.943
Total    11     24.250


Resumen del modelo

                    R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
```

```
1.39386   19.88%      11.87%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  8.000      1.000  (6.611, 9.389)
2        7  6.714      1.604  (5.540, 7.888)

Desv.Est. agrupada = 1.39386
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia       de las       EE de                              Valor p
de niveles       medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            -1.286       0.816  (-3.104, 0.533)    -1.58     0.146

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Slogan vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba de Homocedasticidad para Logotipo

1. Juego de Hipótesis

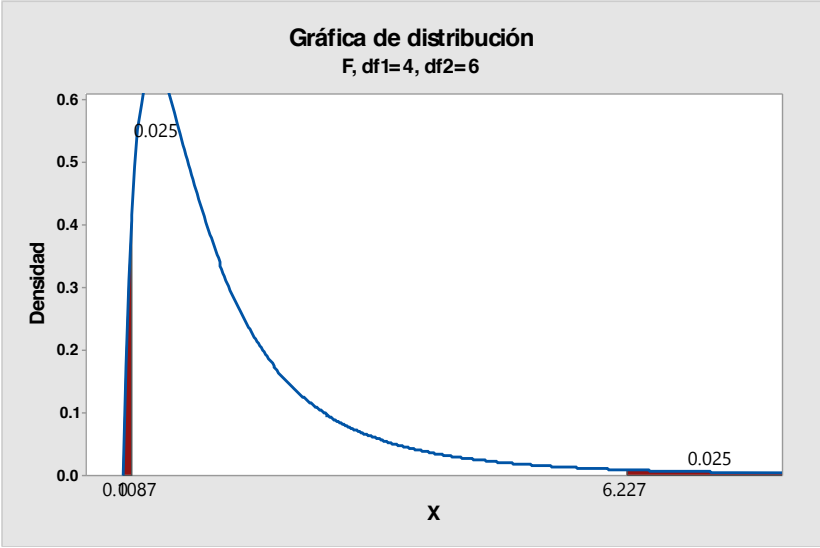
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{31}}^2}{{S_{32}}^2} = \frac{1.4884}{3.5721} = 0.41667$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:

- Estadística: Como 0.1087< 0.41667< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de Confianza para Logotipo

$$M_{31} - M_{32} = \left(\overline{X}_{31} - \overline{X}_{32}\right) \mp t^{\propto}/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = Sp = \sqrt{\frac{(5-1)1.4884 + (7-1)\ 3.5721}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 1.4884 + 6 * 3.5721}{10}}$$

$$Sp = \sqrt{\frac{5.9536 + 214326}{10}}$$

$$Sp = \sqrt{\frac{27.38620}{10}}$$

$$Sp = \sqrt{2.73862}$$

$$Sp = 1.65488$$

$$M_{11} - M_{12} = (4 - 5.71) \mp 2.228 * 1.65488 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (-1.71) \mp 2.228 * 1.65488 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (-1.71) \mp 2.228 * 1.65488 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (-1.71) \mp 2.228 * 1.65488 * 0.58554$$

$$M_{11} - M_{12} = (-1.71) \mp 2.15893 \begin{cases} LIC = -3.869 \\ LSC = 0.449 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la evocación de la marca en la prensa escrita es la misma que la evocación de la marca en las revistas.

Nota: Como la Evocación del Logotipo es el mismo en ambos medios de publicidad es razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se haya rechazado por esta variable.

Comprobación utilizando el Software Minitab 17

Prueba e IC para dos varianzas: Logotipo vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      1.225     1.500  (0.489, 5.050)
2        7      1.890     3.571  (1.204, 4.118)

Relación de desviaciones estándar = 0.648
Relación de varianzas = 0.420
```

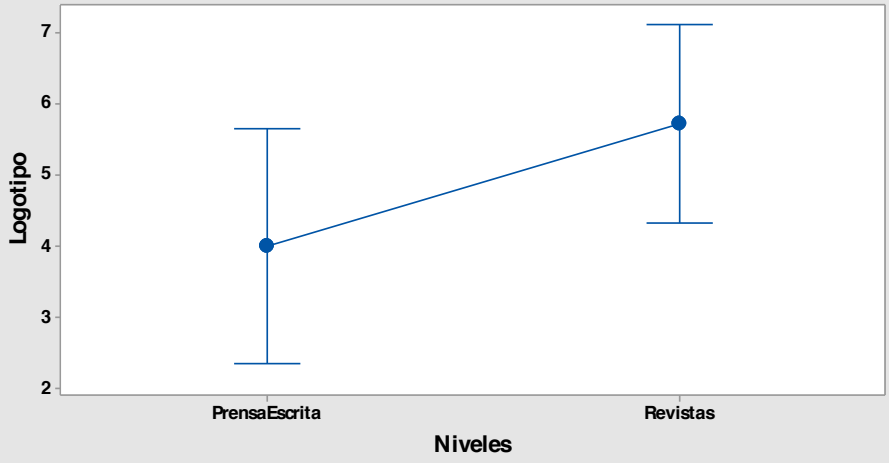
ANOVA unidireccional: Logotipo vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      8.571      8.571     3.13    0.108
Error    10     27.429      2.743
Total    11     36.000


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
1.65616   23.81%      16.19%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  4.000      1.225  (2.350, 5.650)
2        7  5.714      1.890  (4.320, 7.109)

Desv.Est. agrupada = 1.65616
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            1.714       0.970  (-0.446, 3.875)     1.77     0.108

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Logotipo vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

3.2.3.1 ACTIVIDAD DE **APRENDIZAJE**

ACTIVIDAD DE **APRENDIZAJE 3.2.3.1 Anova Multivariada (MANOVA)**

El Director de mercadotecnia de una compañía de Bienes Raíces desea medir el grado de influencia que llega a tener la zona donde se localizan las casas unifamiliares de cierta ciudad en las magnitudes de los atributos de las mismas. Esta influencia es medida a través de tres variables dependientes: Valor comercial de la casa, Tamaño del lote y el Número de habitaciones. Las zonas manejadas son tres: Norte, Sur y Este. El efecto se medirá con base en el grado de influencia que la Zona de la Ciudad ejerce sobre cada una de las tres variables dependientes mediante los siguientes criterios:

1.- Que tanto la Zona de la Ciudad llega a influir sobre el Valor de las casas.

2.- Que tanto la Zona de la Ciudad llega a influir sobre el Tamaño del lote donde están construidas las casas.

3.- Que tanto la Zona de la Ciudad llega a influir sobre el Número de habitaciones que tienen las casas.

Nota importante: Donde se realizó la encuesta el número de habitaciones incluye todos los espacios no solo las recámaras, es decir, sala, comedor, cocina, etc.)

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Valor comercial de la casa (en cientos de dólares) X_2= Tamaño del lote (en cientos de pies cuadrados) y X_3= Número de habitaciones, de 5 propietarios de casas en la zona Norte, 6 en la zona Sur y 7 en la zona Este. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Zona de la ciudad								
Norte			Sur			Este		
Valor	Tamaño	Habita-ciones	Valor	Tamaño	Habita-ciones	Valor	Tamaño	Habita-ciones
1081	367	6	1093	186	4	595	111	4
1214	96	4	931	186	4	905	111	5
1204	226	5	890	130	3	607	111	4
1255	136	4	931	135	3	866	102	5
1053	120	3	911	126	4	758	105	4
			718	130	4	850	102	4
						870	102	4

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de las casas es el mismo en las tres zonas?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si la zona de la ciudad influye sobre las magnitudes promedio de los atributos de las casas.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que zona o zonas los atributos son mayores y en que magnitud.

Resolución del inciso a)

Prueba de hipótesis:

1. Juego de Hipótesis

$$H_0 = \begin{bmatrix}\mu_{11}\\ \mu_{21}\\ \mu_{31}\end{bmatrix} = \begin{bmatrix}\mu_{12}\\ \mu_{22}\\ \mu_{32}\end{bmatrix} = \begin{bmatrix}\mu_{13}\\ \mu_{23}\\ \mu_{33}\end{bmatrix}$$

2. Estadístico de prueba

Como No. de Variables Dependientes : p≥1 y

No. de Niveles del Factor: g=3

Usamos la siguiente fórmula (Caso 4):

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Donde\ £ = \frac{|error|}{|tratamiento|} = \frac{|SCE|}{|SCT|}$$

$$SCE = \sum_{i=1}^{3} SCEi = \sum_{i=1}^{3} D_i' D_i$$

$$SCT = D'D$$

Para SCE:

$$D_1 = \begin{bmatrix} 1081 & 367 & 6 \\ 1214 & 96 & 4 \\ 1204 & 226 & 5 \\ 1255 & 136 & 4 \\ 1053 & 120 & 3 \end{bmatrix} - \begin{bmatrix} 1161.4 & 189 & 4.4 \\ 1161.4 & 189 & 4.4 \\ 1151.4 & 189 & 4.4 \\ 1161.4 & 189 & 4.4 \\ 1161.4 & 189 & 4.4 \end{bmatrix} = \begin{bmatrix} -80.4 & 178 & 1.60 \\ 52.6 & -93 & -0.4 \\ 42.6 & 37 & 0.6 \\ 93.6 & -53 & -0.4 \\ -108.4 & -69 & -1.4 \end{bmatrix}$$

$$D'_1 D_1 = \begin{bmatrix} -80.4 & 52.6 & 42.6 & 93.6 & -108.4 \\ 178 & -93 & 37 & -53 & -69 \\ 1.6 & -0.4 & 0.6 & -0.4 & -1.4 \end{bmatrix} \begin{bmatrix} -80.4 & 178 & 1.60 \\ 52.6 & -93 & -0.4 \\ 42.6 & 37 & 0.6 \\ 93.6 & -53 & -0.4 \\ -108.4 & -69 & -1.4 \end{bmatrix}$$

$$= \begin{bmatrix} 31557.2 & -15108 & -9.8 \\ -15108 & 49272 & 462 \\ -9.8 & 462 & 5.2 \end{bmatrix}$$

$$D^2 = \begin{bmatrix} 1093 & 186 & 4 \\ 931 & 186 & 4 \\ 890 & 130 & 3 \\ 931 & 135 & 3 \\ 911 & 126 & 4 \\ 718 & 130 & 4 \end{bmatrix} - \begin{bmatrix} 912.3 & 148.8 & 3.7 \\ 912.3 & 148.8 & 3.7 \\ 912.3 & 148.8 & 3.7 \\ 912.3 & 148.8 & 3.7 \\ 912.3 & 148.8 & 3.7 \\ 912.3 & 148.8 & 3.7 \end{bmatrix} = \begin{bmatrix} 180.66667 & 37.16667 & 0.33333 \\ 18.66667 & 37.16667 & 0.33333 \\ -22.33333 & -18.83333 & -0.66667 \\ 18.66667 & -13.83333 & -0.66667 \\ -1.33333 & -22.83333 & 0.33333 \\ -194.33333 & -18.83333 & 0.33333 \end{bmatrix}$$

$$D'_2 D_2 = \begin{bmatrix} 180.66667 & 18.66667 & -22.33333 & 18.66667 & -1.33333 & -194.33333 \\ 37.16667 & 37.16667 & -18.83333 & -13.83333 & -22.83333 & -18.83333 \\ 0.33333 & 0.33333 & -0.66667 & -0.66667 & 0.33333 & 0.33333 \end{bmatrix}$$

$$\begin{bmatrix} 180.66667 & 37.16667 & 0.33333 \\ 18.66667 & 37.16667 & 0.33333 \\ -22.33333 & -18.83333 & -0.66667 \\ 18.66667 & -13.83333 & -0.66667 \\ -1.33333 & -22.83333 & 0.33333 \\ -194.33333 & -18.83333 & 0.33333 \end{bmatrix} = \begin{bmatrix} 71603.33333 & 11261.33333 & 3.66667 \\ 11261.33333 & 4184.83333 & 32.66667 \\ 3.66667 & 32.66667 & 1.33333 \end{bmatrix}$$

$$D_3 = \begin{bmatrix} 595 & 111 & 4 \\ 905 & 111 & 5 \\ 607 & 111 & 4 \\ 866 & 102 & 5 \\ 758 & 105 & 4 \\ 850 & 102 & 4 \\ 870 & 102 & 4 \end{bmatrix} \begin{bmatrix} 778.7 & 106.3 & 4.3 \\ 778.7 & 106.3 & 4.3 \\ 778.7 & 106.3 & 4.3 \\ 778.7 & 106.3 & 4.3 \\ 778.7 & 106.3 & 4.3 \\ 778.7 & 106.3 & 4.3 \\ 778.7 & 106.3 & 4.3 \end{bmatrix} = \begin{bmatrix} -183.71429 & 4.71429 & -0.28571 \\ 126.28571 & 4.71429 & 0.71429 \\ -171.71429 & 4.71429 & -0.28571 \\ 87.28571 & -4.28571 & 0.71429 \\ -20.71429 & -1.28571 & -0.28571 \\ 71.28571 & -4.28571 & -0.28571 \\ 91.28571 & -4.28571 & -0.28571 \end{bmatrix}$$

$$D'_3 D_3 = \begin{bmatrix} 183.71429 & 126.28571 & -171.71429 & 87.28571 & -20.71429 & 71.28571 & 91.28571 \\ 4.71429 & 4.71429 & 4.71429 & -4.28571 & -1.28571 & -4.28571 & -4.28571 \\ -0.28571 & 0.71429 & -0.28571 & 0.71429 & -0.28571 & -0.28571 & -0.28571 \end{bmatrix}$$

$$\begin{bmatrix} -183.71429 & 4.71429 & -0.28571 \\ 126.28571 & 4.71429 & 0.71429 \\ -171.71429 & 4.71429 & -0.28571 \\ 87.28571 & -4.28571 & 0.71429 \\ -20.71429 & -1.28571 & -0.28571 \\ 71.28571 & -4.28571 & -0.28571 \\ 91.28571 & -4.28571 & -0.28571 \end{bmatrix} = \begin{bmatrix} 100647.42857 & -2124.42857 & 213.57143 \\ -2124.42857 & 123.42857 & 0.42857 \\ 213.57143 & 0.42857 & 1.42857 \end{bmatrix}$$

$$SCE = \sum_{i=1}^{3} D'_i D_i$$

$$\begin{bmatrix} 31557.2 & -15108 & -9.8 \\ -15108 & 49272 & 462 \\ -9.8 & 462 & 5.2 \end{bmatrix} + \begin{bmatrix} 71603.33333 & 11261.33333 & 3.66667 \\ 11261.33333 & 4184.83333 & 32.66667 \\ 3.66667 & 32.66667 & 1.33333 \end{bmatrix}$$

$$+ \begin{bmatrix} 100647.42857 & -2124.42857 & 213.57143 \\ -2124.42857 & 123.42857 & 0.42857 \\ 213.57143 & 0.42857 & 1.42857 \end{bmatrix}$$

$$SCE = \begin{bmatrix} 203807.42857 & -5971.09524 & 207.43810 \\ -5971.09524 & 53580.26190 & 495.09524 \\ 207.43810 & 495.09524 & 7.96190 \end{bmatrix}$$

Utilizando la regla de Sarrus:

$$|SCE| = \begin{vmatrix} 203807.42857 & -5971.09524 & 207.43810 \\ -5971.09524 & 53580.26190 & 495.09524 \\ 207.43810 & 495.09524 & 7.96190 \\ 203807.96190 & -5971.09524 & 207.43810 \\ -5971.09524 & 53580.26190 & 495.09524 \end{vmatrix} = 3.3171E + 10$$

Para SCT:

$$D\begin{bmatrix}1081 & 367 & 6\\ 1214 & 96 & 4\\ 1204 & 226 & 5\\ 1255 & 136 & 4\\ 1053 & 120 & 3\\ 1093 & 186 & 4\\ 931 & 186 & 4\\ 890 & 130 & 3\\ 931 & 135 & 3\\ 911 & 126 & 4\\ 718 & 130 & 4\\ 595 & 111 & 4\\ 905 & 111 & 5\\ 607 & 111 & 4\\ 866 & 102 & 5\\ 758 & 105 & 4\\ 850 & 102 & 4\\ 870 & 102 & 4\end{bmatrix} - \begin{bmatrix}929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 43.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\\ 929.6 & 143.44 & 4.11\end{bmatrix}$$

$$= \begin{bmatrix} 151.44444 & 223.55556 & 1.88889\\ 284.44444 & -47.44444 & -0.11111\\ 274.44444 & 82.55556 & 0.88889\\ 325.44444 & -7.44444 & -0.11111\\ 123.44444 & -23.44444 & -1.11111\\ 163.44444 & 42.55556 & -0.11111\\ 1.444444 & 42.55556 & -0.111111\\ -39.55556 & -13.44444 & -1.11111\\ 1.44444 & -8.44444 & -1.11111\\ -18.55556 & -17.44444 & -0.11111\\ -211.55556 & -13.44444 & -0.11111\\ -334.55556 & -32.44444 & -0.11111\\ -24.55556 & -32.44444 & 0.88889\\ -322.55556 & -32.44444 & -0.11111\\ -63.55556 & -41.44444 & 0.88889\\ -171.5556 & -38.44444 & -0.11111\\ -79.55556 & -41.44444 & -0.11111\\ -59.55556 & -41.44444 & -0.11111\end{bmatrix}$$

$$SCT = D'D = \begin{bmatrix}151.4444 & 284.44444 & 274.44444 & \dots & -59.55556\\ 223.55556 & -47.44444 & 82.55556 & \dots & -41.44444\\ 1.88889 & -0.11111 & 0.88889 & \dots & -0.11111\end{bmatrix}\begin{bmatrix} 151.44444 & 223.55556 & 1.88889\\ 284.44444 & -47.44444 & -0.11111\\ 274.44444 & 82.55556 & 0.88889\\ \vdots & \vdots & \vdots\\ -59.55556 & -41.44444 & -0.11111\end{bmatrix}$$

$$= \begin{bmatrix}633618.44444 & 85516.55556 & 403.88889\\ 85516.55556 & 73796.44444 & 501.11111\\ 403.88889 & 501.11111 & 9.77778\end{bmatrix}$$

Utilizando la regla de Sarrus:

$$|SCT| = \begin{vmatrix} 633618.44444 & 85516.55556 & 403.88889 \\ 85516.55556 & 73796.44444 & 501.11111 \\ 403.88889 & 501.11111 & 9.77778 \\ 633618.44444 & 85516.55556 & 403.88889 \\ 85516.55556 & 73796.44444 & 501.11111 \end{vmatrix} = 2.4916E + 11$$

$$£ = \frac{|SCE|}{|SCT|} = \frac{3.3171E + 10}{2.4916E + 11} = \underline{0.13313}$$

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p} \frac{(1 - \sqrt{£})}{\sqrt{£}}$$

$$Fcalc = \frac{(18 - 3 - 2)}{3} \frac{(1 - \sqrt{0.13313})}{\sqrt{0.13313}}$$

$$Fcalc = \frac{(13)}{3} \frac{(1 - 0.36487)}{0.36487}$$

$$Fcalc = (4.333)(1.74070)$$

$$Fcalc = 7.54246$$

3. Región de rechazo

$$Fcrit = \left[\propto, 2, p, 2\left(\sum n_1 - p - 2\right)\right]$$

$$Fcrit = \left[0.05,2(3),2\left(\sum 18 - 3 - 2\right)\right]$$

$$Fcrit = [0.05,6,2(13)]$$

$$Fcrit = 0.05,6,26$$

$$F0.05 = 2.47$$

$$F0.01 = 3.59$$

4. Regla de Decisión

Se rechaza H_0 sí $fcal \geq 2.47$

5. Conclusión

Estadista: Como 7.543 > 2.47 > 3.59 ∴ la prueba es (A.S) y se rechaza con mucha potencia.

Administrativa: Existe evidencia suficiente para decir que estadísticamente al menos una de las medias en cada una de las **tres variables dependientes** de las **Tres Zonas de la ciudad** NO son iguales. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Resolución del inciso b)

Prueba de Hipótesis para Valor:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener la ubicación de las casas unifamiliares es el mismo para el **Valor** de las casas .

$$H_0: \mu_{11.} = \mu_{12.} = \mu_{13.}$$

La hipótesis alternativa es No todas las Zonas son iguales para el **Valor** de las casas.

$$H_1: \text{No todas las zonas influyen igual en el } \mathbf{Valor} \text{ de las casas unifamiliares.}$$

También se puede pensar que la hipótesis alternativa sea "al menos una de las zonas influye de manera diferente en el **Valor** de las casas..".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio del **Valor** de las casas con base en la ubicación de las mismas.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos una de las zonas tiene una diferencia significativa para influir en el **Valo**r de las casas.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Zona	Valor							Total fila	Medias filas
	1	2	3	4	5	6	7		
Norte	1081	1214	1204	1255	1053			$X_{1.} = 5807$	$\bar{X}_{1.} = 1161.4$
Sur	1093	931	890	931	911	718		$X_{2.} = 5474$	$\bar{X}_{2.} = 912.33333$
Este	595	905	607	866	758	850	870	$X_{3.} = 5451$	$\bar{X}_{3.} = 778.71429$
					Totales de filas:			$X_{..} = 16732$	$\bar{X}.. = 929.55556$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (1081^2 + 1214^2 + \cdots + 870^2) - \frac{16732^2}{18} = 16´186{,}941.56 - 15'553{,}323.56 = 633{,}618$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{5807^2}{5} + \frac{5451^2}{6} + \frac{5451^2}{7}\right) - \frac{16732^2}{18}$$

$$= 15´983{,}134.04 - 15´553{,}323.56 = 429{,}810.48$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste

la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k}\frac{X_{i.}^2}{n} = 1081^2 + 1214^2 + \cdots + 870^2 - (\frac{5807^2}{5} + \frac{5451^2}{6} + \frac{5451^2}{7})$$
$$= 16\acute{}186{,}941.56 - 15\acute{}983{,}134.04 = 203{,}807.52$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	$F_{calculada}$
Tratamien-tos	$v_1 = k - 1$ $= 3 - 1$ $= \mathbf{2}\ \boldsymbol{g.l.}$	SCt $= 429{,}810.48$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{429{,}810.48}/_{2}$ $= 214{,}905.24$	$F_{Calc.} = {}^{CMt}/_{CME}/$ $= {}^{214{,}905.24}/_{13{,}587.168}$ $= \mathbf{15.81618}$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	SCE $= 203{,}807.52$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{203{,}807.52}/_{15}$ $= 13{,}587.168$	
Total	18-1=17	SCT $= 633{,}618$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de	**Grados de libertad en el numerador**

libertad en el denomina-dor	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **15.82**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener al menos una de las **tres zonas** utilizadas.sobre la **Ubicación** de las casas no es el mismo.

Prueba de Hipótesis para Tamaño de lote:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener la ubicación de las casas unifamiliares es el mismo para el **Tamaño del lote** de las casas .

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todas las Zonas son iguales para el **Tamaño del lote** de las casas.

H_1: No todas las zonas influyen igual en el **Valor** de las casas unifamiliares.

También se puede pensar que la hipótesis alternativa sea "al menos una de las zonas influye de manera diferente en el **Tamaño del lote** de las casas..".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio del **Tamaño del lote** de las casas con base en la ubicación de las mismas.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos una de las zonas tiene una diferencia significativa para influir en el **Tamaño del lote** de las casas.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Zona	Tamaño							Total fila	Medias filas
	1	2	3	4	5	6	7		
Norte	367	96	226	136	120			$X_{1.} = 945$	$\bar{X}_{1.} = 893{,}025$
Sur	186	186	130	135	126	130		$X_{2.} = 893$	$\bar{X}_{2.} = 797{,}449$
Este	111	111	111	102	105	102	102	$X_{3.} = 744$	$\bar{X}_{3.} = 553{,}536$
					Totales de filas:			$X_{..} = 2582$	$\bar{X}.. = 2´244{,}010$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (367^2 + 96^2 + \cdots + 102^2) - \frac{2582^2}{18} = 444{,}170 - 370.373.5556 = 73{,}796.44$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{945^2}{5} + \frac{893^2}{6} + \frac{744^2}{7}\right) - \frac{2582^2}{18} = 390{,}589.7381 - 370{,}373.5556 = 20{,}216.1825$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 367^2 + 96^2 + \cdots + 102^2 - \left(\frac{945^2}{5} + \frac{893^2}{6} + \frac{744^2}{7}\right) = 444{,}170 - 390{,}589.7381$$
$$= 53{,}580.2619$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamientos	$v_1 = k - 1$ $= 3 - 1$ $= \mathbf{2}\ \boldsymbol{g.l.}$	SCt $= 20{,}216.1825$	$CMt = {SCt}/{g.l.}$ $= {20{,}216.1825}/{2}$ $= 10{,}108.09125$	$F_{Calc.} = {CMt}/{CME} =/$ $= {10{,}108.09125}/{3{,}572.01746}$ $= 2.83$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	SCE $= 53{,}580.2619$	$CME = {SCE}/{g.l.}$ $= {53{,}580.2619}/{15}$ $= 3{,}572.01746$	
Total	18-1=17	SCT $= 73{,}796.44$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se

encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **2.83**, que es menor al valor crítico de 3.68; por lo tanto la hipótesis nula **NO** se rechaza y llegamos a la conclusión de que todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llegan a tener las **tres zonas utilizadas** sobre el **Tamaño del lote** es el mismo.

Prueba de Hipótesis para Número de habitaciones:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener la ubicación de las casas unifamiliares es el mismo para el **Número de habitaciones** de las casas .

$$H_0: \mu_{11.} = \mu_{12.} = \mu_{13.}$$

La hipótesis alternativa es No todas las Zonas son iguales para el **Número de habitaciones** de las casas.

H_1: No todas las zonas influyen igual en el **Valor** de las casas unifamiliares.

También se puede pensar que la hipótesis alternativa sea "al menos una de las zonas influye de manera diferente en el **Número de habitaciones** de las casas..".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio del **Número de habitaciones** de las casas con base en la ubicación de las mismas.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos una de las zonas tiene una diferencia significativa para influir en el **Número de habitaciones** de las casas.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Zona	Número de habitaciones							Total fila	Medias filas
	1	2	3	4	5	6	7		
Norte	6	4	5	4	3			$X_{1.} = 22$	$\bar{X}_{1.} = 484$
Sur	4	4	3	3	4	4		$X_{2.} = 22$	$\bar{X}_{2.} = 484$
Este	4	5	4	5	4	4	4	$X_{3.} = 30$	$\bar{X}_{3.} = 900$
					Totales de filas:			$X_{..} = 74$	$\bar{X}.. = 1868$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (6^2 + 4^2 + \cdots + 4^2) - \frac{74^2}{18} = 314 - 304.22222 = 9.77778$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{22^2}{5} + \frac{22^2}{6} + \frac{30^2}{7}\right) - \frac{74^2}{18} = 306.0381 - 304.22222 = 1.8159$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = (6^2 + 4^2 + \cdots + 4^2) - (\frac{22^2}{5} + \frac{22^2}{6} + \frac{30^2}{7}) = 314 - 306.0381 = 7.9619$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	$F_{calculada}$
Tratamientos	$v_1 = k - 1$ $= 3 - 1$ $= \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 1.8159$	CMt $= {SCt}/{g.l.}$ $= {1.8159}/{2}$ $= 0.90795$	$F_{Calc.} = {CMt}/{CME} =/$ $= {0.90795}/{0.5379}$ $= \mathbf{1.68795}$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 7.9619$	CME $= {SCE}/{g.l.}$ $= {7.9619}/{15}$ $= 0.5379$	
Total	18-1=17	$SCT = 9.77778$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **1.69,** que es menor al valor crítico de 3.68; por lo tanto la hipótesis nula NO se rechaza y llegamos a la conclusión de que todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llegan a tener las **tres zonas utilizadas** sobre el **Número de habitaciones** es el mismo en.

Resolución del inciso c)

Prueba de Tukey

Intervalo de confianza para Valor (Norte-Sur). Modelo desbalanceado.

$$M_{11} - M_{12} = (\bar{X}_{11} - \bar{X}_{12}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = (1161.4 - 912.3) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = 249.1 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{203{,}808}{15} = 13{,}587.2$$

$$nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{11} - M_{12} = 249.1 \mp 3.67\sqrt{\frac{13{,}587.2}{5.4545}}$$

$$M_{11} - M_{12} = 249.1 \mp 3.67\sqrt{2491}$$

$$M_{11} - M_{12} = 249.1 \mp 3.67(49.91)$$

$$M_{11} - M_{12} = 249.1 \mp 183.16968 \begin{cases} LIC = 65.93 \\ LSC = 432.27 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de influencia de la zona **Norte** sobre el **Valor** de las casas es mayor que el de la Zona **Sur** por un mínimo de 65.93 y un máximo de 432.27 cientos de dólares.

Intervalo de confianza para Valor (Norte-Este). Modelo desbalanceado

$$M_{11} - M_{13} = (\bar{X}_{11} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = (1161.4 - 778.7) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = 382.7 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = 5.8333$$

$$M_{11} - M_{13} = 382.7 \mp 3.67\sqrt{\frac{13{,}587}{5.8333}}$$

$$M_{11} - M_{13} = 382.7 \mp 3.67(48.26192)$$

$$M_{11} - M_{13} = 382.7 \mp 177.12126 \begin{cases} LIC = 205.6 \\ LSC = 559.82 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de influencia de la zona **Norte** sobre el **Valor** de las casas es mayor que el de la Zona **Este** por un mínimo de 205.6 y un máximo de 559.82 cientos de dólares.

Intervalo de confianza para Valor (Sur-Este). Modelo desbalanceado

$$M_{12} - M_{13} = (\bar{X}_{12} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = (912.3 - 778.7) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = 2.786 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = 6.4615$$

$$M_{12} - M_{13} = 133.6 \mp 3.67\sqrt{\frac{13{,}587}{6.4615}}$$

$$M_{12} - M_{13} = 133.6 \mp 3.67(45.85589)$$

$$M_{12} - M_{13} = 133.6 \mp 168.29111 \begin{cases} LIC = -34.69111 \\ LSC = 301.89 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia de la zona **Sur** sobre el **Valor** de las casas es el mismo que el de la zona **Este**.

Intervalo de confianza para Tamaño del lote (Norte-Sur). Modelo desbalanceado

$$M_{21} - M_{22} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = (189 - 148.8) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = 40.2 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{53{,}580}{15} = 3572$$

$$nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{21} - M_{22} = 40.2 \mp 3.67 \sqrt{\frac{3572}{5.4545}}$$

$$M_{21} - M_{22} = 40.2 \mp 3.67(25.59047)$$

$$M_{21} - M_{22} = 40.2 \mp 93.91702 \begin{cases} -53.71702 \\ 134.11702 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia de la zona **Norte** sobre el **Tamaño del lote** de las casas es el mismo que el de la zona **Sur**.

Intervalo de confianza para Tamaño del lote (Norte –Este). Modelo desbalanceado

$$M_{21} - M_{23} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = (189 - 106.3) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = 82.7 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{21} - M_{23} = 82.7 \mp 3.67\sqrt{\frac{3572}{5.8326}}$$

$$M_{21} - M_{23} = 82.7 \mp 3.67(24.74712)$$

$$M_{21} - M_{23} = 82.7 \mp 90.82\begin{cases} -8.12 \\ 173.52192 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia de la zona **Norte** sobre el **Tamaño del lote** de las casas es el mismo que el de la zona **Este**.

Intervalo de confianza para Tamaño del Lote (Sur-Este). Modelo desbalanceado

$$M_{22} - M_{23} = (\bar{X}_{22} - \bar{X}_{23}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = (148.8 - 106.3) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = 42.5 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{22} - M_{23} = 42.5 \mp 3.67\sqrt{\frac{3572}{6.4620}}$$

$$M_{22} - M_{23} = 42.5 \mp 3.67(23.51106)$$

$$M_{22} - M_{23} = 42.5 \mp 86.28560\begin{cases} LIC = -43.78560 \\ LSC = 128.78560 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia de la zona **Sur** sobre el **Tamaño del lote** de las casas es el mismo que el de la zona **Este**.

Interpretación de confianza para Número de habitaciones (Norte- Sur). Modelo desbalanceado

$$M_{31} - M_{32} = (\bar{X}_{31} - \bar{X}_{32}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = (4.4 - 3.667) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = 0.733 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{7.962}{15} = 0.5308$$

$$nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{31} - M_{32} = 0.733 \mp 3.67\sqrt{\frac{0.5308}{5.4545}}$$

$$M_{31} - M_{32} = 0.733 \mp 3.67(0.31195)$$

$$M_{31} - M_{32} = 0.733 \mp 1.14486 \begin{cases} LIC = -0.41186 \\ LSC = 1.87786 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia de la zona **Norte** sobre el **Número de habitaciones** de las casas es el mismo que el de la zona **Sur**.

Intervalo de confianza para Número de habitaciones (Norte – Este). Modelo desbalanceado

$$M_{31} - M_{33} = (\bar{X}_{31} - \bar{X}_{33}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = (4.4 - 4.286) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = 0.114 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{31} - M_{33} = 0.114 \mp 3.67\sqrt{\frac{0.5308}{5.8326}}$$

$$M_{31} - M_{33} = 0.114 \mp 3.67(0.30167)$$

$$M_{31} - M_{33} = 0.114 \mp 1.10713 \begin{cases} -0.99313 \\ 1.22113 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia de la zona **Norte** sobre el **Número de habitaciones** de las casas es el mismo que el de la zona **Este**.

Intervalo de confianza para Número de habitaciones (Sur-Este). Modelo desbalanceado

$$M_{32} - M_{33} = (\bar{X}_{32} - \bar{X}_{33}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = (3.667 - 4.286) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = -0.619 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{6}+\frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

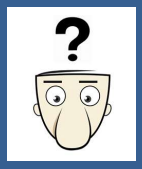
$$M_{32} - M_{33} = -0.619 \mp 3.67\sqrt{\frac{0.5308}{6.4620}}$$

$$M_{32} - M_{33} = -0.619 \mp 3.67(0.2866)$$

$$M_{32} - M_{33} = -0.619 \mp 1.05184 \begin{cases} LIC = -1.67084 \\ LSC = 0.43284 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia de la zona **Sur** sobre el **Número de habitaciones** de las casas es el mismo que el de la zona **Este**.



RESPUESTAS A LOS EJERCICIOS DE AUTOEVALUACIÓN DEL CAPÍTULO 3

3.1.1.1 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 3.1.1.1 Análisis Discriminante

1. Supongamos que cierta empresa desea averiguar si uno de sus nuevos productos, una nueva y mejorada batidora tendrá éxito comercial.

Para llevar a cabo la investigación, la empresa está interesada en identificar si es posible, a aquellos consumidores que comprarían el nuevo producto y aquellos que no lo comprarían.

Para a ayudar a identificar a los compradores potenciales, la empresa ha ideado escalas de valoración para tres características:- duración, funcionamiento y diseño- que son utilizados por los consumidores para evaluar el nuevo producto.

La empresa querría saber que características de su nuevo producto son útiles para diferenciar a los compradores de los NO compradores, es decir, cual separa mejor a cada grupo.

La tabla siguiente refleja la valoración sobre estas tres características con un precio especificado con un grupo de 10 compradores potenciales donde el 1 refleja la valoración más baja y 10 excelente.

No. De sujeto	Grupos basados en la intención de compra	Evaluación del nuevo producto		
		X_1 Duración	X_2 Funcionamiento	X_3 Diseño
1	Compraría	8	9	6

2	Compraría	6	7	5
3	Compraría	10	6	3
4	Compraría	9	4	4
5	Compraría	4	8	2
	Media del grupo 1	7.4	6.8	4
6	No compraría	5	4	7
7	No compraría	3	7	2
8	No compraría	4	5	5
9	No compraría	2	4	3
10	No compraría	2	2	2
	Media del grupo 2	3.2	4.4	3.8

a) Obtenga el modelo discriminante que nos ayude a comprobar la existencia de dos grupos.

b) Pruebe si el modelo es bueno para pronosticar o clasificar a las personas que comprarían de las que no comprarían la nueva y mejorada batidora.

c) Valide el modelo discriminante con los mismos datos de derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados indicando en orden de importancia las variables que logran segmentar y que estrategia se debería seguir.

f) ¿En qué grupo caería una persona que hubiera calificado el nuevo producto de la siguiente manera? : Duración: 7; Funcionamiento: 5 y Diseño: 6

Resolución del inciso a)

La ecuación para derivar el modelo discriminante tendrá la siguiente forma:

$$Z = K_1X_1 + K_2X_2 + K_3X_3$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las *K,* para ello se deben resolver las siguientes ecuaciones normales:

$$\sum X_1'^2 K_1 + \sum X_1' X_2' K_2 + \sum X_1' X_3' K_3 = \bar{X}_1 \,|Grupo\ 1 - \bar{X}_1|Grupo\ 2$$

$$\sum X_1' X_2' K_1 + \sum X_2'^2 K_2 + \sum X_2' X_3' K_3 = \bar{X}_2 \,|Grupo\ 1 - \bar{X}_2|Grupo\ 2$$

$$\sum X_1' X_3' K_1 + \sum X_2' X_3' K_2 + \sum X_3'^2 K_3 = \bar{X}_3 \,|Grupo\ 1 - \bar{X}_3|Grupo\ 2$$

Donde:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\, \bar{X}_2)$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\, \bar{X}_3)$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\, \bar{X}_3)$$

Se sustituyen los valores correspondientes, obteniendo lo siguiente:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_1'^2 = [(8^2 + 6^2 + 10^2 + 9^2 + 4^2) - 5(7.4)^2] + [(5^2 + 3^2 + 4^2 + 2^2 + 2^2) - 5(3.2)^2] = 30$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_2'^2 = [(9^2 + 7^2 + 6^2 + 4^2 + 8^2) - 5(6.8)^2] + [(4^2 + 7^2 + 5^2 + 4^2 + 2^2) - 5(4.4)^2] = 28$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_3'^2 = [(6^2 + 5^2 + 3^2 + 4^2 + 2^2) - 5(4)^2] + [(7^2 + 2^2 + 5^2 + 3^2 + 2^2) - 5(3.8)^2] = 28.80$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\bar{X}_2)$$

$$=[(8x9 + 6x7 + 10x6 + 9x4 + 4x8) - 5(7.4x6.8)] + [(5x4 + 3x7 + 4x5 + 2x4 + 2x2) - 5(3.2x4.4)] = -7$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\bar{X}_3)$$

$$=[(8x6 + 6x5 + 10x3 + 9x4 + 4x2) - 5(7.4x4.0)] + [(5x7 + 3x2 + 4x5 + 2x3 + 2x2) - 5(3.2x3.8)] = 14.2$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\bar{X}_3)$$

$$=[(9x6 + 7x5 + 6x3 + 4x4 + 8x2) - 5(6.8x4.0)] + [(4x7 + 7x2 + 5x5 + 4x3 + 2x2) - 5(4.4x3.8)] = 2.4$$

De lo anterior se obtienen tres ecuaciones, normales:

$$30K_1 - 7K_2 + 14.2K_3 = 4.2$$

$$-7K_1 + 28K_2 + 2.4K_3 = 2.4$$

$$14.2K_1 + 2.4K_2 + 2.8K_3 = 0.20$$

A partir de estas ecuaciones se obtienen los valores de *K* de cada variable. En este caso se utilizará la matriz inversa para la resolución del sistema lineal siguiendo los siguientes pasos:

Paso 1. Expresar el sistema como A *K=D

Donde:

A= Matriz de coeficientes.

K= Coeficientes discriminantes.

D= Términos independientes.

$$\begin{bmatrix} 30 & -7 & 14.2 \\ -7 & 28 & 2.4 \\ 14.2 & 2.4 & 28.8 \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 4.2 \\ 2.4 \\ 0.2 \end{bmatrix}$$

Paso 2. Obtener el determinante de A aplicando la regla de Sarrus para comprobar si el sistema tiene solución, es decir si |A|≠0:

$$|A| = \begin{vmatrix} 30 & -7 & 14.2 \\ -7 & 28 & 2.4 \\ 14.2 & 2.4 & 28.8 \\ 30 & -7 & 14.2 \\ -7 & 28.8 & 2.4 \end{vmatrix} = (30*28*28.8)+ (-7*2.4*14.2)+ (14.2*-7*2.4)- (14.2*28*14.2)- (30*2.4*2.4)- (-7*-7*28.8)=16,484.96$$

Como 16,848.96 es ≠ de 0 el sistema tiene solución.

Paso 3. Obtener la matriz transpuesta de A, que es:

$$A' = \begin{bmatrix} 30 & -7 & 14.2 \\ -7 & 28 & 2.4 \\ 14.2 & 2.4 & 28.8 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A', es decir formamos $\alpha A'$

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 28 & 2.4 \\ 2.4 & 28.8 \end{vmatrix} & -\begin{vmatrix} -7 & 2.4 \\ 14.2 & 28.8 \end{vmatrix} & +\begin{vmatrix} -7 & 28 \\ 14.2 & 2.4 \end{vmatrix} \\ -\begin{vmatrix} -7 & 14.2 \\ 2.4 & 28.8 \end{vmatrix} & +\begin{vmatrix} 30 & 14.2 \\ 14.2 & 28.8 \end{vmatrix} & -\begin{vmatrix} 30 & -7 \\ 14.2 & 2.4 \end{vmatrix} \\ +\begin{vmatrix} -7 & 14.2 \\ 28 & 2.4 \end{vmatrix} & -\begin{vmatrix} 30 & 14.2 \\ -7 & 2.4 \end{vmatrix} & +\begin{vmatrix} 30 & -7 \\ -7 & 28 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(806.4 - 5.76) & -(-201.6 - 34.08) & +(-16.8 - 397.6) \\ -(-201.6 - 34.08) & +(864 - 201.64) & -(72 + 99.4) \\ +((16.8 - 397.6) & -(72 - 99.4) & +(840 - 49) \end{bmatrix} =$$

$$= \begin{bmatrix} 800.64 & 235.68 & -414.4 \\ 235.68 & 662.36 & -171.40 \\ -414.4 & -171.40 & 791 \end{bmatrix}$$

Paso 5. Calculamos la matriz inversa de A:

$$A^{-1} = \frac{1}{|A|} \propto A^{1} = \frac{1}{16{,}484.96} \begin{bmatrix} 800.64 & 235.68 & -414.4 \\ 235.68 & 662.36 & -171.40 \\ -414.4 & -171.40 & 791 \end{bmatrix} = \begin{bmatrix} \frac{800.64}{16{,}484.96} & \frac{235.68}{16{,}484.96} & \frac{-414.4}{16{,}484.96} \\ \frac{235.68}{16{,}484.96} & \frac{662.36}{16{,}484.96} & \frac{-171.40}{16{,}484.96} \\ \frac{-414.4}{16{,}484.96} & \frac{-171.40}{16{,}484.96} & \frac{791}{16{,}484.96} \end{bmatrix}$$

Nota: Es conveniente dejar el resultado en forma de fracción para no perder precisión

Paso 6. Llevamos el valor anterior a la expresión $K = A^{-1} * D$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} \frac{800.64}{16{,}484.96} & \frac{235.68}{16{,}484.96} & \frac{-414.4}{16{,}484.96} \\ \frac{235.68}{16{,}484.96} & \frac{662.36}{16{,}484.96} & \frac{-171.40}{16{,}484.96} \\ \frac{-414.4}{16{,}484.96} & \frac{-171.40}{16{,}484.96} & \frac{791}{16{,}484.96} \end{bmatrix} \begin{bmatrix} 4.2 \\ 2.4 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.23327 \\ 0.15440 \\ -0.12094 \end{bmatrix}$$

Paso 7. Obtenemos el modelo discriminante:

$$\boldsymbol{Z = 0.23327K_1 + 0.15440K_2 - 0.12094K_3}$$

Resolución del inciso b)

A continuación se prueba si el modelo es bueno para discriminar o clasificar a las personas que están de acuerdo con las que no lo están mediante los siguientes pasos:

Paso 1. Plantear el juego de hipótesis:

H_0. El modelo discriminante no es bueno para discriminar o clasificar a las personas que están o no de acuerdo con las características de las laptops analizadas.

H_1. El modelo discriminante es bueno.

Paso 2. Establecer el estadístico de prueba:

$$F\ calc = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2)D^2}{[(n_1 + n_2 - 2)p * (n_1 + n_2)]}$$

$$D^2 = (n_1 + n_2 - 2)\left[\sum K_1(\bar{X}_1|Grupo\ 1 - \bar{X}_1|Grupo2\right]$$

$$D^2 = (5 + 5 - 2)[(0.233327)(4.2) + (0.15440)(2.4) + (-0.12094)(0.2)] = 10.61$$

$$F\ calc = \frac{[(5 + 5 - 2 - 1)(5 * 5)(10.61)]}{[(5 + 5 - 2)2](5 + 5)} = \frac{7x25x10.61}{16x10} = \frac{1856.75}{160} = 11.60$$

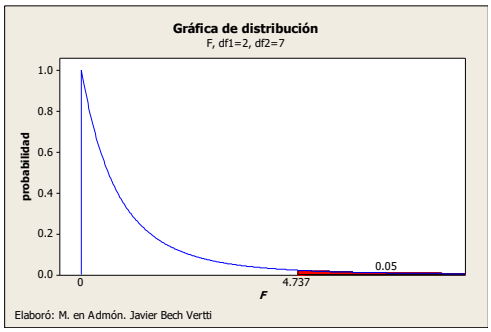
Paso 3. Establecer la región de Rechazo:

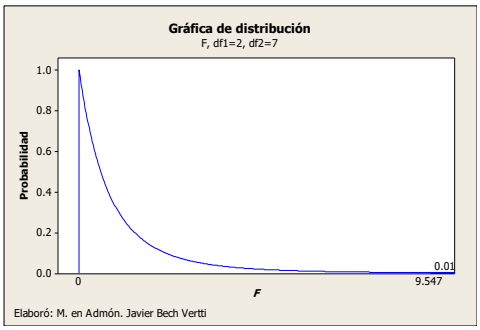
$$Fcrit = [\propto_1 p_1(n_1 + n_2 - p - 1)]$$

$$Fcrit = 0.05, 2, 7$$

$$f_{0.05} = 4.74$$

$$f_{0.01} = 9.55$$





Paso 4. Definir la regla de decisión:

Se rechaza H_0 si $F_{calc.} \geq 4.74$

Paso 5. Conclusiones:

-Estadística: Como 11.60>4.74 y > 9.55 ∴ la prueba es (AS) y se rechaza H_0 con mucha fuerza.

-Administrativa: Existe evidencia suficiente para decir que estadísticamente el modelo es bueno para clasificar o discriminar a los compradores de los no compradores.

Resolución del inciso c)

Cuando se ha concluido que el modelo discriminante es bueno para hacer clasificaciones, se debe proceder a validar el mismo. Esta validación consiste en determinar el grado de confiabilidad que se tendrá al hacer pronósticos con el modelo.

Paso 1. Se obtiene la Z promedio para cada grupo:

$Z = 0.23327K_1 + 0.15440K_2 - 0.12094K_3$

$\bar{Z}_1 = 0.23327(7.4) + 0.15440(6.8) - 0.12094(4.0) = 2.29$

$\bar{Z}_2 = 0.23327(3.2) + 0.15440(4.4) - 0.12094(3.8) = 0.97$

Paso 2. Se determina la Z crítica o de corte:

$$Z^*_{critica} = \frac{n_2\bar{Z}_1 + n_1\bar{Z}_2}{n_1 + n_2} = \frac{5(2.29) + 5(0.97)}{5 + 5} = 1.63$$

*Esta fórmula puede usarse ya sea si $n_1=n_2$ o si $n_1 \neq n_2$

Paso 3. Se obtiene la Z para cada observación:

$Z_1 = 0.23327(8) + 0.15440(9) - 0.12094(6) = 2.5306 > 1.63$

$Z_2 = .0.23327(6) + 0.15440(7) - 0.12094(5) = 1.8761 > 1.63$

$Z_3 = .0.23327(10) + 0.15440(6) - 0.12094(3) = 2.8967 > 1.63$

$Z_4 = 0.23327(9) + 0.15440(4) - 0.12094(4) = 2.2337 > 1.63$

$Z_5 = .0.23327(4) + 0.15440(8) - 0.12094(2) = 1.9266 > 1.63$

$Z_6 = .0.23327(5) + 0.15440(4) - 0.12094(7) = 0.9378 < 1.63$

$Z_7 = .0.23327(3) + 0.15440(7) - 0.12094(2) = 1.3056 < 1.63$

$Z_8 = .0.23327(4) + 0.15440(5) - 0.12094(5) = 1.1007 < 1.63$

$Z_9 = .0.23327(2) + 0.15440(4) - 0.12094(3) = 0.7215 < 1.63$

$Z_{10} = .0.23327(2) + 0.15440(2) - 0.12094(2) = 0.5336 > 1.63$

Paso 4. Se verifica si en las Z observadas la clasificación es correcta o incorrecta

Sujeto No.	Grupos	Duración (X_1)	Funcionamiento (X_2)	Diseño (X_3)	Z obs.	Clas. ✓o X
1	Compraría	8	9	6	$2.5306 > 1.63$	✓
2	Compraría	6	7	5	$1.8761 > 1.63$	✓
3	Compraría	10	6	3	$2.8967 > 1.63$	✓
4	Compraría	9	4	4	$2.2337 > 1.63$	✓
5	Compraría	4	8	2	$1.9266 > 1.63$	✓
	Media Gpo. 1	7.4	6.8	4		
6	No Compraría	5	4	7	$0.9378 < 1.63$	✓
7	No Compraría	3	7	2	$1.3056 < 1.63$	✓
8	No Compraría	4	5	5	$1.1007 < 1.63$	✓
9	No Compraría	2	4	3	$0.7215 < 1.63$	✓
10	No Compraría	2	2	2	$0.5336 > 1.63$	✓
	Media Gpo. 2	3.2	4.4	3.8		

Paso 5. Del paso anterior se obtiene el porcentaje de clasificaciones correctas (PCC):

$$PCC = \frac{\#\checkmark Grupo\ 1 + \#\checkmark Gupo\ 2}{\#\ de\ observaciones} * 100$$

$$PCC = \frac{5+5}{10} * 100 = 100\%$$

Paso 6. Se obtiene el porcentaje mínimo de clasificaciones correctas:

$$C = (p_i)^2 + (1 - p_1)^2$$

$$C = (\frac{5}{10})^2 + (1 - \frac{5}{10})^2$$

$$C = 0.25 + 0.25 = 0.50x100 = 50\%$$

Paso 7. Se compara el porcentaje de clasificaciones correctas contra el porcentaje mínimo de clasificaciones correctas:

Como 100% > 50% se puede validar el modelo para clasificar a los Compradores de los No compradores de la nueva batidora.

Resolución del inciso c)

Para interpretar los resultados obtenidos anteriormente, se evalúa la importancia relativa que posee cada una de las variables independientes para discriminar. El método de los pesos discriminantes estandarizados consiste en estandarizar los coeficientes del modelo discriminante con el fin de que puedan ser comparados. Para jerarquizar las variables se considera solamente la magnitud del peso, no su signo, es decir el valor absoluto.

La fórmula para estandarizar los coeficientes discriminantes es:

$$K_n' = (K_n)\left[\frac{\sum X_n'^2}{(n_1 + n_2 - 2)}\right]^{0.05}$$

En el problema presentado, los coeficientes o constantes estandarizadas son:

4. Duración:

$$K_1 = (0.23327)\left[\frac{30}{(5+5-2)}\right]^{0.5} = 0.45173$$

5. Funcionamiento:

$$K_2 = (0.15440)\left[\frac{28}{(5+5-2)}\right]^{0.5} = 0.28886$$

6. Diseño:

$$K_3 = (-0.12094)\left[\frac{28.8}{(5+5-2)}\right]^{0.5} = -0.22947$$

-Interpretación: De acuerdo a los pesos discriminantes estandarizados **en valores absolutos** (sin considerar el signo), podemos concluir que la variable que más discrimina o diferencia a un grupo de otro para los posibles compradores de la batidora es la variable X_1 (duración) con un peso de .45173 por su mayor *K* estandarizada y la menos importante es la variable X_3 (diseño) con un peso de 0.22947 en valor absoluto.

-Estrategias a seguir: Lo que se pretendería como estrategia es tratar de actuar sobre la variable dependiente (grupos), manipulando las variables independientes (Duración, Funcionamiento y Diseño). Lo que se quisiera lograr es aumentar los valores de Z, para hacer que las personas del grupo ***No Compraría*** se pasen al grupo ***Compraría.*** Para nuestro problema, tenemos que si X_1 (Duración) aumenta, subirá a su vez el valor de Z, pues su signo es positivo. Asímismo tenemos que si X_2 (Funcionamiento) aumenta, subirá el valor de Z pues su signo es también positivo, en cambio para X_3 (Diseño), tenemos que va a afectar en forma inversa al valor de Z, por su signo negativo.

De las tres variables, la primera que se debe cuidar es X_1 (Duración), ya que tiene una mayor importancia para discriminar entre los grupos, aumentando más tal vez la duración de la batidora, después podría pensarse en una estrategia en cuanto a la variable X_2 (Funcionamiento), en un segundo plano, para facilitar el uso de la misma y por último una estrategia para la variable X_3 (Diseño), en un tercer plano, para mejorar la presentación o la forma de la batidora.

Resolución del inciso d)

$X_1 = 7$

$X_2 = 5$

$X_3 = 6$

Sustituimos los valores anteriores en el modelo discriminante:

$Z = 0.23327(7) + 0.15440(5) - 0.12094(6) = 1.68$

Comparamos el valor anterior con la Z crítica:

Como 1.68 > 1.63 ∴ se le clasificaría en el grupo 1 de **Compraría**

3.2.1.1 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 3.2.1.1 Vector de Medias

El Director de Mercadotecnia de una cadena de cines desea medir el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines. Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Para que los potenciales consumidores recuerden la cadena de cines cuando se utiliza la REVISTA como medio masivo para la campaña de publicidad en los consumidores los valores medios de las variables deben ser por lo menos de (6,9,8). Para comprobar si se logró este objetivo, una vez lanzada la campaña en la REVISTA, se tomó una muestra de 7 potenciales clientes y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados son los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo:

$$X = \begin{bmatrix} 4 & 6 & 3 \\ 5 & 4 & 6 \\ 7 & 9 & 8 \\ 3 & 8 & 4 \\ 6 & 7 & 5 \\ 8 & 6 & 8 \\ 7 & 7 & 6 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 9 \\ 8 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

6. Juego de hipótesis

$$H_0 = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 9 \\ 8 \end{bmatrix}$$

$H_1 = Al\ menos\ una\ M\ es\ \neq.$

7. Estadístico de prueba

$$Fcalc = \frac{n-p}{p(n-1)} T^2$$

Donde: $T^2 = n(M - M_0)'[S^2]^{-1}(\hat{M} - \hat{M}_0)$

Donde: $S^2 = \frac{1}{n-1} SCPC = \frac{1}{n-1} D'D$ (Matriz de varianzas y covarianzas)=

$$[s^2]^2 = \frac{1}{|S|} \propto [S^2]'$$

$$D=\begin{bmatrix}4&6&3\\5&4&6\\7&9&8\\3&8&4\\6&7&5\\8&6&8\\7&7&6\end{bmatrix}-\begin{bmatrix}5.71&6.71&5.71\\5.71&6.71&5.71\\5.71&6.71&5.71\\5.71&6.71&5.71\\5.71&6.71&5.71\\5.71&6.71&5.71\\5.71&6.71&5.71\end{bmatrix}=\begin{bmatrix}-1.71&-0.71&-2.71\\-0.71&-2.71&0.29\\1.29&2.29&2.29\\-2.71&1.29&-1.71\\0.29&0.29&-0.71\\2.29&-0.71&2.29\\1.29&0.29&0.29\end{bmatrix}$$

$$D'D=\begin{bmatrix}-1.71&-0.71&1.29&-2.71&0.29&2.29&1.29\\-0.71&-2.71&2.29&1.29&0.29&-0.71&0.29\\-2.71&0.29&2.29&-1.71&-0.71&2.29&0.29\end{bmatrix}\begin{bmatrix}-1.71&-0.71&-2.71\\-0.71&-2.71&0.29\\1.29&2.29&2.29\\-2.71&1.29&-1.71\\0.29&0.29&-0.71\\2.29&-0.71&2.29\\1.29&0.29&0.29\end{bmatrix}$$

$$D'D=\begin{bmatrix}19.42857&1.42857&17.42857\\1.42857&15.42857&2.42857\\17.42857&2.42857&21.42857\end{bmatrix}$$

$$S^2=\frac{1}{n-1}\,D'D=\frac{1}{7-1}D'D=\frac{1}{6}\qquad\begin{bmatrix}19.42857&1.42857&17.42857\\1.42857&15.42857&2.42857\\17.42857&2.42857&21.42857\end{bmatrix}$$

$$S^2=\frac{1}{6}\begin{bmatrix}19.42857&1.42857&17.42857\\1.42857&15.42857&2.42857\\17.42857&2.42857&21.42857\end{bmatrix}$$

$$S^2=\begin{bmatrix}\frac{19.42857}{6}&\frac{1.42857}{6}&\frac{17.42857}{6}\\\frac{1.42857}{6}&\frac{15.42857}{6}&\frac{2.42857}{6}\\\frac{17.42857}{6}&\frac{2.42857}{6}&\frac{21.42857}{6}\end{bmatrix}$$

Utilizando la regla de Sarrus

$$|S^2| = \begin{bmatrix} \frac{19.42857}{6} & \frac{1.42857}{6} & \frac{17.42857}{6} \\ \frac{1.42857}{6} & \frac{15.42857}{6} & \frac{2.42857}{6} \\ \frac{17.42857}{6} & \frac{2.42857}{6} & \frac{21.42857}{6} \\ \frac{19.42857}{6} & \frac{1.42857}{6} & \frac{17.42857}{6} \\ \frac{1.42857}{6} & \frac{15.42287}{6} & \frac{2.42857}{6} \end{bmatrix} =$$

$$\begin{gathered} (19.42857/6 \times 15.42857/6 \times 21.42857/6) + \\ (1.42857/6 \times 2.42857/6 \times 17.42857/6) + \\ (17.42857/6 \times 1.42857/6 \times 2.42857/6) - \\ (17.42857/6 \times 15.42857/6 \times 17.42857/6) - \\ (19.42857/6 \times 2.42857/6 \times 2.42857/6) - \\ (1.42857/6 \times 1.42857/6 \times 21.42857/6) - \end{gathered}$$

$$= 7.868$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 2.5714 & 0.4048 \\ 0.4048 & 3.5714 \end{vmatrix} & -\begin{vmatrix} 0.2381 & 0.4048 \\ 2.9048 & 3.5714 \end{vmatrix} & +\begin{vmatrix} 0.2381 & 2.5714 \\ 2.9048 & 0.4048 \end{vmatrix} \\ -\begin{vmatrix} 0.2381 & 2.9048 \\ 0.4048 & 3.5714 \end{vmatrix} & +\begin{vmatrix} 3.2381 & 2.9048 \\ 2.9048 & 3.5714 \end{vmatrix} & -\begin{vmatrix} 3.2381 & 0.2381 \\ 2.9048 & 0.4048 \end{vmatrix} \\ +\begin{vmatrix} 0.2381 & 2.9048 \\ 2.5714 & 0.4048 \end{vmatrix} & -\begin{vmatrix} 3.2381 & 209048 \\ 0.2381 & 0.4048 \end{vmatrix} & +\begin{vmatrix} 3.2381 & 0.2381 \\ 0.2381 & 2.5714 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(9.1837 - 0.1638) & -(0.8503 - 1.1757) & +(0.0964 - 7.4694) \\ -(-8503 - 1.1757) & +(11.5646 - 8.4376) & -(1.3107 - 0.6916) \\ +((0.0964 - 7.4694) & -(1.3107 - 0.6916) & +(8.3265 - 0.0567) \end{bmatrix} =$$

$$= \begin{bmatrix} 9.0198 & 0.3254 & -7.3730 \\ 0.3254 & 3.1270 & -0.6190 \\ -7.3730 & -0.6190 & 8.2698 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{|S^2|} \propto [S^2]' = \frac{1}{7.868} = \begin{bmatrix} \frac{9.0198}{7.868} & \frac{0.3254}{7.868} & \frac{-7.3730}{7.868} \\ \frac{0.3254}{7.868} & \frac{3.1270}{7.868} & \frac{-0.6190}{7.868} \\ \frac{-7.3730}{7.868} & \frac{-0.6190}{7.868} & \frac{8.2698}{7.868} \end{bmatrix}$$

$$[S^2]^{-1} = \begin{bmatrix} 1.14644 & 0.04136 & -0.93712 \\ 0.04136 & 0.39744 & -0.07868 \\ -0.93712 & -0.07868 & 1.05111 \end{bmatrix}$$

Obtener T^2

$$T^2 = n[\widehat{M} - M_{H_0}]'[S^2]^{-1}[\widehat{M} - M_{H_0}]$$

$$T^2 = 7[5.71 - 6; 6.71 - 9; 5.71 - 8][S^2]^{-1} \begin{bmatrix} 5.71 - 6 \\ 6.71 - 9 \\ 5.71 - 8 \end{bmatrix}$$

$$T^2 = 7[-0.2857; -2.2857; -2.2857][S^2]^{-1} \begin{bmatrix} -0.2857 \\ -2.2857 \\ -2.2857 \end{bmatrix}$$

$$T^2 = [-2; -16; -16] \left[\begin{bmatrix} 1.14644 & 0.04136 & -0.93712 \\ 0.04136 & 0.39744 & -0.07868 \\ -0.93712 & -0.07868 & 1.05111 \end{bmatrix} \right] \begin{bmatrix} -0.2857 \\ -2.2857 \\ -2.2857 \end{bmatrix}$$

$$T^2 = [12.03934 - 5.18292 - 13.68460] \begin{bmatrix} -0.2857 \\ -2.2857 \\ -2.2857 \end{bmatrix}$$

$$T^2 = \underline{39.68594}$$

$$Fcalc = \frac{n - p}{p(n-1)} T^2$$

$$Fcalc = \frac{7 - 3}{3(7-1)} 39.68594 = .2222 * 39.68594 = 8.81910$$

8. Región de rechazo

$$F_{\propto .p.n-p}$$

$F_{0.05.3.4} = 6.59$

$F_{0.01.3.4} = 16.69$

9. Regla de decisión

Se rechaza H_0 si Fcalc >6.59

10. Conclusión

Estadística: Como 6.59 >8.81910 >16.69

∴ la prueba es (S) y se rechaza H_0 con poca potencia.

Adva: Existe evidencia suficiente (E.E.S). para decir que estadísticamente al menos uno de los parámetros de las tres variables estudiadas (X_1= Evocación de la marca, X_2= Evocación del Slogan y X_3= Evocación del logotipo) es diferente.

Resolución del inciso b)

Intervalo de confianza para marca

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 5.71 \mp 2.447 \frac{1.79947}{\sqrt{7}}$$

$$M = 5.71 \mp 2.447 \frac{1.79947}{2.6458}$$

$M = 5.71 \mp 2.447(0.68014)$

$M = 5.71 \mp 1.66429 \begin{cases} LIC = 4.04571 \\ LSC = 7.37429 \end{cases}$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio de la evocación de la marca oscila entre 4.04571 y 7.37429. Como 6 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

Intervalo de confianza para slogan

$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$

$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$

$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$

$M = 6.71 \mp 2.447 \frac{1.60356}{\sqrt{7}}$

$M = 6.71 \mp 2.447 \frac{1.60356}{2.6458}$

$M = 6.71 \mp 2.447(0.60609)$

$M = 6.71 \mp 1.48310 \begin{cases} LIC = 5.22690 \\ LSC = 8.19310 \end{cases}$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio de la evocación del slogan oscila entre 5.22690 y 8.19310 Como 9 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

Intervalo de confianza para logotipo

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 5.71 \mp 2.447 \frac{1.88981}{\sqrt{7}}$$

$$M = 5.71 \mp 2.447 \frac{1.88981}{2.6458}$$

$$M = 5.71 \mp 2.447(.71428)$$

$$M = 5.71 \mp 1.74785 \begin{cases} LIC = 3.99215 \\ LSC = 7.45785 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio de la evocación del logotipo oscila entre 3.992159 y 7.45785. Como 8 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

3.2.2.1 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE AUTOEVALUACIÓN 3.2.2.1 Diferencia de Medias Multivariada

El Director de Mercadotecnia de una cadena de cines desea medir el grado de influencia que llega a tener el medio masivo utilizado para su campaña de publicidad en los consumidores. Esta influencia se maneja por medio del "grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines". Este concepto es medido a través de tres variables dependientes: Marca, Slogan y Logotipo. Los tipos de medios manejados son dos: Televisión y Revistas.

Una vez lanzada la campaña en la TELEVISIÓN y en REVISTAS, se tomó una muestra aleatoria de una

población normal multivariada donde X_1= Evocación de la marca , X_2= Evocación del Slogan y X_3= Evocación del logotipo de 6 potenciales clientes que leyeron la campaña en la Televisión y 7 que la vieron en Revistas y se les pidió que calificarán su grado de evocación o recuerdo sobre la cadena de cines en una escala de 0 a 10 puntos donde el 0 nos indica baja evocación o recuerdo y el 10 alta evocación o recuerdo.

Los resultados fueron los siguientes:

Medio utilizado					
Televisión			Revistas		
Marca	Slogan	Logotipo	Marca	Slogan	Logotipo
9	10	8	4	6	3
8	9	9	5	4	6
7	8	9	7	9	8
9	9	9	3	8	4
8	8	6	6	7	5
			8	6	8
			7	7	6

a) Existe evidencia suficiente que indique que el grado de evocación o recuerdo que tiene el consumidor sobre esta cadena de cines es el mismo en ambos medios utilizados?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de Hipótesis

$$H_0: \begin{bmatrix} M_{11} \\ M_{21} \\ M_{31} \end{bmatrix} = \begin{bmatrix} M_{12} \\ M_{22} \\ M_{32} \end{bmatrix}$$

$$H_1: Al\ menos\ 1\ combinación\ es\ \neq$$

2. Estadístico de prueba

$$Fcalc = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

Donde $T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)^1 [S^2]^{-1} (\widehat{M}_1 - \widehat{M}_2)$

$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2}$

$W_1 = SCPC_1 = D' D_1$

$W_2 = SCPC_2 = D_2' D_2$

$$D_1 = \begin{bmatrix} 9 & 10 & 8 \\ 8 & 9 & 9 \\ 7 & 8 & 9 \\ 9 & 9 & 9 \\ 8 & 8 & 6 \end{bmatrix} - \begin{bmatrix} 8.2 & 8.8 & 8.2 \\ 8.2 & 8.8 & 8.2 \\ 8.2 & 8.8 & 8.2 \\ 8.2 & 8.8 & 8.2 \\ 8.2 & 8.8 & 8.2 \end{bmatrix} = \begin{bmatrix} 0.80 & 1.20 & -0.20 \\ -0.20 & 0.20 & 0.80 \\ -1.20 & -0.80 & 0.80 \\ 0.80 & 0.20 & 0.80 \\ -0.20 & -0.80 & -2.20 \end{bmatrix}$$

$$D_1' D_1 = \begin{bmatrix} 0.80 & -0.20 & -1.20 & 0.80 & -0.20 \\ 1.20 & 0.20 & -0.80 & 0.20 & -0.80 \\ -0.20 & 0.80 & 0.80 & 0.80 & -2.20 \end{bmatrix} \begin{bmatrix} 0.80 & 1.20 & -0.20 \\ -0.20 & 0.20 & 0.80 \\ -1.20 & -0.80 & 0.80 \\ 0.80 & 0.20 & 0.80 \\ -0.20 & -0.80 & -2.20 \end{bmatrix} = \begin{bmatrix} 2.8 & 2.2 & -0.2 \\ 2.2 & 2.8 & 1.2 \\ -0.2 & 1.2 & 6.8 \end{bmatrix}$$

$$W_1 = D_1' D_1 = \begin{bmatrix} 2.8 & 2.2 & -0.2 \\ 2.2 & 2.8 & 1.2 \\ -0.2 & 1.2 & 6.8 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 4 & 6 & 3 \\ 5 & 4 & 6 \\ 7 & 9 & 8 \\ 3 & 8 & 4 \\ 6 & 7 & 5 \\ 8 & 6 & 8 \\ 7 & 7 & 6 \end{bmatrix} - \begin{bmatrix} 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \\ 5.71 & 6.71 & 5.71 \end{bmatrix} = \begin{bmatrix} -1.71 & -0.71 & -2.71 \\ -0.71 & -2.71 & 0.29 \\ 1.29 & 2.29 & 2.29 \\ -2.71 & 1.29 & -1.71 \\ 0.29 & 0.29 & -0.71 \\ 2.29 & -0.71 & 2.29 \\ 1.29 & 0.29 & 0.29 \end{bmatrix}$$

$$D_2' D_2 = \begin{bmatrix} -1.71 & -0.71 & 1.29 & -2.71 & 0.29 & 2.29 & 1.29 \\ -0.71 & -2.71 & 2.29 & 1.29 & 0.29 & -0.71 & 0.29 \\ -2.71 & 0.29 & 2.29 & -1.71 & -0.71 & 2.29 & 0.29 \end{bmatrix} \begin{bmatrix} -1.71 & -0.71 & -2.71 \\ -0.71 & -2.71 & 0.29 \\ 1.29 & 2.29 & 2.29 \\ -2.71 & 1.29 & -1.71 \\ 0.29 & 0.29 & -0.71 \\ 2.29 & -0.71 & 2.29 \\ 1.29 & 0.29 & 0.29 \end{bmatrix} = \begin{bmatrix} 19.42857 & 1.15857 & 17.42857 \\ 1.42857 & 15.42857 & 2.42857 \\ 17.42857 & 2.42857 & 21.42857 \end{bmatrix}$$

$$W_2 = D_2' D_2 = \begin{bmatrix} 19.42857 & 1.15857 & 17.42857 \\ 1.42857 & 15.42857 & 2.42857 \\ 17.42857 & 2.42857 & 21.42857 \end{bmatrix}$$

$$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2} = \frac{\begin{bmatrix} 2.8 & 2.2 & -0.2 \\ 2.2 & 2.8 & 1.2 \\ -0.2 & 1.2 & 6.8 \end{bmatrix} + \begin{bmatrix} 19.42857 & 1.15857 & 17.42857 \\ 1.42857 & 15.42857 & 2.42857 \\ 17.42857 & 2.42857 & 21.42857 \end{bmatrix}}{5 + 7 - 2} = \begin{bmatrix} 2.22286 & 0.36286 & 1.72286 \\ 0.36286 & 1.82286 & 0.36286 \\ 1.72286 & 0.36286 & 2.82286 \end{bmatrix}$$

$$|S^2| = \begin{bmatrix} 2.22286 & 0.36286 & 1.72286 \\ 0.36286 & 1.82286 & 0.36286 \\ 1.72286 & 0.36286 & 2.82286 \\ 2.22286 & 0.36286 & 1.72286 \\ 0.36286 & 1.82286 & 0.36286 \end{bmatrix} = 5.81674$$

Como es ≠ de 0 tiene solución

Transpuesta de S^2

$$[S^2]^1 = \begin{bmatrix} 2.22286 & 0.36286 & 1.72286 \\ 0.36286 & 1.82286 & 0.36286 \\ 1.72286 & 0.36286 & 2.82286 \end{bmatrix}$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} + \begin{vmatrix} 1.82286 & 0.36286 \\ 0.36286 & 2.82286 \end{vmatrix} & - \begin{vmatrix} 0.36286 & 0.36286 \\ 1.72286 & 2.82286 \end{vmatrix} & + \begin{vmatrix} 0.36286 & 1.82286 \\ 1.72286 & 0.36286 \end{vmatrix} \\ - \begin{vmatrix} 0.36286 & 1.72286 \\ 0.36286 & 2.82286 \end{vmatrix} & + \begin{vmatrix} 2.22286 & 1.72286 \\ 1.72286 & 2.82286 \end{vmatrix} & - \begin{vmatrix} 2.22286 & 0.36286 \\ 1.72286 & 0.36286 \end{vmatrix} \\ + \begin{vmatrix} 0.36286 & 1.72286 \\ 1.82286 & 0.36286 \end{vmatrix} & - \begin{vmatrix} 2.22286 & 1.72286 \\ 0.36286 & 0.36286 \end{vmatrix} & + \begin{vmatrix} 2.22286 & 0.36286 \\ 0.36286 & 1.82286 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(5.14567 - 0.13167) & -(1.02429 - 0.62515) & +(0.13167 - 3.14052) \\ -(1.02429 - 0.62515) & +(6.27481 - 2.96824) & -(0.80658 - 0.62515) \\ +((0.13167 - 3.14052) & -(0.80658 - 0.62515) & +(4.05195 - 0.13167) \end{bmatrix} =$$

$$= \begin{bmatrix} 5.01400 & -0.39914 & -3.00886 \\ -0.39914 & 3.30657 & -0.18143 \\ -3.00886 & -0.18143 & 3.92029 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{5.81674} = \begin{bmatrix} 5.01400 & -0.39914 & -3.00886 \\ -0.39914 & 3.30657 & -0.18143 \\ -3.00886 & -0.18143 & 3.92029 \end{bmatrix} = \begin{bmatrix} 0.86199 & -0.06862 & -0.51728 \\ -0.06862 & 0.56846 & -0.03119 \\ -0.51728 & -0.03119 & 0.67397 \end{bmatrix}$$

$$T^2 = \left(\frac{n_1 n_2}{n_1+n_2}\right)\left(\widehat{M}_1 - \widehat{M}_2\right)'[S^2]^{-1}\left(\widehat{M}_1 - \widehat{M}_2\right)$$

$$T^2 = \left(\frac{5*7}{5+7}\right)(8.20-5.71; 8.80-6.71; 8.20-5.71)[S^2]^{-1}\begin{bmatrix} 8.20-5.71 \\ 8.80-6.71 \\ 8.20-5.71 \end{bmatrix}$$

$$T^2 = \left(\frac{35}{12}\right)(2.49; 2.09; 2.49)[S^2]^{-1}\begin{bmatrix} 2.49 \\ 2.09 \\ 2.49 \end{bmatrix}$$

$$T^2 = (2.91667)(2.49; 2.09; 2.49)[S^2]^{-1}\begin{bmatrix} 2.49 \\ 2.09 \\ 2.49 \end{bmatrix}$$

$$T^2 = [7.25; 6.0833; 7.25]\begin{bmatrix} 0.86199 & -0.06862 & -0.51728 \\ -0.06862 & 0.56846 & -0.03119 \\ -0.51728 & -0.03119 & 0.67397 \end{bmatrix}\begin{bmatrix} 2.49 \\ 2.09 \\ 2.49 \end{bmatrix}$$

$$T^2 = [2.0818; 2.7345; 0.94627]\begin{bmatrix} 2.49 \\ 2.09 \\ 2.49 \end{bmatrix}$$

$$T^2 = \underline{13.23021}$$

$$Fcalc = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2$$

$$Fcalc = \frac{5+7-3-1}{3(5+7-2)} 13.23021$$

$$Fcalc = \frac{8}{30} 13.23021$$

$$Fcalc = 0.26667 \times 13.23021$$

$$\boldsymbol{Fcalc = 3.608}$$

3. Región de rechazo

$$F_{\alpha.p.n1+n2-p-1} = F_{0.05.3.5+7-3-1} = F_{0.05.3.8} = 4.07$$

$$F_{\alpha.p.n1+n2-p-1} = F_{0.01.3.5+7-3-1} = F_{0.01.3.8} = 7.59$$

4. Regla de decisión

Se rechaza H_0 si Fcalc > 4.07

5. Conclusión:

- Estadistica: Como 3.608< 4.07 ∴

La prueba es N.S y no se rechaza H_0

- Adva: Existe evidencia suficiente para decir el grado de evocación o recuerdo de la Marca, el Slogan y el Logotipo que tiene el consumidor sobre esta cadena de cines es el mismo en ambos medios utilizados.

Resolución del inciso b)

Nota importante: Para construir el intervalo de confianza se debe probar primero si las varianzas de las muestras son iguales o no para saber que caso de debe usar para construir el intervalo de confianza (Caso 2 para varianzas iguales o Caso 3 para varianzas desiguales) por lo que primero hay que realizar la prueba de homocedasticidad de varianzas.

Prueba de Homocedasticidad para Marca

1. Juego de Hipótesis

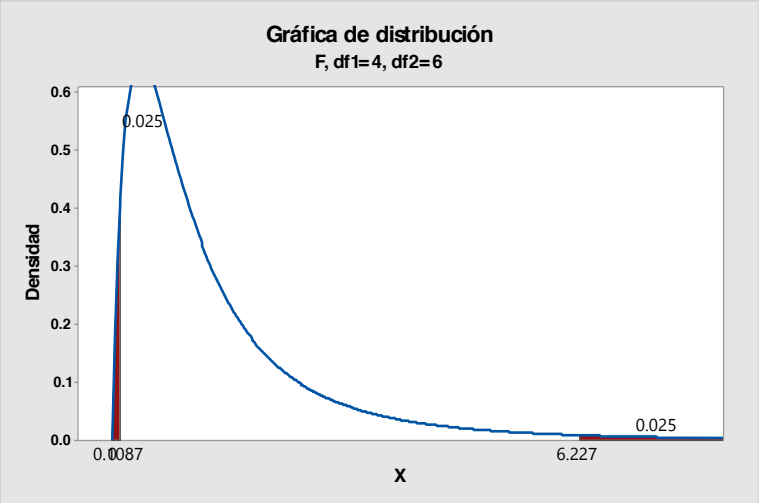
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{11}}^2}{{S_{12}}^2} = \frac{0.70057}{3.24} = 0.21623$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc ≥ 6.23 ó si es ≤ 0.1087

5. Conclusión:

- Estadistica: Como 0.1087< 0.21623< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de Confianza para Marca

$$M_{11} - M_{12} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp (t_{\alpha/2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)0.70057 + (7-1)3.24}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 0.70057 + 6 * 3.24}{10}}$$

$$Sp = \sqrt{\frac{2.80228 + 19.44}{10}}$$

$$Sp = \sqrt{\frac{22.24228}{10}}$$

$Sp = \sqrt{2.22423}$

$Sp = 1.49138$

$$M_{11} - M_{12} = (8.2 - 5.71) \mp 2.228 * 1.49138 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (2.49) \mp 2.228 * 1.49138 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (2.49) \mp 2.228 * 1.49138 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (2.49) \mp 2.228 * 1.49138 * 0.58554$$

$$M_{11} - M_{12} = (2.49) \mp 1.94563 \begin{cases} LIC = 0.54437 \\ LSC = 4.43563 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la evocación de la Marca en la televisión es mayor que la evocación de la marca en las revistas por un mínimo de 0.54437 y un máximo de 4.43563 puntos.

Nota: Como la Evocación de la Marca es diferente en ambos medios de publicidad sería razonable que la prueba de hipótesis de diferencia de medias multivariada se hubiera rechazado por esta variable, sin embargo como el límite inferior de confianza es muy cercano a cero es posible que debido a que las muestras son muy pequeñas los valores críticos de F son muy altos por los grados de libertad tan pequeños por lo que la sugerencia sería tal vez aumentar los tamaños de las muestras para incrementar el valor de F calculada en la prueba de diferencia de medias multivariada bajando asimismo los valores críticos de F con lo cual se podría comprobar la consistencia de la prueba.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Marca vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05
```

```
Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      0.837     0.700  (0.407, 2.829)
2        7      1.799     3.238  (1.055, 4.264)

Relación de desviaciones estándar = 0.465
Relación de varianzas = 0.216
```

