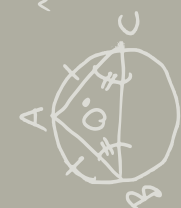


$$\frac{P(x)}{Q(x)} = \frac{G(x) + R(x)}{Q(x)} \quad y \frac{1}{n} = x$$

$$S^2 = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$$

$$\sim \forall x \forall y [p(x, y) \equiv \exists x \exists y [\sim p(x, y)]] \quad (a^m)^n = a^{m \times n}$$

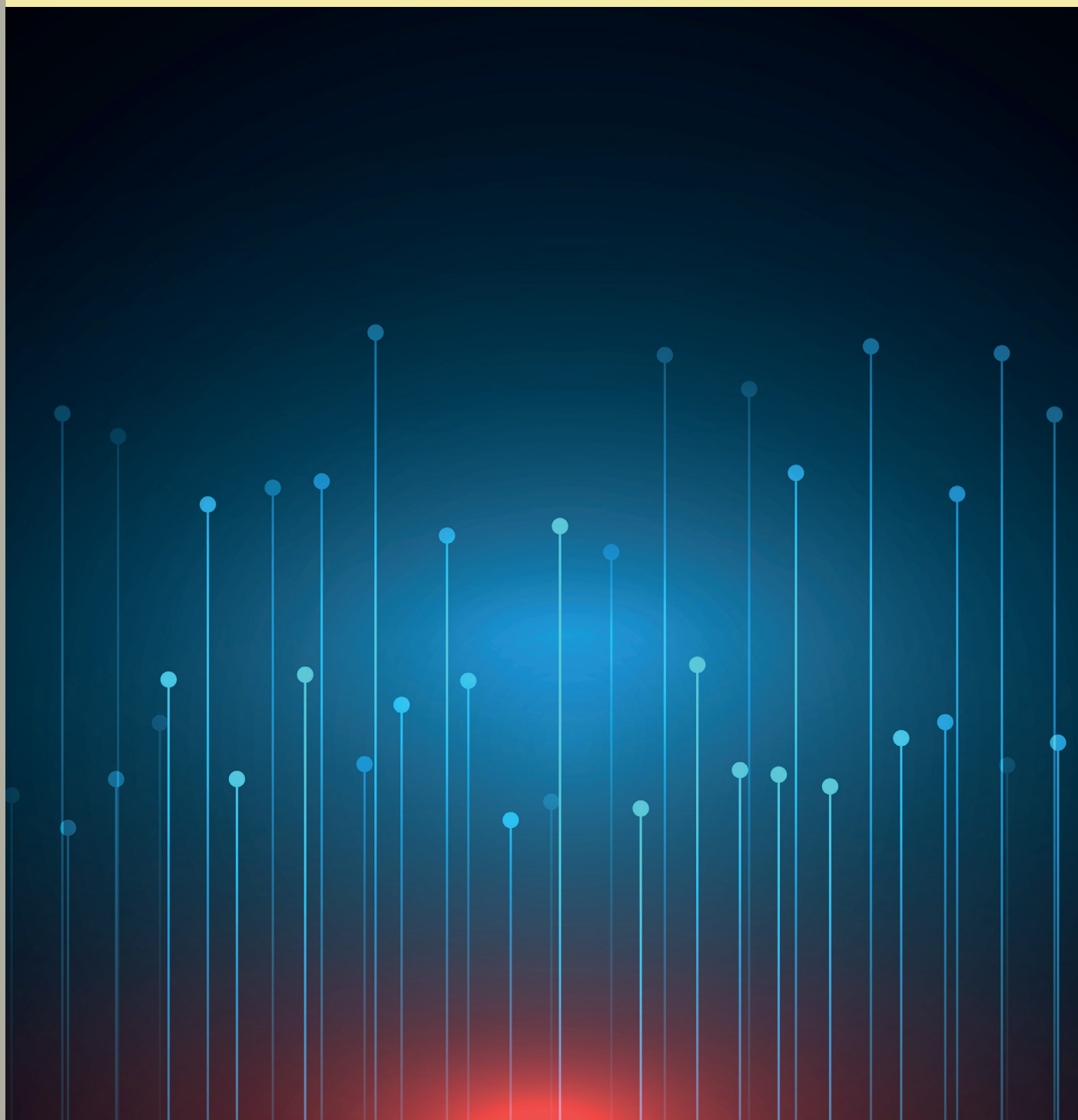


$$S^2 = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{N}}$$

$$\cos \frac{A}{2} = \pm \sqrt{\frac{1 + \cos A}{2}}$$



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



Estadística descriptiva con aplicaciones en SPSS

Santiago Rafael Urquiza Vaca, Luis Fredy Basantes Moreano y
Mónica Jeanette Flores Villafuerte

Estadística descriptiva con aplicaciones en SPSS

Santiago Rafael Urquizo Vaca, Luis Fredy Basantes Moreano y
Mónica Jeanette Flores Villafuerte

Primera edición electrónica: noviembre, 2025

ISBN: 978-9942-652-37-9

Revisión científica:

Mtr. Michael Estefania Játiva Brito - Escuela Superior Politécnica de Chimborazo

Mgtr. Mónica Daniela Barreiro Linzán - Universidad Técnica de Manabí

Universidad de las Fuerzas Armadas-ESPE

Crnl. de C.S.M. Víctor Villavicencio A., Ph. D.

Rector

Publicación autorizada por:

Comisión Editorial de la Universidad de las Fuerzas Armadas-ESPE

Cpfg. Joseph Alexander Guamán Seis, Ph.D. - Vicerrector de Investigación

Presidente

Corrección de estilo y diseño

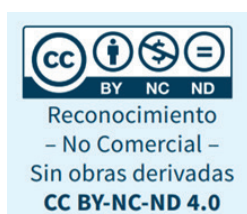
Mtr. Xavier Chinga

Imagen de cubierta: <https://acortar.link/omy4Ee>

Derechos reservados. Se prohíbe la reproducción de esta obra por cualquier medio impreso, reprográfico o electrónico. El contenido, uso de fotografía, gráficos, cuadros, tablas, y referencias es de exclusiva responsabilidad de los autores.

Universidad de las Fuerzas Armadas-ESPE
Av. General Rumiñahui s/n, Sangolquí, Ecuador
www.espe.edu.ec

Los derechos de esta edición electrónica son de la Universidad de las Fuerzas Armadas-ESPE, para consulta de profesores y estudiantes de la universidad e investigadores en www.repositorio.espe.edu.ec.



Departamento de
Ciencias Exactas

Estadística descriptiva con aplicaciones en SPSS

Ing. Santiago Rafael Urquizo Vaca, Msc.

Ing. Luis Fredy Basantes Moreano, Msc.

Ing. Mónica Jeanette Flores Villafuerte, Mg.

EDITORIAL



UNIVERSIDAD DE LAS FUERZAS ARMADAS - ESPE

Santiago Rafael Urquiza Vaca

srurquiza@espe.edu.ec

- Docente Titular del Departamento de Ciencias Exactas de la Universidad de las Fuerzas Armadas – ESPE, sede Latacunga.
- Máster Universitario en Ingeniería Matemática y Computación, Universidad Internacional de la Rioja (UNIR).
- Magister en Tecnología de la Información y Multimedia Educativa, Universidad Técnica de Ambato.
- Diplomado Internacional en Competencias Docentes, Tec. de Monterrey – Cambridge, Tecnológico de Monterrey.
- Diplomado superior en Gestión para el Aprendizaje Universitario, Escuela Politécnica del Ejército.
- Ingeniero en Electrónica, Universidad Técnica de Ambato.

Luis Fredy Basantes Moreano

lfbasantes@espe.edu.ec

- Docente Titular del Departamento de Ciencias Exactas de la Universidad de las Fuerzas Armadas - ESPE, sede Latacunga.
- Magíster en Estadística Aplicada, con Mención en Estadística Aplicada, Universidad Técnica de Manabí.
- Magíster en Docencia Matemática, Universidad Técnica de Ambato.
- Diplomado Internacional en Competencias Docentes, Tec. de Monterrey – Cambridge, Tecnológico de Monterrey.
- Diplomado Internacional en Estadística Aplicada a la Investigación, Corporación Universitaria Remington.
- Ingeniero Automotriz, Escuela Politécnica del Ejército.

Mónica Jeanette Flores Villafuerte

mjfloresvillafuerte@gmail.com

- Ex - Docente del Departamento de Ciencias Exactas de la Universidad de las Fuerzas Armadas - ESPE, sede Latacunga.
- Maestría en Matemática Aplicada mención en Matemática Computacional, Universidad Nacional de Chimborazo (c).
- Magister en Sistemas de Telecomunicaciones, Escuela Superior Politécnica de Chimborazo.
- Ingeniera Electrónica y Computación, Escuela Superior Politécnica de Chimborazo.

Índice

Presentación.....	15
Capítulo I - Introducción a SPSS: Instalación y Configuración.....	17
1.1. ¿Qué es SPSS y para qué se utiliza?.....	19
1.2. Requisitos de hardware y software.....	19
1.3. Acceso al programa.....	20
1.4. Interfaz de usuario.....	20
1.5. Cargar datos.....	26
1.6. Importar datos desde un archivo.....	26
1.7. Guardar un archivo.....	27
1.8. Seleccionar casos.....	27
1.9. Transformar / Calcular Variable.....	31
1.10. Transformar / Recodificación de variables.....	34
1.10.1. Transformar / Recodificar en las mismas variables.....	34
1.10.2. Transformar / Recodificar en distintas variables.....	38
1.10.3. Transformar / Recodificación automática.....	41
Capítulo II - Introducción a la Estadística.....	53
2.1. Definición de Estadística.....	55
2.2. Tipos de Estadística.....	55
2.2.1. Estadística Descriptiva.....	55
2.2.2. Estadística Inferencial.....	55

2.3. Población.....	56
2.4. Muestra.....	56
2.5. Parámetro.....	57
2.6. Estadístico.....	58
2.7. Variables.....	58
2.7.1. Variables cualitativas.....	59
2.7.2. Variables cuantitativas.....	59
2.8. Niveles de medición.....	61
2.8.1. Nivel de tipo nominal.....	61
2.8.2. Nivel de tipo ordinal.....	62
2.8.3. Nivel de tipo intervalo.....	62
2.8.4. Nivel de tipo razón.....	63
2.9. Procedimiento para realizar un análisis estadístico.....	65
2.9.1. Planificación.....	66
2.9.2. Recopilación de la información.....	68
2.9.3. Organización de la información.....	69
2.9.4. Presentación de resultados.....	70
2.9.5. Interpretación de resultados.....	72
2.10. Ejercicios propuestos.....	102
Capítulo III - Métodos gráficos.....	107
3.1. Gráficos para el análisis de variables cualitativas.....	109
3.1.1. Diagrama simple de barras.....	109
3.1.2. Diagrama múltiple de barras.....	109

3.1.3. Diagrama de barras compuesto.....	109
3.1.4. Diagrama circular.....	118
3.1.5. Ejercicios propuestos.....	127
3.2. Gráficos para el análisis de variables cuantitativas.....	131
3.2.1. Histogramas.....	131
3.2.2. Polígonos de frecuencia.....	140
3.2.3. Diagrama de Caja y Bigote.....	144
3.2.4. Ejercicios propuestos.....	161
Capítulo IV - Medidas descriptivas.....	167
4.1. Medidas de Tendencia Central.....	169
4.1.1. Media Aritmética.....	169
4.1.2. Media Geométrica.....	176
4.1.3. Media Armónica.....	179
4.1.4. Mediana.....	182
4.1.5. Moda.....	193
4.1.6. Ejercicios propuestos.....	197
4.2. Medidas de Posición.....	201
4.2.1. Cuartiles.....	201
4.2.2. Deciles.....	202
4.2.3. Percentiles.....	203
4.2.4. Ejercicios propuestos.....	211
4.3. Medidas de Dispersión.....	216
4.3.1. Rango.....	216

4.3.2. Varianza.....	217
4.3.3. Desviación Estándar.....	217
4.3.4. Coeficiente de Variación.....	224
4.3.5. Ejercicios propuestos.....	230
4.4. Medidas de forma.....	233
4.4.1. Asimetría.....	233
4.4.2. Curtosis.....	235
4.4.3. Ejercicios propuestos.....	252
Capítulo V - Teoría de Probabilidades.....	257
5.1. Definición de probabilidad.....	259
5.2. Conceptos básicos.....	259
5.3. Enfoques de Probabilidad.....	260
5.3.1. Probabilidad clásica.....	260
5.3.2. Probabilidad empírica.....	262
5.3.3. Probabilidad subjetiva.....	263
5.4. Ejercicios propuestos.....	264
5.5. Reglas de probabilidad.....	266
5.5.1. Regla del complemento.....	266
5.5.2. Regla de la adición o suma.....	267
5.5.3. Regla del producto.....	271
5.5.4. Regla de la Probabilidad Condicional.....	273
5.5.5. Regla de la Probabilidad Total.....	275
5.5.6. Ejercicios propuestos.....	283

5.6. Teorema de Bayes, Diagrama de Árbol y Tablas de Contingencia.....	286
5.6.1. Teorema de Bayes.....	286
5.6.2. Diagramas de árbol.....	288
5.6.3. Tablas de contingencia.....	290
5.6.4. Ejercicios propuestos.....	293
5.7. Distribución de probabilidades.....	295
5.7.1. Distribución de probabilidades discretas.....	299
5.7.2. Distribución de probabilidades continuas.....	326
Anexos - Respuestas a los Ejercicios Propuestos.....	359
Referencias.....	383

Índice de tablas

Tabla 1	<i>Fórmulas Media Aritmética.....</i>	170
Tabla 2	<i>Fórmulas Media Geométrica.....</i>	177
Tabla 3	<i>Fórmulas Media Armónica.....</i>	180
Tabla 4	<i>Fórmulas Mediana.....</i>	183
Tabla 5	<i>Fórmulas Moda.....</i>	193
Tabla 6	<i>Fórmulas Cuartiles.....</i>	202
Tabla 7	<i>Fórmulas Deciles.....</i>	203
Tabla 8	<i>Fórmulas Percentiles.....</i>	204
Tabla 9	<i>Fórmulas Varianza.....</i>	217
Tabla 10	<i>Fórmulas Desviación Estándar.....</i>	218
Tabla 11	<i>Fórmulas Coeficiente de Variación.....</i>	225
Tabla 12	<i>Fórmulas Coeficiente de Asimetría de Fisher.....</i>	234
Tabla 13	<i>Fórmulas Corrección del Coeficiente de Asimetría de Fisher.....</i>	234
Tabla 14	<i>Fórmulas Coeficiente de Curtosis.....</i>	235
Tabla 15	<i>Fórmulas Exceso de Curtosis.....</i>	236

Índice de figuras

Figura 1	<i>Ventana de inicio de SPSS</i>	20
Figura 2	<i>Interfaz de usuario</i>	21
Figura 3	<i>Ventana vista de datos</i>	22
Figura 4	<i>Ventana vista de variables</i>	22
Figura 5	<i>Tipo de variable</i>	24
Figura 6	<i>Etiquetas de valor</i>	24
Figura 7	<i>Valores perdidos</i>	25
Figura 8	<i>Ingreso de datos por teclado</i>	26
Figura 9	<i>Seleccionar casos donde Sí se satisface la condición</i>	28
Figura 10	<i>Seleccionar casos. Ventana de Edición</i>	29
Figura 11	<i>Casos seleccionados</i>	30
Figura 12	<i>Seleccionar casos. Todos los casos</i>	31
Figura 13	<i>Calcular variable</i>	33
Figura 14	<i>Recodificar en las mismas variables</i>	36
Figura 15	<i>Recodificar en las mismas variables. Valores anti- guos y nuevos</i>	36
Figura 16	<i>Recodificar en distintas variables</i>	39
Figura 17	<i>Recodificar en variables diferentes: valores antiguo y nuevo</i>	40
Figura 18	<i>Recodificación automática</i>	43
Figura 19	<i>Relación entre muestra y población</i>	57
Figura 20	<i>Mapa conceptual del tipo de variables</i>	58

Figura 21	Niveles de medición.....	61
Figura 22	Organización de los datos.....	69
Figura 23	Agrupación visual, selección de variables.....	83
Figura 24	Creación de puntos de corte.....	85
Figura 25	Agrupación visual. Crear etiquetas.....	86
Figura 26	Histograma simétrico.....	132
Figura 27	Histograma con sesgo positivo.....	132
Figura 28	Histograma con sesgo negativo.....	133
Figura 29	Histograma bimodal.....	133
Figura 30	Histograma con valores atípicos.....	134
Figura 31	Diagrama de Caja y Bigote.....	144
Figura 32	Medidas descriptivas.....	169
Figura 33	Cuartiles.....	202
Figura 34	Deciles.....	202
Figura 35	Percentiles.....	203
Figura 36	Asimetría.....	233
Figura 37	Tipos de Curtosis.....	236
Figura 38	Rango de probabilidad.....	259
Figura 39	Lanzamiento de una moneda 3 veces.....	261
Figura 40	Baraja española.....	265
Figura 41	Baraja alemana.....	265
Figura 42	Juego de dominó chino.....	266
Figura 43	Regla de la Adición con diagrama de Venn.....	268

Figura 44 Representación de la probabilidad total entre eventos y sub eventos.....	275
Figura 45 Representación de un diagrama de árbol.....	288
Figura 46 Representación de la función de densidad de probabi- lidad continua.....	327
Figura 47 Gráfica función de distribución acumulada.....	328
Figura 48 Representación de una curva normal.....	337
Figura 49 Representación de una curva normal estándar.....	338

Presentación

Desde la antigüedad, la Estadística Descriptiva ha sido una herramienta matemática muy utilizada para levantar datos poblacionales y económicos de las civilizaciones; actualmente, se ha convertido en una herramienta fundamental para análisis de datos masivos utilizados para interpretar, comprender y comunicar información en múltiples campos del conocimiento. El libro que se presenta a continuación integra la teoría con la práctica y se orienta a la resolución de problemas reales aplicados a la vida estudiantil, a través del uso del software SPSS, una de las herramientas más utilizadas en el análisis de datos a nivel profesional y académico.

Este libro surge de la práctica docente, del contacto directo con alumnos universitarios que, al adentrarse en el mundo de los datos, no solo necesitan herramientas matemáticas, sino también comprensión, aplicación y ejemplos vinculados a su futuro ámbito laboral. En este texto, se busca ofrecer conceptos accesibles, casos prácticos y actividades detalladas que ayuden al estudiante no solo a dominar los procedimientos estadísticos, sino también a cultivar un pensamiento reflexivo y un análisis crítico de la información.

Cada sección del libro se ha diseñado con un enfoque pedagógico claro, guiando al estudiante desde los fundamentos hasta el análisis interpretativo, introduciendo el manejo de SPSS de manera simple, gradual y accesible. No se trata simplemente de dominar cálculos como promedios o gráficos de información, sino de interpretar su significado, descubrir la historia detrás de esos resultados y aprender a aplicar estos conocimientos con criterio ético en la resolución de problemas reales.

Los autores



CAPÍTULO I

Introducción a SPSS: Instalación y Configuración

1.1. ¿Qué es SPSS y para qué se utiliza?

El software SPSS es un paquete de computación que permite la realización de análisis estadísticos de forma sencilla y amigable, se constituye en una potente aplicación dotada de una interfaz gráfica muy fácil de manejar y tiene una capacidad de procesamiento mayor a 30.000 variables, es decir, SPSS maneja grandes volúmenes de datos. Su aplicación puede utilizarse para:

- Manejo y análisis de grandes volúmenes de datos: Permite organizar, limpiar y explorar datos de diversas fuentes.
- Aplicación de técnicas estadísticas avanzadas: Ofrece una amplia gama de procedimientos estadísticos para modelar datos, poner a prueba hipótesis y relacionar variables.
- Realiza predicciones: Puede utilizarse para prever comportamientos de clientes, tendencias del mercado y otros eventos futuros.
- Apoya la toma de decisiones: Facilita la obtención de conclusiones aceptadas y detalladas de la información.

1.2. Requisitos de hardware y software

Para un buen funcionamiento del software SPSS en las versiones más recientes, es necesario cumplir con los siguientes requisitos de software y hardware.

Los requisitos mínimos de hardware son:

- Procesador de 2.0 GHz o superior.
- Memoria RAM de 4 GB
- Espacio libre en el disco duro de 4 GB para su instalación.
- Tarjeta gráfica con resolución mínima de 1024 x 768 píxeles.

Los requisitos mínimos de software son:

- Sistema operativo en Windows 7, 8.1, 10 y 11 (Versión de 64 bits)
- Sistema Operativo en MAC OS 10.13 o superior (Versión de 64 bits).

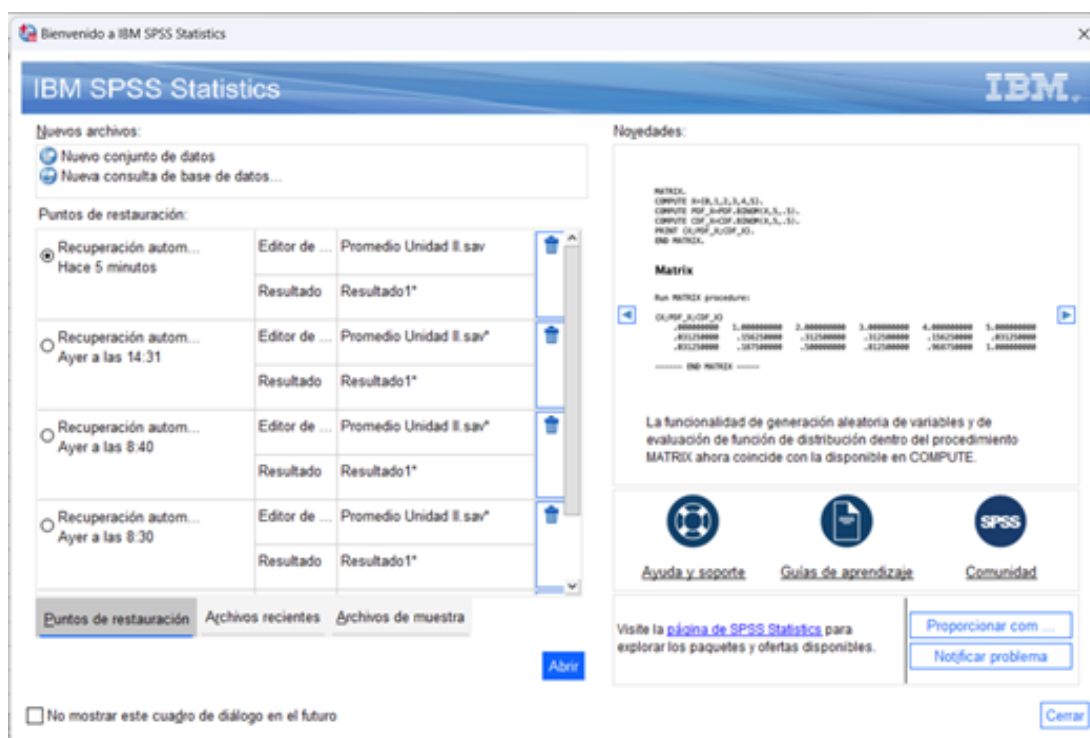
Otro de los requisitos adicionales es la conexión a internet, necesaria para la activación del producto, actualizaciones y accesos de funciones en línea.

1.3. Acceso al programa

Al ejecutar el programa, aparece una ventana con diversas opciones de tareas a realizar en la sesión de trabajo. Esta ventana se superpone a la ventana de Editor de datos; la ejecución que se realiza es cerrar esta ventana.

Figura 1

Ventana de inicio de SPSS

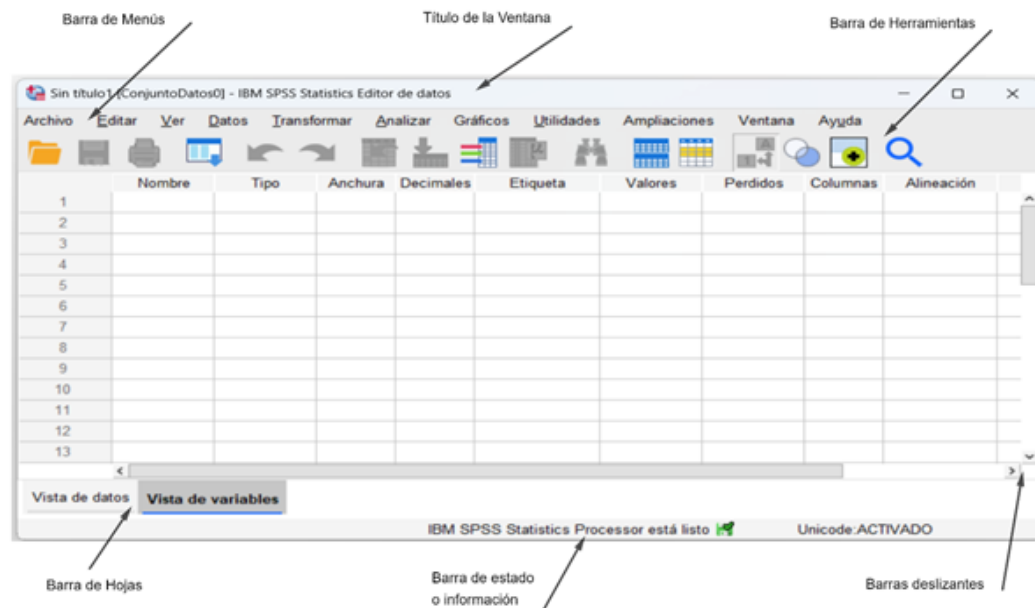


1.4. Interfaz de usuario

La ventana principal de SPSS es el editor de datos; esta es la interfaz de trabajo inicial que se utiliza para introducir los datos y realizar el procedimiento para el análisis estadístico. Esta se divide en varias secciones:

Figura 2

Interfaz de usuario



Barra de menús: Contiene opciones como Archivo, Editar, Ver, Datos, Transformar, Analizar, Gráficos, Utilidades, Ampliaciones, Ventana y Ayuda. Cada una de ellas tiene distintas funcionalidades.

Título de la ventana: Identifica el título del conjunto de datos que se está trabajando.

Barra de herramientas: Proporciona accesos rápidos a funciones comunes. La funcionalidad de estos íconos se muestra ubicando el puntero del mouse en cada uno de ellos.

Barras de estado o información: Aquí se puede visualizar el estado o ejecución del programa.

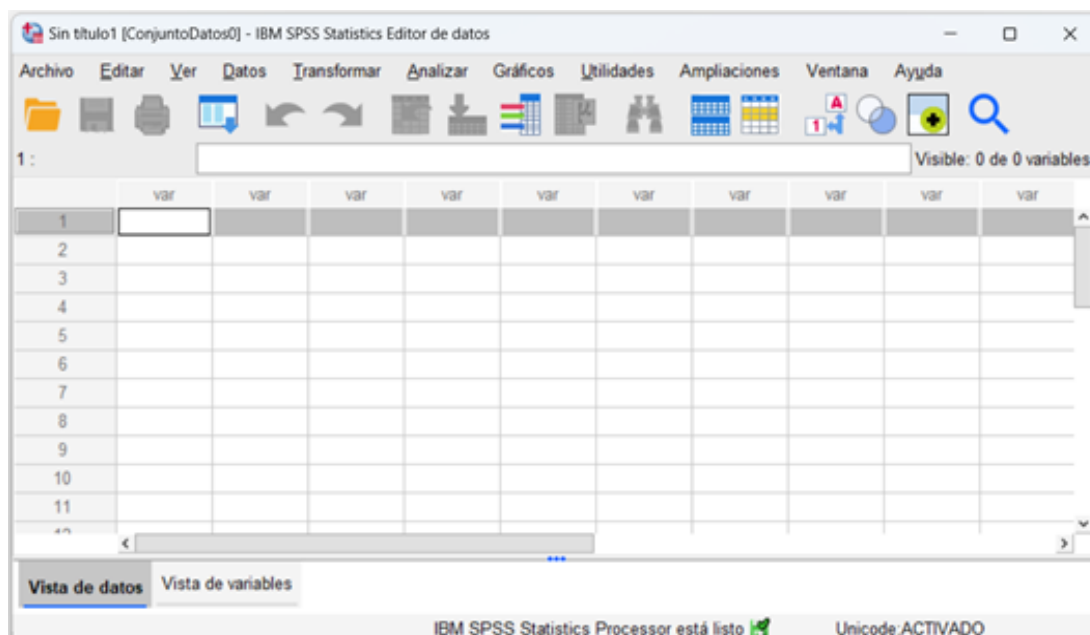
Barras deslizantes: Ayuda a desplazarse de manera horizontal y vertical en el conjunto de datos, permitiendo visualizar los datos y las variables.

Barra de hojas: Proporciona dos pestañas con la Vista de datos y la Vista de variables.

Vista de datos: Muestra los valores de los datos en forma de tabla, aquí se ingresan todos los datos correspondientes a cada variable.

Figura 3

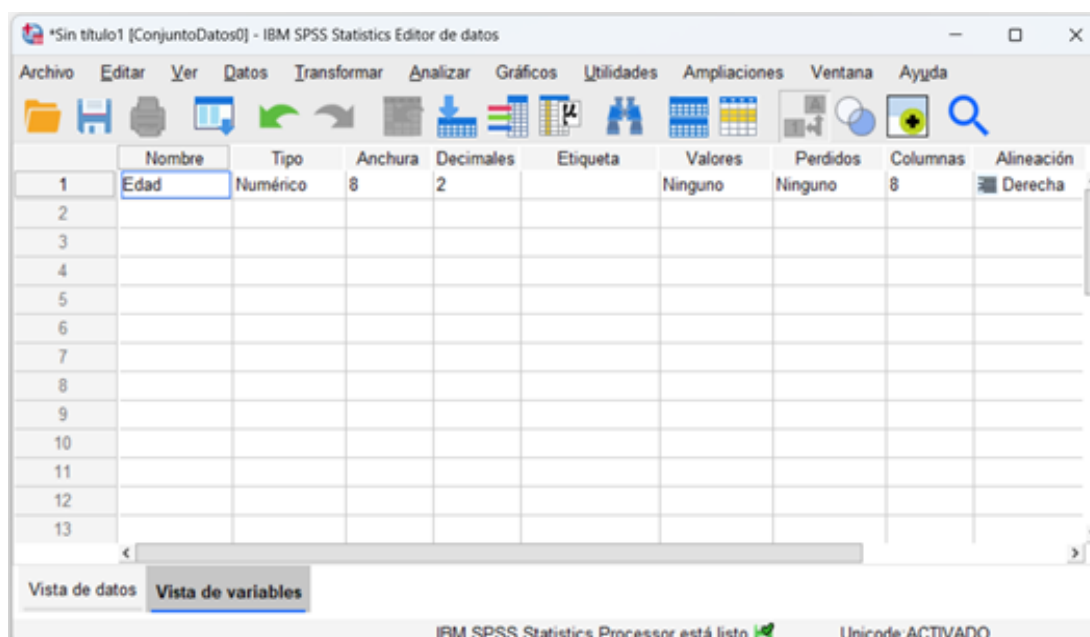
Ventana vista de datos



Vista de variables: Permite definir las propiedades y características de cada variable (nombre, tipo, etiquetas, etc.). Cada una de las filas representa a una variable y cada columna identifica una característica de esta, como se detalla a continuación.

Figura 4

Ventana vista de variables



Contiene los siguientes campos:

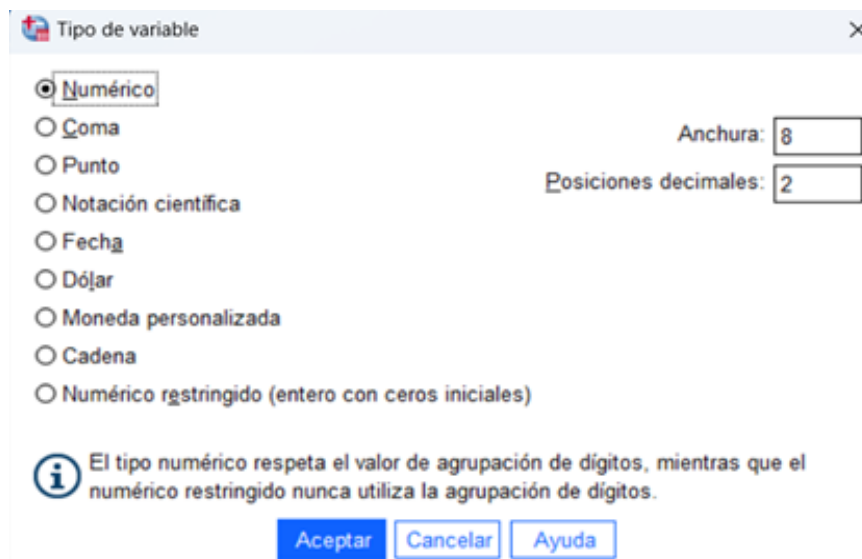
Nombre de la variable: Identifica la variable, tiene que ser único y sin espacios ni caracteres especiales. Ejemplo: edad, género, salario, etc.

Tipo: Se escoge de acuerdo al formato del tipo de variable en estudio, entre ellas se tiene:

- *Numérica:* Una variable cuyos valores son números. Los valores se muestran en formato numérico estándar (se establece la anchura y cifras decimales). Ejemplos: Edad 25 (sin decimales), Peso 68.5 (con decimales)
- *Coma:* Una variable numérica donde las comas determinan los miles. Ejemplo: Ingresos anuales USD 23 500.00 (Se ingresa 23500)
- *Punto:* Una variable numérica donde los puntos determinan los miles. Ejemplo: Altura 1.75 metros
- *Notación científica:* Una variable numérica cuyos valores se muestran con una letra E intercalada y un exponente con signo que representa una potencia de base diez, se utiliza cuando los valores son muy grandes. Ejemplo: Distancia en metros entre Ecuador y Estados Unidos 2 800 000 (Se ingresa 2.8E6)
- *Fecha:* Una variable que almacena fechas en diferentes formatos. Ejemplo: Fecha de nacimiento: 10 de mayo de 2018 (Se ingresa 10-05-2018)
- *Dólar:* Una variable numérica cuyos valores son monetarios. Ejemplo: PrecioUSD 20 (Se ingresa 20)
- *Moneda personalizada:* Una variable numérica cuyos valores se muestran en uno de los diferentes formatos de moneda personalizada que se hayan definido previamente en la pestaña Moneda del cuadro de diálogo Opciones (Edición / Opciones). Ejemplo: Salario en euros €2,500 (Se ingresa 2500)
- *Cadena:* Variable de texto que puede ser palabras, letras o números no calculables. Ejemplo: Nombre (Ana López), Código postal como texto (Se ingresa 05002)
- *Numérico restringido:* Números enteros que contienen ceros a la izquierda. Ejemplo: Código de un producto: 00132 (Se ingresa 132)

Figura 5

Tipo de variable



Anchura: Determina el ancho de la columna.

Decimales: Determina el número de decimales que aparecen en pantalla para datos numéricos.

Etiqueta: Se asigna un nombre que describe a la variable para facilitar la interpretación de los resultados.

Valores: Permite definir categorías o etiquetas a las variables cualitativas. Ejemplo en la variable género: 1 = “Masculino”, 2 = “Femenino”.

Figura 6

Etiquetas de valor

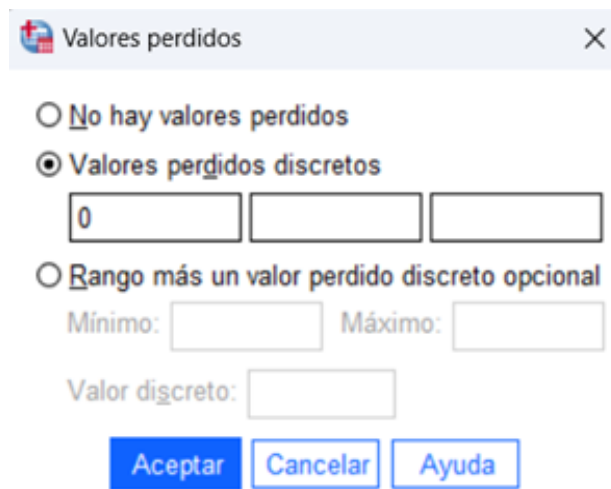


Valores perdidos: Son valores que no se encuentran en la base de datos; esto se debe a la falta de respuesta por parte de los encuestados o por errores en la recopilación.

Esta falta de información de los datos puede llevar a información errónea, por lo cual en SPSS estos datos perdidos se llenan generalmente con un valor específico de 0 si es una variable cuantitativa y 9 o 99999 si es una variable cualitativa codificada.

Figura 7

Valores perdidos



Columnas: Ajusta el ancho de la columna en la Vista de datos.

Alineación: Permite alinear los datos introducidos a la Izquierda, Derecha o Centrado

Medida: Permite definir la variable como Nominal, Ordinal o Escala.

Rol: Es una característica que ayuda a organizar y automatizar el análisis estadístico asignando funciones predefinidas a las variables, se utiliza para procesos de modelos predictivos

- *Entrada:* La variable se utiliza como una entrada (por ejemplo: predictor, variable independiente).
- *Objetivo:* La variable se utiliza como salida o de destino (por ejemplo: variable dependiente).
- *Ambos:* La variable se utiliza como entrada y salida.
- *Ninguna:* La variable no tiene una asignación de funciones.

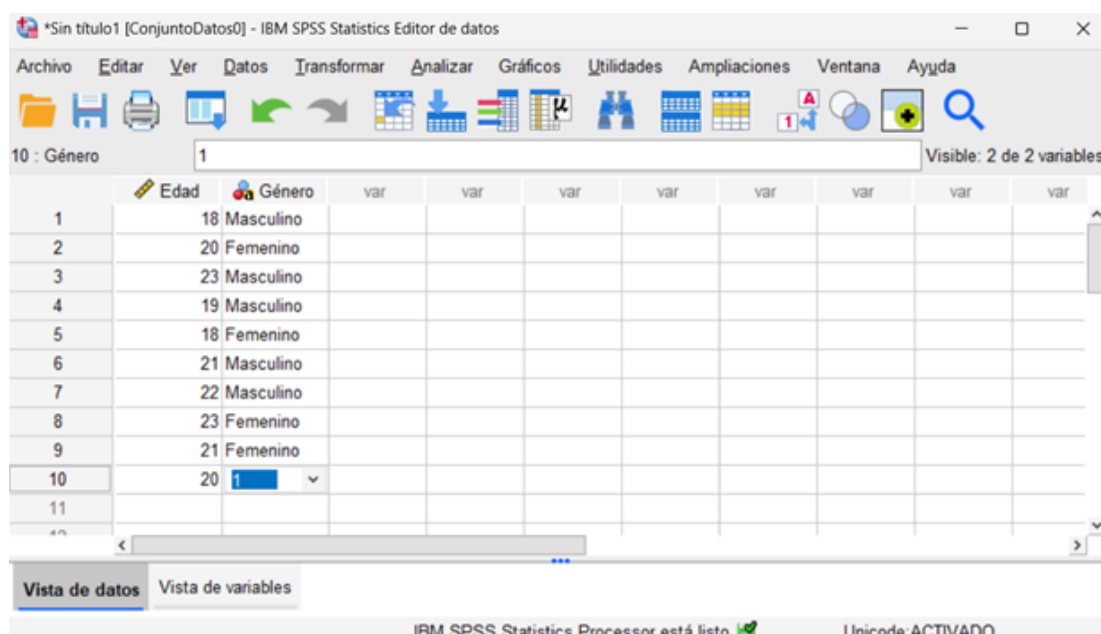
- *Partición:* La variable se utiliza para dividir los datos en muestras separadas.
- *División:* Segmenta los datos en grupos o categorías específicos durante el análisis.

1.5. Cargar datos

Para ingresar los datos manualmente se debe realizar clic en la pestaña Vista de datos, e ingresa los datos por medio del teclado.

Figura 8

Ingreso de datos por teclado



1.6. Importar datos desde un archivo

- Selecciona Archivo / Abrir / Datos.
- Seleccionar el archivo en formato compatible (.sav, Excel o CSV).
- Haz clic en “Abrir”. Si eliges un archivo Excel o CSV, sigue los pasos del asistente para especificar:
 - El rango de datos.
 - Leer nombres de variable desde la primera fila de datos.

- Porcentaje de valores que determinan tipo de datos
- Ignorar filas y columnas ocultas
- Aceptar

Autoevaluación 1.1.

1. Con sus propias palabras explicar qué es el software estadístico SPSS y para qué se utiliza
2. Cuáles son los requisitos de hardware y software para instalar el software estadístico SPSS
3. La interfaz de usuario de SPSS se divide en secciones cuales son ellas
4. Qué tipo de archivos se puede importar en SPSS
5. Explique que se puede observar y para qué sirve la ventana Vista de Datos
6. Explique que se puede observar y para qué sirve la ventana Vista de Variables

1.7. Guardar un archivo

- a. Seleccionar Archivo / Guardar como
- b. Seleccionar la ubicación para guardar el archivo.
- c. Especifica un nombre del archivo y selecciona el formato (.sav).
- d. Haz clic en “Guardar”.

1.8. Seleccionar casos

Esta opción permite filtrar y seleccionar casos específicos de una variable para realizar el análisis estadístico sobre los casos seleccionados; se aplica cuando el estudio necesita ciertas condiciones como seleccionar casos de interés para el estudio.

Ejercicio práctico N°1

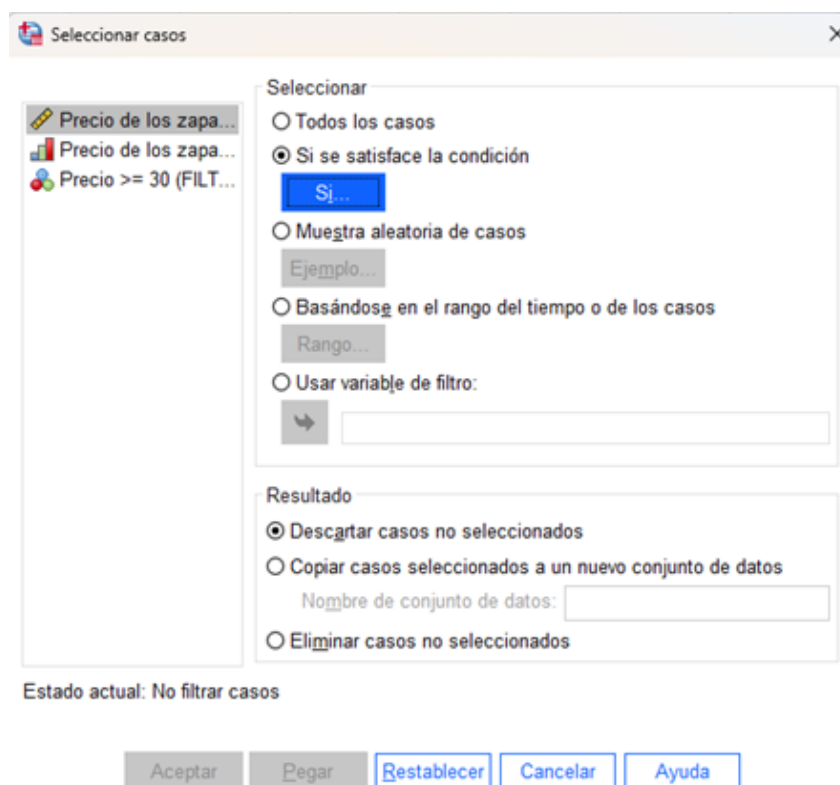
Se realiza un estudio sobre el precio de zapatos deportivos en el centro comercial El Tejar de la ciudad de Quito. Para el estudio, se desea seleccionar solo los precios mayores o iguales a 30 dólares.

Construcción en SPSS

- Descargar el archivo en el siguiente link <https://drive.google.com/file/d/1KH9pdUsn6jIvGEN6FhB5meApUhoAGZx8/view?usp=sharing> y guardarlo con el nombre Precio
- Abrir el SPSS y cargar los datos: Archivo / Abrir / Datos.
- Busca el archivo Precio y haz clic en Abrir
- Seleccionar la variable que se va a usar para filtrar los casos. Por ejemplo, supongamos que quieres seleccionar todos los casos donde la variable "Precio" sea mayor o igual a 30 dólares.
- Seleccionar Datos / Seleccionar Casos.
- En la ventana que aparece, selecciona la opción Si se satisface la condición / Si...

Figura 9

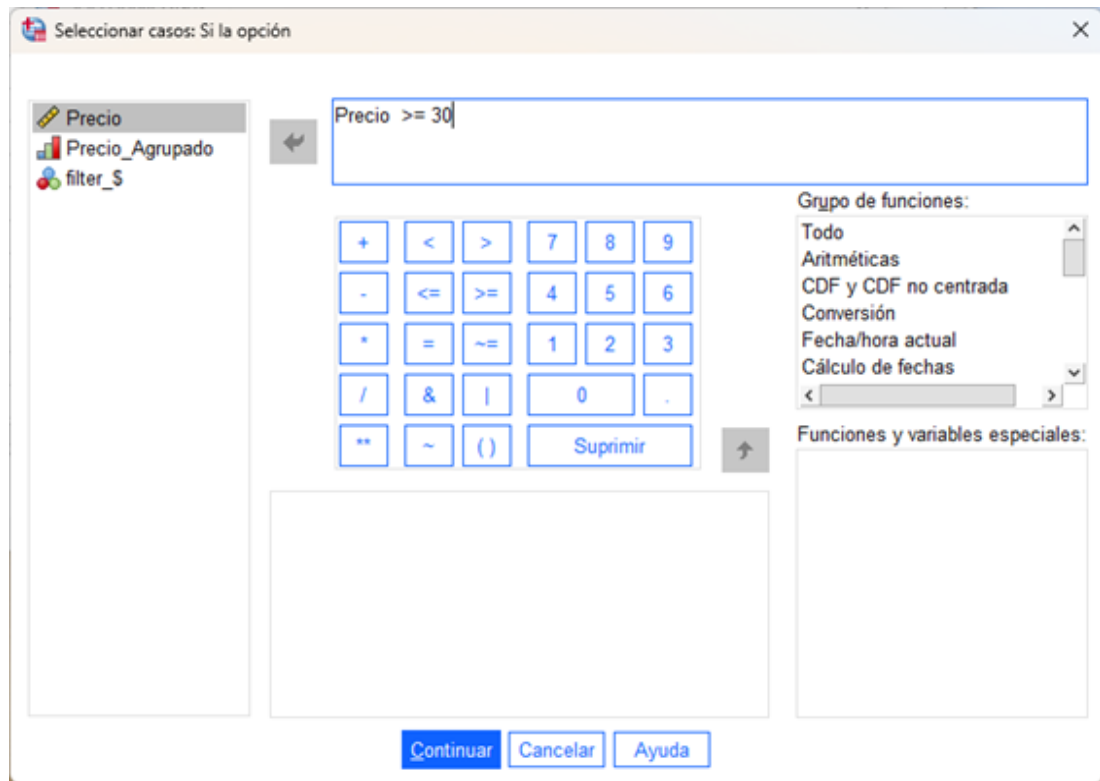
Seleccionar casos donde Sí se satisface la condición



En la ventana Selección de casos, escribe la condición para seleccionar los casos. Por ejemplo: Precio $\geq$ 30, clic en Continuar y luego en Aceptar.

Figura 10

Seleccionar casos. Ventana de Edición



Para verificar la selección de casos después de aplicar el filtro, se crea una nueva variable llamada *filter_\$* que indica cuáles casos están seleccionados (con un valor de 1) y cuáles no (con un valor de 0), así como también aparecerá una línea diagonal marcada sobre el caso indicando que el caso no está seleccionado y los que estén sin la línea diagonal serán los casos seleccionados para el análisis.

Figura 11

Casos seleccionados

11 :

	Precio	filter_\$	var	var	var	var	var	var	var
1	25	0							
2	26	0							
3	27	0							
4	27	0							
5	28	0							
6	28	0							
7	30	1							
8	30	1							
9	30	1							
10	32	1							

Otras opciones

Seleccionar casos. Muestra aleatoria de casos.

Esta opción permite elegir un número específico de casos aleatorios o un porcentaje de los casos totales. Se aplica para obtener una muestra aleatoria de la base de datos.

Seleccionar casos Basándose en el rango de tiempo de los casos.

Esta opción permite seleccionar casos dentro de un rango de valores; además, permite establecer un rango de valores para una variable y solo esos casos serán seleccionados.

Seleccionar casos Usando variable de filtro.

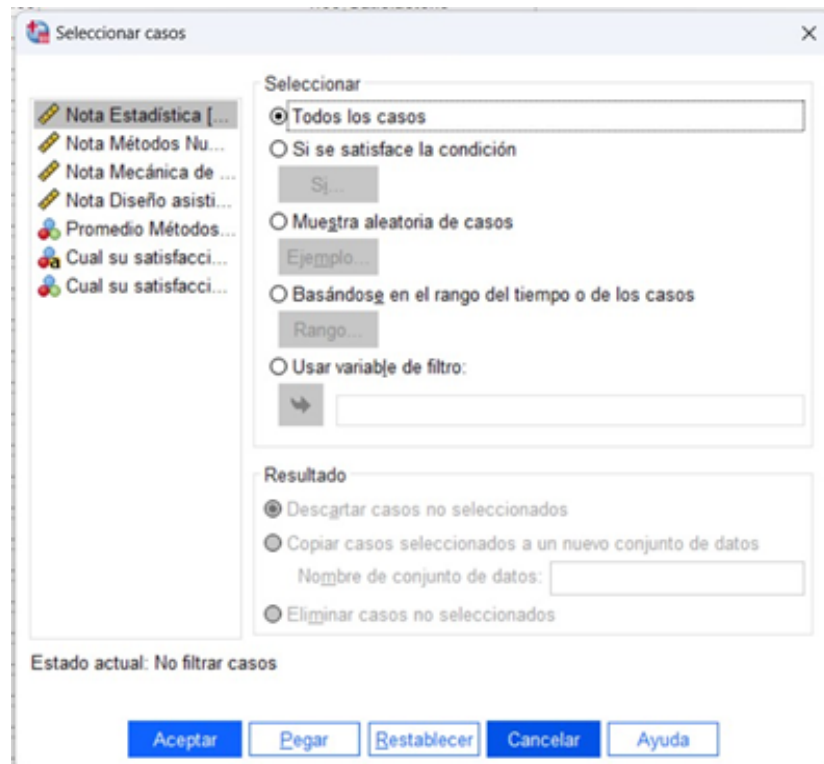
Esta opción permite usar una variable de filtro previamente definida. Si se ha creado una variable binaria (1 = seleccionado, 0 = no seleccionado), esta opción permite usar esa variable para filtrar casos.

Desactivar la selección de casos.

Esta opción permite desactivar la selección de casos; la ruta Datos / Seleccionar casos y elegir la opción Todos los casos. Esto eliminará el filtro y regresará todos los casos a su forma original.

Figura 12

Seleccionar casos. Todos los casos



1.9. Transformar / Calcular Variable

La opción “Calcular variable...” se utiliza para crear una nueva variable a partir de otras variables existentes en la base de datos. Se realizan operaciones matemáticas como (sumas, restas, multiplicaciones, divisiones) o aplicar funciones estadísticas (promedios, máximos, mínimos, etc.).

Ejercicio práctico N°2

Se realiza un estudio respecto al rendimiento académico en las asignaturas de Dominio Matemático y Dominio Lingüístico obteniendo las siguientes notas sobre 10 puntos:

Notas de dominio Matemático:

6.63	6.63	7.00	6.63	5.88	7.94	7.00	5.88	5.88	6.63	7.19	6.25	6.63	5.88
------	------	------	------	------	------	------	------	------	------	------	------	------	------

Notas de dominio Lingüístico:

5.67	5.67	7.18	6.67	7.18	5.67	6.00	4.67	5.33	6.33	5.67	5.67	6.67	6.67
------	------	------	------	------	------	------	------	------	------	------	------	------	------

Calcular el promedio entre estas dos asignaturas

Construcción en SPSS:

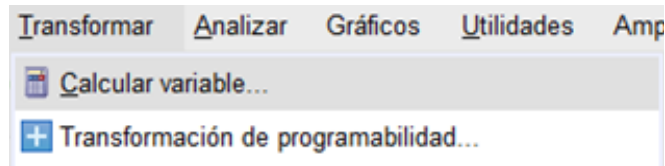
- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” crear nuevas variables llamadas NotaDM, NotaDL con los siguientes parámetros:

- **Nombre:** NotaDM
 - **Tipo:** Numérico
 - **Etiqueta:** Nota Dominio Matemático
 - **Medida:** Escala
-
- **Nombre:** NotaDL
 - **Tipo:** Numérico
 - **Etiqueta:** Nota Dominio Lingüístico
 - **Medida:** Escala

- c. En la ventana “Vista de datos” Ingrese las notas de las dos variables.

	NotaDM	NotaDL
1	6.63	5.67
2	6.63	5.67
3	7.00	7.18
4	6.63	6.67
5	5.88	7.18
6	7.94	5.67
7	7.00	6.00
8	5.88	4.67
9	5.88	5.33
10	6.63	6.33
11	7.19	5.67
12	6.25	5.67
13	6.63	6.67
14	5.88	6.67

d. En el menú principal seleccione: Trasformar / Calcular variable...



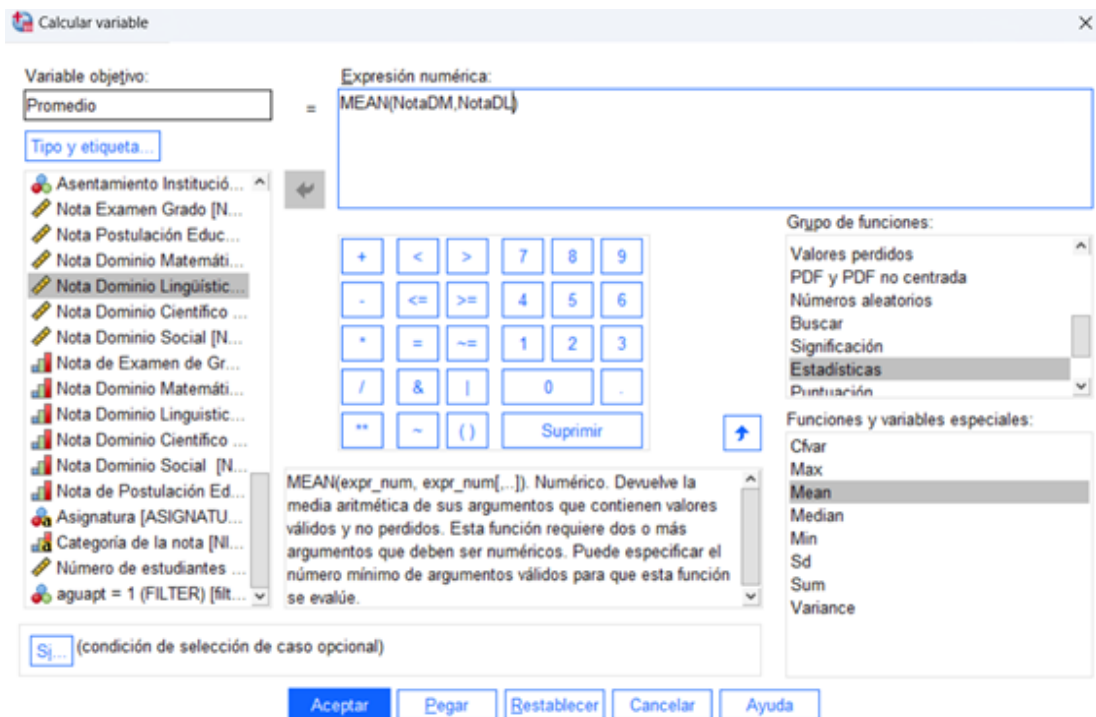
e. Crear una nueva variable llamada (Promedio)

f. De los grupos de funciones seleccionar Estadísticas / Mean, seleccionar las variables que se va a calcular el promedio (NotaDM,NotaDL)

g. Clic en Aceptar.

Figura 13

Calcular variable



Los resultados obtenidos son los promedios de las notas de las asignaturas de Matemáticas y Lingüística.

	 NotaDM	 NotaDL	 Promedio
1	6.63	5.67	6.15
2	7.19	6.00	6.60
3	5.50	6.00	5.75
4	6.63	5.00	5.81
5	7.19	5.67	6.43
6	5.50	5.67	5.59
7	6.25	6.00	6.13
8	5.88	7.35	6.62
9	6.25	6.00	6.13
10	5.88	7.00	6.44
11	6.25	7.35	6.80
12	5.50	5.67	5.59
13	6.63	6.33	6.48
14	7.19	5.67	6.43

1.10. Transformar / Recodificación de variables

Esta opción permite:

- Modificar los valores de una variable existente
- Reemplazar valores específicos por otros
- Crear nuevas categorías o rangos
- Eliminar o sustituir valores perdidos (missing values)
- Realizar transformaciones complejas mediante reglas personalizadas.

Esta opción es de mucha utilidad; se utiliza cuando no se desea crear una nueva variable, sino simplemente actualizar la variable original.

1.10.1. Transformar / Recodificar en las mismas variables

Esta opción permite modificar los valores de una variable existente, reemplazar valores específicos por otros, crear nuevas categorías o rangos, eliminar o sustituir valores perdidos (missing values) y realizar transformaciones complejas mediante reglas personalizadas.

Esta opción es de mucha utilidad, se utiliza cuando no se desea crear una nueva variable, sino simplemente actualizar la variable original.