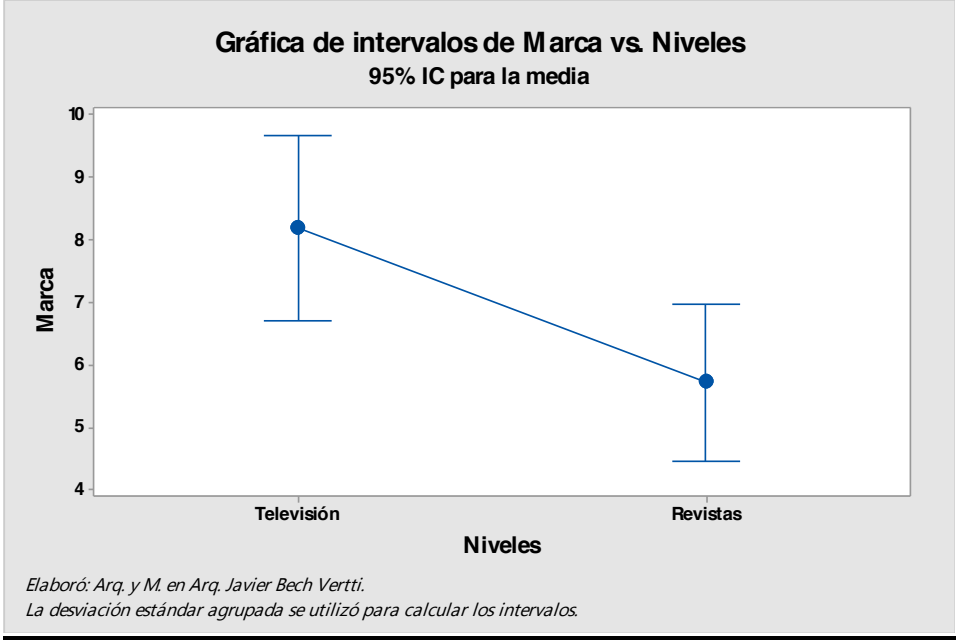
ANOVA unidireccional: Marca vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      18.02     18.021     8.11    0.017
Error    10      22.23      2.223
Total    11      40.25


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
1.49092   44.77%      39.25%   23.43%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  8.200      0.837  (6.714, 9.686)
2        7  5.714      1.799  (4.459, 6.970)

Desv.Est. agrupada = 1.49092
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia      IC de 95%     Valor T  ajustado
2 - 1           -2.486       0.873  (-4.431, -0.541)    -2.85     0.017

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Marca vs. Niveles



Elaboró: Arq. y M. en Arq. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba de Homocedasticidad para Slogan

1. Juego de Hipótesis

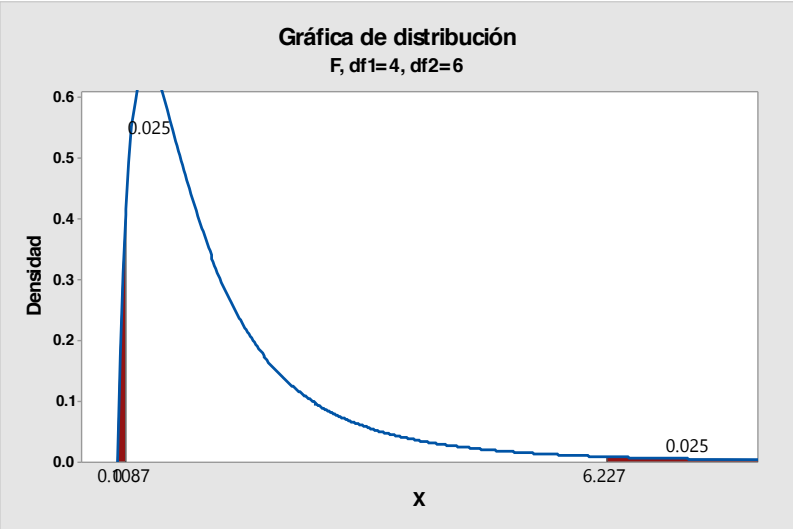
$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$

$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{21}}^2}{{S_{22}}^2} = \frac{0.70057}{2.56} = 0.27366$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:
   - Estadística: Como 0.1087< 0.27366< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
   - Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de Confianza para slogan

$$M_{21} - M_{22} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp t\,{}^{\propto}/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)0.70057 + (7-1)\ 2.56}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 0.70057 + 6 * 2.56}{10}}$$

$$Sp = \sqrt{\frac{2.80228 + 15.36}{10}}$$

$$Sp = \sqrt{\frac{18.16228}{10}}$$

$$Sp = \sqrt{1.81623}$$

$$Sp = 1.34768$$

$$M_{11} - M_{12} = (8.80 - 6.71) \mp 2.228 * 1.34768 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (2.09) \mp 2.228 * 1.34768 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (2.09) \mp 2.228 * 1.34768 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (2.09) \mp 2.228 * 1.34768 * 0.58554$$

$$M_{11} - M_{12} = (2.09) \mp 1.75816 \begin{cases} LIC = 0.33184 \\ LSC = 3.84816 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la evocación del Slogan en la televisión es mayor que la evocación del slogan en las revistas por un mínimo de 0.33184 y un máximo de 3.84816 puntos.

Nota: Como la Evocación del Slogan es diferente en ambos medios de publicidad sería razonable que la prueba de hipótesis de diferencia de medias multivariada se hubiera rechazado por esta variable, sin embargo como el límite inferior de confianza es muy cercano a cero es posible que debido a que las muestras son muy pequeñas los valores críticos de F son muy altos por los grados de libertad tan pequeños por lo que la sugerencia sería tal vez aumentar los tamaños de las muestras para incrementar el valor de F calculada en la prueba de diferencia de medias multivariada bajando asimismo los valores críticos de F con lo cual se podría comprobar la consistencia de la prueba.

Comprobación utilizando el Software Minitab 17

Prueba e IC para dos varianzas: Slogan vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                     IC de 95%
                                       para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      0.837     0.700  (0.407, 2.829)
2        7      1.604     2.571  (0.880, 4.060)

Relación de desviaciones estándar = 0.522
Relación de varianzas = 0.272
```

ANOVA unidireccional: Slogan vs. Niveles

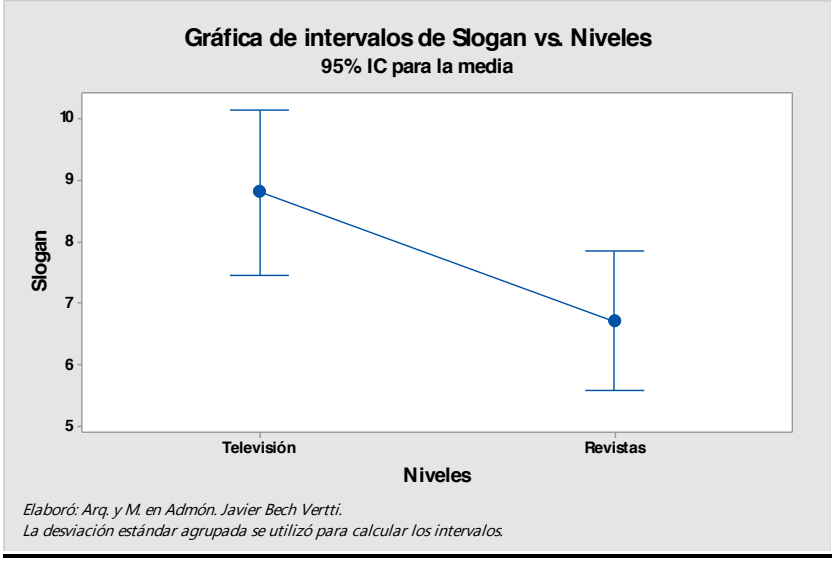
```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.
```

```
Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      12.69     12.688     6.96    0.025
Error    10      18.23      1.823
Total    11      30.92

Resumen del modelo
                     R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
1.35013   41.04%      35.14%   17.92%

Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  8.800      0.837  (7.455, 10.145)
2        7  6.714      1.604  (5.577,  7.851)

Desv.Est. agrupada = 1.35013
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

           Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia      IC de 95%     Valor T  ajustado
2 - 1           -2.086       0.791  (-3.847, -0.324)    -2.64     0.025

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Slogan vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba de Homocedasticidad para Logotipo

1. Juego de Hipótesis

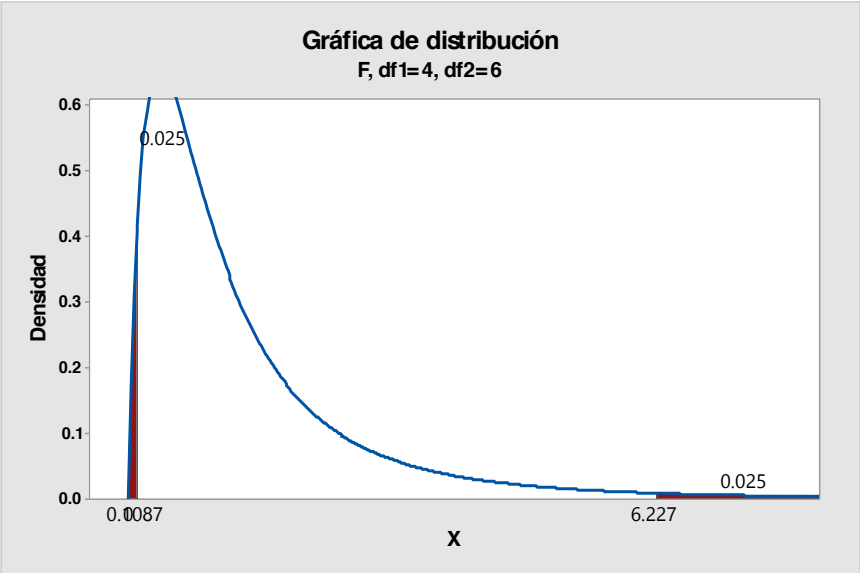
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{31}}^2}{{S_{32}}^2} = \frac{1.69}{3.521} = 0.47998$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:

- Estadística: Como 0.1087< 0.47998< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de Confianza para Logotipo

$$M_{31} - M_{32} = \left(\overline{X}_{31} - \overline{X}_{32}\right) \mp t^{\propto}/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = Sp = \sqrt{\frac{(5-1)1.69 + (7-1)\ 3.521}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 1.69 + 6 * 3.521}{10}}$$

$$Sp = \sqrt{\frac{6.76 + 21.126}{10}}$$

$$Sp = \sqrt{\frac{27.886}{10}}$$

$$Sp = \sqrt{2.78860}$$

$$Sp = 1.66991$$

$$M_{11} - M_{12} = (8.20 - 5.71) \mp 2.228 * 1.66991 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (2.49) \mp 2.228 * 1.66991 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (2.49) \mp 2.228 * 1.66991 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (2.49) \mp 2.228 * 1.66991 * 0.58554$$

$$M_{11} - M_{12} = (2.49) \mp 2.17854 \begin{cases} LIC = 0.31146 \\ LSC = 4.66854 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la evocación del Logotipo en la televisión es mayor que la evocación del Logotipo en las revistas por un mínimo de 0.31146 y un máximo de 4.66854 puntos.

Nota: Como la Evocación del Logotipo es diferente en ambos medios de publicidad sería razonable que la prueba de hipótesis de diferencia de medias multivariada se hubiera rechazado por esta variable, sin embargo como el límite inferior de confianza es muy cercano a cero es posible que debido a que las muestras son muy pequeñas los valores críticos de F son muy altos por los grados de libertad tan pequeños por lo que la sugerencia sería tal vez aumentar los tamaños de las muestras para incrementar el valor de F calculada en la prueba de diferencia de medias multivariada bajando asimismo los valores críticos de F con lo cual se podría comprobar la consistencia de la prueba.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Logotipo vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      1.304     1.700  (0.497, 5.625)
2        7      1.890     3.571  (1.204, 4.118)