Ejercicio práctico N° 1.

En el siguiente link: https://drive.google.com/file/d/11n0vmVTuC-rqpD_sm5BoijmYEv2YSejV/view?usp=sharing

Se realiza una investigación sobre los promedios de cuatro asignaturas de estudiantes de ingeniería, obteniéndose la siguiente base de datos. Codificar los promedios de la asignatura de Estadística en la misma variable, con las categorías: Insuficiente (<14), Elemental (14.01-16), Excelente (16.01-18) y Satisfactorio (>18.01)

Construcción en SPSS

1. Prepare la base de datos

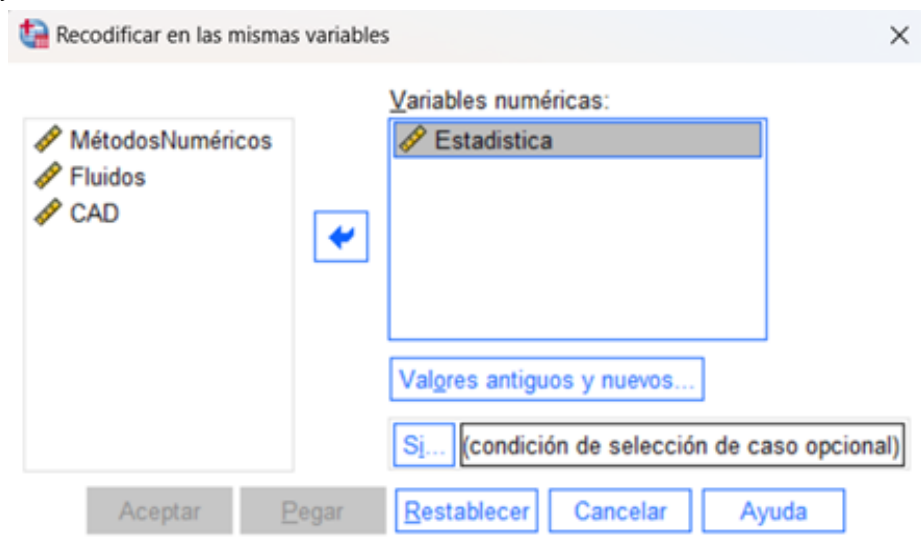
- Abra SPSS y crear la base de datos.
- Los casos en blanco, tratarlos como valores perdidos y colocar la codificación (99999)

	Estadística	Métodos Numéricos	Fluidos	CAD	var	var	var
1	15.50	18.50	99999.00	99999.00			
2	18.89	16.00	17.25	20.00			
3	12.50	15.50	17.50	16.90			
4	15.45	99999.00	99999.00	99999.00			
5	19.99	15.20	18.30	20.00			
6	15.26	99999.00	99999.00	99999.00			
7	16.00	20.00	20.00	10.00			
8	17.68	13.44	17.60	17.13			
9	16.34	15.80	9.00	14.00			
10	17.53	13.56	16.10	16.00			

- En el menú principal seleccionar: Transformar / Recodificar en las mismas variables...
- Seleccionar la variable a ser Recodificada (por ejemplo, Estadística) y clic en Valores antiguos y nuevos.

Figura 14

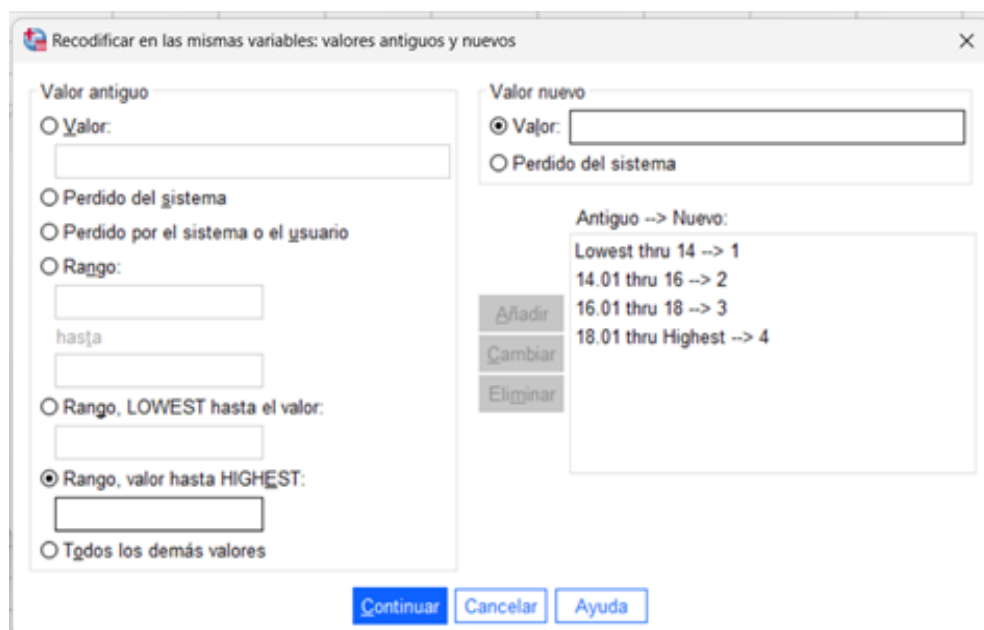
Recodificar en las mismas variables



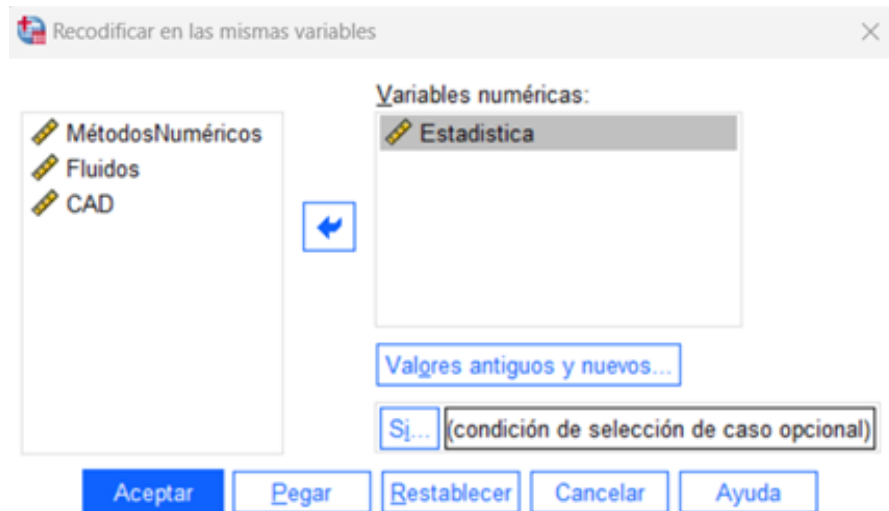
Recodificar la variable con los nuevos valores, en los rangos: Insuficiente (<14), Elemental (14.01-16), Excelente (16.01-18) y Satisfactorio (>18.01) y clic en continuar.

Figura 15

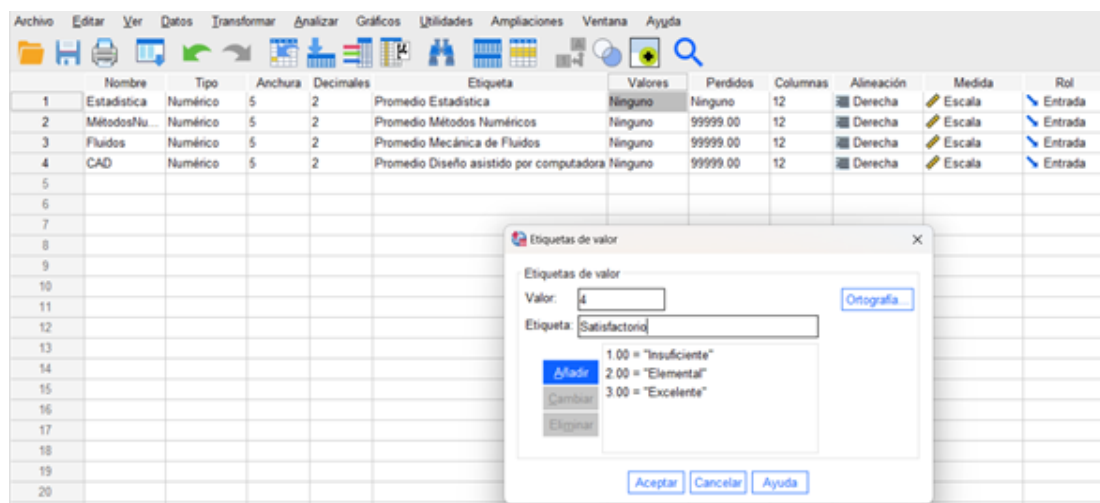
Recodificar en las mismas variables. Valores antiguos y nuevos



Clic en Aceptar.



- Seleccionamos la ventana vista de variables
- En la opción de Valores de la variable estadística, Añadir los nuevos valores de las etiquetas: 1 Insuficiente, 2 Elemental, 3 Excelente, 4 Satisfactorio.
- Clic en Aceptar



La variable numérica Estadística ha sido recodificada en la misma columna.



	Estadística	Métodos Numéricos	Fluidos	CAD	var	var	var
1	Elemental	18.50	99999.00	99999.00			
2	Satisfactorio	16.00	17.25	20.00			
3	Insuficiente	15.50	17.50	16.90			
4	Elemental	99999.00	99999.00	99999.00			
5	Satisfactorio	15.20	18.30	20.00			
6	Elemental	99999.00	99999.00	99999.00			
7	Elemental	20.00	20.00	10.00			
8	Excelente	13.44	17.60	17.13			
9	Excelente	15.80	9.00	14.00			
10	Excelente	13.56	16.10	16.00			

1.10.2. Transformar / Recodificar en distintas variables

Esta opción permite crear una nueva variable a partir de la recodificación de una variable existente, sin modificar los datos originales.

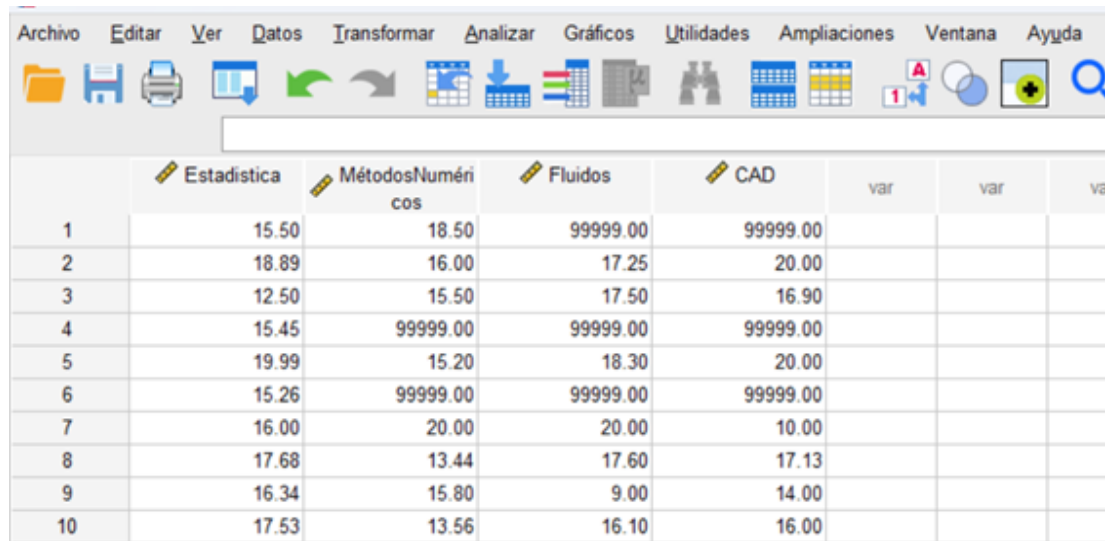
Ejercicio práctico N° 1

Se realiza una investigación sobre los promedios de cuatro asignaturas de estudiantes de ingeniería, obteniéndose la siguiente base de datos. Codificar a los promedios de la asignatura Métodos numéricos, en una nueva variable, con las categorías: Insuficiente (<14), Elemental (14.01-16), Excelente (16.01-18) y Satisfactorio (>18.01)

Construcción en SPSS

Prepare la base de datos

- Abra SPSS y crear la base de datos.
- Tratar los Valores perdidos (99999)



The screenshot shows the SPSS main menu bar with options: Archivo, Editar, Ver, Datos, Transformar, Analizar, Gráficos, Utilidades, Ampliaciones, Ventana, and Ayuda. Below the menu is a toolbar with various icons. Underneath is a data table with 10 rows and 7 columns. The columns are labeled: Estadística, Métodos Numéricos, Fluidos, CAD, var, var, and va. The data values are as follows:

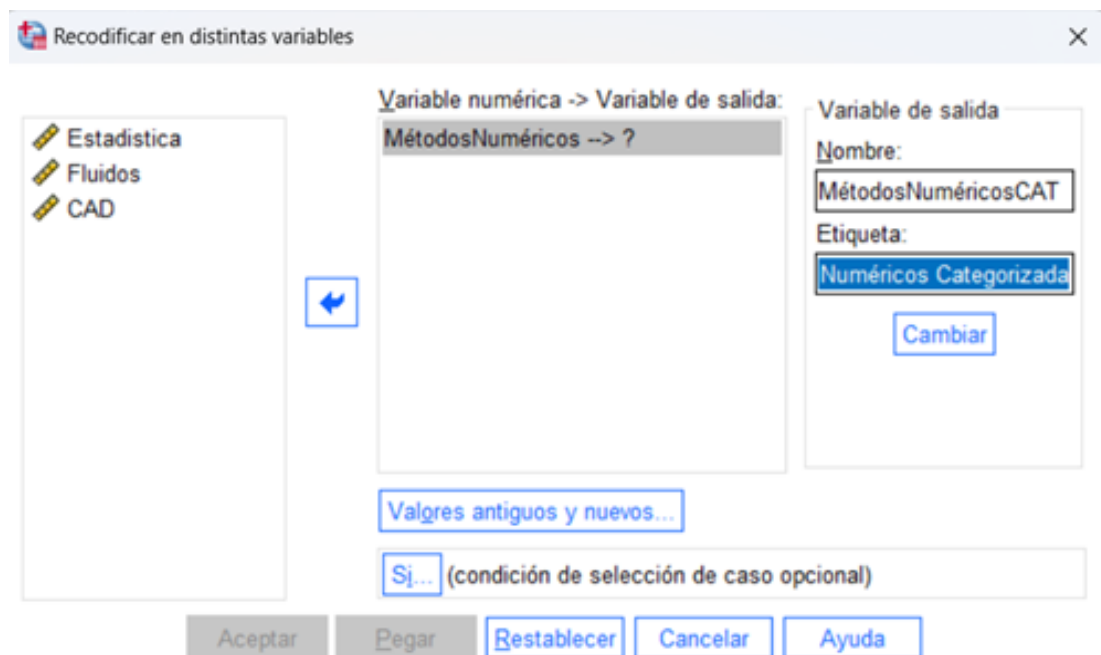
	Estadística	Métodos Numéricos	Fluidos	CAD	var	var	va
1	15.50	18.50	99999.00	99999.00			
2	18.89	16.00	17.25	20.00			
3	12.50	15.50	17.50	16.90			
4	15.45	99999.00	99999.00	99999.00			
5	19.99	15.20	18.30	20.00			
6	15.26	99999.00	99999.00	99999.00			
7	16.00	20.00	20.00	10.00			
8	17.68	13.44	17.60	17.13			
9	16.34	15.80	9.00	14.00			
10	17.53	13.56	16.10	16.00			

En el menú principal seleccionar: Transformar / Recodificar en distintas variables...

Seleccionar la variable a ser Recodificada (por ejemplo, Métodos Numéricos), dar un nuevo Nombre y Etiqueta a la variable de salida, para continuar haciendo clic en Valores antiguos y nuevos.

Figura 16

Recodificar en distintas variables



The screenshot shows the 'Recodificar en distintas variables' dialog box. On the left, a list of variables includes Estadística, Fluidos, and CAD. In the center, the 'Variable numérica -> Variable de salida:' section shows 'MétodosNuméricos --> ?'. On the right, the 'Variable de salida' section has 'Nombre:' set to 'MétodosNuméricosCAT' and 'Etiqueta:' set to 'Numéricos Categorizada'. Below these sections are buttons for 'Valores antiguos y nuevos...', 'Sj...' (condición de selección de caso opcional), 'Aceptar', 'Pegar', 'Restablecer', 'Cancelar', and 'Ayuda'. A 'Cambiar' button is also present next to the label field.

Recodificar la variable con los nuevos valores, en los rangos: Insuficiente (<14), Elemental (14.01-16), Excelente (16.01-18) y Satisfactorio (>18.01) y clic en continuar.

Figura 17

Recodificar en variables diferentes: valores antiguo y nuevo

Recodificar en variables diferentes: valores antiguo y nuevo

Valor antiguo

☐ Valor:

☐ Perdido del sistema

☐ Perdido por el sistema o el usuario

☐ Rango:

hasta

☐ Rango, LOWEST hasta el valor:

☒ Rango, valor hasta HIGHEST:

Todos los demás valores

Valor nuevo

☒ Valor:

☐ Perdido del sistema

☐ Copiar valores antiguos

Antiguo -> Nuevo:

Lowest thru 14 -> 1

14.01 thru 16 -> 2

16.01 thru 18 -> 3

18.01 thru Highest -> 4

Añadir

Cambiar

Eliminar

☐ Las variables de salida son series Anchura: 8

☐ Convertir series numéricas a números (5->5)

Continuar Cancelar Ayuda

Clic en Cambiar y luego en Aceptar.

Recodificar en distintas variables

Variable numérica -> Variable de salida:

MétodosNuméricos -> ?

Variable de salida

Nombre:

MétodosNuméricosCAT

Etiqueta:

Promedio Métodos Num

Cambiar

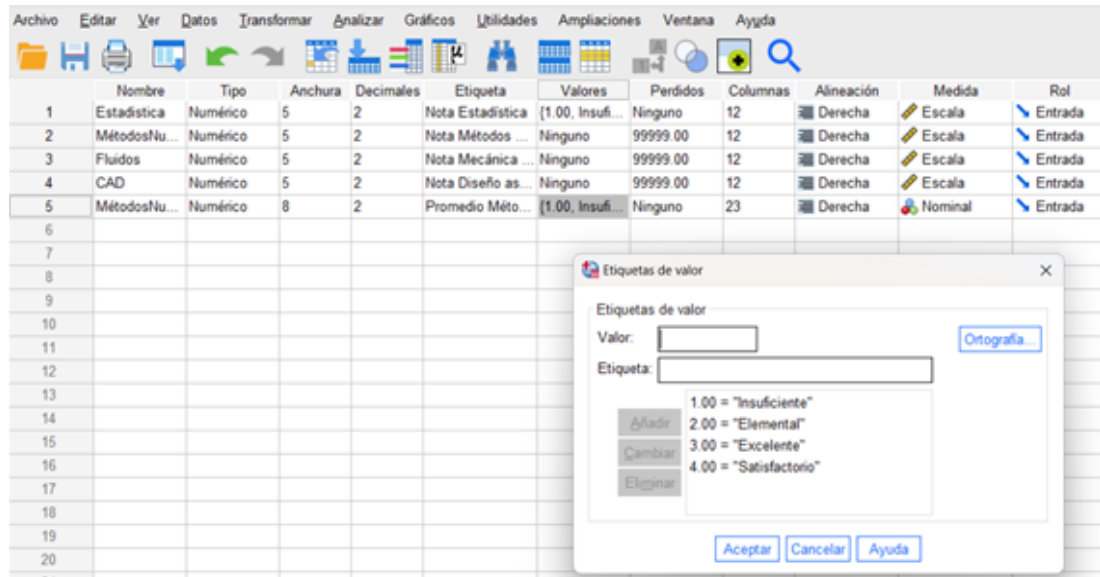
Valores antiguos y nuevos...

Sí... (condición de selección de caso opcional)

Aceptar Pegar Restablecer Cancelar Ayuda

- Seleccionamos la ventana vista de variables
- En la opción de Valores de la variable Métodos Numéricos CAT, Añadir los nuevos valores de las etiquetas: 1 Insuficiente, 2 Elemental, 3 Excelente, 4 Satisfactorio.

c. Clic en Aceptar



La nueva variable Métodos Numéricos CAT ha sido creada y recodificada de la variable Métodos Numéricos.

	Estadística	MétodosNuméricos	Fluidos	CAD	MétodosNuméricosCAT
1	Elemental	18.50	99999.00	99999.00	Satisfactorio
2	Satisfactorio	16.00	17.25	20.00	Elemental
3	Insuficiente	15.50	17.50	16.90	Elemental
4	Elemental	99999.00	99999.00	99999.00	Satisfactorio
5	Satisfactorio	15.20	18.30	20.00	Elemental
6	Elemental	99999.00	99999.00	99999.00	Satisfactorio
7	Elemental	20.00	20.00	10.00	Satisfactorio
8	Excelente	13.44	17.60	17.13	Insuficiente
9	Excelente	15.80	9.00	14.00	Elemental
10	Excelente	13.56	16.10	16.00	Insuficiente

1.10.3. Transformar / Recodificación automática

Esta opción convierte automáticamente valores de texto (strings) o números en códigos numéricos consecutivos, asignando etiquetas descriptivas. Se utiliza en variables categóricas no numéricas para simplificarlas.

Ejercicio práctico N° 1

En el siguiente link: <https://drive.google.com/file/d/1ZKiuE2sUtzdBpTt-5gpI5mfASsyupFYQh/view?usp=sharing>

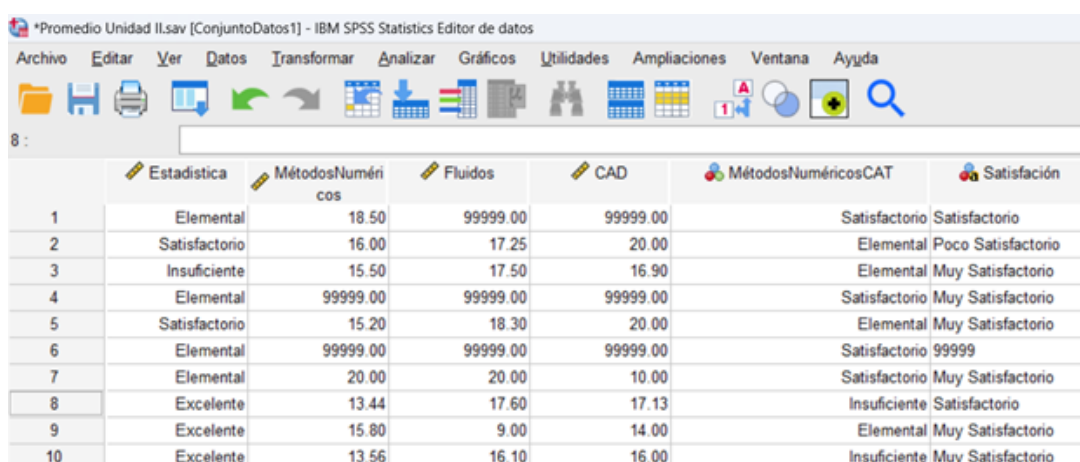
Se realiza una investigación sobre la satisfacción de los promedios de la segunda unidad, a los estudiantes de ingeniería, obteniéndose la siguiente base de datos.

Recodificar automática de la variable Satisfacción, sus categorías son: Poco Satisfactorio, Satisfactorio, Muy Satisfactorio.

Construcción en SPSS

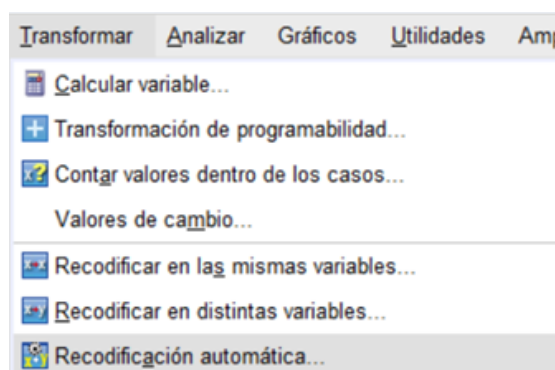
1. Prepare la base de datos

- Abra SPSS y crear la base de datos.
- Tratar los valores perdidos (99999)



	Estadística	MétodosNuméricos	Fluidos	CAD	MétodosNuméricosCAT	Satisfacción
1	Elemental	18.50	99999.00	99999.00	Satisfactorio	Satisfactorio
2	Satisfactorio	16.00	17.25	20.00	Elemental	Poco Satisfactorio
3	Insuficiente	15.50	17.50	16.90	Elemental	Muy Satisfactorio
4	Elemental	99999.00	99999.00	99999.00	Satisfactorio	Muy Satisfactorio
5	Satisfactorio	15.20	18.30	20.00	Elemental	Muy Satisfactorio
6	Elemental	99999.00	99999.00	99999.00	Satisfactorio	99999
7	Elemental	20.00	20.00	10.00	Satisfactorio	Muy Satisfactorio
8	Excelente	13.44	17.60	17.13	Insuficiente	Satisfactorio
9	Excelente	15.80	9.00	14.00	Elemental	Muy Satisfactorio
10	Excelente	13.56	16.10	16.00	Insuficiente	Muy Satisfactorio

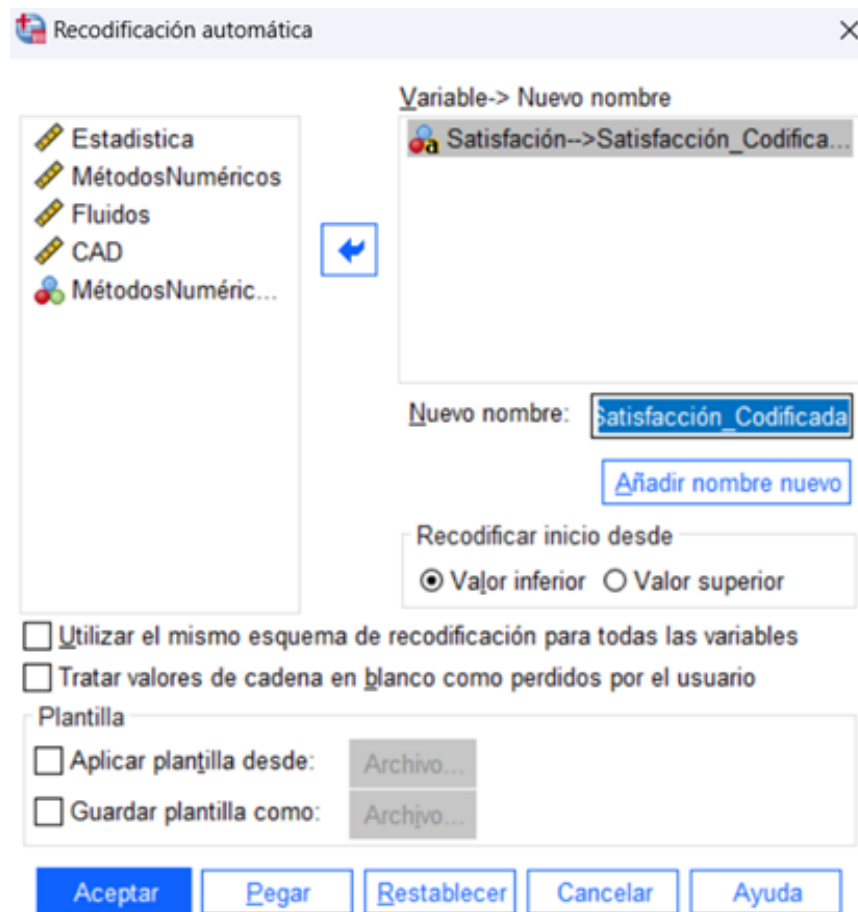
En el menú principal seleccionar: Transformar / Recodificación automática



Seleccionar la variable a ser Recodificada (Satisfacción), dar un Nuevo nombre, para continuar haciendo clic en Añadir nombre nuevo, finalmente dar clic en Aceptar

Figura 18

Recodificación automática



Los resultados indican que se ha codificado en los niveles: 99999 (1), Muy Satisfactorio (2), Poco Satisfactorio (3) y Satisfactorio.

	Estadística	MétodosNuméricos	Fluidos	CAD	MétodosNuméricosCAT	Satisfacción	Satisfacción_Codificada
1	2.00	18.50	99999.00	99999.00		4.00 Satisfactorio	4
2	4.00	16.00	17.25	20.00		2.00 Poco Satisfactorio	3
3	1.00	15.50	17.50	16.90		2.00 Muy Satisfactorio	2
4	2.00	99999.00	99999.00	99999.00		4.00 Muy Satisfactorio	2
5	4.00	15.20	18.30	20.00		2.00 Muy Satisfactorio	2
6	2.00	99999.00	99999.00	99999.00		4.00 99999	1
7	2.00	20.00	20.00	10.00		4.00 Muy Satisfactorio	2
8	3.00	13.44	17.60	17.13		1.00 Satisfactorio	4
9	3.00	15.80	9.00	14.00		2.00 Muy Satisfactorio	2

Ejemplo práctico N° 2: (con base de datos)

En el siguiente link:

Accede aquí:



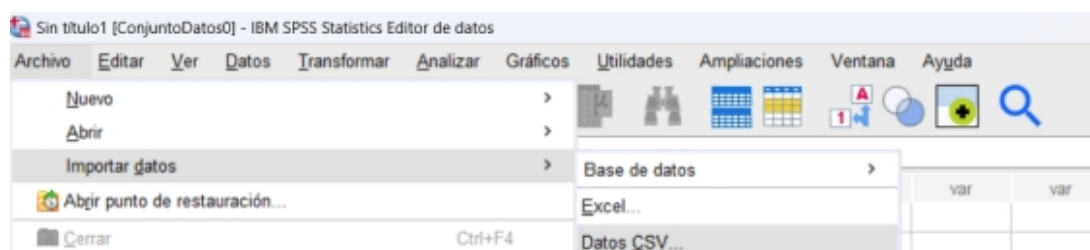
Se encuentra el conjunto de datos que contiene información sobre la estimación de los niveles de obesidad en individuos de los países de México, Perú y Colombia, con base en sus hábitos alimenticios y condición física. Usando el archivo descargado con el nombre **ObesityDataSet_raw_and_data_synthetic.csv**. Aplicar los procedimientos básicos de SPSS aprendidos en el Capítulo I para cargar, explorar, transformar y filtrar datos reales

1. Descarga el archivo “**ObesityDataSet_raw_and_data_synthetic.csv**” desde el enlace indicado anteriormente.

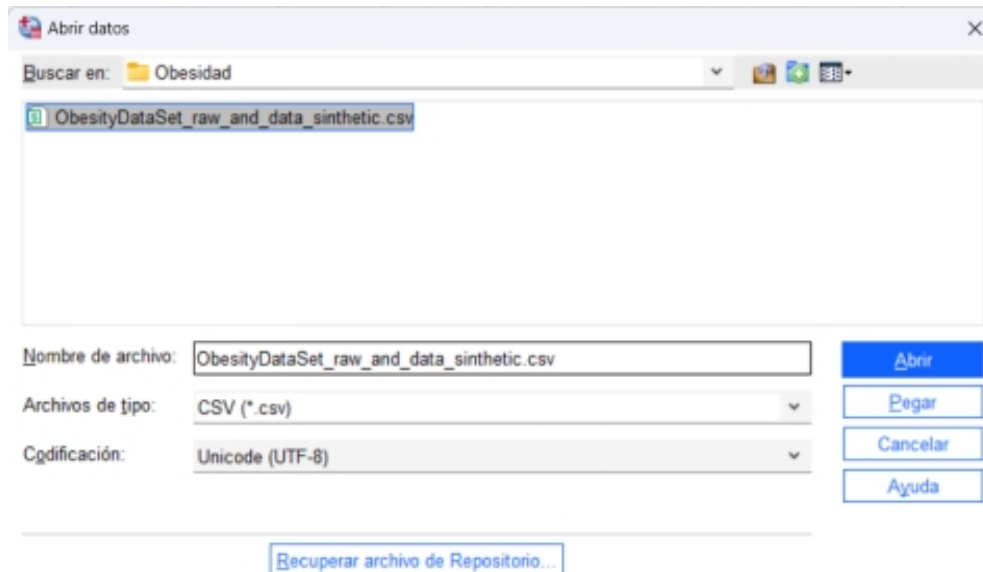
Estimación de los niveles de obesidad en función de los hábitos alimentarios y la condición física		
Donado el 26/08/2019		
Este conjunto de datos incluye datos para la estimación de los niveles de obesidad en individuos de los países de México, Perú y Colombia, con base en sus hábitos alimenticios y condición física.		
Características del conjunto de datos	Área temática	Tareas asociadas
Multivariado	Salud y Medicina	Clasificación, regresión, agrupación en clústeres
Tipo de entidad	# Instancias	# Características
Entero	2111	16

[DESCARGAR \(56.3 KB\)](#)
[IMPORTAR EN PYTHON](#)
[CITAR](#)
 1 citas
 177351 vistas
 DOI: 10.24432/CSH01Z

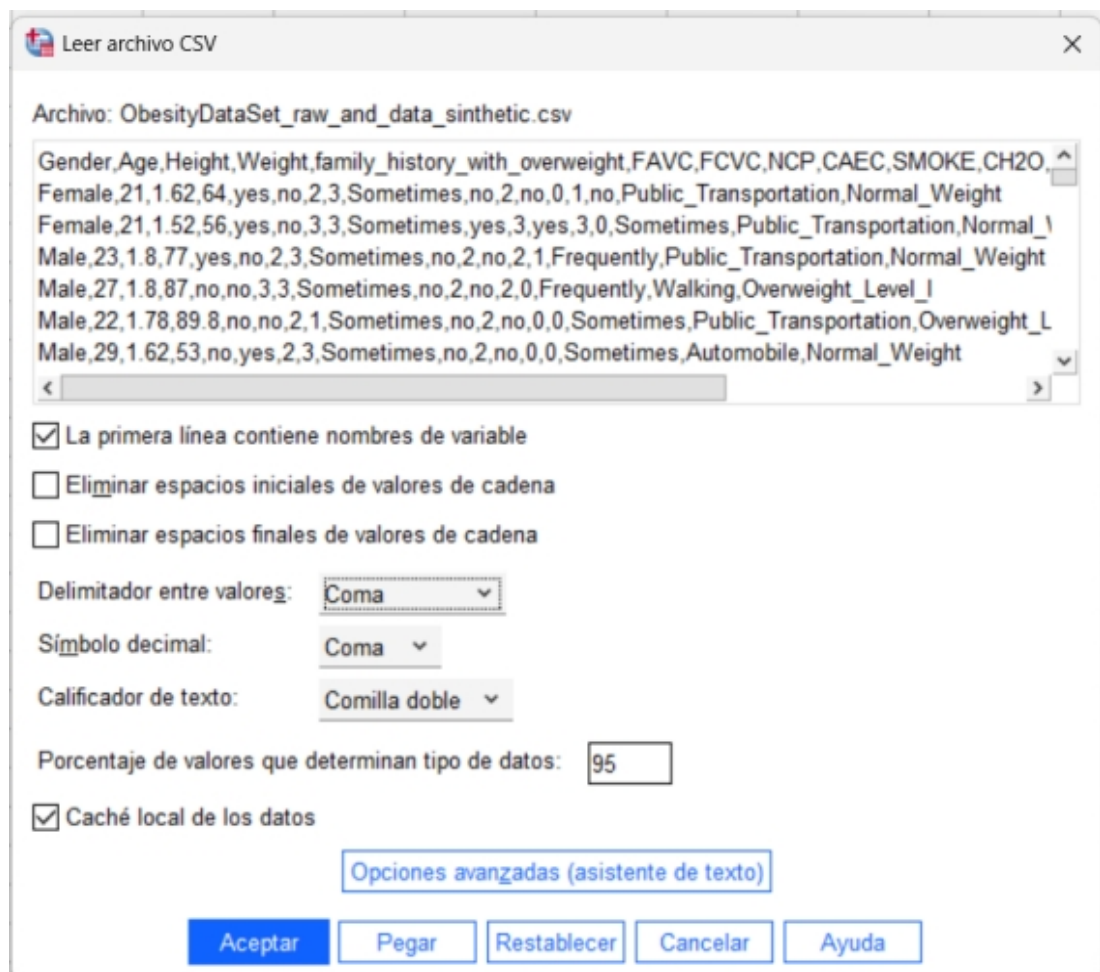
2. Abre el archivo descargado **ObesityDataSet_raw_and_data_synthetic.csv** en SPSS usando la opción de importar datos desde un archivo CSV.



Seleccionar el archivo descargado en el anterior paso **ObesityDataSet_raw_and_data_synthetic.csv** y clic en Abrir



En la opción Delimitador entre variables seleccionar Coma y clic en Aceptar



3. Guarda el archivo en formato nativo de SPSS (`.sav`) con el nombre: `Obesidad.sav`.



4. En la pestaña Vista de Variables etiqueta las variables de acuerdo con la Tabla de variables y clasifícalas según su tipo (Nominal, Ordinal, Escala)

Obesidad.sav [CorporalData01] - IBM SPSS Statistics Editor de datos										
Archivos Editar Ver Datos Transformar Analizar Gráficos Utilidades Aplicaciones Ventana Ayuda										
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Pérdidos	Columnas	Asociación	Medida
1	Gender	Cadena	6	0	¿Cuál es tu género?	(1, Female)	Ninguno	6	Izquierda	Nominal
2	Age	Numérico	9	0	¿Cuántos años tienes?	Ninguno	Ninguno	11	Derecha	Escala
3	Height	Numérico	9	2	¿Cuál es tu altura en metros?	Ninguno	Ninguno	8	Derecha	Escala
4	Weight	Numérico	10	0	¿Cuál es tu peso en kg?	Ninguno	Ninguno	12	Derecha	Escala
5	family_hist...	Cadena	3	0	¿Algun miembro de la familia ha sufrido o sufre de sobrepeso?	(1, yes)	Ninguno	3	Izquierda	Nominal
6	PAVC	Cadena	3	0	¿Comes alimentos ricos en calorías con frecuencia?	(1, yes)	Ninguno	3	Izquierda	Nominal
7	PCVC	Cadena	8	0	¿Sueles comer verduras en tus comidas?	Ninguno	Ninguno	8	Izquierda	Nominal
8	HCPC	Numérico	8	1	¿Cuántas comidas principales tienes al día?	Ninguno	Ninguno	8	Derecha	Escala
9	CAEC	Cadena	10	0	¿Comes algún alimento entre comidas?	(1, no)	Ninguno	10	Izquierda	Nominal
10	SMOKE	Cadena	3	0	¿Usas fumar?	(1, yes)	Ninguno	3	Izquierda	Nominal
11	CHZO	Numérico	8	0	¿Cuánta agua bebes diariamente?	Ninguno	Ninguno	8	Derecha	Escala
12	SDC	Cadena	3	0	¿Controlas las calorías que comes diariamente?	Ninguno	Ninguno	3	Izquierda	Nominal
13	FAF	Numérico	8	0	¿Con qué frecuencia realizas actividad física?	Ninguno	Ninguno	8	Derecha	Escala
14	TUE	Numérico	8	0	¿Cuánto tiempo usas dispositivos tecnológicos como celular, videojuegos, televisión, computadora y otros?	Ninguno	Ninguno	8	Derecha	Escala
15	CALC	Cadena	10	0	¿Con qué frecuencia bebes alcohol?	Ninguno	Ninguno	10	Izquierda	Nominal
16	MTRANS	Cadena	21	0	¿Qué transporte sueles utilizar?	Ninguno	Ninguno	21	Izquierda	Nominal
17	NOObesidad	Cadena	15	0	Nivel de obesidad	Ninguno	Ninguno	15	Izquierda	Nominal

5. Selecciona las observaciones que tienen una edad de mayor o igual a 30 años (Age)

La ruta para seguir es: Datos/Seleccionar casos.../Si se satisface la condición



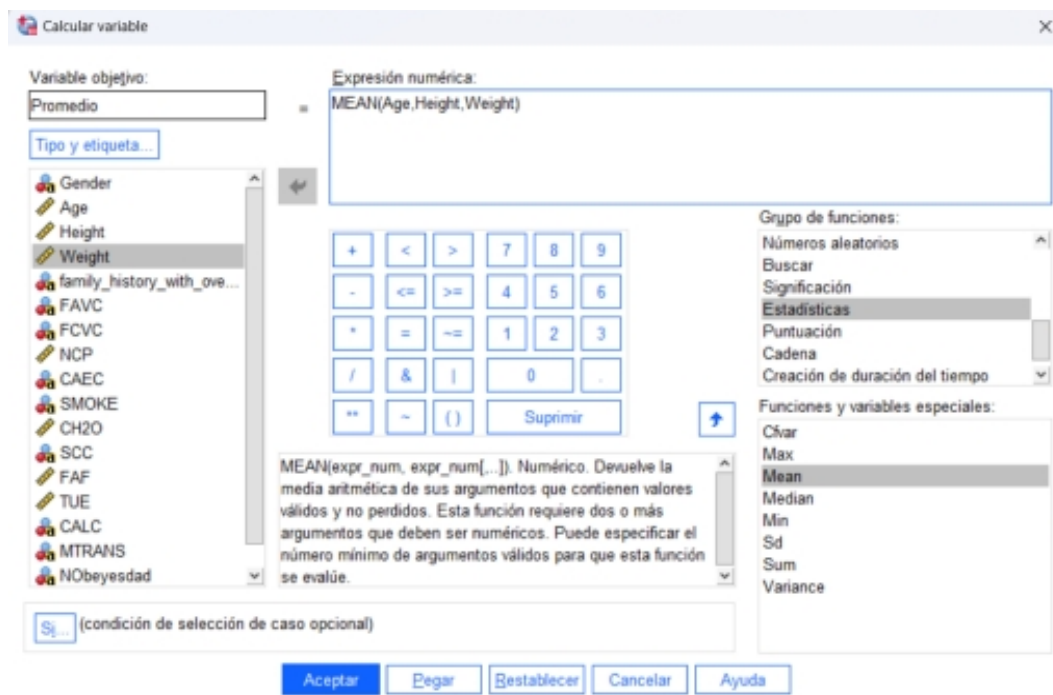
Se obtiene los siguientes resultados

	Gender	Age	Height	Weight	Family History	FAVC	FCVC	NCP	CAEC	SMOKE	CH2O	SCC	FAF	TUE	CALC	MTRANS
163	Female	21	1.63	60	yes	yes	3.000000	3.000	Always	yes	2	no	2	0	Sometimes	Public_Transportation
164	Male	25	1.89	75	no	no	3.000000	3.000	Frequently	no	2	yes	3	0	no	Public_Transportation
165	Male	35	1.77	95	yes	no	3.000000	3.000	Frequently	no	3	no	2	1	Sometimes	Automobile
166	Male	30	1.92	130	yes	no	2.000000	3.000	Sometimes	yes	1	no	1	0	Sometimes	Public_Transportation
167	Female	29	1.74	72	yes	no	3.000000	3.000	Always	no	2	no	0	0	Sometimes	Automobile
168	Male	20	1.65	80	yes	no	2.000000	3.000	Sometimes	no	2	no	1	2	no	Walking
169	Female	22	1.73	79	yes	yes	2.000000	1.000	Sometimes	no	2	no	1	0	Sometimes	Public_Transportation
170	Female	45	1.63	77	yes	yes	2.000000	3.000	Frequently	no	1	no	0	0	no	Automobile
171	Male	22	1.72	82	no	yes	2.000000	1.000	Sometimes	no	2	no	2	1	Sometimes	Public_Transportation
172	Male	18	1.68	58	yes	yes	2.000000	3.000	Frequently	no	2	no	1	0	Sometimes	Public_Transportation
173	Female	23	1.65	59	no	no	3.000000	1.000	Sometimes	no	2	no	1	2	Sometimes	Public_Transportation
174	Male	18	1.74	64	no	no	3.000000	3.000	Sometimes	no	2	yes	0	1	no	Public_Transportation

6. Ve a Transformar → Calcular variable

a) Nombra la variable de destino: `Promedio`.

b) Usa la función `MEAN` para calcular el promedio de `Age`, `Height` y `Weight`.



c) Ejecuta el comando y verifica que la nueva variable aparezca en la Vista de Datos.

*Obesidad.sav [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

	CALC	MTRANS	NObeyesdad	filter_\$	Promedio
163	Sometimes	Public_Transportation	Normal_Weight	0	27.54
164	no	Public_Transportation	Normal_Weight	0	33.96
165	Sometimes	Automobile	Overweight_Level_II	1	40.59
166	Sometimes	Public_Transportation	Obesity_Type_II	1	53.97
167	Sometimes	Automobile	Normal_Weight	0	34.25
168	no	Walking	Overweight_Level_II	0	33.88
169	Sometimes	Public_Transportation	Overweight_Level_I	0	34.24
170	no	Automobile	Overweight_Level_II	1	41.21
171	Sometimes	Public_Transportation	Overweight_Level_II	0	35.24
172	Sometimes	Public_Transportation	Normal_Weight	0	25.87

7. Con la base de datos original. Recodificar en distintas variables

Recodificar la variable `Weight` (peso en kg) en una nueva variable categórica llamada `Categoría peso`, con las siguientes categorías:

Valor	Categoría de peso	Rango de peso (kg)
1	Peso insuficiente	< 53.5 kg
2	Peso normal	53.5 - 71.9 kg
3	Sobrepeso	72.0 - 86.4 kg
4	Obesidad Grado I	86.5 - 100.9 kg
5	Obesidad Grado II	101.0 - 115.4 kg
6	Obesidad Grado III	≥ 115.5 kg

La ruta por seguir es: Transformar/Recodificar en distintas variables.../
Valores antiguos y nuevos

Continuar/ Aceptar

El resultado es:

Obesidad.sav [ConjuntoDatos1] - IBM SPSS Statistics Editor de datos

Archivo

Editar

Ver

Datos

Transformar

Analizar

Gráficos

Utilidades

Ampliaciones

Ventana

Ayuda

</

En la ventana Vista de variables asignar los valores a esta categoría.

ción de toda la base de datos relacionados con el rendimiento académico de estudiantes, Lea el artículo (<https://doi.org/10.1016/j.dib.2019.104344>) para ver la descripción de los atributos.

1. Descarga el archivo `student-mat.csv` desde el enlace de la base de datos.
2. Ábrelo en SPSS usando la opción de importar datos desde un archivo CSV.
3. Guarda el archivo en formato nativo de SPSS (`.sav`) con el nombre: `RendimientoEstudiantil.sav`.
4. Identifica en la Vista de Variables al menos 5 variables y clasifícalas según su tipo (Nominal, Ordinal, Escala).
5. Seleccionar a los estudiantes que tienen un tiempo de estudio semanal (`studytime`) mayor a 2 horas.
6. Ve a Transformar → Calcular variable
 - Nombra la variable de destino: `PromedioFinal`.
 - Usa la función `MEAN` para calcular el promedio de `G1`, `G2` y `G3`.
 - Ejecuta el comando y verifica que la nueva variable aparezca en la Vista de Datos.

7. Literal D: Recodificación de Variables

Recodificar la variable `absences` (inasistencias) en una nueva variable categórica llamada `NivelAusencia`, con las siguientes categorías:

- 1 = Baja: 0 a 2 inasistencias
- 2 = Media: 3 a 5 inasistencias
- 3 = Alta: 6 o más inasistencias

8. Guarda el archivo SPSS con todos los cambios realizados.



<https://n9.cl/ss5s0d>

CAPÍTULO II

Introducción a la Estadística

2.1. Definición de Estadística

La estadística es la ciencia que tiene como objetivo reunir información, organizar, presentar, analizar e interpretar los resultados con el fin de ayudar a tomar decisiones, así como también formular predicciones.

2.2. Tipos de Estadística

La estadística se divide en dos grandes categorías: La estadística descriptiva y la estadística inferencial

2.2.1. Estadística Descriptiva

Describe, analiza y representa un grupo de datos, en forma resumida, utilizando tablas de frecuencia, gráficos estadísticos, así como medidas de tendencia central, posición, dispersión y forma.

2.2.2. Estadística Inferencial

Basado en el cálculo de probabilidades con los resultados de una muestra, efectúa estimaciones, predicciones o inferencias sobre una población, las técnicas que se utilizan son: Regresiones, intervalos de confianza, pruebas de hipótesis entre otras.

Ejemplos:

Investiga un ejemplo aplicado a tus estudios donde se usa la estadística.

.....

.....

.....

En los siguientes ejemplos identifica si se aplica la estadística descriptiva o inferencial

- Informe que presenta gráficas de barras y tablas de frecuencia de las ventas mensuales realizadas en el mes de diciembre.
- Análisis de los tiempos de vida útil promedio de un motor en una muestra de 120 vehículos.

- Representación gráfica de los ciclos de funcionamiento en un brazo robótico, mostrando su eficiencia en diferentes fases del proceso.
- Evaluar las propiedades físicas y químicas de diferentes muestras de petróleo, tomadas de distintos pozos petroleros del Oriente Ecuatoriano, aplicando medidas de tendencia central
- Predecir el rendimiento de un nuevo proceso de refinado basado en datos experimentales de una muestra con una prueba piloto
- Predicción del rendimiento de un nuevo software de acuerdo a la velocidad y eficiencia
- Gráfico de las variaciones de voltaje en un circuito durante un experimento.
- Determinación de la probabilidad de falla de un lote de microcontroladores, de acuerdo al análisis de defectos de una muestra de 100 elementos.
- Tablas de frecuencia que indica el número de turistas que visitan el volcán Cotopaxi en el mes de junio.
- Estimación de las ventas futuras para el próximo año, basándose en las facturas de los meses de marzo, mayo, junio, agosto y diciembre.

2.3. Población

Es el conjunto universo en estudio, de las cuales se desea obtener la información para describir y conseguir conclusiones.

2.4. Muestra

Es un subconjunto seleccionado de la población que interesa ser analizada, con el objetivo de tener información sobre la población en estudio.

Figura 19

Relación entre muestra y población

**Ejemplos:**

Cite cinco ejemplos de población y cinco de muestra aplicados a sus estudios.

Ejemplos de población:

.....

.....

.....

.....

.....

Ejemplos de muestra:

.....

.....

.....

.....

.....

2.5. Parámetro

Es un valor numérico que representa información sobre la población, entre ellas tenemos la media poblacional (μ), desviación estándar poblacional (σ), proporción poblacional (p), etc.

2.6. Estadístico

Es un valor numérico calculado a partir de la muestra, se utiliza para inferir sobre la población, entre ellas tenemos la media muestral ($\bar{x}$), la desviación estándar muestral (s), la proporción muestral ($\hat{p}$), etc.

Autoevaluación 2.1.

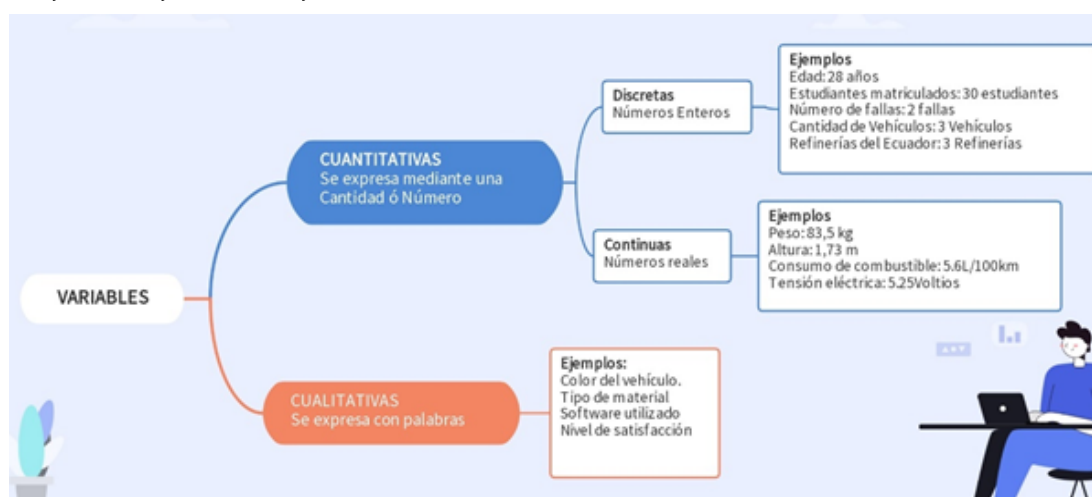
1. Defina los siguientes conceptos y dar un ejemplo de cada uno:
 - a. Estadística descriptiva
 - b. Estadística inferencial
2. Defina los siguientes conceptos y dar un ejemplo de cada uno:
 - a. Población
 - b. Muestra
- 3.Cuál es la diferencia entre un parámetro y un estadístico

2.7. Variables

Variable es cualquier característica, número o cantidad que puede ser medida o cuantificada que puede tomar diferentes valores en diferentes individuos, objetos o fenómenos bajo estudio, se pueden clasificar en variables de tipos cualitativas o variables de tipo cuantitativas.

Figura 20

Mapa conceptual del tipo de variables



2.7.1. Variables cualitativas

Son características o cualidades que representan diferencias entre sujetos y se clasifican en categorías. Estas variables reflejan atributos que pueden variar de un individuo a otro, enfocándose en aspectos no numéricos que describen la naturaleza o el estado de un fenómeno.

2.7.2. Variables cuantitativas

Son características que pueden medirse y expresan cantidades, variando de sujeto a sujeto y representándose mediante valores numéricos que indican magnitudes específicas.

Las variables cuantitativas se subdividen en variables discretas y continuas:

Variables discretas

Son aquellas que pueden tomar únicamente valores enteros y son contables, lo que significa que sus posibles valores se pueden enumerar o listar de manera finita o infinita.

Ejemplos:

- Número de estudiantes matriculados
- Cantidad de productos fabricados
- Número de fallas
- Cantidad de vehículos producidos
- Edad de los estudiantes
- Cantidad de usuarios registrados en una aplicación
- Cantidad de muestras analizadas en un laboratorio
- Número de transacciones realizadas en el día
- Cantidad de reservaciones realizadas en un hotel

Variables continuas

Son aquellas que pueden tomar cualquier valor dentro de un rango determinado, incluyendo valores fraccionarios. Esto significa que pueden representar cantidades que no se limitan a números enteros, permitiendo así una representación más precisa de fenómenos que varían de manera continua.

Ejemplos:

- Estatura de una persona (1,50 metros)
- Consumo de combustible (5.6 L/100 Km)
- Corriente eléctrica (2.75 A)
- Tiempo de respuesta de una aplicación (150.3 ms)
- Tamaño de un archivo (23.5 MB)
- Frecuencia de operación de un motor (50.2 Hz)
- Temperatura de una reacción (75.8 °C)
- Monto Total ingresado (USD1,250.75)
- Tasa de interés (9.25%)
- Gasto promedio por turista en la ciudad de Baños (USD150.30)

Ejercicios:

Identificar cuáles son variables cualitativas y cuales son cuantitativas (discretas o continuas).

- Color de ojos de las personas _____
- Número de libros leídos durante el año 2024 _____
- Altura de las personas _____
- Edad de un grupo de estudiantes universitarios _____
- Tipo de material _____
- Estado del equipo _____
- Nivel de satisfacción del cliente _____
- Tipo de software utilizado _____
- Condición del terreno _____
- Tipo de energía utilizada _____

2.8. Niveles de medición

Figura 21

Niveles de medición



2.8.1. Nivel de tipo nominal

Los datos nominales son un tipo de datos que clasifican observaciones en categorías distintas sin un orden específico. Estos datos se limitan a la identificación o conteo de las categorías, ya que no existe una jerarquía o relación de orden natural entre ellas.

Ejemplos:

- Tipo de combustible que utiliza un vehículo: gasolina, diésel.
- Tipo de componente que utiliza un dispositivo electrónico: resistor, capacitor, transistor, diodo.
- Tipo de lenguaje de programación que utiliza un programador de sistemas: Python, Java, C++, JavaScript.
- Sensores utilizados en un Edificio: sensor de temperatura, sensor de presión, sensor de proximidad.

- Productos químicos utilizados en una Industria Petroquímica: etileno, propileno, butileno.
- Tipo de informe financiero: Balance general, estado de resultados, flujo de efectivo.
- Destino turístico del Ecuador: Islas Galápagos, Baños, Montañita, Cuenca.

2.8.2. Nivel de tipo ordinal

Los datos ordinales son un tipo de datos que se caracterizan por su capacidad de ser clasificados en categorías que poseen un orden jerárquico. Estos datos implican un ordenamiento que se expresa en términos de desigualdades, permitiendo así establecer una secuencia natural entre las categorías.

Ejemplos:

- Clasificación de vehículos según el tipo de motor: 1. Combustión interna, 2. Eléctricos, 3. Híbridos.
- Calidad de señal: 1. Muy baja, 2. Baja, 3. Media, 4. Alta y 5. Muy alta.
- Grado de satisfacción del usuario: 1. Muy insatisfecho, 2. Insatisfecho, 3. Neutral, 4. Satisfecho, 5. Muy satisfecho.
- Nivel de automatización: 1. Manual, 2. Semi-automático, 3. Totalmente automático.
- Calidad del producto: 1. Baja calidad, 2. Calidad media y 3. Alta calidad.
- Grado de cumplimiento normativo: 1. Cumple totalmente, 2. Cumple parcialmente o 3. No cumple.
- Categoría del servicio hotelero: 1. Económico, 2. Turista, 3. De lujo y 4. Premium.

2.8.3. Nivel de tipo intervalo

Los datos de intervalo son una categoría de medición cuantitativa que permite no solo el orden de los valores, sino también la medición precisa de las diferencias entre ellos. Las distancias entre los puntos son significativas y cons-

tantes, carecen de un punto cero absoluto (es decir, el cero no implica ausencia total de la variable medida sino un valor arbitrario o el origen para dar inicio a la medida).

Ejemplo

- Temperatura del motor: Medida en grados Celsius o Fahrenheit.
- Frecuencia de una señal: Medida en hertz (Hz)
- Tiempos de respuesta de una aplicación: Medidos en milisegundos (ms)
- Nivel de vibración de un sistema: Medido en decibelios (dB)
- Temperatura de ebullición de un líquido: Medida en grados Celsius o Fahrenheit
- Fechas contables: Las fechas pueden considerarse variables de intervalo donde se mide el tiempo transcurrido entre eventos financieros
- Calificaciones en encuestas: Las puntuaciones obtenidas en encuestas sobre la satisfacción del cliente

2.8.4. Nivel de tipo razón

Los datos de razón son un tipo de variable cuantitativa que combina las características de las escalas nominal, ordinal y de intervalo, se distinguen por poseer un punto de cero absoluto. (Este cero indica la ausencia total de la medida, lo que permite realizar operaciones matemáticas como suma, resta, multiplicación y división). A diferencia de los datos de intervalo, el cero es arbitrario y pueden tener valores negativos. En la escala de razón no existen valores negativos y las diferencias entre los valores son significativas.

Ejemplos

- Velocidad del vehículo: Medida en kilómetros por hora (km/h)
- Tensión eléctrica: Medida en voltios (V)
- Número de errores por línea de código: Cantidad de errores detectados en relación con el total de líneas de código escritas.
- Torque aplicado: Medido en newton-metros (Nm)
- Producción diaria de petróleo: Medida en barriles por día (bbl/día)
- Ingresos totales: Expresados en unidades monetarias

- Número de visitantes anuales: Cantidad total de turistas que visitan un destino en un año.

Ejercicios

Seleccione la respuesta correcta, identificando el nivel de medición en cada pregunta:

1. *Clasificar las fallas mecánicas en un proceso de manufactura según su gravedad: Leve, Moderada, Crítica, es una variable de tipo:*

- Nominal
- Ordinal
- Intervalo
- Razón

2. *Número de ciclos de carga que soporta un material antes de la fractura, es una variable de tipo:*

- Nominal
- Ordinal
- Intervalo
- Razón

3. *Tipo de aleación usada en la fabricación de turbinas: Inconel, Titanio, Acero inoxidable, es una variable de tipo.*

- Nominal
- Ordinal
- Intervalo
- Razón

4. *Temperatura (°F) en un reactor químico durante un proceso, es una variable de tipo:*

- Nominal
- Ordinal
- Intervalo
- Razón

5. *Velocidad angular (en rad/s) de un motor eléctrico, es una variable de tipo:*
- Nominal
 - Ordinal
 - Intervalo
 - Razón
6. *Satisfacción de usuarios del servicio de una entidad financiera (Escala del 1 al 5: siendo 1 = Muy insatisfecho, 5 = Muy satisfecho), es una variable de tipo:*
- Nominal
 - Ordinal
 - Intervalo
 - Razón

Autoevaluación 2.2.

1. Con sus propias palabras defina variable cualitativa y de dos ejemplos
2. Con sus propias palabras defina variable cuantitativa y de dos ejemplos
3. Con sus propias palabras indique la diferencia entre variables discretas y continuas

2.9. Procedimiento para realizar un análisis estadístico

Las etapas para realizar un análisis estadístico son las siguientes:

1. Planificar el instrumento con los objetivos de la investigación: ¿Qué se desea estudiar?
2. Recopilar la Información: Seleccionar el tamaño de la muestra y el método de recolección de los datos, esta puede ser encuestas, observación, experimentación, etc.
3. Organizar la Información (SPSS): Ingresar los datos en un software estadístico.

4. Presentar de resultados
5. Analizar e interpretar los resultados

2.9.1. Planificación

La etapa de planificación tiene como propósito estudiar los detalles relacionados con la recolección, organización, presentación y análisis de la información. La importancia de esta etapa radica en las características de una determinada población o en la posibilidad de comprobar o comprender la veracidad de las hipótesis de trabajo, pero es importante señalar que, antes de comenzar la planificación, debe haber una definición clara del objetivo de la investigación.

Además, debe llevarse a cabo un análisis oportuno de la información disponible sobre el tema y de las explicaciones ya existentes. Sin estos pasos preliminares no sería posible realizar de manera eficiente la planificación.

Los pasos para considerar una adecuada planificación se resumen en:

- Planteamiento del problema.
- Búsqueda y Evaluación de la información existente.
- Formulación de hipótesis.
- Verificación de las hipótesis.
- Conclusiones y recomendaciones.

El análisis de estos aspectos no es precisamente un asunto estadístico, pero se describirán más adelante, ya que su comprensión y aplicación sistemática representan un método óptimo de trabajo para cualquier investigador.

Planteamiento del problema

Al formular el problema que se investigará, es importante prestar atención a los siguientes aspectos:

- Determinación de la importancia y relevancia del problema de investigación.
- Determinación del objetivo general y de los objetivos específicos de la investigación.

Por ejemplo, el investigador está interesado en conocer la relación entre el número de horas de estudio y el rendimiento académico en la asignatura de Estadística. El planteamiento del problema formulado sería: ¿Existe una relación significativa entre las horas de estudio y el rendimiento académico en la asignatura de Estadística?

La correcta formulación del problema es crucial, ya que guiará la búsqueda de información, la formulación de hipótesis y el análisis posterior.

Búsqueda y evaluación de la información existente

Antes de iniciar un estudio, es importante que el investigador realice una revisión sistemática de la información existente, esto le permitirá conocer el estado actual del conocimiento y entender lo que ya se ha descubierto. Familiarizarse con las técnicas de estudio más adecuadas es fundamental para el éxito del proyecto, al hacerlo, el investigador podrá identificar los métodos más efectivos para abordar su investigación.

Existen varios buscadores bibliográficos que son utilizados para realizar investigaciones académicas y científicas. Estos buscadores permiten acceder a una amplia variedad de fuentes confiables, como artículos científicos, libros, tesis y trabajos de investigación. A continuación, se describen algunos de los buscadores más utilizados en el ámbito académico: Google Scholar (Google Académico), PubMed, Scopus, Web of Science, Dialnet, ERIC (Education Resources Information Center), ScienceDirect, Redalyc, etc.

Formulación de la hipótesis

En toda investigación, el objetivo es probar una hipótesis de trabajo. Una hipótesis se define como una explicación provisional de los hechos, esta explicación se propone con el fin de verificar su veracidad a través del proceso de investigación.

En un estudio de estadística descriptiva, una hipótesis puede ser, por ejemplo: Los alumnos que estudian más de tres horas tienden a tener un mayor rendimiento académico. Esta afirmación será sometida a pruebas estadísticas utilizando los datos recopilados.

El software estadístico SPSS será la herramienta que nos permita analizar los datos y comprobar la veracidad de las hipótesis formuladas. En este sentido, la formulación correcta de las hipótesis es clave, ya que determinará el tipo de análisis estadístico que se debe realizar (correlación, regresión, prueba t de Student, etc.).

Verificación de la hipótesis

La verificación de la hipótesis constituye la investigación propiamente dicha, para esto se tienen en cuenta dos aspectos importantes como son el diseño de la investigación y su ejecución.

En esta sección, el análisis estadístico toma un rol central, utilizando herramientas como SPSS, se pueden aplicar diversas técnicas para analizar los datos de manera precisa.

El uso de SPSS permite realizar análisis descriptivos (media, moda, mediana, desviación estándar, entre otros) que resumen las características de los datos. Por otra parte, se pueden realizar pruebas de hipótesis, como las pruebas de correlación o de regresión, para evaluar si los datos respaldan o no las hipótesis planteadas.

Por ejemplo, si se quiere probar la relación entre las horas de estudio y rendimiento académico, se podría realizar un análisis de correlación en SPSS. Dependiendo del resultado del p-valor, se podrá aceptar o rechazar la hipótesis con base en el nivel de significancia estadística α .

Conclusiones y recomendaciones

Las conclusiones deben basarse directamente en los resultados obtenidos durante el análisis y deben responder de manera clara y precisa a las preguntas formuladas en el planteamiento del problema. Además, deben abordar las hipótesis y señalar si fueron aceptadas o rechazadas.

Las recomendaciones deben estar orientadas a la acción. Pueden incluir sugerencias para mejorar procesos, tomar decisiones informadas o proponer nuevas investigaciones que profundicen en áreas no cubiertas en el estudio actual.

Por ejemplo, si se encuentra que existe una correlación significativa entre las horas de estudio y el rendimiento académico, una recomendación podría ser fomentar hábitos de estudio más extensos y organizados entre los estudiantes.

2.9.2. Recopilación de la información

La recopilación de la información es el proceso en donde se obtienen los datos para responder las preguntas de investigación planteadas y probar la hipótesis en base a los objetivos de estudio. Estos datos pueden ser primarios (encuestas o experimentos) o secundarios (bases de datos, datos existentes).

Los métodos para recopilar los datos son técnicas cuantitativas como las encuestas, cuestionarios, mientras que las cualitativas son las entrevistas y observaciones, pero también pueden ser mixtas, todo depende de los objetivos de la investigación y su correcta aplicación de estos instrumentos garantizará la validez y confiabilidad de los resultados.

El procedimiento para recolectar la información es:

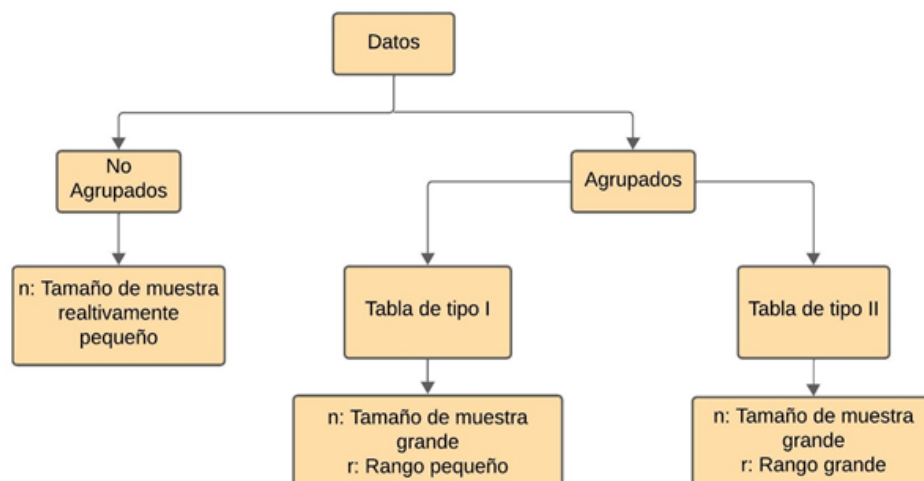
- Definir los objetivos de la investigación.
- Seleccionar el método adecuado
- Diseñar el instrumento y validarlo con expertos. (Encuestas, Cuestionario, Entrevista)
- Cálculo y selección de la muestra
- Recolectar datos
- Validarlo y almacenarlos

2.9.3. Organización de la información

La organización de la información es un paso fundamental en cualquier análisis estadístico ya que facilita la transformación de los datos en bruto a un formato apropiado para su estudio. En el ámbito de la estadística descriptiva aplicada con SPSS, organizar la información significa definir las variables, identificar su nivel de medición y preparar los datos de tal manera que pudieran ser analizados de manera efectiva.

Figura 22

Organización de los datos



2.9.4. Presentación de resultados

Datos no agrupados

Los datos no agrupados se consideran cuando se tiene un conjunto de datos relativamente pequeño (n); para realizar el análisis estadístico no es necesario agruparlos. La ventaja de trabajar con datos no agrupados es que sus resultados son más precisos y con esto evitar que se pierda información.

Ejemplos.

- El consumo de combustible (en litros por kilómetro) de 5 marcas de vehículos con el mismo cilindraje son: Toyota 0.08(L/km), Chevrolet 0.09 (L/km), Hyundai 0.07(L/km), Nissan 0.10(L/km), Kía 0.08(L/km)
- Las mediciones del voltaje exacto (en voltios) en 5 puntos de un circuito electrónico en condiciones de operación normales son: 12.05, 11.98, 12.02, 12.10, 11.95
- El tiempo exacto (en segundos) que un brazo robótico tarda en completar una cierta tarea es: 12.3, 11.8, 12.1, 12.5, 11.9, 12.0, 12.1, 11.8
- La densidad de 10 muestras de crudo (en kg/m^3) tomadas de diferentes puntos del proceso de producción son: 850, 870, 860, 880, 850, 860, 845, 850, 870, 870
- Las transacciones realizadas (dólares) por una empresa durante un día específico son los siguientes valores: 150, 500, 450, 1000, 250, 150, 150
- El gasto en dólares realizado por 10 turistas en un día específico durante su visita a la laguna del Quilotoa. 50, 40, 20, 30, 50, 20, 30, 45, 30, 40

Datos Agrupados

Tabla tipo I

La tabla tipo I se usa cuando el tamaño de muestra (n) es grande, pero el rango (r) es pequeño. En esta tabla las clases no se agrupan en intervalos lo que facilita visualizar tendencias y patrones. Además, las gráficas estadísticas que se pueden realizar son diagramas de barras y diagramas circulares.

Tablas de frecuencia

Las tablas de frecuencia se utilizan para organizar los datos en categorías o intervalos, y permiten resumir un conjunto de datos de manera clara y concisa.

Frecuencia absoluta (f_i): Indica el número de veces que un evento o valor específico ocurre en un conjunto de datos, perteneciente a esta clase.

La sumatoria de las frecuencias absolutas da como resultado el tamaño de muestra (n)

$$f_1 + f_2 + f_3 + \dots + f_k = n$$

Frecuencia relativa (f_r): Indica la proporción de veces que un valor específico aparece en un conjunto de datos en relación con la muestra.

Se calcula dividiendo la frecuencia absoluta por el número total de observaciones, esta medida se expresa comúnmente de manera decimal o porcentaje.

$$f_r = \frac{f_i}{n} \times 100 \%$$

f_i = frecuencia absoluta

n = tamaño de muestra

Frecuencia absoluta acumulada (f_{ia}): Es la suma de las frecuencias absolutas de todos los valores anteriores más la frecuencia actual.

$$f_{a1} = f_1$$

$$f_{a2} = f_1 + f_2$$

$$f_{a3} = f_{a2} + f_3$$

.

.

.

$$f_{ak} = f_{a(k-1)} + f_k \quad k: \text{Número de clases}$$

Frecuencia relativa acumulada (f_{ra}): Es la suma acumulativa de las frecuencias relativas de un conjunto de datos, la frecuencia de la última clase alcanza un valor de 1 ó 100%, es útil para entender la proporción de datos que se encuentran por debajo de un determinado intervalo de una distribución.

2.9.5. Interpretación de resultados

La interpretación de resultados es la parte más importante en la estadística, porque permite comprender el significado de los datos analizados y poder tomar decisiones correctas basadas en evidencias. Mediante la interpretación se puede responder a las preguntas de investigación, validar hipótesis, identificar relaciones entre variables, hallar patrones y poder tener conclusiones correctas.

Autoevaluación 2.3.

1. ¿Cuáles son las etapas para realizar un análisis estadístico completo?
2. Con sus propias palabras indicar cuando utilizar una tabla de frecuencias tipo I y II
3. Los siguientes datos representan la edad de los estudiantes de tercer nivel de una universidad pública: 18, 19, 20, 20, 19, 20, 21, elaborar una tabla de frecuencias e interpretar los resultados.

Ejercicio práctico N°1

A un grupo de 20 estudiantes de la Universidad de las Fuerzas Armadas ESPE se le consultó cuántas veces a la semana visitan la biblioteca. Las respuestas obtenidas son: 5; 3; 1; 3; 2; 4; 2; 5; 2; 3; 5; 5; 5; 4; 4; 3; 5; 2; 5; 3. Elaborar una tabla de frecuencias.

Cálculo del rango (r): (Es la diferencia entre el dato mayor (H) y el dato menor (L))

$$r = H - L$$

$$r = 5 - 1 = 4$$

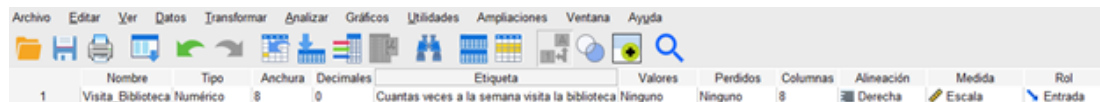
Visita Biblioteca	Frecuencia absoluta	Frecuencia absoluta acumulada	Frecuencia relativa	Frecuencia relativa acumulada
1	1	1	0,05	0.05
2	4	5	0.20	0.25
3	5	10	0.25	0.50
4	3	13	0.15	0.65
5	7	20	0.35	1
Total	20			

Construcción en SPSS

1. Prepare los datos:

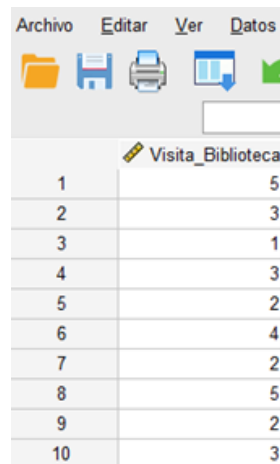
- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree una nueva variable llamada Visita_Biblioteca con los siguientes parámetros:

- **Nombre:** Visita_Biblioteca
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Cuantas veces a la semana visita la biblioteca
- **Medida:** Escala



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Visita_Biblioteca	Numérico	8	0	Cuantas veces a la semana visita la biblioteca	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese los datos

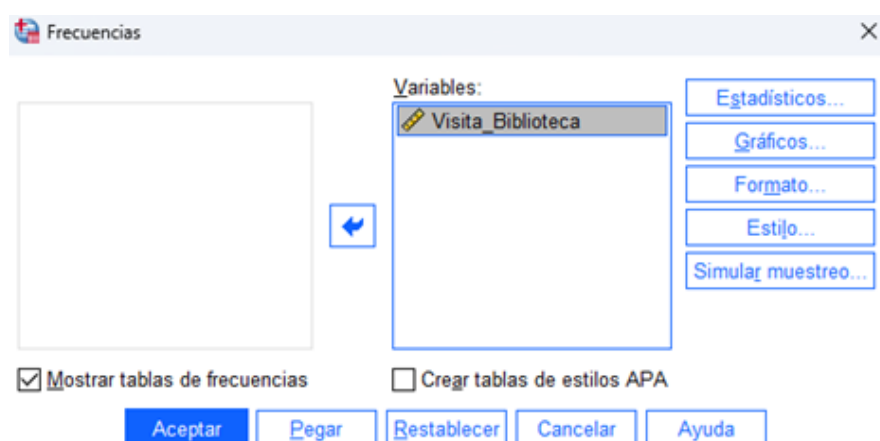


	Visita_Biblioteca
1	5
2	3
3	1
4	3
5	2
6	4
7	2
8	5
9	2
10	3

Crear tabla de frecuencias:

En el menú principal seleccionar: **Analizar / Estadísticos descriptivos / Frecuencias**.

En la ventana que aparece seleccionar la variable de interés (Visita_Biblioteca), luego dar un clic en la opción **Mostrar tablas de frecuencia** y presionar **Aceptar**



Resultados obtenidos

En la Ventana de Resultados, encontrará la tabla de frecuencias generada.

Cuántas veces a la semana visita la biblioteca

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1	1	5.0	5.0	5.0
	2	4	20.0	20.0	25.0
	3	5	25.0	25.0	50.0
	4	3	15.0	15.0	65.0
	5	7	35.0	35.0	100.0
	Total	20	100.0	100.0	

Interpretación

La tabla muestra la visita de 20 estudiantes a la biblioteca durante la semana. El 35% de ellos la visitan 5 veces a la semana, seguido por el 25% que la visitan 3 veces, el 20% que la visitan 2 veces, el 15% que la visitan 4 veces, y solo el 5% que la visitan una vez. Por otra parte, el 50% de los estudiantes visitan la biblioteca al menos 3 veces por semana, y el 65% la visitan al menos 4 veces.

Esta información indica que la gran mayoría de los estudiantes, un 35%, visitan la biblioteca 5 veces a la semana, lo que muestra que muchos estudiantes tienen un hábito frecuente de ir a la biblioteca. Además, el 50% de ellos visitan la biblioteca al menos 3 veces por semana.

Ejercicio práctico N°2

En un estudio realizado a 50 familias de una localidad, se les pregunta sobre el número de hijos que tiene. Los resultados son los siguientes: 2 ; 4 ; 2 ; 3 ; 1 ; 2

; 4 ; 2 ; 3 ; 0 ; 2 ; 2 ; 2 ; 3 ; 2 ; 6 ; 2 ; 3 ; 2 ; 2 ; 3 ; 2 ; 3 ; 3 ; 4 ; 1 ; 3 ; 3 ; 3 ; 4 ; 5 ; 2 ; 0 ; 3 ; 2 ; 1 ; 2 ; 3 ; 2 ; 2 ; 3 ; 1 ; 4 ; 2 ; 3 ; 2 ; 4 ; 3 ; 3 ; Elaborar un tabla de frecuencias e interpretar los resultados.

Cálculo de rango:

$$r = H - L = 6 - 0 = 6$$

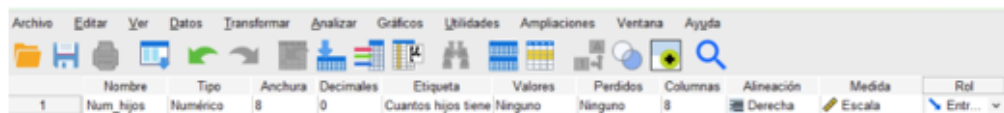
Número hijos	Frecuencia absoluta f_i	Frecuencia absoluta acumulada f_{ia}	Frecuencia Relativa f_r	Frecuencia relativa acumulada f_{ra}
0	2	2	0.04	0.04
1	4	6	0.08	0.12
2	20	26	0.40	0.52
3	16	42	0.32	0.84
4	6	48	0.12	0.96
5	1	49	0.02	0.98
6	1	50	0.02	1
Total	50			

Construcción en SPSS:

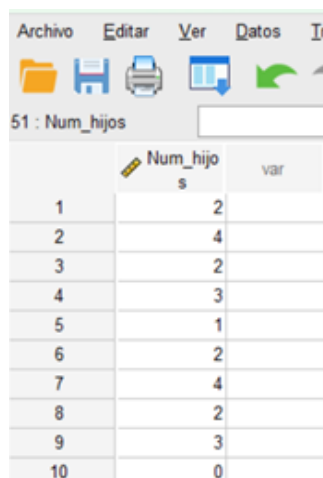
1. *Prepare los datos:*

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Número de hijos con los siguientes parámetros:

- **Nombre:** Num_hijos
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Cuantos hijos tiene
- **Medida:** Escala



- c. En la ventana “Vista de datos” ingrese las edades de los encuestados

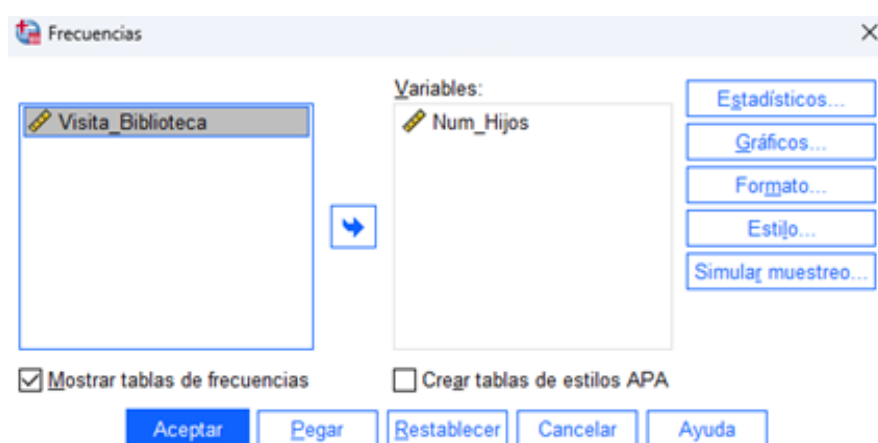


	Num_hijos	var
1	2	
2	4	
3	2	
4	3	
5	1	
6	2	
7	4	
8	2	
9	3	
10	0	

Crear tabla de frecuencias:

En el menú principal seleccionar: Analizar / Estadísticos descriptivos / Frecuencias.

Seleccionar la variable de interés (Num_hijos), luego dar un clic en la opción Mostrar tablas de frecuencia y luego presionar Aceptar.



2. Resultados obtenidos

En la Ventana de Resultados, encontrará la tabla de frecuencias generada

Cuantos hijos tiene

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	0	2	4.0	4.0	4.0
	1	4	8.0	8.0	12.0
	2	20	40.0	40.0	52.0
	3	16	32.0	32.0	84.0
	4	6	12.0	12.0	96.0
	5	1	2.0	2.0	98.0
	6	1	2.0	2.0	100.0
	Total	50	100.0	100.0	

Interpretación

La tabla de frecuencias muestra los resultados de una encuesta sobre el número de hijos en 50 familias en cierta localidad. La mayoría de las familias tienen entre 2 y 3 hijos, siendo 2 hijos el más común con el 40.0%. Un 52.0% de las familias tienen 2 hijos o menos, mientras que el 48.0% tienen más de 2 hijos. Tan solo un 4.0% de las familias no tienen hijos, y un pequeño porcentaje de 2.0% cada uno, tiene 5 o 6 hijos.

Esta información indica que las familias en esta localidad tienden a tener 2 o 3 hijos, hay pocos casos de familias numerosas o que no tienen hijos.

Ejercicio práctico N°3

En una fábrica se elabora piezas metálicas de forma cilíndrica, para verificar si la máquina está trabajando en perfectas condiciones, se toma una muestra de 30 piezas de un lote, si el diámetro está entre 10.00 mm y 10.2 mm no necesita mantenimiento. Los datos obtenidos son los siguientes:

10.1; 10.2; 10.0; 10.3; 10.1; 10.2; 10.0; 10.1; 10.3; 10.2; 10.1; 10.0; 10.2; 10.1; 10.3; 10.0; 10.1; 10.2; 10.1; 10.0; 10.3; 10.2; 10.1; 10.0; 10.2; 10.1; 10.3; 10.0; 10.1; 10.2

Elaborar una tabla de frecuencias e interpretar los resultados.

Cálculo de rango:

$$r = H - L = 10.3 - 10 = 0.3$$

Diámetro	Frecuencia absoluta f_i	Frecuencia absoluta acumulada f_{ia}	Frecuencia Relativa f_r	Frecuencia relativa acumulada f_{ra}
10	7	7	0.233	0.233
10.1	10	17	0.333	0.566
10.2	8	25	0.267	0.833
10.3	5	30	0.167	1
Total	30			

Construcción en SPSS:

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Diámetro (mm) con los siguientes parámetros:

- **Nombre:** Diámetro(mm)
- **Tipo:** Numérico
- **Etiqueta:** Diámetro de la pieza
- **Medida:** Escala

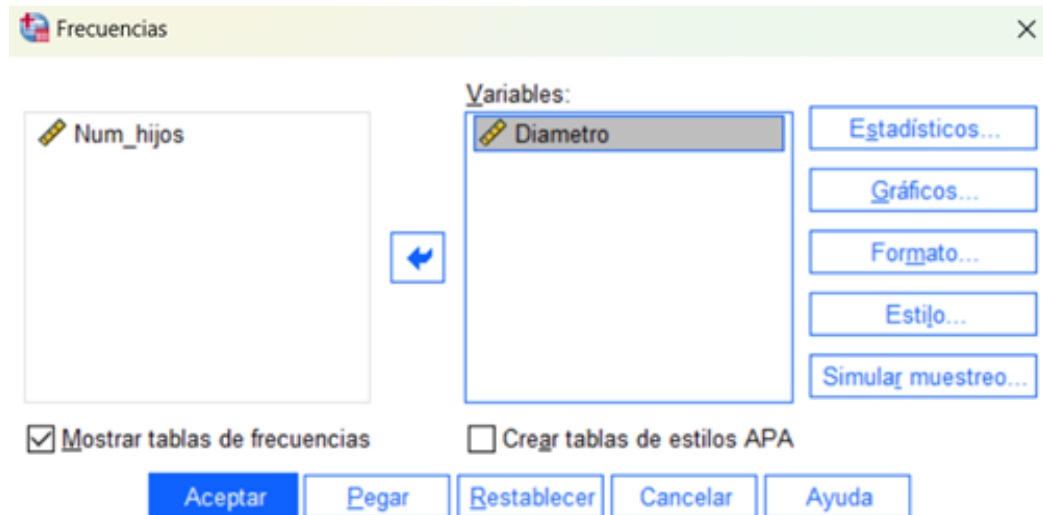


- En la ventana “Vista de datos” ingrese los diámetros de las piezas

Tabla de frecuencias:

En el menú principal seleccionar: Analizar / Estadísticos descriptivos / Frecuencias.

Seleccionar la variable de interés (Diámetro), de un clic en la opción Mostrar tablas de frecuencia y luego presionar Aceptar.



Resultados obtenidos

En la Ventana de Resultados, encontrará la tabla de frecuencias generada

Diámetro de la pieza

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	10.0	7	23.3	23.3	23.3
	10.1	10	33.3	33.3	56.7
	10.2	8	26.7	26.7	83.3
	10.3	5	16.7	16.7	100.0
	Total	30	100.0	100.0	

Interpretación

La tabla de frecuencias muestra que el diámetro más común de las piezas metálicas es 10.1 mm que representa el 33.3%, seguido de 10.2 mm con el 26.7% y 10.0 mm con el 23.3%. El 83.3% de las piezas cumplen con las especificaciones de calidad al tener un diámetro entre 10.0 mm y 10.2 mm, mientras que el 16.7% restante, con un diámetro de 10.3 mm, podrían considerarse fuera de las recomendaciones.

Esto sugiere la necesidad de revisar la maquinaria para reducir la variabilidad y mejorar la precisión en la elaboración de las piezas cilíndricas.

Tabla tipo II

La tabla Tipo II se aplica cuando el tamaño de muestra (n) y el rango (r) son grande, En esta tabla los datos se agrupan en intervalos, se utiliza para datos

discretos o continuos, además permite crear gráficos como histogramas o polígonos de frecuencias, que son herramientas visuales útiles para interpretar la distribución de los datos.

Para realizar una correcta distribución de los datos se recomienda realizar los siguientes pasos a seguir.

1. Identificar el tamaño de muestra (n)

2. Encontrar el número de clases (k)

El número de clases se puede encontrar con las siguientes fórmulas

$k = \sqrt{n}$: Si n no es muy grande.

$k = 1 + 3,33 \cdot \log(n)$: Cualquier otro caso

3. Cálculo de la amplitud inicial o rango (AI)

Es la diferencia en el dato mayor H menos el dato menor L

$$AI = H - L$$

4. Cálculo del intervalo de amplitud (i)

Es el cociente entre la amplitud inicial y el número de clases, si este resultado tiene al menos un decimal sin importar su valor, se redondea al inmediato superior.

$$i \geq \frac{H - L}{k}$$

5. Cálculo de la amplitud total (AT)

Es el producto entre el número de clases y el intervalo de amplitud.

$$AT = k \times i$$

6. Cálculo del exceso (E)

Es la diferencia entre la amplitud total menos la amplitud inicial

$$E = AT - AI$$

7. Cálculo del límite real inferior (LRI)

Es la diferencia entre el dato menor menos la mitad del exceso.

$$LRI = L - \frac{E}{2}$$

Ejercicio práctico N° 4

Para entender cuánto tiempo tardan los estudiantes en viajar desde su lugar de residencia hasta el campus Universitario, se hizo una encuesta a 30 estudiantes de diferentes carreras, obteniendo los siguientes resultados.

Nota: Los resultados del tiempo están dados en minutos.

60	10	60	60	40	60	30	60	30	30
30	120	90	90	30	60	45	45	25	75
30	45	80	30	60	60	20	60	60	25

1.- Identificar el número de datos (n)

$$n = 30$$

2.- Encontrar el número de clases (k)

$$k = \sqrt{n} = \sqrt{30} = 5,5 \approx 6$$

$$k = 1 + 3,33 \cdot \log(30) = 5,92 \approx 6$$

3.- Rango o amplitud inicial (AI)

$$AI = H - L = 120 - 10 = 110$$

4.- Intervalo de amplitud (i)

$$i \geq \frac{H-L}{k} = \frac{120-10}{6} = 18.33 \gg 19$$

5.- Amplitud total (AT)

$$AT = k \times i = 6 \times 19 = 114$$

6.- Exceso (E)

$$E = AT - AI = 114 - 110 = 4$$

7.- Limite real inferior (LRI)

$$LRI = L - \frac{E}{2} = 10 - \frac{4}{2} = 8$$

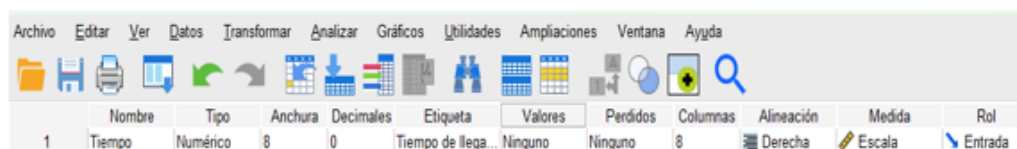
Número de clase k	Clase	Frecuencia absoluta f_i	Frecuencia absoluta acumulada f_{ia}	Frecuencia relativa f_r	Frecuencia relativa acumulada f_{ra}
1	8 - 26	4	4	0.133	0.133
2	27 - 45	11	15	.367	0.50
3	46 - 64	10	25	.333	0.833
4	65 - 83	2	27	.067	0.9
5	84 - 102	2	29	.067	0.967
6	103 - 122	1	30	.033	1
Total		30			

Construcción en SPSS:

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Tiempo (min) con los siguientes parámetros:

- **Nombre:** Tiempo
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Tiempo que tarda desde su vivienda hasta la Universidad
- **Medida:** Escala

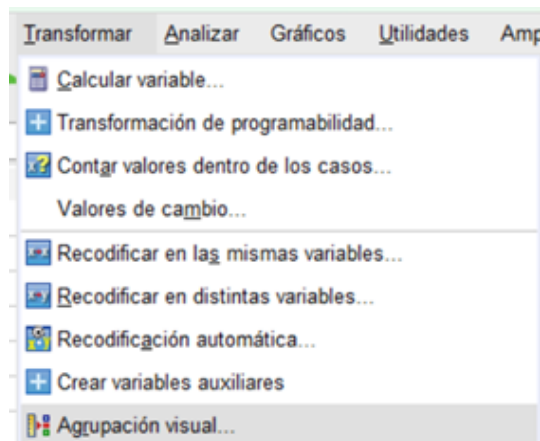


- En la ventana “Vista de datos” ingrese los tiempos obtenidos

Archivo Editar Ver	
14 :	
	Tiempo
1	60
2	10
3	60
4	60
5	40

Crear intervalos (Agrupar los datos)

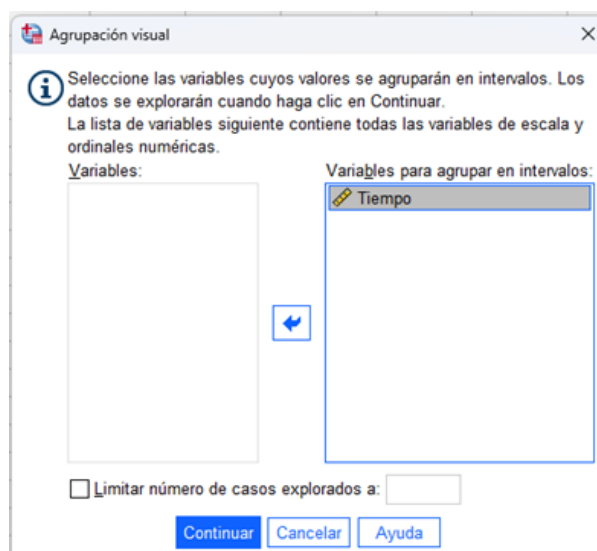
- a. En el menú principal seleccionar: Transformar / Agrupación visual.



- b. Seleccionar la variable que se necesita agrupar (Tiempo), y dar un clic en Continuar

Figura 23

Agrupación visual, selección de variables



- c. En la ventana “Agrupación Visual” ajustar los intervalos manualmente
Ingresar el nombre de la variable agrupada (Por ejemplo, Tiempo_Agrupado) y dar un clic en el botón “Crear puntos de corte”

Lista de variables exploradas: **Tiempo**

Nombre: **Tiempo** Etiqueta: **Tiempo de llegada**

Variable actual: **Tiempo** Variable agrupada: **Tiempo de llegada (Agrupada)**

Mínimo: **10** Valores no perdidos Máximo: **120**

Especifique puntos de corte de intervalo o haga clic en Crear puntos de corte para intervalos automáticos. Un valor de punto de corte de 10, por ejemplo, define un intervalo que empieza por encima del intervalo anterior y acaba en 10.

Valor	Etiqueta
1	HIGH
2	

Puntos finales super...
☒ Incluidos (<=)
☐ Excluidos (<)

Crear puntos de corte...
 Crear etiquetas
☐ Invertir la escala

Aceptar Pegar Restablecer Cancelar Ayuda

- d. Definir los puntos de corte de los intervalos.
- Ingresar estos datos de acuerdo al cálculo previo realizado con anterioridad (Tabla Tipo II)
 - Clic en Aplicar

Figura 24

Creación de puntos de corte

Crear puntos de corte

☒ Intervalos de ancho iguales

Intervalos - rellenar al menos dos campos

Primera ubicación de punto de corte: 26

Número de puntos de corte: 6

Anchura: 19

Última ubicación de punto de corte: 121

☐ Percentiles iguales basados en casos explorados

Intervalos - rellene cualquier campo

Número de puntos de corte:

Ancho (%):

☐ Puntos de corte en media y desviaciones estándar seleccionadas basadas en casos explorados

☐ Desviación estándar +/- 1

☐ Desviación estándar +/- 2

☐ Desviación estándar +/- 3

i La aplicación sustituirá las definiciones de punto de corte actuales con esta especificación. Un intervalo final incluirá todos los valores restantes: N puntos de corte generan N+1 intervalos.

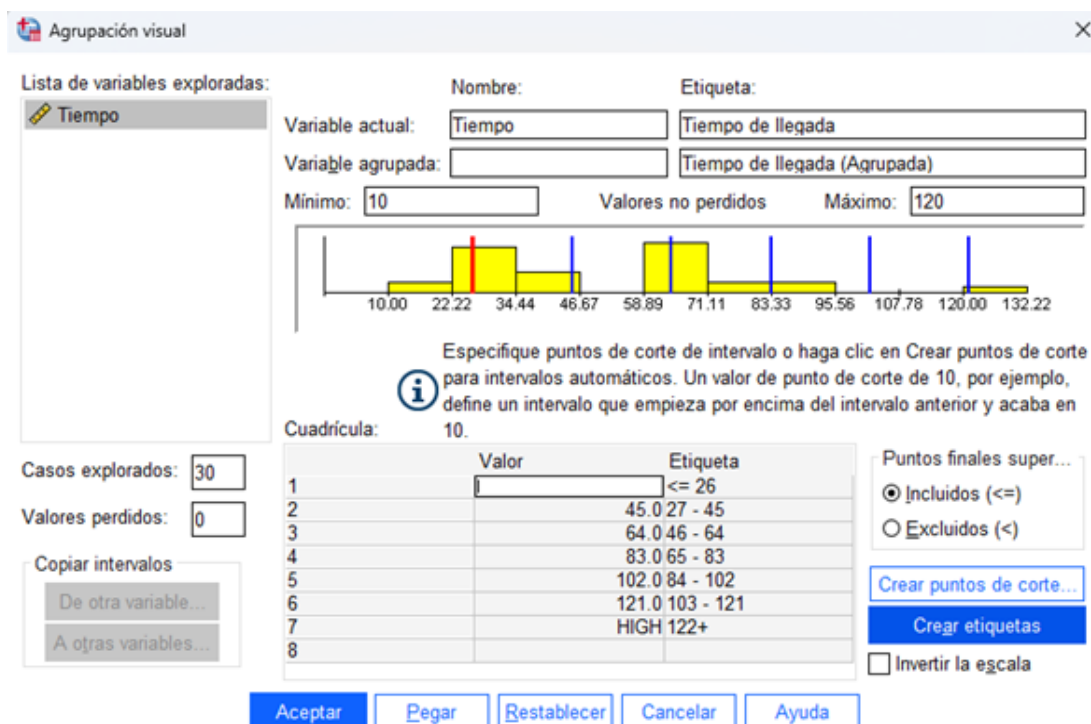
Aplicar Cancelar Ayuda

e. Crear Etiquetas

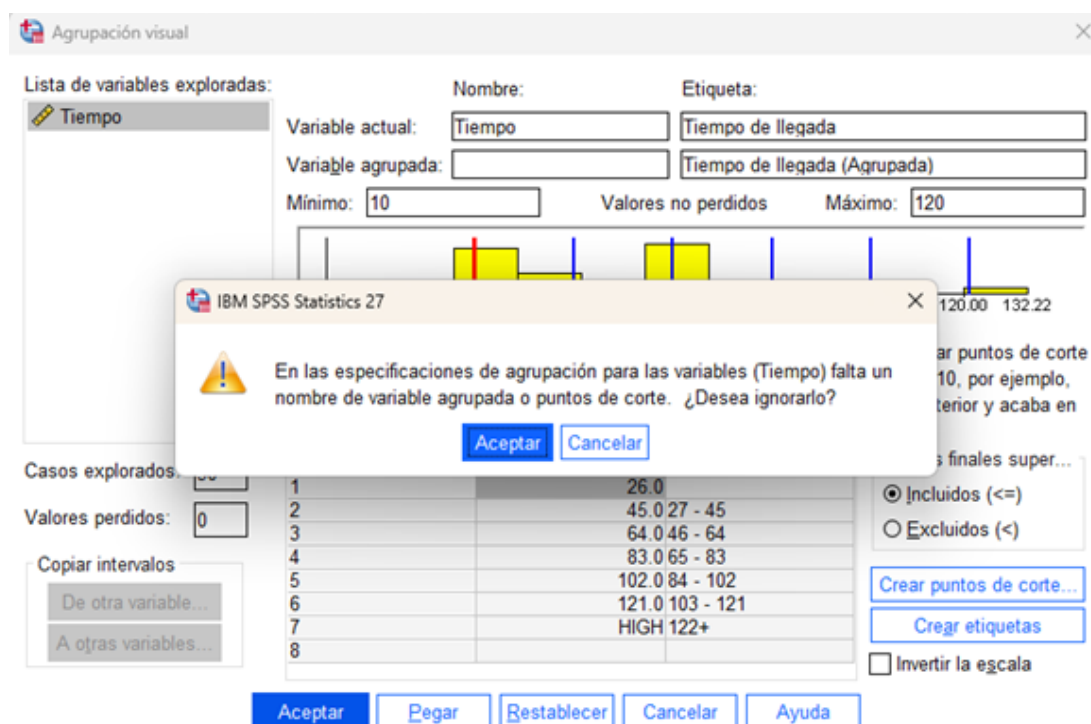
- Clic en crear etiquetas
- Clic en Aceptar

Figura 25

Agrupación visual. Crear etiquetas



f. Para crear la nueva variable agrupada hacer clic en Aceptar



- g. Interpretar los resultados de la tabla de frecuencia agrupada (tabla tipo II) obtenidos en la hoja de salida.

1. Resultados obtenidos

En la Ventana de Resultados, encontrará la tabla de frecuencias generada.

Tiempo de llegada (Agrupada)					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	<= 26	4	13.3	13.3	13.3
	27 - 45	11	36.7	36.7	50.0
	46 - 64	10	33.3	33.3	83.3
	65 - 83	2	6.7	6.7	90.0
	84 - 102	2	6.7	6.7	96.7
	103 - 121	1	3.3	3.3	100.0
	Total	30	100.0	100.0	

Interpretación

La mayoría de los estudiantes universitarios, el 83.3%, tardan 64 minutos o menos en llegar al campus universitario; el intervalo más común está entre los 27 y 45 minutos, con el 36.7%. Tan solo un 10% de los estudiantes toma tiempos de viaje más largos, con un 6.7% tardando entre 84-102 minutos y un 3.3% entre 103-121 minutos.

Esto indica que la mayoría de los estudiantes universitarios tiene un tiempo de viaje razonable; un pequeño grupo enfrenta trayectos significativamente más largos, posiblemente debido a la distancia de sus residencias que se encuentran fuera del cantón o provincia.

Ejercicio práctico N° 5

Se realiza una investigación sobre el precio de zapatos deportivos, de similares características en diversos almacenes de la ciudad, obteniéndose los siguientes datos (dólares).

50	43	39	43	40	38	35	25	37	32
49	43	39	44	40	38	33	26	36	30
49	43	39	44	40	38	33	27	36	30
47	41	39	45	40	37	33	27	35	30
46	41	38	46	40	37	32	28	35	28

Elabore una tabla de frecuencias con datos agrupados por clase.

1. Identificar el número de datos (n)

$$n=50$$

2. Encontrar el número de clases (k)

$$k=1+3,33 \cdot \log(50)=6.66 \approx 7$$

3. Rango o amplitud inicial (AI)

$$AI=H-L=50-25=25$$

4. Intervalo de amplitud (i)

$$i \geq \frac{H-L}{k} = \frac{50-25}{7} = 3.57 \gg 4$$

5. Amplitud total (AT)

$$AT=k \times i=7 \times 4=28$$

6.- Exceso(E)

$$E=AT-AI=28-25=3$$

7.- Limite real inferior (LRI)

$$LRI = L - \frac{E}{2} = 25 - \frac{3}{2} = 23.5$$

Número de clase k	Clase	Frecuencia absoluta f_i	Frecuencia absoluta acumulada f_{ia}	Frecuencia Relativa f_r	Frecuencia relativa acumulada f_{ra}
1	23.5 - 27.4	4	4	0.08	0.08
2	27.5 - 31.4	5	9	0.1	0.18
3	31.5 - 35.4	8	17	0.16	0.34
4	35.5 - 39.4	13	30	0.26	0.6
5	39.5 - 43.4	11	41	0.22	0.82
6	43.5 - 47.4	6	47	0.12	0.94
7	47.5 - 51.5	3	50	0.06	1
Total		50			

Construcción en SPSS:**1. Prepare los datos:**

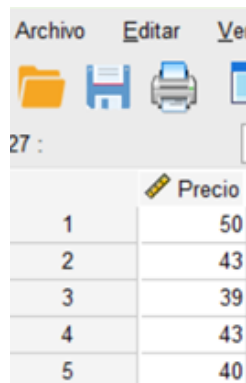
- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Precio (dólares) con los siguientes parámetros:

- **Nombre:** Precio
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Precio de los zapatos deportivos
- **Medida:** Escala



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Precio	Numérico	6	0	Precio de los zapatos deportivos	Ninguno	Ninguno	6	Derecha	Escala	Entr...

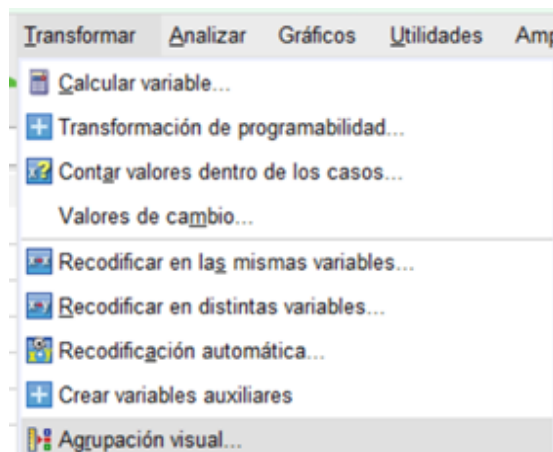
- En la ventana “Vista de datos” ingrese los tiempos obtenidos



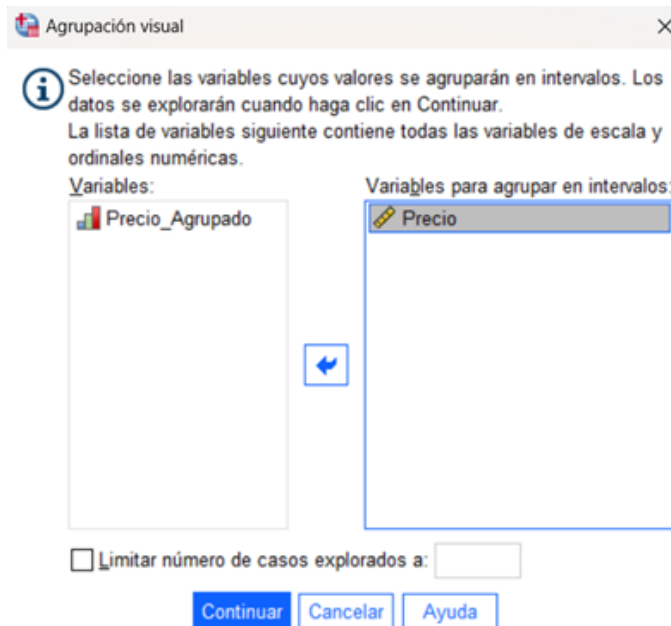
	Precio
1	50
2	43
3	39
4	43
5	40

2. Crear intervalos (Agrupar los datos)

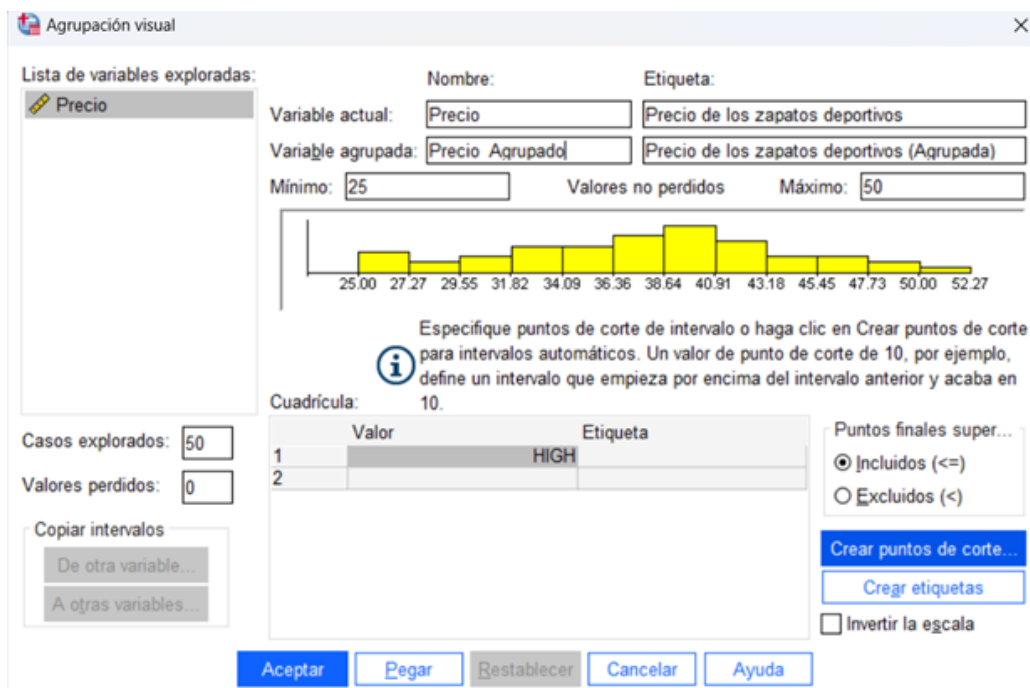
- En el menú principal seleccionar: Transformar / Agrupación visual.



- b. Seleccionar la variable que se necesita agrupar (Precio), y clic en continuar



- c. En la ventana “Agrupación Visual” ajustar los intervalos manualmente
Ingresar el nombre de la variable agrupada (Por ejemplo, Precio_Agrupado) y dar un clic en “Crear puntos de corte”



- Definir los puntos de corte de los intervalos.
- Ingresar estos datos de acuerdo al cálculo previo realizado con anterioridad (Tabla Tipo I) Clic en Aplicar

Crear puntos de corte

☒ **Intervalos de ancho iguales**

Intervalos - rellenar al menos dos campos

Primera ubicación de punto de corte: 28

Número de puntos de corte: 7

Anchura: 4

Última ubicación de punto de corte: 52

☐ **Percentiles iguales basados en casos explorados**

Intervalos - rellene cualquier campo

Número de puntos de corte:

Ancho (%):

☐ **Puntos de corte en media y desviaciones estándar seleccionadas basadas en casos explorados**

☐ Desviación estándar +/- 1

☐ Desviación estándar +/- 2

☐ Desviación estándar +/- 3

i La aplicación sustituirá las definiciones de punto de corte actuales con esta especificación. Un intervalo final incluirá todos los valores restantes: N puntos de corte generan N+1 intervalos.

Aplicar Cancelar Ayuda

d. Crear Etiquetas

- Clic en crear etiquetas
- Clic en Aceptar

Agrupación visual

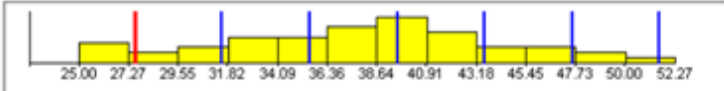
Lista de variables exploradas:

Precio

Nombre: Variable actual: Precio Etiqueta: Precio de los zapatos deportivos

Variable agrupada: Precio Agrupado Etiqueta: Precio de los zapatos deportivos (Agrupada)

Mínimo: 25 Valores no perdidos Máximo: 50



i Especifique puntos de corte de intervalo o haga clic en Crear puntos de corte para intervalos automáticos. Un valor de punto de corte de 10, por ejemplo, define un intervalo que empieza por encima del intervalo anterior y acaba en 10.

Cuadrícula:

	Valor	Etiqueta
1		<= 28
2		31.5 29 - 32
3		35.5 33 - 36
4		39.5 37 - 40
5		43.5 41 - 44
6		47.5 45 - 48
7		51.5 49 - 52
8		HIGH 53+
9		

Puntos finales super...

☒ Incluidos (<=)

☐ Excluidos (<)

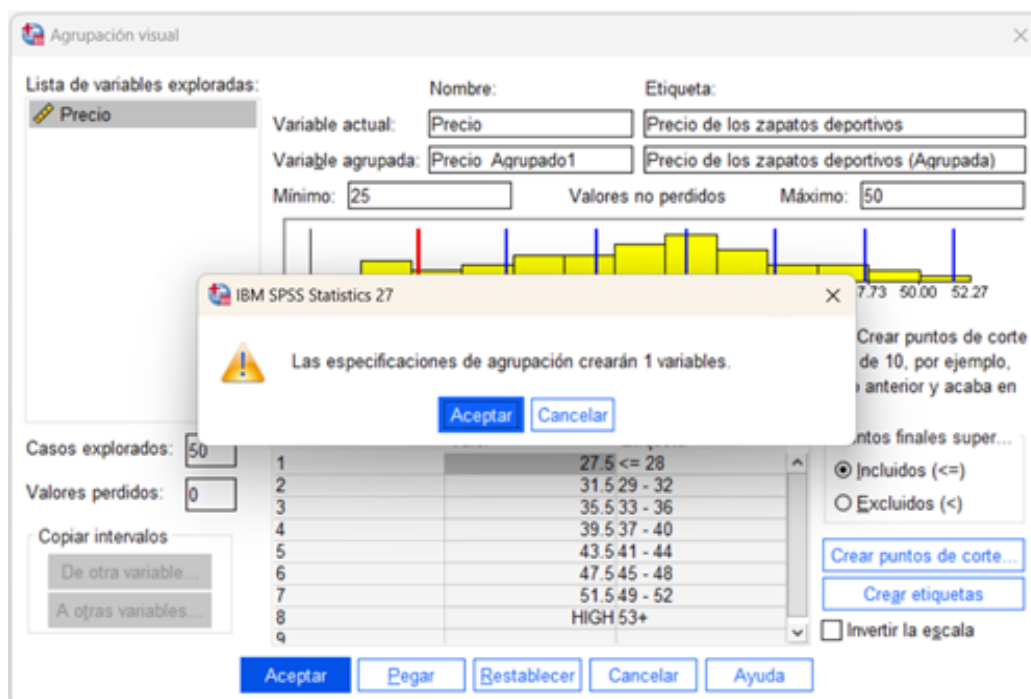
Crear puntos de corte...

Crear etiquetas

☐ Invertir la escala

Aceptar Pegar Restablecer Cancelar Ayuda

e. Para crear la nueva variable agrupada hacer clic en Aceptar



f. Interpretar los resultados de la tabla de frecuencia agrupada (tabla tipo II) obtenidos en la hoja de Salida.

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la tabla de frecuencias.

Precio de los zapatos deportivos (Agrupada)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	<= 27	4	8.0	8.0	8.0
	28 - 31	5	10.0	10.0	18.0
	32 - 35	8	16.0	16.0	34.0
	36 - 39	13	26.0	26.0	60.0
	40 - 43	11	22.0	22.0	82.0
	44 - 47	6	12.0	12.0	94.0
	48 - 51	3	6.0	6.0	100.0
	Total	50	100.0	100.0	

4. Interpretación

La tabla muestra que los precios más comunes de zapatos deportivos están en el rango de 36 a 39 dólares, con un 26%, seguido por el rango de precios de 40 a 43 dólares con un 22%. Los precios más bajos se encuentran por debajo de los 27 dólares y los precios más altos están en el rango de 48 a 51 dólares, son los menos frecuentes tan solo con un 8% y 6%, respectivamente.