Relación de desviaciones estándar = 0.690
Relación de varianzas = 0.476
```

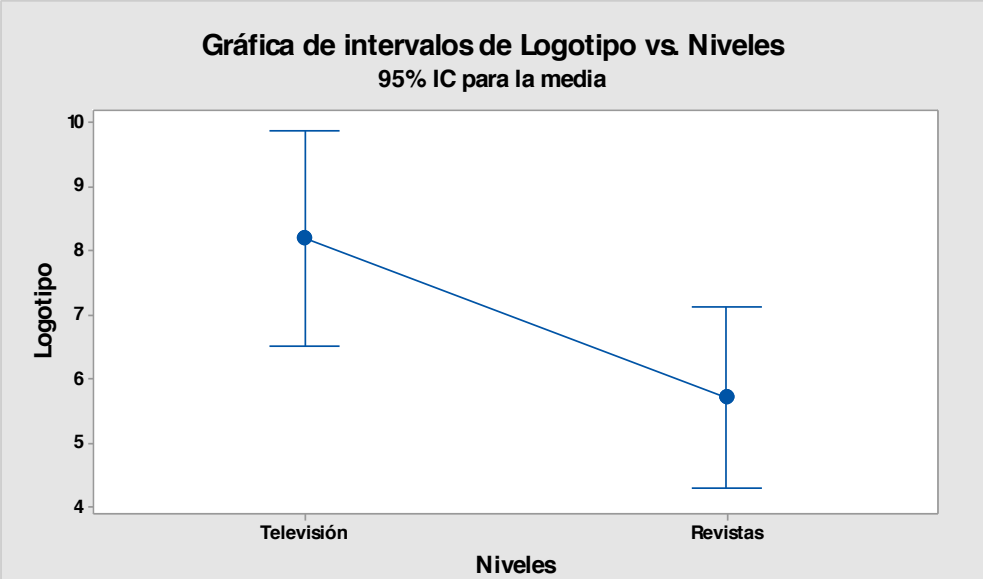
ANOVA unidireccional: Logotipo vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05


Se presupuso igualdad de varianzas para el análisis.



Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2



Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      18.02     18.021     6.38    0.030
Error    10      28.23      2.823
Total    11      46.25


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
1.68014   38.97%      32.86%   13.96%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  8.200      1.304  (6.526, 9.874)
2        7  5.714      1.890  (4.299, 7.129)

Desv.Est. agrupada = 1.68014
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

           Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia      IC de 95%     Valor T  ajustado
2 - 1           -2.486       0.984  (-4.678, -0.294)    -2.53     0.030

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Logotipo vs. Niveles



ElaborÓ. Arq. y M. en Admóm. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

3.2.3.1 EJERCICIO DE AUTOEVALUACIÓN

EJERCICIO DE **AUTOEVALUACIÓN 3.2.3.1 Análisis de Varianza Multivariada (MANOVA)**

El Director de Industrias México desea medir el grado de influencia que llega a tener el área de trabajo de los empleados de Industrias México sobre los ingresos que reciben "antes de impuestos" (en millones de pesos), los años de escolaridad concluidos y la edad de los empleados. Las áreas de trabajo que se manejaron son tres: Gerencial, Servicios y Obreros.

El efecto se evalúa mediante tres criterios:

1.- Que tanto el Área de Trabajo llega a influir sobre el Ingreso de los empleados de Industrias México.

2.- Que tanto el Área de Trabajo llega a influir sobre la Escolaridad de los empleados de Industrias México.

3.- Que tanto el Área de Trabajo llega a influir sobre la Edad de los empleados de Industrias México.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Ingresos anuales "antes de impuestos" (en millones de pesos), X_2= Años de escolaridad concluidos y X_3= Edad de los empleados (En años cumplidos), de 5 empleados del área GerencialLos resultados fueron los siguientes:

Área de trabajo								
Gerencial			Servicios			Obreros		
Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad
1.234	16	43	0.542	12	40	0.210	11	20
0.894	15	62	0.326	14	21	0.230	12	18
0.950	16	43	0.380	14	34	0.204	9	53
0.702	16	28	0.436	12	65	0.292	13	27
0.718	14	36	0.404	10	63	0.240	10	22
			0.442	10	40	0.238	12	17
						0.214	11	45

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de los empleados es el mismo en las tres áreas de trabajo?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el área de trabajo influye sobre las magnitudes promedio de los empleados.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que área o áreas de trabajo los atributos son mayores y en que magnitud.

Resolución del inciso a)

Prueba de hipótesis:

1. Juego de Hipótesis

$$H_0 = \begin{bmatrix}\mu_{11}\\ \mu_{21}\\ \mu_{31}\end{bmatrix} = \begin{bmatrix}\mu_{12}\\ \mu_{22}\\ \mu_{32}\end{bmatrix} = \begin{bmatrix}\mu_{13}\\ \mu_{23}\\ \mu_{33}\end{bmatrix}$$

2. Estadístico de prueba

Como No. de Variables Dependientes : p≥1 y

No. de Niveles del Factor: g=3

Usamos la siguiente fórmula (Caso 4):

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Donde\ £ = \frac{|error|}{|tratamiento|} = \frac{|SCE|}{|SCT|}$$

$$SCE = \sum_{i=1}^{3} SCEi = \sum_{i=1}^{3} D_i' D_i$$

$$SCT = D'D$$

Para SCE:

$$D_1 = \begin{bmatrix} 1.234 & 16 & 43 \\ 0.894 & 15 & 62 \\ 0.950 & 16 & 43 \\ 0.702 & 16 & 28 \\ 0.718 & 14 & 36 \end{bmatrix} - \begin{bmatrix} 0.9 & 15.4 & 42.4 \\ 0.9 & 15.4 & 42.4 \\ 0.9 & 15.4 & 42.4 \\ 0.9 & 15.4 & 42.4 \\ 0.9 & 15.4 & 42.4 \end{bmatrix} = \begin{bmatrix} 0.3344 & 0.6 & 0.6 \\ -0.0056 & -0.4 & 19.6 \\ 0.0504 & 0.6 & 0.6 \\ -0.19760 & 0.6 & -14.4 \\ -0.1816 & -1.4 & -6.4 \end{bmatrix}$$

$$D'_1 D_1 = \begin{bmatrix} 0.3344 & -0.0056 & 0.0504 & -0.1976 & -0.1816 \\ 0.6 & -0.4 & 0.6 & 0.6 & -1.4 \\ 0.6 & 19.6 & 0.6 & -14.4 & -6.4 \end{bmatrix} \begin{bmatrix} 0.3344 & 0.6 & 0.6 \\ -0.0056 & -0.4 & 19.6 \\ 0.0504 & 0.6 & 0.6 \\ -0.19760 & 0.6 & -14.4 \\ -0.1816 & -1.4 & -6.4 \end{bmatrix}$$

$$= \begin{bmatrix} 0.18642 & 0.3688 & 4.1288 \\ 0.3688 & 3.2 & -6.8 \\ 4.1288 & -6.8 & 633.2 \end{bmatrix}$$

$$D^2 = \begin{bmatrix} 0.542 & 12 & 40 \\ 0.326 & 14 & 21 \\ 0.380 & 14 & 34 \\ 0.436 & 12 & 65 \\ 0.404 & 10 & 63 \\ 0.442 & 10 & 40 \end{bmatrix} - \begin{bmatrix} 0.422 & 12 & 43.833 \\ 0.422 & 12 & 43.833 \\ 0.422 & 12 & 43.833 \\ 0.422 & 12 & 43.833 \\ 0.422 & 12 & 43.833 \\ 0.422 & 12 & 43.833 \end{bmatrix} = \begin{bmatrix} 0.12033 & 0.00000 & -3.83333 \\ -0.09567 & 2 & -22.83333 \\ -0.04167 & 2 & -9.83333 \\ 0.01433 & 0.00000 & 21.16667 \\ -0.01767 & -2 & 19.16667 \\ 0.02033 & -2 & -3.83333 \end{bmatrix}$$

$D'_2 D_2$

$$= \begin{bmatrix} 0.12033 & -0.09567 & -0.04167 & 0.01433 & -0.01767 & 0.02033 \\ 0 & 2 & 2 & 0 & -2 & -2 \\ -3.83333 & -22.83333 & -9.83333 & 21.16667 & 19.16667 & -3.83333 \end{bmatrix} \begin{bmatrix} 0.12033 & 0.00000 & -3.83333 \\ -0.09567 & 2 & -22.83333 \\ -0.04167 & 2 & -9.83333 \\ 0.01433 & 0.00000 & 21.16667 \\ -0.01767 & -2 & 19.16667 \\ 0.02033 & -2 & -3.83333 \end{bmatrix}$$

$$= \begin{bmatrix} 0.02630 & -0.28 & 2.01967 \\ -0.28 & 16 & -96 \\ 2.01967 & -96 & 1462.83333 \end{bmatrix}$$

$$D_3 = \begin{bmatrix} 0.21 & 11 & 20 \\ 0.230 & 12 & 18 \\ 0.204 & 9 & 53 \\ 0.292 & 13 & 27 \\ 0.240 & 10 & 22 \\ 0.238 & 12 & 17 \\ 0.214 & 11 & 45 \end{bmatrix} \begin{bmatrix} 0.233 & 11.143 & 28.857 \\ 0.233 & 11.143 & 28.857 \\ 0.233 & 11.143 & 28.857 \\ 0.233 & 11.143 & 28.857 \\ 0.233 & 11.143 & 28.857 \\ 0.233 & 11.143 & 28.857 \\ 0.233 & 11.143 & 28.857 \end{bmatrix} = \begin{bmatrix} -0.02257 & -0.14286 & -8.85714 \\ -0.00257 & 0.85714 & -10.85714 \\ -0.02857 & -2.14286 & 24.14286 \\ 0.05943 & 1.85714 & -1.85714 \\ 0.00743 & -1.14286 & -6.85714 \\ 0.00543 & 0.85714 & -11.85714 \\ -0.01857 & -0.14286 & 16.14286 \end{bmatrix}$$

$$D'_3 D_3 = \begin{bmatrix} -0.02257 & -0.00257 & -0.02857 & 0.05943 & 0.00743 & 0.00543 & -0.01857 \\ -0.14286 & 0.85714 & -2.14286 & 1.85714 & -1.14286 & 0.85714 & -0.14286 \\ -8.85714 & -10.85714 & 24.14286 & -1.85714 & -6.85714 & -11.85714 & 16.14286 \end{bmatrix} \begin{bmatrix} -0.02257 & -0.14286 & -8.85714 \\ -0.00257 & 0.85714 & -10.85714 \\ -0.02857 & -2.14286 & 24.14286 \\ 0.05943 & 1.85714 & -1.85714 \\ 0.00743 & -1.14286 & -6.85714 \\ 0.00543 & 0.85714 & -11.85714 \\ -0.01857 & -0.14286 & 16.14286 \end{bmatrix}$$

$$= \begin{bmatrix} 0.00529 & 0.17143 & -0.98743 \\ 0.17143 & 10.85714 & -67.85714 \\ -0.98743 & -67.85714 & 1230.85714 \end{bmatrix}$$

$$SCE = \sum_{i=1}^{3} D'_i D_i$$

$$\begin{bmatrix} 0.18642 & 0.3688 & 4.1288 \\ 0.3688 & 3.2 & -6.8 \\ 4.1288 & -6.8 & 633.2 \end{bmatrix} + \begin{bmatrix} 0.02630 & -0.28 & 2.01967 \\ -0.28 & 16 & -96 \\ 2.01967 & -96 & 1462.83333 \end{bmatrix} + \left[\begin{bmatrix} 0.00529 & 0.17143 & -0.98743 \\ 0.17143 & 10.85714 & -67.85714 \\ -0.98743 & -67.85714 & 1230.85714 \end{bmatrix} \right]$$

$$SCE = \begin{bmatrix} 0.21801 & 0.26023 & 5.16104 \\ 0.26023 & 30.05714 & -170.65714 \\ 5.16104 & -170.65714 & 3326.89048 \end{bmatrix}$$

Utilizando la regla de Sarrus:

$$|SCE| = \begin{vmatrix} 0.21801 & 0.26023 & 5.16104 \\ 0.26023 & 30.05714 & -170.65714 \\ 5.16104 & -170.65714 & 3326.89048 \\ 0.21801 & 0.26023 & 5.16104 \\ 0.26023 & 30.05714 & -170.65714 \end{vmatrix} =$$

$$(0.21801 \times 30.05714 \times 3326.89048) + (0.26023 \times -170.65714 \times 5.16104) + (5.16104 \times 0.26023 \times -170.65714) - (5.16104 \times 30.05714 \times 5.16104) - (0.21801 \times -170.65714 \times -170.65714) - (0.26023 \times 0.26023 \times 3326.89048) = 13{,}966.8668$$

Para SCT:

$$D\begin{bmatrix}1.234 & 16 & 43\\ 0.894 & 15 & 62\\ 0.950 & 16 & 43\\ 0.702 & 16 & 28\\ 0.718 & 14 & 36\\ 0.542 & 12 & 40\\ 0.326 & 14 & 21\\ 0.380 & 14 & 34\\ 0.436 & 12 & 65\\ 0.404 & 10 & 63\\ 0.442 & 10 & 40\\ 0.210 & 11 & 20\\ 0.230 & 12 & 18\\ 0.204 & 9 & 53\\ 0.292 & 13 & 27\\ 0.240 & 10 & 22\\ 0.238 & 12 & 17\\ 0.214 & 11 & 45\end{bmatrix} - \begin{bmatrix}0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\\ 0.481 & 12.611 & 37.611\end{bmatrix} = \begin{bmatrix}0.75311 & 3.38889 & 5.38889\\ 0.41311 & 2.38889 & 24.38889\\ 0.46911 & 3.38889 & 5.38889\\ 0.22111 & 3.38889 & -9.61111\\ 0.23711 & 1.38889 & -1.61111\\ 0.06111 & -0.61111 & 2.38889\\ -0.15489 & 1.38889 & -16.61111\\ -0.10089 & 1.38889 & -3.61111\\ -0.04489 & -0.61111 & 27.38889\\ -0.07689 & -2.61111 & 25.38889\\ -0.03889 & -2.61111 & 2.38889\\ -0.27089 & -1.61111 & -17.61111\\ -0.25089 & -0.61111 & -19.61111\\ -0.27689 & -3.61111 & 15.38889\\ -0.18889 & 0.38889 & -10.61111\\ -0.24089 & -2.61111 & -15.61111\\ -0.24289 & -0.61111 & -20.61111\\ -0.26689 & -1.61111 & 7.38889\end{bmatrix}$$

$$SCT = D'D = \begin{bmatrix}0.75311 & 0.41311 & 0.46911 & \dots & -0.26689\\ 3.38889 & 2.38889 & 3.38889 & \dots & -1.61111\\ 5.38889 & 24.38889 & 5.38889 & \dots & 7.38889\end{bmatrix}\begin{bmatrix}0.75311 & 3.38889 & 5.38889\\ 0.41311 & 2.38889 & 24.38889\\ 0.46911 & 3.38889 & 5.38889\\ \vdots & \vdots & \vdots\\ -0.26689 & -1.61111 & 7.38889\end{bmatrix}$$

$$= \begin{bmatrix}1.54728 & 8.86822 & 28.19222\\ 8.86822 & 86.27778 & -36.72222\\ 28.19222 & -36.72222 & 4210.27778\end{bmatrix}$$

Utilizando la regla de Sarrus:

$$|SCT| = \begin{vmatrix}1.54728 & 8.86822 & 28.19222\\ 8.86822 & 86.27778 & -36.72222\\ 28.19222 & -36.72222 & 4210.27778\\ 1.54728 & 8.86822 & 28.19222\\ 8.86822 & 86.27778 & -36.72222\end{vmatrix}$$

$$= \begin{matrix}(1.54728 \times 86.27778 \times 4210.27778) + (8.86822 \times -36.72222 \times 28.19222) + \\ (28.19222 \times 8.86822 \times -36.72222) - (28.19222 \times 86.27778 \times 28.19222) - \\ (1.54728 \times -36.72222 \times -36.72222) - (8.86822 \times 8.86822 \times 4210.27778)\end{matrix} = 141{,}914.103$$

$$£ = \frac{|SCE|}{|SCT|} = \frac{13{,}966.8668}{141{,}914.103} = \underline{0.09842}$$

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1 - \sqrt{£})}{\sqrt{£}}$$

$$Fcalc = \frac{(18 - 3 - 2)}{3}\frac{(1 - \sqrt{0.09842})}{\sqrt{0.09842}}$$

$$Fcalc = \frac{(13)}{3}\frac{(1 - 0.31372)}{0.31372}$$

$$Fcalc = (4.333)(2.18756)$$

$$Fcalc = 9.47868$$

3. Región de rechazo

$$Fcrit = \left[\propto, 2, p, 2\left(\sum n_1 - p - 2\right)\right]$$

$$Fcrit = \left[0.05{,}2(3){,}2\left(\sum 18 - 3 - 2\right)\right]$$

$$Fcrit = [0.05{,}6{,}2(13)]$$

$$Fcrit = 0.05{,}6{,}26$$

$$F0.05 = 2.47$$

$$F0.01 = 3.59$$

4. Regla de Decisión

Se rechaza H_0 sí $fcal \geq 2.47$

5. Conclusión

Estadista: Como 9.48 > 2.47 > 3.59 ∴ la prueba es (A.S) y se rechaza con mucha potencia.

Administrativa: Existe evidencia suficiente para decir que estadísticamente al menos una de las medias en cada una de las **tres variables dependientes** de las **Tres Áreas de trabajo** NO son iguales. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Resolución del inciso b)

Prueba de Hipótesis para Ingresos:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el área de trabajo de los empleados de Industrias México es el mismo sobre los **Ingresos** de los empleados.

$$H_0: \mu_{11.} = \mu_{12.} = \mu_{13.}$$

La hipótesis alternativa es No todas las **Áreas de trabajo** son iguales sobre los **Ingresos** de los empleados.

H_1: No todas las Áreas de trabajo utilizadas para medir los Ingresos, la Escolaridad y la Edad son iguales.

También se puede pensar que la hipótesis alternativa sea "al menos una de las Áreas de trabajo utilizadas para medir la influencia sobre los **Ingresos** de los empleados es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de los **Ingresos** con base en el Área de trabajo es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos una de las **Áreas de trabajo** tiene una diferencia significativa sobre los **Ingresos** de los empleados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Área	Millones de Pesos							Total fila	Medias filas
	1	2	3	4	5	6	7		
Gerencial	1.234	0.894	0.950	0.702	0.718			$X_{1.} = 4.498$	$\bar{X}_{1.} = 0.8996$
Servicios	0.542	0.326	0.380	0.436	0.404	0.442		$X_{2.} = 2.53$	$\bar{X}_{2.} = 0.42167$
Obreros	0.210	0.230	0.204	0.292	0.240	0.238	0.214	$X_{3.} = 1.628$	$\bar{X}_{3.} = 0.23257$
					Totales de filas:			$X_{..} = 8.656$	$\bar{X}.. = 0.48089$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (1.234^2 + 0.894^2 + \cdots + 0.214^2) - \frac{8.656^2}{18} = 5.70986 - 4.16257 = \mathbf{1.54729}$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{4.498^2}{5} + \frac{2.53^2}{6} + \frac{1.628^2}{7}\right) - \frac{8.656^2}{18} = 5.49184 - 4.16257 = \mathbf{1.32927}$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k}\frac{X_{i.}^2}{n} = 1234^2 + 0894^2 + \cdots + 0214^2 - (\frac{25^2}{5} + \frac{51^2}{6} + \frac{40^2}{7}) = 5.70986 - 5.49184$$
$$= \mathbf{0.21802}$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamien-tos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 1.32927$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{1.32927}/_{2}$ $= 0.66464$	$F_{Calc.} = {}^{CMt}/_{CME} = /$ $= {}^{0.66464}/_{0.01453}$ $= 45.7426$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 0.21802$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{0.21802}/_{15}$ $= 0.01453$	
Total	18-1=17	$SCT = 1.54729$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denomina-dor	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **45.7426**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Área de trabajo sobre los **Ingresos** no es el mismo en al menos una de las tres Áreas de trabajo.

Prueba de Hipótesis para Escolaridad:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el área de trabajo de los empleados de Industrias México es el mismo sobre la **Escolaridad** de los empleados.

$$H_0\colon\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todas las **Áreas de trabajo** son iguales sobre la **Escolaridad** de los empleados.

H_1: No todas las Áreas de trabajo utilizadas para medir los Ingresos, la Escolaridad y la Edad son iguales.

También se puede pensar que la hipótesis alternativa sea "al menos una de las Áreas de trabajo utilizadas para medir la influencia sobre la **Escolaridad** de los empleados es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de la **Escolaridad** con base en el Área de trabajo es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos una de las **Áreas de trabajo** tiene una diferencia significativa sobre la **Escolaridad** de los empleados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Área	Años							Total fila	Medias filas
	1	2	3	4	5	6	7		
Gerencial	16	15	16	16	14			$X_{1.} = 77$	$\bar{X}_{1.} = 15.4$
Servicios	12	14	14	12	10	10		$X_{2.} = 72$	$\bar{X}_{2.} = 12$
Obreros	11	12	9	13	10	12	11	$X_{3.} = 78$	$\bar{X}_{3.} = 11.14286$
					Totales de filas:			$X_{..} = 227$	$\bar{X}.. = 12.61111$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (16^2 + 15^2 + \cdots + 11^2) - \frac{227^2}{18} = 2949 - 2862.72222 = \mathbf{86.27778}$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{77^2}{5} + \frac{72^2}{6} + \frac{78^2}{7}\right) - \frac{227^2}{18} = 2918.94286 - 2862.72222 = \mathbf{56.22064}$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 16^2 + 15^2 + \cdots + 11^2 - (\frac{77^2}{5} + \frac{72^2}{6} + \frac{78^2}{7}) = 2949 - 2918.94286 = \mathbf{30.05714}$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamientos	$v_1 = k - 1 = 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 56.22064$	$CMt = {}^{SCt}/_{g.l.} = {}^{56.22064}/_{2} = 28.11032$	$F_{Calc.} = {}^{CMt}/_{CME} = {}^{28.11032}/_{2.00381} = \mathbf{14.02844}$
Error	$v_2 = N - k = 18 - 3 = \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 30.05714$	$CME = {}^{SCE}/_{g.l.} = {}^{30.05714}/_{15} = 2.00381$	
Total	18-1=17	$SCT = 56.22064$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **14.028**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son

iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Área de trabajo sobre la **Escolaridad** no es el mismo en al menos una de las tres Áreas de trabajo.

Prueba de Hipótesis para Edad:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el área de trabajo de los empleados de Industrias México es el mismo sobre la **Edad** de los empleados.

$$H_0\colon\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todas las **Áreas de trabajo** son iguales sobre la **Edad** de los empleados.

H_1: No todas las Áreas de trabajo utilizadas para medir los Ingresos, la Escolaridad y la Edad son iguales.

También se puede pensar que la hipótesis alternativa sea "al menos una de las Áreas de trabajo utilizadas para medir la influencia sobre la **Edad** de los empleados es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de la **Edad** con base en el Área de trabajo es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos una de las **Áreas de trabajo** tiene una diferencia significativa sobre la **Edad** de los empleados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Empiece por calcular las sumas y medias para cada tratamiento así como la suma y media total:

Área	Años							Total fila	Medias filas
	1	2	3	4	5	6	7		
Gerencial	43	62	43	28	36			$X_{1.} = 212$	$\bar{X}_{1.} = 42.4$
Servicios	40	21	34	65	63	40		$X_{2.} = 263$	$\bar{X}_{2.} = 43.83333$
Obreros	20	18	53	27	22	17	45	$X_{3.} = 202$	$\bar{X}_{3.} = 28.85714$
					Totales de filas:			$X_{..} = 677$	$\bar{X}.. = 12.61111$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = (43^2 + 62^2 + \cdots + 45^2) - \frac{677^2}{18} = 29{,}673 - 25{,}462.72222 = \mathbf{4{,}210.3}$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = \left(\frac{212^2}{5} + \frac{263^2}{6} + \frac{202^2}{7}\right) - \frac{677^2}{18} = 26{,}346.10952 - 25{,}462.72222 = \mathbf{883.4}$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = (43^2 + 62^2 + \cdots + 45^2) - \left(\frac{212^2}{5} + \frac{263^2}{6} + \frac{202^2}{7}\right) = 29{,}673 - 26{,}346.10952$$
$$= \mathbf{3{,}326.9}$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	$F_{calculada}$
Tratamientos	$v_1 = k - 1$ $= 3 - 1$ $= \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 883.4$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{883.4}/_{2} = 441.7$	$F_{Calc.} = {}^{CMt}/_{CME} = /$ $= {}^{441.7}/_{221.79333}$ $= 1.99$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 3{,}326.9$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{3{,}326.9}/_{15}$ $= 221.79333$	
Total	18-1=17	$SCT = 4{,}210.3$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **1.99**, que es menor al valor crítico de 3.68; por lo tanto la hipótesis nula NO se rechaza y llegamos a la conclusión de que todas las medias de la población son iguales..

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Área de trabajo sobre la **Edad** de los empleados es el mismo en las tres Áreas de trabajo utilizadas.

Resolución del inciso c)

Prueba de Tukey

Intervalo de confianza para Ingresos (Gerencial-Servicios) .Modelo desbalanceado.

$$M_{11} - M_{12} = (\bar{X}_{11} - \bar{X}_{12}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = (0.9 - 0.422) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = 0.478 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{0.218}{15} = 0.01453$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{11} - M_{12} = 0.478 \mp 3.67\sqrt{\frac{0.01453}{5.4545}}$$

$$M_{11} - M_{12} = 0.478 \mp 3.67\sqrt{0.00266}$$

$$M_{11} - M_{12} = 0.478 \mp 3.67(0.05161)$$

$$M_{11} - M_{12} = 0.478 \mp 0.18942 \begin{cases} LIC = 0.28858 \\ LSC = 0.66742 \end{cases}$$

Interpretación:

Como ambos límites son positivos podemos decir que estadísticamente el grado de influencia que tiene el área gerencial sobre los ingresos es mayor que el área de servicios sobre los ingresos por un mínimo de 0.28858 y un máximo de 0.66742 millones de pesos.

Intervalo de confianza para Ingresos (Gerencial-Obreros)

$$M_{11} - M_{13} = (\bar{X}_{11} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = (0.8996 - 0.2326) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = 0.667 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = 5.8333$$

$$M_{11} - M_{13} = 0.667 \mp 3.67\sqrt{\frac{0.01453}{5.8333}}$$

$$M_{11} - M_{13} = 0.667 \mp 3.67(0.04991)$$

$$M_{11} - M_{13} = 0.667 \mp 0.18316 \begin{cases} LIC = 0.48384 \\ LSC = 0.85016 \end{cases}$$

Interpretación:

Como ambos límites son positivos podemos decir que estadísticamente el grado de influencia que tiene el área gerencial sobre los ingresos es mayor que el área de obreros sobre los ingresos por un mínimo de 0.48384 y un máximo de 0.85016 millones de pesos.

Intervalo de confianza para Ingresos (Servicios-Obreros)

$$M_{12} - M_{13} = (\bar{X}_{12} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = (0.4217 - 0.2326) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = 0.1891 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = 6.4615$$

$$M_{12} - M_{13} = 0.1891 \mp 3.67 \sqrt{\frac{0.01453}{6.4615}}$$

$$M_{12} - M_{13} = 0.1891 \mp 3.67(0.04742)$$

$$M_{12} - M_{13} = 0.1891 \mp 0.17403 \begin{cases} LIC = 0.01507 \\ LSC = 0.36313 \end{cases}$$

Interpretación:

Como ambos límites son positivos podemos decir que estadísticamente el grado de influencia que tiene el área de servicios sobre los ingresos es mayor que el área de obreros sobre los ingresos por un mínimo de 0.01507 y un máximo de 0.36313 millones de pesos.

Intervalo de confianza para Escolaridad (Gerencial-Servicios)

$$M_{21} - M_{22} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = (15.4 - 12) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = 3.4 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N - k} = \frac{SCE}{18 - 3} = \frac{30.06}{15} = \underline{2.004}$$

$$nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{21} - M_{22} = 3.4 \mp 3.67 \sqrt{\frac{2.004}{5.4545}}$$

$$M_{21} - M_{22} = 3.4 \mp 3.67(0.60614)$$

$$M_{21} - M_{22} = 3.4 \mp 2.22453 \begin{cases} 1.17547 \\ 5.62453 \end{cases}$$

Interpretación:

Como ambos límites son positivos podemos decir que estadísticamente el grado de influencia que tiene el área gerencial sobre la escolaridad es mayor que el área de servicios sobre la escolaridad por un mínimo de 1.17547 y un máximo de 5.62453 años.

Intervalo de confianza para Escolaridad (Gerencial – Obreros)

$$M_{21} - M_{23} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = (15.4 - 11.143) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = 4.257 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{21} - M_{23} = 4.257 \mp 3.67 \sqrt{\frac{2.004}{5.8326}}$$

$$M_{21} - M_{23} = 4.257 \mp 3.67(0.58616)$$

$$M_{21} - M_{23} = 4.257 \mp 2.15122 \begin{cases} 2.10578 \\ 6.40822 \end{cases}$$

Interpretación:

Como ambos límites son positivos podemos decir que estadísticamente el grado de influencia que tiene el área gerencial sobre la escolaridad es mayor que el área de obreros sobre la escolaridad por un mínimo de 2.10578 y un máximo de 6.40822 años.

Intervalo de confianza para Escolaridad (Servicios-Obreros)

$$M_{22} - M_{23} = (\bar{X}_{22} - \bar{X}_{23}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = (12 - 11.143) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = 0.857 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{6}+\frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{22} - M_{23} = 0.857 \mp 3.67\sqrt{\frac{2.004}{6.4620}}$$

$$M_{22} - M_{23} = 0.857 \mp 3.67(0.55688)$$

$$M_{22} - M_{23} = 0.857 \mp 2.04377 \begin{cases} LIC = -1.18677 \\ LSC = 2.90077 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente el grado de influencia que tiene el área de servicios sobre la escolaridad es el mismo que el área de obreros sobre la escolaridad.

Interpretación de confianza para Edad (Gerencial- Servicios)

$$M_{31} - M_{32} = (\bar{X}_{31} - \bar{X}_{32}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = (42.4 - 43.83) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = -1.43 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{3{,}326.9}{15} = 221.79333$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{31} - M_{32} = -1.43 \mp 3.67\sqrt{\frac{221.79333}{5.4545}}$$

$$M_{31} - M_{32} = -1.43 \mp 3.67(6.37671)$$

$$M_{31} - M_{32} = -1.43 \mp 23.40253 \begin{cases} LIC = -24.83253 \\ LSC = 21.97253 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente el grado de influencia que tiene el área gerencial sobre la edad es el mismo que el área de servicios sobre la edad.

Intervalo de confianza para Edad (Gerencial – Obreros)

$$M_{31} - M_{33} = (\bar{X}_{31} - \bar{X}_{33}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = (42.40 - 28.86) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = 13.54 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{31} - M_{33} = 13.54 \mp 3.67 \sqrt{\frac{221.79333}{5.8326}}$$

$$M_{31} - M_{33} = 13.54 \mp 3.67(6.16656)$$

$$M_{31} - M_{33} = 13.54 \mp 22.63128 \begin{cases} -9.09128 \\ 36.17128 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente el grado de influencia que tiene el área gerencial sobre la edad es el mismo que el área de obreros sobre la edad.

Intervalo de confianza para Edad (Servicios-Obreros)

$$M_{32} - M_{33} = (\bar{X}_{32} - \bar{X}_{33}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = (43.83 - 28.86) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = 14.97 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

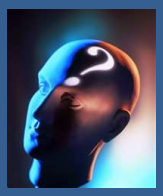
$$M_{32} - M_{33} = 14.97 \mp 3.67\sqrt{\frac{221.79333}{6.4620}}$$

$$M_{32} - M_{33} = 14.97 \mp 3.67(5.85856)$$

$$M_{32} - M_{33} = 14.97 \mp 21.50091 \begin{cases} LIC = -6.53091 \\ LSC = 36.47091 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente el grado de influencia que tiene el área de servicios sobre la edad es el mismo que el área de obreros sobre la edad.



RESPUESTAS A LOS EJERCICIOS DE **REFUERZO** DEL CAPÍTULO 3

3.1.1.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.1.1.1.1. Análisis Discriminante**

3.1.1.1.1 Supongamos que una compañía de Piscinas ha reunido datos de ingreso anual y tamaño de terreno de grupos de propietarios de piscina y no propietarios, quienes viven en el sureste de Mèxico. Además, hay datos disponibles para cada grupo sobre actitudes hacia los baños de sol, en escala de 0= *detestar los baños de sol* hasta 10= *extremadamente aficionado a tomar baños de sol.* Los datos se resumen a continuación:

No. De Sujeto	**Grupo basado en Propiedad de Piscina**	**Ingreso (000'000'pesos) (X_1)**	**Tamaño del lote (Miles de m^2) ($X_{2)}$)**	**Medida de Actitud ($X_{3)}$)**
1	Propietario	0.7	4.7	9
2	Propietario	0.9	5.6	10
3	Propietario	1.4	9.3	8
4	Propietario	0.7	2.3	5
5	Propietario	0.5	3.8	7
	Media Grupo 1	**0.84**	**5.14**	**7.8**
6	No Propietario	0.5	3.3	4
7	No Propietario	0.7	4.0	3
8	No Propietario	0.6	3.6	2

9	No Propietario	0.4	2.8	4
10	No Propietario	0.5	2.4	5
	Media Grupo 2	**0.54**	**3.22**	**3.6**

a) Obtener el modelo discriminante.
b) Probar si el modelo es bueno para pronosticar.
c) Validar el modelo con los mismos datos de derivación.
d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados.
e) En que grupo caería una persona con las siguientes características:

Ingreso anual = $ 600,000

Tamaño del terreno= 4,180 metros cuadrados.

Actitud=8

Resolución del inciso a)

La ecuación para derivar el modelo discriminante tendrá la siguiente forma:

$$Z = K_1X_1 + K_2X_2 + K_3X_3$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las *K,* para ello se deben resolver las siguientes ecuaciones normales:

$$\sum X_1'^2 K_1 + \sum X_1'X_2'K_2 + \sum X_1'\,X_3'K_3 = \bar{X}_1\,|Grupo\ 1 - \bar{X}_1|Grupo\ 2$$

$$\sum X_1'X_2'K_1 + \sum X_2'^2 K_2 + \sum X_2'\,X_3'K_3 = \bar{X}_2\,|Grupo\ 1 - \bar{X}_2|Grupo\ 2$$

$$\sum X_1'X_3'K_1 + \sum X_2'X_3'K_2 + \sum X_3'^2\,K_3 = \bar{X}_3\,|Grupo\ 1 - \bar{X}_3|Grupo\ 2$$

Donde:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1 \bar{X}_2)$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1 \bar{X}_3)$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2 \bar{X}_3)$$

Se sustituyen los valores correspondientes, obteniendo lo siguiente:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_1'^2 = [(0.7^2 + 0.9^2 + 1.4^2 + 0.7^2 + 0.5^2) - 5(0.84)^2] + [(0.5^2 + 0.7^2 + 0.6^2 + 0.4^2 + 0.5^2) - 5(0.54)^2] = 0.524$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_2'^2 = [(4.7^2 + 5.6^2 + 9.3^2 + 2.3^2 + 3.8^2) - 5(5.14)^2] + [(3.3^2 + 4.0^2 + 3.6^2 + 2.8^2 + 2.4^2) - 5(3.22)^2] = 29.18$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_3'^2 = [(9^2 + 10^2 + 8^2 + 5^2 + 7^2) - 5(7.8)^2] + [(4^2 + 3^2 + 2^2 + 4^2 + 5^2) - 5(3.6)^2] = 20$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\bar{X}_2)$$

$$=[(0.7x4.7 + 0.9x5.6 + 1.4x9.3 + 0.7x2.3 + 0.5x3.8) - 5(0.84x5.14)] + [(0.5x3.3 + 0.7x4.0 + 0.6x3.6 + 0.4x2.8 + 0.5x2.4) - 5(0.54x3.22)] = 3.508$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\bar{X}_3)$$

$$=[(0.7x9 + 0.9x10 + 1.4x8 + 0.7x5 + 0.5x7) - 5(0.84x7.8)] + [(0.5x4 + 0.7x3 + 0.6x2 + 0.4x4 + 0.5x5) - 5(0.54x3.6)] = 0.420$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\bar{X}_3)$$

$$=[(4.7x9 + 5.6x10 + 9.3x8 + 2.3x5 + 3.8x7) - 5(5.14x7.8)] + [(3.3x4 + 4.0x3 + 3.6x2 + 2.8x4 + 2.4x5) - 5(3.22x3.6)] = 7.980$$

De lo anterior se obtienen tres ecuaciones, normales:

$$0.524K_1 + 3.508K_2 + 0.420K_3 = 0.3$$

$$3.508K_1 + 29.180K_2 + 7.980K_3 = 1.92$$

$$0.420K_1 + 7.980K_2 + 20K_3 = 4.2$$

A partir de estas ecuaciones se obtienen los valores de *K* de cada variable. En este caso se utilizará la matriz inversa para la resolución del sistema lineal siguiendo los siguientes pasos:

Paso 1. Expresar el sistema como A *K=D

Donde:

A= Matriz de coeficientes.

K= Coeficientes discriminantes.

D= Términos independientes.

$$\begin{bmatrix} 0.524 & 3.508 & 0.420 \\ 3.508 & 29.180 & 7.980 \\ 0.420 & 7.980 & 20 \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 0.3 \\ 1.92 \\ 4.2 \end{bmatrix}$$

Paso 2. Obtener el determinante de A aplicando la regla de Sarrus para comprobar si el sistema tiene solución, es decir si |A|≠0:

$$|A| = \begin{vmatrix} 0.524 & 3.508 & 0.420 \\ 3.508 & 29.180 & 7.980 \\ 0.420 & 7.980 & 20 \\ 0.524 & 3.508 & 0.420 \\ 3.508 & 29.180 & 7.980 \end{vmatrix} =$$ (0.524*29.18*20)+ (3.508*7.98*0.42)+ (0.42*3.508*7.98)- (0.42*29.18*0.42)- (0.524*7.98*7.98)- (3.508*3.508*20)=44.684

Como 44.684 es ≠ de 0 el sistema tiene solución.

Paso 3. Obtener la matriz transpuesta de A, que es:

$$A' = \begin{bmatrix} 0.524 & 3.508 & 0.42 \\ 3.508 & 29.18 & 7.98 \\ 0.42 & 7.98 & 20 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A', es decir formamos $\alpha A'$

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 29.18 & 7.98 \\ 7.98 & 20 \end{vmatrix} & -\begin{vmatrix} 3.508 & 7.98 \\ 0.42 & 20 \end{vmatrix} & +\begin{vmatrix} 3.508 & 29.18 \\ 0.42 & 7.98 \end{vmatrix} \\ -\begin{vmatrix} -1 & 0 \\ -4 & 8 \end{vmatrix} & +\begin{vmatrix} 0.524 & 0.42 \\ 0.42 & 20 \end{vmatrix} & -\begin{vmatrix} 0.524 & 3.508 \\ 0.42 & 7.98 \end{vmatrix} \\ +\begin{vmatrix} -1 & 0 \\ 14 & -4 \end{vmatrix} & -\begin{vmatrix} 0.524 & 0.42 \\ 3.508 & 7.98 \end{vmatrix} & +\begin{vmatrix} 0.524 & 3.508 \\ 3.508 & 29.18 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(583.6 - 63.68) & -(70.16 + 3.352) & +(27.994 - 12.256) \\ -(70.16 - 3.352) & +(10.48 - 0.176) & -(4.182 + 1.473) \\ +((27.994 - 12.256) & -(4.182 - 1.473) & +(15.29 - 12.306) \end{bmatrix} =$$

$$= \begin{bmatrix} 519.920 & -66.808 & 15.738 \\ -66.808 & 10.304 & -2.708 \\ 15.738 & -2.708 & 2.984 \end{bmatrix}$$

Paso 5. Calculamos la matriz inversa de A:

$$A^{-1} = \frac{1}{|A|} \propto A^{1} = \frac{1}{44.684}\begin{bmatrix} 519.920 & -66.808 & 15.738 \\ -66.808 & 10.304 & -2.708 \\ 15.738 & -2.708 & 2.984 \end{bmatrix} = \begin{bmatrix} \frac{519.920}{44.684} & \frac{-66.808}{44.684} & \frac{15.738}{44.684} \\ \frac{-66.808}{44.684} & \frac{10.304}{44.684} & \frac{-2.708}{44.684} \\ \frac{15.738}{44.684} & \frac{-2.708}{44.684} & \frac{2.984}{44.684} \end{bmatrix}$$

Nota: Es conveniente dejar el resultado en forma de fracción para no perder precisión

Paso 6. Llevamos el valor anterior a la expresión $K = A^{-1} * D$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} \frac{519.920}{44.684} & \frac{-66.808}{44.684} & \frac{15.738}{44.684} \\ \frac{-66.808}{44.684} & \frac{10.304}{44.684} & \frac{-2.708}{44.684} \\ \frac{15.738}{44.684} & \frac{-2.708}{44.684} & \frac{2.984}{44.684} \end{bmatrix} \begin{bmatrix} 0.3 \\ 1.92 \\ 4.2 \end{bmatrix} = \begin{bmatrix} 2.09928 \\ -0.26036 \\ 0.26980 \end{bmatrix}$$

Paso 7. Obtenemos el modelo discriminante:

$$\boldsymbol{Z = 0.2.09928K_1 - 0.26036K_2 + 0.26980K_3}$$

Resolución del inciso b)

A continuación se prueba si el modelo es bueno para discriminar o clasificar a las personas que están de acuerdo con las que no lo están mediante los siguientes pasos:

Paso 1. Plantear el juego de hipótesis:

H_0. El modelo discriminante no es bueno para discriminar o clasificar a las personas que están o no de acuerdo con las características de las laptops analizadas.

H_1. El modelo discriminante es bueno.

Paso 2. Establecer el estadístico de prueba:

$$F\ calc = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2)D^2}{[(n_1 + n_2 - 2)p * (n_1 + n_2)]}$$

$$D^2 = (n_1 + n_2 - 2)\left[\sum K_1(\bar{X}_1|Grupo\ 1 - \bar{X}_1|Grupo2\right]$$

$$D^2 = (5 + 5 - 2)[(2.09928)(0.3) + (-0.26036)(1.92) + 0.26980(4.2)] = 10.10$$

$$F\ calc = \frac{[(5 + 5 - 2 - 1)(5 * 5)(10.10)]}{[(5 + 5 - 2)2](5 + 5)} = \frac{7x25x10.10}{16x10} = \frac{11727.45}{160} = 11.05$$

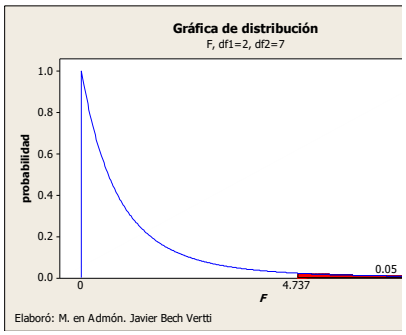
Paso 3. Establecer la región de Rechazo:

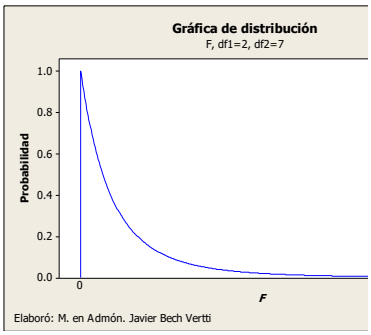
$$Fcrit = [\propto_1 p_1(n_1 + n_2 - p - 1)]$$

$$Fcrit = 0.05, 2, 7$$

$$f_{0.05} = 4.74$$

$$f_{0.01} = 9.55$$





Paso 4. Definir la regla de decisión:

Se rechaza H_0 si $F_{calc.} \geq 4.74$

Paso 5. Conclusiones:

-Estadística: Como 11.05>4.74 y > 9.55 ∴ la prueba es (AS) y se rechaza H_0 con mucha fuerza.

-Administrativa: Existe evidencia suficiente para decir que estadísticamente el modelo es bueno para clasificar o discriminar a los propietarios de piscina de los no propietarios.

Resolución del inciso c)

Cuando se ha concluido que el modelo discriminante es bueno para hacer clasificaciones, se debe proceder a validar el mismo. Esta validación consiste en determinar el grado de confiabilidad que se tendrá al hacer pronósticos con el modelo.

Paso 1. Se obtiene la Z promedio para cada grupo:

$Z = 2.09928K_1 - 0.26036K_2 + 0.26980K_3$

$\bar{Z}_1 = 2.09928(0.84) - 0.26036(5.14) + 0.26980(7.8) = 2.53$

$\bar{Z}_2 = 2.09928(0.54) - 0.26036(3.22) + 0.26980(3.6) = 1.27$

Paso 2. Se determina la Z crítica o de corte:

$$Z^*_{critica} = \frac{n_2\bar{Z}_1 + n_1\bar{Z}_2}{n_1 + n_2} = \frac{5(2.53) + 5(1.27)}{5 + 5} = 1.9$$

*Esta fórmula puede usarse ya sea si $n_1=n_2$ o si $n_1 \neq n_2$

Paso 3. Se obtiene la Z para cada observación:

$Z_1 = 2.0992(0.7) + 0.26036(4.7) + 0.26980(9) = 2.67 > 1.9$

$Z_2 = 2.0992(0.9) + 0.26036(5.6) + 0.26980(10) = 3.13 > 1.9$

$Z_3 = 2.0992(1.4) + 0.26036(9.3) + 0.26980(8) = 2.68 > 1.9$

$Z_4 = 2.0992(0.7) + 0.26036(2.3) + 0.26980(5) = 2.22 > 1.9$

$Z_5 = 2.0992(0.5) + 0.26036(3.8) + 0.26980(7) = 1.95 > 1.9$

$Z_6 = 2.0992(0.5) + 0.26036(3.3) + 0.26980(4) = 1.27 < 1.9$

$Z_7 = 2.0992(0.7) + 0.26036(4) + 0.26980(3) = 1.24 > 1.9$

$Z_8 = 2.0992(0.6) + 0.26036(3.6) + 0.26980(2) = 0.86 < 1.9$

$Z_9 = 2.0992(0.4) + 0.26036(2.8) + 0.26980(4) = 1.19 < 1.9$

$Z_{10} = 2.0992(0.5) + 0.26036(2.4) + 0.26980(5) = 1.77 > 1.9$

Paso 4. Se verifica si en las Z observadas la clasificación es correcta o incorrecta

Sujeto No.	**Grupos**	**Ingreso (000,000)pesos (X_1)**	**Tamaño del lote M^2 (X_2)**	**Actitud (X_3)**	**Z obs.**	**Clas.** ✓o **X**
1	Propietario	0.7	4.7	9	$2.67 > 1.9$	✓
2	Propietario	0.9	5.6	10	$3.13 > 1.9$	✓
3	Propietario	1.4	9.3	8	$2.68 > 1.9$	✓
4	Propietario	0.7	2.3	5	$2.22 > 1.9$	✓
5	Propietario	0.5	3.8	7	$1.95 > 1.9$	✓
	Media Gpo. 1	0.84	5.14	7.8		
6	No propietario	0.5	3.3	4	$1.27 < 1.9$	✓
7	No propietario	0.7	4.0	3	$1.24 > 1.9$	✓
8	No propietario	0.6	3.6	2	$0.86 < 1.9$	✓
9	No propietario	0.4	2.8	4	$1.19 < 1.9$	✓
10	No propietario	0.5	2.4	5	$1.77 > 1.9$	✓
	Media Gpo. 2	0.54	3.22	3.6		

Paso 5. Del paso anterior se obtiene el porcentaje de clasificaciones correctas (PCC):

$$PCC = \frac{\#\checkmark Grupo\ 1 + \#\checkmark Gupo\ 2}{\#\ de\ observaciones} * 100$$

$$PCC = \frac{5+5}{10} * 100 = 100\%$$

Paso 6. Se obtiene el porcentaje mínimo de clasificaciones correctas:

$$C = (p_i)^2 + (1 - p_1)^2$$

$$C = (\frac{5}{10})^2 + (1 - \frac{5}{10})^2$$

$$C = 0.25 + 0.25 = 0.50x100 = 50\%$$

Paso 7. Se compara el porcentaje de clasificaciones correctas contra el porcentaje mínimo de clasificaciones correctas:

Como 100% > 50% se puede validar el modelo para clasificar a los que les agradan o no los baños de sol al analizar ambos grupos.

Resolución del inciso d)

Para interpretar los resultados obtenidos anteriormente, se evalúa la importancia relativa que posee cada una de las variables independientes para discriminar. El método de los pesos discriminantes estandarizados consiste en estandarizar los coeficientes del modelo discriminante con el fin de que puedan ser comparados. Para jerarquizar las variables se considera solamente la magnitud del peso, no su signo, es decir el valor absoluto.

La fórmula para estandarizar los coeficientes discriminantes es:

$$K_n' = (K_n)\left[\frac{\sum X_n'^2}{(n_1 + n_2 - 2)}\right]^{0.05}$$

En el problema presentado, los coeficientes o constantes estandarizadas son:

1. Ingreso:
   $K_1 = (2.09928)\left[\frac{0.524}{(5+5-2)}\right]^{0.5} = 0.53727$
2. Tamaño:
   $K_2 = (-0.26036)\left[\frac{29.18}{(5+5-2)}\right]^{0.5} = -0.49724$
3. Actitud:
   $K_3 = (0.26980)\left[\frac{20}{(5+5-2)}\right]^{0.5} = 0.42659$

-Interpretación: De acuerdo a los pesos discriminantes estandarizados **en valores absolutos** (sin considerar el signo), podemos concluir que la variable que más discrimina o diferencia a un grupo de otro para los que les agradan o no los baños de sol es la variable X_1 (Ingreso) con un peso de .53727 por su mayor *K* estandarizada y la menos importante es la variable X_3 (Actitud) con un peso de 0.42659.

-Estrategias a seguir: Lo que se pretendería como estrategia es tratar de actuar sobre la variable dependiente (grupos), manipulando las variables independientes (Ingreso, Tamaño del lote y Actitud hacia los baños de sol). Lo que se quisiera lograr es aumentar los valores de Z, para hacer que las personas del grupo ***No propietaria*** se pasen al grupo ***Propietario.*** Para nuestro problema, tenemos que si X_1 (Ingreso) aumenta, subirá a su vez el valor de Z, pues su signo es positivo. Asímismo tenemos que si X_3 (Actitud) aumenta, subirá el valor de Z pues su signo es también positivo, en cambio para X_2 (Tamaño), tenemos que va a afectar en forma inversa al valor de Z, por su signo negativo.

De las tres variables, la primera que se debe cuidar es X_1 (Ingreso), ya que tiene una mayor importancia para discriminar entre los grupos, después podría pensarse en una estrategia en cuanto a la variable X_3 (Actitud), en un segundo plano, para incrementar la posibilidad de ser propietario de una piscina y por último una estrategia para la variable X_2 (Tamaño), en un tercer plano, que no es muy relevante en cuanto ser o no propietario de una piscina.

Resolución del inciso e)

$X_1 = 0.6$

$X_2 = 4.18$

$X_3 = 8$

Sustituimos los valores anteriores en el modelo discriminante:

$Z = 2.09928(0.6) - 0.26036(4.18) + 0.26980(8) = 2.33$

Comparamos el valor anterior con la Z crítica:

Como 2.33> 1.9 ∴ se le clasificaría en el grupo 1 de **Propietario**

3.1.1.1.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.1.1.1.2 Análisis Discriminante**

3.1.1.1.2. Un departamento de créditos bancarios está interesado en el posible uso del análisis discriminante en la selección preliminar de solicitudes de crédito. De registros anteriores la Institución ha reunido información sobre dos grupos de adjudicatarios de crédito: a) Riesgos buenos y b) Riesgos dudosos; con base en tres características: Ingresos anuales (en millones de pesos) (X_1), Número de tarjetas de Crédito (X_2) y Edad del solicitante (X_3). Reúne información sobre una muestra aleatoria de 10 adjudicatarios de crédito que a continuación se presenta:

No. De sujeto	Grupos basados en la magnitud del riesgo	Evaluación del riesgo		
		X_1 Ingreso anual (en millones de	X_2 Tarjetas de crédito (número)	X_3 Edad (años)

		pesos)		
1	Riesgo bueno	0.376	7	42
2	Riesgo bueno	0.348	6	47
3	Riesgo bueno	0.452	4	41
4	Riesgo bueno	0.486	5	37
5	Riesgo bueno	0.388	1	43
	Media del grupo 1	**0.410**	**5**	**42**
6	Riesgo dudoso	0.286	4	33
7	Riesgo dudoso	0.292	7	32
8	Riesgo dudoso	0.270	3	40
9	Riesgo dudoso	0.204	1	36
10	Riesgo dudoso	0.298	6	39
	Media del grupo 2	**0.270**	**4**	**36**

a) Obtenga el modelo discriminante que nos ayude a comprobar la existencia de dos grupos.

b) Pruebe si el modelo es bueno para pronosticar o clasificar el riesgo en la selección de solicitudes de crédito.

c) Valide el modelo discriminante con los mismos datos de derivación.

d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados indicando en orden de importancia las variables que logran segmentar y que estrategia se debería seguir.

f) ¿En qué grupo caería una persona con un ingreso de $ 340,000 pesos, 3 tarjetas y 50 años?

Resolución del inciso a)

La ecuación para derivar el modelo discriminante tendrá la siguiente forma:

$$Z = K_1X_1 + K_2X_2 + K_3X_3$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las *K,* para ello se deben resolver las siguientes ecuaciones normales:

$$\sum X_1'^2 K_1 + \sum X_1' X_2' K_2 + \sum X_1' X_3' K_3 = \bar{X}_1 \,|Grupo\ 1 - \bar{X}_1|Grupo\ 2$$

$$\sum X_1' X_2' K_1 + \sum X_2'^2 K_2 + \sum X_2' X_3' K_3 = \bar{X}_2 \,|Grupo\ 1 - \bar{X}_2|Grupo\ 2$$

$$\sum X_1' X_3' K_1 + \sum X_2' X_3' K_2 + \sum X_3'^2 K_3 = \bar{X}_3 \,|Grupo\ 1 - \bar{X}_3|Grupo\ 2$$

Donde:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1 \bar{X}_2)$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1 \bar{X}_3)$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2 \bar{X}_3)$$

Se sustituyen los valores correspondientes, obteniendo lo siguiente:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_1'^2 = [(0.376^2 + 0.348^2 + 0.452^2 + 0.486^2 + 0.388^2) - 5(0.410)^2] + [(0.286^2 + 0.292^2 + 0.270^2 + 0.204^2 + 0.298^2) - 5(0.270)^2] = 0.019$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_2'^2 = [(7^2 + 6^2 + 4^2 + 5^2 + 1^2) - 5(5)^2] + [(4^2 + 7^2 + 3^2 + 1^2 + 6^2) - 5(4)^2] = 44$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_3'^2 = [(42^2 + 47^2 + 41^2 + 37^2 + 43^2) - 5(42)^2] + [(33^2 + 32^2 + 40^2 + 36^2 + 39^2) - 5(36)^2] = 102$$

$$\sum X_1' X_2' = \sum (\sum X_1 X_2 - n\bar{X}_1\bar{X}_2)$$

$$=[(0.376x7 + 0.348x6 + 0.452x4 + 0.486x5 + 0.388x1) - 5(0.410x5)] + [(0.286x4 + 0.292x7 + 0.270x3 + 0.204x1 + 0.298x6) - 5(0.270x4)] = 0.236$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\bar{X}_3)$$

$$=[(0.376x42 + 0.348x47 + 0.452x41 + 0.486x37 + 0.388x43) - 5(0.410x42)] + [(0.286x33 + 0.292x32 + 0.270x40 + 0.204x36 + 0.298x39) - 5(0.270x36)] = -0.806$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\bar{X}_3)$$

$$=[(7x42 + 6x47 + 4x41 + 5x37 + 1x43) - 5(5x42)] + [(4x33 + 7x32 + 3x40 + 1x36 + 6x39) - 5(4x36)] = -8$$

De lo anterior se obtienen tres ecuaciones, normales:

$$0.019K_1 + 0.236K_2 - 0.806K_3 = 0.14$$

$$0.236K_1 + 44K_2 - 8K_3 = 0.4$$

$$-0.806K_1 - 8K_2 + 102K_3 = 6$$

A partir de estas ecuaciones se obtienen los valores de *K* de cada variable. En este caso se utilizará la matriz inversa para la resolución del sistema lineal siguiendo los siguientes pasos:

Paso 1. Expresar el sistema como A *K=D

Donde:

A= Matriz de coeficientes.

K= Coeficientes discriminantes.

D= Términos independientes.

$$\begin{bmatrix} 0.019 & 0.236 & -0.806 \\ 0.236 & 44 & -8 \\ -0.806 & -8 & 102 \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.4 \\ 6 \end{bmatrix}$$

Paso 2. Obtener el determinante de A aplicando la regla de Sarrus para comprobar si el sistema tiene solución, es decir si |A|≠0:

$$|A| = \begin{vmatrix} 0.019 & 0.236 & -0.806 \\ 0.236 & 44 & -8 \\ -0.806 & -8 & 102 \\ 0.019 & 0.236 & -0.806 \\ 0.236 & 44 & -8 \end{vmatrix} =$$ (0.019*44*102)+ (0.236*-8*-0.806)+ (-0.806*0.236*-8)- (-0.806*44*-0.806)- (0.019*-8*-8)- (0.236*0.236*102)=52.83 ≅**52.41**

Como 52.41 es ≠ de 0 el sistema tiene solución.

Paso 3. Obtener la matriz transpuesta de A, que es:

$$A' = \begin{bmatrix} 0.019 & 0.236 & -0.806 \\ 0.236 & 44 & -8 \\ -0.806 & -8 & 102 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A', es decir formamos $\alpha A'$

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 44 & -8 \\ -8 & 102 \end{vmatrix} & -\begin{vmatrix} 0.236 & -8 \\ -0.806 & 102 \end{vmatrix} & +\begin{vmatrix} 0.236 & 44 \\ -0.806 & -8 \end{vmatrix} \\ -\begin{vmatrix} 0.236 & -0.806 \\ -8 & 102 \end{vmatrix} & +\begin{vmatrix} 0.019 & -0.806 \\ -0.806 & 102 \end{vmatrix} & -\begin{vmatrix} 0.019 & 0.236 \\ -0.806 & -8 \end{vmatrix} \\ +\begin{vmatrix} 0.236 & -0.806 \\ 44 & -8 \end{vmatrix} & -\begin{vmatrix} 0.019 & -0.806 \\ 0.236 & -8 \end{vmatrix} & +\begin{vmatrix} 0.019 & 0.236 \\ 0.236 & 44 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(4488-64) & -(24.072-6.448) & +(-1.888+35.464) \\ -(24.072-6.448) & +(1.928-0.650) & -(-0.151+0.190) \\ +((-1.888+35.464) & -(-0.151+0.190) & +(0.832-0.056) \end{bmatrix} =$$

$$= \begin{bmatrix} 4424 & -17.624 & 33.576 \\ -17.624 & 1.279 & -0.039 \\ 33.576 & -0.039 & 0.776 \end{bmatrix}$$

Paso 5. Calculamos la matriz inversa de A:

$$A^{-1} = \frac{1}{|A|} \propto A^{1} = \frac{1}{52.41}\begin{bmatrix} 4424 & -17.624 & 33.576 \\ -17.624 & 1.279 & -0.039 \\ 33.576 & -0.039 & 0.776 \end{bmatrix} = \begin{bmatrix} \frac{4424}{52.41} & \frac{-17.624}{52.41} & \frac{33.576}{52.41} \\ \frac{-17.624}{52.41} & \frac{1.279}{52.41} & \frac{-0.039}{16{,}484.96} \\ \frac{33.576}{52.41} & \frac{-0.039}{52.41} & \frac{0.776}{52.41} \end{bmatrix}$$

Nota: Es conveniente dejar el resultado en forma de fracción para no perder precisión

Paso 6. Llevamos el valor anterior a la expresión $K = A^{-1} * D$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} \frac{4424}{52.41} & \frac{-17.624}{52.41} & \frac{33.576}{52.41} \\ \frac{-17.624}{52.41} & \frac{1.279}{52.41} & \frac{-0.039}{16{,}484.96} \\ \frac{33.576}{52.41} & \frac{-0.039}{52.41} & \frac{0.776}{52.41} \end{bmatrix}\begin{bmatrix} 0.14 \\ 0.4 \\ 6 \end{bmatrix} = \begin{bmatrix} 15.52700 \\ -0.04178 \\ 0.17824 \end{bmatrix}$$

Paso 7. Obtenemos el modelo discriminante:

$$\boldsymbol{Z = 15.52700K_1 - 0.04178K_2 + 0.17824K_3}$$

Resolución del inciso b)

A continuación se prueba si el modelo es bueno para discriminar o clasificar a las personas que están de acuerdo con las que no lo están mediante los siguientes pasos:

Paso 1. Plantear el juego de hipótesis:

H_0. El modelo discriminante no es bueno para discriminar o clasificar a las personas que están o no de acuerdo con las características de las laptops analizadas.

H_1. El modelo discriminante es bueno.

Paso 2. Establecer el estadístico de prueba:

$$F\ calc = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2)D^2}{[(n_1 + n_2 - 2)p * (n_1 + n_2)]}$$

$$D^2 = (n_1 + n_2 - 2)\left[\sum K_1(\bar{X}_1|Grupo\ 1 - \bar{X}_1|Grupo2\right]$$

$$D^2 = (5 + 5 - 2)[(15.52700)(0.14) - (0.04178)(1) + (0.17824)(6)] = 25.81$$

$$F\ calc = \frac{[(5 + 5 - 2 - 1)(5 * 5)(25.81)]}{[(5 + 5 - 2)2](5 + 5)} = \frac{7x25x25.81}{16x10} = \frac{1856.75}{160} = 28.23$$

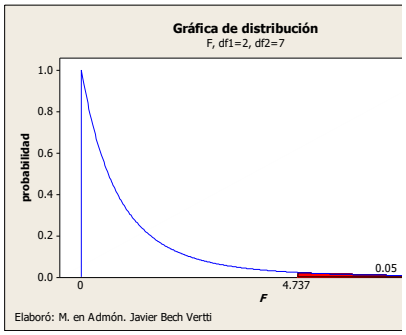
Paso 3. Establecer la región de Rechazo:

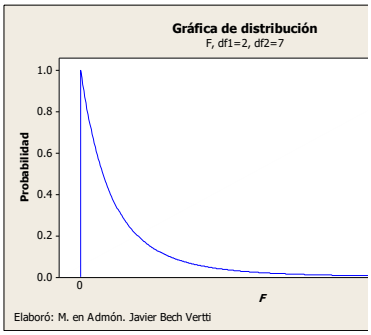
$$Fcrit = [\propto_1 p_1(n_1 + n_2 - p - 1)]$$

$$Fcrit = 0.05, 2, 7$$

$$f_{0.05} = 4.74$$

$$f_{0.01} = 9.55$$





Paso 4. Definir la regla de decisión:

Se rechaza H_0 si $F_{calc.} \geq 4.74$

Paso 5. Conclusiones:

-Estadística: Como 28.23>4.74 y > 9.55 ∴ la prueba es (AS) y se rechaza H_0 con mucha fuerza.

-Administrativa: Existe evidencia suficiente para decir que estadísticamente el modelo es bueno para clasificar o discriminar la magnitud del riesgo.

Resolución del inciso c)

Cuando se ha concluido que el modelo discriminante es bueno para hacer clasificaciones, se debe proceder a validar el mismo. Esta validación consiste en determinar el grado de confiabilidad que se tendrá al hacer pronósticos con el modelo.

Paso 1. Se obtiene la Z promedio para cada grupo:

$$Z = 15.527K_1 - 0.04178K_2 + 0.17824K_3$$

$$\bar{Z}_1 = 15.527(0.41) - 0.04178(5) + 0.17824(42) = 13.66$$

$$\bar{Z}_2 = 15.527(0.27) - 0.04178(4) + 0.17824(36) = 10.44$$

Paso 2. Se determina la Z crítica o de corte:

$$Z^*_{critica} = \frac{n_2\bar{Z}_1 + n_1\bar{Z}_2}{n_1 + n_2} = \frac{5(13.66) + 5(10.44)}{5 + 5} = 12.05$$

*Esta fórmula puede usarse ya sea si $n_1=n_2$ o si $n_1 \neq n_2$

Paso 3. Se obtiene la Z para cada observación:

$Z_1 = 15.527(0.376) - 0.04178(7) + 0.17824(42) = 13.03 > 12.05$

$Z_2 = 15.527(0.348) - 0.04178(6) + 0.17824(47) = 13.53 > 12.05$

$Z_3 = 15.527(0.452) - 0.04178(4) + 0.17824(41) = 14.16 > 12.05$

$Z_4 = 15.527(0.486) - 0.04178(5) + 0.17824(37) = 13.93 > 12.05$

$Z_5 = 15.527(0.388) - 0.04178(1) + 0.17824(43) = 13.65 > 12.05$

$Z_6 = 15.527(0.286) - 0.04178(4) + 0.17824(33) = 10.16 < 12.05$

$Z_7 = .15.527(0.292) - 0.04178(7) + 0.17824(32) = 9.95 < 12.05$

$Z_8 = .15.527(0.270) - 0.04178(3) + 0.17824(40) = 11.20 < 12.05$

$Z_9 = 15.527(0.204) - 0.04178(1) + 0.17824(36) = 9.54 < 12.05$

$Z_{10} = 15.527(0.298) - 0.04178(6) + 0.17824(39) = 11.33 < 12.05$

Paso 4. Se verifica si en las Z observadas la clasificación es correcta o incorrecta

Sujeto No.	**Grupos**	**Duración (X_1)**	**Funcionamiento (X_2)**	**Diseño (X_3)**	**Z obs.**	**Clas.** ✓o **X**
1	Riesgo bueno	0.376	7	42	13.03 > 12.05	✓
2	Riesgo bueno	0.348	6	47	13.53 > 12.05	✓
3	Riesgo bueno	0.452	4	41	14.16 > 12.05	✓
4	Riesgo bueno	0.486	5	37	13.93 > 12.05	✓
5	Riesgo bueno	0.388	1	43	13.65 > 12.05	✓
	Media Gpo. 1	0.410	5	42		
6	Riesgo dudoso	0.286	4	33	10.16 < 12.05	✓
7	Riesgo dudoso	0.292	7	32	9.95 < 12.05	✓
8	Riesgo dudoso	0.270	3	40	11.20 < 12.05	✓
9	Riesgo dudoso	0.204	1	36	9.54 < 12.05	✓
10	Riesgo dudoso	0.298	6	39	11.33 < 12.05	✓
	Media Gpo. 2	0.270	4	36		

Paso 5. Del paso anterior se obtiene el porcentaje de clasificaciones correctas (PCC):

$$PCC = \frac{\#\checkmark Grupo\ 1 + \#\checkmark Gupo\ 2}{\#\ de\ observaciones} * 100$$

$$PCC = \frac{5+5}{10} * 100 = 100\%$$

Paso 6. Se obtiene el porcentaje mínimo de clasificaciones correctas:

$$C = (p_i)^2 + (1-p_1)^2$$

$$C = (\frac{5}{10})^2 + (1-\frac{5}{10})^2$$

$$C = 0.25 + 0.25 = 0.50x100 = 50\%$$

Paso 7. Se compara el porcentaje de clasificaciones correctas contra el porcentaje mínimo de clasificaciones correctas:

Como 100% > 50% se puede validar el modelo para clasificar la magnitud del riesgo en la selección preliminar de solicitudes de crédito.

Resolución del inciso d)

Para interpretar los resultados obtenidos anteriormente, se evalúa la importancia relativa que posee cada una de las variables independientes para discriminar. El método de los pesos discriminantes estandarizados consiste en estandarizar los coeficientes del modelo discriminante con el fin de que puedan ser comparados. Para jerarquizar las variables se considera solamente la magnitud del peso, no su signo, es decir el valor absoluto.

La fórmula para estandarizar los coeficientes discriminantes es:

$$K_n' = (K_n)\left[\frac{\sum X_n'^2}{(n_1+n_2-2)}\right]^{0.05}$$

En el problema presentado, los coeficientes o constantes estandarizadas son:

1. Ingreso:

$$K_1 = (15.527)\left[\frac{0.019}{(5+5-2)}\right]^{0.5} = 0.75478$$

2. Tarjetas:

$$K_2 = (-0.04178)\left[\frac{44}{(5+5-2)}\right]^{0.5} = -0.09799$$

3. Edad:

$$K_3 = (0.17824)\left[\frac{102}{(5+5-2)}\right]^{0.5} = 0.63644$$

-Interpretación: De acuerdo a los pesos discriminantes estandarizados **en valores absolutos** (sin considerar el signo), podemos concluir que la variable que más discrimina o diferencia a un grupo de otro para evaluar el riesgo en la selección preliminar de solicitudes de crédito es la variable X_1 (Ingreso) con un peso de .75478 por su mayor *K* estandarizada y la menos importante es la variable X_2 (Tarjetas) con un peso de 0.09799 **en valor absoluto**.

-Estrategias a seguir: Lo que se pretendería como estrategia es tratar de actuar sobre la variable dependiente (grupos), manipulando las variables independientes (Ingreso, Tarjetas y Edad). Lo que se quisiera lograr es aumentar los valores de Z, para hacer que las personas del grupo ***Riesgo dudoso*** se pasen al grupo ***Riesgo bueno.*** Para nuestro problema, tenemos que si X_1 (Ingreso) aumenta, subirá a su vez el valor de Z, pues su signo es positivo. Asímismo tenemos que si X_3 (Edad) aumenta, subirá el valor de Z pues su signo es también positivo, en cambio para X_2 (Tarjetas), tenemos que va a afectar en forma inversa al valor de Z, por su signo negativo.

De las tres variables, la primera que se debe cuidar es X_1 (Ingreso), ya que tiene una mayor importancia para discriminar entre los grupos, después podría pensarse en una estrategia en cuanto a la variable X_3 (Edad), en un segundo plano, para facilitar el trámite de la solicitud y por último una estrategia para la variable X_2 (Tarjetas), en un tercer plano, para disminuir el Riesgo.

Resolución del inciso e)

$X_1 = 0.34$ (en millones de pesos)

$X_2 = 3$

$X_3 = 50$

Sustituimos los valores anteriores en el modelo discriminante:

$Z = 15.527(.34) - 0.04178(3) + 0.17824(50) = 14.065$

Comparamos el valor anterior con la Z crítica:

Como 1.68 > 1.63 ∴ se le clasificaría en el grupo 1 de **Riesgo bueno**

3.1.1.1.3 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 3.1.1.1.3 Análisis Discriminante

3.1.1.1.3. Se está interesado en el posible uso del análisis discriminante para conocer si en una cierta Universidad no existen profesores intermedios, es decir , o son buenos o son malos; por lo que se analizaron 10 profesores con base en su capacidad docente (X_1) (metodología de enseñanza, exámenes representativos, calificaciones justas), experiencia en el área de la materia (X_2) y conocimiento en la misma (X_3); todo esto en una escala de 1-5 donde el uno significa malo y cinco bueno, para ver si realmente se segmentaban de esta forma. Los resultados obtenidos se muestran a continuación:

Profesor	Opinión	Cap.docente	Experiencia	Conocimiento
1	Bueno	3	4	3
2	Bueno	4	5	5
3	Bueno	4	5	2
4	Bueno	5	4	5
5	Bueno	5	4	2
	Media del grupo 1	**4.2**	**4.4**	**3.4**
6	Malo	1	2	1
7	Malo	1	1	1
8	Malo	2	1	2
9	Malo	1	1	1
10	Malo	3	2	1
	Media del grupo 2	**1.6**	**1.4**	**1.2**

a) Obtener el modelo discriminante
b) Probar si el modelo es bueno para segmentar
c) Valide el modelo con los mismos datos de derivación.
d) Interprete y comente los resultados obtenidos mediante el método de los pesos discriminantes estandarizados.
e) ¿En qué grupo caería un profesor que hubiera sido evaluado con con una Capacidad docente de 4, una experiencia de 3 y un conocimiento de 3?

Resolución del inciso a)

La ecuación para derivar el modelo discriminante tendrá la siguiente forma:

$$Z = K_1X_1 + K_2X_2 + K_3X_3$$

Para la derivación del modelo, lo que se busca es encontrar los valores de los pesos de las variables para discriminar, es decir los valores de las *K,* para ello se deben resolver las siguientes ecuaciones normales:

$$\sum X_1'^2 K_1 + \sum X_1'X_2'K_2 + \sum X_1' X_3'K_3 = \bar{X}_1\,|Grupo\ 1 - \bar{X}_1|Grupo\ 2$$

$$\sum X_1'X_2'K_1 + \sum X_2'^2 K_2 + \sum X_2' X_3'K_3 = \bar{X}_2\,|Grupo\ 1 - \bar{X}_2|Grupo\ 2$$

$$\sum X_1'X_3'K_1 + \sum X_2'X_3'K_2 + \sum X_3'^2\, K_3 = \bar{X}_3\,|Grupo\ 1 - \bar{X}_3|Grupo\ 2$$

Donde:

$$\sum X_1'^2 = \sum(\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_2'^2 = \sum(\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_3'^2 = \sum(\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_1'X_2' = \sum(\sum X_1X_2 - n\bar{X}_1\,\bar{X}_2)$$

$$\sum X_1' X_3' = \sum (\sum X_1 X_3 - n\bar{X}_1\,\bar{X}_3)$$

$$\sum X_2' X_3' = \sum (\sum X_2 X_3 - n\bar{X}_2\,\bar{X}_3)$$

Se sustituyen los valores correspondientes, obteniendo lo siguiente:

$$\sum X_1'^2 = \sum (\sum X_1^2 - n\bar{X}_1^2)$$

$$\sum X_1'^2 = [(3^2 + 4^2 + 4^2 + 5^2 + 5^2) - 5(4.2)^2] + [(1^2 + 1^2 + 2^2 + 1^2 + 3^2) - 5(1.6)^2] = 6$$

$$\sum X_2'^2 = \sum (\sum X_2^2 - n\bar{X}_2^2)$$

$$\sum X_2'^2 = [(4^2 + 5^2 + 5 + 4^2 + 4^2) - 5(4.4)^2] + [(2^2 + 1^2 + 1^2 + 1^2 + 2^2) - 5(1.4)^2] = 2.4$$

$$\sum X_3'^2 = \sum (\sum X_3^2 - n\bar{X}_3^2)$$

$$\sum X_3'^2 = [(3^2 + 5^2 + 2^2 + 5^2 + 2^2) - 5(3.4)^2] + [(1^2 + 1^2 + 2^2 + 1^2 + 1^2) - 5(1.2)^2] = 10$$

$$\sum X_1'\,X_2' = \sum (\sum X_1\,X_2 - n\bar{X}_1\bar{X}_2)$$

$$= [(3x4 + 4x5 + 4x5 + 5x4 + 5x4) - 5(4.2x4.4)] + [(1x2 + 1x1 + 2x1 + 1x1 + 3x2) - 5(1.6x1.4)] = 0.4$$

$$\sum X_1'\,X_3' = \sum (\sum X_1\,X_3 - n\bar{X}_1\bar{X}_3)$$

$$= [(3x3 + 4x5 + 4x2 + 5x5 + 5x2) - 5(4.2x3.4)] + [(1x1 + 1x1 + 2x2 + 1x1 + 3x1) - 5(1.6x1.2)] = 1$$

$$\sum X_2'\,X_3' = \sum (\sum X_2\,X_3 - n\bar{X}_2\bar{X}_3)$$

$$= [(4x3 + 5x5 + 5x2 + 4x5 + 4x2) - 5(4.4x3.4)] + [(2x1 + 1x1 + 1x2 + 1x1 + 2x1) - 5(1.4x1.2)] = -0.2$$

De lo anterior se obtienen tres ecuaciones, normales:

$6K_1 + 0.4K_2 + 1K_3 = 2.6$

$0.4K_1 + 2.4K_2 - 0.2K_3 = 3$

$1K_1 - 0.2K_2 + 10K_3 = 2.2$

A partir de estas ecuaciones se obtienen los valores de *K* de cada variable. En este caso se utilizará la matriz inversa para la resolución del sistema lineal siguiendo los siguientes pasos:

Paso 1. Expresar el sistema como A *K=D

Donde:

A= Matriz de coeficientes.

K= Coeficientes discriminantes.

D= Términos independientes.

$$\begin{bmatrix} 6 & 0.4 & 1 \\ 0.4 & 2.4 & -0.2 \\ 1 & -0.2 & 10 \end{bmatrix} \begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} 2.6 \\ 3 \\ 2.2 \end{bmatrix}$$

Paso 2. Obtener el determinante de A aplicando la regla de Sarrus para comprobar si el sistema tiene solución, es decir si |A|≠0:

$$|A| = \begin{vmatrix} 6 & 0.4 & 1 \\ 0.4 & 2.4 & -0.2 \\ 1 & -0.2 & 10 \\ 6 & 0.4 & 1 \\ 0.4 & 2.4 & -0.2 \end{vmatrix} =$$ (6*2.4*10)+ (0.4*-0.2*1)+ (1*0.4*-0.2)- (1*2.4*1)- (6*-0.2*-0.2)- (0.4*0.4*10)=139.6

Como 1144 es ≠ de 0 el sistema tiene solución.

Paso 3. Obtener la matriz transpuesta de A, que es:

$$A' = \begin{bmatrix} 6 & 0.4 & 1 \\ 0.4 & 2.4 & -0.2 \\ 1 & -0.2 & 10 \end{bmatrix}$$

Paso 4. Formar la matriz de cofactores de A', es decir formamos $\alpha A'$

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 2.4 & -0.2 \\ -0.2 & 10 \end{vmatrix} & -\begin{vmatrix} 0.4 & -0.2 \\ 1 & 10 \end{vmatrix} & +\begin{vmatrix} 0.4 & 2.4 \\ 1 & -0.2 \end{vmatrix} \\ -\begin{vmatrix} 0.4 & 1 \\ -0.2 & 10 \end{vmatrix} & +\begin{vmatrix} 6 & 1 \\ 1 & 10 \end{vmatrix} & -\begin{vmatrix} 6 & 0.4 \\ 1 & -0.2 \end{vmatrix} \\ +\begin{vmatrix} 0.4 & 1 \\ 2.4 & -0.2 \end{vmatrix} & -\begin{vmatrix} 6 & 1 \\ 0.4 & -0.2 \end{vmatrix} & +\begin{vmatrix} 6 & 0.4 \\ 0.4 & 2.4 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(24 - 0.04) & -(4 + 0.2) & +(-0.08 - 2.4) \\ -(4 + 0.2) & +(60 - 1) & -(-1.2 - 0.4) \\ +((-0.08 - 2.4 & -(-1.2 - 0.4) & +(14.4 - 0.16) \end{bmatrix} =$$

$$= \begin{bmatrix} 23.96 & -4.2 & -2.48 \\ -4.2 & 59 & 1.6 \\ -2.48 & 1.6 & 14.24 \end{bmatrix}$$

Paso 5. Calculamos la matriz inversa de A:

$$A^{-1} = \frac{1}{|A|} \propto A^1 = \frac{1}{139.6}\begin{bmatrix} 23.96 & -4.2 & -2.48 \\ -4.2 & 59 & 1.6 \\ -2.48 & 1.6 & 14.24 \end{bmatrix} = \begin{bmatrix} \frac{23.96}{139.6} & \frac{-4.2}{139.6} & \frac{-2.48}{139.6} \\ \frac{-4.2}{139.6} & \frac{59}{139.6} & \frac{1.6}{139.6} \\ \frac{-2.48}{139.6} & \frac{1.6}{139.6} & \frac{14.24}{139.6} \end{bmatrix}$$

Nota: Es conveniente dejar el resultado en forma de fracción para no perder precisión

Paso 6. Llevamos el valor anterior a la expresión $K = A^{-1} * D$

$$\begin{bmatrix} K_1 \\ K_2 \\ K_3 \end{bmatrix} = \begin{bmatrix} \frac{23.96}{139.6} & \frac{-4.2}{139.6} & \frac{-2.48}{139.6} \\ \frac{-4.2}{139.6} & \frac{59}{139.6} & \frac{1.6}{139.6} \\ \frac{-2.48}{139.6} & \frac{1.6}{139.6} & \frac{14.24}{139.6} \end{bmatrix}\begin{bmatrix} 2.6 \\ 3 \\ 2.2 \end{bmatrix} = \begin{bmatrix} 0.31691 \\ 1.21490 \\ 0.21261 \end{bmatrix}$$

Paso 7. Obtenemos el modelo discriminante:

$$\boldsymbol{Z = 0.31691K_1 + 1.21490K_2 + 0.21261K_3}$$

Resolución del inciso b)

A continuación se prueba si el modelo es bueno para discriminar o clasificar a las personas que están de acuerdo con las que no lo están mediante los siguientes pasos:

Paso 1. Plantear el juego de hipótesis:

H_0. El modelo discriminante no es bueno para discriminar o clasificar a las personas que están o no de acuerdo con las características de las laptops analizadas.

H_1. El modelo discriminante es bueno.

Paso 2. Establecer el estadístico de prueba:

$$F\ calc = \frac{(n_1 + n_2 - p - 1)(n_1 * n_2)D^2}{[(n_1 + n_2 - 2)p * (n_1 + n_2)]}$$

$$D^2 = (n_1 + n_2 - 2)\left[\sum K_1(\bar{X}_1|Grupo\ 1 - \bar{X}_1|Grupo2\right]$$

$$D^2 = (5 + 5 - 2)[(0.31691)(2.6) + (1.2149)(3) + (0.21261)(2.2)] = 39.49$$

$$F\ calc = \frac{[(5 + 5 - 2 - 1)(5 * 5)(39.49)]}{[(5 + 5 - 2)2](5 + 5)} = \frac{7x25x39.49}{16x10} = \frac{6910.75}{160} = 43.19$$

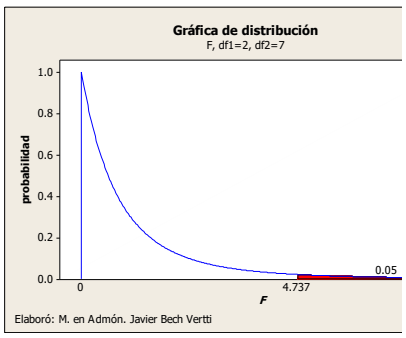
Paso 3. Establecer la región de Rechazo:

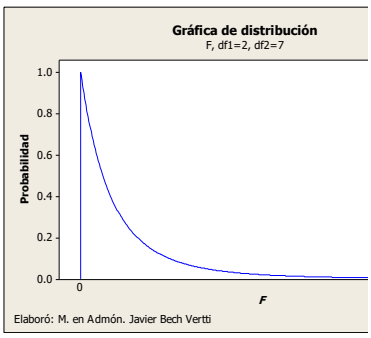
$Fcrit = [\propto_1 p_1(n_1 + n_2 - p - 1)]$

$Fcrit = 0.05, 2, 7$

$f_{0.05} = 4.74$

$f_{0.01} = 9.55$





Paso 4. Definir la regla de decisión:

Se rechaza H_0 si $F_{calc.} \geq 4.74$

Paso 5. Conclusiones:

-Estadística: Como 43.19>4.74 y > 9.55 ∴ la prueba es (AS) y se rechaza H_0 con mucha fuerza.

-Administrativa: Existe evidencia suficiente para decir que estadísticamente el modelo es bueno para clasificar o discriminar la opinión sobre la calidad de los profesores de cierta Universidad.

Resolución del inciso c)

Cuando se ha concluido que el modelo discriminante es bueno para hacer clasificaciones, se debe proceder a validar el mismo. Esta validación consiste en determinar el grado de confiabilidad que se tendrá al hacer pronósticos con el modelo.

Paso 1. Se obtiene la Z promedio para cada grupo:

$Z = 0.31691K_1 + 1.21490K_2 + 0.21261K_3$

$\bar{Z}_1 = 0.31691(4.2) + 1.21490(4.4) + 0.21261(3.4) = 7.40$

$\bar{Z}_2 = 0.31691(1.6) + 1.21490(1.4) + 0.21261(1.2) = 2.46$

Paso 2. Se determina la Z crítica o de corte:

$$Z^*_{critica} = \frac{n_2\bar{Z}_1 + n_1\bar{Z}_2}{n_1 + n_2} = \frac{5(7.40) + 5(2.46)}{5 + 5} = 4.93$$

*Esta fórmula puede usarse ya sea si $n_1=n_2$ o si $n_1 \neq n_2$

Paso 3. Se obtiene la Z para cada observación:

$Z_1 = 0.31691(3) + 1.21490(4) + 0.21261(3) = 6.45 > 4.93$

$Z_2 = 0.31691(4) + 1.21490(5) + 0.21261(5) = 8.41 > 4.93$

$Z_3 = 0.31691(4) + 1.21490(5) + 0.21261(2) = 7.77 > 4.93$

$Z_4 = 0.31691(5) + 1.21490(4) + 0.21261(5) = 7.51 > 4.93$

$Z_5 = 0.31691(5) + 1.21490(4) + 0.21261(2) = 6.87 > 4.93$

$Z_6 = 0.31691(1) + 1.21490(2) + 0.21261(1) = 2.96 < 4.93$

$Z_7 = 0.31691(1) + 1.21490(1) + 0.21261(1) = 1.74 < 4.93$

$Z_8 = 0.31691(2) + 1.21490(1) + 0.21261(2) = 2.27 < 4.93$

$Z_9 = 0.31691(1) + 1.21490(1) + 0.21261(1) = 1.74 < 4.93$

$Z_{10} = 0.31691(3) + 1.21490(2) + 0.21261(1) = 3.59 < 4.93$

Paso 4. Se verifica si en las Z observadas la clasificación es correcta o incorrecta

Pro-fesor No.	**Opinión**	**Capacidad docente (X_1)**	**Experiencia (X_2)**	**Conocimiento (X_3)**	**Z obs.**	**Clas.** ✓o **X**
1	Bueno	3	4	3	8.41 > 4.93	✓
2	Bueno	4	5	5	8.41 > 4.93	✓
3	Bueno	4	5	2	7.77 > 4.93	✓
4	Bueno	5	4	5	7.51 > 4.93	✓
5	Bueno	5	4	2	6.87 > 4.93	✓
	Media Gpo. 1	4.2	4.4	3.4		
6	Malo	1	2	1	2.96 < 4.93	✓
7	Malo	1	1	1	1.74 < 4.93	✓
8	Malo	2	1	2	2.27 < 4.93	✓
9	Malo	1	1	1	1.74 < 4.93	✓
10	Malo	3	2	1	3.59 < 4.93	✓
	Media Gpo. 2	1.6	1.4	1.2		

Paso 5. Del paso anterior se obtiene el porcentaje de clasificaciones correctas (PCC):

$$PCC = \frac{\#\checkmark Grupo\ 1 + \#\checkmark Gupo\ 2}{\#\ de\ observaciones} * 100$$

$$PCC = \frac{5+5}{10} * 100 = 100\%$$

Paso 6. Se obtiene el porcentaje mínimo de clasificaciones correctas:

$C = (p_i)^2 + (1 - p_1)^2$

$C = (\frac{5}{10})^2 + (1 - \frac{5}{10})^2$

$C = 0.25 + 0.25 = 0.50x100 = 50\%$

Paso 7. Se compara el porcentaje de clasificaciones correctas contra el porcentaje mínimo de clasificaciones correctas:

Como 100% > 50% se puede validar el modelo para clasificar a los que opinan que el profesor es **Bueno** de los que opinan que el profesor es **Malo**.

Resolución del inciso d)

Para interpretar los resultados obtenidos anteriormente, se evalúa la importancia relativa que posee cada una de las variables independientes para discriminar. El método de los pesos discriminantes estandarizados consiste en estandarizar los coeficientes del modelo discriminante con el fin de que puedan ser comparados. Para jerarquizar las variables se considera solamente la magnitud del peso, no su signo, es decir el valor absoluto.

La fórmula para estandarizar los coeficientes discriminantes es:

$$K_n' = (K_n)\left[\frac{\sum X_n'^2}{(n_1 + n_2 - 2)}\right]^{0.05}$$

En el problema presentado, los coeficientes o constantes estandarizadas son:

1. Capacidad docente:
   $K_1 = (0.31691)\left[\frac{6}{(5+5-2)}\right]^{0.5} = 0.27445$
2. Experiencia:
   $K_2 = (1.21490)\left[\frac{2.4}{(5+5-2)}\right]^{0.5} = 0.66543$
3. Conocimiento :
   $K_3 = (0.21261)\left[\frac{10}{(5+5-2)}\right]^{0.5} = 0.23771$

-Interpretación: De acuerdo a los pesos discriminantes estandarizados **en valores absolutos** (sin considerar el signo), podemos concluir que la variable que más discrimina o diferencia a un grupo de otro de profesores es la variable X_2 (Experiencia) con un peso de .66543 por su mayor *K* estandarizada y la menos importante al parecer, es la variable X_3 (Conocimiento) con un peso de 0.23771 en valor absoluto.

-Estrategias a seguir: Lo que se pretendería como estrategia es tratar de actuar sobre la variable dependiente (grupos), manipulando las variables independientes (Capacidad, Experiencia y Conocimiento). Lo que se quisiera lograr es aumentar los valores de Z, para hacer que los profesores del grupo ***Malo*** se pasen al grupo

Bueno. Para nuestro problema, tenemos que si X_2 (Experiencia) aumenta, subirá a su vez el valor de Z, pues su signo es positivo. Tenemos que si X_1 (Capacidad docente) aumenta, subirá el valor de Z pues su signo es positivo y para X_3 (Conocimiento), tenemos que va a afectar también en forma directa al valor de Z, por su signo positivo.

De las tres variables, la primera que se debe cuidar es X_2 (Experiencia), ya que tiene una mayor importancia para discriminar entre los grupos, con más conocimiento en el área de la materia, después podría pensarse en una estrategia en cuanto a la variable X_1 (Capacidad docente), en un segundo plano, con más cursos de metodología de la enseñanza, mayor imparcialidad en las calificaciones y por último una estrategia para la variable X_3 (Conocimiento), en un tercer plano, aumentando el conocimiento en la materia a impartir con cursos de especialidad.

Resolución del inciso e)

$X_1 = 4$

$X_2 = 3$

$X_3 = 3$

Sustituimos los valores anteriores en el modelo discriminante:

$Z = 0.31691(4) + 1.21490(3) + 0.21261(3) = 5.55$

Comparamos el valor anterior con la Z crítica:

Como 5.55> 4.93 ∴ se le clasificaría en el grupo 1 de **Bueno**

3.2.1.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.1.1.1. Vector de Medias.**

3.2.1.1.1 El Director de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia de la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares) , el tamaño del lote (en pies cuadrados) y el número de recámaras. Para este experimento se manejo uno de los niveles de la ubicación geográfica: **Norte**. Se tomaron lecturas en 7 casas unifamiliares y los resultados fueron los siguientes

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada denominada **"BIENES1"** donde X_1= Valor de avalúo en la zona **Norte** , X_2= Tamaño del lote en la zona **Norte** y X_3= Número de recámaras en la zona **Norte**:

$$X = \begin{bmatrix} 174.5 & 6.00 & 3 \\ 275.0 & 6.00 & 3 \\ 185.0 & 6.00 & 4 \\ 220.0 & 6.00 & 3 \\ 218.0 & 7.40 & 3 \\ 190.0 & 3.55 & 4 \\ 218.0 & 6.00 & 3 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 190.8 \\ 7.50 \\ 3.60 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de hipótesis

$$H_0 = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} 190.8 \\ 7.50 \\ 3.60 \end{bmatrix}$$

$H_1 = Al\ menos\ una\ M\ es\ \neq.$

2. Estadístico de prueba

$$Fcalc = \frac{n-p}{p(n-1)} T^2$$

Donde: $T^2 = n(M - M_0)'[S^2]^{-1}(\widehat{M} - \widehat{M}_0)$

Donde: $S^2 = \frac{1}{n-1}\, SCPC = \frac{1}{n-1}\, D'D$ (Matriz de varianzas y covarianzas)=

$$[s^2]^2 = \frac{1}{|S|} \propto [S^2]'$$

$$D = \begin{bmatrix} 174.5 & 6.00 & 3 \\ 275.0 & 6.00 & 3 \\ 185.0 & 6.00 & 4 \\ 220.0 & 6.00 & 3 \\ 218.0 & 7.40 & 3 \\ 190.0 & 3.55 & 4 \\ 218.0 & 6.00 & 3 \end{bmatrix} - \begin{bmatrix} 211.5 & 5.85 & 3.29 \\ 211.5 & 5.85 & 3.29 \\ 211.5 & 5.85 & 3.29 \\ 211.5 & 5.85 & 3.29 \\ 211.5 & 5.85 & 3.29 \\ 211.5 & 5.85 & 3.29 \\ 211.5 & 5.85 & 3.29 \end{bmatrix} = \begin{bmatrix} -37.00 & 0.15 & -0.29 \\ 63.50 & 0.15 & -0.29 \\ -26.50 & 0.15 & 0.71 \\ 8.50 & 0.15 & -0.29 \\ 6.50 & 1.55 & -0.29 \\ -21.50 & -2.30 & 0.71 \\ 6.50 & 0.15 & -0.29 \end{bmatrix}$$

$$D'D = \begin{bmatrix} -37.00 & 63.50 & -26.50 & 8.50 & 6.50 & -21.50 & 6.50 \\ 0.15 & 0.15 & 0.15 & 0.15 & 1.55 & -2.30 & 0.15 \\ -0.29 & -0.29 & 0.71 & -0.29 & -0.29 & 0.71 & -0.29 \end{bmatrix} \begin{bmatrix} -37.00 & 0.15 & -0.29 \\ 63.50 & 0.15 & -0.29 \\ -26.50 & 0.15 & 0.71 \\ 8.50 & 0.15 & -0.29 \\ 6.50 & 1.55 & -0.29 \\ -21.50 & -2.30 & 0.71 \\ 6.50 & 0.15 & -0.29 \end{bmatrix}$$

$$D'D = \begin{bmatrix} 6722.5 & 61775 & -48 \\ 61.775 & 7.805 & -2.15 \\ -48 & -2.15 & 1.42857 \end{bmatrix}$$

$$S^2 = \frac{1}{n-1}\, D'D = \frac{1}{7-1} D'D = \frac{1}{6} \quad \begin{bmatrix} 6722.5 & 61775 & -48 \\ 61.775 & 7.805 & -2.15 \\ -48 & -2.15 & 1.42857 \end{bmatrix}$$

$$S^2 = \frac{1}{6}\begin{bmatrix} 6722.5 & 61775 & -48 \\ 61.775 & 7.805 & -2.15 \\ -48 & -2.15 & 1.42857 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} \frac{6722.5}{6} & \frac{61.775}{6} & \frac{-48}{6} \\ \frac{61.775}{6} & \frac{7.805}{6} & \frac{-2.15}{6} \\ \frac{-48}{6} & \frac{-2.15}{6} & \frac{1.42857}{6} \end{bmatrix}$$

Utilizando la regla de Sarrus

$$|S^2| = \begin{bmatrix} \frac{6722.5}{6} & \frac{61.775}{6} & \frac{-48}{6} \\ \frac{61.775}{6} & \frac{7.805}{6} & \frac{-2.15}{6} \\ \frac{-48}{6} & \frac{-2.15}{6} & \frac{1.42857}{6} \\ \frac{6722.5}{6} & \frac{61.775}{6} & \frac{-48}{6} \\ \frac{61.775}{6} & \frac{7.805}{6} & \frac{-2.15}{6} \end{bmatrix} = 153.69$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 1.30083 & -0.35833 \\ -0.35833 & 0.23810 \end{vmatrix} & -\begin{vmatrix} 10.29583 & -0.35833 \\ -8 & 0.23810 \end{vmatrix} & +\begin{vmatrix} 10.29583 & 1.30083 \\ -8 & -0.35833 \end{vmatrix} \\ -\begin{vmatrix} 10.29583 & -8 \\ -0.35833 & 0.23810 \end{vmatrix} & +\begin{vmatrix} 1120.41667 & -8 \\ -8 & 0.23810 \end{vmatrix} & -\begin{vmatrix} 1120.41667 & 10.29583 \\ -8 & -0.35833 \end{vmatrix} \\ +\begin{vmatrix} 10.29583 & -8 \\ 1.30083 & -0.35833 \end{vmatrix} & -\begin{vmatrix} 1120.41667 & -8 \\ 10.29583 & -0.35833 \end{vmatrix} & +\begin{vmatrix} 1120.41667 & 10.29583 \\ 10.29583 & 1.30083 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(0.30972 - 0.12840) & -(2.45139 - 2.86667) & +(-3.68934 + 10.40667) \\ -(2.45139 - 2.86667) & +(266.76587 - 64) & -(-401.48264 + 82.36667) \\ +((-3.68934 + 10.40667) & -(-401.48264 + 82.36667) & +(1457.47535 - 106.00418) \end{bmatrix} =$$

$$= \begin{bmatrix} 0.18132 & 0.41528 & 6.71733 \\ 0.41528 & 202.76587 & 319.11597 \\ 6.71733 & 319.11597 & 1351.47116 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{|S^2|} \propto [S^2]' = \frac{1}{153.69} = \begin{bmatrix} \frac{0.18132}{153.69} & \frac{0.41528}{153.69} & \frac{6.71733}{153.69} \\ \frac{0.41528}{153.69} & \frac{202.76587}{153.69} & \frac{319.11597}{153.69} \\ \frac{6.71733}{153.69} & \frac{319.11597}{153.69} & \frac{1351.47116}{153.69} \end{bmatrix}$$

$$[S^2]^{-1} = \begin{bmatrix} 0.00118 & 0.00270 & 0.04370 \\ 0.00270 & 1.31931 & 2.07635 \\ 0.04371 & 2.07636 & 8.79346 \end{bmatrix}$$

Obtener T^2

$$T^2 = n[\widehat{M} - M_{H_0}]'[S^2]^{-1}[\widehat{M} - M_{H_0}]$$

$$T^2 = 7[190.8 - 211.5; 7.50 - 5.85; 3.60 - 3.29][S^2]^{-1} \begin{bmatrix} 190.8 - 211.5 \\ 7.50 - 5.85 \\ 3.60 - 3.29 \end{bmatrix}$$

$$T^2 = 7[20.7; -1.65; - - 0.3143][S^2]^{-1} \begin{bmatrix} 20.7 \\ -1.65 \\ -0.3143 \end{bmatrix}$$

$$T^2 = [144.9; -11.55; -2.2]\begin{bmatrix}[0.00118 & 0.00270 & 0.04370]\\ 0.00270 & 1.31931 & 2.07635\\ 0.04371 & 2.07636 & 8.79346\end{bmatrix}\begin{bmatrix}20.7\\ -1.65\\ -0.3143\end{bmatrix}$$

$$T^2 = [0.04359; -19.41454; -36.99442]\begin{bmatrix}20.7\\ -1.65\\ -0.3143\end{bmatrix}$$

$$T^2 = \underline{44.56302}$$

$$Fcalc = \frac{n-p}{p(n-1)}T^2$$

$$Fcalc = \frac{7-3}{3(7-1)}44.56302 = .2222 * 44.56302 = 9.90$$

3. Región de rechazo

$$F_{\propto.p.n-p}$$

$$F_{0.05.3.4} = 6.59$$

$$F_{0.01.3.4} = 16.69$$

4. Regla de decisión

Se rechaza H_0 si Fcalc >6.59

5. Conclusión

Estadística: Como 6.59 >9.90 >16.69

$\therefore$ la prueba es (S) y se rechaza H_0 con poca potencia.

Adva: Existe evidencia suficiente (E.E.S). para decir que estadísticamente al menos uno de los parámetros de las tres variables estudiadas (X_1= Valor de avalúo en la zona Norte, X_2= Tamaño del lote en la zona Norte y X_3= Número de recámaras en la zona Norte es diferente.

Resolución del inciso b)

Intervalo de confianza para valor

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 211.5 \mp 2.447 \frac{33.5}{\sqrt{7}}$$

$$M = 211.5 \mp 2.447 \frac{33.5}{2.6458}$$

$$M = 211.5 \mp 2.447(12.66181)$$

$$M = 211.5 \mp 30.98 \begin{cases} LIC = 180.52 \\ LSC = 242.48 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del VALOR de avalúo en la zona Norte oscila entre 180.52 y 242.48 Como 190.8 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

Intervalo de confianza para tamaño

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 5.85 \mp 2.447 \frac{1.141}{\sqrt{7}}$$

$$M = 5.85 \mp 2.447 \frac{1.141}{2.6458}$$

$$M = 5.85 \mp 2.447(0.43126)$$

$$M = 5.85 \mp 1.05529 \begin{cases} LIC = 4.795 \\ LSC = 6.905 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del TAMAÑO del lote en la zona Norte oscila entre 4.795 y 6.905 Como 7.5 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

Intervalo de confianza para Número de recámaras

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 3.29 \mp 2.447 \frac{0.488}{\sqrt{7}}$$

$$M = 3.29 \mp 2.447 \frac{0.488}{2.6458}$$

$$M = 3.29 \mp 2.447(0.184)$$

$M = 3.29 \mp 0.45025 \begin{cases} LIC = 2.834 \\ LSC = 3.74 \end{cases}$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del Número de recamaras en la zona Norte oscila entre 2.834 y 3.74. Como 3.60 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

3.2.1.1.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.2.1.2 Vector de Medias Multivariado**

3.2.1.1.2 El Director de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia de la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares) , el tamaño del lote (en pies cuadrados) y el número de recámaras. Para este experimento se manejo uno de los niveles de la ubicación geográfica: **Sur**. Se tomaron lecturas en 7 casas unifamiliares y los resultados fueron los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada denominada **"BIENES1"** donde X_1= Valor de avalúo en la zona **Sur** , X_2= Tamaño del lote en la zona **Sur** y X_3= Número de recámaras en la zona **Sur**:

$$X = \begin{bmatrix} 189.9 & 7.20 & 3 \\ 305.0 & 12.10 & 3 \\ 189.9 & 7.00 & 5 \\ 115.0 & 18.92 & 3 \\ 210.0 & 10.00 & 3 \\ 181.0 & 7.00 & 4 \\ 180.0 & 7.00 & 4 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 155 \\ 7 \\ 3 \end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de hipótesis

$$H_0 = \begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = \begin{bmatrix} 155 \\ 7 \\ 3 \end{bmatrix}$$

$H_1 = Al\ menos\ una\ M\ es\ \neq.$

2. Estadístico de prueba

$$Fcalc = \frac{n-p}{p(n-1)} T^2$$

Donde: $T^2 = n(M - M_0)'[S^2]^{-1}(\widehat{M} - \widehat{M}_0)$

Donde: $S^2 = \frac{1}{n-1} SCPC = \frac{1}{n-1} D'D$ (Matriz de varianzas y covarianzas)=

$$[s^2]^2 = \frac{1}{|S|} \propto [S^2]'$$

$$D = \begin{bmatrix} 189.9 & 7.20 & 3 \\ 305.0 & 12.10 & 3 \\ 189.9 & 7.00 & 5 \\ 115.0 & 18.92 & 3 \\ 210.0 & 10.00 & 3 \\ 181.0 & 7.00 & 4 \\ 180.0 & 7.00 & 4 \end{bmatrix} - \begin{bmatrix} 195.8 & 9.89 & 3.57 \\ 195.8 & 9.89 & 3.57 \\ 195.8 & 9.89 & 3.57 \\ 195.8 & 9.89 & 3.57 \\ 195.8 & 9.89 & 3.57 \\ 195.8 & 9.89 & 3.57 \\ 195.8 & 9.89 & 3.57 \end{bmatrix} = \begin{bmatrix} -5.93 & -2.69 & -0.57 \\ 109.17 & 2.21 & -0.57 \\ -5.93 & -2.89 & 1.43 \\ -80.83 & 9.03 & -0.57 \\ 14.17 & 0.11 & -0.57 \\ -14.83 & -2.89 & 0.43 \\ -15.83 & -2.89 & 0.43 \end{bmatrix}$$

$$D'D = \begin{bmatrix} -5.93 & 109.17 & -5.93 & -80.83 & 14.17 & -14.83 & -15.83 \\ -2.69 & 2.21 & -2.89 & 9.03 & 0.11 & -2.89 & -2.89 \\ -0.57 & -0.57 & 1.43 & -0.57 & -0.57 & 0.43 & 0.43 \end{bmatrix} \begin{bmatrix} -5.93 & -2.69 & -0.57 \\ 109.17 & 2.21 & -0.57 \\ -5.93 & -2.89 & 1.43 \\ -80.83 & 9.03 & -0.57 \\ 14.17 & 0.11 & -0.57 \\ -14.83 & -2.89 & 0.43 \\ -15.83 & -2.89 & 0.43 \end{bmatrix}$$

$$D'D = \begin{bmatrix} 19193.21429 & -365.37371 & -42.51429 \\ -355.37371 & 118.72949 & -11.55429 \\ -42.51429 & -11.55429 & 3.71429 \end{bmatrix}$$

$$S^2 = \frac{1}{n-1} D'D = \frac{1}{7-1} D'D = \frac{1}{6} \begin{bmatrix} 19193.21429 & -365.37371 & -42.51429 \\ -355.37371 & 118.72949 & -11.55429 \\ -42.51429 & -11.55429 & 3.71429 \end{bmatrix}$$

$$S^2 = \frac{1}{6} \begin{bmatrix} 19193.21429 & -365.37371 & -42.51429 \\ -365.37371 & 118.72949 & -11.55429 \\ -42.51429 & -11.55429 & 3.71429 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} \frac{19193.21429}{6} & \frac{-365.37371}{6} & \frac{-42.51429}{6} \\ \frac{-365.37371}{6} & \frac{118.72949}{6} & \frac{-11.55429}{6} \\ \frac{-42.51429}{6} & \frac{-11.55429}{6} & \frac{3.71429}{6} \end{bmatrix}$$

Utilizando la regla de Sarrus

$$|S^2| = \begin{bmatrix} \frac{198193.21429}{6} & \frac{-365.37371}{6} & \frac{-42.51429}{6} \\ \frac{-365.37371}{6} & \frac{118.72949}{6} & \frac{-11.55429}{6} \\ \frac{-42.51429}{6} & \frac{-11.55429}{6} & \frac{3.71429}{6} \\ \frac{19193.21429}{6} & \frac{-365.37371}{6} & \frac{-42.51429}{6} \\ \frac{-365.37371}{6} & \frac{118.72949}{6} & \frac{-11.55429}{6} \end{bmatrix} = 22372.1506$$

Matriz de cofactores

$$\propto A\begin{bmatrix} +\begin{vmatrix} 19.78825 & -1.92571 \\ -1.92571 & 0.61905 \end{vmatrix} & -\begin{vmatrix} -60.89562 & -1.92571 \\ -7.08571 & 0.61905 \end{vmatrix} & +\begin{vmatrix} -60.89562 & 19.78825 \\ -7.08571 & -1.92571 \end{vmatrix} \\ -\begin{vmatrix} -60.89562 & -7.08571 \\ -1.92571 & 0.61905 \end{vmatrix} & +\begin{vmatrix} 3198.86905 & -7.08571 \\ -7.08571 & 0.61905 \end{vmatrix} & -\begin{vmatrix} 3198.86905 & -60.89562 \\ -7.08571 & -1.92571 \end{vmatrix} \\ +\begin{vmatrix} -60.89562 & -7.08571 \\ 19.78825 & -1.92571 \end{vmatrix} & -\begin{vmatrix} 3198.86905 & -7.08571 \\ -60.89562 & -1.92571 \end{vmatrix} & +\begin{vmatrix} 3198.86905 & -60.89562 \\ -60.89562 & 19.78825 \end{vmatrix} \end{bmatrix}' =$$

$$= \begin{bmatrix} +(12.24987 - 3.70838) & -(-37.69729 - 13.64506) & +(117.26756 + 140.21387) \\ -(-37.69729 - 13.64506) & +(1980.25227 - 50.20735) & -(-6160.10782 - 431.48896) \\ +((117.26756 + 140.21387) & -(-6160.10782 - 431.48896) & +(63300.01282 - 3708.27642) \end{bmatrix} =$$

$$= \begin{bmatrix} 8.54149 & 51.34235 & 257.48143 \\ 51.34235 & 1930.04492 & 6591.59678 \\ 257.48143 & 6591.59678 & 59591.73640 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{|S^2|} \propto [S^2]' = \frac{1}{22372.1506} = \begin{bmatrix} \frac{8.54149}{22372.1506} & \frac{51.34235}{22372.1506} & \frac{257.48143}{22372.1506} \\ \frac{51.34235}{22372.1506} & \frac{1930.04492}{22372.1506} & \frac{6591.59678}{22372.1506} \\ \frac{257.48143}{22372.1506} & \frac{6591.59678}{22372.1506} & \frac{59591.73640}{22372.1506} \end{bmatrix}$$

$$[S^2]^{-1} = \begin{bmatrix} 0.00038 & 0.00229 & 0.01151 \\ 0.00229 & 0.08627 & 0.29463 \\ 0.01151 & 0.29463 & 2.66366 \end{bmatrix}$$

Obtener T^2

$$T^2 = n[\widehat{M} - M_{H_0}]'[S^2]^{-1}[\widehat{M} - M_{H_0}]$$

$$T^2 = 7[155 - 195.8; 7.00 - 9.89; 3.00 - 3.57][S^2]^{-1} \begin{bmatrix} 155 - 195.8 \\ 7.00 - 9.89 \\ 3.00 - 3.57 \end{bmatrix}$$

$$T^2 = 7[40.8286; 2.8886; 0.5714][S^2]^{-1}\begin{bmatrix}40.8286\\2.8886\\0.5714\end{bmatrix}$$

$$T^2 = [285.8; 20.22; 4]\left[\begin{bmatrix}0.00038 & 0.00229 & 0.01151\\0.00229 & 0.08627 & 0.29463\\0.01151 & 0.29463 & 2.66366\end{bmatrix}\right]\begin{bmatrix}40.8286\\2.8886\\0.5714\end{bmatrix}$$

$$T^2 = [0.20156; 3.57880; 19.90140]\begin{bmatrix}40.8286\\2.8886\\0.5714\end{bmatrix}$$

$T^2 = \underline{29.93908}$

$$Fcalc = \frac{n-p}{p(n-1)}T^2$$

$$Fcalc = \frac{7-3}{3(7-1)}29.93908 = .2222 * 29.93908 = 6.65$$

3. Región de rechazo

$F_{\propto.p.n-p}$

$F_{0.05.3.4} = 6.59$

$F_{0.01.3.4} = 16.69$

4. Regla de decisión

Se rechaza H_0 si Fcalc >6.59

5. Conclusión

Estadística: Como 6.59 >6.65 >16.69

∴ la prueba es (S) y se rechaza H_0 con poca potencia.

Adva: Existe evidencia suficiente (E.E.S). para decir que estadísticamente al menos uno de los parámetros de las tres variables estudiadas (X_1= Valor de avalúo en la zona Sur, X_2= Tamaño del lote en la zona Sur y X_3= Número de recámaras en la zona Sur es diferente.

Resolución del inciso b)

Intervalo de confianza para valor

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 195.8 \mp 2.447 \frac{56.6}{\sqrt{7}}$$

$$M = 195.8 \mp 2.447 \frac{56.6}{2.6458}$$

$$M = 195.8 \mp 2.447(21.39240)$$

$$M = 195.8 \mp 52.34 \begin{cases} LIC = 143.46 \\ LSC = 248.14 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del VALOR de avalúo en la zona Sur oscila entre 143.46 y 248.14 Como 155 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

Intervalo de confianza para tamaño

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 9.89 \mp 2.447 \frac{4.45}{\sqrt{7}}$$

$$M = 9.89 \mp 2.447 \frac{4.45}{2.6458}$$

$$M = 9.89 \mp 2.447(1.68)$$

$$M = 9.89 \mp 4.11096 \begin{cases} LIC = 5.77 \\ LSC = 14.00 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del TAMAÑO del lote en la zona Sur oscila entre 5.77 y 14 Como 7 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

Intervalo de confianza para Número de recámaras

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 3.57 \mp 2.447 \frac{0.787}{\sqrt{7}}$$

$$M = 3.57 \mp 2.447 \frac{0.787}{2.6458}$$

$$M = 3.57 \mp 2.447(0.297)$$

$$M = 3.57 \mp 0.72676 \begin{cases} LIC = 2.843 \\ LSC = 4.297 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del Número de recamaras en la zona Sur oscila entre 2.843 y 4.297. Como 3 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

3.2.1.1.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.1.1.3. Vector de Medias Multivariado**

3.2.1.1.3 El Director de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia de la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares) , el tamaño del lote (en pies cuadrados) y el número de recámaras. Para este experimento se manejo uno de los niveles de la ubicación geográfica: **Este**. Se tomaron lecturas en 7 casas unifamiliares y los resultados fueron los siguientes:

Considere la siguiente muestra de tamaño n=7 extraída de una población normal multivariada denominada **"BIENES1"** donde X_1= Valor de avalúo en la zona **Este** , X_2= Tamaño del lote en la zona **Este** y X_3= Número de recámaras en la zona **Este**:

$$X = \begin{bmatrix} 180.0 & 6.00 & 4 \\ 152.0 & 6.00 & 4 \\ 167.8 & 6.00 & 3 \\ 180.0 & 11.00 & 4 \\ 163.0 & 6.00 & 3 \\ 223.5 & 6.00 & 4 \\ 160.0 & 6.00 & 4 \end{bmatrix}$$

a) Desarrolle el procedimiento de inferencia correspondiente para las medias de los tres vectores suponiendo la siguiente hipótesis nula :

$$H_o: \begin{bmatrix}\mu_1\\ \mu_2\\ \mu_3\end{bmatrix} = \begin{bmatrix}195\\ 7.5\\ 3\end{bmatrix}$$

b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada variable en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y las hipotéticas e interprete.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesi

1. Juego de hipótesis

$$H_0 = \begin{bmatrix}M_1\\ M_2\\ M_3\end{bmatrix} = \begin{bmatrix}195\\ 7.5\\ 3\end{bmatrix}$$

$H_1 = Al\ menos\ una\ M\ es\ \neq.$

2. Estadístico de prueba

$$Fcalc = \frac{n-p}{p(n-1)}T^2$$

Donde: $T^2 = n(M - M_0)'[S^2]^{-1}(\widehat{M} - \widehat{M}_0)$

Donde: $S^2 = \frac{1}{n-1}\ SCPC = \frac{1}{n-1}\ D'D$ (Matriz de varianzas y covarianzas)=

$[s^2]^2 = \frac{1}{|S|} \propto [S^2]'$

$$D = \begin{bmatrix} 180.0 & 6.00 & 4 \\ 152.0 & 6.00 & 4 \\ 167.8 & 6.00 & 3 \\ 180.0 & 11.00 & 4 \\ 163.0 & 6.00 & 3 \\ 223.5 & 6.00 & 4 \\ 160.0 & 6.71 & 4 \end{bmatrix} - \begin{bmatrix} 175.2 & 6.71 & 3.71 \\ 175.2 & 6.71 & 3.71 \\ 175.2 & 6.71 & 3.71 \\ 175.2 & 6.71 & 3.71 \\ 175.2 & 6.71 & 3.71 \\ 175.2 & 6.71 & 3.71 \\ 175.2 & 6.71 & 3.71 \end{bmatrix} = \begin{bmatrix} 4.81 & -0.71 & 0.29 \\ -23.19 & -0.71 & 0.29 \\ -7.39 & -0.71 & -0.71 \\ 4.81 & 4.29 & 0.29 \\ -12.19 & -0.71 & -0.71 \\ 48.31 & -0.71 & 0.29 \\ -15.19 & -0.71 & 0.29 \end{bmatrix}$$

$$D'D = \begin{bmatrix} 4.81 & -23.19 & -7.39 & 4.81 & -12.19 & 48.31 & -15.19 \\ -0.71 & -0.71 & -0.71 & 4.29 & -0.71 & -0.71 & -0.71 \\ 0.29 & 0.29 & -0.71 & 0.29 & -0.71 & 0.29 & 0.29 \end{bmatrix} \begin{bmatrix} 4.81 & -0.71 & 0.29 \\ -23.19 & -0.71 & 0.29 \\ -7.39 & -0.71 & -0.71 \\ 4.81 & 4.29 & 0.29 \\ -12.19 & -0.71 & -0.71 \\ 48.31 & -0.71 & 0.29 \\ -15.19 & -0.71 & 0.29 \end{bmatrix}$$

$$D'D = \begin{bmatrix} 3351.84857 & 24.07143 & 19.57143 \\ 24.07143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}$$

$$S^2 = \frac{1}{n-1} D'D = \frac{1}{7-1} D'D = \frac{1}{6} \quad \begin{bmatrix} 3351.84857 & 24.07143 & 19.57143 \\ 24.07143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}$$

$$S^2 = \frac{1}{6} \begin{bmatrix} 3351.84857 & 24.07143 & 19.57143 \\ 24.07143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}$$

$$S^2 = \begin{bmatrix} \frac{3351.84857}{6} & \frac{24.07143}{6} & \frac{19.57143}{6} \\ \frac{24.07143}{6} & \frac{21.42857}{6} & \frac{1.42857}{6} \\ \frac{19.57143}{6} & \frac{1.42857}{6} & \frac{1.42857}{6} \end{bmatrix}$$

Utilizando la regla de Sarrus

$$|S^2| = \begin{bmatrix} \frac{3351.84857}{6} & \frac{24.07143}{6} & \frac{19.57143}{6} \\ \frac{24.07143}{6} & \frac{21.42857}{6} & \frac{1.42857}{6} \\ \frac{19.57143}{6} & \frac{1.42857}{6} & \frac{1.42857}{6} \\ \frac{3351.84857}{6} & \frac{24.07143}{6} & \frac{19.57143}{6} \\ \frac{24.07143}{6} & \frac{21.42857}{6} & \frac{1.42857}{6} \end{bmatrix} = 407.76554$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 3.57143 & 0.23810 \\ 0.2810 & 0.23810 \end{vmatrix} & -\begin{vmatrix} 4.01190 & 0.23810 \\ 3.26190 & 0.23810 \end{vmatrix} & +\begin{vmatrix} 4.01190 & 3.57143 \\ 3.26190 & 0.23810 \end{vmatrix} \\ -\begin{vmatrix} 4.01190 & 3.26190 \\ 0.23810 & 0.23810 \end{vmatrix} & +\begin{vmatrix} 558.64143 & 3.26190 \\ 3.26190 & 0.23810 \end{vmatrix} & -\begin{vmatrix} 558.64143 & 4.01190 \\ 3.26190 & 0.23810 \end{vmatrix} \\ +\begin{vmatrix} 4.01190 & 3.26190 \\ 3.57143 & 0.23810 \end{vmatrix} & -\begin{vmatrix} 558.64143 & 3.26190 \\ 4.01190 & 0.23810 \end{vmatrix} & +\begin{vmatrix} 558.64143 & 4.01190 \\ 4.01190 & 3.57143 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(0.85034 - 0.05669) & -(0.95522 - 0.77664) & +(0.95522 - 11.64966) \\ -(0.95522 - 0.77664) & +(133.00986 - 10.64002) & -(133.00986 - 13.08645) \\ +((0.95522 - 11.64966) & -(133.00986 - 13.08645) & +(1995.14796 - 16.09538) \end{bmatrix} =$$

$$= \begin{bmatrix} 0.79365 & -0.17857 & -10.69444 \\ -0.17857 & 122.36984 & -119.92341 \\ -10.69444 & -119.92341 & 1979.05258 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{|S^2|} \propto [S^2]' = \frac{1}{407.76554} = \begin{bmatrix} \frac{0.79365}{407.76554} & \frac{-0.17857}{407.76554} & \frac{-10.69444}{407.76554} \\ \frac{-0.17857}{407.76554} & \frac{122.36984}{407.76554} & \frac{-119.92341}{407.76554} \\ \frac{-10.69444}{407.76554} & \frac{-119.92341}{407.76554} & \frac{1979.05258}{407.76554} \end{bmatrix}$$

$$[S^2]^{-1} = \begin{bmatrix} 0.00195 & -0.00044 & -0.02623 \\ -0.00044 & 0.30010 & -0.29410 \\ -0.02623 & -0.29410 & 4.85341 \end{bmatrix}$$

Obtener T^2

$$T^2 = n[\widehat{M} - M_{H_0}]'[S^2]^{-1}[\widehat{M} - M_{H_0}]$$

$$T^2 = 7[195.0 - 175.2; 7.50 - 6.71; 3.00 - 3.71][S^2]^{-1} \begin{bmatrix} 195.0 - 175.2 \\ 7.50 - 6.71 \\ 3.00 - 3.71 \end{bmatrix}$$

$$T^2 = 7[-19.8143; -0.7857; 0.7143][S^2]^{-1} \begin{bmatrix} -19.8142 \\ -0.7857 \\ 0.7143 \end{bmatrix}$$

$$T^2 = [-138.7; -5.5; 5] \left[\begin{bmatrix} 0.00195 & -0.00044 & -0.02623 \\ -0.00044 & 0.30010 & -0.29410 \\ -0.02623 & -0.29410 & 4.85341 \end{bmatrix} \right] \begin{bmatrix} -19.8142 \\ -0.7857 \\ 0.7143 \end{bmatrix}$$

$$T^2 = [-0.39868; 3.06030; 29.52262] \begin{bmatrix} -19.8142 \\ -0.7857 \\ 0.7143 \end{bmatrix}$$

$$T^2 = \underline{31.39148}$$

$$Fcalc = \frac{n-p}{p(n-1)} T^2$$

$$Fcalc = \frac{7-3}{3(7-1)} 31.39148 = .2222 * 31.39148 = 6.98$$

3. Región de rechazo

$F_{\propto.p.n-p}$

$F_{0.05.3.4} = 6.59$

$F_{0.01.3.4} = 16.69$

4. Regla de decisión

Se rechaza H_0 si Fcalc >6.59

5. Conclusión

Estadística: Como 6.59 >6.98 >16.69

∴ la prueba es (S) y se rechaza H_0 con poca potencia.

Adva: Existe evidencia suficiente (E.E.S). para decir que estadísticamente al menos uno de los parámetros de las tres variables estudiadas (X_1= Valor de avalúo en la zona Este, X_2= Tamaño del lote en la zona Este y X_3= Número de recámaras en la zona Este es diferente.

Resolución del inciso b)

Intervalo de confianza para valor

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 175.2 \mp 2.447 \frac{23.64}{\sqrt{7}}$$

$$M = 175.2 \mp 2.447 \frac{23.64}{2.6458}$$

$$M = 175.2 \mp 2.447(8.93508)$$

$$M = 175.2 \mp 21.86414 \begin{cases} LIC = 153.33 \\ LSC = 197.06 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del VALOR de avalúo en la zona Este oscila entre 153.33 y 197.06 Como 195 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

Intervalo de confianza para tamaño

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 6.71 \mp 2.447 \frac{1.89}{\sqrt{7}}$$

$$M = 6.71 \mp 2.447 \frac{1.89}{2.6458}$$

$$M = 6.71 \mp 2.447(0.71435)$$

$$M = 6.71 \mp 1.74802 \begin{cases} LIC = 4.96 \\ LSC = 8.46 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del TAMAÑO del lote en la zona Este oscila entre 4.96 y 8.46 Como 7.5 se encuentra dentro del intervalo de confianza, la prueba de hipótesis multivariada no se debió haber rechazado por esta variable.

Intervalo de confianza para Número de recámaras

$$M = \bar{x} \mp t_{\propto/2.n-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.7-1} \frac{S}{\sqrt{n}}$$

$$M = \bar{x} \mp t_{0.05/2.6} \frac{S}{\sqrt{n}}$$

$$M = 3.71 \mp 2.447 \frac{0.488}{\sqrt{7}}$$

$$M = 3.71 \mp 2.447 \frac{0.488}{2.6458}$$

$$M = 3.71 \mp 2.447(0.184)$$

$$M = 3.71 \mp 0.45025 \begin{cases} LIC = 3.26 \\ LSC = 4.16 \end{cases}$$

Interpretación:

El 95% de cada 100 muestras de tamaño 7, el verdadero valor promedio del Número de recamaras en la zona Este oscila entre 3.26 y 4.16. Como 3 se encuentra fuera del intervalo de confianza, la prueba de hipótesis multivariada se debió haber rechazado por esta variable.

3.2.2.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.2.1.1. Diferencia de Medias Multivariada**

3.2.2.1.1. El Director de Mercadotecnia de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia que llega a tener la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares), el tamaño del lote (en pies cuadrados) y número de recámaras. Para este experimento se manejaron dos de los niveles de la ubicación geográfica: **Norte y Sur**. Se tomaron lecturas en 5 casas unifamiliares ubicadas en la zona Norte y 7 ubicadas en la zona Sur y los resultados fueron los siguientes:

Considere la siguiente muestra extraída de una población normal bivariada denominada **"BIENES1"** donde X_{11}= Valor de avalúo en la zona Norte, X_{12}= Tamaño del lote en la zona norte, X_{31}= Número de recámaras en la zona Norte, X_{12}= Valor del avalúo en la zona Sur, X_{22}= Tamaño del Lote en la zona Sur y X_{32}= Número de recámaras en la zona Sur:

Nota: El subíndice indica número de variable y zona. Ejemplo: X_{11} indica Variable 1 Valor en zona 1 Norte

ZONA NORTE			**ZONA SUR**		
X_{11}	**X_{21}**	**X_{31}**	**X_{12}**	**X_{22}**	**X_{32}**
Valor	**Tamaño**	**Número**	**Valor**	**Tamaño**	**Número**
265	26	2	181	7	4
245	6	5	220	9	4
210	16	2	190	7	4
239	9	3	160	6	2
299	11	4	176	6	4
			181	7	3
			185	6	3

a) Existe evidencia suficiente que indique que la influencia que llega a tener la ubicación geográfica de las casas unifamiliares de las zonas Norte y Sur sobre el valor de las mismas, el tamaño del lote y el número de recámaras es el mismo en ambas ubicaciones?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de Hipótesis

$$H_0: \begin{bmatrix} M_{11} \\ M_{21} \\ M_{31} \end{bmatrix} = \begin{bmatrix} M_{12} \\ M_{22} \\ M_{32} \end{bmatrix}$$

$H_1: Al\ menos\ 1\ combinación\ es\ \neq$

2. Estadístico de prueba

$$Fcalc = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

Donde $T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)^1 [S^2]^{-1} (\widehat{M}_1 - \widehat{M}_2)$

$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2}$

$W_1 = SCPC_1 = D'D_1$

$W_2 = SCPC_2 = D_2' D_2$

$$D_1 = \begin{bmatrix} 265 & 26 & 2 \\ 245 & 6 & 5 \\ 210 & 16 & 2 \\ 239 & 9 & 3 \\ 299 & 11 & 6 \end{bmatrix} - \begin{bmatrix} 251.6 & 13.6 & 3.6 \\ 251.6 & 13.6 & 3.6 \\ 251.6 & 13.6 & 3.6 \\ 251.6 & 13.6 & 3.6 \\ 251.6 & 13.6 & 3.6 \end{bmatrix} = \begin{bmatrix} 13.4 & 12.4 & -1.6 \\ -6.6 & -7.6 & 1.4 \\ -41.6 & 2.4 & -1.6 \\ -12.6 & -4.6 & -0.6 \\ 47.4 & -2.6 & 2.4 \end{bmatrix}$$

$$D_1'D_1 = \begin{bmatrix} 13.4 & -6.6 & -41.6 & -12.6 & 47.4 \\ 12.4 & -7.6 & 2.4 & -4.6 & -2.6 \\ -1.6 & 1.4 & -1.6 & -0.6 & 2.4 \end{bmatrix} \begin{bmatrix} 0.80 & 1.20 & -0.20 \\ -0.20 & 0.20 & 0.80 \\ -1.20 & -0.80 & 0.80 \\ 0.80 & 0.20 & 0.80 \\ -0.20 & -0.80 & -2.20 \end{bmatrix} = \begin{bmatrix} 4359.2 & 51.2 & 157.2 \\ 51.2 & 245.2 & -37.8 \\ 157.2 & -37.8 & 13.2 \end{bmatrix}$$

$$W_1 = D_1'D_1 = \begin{bmatrix} 4359.2 & 51.2 & 157.2 \\ 51.2 & 245.2 & -37.8 \\ 157.2 & -37.8 & 13.2 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 181 & 7 & 4 \\ 220 & 9 & 4 \\ 190 & 7 & 4 \\ 160 & 6 & 2 \\ 176 & 6 & 4 \\ 181 & 7 & 3 \\ 185 & 6 & 3 \end{bmatrix} - \begin{bmatrix} 184.71 & 6.86 & 3.43 \\ 184.71 & 6.86 & 3.43 \\ 184.71 & 6.86 & 3.43 \\ 184.71 & 6.86 & 3.43 \\ 184.71 & 6.86 & 3.43 \\ 184.71 & 6.86 & 3.43 \\ 184.71 & 6.86 & 3.43 \end{bmatrix} = \begin{bmatrix} -3.71 & 0.14 & 0.57 \\ 35.29 & 2.14 & 0.57 \\ 5.29 & 0.14 & 0.57 \\ -24.71 & -0.86 & -1.43 \\ -8.71 & -0.86 & 0.57 \\ -3.71 & 0.14 & -0.43 \\ 0.29 & -0.86 & -0.43 \end{bmatrix}$$

$$D_2'D_2 = \begin{bmatrix} -3.71 & 35.29 & 5.29 & -24.71 & -8.71 & -3.71 & 0.29 \\ 0.14 & 2.14 & 0.14 & -0.86 & -0.86 & 0.14 & -0.86 \\ 0.57 & 0.57 & 0.57 & -1.43 & 0.57 & -0.43 & -0.43 \end{bmatrix} \begin{bmatrix} -3.71 & 0.14 & 0.57 \\ 35.29 & 2.14 & 0.57 \\ 5.29 & 0.14 & 0.57 \\ -24.71 & -0.86 & -1.43 \\ -8.71 & -0.86 & 0.57 \\ -3.71 & 0.14 & -0.43 \\ 0.29 & -0.86 & -0.43 \end{bmatrix} = \begin{bmatrix} 1987.42857 & 103.71429 & 52.85714 \\ 103.71429 & 6.85714 & 2.42857 \\ 52.85714 & 2.42857 & 3.71429 \end{bmatrix}$$

$$W_2 = D_2'D_2 = \begin{bmatrix} \begin{bmatrix} 1987.42857 & 103.71429 & 52.85714 \\ 103.71429 & 6.85714 & 2.42857 \\ 52.85714 & 2.42857 & 3.71429 \end{bmatrix} \end{bmatrix}$$

$$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2} = \frac{\begin{bmatrix} 4359.2 & 51.2 & 157.2 \\ 51.2 & 245.2 & -37.8 \\ 157.2 & -37.8 & 13.2 \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} 1987.42857 & 103.71429 & 52.85714 \\ 103.71429 & 6.85714 & 2.42857 \\ 52.85714 & 2.42857 & 3.71429 \end{bmatrix} \end{bmatrix}}{5 + 7 - 2}$$

$$= \begin{bmatrix} 634.66286 & 15.49143 & 21.00571 \\ 15.49143 & 25.20571 & -3.53714 \\ 21.00571 & -3.53714 & 1.69143 \end{bmatrix}$$

$$|S^2| = \begin{bmatrix} 634.66286 & 15.49143 & 21.00571 \\ 15.49143 & 25.20571 & -3.53714 \\ 21.00571 & -3.53714 & 1.69143 \\ 634.66286 & 15.49143 & 21.00571 \\ 15.49143 & 25.20571 & -3.53714 \end{bmatrix} = 5287.77646$$

Como es ≠ de 0 tiene solución

Transpuesta de S^2

$$[S^2]^1 = \begin{bmatrix} 634.66286 & 15.49143 & 21.00571 \\ 15.49143 & 25.20571 & -3.53714 \\ 21.00571 & -3.53714 & 1.69143 \end{bmatrix}$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 25.20571 & -3.53714 \\ -3.53714 & 1.69143 \end{vmatrix} & -\begin{vmatrix} 15.49143 & -3.53714 \\ 21.00571 & 1.69143 \end{vmatrix} & +\begin{vmatrix} 15.49143 & 25.20571 \\ 21.00571 & -3.53714 \end{vmatrix} \\ -\begin{vmatrix} 15.49143 & 21.00571 \\ -3.53714 & 1.69143 \end{vmatrix} & +\begin{vmatrix} 634.66286 & 21.00571 \\ 21.00571 & 1.69143 \end{vmatrix} & -\begin{vmatrix} 634.66286 & 15.49143 \\ 21.00571 & -3.53714 \end{vmatrix} \\ +\begin{vmatrix} 15.49143 & 21.00571 \\ 25.20571 & -3.53714 \end{vmatrix} & -\begin{vmatrix} 634.66286 & 21.00571 \\ 15.49143 & -3.53714 \end{vmatrix} & +\begin{vmatrix} 634.66286 & 15.49143 \\ 15.49143 & 25.20571 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(42.63367 - 12.51138) & -(26.20264 + 74.30021) & +(-54.79540 - 529.46403) \\ -(26.20264 + 74.30021) & +(1073.48689 - 441.24003) & -(-2244.89319 - 325.40852) \\ +((-54.79540 - 529.46403) & -(-2244.89319 - 325.40852) & +(15997.13064 - 239.98436) \end{bmatrix} =$$

$$= \begin{bmatrix} 30.12229 & -100.50286 & -584.25943 \\ -100.50286 & 632.24686 & 2570.30171 \\ -584.25943 & 2570.30171 & 15757.14629 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{5287.77646} = \begin{bmatrix} 30.12229 & -100.50286 & -584.25943 \\ -100.50286 & 632.24686 & 2570.30171 \\ -584.25943 & 2570.30171 & 15757.14629 \end{bmatrix}$$

$$= \begin{bmatrix} 0.00570 & -0.01901 & -0.11049 \\ -0.01901 & 0.11957 & 0.48608 \\ -0.11049 & 0.48608 & 2.97992 \end{bmatrix}$$

$$T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\hat{M}_1 - \hat{M}_2)'[S^2]^{-1}(\hat{M}_1 - \hat{M}_2)$$

$$T^2 = \left(\frac{5*7}{5+7}\right)(251.6-184.71; 13.6-6.86; 3.6-3.43)[S^2]^{-1}\begin{bmatrix}251.6-184.71\\13.6-6.86\\3.6-3.43\end{bmatrix}$$

$$T^2 = \left(\frac{35}{12}\right)(66.89; 6.74; 0.17)[S^2]^{-1}\begin{bmatrix}66.89\\6.74\\0.17\end{bmatrix}$$

$$T^2 = (2.91667)(66.89; 6.74; 0.17)[S^2]^{-1}\begin{bmatrix}66.89\\6.74\\0.17\end{bmatrix}$$

$$T^2 = [195.0833; 19.6667; 0.5]\begin{bmatrix}0.00570 & -0.01901 & -0.11049\\-0.01901 & 0.11957 & 0.48608\\-0.11049 & 0.48608 & 2.97992\end{bmatrix}\begin{bmatrix}66.89\\6.74\\0.17\end{bmatrix}$$

$$T^2 = [0.6823; -1.1133; -10.50563]\begin{bmatrix}66.89\\6.74\\0.17\end{bmatrix}$$

$$T^2 = \underline{36.32579}$$

$$Fcalc = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2$$

$$Fcalc = \frac{5+7-3-1}{3(5+7-2)}(36.32579)$$

$$Fcalc = \frac{8}{30}(36.32579)$$

$$Fcalc = 0.26667 \times 36.32579$$

$$\boldsymbol{Fcalc = 9.69}$$

3. Región de rechazo

$F_{\alpha.p.n1+n2-p-1} = F_{0.05.3.5+7-3-1} = F_{0.05.3.8} = 4.07$

$F_{\alpha.p.n1+n2-p-1} = F_{0.01.3.5+7-3-1} = F_{0.01.3.8} = 7.59$

4. Regla de decisión

Se rechaza H_0 si Fcalc > 4.07

5. Conclusión:

- Estadistica: Como 9.69> 4.07 y >7.59 ∴

La prueba es A.S y se rechaza H_0 con mucha potencia.

- Adva: Existe evidencia suficiente para decir que la influencia que llega a tener la ubicación geográfica de las casas sobre el valor del avalúo, el tamaño del lote y el número de recámaras no es el mismo en ambas zonas estudiadas en al menos una combinación de ellas.

Resolución del inciso b)

Nota importante: Para construir el intervalo de confianza se debe probar primero si las varianzas de las muestras son iguales o no para saber que caso de debe usar para construir el intervalo de confianza (Caso 2 para varianzas iguales o Caso 3 para varianzas desiguales) por lo que primero hay que realizar la prueba de homocedasticidad de varianzas.

Prueba de Homocedasticidad para Valor

1. Juego de Hipótesis

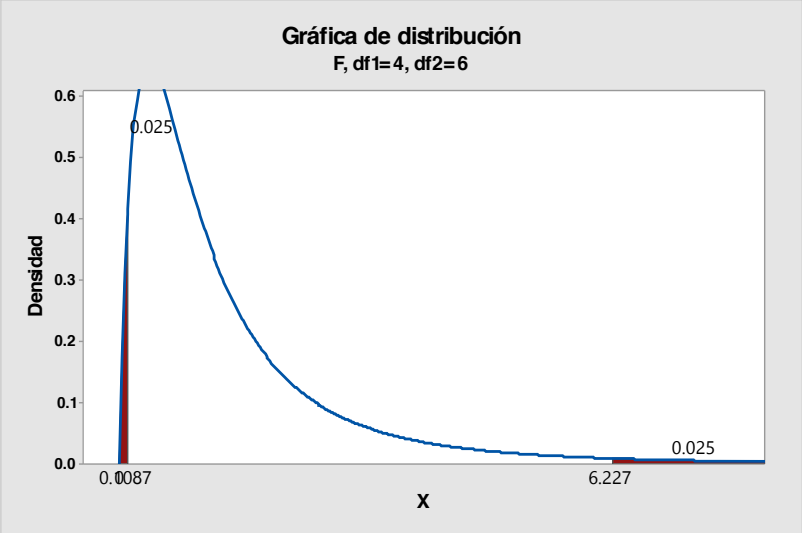
$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$

$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{11}}^2}{{S_{12}}^2} = \frac{1089}{331.24} = 3.28765$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc ≥ 6.23 ó si es ≤ 0.1087

5. Conclusión:

- Estadistica: Como 0.1087< 3.28765< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de Valor de ambas medias son iguales.

Intervalo de Confianza para Valor

$$M_{11} - M_{12} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp (t_{\alpha/2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)1089 + (7-1)331.24}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 1089 + 6 * 331.24}{10}}$$

$$Sp = \sqrt{\frac{4356 + 1987.44}{10}}$$

$$Sp = \sqrt{\frac{6343.44}{10}}$$

$$Sp = \sqrt{634.344}$$

$$Sp = 25.18619$$

$$M_{11} - M_{12} = (251.6 - 184.71) \mp 2.228 * 25.18619 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (66.89) \mp 2.228 * 25.18619 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (66.89) \mp 2.228 * 25.18619 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (66.89) \mp 2.228 * 25.18619 * 0.58554$$

$$M_{11} - M_{12} = (66.89) \mp 32.85748 \begin{cases} LIC = 34.03252 \\ LSC = 99.74748 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la influencia de la zona Norte sobre el valor de las casas es mayor que la influencia de la zona Sur por un mínimo de 34.03252 y un máximo de 99.74748 miles de dólares.

Nota: Como la influencia de la zona sobre el **Valor** de las casas es diferente en ambas zonas sería razonable que la prueba de hipótesis de diferencia de medias multivariada se hubiera rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Valor vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                  IC de 95% para
Niveles  N  Desv.Est.  Varianza      Desv.Est.
1        5     33.012  1089.800  (14.660, 122.268)
2        7     18.200   331.238  ( 7.643,  60.193)

Relación de desviaciones estándar = 1.814
Relación de varianzas = 3.290
```