Los valores de los zapatos deportivos en su gran mayoría, con el 82%, se encuentran entre 28 a 43 dólares, lo que evidencia una tendencia hacia precios moderados dentro del mercado de esta ciudad.

Ejemplo práctico N° 6 (con Base de Datos)

Con la base de datos del capítulo I, correspondiente al link:

Accede aquí:

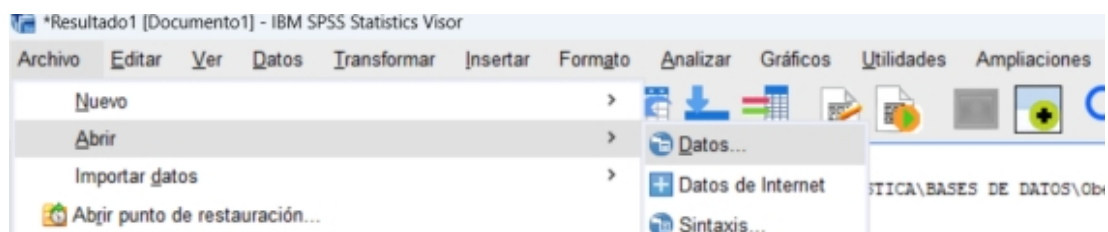


Niveles de obesidad en individuos de los países de México, Perú y Colombia, con base en sus hábitos alimenticios y condición física. Usando el archivo descargado con el nombre ObesityDataSet_raw_and_data_synthetic.csv. Realizar las siguientes actividades:

Realizar una tabla de frecuencias tipo I de la variable (CAEC: ¿Comes algún alimento entre comidas?)

1. En SPSS abrir el archivo guardado con el nombre de `Obesidad.sav`, que se realizó en el Ejercicio práctico N°2 del Capítulo I.

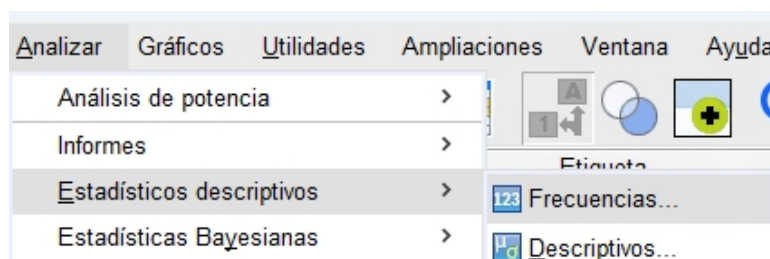
Seleccionar la siguiente ruta Archivo/ Abrir/ Datos



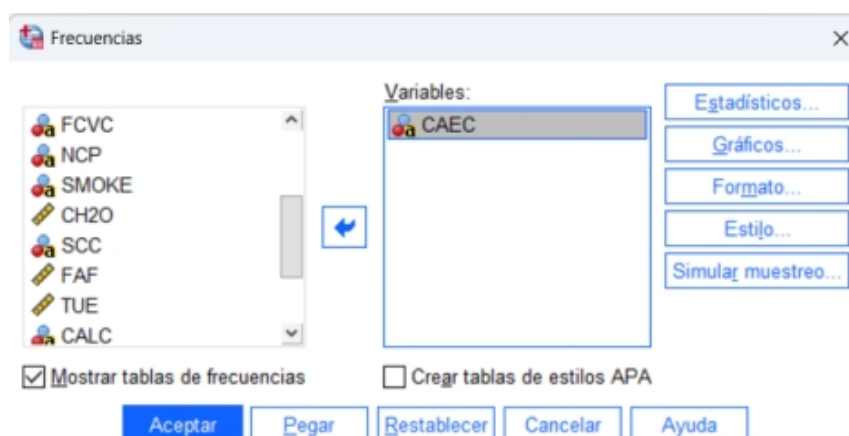


- a. Tabla de frecuencias tipo I de la variable (CAEC: ¿Comes algún alimento entre comidas?)

En el menú principal seleccionar: Analizar / Estadísticos descriptivos / Frecuencias.



Seleccionar la variable de interés (CAEC), Mostrar tablas de frecuencia y Aceptar



Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la tabla de frecuencias

¿Comes algún alimento entre comidas?

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Always	53	2,5	2,5	2,5
	Frequently	242	11,5	11,5	14,0
	no	51	2,4	2,4	16,4
	Sometimes	1765	83,6	83,6	100,0
	Total	2111	100,0	100,0	

En una muestra de 2111 personas encuestadas de México, Colombia y Perú, 1765 que corresponde al 83,6% indicó que come algo entre comidas (Sometimes), el 11,5% frecuentemente, mientras que solo tan solo el 2,5% lo hace siempre y el 2,4% nunca; de igual manera, el 14% del total lo hace de forma frecuente y siempre, lo que se concluye que un pequeño grupo está relacionado con niveles elevados de obesidad, en relación con la gran mayoría que solo come algún alimento entre comidas ocasionalmente.

Ejemplo práctico N° 7 (con Base de Datos)

La siguiente base de datos, Corresponde a los resultados del Examen Ser Bachiller durante la Pandemia, lo puede encontrar en el siguiente link:

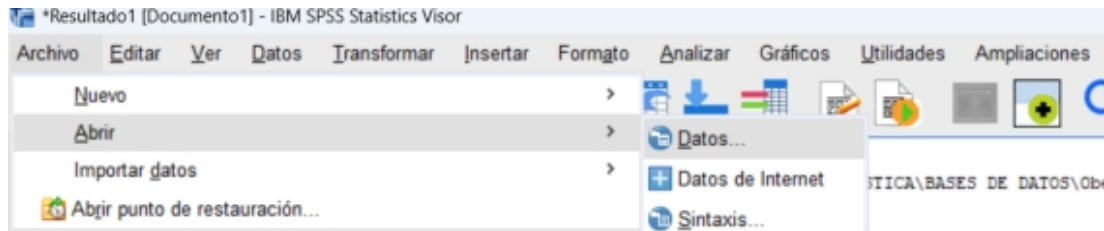
Accede aquí:



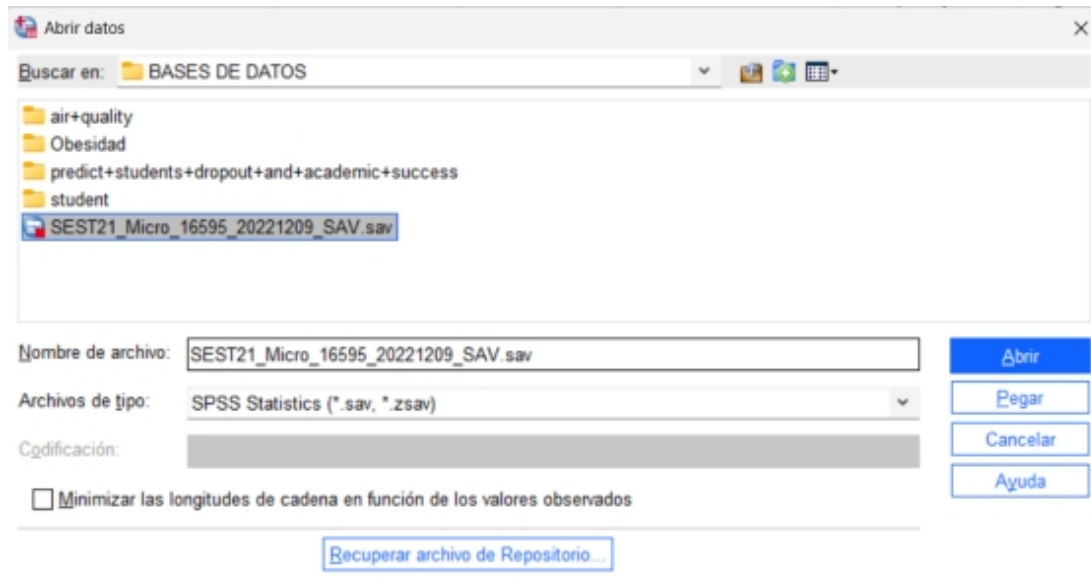
Usando esta base de datos realizar una tabla de frecuencia Tipo II (Agrupada por intervalos) de la variable inev (Promedio global (Nota global obtenida por el estudiante)).

- Abrir el archivo descargado

Seleccionar la ruta Archivo/ Abrir/ Datos



Seleccionar la el archivo **SEST21_Micro_16595_20221209_SAV** y seleccionar **Abrir**



Se obtiene la siguiente ventana

IBM SPSS Statistics Editor de datos											
Archivo	Editar	Ver	Datos	Transformar	Análisis	Gráficos	Herramientas	Aplicaciones	Ventanas	Ayuda	
											Visible: 49 de 49 variables
ciclo	grado	estado_eval	codigo	serie	nro_reg	es_regno	id_zona	id_dist	id_circ	id_prov	id_iniv
2020-2021	4° de EGB	Ausente	0081LEW9D	Escuela de Educaci...	Siena	Siena - Amazonia Zona 6 (Azuay - Cañ...	Cuenca 2	Cuenca - Manabí	Azuay	Cue...	
2020-2021	3° de BG	Evaluated	00823R1PG	Glenn Doman	Siena	Siena - Amazonia Zona 3 (Cotacachi - C...	Ambato 2	Picagua - Totoras	Tangarashua	Am...	
2020-2021	10° de EGB	Ausente	00F58D7W3	Escuela de Educaci...	Siena	Siena - Amazonia Zona 4 (Azuay - Cañ...	Azogues - Billa	Azogues	Cañar	Azo...	
2020-2021	10° de EGB	Evaluated	002W0299T	Unidad Educativa E...	Costa	Costa - Galapagos Zona 4 (Manabí - Sa...	Portoviejo	Portoviejo	Manabí	Port...	
2020-2021	4° de EGB	Evaluated	0057BL0MT	Unidad Educativa M...	Costa	Costa - Galapagos Zona 5 (Bolívar - Gu...	Babahoyo - Baba...	La Unión	Los Rios	Bab...	
2020-2021	3° de BG	Evaluated	000TCDSP1	Glenn Doman	Siena	Siena - Amazonia Zona 3 (Cotacachi - C...	Ambato 2	Picagua - Totoras	Tangarashua	Am...	
2020-2021	10° de EGB	Evaluated	00E2H21R3	Unidad Educativa P...	Costa	Costa - Galapagos Zona 6 (Guayasquí - ...	Gómez Rendón	Monro	Guayas	Gua...	
2020-2021	4° de EGB	Evaluated	00EFDMA7X	Escuela de Educaci...	Costa	Costa - Galapagos Zona 4 (Manabí - Sa...	Portoviejo	Portoviejo	Manabí	Port...	
2020-2021	7° de EGB	Evaluated	00F58D8E4B	Rocafuerte	Costa	Costa - Galapagos Zona 1 (Esmeraldas...	Rio Verde	Montalvo (Cab. en H...	Esmeraldas	Rio...	
2020-2021	3° de BG	Ausente	00K0G47VFS	Unidad Educativa Pr...	Costa	Costa - Galapagos Zona 5 (Bolívar - Gu...	El Triunfo - Gral...	Gral. Antonio Elizald...	Guayas	Gen...	
2020-2021	7° de EGB	Evaluated	00ANF0R3H	Jorge Alfaro Castillo	Oriente	Siena - Amazonia Zona 1 (Esmeraldas...	Lago Agrio	Nueva Loja	Sucumbales	Lago...	

b. Agrupación en intervalos de la variable **inev** (Promedio global (Nota global obtenida por el estudiante)).

- Con la ayuda del SPSS Realizar una tabla de frecuencias de la variable **inev**, para identificar la muestra (n), el dato mayor (H) y menor (L)

Muestra n = 11657, Dato mayor 945, Dato menor 408

- Calcular el número de clases, el intervalo de amplitud, el límite real superior y el valor de la primera ubicación del punto de corte

1. Identificar el número de datos

$$n = 11657$$

2. Encontrar el número de clases

$$k = 1 + 3,33 \cdot \log(11657) = 14.54 \approx 15$$

3. Rango o amplitud inicial.

$$AI = H - L = 945 - 408 = 537$$

4. Intervalo de amplitud

$$i \geq \frac{H-L}{k} = \frac{945-408}{15} = 35.8 \gg 36$$

5. Amplitud total

$$AT = k \times i = 15 \times 36 = 540$$

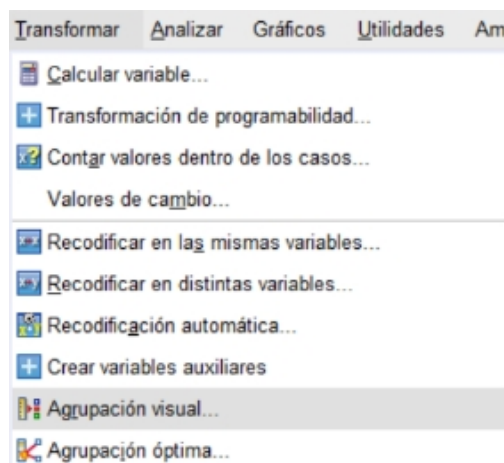
6. El exceso

$$E = AT - AI = 540 - 537 = 3$$

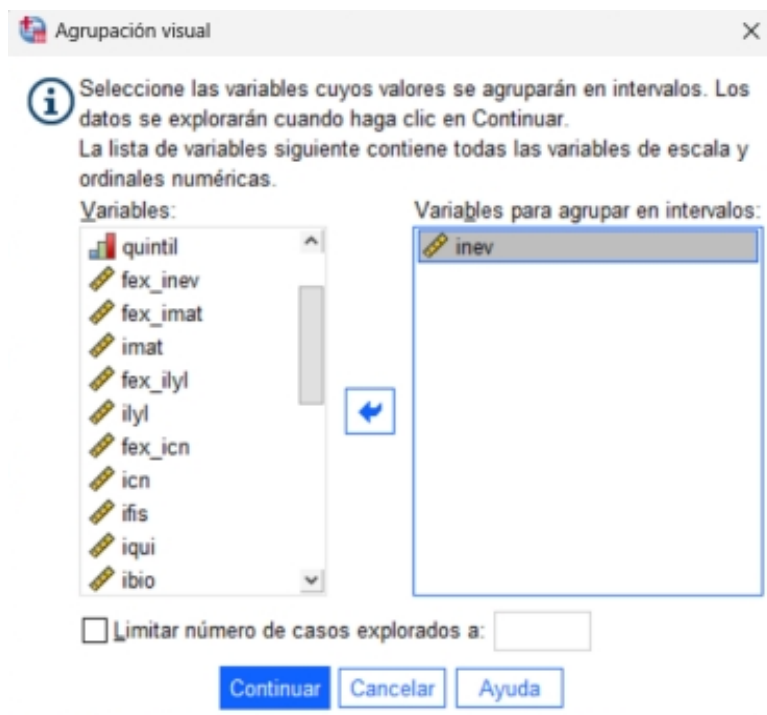
7. Limite real inferior

$$LRI = L - \frac{E}{2} = 408 - \frac{3}{2} = 406.5$$

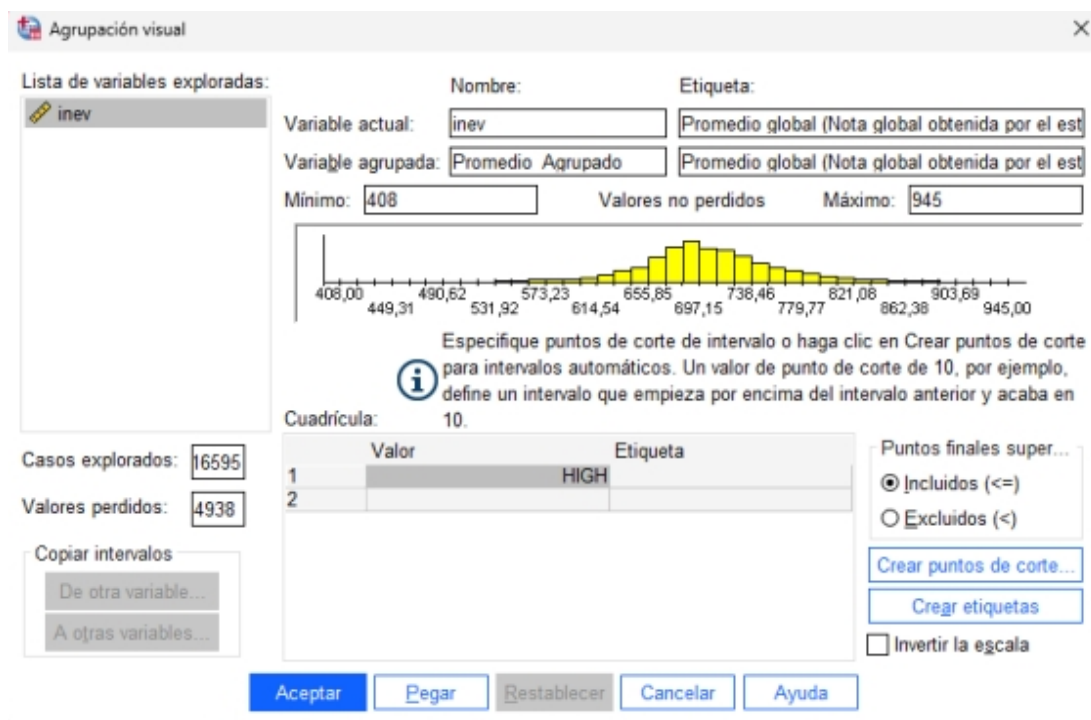
- En el SPSS agrupar la variable de estudio (inev) en intervalos
Seleccionar la ruta: Transformar/Agrupación visual



Seleccionar la variable (inev) a ser agrupada en intervalos y clic en Continuar



En la opción Variable agrupada: proporcionar el nuevo nombre de la variable Promedio_Agrupado y dar clic en Crear puntos de corte...



Ingresar los valores calculados anteriormente como se indica en la figura y clic en Aplicar

Crear puntos de corte

☒ **Intervalos de ancho iguales**

Intervalos - rellenar al menos dos campos

Primera ubicación de punto de corte: 442

Número de puntos de corte: 15

Anchura: 36

Última ubicación de punto de corte: 946

☐ **Percentiles iguales basados en casos explorados**

Intervalos - rellene cualquier campo

Número de puntos de corte:

Ancho (%):

☐ **Puntos de corte en media y desviaciones estándar seleccionadas basadas en casos explorados**

☐ Desviación estándar +/- 1

☐ Desviación estándar +/- 2

☐ Desviación estándar +/- 3

i La aplicación sustituirá las definiciones de punto de corte actuales con esta especificación. Un intervalo final incluirá todos los valores restantes: N puntos de corte generan N+1 intervalos.

Aplicar Cancelar Ayuda

Finalmente dar clic en Crear etiquetas / Aceptar y se crea una nueva variable llamada Promedio_Agrupado

Agrupación visual

Lista de variables exploradas: ineq

Nombre: Variable actual: ineq Etiqueta: Promedio global (Nota global obtenida por el est)

Variable agrupada: Promedio_Agrupado Etiqueta: Promedio global (Nota global obtenida por el est)

Mínimo: 408 Valores no perdidos Máximo: 945

i Especifique puntos de corte de intervalo o haga clic en Crear puntos de corte para intervalos automáticos. Un valor de punto de corte de 10, por ejemplo, define un intervalo que empieza por encima del intervalo anterior y acaba en 10.

Cuadrícula:

	Valor	Etiqueta
1		<= 443
2	479,0444 - 479	
3	515,0480 - 515	
4	551,0516 - 551	
5	587,0552 - 587	
6	623,0588 - 623	
7	659,0624 - 659	
8	695,0660 - 695	

Casos explorados: 16595

Valores perdidos: 4938

Copiar intervalos

De otra variable...

A otras variables...

Puntos finales super...

☒ Incluidos (<=)

☐ Excluidos (<)

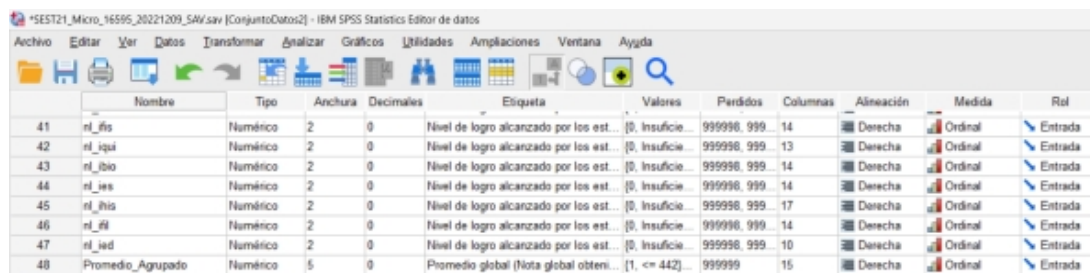
Crear puntos de corte...

Crear etiquetas

☐ Invertir la escala

Aceptar Pegar Restablecer Cancelar Ayuda

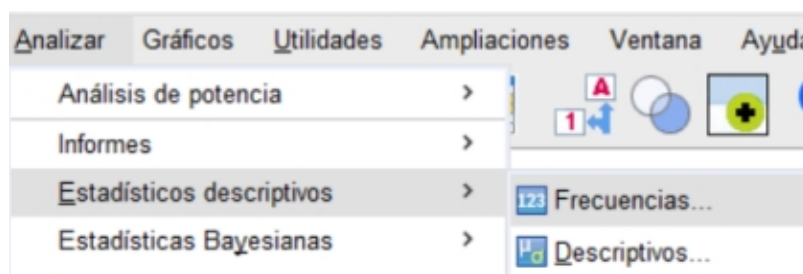
Ahora se crea una nueva variable llamada Promedio_Agrupado



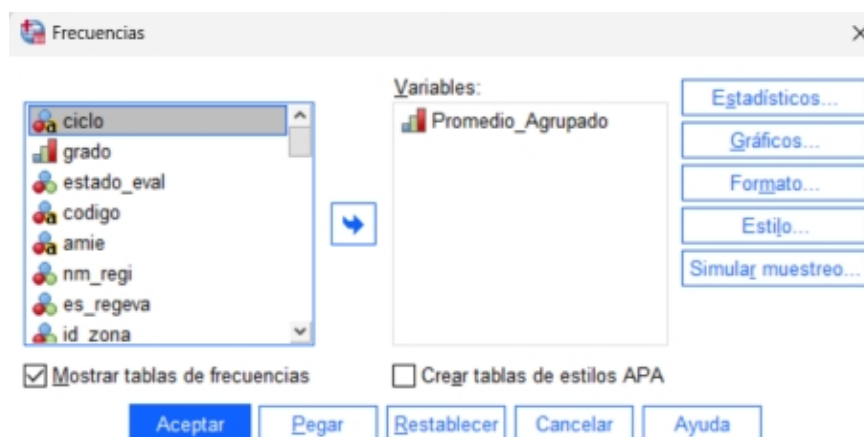
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rel
41	ni_efs	Numérico	2	0	Nivel de logro alcanzado por los est...	[0, Insuficie...	999998, 999...	14	Derecha	Ordinal	Entrada
42	ni_igu	Numérico	2	0	Nivel de logro alcanzado por los est...	[0, Insuficie...	999998, 999...	13	Derecha	Ordinal	Entrada
43	ni_bio	Numérico	2	0	Nivel de logro alcanzado por los est...	[0, Insuficie...	999998, 999...	14	Derecha	Ordinal	Entrada
44	ni_ies	Numérico	2	0	Nivel de logro alcanzado por los est...	[0, Insuficie...	999998, 999...	14	Derecha	Ordinal	Entrada
45	ni_its	Numérico	2	0	Nivel de logro alcanzado por los est...	[0, Insuficie...	999998, 999...	17	Derecha	Ordinal	Entrada
46	ni_iti	Numérico	2	0	Nivel de logro alcanzado por los est...	[0, Insuficie...	999998, 999...	14	Derecha	Ordinal	Entrada
47	ni_jed	Numérico	2	0	Nivel de logro alcanzado por los est...	[0, Insuficie...	999998, 999...	10	Derecha	Ordinal	Entrada
48	Promedio_Agrupado	Numérico	5	0	Promedio global (Nota global obteni...	[1, <= 442]...	999999	15	Derecha	Ordinal	Entrada

c. Creación de la tabla de frecuencias Tipo II de la variable Promedio_Agrupado

- Seleccionar la siguiente ruta **Analizar/Estadísticos descriptivos/Frecuencias**



- Seleccionar la variable Promedio_Agrupado y clic en Aceptar



- En la hoja de resultados de SPSS se obtiene la tabla de Frecuencias tipo II.

Promedio global (Nota global obtenida por el estudiante) (Agrupada)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	<= 442	3	,0	,0	,0
	443 - 478	1	,0	,0	,0
	479 - 514	10	,1	,1	,1
	515 - 550	51	,3	,4	,6
	551 - 586	199	1,2	1,7	2,3
	587 - 622	422	2,5	3,6	5,9
	623 - 658	1256	7,6	10,8	16,7
	659 - 694	3300	19,9	28,3	45,0
	695 - 730	3199	19,3	27,4	72,4
	731 - 766	1844	11,1	15,8	88,2
	767 - 802	885	5,3	7,6	95,8
	803 - 838	333	2,0	2,9	98,7
	839 - 874	127	,8	1,1	99,8
	875 - 910	25	,2	,2	100,0
	911 - 946	2	,0	,0	100,0
	Total	11657	70,2	100,0	
Perdidos	No se dispone de información	4938	29,8		
Total		16595	100,0		

Interpretación

El análisis de los resultados del examen “Ser Bachiller” durante el periodo de pandemia en el Ecuador revela que se contó con 11657 calificaciones válidas (70.2%), observándose que más de la mitad de los evaluados (55.7%) se ubicaron entre los 659 y 730 puntos. La frecuencia de notas disminuye de forma progresiva hacia ambos extremos, con muy pocos estudiantes en los intervalos inferiores (≤ 550) y superiores (≥ 875), donde solo dos alumnos alcanzaron el rango máximo de 911 a 946 puntos. Cabe destacar que un 29.8% de los registros corresponden a datos perdidos, lo que podría estar asociado a las circunstancias excepcionales del contexto pandémico. En síntesis, entre los estudiantes que sí lograron rendir la prueba, prevaleció un desempeño académico moderado-alto.

2.10. Ejercicios propuestos

1. Los siguientes datos representan el número de hermanos que tiene cada estudiante:

2	3	3	2	2	1	4	0	3	2
1	1	1	3	1	3	3	5	2	5
4	3	2	4	3	2	3	4	1	0
0	1	1	1	3	1	2	2	2	1

- Elabore una tabla de frecuencias.
 - ¿Cuántos estudiantes no tienen hermanos?
 - ¿Qué porcentaje de estudiantes tiene entre 1 y 2 hermanos?
 - ¿Cuántos estudiantes tienen más de 3 hermanos?
2. Una máquina produce pernos roscables; para control de calidad se toma una muestra de 20 pernos y se mide su diámetro (mm). Los datos que se obtienen son:

5.2	5.0	5.1	5.3	5.0	5.2	5.1	5.4	5.0	5.2
5.3	5.1	5.2	5.0	5.4	5.1	5.3	5.2	5.0	5.1

- Elaborar una tabla de frecuencias sin agrupar
 - ¿Cuál es el diámetro más frecuente que se puede encontrar?
 - ¿Qué porcentaje de tornillos tiene el diámetro ≤ 5.2 mm?
3. En la revisión vehicular se midieron las emisiones de CO₂ (g/km) de 15 vehículos con motor diésel, obteniendo los siguientes resultados:

120	125	130	125	135	120	130	125	140	135
125	130	120	135	130					

- a. Construya una tabla de frecuencias
- b. ¿Cuál es la emisión más frecuente que se encuentran en la revisión vehicular?
- c. ¿Qué porcentaje de vehículos emiten CO₂ menor o igual a 130 (g/km)?

4. Se midió el error de precisión en milímetros de 12 brazos robóticos de diferentes marcas, obteniendo los siguientes datos.

0.5	0.7	0.6	0.5	0.8	0.7	0.6	0.5	0.9	0.7
0.5	0.6								

- a. Elaborar una tabla de frecuencias
- b. ¿Cuál es el error de precisión más frecuente (en milímetros) de los brazos robóticos?
- c. ¿Qué porcentaje de los brazos robóticos tienen un error de precisión menor o igual a 0.7 mm?
- d. ¿Cuántos brazos robóticos tienen un error de precisión mayor a 0.8 mm?
- e. ¿Cuál es la frecuencia absoluta acumulada para un error de precisión de 0.6 mm?

5. En el departamento de contabilidad se registran los gastos mensuales en dólares de una empresa durante 10 meses. En donde se registraron los siguientes resultados.

1200	1500	1300	1200	1800	1500	1300	1200	1600	1500
------	------	------	------	------	------	------	------	------	------

- a. Elaborar una tabla de frecuencias.
- b. ¿Cuáles son los gastos mensuales más comunes registrados en la empresa?
- c. ¿Qué porcentaje de los meses tiene un gasto mensual menor o igual a 1500 dólares?

- d. ¿Cuál es el gasto mensual con menor frecuencia registrada en la tabla?
- e. ¿Qué porcentaje de meses tuvo gastos < 1600 USD?

6. En una empresa se registró la edad (en años completos) de sus empleados, obteniendo lo siguientes registros

31	49	36	39	56	29	57	41	40	51
45	60	40	39	47	27	36	37	16	37
51	18	29	34	42	38	62	31	28	25
36	40	46	37	49	25	21	39	35	37
56	35	48	44	42	43	49	22	25	28

- a. Elaborar una tabla de frecuencias agrupada
- b. ¿Cuántos empleados tienen una edad menor a 35 años?
- c. ¿Qué porcentaje de empleados tiene una edad mayor a 43 años?
- d. ¿Cuál es la frecuencia relativa (fr) de la clase de edades 35.5 - 42.5 años?
- e. ¿Cuántos empleados tienen una edad entre 42.5 y 56.5 años?

7. En el laboratorio de Ingeniería Automotriz, se evaluó el consumo de combustible de 30 vehículos en condiciones de ruta normal. Los datos que se registran en (km/L) son los que se detallan a continuación:

12.1	13.5	14.0	11.8	15.2	12.7	13.9	14.5	11.2	15.8
13.1	11.5	15.0	12.9	13.7	14.8	11.0	15.5	12.5	13.3
14.1	11.9	12.8	13.6	14.7	11.4	15.1	14.2	12.3	15.3

- a. Construir una tabla de frecuencia agrupada
- b. ¿Qué porcentaje de vehículos tiene el mayor consumo de combustible?
- c. ¿Cuántos vehículos tiene un consumo de combustible menor a 13.4 km/L?
- d. ¿Cuántos vehículos tienen un consumo de combustible mayor o igual a 13.4 km/L?

8. Los registros de la inversión anual, en miles de dólares, de una muestra de 40 empresas fueron:

36	19	29	37	33	22	29	31	21	35
20	15	25	34	24	27	27	24	26	31
27	17	31	12	28	15	41	30	18	39
46	26	12	23	18	33	25	28	23	28

- Construya una tabla de frecuencias
 - ¿Cuál es la clase de inversión anual con la mayor frecuencia absoluta?
 - ¿Qué porcentaje de empresas tienen una inversión anual mayor o igual a 29 mil dólares?
 - ¿Qué porcentaje de empresas tienen una inversión anual menor a 23 mil dólares?
9. En la siguiente tabla de frecuencias se muestran los tiempos de espera registrados en la estación de autobuses.

Tiempo (min)	Frecuencia Absoluta	Frecuencia Relativa
0 - 3	32	f_{r1}
3 - 6	f_2	0.30
6 - 9	f_3	f_{r2}
9 - 12	8	0.05
12 - 15	f_5	0.10
Total	n	

- Complete la tabla de distribución de frecuencias.
- ¿Cuál es el tamaño de muestra seleccionado?
- ¿Cuál es el intervalo de tiempo con más frecuencia que los pasajeros tienen que esperar?
- ¿Qué porcentaje de personas tiene un tiempo de espera menor o igual a 6 minutos?
- ¿Cuántas personas tienen un tiempo de espera mayor a 9 minutos?

10. Una empresa de servicios contables registró los gastos operativos mensuales (en dólares) de 60 compañías, los mismos que se registran a continuación.

120	350	480	220	610	730	890	150	420	540
310	680	790	250	380	510	650	950	180	440
570	700	820	290	530	670	900	130	360	490
630	750	200	330	460	720	850	170	400	520
660	780	240	370	500	770	910	160	390	510
620	880	210	340	470	410	590	930	640	740

- Elaborar una tabla de frecuencias agrupada tipo II.
- ¿Cuál es el intervalo de gastos con más frecuencia en las compañías?
- ¿Cuántas empresas tienen mayor gasto?
- ¿Cuántas empresas tienen un gasto operativo menor a USD 594.5?

Ejercicio propuesto (con base de datos)

Usando la base de datos del Examen Ser Bachiller la cual se puede encontrar en el siguiente link.



Aplicar los conocimientos aprendidos en el capítulo II.

- Abrir el archivo descargado anteriormente (SEST21_Micro_16595_20221209_SAV)
- Realizar una tabla de frecuencias tipo I de la variable sostenimiento (Sostenimiento de la institución educativa) e interpretar la misma
- Realizar una tabla de frecuencias tipo II (Agrupada por intervalos) de la variable imat (Promedio obtenido en el campo de Matemática) e interpretar la misma



<https://n9.cl/x832j>

CAPÍTULO III

Métodos gráficos

3.1. Gráficos para el análisis de variables cualitativas

3.1.1. Diagrama simple de barras

Consiste en representar los datos mediante barras verticales u horizontales, los anchos de cada barra son iguales y la longitud de cada una de ellas varía según la frecuencia de cada observación. Su construcción es muy sencilla, el diagrama de barras muestra el comportamiento de una variable y su tendencia a lo largo del tiempo y facilita la comparación de datos ubicados en diferentes categorías. Si el diagrama se satura con una gran cantidad de barras, lo mejor será reemplazarlo por un diagrama de líneas.

3.1.2. Diagrama múltiple de barras

El diagrama de barras múltiple está compuesto por un conjunto de barras agrupadas según las categorías formadas y compara las mismas en función de diversas variables a lo largo del tiempo.

3.1.3. Diagrama de barras compuesto

El diagrama de barras compuesto también recibe el nombre de diagrama de barras apiladas, cada barra representa un conjunto de datos apilados uno encima de otro, se compone de segmentos divididos por colores y su longitud muestra el tamaño total de la categoría, adicional compara un conjunto de variables e indica el promedio o el total de cada categoría.

Ejercicio práctico N°1

Utilizar un diagrama de barras para analizar la información correspondiente al grupo sanguíneo de 30 pacientes.

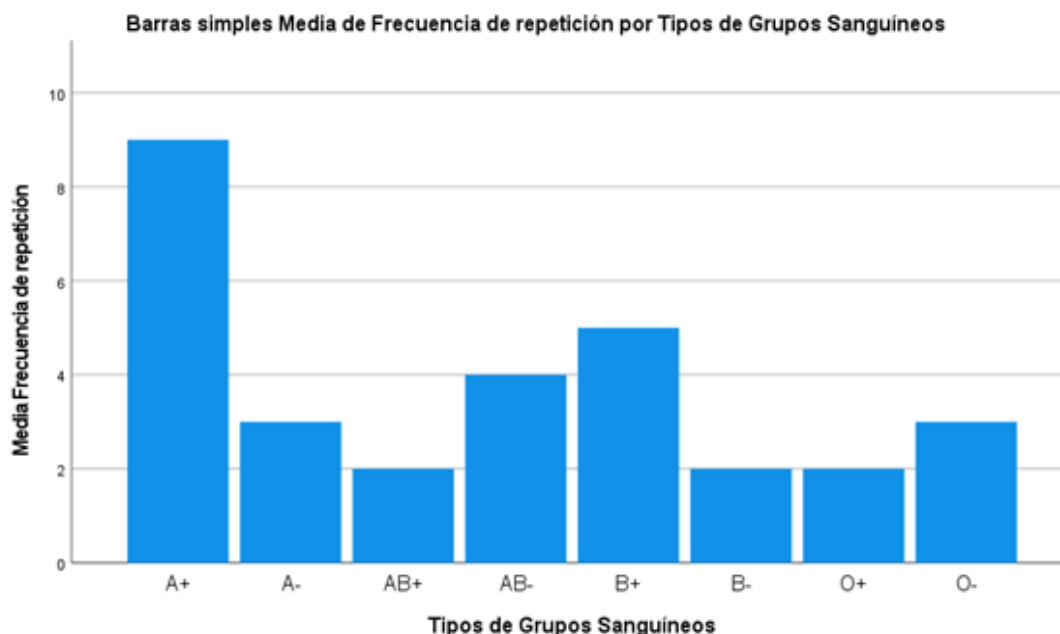
Grupos Sanguíneos					
A+	AB-	B+	O+	O-	AB-
A+	O+	O-	A+	O-	A-
A+	AB+	B+	AB+	A+	B+
A+	B-	A-	AB-	A+	A-
B+	AB-	A+	B+	A+	B-

Construcción en SPSS:

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** Tipos
 - **Tipo:** Numérico
 - **Decimales:** Cero
 - **Etiqueta:** Tipos de grupos sanguíneos
 - **Valores:** Asignar el valor de 1 a la categoría “A+”, 2 a la categoría “A-”, 3 a la categoría “AB+” y así sucesivamente hasta el valor 8 a la categoría “O-”
 - **Medida:** Ordinal
-
- **Nombre:** Datos
 - **Tipo:** Numérico
 - **Decimales:** Cero
 - **Etiqueta:** Frecuencia de repetición
 - **Medida:** Escala



4. Interpretación

La investigación indica que 9 pacientes tienen el grupo sanguíneo A+, 5 pacientes tienen el grupo sanguíneo B+, 4 pacientes tienen el grupo sanguíneo AB-, el A+ y el O- tienen 3 pacientes cada uno y el grupo sanguíneo que tiene la menor cantidad de pacientes es el grupo AB+, B+, O+ con 2 pacientes cada uno. Siendo el tipo sanguíneo A+ el grupo predominante en la investigación realizada.

Ejercicio práctico N°2

En la siguiente tabla se muestra el número de estudiantes matriculados para primer nivel en algunas Carreras de la Universidad y su distribución por género. Interprete la información mediante de un Diagrama de Barras.

Carreras de la Universidad	Matriculados	Hombres	Mujeres
Ingeniería Mecatrónica	125	72	53
Ingeniería Electromecánica	100	61	39
Ingeniería Automotriz	154	57	97
Ingeniería en Petroquímica	66	33	33
Ingeniería en Software	71	33	38

Construcción en SPSS

1. Prepare los datos:

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** Carreras
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Carreras que oferta la Universidad
- **Valores:** asignar el valor de 1 a la categoría “Mecatrónica”, 2 a la categoría “Electromecánica” y así sucesivamente hasta el valor 5 a la categoría “Software”
- **Medida:** Nominal
- **Nombre:** Matriculados
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Estudiantes matriculados por Carrera
- **Medida:** Escala
- **Nombre:** Género
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Género de los estudiantes
- **Valores:** asignar el valor de 1 a la categoría “Hombres” y 2 a la categoría “Mujeres”
- **Medida:** Nominal

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Carreras	Numérico	8	0	Carreras que oferta la Universidad	(1, Mecatrón...	Ninguno	8	Centro	Nominal	Entrada
2	Matriculados	Numérico	8	0	Estudiantes matriculados por Carrera	Ninguno	Ninguno	8	Centro	Escala	Entrada
3	Género	Numérico	8	0	Género de los estudiantes	(1, Hombres...	Ninguno	8	Centro	Nominal	Entrada

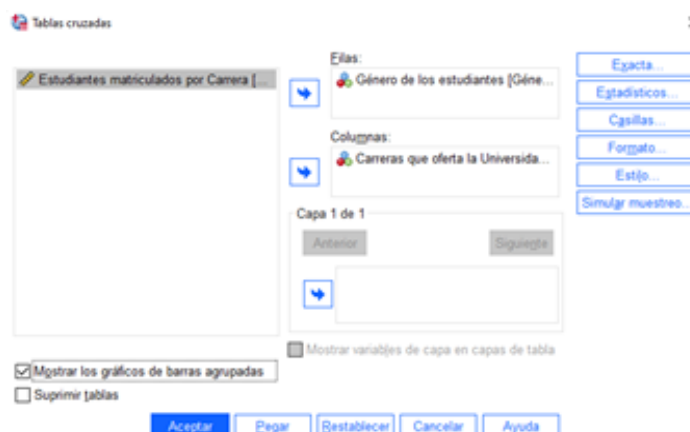
- c. En la ventana “Vista de datos” ingrese las categorías correspondientes a las Carreras y al Género, así como también ingresar el numérico de alumnos Matriculados en distintas carreras. (En la pantalla capturada se muestran 22 de los 516 casos)

	Carreras	Matriculados	Género
1	Mecatrónica	1	Hombres
2	Mecatrónica	1	Hombres
3	Mecatrónica	1	Hombres
4	Mecatrónica	1	Hombres
5	Mecatrónica	1	Hombres
6	Mecatrónica	1	Hombres
7	Mecatrónica	1	Hombres
8	Mecatrónica	1	Hombres
9	Mecatrónica	1	Hombres
10	Mecatrónica	1	Hombres
11	Mecatrónica	1	Hombres
12	Mecatrónica	1	Hombres
13	Mecatrónica	1	Hombres
14	Mecatrónica	1	Hombres
15	Mecatrónica	1	Hombres
16	Mecatrónica	1	Hombres
17	Mecatrónica	1	Hombres
18	Mecatrónica	1	Hombres
19	Mecatrónica	1	Hombres
20	Mecatrónica	1	Hombres
21	Mecatrónica	1	Hombres
22	Mecatrónica	1	Hombres

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Tablas cruzadas

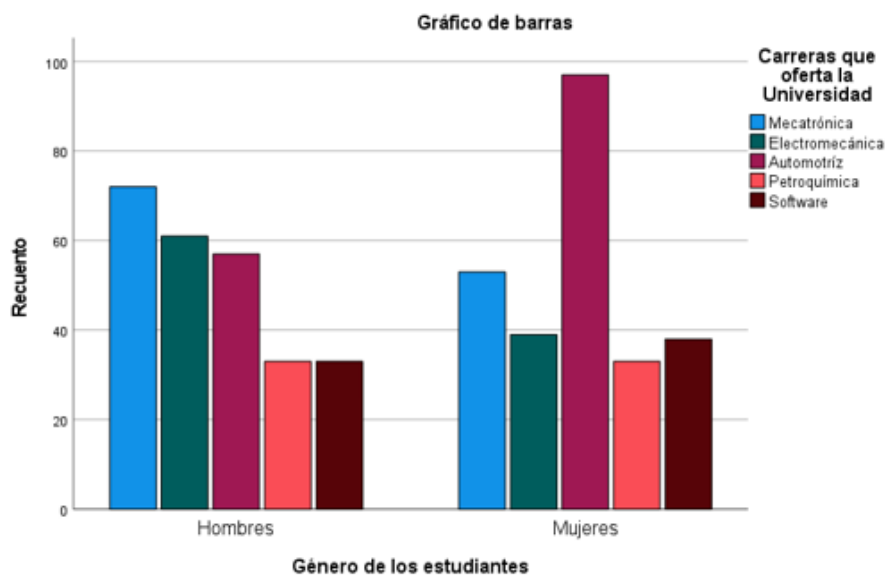
- En la ventana que aparece seleccione la variable Género a Filas y la variable Carreras a Columnas
- De un clic en la opción “Mostrar los gráficos de Barras Agrupadas”
- Presione Aceptar



3. Resultados obtenidos

En la Ventana de Resultados, encontrará la Tabla Cruzada entre las variables Género y Carreras, así como también el Diagrama de Barras generado.

Tabla cruzada Género de los estudiantes*Carreras que oferta la Universidad							
Recuento		Carreras que oferta la Universidad					
Género de los estudiantes	Hombres	Mecatrónica	Electromecánica	Automotriz	Petroquímica	Software	Total
		72	61	57	33	33	256
	Mujeres	53	39	97	33	38	260
Total		125	100	154	66	71	516



Otra forma de obtener el Diagrama de Barras es en:

En el menú principal seleccione: Gráficos / Generador de Gráficos

- Selecione la opción Barras agrupadas y arrastre a la ventana de edición
- Arrastre la variable Género al Eje X y la variable Carreras a la opción "agrupar en x, establecer color"



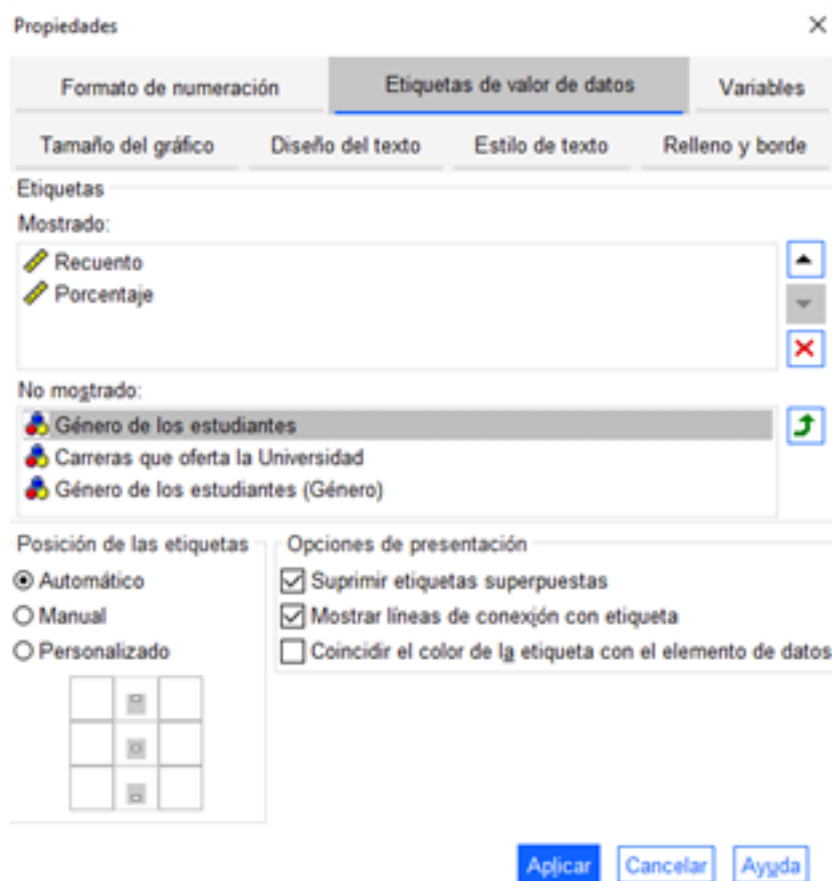
c. Presione Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el Diagrama de Barras generado

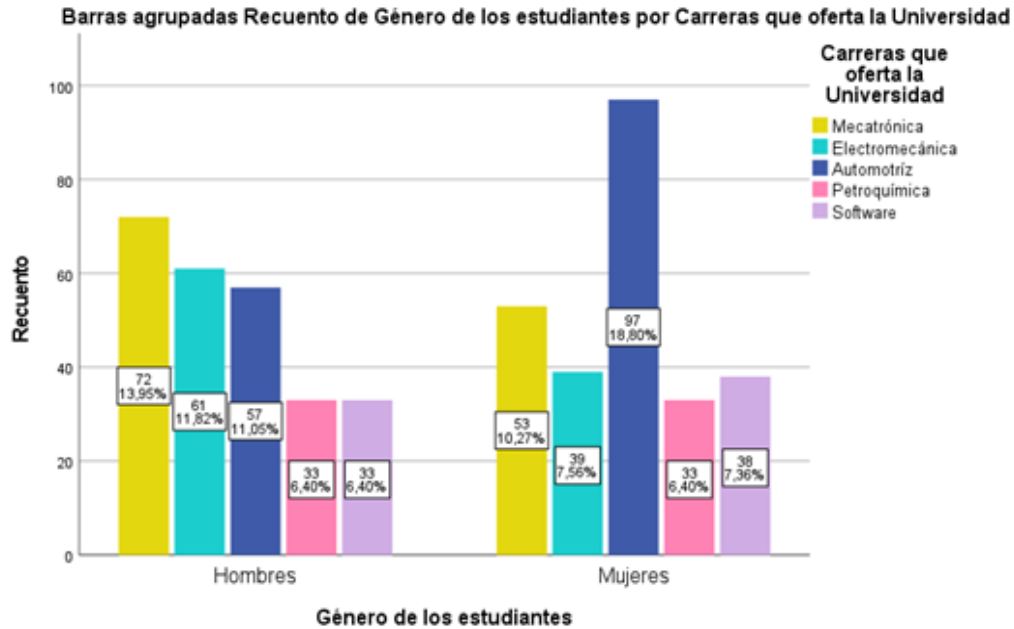
Se puede editar el gráfico obtenido con las siguientes opciones:

- a. Para editar las barras, presionar dos veces un clic izquierdo en cualquier parte del gráfico, dé un clic derecho en una de las barras y tomar la opción Mostrar etiqueta de datos. En la opción Mostrado se encuentra activado la opción Recuento y en No mostrado tomar la opción Porcentaje, esto permitirá observar la frecuencia y el porcentaje respectivo de cada uno de los valores.

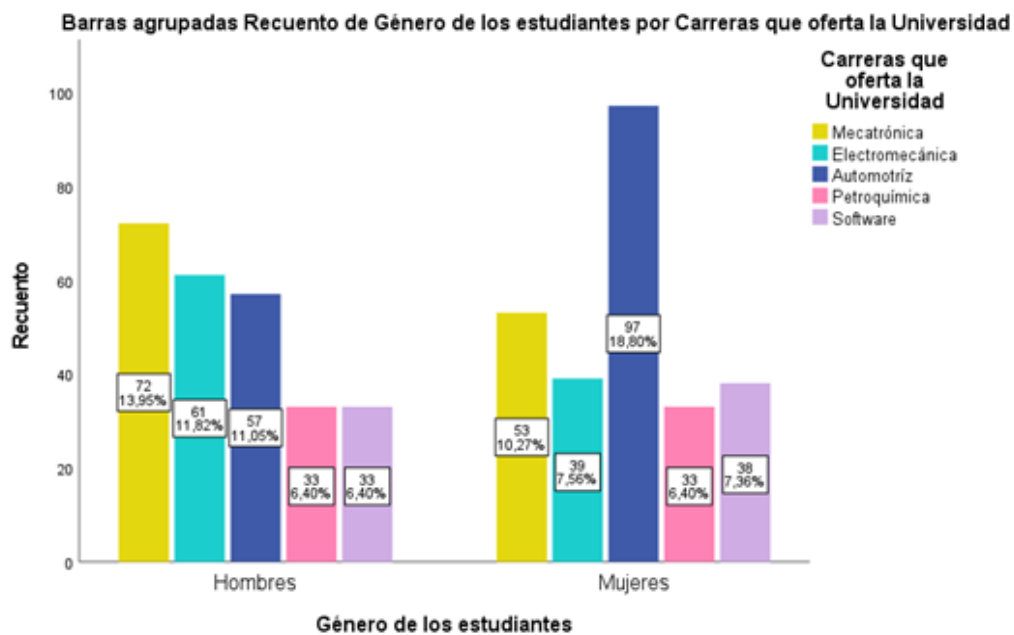


b. Presione Aplicar y luego Cerrar

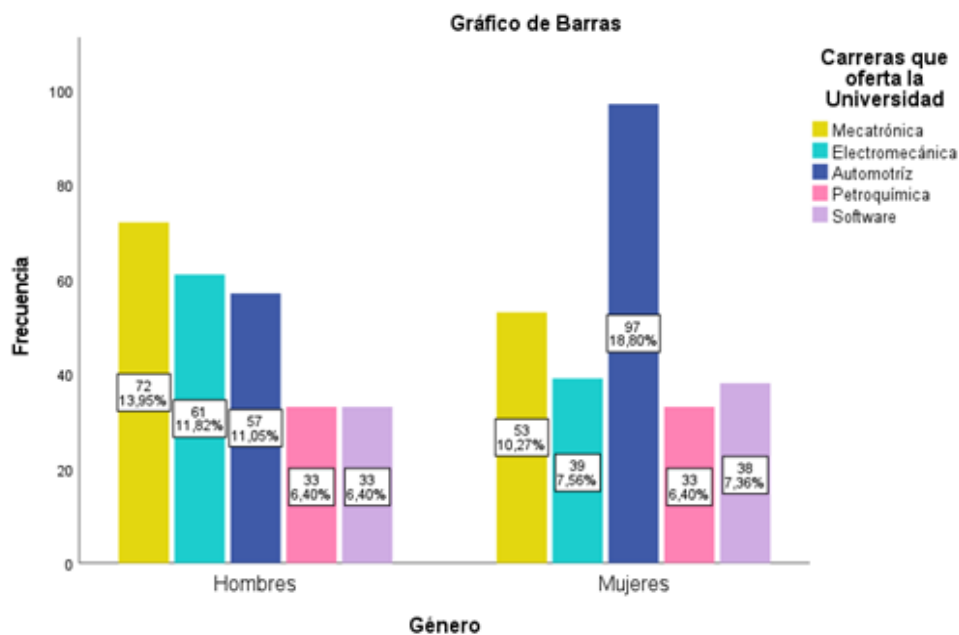
- c. Se puede también modificar el color de las barras realizando nuevamente un doble clic en cada una de las barras y en la ventana de propiedades / relleno de borde.



- d. Se puede también ocultar las líneas de división con un doble clic en la figura, clic derecho en las líneas de división y seleccionar Quitar líneas de cuadrícula.



- e. Se puede modificar el texto de los ejes solo con dar doble clic izquierdo en la figura y doble clic en el texto a modificar.



4. Interpretación

La carrera que más agrado tiene en las mujeres es la Ingeniería Automotriz con un 18.80% de inscritos, mientras que la de menor aceptación tiene es la carrera de Ingeniería en Petroquímica con un 6.40% de inscritos.

La carrera que más agrado tiene en los hombres es la carrera de Ingeniería en Mecatrónica con un 13.95% de inscritos mientras que la carrera que menos les gusta es la carrera de Ingeniería en Software con un 6.40%.

3.1.4. Diagrama circular

También llamado gráfico de sectores o diagrama de pastel, se utiliza para mostrar la proporción o sector que le corresponde a cada categoría, los datos se representan en forma porcentual y la suma de todos los sectores da un total de 100%. En el diagrama circular se representan las categorías más grandes, en el caso de haber categorías pequeñas y a medida que el estudio lo permita juntar todas estas categorías en una sola categoría llamada Otros.

Ejercicio práctico N°3

Construya un gráfico circular con información que contenga las Provincias del Oriente Ecuatoriano y las superficies correspondientes a cada una de ellas en km².

Provincias	Superficie (km ²)
Sucumbíos	18084
Napo	12513
Orellana	21601
Pastaza	29514
Morona Santiago	25690
Zamora Chinchipe	10559

Construcción en SPSS

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Provincias con los siguientes parámetros:

- **Nombre:** Provincias
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Provincias del Oriente Ecuatoriano
- **Valores:** asignar el valor de 1 a la categoría “Sucumbíos”, 2 a la categoría “Napo”, 3 a la categoría “Orellana” y así sucesivamente hasta el valor 6 a la categoría “Zamora Chinchipe”
- **Medida:** Nominal

- **Nombre:** Superficie
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Superficie en km²
- **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Provincias	Númérico	8	0	Provincias del ...	{1, Sucumb...	Ninguno	15	Centro	Nominal	Entrada
2	Superficie	Númérico	8	0	Superficie en k...	Ninguno	Ninguno	8	Centro	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese las Provincias y las Superficies en km² correspondientes a cada una de ellas.

	Provincias	Superficie
1	Sucumbios	18084
2	Napo	12513
3	Orellana	21601
4	Pastaza	29514
5	Morona Santiago	25690
6	Zamora Chinchipe	10559

2. Análisis descriptivo:

En el menú principal seleccione: Gráficos / Generador de Gráficos

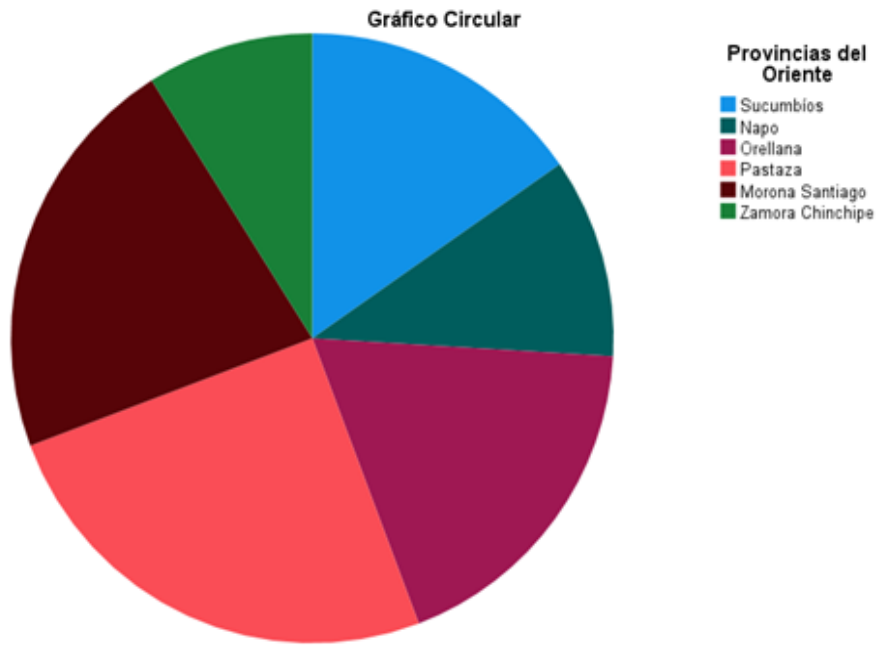
- a. Seleccionar la opción Circular/Polar y mueva el Gráfico circular a la vista previa. En la opción “¿Porciones por?” coloque la variable Provincias del Oriente y en la opción “Recuento” coloque la variable Superficie.



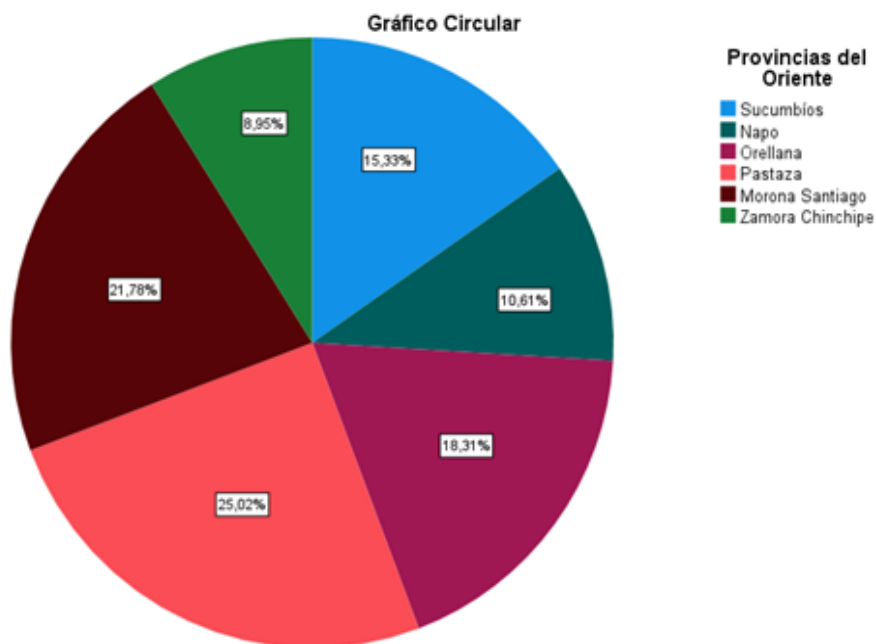
b. Presione el botón Aceptar

3. Resultados obtenidos

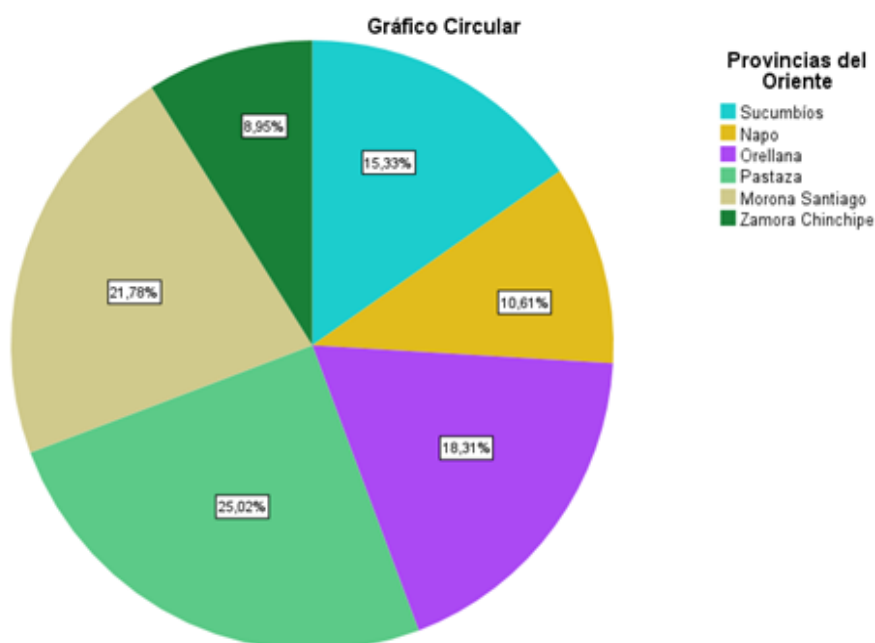
En la Ventana de Resultados, encontrará el gráfico circular generado



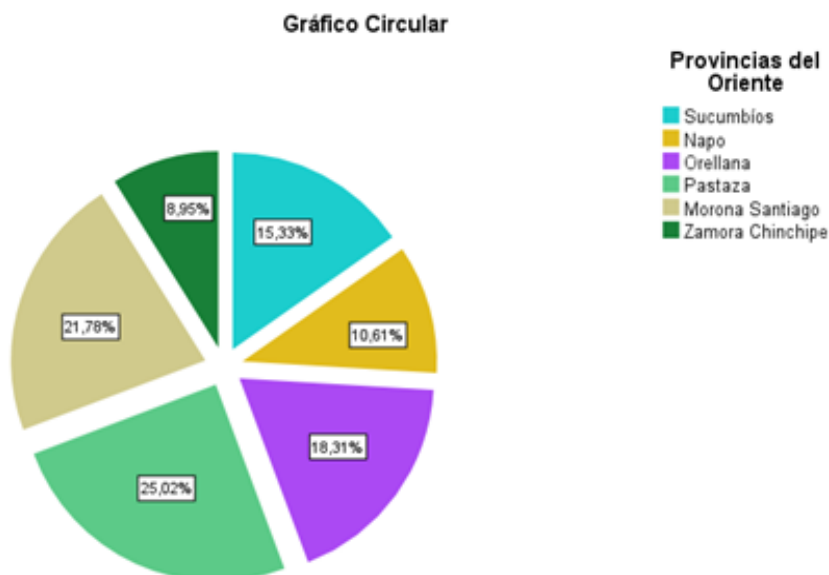
El gráfico obtenido se puede editar realizando doble clic izquierdo en la zona gráfica y luego un clic derecho, tomar la opción Mostrar etiquetas de datos y seleccionar la opción porcentaje.



Se puede cambiar los colores de cada sector, seleccionando por separado cada uno de ellos y escogiendo el color a su gusto.



También se puede crear separaciones entre los sectores del diagrama para generar un gráfico mucho más vistoso, para ello se selecciona el gráfico con doble clic izquierdo y luego clic derecho para tomar la opción desgajar porción.



4. Interpretación

Pastaza es la provincia del Oriente que mayor superficie tiene con un total 29514 km² correspondiente al 25.02%, le sigue la provincia de Morona Santia-

go con una superficie de 25690 km² correspondiente a un 21.78%, la provincia de Orellana con una superficie de 21601 km² correspondiente a 18.31%, la provincia de Sucumbíos con una superficie de 18084 correspondiente a 15.33%, la provincia de Napo con una superficie de 12513 km² correspondiente a 10.61% y finalmente la provincia de Zamora Chinchipe con una superficie de 10559 correspondiente a 8.95%.

Ejercicio práctico N°4

Se realizó una encuesta a 100 personas y se les preguntó: ¿Cuál es el nivel de satisfacción de un cliente al hacer uso de los servicios que brinda una empresa pública?, se obtuvieron las siguientes respuestas:

Nivel de satisfacción	N° personas
Muy satisfecho	20
Satisfecho	30
Neutral	25
Insatisfecho	15
Muy insatisfecho	10

Construya un gráfico circular e interprete sus resultados

Cálculos previos:

$$\text{Muy satisfecho: } \frac{20}{100} * 100\% = 20\%$$

$$\text{Satisfecho: } \frac{30}{100} * 100\% = 30\%$$

$$\text{Neutral: } \frac{25}{100} * 100\% = 25\%$$

$$\text{Insatisfecho: } \frac{15}{100} * 100\% = 15\%$$

$$\text{Muy insatisfecho: } \frac{10}{100} * 100\% = 10\%$$

Construcción en SPSS

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** Satisfacción
 - **Tipo:** Numérico
 - **Decimales:** 0
 - **Etiqueta:** Nivel de Satisfacción
 - **Valores:** asignar el valor de 1 a la categoría “Muy Satisfecho”, 2 a la categoría “Satisfecho” y así sucesivamente hasta el valor 5 a la categoría “Muy insatisfecho”
 - **Medida:** Ordinal
-
- **Nombre:** Datos
 - **Tipo:** Numérico
 - **Decimales:** 0
 - **Etiqueta:** Número de personas encuestadas
 - **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Satisfacción	Numérico	8	0	Nivel de Satisfacción	(1, Muy Sati...	Ninguno	8	Centro	Ordinal	Entrada
2	Datos	Numérico	8	0	Número de personas encuestadas	Ninguno	Ninguno	8	Centro	Escala	Entrada

- En la ventana “Vista de Datos” ingrese el nivel de satisfacción y el número de personas encuestadas que respondieron a cada nivel respectivamente.

Archivo Editar Ver Datos Transformar		
	Satisfacción	Datos
1	Muy Satisfecho	20
2	Satisfecho	30
3	Neutral	25
4	Insatisfecho	15
5	Muy insatisfecho	10

2. Análisis descriptivo:

En el menú principal seleccione: Gráficos / Generador de Gráficos

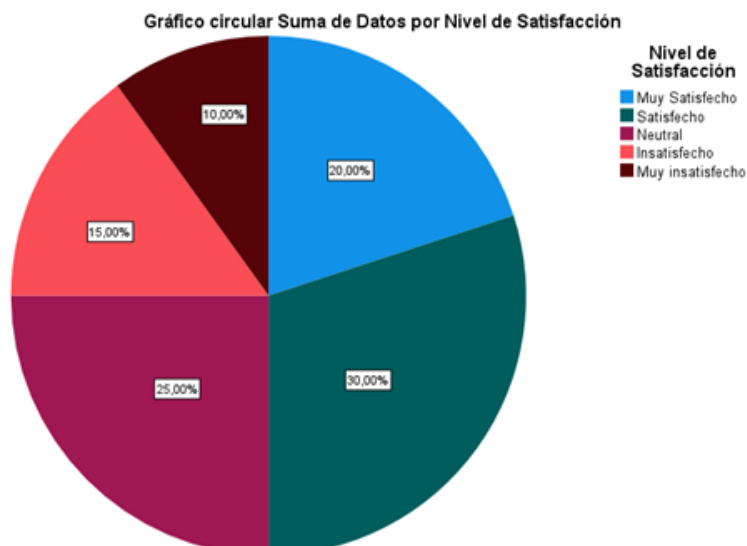
- Seleccionar la opción Circular/Polar y mueva el Gráfico circular a la vista previa. En la opción “¿Porciones por?” coloque la variable Satisfacción y en la opción “Recuento” coloque la variable Datos.



- Presione el botón Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el gráfico circular generado



4. Interpretación:

A la pregunta realizada sobre el nivel de satisfacción de un cliente al hacer uso de los servicios que brinda una Empresa Pública se puede indicar que un 20% de personas mencionaron que están muy satisfechos con los servicios que brinda la empresa, el 30% de personas dicen estar satisfechas, el 25% de personas prefieren ser neutrales y existe un 15% y 10% de personas que dicen estar insatisfechas y muy insatisfechas respectivamente.

Existe un mayor porcentaje de personas encuestadas que manifiestan estar satisfechas con los servicios que brinda la empresa pública.

Autoevaluación 3.1.

1. ¿Cuál es la diferencia básica entre un diagrama de barras simple, diagrama de barras múltiple y un diagrama de barras compuesta.?
2. Si cuentas con la información de 10 laboratorios clínicos que funcionan en el sector centro de la ciudad de Latacunga y quieres comparar la frecuencia de visita a cada uno de ellos, qué diagrama sería el más adecuado para presentar esta información: un diagrama circular o un diagrama de barras y porqué.
3. Se realizó una encuesta a 50 adultos mayores sobre las enfermedades más comunes en su vejez. Los resultados fueron: 20 sufren de diabetes, 15 sufren artritis, 10 sufren Alzheimer y 5 sufren de depresión.
 - Calcula la frecuencia relativa (en porcentaje) para cada enfermedad
 - Con estos datos realiza un diagrama de barras

3.1.5. Ejercicios propuestos

1. Una empresa de electrodomésticos registra sus ventas mensuales en dólares:

Mes	Ventas (USD)
Enero	5313
Febrero	6354
Marzo	6120
Abril	6932
Mayo	8215

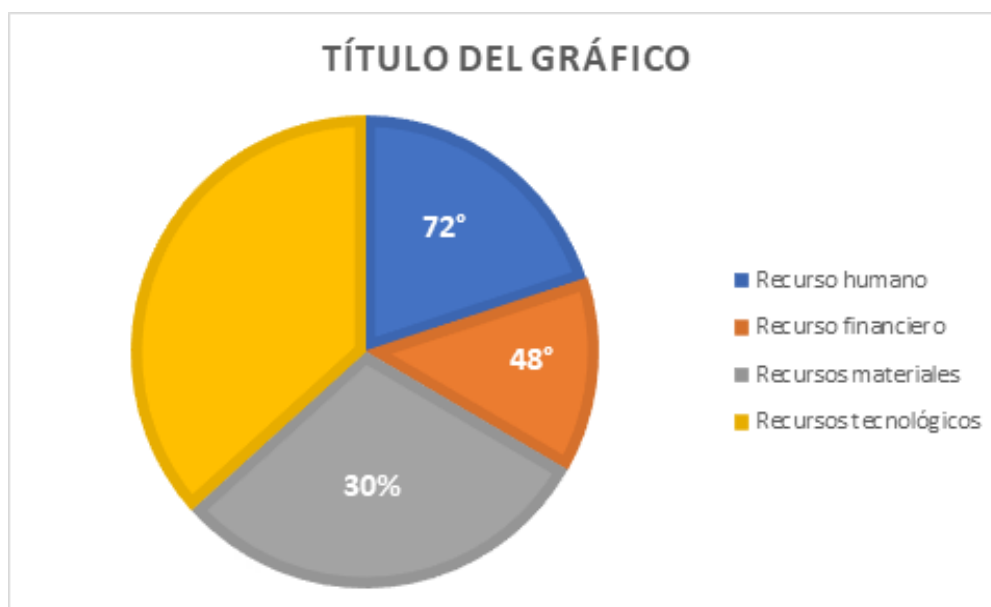
Grafique un diagrama de barras e interprete

2. Se registra la producción en una fábrica de textiles de la ciudad durante tres meses

Mes	Fábrica A	Fábrica B
Enero	4	2
Febrero	5	4
Marzo	6	3

Grafique un diagrama de barras compuesta e interprete la información

3. Entre los trabajadores de una empresa se realizó una encuesta para determinar cuál es el recurso más importante para llevar a cabo un proyecto. Cada trabajador señaló un recurso, formándose el siguiente diagrama de sectores:



Si 33 personas indicaron que los recursos tecnológicos son los más importantes para llevar a cabo el proyecto, ¿cuántos eligieron los recursos materiales o los recursos humanos?

4. La calificación que dieron 20 clientes sobre el servicio de Catering en un evento realizado por la municipalidad de la ciudad es el siguiente:

Excelente, Bueno, Bueno, Excelente, Malo, Bueno, Bueno, Excelente, Excelente, Bueno, Excelente, Bueno, Bueno, Malo, Bueno, Bueno, Bueno, Malo, Bueno, Bueno.

Construya el gráfico de sectores circulares e interprete la información.

5. Una encuesta realizada al azar a 30 jóvenes entre hombres y mujeres de un prestigioso colegio de la ciudad sobre la red social que más utiliza es el siguiente:

Red Social	N° Estudiantes	
	Hombres	Mujeres
Facebook	3	6
WhatsApp	5	2
TikTok	7	7

- a. Grafique un diagrama de barras
- b. ¿Qué porcentaje del presupuesto total de la institución educativa representan las categorías “Laboratorios” y “Equipos de cómputo”?
- c. ¿De los cinco principales rubros de inversión, ¿qué porcentaje del total representan “Laboratorios” y “Equipos de cómputo”?

7. Se entrevistaron a 100 personas sobre su intención de voto para las elecciones presidenciales. La pregunta específica era “Votaría Ud. por el candidato de la derecha”, a lo que los votantes respondieron:

Sí: 60

No: 30

No contestan: 10

- a. En la construcción del gráfico circular, ¿cuántos grados le corresponderían del diagrama circular a la respuesta Sí?
- b. ¿Cuántos grados le corresponderían del diagrama circular a la respuesta No?
- c. Construya una gráfica de pastel

8. En una fábrica de componentes electrónicos se reciben lotes para el ensamblaje de circuitos. Para facilitar su identificación, los componentes vienen en diferentes colores. El último lote de componentes entregado a la fábrica es el siguiente:

- 46 microchips con carcasa negra
- 50 leds de color azul
- 75 transistores con encapsulado café
- 60 diodos de color rojo
- 95 resistencias con bandas verdes

Elabore una gráfica que describa esta información y redacte un párrafo que describa los resultados.

9. Una empresa de desarrollo de software ha utilizado los siguientes lenguajes de programación en diversos proyectos durante los últimos tres años 2022, 2023, 2024. La cantidad de proyectos con los diversos lenguajes de programación utilizados en cada año son los siguientes:

Lenguajes de programación	2022	2023	2024
Python	5	11	12
Java	9	12	19
C++	7	11	15
SQL	6	9	11
HTML	6	8	10

- ¿Qué lenguaje de programación fue el más utilizado conforme pasan los años?
 - ¿Cuál es el lenguaje menos utilizado en el año 2023?
10. El primer nivel de la carrera de Ingeniería en Electrónica tiene un total de 32 estudiantes entre hombres y mujeres, los cuales vienen de distintas ciudades tal como se muestra la siguiente tabla

Ciudad	N° Estudiantes
Latacunga	12
Quito	6
Ambato	7
Riobamba	5
Quevedo	2

Realice un diagrama circular y exprese en porcentajes el numérico de estudiantes por cada ciudad

3.2. Gráficos para el análisis de variables cuantitativas

3.2.1. Histogramas

Un histograma es la representación gráfica de una variable continua en forma de barras unidas del mismo ancho, se utiliza para comparar gran cantidad de información, identificar patrones y tendencias, así como obtener conclusiones significativas de manera rápida y efectiva. Lo usual es emplear entre 5 y 15 barras, aunque esto dependerá de la serie de datos que se va a analizar, usar pocas barras hará que se pierdan detalles en el análisis y utilizar muchas barras dificultará ver el patrón de observación.

Los aspectos más importantes para interpretar un histograma son:

1. Centrado del proceso

Se define en qué lugar se encuentra la posición central del Histograma, el promedio o valores típicos de los valores centrales y dónde se espera que se encuentren la mayoría de las observaciones.

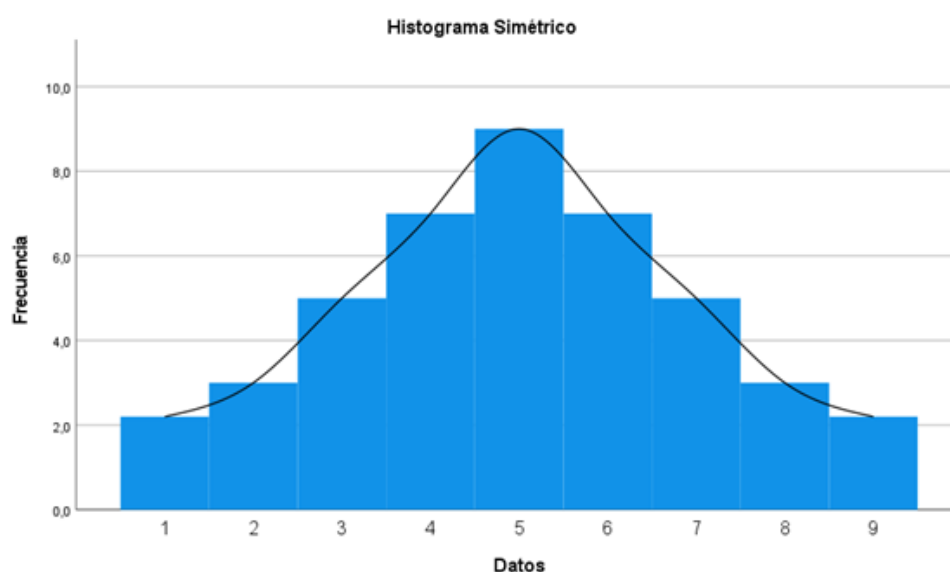
2. Variabilidad

Se analiza el rango de variación, es decir en qué valor inicia y en qué valor finaliza la variable numérica.

3. Forma del Histograma

La forma del histograma es Simétrica cuando los datos se distribuyen uniformemente alrededor de la media y se ajustan a una distribución normal en forma de campana. Se considera un proceso estable (mayor concentración de los datos entorno a la media).

Figura 26

Histograma simétrico

Si la forma es asimétrica los datos no se distribuyen uniformemente, la calidad de un proceso no es buena y eso admite un análisis de calidad. La forma del histograma tiene una asimetría a la derecha, en este caso los valores más cercanos se encuentran alrededor del pico y la mayor parte de los datos se colocan al costado derecho del histograma, mientras que cuando la forma del histograma tiene una asimetría a la izquierda, los valores más cercanos se encuentran alrededor del pico y la mayor parte de los datos se colocan al costado izquierdo del histograma.

Figura 27

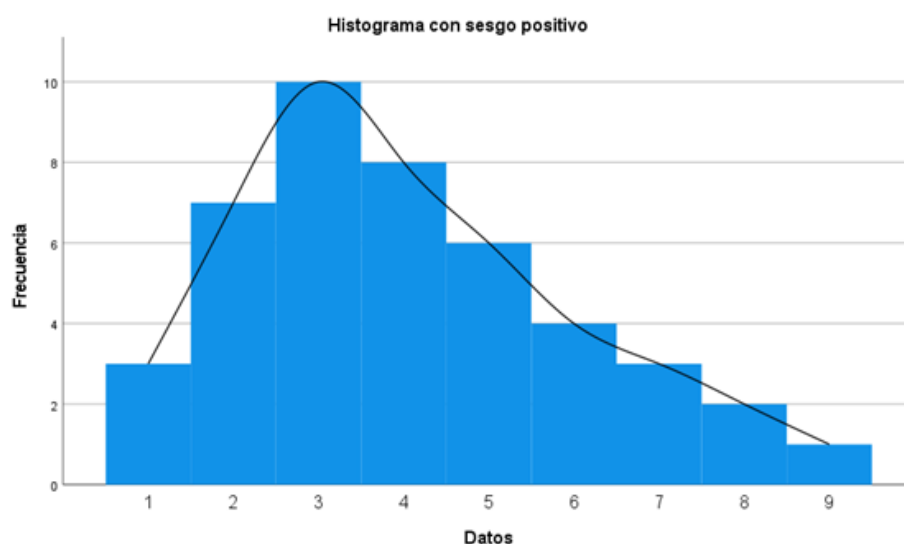
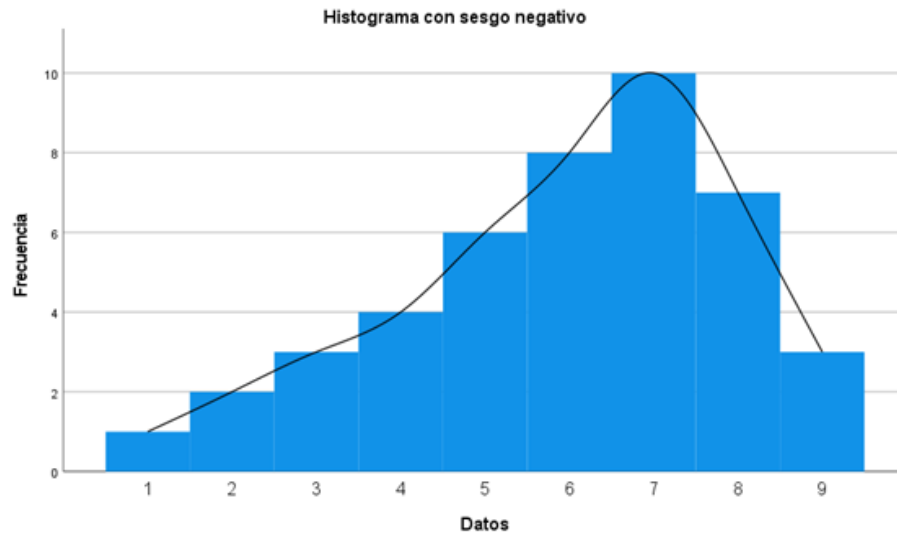
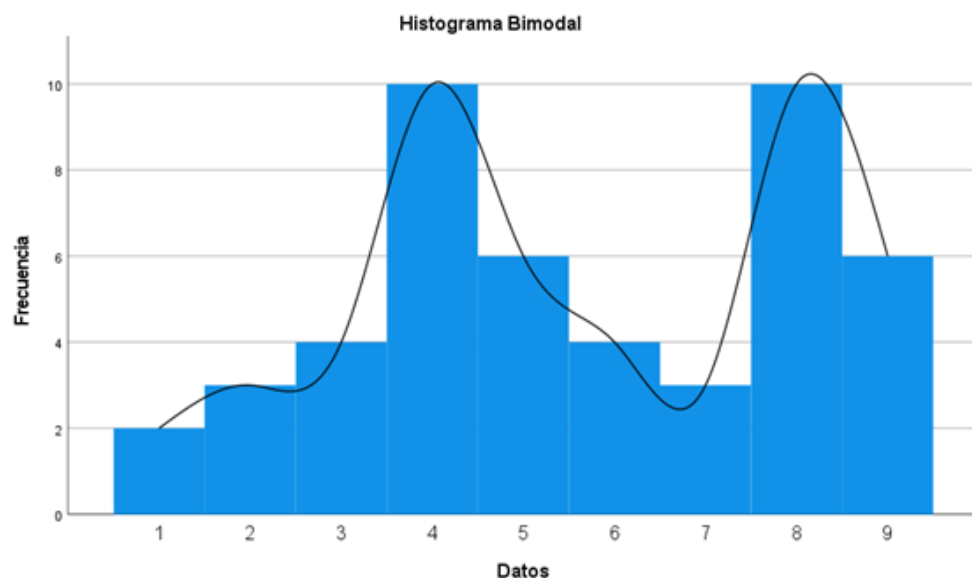
Histograma con sesgo positivo

Figura 28

Histograma con sesgo negativo

Si los datos son bimodales o multimodales, suele encontrarse con una serie de datos de 2 o más picos respectivamente, ocurre cuando los datos se recopilan a partir de más de un proceso, esto significa que existe una variable (que no se visualiza correctamente) que separa en grupos o categorías la variable numérica porque no está fija y es necesario poder identificarla para tratarla por separado, como por ejemplo las partes de un vehículo ensamblados por la fábrica 1 y por la fábrica 2, el número de viajes en el Metro en una semana y el número de viajes realizados el fin de semana, etc.

Figura 29

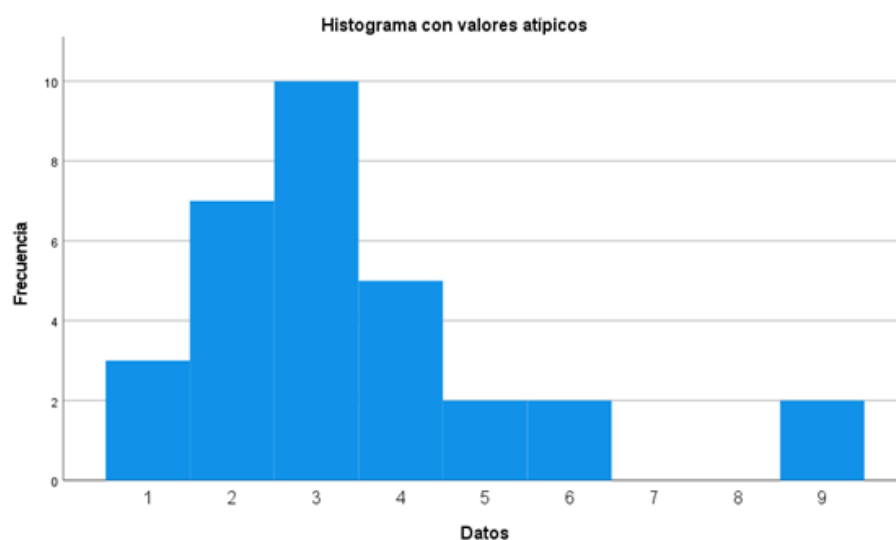
Histograma bimodal

4. Valores atípicos

Son valores inusuales o muy extremos que están bastante dispersos y suceden con poca frecuencia, es importante determinar porqué ocurren estos valores, corregirlos o eliminar aquellos que estén relacionados con eventos fuera de lo normal.

Figura 30

Histograma con valores atípicos



Ejercicio práctico N°1

Se muestran las edades de personas que trabajan en una empresa florícola de la provincia de Cotopaxi. Construya un histograma de frecuencias e interprete.

Intervalos de clase	Marca de clase	Frecuencia
22 – 25	23.5	5
26 – 29	27.5	11
30 – 33	31.5	27
34 – 37	35.5	15
38 – 41	39.5	7

42 – 45	43.5	5
46 – 49	47.5	2
50 - 53	51.5	1

Construcción en SPSS:

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** M_Clase
 - **Tipo:** Numérico
 - **Decimales:** 1
 - **Etiqueta:** Marca de clase
 - **Medida:** Nominal
- **Nombre:** Frecuencia
 - **Tipo:** Numérico
 - **Decimales:** 0
 - **Etiqueta:** Frecuencia de repetición
 - **Medida:** Escala

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda											
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Marca	Númérico	8	1	Marcas de clase	Ninguno	Ninguno	8	Centro	Nominal	Entrada
2	Frecuencia	Númérico	8	0	Frecuencia de repetición	Ninguno	Ninguno	8	Centro	Escala	Entrada

- En la ventana “Vista de datos” ingrese los valores correspondientes a las Marcas de clase y las Frecuencias correspondientes al numérico de Pacientes según la Edad.

	Marca	Frecuencia
1	23,5	5
2	27,5	11
3	31,5	27
4	35,5	15
5	39,5	7
6	43,5	5
7	47,5	2
8	51,5	1

2. Análisis descriptivo:

En el menú principal seleccione: Gráficos / Generador de Gráficos

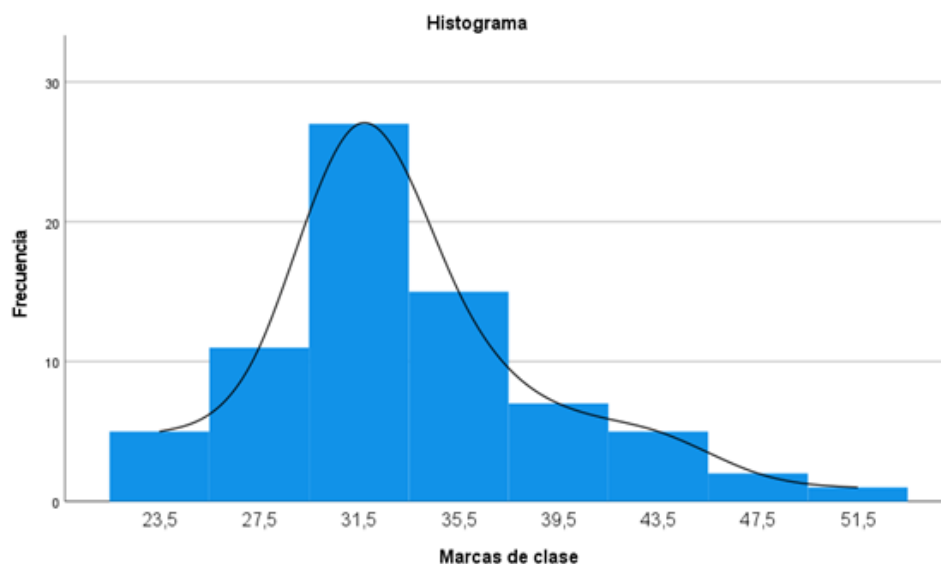
- Seleccionar la opción Histograma y mueva la opción Histograma simple a la vista previa. En el Eje x colocamos las marcas de clase y en el Eje y colocamos la variable Frecuencias



- Presione el botón Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el histograma generado



4. Interpretación

Se puede observar un histograma asimétrico en donde los datos se encuentran sesgados a la derecha, los datos no se distribuyen de manera equitativa para ambos lados de la media y el centro del proceso, así como las frecuencias más altas están a la izquierda de la distribución.

La empresa cuenta con 16 trabajadores menores que 29 años y con al menos 8 trabajadores mayores de 45 años, las edades de los trabajadores varían de 22 a 53 años, no habiendo trabajadores en la empresa con edades fuera del rango normal.

El pico de la distribución se encuentra en el rango de edades de 30 a 33 años, esto indica que la mayor parte de trabajadores de esta empresa se encuentra en este rango de edades.

Ejercicio práctico N°2

Se muestra una cantidad total de 200 departamentos en una zona costera del país, categorizados según su costo de arriendo, analizar dicha información mediante un histograma y determinar si la cantidad de departamentos respecto a su costo de arriendo es una distribución simétrica.

Intervalos de Clase	Marca de clase	Frecuencia
80 – 99	89.5	5
100 – 119	109.5	20
120 – 139	129.5	40
140 – 159	149.5	60
160 – 179	169.5	40
180 – 199	189.5	20
200 – 219	209.5	10
220 – 239	229.5	5
240 – 259	249.5	0

Construcción en SPSS:

1. Prepare los datos:

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** M_Clase
 - **Tipo:** Numérico
 - **Decimales:** 1
 - **Etiqueta:** Marca de clase
 - **Valores:** asignar el valor de 1 a la categoría “80 a 99”, 2 a la categoría “100 a 119”, 3 a la categoría “120 a 139” y así sucesivamente hasta el valor 9 a la categoría “240 a 259”
 - **Medida:** Ordinal
-
- **Nombre:** Frecuencia
 - **Tipo:** Numérico
 - **Decimales:** 0
 - **Etiqueta:** Número de pacientes
 - **Medida:** Escala

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda											
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	M_Clase	Numérico	8	1	Marca de Clase	{1,0, 80 - 99...	Ninguno	8	Derecha	Ordinal	Entrada
2	Frecuencia	Numérico	8	0	Número de Departamentos	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese las categorías correspondientes al costo de arrendamiento y al número de Departamentos según corresponda.

	M. Clase	Frecuencia
1	89,5	5
2	109,5	20
3	129,5	40
4	149,5	60
5	169,5	40
6	189,5	20
7	209,5	10
8	229,5	5
9	249,5	0

2. Análisis descriptivo:

En el menú principal seleccione: Gráficos / Generador de Gráficos

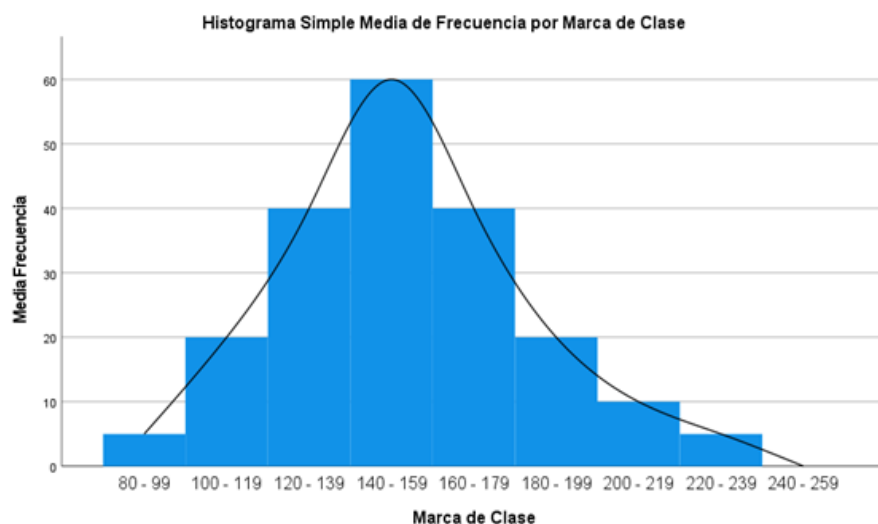
- Seleccionar la opción Histograma y mueva la opción Histograma simple a la vista previa. En el Eje x colocamos las marcas de clase y en el Eje y colocamos la variable Frecuencias



- Presione el botón Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el histograma generado



Interpretación:

El proceso es centrado puesto que la zona donde hay más departamentos es aquella que se encuentra en el rango de 140 a 179 dólares siendo la zona con los precios más representativos de la muestra. La distribución se ve más balanceada, existe aproximadamente la misma cantidad de departamentos con arriendos que van desde los 120 a 159 dólares que con los que van de 160 a 199 dólares, es decir los precios de arriendo se encuentran distribuidos de manera más uniforme, por esta razón es menos probable la existencia de valores atípicos.

3.2.2. Polígonos de frecuencia

Un polígono de frecuencias es un gráfico que permite visualizar las frecuencias con la que cambian cada una de las categorías o variables de estudio, para construirlo se colocan sobre el eje horizontal las marcas de clase o puntos medios y sobre el eje vertical las frecuencias, los puntos formados se unen por medio de segmentos de recta. Se agregan adicionalmente clases con frecuencias 0 (cero) en el eje x tanto al inicio como al final de la distribución para enlazar el diagrama con el eje horizontal.

Otra forma de construirlo es uniendo con líneas rectas las marcas de clase colocados en las cimas de las barras de un Histograma y agregando clases con frecuencias 0 (cero) en el eje x tanto al inicio como al final de la distribución para enlazar el diagrama con el eje horizontal.

Ejemplo práctico N°3:

Construir un polígono de frecuencias para el peso (en kilogramos) de 40 estudiantes e interpretar sus resultados.

Peso en Kg. (Marca de clase)	Frecuencia
112,5	1
118,5	5
124,5	6

130,5	8
136,5	12
142,5	9
148,5	4
154,5	2
160,5	1
166,5	3

Construcción en SPSS:

1. Prepare los datos:

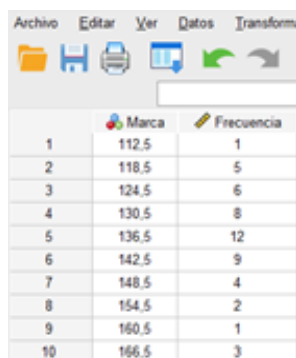
- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Marcas de clase con los siguientes parámetros:

- **Nombre:** Marcas de clase
- **Tipo:** Numérico
- **Decimales:** 1
- **Etiqueta:** Peso en kilogramos
- **Medida:** Escala

- **Nombre:** Frecuencia
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Frecuencia de repetición
- **Medida:** Escala

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda												
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol	
1	Marca	Numérico	8	1	Marcas de clase	Ninguno	Ninguno	8	Centro	Nominal	Entrada	
2	Frecuencia	Numérico	8	0	Frecuencia de repetición	Ninguno	Ninguno	8	Centro	Escala	Entrada	

- c. En la ventana “Vista de datos” ingrese los pesos en kilogramos y el número de personas que se ubican en cada categoría.



	Marca	Frecuencia
1	112.5	1
2	118.5	5
3	124.5	6
4	130.5	8
5	136.5	12
6	142.5	9
7	148.5	4
8	154.5	2
9	160.5	1
10	166.5	3

2. Análisis descriptivo:

En el menú principal seleccione: Gráficos / Generador de Gráficos

- a. Seleccionar la opción Histograma y mueva la opción Polígono de frecuencia a la vista previa. En el Eje x colocamos las marcas de clase y en el Eje y colocamos la variable Frecuencias



- b. Presione Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el Polígono de Frecuencia generado.



Se puede editar el gráfico y eliminar el relleno realizando doble clic izquierdo en la zona gráfica y luego nuevamente doble clic, en relleno seleccionar transparente, en ponderación elegir 1.5 y colocar aplicar.



4. Interpretación:

El Polígono de Frecuencia alcanza su punto máximo en 136.5kg con la mayor cantidad de encuestados encasillados en este peso, a partir de este punto se observa una disminución progresiva de la curva presentándose una asimetría sesgada positivamente. Se puede inferir gráficamente que la mayor cantidad de personas se encuentran en el rango entre 130.50 a 142.50 kg, observando una pequeña elevación al final de la distribución lo cual hace pensar que podrían presentarse valores atípicos.

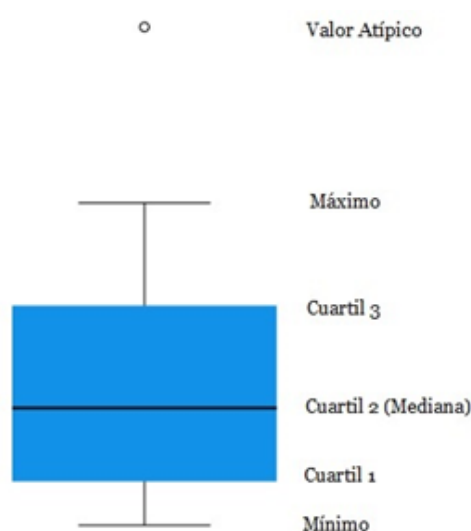
3.2.3. Diagrama de Caja y Bigote

También llamado Boxplot, un diagrama de cajas es una representación gráfica que visualiza tanto la dispersión de los datos como la forma de su distribución, detecta valores atípicos, hace énfasis en las medidas de posición y es muy útil para comparar diferentes distribuciones de manera simultánea.

Un diagrama de caja y bigote consiste en un rectángulo cuya longitud es su rango intercuartílico, se encuentra dividido por un segmento llamado mediana el cual representa el valor central de los datos y se complementa por dos líneas llamados Bigotes que parten de los extremos del rectángulo e intentan alcanzar los valores máximo y mínimo observados, los valores que sobrepasan los bigotes son considerados como atípicos (datos anormales o que son fuera de lo común).

Figura 31

Diagrama de Caja y Bigote



Los cuartiles dividen al diagrama de caja en cuatro partes iguales, el primer cuartil (Q1) es el valor que deja el 25 % de los datos por debajo de la distribución, el segundo cuartil (Q2) es la mediana y deja el 50% de los datos y el tercer cuartil deja el 75% de los datos por debajo de la serie.

Para identificar los valores extremos de la serie se calculan los límites superior e inferior y serán considerados como valores atípicos todos aquellos valores mayores al límite superior y aquellos menores al límite inferior.

$$L_s = Q_3 + 1.5 \times \text{Rango Intercuartil}$$

$$L_i = Q_1 - 1.5 \times \text{Rango Intercuartil}$$

Ejemplo práctico N°4

El tiempo en minutos en que 25 estudiantes de secundaria se demoran en resolver una prueba de factorización se muestra a continuación.

28 38 35 31 32 26 36 23 16 38 41 48 28
33 42 43 25 21 31 29 23 26 19 37 25

Realizar un diagrama de Caja y Bigote e interprete sus resultados

Construcción en SPSS:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Tiempos con los siguientes parámetros:

- **Nombre:** Tiempo
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Tiempo en que se demoran en resolver una prueba
- **Medida:** Escala



- c. En la ventana “Vista de datos” ingrese los tiempos que se demoran en resolver una prueba de factorización (la pantalla capturada muestra 21 de 25 casos).

	Tiempo
1	28
2	38
3	35
4	31
5	32
6	26
7	36
8	23
9	16
10	38
11	41
12	48
13	28
14	33
15	42
16	43
17	25
18	21
19	31
20	29
21	23

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- a. En la ventana que aparece mueva la variable Tiempo al cuadro de variables seleccionadas



- b. En el botón Estadísticos seleccionar Cuartiles, Mínimo y Máximo



- c. Presione el botón Continuar y luego Aceptar

3. Análisis gráfico:

En el menú principal seleccione: Gráficos / Generador de Gráficos / Diagrama de Cajas.

- a. En la ventana que aparece mueva la opción “Diagrama de Caja 1-D” a la vista previa y en el Eje X colocar la variable Tiempo

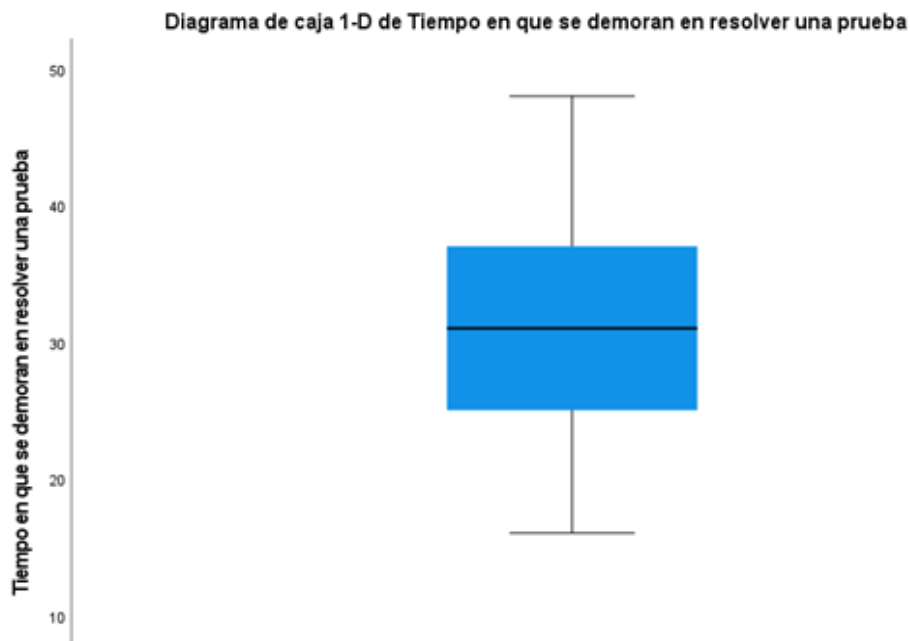


- b. Presione el botón Aceptar

4. Resultados obtenidos

En la Ventana de Resultados, encontrará los estadísticos calculados y la gráfica generada.

Estadísticos		
Tiempo en que se demoran en resolver una prueba		
N	Válido	25
	Perdidos	0
Mínimo		16
Máximo		48
Percentiles	25	25,00
	50	31,00
	75	37,50



5. Cálculos adicionales

Se calculan los límites superior e inferior para determinar la existencia de valores atípicos.

Rango:

$$R = V_{\text{máx}} - V_{\text{mín}}$$

$$R = 48 - 16$$

$$R = 32$$

Rango intercuartil:

$$IQR = Q_3 - Q_1$$

$$IQR = 36 - 25$$

$$IQR = 11$$

Límites Superior e Inferior

$$L_s = 36 + 1.5 * 11$$

$$L_s = 52.5$$

$$L_i = 25 - 1.5 * 11$$

$$L_i = 8.5$$

Para que existan valores atípicos algún dato de la distribución debe sobrepasar los límites superior e inferior calculados por lo que se puede observar que en el presente ejemplo ningún valor es superior a 52.5 tampoco es inferior a 8.5 por lo tanto no existen valores atípicos.

6. Interpretación:

- Hay un 25% de estudiantes que se demoran en resolver la prueba entre 16 y 25 minutos
- El 75% de los estudiantes se demoran en resolver la prueba hasta 37 minutos aproximadamente
- Todos los estudiantes se demoran en resolver la prueba menos de 48 minutos
- Al menos el 75% de los estudiantes se demoran en resolver la prueba 36 minutos o más
- Solo hay un estudiante que resuelve la prueba en 16 minutos
- Los tiempos de resolución de la prueba por parte de los estudiantes está en un rango de entre 16 y 48 minutos.

Ejemplo práctico N°5

Diagrama de cajas simple

Se muestran las edades de 20 trabajadores clasificados por su género en una empresa que fabrica calzado para las comunidades rurales de la provincia de Cotopaxi.

Género	Edades (años)									
Hombres	25	27	29	30	32	34	36	38	40	45
Mujeres	22	24	25	26	27	28	30	32	35	40

Realizar un diagrama de Caja y Bigote e interpretar sus resultados

Construcción en SPSS:

1. Prepare los datos:

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree una nueva variable llamada Edad con los siguientes parámetros:

- **Nombre:** Edad
 - **Tipo:** Numérico
 - **Decimales:** Cero
 - **Etiqueta:** Edad
 - **Medida:** Escala
-
- **Nombre:** Género
 - **Tipo:** Numérico
 - **Decimales:** Cero
 - **Etiqueta:** Género
 - **Valores:** asignar el valor de 1 a la categoría “Hombres” y 2 a la categoría “Mujer”
 - **Medida:** Nominal

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda											
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Edad	Numérico	8	0	Edad	Ninguno	Ninguno	8	Centro	Escala	Entrada
2	Género	Numérico	8	0	Género	{1, Hombres...}	Ninguno	8	Centro	Nominal	Entrada

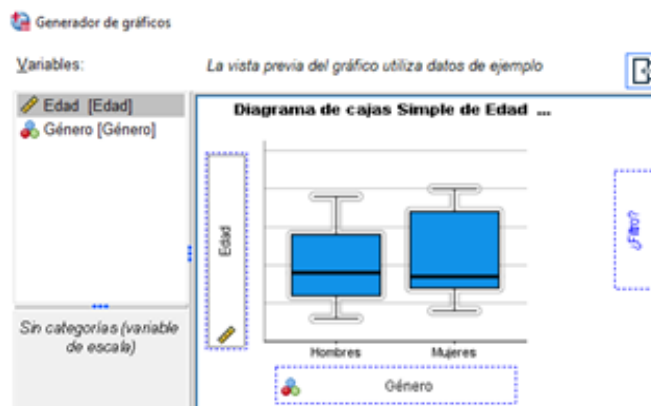
- c. En la ventana “Vista de datos” ingrese las Edades de los Trabajadores y el Género correspondiente a cada uno de ellos.

	Edad	Género
1	25	Hombres
2	27	Hombres
3	29	Hombres
4	30	Hombres
5	32	Hombres
6	34	Hombres
7	36	Hombres
8	38	Hombres
9	40	Hombres
10	45	Hombres
11	22	Mujeres
12	24	Mujeres
13	25	Mujeres
14	26	Mujeres
15	27	Mujeres
16	28	Mujeres
17	30	Mujeres
18	32	Mujeres
19	35	Mujeres
20	40	Mujeres

2. Análisis gráfico:

En el menú principal seleccione: Gráficos / Generador de Gráficos.

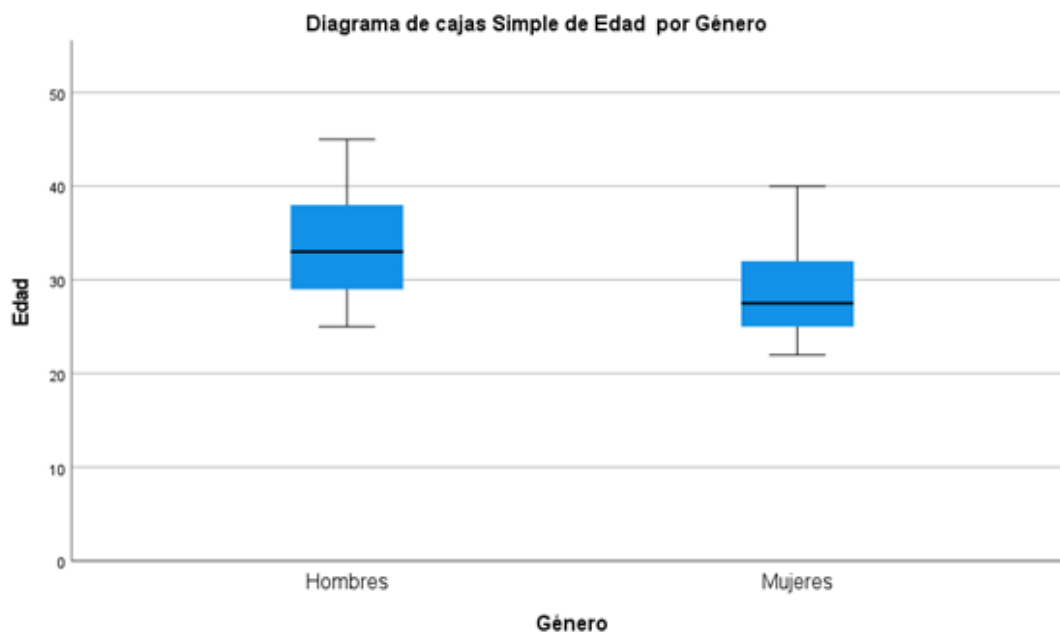
- En la ventana que aparece seleccione Diagrama de Cajas y mueva la opción Diagrama de cajas simple a la vista previa. En el Eje X coloque la variable Género y en el Eje Y coloque la variable Edad.



- Presione el botón Aceptar

3. Resultados obtenidos

Se obtiene el Diagrama de Caja y Bigote



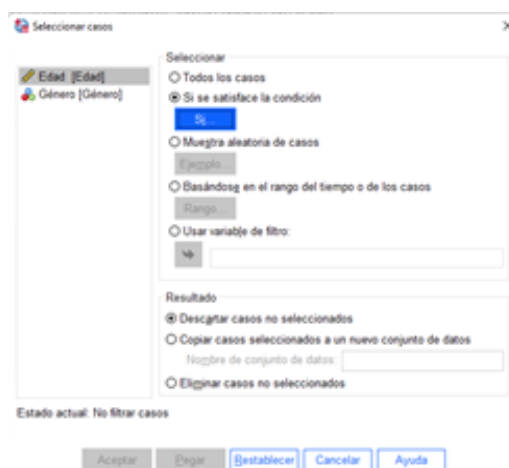
4. Análisis

Para interpretar el diagrama obtenido, se procede a realizar un análisis estadístico descriptivo por separado de cada categoría de la variable Género en función de la variable Edad.

Análisis variable Hombre – Edad

En el menú principal seleccione: Datos / Seleccionar casos

- a. En la ventana que aparece seleccione la opción “Si satisface la condición”



- b. Seleccione la opción “Si se satisface la condición” y presione el botón “Si...” mueva la variable Género a la ventana de edición y escriba = 1



- c. Presione Continuar y luego Aceptar
- d. En la ventana "Vista de Datos" se observa que únicamente aparece activa la opción Hombres

	Edad	Género	filter_\$
1	25	Hombres	Selected
2	27	Hombres	Selected
3	29	Hombres	Selected
4	30	Hombres	Selected
5	32	Hombres	Selected
6	34	Hombres	Selected
7	36	Hombres	Selected
8	38	Hombres	Selected
9	40	Hombres	Selected
10	45	Hombres	Selected
11	22	Mujeres	Not Selected
12	24	Mujeres	Not Selected
13	25	Mujeres	Not Selected
14	26	Mujeres	Not Selected
15	27	Mujeres	Not Selected
16	28	Mujeres	Not Selected
17	30	Mujeres	Not Selected
18	32	Mujeres	Not Selected
19	35	Mujeres	Not Selected
20	40	Mujeres	Not Selected

- e. En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias y seleccionamos la variable Edad



- f. Presione el botón Estadístico y damos un clic en Cuartiles, desv. Estándar, máximo, mínimo, media y mediana.



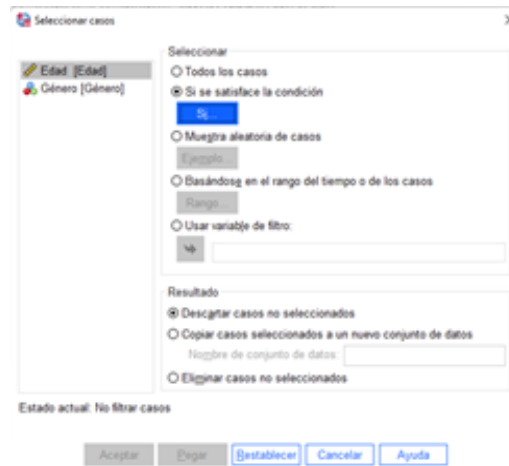
- g. Presione el botón Continuar y Aceptar y se obtiene la siguiente información

Estadísticos		
Edad		
N	Válido	10
	Perdidos	0
Media		33,60
Mediana		33,00
Desv. Desviación		6,240
Mínimo		25
Máximo		45
Percentiles	25	28,50
	50	33,00
	75	38,50

Análisis variable Mujer – Edad

En el menú principal seleccione: Datos / Seleccionar casos

- a. En la ventana que aparece seleccione la opción “Si satisface la condición”



- b. Seleccione la opción “Si se satisface la condición” y presione el botón “Si...” mueva la variable Género a la ventana de edición y escriba = 2



- c. Presione Continuar y luego Aceptar
- d. En la ventana “Vista de Datos” se observa que únicamente aparece activa la opción Mujeres

	Edad	Género	Filtrar
1	25	Hombres	Not Selected
2	27	Hombres	Not Selected
3	29	Hombres	Not Selected
4	30	Hombres	Not Selected
5	32	Hombres	Not Selected
6	34	Hombres	Not Selected
7	36	Hombres	Not Selected
8	38	Hombres	Not Selected
9	40	Hombres	Not Selected
10	45	Hombres	Not Selected
11	22	Mujeres	Selected
12	24	Mujeres	Selected
13	25	Mujeres	Selected
14	26	Mujeres	Selected
15	27	Mujeres	Selected
16	28	Mujeres	Selected
17	30	Mujeres	Selected
18	32	Mujeres	Selected
19	35	Mujeres	Selected
20	40	Mujeres	Selected

- e. En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias y seleccionamos la variable Edad



- f. Presione el botón Estadístico y damos un clic en Cuartiles, desv. Estándar, máximo, mínimo, media y mediana.