ANOVA unidireccional: Valor vs. Niveles

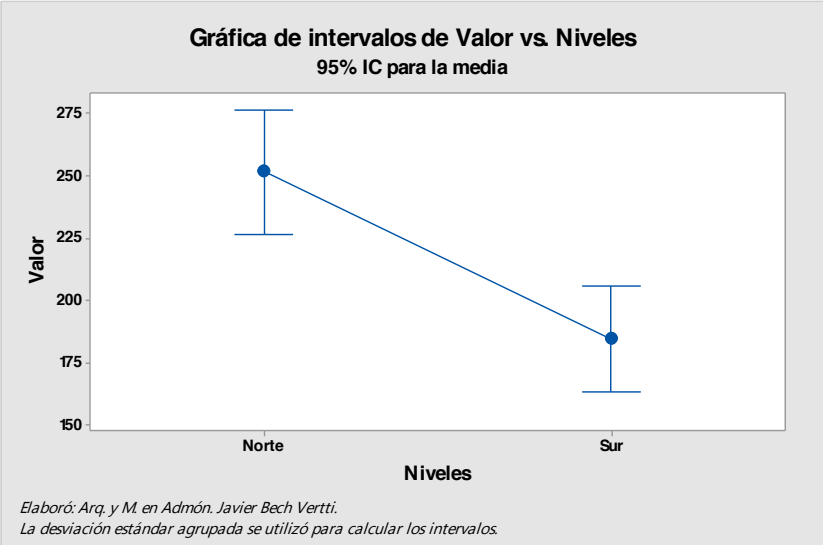
```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2
```

```
Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      13048    13048.3    20.56    0.001
Error    10       6347      634.7
Total    11      19395


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
25.1925   67.28%      64.00%   50.93%


Medias

Niveles  N   Media  Desv.Est.      IC de 95%
1        5   251.6       33.0  ( 226.5,  276.7)
2        7  184.71      18.20  (163.50, 205.93)

Desv.Est. agrupada = 25.1925
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

           Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%    Valor T  ajustado
2 - 1            -66.9        14.8  (-99.8, -34.0)    -4.53     0.001

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Valor vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba de Homocedasticidad para Tamaño

1. Juego de Hipótesis

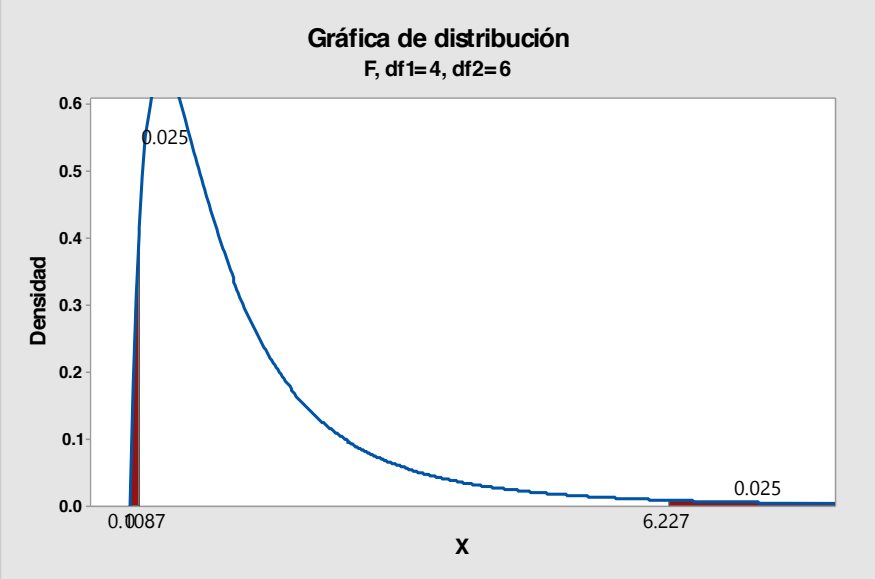
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{21}}^2}{{S_{22}}^2} = \frac{61.30890}{1.14490} = 53.54957$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:
   - Estadística: Como 53.54957>6.23 ∴ se rechaza H_0 y se utiliza el caso 3.
   - Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas del Tamaño de ambas medias son diferentes.

Intervalo de Confianza para Tamaño

$$M_{21} - M_{22} = \left(\overline{X}_{21} - \overline{X}_{22}\right) \mp t^{\propto}/_{2} \cdot v\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Donde (v), los grados de libertad se determinan de la siguiente manera:

$$v = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}}$$

$$v = \frac{\left[\frac{61.30890}{5} + \frac{1.14490}{7}\right]^2}{\frac{\left[\frac{61.30890}{5}\right]^2}{4} + \frac{\left[\frac{1.14490}{7}\right]^2}{6}} = \frac{[12.26178 + 0.16356]^2}{\frac{150.35125}{4} + \frac{0.02675}{6}} =$$

$$\frac{154.38907}{37.58781+0.00446} = \frac{154.38907}{37.59227} = 4.10694 \sim 4\ g.l$$

$$M_{21} - M_{22} = (13.6 - 6.86) \mp 2.776\sqrt{\frac{61.30890}{5} + \frac{1.14490}{7}}$$

$$M_{21} - M_{22} = (6.74) \mp 2.776\sqrt{12.26178 + 0.16356}$$

$$M_{21} - M_{22} = (6.74) \mp 2.776\sqrt{12.42534}$$

$$M_{21} - M_{22} = (6.74) \mp 2.776(3.52496)$$

$$M_{21} - M_{22} = (6.74) \mp 9.78529 \begin{cases} LIC = -3.04529 \\ LSC = 16.52529 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la influencia de la zona Norte sobre el tamaño del lote es el mismo que la influencia de la zona Sur sobre el Valor de las casas.

Nota: Como la influencia de la zona Norte sobre el tamaño del lote es el misma que la influencia de la zona Sur sobre el Valor de las casas hubiera sido razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se pudiera haber rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Tamaño vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                       para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      7.829    61.300  (3.319, 30.382)
2        7      1.069     1.143  (0.419,  3.785)

Relación de desviaciones estándar = 7.324
Relación de varianzas = 53.637
```

Prueba T e IC de dos muestras: Tamaño, Niveles

```
T de dos muestras para Tamaño

                               Error
                            estándar
                               de la
Niveles  N  Media  Desv.Est.     media
1        5  13.60       7.83       3.5
2        7   6.86       1.07      0.40
```

```
Diferencia = μ (1) - μ (2)
Estimación de la diferencia:  6.74
IC de 95% para la diferencia:  (-3.04, 16.53)
Prueba T de diferencia = 0 (vs. ≠): Valor T = 1.91  Valor p = 0.128  GL = 4
```

Prueba de Homocedasticidad para Número de Recámaras

1. Juego de Hipótesis

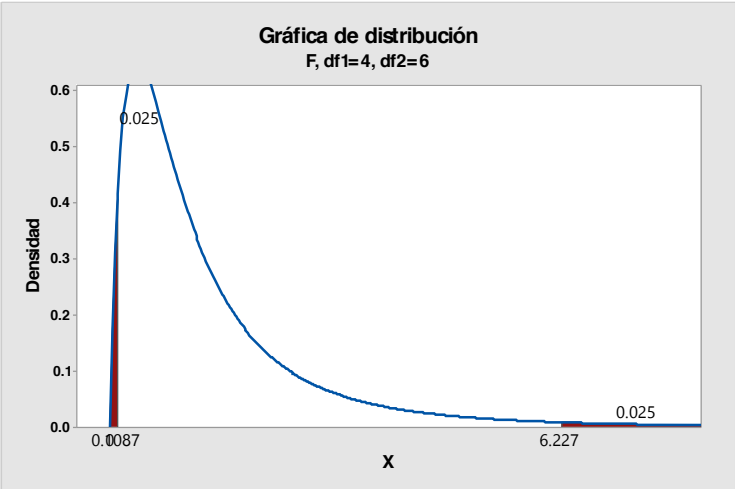
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{31}}^2}{{S_{32}}^2} = \frac{17}{0.61937} = 2.74472$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:

- Estadística: Como 0.1087< 2.74472< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de marca de ambas medias son iguales.

Intervalo de Confianza para Número de Recámaras

$$M_{31} - M_{32} = \left(\overline{X}_{31} - \overline{X}_{32}\right) \mp t^{\propto}/_{2} \cdot n_1 + n_2 - 2 \quad Sp\sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = Sp = \sqrt{\frac{(5-1)1.7 + (7-1)\,0.61937}{5+7-2}}$$

$$Sp = \sqrt{\frac{4*1.7 + 6*0.61937}{10}}$$

$$Sp = \sqrt{\frac{6.8 + 3.71622}{10}}$$

$$Sp = \sqrt{\frac{10.51622}{10}}$$

$$Sp = \sqrt{1.051622}$$

$Sp = 1.02549$

$$M_{31} - M_{32} = (3.2 - 3.429) \mp 2.228 * 1.02549 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{31} - M_{32} = (-0.229) \mp 2.228 * 1.02549 * \sqrt{0.20 + 0.14286}$$

$$M_{31} - M_{32} = (-0.229) \mp 2.228 * 1.02549 * \sqrt{0.34286}$$

$$M_{31} - M_{32} = (-0.229) \mp 2.228 * 1.02549 * 0.58554$$

$$M_{31} - M_{32} = (-0.229) \mp 1.33784 \begin{cases} LIC = -1.56684 \\ LSC = 1.10884 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la influencia de la zona Norte sobre el Número de Recámaras es el mismo que la influencia de la zona Sur sobre el Número de Recámaras de las casas.

Nota: Como la influencia de la zona Norte sobre el Número de Recámaras es el misma que la influencia de la zona Sur sobre el Número de recámaras de las casas hubiera sido razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se pudiera haber rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Número vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      1.304     1.700  (0.685, 4.084)
2        7      0.787     0.619  (0.323, 2.659)
```

```
Relación de desviaciones estándar = 1.657
Relación de varianzas = 2.746
```

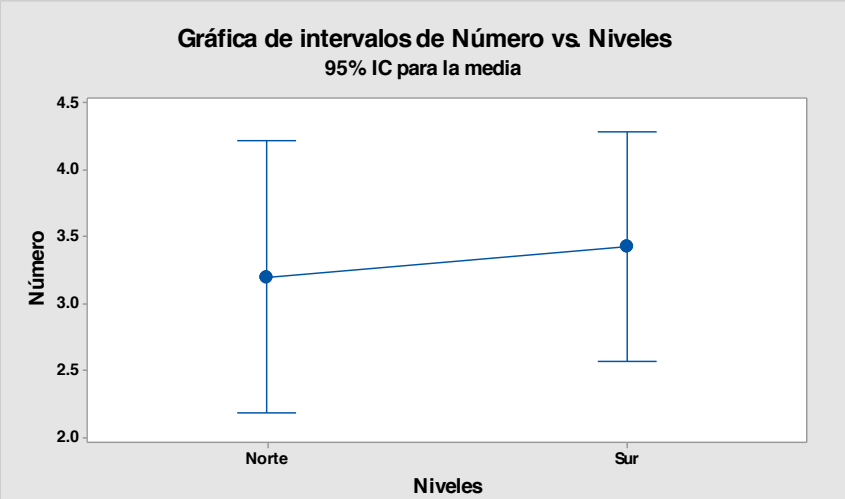
ANOVA unidireccional: Número vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1     0.1524     0.1524     0.14    0.711
Error    10    10.5143     1.0514
Total    11    10.6667


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
1.02539    1.43%       0.00%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  3.200      1.304  (2.178, 4.222)
2        7  3.429      0.787  (2.565, 4.292)

Desv.Est. agrupada = 1.02539
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

           Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            0.229       0.600  (-1.109, 1.566)     0.38     0.711

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Número vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

3.2.2.1.2 EJERCICIO DE REFUERZO

EJERCICIO DE REFUERZO 3.2.2.1.2. Diferencia de Medias Multivariada

3.2.2.1.2. El Director de Mercadotecnia de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia que llega a tener la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares), el tamaño del lote (en pies cuadrados) y número de recámaras. Para este experimento se manejaron dos de los niveles de la ubicación geográfica: **Norte y Este**. Se tomaron lecturas en 5 casas unifamiliares ubicadas en la zona Norte y 7 ubicadas en la zona Este y los resultados fueron los siguientes:

Considere la siguiente muestra extraída de una población normal bivariada denominada **"BIENES1"** donde X_{11}= Valor de avalúo en la zona Norte, X_{12}= Tamaño del lote en la zona Norte, X_{31}= Número de recámaras en la zona Norte, X_{12}= Valor del avalúo en la zona Este, X_{22}= Tamaño del Lote en la zona Este y X_{32}= Número de recámaras en la zona Este:

Nota: El subíndice indica número de variable y zona. Ejemplo: X_{11} indica Variable 1 Valor en zona 1 Norte

ZONA NORTE			**ZONA ESTE**		
X_{11}	X_{21}	X_{31}	X_{12}	X_{22}	X_{32}
Valor	**Tamaño**	**Número**	**Valor**	**Tamaño**	**Número**
265	26	2	180	6	4
245	6	5	152	6	4
210	16	2	168	6	3
239	9	3	180	11	4
299	11	4	163	6	3
			224	6	4
			160	6	4

a) Existe evidencia suficiente que indique que la influencia que llega a tener la ubicación geográfica de las casas unifamiliares de las zonas Norte y Este sobre el valor de las mismas, el tamaño del lote y el número de recámaras es el mismo en ambas ubicaciones?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

1. Juego de Hipótesis

$$H_0: \begin{bmatrix} M_{11} \\ M_{21} \\ M_{31} \end{bmatrix} = \begin{bmatrix} M_{12} \\ M_{22} \\ M_{32} \end{bmatrix}$$

$H_1: Al\ menos\ 1\ combinación\ es\ \neq$

2. Estadístico de prueba

$$Fcalc = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

Donde $T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)^1 [S^2]^{-1} (\widehat{M}_1 - \widehat{M}_2)$

$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2}$

$W_1 = SCPC_1 = D^{'} D_1$

$W_2 = SCPC_2 = D_2^{'} D_2$

$$D_1 = \begin{bmatrix} 265 & 26 & 2 \\ 245 & 6 & 5 \\ 210 & 16 & 2 \\ 239 & 9 & 3 \\ 299 & 11 & 4 \end{bmatrix} - \begin{bmatrix} 251.6 & 13.6 & 3.2 \\ 251.6 & 13.6 & 3.2 \\ 251.16 & 13.6 & 3.2 \\ 251.16 & 13.6 & 3.2 \\ 251.16 & 13.6 & 3.2 \end{bmatrix} = \begin{bmatrix} 13.4 & 12.4 & -1.2 \\ -6.6 & -7.60 & 1.8 \\ -41.6 & 2.4 & -1.2 \\ -12.6 & -4.6 & -0.2 \\ 47.4 & -2.6 & 0.8 \end{bmatrix}$$

$$D_1^{'} D_1 = \begin{bmatrix} 13.4 & -6.6 & -41.6 & -12.6 & 47.4 \\ 12.4 & -7.6 & 2.4 & -4.6 & -2.6 \\ -1.2 & 1.8 & -1.2 & -0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 13.4 & 12.4 & -1.2 \\ -6.6 & -7.60 & 1.8 \\ -41.6 & 2.4 & -1.2 \\ -12.6 & -4.6 & -0.2 \\ 47.4 & -2.6 & 0.8 \end{bmatrix} = \begin{bmatrix} 4359.2 & 51.2 & 62.4 \\ 51.2 & 245.2 & -32.6 \\ 62.4 & -32.6 & 6.8 \end{bmatrix}$$

$$W_1 = D_1^{'} D_1 = \begin{bmatrix} 4359.2 & 51.2 & 62.4 \\ 51.2 & 245.2 & -32.6 \\ 62.4 & -32.6 & 6.8 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 180 & 6 & 4 \\ 152 & 6 & 4 \\ 168 & 6 & 3 \\ 180 & 11 & 4 \\ 163 & 6 & 3 \\ 224 & 6 & 4 \\ 160 & 6 & 4 \end{bmatrix} - \begin{bmatrix} 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \end{bmatrix} = \begin{bmatrix} 4.71 & -0.71 & 0.29 \\ -23.29 & -0.71 & 0.29 \\ -7.29 & -0.71 & -0.71 \\ 4.71 & 4.29 & 0.29 \\ -12.29 & -0.71 & -0.71 \\ 48.71 & -0.71 & 0.29 \\ -15.29 & -0.71 & 0.29 \end{bmatrix}$$

$$D_2^{'} D_2 = \begin{bmatrix} 4.71 & -23.29 & -7.29 & 4.71 & -12.29 & 48.71 & -15.29 \\ -0.71 & -0.71 & -0.71 & 4.29 & -0.71 & -0.71 & -0.71 \\ 0.29 & 0.29 & -0.71 & 0.29 & -0.71 & 0.29 & 0.29 \end{bmatrix} \begin{bmatrix} 4.71 & -0.71 & 0.29 \\ -23.29 & -0.71 & 0.29 \\ -7.29 & -0.71 & -0.71 \\ 4.71 & 4.29 & 0.29 \\ -12.29 & -0.71 & -0.71 \\ 48.71 & -0.71 & 0.29 \\ -15.29 & -0.71 & 0.29 \end{bmatrix} = \begin{bmatrix} 3397.42857 & 23.57143 & 19.57143 \\ 23.57143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}$$

$$W_2 = D_2^{'}D_2 = \begin{bmatrix} 3397.42857 & 23.57143 & 19.57143 \\ 23.57143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}$$

$$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2} = \frac{\begin{bmatrix} 4359.2 & 51.2 & 62.4 \\ 51.2 & 245.2 & -32.6 \\ 62.4 & -32.6 & 6.8 \end{bmatrix} + \begin{bmatrix} 3397.42857 & 23.57143 & 19.57143 \\ 23.57143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}}{5 + 7 - 2}$$

$$= \begin{bmatrix} 775.66286 & 7.47714 & 8.19714 \\ 7.47714 & 26.66286 & -3.11714 \\ 8.19714 & -3.11714 & 0.82286 \end{bmatrix}$$

$$|S^2| = \begin{bmatrix} 775.66286 & 7.47714 & 8.19714 \\ 7.47714 & 26.66286 & -3.11714 \\ 8.19714 & -3.11714 & 0.82286 \\ 775.66286 & 7.47714 & 8.19714 \\ 7.47714 & 26.66286 & -3.11714 \end{bmatrix} = 7261.36560$$

Como es ≠ de 0 tiene solución

Transpuesta de S^2

$$[S^2]^1 = \begin{bmatrix} 7756.62857 & 74.77143 & 81.97143 \\ 74.77143 & 266.62857 & -31.17143 \\ 81.97143 & -31.17143 & 8.22857 \end{bmatrix}$$

Matriz de cofactores

$$\alpha A^{'} = \begin{bmatrix} +\begin{vmatrix} 26.66286 & -3.11714 \\ -3.11714 & 0.82286 \end{vmatrix} & -\begin{vmatrix} 7.47714 & -3.11714 \\ 8.19714 & 0.82286 \end{vmatrix} & +\begin{vmatrix} 7.47714 & 26.66286 \\ 8.19714 & -3.11714 \end{vmatrix} \\ -\begin{vmatrix} 7.47714 & 8.19714 \\ -3.11714 & 0.82286 \end{vmatrix} & +\begin{vmatrix} 775.66286 & 8.19714 \\ 8.19714 & 0.82286 \end{vmatrix} & -\begin{vmatrix} 775.66286 & 7.47714 \\ 8.19714 & -3.11714 \end{vmatrix} \\ +\begin{vmatrix} 7.47714 & 8.19714 \\ 26.66286 & -3.11714 \end{vmatrix} & -\begin{vmatrix} 775.66286 & 8.19714 \\ 7.47714 & -3.11714 \end{vmatrix} & +\begin{vmatrix} 775.66286 & 7.47714 \\ 7.47714 & 26.66286 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(5.14567 - 0.13167) & -(1.02429 - 0.62515) & +(0.13167 - 3.14052) \\ -(1.02429 - 0.62515) & +(6.27481 - 2.96824) & -(0.80658 - 0.62515) \\ +((0.13167 - 3.14052) & -(0.80658 - 0.62515) & +(4.05195 - 0.13167) \end{bmatrix} =$$

$$= \begin{bmatrix} 12.22314 & -31.70429 & -241.86657 \\ -31.70429 & 571.06657 & 2479.14314 \\ -241.86657 & 2479.14314 & 20625.48029 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{7261.36560} = \begin{bmatrix} 12.22314 & -31.70429 & -241.86657 \\ -31.70429 & 571.06657 & 2479.14314 \\ -241.86657 & 2479.14314 & 20625.48029 \end{bmatrix}$$

$$= \begin{bmatrix} 0.00168 & -0.00437 & -0.03331 \\ -0.00437 & 0.07864 & 0.34142 \\ -0.03331 & 0.34142 & 2.84044 \end{bmatrix}$$

$$T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)'[S^2]^{-1}(\widehat{M}_1 - \widehat{M}_2)$$

$$T^2 = \left(\frac{5 * 7}{5 + 7}\right)(251.6 - 175.29; 13.6 - 6.71; 3.2 - 3.71)[S^2]^{-1} \begin{bmatrix} 251.6 - 175.29 \\ 13.6 - 6.71 \\ 3.2 - 3.71 \end{bmatrix}$$

$$T^2 = \left(\frac{35}{12}\right)(76.31429; 6.88571; -0.51429)[S^2]^{-1} \begin{bmatrix} 76.31429 \\ 6.88571 \\ 0.51429 \end{bmatrix}$$

$$T^2 = (2.91667)(66.89; 6.74; 0.17)[S^2]^{-1} \begin{bmatrix} 76.31429 \\ 6.88571 \\ 0.51429 \end{bmatrix}$$

$$T^2 = [22.58333; 20.08333; -1.5] \begin{bmatrix} 0.00168 & -0.00437 & -0.03331 \\ -0.00437 & 0.07864 & 0.34142 \\ -0.03331 & 0.34142 & 2.84044 \end{bmatrix} \begin{bmatrix} 76.31429 \\ 6.88571 \\ 0.51429 \end{bmatrix}$$

$$T^2 = [0.33695; 0.09549; -4.81786] \begin{bmatrix} 76.31429 \\ 6.88571 \\ 0.51429 \end{bmatrix}$$

$T^2 = \underline{28.84958}$

$$Fcalc = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2$$

$$Fcalc = \frac{5+7-3-1}{3(5+7-2)}(28.84958)$$

$$Fcalc = \frac{8}{30}(28.84958)$$

$$Fcalc = 0.26667 \times 28.84958$$

$$\boldsymbol{Fcalc = 7.87}$$

3. Región de rechazo

$$F_{\alpha.p.n1+n2-p-1} = F_{0.05.3.5+7-3-1} = F_{0.05.3.8} = 4.07$$

$$F_{\alpha.p.n1+n2-p-1} = F_{0.01.3.5+7-3-1} = F_{0.01.3.8} = 7.59$$

4. Regla de decisión

Se rechaza H_0 si Fcalc > 4.07

5. Conclusión:

- Estadistica: Como 7.87>4.07 y >7.59 ∴

La prueba es A.S y se rechaza H_0 con mucha potencia.

- Adva: Existe evidencia suficiente para decir que la influencia que llega a tener la ubicación geográfica de las casas sobre el valor del avalúo, el tamaño del lote y el número de recámaras no es el mismo en ambas zonas estudiadas en al menos una combinación de ellas.

Resolución del inciso b)

Nota importante: Para construir el intervalo de confianza se debe probar primero si las varianzas de las muestras son iguales o no para saber que caso de debe usar para construir el intervalo de confianza (Caso 2 para varianzas iguales o Caso 3 para varianzas desiguales) por lo que primero hay que realizar la prueba de homocedasticidad de varianzas.

Prueba de Homocedasticidad para Valor

1. Juego de Hipótesis

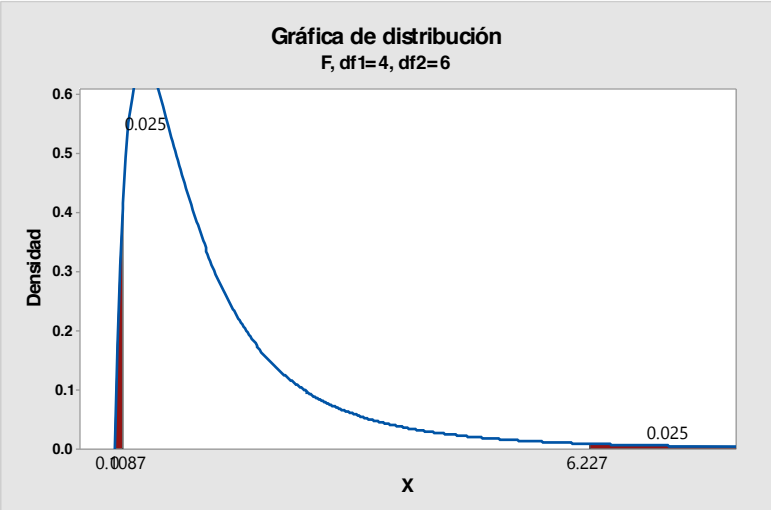
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{11}}^2}{{S_{12}}^2} = \frac{1089.8}{566.44} = 1.92395$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc ≥ 6.23 ó si es ≤ 0.1087

5. Conclusión:

- Estadistica: Como 0.1087< 1.92395< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de Valor de ambas medias son iguales.

Intervalo de Confianza para Valor

$$M_{11} - M_{12} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp (t_{\alpha/2} \cdot n_1 + n_2 - 2 \quad Sp \sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)1089.8 + (7-1)566.44}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 1089.8 + 6 * 566.64}{10}}$$

$$Sp = \sqrt{\frac{4359.2 + 3398.64}{10}}$$

$$Sp = \sqrt{\frac{7757.84}{10}}$$

$$Sp = \sqrt{775.784}$$

$$Sp = 27.8529$$

$$M_{11} - M_{12} = (251.6 - 175.29) \mp 2.228 * 27.8529 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (76.31) \mp 2.228 * 27.8529 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (76.31) \mp 2.228 * 27.8529 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (76.31) \mp 2.228 * 27.8529 * 0.58554$$

$$M_{11} - M_{12} = (76.31) \mp 36.33642 \begin{cases} LIC = 40 \\ LSC = 112.7 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la influencia de la zona Norte sobre el valor de las casas es mayor que la influencia de la zona Este por un mínimo de 40 y un máximo de 112.7 miles de dólares.

Nota: Como la influencia de la zona sobre el **Valor** de las casas es diferente en ambas zonas sería razonable que la prueba de hipótesis de diferencia de medias multivariada se hubiera rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Valor vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05
```

```
Estadísticas

                                    IC de 95% para
Niveles  N  Desv.Est.  Varianza      Desv.Est.
1        5     33.012  1089.800  (14.660, 122.268)
2        7     23.796   566.238  ( 8.569,  91.779)

Relación de desviaciones estándar = 1.387
Relación de varianzas = 1.925
```

ANOVA unidireccional: Valor vs. Niveles

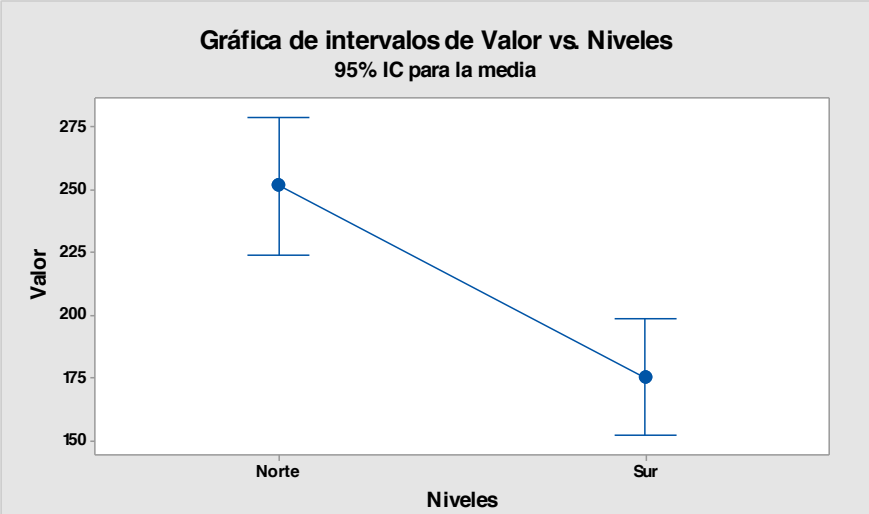
```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      16986    16986.3    21.90    0.001
Error    10       7757      775.7
Total    11      24743


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
27.8507   68.65%      65.52%   53.78%


Medias

Niveles  N   Media  Desv.Est.      IC de 95%
1        5   251.6       33.0  ( 223.8,  279.4)
2        7  175.29      23.80  (151.83, 198.74)

Desv.Est. agrupada = 27.8507
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
```

```
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            -76.3        16.3  (-112.7, -40.0)    -4.68     0.001

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Valor vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

Prueba T e IC de dos muestras: Valor, Niveles

```
T de dos muestras para Valor

                              Error
                           estándar
                              de la
Niveles  N  Media  Desv.Est.  media
1        5  251.6       33.0     15
2        7  175.3       23.8    9.0


Diferencia = μ (1) - μ (2)
Estimación de la diferencia:  76.3
IC de 95% para la diferencia:  (40.0, 112.7)
Prueba T de diferencia = 0 (vs. ≠): Valor T = 4.68  Valor p = 0.001  GL = 10
Ambos utilizan Desv.Est. agrupada = 27.8507
```

Prueba de Homocedasticidad para Tamaño

1. Juego de Hipótesis

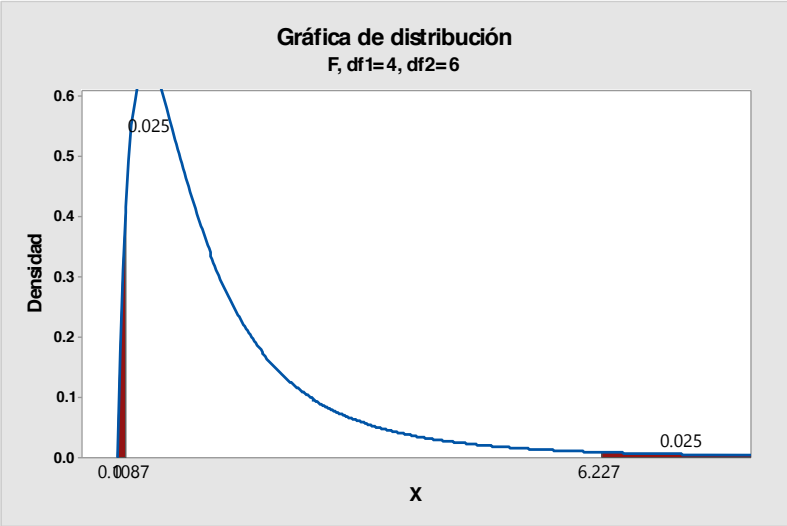
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{21}}^2}{{S_{22}}^2} = \frac{61.30890}{3.27210} = 17.16327$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:
   - Estadística: Como 17.16327>6.23 ∴ se rechaza H_0 y se utiliza el caso 3.
   - Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas del Tamaño del lote de ambas medias son diferentes.

Intervalo de Confianza para Tamaño

$$M_{21} - M_{22} = \left(\overline{X}_{21} - \overline{X}_{22}\right) \mp t^{\propto}/_{2} \cdot v\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Donde (v), los grados de libertad se determinan de la siguiente manera:

$$v = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}}$$

$$v = \frac{\left[\frac{61.30890}{5} + \frac{3.57210}{7}\right]^2}{\frac{\left[\frac{61.30890}{5}\right]^2}{4} + \frac{\left[\frac{3.57210}{7}\right]^2}{6}} = \frac{[12.26178 + 0.51030]^2}{\frac{150.35125}{4} + \frac{0.26041}{6}} =$$

$$\frac{163.12603}{37.58781+0.04340} = \frac{163.12603}{37.63121} = 4.33486 \sim 4\ g.l$$

$$M_{21} - M_{22} = (13.6 - 6.71) \mp 2.776\sqrt{\frac{61.30890}{5} + \frac{3.57210}{7}}$$

$$M_{21} - M_{22} = (6.89) \mp 2.776\sqrt{12.26178 + 0.51030}$$

$$M_{21} - M_{22} = (6.89) \mp 2.776\sqrt{12.77208}$$

$$M_{21} - M_{22} = (6.89) \mp 2.776(3.57380)$$

$$M_{21} - M_{22} = (6.89) \mp 9.92088 \begin{cases} LIC = -3.03088 \\ LSC = 16.81088 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la influencia de la zona Norte sobre el tamaño del lote es el mismo que la influencia de la zona Este sobre el Valor de las casas.

Nota: Como la influencia de la zona Norte sobre el Tamaño del lote es el misma que la influencia de la zona Este sobre el Tamaño del lote de las casas hubiera sido razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se pudiera haber rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Tamaño vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                       para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      7.829    61.300  (3.319, 30.382)
2        7      1.890     3.571  (0.422, 11.748)

Relación de desviaciones estándar = 4.143
Relación de varianzas = 17.164
```

Prueba T e IC de dos muestras: Tamaño, Niveles

```
T de dos muestras para Tamaño

                               Error
                            estándar
                               de la
Niveles  N  Media  Desv.Est.   media
1        5  13.60       7.83     3.5
2        7   6.71       1.89    0.71


Diferencia = μ (1) - μ (2)
Estimación de la diferencia:  6.89
IC de 95% para la diferencia:  (-3.04, 16.81)
```

```
Prueba T de diferencia = 0 (vs. ≠): Valor T = 1.93  Valor p = 0.126  GL = 4
```

Prueba de Homocedasticidad para Número de Recámaras

1. Juego de Hipótesis

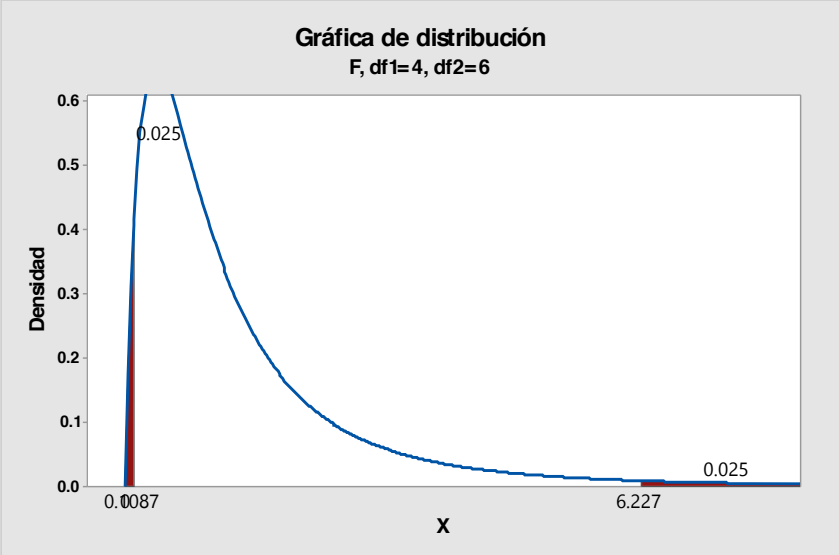
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{31}}^2}{{S_{32}}^2} = \frac{1.7}{0.238} = 7.140$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:
   - Estadística: Como 7.14>6.23 ∴ se rechaza H_0 y se utiliza el caso 3.
   - Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas del Número de Recamaras de ambas medias son diferentes.