- g. Presione el botón Continuar y Aceptar y se obtiene la siguiente información

Estadísticos		
Edad		
N	Válido	10
	Perdidos	0
Media		28,90
Mediana		27,50
Desv. Desviación		5,486
Mínimo		22
Máximo		40
Percentiles	25	24,75
	50	27,50
	75	32,75

5. Interpretación:

Medianas

La mediana de las edades de los trabajadores hombres es de 33 años, mientras que la mediana de las edades de las trabajadoras mujeres es de 27.5 años, esto significa que en promedio los hombres son mayores que las mujeres.

Dispersión

En el diagrama de caja y bigote, las trabajadoras mujeres tienen una caja más estrecha que la de los trabajadores hombres, esto indica que las edades de las mujeres están más concentradas alrededor de la mediana mientras que en la caja de los trabajadores hombres sus edades están más dispersas alrededor de la mediana indicando una mayor variabilidad.

Simetría y Sesgo

Para el grupo de los Hombres, se tiene una media 33.6 años y una mediana de 33 años, estas medidas son muy cercanas, apenas tienen 0.6 años de diferencia, lo que sugiere que la distribución de edades es aproximadamente simétrica y los bigotes del diagrama de caja se extienden de manera equilibrada

Para el grupo de las Mujeres, se tiene una media 28.9 años y una mediana de 27.5 años, la media es ligeramente mayor que la mediana lo que implica un leve sesgo hacia la derecha, es decir la mayoría de edades están concentradas en el rango inferior

Bigotes

El rango de edad para los hombres es de 20 años mientras que el de las mujeres es de 18 años, es decir las edades de los hombres se distribuyen en un rango ligeramente más amplio que el de las mujeres.

Valores atípicos

Los datos que se encuentran fuera de los bigotes del boxplot pueden considerarse valores atípicos. Para determinar si existen Edades fuera de lo común se calculan los límites de los bigotes de la siguiente forma:

Trabajadores hombres	
$Bigote\ inferior = Q_1 - 1.5\ IQR$	$Bigote\ superior = Q_3 + 1.5\ IQR$
$Bigote\ inferior = 28.5 - 1.5(Q_3 - Q_1)$	$Bigote\ superior = 38.5 + 1.5(Q_3 - Q_1)$
$Bigote\ inferior$ $= 28.5 - 1.5(38.5 - 28.5)$	$Bigote\ superior$ $= 38.5 + 1.5(38.5 - 28.5)$
$Bigote\ inferior = 13.5$ (no hay edades que caigan por debajo de este valor)	$Bigote\ superior = 53.5$ (no hay edades que excedan este valor)
Trabajadoras mujeres	
$Bigote\ inferior = Q_1 - 1.5\ IQR$	$Bigote\ superior = Q_3 + 1.5\ IQR$
$Bigote\ inferior = 25.5 - 1.5(Q_3 - Q_1)$	$Bigote\ superior = 32.75 + 1.5(Q_3 - Q_1)$
$Bigote\ inferior$ $= 24.75$ $- 1.5(32.75 - 24.75)$	$Bigote\ superior$ $= 32.75$ $+ 1.5(32.75 - 24.75)$
$Bigote\ inferior = 12.75$ (no hay edades que caigan por debajo de este valor)	$Bigote\ superior = 44.75$ (no hay edades que excedan este valor)

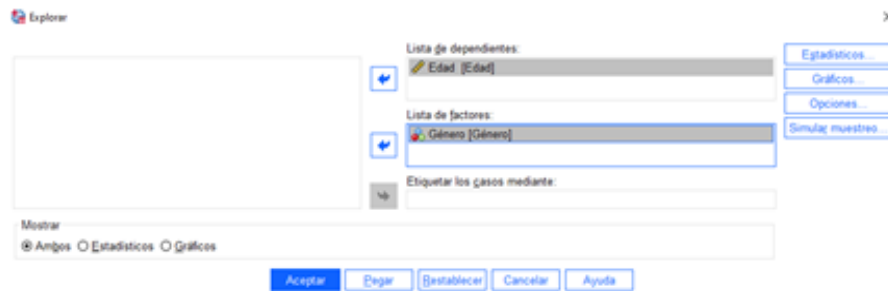
En ambos casos tanto para hombres como para mujeres no existen valores atípicos. Su demostración en SPSS a continuación:

Determinación de valores atípicos en SPSS

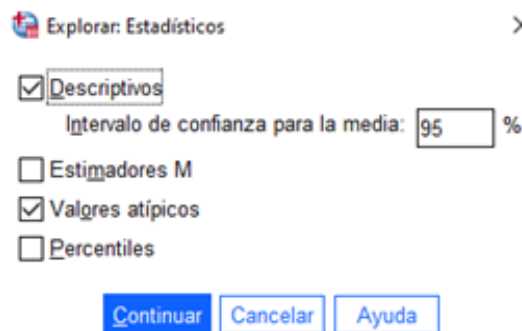
Construcción en SPSS:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Explorar

- En la ventana que aparece mueva la variable Edades a Lista de dependientes y la variable Género a Lista de factores.



- Presione el botón Estadísticos y seleccione la opción Valores Atípicos



- Presione el botón Aceptar

Resultados obtenidos

En la Ventana de Resultados, encontrará la tabla de Valores extremos

Valores extremos				
Género		Número del caso		Valor
Edades	Mujer	Mayor	1	24
				48,00
			2	23
				43,00
			3	22
				42,00
			4	20
				38,00
			5	18
				36,00
		Menor	1	1
				16,00

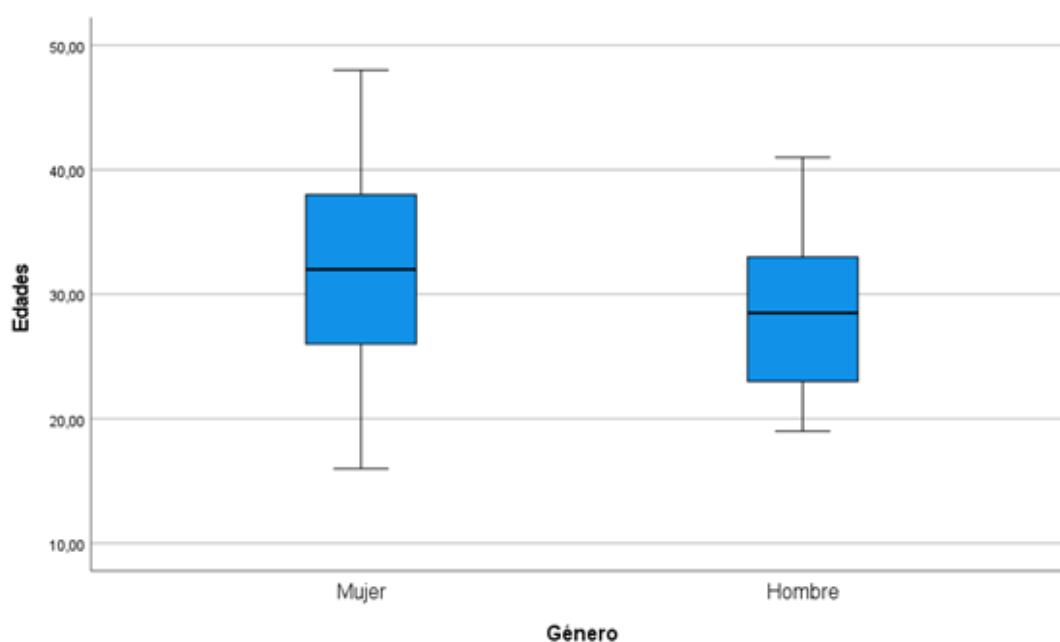
Hombre		2	6	25,00
		3	5	25,00
		4	8	26,00
		5	10	28,00
		Mayor 1	21	41,00
		2	19	37,00
		3	17	35,00
		4	12	31,00
		5	13	31,00
		Menor 1	2	19,00
		2	3	21,00
		3	25	23,00
		4	4	23,00
		5	7	26,00

La tabla muestra los valores extremos de las Edades en función del género.

La tabla muestra que los valores extremos tanto para trabajadores mujeres como para trabajadores hombres se encuentran dentro de los rangos de los bigotes inferior y superior.

El menor valor obtenido en la tabla para los hombres es el 19 y el valor obtenido en los cálculos es de 13.5

Esto se refleja también en el gráfico, en el mismo se observa solo los diagramas de caja, pero no se observan puntos extremos.



Autoevaluación 3.2.

¿En qué casos conviene utilizar un polígono de frecuencias en lugar de un histograma?

Si se realiza una investigación sobre el consumo eléctrico de 50 viviendas en el sector centro de la ciudad de Latacunga y la mayor parte de encuestados reportan consumos altos, ¿qué puedes concluir sobre la distribución?

Los datos de 20 notas de un examen de Estadística son los siguientes: 20, 16, 12, 16, 4, 12, 4, 8, 20, 6, 19, 8, 6, 10, 18, 10, 14, 19, 18, 14, construya una tabla de frecuencias Tipo I y determine la frecuencia absoluta y la frecuencia relativa, realice un diagrama de caja y bigote con los datos proporcionados e interprete la forma de su distribución.

3.2.4. Ejercicios propuestos

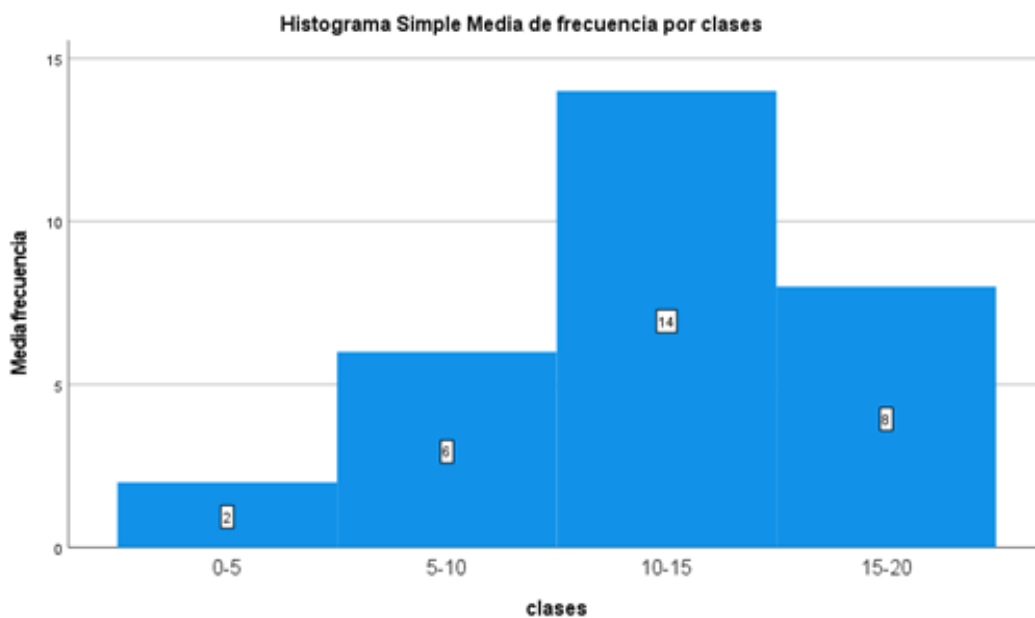
11. Las exportaciones de banano en la época de los 80's en millones de dólares fueron:

Año	Exportaciones en millones de dólares
1980	1522,2
1981	1669,3
1982	1702,4
1983	1712,5
1984	1824,4
1985	1857,1
1986	1986,3

1987	1831,7
1988	1712,3
1989	1506,8

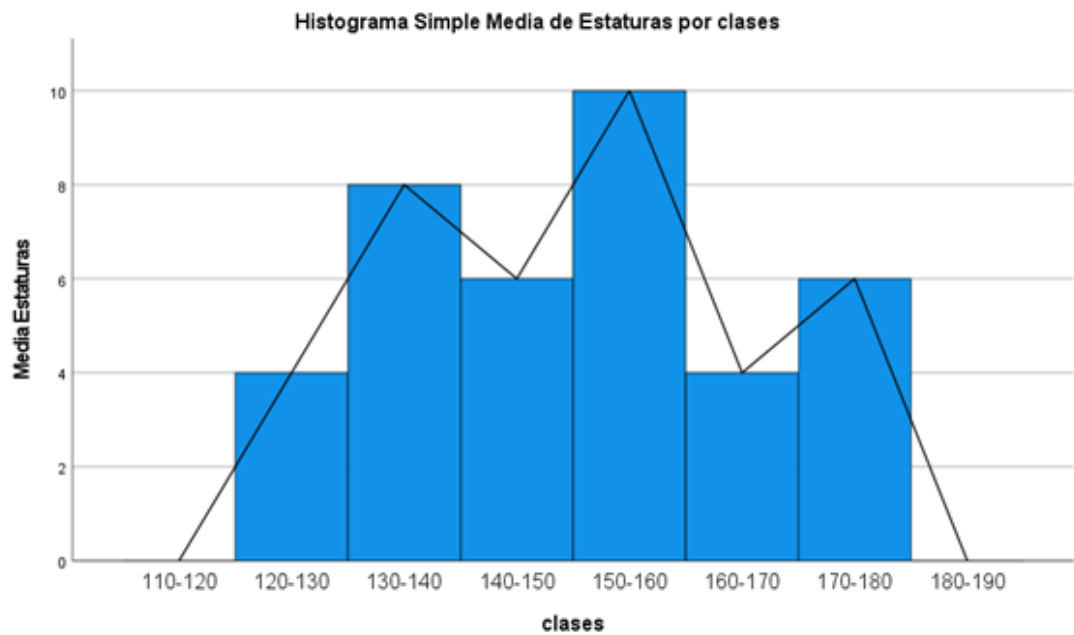
Construya un polígono de frecuencias e interprete su gráfica

12. El siguiente Histograma muestra los resultados del examen de la primera unidad de Estadística



- ¿Cuántos estudiantes se presentaron al examen?
- ¿Cuál es el intervalo de clase?
- ¿Cuántos estudiantes obtuvieron un resultado inferior a 10 puntos?

13. El diagrama siguiente muestra las estaturas de los pacientes atendidos en un hospital



¿Qué pacientes tienen una estatura entre 145 y 175 cm?

14. Los pesos en kilogramos de 30 señoras adultas son:

Peso (kg)	42	45	47	50	53	55	58	61
Frecuencia	1	3	4	6	7	3	4	2

Trazar el diagrama de caja y bigote, determine la presencia de valores atípicos e interprete sus resultados.

15. Se recogieron los datos de los ingresos mensuales de 145 hombres y 90 mujeres que laboran en distintas empresas públicas del país, obteniéndose los siguientes valores:

Ingresos mensuales	Hombres	Mujeres
2100	7	13
1400	23	18
1700	35	7
1100	25	13

1500	52	11
2500	2	19
3000	1	9

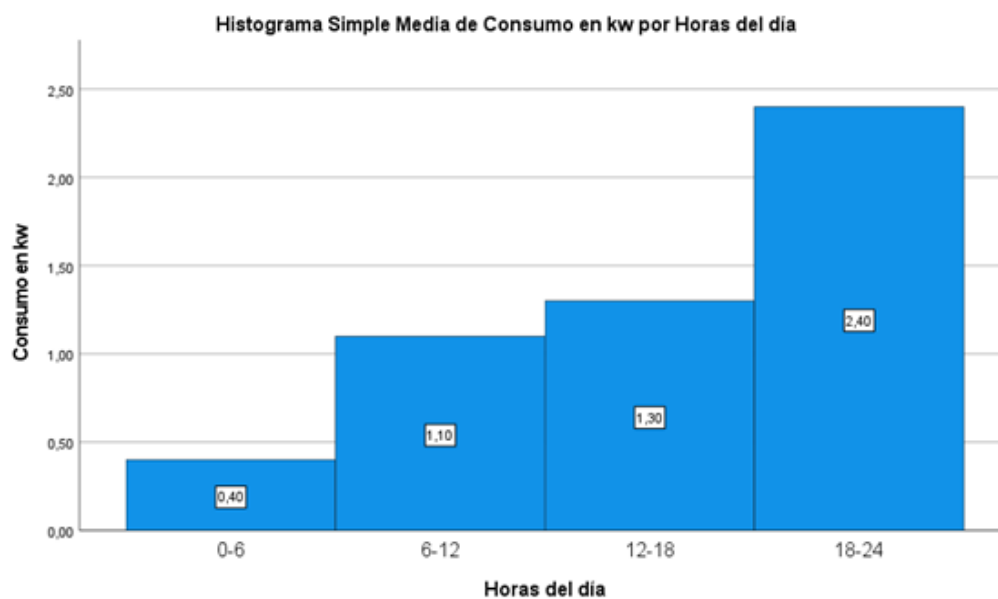
Comparar los ingresos de los hombres y las mujeres mediante sus diagramas de caja

16. Dada la siguiente tabla de frecuencias:

Intervalos de clase	Frecuencia
20-25	4
25-30	5
30-35	13
35-40	28
40-45	31
45-50	10

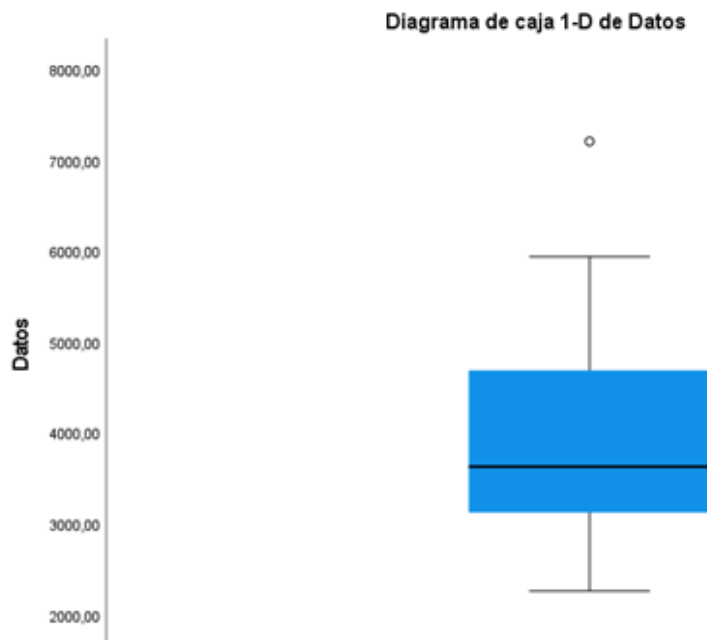
Construya un Histograma de frecuencias

17. En el siguiente gráfico se muestra el consumo de energía en una casa



¿Qué porcentaje del consumo diario se utiliza desde las 18h hasta las 24h?

18. El diagrama de caja siguiente indica el gasto realizado en promedio por la adquisición de alarmas de distintas marcas para vehículos durante un mes por la Gerencia de Auto lujos “M&M”



- Calcule la mediana del gasto realizado
 - Calcule el primero y el tercer cuartil del gasto realizado
 - Calcule el rango intercuartil (IQR)
 - ¿Más allá de qué punto se considera un dato como valor atípico?, si existe, identifique uno y calcule su valor
 - ¿La distribución es simétrica?, si no lo es determine su tipo de asimetría
19. Un estudio realizado sobre el consumo de corriente en circuitos integrados de microcontroladores en distintos dispositivos electrónicos determinó una mediana de consumo de 29.50mA, un valor pequeño de 12.00 mA y un valor alto de 44.00mA. El primer y tercer cuartil fueron de 20.00mA y 36.75 mA, respectivamente. Elabore un diagrama de caja e indique cómo es la forma de su distribución.

20. El equipo líder de Ingenieros Automotrices de una importante Empresa de la ciudad se encuentra analizando el consumo de combustible (en kilómetros por litro) de un lote de 50 vehículos SUV medianos con diferentes tipos de motor y transmisión. Tras realizar pruebas en condiciones controladas, los datos se agruparon en la siguiente tabla de frecuencia.

Consumo de combustible (km/L)	Frecuencia
10 - 13	4
13 - 16	8
16 - 19	15
19 - 22	12
22 - 25	7
25 - 28	4

Construya un polígono de frecuencia



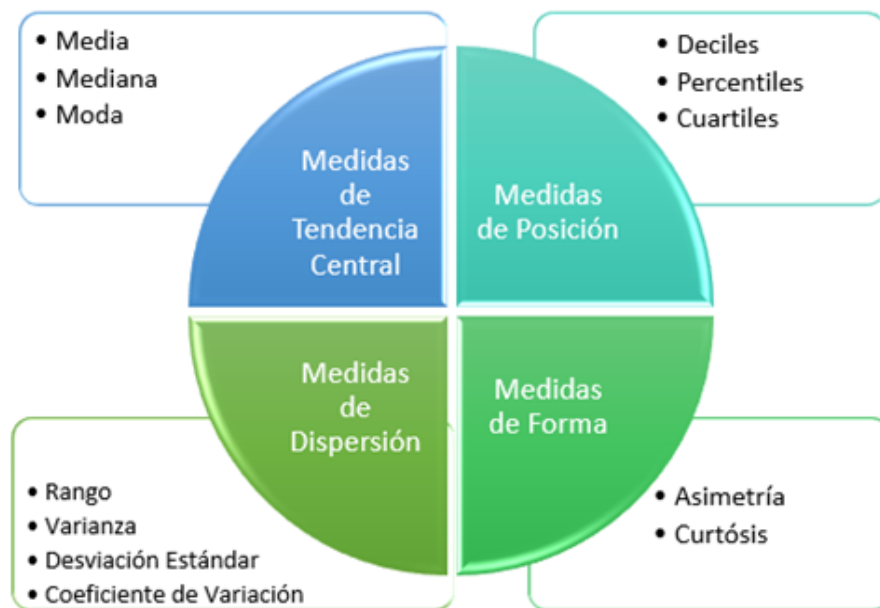
CAPÍTULO IV

Medidas descriptivas

Las medidas descriptivas son un conjunto de técnicas y métodos que se utilizan para describir, resumir y organizar los datos recogidos de un conjunto de observaciones. Estas medidas se dividen en cuatro categorías principales y cada una de ellas en subcategorías tal como lo muestra el siguiente diagrama:

Figura 32

Medidas descriptivas



4.1. Medidas de Tendencia Central

Las medidas de Tendencia central son aquellas que se utilizan para describir la posición central de un conjunto de datos. Las principales medidas son:

4.1.1. Media Aritmética

La media aritmética es el valor promedio de un conjunto de observaciones, se obtiene sumando todos los valores de cada observación y dividiendo para el número total de datos

Características:

- Son fáciles de calcular e interpretar.

- Se utilizan principalmente cuando los datos se distribuyen de manera uniforme.
- Si se tiene un valor muy alto o muy bajo, la media puede verse afectada significativamente.
- No se recomiendan en distribuciones asimétricas.

Fórmulas:

Tabla 1

Fórmulas Media Aritmética

Datos no agrupados	Datos agrupados
$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	$\bar{x} = \frac{\sum_{i=1}^n f_i x_m}{n}$
$\bar{x}$: media aritmética	f_i : frecuencia absoluta
x_i : cada observación	x_m : marca de clase
n : número total de observaciones	n : número total de observaciones

Ejercicio práctico N°1:

Los salarios en dólares de 10 funcionarios públicos de una determinada Empresa son los siguientes:

1554	1217	1416	1986	1521
1088	894	817	1210	1818

Encuentre el salario promedio que perciben los trabajadores de esta Empresa.

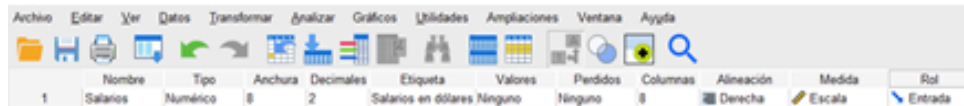
$$\bar{x} = \frac{1554 + 1217 + 1416 + 1986 + 1521 + 1088 + 894 + 817 + 1210 + 1818}{10}$$

$$\bar{x} = 1352.10$$

Construcción en SPSS:**1. Prepare los datos:**

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree una nueva variable llamada Salarios con los siguientes parámetros:

- **Nombre:** Salarios
- **Tipo:** Numérico
- **Etiqueta:** Salarios en dólares
- **Medida:** Escala



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Salarios	Numérico	8	2	Salarios en dólares	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese los salarios de los funcionarios públicos

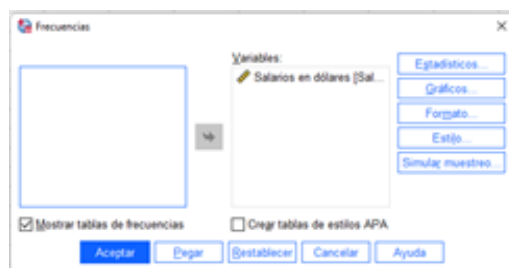


	Salarios
1	1554,00
2	1217,00
3	1416,00
4	1986,00
5	1521,00
6	1088,00
7	894,00
8	817,00
9	1210,00
10	1818,00

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- a. En la ventana que aparece mueva la variable Salarios al cuadro de variables seleccionadas



b. Presione el botón Estadísticos, dé un clic en el casillero “Media”



c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

Estadísticos		
Salarios en dólares		
N	Válido	10
	Perdidos	0
Media		1352,10

4. Interpretación

El salario promedio de los trabajadores de la Empresa es de 1352.10 usd.

Ejercicio práctico N°2:

En la siguiente tabla de frecuencias se muestran las estaturas de 36 jóvenes universitarios distribuidos en 9 categorías. Mediante un análisis de datos agrupados determine la estatura promedio.

N°	Estatura	f_i	x_m	$f_i x_m$
1	150 – 152	3	151	453
2	153 – 155	2	154	308
3	156 – 158	1	157	157
4	159 – 161	2	160	320
5	162 – 164	6	163	978
6	165 – 167	4	166	664
7	168 – 170	7	169	1183
8	171 – 173	8	172	1376
9	174 – 176	3	175	525
TOTAL		36		5964

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_m}{n}$$

$$\bar{x} = \frac{5964}{36}$$

$$\bar{x} = 165.67$$

Construcción en SPSS:**1. Prepare los datos:**

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** Estatura
- **Tipo:** Numérico
- **Decimales:** 2
- **Etiqueta:** Estatura de jóvenes universitarios en cm
- **Valores:** Asignar el valor de 1 a la categoría “150-152”, 2 a la categoría “153-155”, 3 a la categoría “156-158” y así sucesivamente hasta el valor 9 a la categoría “174-176”
- **Medida:** Ordinal
- Nombre:** Frecuencia
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Número de estudiantes en cada categoría
- **Medida:** Escala
- Nombre:** M_Clase
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Marca de clase
- **Medida:** Escala

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda										
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida
1	Estatura	Número	8	2	Estatura de jóvenes universitarios en cm	(1.00, 150 - ...	Ninguno	8	Centro	Ordinal
2	Frecuencia	Número	8	0	Número de estudiantes en cada categoría	Ninguno	Ninguno	8	Centro	Escala
3	M_Clase	Número	8	0	Marca de clase	Ninguno	Ninguno	8	Centro	Escala

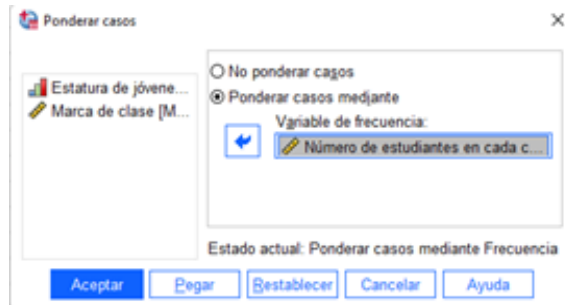
- c. En la ventana “Vista de datos” ingrese los datos correspondientes a la Estatura, la Frecuencia y la Marca de clase de cada Estudiante

Archivo Editar Ver Datos Transformar Analizar			
14 :			
	Estatura	Frecuencia	M_Clase
1	150 - 152	3	151
2	153 - 155	2	154
3	156 - 158	1	157
4	159 - 161	2	160
5	162 - 164	6	163
6	165 - 167	4	166
7	168 - 170	7	169
8	171 - 173	8	172
9	174 - 176	3	175

2. Análisis descriptivo:

En el menú principal seleccione: Datos / Ponderar casos.

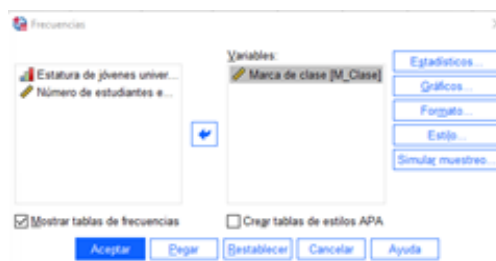
- a. En la ventana que aparece seleccione la opción *Ponderar casos mediante* y ahora seleccione la variable Frecuencias, arrastre esta variable al recuadro que dice **Variable de frecuencia**:



- b. Presione el botón Aceptar

En el menú principal Analizar / Estadísticos Descriptivos / Frecuencias

- a. En la ventana que aparece mueve la variable Marca de clase al cuadro de variables seleccionadas.



- b. Presione el botón Estadísticos, dé un clic en el casillero "Media"



- c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos:

Estadísticos		
Marca de clase		
N	Válido	36
	Perdidos	0
Media		165,67

4. Interpretación

La estatura promedio de los jóvenes universitarios es de 165.67 cm.

4.1.2. Media Geométrica

Es una medida de tendencia central que se utiliza para promediar porcentajes que aumentan a razón constante y se emplea para calcular el valor típico de un conjunto de datos.

Características:

- Es útil para calcular tasas de crecimiento promedio
- Trabaja con series de datos grandes o muy sesgados
- Solo se puede calcular para datos positivos
- La media geométrica es menos intuitiva que la media aritmética, puede ser más difícil de entender
- Es menos sensible a valores extremadamente altos, pero más sensible a valores extremadamente bajos

Fórmulas:

Tabla 2

Fórmulas Media Geométrica

Datos no agrupados	Datos agrupados
$\bar{x}_G = \sqrt[n]{\prod_{i=1}^n x_i}$	$\bar{x}_G = \sqrt[n]{\sum_{i=1}^n xm^{f_i}}$
$\bar{x}_G$: Media geométrica	f_i : frecuencia absoluta
x_i : cada observación	xm : marca de clase
n : número total de datos	

Ejercicio práctico N°3:

La siguiente tabla muestra los porcentajes de utilidad sobre las ventas anuales de la Corporación Resplandecer del 2013 – 2017. Calcule la media geométrica.

Año	Porcentaje de utilidad
2013	22.90%
2014	25.00%
2015	25.90%
2016	23.60%
2017	26.50%

La media geométrica es:

$$\bar{x}_G = \sqrt[5]{22.90 \times 25.00 \times 25.90 \times 23.60 \times 26.50}$$

$$\bar{x}_G = 24.74$$

Construcción en SPSS.**1. Prepare los datos:**

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** Año
- **Tipo:** Numérico
- **Etiqueta:** Años de estudio
- **Valores:** Asignamos el valor de 1 para el año 2013, el valor 2 para el año 2014 y así sucesivamente hasta llegar al valor 5 para el año 2017
- **Medida:** Ordinal
- **Nombre:** Utilidad
- **Tipo:** Numérico
- **Etiqueta:** Porcentaje de Utilidad
- **Medida:** Escala

Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1 Año	Númerico	8	2	Año de estudio	(1,00, 2013)	Ninguno	8	Derecha	Ordinal	Entrada
2 Utilidad	Númerico	8	2	Porcentaje de Utilidad	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese los datos correspondientes a Año y Porcentaje de Utilidad

Año	Utilidad
1 2013	22.90
2 2013	25.00
3 2015	25.90
4 2016	23.60
5 2017	26.50

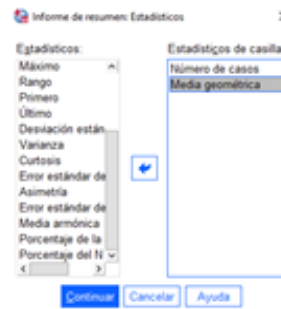
2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Informes / Resúmenes de casos.

- a. En la ventana que aparece mueve la variable Utilidad al cuadro de variables seleccionadas



b. Presionamos el botón Estadísticos y seleccionamos Media geométrica



c. Presionamos el botón Continuar y luego Aceptar

3. Resultados obtenidos:

Resúmenes de casos^a

	Porcentaje de Utilidad
1	22,90
2	25,00
3	25,90
4	23,60
5	26,50
Total	
Media geométrica	24,7427
N	5

a. Limitado a los primeros 100 casos.

4. Interpretación

Aunque se ha tenido porcentajes de utilidad variante en cada año, el desempeño equivalente de la Corporación Resplandecer en esos 5 años ha sido como si se hubiere crecido consistentemente un 24.74 % cada año.

4.1.3. Media Armónica

La media armónica se obtiene dividiendo el número de valores entre la suma de los recíprocos de dichos valores

Características:

- Ningún valor de la muestra debe ser cero
- Es el valor inverso o recíproco de la media aritmética
- Es útil cuando se manejan tasas o proporciones

Fórmulas:

Tabla 3

Fórmulas Media Armónica

Datos no agrupados	Datos agrupados
$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$	$\bar{x}_H = \frac{n}{\sum_{i=1}^n \frac{f_i}{x_m}}$
$\bar{x}_H$: media armónica	f_i : frecuencia absoluta
n : total de observaciones	x_m : marca de clase
x_i : cada observación	

Se debe cumplir siempre la relación:

$$\bar{x}_H \leq \bar{x}_G \leq \bar{x}$$

Ejercicio práctico N°4

Los valores siguientes representan las velocidades en km/h de un vehículo en diferentes tramos de una carretera.

1	3	5	7	9
---	---	---	---	---

Encontrar la velocidad promedio real del viaje y comparar su respuesta con la media aritmética.

$$\bar{x}_H = \frac{5}{\frac{1}{1} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9}}$$

$$\bar{x}_H = 2.80$$

Construcción en SPSS**1. Prepare los datos:**

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** Datos
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Datos
- **Medida:** Escala



- c. En la ventana "Vista de datos" ingrese los valores correspondientes a la variable Datos

	Datos
1	1
2	3
3	5
4	7
5	9

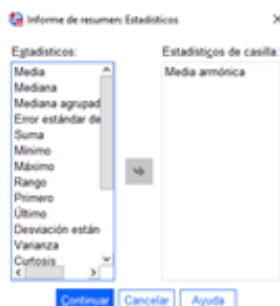
2. *Análisis descriptivo:*

En el menú principal seleccione: Analizar / Informes / Resúmenes de casos.

- a. En la ventana que aparece mueve la variable Datos al cuadro de variables seleccionadas



- b. Presionamos el botón Estadísticos y seleccionamos Media armónica



c. Presionamos el botón Continuar y luego Aceptar

3. Resultados obtenidos:

Resúmenes de casos^a

		datos
1		1,00
2		3,00
3		7,00
4		5,00
5		9,00
Total	Media armónica	2,7975

a. Limitado a los primeros 100 casos.

4. Interpretación

La media armónica representa la velocidad promedio de los vehículos sobre una distancia constante e indica cuán rápido como grupo cubrieron los vehículos la distancia fija sobre la carretera, por cuanto los valores representan velocidades en km/h en los diferentes tramos de un viaje, la velocidad promedio real del viaje sería de 2.80 km/h en lugar de 5 km/h que sería el valor de la media aritmética. En este tipo de ejercicios tiene una mejor estimación central la media armónica.

4.1.4. Mediana

La Mediana es un valor que ocupa la posición central de un conjunto de datos ordenados de menor a mayor.

Características:

- Requiere un ordenamiento de los datos

- Es fácil de calcular y entender
- Es estable a valores extremos
- Es recomendable para distribuciones asimétricas o sesgadas

Tabla 4

Fórmulas Mediana

Datos no agrupados	Datos agrupados
Posición:	
Datos impares: $\frac{n+1}{2}$	
Datos pares: $\frac{\left(\frac{n}{2}\right) + \left(\frac{n}{2} + 1\right)}{2}$	
Valor:	$\tilde{x} = L_i + i \left(\frac{\frac{n}{2} - f_{ia}}{f_i} \right)$
Datos impares: la mediana corresponde a la posición central de la distribución	
Datos pares: la mediana corresponde al promedio entre las posiciones centrales	
<i>n</i> : número total de observaciones	<i>$\tilde{x}$</i> : mediana
	<i>L_i</i> : Límite inferior
	<i>i</i> : Ancho del intervalo
	<i>f_{ia}</i> : Frecuencia acumulada anterior
	<i>f_i</i> : frecuencia absoluta

Ejercicio práctico N°5:

Las notas obtenidas de 17 estudiantes en el examen final de la asignatura de Estadística se muestran a continuación, encuentre la nota que represente la posición central de la distribución (mediana).

12	17	15	10	20	13	19	11	16	14	18	10	17	15	13	12	14
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

En primer lugar, se deben ordenar los datos.

Posición	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Datos	10	10	11	12	12	13	13	14	14	15	15	16	17	17	18	19	20

Se encuentra la posición central:

$$posición = \frac{n + 1}{2}$$

$$posición = \frac{17 + 1}{2}$$

$$posición = 9$$

Se identifica la mediana:

La nota correspondiente a la novena posición es 14 puntos, por lo tanto, la mediana de la distribución es 14

$$\tilde{x} = 14$$

Construcción en SPSS:

1. Prepare los datos:

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree una nueva variable llamada Notas con los siguientes parámetros:

- **Nombre:** Notas
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Notas obtenidas en el Examen
- **Medida:** Escala



- c. En la ventana “Vista de datos” ingrese las notas obtenidas en el Examen de Estadística

Archivo Editar Ver	
13 :	
	Notas
1	12
2	17
3	15
4	10
5	20
6	13
7	19
8	11
9	16
10	14
11	18
12	10
13	17
14	15
15	13
16	12
17	14

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- En la ventana que aparece mueva la variable Notas al cuadro de variables seleccionadas



- Presione el botón Estadísticos, dé un clic en el casillero “Mediana”



- Presione el botón Continuar y luego Aceptar

4. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada:

Estadísticos		
Notas obtenidas en el		
Examen		
N	Válido	17
	Perdidos	0
Mediana		14,00

5. Interpretación

El 50% de las notas obtenidas en el examen final de Estadística están alrededor de 14 puntos

Ejercicio práctico N°6:

En el ejercicio anterior, se obtuvo el retiro de un estudiante antes de la evaluación final, de tal manera que solo se quedaron 16 estudiantes, encuentre la Mediana.

12	17	15	10	20	13	19	11	16	14	18	17	15	13	12	14
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Se ordenan los datos:

Posición	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
----------	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----

Datos	10	11	12	12	13	13	14	14	15	15	16	17	17	18	19	20
-------	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Se encuentra la posición central:

$$posición = \frac{\left(\frac{n}{2}\right) + \left(\frac{n}{2} + 1\right)}{2}$$

$$posición = \frac{8 + 9}{2}$$

$$posición = 8.5$$

Se identifica la mediana:

La posición 8.5 corresponde al promedio entre valores 14 y 15 respectivamente:

$$\tilde{x} = \frac{14 + 15}{2}$$

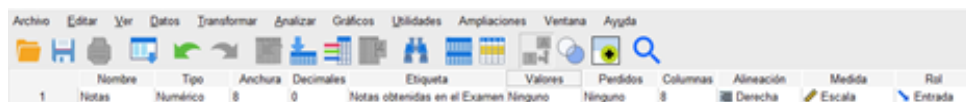
$$\tilde{x} = 14.5$$

Construcción en SPSS:

1. Prepare los datos:

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree una nueva variable llamada Notas con los siguientes parámetros:

- **Nombre:** Notas
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Notas obtenidas en el Examen
- **Medida:** Escala



- c. En la ventana “Vista de datos” ingrese las notas obtenidas en el Examen de Estadística

	Notas
1	12
2	17
3	15
4	20
5	13
6	19
7	11
8	16
9	14
10	18
11	10
12	17
13	15
14	13
15	12
16	14

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- d. En la ventana que aparece mueva la variable Notas al cuadro de variables seleccionadas



- e. Presione el botón Estadísticos, dé un clic en el casillero "Mediana"



- e. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada

Estadísticos

Notas obtenidas en el

Examen

N	Válido	16
	Perdidos	0
Mediana		14,50

4. Interpretación

El 50% de las notas que obtuvieron los estudiantes en el examen final de Estadística están alrededor de 14.50 puntos.

Ejercicio práctico N°7

A continuación, se muestran las calificaciones obtenidas sobre 100 puntos de estudiantes que rindieron un examen de ingreso a la Universidad, encuentre la Mediana de las notas obtenidas.

49	53	56	58	60	62	63	64	64	65	65	67	68	68
69	69	70	70	70	72	72	72	72	73	74	74	75	75
75	76	76	76	77	77	78	78	79	79	80	81	82	83
84	85	85	85	86	88	89	90	90	91	93	95	97	100

Los datos mostrados se agrupan en intervalos de clase.

N°	Intervalos de Clase	f_i	f_{ia}
1	46.5 – 54.5	2	2
2	54.5 – 62.5	4	6
3	62.5 – 70.5	13	19
4	70.5 – 78.5	17	36
5	78.5 – 86.5	11	47
6	86.5 – 94.5	6	53
7	94.5 – 102.5	3	56

Para determinar con cuál intervalo de clase trabajar de la Tabla de distribución de Frecuencias, se obtiene $N/2$, es decir el número total de datos dividido entre dos, con objeto de observar donde se encuentra la clase mediana:

$$\frac{N}{2} = \frac{56}{2} = 28$$

Las frecuencias acumuladas hasta la tercera clase suman 19 y las frecuencias acumuladas hasta la cuarta clase suman 36, es evidente que este último valor supera al 28 obtenido anteriormente por lo tanto la mediana se encontrará en la cuarta clase, ésta será la clase mediana.

Se encuentra la Mediana aplicando la siguiente fórmula:

$$\tilde{x} = L_i + i \left(\frac{\frac{n}{2} - f_{ia}}{f_i} \right)$$

$$\tilde{x} = 70.5 + 8 \left(\frac{\frac{56}{2} - 19}{17} \right)$$

$$\tilde{x} = 74.7 \approx 75$$

Construcción en SPSS:

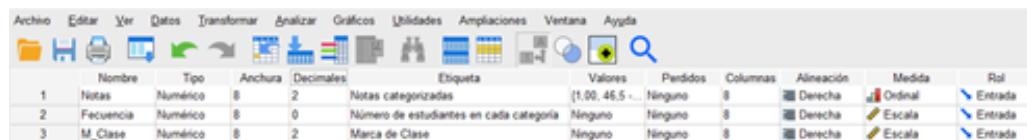
1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana "Vista de variables" cree las siguientes variables:

- **Nombre:** Notas
 - **Tipo:** Numérico
 - **Decimales:** 2
 - **Etiqueta:** Notas categorizadas
 - **Valores:** asignar el valor de 1 a la categoría "46,5 – 54,5", 2 a la categoría "54,5 – 62,5", 3 a la categoría "62,5 – 70,5" y así sucesivamente hasta el valor 7 a la categoría "94,5 – 102,5"
 - **Medida:** Ordinal
-
- **Nombre:** Frecuencia
 - **Tipo:** Numérico
 - **Decimales:** 0
 - **Etiqueta:** Número de estudiantes en cada categoría
 - **Medida:** Escala

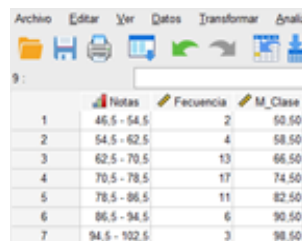
Nombre: M_Clase

- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Marca de clase
- **Medida:** Escala



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Notas	Número	8	2	Notas categorizadas	{1.00, 46.5 - ...}	Ninguno	8	Derecha	Ordinal	Entrada
2	Frecuencia	Número	8	0	Número de estudiantes en cada categoría	Ninguno	Ninguno	8	Derecha	Escala	Entrada
3	M_Clase	Número	8	2	Marca de Clase	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese las categorías correspondientes a las Notas, la Frecuencia y la Marca de clase de cada Estudiante



	Notas	Frecuencia	M_Clase
1	46.5 - 54.5	2	50.50
2	54.5 - 62.5	4	58.50
3	62.5 - 70.5	13	66.50
4	70.5 - 78.5	17	74.50
5	78.5 - 86.5	11	82.50
6	86.5 - 94.5	6	90.50
7	94.5 - 102.5	3	98.50

2. Análisis descriptivo:

En el menú principal Analizar / Estadísticos Descriptivos / Frecuencias

- a. En la ventana que aparece mueve la variable Marca de clase al cuadro de Variables seleccionadas



- b. Presione el botón Estadísticos, dé un clic en el casillero “Mediana”



c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada

Estadísticos		
Marca de Clase		
N	Válido	56
	Perdidos	0
Mediana		74,5000

SPSS calcula el valor de la Mediana realizando un promedio entre los límites inferior y superior de la clase mediana

Para el presente ejercicio:

$$Me = \frac{70.5 + 78.5}{2}$$

$$Me = 74.5$$

4. Interpretación

El 50% de las notas sobre 100 puntos obtenidas por los estudiantes que rindieron el examen de ingreso a la Universidad es 74,5 equivalente a 75 puntos.

4.1.5. Moda

La Moda es el valor que con más frecuencia se repite en la serie

Características:

- Se puede encontrar por simple inspección para datos no agrupados
- Su uso puede ser tanto en datos cuantitativos como en datos cualitativos
- No se ve influenciada por la presencia de valores atípicos
- No siempre existe, depende del conjunto de datos que se dispone

Fórmulas:

Tabla 5

Fórmulas Moda

Datos no agrupados	Datos agrupados
Por simple inspección, se determina el elemento que con más frecuencia se repite	$\hat{x} = L_i + i \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right)$ <i>L_i: Límite inferior</i> <i>i: Ancho del intervalo</i> $\Delta_1 =$ frecuencia de la clase menos la frecuencia de la clase anterior $\Delta_2 =$ frecuencia de clase menos la frecuencia de la clase siguiente

Ejemplo:

Se encuesta a 12 personas sobre el deporte favorito que más practica dándonos los siguientes resultados:

- Fútbol: 6

- Basquetbol: 3
- Voleibol: 1
- Ajedrez: 2
- Natación: 1

El deporte de mayor preferencia entre los encuestados es el fútbol con 6 respuestas favorables.

Ejercicio práctico N°8

Determinar la Moda de las edades de pacientes atendidos en un hospital en base a la tabla que se muestra a continuación.

N°	Clases	f_i
1	2 - 8	1
2	8 - 14	2
3	14 - 20	3
4	20 - 26	2
5	26 - 32	8
6	32 - 38	1
7	34 - 44	3
8	44 - 50	4

La clase Modal será aquella que contiene la mayor frecuencia de repetición, en este caso será la clase N° 5 con 8 personas que están en un rango de 26 a 32 años de edad.

$$\hat{x} = L_i + i \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right)$$

$$\hat{x} = 26 + 6 \left(\frac{(8 - 2)}{(8 - 2) + (8 - 1)} \right)$$

$$\hat{x} = 28.8 \approx 29$$

La edad que con más frecuencia se repite es de 29 años con un total de 8 pacientes atendidos en el Hospital.

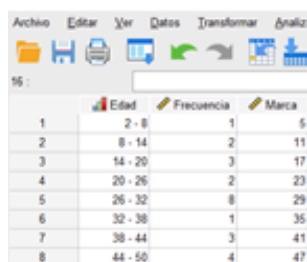
Construcción en SPSS:**1. Prepare los datos:**

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana "Vista de variables" cree las siguientes variables:

- **Nombre:** Edad
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Edad de los pacientes
- **Valores:** asignar el valor de 1 a la categoría "2 - 8", 2 a la categoría "8 - 14", 3 a la categoría "14 - 20" y así sucesivamente hasta el valor 8 a la categoría "44 - 50"
- **Medida:** Ordinal
- **Nombre:** Frecuencia
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Número de pacientes
- **Medida:** Escala
- **Nombre:** M_Clase
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Marca de clase
- **Medida:** Escala

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda											
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Edad	Númérico	8	2	Edad de pacientes	(1,00, 2 + 8)...	Ninguno	8	Derecha	Ordinal	Entrada
2	Frecuencia	Númérico	8	0	Número de pacientes	Ninguno	Ninguno	8	Derecha	Escala	Entrada
3	Marca	Númérico	8	0	Marca de clase	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese las categorías correspondientes a las Edades, la Frecuencia y la Marca de clase de cada Paciente.

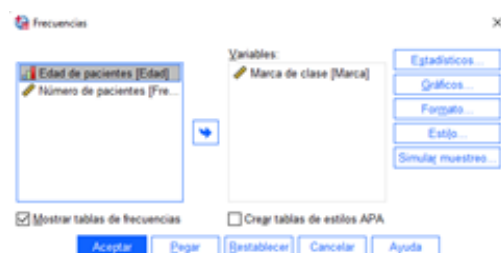


	Edad	Frecuencia	Marca
1	2 - 8	1	5
2	8 - 14	2	11
3	14 - 20	3	17
4	20 - 26	2	23
5	26 - 32	8	29
6	32 - 38	1	35
7	38 - 44	3	41
8	44 - 50	4	47

2. Análisis descriptivo:

En el menú principal Analizar / Estadísticos Descriptivos / Frecuencias

- a. En la ventana que aparece mueve la variable Marca de clase al cuadro de variables seleccionadas.



- b. Presione el botón Estadísticos, dé un clic en el casillero “Moda”



3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

Estadísticos		
Marca de clase		
N	Válido	24
	Perdidos	0
Moda		29

SPSS calcula el valor de la Moda realizando un promedio entre los límites inferior y superior de la clase modal

Para el presente ejercicio $Mo = \frac{26+32}{2} = 29$

4. Interpretación

La edad de los pacientes que con más frecuencia se repite es 29 años.

Autoevaluación 4.1.

1. ¿En qué casos se recomienda utilizar la media geométrica y la media armónica? Da un ejemplo de cada uno
2. ¿Qué ventaja tiene la mediana respecto a la media aritmética ante la presencia de valores atípicos?
3. En una cafetería se registran el número de cafés consumidos al día por 10 personas: 2,2,3,3,3,3,4,5,6,6. Determina la media aritmética, la mediana y la moda e interprete sus resultados.

4.1.6. Ejercicios propuestos

1. Calcule la media, mediana y moda de las edades de 28 estudiantes universitarios

19	20	18	18	21	22	17	19	21	20	18	18	22	22
21	21	22	22	21	18	22	21	17	22	22	22	21	20

2. La siguiente tabla muestra la temperatura nocturna (en °C) en la época de invierno en la ciudad de Quito durante 180 días, encuentre la media, mediana y moda.

Intervalo	Frecuencia
8 – 10	20
10 – 12	43
12 – 14	25
14 – 16	29
16 – 18	35
18 – 20	28

3. La siguiente tabla muestra la distribución de los sueldos de 150 trabajadores de una empresa florícola de la provincia de Cotopaxi. Calcule la Moda.

Sueldos	Trabajadores
400 – 500	50
500 – 600	35
600 – 700	22
700 – 800	20
800 – 900	13
900 – 1000	10

4. El profesor de Estadística evalúa el desempeño a 27 estudiantes en una práctica de laboratorio obteniendo los siguientes puntajes.

17	18	20	18	18	18	15	17	15	19	17	15	15	15
15	15	20	19	16	15	19	15	15	18	20	15	17	

- a. Calcula la media, mediana y moda
 - b. Qué medidas tomaría para mejorar el rendimiento de la clase
5. Como parte de un proyecto, los estudiantes de la carrera de Ingeniería Automotriz quieren analizar el rendimiento promedio en kilómetros por galón de 12 prototipos, los datos obtenidos en las pruebas de laboratorio son los siguientes:

91	87	82	75	72	76	73	77	83	95	92	75
----	----	----	----	----	----	----	----	----	----	----	----

En la etapa final de revisión, los estudiantes notaron que el dato correspondiente al vehículo con mejor desempeño había sido registrado de manera incorrecta, es decir en lugar de marcar 95km/g, el rendimiento real era de 155 km/g, es decir hubo un incremento de 60 unidades

- a. ¿Qué efecto tiene sobre la media de los datos?
 - b. ¿Qué efecto tiene sobre la mediana de los datos?
6. Una empresa multinacional de telecomunicaciones cuenta con tres centros de operación en distintas ciudades del país. El primer centro emplea 421 técnicos con un salario promedio de 1200 dólares, el segundo emplea a 645 técnicos con un salario promedio de 1000 dólares y el tercero emplea a 815 técnicos con un salario promedio de 1350 dólares. ¿Cuál es el salario promedio de los empleados de la empresa multinacional de telecomunicaciones?
7. Una empresa tecnológica, muestra los tiempos de respuesta de 50 servidores utilizados para alojar aplicaciones web durante un período de prueba. Los tiempos de respuesta (en milisegundos) registrados de cada servidor son los siguientes:

138, 184, 163, 169, 117, 106, 191, 134, 188, 149, 200, 161, 166, 130, 130, 145, 134,
158, 183, 183, 117, 131, 149, 156, 180, 170, 117, 132, 187, 165, 153, 107, 137, 117,
179, 174, 147, 112, 173, 166, 115, 181, 152, 193, 154, 131, 117, 199, 167, 165

- Calcule la media, mediana y moda (sin agrupar los datos)
 - Grafique la distribución en un diagrama de Caja y Bigote
 - Agrupe los datos en 7 intervalos de clase, cada uno de amplitud 14
 - Calcule las medidas solicitadas en el inciso a
8. Los salarios de los empleados de una empresa de Telecomunicaciones son los siguientes:

1 Gerente	USD 5200
1 Presidente Ejecutivo	USD 1500
1 Contabilidad	USD 980
1 Secretaria	USD 900
5 Empleados	USD 550 (c/u)
1 Conserje	USD 300

¿Cuál es la medida de localización más representativa?

9. La siguiente tabla de distribución de frecuencias muestra las notas obtenidas por 30 estudiantes en una prueba de Cálculo Diferencial e Integral

Clases	Frecuencia
6 - 8	1
8 - 10	2
10 - 12	5
12 - 14	6
14 - 16	8
16 - 18	7
18 - 20	1

Determine:

- a. La media aritmética
 - b. La media geométrica
 - c. La media armónica
 - d. ¿Cuál es la relación entre las tres medidas?
10. La siguiente tabla de frecuencias presenta el número de estudiantes que viajan en un bus desde la ciudad de Ambato a la ciudad de Latacunga para asistir a la Universidad de las Fuerzas Armadas - ESPE en una semana.

Pasajeros	2 - 6	6 - 10	10 - 14	14 - 18	18 - 22
Frecuencia	3	5	8	7	3

- a. Encuentre el número promedio de pasajeros que se transporta por cada viaje
- b. Encuentre la moda y la mediana del número de pasajeros

4.2. Medidas de Posición

Las medidas de Posición son aquellas que dividen a una serie de datos en partes iguales e indican el lugar que ocupan los mismos dentro de una distribución, para calcularlos se recomienda que la serie de datos se ordenen en forma ascendente.

Se clasifican en:

4.2.1. Cuartiles

Son aquellos valores que dividen a una serie de datos en 4 partes iguales, es así que el Q1, Q2 (Mediana) y Q3 establecen los valores correspondientes al 25%, al 50% y al 75% de los datos.

Figura 33
Cuartiles



Tabla 6
Fórmulas Cuartiles

Datos No Agrupados	Datos Agrupados
$Q_1 = \frac{1(n + 1)}{4}$	$Q_1 = L_i + i \left(\frac{\frac{1n}{4} - f_{aa}}{f} \right)$
$Q_2 = \frac{2(n + 1)}{4}$	$Q_2 = L_i + i \left(\frac{\frac{2n}{4} - f_{aa}}{f} \right)$
$Q_3 = \frac{3(n + 1)}{4}$	$Q_3 = L_i + i \left(\frac{\frac{3n}{4} - f_{aa}}{f} \right)$
n : número de observaciones	L_i : Límite inferior
	i : ancho del intervalo
	f_{aa} : Frecuencia acumulada anterior a la posición
	f : Frecuencia absoluta

4.2.2. Deciles

Son aquellos valores que dividen a una serie de datos en 10 partes iguales, es así que el D1, D2, ..., D9 establecen los valores correspondientes al 10%, 20%, ..., 90% de los datos.

Figura 34
Deciles

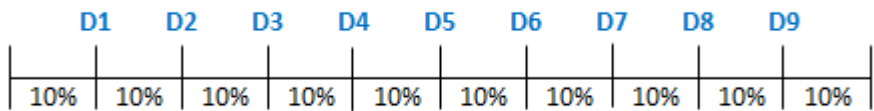


Tabla 7
Fórmulas Deciles

Datos No Agrupados	Datos Agrupados
$D_1 = \frac{1(n + 1)}{10}$	$D_1 = L_i + i \left(\frac{\frac{1n}{10} - f_{aa}}{f} \right)$
$D_2 = \frac{2(n + 1)}{10}$	$D_2 = L_i + i \left(\frac{\frac{2n}{10} - f_{aa}}{f} \right)$
.....	
$D_9 = \frac{9(n + 1)}{10}$	$D_9 = L_i + i \left(\frac{\frac{9n}{10} - f_{aa}}{f} \right)$
<i>n</i> : número de observaciones	<i>L_i</i> : Límite inferior
	<i>i</i> : ancho del intervalo
	<i>f_{aa}</i> : Frecuencia acumulada anterior a la posición
	<i>f</i> : Frecuencia absoluta

4.2.3. Percentiles

Son aquellos valores que dividen a una serie de datos en 100 partes iguales, es así que el P1, P2, ..., P99 establecen los valores correspondientes al 1%, 2%, ..., 99% de los datos.