Intervalo de Confianza para Número de Recámaras

$$M_{31} - M_{32} = \left(\overline{X}_{31} - \overline{X}_{32}\right) \mp t^{\propto}/_{2} \cdot v\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Donde (v), los grados de libertad se determinan de la siguiente manera:

$$v = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}}$$

$$v = \frac{\left[\frac{1.7}{5} + \frac{0.238}{7}\right]^2}{\frac{\left[\frac{1.7}{5}\right]^2}{4} + \frac{\left[\frac{0.238}{7}\right]^2}{6}} = \frac{[0.34 + 0.034]^2}{\frac{0.11560}{4} + \frac{0.00116}{6}} =$$

$$\frac{0.13988}{0.0289+0.00019} = \frac{0.13989}{0.02909} = 4.8 \sim 4\ g.l$$

$$M_{31} - M_{32} = (3.2 - 3.71) \mp 2.776\sqrt{\frac{1.7}{5} + \frac{0.238}{7}}$$

$$M_{31} - M_{32} = (-0.51) \mp 2.776\sqrt{0.34 + 0.034}$$

$$M_{31} - M_{32} = (-0.51) \mp 2.776\sqrt{0.374}$$

$$M_{31} - M_{32} = (-0.51) \mp 2.776(0.61156)$$

$$M_{31} - M_{32} = (-0.51) \mp 1.69768 \begin{cases} LIC = -2.20768 \\ LSC = 1.18768 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero, podemos decir que estadísticamente la influencia de la zona Norte sobre el Número de Recámaras es el mismo que la influencia de la zona Este sobre el Número de Recámaras

Nota: Como la influencia de la zona Norte sobre el Número de Recámaras es el misma que la influencia de la zona Este sobre el Número de Recámaras de las casas hubiera sido razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se pudiera haber rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Número vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      1.304     1.700  (0.685, 4.084)
2        7      0.488     0.238  (0.143, 2.317)

Relación de desviaciones estándar = 2.672
Relación de varianzas = 7.140
```

Prueba T e IC de dos muestras: Número, Niveles

```
T de dos muestras para Número

                               Error
                            estándar
                               de la
Niveles  N  Media  Desv.Est.     media
1        5   3.20       1.30      0.58
2        7  3.714      0.488      0.18


Diferencia = μ (1) - μ (2)
Estimación de la diferencia:  -0.514
IC de 95% para la diferencia:  (-2.212, 1.184)
Prueba T de diferencia = 0 (vs. ≠): Valor T = -0.84  Valor p = 0.448  GL = 4
```

3.2.2.1.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.2.1.3. Diferencia de Medias Multivariada**

3.2.2.1.3. El Director de Mercadotecnia de una compañía de Bienes Raíces desea realizar un experimento para investigar la posible influencia que llega a tener la ubicación geográfica de las casas unifamiliares de una encuesta denominada **"BIENES1"** sobre valor del avalúo de las casas unifamiliares (en miles de dólares), el tamaño del lote (en pies cuadrados) y número de recámaras. Para este experimento se manejaron dos de los niveles de la ubicación geográfica: **Sur y Este**. Se tomaron lecturas en 5 casas unifamiliares ubicadas en la zona Sur y 7 ubicadas en la zona Este y los resultados fueron los siguientes:

Considere la siguiente muestra extraída de una población normal bivariada denominada **"BIENES1"** donde X_{11}= Valor de avalúo en la zona Sur, X_{12}= Tamaño del lote en la zona Sur, X_{31}= Número de recámaras en la zona Sur, X_{12}= Valor del avalúo en la zona Este, X_{22}= Tamaño del Lote en la zona Este y X_{32}= Número de recámaras en la zona Este:

Nota: El subíndice indica número de variable y zona. Ejemplo: X_{11} indica Variable 1 Valor en zona 1 Sur

ZONA SUR			**ZONA ESTE**		
X_{11}	**X_{21}**	**X_{31}**	**X_{12}**	**X_{22}**	**X_{32}**
Valor	**Tamaño**	**Número**	**Valor**	**Tamaño**	**Número**
190	7	4	180	6	4
160	6	2	152	6	4
176	6	4	168	6	3
181	7	3	180	11	4
185	6	3	163	6	3
			224	6	4
			160	6	4

a) Existe evidencia suficiente que indique que la influencia que llega a tener la ubicación geográfica de las casas unifamiliares de las zonas Sur y Este sobre el valor de las mismas, el tamaño del lote y el número de recámaras es el mismo en ambas ubicaciones?. Pruébelo.
b) Utilice el método del intervalo de confianza para probar la hipótesis nula para cada par de variables en forma aislada y verifique si existen diferencias significativas entre las medias muestrales y su magnitud.

Nota importante: Los cálculos para la resolución de los incisos se realizó con Excel y Minitab por lo que puede haber variaciones en los resultados. Se sugiere trabajar en la calculadora con cinco decimales para que la diferencia sea mínima.

Resolución del inciso a)

Prueba de hipótesis

6. Juego de Hipótesis

$$H_0: \begin{bmatrix} M_{11} \\ M_{21} \\ M_{31} \end{bmatrix} = \begin{bmatrix} M_{12} \\ M_{22} \\ M_{32} \end{bmatrix}$$

$$H_1: Al\ menos\ 1\ combinación\ es\ \neq$$

7. Estadístico de prueba

$$Fcalc = \frac{n_1 + n_2 - p - 1}{p(n_1 + n_2 - 2)} T^2$$

Donde $T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)^1 [S^2]^{-1} (\widehat{M}_1 - \widehat{M}_2)$

$$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2}$$

$$W_1 = SCPC_1 = D^{'} D_1$$

$$W_2 = SCPC_2 = D_2^{'} D_2$$

$$D_1 = \begin{bmatrix} 190 & 7 & 4 \\ 160 & 6 & 2 \\ 176 & 6 & 4 \\ 181 & 7 & 3 \\ 185 & 6 & 3 \end{bmatrix} - \begin{bmatrix} 178.4 & 6.4 & 3.2 \\ 178.4 & 6.4 & 3.2 \\ 178.4 & 6.4 & 3.2 \\ 178.4 & 6.4 & 3.2 \\ 178.4 & 6.4 & 3.2 \end{bmatrix} = \begin{bmatrix} 11.6 & 0.6 & 0.8 \\ -18.4 & -0.4 & -1.2 \\ -2.4 & -0.4 & 0.8 \\ 2.6 & 0.6 & -0.2 \\ 6.6 & -0.4 & -0.2 \end{bmatrix}$$

$$D_1'D_1 = \begin{bmatrix} 11.6 & -18.4 & -2.4 & 2.6 & 6.6 \\ 0.6 & -0.4 & -0.4 & 0.6 & -0.4 \\ 0.8 & -1.2 & 0.8 & -0.2 & -0.2 \end{bmatrix} \begin{bmatrix} 11.6 & 0.6 & 0.8 \\ -18.4 & -0.4 & -1.2 \\ -2.4 & -0.4 & 0.8 \\ 2.6 & 0.6 & -0.2 \\ 6.6 & -0.4 & -0.2 \end{bmatrix} = \begin{bmatrix} 529.2 & 14.2 & 27.6 \\ 14.2 & 1.2 & 0.6 \\ 27.6 & 0.6 & 2.8 \end{bmatrix}$$

$$W_1 = D_1'D_1 = \begin{bmatrix} 529.2 & 14.2 & 27.6 \\ 14.2 & 1.2 & 0.6 \\ 27.6 & 0.6 & 2.8 \end{bmatrix}$$

$$D_2 = \begin{bmatrix} 180 & 6 & 4 \\ 152 & 6 & 4 \\ 168 & 6 & 3 \\ 180 & 11 & 4 \\ 163 & 6 & 3 \\ 224 & 6 & 4 \\ 160 & 6 & 4 \end{bmatrix} - \begin{bmatrix} 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \\ 175.29 & 6.71 & 3.71 \end{bmatrix} = \begin{bmatrix} 4.71 & -0.71 & 0.29 \\ -23.29 & -0.71 & 0.29 \\ -7.29 & -0.71 & -0.71 \\ 4.71 & 4.29 & 0.29 \\ -12.29 & -0.71 & -0.71 \\ 48.71 & -0.71 & 0.29 \\ -15.29 & -0.71 & 0.29 \end{bmatrix}$$

$$D_2'D_2 = \begin{bmatrix} 4.71 & -23.29 & -7.29 & 4.71 & -12.29 & 48.71 & -15.29 \\ -0.71 & -0.71 & -0.71 & 4.29 & -0.71 & -0.71 & -0.71 \\ 0.29 & 0.29 & -0.71 & 0.29 & -0.71 & 0.29 & 0.29 \end{bmatrix} \begin{bmatrix} 4.71 & -0.71 & 0.29 \\ -23.29 & -0.71 & 0.29 \\ -7.29 & -0.71 & -0.71 \\ 4.71 & 4.29 & 0.29 \\ -12.29 & -0.71 & -0.71 \\ 48.71 & -0.71 & 0.29 \\ -15.29 & -0.71 & 0.29 \end{bmatrix} = \begin{bmatrix} 3397.42857 & 23.57143 & 19.57143 \\ 23.57143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}$$

$$W_2 = D_2'D_2 = \begin{bmatrix} 3397.42857 & 23.57143 & 19.57143 \\ 23.57143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}$$

$$S^2 = \frac{W_1 + W_2}{n_1 + n_2 - 2} = \frac{\begin{bmatrix} 529.2 & 14.2 & 27.6 \\ 14.2 & 1.2 & 0.6 \\ 27.6 & 0.6 & 2.8 \end{bmatrix} + \begin{bmatrix} 3397.42857 & 23.57143 & 19.57143 \\ 23.57143 & 21.42857 & 1.42857 \\ 19.57143 & 1.42857 & 1.42857 \end{bmatrix}}{5 + 7 - 2}$$

$$= \begin{bmatrix} 392.66286 & 3.77714 & 4.71714 \\ 3.77714 & 2.26286 & 0.20286 \\ 4.71714 & 0.20286 & 0.42286 \end{bmatrix}$$

Aplicando la regla de Sarrus:

$$|S^2| = \begin{bmatrix} 392.66286 & 3.77714 & 4.71714 \\ 3.77714 & 2.26286 & 0.20286 \\ 4.71714 & 0.20286 & 0.42286 \\ 392.66286 & 3.77714 & 4.71714 \\ 3.77714 & 2.26286 & 0.20286 \end{bmatrix} = 310.41109$$

Como es ≠ de 0 tiene solución

Transpuesta de S^2

$$[S^2]^1 = \begin{bmatrix} 392.66286 & 3.77714 & 4.71714 \\ 3.77714 & 2.26286 & 0.20286 \\ 4.71714 & 0.20286 & 0.42286 \end{bmatrix}$$

Matriz de cofactores

$$\alpha A' = \begin{bmatrix} +\begin{vmatrix} 2.26286 & 0.20286 \\ 0.20286 & 0.42286 \end{vmatrix} & -\begin{vmatrix} 3.77714 & 0.20286 \\ 4.71714 & 0.42286 \end{vmatrix} & +\begin{vmatrix} 3.77714 & 2.26286 \\ 4.71714 & 0.20286 \end{vmatrix} \\ -\begin{vmatrix} 3.77714 & 4.71714 \\ 0.20286 & 0.42286 \end{vmatrix} & +\begin{vmatrix} 392.66286 & 4.71714 \\ 4.71714 & 0.42286 \end{vmatrix} & -\begin{vmatrix} 392.66286 & 3.77714 \\ 4.71714 & 0.20286 \end{vmatrix} \\ +\begin{vmatrix} 3.77714 & 4.71714 \\ 2.26286 & 0.20286 \end{vmatrix} & -\begin{vmatrix} 392.66286 & 4.71714 \\ 3.77714 & 0.20286 \end{vmatrix} & +\begin{vmatrix} 392.66286 & 3.77714 \\ 3.77714 & 2.26286 \end{vmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} +(0.95687 - 0.04115) & -(1.59719 - 0.95691) & +(0.76622 - 10.67422) \\ -(1.59719 - 0.95691) & +(166.04029 - 22.25144) & -(79.65447 - 17.81732) \\ +((0.76622 - 10.67422) & -(79.65447 - 17.81732) & +(888.53995 - 14.26681) \end{bmatrix} =$$

$$= \begin{bmatrix} 0.91571 & -0.64029 & -9.90800 \\ -0.64029 & 143.78886 & -61.83714 \\ -9.90800 & -61.83714 & 874.27.14 \end{bmatrix}$$

Matriz inversa

$$[S^2]^{-1} = \frac{1}{310.41109} = \begin{bmatrix} 0.91571 & -0.64029 & -9.90800 \\ -0.64029 & 143.78886 & -61.83714 \\ -9.90800 & -61.83714 & 874.27.14 \end{bmatrix}$$

$$= \begin{bmatrix} 00295 & -0.00206 & -0.03192 \\ -0.00206 & 0.46322 & -0.19921 \\ -0.03192 & -0.19921 & 2.81650 \end{bmatrix}$$

$$T^2 = \left(\frac{n_1 n_2}{n_1 + n_2}\right)(\widehat{M}_1 - \widehat{M}_2)'[S^2]^{-1}(\widehat{M}_1 - \widehat{M}_2)$$

$$T^2 = \left(\frac{5*7}{5+7}\right)(178.4-175.29; -6.40-6.71; 3.2-3.71)[S^2]^{-1}\begin{bmatrix}178.4-175.29\\6.4-6.71\\3.2-3.71\end{bmatrix}$$

$$T^2 = \left(\frac{35}{12}\right)(3.11429; -0.31429; -0.51429)[S^2]^{-1}\begin{bmatrix}3.11429\\-0.31429\\-0.51429\end{bmatrix}$$

$$T^2 = (2.91667)(3.11429; -0.31429; -0.51429)[S^2]^{-1}\begin{bmatrix}3.11429\\-0.31429\\-0.51429\end{bmatrix}$$

$$T^2 = [9.08333; -0.91667; -1.5]\begin{bmatrix}00295 & -0.00206 & -0.03192\\-0.00206 & 0.46322 & -0.19921\\-0.03192 & -0.19921 & 2.81650\end{bmatrix}\begin{bmatrix}3.11429\\-0.31429\\-0.51429\end{bmatrix}$$

$$T^2 = [0.0766; -0.14454; -4.33207]\left[\begin{bmatrix}3.11429\\-0.31429\\-0.51429\end{bmatrix}\right]$$

$$T^2 = \underline{2.5118}$$

$$Fcalc = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T^2$$

$$Fcalc = \frac{5+7-3-1}{3(5+7-2)}(28.84958)$$

$$Fcalc = \frac{8}{30}(2.5118)$$

$$Fcalc = 0.26667 \times 2.5118$$

$$\boldsymbol{Fcalc = 0.685}$$

8. Región de rechazo

$$F_{\alpha.p.n1+n2-p-1} = F_{0.05.3.5+7-3-1} = F_{0.05.3.8} = 4.07$$

$$F_{\alpha.p.n1+n2-p-1} = F_{0.01.3.5+7-3-1} = F_{0.01.3.8} = 7.59$$

9. Regla de decisión

Se rechaza H_0 si Fcalc > 4.07

10. Conclusión:

- Estadistica: Como 0.685<4.07 ∴

La prueba es (N.S) y NO se rechaza H_0

- Adva: Existe evidencia suficiente para decir que la influencia que llega a tener la ubicación geográfica de las casas sobre el valor del avalúo, el tamaño del lote y el número de recámaras es el mismo en ambas zonas estudiadas.

Resolución del inciso b)

Nota importante: Para construir el intervalo de confianza se debe probar primero si las varianzas de las muestras son iguales o no para saber que caso de debe usar para construir el intervalo de confianza (Caso 2 para varianzas iguales o Caso 3 para varianzas desiguales) por lo que primero hay que realizar la prueba de homocedasticidad de varianzas.

Prueba de Homocedasticidad para Valor

1. Juego de Hipótesis

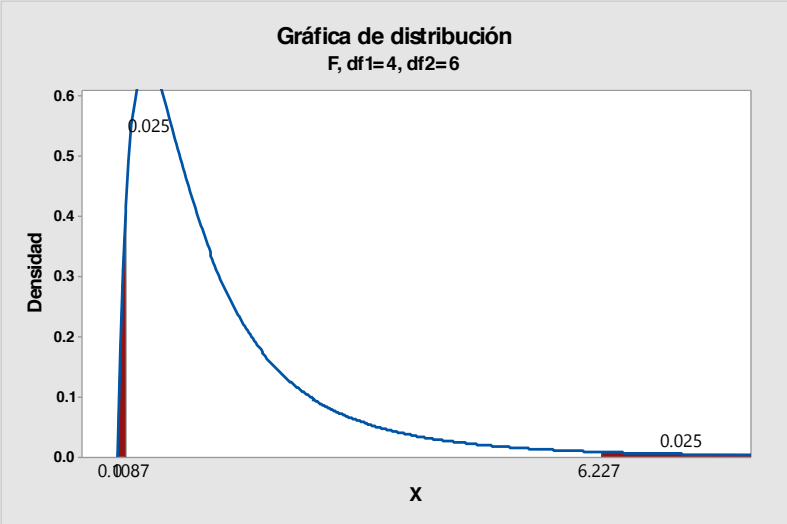
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{11}}^2}{{S_{12}}^2} = \frac{132.25}{566.44} = 0.23348$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc ≥ 6.23 ó si es ≤ 0.1087

5. Conclusión:

- Estadistica: Como 0.1087< 0.23348< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de Valor de ambas medias son iguales.

Intervalo de Confianza para Valor

$$M_{11} - M_{12} = \left(\overline{X}_{11} - \overline{X}_{12}\right) \mp (t_{\alpha/2} \cdot n_1 + n_2 - 2 \quad Sp \sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)132.25 + (7-1)566.44}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 132.25 + 6 * 566.64}{10}}$$

$$Sp = \sqrt{\frac{529 + 3398.64}{10}}$$

$$Sp = \sqrt{\frac{3927.64}{10}}$$

$$Sp = \sqrt{392.764}$$

$$Sp = 19.81827$$

$$M_{11} - M_{12} = (178.4 - 175.29) \mp 2.228 * 19.81827 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{11} - M_{12} = (3.11) \mp 2.228 * 19.81827 * \sqrt{0.20 + 0.14286}$$

$$M_{11} - M_{12} = (3.11) \mp 2.228 * 19.81827 * \sqrt{0.34286}$$

$$M_{11} - M_{12} = (3.11) \mp 2.228 * 19.81827 * 0.58554$$

$$M_{11} - M_{12} = (3.11) \mp 25.85458 \begin{cases} LIC = -22.7 \\ LSC = 29 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la influencia de la zona Sur es igual que la influencia de la zona Este sobre el valor de las casas.

Nota: Como la influencia de la zona sobre el **Valor** de las casas es el mismo en ambas zonas sería razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se hubiera rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Valor vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                       para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5     11.502   132.300  (4.712, 46.176)
2        7     23.796   566.238  (8.569, 91.779)

Relación de desviaciones estándar = 0.483
Relación de varianzas = 0.234
```

ANOVA unidireccional: Valor vs. Niveles

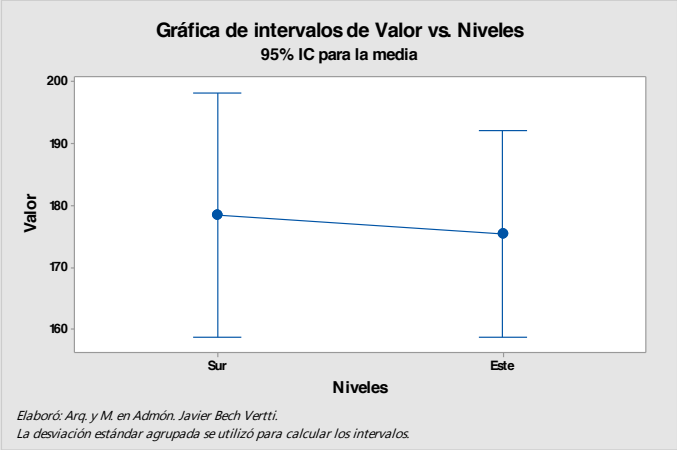
```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.
```

```
Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1      28.29      28.29     0.07    0.794
Error    10    3926.63     392.66
Total    11    3954.92


Resumen del modelo

                    R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
19.8157    0.72%       0.00%    0.00%


Medias

Niveles  N   Media  Desv.Est.      IC de 95%
1        5  178.40      11.50  (158.65, 198.15)
2        7  175.29      23.80  (158.60, 191.97)

Desv.Est. agrupada = 19.8157
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

           Diferencia
Diferencia      de las       EE de                          Valor p
de niveles      medias  diferencia    IC de 95%    Valor T  ajustado
2 - 1             -3.1        11.6  (-29.0, 22.7)    -0.27     0.794

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Valor vs. Niveles



Prueba de Homocedasticidad para Tamaño

1. Juego de Hipótesis

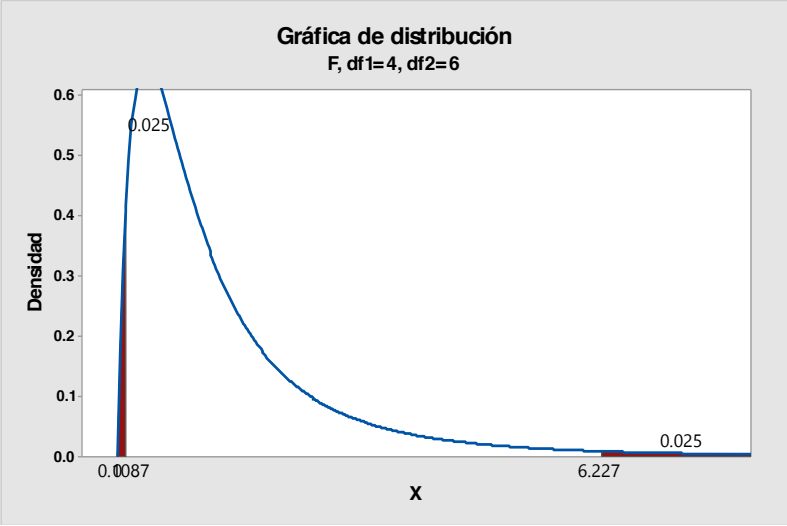
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{21}}^2}{{S_{22}}^2} = \frac{0.30030}{3.57210} = 0.08407$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:
   - Estadística: Como 0.08407<0.1087 ∴ se rechaza H_0 y se utiliza el caso 3.
   - Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas del Tamaño del lote de ambas medias son diferentes.

Intervalo de Confianza para Tamaño

$$M_{21} - M_{22} = \left(\overline{X}_{21} - \overline{X}_{22}\right) \mp t^{\propto}/_{2} \cdot v\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Donde (v), los grados de libertad se determinan de la siguiente manera:

$$v = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right]^2}{\frac{\left[\frac{S_1^2}{n_1}\right]^2}{n_1 - 1} + \frac{\left[\frac{S_2^2}{n_2}\right]^2}{n_2 - 1}}$$

$$v = \frac{\left[\frac{0.30030}{5} + \frac{3.57210}{7}\right]^2}{\frac{\left[\frac{0.30030}{5}\right]^2}{4} + \frac{\left[\frac{3.57210}{7}\right]^2}{6}} = \frac{[0.06006 + 0.51030]^2}{\frac{0.00361}{4} + \frac{0.26041}{6}} =$$

$$\frac{0.32531}{0.0009+0.0434} = \frac{0.32531}{0.04430} = 7.34334 \sim 7\ g.l$$

$$M_{21} - M_{22} = (6.4 - 6.71) \mp 2.365\sqrt{\frac{0.30030}{5} + \frac{3.57210}{7}}$$

$$M_{21} - M_{22} = (-0.31) \mp 2.365\sqrt{0.06006 + 0.51030}$$

$$M_{21} - M_{22} = (-0.31) \mp 2.365\sqrt{0.57036}$$

$$M_{21} - M_{22} = (-0.31) \mp 2.365(0.75522)$$

$$M_{21} - M_{22} = (-0.31) \mp 1.78610 \begin{cases} LIC = -2.1 \\ LSC = 1.47 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la influencia de la zona Sur es igual que la influencia de la zona Este sobre el Tamaño del lote de las casas.

Nota: Como la influencia de la zona sobre el **Tamaño del lote** de las casas es el mismo en ambas zonas sería razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se hubiera rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Tamaño vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05


Estadísticas

                                    IC de 95%
                                       para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      0.548     0.300  (0.347,  1.423)
2        7      1.890     3.571  (0.422, 11.748)

Relación de desviaciones estándar = 0.290
Relación de varianzas = 0.084
```

Prueba T e IC de dos muestras: Tamaño, Niveles

```
T de dos muestras para Tamaño

                               Error
                            estándar
                               de la
Niveles  N  Media  Desv.Est.   media
```

```
1        5   6.400      0.548      0.24
2        7    6.71       1.89      0.71


Diferencia = μ (1) - μ (2)
Estimación de la diferencia:  -0.314
IC de 95% para la diferencia:  (-2.100, 1.471)
Prueba T de diferencia = 0 (vs. ≠): Valor T = -0.42  Valor p = 0.690  GL = 7
```

Prueba de Homocedasticidad para Número de Recámaras

1. Juego de Hipótesis

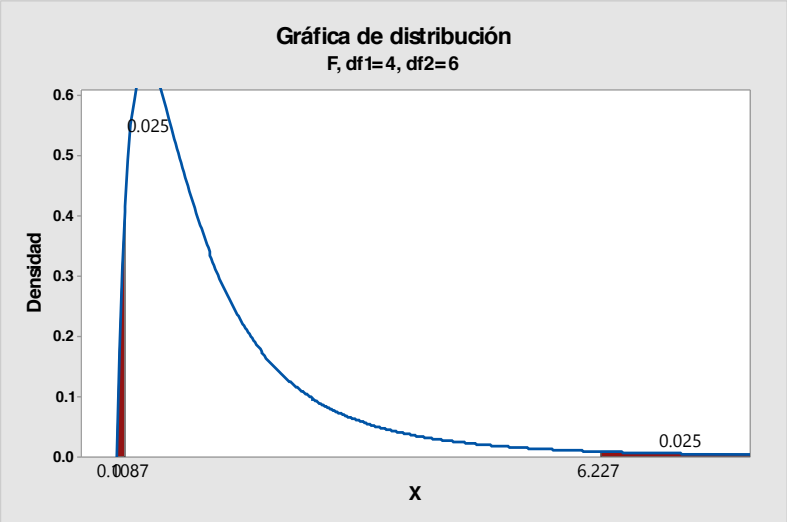
$$H_0: \frac{{\sigma_1}^2}{{\sigma_2}^2} = 1; ({\sigma_1}^2 = {\sigma_2}^2) homoscedasticidad \ "Caso\ 2"$$

$$H_1: \frac{{\sigma_1}^2}{{\sigma_2}^2} \neq 1; ({\sigma_1}^2 \neq {\sigma_2}^2) heteroscedasticos \ "Caso\ 3"$$

2. Estadístico de prueba

$$Fcalc = \frac{{S_{31}}^2}{{S_{32}}^2} = \frac{0.70057}{0.23814} = 2.94$$

3. Región de rechazo



$$F_{\frac{\alpha}{2}.n_1-1.n_2-1} = F_{0.025.5-1.7-1} = F_{0.025.4.6} = 6.23$$

$$F_{1-\alpha/2.n_1-1.n_2-1} = \frac{1}{F_{\frac{\alpha}{2}.n2-1.n1-1}} = \frac{1}{F_{0.025.7-1.5-1}} = \frac{1}{F_{0.025-6.4}} = \frac{1}{9.20} = 0.1087$$

4. Regla de Decisión

Se rechaza H_0 si Fcalc > 7.39 ó si es ≤ 0.1068

5. Conclusión:

- Estadistica: Como 0.1087< 2.94< 6.23 ∴ no se rechaza H_0 y se utiliza el caso 2.
- Adva: Existe evidencia suficiente para decir que estadísticamente las varianzas de Valor de ambas medias son iguales.

Intervalo de Confianza para Número de Recámaras

$$M_{31} - M_{32} = \left(\overline{X}_{31} - \overline{X}_{32}\right) \mp (t_{\alpha/2} \cdot n_1 + n_2 - 2 \quad Sp \sqrt{\frac{1}{n} + \frac{1}{n_2}}$$

Donde: $Sp = \sqrt{\frac{(n_1-1){S_1}^2+(n_2-1){S_2}^2}{n_1+n_2-2}}$

$$Sp = \sqrt{\frac{(5-1)0.70057 + (7-1)0.23814}{5+7-2}}$$

$$Sp = \sqrt{\frac{4 * 0.70057 + 6 * 0.23814}{10}}$$

$$Sp = \sqrt{\frac{2.80228 + 1.42884}{10}}$$

$$Sp = \sqrt{\frac{4.23112}{10}}$$

$$Sp = \sqrt{0.42311}$$

$$Sp = 0.65047$$

$$M_{31} - M_{32} = (3.2 - 3.71) \mp 2.228 * 0.65047 * \sqrt{\frac{1}{5} + \frac{1}{7}}$$

$$M_{31} - M_{32} = (-0.51) \mp 2.228 * 0.65047 * \sqrt{0.20 + 0.14286}$$

$$M_{31} - M_{32} = (-0.51) \mp 2.228 * 0.65047 * \sqrt{0.34286}$$

$$M_{31} - M_{32} = (-0.51) \mp 2.228 * 0.65047 * 0.58554$$

$$M_{31} - M_{32} = (-0.51) \mp 0.84859 \begin{cases} LIC = -1.36 \\ LSC = 0.33 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos, podemos decir que estadísticamente la influencia de la zona Sur es igual que la influencia de la zona Este sobre el Número de Recámaras de las casas.

Nota: Como la influencia de la zona sobre el **Número de Recámaras** de las casas es el mismo en ambas zonas sería razonable que la prueba de hipótesis de diferencia de medias multivariada **NO** se hubiera rechazado por esta variable.

Comprobación utilizando el software Minitab 17

Prueba e IC para dos varianzas: Número vs Niveles

```
Método

Hipótesis nula          σ(1) / σ(2) = 1
Hipótesis alterna       σ(1) / σ(2) ≠ 1
Nivel de significancia  α = 0.05
```

```
Estadísticas

                                    IC de 95%
                                      para
Niveles  N  Desv.Est.  Varianza     Desv.Est.
1        5      0.837     0.700  (0.407, 2.829)
2        7      0.488     0.238  (0.143, 2.317)

Relación de desviaciones estándar = 1.715
Relación de varianzas = 2.940
```

ANOVA unidireccional: Número vs. Niveles

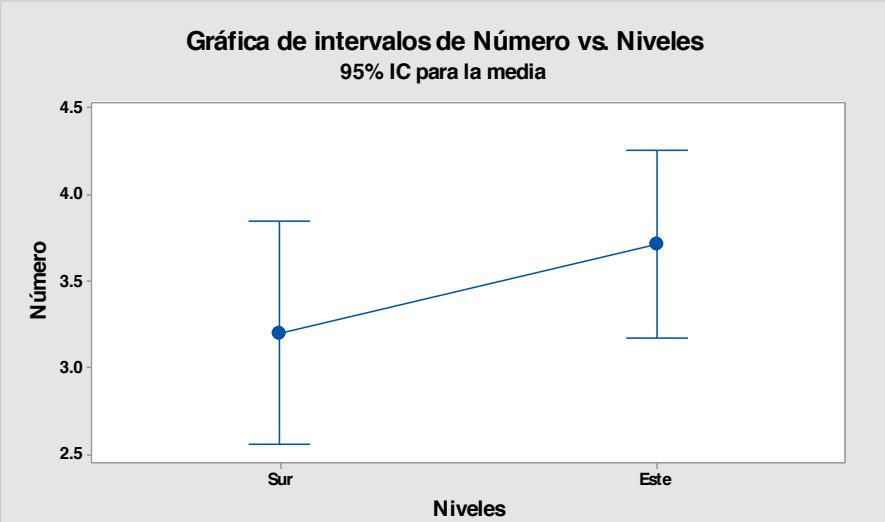
```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        2  1, 2


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   1     0.7714     0.7714     1.82    0.207
Error    10     4.2286     0.4229
Total    11     5.0000


Resumen del modelo

                      R-cuad.  R-cuad.
       S  R-cuad.  (ajustado)   (pred)
0.650275   15.43%       6.97%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  3.200      0.837  (2.552, 3.848)
2        7  3.714      0.488  (3.167, 4.262)

Desv.Est. agrupada = 0.650275
```

Comparaciones en parejas de Tukey

```
Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
```

```
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            0.514       0.381  (-0.334, 1.363)     1.35     0.207

Nivel de confianza individual = 95.00%
```

Gráfica de intervalos de Número vs. Niveles



Elaboró: Arq. y M. en Arq. Javier Bech Vertti.
La desviación estándar agrupada se utilizó para calcular los intervalos.

3.2.3.1.1 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.3.1.1. Análisis de Varianza Multivariada (MANOVA)**

3.2.3.1.1. En cierta empresa, el departamento de recursos humanos está midiendo la efectividad de 3 distintos programas de inducción para su personal, los cuales consisten en Dinámicas, Casos o Seminarios. La efectividad será medida basándose en el grado de identificación que el personal logra con su empresa. Este último se evalúa a través de 3 criterios:

1. ¿Qué tanto el empleado llega a conocer los objetivos de la empresa? Y,
2. ¿Qué tanta motivación se logró dar a la persona durante el programa de capacitación?
3. ¿Qué tanta habilidad obtuvo el empleado para desempeñar el nuevo puesto?

Los datos obtenidos son los siguientes:

Método de Inducción								
Dinámica			Casos			Seminarios		
Conocimiento	Motivación	Habilidad	Conocimiento	Motivación	Habilidad	Conocimiento	Motivación	Habilidad
9	3	6	1	3	0	4	8	1
6	2	9	1	1	2	1	9	0
8	5	7	2	2	1	2	7	3
9	4	9	0	3	1	2	6	2
8	6	4	2	2	0	1	8	3
			0	1	2	1	9	3
						3	9	2

a) ¿Existe evidencia suficiente que indique que el grado de identificación que el personal logra con su empresa es el mismo en los tres Métodos de Inducción utilizados?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable independiente para ver si el Método utilizado influye sobre el "grado de identificación que el personal logra con su empresa".
c) Realice las pruebas Post-hoc de Tukey para cada variable independiente y construya los intervalos de confianza que indique, en su caso, cual Método de Inducción utilizado logró mayor "grado de identificación del personal con su empresa" y por cuanto es la diferencia.

Resolución del inciso a)

Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

Prueba de hipótesis:

1. Juego de Hipótesis

$$H_0 = \begin{bmatrix}\mu_{11}\\ \mu_{21}\\ \mu_{31}\end{bmatrix} = \begin{bmatrix}\mu_{12}\\ \mu_{22}\\ \mu_{32}\end{bmatrix} = \begin{bmatrix}\mu_{13}\\ \mu_{23}\\ \mu_{33}\end{bmatrix}$$

2. Estadístico de prueba

Como No. de Variables Dependientes : p≥1 y

No. de Niveles del Factor: g=3

Usamos la siguiente fórmula (Caso 4):

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Donde\ £ = \frac{|error|}{|tratamiento|} = \frac{|SCE|}{|SCT|} = 0.00603$$

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Fcalc = \frac{(18-3-2)}{3}\frac{(1-\sqrt{0.00603})}{\sqrt{0.00603}}$$

$$Fcalc = \frac{(13)}{3}\frac{(1-0.07765}{0.07765}$$

$$Fcalc = (4.333)(11.87779)$$

$$Fcalc = 51.47$$

3. Región de rechazo

$$Fcrit = \left[\propto, 2, p, 2\left(\sum n_1 - p - 2\right)\right]$$

$Fcrit = \left[0.05,2(3),2\left(\sum 18-3-2\right)\right]$

$Fcrit = [0.05,6,2(13)]$

$Fcrit = 0.05,6,26$

$F0.05 = 2.47$

$F0.01 = 3.59$

4. Regla de Decisión

Se rechaza H_0 sí $fcal \geq 2.47$

5. Conclusión

Estadista: Como 51.47 > 2.47 > 3.59 ∴ la prueba es (A.S) y se rechaza con mucha potencia.

Administrativa: Existe evidencia suficiente para decir que estadísticamente que al menos una de las medias de los 3 grupos NO es igual.

Resolución del inciso b)

Prueba de Hipótesis para Conocimiento:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de efectividad que llega a tener el Método de Inducción utilizado para los nuevos empleados es el mismo para el **Conocimiento** de los objetivos de la empresa.

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los Métodos de Inducción utilizados para los nuevos empleados son iguales para el **Conocimiento** de los objetivos de la empresa.

$$H_1: \text{No todos los Metodos de Inducción son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Métodos de Inducción utilizados para los nuevos empleados para el **Conocimiento** de los objetivos de la empresa es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de la puntuación asignada al **Conocimiento** con base los Métodos de Inducción y el grado de efectividad es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Métodos de Inducción tiene una diferencia significativa para el **Conocimiento** de los objetivos de la empresa sobre los otros Métodos de Inducción utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 172$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 154$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 18$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamien-tos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 154$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{154}/_{2} = 77$	$F_{Calc.} = {}^{CMt}/_{CM}$ $= {}^{77}/_{1.2}$ $= \mathbf{64.17}$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 18$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{18}/_{15} = 1.2$	
Total	18-1=17	$SCT = 172$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **64.17**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de efectividad que llega a tener el Método de Inducción utilizado para los nuevos empleados no es el mismo en al menos uno de los tres Métodos de Inducción utilizados.

Prueba de Hipótesis para Motivación:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de efectividad que llega a tener el Método de Inducción utilizado para los nuevos empleados es el mismo para la **Motivación** de los nuevos empleados

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los Métodos de Inducción utilizados para los nuevos empleados son iguales para la **Motivación** de los nuevos empleados

$$H_1: \text{No todos los Metodos de Inducción son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Métodos de Inducción utilizados para los nuevos empleados para la **Motivación** de los nuevos empleados es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio de la puntuación asignada a la **Motivación** con base los Métodos de Inducción y el grado de efectividad es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Métodos de Inducción tiene una diferencia significativa para la **Motivación** de los nuevos empleados sobre los otros Métodos de Inducción utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 143.78$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 121.78$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 22$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	$F_{calculada}$
Tratamien-tos	$v_1 = k - 1 = 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 121.78$	$CMt = {}^{SCt}/_{g.l.} = {}^{121.78}/_{2} = 60.889$	$F_{Calc.} = {}^{CMt}/_{CM} = {}^{60.889}/_{1.46} = \mathbf{41.51}$
Error	$v_2 = N - k = 18 - 3 = \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 22$	$CME = {}^{SCE}/_{g.l.} = {}^{22}/_{15} = 1.467$	
Total	18-1=17	$SCT = 143.78$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denomi nador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48

11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **41.51**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que NO todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de efectividad que llega a tener el Método de Inducción utilizado para los nuevos empleados NO es el mismo en los tres Métodos utilizados.

Prueba de Hipótesis para Habilidad:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de efectividad que llega a tener el Método de Inducción utilizado para los nuevos empleados es el mismo para la **Habilidad** de los nuevos empleados

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los Métodos de Inducción utilizados para los nuevos empleados son iguales para la **Habilidad** de los nuevos empleados

$$H_1: \text{No todos los Metodos de Inducción son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Métodos de Inducción utilizados para los nuevos empleados para la **Habilidad** de los nuevos empleados es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias

significativas en el promedio de la puntuación asignada a la **Habilidad** con base los Métodos de Inducción y el grado de efectividad es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Métodos de Inducción tiene una diferencia significativa para la **Habilidad** de los nuevos empleados sobre los otros Métodos de Inducción utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 140.94$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 110.94$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 30$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamientos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 110.94$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{110.94}/_{2}$ $= 55.472$	$F_{Calc.} = {}^{CMt}/_{CM}$ $= {}^{55.472}/_{2}$ $= 27.74$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 30$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{30}/_{15} = 2$	
Total	18-1=17	$SCT = 140.94$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	Grados de libertad en el numerador			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **27.74**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de efectividad que llega a tener el Método de Inducción utilizado para los nuevos empleados no es el mismo en al menos uno de los tres Métodos de Inducción utilizados.

Resolución del inciso c)

Prueba de Tukey

Intervalo de confianza para Conocimiento (Dinámica-Casos) .Modelo desbalanceado.

$$M_{11} - M_{12} = (\bar{X}_{11} - \bar{X}_{12}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = (8-1) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = 7 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{18}{15} = \underline{1.20}$$

$$nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{11} - M_{12} = 7 \mp 3.67 \sqrt{\frac{1.2}{5.4545}}$$

$$M_{11} - M_{12} = 7 \mp 3.67 \sqrt{0.22}$$

$$M_{11} - M_{12} = 7 \mp 3.67(0.46904)$$

$$M_{11} - M_{12} = 7 \mp 1.72138 \begin{cases} LIC = 5.279 \\ LSC = 8.721 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica sobre el Conocimiento** es mayor que el del **Método de Casos sobre el Conocimiento** por un mínimo de 5.279 y un máximo de 8.721 puntos.

Intervalo de confianza para Conocimiento(Dinámica-Seminarios). Modelo desbalanceado

$$M_{11} - M_{13} = (\bar{X}_{11} - \bar{X}_{13}) \mp q \propto, (k.N-k) \sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = (8-2) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = 6 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = 5.8333$$

$$M_{11} - M_{13} = 6 \mp 3.67 \sqrt{\frac{1.2}{5.8333}}$$

$$M_{11} - M_{13} = 6 \mp 3.67(0.45356)$$

$$M_{11} - M_{13} = 6 \mp 1.66456 \begin{cases} LIC = 4.335 \\ LSC = 7.665 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica sobre el Conocimiento** es mayor que el del **Método de Seminarios sobre el Conocimiento** por un mínimo de 4.335 y un máximo de 7.665 puntos.

Intervalo de confianza para Conocimiento (Casos-Seminarios). Modelo desbalanceado

$$M_{12} - M_{13} = (\bar{X}_{12} - \bar{X}_{13}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = (1 - 2) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = -1 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = 6.4615$$

$$M_{12} - M_{13} = -1 \mp 3.67 \sqrt{\frac{1.2}{6.4615}}$$

$$M_{12} - M_{13} = -1 \mp 3.67(0.43095)$$

$$M_{12} - M_{13} = -1 \mp 1.58158 \begin{cases} LIC = -2.582 \\ LSC = 0.582 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de efectividad del **Método de Casos sobre el Conocimiento** es el mismo que el del **Método de Seminarios sobre el Conocimiento.**

Intervalo de confianza para Motivación (Dinámica-Casos). Modelo desbalanceado

$$M_{21} - M_{22} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = (4-2) \mp 0.05,3,15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = 2 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{22}{15} = \underline{1.467}$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{21} - M_{22} = 2 \mp 3.67\sqrt{\frac{1.467}{5.4545}}$$

$$M_{21} - M_{22} = 2 \mp 3.67(0.51861)$$

$$M_{21} - M_{22} = 2 \mp 1.90328\begin{cases}0.09672\\3.90328\end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica sobre la Motivación** es mayor que el del **Método de Casos sobre la Motivación** por un mínimo de 0.09672 y un máximo de 3.90328 puntos.

Intervalo de confianza para Motivación (Dinámica-Seminarios). Modelo desbalanceado.

$$M_{21} - M_{23} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto,(k.N-k)\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = (4-8) \mp 0.05,3,15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = -4 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{21} - M_{23} = -4 \mp 3.67\sqrt{\frac{1.467}{5.8326}}$$

$$M_{21} - M_{23} = -4 \mp 3.67(0.50152)$$

$$M_{21} - M_{23} = -4 \mp 1.84056 \begin{cases} -5.84056 \\ -2.15944 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos podemos decir que estadísticamente el grado de efectividad del **Método de Seminarios sobre la Motivación** es mayor que el del **Método de Dinámica sobre la Motivación** por un mínimo de 2.15944 y un máximo de 5.84056 puntos. (En valor absouto).

Intervalo de confianza para Motivación (Casos-Seminarios). Modelo desbalanceado.

$$M_{22} - M_{23} = (\bar{X}_{22} - \bar{X}_{23}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = (2 - 8) \mp 0.05,3,15 \sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = -6 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{22} - M_{23} = -6 \mp 3.67 \sqrt{\frac{1.467}{6.4620}}$$

$$M_{22} - M_{23} = -6 \mp 3.67(0.47647)$$

$$M_{22} - M_{23} = -6 \mp 1.74863 \begin{cases} LIC = -7.74863 \\ LSC = -4.25137 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos podemos decir que estadísticamente el grado de efectividad del **Método de Seminarios sobre la Motivación** es mayor que el del **Método de Casos sobre la Motivación** por un mínimo de 4.25137 y un máximo de 7.74863 puntos. (En valor absouto).

Interpretación de confianza para Habilidad (Dinámica-Casos). Modelo desbalanceado

$$M_{31} - M_{32} = (\bar{X}_{31} - \bar{X}_{32}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = (7 - 1) \mp 0.05,3,15 \sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = 6 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{30}{15} = 2$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{31} - M_{32} = 6 \mp 3.67\sqrt{\frac{2}{5.4545}}$$

$$M_{31} - M_{32} = 6 \mp 3.67(0.60553)$$

$$M_{31} - M_{32} = 6 \mp 2.2223 \begin{cases} LIC = 3.7777 \\ LSC = 8.2223 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica sobre la Habilidad** es mayor que el del **Método de Casos sobre la Habilidad** por un mínimo de 3.7777 y un máximo de 8.2223 puntos.

Intervalo de confianza para Habilidad (Dinámica-Seminarios). Modelo desbalanceado.

$$M_{31} - M_{33} = (\bar{X}_{31} - \bar{X}_{33}) \mp q \propto, (k.N-k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = (7-2) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = 5 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{31} - M_{33} = 5 \mp 3.67\sqrt{\frac{2}{5.8326}}$$

$$M_{31} - M_{33} = 5 \mp 3.67(0.58558)$$

$$M_{31} - M_{33} = 5 \mp 2.14907 \begin{cases} 2.85093 \\ 7.14907 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica sobre la Habilidad** es mayor que el del **Método de Seminarios sobre la Habilidad** por un mínimo de 2.85093 y un máximo de 7.14907 puntos.

Intervalo de confianza para Habilidad (Casos-Seminarios). Modelo desbalanceado.

$$M_{32} - M_{33} = (\bar{X}_{32} - \bar{X}_{33}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = (1 - 2) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = -1 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{32} - M_{33} = -1 \mp 3.67\sqrt{\frac{2}{6.4620}}$$

$$M_{32} - M_{33} = -1 \mp 3.67(0.55633)$$

$$M_{32} - M_{33} = -1 \mp 2.04173 \begin{cases} LIC = -3.04173 \\ LSC = 1.04173 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de efectividad del **Método de Casos sobre la Habilidad** es el mismo que el del **Método de Seminarios sobre la Habilidad.**

3.2.3.1.1 EJERCICIO DE **REFUERZO con Minitab 17**

EJERCICIO DE **REFUERZO 3.2.3.1.1. Análisis de Varianza Multivariada (MANOVA) con Minitab 17**

3.2.3.1.1. En cierta empresa, el departamento de recursos humanos está midiendo la efectividad de 3 distintos programas de inducción para su personal, los cuales consisten en Dinámicas, Casos o Seminarios. La efectividad será medida basándose en el grado de identificación que el personal logra con su empresa. Este último se evalúa a través de 3 criterios:

1. ¿Qué tanto el empleado llega a conocer los objetivos de la empresa? Y,
2. ¿Qué tanta motivación se logró dar a la persona durante el programa de capacitación?
3. ¿Qué tanta habilidad obtuvo el empleado para desempeñar el nuevo puesto?

Los datos obtenidos son los siguientes:

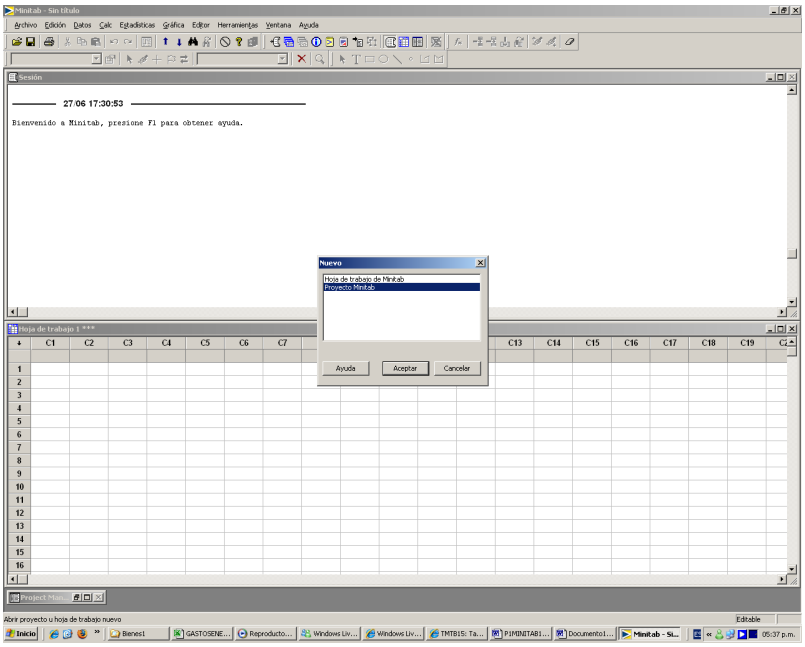
Método de Inducción								
Dinámica			Casos			Seminarios		
Conocimiento	Motivación	Habilidad	Conocimiento	Motivación	Habilidad	Conocimiento	Motivación	Habilidad
9	3	6	1	3	0	4	8	1
6	2	9	1	1	2	1	9	0
8	5	7	2	2	1	2	7	3
9	4	9	0	3	1	2	6	2
8	6	4	2	2	0	1	8	3
			0	1	2	1	9	3
						3	9	2

a) ¿Existe evidencia suficiente que indique que el grado de identificación que el personal logra con su empresa es el mismo en los tres Métodos de Inducción utilizados?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable independiente para ver si el Método utilizado influye sobre el "grado de identificación que el personal logra con su empresa". Asimismo realice las pruebas Post-hoc de Tukey para cada variable independiente y construya los intervalos de confianza que indique, en su caso, cual Método de Inducción utilizado logró mayor "grado de identificación del personal con su empresa" y por cuanto es la diferencia.

Resolución del inciso a)

Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

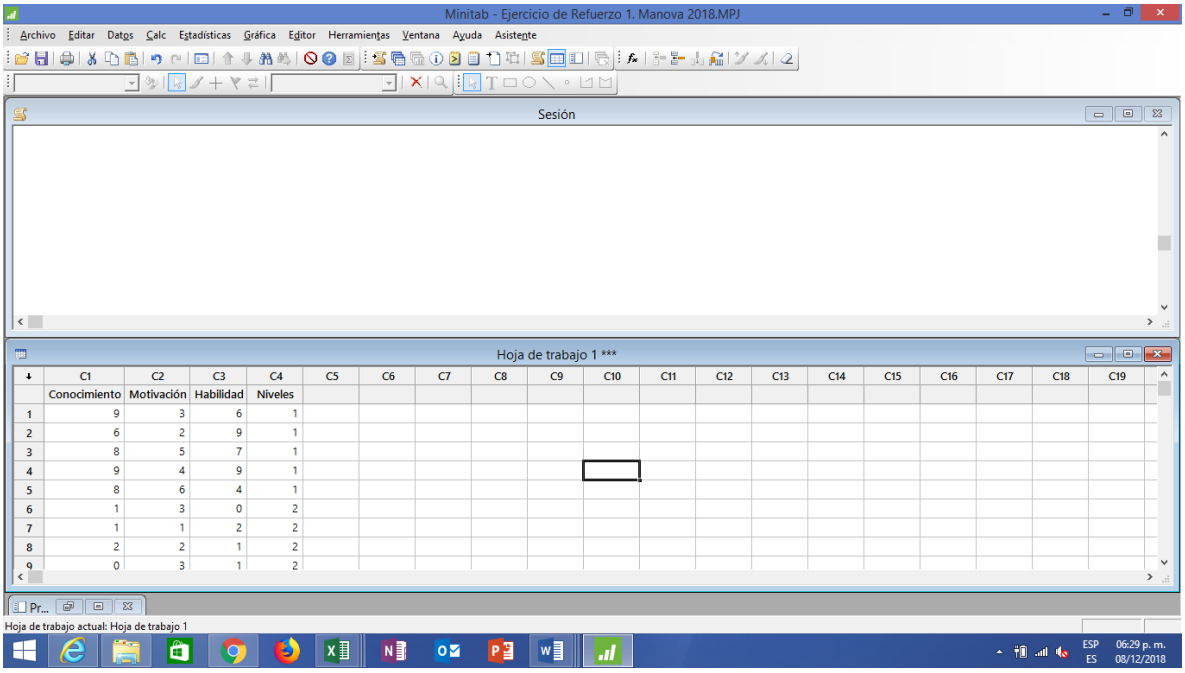
Para crear la base de datos del ejemplo ilustrativo en **MINITAB17**, accede a **MINITAB VERSIÓN 17.** Cuando aparezca la pantalla da un clic en la opción **Archivo** de la barra menú y selecciona **Nuevo**. Aparecerá la ventana **Nuevo. Selecciona la opción Proyecto Minitab y Aceptar**



Procede a editar la base de datos de la siguiente manera:

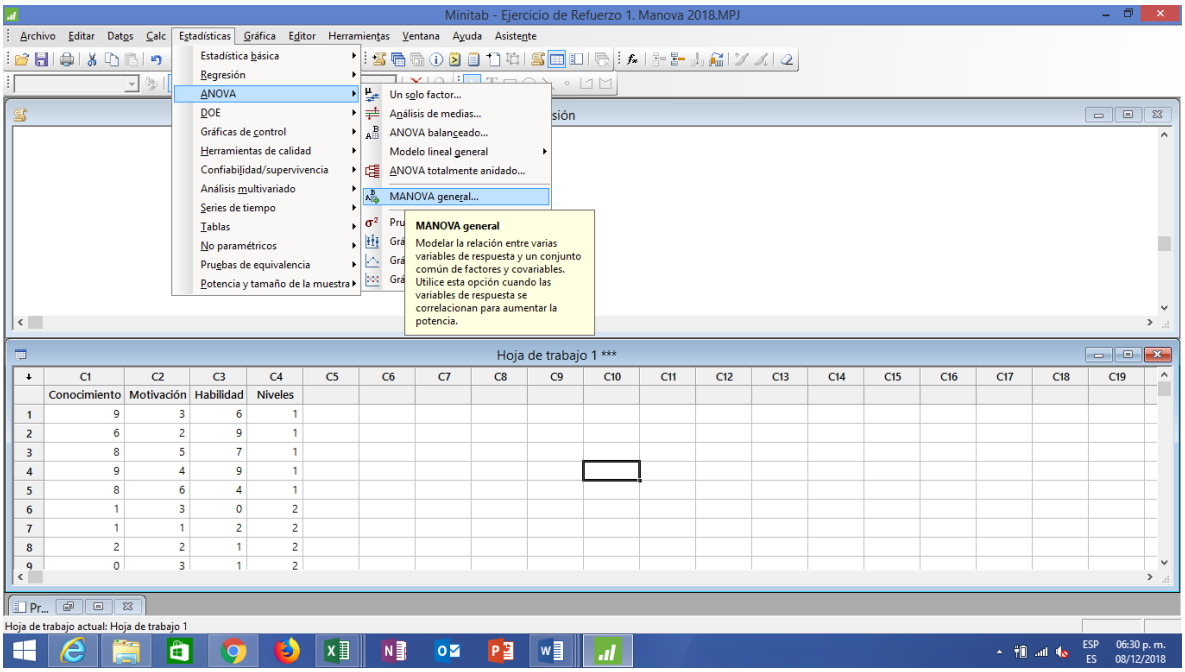
Coloca los nombres de cada **variable (En este caso Conocimiento, Motivación, Habilidad y Niveles**), debajo de cada número de columna, en este caso **C1, C2, C3 y C4**.

Una vez termines de colocar los nombres de cada **variable (En este caso Marca, Slogan, Logotipo y Niveles**, coloca el cursor en el primer nombre de la variable **(Marca)** y llena los datos de Marca en forma vertical de los tres medios utilizados (Prensa escrita, Televisión y Revistas). Cuando se terminen los 18 datos coloca el cursor en el segundo nombre de la variable (**Slogan**) y llena los 18 datos. Cuando se terminen los 18 datos de Slogan coloca el cursor en el tercer nombre de la variable **(Logotipo**) y llena los 18 datos. Cuando se terminen los 18 datos de Logotipo coloca el cursor en la cuarta columna (**Niveles**) y coloca en forma vertical cinco 1, seis 2 y siete 3. El número 1 corresponde al medio **Prensa escrita,** el 2 corresponde al medio **Televisión** y el 3 corresponde al Medio **Revistas**.

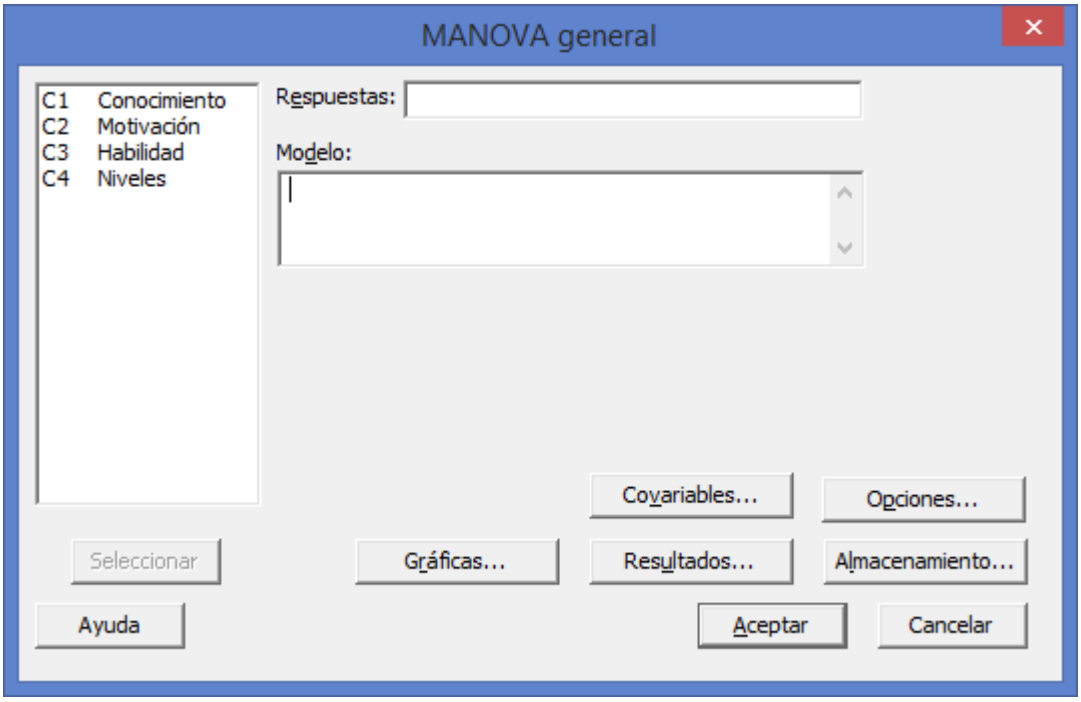


	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
	Conocimiento	Motivación	Habilidad	Niveles															
1	9	3	6	1															
2	6	2	9	1															
3	8	5	7	1															
4	9	4	9	1															
5	8	6	4	1															
6	1	3	0	2															
7	1	1	2	2															
8	2	2	1	2															
9	0	3	1	2															

2 Elija **Estadísticas > ANOVA > MANOVA general**.

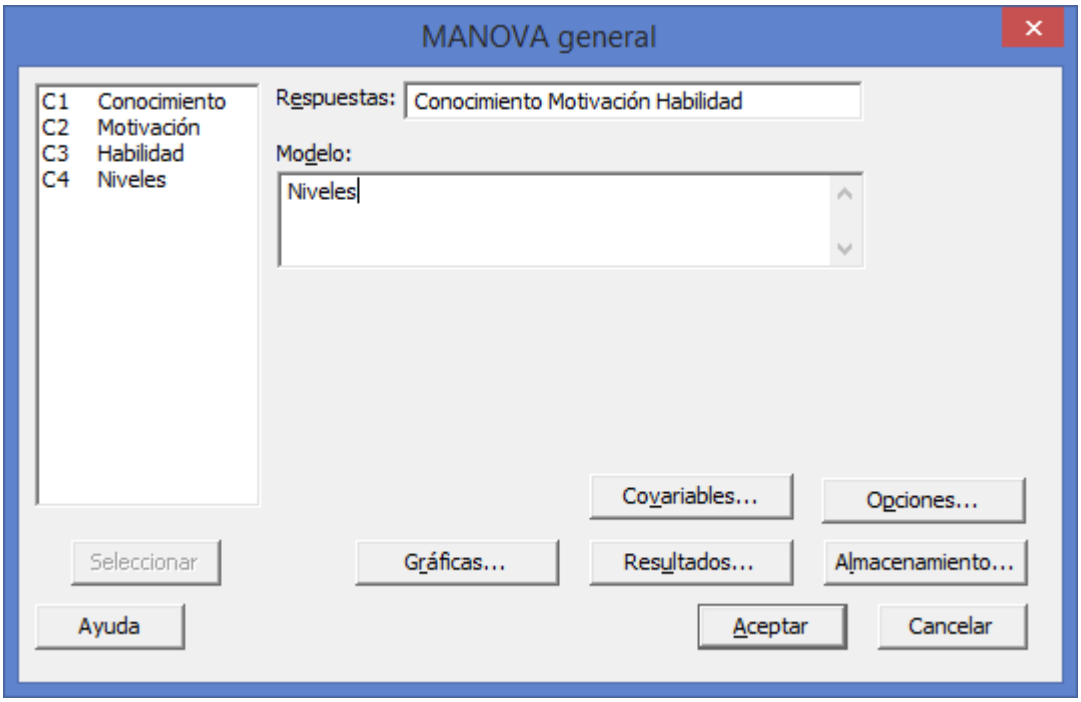


	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19
	Conocimiento	Motivación	Habilidad	Niveles															
1	9	3	6	1															
2	6	2	9	1															
3	8	5	7	1															
4	9	4	9	1															
5	8	6	4	1															
6	1	3	0	2															
7	1	1	2	2															
8	2	2	1	2															
9	0	3	1	2															



2 En **Respuestas**, ingrese las 3 variables o columnas numéricas que contengan las variables de respuesta (**Conocimiento, Motivación y Habilidad**).

3 En **Factores**, ingresa la cuarta columna (**Niveles**)



4 Haga clic en **Aceptar**.

Salida de la ventana Sesión

Modelo lineal general: Conocimiento, Motivación, Habilidad vs. Niveles

```
MANOVA para Niveles
s = 2    m = 0.0    n = 5.5

                  Estadística              GL
Criterio            de prueba       F  Núm  Denom      P
De Wilks              0.00603  51.483    6     26  0.000
Lawley-Hotelling     27.11777  54.236    6     24  0.000
De Pillai             1.82450  48.515    6     28  0.000
De Roy               20.34475
```

Interpretación de los resultados

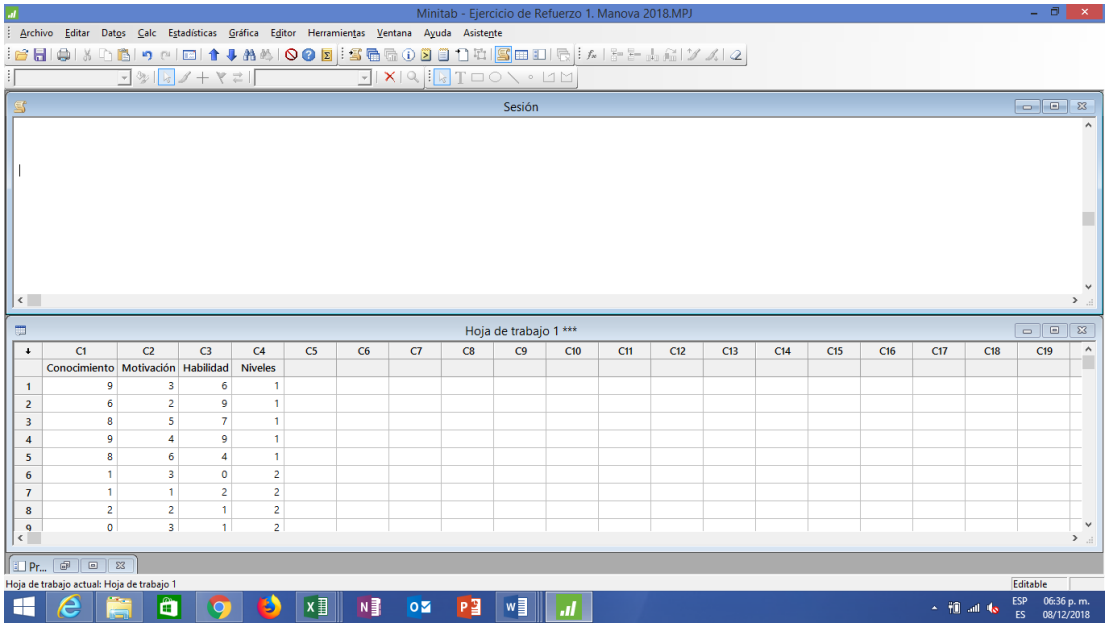
La salida de la ventana sesión presenta los resultados de cuatro criterios o pruebas entre ellas el de **Lambda de Wilks** que arroja una Estadística de prueba de **0.00603** con una Fcalculada de 51.482. El valor de p de 0.000 (nivel de significancia observado) indica que hay suficiente evidencia de que no todas las medias en cada variable son iguales en los tres métodos utilizados cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Resolución del inciso b)

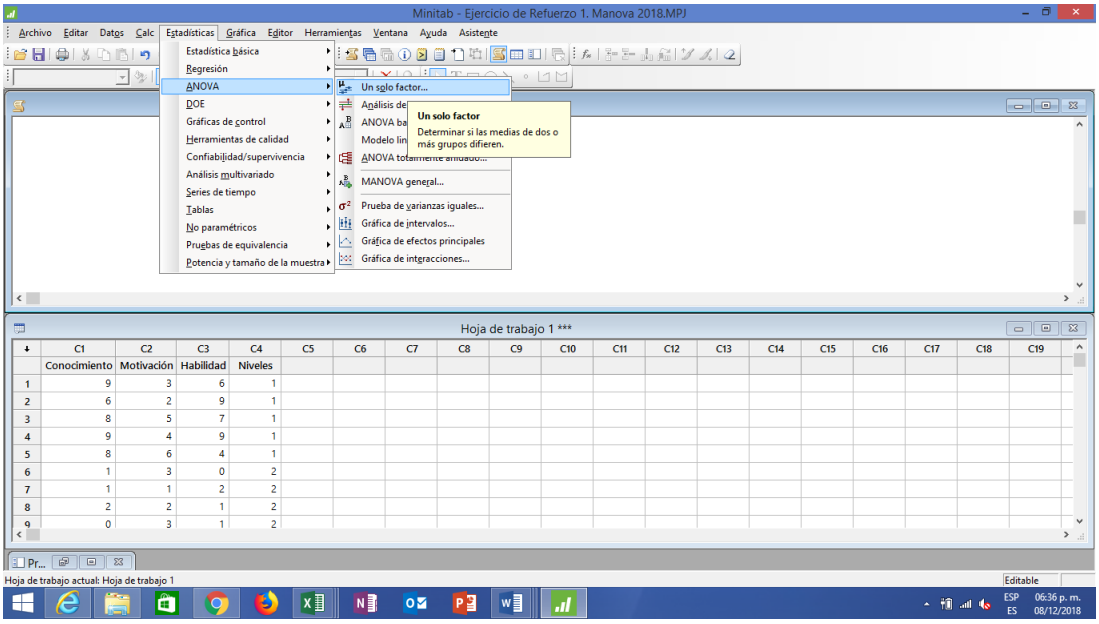
1. **Prueba ANOVA de Hipótesis para Conocimiento y pruebas Post-hoc de Tukey para Conocimiento:**

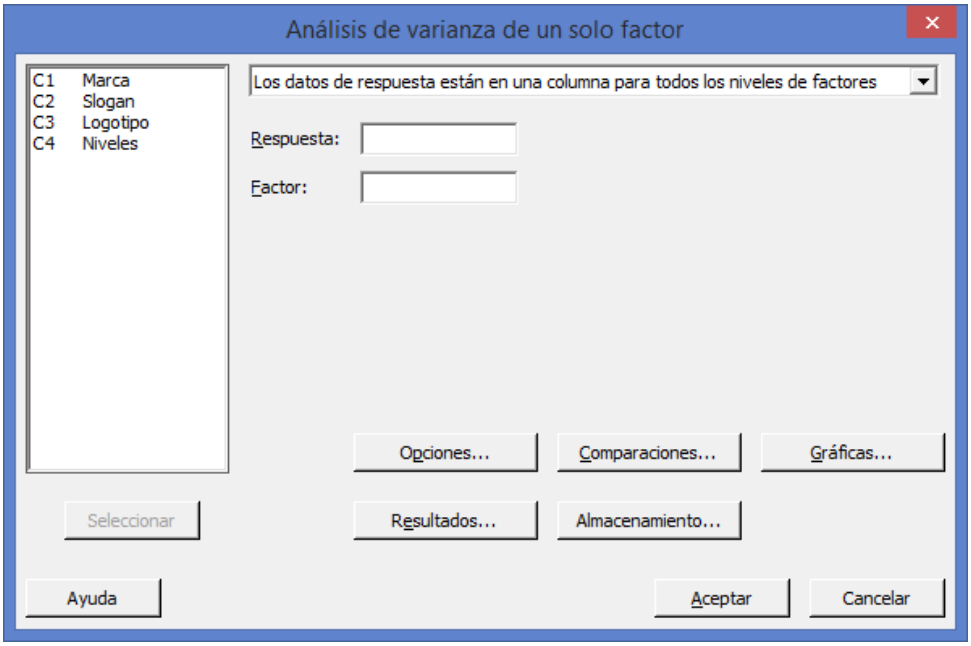
Cuando el número de observaciones en cada tratamiento es extenso y/o existen muchos tratamientos, los cálculos manuales son tediosos. Existen muchos paquetes de software que pueden mostrar los resultados entre ellos Minitab (Versión 17).

Comenzamos introduciendo los datos en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:



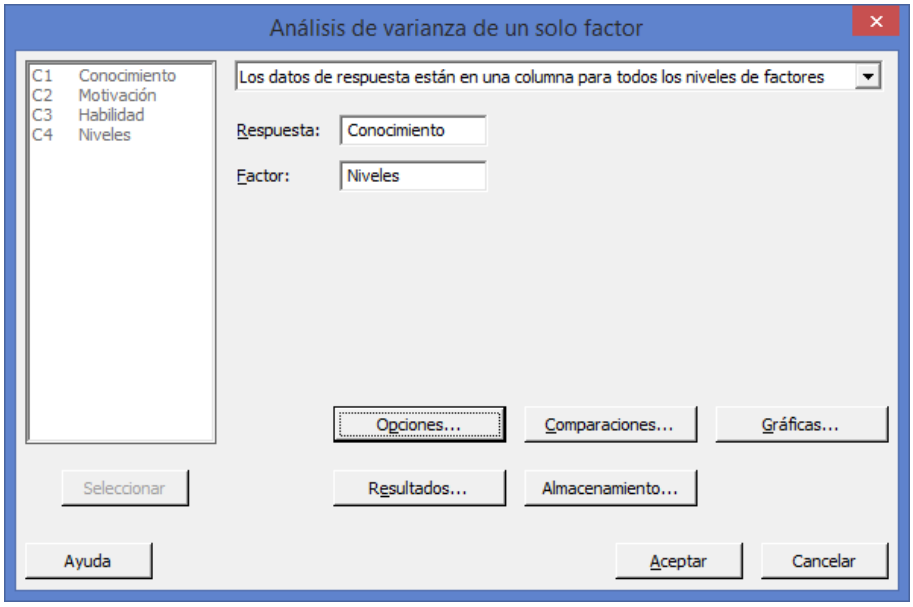
Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**



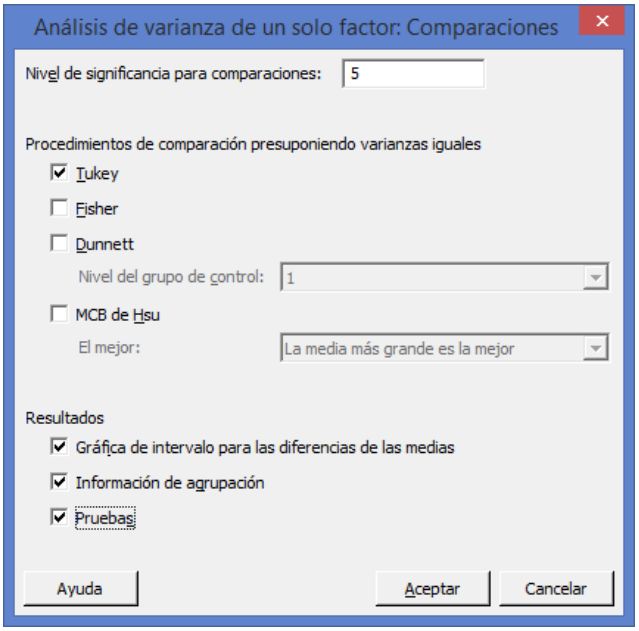


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Conocimiento**. En Factor, ingrese **Niveles**.



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Conocimiento vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2     154.00     77.000    64.17    0.000
Error    15      18.00      1.200
Total    17     172.00


Resumen del modelo

                    R-cuad.  R-cuad.
```

```
      S  R-cuad.  (ajustado)   (pred)
1.09545   89.53%      88.14%   84.87%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  8.000      1.225  (6.956, 9.044)
2        6  1.000      0.894  (0.047, 1.953)
3        7  2.000      1.155  (1.117, 2.883)

Desv.Est. agrupada = 1.09545
```

Interpretación de los resultados

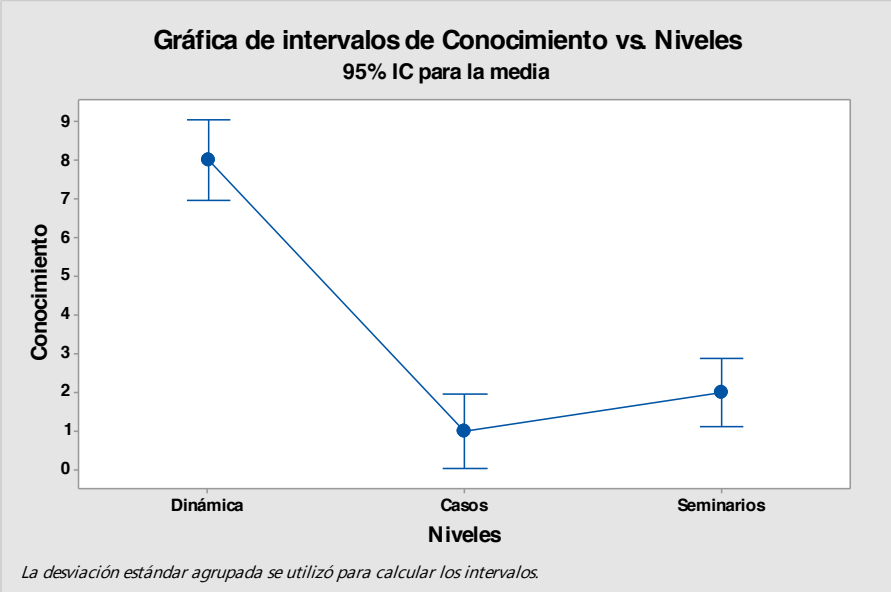
En la Tabla de ANOVA, para **Conocimiento** en **los tres Métodos de Inducción** utilizados el valor de F calculada de 64.17 con un valor p (0.000) indica que hay suficiente evidencia de que no todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

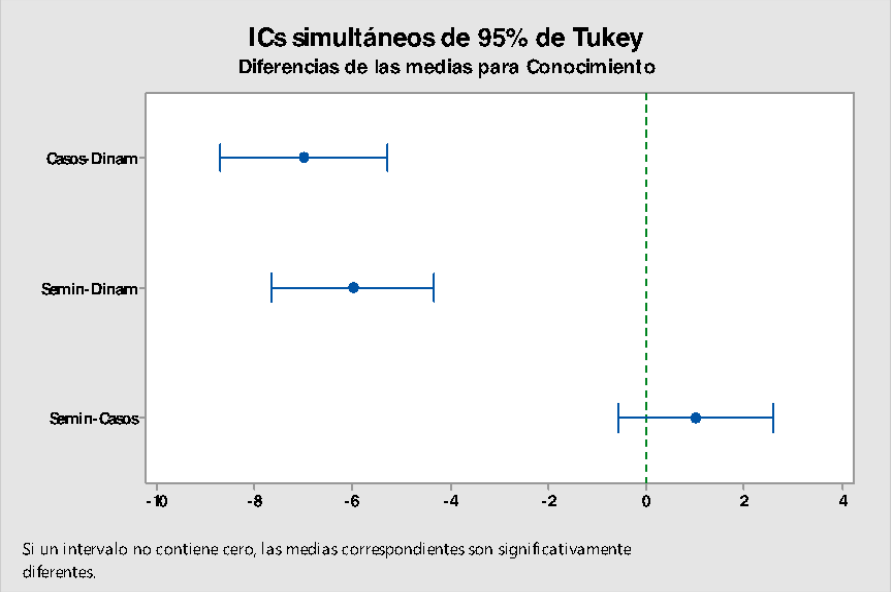
Comparaciones en parejas de Tukey

```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N  Media  Agrupación
1        5  8.000  A
3        7  2.000    B
2        6  1.000    B

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

           Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia      IC de 95%     Valor T  ajustado
2 - 1           -7.000       0.663  (-8.721, -5.279)   -10.55     0.000
3 - 1           -6.000       0.641  (-7.665, -4.335)    -9.35     0.000
3 - 2            1.000       0.609  (-0.582,  2.582)     1.64     0.260

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Conocimiento vs. Niveles





Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Conocimiento** en **Casos** (Nivel 2) con **Conocimiento** en **Dinámica** (Nivel 1) oscila entre -8.721 y -5.279. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica** sobre **Conocimiento** es mayor que el **Método de Casos** sobre **Conocimiento** por un mínimo de 5.279 y un máximo de 8.721 puntos (en valor absoluto).

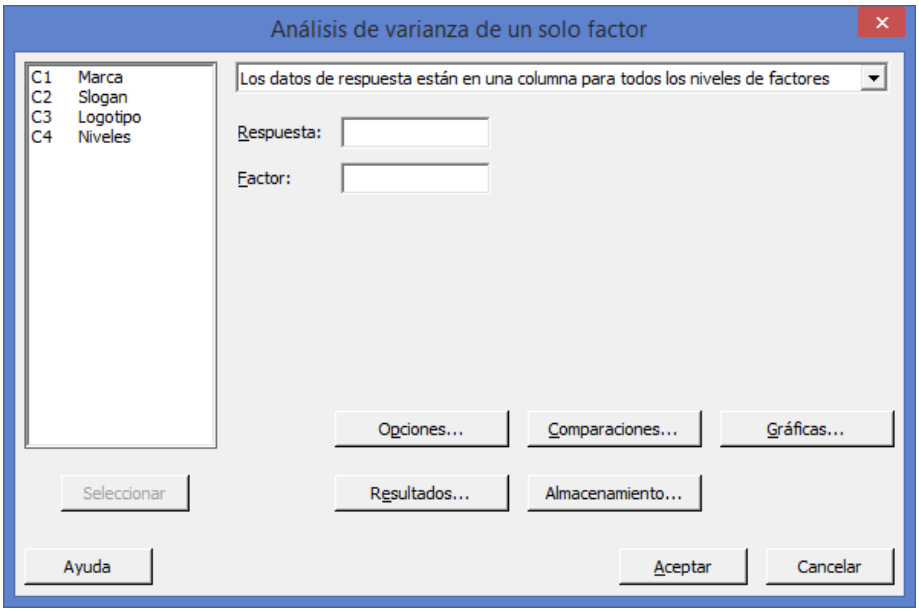
En cuanto al intervalo de confianza de **Conocimiento** en **Seminarios** (Nivel 3) con **Conocimiento** en **Dinámica** (Nivel 1) oscila entre -7.665 y -4.335. Como ambos límites de confianza son negativos podemos

decir que estadísticamente el grado de efectividad del **Método Dinámica** sobre **Conocimiento** es mayor que el **Método de Seminarios** sobre **Conocimiento** por un mínimo de 4.335 y un máximo de 7.665 puntos (en valor absoluto).

Por último podemos observar que el intervalo de confianza de **Conocimiento** en **Seminarios** (Nivel 3) con **Conocimiento** en **Casos** (Nivel 2) oscila entre -0.582 y 2.582. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de efectividad del **Método de Seminarios** sobre **Conocimiento** es igual que el del **Método de Casos** sobre **Conocimiento.**

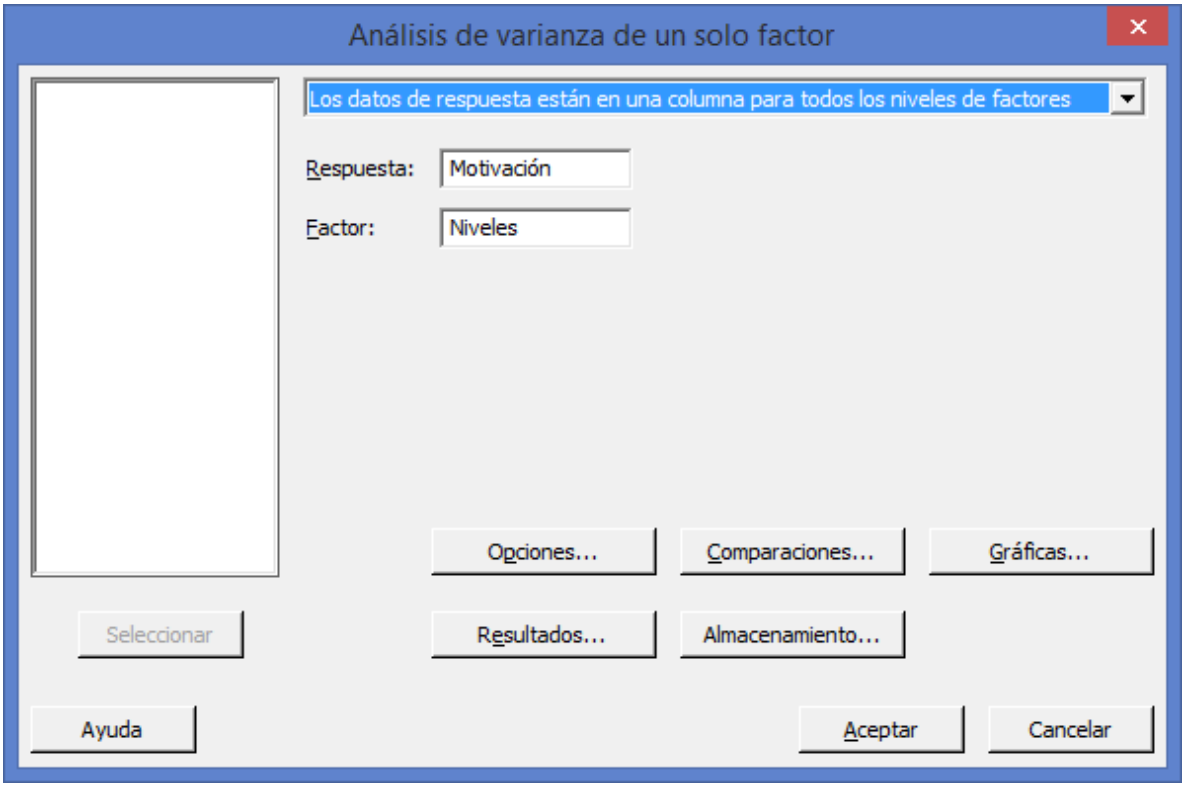
2. Prueba ANOVA de Hipótesis para Motivación y pruebas Post-hoc de Tukey para Motivación:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

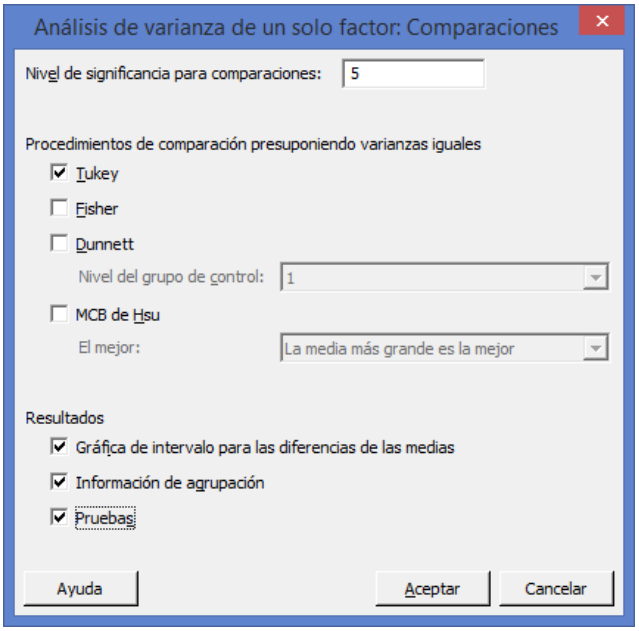


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Motivación**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Motivación vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2     121.78     60.889    41.52    0.000
Error    15      22.00      1.467
Total    17     143.78


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
1.21106   84.70%      82.66%   77.55%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  4.000      1.581  (2.846, 5.154)
2        6  2.000      0.894  (0.946, 3.054)
3        7  8.000      1.155  (7.024, 8.976)

Desv.Est. agrupada = 1.21106
```

Interpretación de los resultados

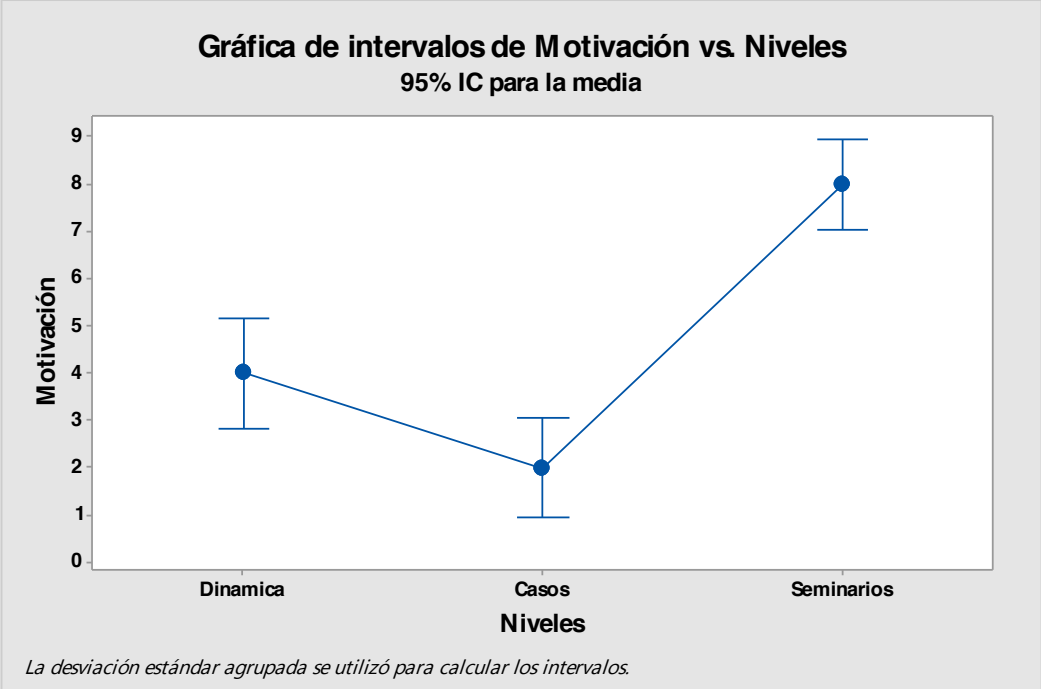
En la Tabla de ANOVA, para **Motivación** en **los tres Métodos de Inducción** utilizados el valor de F calculada de 41.52 con un valor p (0.000) indica que hay suficiente evidencia de que no todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

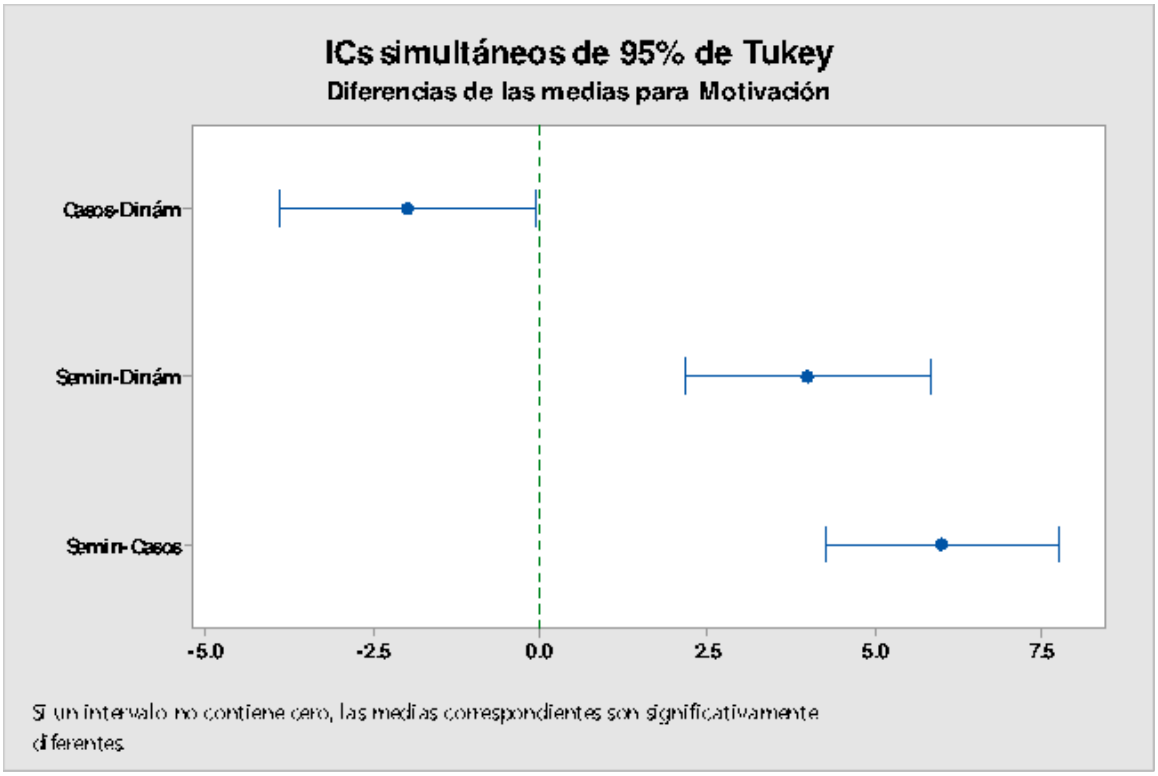
Comparaciones en parejas de Tukey

```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N  Media  Agrupación
3        7  8.000  A
1        5  4.000    B
2        6  2.000      C

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                                 Valor p
de niveles      medias  diferencia      IC de 95%     Valor T  ajustado
2 - 1           -2.000       0.733  (-3.903, -0.097)    -2.73     0.039
3 - 1            4.000       0.709  ( 2.160,  5.840)     5.64     0.000
3 - 2            6.000       0.674  ( 4.252,  7.748)     8.91     0.000

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Motivación vs. Niveles





Interpretación de los resultados

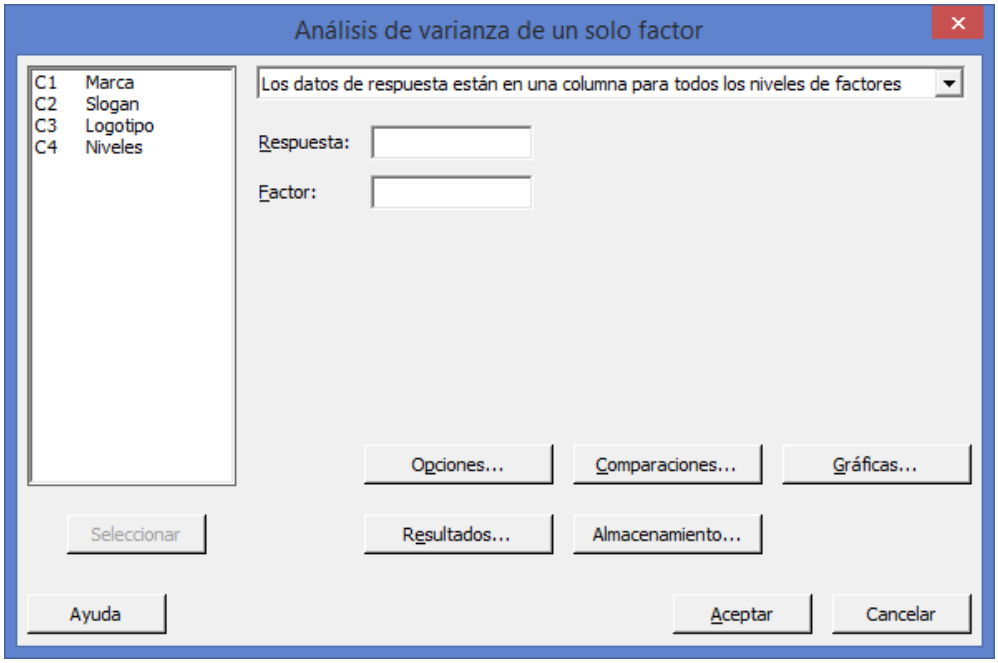
De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Motivación** en **Casos** (Nivel 2) con **Motivación** en **Dinámica** (Nivel 1) oscila entre -3.903 y -0.097. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica** sobre **Motivación** es mayor que el **Método de Seminarios** sobre **Motivación** por un mínimo de de 0.097 y un máximo de 3.903 puntos (en valor absoluto).

En cuanto al intervalo de confianza de **Motivación** en **Seminarios** (Nivel 3) con **Motivación** en **Dinámica** (Nivel 1) oscila entre 2.160 y 5.840. Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de efectividad del **Método Dinámico** sobre la **Motivació**n es mayor que el del **Método de Seminarios** sobre la **Motivación** por un mínimo de 2.160 y un máximo de 5.840.

Por último podemos observar que el intervalo de confianza de **Motivación** en **Seminarios** (Nivel 3) con **Motivación** en **Casos** (Nivel 2) oscila entre 4.252 y 7.748. Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de efectividad del **Método de Seminarios** sobre la **Motivación** es mayor que el del **Método de Casos** sobre la **Motivación** por un mínimo de 4.252 y un máximo de 7.748.

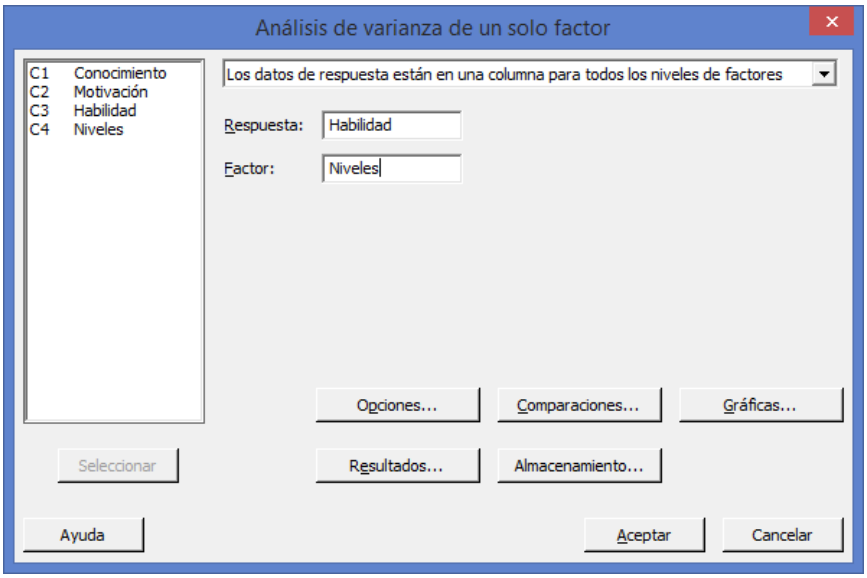
3. Prueba ANOVA de Hipótesis para Habilidad y pruebas Post-hoc de Tukey para Habilidad:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

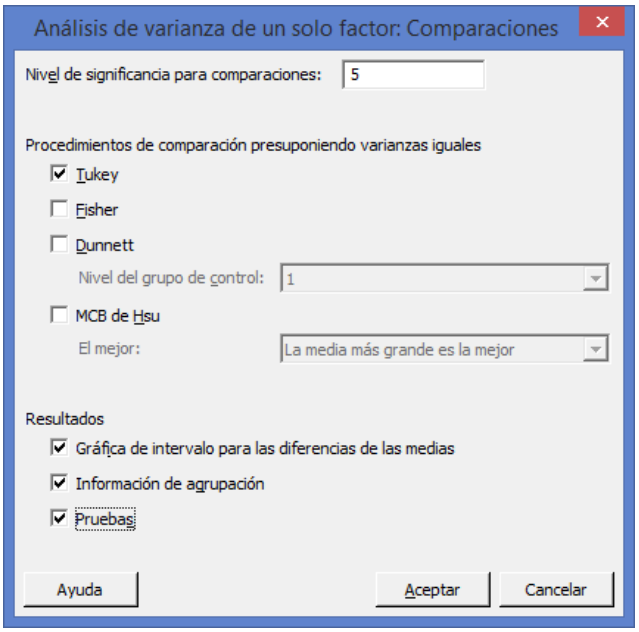


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Habilidad**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Habilidad vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2     110.94     55.472    27.74    0.000
Error    15      30.00      2.000
Total    17     140.94
```

```
Resumen del modelo

                      R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
1.41421   78.72%      75.88%   68.23%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  7.000      2.121  ( 5.652, 8.348)
2        6  1.000      0.894  (-0.231, 2.231)
3        7  2.000      1.155  ( 0.861, 3.139)

Desv.Est. agrupada = 1.41421
```

Interpretación de los resultados

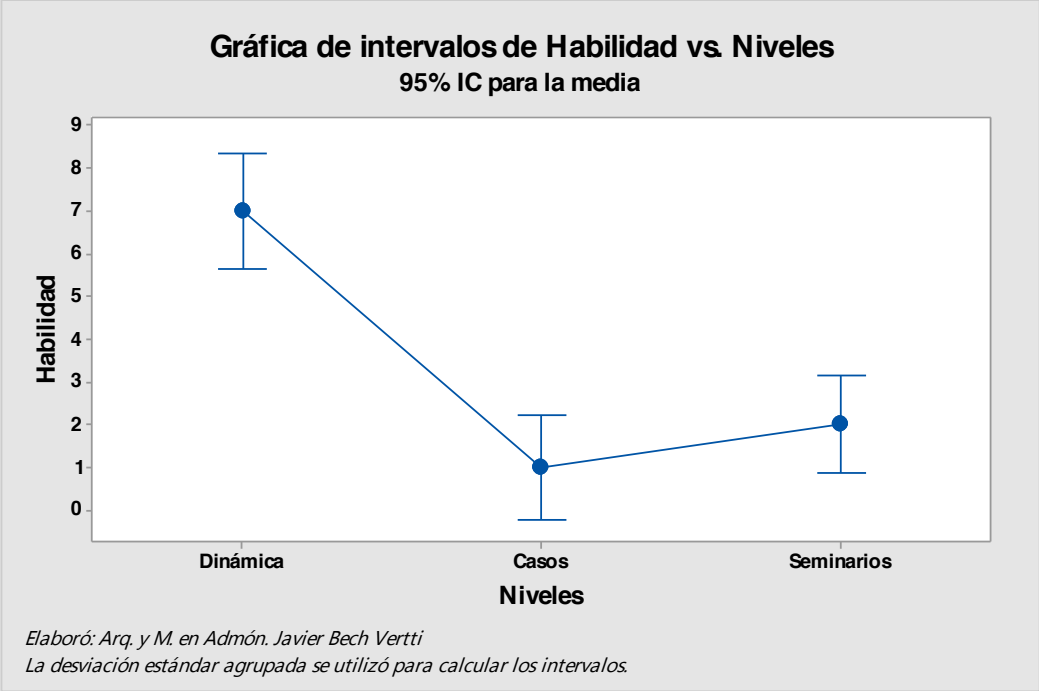
En la Tabla de ANOVA, para **Habilidad** en **los tres Métodos de Inducción** utilizados el valor de F calculada de 27.74 con un valor p (0.000) indica que hay suficiente evidencia de que no todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Comparaciones en parejas de Tukey

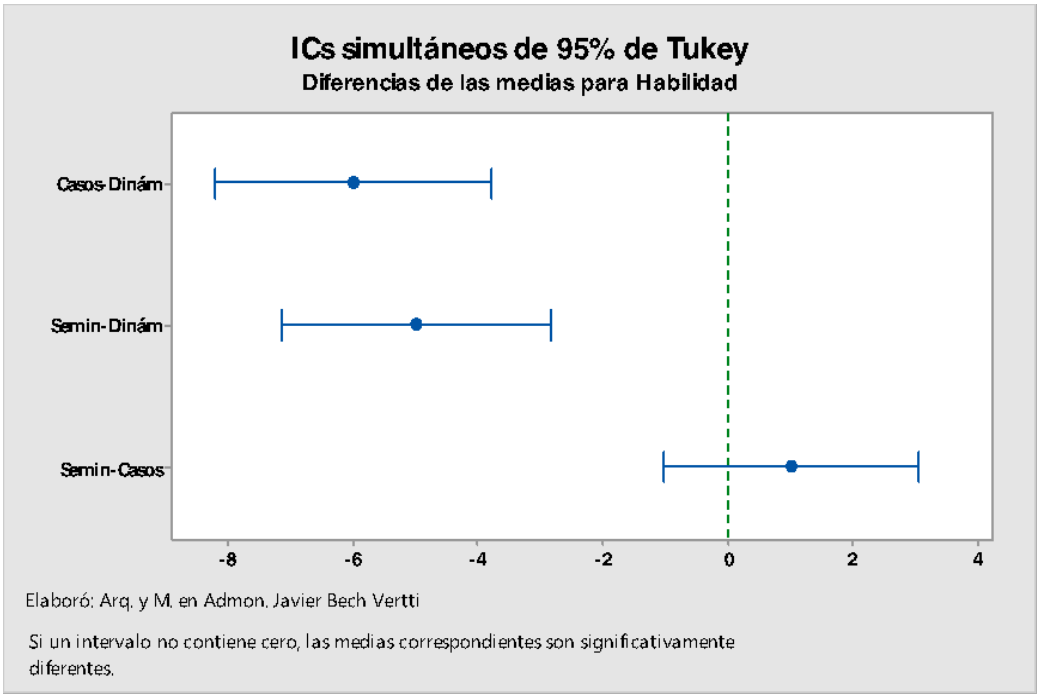
```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N  Media  Agrupación
1        5  7.000  A
3        7  2.000    B
2        6  1.000    B

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia      IC de 95%     Valor T  ajustado
2 - 1           -6.000       0.856  (-8.222, -3.778)    -7.01     0.000
3 - 1           -5.000       0.828  (-7.149, -2.851)    -6.04     0.000
3 - 2            1.000       0.787  (-1.042,  3.042)     1.27     0.432

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Habilidad vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.



Elaboró: Arq. y M. en Admon. Javier Bech Vertti

Si un intervalo no contiene cero, las medias correspondientes son significativamente diferentes.

Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Habilidad** en **Casos** (Nivel 2) con **Habilidad** en **Dinámica** (Nivel 1) oscila entre -8.222 y -3.778. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de efectividad del **Método Dinámica** sobre la **Habilidad** es mayor que el del **Método de Casos** sobre la **Habilidad** por un mínimo de 3.778 y un máximo de 8.222 puntos (en valor absoluto).

En cuanto al intervalo de confianza de **Habilidad** en **Seminarios** (Nivel 3) con **Habilidad** en **Dinámica** (Nivel 1) oscila entre -7.149 y -2.851. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de efectividad del **Método de Dinámica** sobre la **Habilidad** es mayor que el del **Método de Seminarios** sobre la **Habilidad** por un mínimo de 2.851 y un máximo de 7.149 (en valor absoluto).

Por último podemos observar que el intervalo de confianza de **Habilidad** en **Seminarios** (Nivel 3) con **Habilidad** en **Casos** (Nivel 2) oscila entre -1.042 y 3.042. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de efectividad del **Método de Seminarios** sobre la **Habilidad** es igual que el del **Método de Casos** sobre la **Habilidad**.

3.2.3.1.2 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.3.1.2. Análisis de Varianza Multivariada (MANOVA)**

3.2.3.1.2. El Director de mercadotecnia de una compañía de Bienes Raíces desea medir el grado de influencia que llega a tener el estilo arquitectónico donde se localizan las casas unifamiliares de cierta ciudad en las magnitudes de los atributos de las mismas. Esta influencia es medida a través de tres variables dependientes: Valor comercial de la casa, Tamaño del lote y el Número de Recamaras. Los estilos arquitectónicos manejados son tres: Moderno, Hacienda y Colonial

El efecto se evalúa mediante tres criterios:

1.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Valor de las Casas?.

2.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Tamaño del Lote donde se ubican las casas?.

3.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Número de recámaras de las casas?.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Valor comercial de la casa(en millones de pesos) X_2= Tamaño del lote (en metros cuadrados) y X_3= Número de Recámaras, de 5 propietarios de casas Con estilo arquitectónico Moderno, 6 con estilo Hacienda y 7 Con estilo Colonial. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Estilo Arquitectónico de las Casas								
Moderno			Hacienda			Colonial		
Valor	Tamaño	Recámaras	Valor	Tamaño	Recámaras	Valor	Tamaño	Recámaras
3.74	600	4	6.10	600	5	4.58	1700	4
3.80	1100	4	5.50	750	3	3.50	1518	4
3.60	618	3	5.58	600	5	4.78	1218	4
4.36	600	4	4.98	840	4	5.34	1625	4
3.36	800	4	4.36	600	4	3.80	1000	3
			4.20	700	4	3.60	1125	5
						3.96	1600	4

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de las casas es el mismo en los tres estilos arquitectónicos?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el estilo arquitectónico influye sobre las magnitudes promedio de los atributos de las casas.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que estilo o estilos los atributos son mayores y en que magnitud.

Resolución del inciso a)

Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

Prueba de hipótesis:

1. Juego de Hipótesis

$$H_0 = \begin{bmatrix}\mu_{11}\\ \mu_{21}\\ \mu_{31}\end{bmatrix} = \begin{bmatrix}\mu_{12}\\ \mu_{22}\\ \mu_{32}\end{bmatrix} = \begin{bmatrix}\mu_{13}\\ \mu_{23}\\ \mu_{33}\end{bmatrix}$$

2. Estadístico de prueba

Como No. de Variables Dependientes : p≥1 y

No. de Niveles del Factor: g=3

Usamos la siguiente fórmula (Caso 4):

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Donde\ £ = \frac{|error|}{|tratamiento|} = \frac{|SCE|}{|SCT|} = 0.13562$$

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Fcalc = \frac{(18-3-2)}{3}\frac{(1-\sqrt{0.13562})}{\sqrt{0.13562}}$$

$$Fcalc = \frac{(13)}{3}\frac{(1-0.36827}{0..36827}$$

$$Fcalc = (4.333)(1.71540)$$

$$Fcalc = 7.43$$

3. Región de rechazo

$$Fcrit = \left[\propto ,2,p,2\left(\sum n_1 - p - 2\right)\right]$$

$$Fcrit = \left[0.05,2(3),2\left(\sum 18-3-2\right)\right]$$

$$Fcrit = [0.05,6,2(13)]$$

$$Fcrit = 0.05,6,26$$

$$F0.05 = 2.47$$

$$F0.01 = 3.59$$

4. Regla de Decisión

Se rechaza H_0 sí $fcal \geq 2.47$

5. Conclusión

Estadista: Como 7.43 > 2.47 > 3.59 ∴ la prueba es (A.S) y se rechaza con mucha potencia.

Administrativa: Existe evidencia suficiente para decir que estadísticamente que al menos una de las medias de los 3 grupos NO es igual.

Resolución del inciso b)

Prueba de Hipótesis para Valor:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el Estilo Arquitectónico de las casas es el mismo para el **Valor** de las casas.

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los Estilos Arquitectónicos utilizados en las casas son iguales para el **Valor** de las casas.

$$H_1: \text{No todos los Estilos Arquitectónicos son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Estilos Arquitectónicos utilizados en las casas para el **Valor** de las mismas es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio asignado al **Valor** con base los Estilos Arquitectónicos y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Estilos Arquitectónicos tiene una diferencia significativa para el **Valor** de las casas sobre los otros Estilos Arquitectónicos utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 11.461$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 5.304$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 6.158$$

Para encontrar el valor calculado de ***F,*** trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamien-tos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 5.304$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{5.304}/_{2} = 2.6519$	$F_{Calc.} = {}^{CMt}/_{CM}$ $= {}^{2.6519}/_{0.41}$ $= \mathbf{6.46}$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 6.158$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{6.158}/_{15}$ $= 0.4105$	
Total	18-1=17	$SCT = 11.461$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **6.46**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Estilo Arquitectónico utilizado para las casas no es el mismo en al menos uno de los tres Estilos Arquitectónicos utilizados.

Prueba de Hipótesis para Tamaño:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el Estilo Arquitectónico de las casas es el mismo para el **Tamaño** del lote.

$$H_0: \mu_{11.} = \mu_{12.} = \mu_{13.}$$

La hipótesis alternativa es No todos los Estilos Arquitectónicos utilizados en las casas son iguales para el **Tamaño** del lote.

$$H_1: \text{No todos los Estilos Arquitectónicos son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Estilos Arquitectónicos utilizados en las casas para el **Tamaño** del lote es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio asignado al **Tamaño** del lote con base los Estilos Arquitectónicos y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Estilos Arquitectónicos tiene una diferencia significativa para el **Tamaño** del lote sobre los otros Estilos Arquitectónicos utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 2\acute{}736{,}964$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 2\grave{}036{,}392$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 700{,}573$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	$F_{calculada}$
Tratamien-tos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 2\acute{}036{,}392$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{2\acute{}036{,}392}/_{2}$ $= 1\acute{}018{,}196$	$F_{Calc.} = {}^{CMt}/_{CME} = /$ $= {}^{1\acute{}018{,}196}/_{46{,}705}$ $= \mathbf{21.80}$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 700{,}573$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{700{,}573}/_{15}$ $= 46{,}705$	
Total	18-1=17	SCT $= 2\acute{}736{,}964$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **21.8**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que NO todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Estilo Arquitectónico utilizado en las casas NO es el mismo en los tres Estilos Arquitectónicos utilizados.

Prueba de Hipótesis para Recámaras:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el Estilo Arquitectónico de las casas es el mismo para el Número de **Recámaras** de las casas.

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los Estilos Arquitectónicos utilizados en las casas son iguales para el Número de **Recámaras** de las casas.

$$H_1: \text{No todos los Estilos Arquitectónicos son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Estilos Arquitectónicos utilizados en las casas para el Número de **Recámaras** de las mismas es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio asignado al Número de **Recámaras** con base los Estilos Arquitectónicos y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Estilos Arquitectónicos tiene una diferencia significativa para el Número de **Recámaras** de las casas sobre los otros Estilos Arquitectónicos utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 6$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 0.3667$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k} \sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 5.6333$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamientos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 0.3667$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{0.3667}/_{2}$ $= 0.1833$	$F_{Calc.} = {}^{CMt}/_{CME} = /$ $= {}^{0.1833}/_{0.37562}$ $= 0.49$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 5.6333$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{5.6333}/_{15}$ $= 0.3756$	
Total	18-1=17	$SCT = 6$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de 0.49, que es menor al valor crítico de 3.68; por lo tanto la hipótesis nula NO se rechaza y llegamos a la conclusión de que todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Estilo Arquitectónico utilizado para casas es el mismo en los tres Estilos Arquitectónicos utilizados.

Resolución del inciso c)

Prueba de Tukey

Intervalo de confianza para Valor (Moderno-Hacienda) .Modelo desbalanceado.

$$M_{11} - M_{12} = (\bar{X}_{11} - \bar{X}_{12}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = (3.771 - 5.120) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = -1.349 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{6.158}{15} = \underline{0.4105}$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{11} - M_{12} = -1.349 \mp 3.67\sqrt{\frac{0.4105}{5.4545}}$$

$$M_{11} - M_{12} = -1.349 \mp 3.67\sqrt{0.07526}$$

$$M_{11} - M_{12} = -1.349 \mp 3.67(0.27433)$$

$$M_{11} - M_{12} = -1.349 \mp 1.00680 \begin{cases} LIC = -2.3558 \\ LSC = -0.34220 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos podemos decir que estadísticamente el grado de efectividad del **Estilo Hacienda sobre el Valor** es mayor que el del **Estilo Moderno sobre el Valor** por un mínimo de 0.34220 y un máximo de 2.3558 millones de pesos. (en Valor absoluto)

Intervalo de confianza para Valor (Moderno-Colonial). Modelo desbalanceado

$$M_{11} - M_{13} = (\bar{X}_{11} - \bar{X}_{13}) \mp q \propto, (k.N-k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = (3.771 - 4.223) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = -0.452 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = 5.8333$$

$$M_{11} - M_{13} = -0.452 \mp 3.67\sqrt{\frac{0.4105}{5.8333}}$$

$$M_{11} - M_{13} = -0.452 \mp 3.67(0.26528)$$

$$M_{11} - M_{13} = -0.452 \mp 0.97357 \begin{cases} LIC = -1.42557 \\ LSC = 0.52157 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de influencia del **Estilo Moderno sobre el Valor** es el mismo que el del **Estilo Colonial sobre el Valor.**

Intervalo de confianza para Valor (Hacienda-Colonial). Modelo desbalanceado

$$M_{12} - M_{13} = (\bar{X}_{12} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = (5.120 - 4.223) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = 0.897 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = 6.4615$$

$$M_{12} - M_{13} = 0.897 \mp 3.67\sqrt{\frac{0.4105}{6.4615}}$$

$$M_{12} - M_{13} = 0.897 \mp 3.67(0.25205)$$

$$M_{12} - M_{13} = 0.897 \mp 0.92503 \begin{cases} LIC = -0.02803 \\ LSC = 1.82203 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de influencia del **Estilo Hacienda sobre el Valor** es el mismo que el del **Estilo Colonial sobre el Valor**

Intervalo de confianza para Tamaño (Moderno-Hacienda). Modelo desbalanceado

$$M_{21} - M_{22} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = (743.6 - 681.7) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = 61.9 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{700{,}573}{15} = \underline{46{,}705}$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{21} - M_{22} = 61.9 \mp 3.67\sqrt{\frac{46{,}705}{5.4545}}$$

$$M_{21} - M_{22} = 61.9 \mp 3.67(92.53461)$$

$$M_{21} - M_{22} = 61.9 \mp 339.60203 \begin{cases} 277.70203 \\ 401.50203 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de influencia del **Estilo Arquitectónico Moderno sobre el Tamaño del lote** es mayor que el del **Estilo Arquitectónico Hacienda sobre el Tamaño del Lote** por un mínimo de 277.7020. y un máximo de 401.50203 metros cuadrados.

Intervalo de confianza para Tamaño (Moderno-Colonial). Modelo desbalanceado.

$$M_{21} - M_{23} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N-k)\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = (743.6 - 1398) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = -654.4 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{21} - M_{23} = -654.4 \mp 3.67\sqrt{\frac{46{,}705}{5.8326}}$$

$$M_{21} - M_{23} = -654.4 \mp 3.67(89.48507)$$

$$M_{21} - M_{23} = -654.4 \mp 328.41021 \begin{cases} -982.81021 \\ -325.98979 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos podemos decir que estadísticamente el grado de influencia del **Estilo Arquitectónico sobre el Tamaño del Lote** es mayor que el del **Estilo Arquitectónico Colonial sobre el Tamaño del Lote** por un mínimo de 325.98979 y un máximo de 982.81021 metros cuadrados. (En valor absoluto).

Intervalo de confianza para Tamaño (Hacienda-Colonial). Modelo desbalanceado.

$$M_{22} - M_{23} = (\bar{X}_{22} - \bar{X}_{23}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = (681.7 - 1398) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = -716.3 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{22} - M_{23} = -716.3 \mp 3.67\sqrt{\frac{46{,}705}{6.4620}}$$

$$M_{22} - M_{23} = -716.3 \mp 3.67(85.01552)$$

$$M_{22} - M_{23} = -716.3 \mp 312.00696 \begin{cases} LIC = -1{,}028.30696 \\ LSC = -404.29304 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos podemos decir que estadísticamente el grado de influencia del **Estilo Arquitectónico Colonial sobre el Tamaño del Lote** es mayor que el del **Estilo Arquitectónico Hacienda** por un mínimo de 404.29304 y un máximo de 1,028.30696 metros cuadrados. (En valor absoluto).

Interpretación de confianza para Recámaras (Moderno-Hacienda). Modelo desbalanceado

$$M_{31} - M_{32} = (\bar{X}_{31} - \bar{X}_{32}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = (3.8 - 4.167) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = -0.367 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{5.6333}{15} = 0.3756$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{31} - M_{32} = -0.367 \mp 3.67\sqrt{\frac{0.3756}{5.4545}}$$

$$M_{31} - M_{32} = -0.367 \mp 3.67(0.26241)$$

$$M_{31} - M_{32} = -0.367 \mp 0.96303 \begin{cases} LIC = -1.33003 \\ LSC = 0.59603 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Arquitectónico Moderno sobre el Número de Recamaras** es el mismo que el del **Estilo Arquitectónico Hacienda sobre el Número de recámaras**.

Intervalo de confianza para Recámaras (Moderno-Colonial). Modelo desbalanceado.

$$M_{31} - M_{33} = (\bar{X}_{31} - \bar{X}_{33}) \mp q \propto, (k.N-k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = (3.8-4) \mp 0.05,3,15\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = -0.2 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{31} - M_{33} = -0.2 \mp 3.67\sqrt{\frac{0.3756}{5.8326}}$$

$$M_{31} - M_{33} = -0.2 \mp 3.67(0.25376)$$

$$M_{31} - M_{33} = -0.2 \mp 0.93132 \begin{cases} -1.13132 \\ 0.73132 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Arquitectónico Moderno sobre el Número de Recamaras** es el mismo que el del **Estilo**

Arquitectónico Colonial sobre el Número de recámaras.

Intervalo de confianza para Recámaras (Hacienda-Colonial). Modelo desbalanceado.

$$M_{32} - M_{33} = (\bar{X}_{32} - \bar{X}_{33}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = (4.167 - 4) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = 0.167 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{32} - M_{33} = 0.167 \mp 3.67\sqrt{\frac{0.3752}{6.4620}}$$

$$M_{32} - M_{33} = 0.167 \mp 3.67(0.24096)$$

$$M_{32} - M_{33} = 0.167 \mp 0.88433 \begin{cases} LIC = -0.71733 \\ LSC = 1.05133 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de influencia del **Estilo Arquitectónico Hacienda sobre el Número de Recámaras** es el mismo que el del **Estilo Arquitectónico Colonial sobre el Número de Recámaras.**

3.2.3.1.2 EJERCICIO DE **REFUERZO con Minitab 17**

EJERCICIO DE **REFUERZO 3.2.3.1.2. Análisis de Varianza Multivariada (MANOVA) con Minitab 17**

3.2.3.1.2. El Director de mercadotecnia de una compañía de Bienes Raíces desea medir el grado de influencia que llega a tener el estilo arquitectónico donde se localizan las casas unifamiliares de cierta ciudad en las magnitudes de los atributos de las mismas. Esta influencia es medida a través de tres variables dependientes: Valor comercial de la casa, Tamaño del lote y el Número de Recamaras. Los estilos arquitectónicos manejados son tres: Moderno, Hacienda y Colonial

El efecto se evalúa mediante tres criterios:

1.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Valor de las Casas?.

2.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Tamaño del Lote donde se ubican las casas?.

3.- ¿Que tanto el Estilo Arquitectónico de las Casas llega a influir sobre el Número de recámaras de las casas?.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Valor comercial de la casa(en millones de pesos) X_2= Tamaño del lote (en metros cuadrados) y X_3= Número de Recámaras, de 5 propietarios de casas Con estilo arquitectónico Moderno, 6 con estilo Hacienda y 7 Con estilo Colonial. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

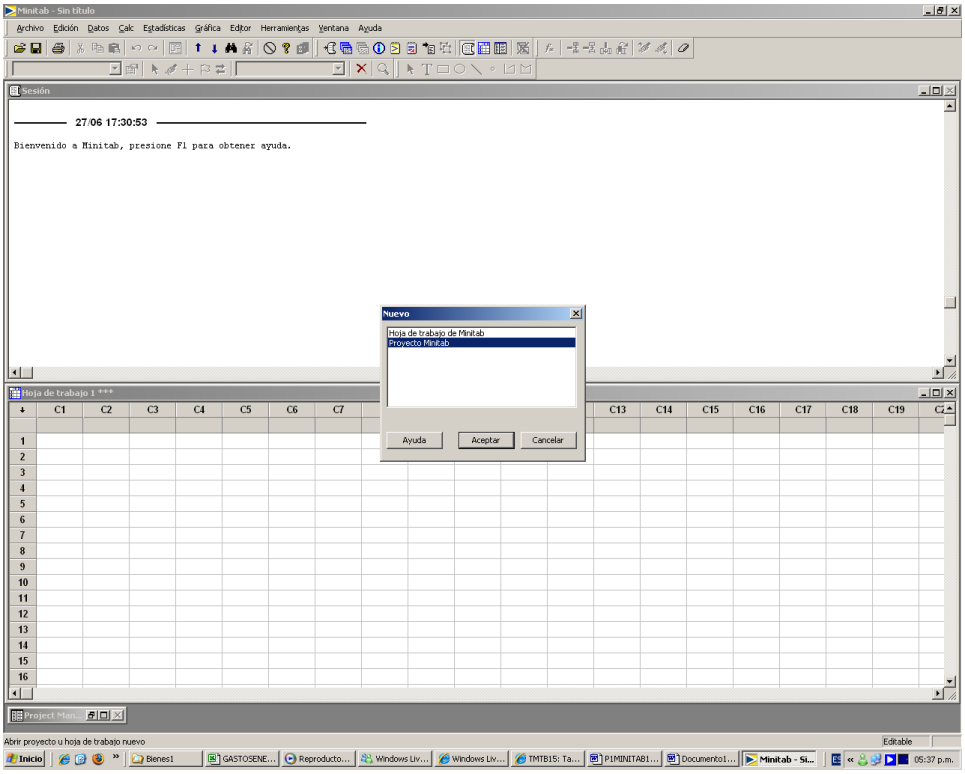
Estilo Arquitectónico de las Casas								
Moderno			Hacienda			Colonial		
Valor	Tamaño	Recámaras	Valor	Tamaño	Recámaras	Valor	Tamaño	Recámaras
3.74	600	4	6.10	600	5	4.58	1700	4
3.80	1100	4	5.50	750	3	3.50	1518	4
3.60	618	3	5.58	600	5	4.78	1218	4
4.36	600	4	4.98	840	4	5.34	1625	4
3.36	800	4	4.36	600	4	3.80	1000	3
			4.20	700	4	3.60	1125	5
						3.96	1600	4

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de las casas es el mismo en los tres estilos arquitectónicos?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el estilo arquitectónico influye sobre las magnitudes promedio de los atributos de las casas. Y realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que estilo o estilos los atributos son mayores y en que magnitud.

Resolución del inciso a)

Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

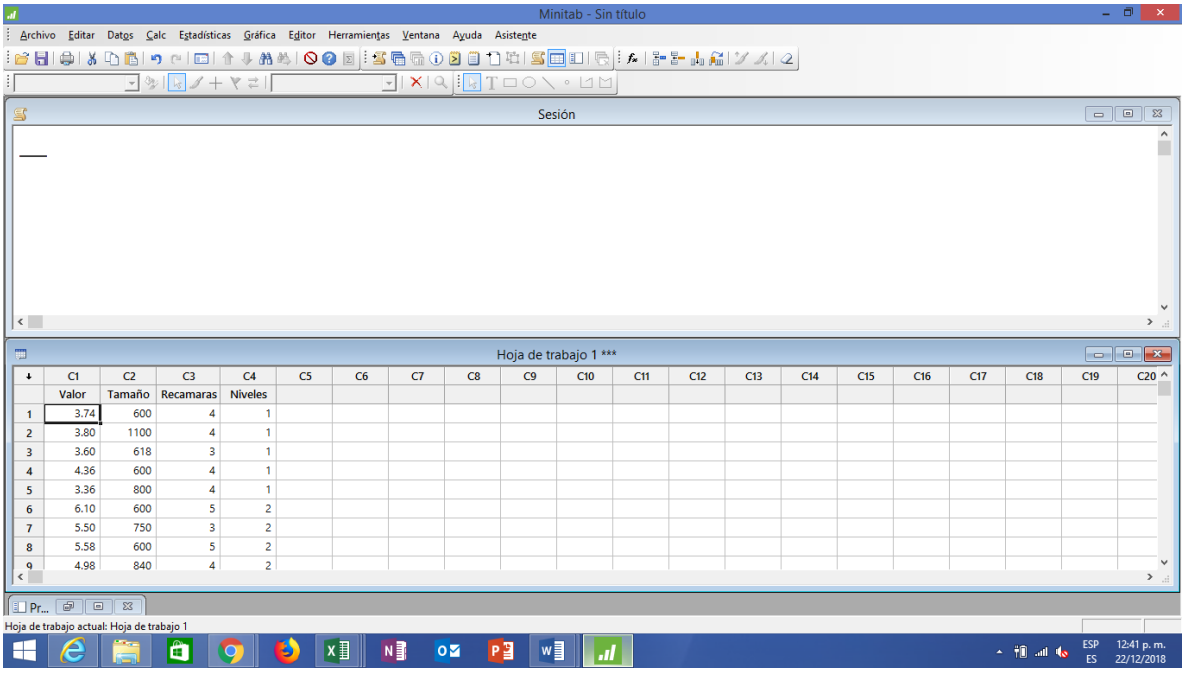
Para crear la base de datos del ejemplo ilustrativo en **MINITAB17**, accede a **MINITAB VERSIÓN 17.** Cuando aparezca la pantalla da un clic en la opción **Archivo** de la barra menú y selecciona **Nuevo**. Aparecerá la ventana **Nuevo. Selecciona la opción Proyecto Minitab y Aceptar**



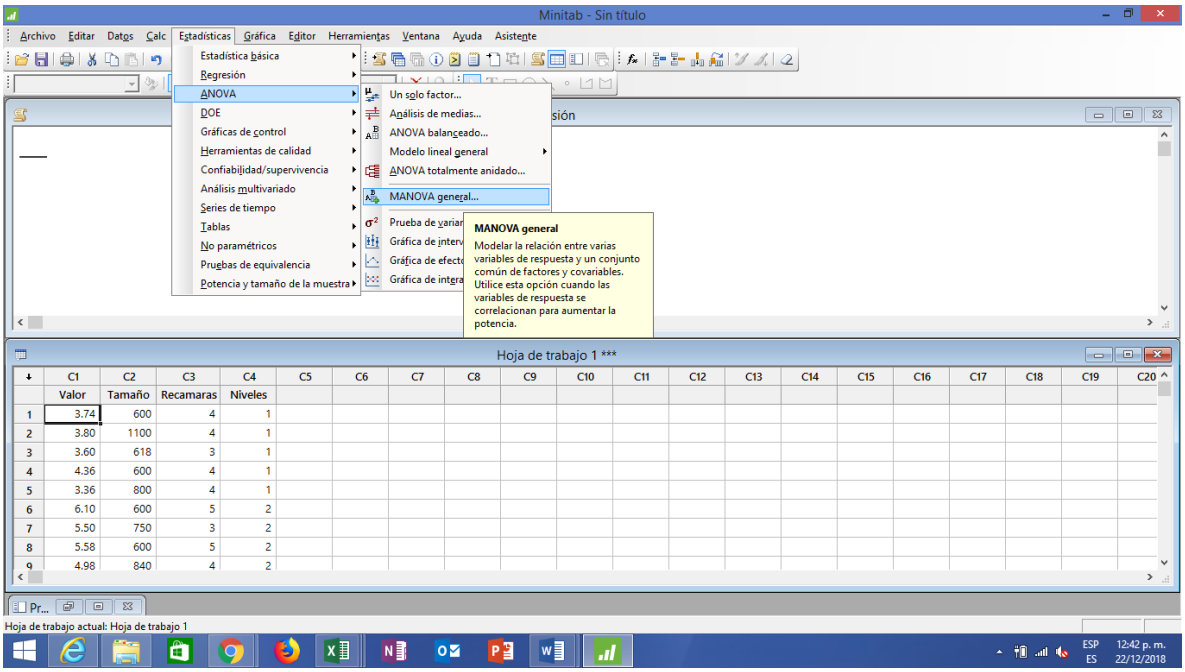
Procede a editar la base de datos de la siguiente manera:

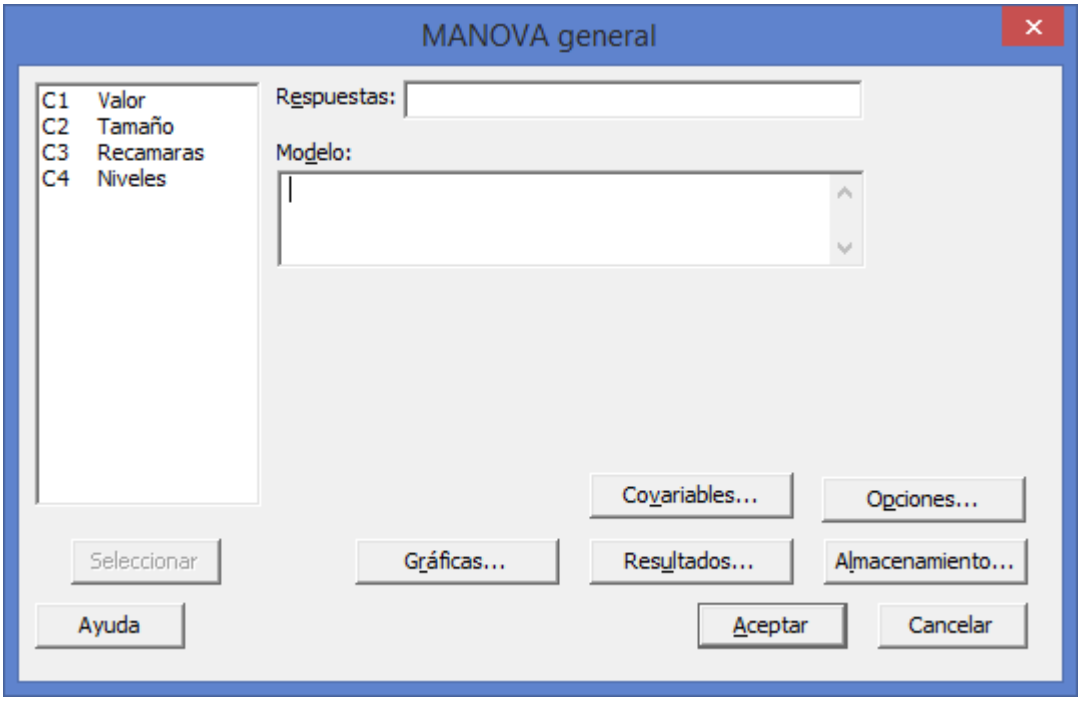
Coloca los nombres de cada **variable (En este caso Valor, Tamaño, Recámaras y Niveles**), debajo de cada número de columna, en este caso **C1, C2, C3 y C4**.

Una vez termines de colocar los nombres de cada **variable (En este caso Valor, Tamaño, Recámaras y Niveles**, coloca el cursor en el primer nombre de la variable **(Valor)** y llena los datos de Valor en forma vertical de los tres Estilos Arquitectónicos utilizados (Moderno, Hacienda y Colonial). Cuando se terminen los 18 datos coloca el cursor en el segundo nombre de la variable (**Tamaño**) y llena los 18 datos. Cuando se terminen los 18 datos de Tamaño coloca el cursor en el tercer nombre de la variable **(Recámaras**) y llena los 18 datos. Cuando se terminen los 18 datos de Recámaras coloca el cursor en la cuarta columna (**Niveles**) y coloca en forma vertical cinco 1, seis 2 y siete 3. El número 1 corresponde al Estilo **Moderno,** el 2 corresponde al Estilo **Hacienda** y el 3 corresponde al Estilo **Colonial.**



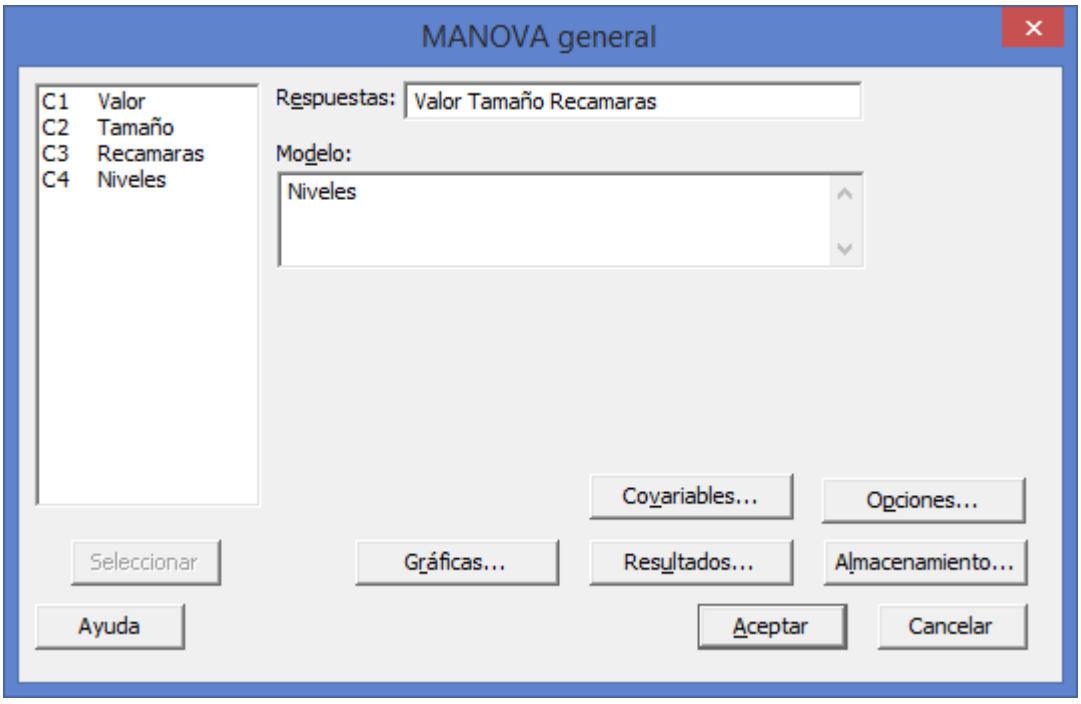
5 Elija **Estadísticas > ANOVA > MANOVA general**.