Figura 35
Percentiles



Tabla 8

Fórmulas Percentiles

Datos No Agrupados	Datos Agrupados
$P_1 = \frac{1(n+1)}{100}$	$P_1 = L_i + i \left(\frac{\frac{n}{100} - f_{aa}}{f} \right)$
$P_2 = \frac{2(n+1)}{100}$	$P_2 = L_i + i \left(\frac{\frac{2n}{100} - f_{aa}}{f} \right)$
.....	
$P_{99} = \frac{99(n+1)}{100}$	$P_{99} = L_i + i \left(\frac{\frac{99n}{100} - f_{aa}}{f} \right)$
<i>n</i> : número de observaciones	L_i : Límite inferior
	<i>i</i> : ancho del intervalo
	f_{aa} : Frecuencia acumulada anterior a la posición
	<i>f</i> : Frecuencia absoluta

Ejercicio Práctico N°1

Los tiempos de uso de Internet medidos por los estudiantes de Bachillerato para realizar un trabajo de Programación en el Laboratorio de su Institución son los siguientes: 41, 46, 43, 57, 85, 103, 92, 27, 22, 19, 16, 28, 35, 33, 39, 89, 83, 68, 76, 65, 48, 53, 60, 54 minutos. Encuentre el Q1, D7, P65.

Recordemos que, para encontrar las medidas de posición, es necesario ordenar los datos de manera ascendente:

16, 19, 22, 27, 28, 33, 35, 39, 41, 43, 46, 48, 53, 54, 57, 60, 65, 68, 76, 83, 85, 89, 92 y 103

Cuartil 1 (Q1)

Para encontrar el Cuartil 1, realizamos el siguiente proceso:

Para encontrar la posición del cuartil 1:

$$Q_1 = \frac{1(n+1)}{4}$$

$$Q_1 = \frac{1(24 + 1)}{4}$$

$$Q_1 = 6.25$$

Para encontrar el valor del cuartil 1, se debe interpolar:

Posición 6: 33

Posición 7: 35

Q1: posición 6+(posición 7-posición 6)*decimal obtenido

Q1: 33+(35-33)*0.25

Q1 = 33.5

Decil 7 (D7)

Para encontrar el Decil 7, realizamos el siguiente proceso:

Para encontrar la posición del decil 7:

$$D_7 = \frac{7(n + 1)}{10}$$

$$D_7 = \frac{7(24 + 1)}{10}$$

$$D_7 = 17.5$$

Para encontrar el valor del decil 7, se debe interpolar:

Posición 17: 65

Posición 18: 68

D7: posición 17+(posición 18-posición 7)*decimal obtenido

D7: 65+(68-65)*0.25

D7 = 66.5

Percentil 65 (P65)

Para encontrar el Percentil 65, realizamos el siguiente proceso:

Para encontrar la posición del percentil 65:

$$P_{65} = \frac{65(n+1)}{100}$$

$$P_{65} = \frac{65(24+1)}{100}$$

$$P_{65} = 16.25$$

Para encontrar el valor del percentil 65, se debe interpolar:

Posición 16: 60

Posición 17: 65

P65: posición 16+(posición 17-posición 16)*decimal obtenido

P65: 60+(65-60)*0.25

P65 = 61.25

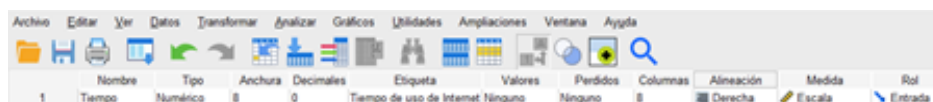
Construcción en SPSS:

1. Prepare los datos:

Abra SPSS y cree un nuevo archivo de datos

- En la ventana “Vista de variables” cree una nueva variable llamada Tiempo con los siguientes parámetros:

- **Nombre:** Tiempo
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Tiempo de uso de Internet
- **Medida:** Escala



- En la ventana “Vista de datos” ingrese los tiempos de uso de Internet.

	Tiempo
1	16
2	19
3	22
4	27
5	28
6	33
7	36
8	39
9	41
10	43
11	46
12	48
13	53
14	54
15	57
16	60
17	65
18	68
19	76
20	83
21	85
22	89

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- En la ventana que aparece mueva la variable Tiempo al cuadro de variables seleccionadas.



- Presione el botón Estadísticos, dé un clic en el casillero "Percentiles" y seleccione los percentiles 25, 65 y 70 que corresponden al Q1, P65 y al D7 respectivamente.



c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

Estadísticos		
Datos		
N	Válido	24
	Perdidos	0
Percentiles	25	33,5000
	65	61,2500
	70	66,5000

4. Interpretación

El 25% de los tiempos de uso de Internet medidos por los estudiantes de Bachillerato para realizar un trabajo de Programación se encuentran por debajo de 33.5 minutos, el 65% de los tiempos se encuentran por debajo de 61.25 minutos y el 70% de los tiempos se encuentran por debajo de 66.5 minutos.

Ejercicio N°2

A continuación, se muestran las calificaciones ordenadas sobre 100 puntos obtenidas de estudiantes que rindieron un examen de ingreso a la universidad, encuentre la mediana de las notas obtenidas, el decil 8 y el percentil 45.

49	53	56	59	61	63	65	66	66	67	67	68	69	69	70	70	71	71	71
72	72	72	72	73	74	76	76	76	77	77	77	78	78	79	79	80	80	81
82	82	83	84	85	86	86	86	87	89	91	91	93	95	97	98	100		

N°	Intervalos	f	fa
1	[46,5-54,5)	2	2
2	[54,5-62,5)	4	6
3	[62,5-70,5)	13	19
4	[70,5-78,5)	17	36 (P45)
5	[78,5-86,5)	11	47 (D8)
6	[86,5-94,5)	6	53
7	[94,5-102,5]	3	56
		56	

Mediana:

$$\frac{N}{2} = \frac{56}{2} = 28 \quad (\text{posición})$$

$$\tilde{x} = L_i + i \left(\frac{\frac{n}{2} - f_a a}{f} \right)$$

$$\tilde{x} = 70.5 + 8 \left(\frac{\frac{56}{2} - 19}{17} \right)$$

$$\tilde{x} = 74.7$$

Decil:

$$D_8 = \frac{8(56)}{10} = 44.8 \quad (\text{posición})$$

$$D_8 = L_i + i \left(\frac{\frac{8n}{10} - f_a a}{f} \right)$$

$$D_8 = 78.5 + 8 \left(\frac{44.8 - 36}{11} \right)$$

$$D_8 = 84.9$$

Percentil:

$$P_{45} = \frac{45(56)}{100} = 25.2 \quad (\text{posición})$$

$$P_{45} = L_i + i \left(\frac{\frac{45n}{100} - f_a a}{f} \right)$$

$$P_{45} = 78.5 + 8 \left(\frac{25.2 - 19}{17} \right)$$

$$D_8 = 73.42$$

Ejercicio N°3

Encuentre el cuartil 1 y el cuartil de 2 de los siguientes 100 datos obtenidos de un estudio.

50	52	55	57	60	58	61	62	100	91
67	67	68	68	61	70	70	70	60	78
74	76	74	76	75	75	76	76	77	72
82	83	82	83	84	85	87	90	88	84

Nº	Intervalos	f	fa
1	[47-54]	2	2
2	[55-62]	8	10
3	[63-70]	8	18
4	[71-78]	11	29
5	[79-86]	6	35
6	[87-94]	4	39
7	[95-103]	1	40
		40	

$$Q_1 = \frac{n}{4} \quad (\text{posición})$$

$$Q_1 = \frac{40}{4} = 10$$

$$Q_1 = L_i + i \left(\frac{\frac{n}{4} - f_a}{f} \right)$$

$$Q_1 = 55 + 8 \left(\frac{10 - 10}{8} \right)$$

$$Q_1 = 63$$

$$Q_2 = \frac{2n}{4} \quad (\text{posición})$$

$$Q_2 = \frac{2 * 40}{4} = 20$$

$$Q_2 = L_i + i \left(\frac{\frac{2n}{4} - f_a}{f} \right)$$

$$Q_2 = 71 + 8 \left(\frac{20 - 18}{11} \right)$$

$$Q_2 = 72.45$$

Autoevaluación 4.2.

1. ¿A qué decil equivale el cuartil 2 y el percentil 50?
2. Interpreta el siguiente resultado: el percentil 25 de las estaturas de los adolescentes de un colegio es 1.50m
3. El siguiente es un conjunto de datos correspondiente a los tiempos en minutos que demoran los estudiantes en responder un test de evaluación de la asignatura de Álgebra Lineal: 60, 85, 72, 93, 68, 81, 110, 95, 105, 118.
Determinar el Q1, D6 y el P82

4.2.4. Ejercicios propuestos

11. Determinar la mediana y los valores correspondientes al primer y tercer cuartil en los siguientes datos:

7	5	6	14	10	7	13	11	12	12
12	8	14	9	8	15	7	10	15	6

12. Dados los siguientes datos:

53	63	76	82	27	49	52	98	57	29
50	46	44	43	57	75	43	86	88	63
32	74	95	74	36	45	92	33	59	46
95	34	93	89	19	69	55	96	12	44
93	14	51	74	36	21	59	74	40	11

- a. Construya una tabla de frecuencias
- b. Encuentre los valores bajo los cuales se encuentra el 16%, el 45% y el 85% de los datos

13. Un investigador registró los tiempos (en minutos) que tardaron 30 estudiantes en resolver un examen de Estadística.

90	109	110	90	77	67	115	111	101	114
90	91	78	113	78	63	90	111	84	88
85	79	87	108	118	112	115	105	103	68

- Construya una tabla de frecuencias que contemple 6 intervalos
 - Construya un histograma
 - Halle el promedio y la mediana
 - Determine el valor por debajo del cual se hallan el 30% de los tiempos registrados
 - Determine el valor por encima del cual se hallan el 20% de los tiempos registrados
14. La siguiente tabla de frecuencias registra las calificaciones de 26 estudiantes en el examen de Estadística Inferencial del tercer parcial. Determine:
- La calificación más baja lograda por el 30% superior de la clase
 - La calificación más alta lograda por el 25% inferior de la clase
 - Cuál es la calificación del 50% de los alumnos.

Calificaciones	Frecuencia
10 - 12	4
12 - 14	5
14 - 16	6
16 - 18	8
18 - 20	3

15. La Agencia Nacional de Tránsito registró la velocidad de circulación (en kilómetros por hora) de cincuenta y seis automóviles en la zona urbana de una ciudad dando los siguientes valores:

47	57	40	53	45	43	60	49	54	44	51	35	37	46
44	37	31	30	38	45	37	52	43	32	34	54	32	61
39	32	51	55	38	34	53	44	54	43	43	31	48	44
30	39	56	36	36	40	42	59	39	42	39	33	51	54

- Resuma los datos en una tabla de frecuencia
 - Encuentra las velocidades más baja y más alta de circulación
 - Calcule la velocidad promedio de circulación
 - Calcule los cuartiles inferior y superior de la muestra
 - Encuentre el rango intercuartil
 - Grafique la distribución mediante un Diagrama de Cajas
16. El Docente de la asignatura de Matemática Financiera registró las notas finales del primer parcial de sus estudiantes. Él desea identificar qué estudiantes tienen que presentarse a Tutorías para el segundo parcial si se encuentran en el 40% inferior y quienes pueden recibir algún tipo de reconocimiento si se encuentran en el 20% superior del grupo.

4	19	5	12	14	20	19	14	15	17	17	4
20	20	16	16	15	9	8	19	18	7	5	18

Determine:

- ¿Cuál es la calificación máxima para estar en el 40% inferior?
- ¿Desde qué calificación se considera que un estudiante pertenece al 20% superior?
- ¿Cuál es la mediana de las notas del grupo?

- d. ¿Quiénes deben recibir Tutorías de reforzamiento y quienes deben recibir algún tipo de reconocimiento?

17. En un prestigioso colegio de la ciudad se realizaron juegos deportivos internos con la disciplina de atletismo 5 km como su prueba con más demanda, para ello se registró la participación de un pelotón de 15 estudiantes con los siguientes tiempos.

23	34	34	20	22	35	34	21	21	32	33	22	23	31	32
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

- Calcule el tercer cuartil para determinar el grupo de estudiantes más lentos en completar la carrera
- Calcule el primer decil para identificar a los corredores que se encuentran dentro del 10% con los mejores tiempos
- Qué valor debe superar un participante para ubicarse dentro del 60% con mejor rendimiento (percentil 40)

18. Se estudió el gasto mensual en dólares de 60 jóvenes que manejan pequeños negocios digitales. La distribución es la siguiente:

Gasto mensual (USD)	Frecuencia
100 – 200	5
200 – 300	9
300 – 400	13
400 – 500	16
500 – 600	10
600 – 700	7

- Determine el decil 2

19. Una empresa desea analizar los salarios semanales de sus 60 empleados. Para ello, ha agrupado los datos en la siguiente tabla de distribución de frecuencias:

Nº	Intervalo	Frecuencia
1	10-20	5
2	20-30	8
3	30-40	12
4	40-50	20
5	50-60	10
6	60-70	5
		60

Con base en esta información, calcule los tres cuartiles Q_1 , Q_2 , Q_3 que dividen la distribución en cuatro partes iguales.

20. Una compañía de distribución recopiló datos sobre el número de entregas semanales realizadas por sus 80 repartidores. Los resultados se agruparon de la siguiente manera:

Nº	Intervalo	Frecuencia
1	10-20	6
2	20-30	14
3	30-40	20
4	40-50	18
5	50-60	12
6	60-70	10
		80

Calcule:

- El séptimo decil (D_7)
- El percentil 30 (P_{30})

4.3. Medidas de Dispersión

Las Medidas de Dispersión son aquellas que indican qué tan dispersos se encuentran los datos respecto a la media, se clasifican en:

4.3.1. Rango

El Rango es la diferencia entre el valor máximo y el valor mínimo de una serie de datos.

Características:

- Es fácil de calcular e interpretar.
- Se consideran únicamente los valores extremos (máximo y mínimo) sin importar los demás datos que tiene la serie
- El rango puede ser el mismo en dos distribuciones diferentes si tiene los mismos valores extremos
- Al tener en cuenta únicamente los valores máximo y mínimo, no se tiene información de cómo es la forma de su distribución

Fórmula:

$$\text{Rango} = \text{valor máximo} - \text{valor mínimo}$$

Ejemplo:

Los días tomados de vacaciones por los empleados de una empresa en el primer semestre del año son los siguientes:

5, 20, 15, 9, 2, 3, 18, 16, 13, 1, 8, 14

Encuentre el Rango:

$$\text{Rango} = 18 - 1$$

$$\text{Rango} = 17$$

El rango de 17 indica la diferencia entre la persona que tiene menos días de vacaciones con la persona que tiene más días de vacaciones, esta diferencia es significativa y muestra una dispersión considerable en el número de días tomados de vacaciones.

4.3.2. Varianza

Es una medida de dispersión que indica el grado de variación de un conjunto de datos alrededor de la media.

Características

- Dificulta su interpretación por sus unidades elevadas al cuadrado.
- Se obtendrá siempre un valor positivo por la sumatoria del cuadrado de las diferencias.
- Para su cálculo utiliza todos los datos de la distribución proporcionando así una visión más completa de la dispersión.

Tabla 9

Fórmulas Varianza

	Datos No agrupados	Datos agrupados
Poblacional	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$	$\sigma^2 = \frac{\sum_{i=1}^N f(x_m - \mu)^2}{N}$
Muestral	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$	$s^2 = \frac{\sum_{i=1}^n f(x_m - \bar{x})^2}{n - 1}$
	x_i : cada observación	f : frecuencia
	$\bar{x}$: media aritmética	x_m : marca de clase
	N : tamaño de la población	n : tamaño de la muestra

4.3.3. Desviación Estándar

Es una medida de dispersión que indica el grado de variación de un conjunto de datos con respecto a la media, una desviación baja indica una mayor concentración de los datos en torno a la media, una desviación alta indica una mayor dispersión.

Características

- La desviación estándar es la raíz cuadrada de la varianza, su valor es positivo y se expresa en las mismas unidades que los datos originales lo que facilita su interpretación.
- Se ve afectado significativamente por la presencia de valores atípicos, si se presentan desviaciones altas
- Se extiende su uso en la Estadística Inferencial, en la estimación de intervalos de confianza, pruebas de hipótesis y como un parámetro de la Distribución Normal.

Tabla 10

Fórmulas Desviación Estándar

	Datos No agrupados	Datos agrupados
Poblacional	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$	$\sigma = \sqrt{\frac{\sum_{i=1}^N f(x_m - \mu)^2}{N}}$
Muestral	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$	$s = \sqrt{\frac{\sum_{i=1}^n f(x_m - \bar{x})^2}{n - 1}}$
	x_i : cada observación	f : frecuencia
	$\bar{x}$: media aritmética	x_m : marca de clase
	N : tamaño de la población	n : tamaño de la muestra

Ejercicio Práctico N°1:

Los tiempos de respuesta en milisegundos de un servidor a diferentes requerimientos son: 121, 127, 119, 117, 109, 105, 123 y 115. Encuentre la varianza y la desviación estándar.

$$\bar{x} = \frac{121 + 127 + 119 + 117 + 109 + 105 + 123 + 115}{8}$$

$$\bar{x} = 117$$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
-------	-----------------	---------------------

121	4	16
117	10	100
119	2	4
117	0	0
109	-8	64
105	-12	144
123	6	36
115	-2	4
$\sum (x_i - \bar{x})^2$		368

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

$$s^2 = \frac{368}{8 - 1}$$

$$s^2 = 52.57 \text{ (Varianza)}$$

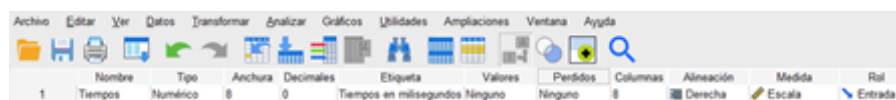
$$s = 7.25 \text{ (Desviación estándar)}$$

Construcción en SPSS:

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Tiempos con los siguientes parámetros:

- **Nombre:** Tiempos
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Tiempos en milisegundos
- **Medida:** Escala



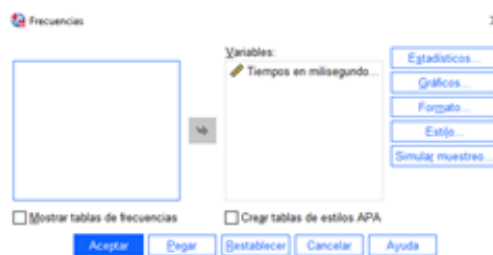
- c. En la ventana “Vista de datos” ingrese los tiempos de respuesta del servidor a los diferentes requerimientos.

	Tiempos
1	121
2	127
3	119
4	117
5	109
6	105
7	123
8	115

Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- a. En la ventana que aparece mueva la variable Tiempos al cuadro de variables seleccionadas.



- b. Presione el botón Estadísticos, dé un clic en el casillero “Desv. estándar” y en “Varianza”



- c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado obtenido de la desviación estándar y de la varianza.

Estadísticos		
datos		
N	Válido	8
	Perdidos	0
Desv. Desviación		7,25062
Varianza		52,571

4. Interpretación

La mayoría de los tiempos de respuesta de un servidor a diferentes requerimientos se encuentran entre 109.75 y 124.25 milisegundos (117 ± 7.25), en términos generales hay poca dispersión de los datos.

Ejercicio práctico N°2

Las horas extras semanales laboradas por los empleados de una empresa se muestran en la siguiente tabla de frecuencias. Encuentre la varianza y la desviación estándar.

N° clases	LI	LS	f_i	x_m	$f_i * x_m$	$x_m - \bar{x}$	$(x_m - \bar{x})^2$	$f_i * (x_m - \bar{x})^2$
1	0	3	3	1.5	4.5	-9.83	96.53	289.59
2	3	6	5	4.5	22.5	-6.83	46.58	232.90
3	6	9	6	7.5	45	-3.83	14.63	87.78
4	9	12	6	10.5	63	-0.82	0.68	4.08
5	12	15	8	13.5	108	2.18	4.73	37.85
6	15	18	8	16.5	132	5.18	26.78	214.25
7	18	21	4	19.5	78	8.18	66.83	267.32
					453			1133.78

Promedio:

$$\bar{x} = \frac{\sum f_i * x_m}{n}$$

$$\bar{x} = \frac{453}{40}$$

$$\bar{x} = 11.33$$

Cálculo de la Varianza:

$$s^2 = \frac{\sum_{i=1}^n f_i (x_m - \bar{x})^2}{n - 1}$$

$$s^2 = \frac{1133.78}{39}$$

$$s^2 = 29.07$$

Cálculo de la Desviación estándar

$$s = \sqrt{s^2}$$

$$s = \sqrt{29.07}$$

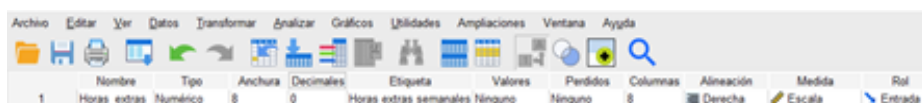
$$s = 5.39$$

Construcción en SPSS

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada horas extras con los siguientes parámetros:

- **Nombre:** Tiempo
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Horas extras semanales
- **Medida:** Escala



- En la ventana “Vista de datos” ingrese las horas extras semanales laboradas por los empleados de la empresa

Archivo Editar Ver Data																																													
Horas extras <table border="1"> <tr><td>1</td><td>4</td></tr> <tr><td>2</td><td>14</td></tr> <tr><td>3</td><td>4</td></tr> <tr><td>4</td><td>13</td></tr> <tr><td>5</td><td>18</td></tr> <tr><td>6</td><td>16</td></tr> <tr><td>7</td><td>16</td></tr> <tr><td>8</td><td>20</td></tr> <tr><td>9</td><td>9</td></tr> <tr><td>10</td><td>5</td></tr> <tr><td>11</td><td>7</td></tr> <tr><td>12</td><td>12</td></tr> <tr><td>13</td><td>13</td></tr> <tr><td>14</td><td>13</td></tr> <tr><td>15</td><td>11</td></tr> <tr><td>16</td><td>18</td></tr> <tr><td>17</td><td>2</td></tr> <tr><td>18</td><td>15</td></tr> <tr><td>19</td><td>4</td></tr> <tr><td>20</td><td>13</td></tr> <tr><td>21</td><td>19</td></tr> <tr><td>22</td><td>16</td></tr> </table>		1	4	2	14	3	4	4	13	5	18	6	16	7	16	8	20	9	9	10	5	11	7	12	12	13	13	14	13	15	11	16	18	17	2	18	15	19	4	20	13	21	19	22	16
1	4																																												
2	14																																												
3	4																																												
4	13																																												
5	18																																												
6	16																																												
7	16																																												
8	20																																												
9	9																																												
10	5																																												
11	7																																												
12	12																																												
13	13																																												
14	13																																												
15	11																																												
16	18																																												
17	2																																												
18	15																																												
19	4																																												
20	13																																												
21	19																																												
22	16																																												

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- En la ventana que aparece mueva la variable Horas extras al cuadro de variables seleccionadas.



- Presione el botón Estadísticos, dé un clic en el casillero “Desv. Estándar” y “Varianza”.



- Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

Estadísticos		
Horas extras semanales		
N	Válido	40
	Perdidos	0
Desv. Desviación		5,425
Varianza		29,436

4. Interpretación

Las horas extras semanales tienen una desviación de 5.42 horas respecto a la media. La mayoría de los tiempos de respuesta de un servidor a diferentes requerimientos se encuentran entre 109.75 y 124.25 milisegundos (117 ± 7.25), en términos generales hay poca dispersión de los datos.

4.3.4. Coeficiente de Variación

El coeficiente de variación se utiliza para comparar la dispersión de diferentes conjuntos de datos respecto a la media.

Características

- El coeficiente de variación solo muestra la relación entre la desviación estándar y la media y no da información sobre la forma de su distribución, sin embargo, es fácil de calcular y facilita su interpretación en términos de porcentaje.
- Si la media es cero no existe coeficiente de variación ya que la división para cero no está definida, si la media es aproximadamente cero, el coeficiente de variación puede volverse extremadamente grande y su resultado ser poco fiable e interpretable.
- Si los datos tienen poca variabilidad en torno a la media, se obtendrá un coeficiente de variación bajo, esto significa una mayor concentración de los datos alrededor de la media.

- Si la variación de los datos es alta en torno a la media, se obtendrá un coeficiente de variación mayor, esto significa que los datos se encuentren más dispersos alrededor de la media.

Tabla 11

Fórmulas Coeficiente de Variación

	Población	Muestra
Coeficiente de Variación	$CV = \left(\frac{\sigma}{\mu} \right) * 100\%$	$V = \left(\frac{s}{\bar{x}} \right) * 100\%$
	σ : desviación estándar	s : desviación estándar
	μ : media aritmética	$\bar{x}$: media aritmética

Ejercicio Práctico N°1

La edad de los estudiantes que acuden a un prestigiosa Universidad del centro del País para cursar la carrera de Contabilidad y Auditoría en los turnos matutino y vespertino, se describe a continuación.

Turno matutino	19	22	17	21	17	19	20	22	18	17
Turno vespertino	17	26	24	23	27	21	21	25	24	27

¿Si la homogeneidad de la clase es un factor positivo en el aprendizaje, en qué grupo será más fácil enseñar?

Turno matutino:

$$\bar{x} = 19.20$$

$$s = 1.99$$

$$CV = \frac{1.99}{19.20} * 100\%$$

$$CV = 10.36\%$$

Turno vespertino:

$$\bar{x} = 23.50$$

$$s = 3.14$$

$$CV = \frac{3.14}{23.50} \times 100\%$$

$$CV = 13.36\%$$

El resultado nos indica que las edades de los estudiantes en el turno matutino son más homogéneas lo cual facilitaría el proceso de aprendizaje.

Construcción en SPSS:

1. Prepare los datos:

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree las variables:

- **Nombre:** Matutino
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Edad de los estudiantes en el turno matutino
- **Medida:** Escala
- **Nombre:** Vespertino
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Edad de los estudiantes en el turno vespertino
- **Medida:** Escala
- **Nombre:** Unidades
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Variable temporal
- **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Matutino	Númerico	8	0	Edad de los estudiantes en el turno matutino	Ninguno	Ninguno	8	Centro	Escala	Entrada
2	Vespertino	Númerico	8	0	Edad de los estudiantes en el turno vespertino	Ninguno	Ninguno	8	Centro	Escala	Entrada
3	Unidad	Númerico	8	0	Variable temporal	Ninguno	Ninguno	8	Centro	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese las edades correspondientes a los dos turnos y en la variable unidad ingrese solamente números “1” (esto servirá para cálculos posteriores).

	Matutino	Vespertino	Unidad
1	19	17	1
2	22	26	1
3	17	24	1
4	21	23	1
5	17	27	1
6	19	21	1
7	20	21	1
8	22	25	1
9	18	24	1
10	17	27	1

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Razón.

- a. En la ventana que aparece mueva las variables Matutino al casillero Numerador y la variable Unidad al casillero Denominador.

- b. Presione el botón Estadísticos, dé un clic en el casillero “Media”, “CDV centrado en la media” y “Desviación estándar”.

- c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

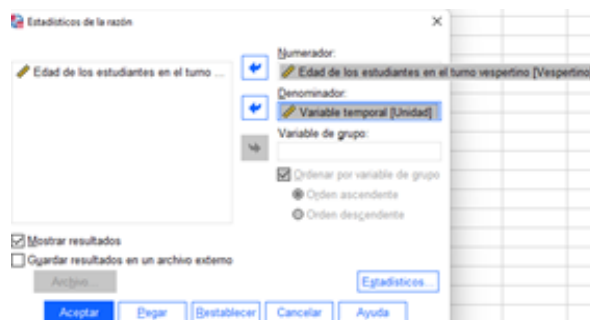
Estadísticos de la razón para Edad de los estudiantes en el turno matutino / Variable temporal		
	Desv.	Coefficiente de variación
Media	Desviación	Media centrada
19,200	1,989	10,4%

Repetimos el punto 2 del análisis, pero ahora con la variable Turno vespertino.

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Razón.

- a. En la ventana que aparece mueva las variables Vespertino al casillero Numerador y la variable Unidad al casillero Denominador.



- b. Presione el botón Estadísticos, dé un clic en el casillero “Media”, “CDV centrado en la media” y “Desviación estándar”.



c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

Estadísticos de la razón para Edad de
los estudiantes en el turno vespertino /
Variable temporal

		Coficiente de variación
	Desv.	Media
Media	Desviación	centrada
23,500	3,136	13,3%

4. Interpretación

El resultado nos indica que las edades de los estudiantes en el turno matutino

CV=10.4% son más homogéneas que las del turno vespertino CV=13.3% lo cual facilitaría el proceso de aprendizaje.

Autoevaluación 4.3.

1. En los salarios de la Empresa pública de agua potable, qué significa una desviación estándar baja en el conjunto de datos
2. ¿Qué medida es más sensible ante la presencia de outliers?, el rango o el rango intercuartílico y por qué
3. Luego de una evaluación escrita de la asignatura de Matemática Financiera se han formado dos grupos con la siguiente información:

Grupo A: media = 17, desviación estándar = 5

Grupo B: media = 13, desviación estándar = 8

Calcule el coeficiente de variación de ambos grupos e indique cuál de los dos presentan notas más homogéneas y por qué

4.3.5. Ejercicios propuestos

21. Una pequeña empresa de la ciudad de Latacunga, registra las edades de 10 empleados: 22, 34, 29, 31, 40, 27, 35, 30, 24, 30. Calcular:
 - a. El rango de las edades
 - b. La varianza y la desviación estándar
 - c. El coeficiente de variación
22. Las calificaciones obtenidas por un curso de 25 estudiantes en el examen final del parcial de la materia de matemática básica son: 8, 7, 9, 6, 10, 6, 8, 9, 3, 8, 10, 7, 7, 8, 9, 5, 6, 7, 8, 3, 4, 3, 8, 9, 7. Calcula la desviación estándar de las calificaciones.
23. La batería de un dispositivo electrónico muy costoso presenta problemas, por lo que un técnico decide medir el voltaje de la batería cinco veces y obtiene los siguientes resultados: 3.0V, 3.2V, 3.3V, 3.1V, 3.1V. Determina la varianza de los valores medidos para así encontrar la falla de la batería.

24. Una importante fábrica productora de leche tiene en nómina a 100 trabajadores cuyos salarios son los que se muestran en la siguiente tabla:

Salarios	# trabajadores
[550 - 600)	29
[600 - 650)	17
[650 - 700)	33
[700 - 750)	11
[750 - 800)	10

Calcula la desviación estándar y el coeficiente de variación de los salarios.

25. Una fábrica ensambladora de partes automotrices posee dos robots que se encargan de armar una pieza; se realizaron 6 mediciones del tiempo en segundos que tardan en armar la pieza los dos robots:

- Alfa Robot: 12, 13, 14, 12, 13, 14 segundos
- Beta Robot: 14, 12, 10, 16, 8, 18 segundos

Usa la desviación estándar para conocer que robot muestra mayor estabilidad en los tiempos de armado de la pieza.

26. Las temperaturas máximas en grados centígrados registrados en la ciudad de Ambato la semana pasada fueron: 21, 18, 18, 17, 19, 17, 18 grados centígrados. Calcula el rango, la varianza y el coeficiente de variación de las temperaturas.
27. La empresa CrosBoarding recoge de la aduana 30 paquetes cuyos pesos en kilogramos son:

5	1	2	2	3	6	1	5	1	1
7	4	4	3	4	1	4	5	7	4
3	1	3	7	3	3	2	5	1	1

Calcula la desviación estándar y el coeficiente de variación de los pesos de los 30 paquetes.

28. Los estudiantes de primer semestre de la carrera de ingeniería electrónica están aprendiendo a medir las resistencias (k-ohmios) de algunos componentes básicos, por lo que 4 estudiantes realizan las mediciones 4 veces, obteniendo los siguientes resultados:

- Andrés: 9.6, 9.7, 9.8, 9.9
- Camila: 9.7, 9.4, 10.2, 9.8
- Esteban: 9.5, 9.4, 9.9, 10.8
- María: 9.6, 10.3, 9.8, 9.9

Calcula la desviación estándar para saber quién tiene más precisión al medir la resistencia de los componentes básicos.

29. Una fábrica de peluches trabaja en diversos turnos distribuidos semanalmente, en cada turno se producían una cierta cantidad de peluches defectuosos, los mismos que se muestran en la siguiente tabla:

Peluches Defectuosos	Turnos
0	4
1	6
2	5
3	3
4	2

Calcula la varianza y desviación estándar de peluches defectuosos por turno.

30. En el mercado de la electrónica existen dos marcas de resistencias eléctricas SMD y AXC cada una tiene resistencias con mediciones (en ohmios) diferentes como se muestra a continuación:

- SMD: 99, 101, 100, 102, 98
- AXC: 200, 198, 202, 195, 205,

Calcular el coeficiente de variación para saber ¿qué marca de resistencias tiene menor variabilidad relativa?

4.4. Medidas de forma

Son aquellas que describen la forma de una distribución, hacen referencia a la asimetría (uniformidad de los datos respecto a la media sesgo) y la curtosis (picos y colas) de una curva.

4.4.1. Asimetría

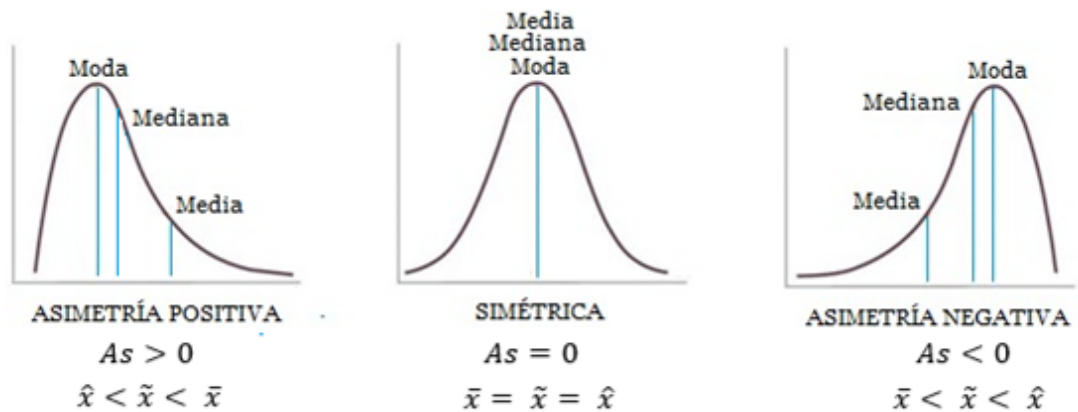
Si los datos se distribuyen de manera equitativa respecto a la media, la distribución de frecuencias es simétrica, caso contrario se puede destacar dos tipos de distribución asimétrica:

Asimetría positiva: Las frecuencias más altas se encuentran en el lado izquierdo de la media, en el derecho se encuentran las más pequeñas.

Asimetría negativa: Las frecuencias más altas se encuentran en el lado derecho de la media, en el lado izquierdo se encuentran las más pequeñas.

Figura 36

Asimetría



Coefficiente de Asimetría de Fisher

Mide el grado de asimetría de la distribución con respecto a la media.

Tabla 12

Fórmulas Coeficiente de Asimetría de Fisher

Datos No agrupados	Datos agrupados
$As = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n s^3}$	$As = \frac{\sum_{i=1}^n f_i (x_m - \bar{x})^3}{n s^3}$
x_i : cada observación	x_m : marca de clase
$\bar{x}$: media muestral	$\bar{x}$: media muestral
n : tamaño de la muestra	n : tamaño de la muestra
s : desviación estándar	s : desviación estándar
	f_i : frecuencia absoluta

La fórmula establecida presenta un sesgo para muestras pequeñas, por lo que se muestra la versión corregida del Coeficiente de Asimetría de Fisher también conocida como asimetría muestral.

Corrección del Coeficiente de Asimetría de Fisher

Tabla 13

Fórmulas Corrección del Coeficiente de Asimetría de Fisher

Datos No agrupados	Datos agrupados
$As = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$	$As = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^n f_i * (x_m - \bar{x})^3}{s^3}$
x_i : cada observación	x_m : marca de clase
$\bar{x}$: media muestral	$\bar{x}$: media muestral
n : tamaño de la muestra	n : tamaño de la muestra
s : desviación estándar	s : desviación estándar
	f_i : frecuencia absoluta

Nota. Tomado de Lind Marshall, pág. 120

Esta fórmula incluye un factor de corrección $\frac{n}{(n-1)(n-2)}$, esta corrección permite mejorar la estimación en muestras pequeñas reduciendo el sesgo. Programas como Excel, Python y SPSS utilizan esta corrección para realizar los cálculos internos y encontrar el valor del coeficiente de asimetría.

4.4.2. Curtosis

La Curtosis (Apuntamiento) es una medida estadística que determina el alargamiento o estreches de las colas y describe el nivel de concentración de los datos respecto de su media

Coeficiente de Curtosis de Fisher

Tabla 14

Fórmulas Coeficiente de Curtosis

Datos No agrupados	Datos agrupados
$K = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}}{s^4}$	$K = \frac{\frac{\sum_{i=1}^n (x_i - \bar{x})^4 * f_i}{n}}{s^4}$
x_i : cada observación	x_i : cada observación
$\bar{x}$: media aritmética	$\bar{x}$: media aritmética
n : tamaño de la población	n : tamaño de la población
s : desviación estándar	s : desviación estándar
	f_i : frecuencia absoluta

Curtosis Leptocúrtica (K>3)

Los datos tienen una alta concentración alrededor de la media, tiene un pico alto y pronunciado y las colas son más largas, es común tener la presencia de valores extremos.

Curtosis Mesocúrtica (K=3)

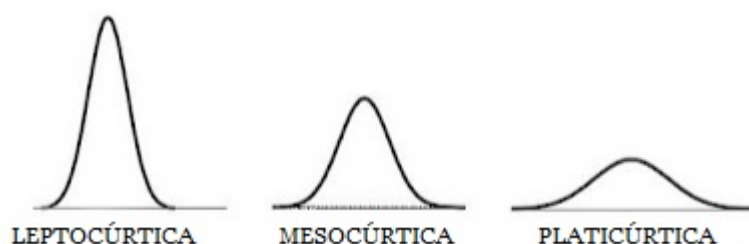
Los datos se concentran en una forma “normal”, haciendo referencia a una distribución de probabilidad normal, el pico de la curva y las colas son moderadas.

Curtosis Platicúrtica ($K < 3$)

Los datos tienen una concentración baja alrededor de la media haciendo que la forma de la curva sea más achatada y las colas son más cortas y planas, se tiene poca probabilidad de encontrar valores extremos o atípicos.

Figura 37

Tipos de Curtosis



Exceso de Curtosis

El exceso de curtosis es una medida estadística que se utiliza para comparar las colas de una distribución de probabilidad con las de la distribución normal.

El exceso de curtosis se obtiene restando 3 la curtosis de Fisher:

$$K_{exceso} = K - 3$$

Tabla 15

Fórmulas Exceso de Curtosis

Datos No agrupados	Datos agrupados
$K_{exceso} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n s^4} - 3$	$K_{exceso} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4 * f_i}{n s^4} - 3$
x_i : cada observación	x_i : cada observación
$\bar{x}$: media aritmética	$\bar{x}$: media aritmética
n : tamaño de la población	n : tamaño de la población
s : desviación estándar	s : desviación estándar
	f_i : frecuencia absoluta

Curtosis Leptocúrtica ($K_{\text{exceso}} > 0$)

La distribución tiene colas más pesadas que la distribución normal, esto significa que la distribución tiene valores extremos más frecuentes o valores más alejados de la media que en una distribución normal.

Curtosis Mesocúrtica ($K_{\text{exceso}} = 0$)

La distribución tiene colas similares a las de la distribución normal, es decir no hay la presencia de valores tan extremos de los que se espera tener en una distribución normal.

Curtosis Platicúrtica ($K_{\text{exceso}} < 0$)

La distribución tiene colas más ligeras que la distribución normal, los datos se encuentran más dispersos y hay menos valores extremos de los que se espera tener en una distribución normal.

Corrección del Coeficiente de Curtosis de Fisher

La versión corregida del coeficiente de curtosis reduce el sesgo en muestras pequeñas con el objetivo de tener una mejor estimación al valor poblacional real.

$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^n (x_m - \bar{x})^4 * f_i}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Ejercicio práctico N°1

Los tiempos en minutos que se toman 10 empleados de una empresa para realizar un Informe diario de actividades son los siguientes:

9	10	10	12	11	13	11	9	10	14
---	----	----	----	----	----	----	---	----	----

Encuentre el coeficiente de asimetría corregido de Fisher e interprete sus resultados.

Generamos la tabla de frecuencias:

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$
9	-1.90	3.61	-6.86
10	-0.90	0.81	-0.73
10	-0.90	0.81	-0.73
12	1.10	1.21	1.33
11	0.10	0.01	0.00
13	2.10	4.41	9.26
11	0.10	0.01	0.00
9	-1.90	3.61	-6.86
10	-0.90	0.81	-0.73
14	3.10	9.61	29.79
		24.90	24.48

Media aritmética:

$$\bar{x} = \frac{9 + 10 + 10 + 12 + 11 + 13 + 11 + 9 + 10 + 14}{10}$$

$$\bar{x} = 10.9$$

Desviación estándar:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

$$s = \sqrt{\frac{24.90}{9}}$$

$$s = 1.66$$

Coefficiente de Asimetría corregido de Fisher:

$$As = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{s^3}$$

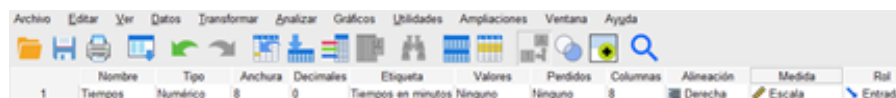
$$As = \frac{10}{9 * 8} * \frac{24.48}{(1.66)^3}$$

$$As = 0.74$$

Construcción en SPSS:**1. Prepare los datos:**

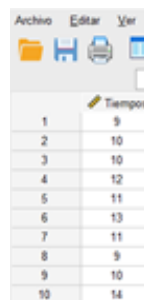
- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree una nueva variable llamada Tiempo con los siguientes parámetros:

- **Nombre:** Tiempos
- **Tipo:** Numérico
- **Decimales:** Cero
- **Etiqueta:** Tiempos en minutos
- **Medida:** Escala



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Pérdidos	Columnas	Alineación	Medida	Rol
1	Tiempos	Numérico	8	0	Tiempos en minutos	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- En la ventana “Vista de datos” ingrese los tiempos en que se demoran en hacer un informe los empleados de la empresa.



	Tiempos
1	9
2	10
3	10
4	12
5	11
6	13
7	11
8	9
9	10
10	14

2. Análisis descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

- En la ventana que aparece mueva la variable Tiempos al cuadro de variables seleccionadas



- b. Presione el botón Estadísticos, dé un clic en el casillero “Media”, “Desv. estándar” y “Asimetría”



- c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

Estadísticos		
Tiempos en minutos		
N	Válido	10
	Perdidos	0
Media		10,90
Desv. Desviación		1,663
Asimetría		,739
Error estándar de asimetría		,687

4. Interpretación

Se tiene una leve asimetría positiva, esto significa que los datos más frecuentes se concentran en el extremo inferior de la curva, mientras que los datos más grandes son menos repetitivos y generan la cola que se extiende hacia la derecha.

Ejercicio práctico N°2

Las temperaturas en grados celsius que se registraron en la ciudad de Guayaquil en el mes de diciembre 2024 fueron las siguientes:

31	26	33	26	32	27	30	32	33	25	25	26	25	33
33	32	30	27	26	31	31	23	28	32	33	23	33	30
27	28	29	31	25	31	33	31	28	29	26	22	30	28

Determine los coeficientes de asimetría y curtosis corregidos de Fisher e interprete sus resultados.

Generamos la Tabla de frecuencia

K	L_i	L_s	f_i	X_m	$f_i X_m$	$f_i(X_m - \bar{x})^2$	$f_i(X_m - \bar{x})^3$	$f_i(X_m - \bar{x})^4$
1	21,5	23,5	3	22,5	67,5	120,33	-762,11	4826,70
2	23,5	25,5	4	24,5	98	75,11	-325,48	1410,42
3	25,5	27,5	8	26,5	212	43,56	-101,63	237,14
4	27,5	29,5	6	28,5	171	0,67	-0,22	0,07
5	29,5	31,5	10	30,5	305	27,78	46,30	77,16
6	31,5	33,5	11	32,5	357,5	147,89	542,26	1988,28
					1211	415,33	-600,89	8539,78

La media aritmética es:

$$\bar{x} = \frac{\sum_{i=1}^n f_i x_m}{n}$$

$$\bar{x} = \frac{1211}{42}$$

$$\bar{x} = 28,83$$

La desviación estándar es:

$$s = \sqrt{\frac{\sum_{i=1}^n f_i (x_m - \bar{x})^2}{n - 1}}$$

$$s = \sqrt{\frac{415,33}{42-1}}$$

$$s = 3,18$$

El coeficiente de asimetría corregido de Fisher es:

$$As = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^n f_i (x_m - \bar{x})^3}{s^3}$$

$$As = \frac{42}{(42-1)(42-2)} \frac{-600,89}{(3,18)^3}$$

$$As = -0.48$$

El coeficiente de curtosis corregido es:

$$K = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \frac{(x_m - \bar{x})^4 f_i}{s^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

$$K = \frac{42(43)}{(41)(40)(39)} \cdot \frac{8539,78}{(3,18)^4} - \frac{3(41)^2}{(40)(39)}$$

$$K = -0.87$$

Construcción en SPSS:

1. Prepare los datos:

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana "Vista de variables" cree las siguientes variables:

- **Nombre:** Temperatura
- **Tipo:** Numérico
- **Decimales:** 0
- **Etiqueta:** Temperatura en grados Celsius
- **Valores:** asignar el valor de 1 a la categoría "21.5 - 23.5", 2 a la categoría "23.5 - 25.5", 3 a la categoría "25.5 - 27.5" y así sucesivamente hasta el valor 6 a la categoría "31.5 - 33.5"
- **Medida:** Ordinal

- **Nombre:** Frecuencia
 - **Tipo:** Numérico
 - **Decimales:** 0
 - **Etiqueta:** Frecuencia de repetición
 - **Medida:** Escala
-
- **Nombre:** M_Clase
 - **Tipo:** Numérico
 - **Decimales:** 0
 - **Etiqueta:** Marca de clase
 - **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Temperatura	Númerico	8	0	Temperatura en grados celcius (1, 21.5 a 2...	Ninguno	Ninguno	8	Centro	Ordinal	Entrada
2	Frecuencia	Númerico	8	0	Frecuencia de repetición	Ninguno	Ninguno	8	Centro	Escala	Entrada
3	M_clase	Númerico	8	1	Marca de clase	Ninguno	Ninguno	8	Centro	Escala	Entrada

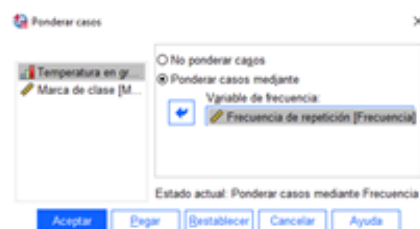
c. En la ventana “Vista de datos” ingrese las temperaturas registradas

	Temperatura	Frecuencia	M_clase
1	21.5 a 23.5	3	22.5
2	23.5 a 25.5	4	24.5
3	25.5 a 27.5	8	26.5
4	27.5 a 29.5	6	28.5
5	29.5 a 31.5	10	30.5
6	31.5 a 33.5	11	32.5

2. Análisis descriptivo:

En el menú principal seleccione: Datos / Ponderar casos.

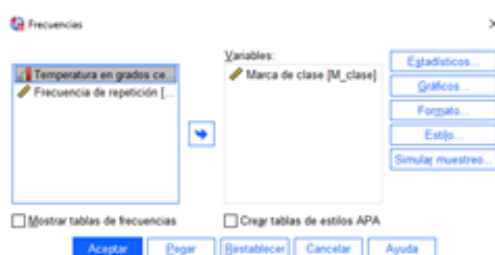
- a. En la ventana que aparece seleccione la opción Ponderar casos y ahora seleccione la variable Frecuencias, arrastre esta variable al recuadro que dice Variable de frecuencia:



- b. Presione el botón Aceptar

En el menú principal Analizar / Estadísticos Descriptivos / Frecuencias

- a. En la ventana que aparece mueve la variable Marca de clase al cuadro de variables seleccionadas



- b. Presione el botón Estadísticos, dé un clic en el casillero “Media”, “Desv. estándar”, “Asimetría” y “Curtosis”



- c. Presione el botón Continuar y luego Aceptar

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la medida calculada.

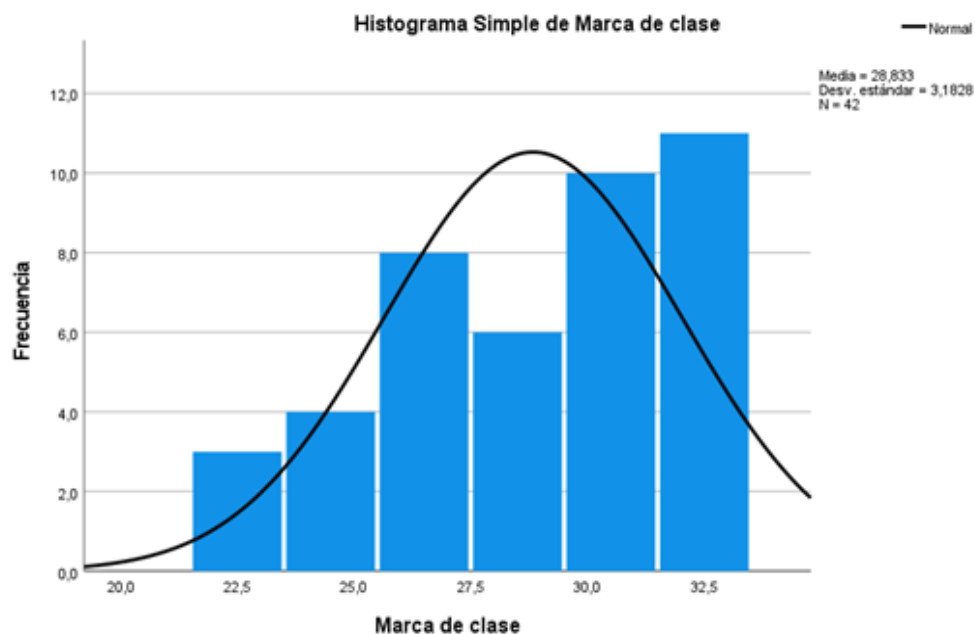
Estadísticos		
Marca de clase		
N	Válido	42
	Perdidos	0
Media		28,833
Desv. Desviación		3,1828

Asimetría	-,477
Error estándar de asimetría	,365
Curtosis	-,883
Error estándar de curtosis	,717

4. Interpretación

La asimetría negativa se puede evidenciar en el histograma que se encuentra a continuación, en él se puede encontrar que las temperaturas predominantes en el mes de diciembre en la ciudad de Guayaquil son aquellas que oscilan entre 30°C y 33°C, presentándose temperaturas dispersas de 21°C a 23°C.

Un coeficiente de curtosis de -0.88 muestra una curva Platicúrtica (la curva es más achatada respecto a una curva normal), esto indica que existen menos cantidad de valores concentrados en torno a la media aritmética, así como también la poca presencia de valores muy extremos en la distribución.



Ejercicio práctico N° 3 (con Base de Datos)

Utilizando la base de datos Wine Quality White que se encuentra en el siguiente enlace:

Accede aquí:



Realiza un análisis estadístico descriptivo de la variable dióxido de azufre total (total_sulfur_dioxide) e interpreta sus resultados:

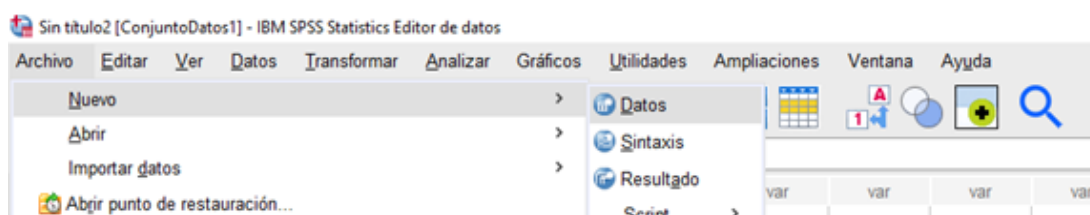
Solución:

1. Descargar la base de datos:

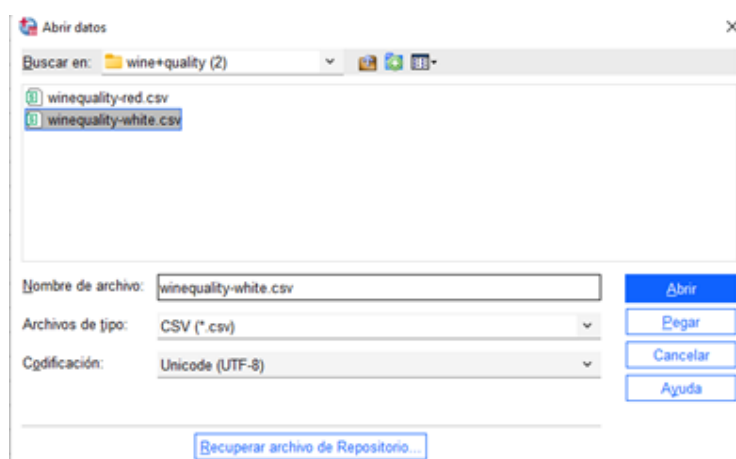
Dirigirse al enlace propuesto y en la parte superior derecha dar clic en descargar.



2. Abrir el SPSS y dar clic en abrir, datos:



3. En esta ventana cambiamos el tipo de archivo a *.csv y buscamos la ubicación de la base de datos descargada y clic en abrir:



4. Configuramos los parámetros de visualización y se finaliza la carga de los datos:

Asistente para la importación de texto - Paso 6 de 6

Ha definido el formato del archivo de texto correctamente...

¿Desea guardar este formato de archivo para utilizarlo en el futuro?

☐ Sí ☐ No [Guardar como...](#)

¿Desea pegar la sintaxis?

☐ Sí ☒ No ☒ Caché local ...

Pulse en el botón Finalizar para finalizar el Asistente p...

Vista previa de datos

	fixedacidity	volatileacidity	citricacid	residualsugar	chlorides	free sulfur dioxide	total sulfur dioxide
7	0.27	0.36	20.7	0.045	45	170	1.001
6.3	0.3	0.34	1.6	0.049	14	132	0.994
8.1	0.28	0.4	6.9	0.05	30	97	0.9951
7.2	0.23	0.32	8.5	0.058	47	186	0.9956
7.2	0.23	0.32	8.5	0.058	47	186	0.9956
8.1	0.28	0.4	6.9	0.05	30	97	0.9951

< >

< Anterior Siguiente > Finalizar Cancelar Ayuda

5. Ahora se puede visualizar la información de todos los datos que contiene la base de datos calidad de vino blanco.

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

Visible: 12 de 12 variables

	fixed acidity	volatile acidity	citric acid	residualsugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol	quality
1	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6
2	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6
3	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6
4	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6
5	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9	6
6	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1	6
7	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.18	0.47	9.6	6
8	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8	6
9	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5	6
10	8.1	0.22	0.43	1.5	0.044	28	129	0.9938	3.22	0.45	11	6
11	8.1	0.27	0.41	1.45	0.033	11	63	0.9908	2.99	0.56	12	5
12	8.6	0.23	0.04	4.2	0.035	17	109	0.9947	3.14	0.53	9.7	5
13	7.9	0.18	0.37	1.2	0.04	16	75	0.992	3.18	0.63	10.8	5
14	6.6	0.16	0.04	1.5	0.044	48	143	0.9912	3.54	0.52	12.4	7
15	8.3	0.42	1.02	19.25	0.04	41	172	1.0002	2.98	0.67	9.7	5
16	6.6	0.17	0.38	1.5	0.032	28	112	0.9914	3.25	0.55	11.4	7
17	6.3	0.48	0.04	1.1	0.046	30	99	0.9928	3.24	0.36	9.6	6
18	6.2	1.06	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8
19	7.4	0.34	0.42	1.1	0.033	17	171	0.9917	3.12	0.53	11.3	6
20	6.5	0.31	0.14	7.5	0.044	34	133	0.9955	3.22	0.5	9.5	5
21	6.2	1.06	0.48	1.2	0.029	29	75	0.9892	3.33	0.39	12.8	8

Vista de datos Vista de variables

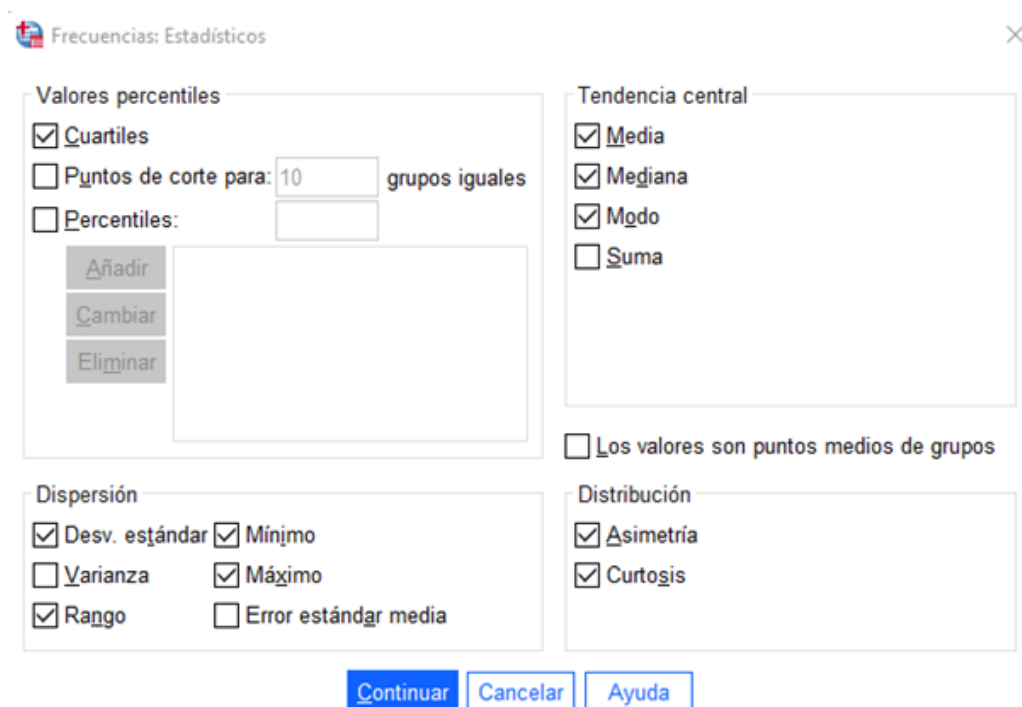
6. Análisis Descriptivo:

En el menú principal seleccione: Analizar / Estadísticos Descriptivos / Frecuencias.

a. En la ventana que aparece mueva la variable totalsulfurdioxide.