2 En **Respuestas**, ingrese las 3 variables o columnas numéricas que contengan las variables de respuesta (**Valor, Tamaño , Recámaras**).

3 En **Factores**, ingresa la cuarta columna (**Niveles**)



4 Haga clic en **Aceptar**.

Salida de la ventana Sesión

Modelo lineal general: Valor, Tamaño, Recamaras vs. Niveles

```
MANOVA para Niveles
s = 2    m = 0.0    n = 5.5

                 Estadística             GL
Criterio           de prueba      F  Núm  Denom      P
De Wilks             0.13562  7.434    6     26  0.000
Lawley-Hotelling     4.03518  8.070    6     24  0.000
De Pillai            1.18151  6.736    6     28  0.000
De Roy               3.33377
```

Interpretación de los resultados

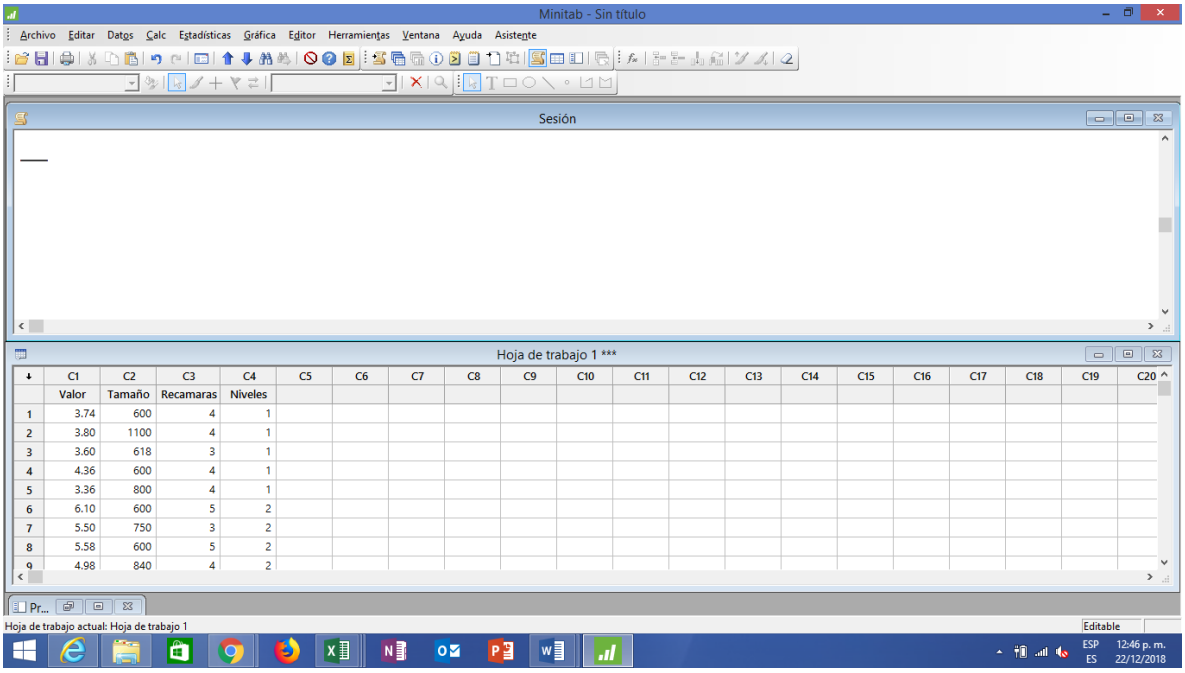
La salida de la ventana sesión presenta los resultados de cuatro criterios o pruebas entre ellas el de **Lambda de Wilks** que arroja una Estadística de prueba de **0.13562** con una Fcalculada de 7.434. El valor de p de 0.000 (nivel de significancia observado) indica que hay suficiente evidencia de que no todas las medias en cada variable son iguales en los tres estilos utilizados cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Resolución del inciso b)

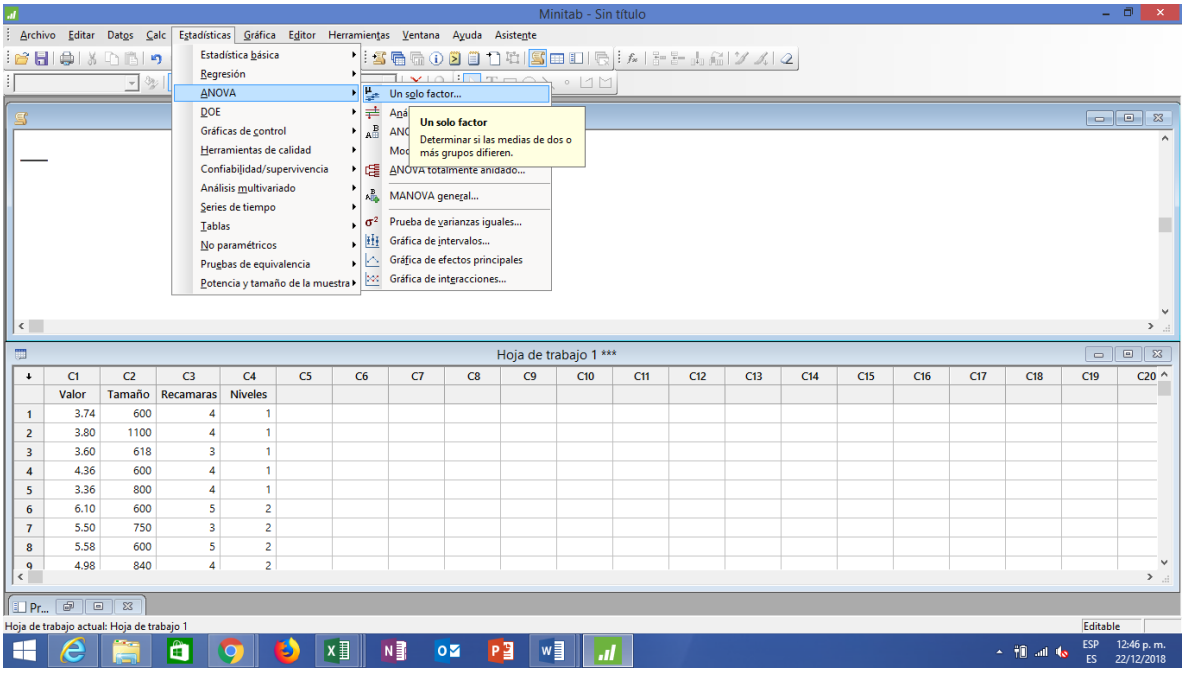
1. Prueba ANOVA de Hipótesis para Valor y pruebas Post-hoc de Tukey para Valor:

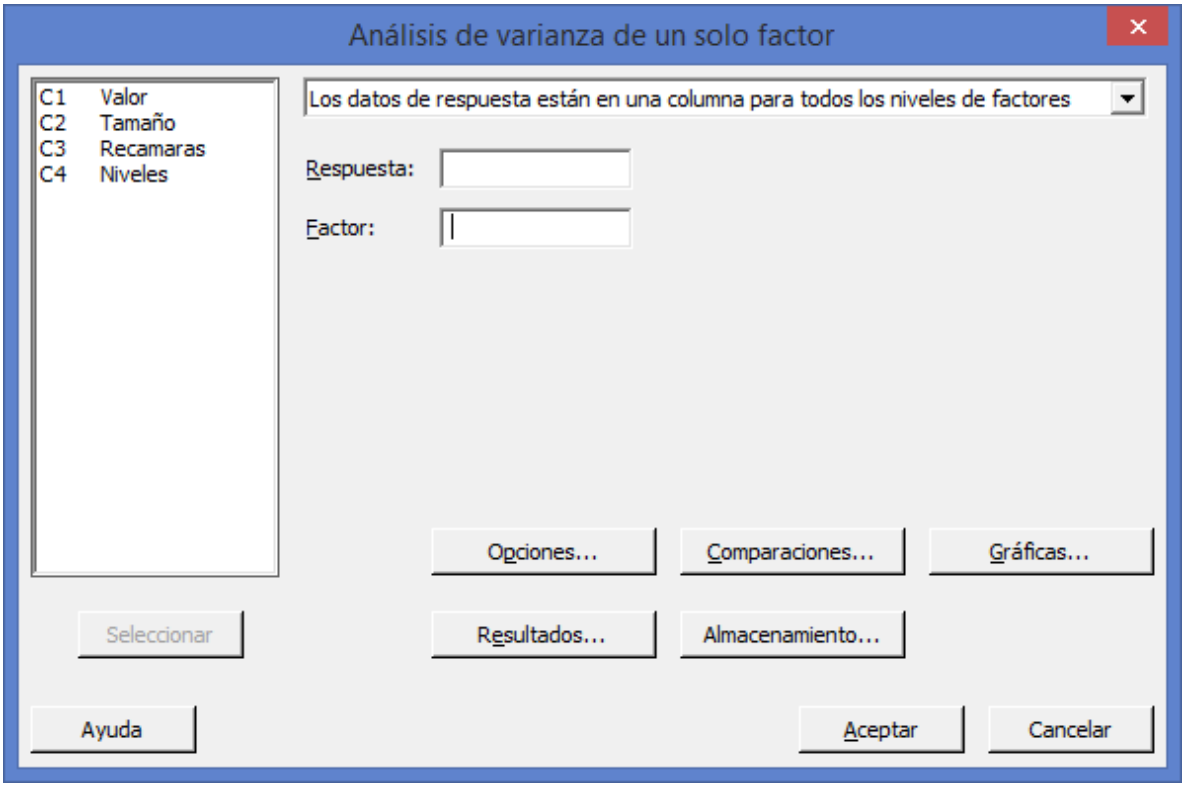
Cuando el número de observaciones en cada tratamiento es extenso y/o existen muchos tratamientos, los cálculos manuales son tediosos. Existen muchos paquetes de software que pueden mostrar los resultados entre ellos Minitab (Versión 17).

Comenzamos introduciendo los datos en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:



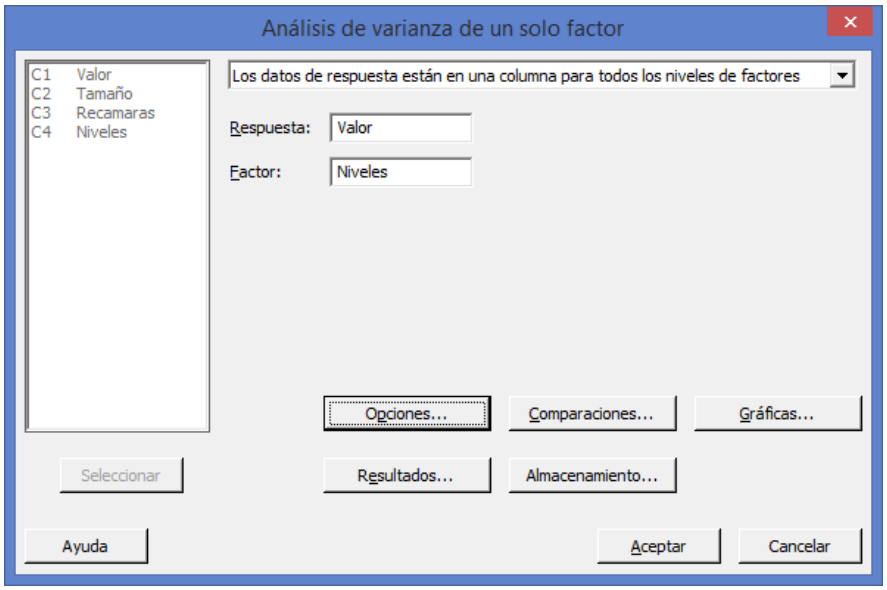
Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**



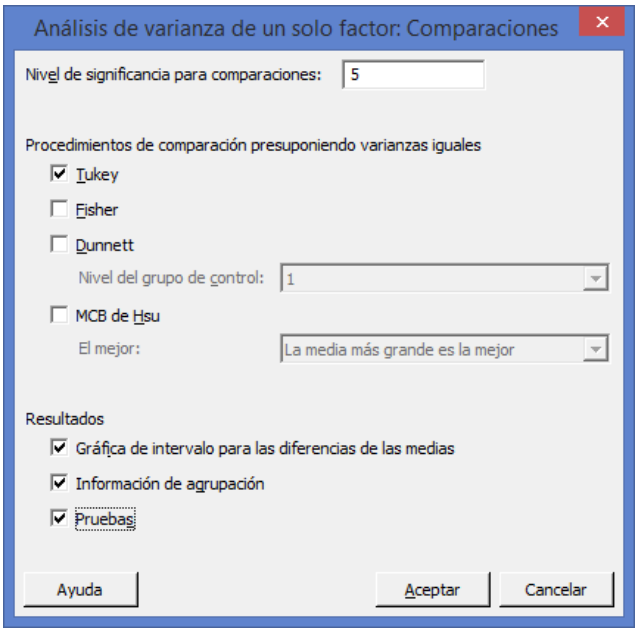


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Valor**. En Factor, ingrese **Niveles**.



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Valor vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2      5.304     2.6519     6.46    0.009
Error    15      6.158     0.4105
Total    17     11.461
```

```
Resumen del modelo

                      R-cuad.  R-cuad.
       S  R-cuad.  (ajustado)   (pred)
0.640707   46.28%      39.11%   24.01%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  3.771      0.371  (3.160, 4.382)
2        6  5.120      0.743  (4.562, 5.677)
3        7  4.223      0.689  (3.707, 4.739)

Desv.Est. agrupada = 0.640707
```

Interpretación de los resultados

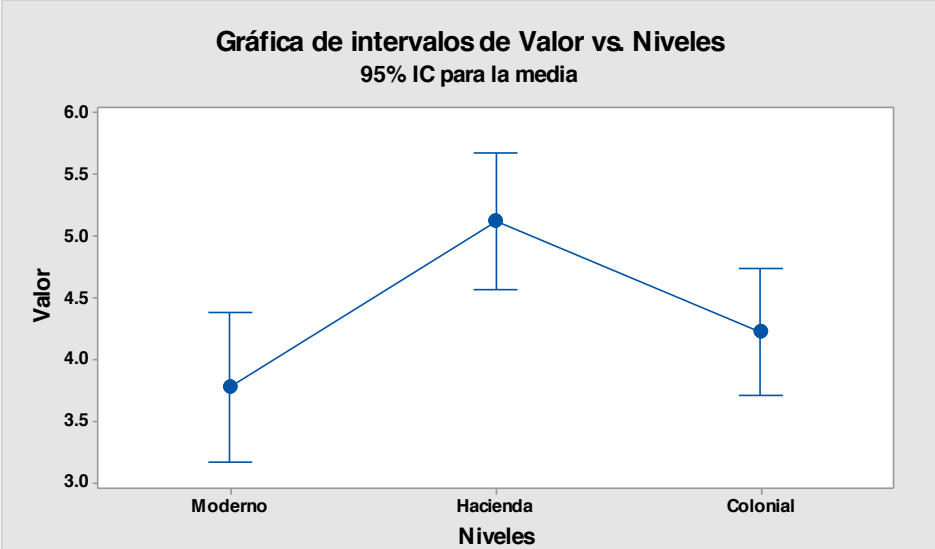
En la Tabla de ANOVA, para **Valor** en **los tres Estilos Arquitectónicos** utilizados el valor de F calculada de 6.46 con un valor p (0.009) indica que hay suficiente evidencia de que no todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Comparaciones en parejas de Tukey

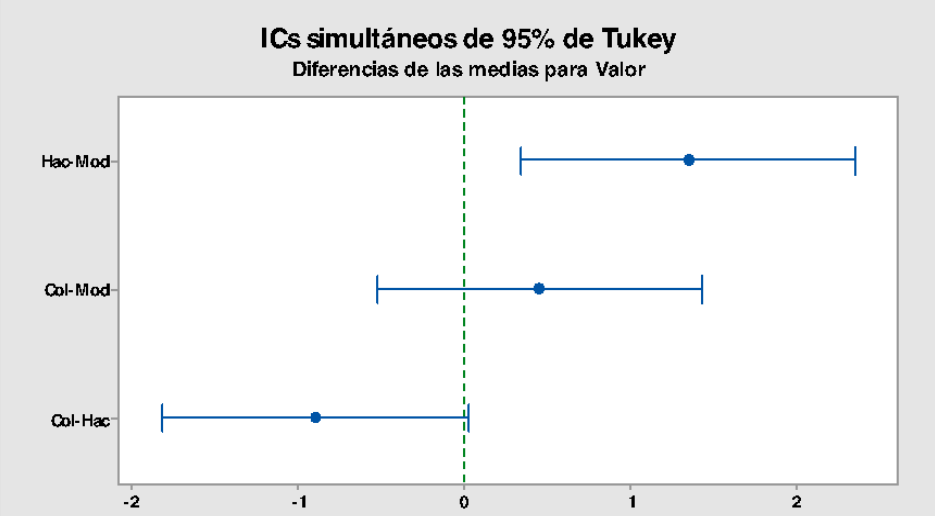
```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N  Media  Agrupación
2        6  5.120  A
3        7  4.223  A B
1        5  3.771    B

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            1.348       0.388  ( 0.342, 2.355)     3.48     0.009
3 - 1            0.452       0.375  (-0.522, 1.425)     1.20     0.469
3 - 2           -0.897       0.356  (-1.822, 0.028)    -2.52     0.058

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Valor vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

Si un intervalo no contiene cero, las medias correspondientes son significativamente diferentes.

Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Valor** en **Hacienda** (Nivel 2) con **Valor** en **Moderno** (Nivel 1) oscila entre 0.342 y 2.355. Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de influencia del **Estilo Hacienda** sobre **Valor** es mayor que el **Estilo Moderno** sobre **Valor** por un mínimo de 0.342 y un máximo

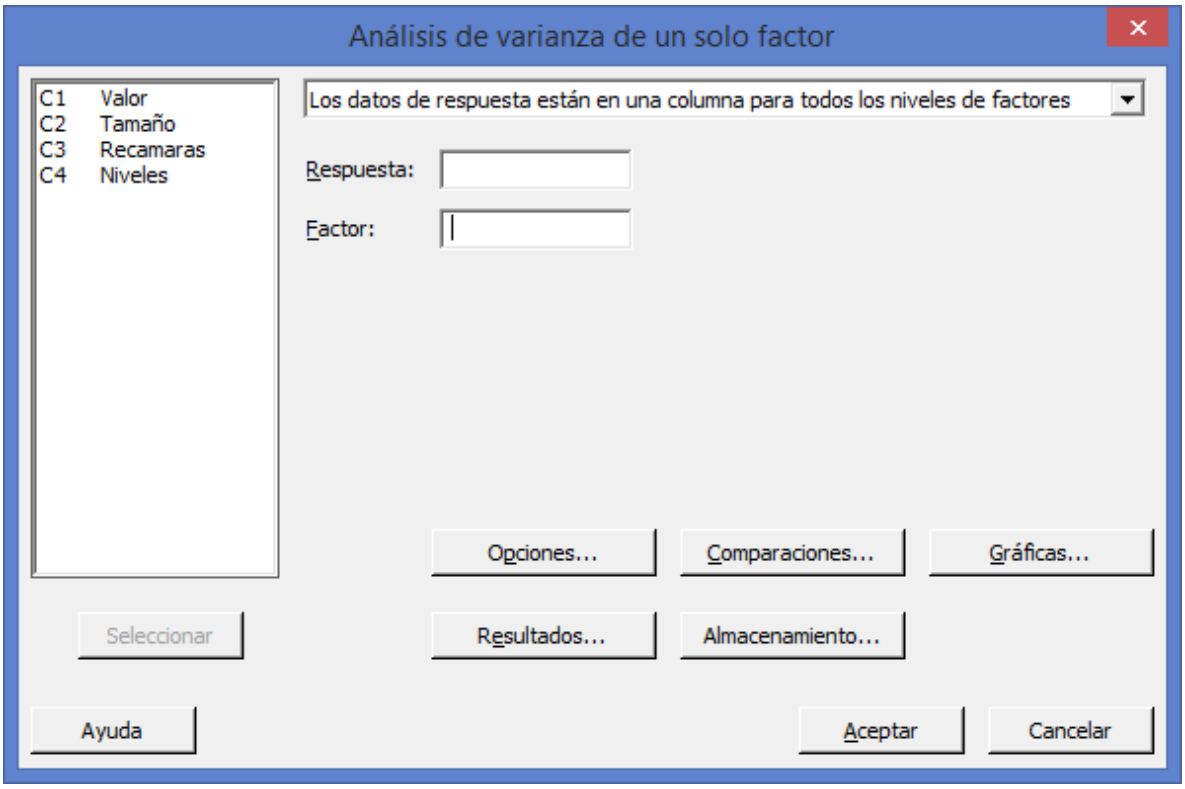
de 2.355 millones de pesos.

En cuanto al intervalo de confianza de **Valor** en **Colonial** (Nivel 3) con **Valor** en **Moderno** (Nivel 1) oscila entre -0.522 y 1.425. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Colonial** sobre **Valor** es igual que el del **Estilo Moderno** sobre **Valor.**

Por último podemos observar que el intervalo de confianza de **Valor** en **Colonial** (Nivel 3) con **Valor** en **Hacienda** (Nivel 2) oscila entre -1.822 y 0.028. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Colonial** sobre **Valor** es igual que el del **Estilo Hacienda** sobre **Valor**

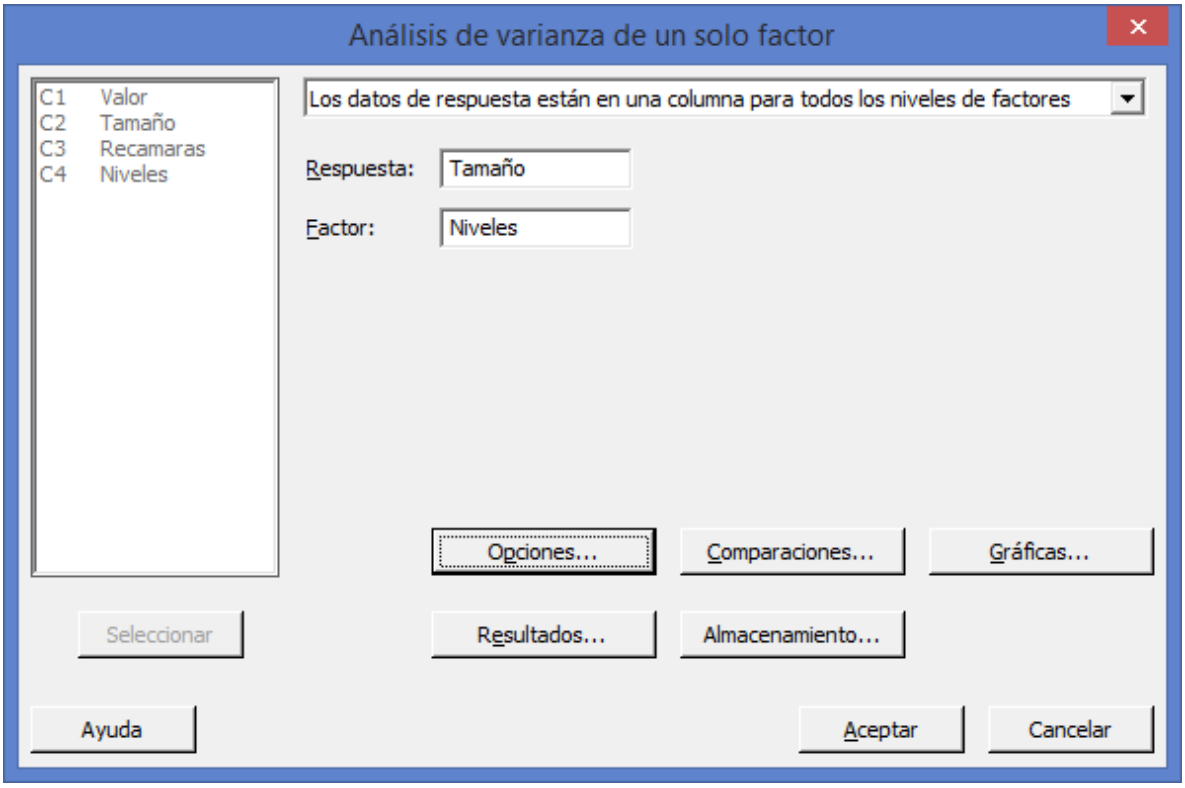
2. Prueba ANOVA de Hipótesis para Tamaño y pruebas Post-hoc de Tukey para Tamaño:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

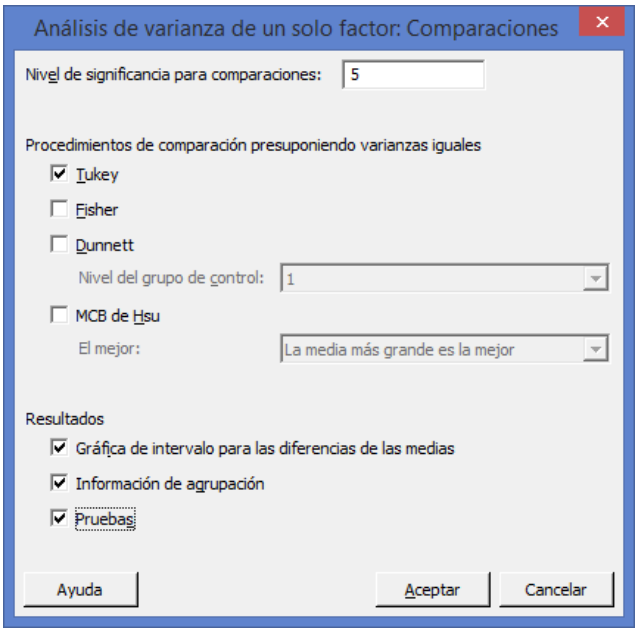


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Tamaño**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Tamaño vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2    2036392    1018196    21.80    0.000
Error    15     700573      46705
Total    17    2736964


Resumen del modelo

                    R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
216.113   74.40%      70.99%   63.64%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  743.6      216.3  (537.6, 949.6)
2        6  681.7      100.1  (493.6, 869.7)
3        7   1398        278  ( 1224,  1572)

Desv.Est. agrupada = 216.113
```

Interpretación de los resultados

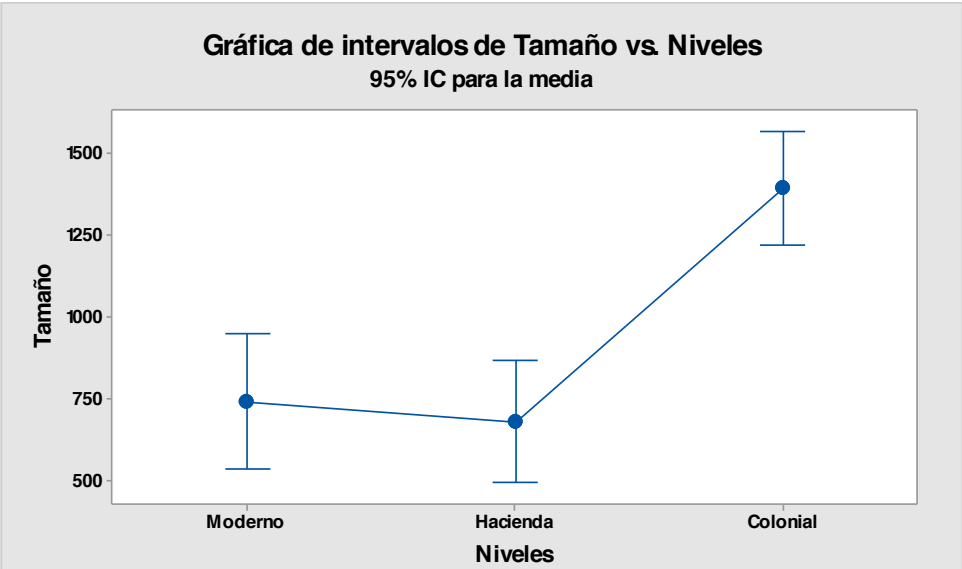
En la Tabla de ANOVA, para **Tamaño** en **los tres Estilos Arquitectónicos** utilizados el valor de F calculada de 21.8 con un valor p (0.000) indica que hay suficiente evidencia de que no todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Comparaciones en parejas de Tukey

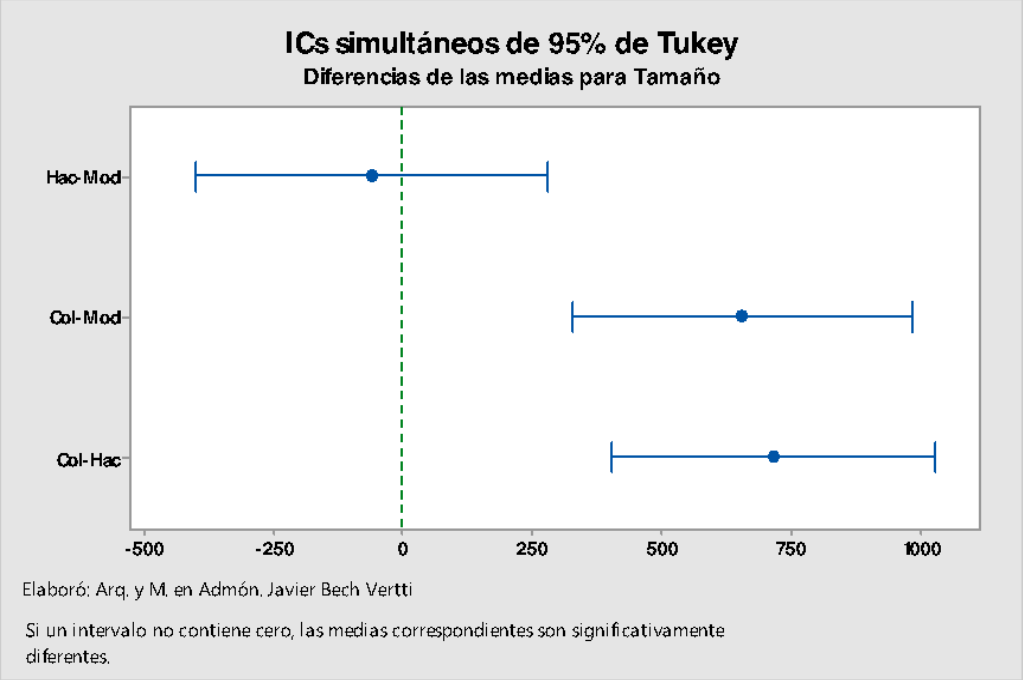
```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N  Media  Agrupación
3        7   1398  A
1        5  743.6    B
2        6  681.7    B

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                          Valor p
de niveles      medias  diferencia    IC de 95%   Valor T  ajustado
2 - 1              -62         131  (-402,  278)    -0.47     0.885
3 - 1              654         127  ( 326,  983)     5.17     0.000
3 - 2              716         120  ( 404, 1028)     5.96     0.000

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Tamaño vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.



Interpretación de los resultados

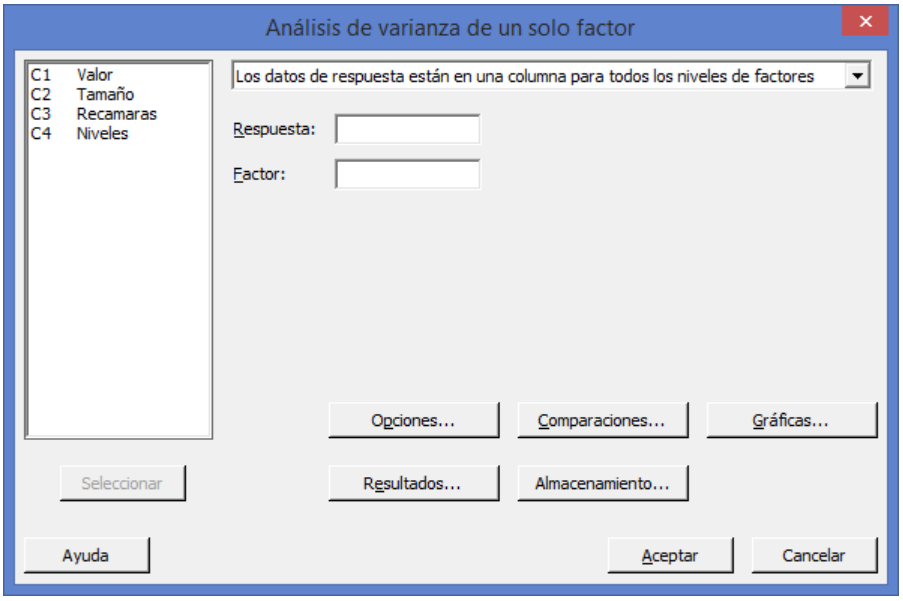
De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Tamaño** en **Hacienda** (Nivel 2) con **Tamaño** en **Moderno** (Nivel 1) oscila entre -402 y 278. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Hacienda** sobre **Tamaño** es igual que el del **Estilo Moderno** sobre **Tamaño.**

En cuanto al intervalo de confianza de **Tamaño** en **Colonial** (Nivel 3) con **Tamaño** en **Moderno** (Nivel 1) oscila entre 326 y 983. Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de influencia del **Estilo Colonial** sobre **Tamaño** es mayor que el **Estilo Moderno** sobre **Tamaño** por un mínimo de 326 y un máximo de 983 metros cuadrados.

Por último podemos observar que el intervalo de confianza de **Tamaño** en **Colonial** (Nivel 3) con **Tamaño** en **Hacienda** (Nivel 2) oscila entre 404 y 1028. Como ambos límites de confianza son positivos podemos decir que estadísticamente el grado de influencia del **Estilo Colonial** sobre **Tamaño** es mayor que el **Estilo Hacienda** sobre **Tamaño** por un mínimo de 404 y un máximo de 1028 metros cuadrados.

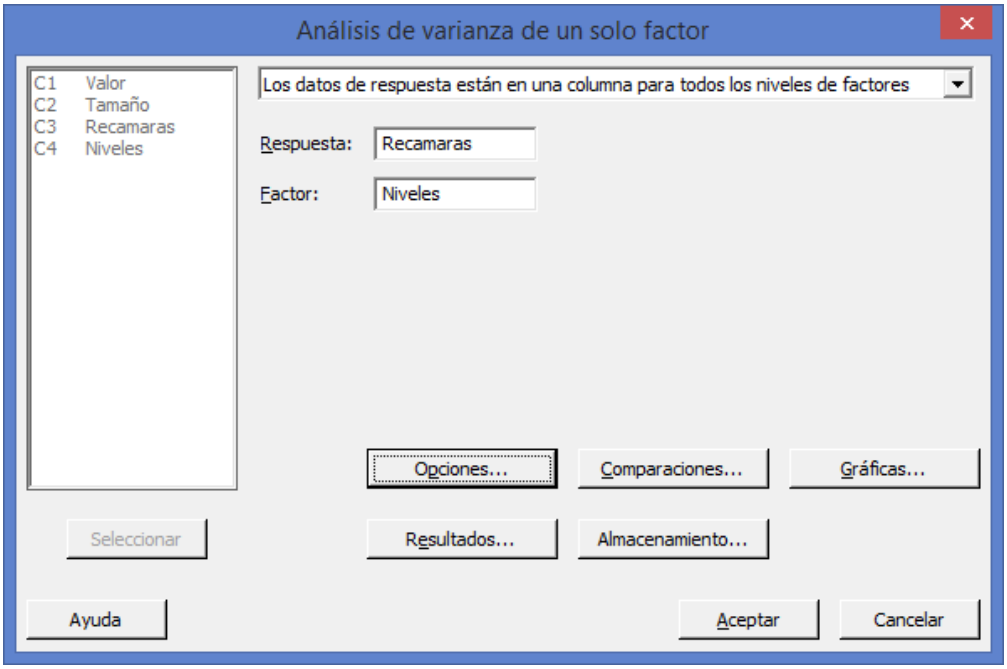
3. Prueba ANOVA de Hipótesis para Recámaras y pruebas Post-hoc de Tukey para Recámaras:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

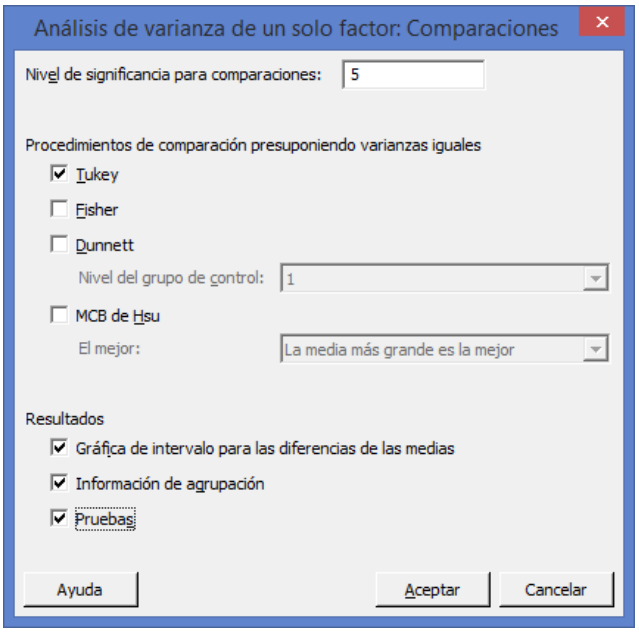


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Recámaras**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Recamaras vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2     0.3667     0.1833     0.49    0.623
Error    15     5.6333     0.3756
Total    17     6.0000
```

```
Resumen del modelo

                      R-cuad.  R-cuad.
       S  R-cuad.   (ajustado)   (pred)
0.612826    6.11%       0.00%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  3.800      0.447  (3.216, 4.384)
2        6  4.167      0.753  (3.633, 4.700)
3        7  4.000      0.577  (3.506, 4.494)

Desv.Est. agrupada = 0.612826
```

Interpretación de los resultados

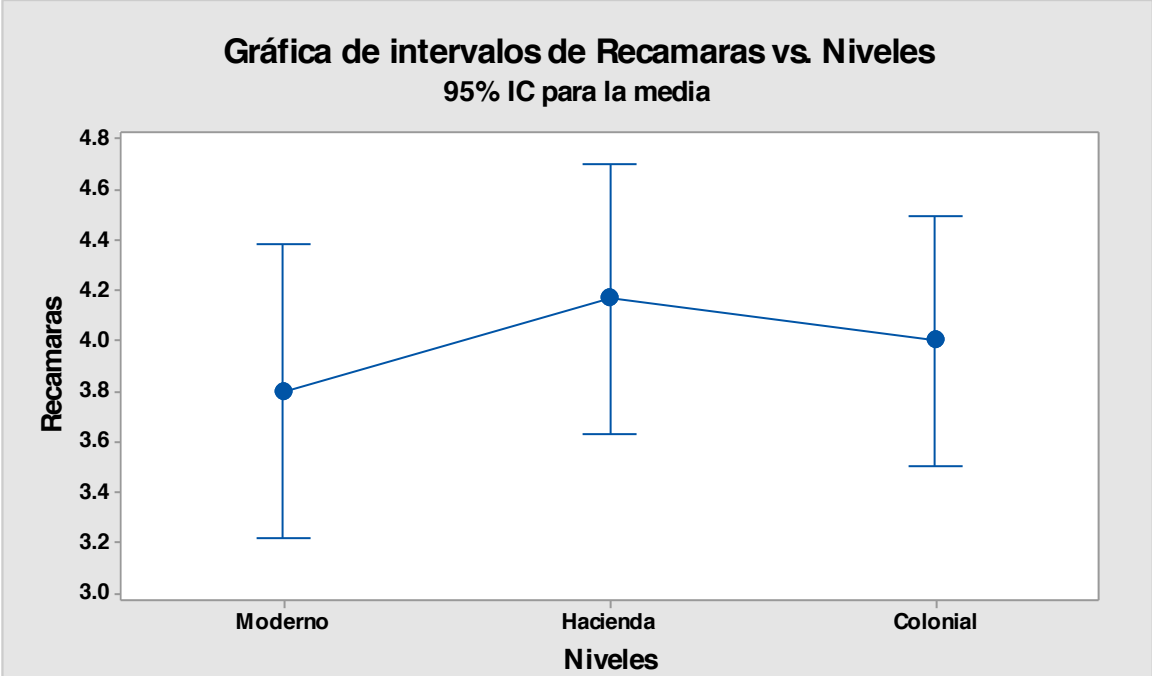
En la Tabla de ANOVA, para **Recámaras** en **los tres Estilos Arquitectónicos** utilizados el valor de F calculada de 0.49 con un valor p (0.623) indica que hay suficiente evidencia de que todas las medias son iguales cuando alfa se establece en 0.05. En este caso en particular como no hay diferencias entre medias no se deberían examinar los resultados de las comparaciones múltiples, sin embargo como ejercicio se llevará a cabo para comprobar que los tres intervalos de confianza deben pasar por cero.

Comparaciones en parejas de Tukey

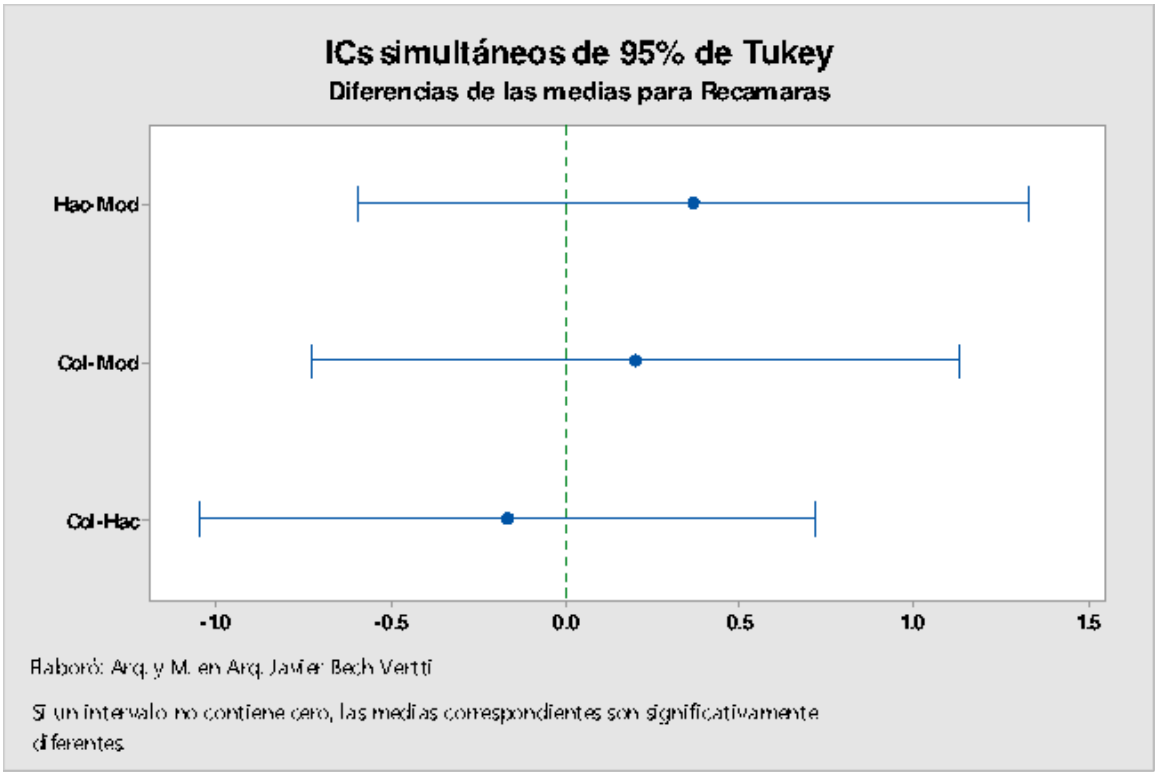
```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N  Media  Agrupación
2        6  4.167  A
3        7  4.000  A
1        5  3.800  A

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1            0.367       0.371  (-0.596, 1.330)     0.99     0.595
3 - 1            0.200       0.359  (-0.731, 1.131)     0.56     0.844
3 - 2           -0.167       0.341  (-1.051, 0.718)    -0.49     0.878

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Recamaras vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.



Elaboró: Arq. y M. en Arq. Javier Bech Vertti

Si un intervalo no contiene cero, las medias correspondientes son significativamente diferentes.

Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Recámaras** en **Hacienda** (Nivel 2) con **Recamaras** en **Moderno** (Nivel 1) oscila entre -0.596 y 1.33. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Hacienda** sobre **Recámaras** es igual que el del **Estilo Moderno** sobre **Recámaras.**

En cuanto al intervalo de confianza de **Recámaras** en **Colonial** (Nivel 3) con **Recámaras** en **Moderno** (Nivel 1) oscila entre -731 y 1.131. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Colonial** sobre **Recámaras** es igual que el del **Estilo Moderno** sobre **Recámaras**

Por último podemos observar que el intervalo de confianza de **Recámaras** en **Colonial** (Nivel 3) con **Recámaras** en **Hacienda** (Nivel 2) oscila entre -1.051 y 0.718. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Estilo Colonial** sobre **Recámaras** es igual que el del **Estilo Hacienda** sobre **Recámaras**

3.2.3.1.3 EJERCICIO DE **REFUERZO**

EJERCICIO DE **REFUERZO 3.2.3.1.3. Análisis de Varianza Multivariada (MANOVA)**

3.2.3.1.3. El Director de Industrias México desea medir el grado de influencia que llega a tener el Nivel de Satisfacción con la empresa de los empleados de Industrias México sobre los ingresos que reciben "antes de impuestos" (en millones de pesos), los años de escolaridad concluidos y la edad de los empleados. Los niveles de satisfacción que se manejaron son tres: Muy Satisfecho, Moderadamente Satisfecho y Muy Insatisfecho

El efecto se evalúa mediante tres criterios:

1.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre el Ingreso de los empleados de Industrias México.

2.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre la Escolaridad de los empleados de Industrias México.

3.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre la Edad de los empleados de Industrias México.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Ingresos anuales "antes de impuestos" (en millones de pesos), X_2= Años de escolaridad concluidos y X_3= Edad de los empleados (En años

cumplidos), de 5 empleados muy satisfechos, 6 empleados moderadamente satisfechos y 7 empleados muy insatisfechos. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Nivel de Satisfacción								
Muy Satisfecho			Moderadamente Satisfecho			Muy Insatisfecho		
Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad
0.854	15	35	0.550	14	49	0.320	18	36
0.974	19	38	0.656	17	48	0.540	19	47
0.894	16	26	0.472	16	29	0.332	15	43
1.284	16	39	0.646	13	30	0.570	12	36
0.626	12	34	0.358	14	27	0.272	13	30
			0.744	16	36	0.310	11	27
						0.462	16	42

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de los empleados es el mismo en los tres niveles de satisfacción de los empleados con la empresa?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el nivel de satisfacción de los empleados con la empresa influye sobre las magnitudes promedio de los empleados.
c) Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que nivel o niveles de satisfacción de los empleados con la empresa, los atributos son mayores y en que magnitud.

Resolución del inciso a)

Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

Prueba de hipótesis:

1. Juego de Hipótesis

$$H_0 = \begin{bmatrix}\mu_{11}\\ \mu_{21}\\ \mu_{31}\end{bmatrix} = \begin{bmatrix}\mu_{12}\\ \mu_{22}\\ \mu_{32}\end{bmatrix} = \begin{bmatrix}\mu_{13}\\ \mu_{23}\\ \mu_{33}\end{bmatrix}$$

2. Estadístico de prueba

Como No. de Variables Dependientes : p≥1 y

No. de Niveles del Factor: g=3

Usamos la siguiente fórmula (Caso 4):

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Donde\ £ = \frac{|error|}{|tratamiento|} = \frac{|SCE|}{|SCT|} = 0.27155$$

$$Fcalc = \frac{(\sum n_1 - p - 2)}{p}\frac{(1-\sqrt{£})}{\sqrt{£}}$$

$$Fcalc = \frac{(18-3-2)}{3}\frac{(1-\sqrt{0.27155})}{\sqrt{0..27155}}$$

$$Fcalc = \frac{(13)}{3}\frac{(1-0.52110}{0..52110}$$

$$Fcalc = (4.333)(0.91902)$$

$$Fcalc = 3.9821$$

3. Región de rechazo

$$Fcrit = \left[\propto, 2, p, 2\left(\sum n_1 - p - 2\right)\right]$$

$$Fcrit = \left[0.05,2(3),2\left(\sum 18-3-2\right)\right]$$

$Fcrit = [0.05,6,2(13)]$

$Fcrit = 0.05,6,26$

$F0.05 = 2.47$

$F0.01 = 3.59$

4. Regla de Decisión

Se rechaza H_0 sí $fcal \geq 2.47$

5. Conclusión

Estadista: Como 3.982 > 2.47 > 3.59 ∴ la prueba es (A.S) y se rechaza con mucha potencia.

Administrativa: Existe evidencia suficiente para decir que estadísticamente que al menos una de las medias de los 3 grupos NO es igual.

Resolución del inciso b)

Prueba de Hipótesis para Ingresos:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el Nivel de Satisfacción de los empleados es el mismo para los **Ingresos** en los tres niveles.

$$H_0: \mu_{11.} = \mu_{12.} = \mu_{13.}$$

La hipótesis alternativa es No todos los Niveles de Satisfacción utilizados en los empleados son iguales para los **Ingresos**.

$$H_1: \text{No todos los Niveles de Satisfacción son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Niveles de Satisfacción" utilizados en los empleados para los **Ingresos** es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio asignado a los **Ingresos** con base los Niveles de Satisfacción y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Niveles de Satisfacción tiene una diferencia significativa para los **Ingresos** de los empleados sobre los otros Niveles de Satisfacción utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SC$$

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 1.2280$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 0.8151$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 0.4129$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	***$F_{calculada}$***
Tratamien-tos	$v_1 = k - 1 = 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 0.8151$	$CMt = {}^{SCt}/_{g.l.} = {}^{0.8151}/_{2} = 0.40755$	$F_{Calc.} = {}^{CMt}/_{CME} = / = {}^{0.40755}/_{0.02753} = \mathbf{14.81}$
Error	$v_2 = N - k = 18 - 3 = \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 0.4129$	$CME = {}^{SCE}/_{g.l.} = {}^{0.4129}/_{15} = 0.02753$	
Total	18-1=17	$SCT = 1.2280$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **14.81**, que es mayor al valor crítico de 3.68; por lo tanto la hipótesis nula se rechaza y llegamos a la conclusión de que no todas las medias de la población son iguales, es decir al menos una de ellas es diferente.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Nivel de Satisfacción utilizado no es el mismo en al menos uno de los tres Niveles de Satisfacción utilizados.

Prueba de Hipótesis para Escolaridad

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el Nivel de Satisfacción de los empleados es el mismo para la **Escolaridad** en los tres niveles.

$$H_0:\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los Niveles de Satisfacción utilizados en los empleados son iguales para la **Escolaridad**.

$$H_1: \text{No todos los Niveles de Satisfacción son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Niveles de Satisfacción" utilizados en los empleados para la **Escolaridad** es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio asignado a la **Escolaridad** con base los Niveles de Satisfacción y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Niveles de Satisfacción tiene una diferencia significativa para la **Escolaridad** de los empleados sobre los otros Niveles de Satisfacción utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 93.778$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 1.721$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 92.05$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	$F_{calculada}$
Tratamien-tos	$v_1 = k - 1 = 3 - 1 = \mathbf{2\ g.l.}$	$SCt = 1.721$	$CMt = {}^{SCt}/_{g.l.} = {}^{1.721}/_{2} = 0.8603$	$F_{Calc.} = {}^{CMt}/_{CME} = / = {}^{0.8603}/_{6.1371} = \mathbf{0.14}$
Error	$v_2 = N - k = 18 - 3 = \mathbf{15\ g.l.}$	$SCE = 92.057$	$CME = {}^{SCE}/_{g.l.} = {}^{92.057}/_{15} = 6.1371$	
Total	18-1=17	$SCT = 93.778$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de **0.14**, que es menor al valor crítico de 3.68; por lo tanto la hipótesis nula NO se rechaza y llegamos a la conclusión de que todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Nivel de Satisfacción utilizado en los empleados es el mismo en los tres Niveles de Satisfacción utilizados.

Prueba de Hipótesis para Edad:

Se usa el proceso de prueba de hipótesis de cinco pasos.

Paso 1.- Establecer la hipótesis nula (H_0) y la hipótesis alternativa (H_1).

La hipótesis nula es que el grado de influencia que llega a tener el Nivel de Satisfacción de los empleados es el mismo para la **Edad** en los tres niveles.

$$H_0\colon\ \mu_{11.} = \mu_{12.} = \ \mu_{13.}$$

La hipótesis alternativa es No todos los Niveles de Satisfacción utilizados en los empleados son iguales para la **Edad**.

$$H_1\colon \text{No todos los Niveles de Satisfacción son iguales.}$$

También se puede pensar que la hipótesis alternativa sea "al menos uno de los Niveles de Satisfacción" utilizados en los empleados para la **Edad** es diferente.".

En este caso si la hipótesis nula no se rechaza, podemos llegar a la conclusión de que no existen diferencias significativas en el promedio asignado a la **Edad** con base los Niveles de Satisfacción y el grado de influencia es el mismo.

Por otro lado si la hipótesis nula se rechaza, podemos concluir que al menos uno de los Niveles de Satisfacción tiene una diferencia significativa para la **Edad** de los empleados sobre los otros Niveles de Satisfacción utilizados.

Paso 2.- Seleccionar y calcular el valor del estadístico de prueba apropiado.

El estadístico de prueba sigue una distribución ***F***

$$F_{Calculada} = \frac{CMt}{CME} = \frac{SCt/_{g.l.}}{SCE/_{g.l.}}$$

Donde:

$$SCT = SCt + SCE$$

Si utilizamos la fórmula abreviada, primero debe sumar el cuadrado de totas las observaciones y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCT = \sum_{I=1}^{3}\sum_{J=1}^{8} X_{ij}^2 - \frac{X_{..}^2}{N} = 919.11$$

Si usamos la fórmula abreviada primero debe sumar el cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento y al final restar el promedio del cuadrado de la suma total de la siguiente manera:

$$SCt = \sum_{i=1}^{3} \frac{X_{i.}^2}{n_i} - \frac{X_{..}^2}{N} = 24.98$$

Si usamos la fórmula abreviada, primero debe sumar el cuadrado de todas las observaciones y al final reste la suma del cuadrado de cada total de tratamiento entre el tamaño de la muestra correspondiente a dicho tratamiento de la siguiente manera:

$$SCE = \sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij}^2 - \sum_{i=1}^{k} \frac{X_{i.}^2}{n} = 894.13$$

Para encontrar el valor calculado de ***F***, trabaje con la tabla de **ANOVA**. El término de cuadrado de la media es otra expresión que se utiliza para un cálculo de la varianza. El cuadrado de la media para los tratamientos es ***SCt*** dividida entre sus grados de libertad. El resultado es el cuadrado de la media para los tratamientos y se escribe ***CMt.***

Fuente de Variación	***Grados de Libertad***	***Suma de Cuadrados***	***Cuadrado Medio***	$F_{calculada}$
Tratamientos	$v_1 = k - 1$ $= 3 - 1 = \mathbf{2}\ \boldsymbol{g.l.}$	$SCt = 24.98$	$CMt = {}^{SCt}/_{g.l.}$ $= {}^{24.98}/_{2} = 12.49$	$F_{Calc.} = {}^{CMt}/_{CME} = /$ $= {}^{12.49}/_{59.61}$ $= \mathbf{0.21}$
Error	$v_2 = N - k$ $= 18 - 3$ $= \mathbf{15}\ \boldsymbol{g.l.}$	$SCE = 894.13$	$CME = {}^{SCE}/_{g.l.}$ $= {}^{894.13}/_{15}$ $= 59.61$	
Total	18-1=17	$SCT = 919.11$		

Paso 3.- Establecer la región de rechazo de (H_0).

Para determinar la región de rechazo, se necesita el valor crítico. El valor crítico en el estadístico ***F*** se encuentra en las tablas ***F*** del apéndice (). Para utilizar esta tabla se necesita conocer los grados de libertad en el numerador y en el denominador. Los grados de libertad en el numerador son iguales al número de tratamientos, designados como k, menos 1. Los grados de libertad en el denominador son el número total de observaciones, N, menos el número de tratamientos. Para este problema existen 3 tratamientos y un total de 18 observaciones, por lo tanto los grados de libertad en el numerador son: k-1=3-1= 2 g.l. y los grados de libertad del denominador son : N-k=18-3=15 g.l.

Como existen tablas para niveles de Alfa diferentes, busque la que corresponda al nivel de significancia solicitada, en este caso 0.05, y desplácese horizontalmente sobre la parte superior de la página hasta llegar a las 2 grados de libertad del numerador. Luego descienda en esa columna hasta llegar a la fila que presenta 15 grados de libertad. El valor en esta intersección es **3.68** que en este caso es el valor crítico.

Grados de libertad en el denominador	**Grados de libertad en el numerador**			
	1	**2**	3	4
8	5.32	4.46	4.07	3.84
9	5.12	4.26	3.86	3.63
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26

13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	**3.68**	3.29	3.06

Paso 4.- Formular una regla de decisión basada en los pasos 1,2 y 3 anteriores.

La regla de decisión es rechazar H_0 si el valor calculado de ***F*** es mayor a 3.68

Paso 5.- Tomar una decisión en cuanto a la hipótesis nula con base en la información de la muestra (conclusión estadística). Interpretar los resultados de la prueba (conclusión administrativa).

Estadística: Como el valor calculado de ***F*** es de 0.21, que es menor al valor crítico de 3.68; por lo tanto la hipótesis nula NO se rechaza y llegamos a la conclusión de que todas las medias de la población son iguales.

Administrativa: Existe evidencia suficiente para concluir que estadísticamente el grado de influencia que llega a tener el Nivel de Satisfacción utilizado para los empleados es el mismo en los tres Niveles de Satisfacción utilizados.

Resolución del inciso c)

Prueba de Tukey

Intervalo de confianza para Ingresos (Muy Satisfecho-Moderadamente Satisfecho) .Modelo desbalanceado.

$$M_{11} - M_{12} = (\bar{X}_{11} - \bar{X}_{12}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = (0.926 - 0.571) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{12} = 0.355 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{0.4129}{15} = \underline{0.02753}$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{11} - M_{12} = 0.355 \mp 3.67\sqrt{\frac{0.02753}{5.4545}}$$

$$M_{11} - M_{12} = 0.355 \mp 3.67\sqrt{0.00505}$$

$$M_{11} - M_{12} = 0.355 \mp 3.67(0.07104)$$

$$M_{11} - M_{12} = 0.355 \mp 0.26073 \begin{cases} LIC = 0.09427 \\ LSC = 0.61573 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de influencia del **Nivel Muy Satisfecho sobre Ingresos** es mayor que el del **Nivel Moderadamente Satisfecho sobre Ingresos** por un mínimo de **0.09424** y un máximo de **0.61573** millones de pesos.

Intervalo de confianza para Ingresos (Muy Satisfecho-Muy Insatisfecho). Modelo desbalanceado

$$M_{11} - M_{13} = (\bar{X}_{11} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = (0.926 - 0.4009) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{11} - M_{13} = 0.5251 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = 5.8333$$

$$M_{11} - M_{13} = 0.5251 \mp 3.67\sqrt{\frac{0.02753}{5.8333}}$$

$$M_{11} - M_{13} = 0.5251 \mp 3.67(0.0687)$$

$$M_{11} - M_{13} = 0.5251 \mp 0.25212 \begin{cases} LIC = 0.27298 \\ LSC = 0.77722 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son positivos podemos decir que estadísticamente el grado de influencia del **Nivel Muy Satisfecho sobre Ingresos** es mayor que el del **Nivel Muy Insatisfecho sobre Ingresos** por un mínimo de **0.27298** y un máximo de **0.77722** millones de pesos.

Intervalo de confianza para Ingresos (Moderadamente Satisfecho-Muy Insatisfecho). Modelo desbalanceado

$$M_{12} - M_{13} = (\bar{X}_{12} - \bar{X}_{13}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = (0.571 - 0.4009) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{12} - M_{13} = 0.1701 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = 6.4615$$

$$M_{12} - M_{13} = 0.1701 \mp 3.67\sqrt{\frac{0.02753}{6.4615}}$$

$$M_{12} - M_{13} = 0.1701 \mp 3.67(0.06527)$$

$$M_{12} - M_{13} = 0.1701 \mp 0.23955 \begin{cases} LIC = -0.06945 \\ LSC = 0.40965 \end{cases}$$

I**nterpretación:**

Como el intervalo de confianza pasa por cero podemos decir que el grado de influencia del **Nivel Moderadamente Satisfecho sobre Ingresos** es el mismo que el del **Nivel Muy Insatisfecho sobre Ingresos.**

Intervalo de confianza para Escolaridad (Muy Satisfecho-Moderadamente Satisfecho). Modelo desbalanceado

$$M_{21} - M_{22} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = (15.6 - 15) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{22} = 0.6 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{92.057}{15} = \underline{6.1371}$$

$$nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{21} - M_{22} = 0.6 \mp 3.67\sqrt{\frac{6.1371}{5.4545}}$$

$$M_{21} - M_{22} = 0.6 \mp 3.67(1.06073)$$

$$M_{21} - M_{22} = 0.6 \mp 3.89287 \begin{cases} LIC = -3.29287 \\ LSC = 4.49287 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de influencia del **Nivel Muy Satisfecho sobre Escolaridad** es el mismo que el del **Nivel Moderadamente Satisfecho sobre Escolaridad.**

Intervalo de confianza para Escolaridad (Muy Satisfecho-Muy Insatisfecho). Modelo desbalanceado.

$$M_{21} - M_{23} = (\bar{X}_{21} - \bar{X}_{22}) \mp q \propto, (k.N-k)\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = (15.6 - 14.86) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{21} - M_{23} = 0.74 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1}+\frac{1}{n_2}} = \frac{2}{\frac{1}{5}+\frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{21} - M_{23} = 0.74 \mp 3.67\sqrt{\frac{6.1371}{5.8326}}$$

$$M_{21} - M_{23} = 0.74 \mp 3.67(1.02577)$$

$$M_{21} - M_{23} = 0.74 \mp 3.76458 \begin{cases} LIC = -3.02458 \\ LSC = -4.50458 \end{cases}$$

Interpretación:

Como ambos límites del intervalo son negativos podemos decir que estadísticamente el grado de influencia del **Nivel Muy Satisfecho sobre Escolaridad** es mayor que el del **Nivel Muy Insatisfecho sobre Escolaridad** por un mínimo de **3.02458** y un máximo de **4.50458** años. (En valor absoluto).

Intervalo de confianza para Escolaridad (Moderadamente Satisfecho-Muy Insatisfecho). Modelo desbalanceado.

$$M_{22} - M_{23} = (\bar{X}_{22} - \bar{X}_{23}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = (15 - 14.86) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{22} - M_{23} = 0.14 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{22} - M_{23} = 0.14 \mp 3.67\sqrt{\frac{6.1371}{6.4620}}$$

$$M_{22} - M_{23} = 0.14 \mp 3.67(0.97454)$$

$$M_{22} - M_{23} = 0.14 \mp 3.57655 \begin{cases} LIC = -3.43655 \\ LSC = 3.71655 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de influencia del **Nivel Moderadamente Satisfecho sobre Escolaridad** es el mismo que el del **Nivel Muy Insatisfecho sobre Escolaridad.**

Intervalo de confianza para Edad (Muy Satisfecho-Moderadamente Satisfecho). Modelo desbalanceado

$$M_{31} - M_{32} = (\bar{X}_{31} - \bar{X}_{32}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = (34.4 - 36.5) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{32} = -2.1 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ CME = \frac{SCE}{N-k} = \frac{SCE}{18-3} = \frac{894.13}{15} = 59.61$$

$$nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{6}} = \frac{2}{0.3667} = 5.4545$$

$$M_{31} - M_{32} = -2.1 \mp 3.67 \sqrt{\frac{59.61}{5.4545}}$$

$$M_{31} - M_{32} = -2.1 \mp 3.67(3.30584)$$

$$M_{31} - M_{32} = -2.1 \mp 12.13244 \begin{cases} LIC = -14.23244 \\ LSC = 10.03244 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de satisfacción del **Nivel Muy Satisfecho sobre Edad** es el mismo que el del **Nivel Moderadamente Satisfecho sobre Edad.** (en valor absoluto**).**

Intervalo de confianza para Edad (Muy Satisfecho-Muy Insatisfecho). Modelo desbalanceado.

$$M_{31} - M_{33} = (\bar{X}_{31} - \bar{X}_{33}) \mp q \propto, (k.N - k) \sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = (34.4 - 37.29) \mp 0.05{,}3{,}15 \sqrt{\frac{CME}{nh}}$$

$$M_{31} - M_{33} = -2.89 \mp 3.67 \sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{5} + \frac{1}{7}} = \frac{2}{0.3429} = 5.8326$$

$$M_{31} - M_{33} = -2.89 \mp 3.67 \sqrt{\frac{59.61}{5.8326}}$$

$$M_{31} - M_{33} = -2.89 \mp 3.67(3.1969)$$

$$M_{31} - M_{33} = -2.89 \mp 11.73261 \begin{cases} LIC = -14.62261 \\ LSC = 8.84261 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del **Nivel de Satisfacción Muy Satisfecho sobre la Edad** es el mismo que el del **Nivel de Satisfacción Muy Insatisfecho sobre la Edad.**

Intervalo de confianza para Edad (Moderadamente Satisfecho-Muy Insatisfecho). Modelo desbalanceado.

$$M_{32} - M_{33} = (\bar{X}_{32} - \bar{X}_{33}) \mp q \propto, (k.N - k)\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = (36.5 - 37.29) \mp 0.05{,}3{,}15\sqrt{\frac{CME}{nh}}$$

$$M_{32} - M_{33} = -0.79 \mp 3.67\sqrt{\frac{CME}{nh}}$$

$$Donde\ nh = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2}} = \frac{2}{\frac{1}{6} + \frac{1}{7}} = \frac{2}{0.3095} = 6.4620$$

$$M_{32} - M_{33} = -0.79 \mp 3.67\sqrt{\frac{59.61}{6.4620}}$$

$$M_{32} - M_{33} = -0.79 \mp 3.67(3.03722)$$

$$M_{32} - M_{33} = -0.79 \mp 11.14659 \begin{cases} LIC = -11.93659 \\ LSC = 10.35659 \end{cases}$$

Interpretación:

Como el intervalo de confianza pasa por cero podemos decir que el grado de influencia del **Nivel Moderadamente Satisfecho sobre la Edad** es el mismo que el del **Nivel Muy Insatisfecho sobre la Edad.**

3.2.3.1.3 EJERCICIO DE REFUERZO con Minitab 17

EJERCICIO DE REFUERZO 3.2.3.1.3. Análisis de Varianza Multivariada (MANOVA) con Minitab 17

3.2.3.1.3. El Director de Industrias México desea medir el grado de influencia que llega a tener el Nivel de Satisfacción con la empresa de los empleados de Industrias México sobre los ingresos que reciben "antes de impuestos" (en millones de pesos), los años de escolaridad concluidos y la edad de los empleados. Los niveles de satisfacción que se manejaron son tres: Muy Satisfecho, Moderadamente Satisfecho y Muy Insatisfecho

El efecto se evalúa mediante tres criterios:

1.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre el Ingreso de los empleados de Industrias México.

2.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre la Escolaridad de los empleados de Industrias México.

3.- Que tanto el Nivel de Satisfacción de los empleados con la empresa llega a influir sobre la Edad de los empleados de Industrias México.

Se tomó una muestra aleatoria de una población normal multivariada donde X_1= Ingresos anuales "antes de impuestos" (en millones de pesos), X_2= Años de escolaridad concluidos y X_3= Edad de los empleados (En años cumplidos), de 5 empleados muy satisfechos, 6 empleados moderadamente satisfechos y 7 empleados muy insatisfechos. Se les aplicó el respectivo cuestionario.

Los resultados fueron los siguientes:

Nivel de Satisfacción								
Muy Satisfecho			Moderadamente Satisfecho			Muy Insatisfecho		
Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad	Ingresos	Escolaridad	Edad
0.854	15	35	0.550	14	49	0.320	18	36
0.974	19	38	0.656	17	48	0.540	19	47
0.894	16	26	0.472	16	29	0.332	15	43

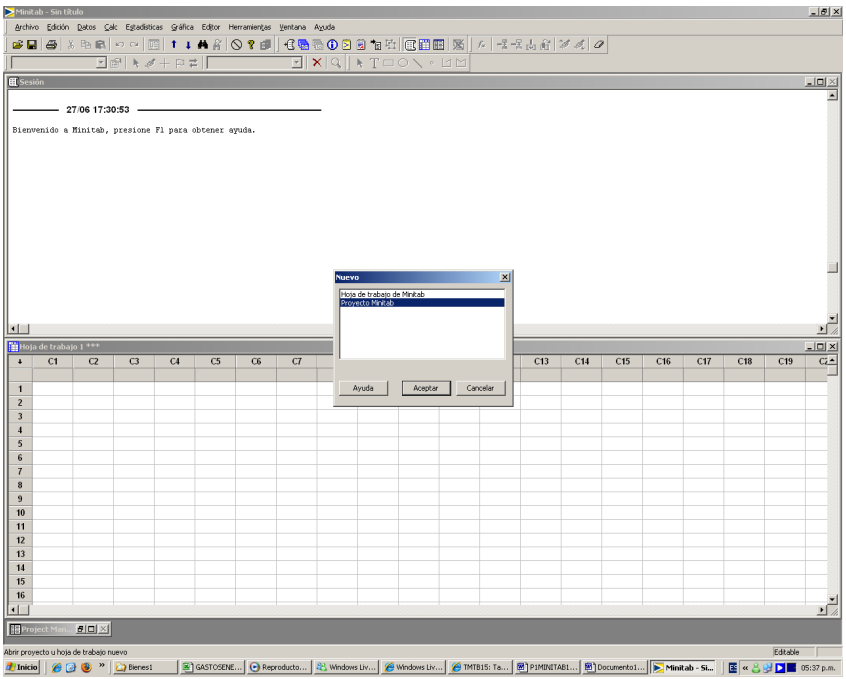
1.284	16	39	0.646	13	30	0.570	12	36
0.626	12	34	0.358	14	27	0.272	13	30
			0.744	16	36	0.310	11	27
						0.462	16	42

a) ¿Existe evidencia suficiente que indique que las magnitudes promedio de los atributos de los empleados es el mismo en los tres niveles de satisfacción de los empleados con la empresa?. Pruébelo.
b) Realice las pruebas ANOVA para cada variable dependiente por separado para ver si el nivel de satisfacción de los empleados con la empresa influye sobre las magnitudes promedio de los empleados. Realice las pruebas Post-hoc de Tukey para cada variable dependiente por separado y construya los intervalos de confianza e indique, si es el caso, en que nivel o niveles de satisfacción de los empleados con la empresa, los atributos son mayores y en que magnitud.

Resolución del inciso a)

Utilice **MANOVA general** para realizar un análisis multivariado de la varianza (MANOVA) con diseños balanceados y no balanceados. Este procedimiento saca provecho de la estructura de covarianza de los datos para probar simultáneamente la igualdad de las medias de diferentes respuestas.

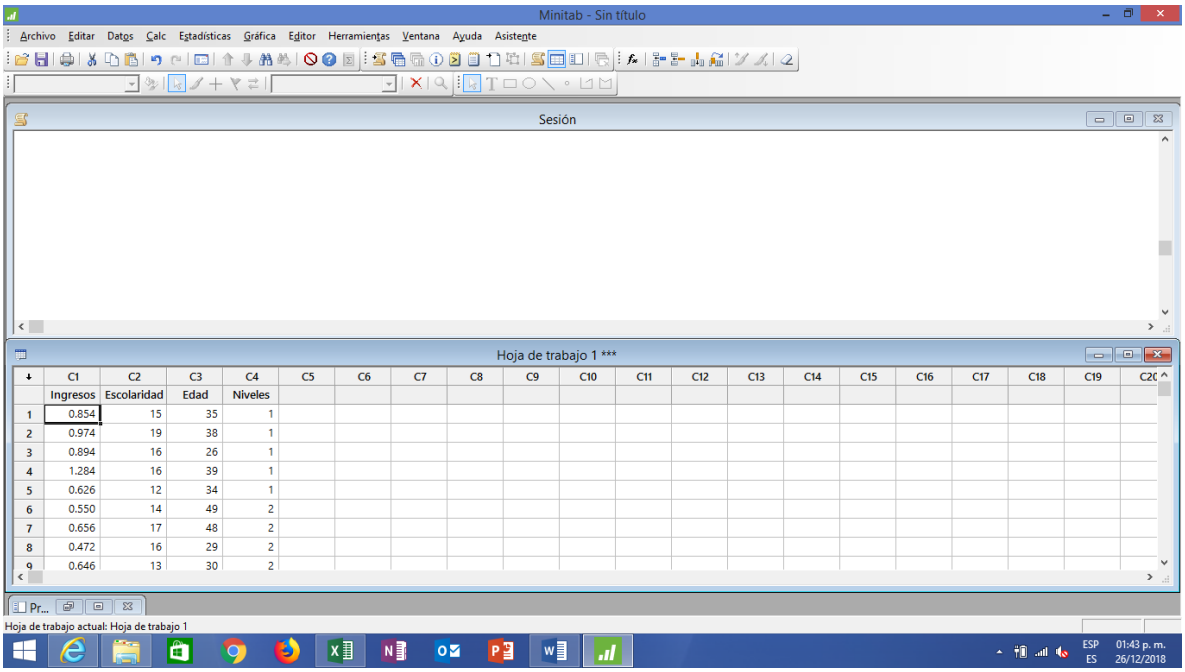
Para crear la base de datos del ejemplo ilustrativo en **MINITAB17**, accede a **MINITAB VERSIÓN 17.** Cuando aparezca la pantalla da un clic en la opción **Archivo** de la barra menú y selecciona **Nuevo**. Aparecerá la ventana **Nuevo. Selecciona la opción Proyecto Minitab y Aceptar**



Procede a editar la base de datos de la siguiente manera:

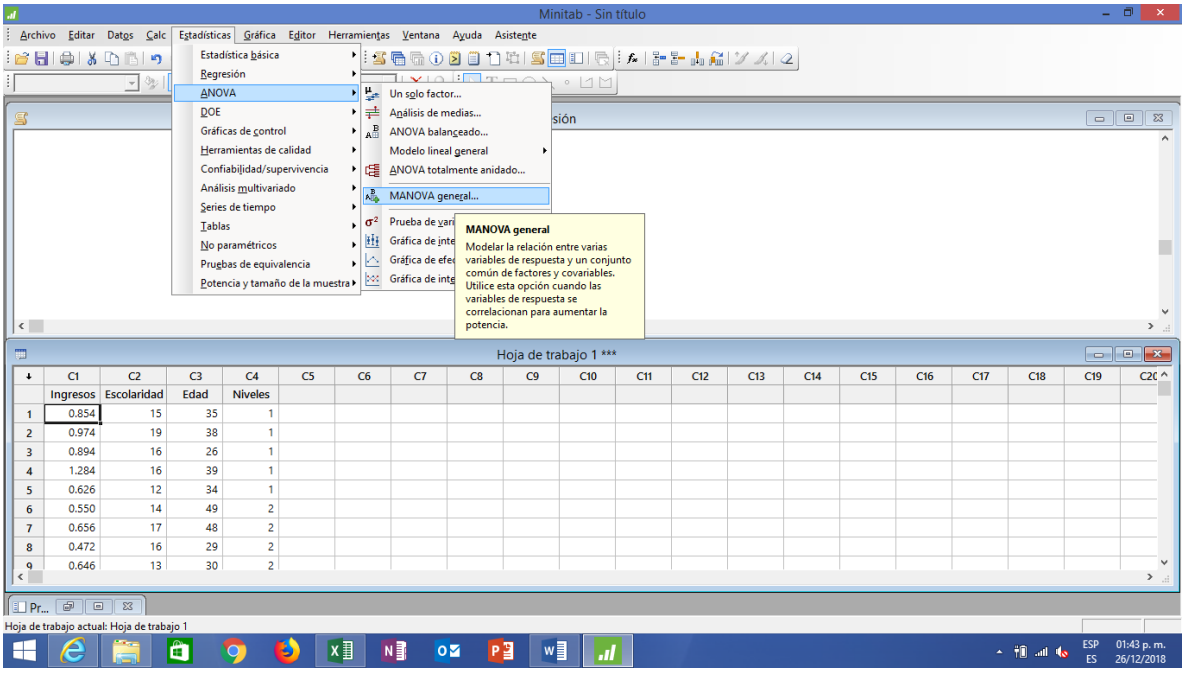
Coloca los nombres de cada **variable (En este caso Ingresos, Escolaridad, Edad y Niveles**), debajo de cada número de columna, en este caso **C1, C2, C3 y C4**.

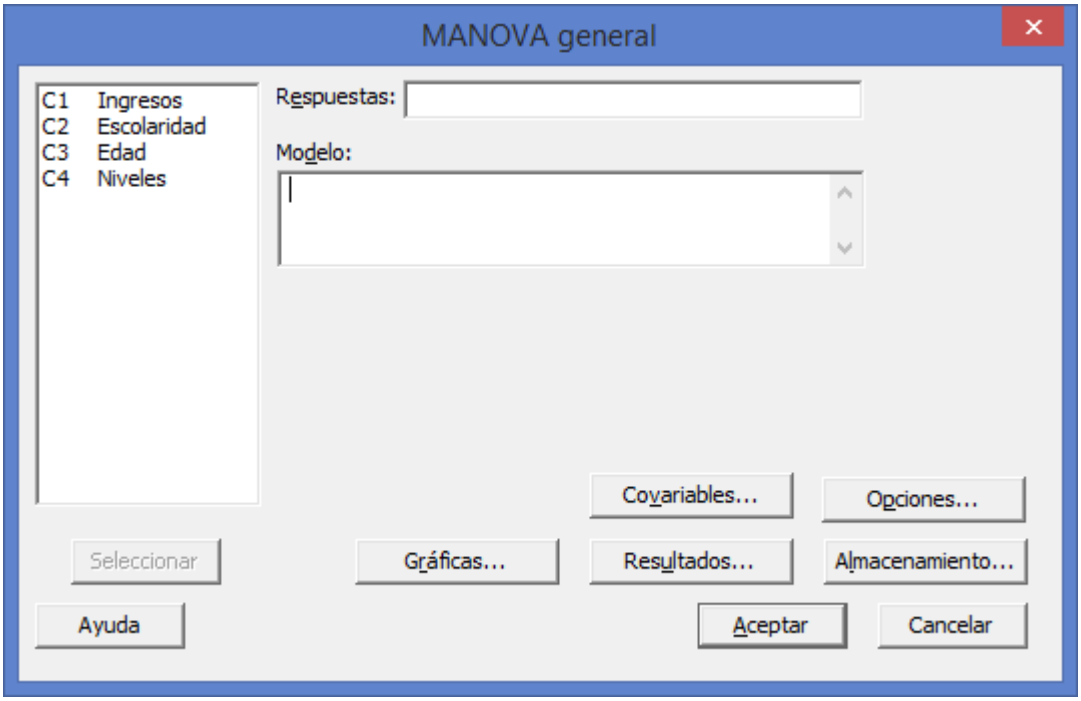
Una vez termines de colocar los nombres de cada **variable (En este caso Ingresos, Escolaridad, Edad y Niveles**, coloca el cursor en el primer nombre de la variable **(Ingresos)** y llena los datos de Ingresos en forma vertical de los tres Niveles de Satisfacción utilizados (Muy Satisfecho, Moderadamente Satisfecho y Muy Insatisfecho). Cuando se terminen los 18 datos coloca el cursor en el segundo nombre de la variable (**Escolaridad)** y llena los 18 datos. Cuando se terminen los 18 datos de Escolaridad coloca el cursor en el tercer nombre de la variable (**Edad**) y llena los 18 datos. Cuando se terminen los 18 datos de Edad coloca el cursor en la cuarta columna (**Niveles**) y coloca en forma vertical cinco 1, seis 2 y siete 3. El número 1 corresponde al Nivel de Satisfacción **Muy Satisfecho,** el 2 corresponde al Nivel de Satisfacción **Moderadamente Satisfecho** y el 3 corresponde al Nivel de Satisfacción **Muy Insatisfecho.**



	C1	C2	C3	C4
	Ingresos	Escolaridad	Edad	Niveles
1	0.854	15	35	1
2	0.974	19	38	1
3	0.894	16	26	1
4	1.284	16	39	1
5	0.626	12	34	1
6	0.550	14	49	2
7	0.656	17	48	2
8	0.472	16	29	2
9	0.646	13	30	2

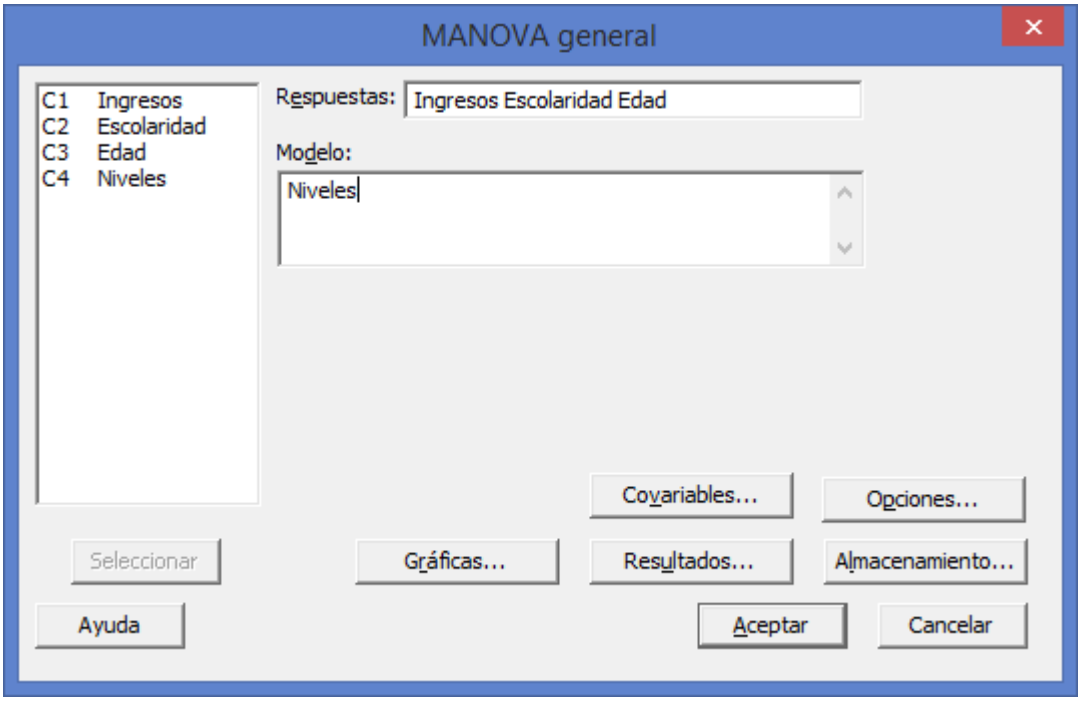
5 Elija **Estadísticas > ANOVA > MANOVA general**.





2 En **Respuestas**, ingrese las 3 variables o columnas numéricas que contengan las variables de respuesta (**Ingresos, Escolaridad y Edad**).

6 En **Modelo**, ingresa la cuarta columna (**Niveles**)



7 Haga clic en **Aceptar**.

Salida de la ventana Sesión

Modelo lineal general: Ingresos, Escolaridad, Edad vs. Niveles

```
MANOVA para Niveles
s = 2    m = 0.0    n = 5.5

                 Estadística              GL
Criterio           de prueba      F  Núm  Denom      P
De Wilks             0.27155  3.982    6     26  0.006
Lawley-Hotelling     2.68020  5.360    6     24  0.001
De Pillai            0.72909  2.677    6     28  0.035
De Roy               2.67931
```

Interpretación de los resultados

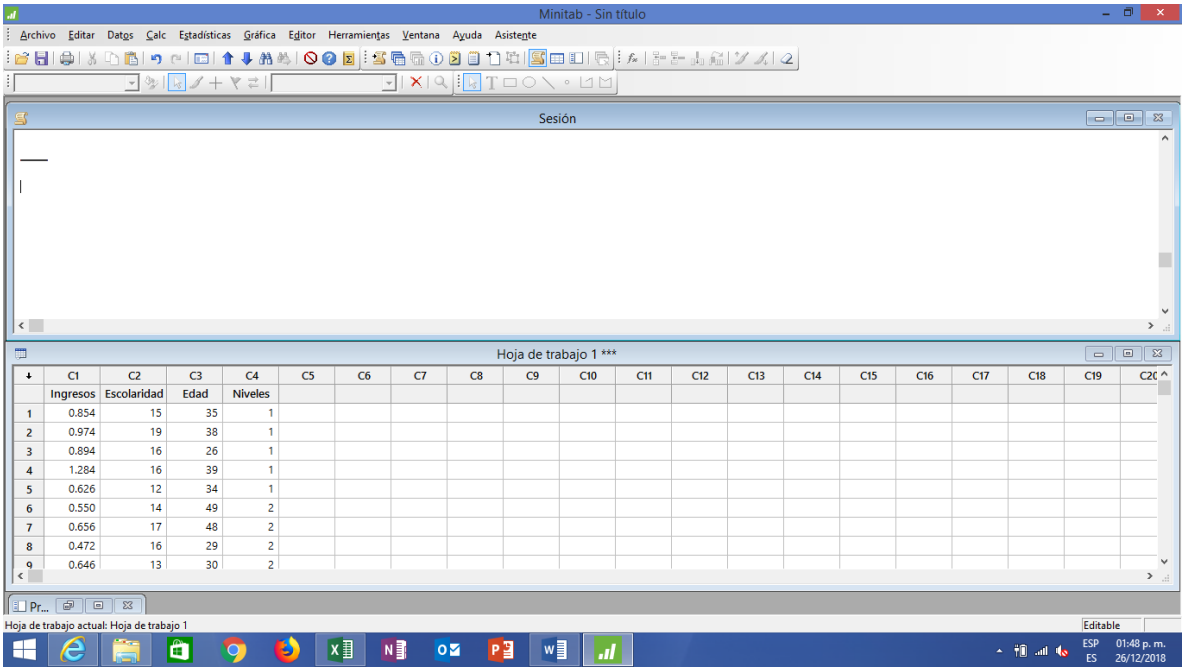
La salida de la ventana sesión presenta los resultados de cuatro criterios o pruebas entre ellas el de **Lambda de Wilks** que arroja una Estadística de prueba de **0.27155** con una Fcalculada de **3.982**. El valor de p de **0.006** (nivel de significancia observado) indica que hay suficiente evidencia de que no todas las medias en cada variable son iguales en los tres estilos utilizados cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Resolución del inciso b)

1. Prueba ANOVA de Hipótesis para Ingresos y pruebas Post-hoc de Tukey para Ingresos:

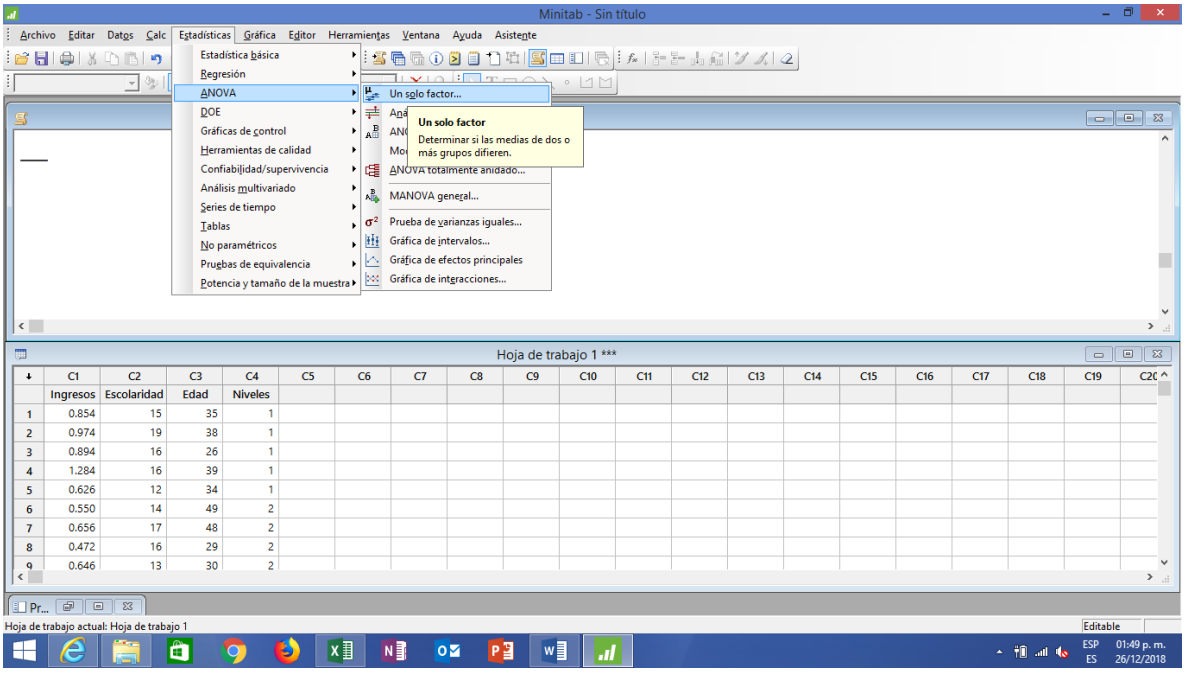
Cuando el número de observaciones en cada tratamiento es extenso y/o existen muchos tratamientos, los cálculos manuales son tediosos. Existen muchos paquetes de software que pueden mostrar los resultados entre ellos Minitab (Versión 17).

Comenzamos introduciendo los datos en la hoja de Trabajo 1 de Minitab, tal y como se muestra a continuación:

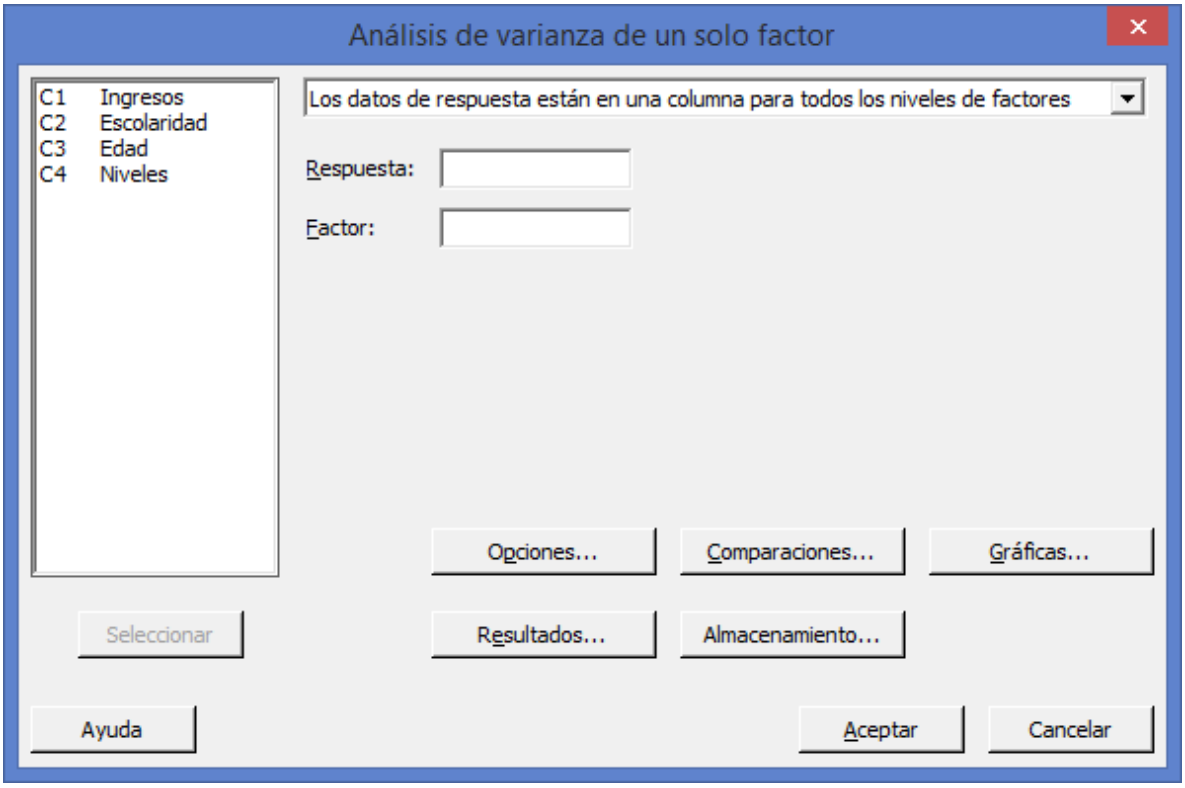


	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20
	Ingresos	Escolaridad	Edad	Niveles																
1	0.854	15	35	1																
2	0.974	19	38	1																
3	0.894	16	26	1																
4	1.284	16	39	1																
5	0.626	12	34	1																
6	0.550	14	49	2																
7	0.656	17	48	2																
8	0.472	16	29	2																
9	0.646	13	30	2																

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**



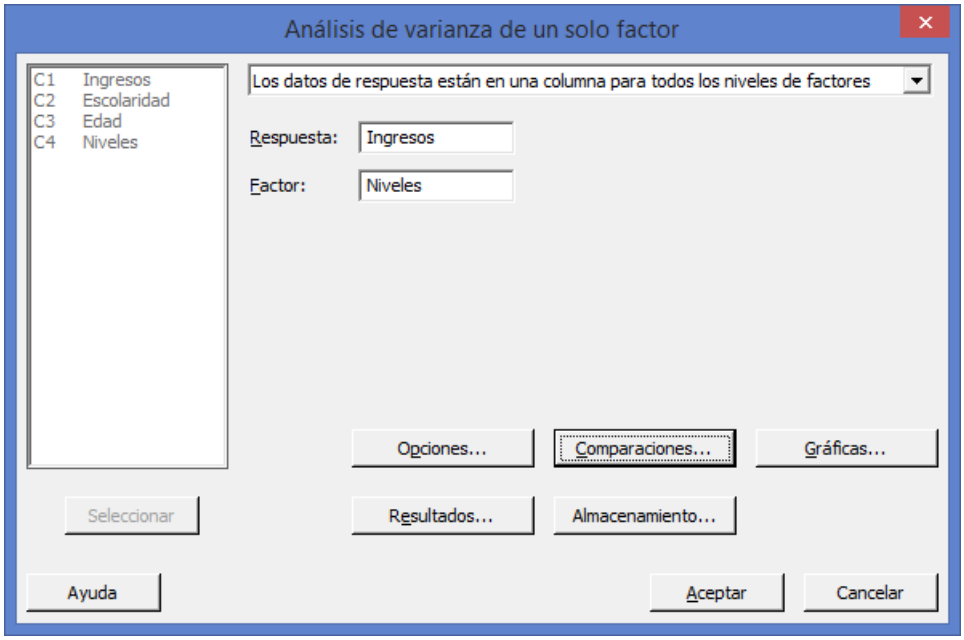
	C1	C2	C3	C4
	Ingresos	Escolaridad	Edad	Niveles
1	0.854	15	35	1
2	0.974	19	38	1
3	0.894	16	26	1
4	1.284	16	39	1
5	0.626	12	34	1
6	0.550	14	49	2
7	0.656	17	48	2
8	0.472	16	29	2
9	0.646	13	30	2



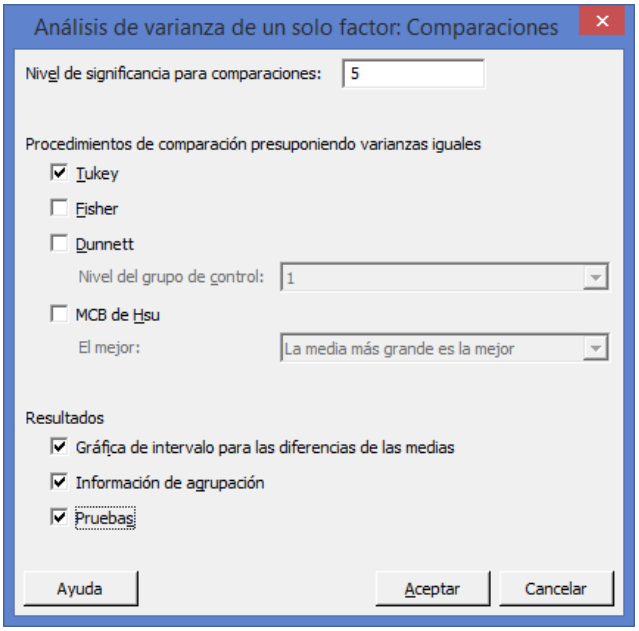
El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la

da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Ingresos**. En Factor, ingrese **Niveles**.



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Ingresos vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2     0.8151    0.40755    14.81    0.000
Error    15     0.4129    0.02753
Total    17     1.2280


Resumen del modelo

                      R-cuad.  R-cuad.
       S  R-cuad.  (ajustado)   (pred)
0.165914   66.38%      61.89%   49.88%


Medias

Niveles  N   Media  Desv.Est.      IC de 95%
1        5   0.926      0.238  ( 0.768,  1.085)
2        6  0.5710     0.1403  (0.4266, 0.7154)
3        7  0.4009     0.1210  (0.2672, 0.5345)

Desv.Est. agrupada = 0.165914
```

Interpretación de los resultados

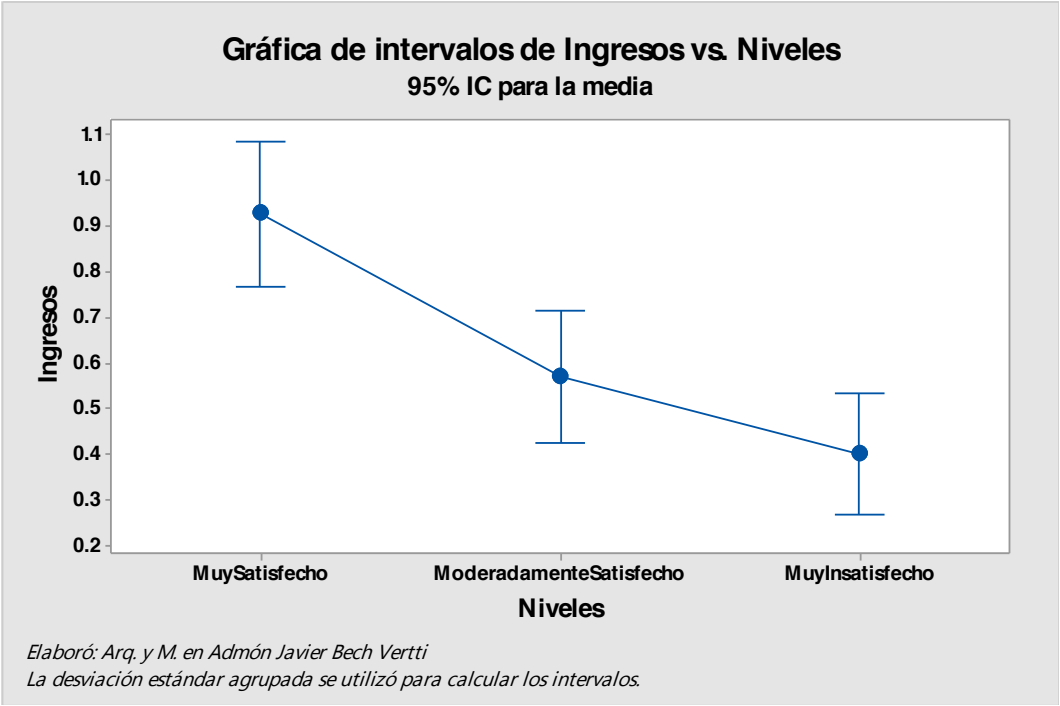
En la Tabla de ANOVA, para **Ingresos** en **los tres Niveles de Satisfacción** utilizados el valor de F calculada de **14.81** con un valor p (**0.000**) indica que hay suficiente evidencia de que no todas las medias son iguales cuando alfa se establece en 0.05. Para explorar las diferencias entre medias, examine los resultados de las comparaciones múltiples.

Comparaciones en parejas de Tukey

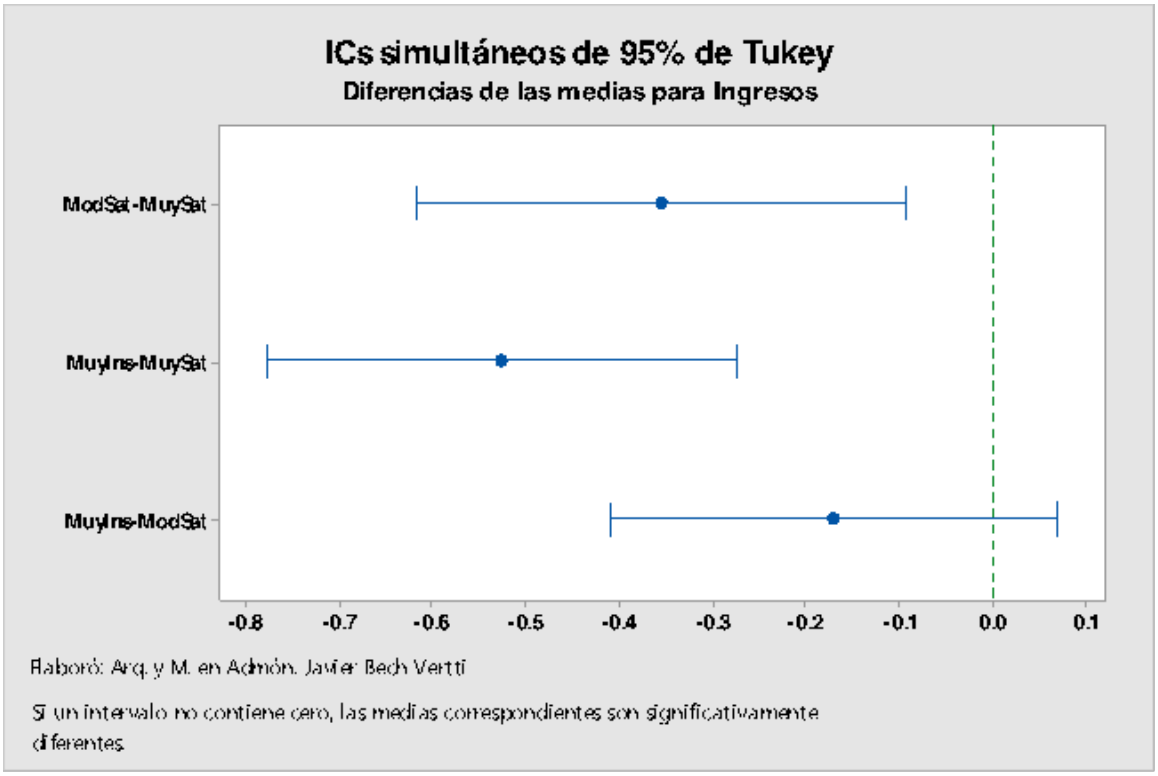
```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N   Media  Agrupación
1        5   0.926  A
2        6  0.5710    B
3        7  0.4009    B

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                                  Valor p
de niveles      medias  diferencia       IC de 95%      Valor T  ajustado
2 - 1           -0.355       0.100  ( -0.616,  -0.095)    -3.54     0.008
3 - 1          -0.5255      0.0971  (-0.7777, -0.2734)    -5.41     0.000
3 - 2          -0.1701      0.0923  (-0.4097,  0.0694)    -1.84     0.190

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Ingresos vs. Niveles



Elaboró: Arq. y M. en Admón Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.



Interpretación de los resultados

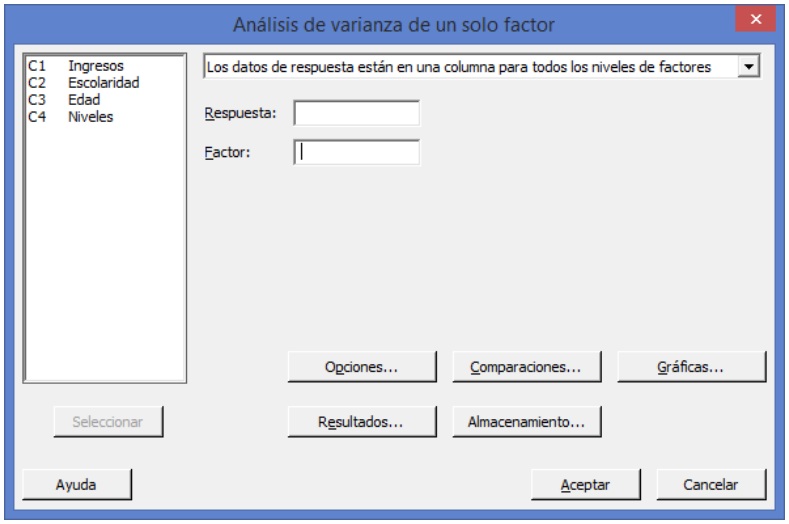
De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Ingresos** en **Moderadamente Satisfecho** (Nivel 2) con **Ingresos** en **Muy Satisfecho** (Nivel 1) oscila entre **-0.616 y -0.095**. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de influencia del nivel de Satisfacción **Muy Satisfecho** sobre **Ingresos** es mayor que el nivel de Satisfacción **Moderadamente Satisfecho** sobre **Ingresos** por un mínimo de **0.0.95** y un máximo de **0.616** millones de pesos. (Valor absoluto)

En cuanto al intervalo de confianza de **Ingresos** en **Muy Insatisfecho** (Nivel 3) con **Ingreso** en **Muy Satisfecho** (Nivel 1) oscila entre **-0.7777 y -0.27341**. Como ambos límites de confianza son negativos podemos decir que estadísticamente el grado de influencia del nivel de Satisfacción **Muy Satisfecho** sobre **Ingresos** es mayor que el nivel de Satisfacción **Muy Insatisfecho** sobre **Ingresos** por un mínimo de **0.27341** y un máximo de **0.7777** millones de pesos. (Valor absoluto)

Por último podemos observar que el intervalo de confianza de **Ingresos** en **Muy Insatisfecho** (Nivel 3) con **Ingresos** en **Moderadamente Satisfecho** (Nivel 2) oscila entre -**0.4097 y 0.06941**. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del Nivel de Satisfacción **Muy Insatisfecho** sobre **Ingresos** es igual que el Nivel de Satisfacción **Moderadamente Satisfecho** sobre **Ingresos.**

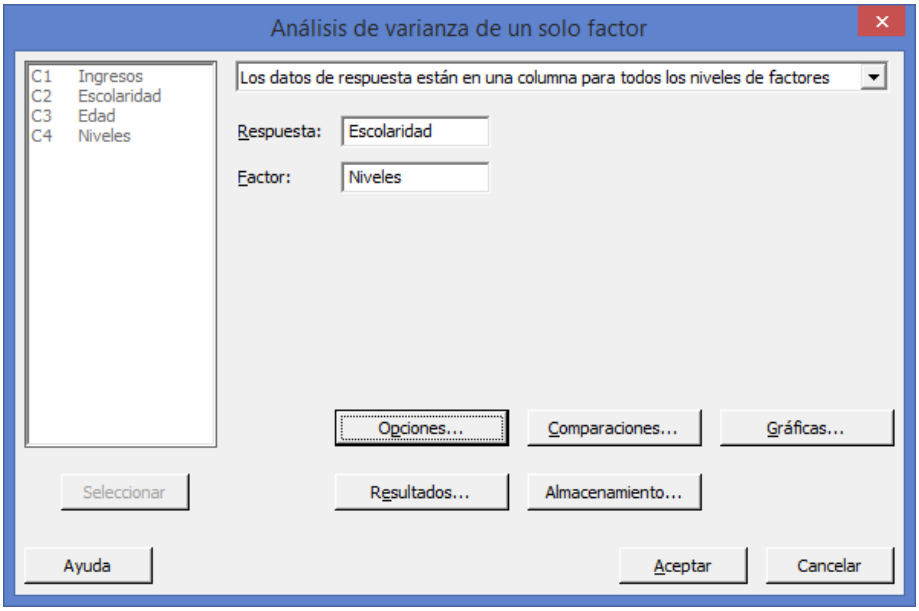
2. Prueba ANOVA de Hipótesis para Escolaridad y pruebas Post-hoc de Tukey para Escolaridad:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

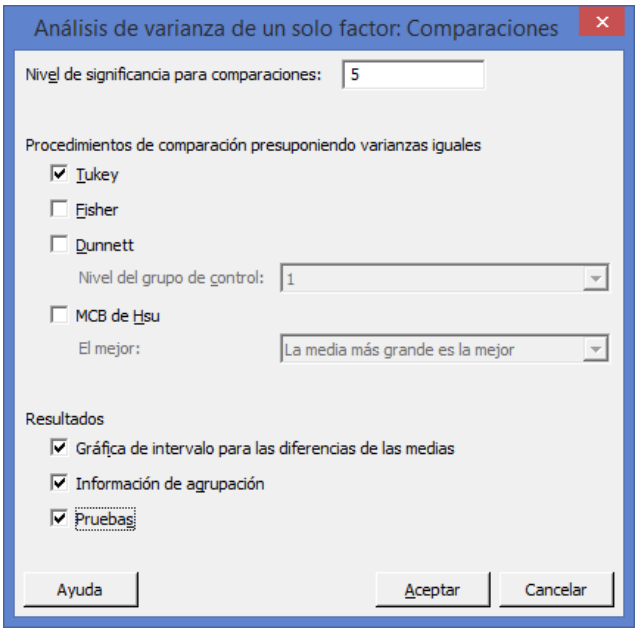


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Escolaridad**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en **Aceptar** en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Escolaridad vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2      1.721     0.8603     0.14    0.870
Error    15     92.057     6.1371
Total    17     93.778
```

```
Resumen del modelo

                      R-cuad.  R-cuad.
      S  R-cuad.   (ajustado)   (pred)
2.47733    1.83%       0.00%    0.00%


Medias

Niveles  N   Media  Desv.Est.      IC de 95%
1        5   15.60       2.51  ( 13.24,  17.96)
2        6  15.000      1.549  (12.844, 17.156)
3        7   14.86       3.02  ( 12.86,  16.85)

Desv.Est. agrupada = 2.47733
```

Interpretación de los resultados

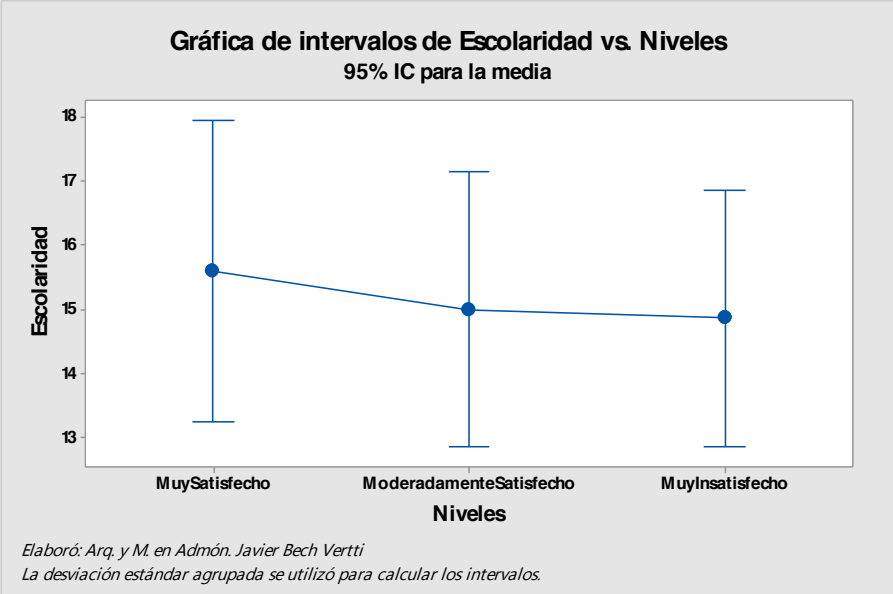
En la Tabla de ANOVA, para **Escolaridad** en **los tres Niveles de Satisfacción** utilizados el valor de F calculada de **0.14** con un valor p (**0.870**) indica que hay suficiente evidencia de que todas las medias son iguales cuando alfa se establece en 0.05. En este caso en particular no se deberían explorar las diferencias entre medias porque estadísticamente no las hay, pero por ser un ejercicio académico se llevará a cabo dicha exploración de las diferencias entre medias, examinando los resultados de las comparaciones múltiples ya que los tres intervalos de confianza deben pasar por cero.

Comparaciones en parejas de Tukey

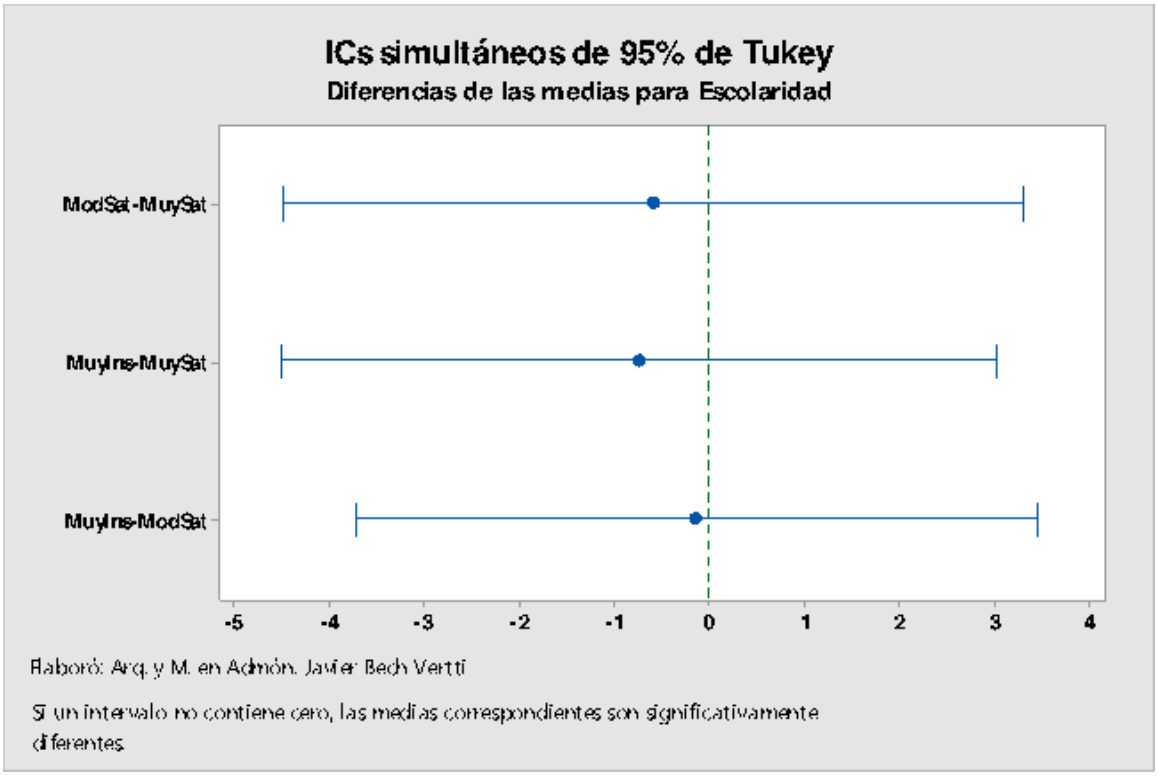
```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N   Media  Agrupación
1        5   15.60  A
2        6  15.000  A
3        7   14.86  A

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%    Valor T  ajustado
2 - 1            -0.60        1.50  (-4.49, 3.29)    -0.40     0.916
3 - 1            -0.74        1.45  (-4.51, 3.02)    -0.51     0.867
3 - 2            -0.14        1.38  (-3.72, 3.43)    -0.10     0.994

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Escolaridad vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.



Elaboró: Arq. y M. en Admón. Javier Bech Vertti

Si un intervalo no contiene cero, las medias correspondientes son significativamente diferentes

Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Escolaridad** en **Moderadamente Satisfecho** (Nivel 2) con **Escolaridad** en **Muy Satisfecho** (Nivel 1) oscila entre **-4.49 y 3.29**. Como el intervalo de confianza pasa por cero podemos decir que

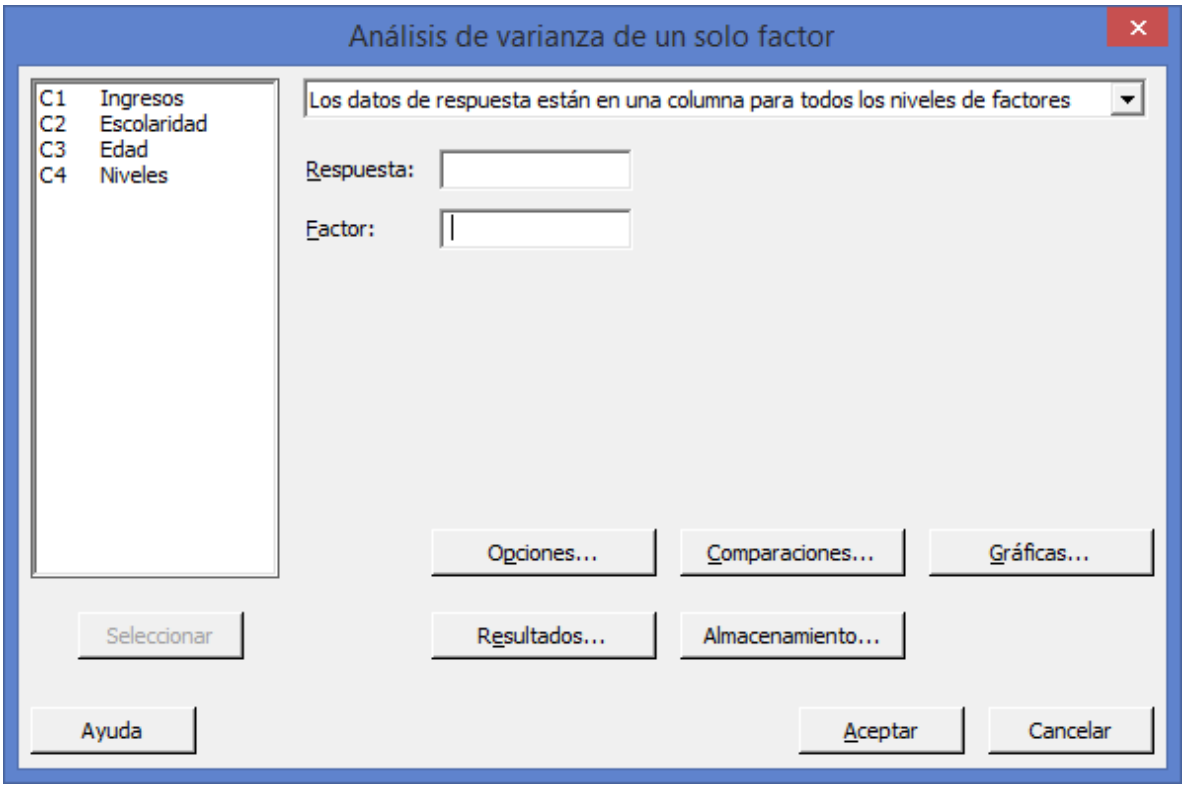
estadísticamente el grado de influencia del Nivel de Satisfacción **Moderadamente Satisfecho** sobre **Escolaridad** es igual que el Nivel de Satisfacción **Muy Satisfecho** sobre **Escolaridad.**

En cuanto al intervalo de confianza de **Escolaridad** en **Muy Insatisfecho** (Nivel 3) con **Escolaridad** en **Muy Satisfecho** (Nivel 1) oscila entre **-4.51 y 0.3.02**. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del Nivel de Satisfacción **Muy Insatisfecho** sobre **Escolaridad** es igual que el Nivel de Satisfacción **Muy Satisfecho** sobre **Escolaridad.**

Por último podemos observar que el intervalo de confianza de **Escolaridad** en **Muy Insatisfecho** (Nivel 3) con **Escolaridad** en **Moderadamente Satisfecho** (Nivel 2) oscila entre **-3.72 y 3.43**. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del Nivel de Satisfacción **Muy Insatisfecho** sobre **Escolaridad** es igual que el Nivel de Satisfacción **Moderadamente Satisfecho** sobre **Escolaridad.**

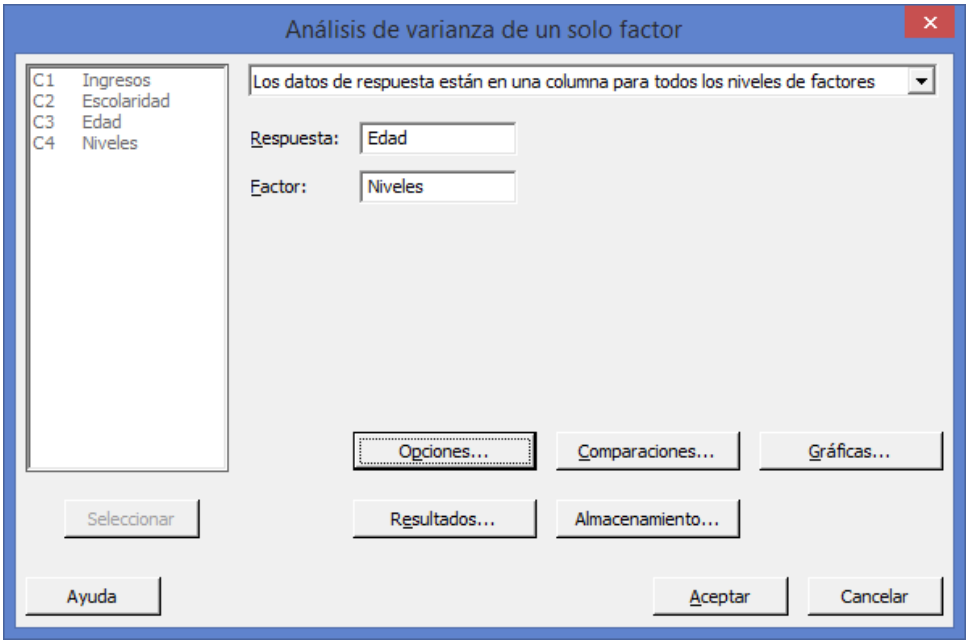
3. Prueba ANOVA de Hipótesis para Edad y pruebas Post-hoc de Tukey para Edad:

Como tenemos un **modelo con un solo factor fijo** seleccionamos la opción **Anova** y **un solo factor** del menú **Estadìsticas,**

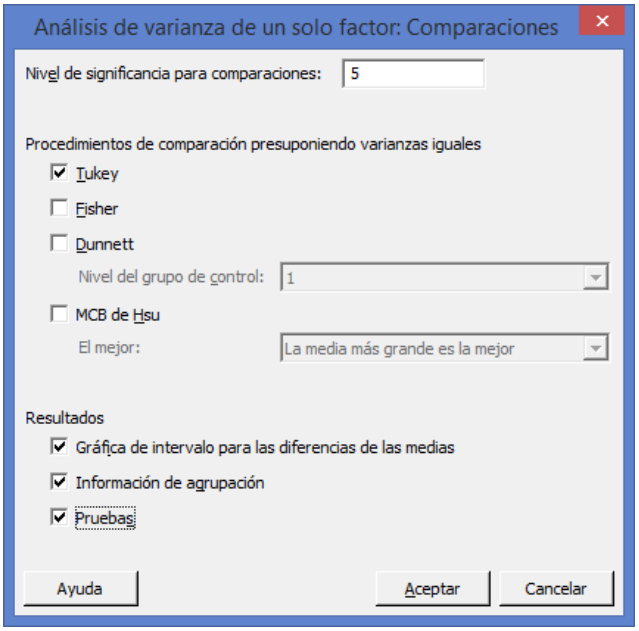


El primer elemento del cuadro de dialogo debe decir: **Los datos de respuesta están en una columna para todos los niveles de factores** debido a que así fueron introducidos los datos. Esta opción por regla general la da por default, sino hay que seleccionarla.

En Respuesta, ingrese **Edad**. En Factor, ingrese **Niveles**



Haga clic en el botón Comparaciones. En Procedimientos de comparaciones marque T**ukey**, y en Resultados marque **Pruebas.**



Haga clic en Aceptar en cada cuadro de dialogo.

Salida de la ventana Sesión

ANOVA unidireccional: Edad vs. Niveles

```
Método

Hipótesis nula          Todas las medias son iguales
Hipótesis alterna       Por lo menos una media es diferente
Nivel de significancia  α = 0.05

Se presupuso igualdad de varianzas para el análisis.


Información del factor

Factor   Niveles  Valores
Niveles        3  1, 2, 3


Análisis de Varianza

Fuente   GL  SC Ajust.  MC Ajust.  Valor F  Valor p
Niveles   2      24.98      12.49     0.21    0.813
Error    15     894.13      59.61
Total    17     919.11


Resumen del modelo

                     R-cuad.  R-cuad.
      S  R-cuad.  (ajustado)   (pred)
7.72066    2.72%       0.00%    0.00%


Medias

Niveles  N  Media  Desv.Est.     IC de 95%
1        5  34.40       5.13  (27.04, 41.76)
2        6  36.50       9.77  (29.78, 43.22)
3        7  37.29       7.20  (31.07, 43.51)

Desv.Est. agrupada = 7.72066
```

Interpretación de los resultados

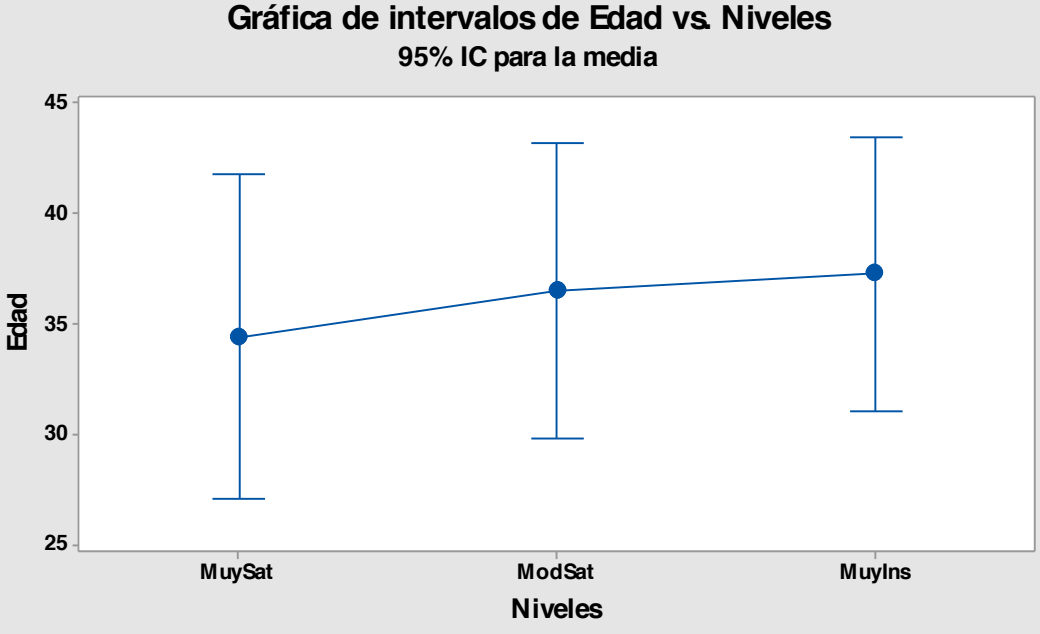
En la Tabla de ANOVA, para **Edad** en **los tres Niveles de Satisfacción** utilizados el valor de F calculada de **0.21** con un valor p (**0.813**) indica que hay suficiente evidencia de que todas las medias son iguales cuando alfa se establece en 0.05. En este caso en particular como no hay diferencias entre medias no se deberían examinar los resultados de las comparaciones múltiples, sin embargo como ejercicio se llevará a cabo para comprobar que los tres intervalos de confianza deben pasar por cero.

Comparaciones en parejas de Tukey

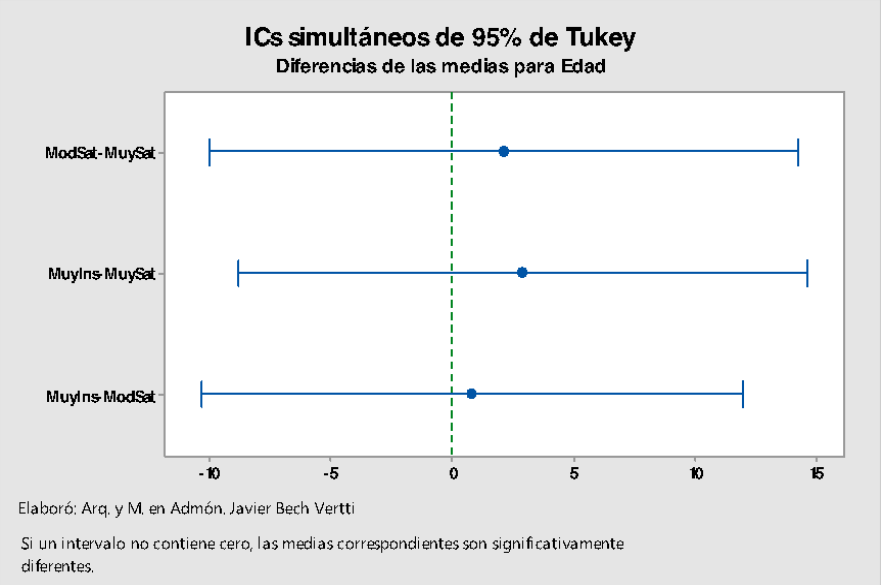
```
Agrupar información utilizando el método de Tukey y una confianza de 95%

Niveles  N  Media  Agrupación
3        7  37.29  A
2        6  36.50  A
1        5  34.40  A

Las medias que no comparten una letra son significativamente diferentes.


Pruebas simultáneas de Tukey para diferencias de las medias

            Diferencia
Diferencia      de las       EE de                              Valor p
de niveles      medias  diferencia     IC de 95%     Valor T  ajustado
2 - 1             2.10        4.68  (-10.03, 14.23)     0.45     0.896
3 - 1             2.89        4.52  ( -8.85, 14.62)     0.64     0.802
3 - 2             0.79        4.30  (-10.36, 11.93)     0.18     0.982

Nivel de confianza individual = 97.97%
```

ICs simultáneos de 95% de Tukey

Gráfica de intervalos de Edad vs. Niveles



Elaboró: Arq. y M. en Admón. Javier Bech Vertti
La desviación estándar agrupada se utilizó para calcular los intervalos.



Interpretación de los resultados

De las pruebas simultaneas de Tukey para diferencias de medias podemos observar que el intervalo de confianza de **Edad** en **Moderadamente Satisfecho** (Nivel 2) con **Edad** en **Muy Satisfecho** (Nivel 1) oscila entre **-10.03** y **14.23**. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del Nivel de Satisfacción **Moderadamente Satisfecho** sobre **Edad** es igual que el Nivel de Satisfacción **Muy Satisfecho** sobre **Edad.**

En cuanto al intervalo de confianza de **Edad** en **Muy Insatisfecho** (Nivel 3) con **Edad** en **Muy Satisfecho** (Nivel 1) oscila entre **-8.85 y 14.62** . Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del Nivel de Satisfacción **Muy Insatisfecho** sobre **Edad** es igual que el Nivel de Satisfacción **Muy Satisfecho** sobre **Edad.**

Por último podemos observar que el intervalo de confianza de **Edad** en **Muy Insatisfecho** (Nivel 3) con **Edad** en **Moderadamente Satisfecho** (Nivel 2) oscila entre **-10.36 y 11.93**. Como el intervalo de confianza pasa por cero podemos decir que estadísticamente el grado de influencia del Nivel de Satisfacción **Muy Insatisfecho** sobre **Edad** es igual que el Nivel de Satisfacción **Moderadamente Satisfecho** sobre **Edad.**

RESPUESTAS A LA AUTOEVALUACIÓN CON REACTIVOS DE **FALSO Ó VERDADERO** DEL CAPÍTULO 3

1.F	2.F	3.F	4.F	5.V	6.F	7.V	8.F	9.F
10.F	11.V	12.V	13.F	14.F	15.F	16.V	17.F	18.V
19.V	20.V	21.V	22.F	23.V	24.F	25.V	26.V	27.V
28.F	29.V	30.F	31.V	32.F	33.F	34.V	35.F	36.F
37.F	38.V	39.V	40.F	41.F	42.V	43.F	44.V	45.V

RESPUESTAS A LA AUTOEVALUACIÓN CON REACTIVOS DE **OPCIÓN MÚLTIPLE** DEL CAPÍTULO 3

1.a	2.c	3.b	4.b	5.d	6.c	7.a	8.d	9.b
10.c	11.b	12.a	13.d	14.a	15.b	16.a	17.c	18.c
19.b	20.d	21.d	22.c	23.b	24.b	25.c		