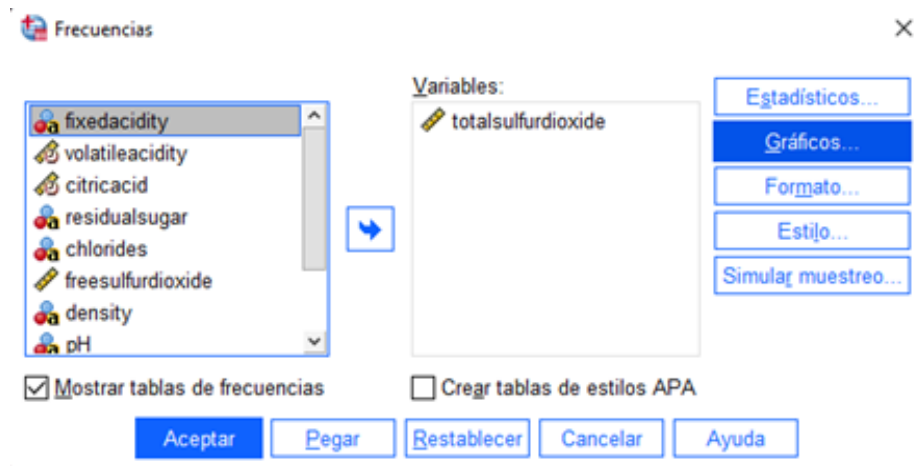
b. Presione el botón Estadísticos, dé un clic en los casilleros siguientes:



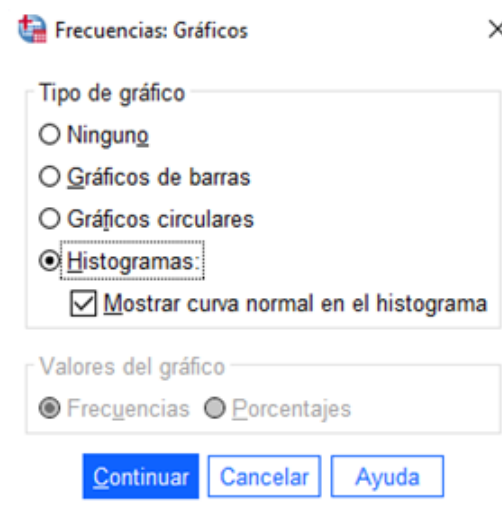
c. Presione el botón Continuar y luego Aceptar

7. Resultados obtenidos:

En la misma ventana de Frecuencias elegir ahora la opción Gráficos:



En la ventana Frecuencias: Gráficos, seleccionar:



Presionar Continuar y Aceptar

Resultados obtenidos

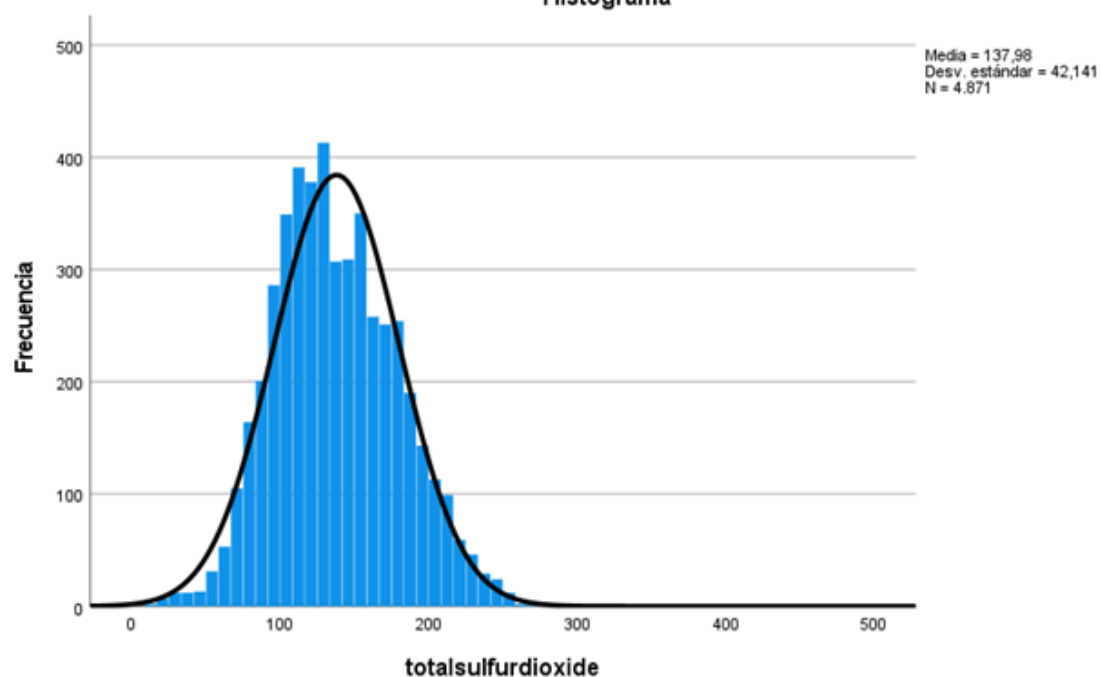
En la Ventana de Resultados, encontrará el resultado de las medidas calculadas.

Estadísticos

totalsulfurdioxide

N	Válido	4871
	Perdidos	27
Media		137,98
Mediana		134,00
Moda		111
Desv. Desviación		42,141
Asimetría		,359
Error estándar de asimetría		,035
Curtosis		,446
Error estándar de curtosis		,070
Rango		431
Mínimo		9
Máximo		440
Percentiles	25	108,00
	50	134,00
	75	167,00

Histograma



Interpretación:

La información contenida en la base de datos está compuesta por 4898 datos, de los cuales 27 de ellos (menos del 1%) no revelan información y se consideran como datos perdidos, este porcentaje no genera una mayor representatividad en el análisis. Los resultados indican que en promedio las muestras de vino contienen un 137.98 mg/L de dióxido de azufre total con una variación promedio de 42 mg/L en relación a la media calculada, el 50% de las observaciones tienen valores iguales o menores a 134 mg/L y los datos más frecuentes son 111, el 25% de las muestras tienen valores iguales o menores a 108 mg/L y el 75% con valores menores o iguales a 167 mg/L. La distribución tiene valores registrados mínimos de 9 mg/L y máximos de 440 mg/L, mostrando una amplia variabilidad equivalente a 431 mg/L.

En cuanto a la forma de su distribución sugiere una curva levemente Leptocúrtica (coeficiente de curtosis de 0.446), esto indica que los datos tienden a concentrarse más alrededor de la media a diferencia de una distribución normal, mostrando una asimetría positiva (coeficiente de asimetría de 0.359), con una cola extendida hacia la derecha por la posible existencia de valores atípicos.

Autoevaluación 4.4.

1. ¿Qué representan las medidas de forma y cómo se complementan con el análisis de las medidas de tendencia central y dispersión en un conjunto de datos?
2. ¿Qué explicación daría si en el análisis de una distribución de datos como resultado se presenta una curva leptocúrtica y un sesgo positivo? ¿Qué sucede con las medidas de tendencia central?
3. Los salarios (en dólares) de 10 funcionarios de una empresa pública son los siguientes: 1100, 950, 2200, 2500, 1800, 900, 2300, 1500, 1200, 1250.
¿Cómo es la forma de su distribución: simétrica o asimétrica, en el caso de ser asimétrica, qué tipo de sesgo tiene? Interpreta su resultado

4.4.3. Ejercicios propuestos

31. Una empresa que fabrica muebles en el centro del país, realiza un estudio sobre los ingresos mensuales de sus empleados (en cientos de dólares), los que se muestran ordenados en la siguiente tabla:

Ingresos	# Empleados
[4 - 6)	5
[6 - 8)	12
[8 - 10)	20
[10 - 12)	10
[12 - 14]	3
	50

Calcula el coeficiente de asimetría de Pearson y determina si la distribución es simétrica, asimétrica positiva o asimétrica negativa.

32. En la siguiente tabla se muestran las estaturas de un grupo de estudiantes de la carrera de Ingeniería de Software de la ESPE, los que saldrán a una gira fuera del país:

Estatura (cm)	# estudiantes
[150 - 155)	2
[155 - 160)	8
[160 - 165)	15
[165 - 170)	10
[170 - 175)	5
	40

Se desea calcular el coeficiente de Curtosis y el coeficiente de asimetría

33. Una auditoría anual realizada en una importante empresa de la ciudad de Ambato analiza el número de accidentes reportados por sus trabajadores, los mismos que se muestran en la siguiente tabla:

Accidentes	# trabajadores
0 - 1	30
1 - 2	40
2 - 3	20
3 - 4	5
4 - 5	5
	100

Para analizar el riesgo de accidentes de la empresa, se desea calcular el coeficiente de asimetría de Fisher, e interpretar el sesgo del riesgo.

34. Se realizó una evaluación psicológica al personal de trabajadores de una Universidad de centro del país, se aplicó la evaluación a 50 trabajadores de diferentes áreas, con los siguientes resultados (sobre 100 puntos):

Puntaje	# trabajadores
[40 - 50)	3
[50 - 60)	7
[60 - 70)	20
[70 - 80)	15
[80 - 90]	5
	50

Para determinar el rendimiento de los trabajadores en la evaluación se desea determinar la asimetría y la curtosis. Interpreta los resultados respecto al rendimiento de los trabajadores.

35. En una investigación realizada por los estudiantes de la carrera de software se midió el tiempo de respuesta (en milisegundos) de los servidores de la institución durante el periodo de matrículas, con los siguientes resultados:

Tiempo (ms)	mediciones
[100 - 120)	5
[120 - 140)	10
[140 - 160)	15
[160 - 180)	8
[180 - 200)	2
	40

Se desea analizar el tipo de asimetría y curtosis que presenta la distribución de los tiempos de respuesta de los servidores institucionales.

36. Una pequeña tienda ubicada frente a la Universidad, registra los precios (en dólares) de un producto nuevo en un mes, como se muestra a continuación:

Precio (USD)	# de productos
3.5	2
4.0	4
4.5	6
5.0	8
5.5	5
6.0	3
6.5	2

Calcula el coeficiente de asimetría y curtosis para determinar el comportamiento de los precios.

37. Una docente de estadística de la carrera de Logística registró las notas de 30 de sus estudiantes en el examen de fin de parcial. Las mismas que se ordenaron en la siguiente tabla:

Calificación	# de estudiantes
2	1
3	2
4	4
5	5
6	7
7	6
8	3
9	2

Determina si las calificaciones presentan asimetría significativa y si hay una concentración en torno a la media o en los extremos (curtosis).

38. Debido al temporal que se está viviendo en Ecuador, se registró el número de días con lluvia en 20 ciudades con mayores precipitaciones durante un mes, obteniéndose los siguientes resultados:

Días de lluvia	# ciudades
3	1
6	3
9	4
12	6
15	4
18	1
21	1

Calcula el coeficiente de asimetría y curtosis para identificar si existen más o menos lluvias.

39. Una empresa telefónica para cumplir las metas propuestas realiza evaluaciones frecuentes donde evalúa el tiempo de conexión diaria (en horas) de sus clientes:

Tiempo (h)	# clientes
[1 - 2)	100
[2 - 3)	200
[3 - 4)	150
[4 - 5)	30
[5 - 6]	20

Analiza el grado de asimetría y su relación con el comportamiento de los usuarios.

40. Una entidad bancaria tiene una gran cantidad de usuarios los días lunes, por esta razón se registraron los tiempos de espera (en minutos) para su atención, los mismos se agruparon en la siguiente tabla:

Tiempo (min)	# clientes
[0 - 5)	3
[5 - 10)	8
[10 - 15)	20
[15 - 20)	10
[20 - 25]	4

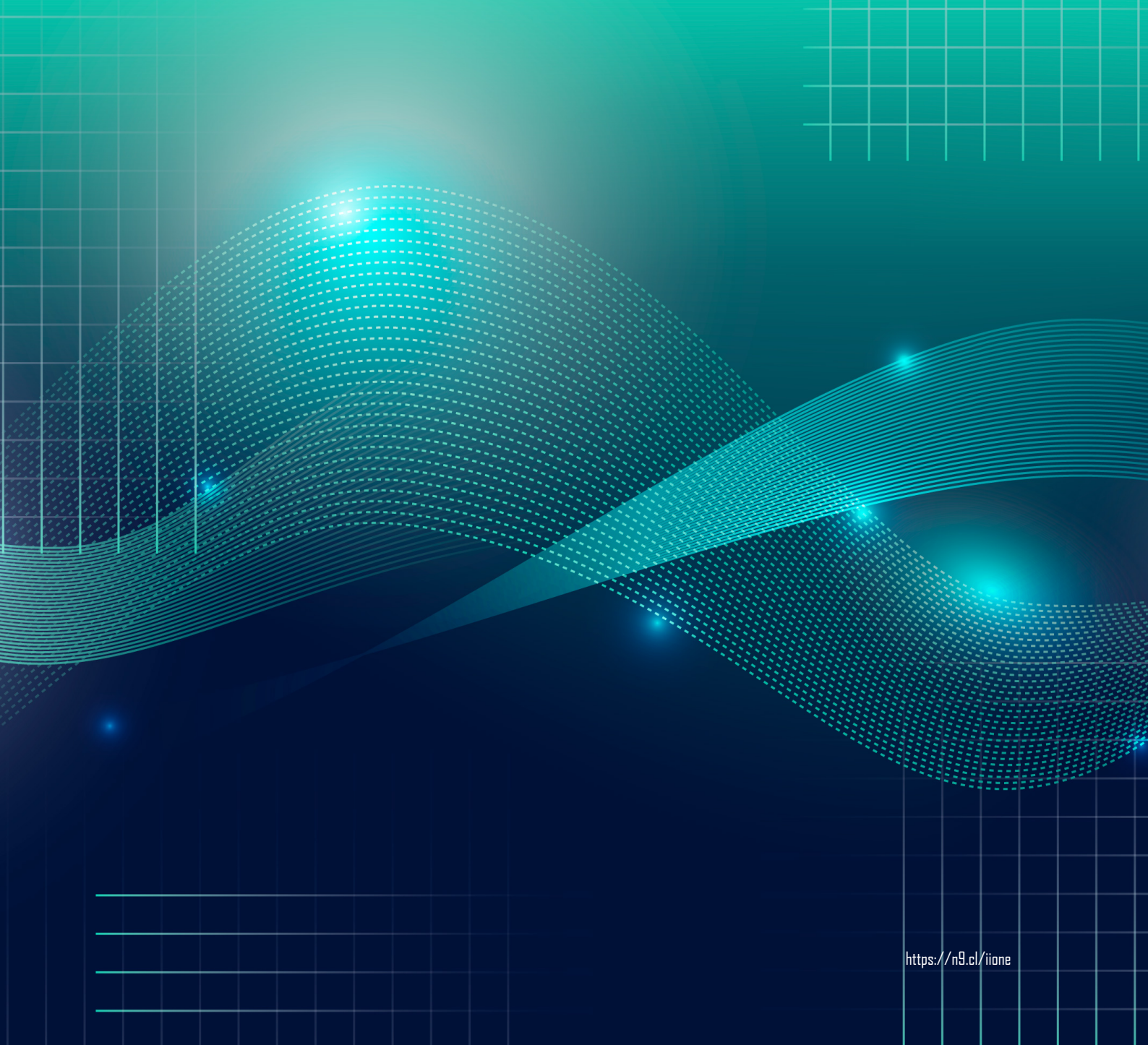
¿Existe evidencia de asimetría positiva o negativa?

Ejercicio propuesto (con Base de Datos)

Utilizando la base de datos student performance que se encuentra en el siguiente enlace:



Describe un análisis estadístico descriptivo de la variable edad (age)



<https://n9.cl/iione>

CAPÍTULO V

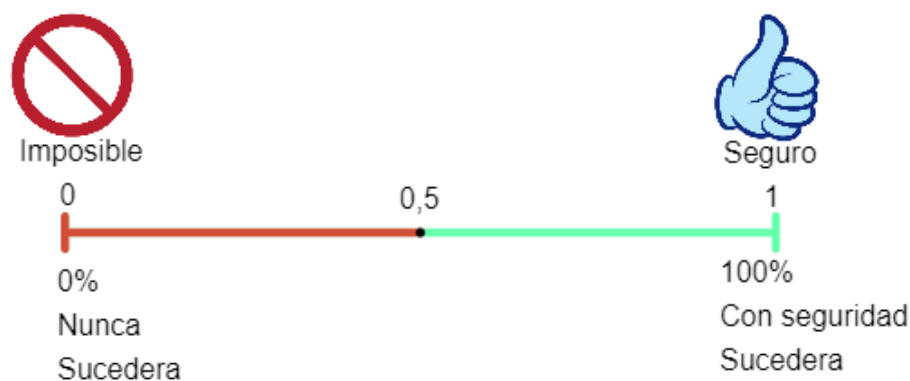
Teoría de Probabilidades

5.1. Definición de probabilidad

La posibilidad de que ocurra un evento producto del azar se lo conoce como probabilidad. En términos matemáticos, se define como los eventos favorables dividido para los eventos posibles en un experimento (Lind et al., 2012).

Figura 38

Rango de probabilidad



5.2. Conceptos básicos

Experimento aleatorio: Un experimento se dice aleatorio cuando al repetirse bajo las mismas condiciones, puede producir diferentes resultados.

Ejemplo: lanzar una moneda al aire

Espacio Muestral (S): Conjunto de todos los resultados posibles en un experimento.

Ejemplo: el espacio muestral es $S=\{\text{cara}, \text{sello}\}$

Evento (E): Un subconjunto del espacio muestral. Es el resultado o conjunto de resultados específicos que nos interesa, Puede estar formado por uno o varios resultados.

Ejemplo: Obtener cara al lanzar una moneda, el evento sería $E=\{\text{cara}\}$

Evento Seguro: Un evento que siempre ocurre, es decir, coincide con todo el espacio muestral. Su probabilidad es $P(S)=1$

Evento Imposible: Un evento que nunca ocurre, como obtener un 7 al lanzar un dado estándar. Su probabilidad es $P(\emptyset)=0$

Eventos mutuamente excluyentes: Cuando dos eventos no pueden ocurrir al mismo tiempo (no tienen resultados comunes) (Lind et al., 2012).

Eventos no mutuamente excluyentes: Dos eventos no son mutuamente excluyentes si pueden ocurrir al mismo tiempo (Lind et al., 2012).

Eventos independientes: Que ocurra un evento no afecta la ocurrencia de otro evento.

Eventos dependientes: Dos eventos son dependientes cuando la ocurrencia de uno afecta la ocurrencia del otro.

Para el cálculo de probabilidades se debe tener en cuenta tres enfoques:

5.3. Enfoques de Probabilidad

5.3.1. Probabilidad clásica

Todos los eventos tienen la misma posibilidad de ocurrir, se calcula:

$$P(E) = \frac{\text{Número de resultados favorables del evento (E)}}{\text{Número total de resultados en el espacio muestral (S)}}$$

Ejemplo

Se lanza una moneda al aire tres veces. ¿Cuál es la probabilidad de que salga dos sellos?

1. Prepare los datos:

- a. Asigne letras a cada variable:

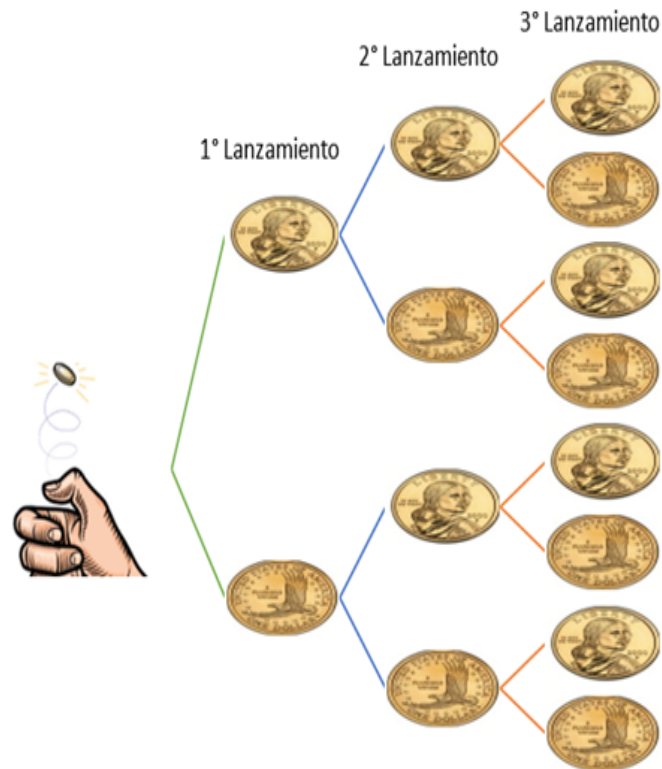
Cara=c

Sello=s

- b. Represente gráficamente el espacio muestral si es posible para entender mejor el ejercicio:

Figura 39

Lanzamiento de una moneda 3 veces

**2. Análisis:**

- a. Sacar el espacio muestral:

$$S=\{(c,c,c),(c,c,s),(c,s,c),(c,s,s),(s,c,c),(s,c,s),(s,s,c),(s,s,s)\}$$

- b. Obtener el evento “salga al menos dos sellos”:

$$E=\{(c,s,s),(s,c,s),(s,s,c),(s,s,s)\}$$

- c. Calcular la probabilidad clásica

$$P(E) = \frac{E}{S} = \frac{4}{8} = \frac{1}{2}$$

3. Interpretación:

La probabilidad de que al lanzar tres monedas al aire salga al menos dos sellos es del 50%.

Ejemplo

Se colocan dentro de una caja 11 bolas rojas, 2 blancas y 7 azules cajas. ¿Cuál es la probabilidad de que, al sacar una bola, esta sea azul?

1. Prepare los datos:

- a. Sacar el espacio muestral:

$$S = \{ \text{11 bolas rojas, 2 blancas, 7 azules} \} = 20$$

- b. Obtener el evento “sacar una bola azul”:

$$E = \{ \text{7 bolas azules} \} = 7$$

2. Análisis:

- a. Calcular la probabilidad

$$P(E) = \frac{E}{S} = \frac{7}{20} = 0,35$$

3. Interpretación:

La probabilidad de que, al sacar una bola de la caja, esta sea azul es del 35%

5.3.2. Probabilidad empírica

Estima la probabilidad de un evento. Se calcula como la relación entre las veces que ocurre un evento y el número total de observaciones, como:

$$P(E) = \frac{\text{Número de veces que ocurre un evento}}{\text{Número total de observaciones}}$$

El enfoque empírico de la probabilidad se basa en la llamada ley de los grandes números, es decir, que al realizar un experimento una gran cantidad de intentos, la probabilidad empírica de este evento se aproximará a su probabilidad real. (Lind et al., 2012).

Ejemplo

Un local de venta de helados desea saber cuál es el sabor de helado más popular entre sus clientes; durante un día, encuestan a 500 personas, obteniendo los siguientes resultados:

- Coco: 120 personas
- Chocolate: 200 personas
- Mora: 100 personas
- Otros sabores: 80 personas

Se desea saber la probabilidad empírica de que un cliente prefiera el sabor de chocolate.

1. Preparación y análisis:

Se aplica la fórmula de la probabilidad empírica con la información del problema.

$$P(E) = \frac{\text{Número de veces que ocurre un evento}}{\text{Número total de observaciones}}$$

$$P(\text{Helado Chocolate}) = \frac{200}{500} = \frac{2}{5} = 0,40$$

2. Interpretación:

Se puede concluir que el 40% de las personas prefieren el helado de chocolate.

5.3.3. Probabilidad subjetiva

Cuando existe poca información sobre un evento en particular el investigador puede evaluar la información disponible y luego calcular o asignar una probabilidad (Lind et al., 2012).

Ejemplos:

1. Un experto en meteorología piensa que hay un 75% de probabilidad de que llueva mañana en la capital de los ecuatorianos.
2. Un hincha del Macará cree que su equipo tiene un 80% de probabilidad de ganar el próximo partido contra Barcelona.

3. Un degustador culinario muy reconocido estima que hay un 83% de probabilidad de que el restaurante “Las Delicias” cree un nuevo menú y sea bueno.
4. Un estudiante de la carrera de Ingeniería de software cree que tiene un 55% de probabilidad de aprobar un examen final de Estadística I, el cual es muy difícil.
5. Diego está jugando a lanzar la moneda al aire, él cree que hay un 80% de probabilidad de que salga cara en el próximo lanzamiento.

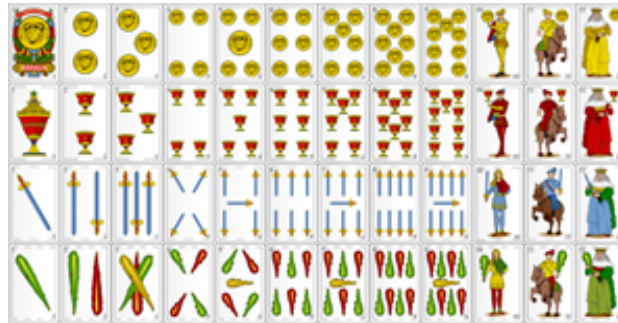
Autoevaluación 5.1.

1. Define los siguientes conceptos y da un ejemplo de cada uno:
 - a. Espacio Muestral
 - b. Evento o Suceso
2. Si la probabilidad de que llueva hoy es del 30% ($P(A) = 0.3$), ¿cuál es la probabilidad de que NO llueva hoy? Nombra el concepto clave que utilizaste para resolverlo.
3. En una caja hay 5 bolas rojas, 3 bolas verdes y 2 bolas azules. Todas son del mismo tamaño y peso. Si se saca una bola al azar, calcula:
 - a) La probabilidad de sacar una bola roja.
 - b) La probabilidad de sacar una bola que NO sea azul.

5.4. Ejercicios propuestos

1. En una caja de 32 manzanas hay 24 maduras, 2 verdes y 6 dañadas ¿Cuál es la probabilidad que, al sacar una manzana, esta no esté dañada?
2. En el juego de Monopolio, se lanzan dos dados de seis caras. ¿Cuál es la probabilidad de que al sumar los puntos de la cara superior de los dados se obtenga un número mayor que 5?
3. En la figura siguiente se muestra una baraja española de 48 cartas, ¿cuál es la probabilidad de sacar una Sota (baraja con el numero 10)?

Figura 40

Baraja española

4. María está jugando Parchis con sus amigos, se lanzan dos dados al aire, para ganar María necesita un 8, pero los números en las caras superiores no deben ser iguales. ¿Cuál es la probabilidad que María gane el juego de Parchis?
5. En la figura siguiente se muestra los naipes de una baraja alemana, que consta de 36 naipes divididos en cuatro palos campanas, hojas, corazones y bellotas. ¿Cuál es la probabilidad de sacar un campesino de bellota o un rey, pero no de corazones?

Figura 41

Baraja alemana

6. En una empresa de fabricación de zapatos se tiene un total de 350 trabajadores, de los cuales 23 han sufrido un accidente laboral. ¿Cuál es la probabilidad de no sufrir un accidente laboral en la empresa?
7. En una canasta hay 10 duraznos, 15 mandarinas y 8 claudias. Se selecciona al azar una fruta de la canasta, ¿cuál es la probabilidad de elegir una mandarina?

8. En el área de Neonatología de un Hospital, al mes ingresaron 15 recién nacidos hombres y 25 recién nacidos mujeres. Si se selecciona un recién nacido al azar, ¿cuál es la probabilidad de que sea hombre?
9. Si se lanzan dos dados al aire, ¿Cuál es la probabilidad de obtener como máximo 7 puntos si en el primer dado sale un 3?
10. Vamos a considerar el siguiente juego de cartas de dominó chino. Que consta de dos palos, en la parte superior el civil y en la parte inferior el militar, de acuerdo a la figura siguiente.

Figura 42

Juego de dominó chino



¿Cuál es la probabilidad de sacar una carta de dominó chino civil mixta (puntos negros y rojos)?

5.5. Reglas de probabilidad

Son principios fundamentales que nos permiten calcular y entender la probabilidad de eventos en diferentes situaciones. A continuación, se detallan las reglas más importantes:

5.5.1. Regla del complemento

La probabilidad de que ocurra el complemento de un evento A, es decir, de que A no ocurra, se expresa como:

$$P(A^c) = 1 - P(A)$$

Esta regla es útil cuando es más fácil calcular la probabilidad de que algo no ocurra.

Ejemplo

Un estudio del Instituto Geográfico Militar indica que la probabilidad que el día de hoy suceda un temblor es 0,015, ¿cuál es la probabilidad que no haya temblores el día de hoy?



1. Análisis de los datos:

Aplicar la fórmula de la regla del complemento:

$$P(\text{no haya temblor}) = 1 - P(\text{haya temblor}) = 1 - 0,015 = 0,985$$

2. Interpretación:

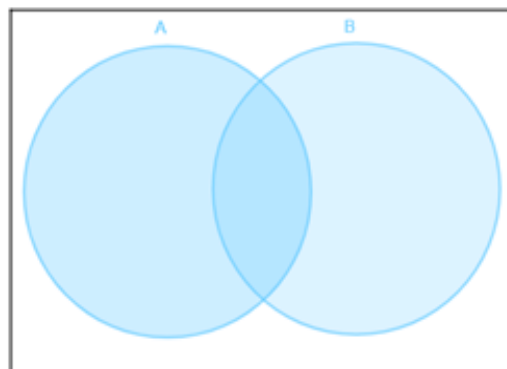
La probabilidad que no haya temblores el día de hoy es del 98,5%

5.5.2. Regla de la adición o suma

Se puede aplicar la regla de la adición cuando se tiene eventos mutuamente excluyentes o, eventos mutuamente no excluyentes. (Mood et al., 1973).

Figura 43

Regla de la Adición con diagrama de Venn



Eventos mutuamente excluyentes

La probabilidad de que ocurra un evento A u otro evento B es igual a:

$$P(A \cup B) = P(A) + P(B)$$

Donde:

- $P(A)$: Probabilidad de que ocurra el evento A
- $P(B)$: Probabilidad de que ocurra el evento B
- $P(A \cup B)$: Probabilidad que ocurra el evento A o el evento B

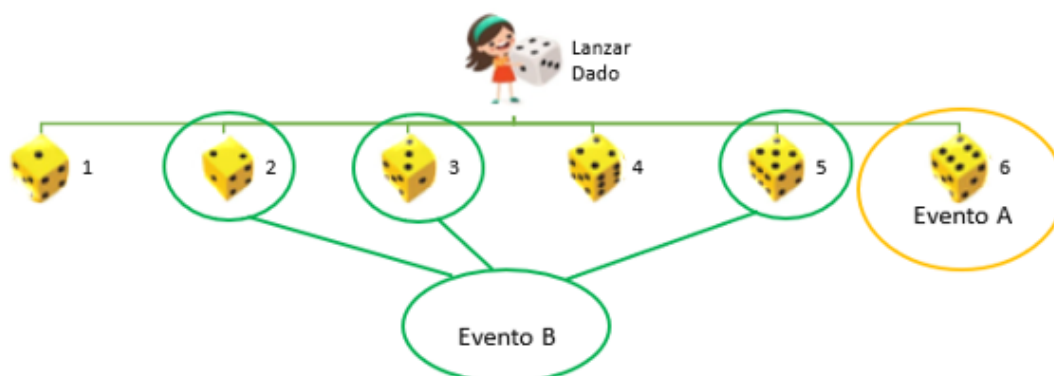
Ejemplo:

Si se lanza un dado al aire, ¿cuál es la probabilidad de obtener un seis o un número primo?

1. Prepare los datos:

Evento A: "Obtener un 6"

Evento B: "Obtener un número primo"



Estos dos eventos como se observa en la figura son mutuamente excluyentes, ya que el 6 no es número primo.

2. Análisis:

- a. Se calculan las probabilidades de que ocurra cada evento por separado y se suman los resultados.

$$P(A) = \frac{1}{6}$$

$$P(B) = \frac{3}{6}$$

$$P(A \cup B) = \frac{1}{6} + \frac{3}{6} = \frac{4}{6} = \frac{2}{3} = 0,6667$$

3. Interpretación:

Se puede concluir entonces que la probabilidad de obtener un 6 o un número primo al lanzar un dado es del 66,67%.

Eventos no mutuamente excluyentes

Es decir, dos eventos pueden ocurrir al mismo tiempo, para que esto suceda se debe restar la probabilidad de que ocurran ambos eventos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Donde:

- $P(A)$: Probabilidad de que ocurra el evento A
- $P(B)$: Probabilidad de que ocurra el evento B
- $P(A \cap B)$: Probabilidad que ocurra el evento A y B al mismo tiempo
- $P(A \cup B)$: Probabilidad que ocurra el evento A o el evento B

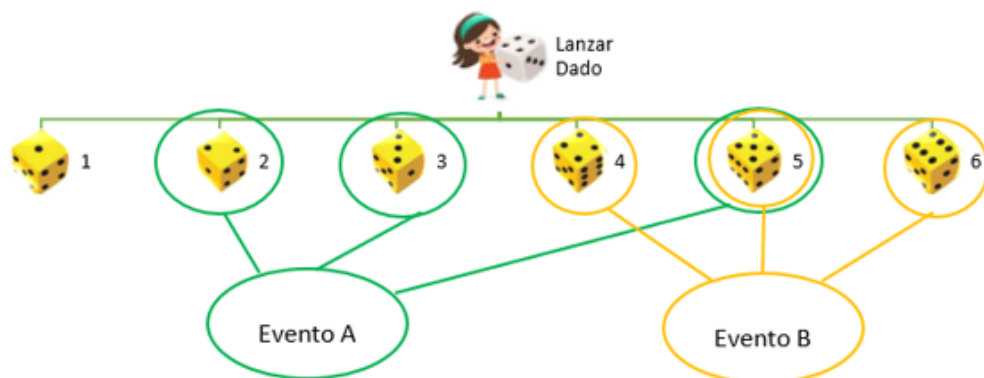
Ejemplo

Si se lanza un dado al aire, ¿cuál es la probabilidad de obtener un número primo o un número mayor que tres?

1. Prepare los datos:

Evento A: "Obtener un número primo"

Evento B: “Obtener un número mayor que tres”



Estos dos eventos como se observa en la imagen son no mutuamente excluyentes, ya que el 5 es un número primo, pero también es mayor que tres.

2. Análisis:

Se calcula la probabilidad de que ocurra cada evento por separado y también que ocurran al mismo tiempo, para finalmente aplicar la fórmula de la probabilidad para eventos no mutuamente excluyentes.

$$P(A) = \frac{3}{6}$$

$$P(B) = \frac{3}{6}$$

$$P(A \cap B) = \frac{1}{6}$$

$$P(A \cup B) = \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6} = 0,8333$$

3. Interpretación:

Se puede concluir entonces que la probabilidad de obtener un número primo o mayor que 3 es 83,33%.

5.5.3. Regla del producto

Nos permite calcular la probabilidad de que dos eventos ocurran de forma simultánea; se puede aplicar esta regla, cuando se tienen eventos dependientes o independientes (Mood et al., 1973).

Eventos independientes

Si se tiene dos eventos A y B, la ocurrencia del evento A no afecta la ocurrencia del evento B, la fórmula es:

$$P(A \cap B) = P(A) \cdot P(B)$$

Ejemplo:

Si se lanza un dado y una moneda al mismo tiempo, ¿cuál es la probabilidad de obtener un número impar en el dado y una cara en la moneda?

1. Prepare los datos:

Representar el espacio muestral:

En la figura se observa el número de caras que tiene una moneda y un dado; y las combinaciones que se pueden tener al lanzar la moneda y el dado juntos.

Evento A: "Sacar un número impar en el dado"

Evento B: "Sacar cara al lanzar una moneda"

2. Análisis:

Calcular la probabilidad que suceda cada evento.

$$P(A) = \frac{6}{12} = \frac{1}{2}$$

$$P(B) = \frac{6}{12} = \frac{1}{2}$$

$$P(A \cap B) = P(A).P(B)$$

$$P(A \cap B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

3. Interpretación:

En este ejemplo, los eventos lanzar un dado y una moneda son totalmente independientes, por lo tanto, existe una probabilidad de 25% de obtener un número impar al lanzar un dado y una cara al lanzar una moneda al mismo.

Eventos dependientes

Si se tiene dos eventos A y B y no son independientes, la ocurrencia del evento A afecta la ocurrencia del evento B; entonces debemos usar la probabilidad condicional, para calcular la intersección de ambos eventos:

$$P(A \cap B) = P(A).P(B|A)$$

Donde $P(B | A)$ es la probabilidad que ocurra el evento B dado que ya ocurrió el evento A.

Ejemplo

En una urna hay 8 bolas rojas y 5 verdes. Si sacamos dos bolas consecutivas sin devolución, ¿cuál es la probabilidad de sacar dos bolas verdes consecutivas?



1. Prepare los datos:

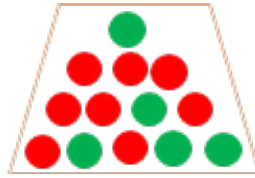
Evento A: "Sacar la primera bola verde"

Evento B: "Sacar la segunda bola verde"

2. Análisis:

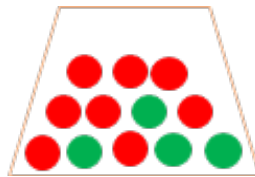
a. Calcular la probabilidad de que la primera bola sea verde

$$P(A) = \frac{5}{13}$$



- b. Calcular la probabilidad condicional, es decir que la segunda sea verde si ya la primera también lo fue.

$$P(B|A) = \frac{4}{12} = \frac{1}{3}$$



- c. Cálculo de la probabilidad de sacar dos bolas verdes

Para el evento A, se tienen 5 bolas verdes del total de 13 bolas, y se saca de la urna, esto afecta para encontrar el segundo evento, ya que la segunda bola también es verde, ahora ya no tenemos 5 bolas verdes sino 4, y en total ya no son 13 bolas sino 12 bolas, por lo tanto, esto afecta el cálculo de la probabilidad del segundo evento.

$$P(A \cap B) = P(A) \cdot P(B|A)$$

$$P(A \cap B) = \frac{5}{13} \times \frac{1}{3} = \frac{5}{39} = 0,1282$$

3. Interpretación:

La probabilidad de sacar dos bolas verdes consecutivas de una urna sin devolución es 12,82%.

5.5.4. Regla de probabilidad condicional

La probabilidad condicional $P(A | B)$ mide la probabilidad de que ocurra un evento A, sabiendo que ya ocurrió un evento B (Mood et al., 1973), se define como:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Ejemplo:

Si sabemos que, al lanzar dos dados, la suma de las caras superiores es mayor que 7, ¿cuál es la probabilidad de que por lo menos en uno de los dados la cara superior sea un 5?

1. Prepare los datos:

Encuentre el espacio Muestral de lanzar dos dados:

En la figura se representa el espacio muestral de lanzar dos dados, y todas las combinaciones que se pueden tener.

Evento A: "Suma sea mayor a siete"

Evento B: "Cara superior de un dado igual cinco"

2. Análisis:

a. Calcule las probabilidades solicitadas:

$$P(A) = \frac{15}{36} = \frac{5}{12}$$

$$P(A \cap B) = \frac{7}{36}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

$$P(A|B) = \frac{\frac{7}{36}}{\frac{5}{12}} = \frac{7}{15} = 0,4667$$

3. Interpretación:

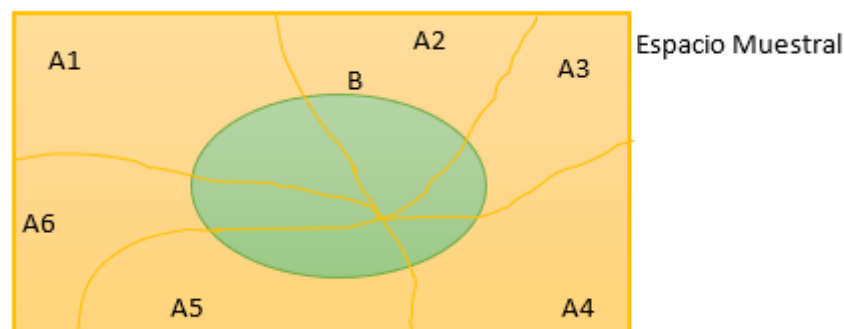
La probabilidad de que en uno de los dados la cara superior sea 5, sabiendo que la suma de las caras de los dados es mayor que siete es 46,67%.

5.5.5. Regla de la probabilidad Total

Un evento puede descomponerse en varios sub eventos ($A_1, A_2, A_3, A_4, \dots, A_n$) mutuamente excluyentes y se tiene el evento B que es un suceso cualquiera del espacio muestral como se observa en la figura siguiente; la probabilidad total del evento P(B), puede calcularse sumando las probabilidades de los sub eventos (Lind et al., 2012).

Figura 44

Representación de la probabilidad total entre eventos y sub eventos



$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) + \dots + P(A_n \cap B)$$

La fórmula de la probabilidad total también puede escribirse como:

$$P(B) = P(A_1) \times P(B|A_1) + P(A_2) \times P(B|A_2) + P(A_3) \times P(B|A_3) + \dots + P(A_n) \times P(B|A_n)$$

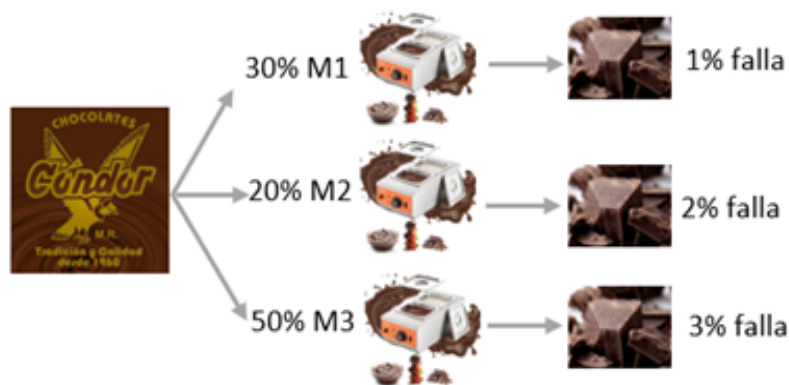
La suma de las probabilidades de todos los sub eventos es igual a 1.

Ejemplo:

La fábrica de caramelos “El Cóndor” tiene tres máquinas que elaboran chocolates, y producen el 30%, 20%, y 50% del total de chocolates. Además, se conoce que las maquinas producen el 1%, 2% y 3% de chocolates con fallas que no se pueden vender al público. Si se selecciona un chocolate al azar, ¿cuál es la probabilidad de que tenga falla?

1. Prepare los datos:

Represente los datos gráficamente:



Asignar los nombres a cada uno de los eventos de la fabricación de chocolates.

- Evento A: Producto fabricado en la máquina 1
- Evento B: Producto fabricado en la máquina 2
- Evento C: Producto fabricado en la máquina 3
- Evento D: Producto defectuoso

2. Análisis:

a. Calcular las probabilidades por separado de que ocurra cada evento.

$$P(A) = \frac{30}{100} = 0,30 \quad P(B) = \frac{20}{100} = 0,20 \quad P(C) = \frac{50}{100} = 0,50$$

- b. Calcular las probabilidades condicionales necesarias para el cálculo de la probabilidad total.

$$P(D|A) = \frac{1}{100} = 0,01 \quad P(D|B) = \frac{2}{100} = 0,02 \quad P(D|C) = \frac{3}{100} = 0,03$$

- c. Calcular la probabilidad total de los chokolatines defectuosos

$$P(D) = P(A) \times P(D|A) + P(B) \times P(D|B) + P(C) \times P(D|C)$$

$$P(D) = 0,30 * 0,01 + 0,20 * 0,02 + 0,50 * 0,03 = 0,022$$

3. Interpretación:

La probabilidad que al elegir un chokolatín este tenga falla es 2,2%

Ejercicio práctico (con Base de Datos)

Utilizando la base de datos Wine Quality Red que se encuentra en el siguiente enlace:

Accede aquí:



Calcula las siguientes probabilidades:

- ¿Cuál es la probabilidad marginal de que un vino sea de alta calidad?
- ¿Cuál es la probabilidad conjunta de que un vino tenga bajo alcohol y baja calidad?
- Dado que un vino tiene alto alcohol, ¿cuál es la probabilidad de que sea de alta calidad?

Solución:

1. *Descargar la base de datos:*

Dirigirse al enlace propuesto y en la parte superior derecha dar clic den descargar.



Wine Quality

Donated on 10/6/2009

Two datasets are included, related to red and white vinho verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests (see [Cortez et al., 2009], <http://www3.dsi.uminho.pt/pcortez/wine/>).

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Business	Classification, Regression

Feature Type	# Instances	# Features
Real	4898	11

DOWNLOAD (89.2 KB)

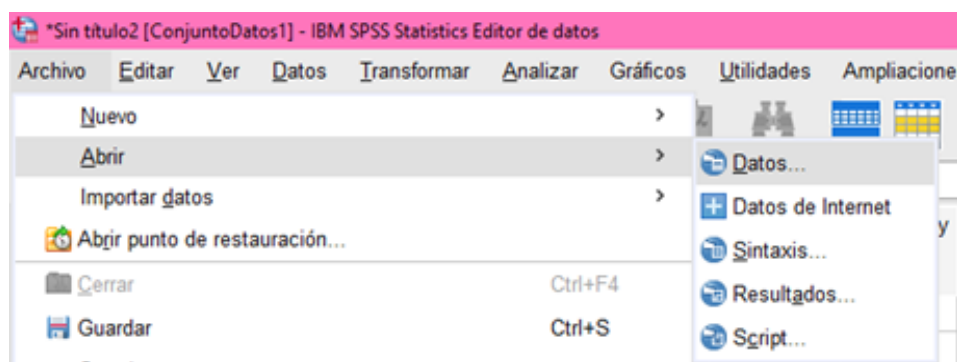
IMPORT IN PYTHON

CITE

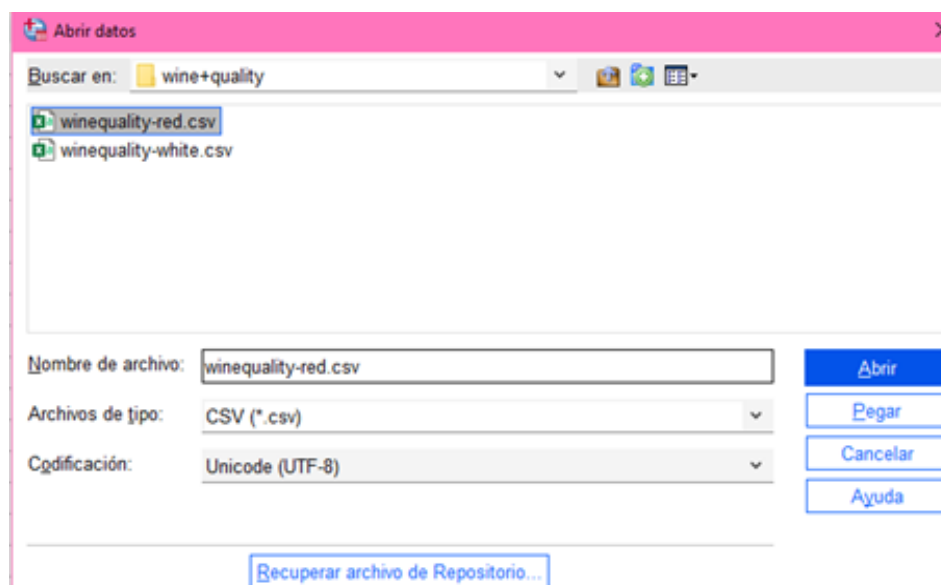
1 citations
606348 views

Keywords
Chemistry

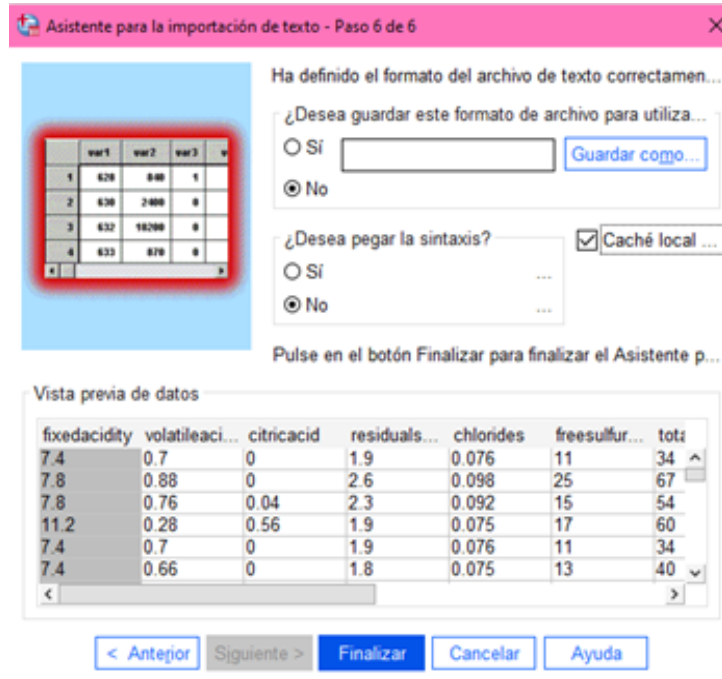
2. Abrir el SPSS y dar clic en abrir, datos:



3. En esta ventana cambiamos el tipo de archivo a *.csv y buscamos la ubicación de la base de datos descargada y clic en abrir:



4. Una vez aquí configuramos ciertos parámetros de visualización de la base de datos y se finaliza la carga de los datos:

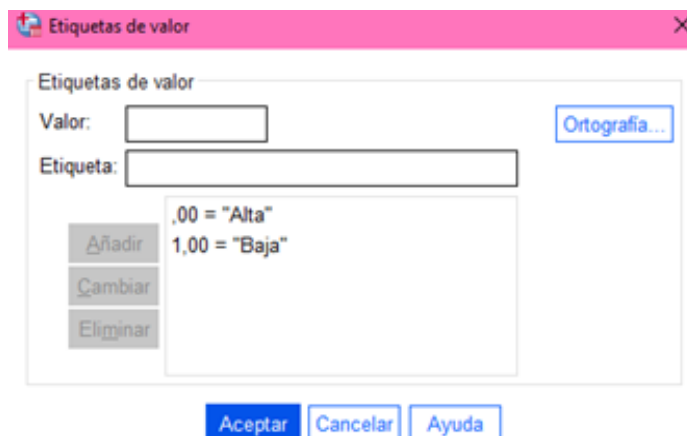


5. Ahora se puede visualizar la información de todos los datos que contiene la base de datos calidad de vino rojo.

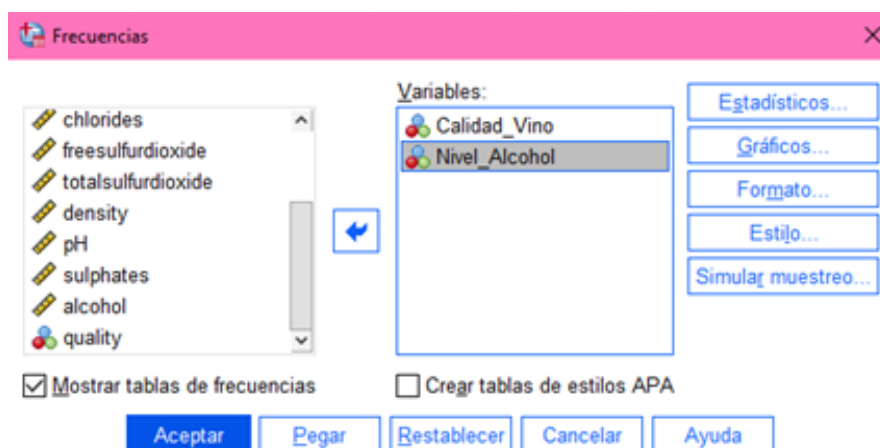
	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulfates	alcohol	quality
1	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
2	7.8	0.88	0	2.6	0.098	25	67	0.9968	3.02	0.68	9.8	5
3	7.8	0.76	0.04	2.3	0.092	15	54	0.997	3.26	0.65	9.8	5
4	11.2	0.28	0.56	1.9	0.075	17	60	0.998	3.16	0.58	9.8	6
5	7.4	0.7	0	1.9	0.076	11	34	0.9978	3.51	0.56	9.4	5
6	7.4	0.66	0	1.8	0.075	13	40	0.9978	3.51	0.56	9.4	5
7	7.9	0.6	0.06	1.6	0.069	15	59	0.9964	3.03	0.46	9.4	5
8	7.3	0.65	0	1.2	0.065	15	21	0.9946	3.39	0.47	10	7
9	7.8	0.58	0.02	2	0.073	9	18	0.9968	3.36	0.57	9.5	7
10	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5
11	6.7	0.58	0.08	1.8	0.097	15	65	0.9959	3.28	0.54	9.2	5
12	7.5	0.5	0.36	6.1	0.071	17	102	0.9978	3.35	0.8	10.5	5

a. ¿Cuál es la probabilidad de que un vino sea de alta calidad?

Para encontrar esta probabilidad lo primero que se debe hacer es crear una variable categórica para saber si la calidad del vino es alta o baja, por lo que se le asigna 1 a baja y 0 a alta, en la pestaña vista de variables, clic en la columna valores de calidad vino:



Para calcular la probabilidad de que el vino sea de alta calidad, dar clic en Analizar > Estadísticos Descriptivos > Frecuencias. Se selecciona Calidad_Vino y clic en Aceptar.



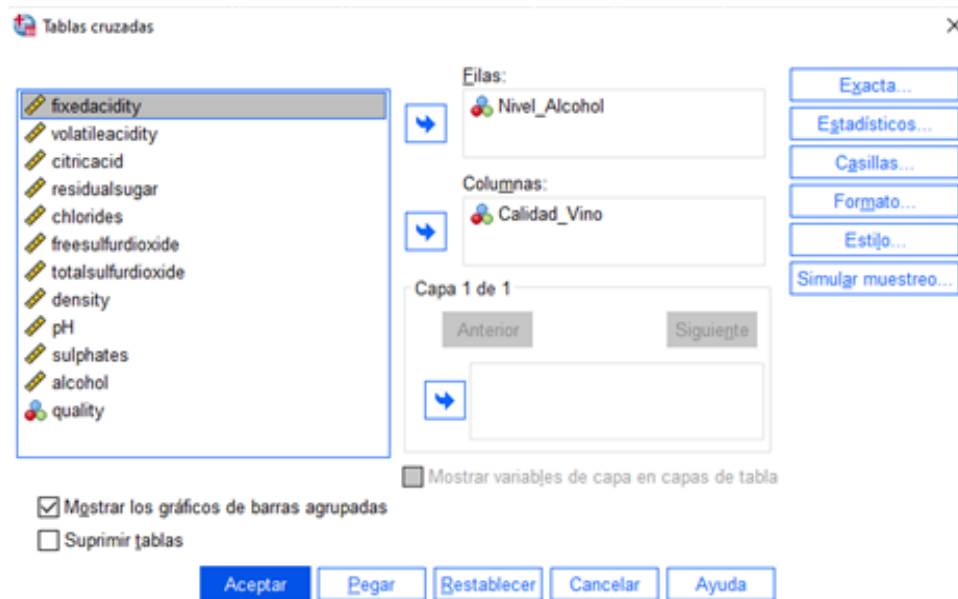
Finalmente se visualiza el porcentaje de la calidad del vino.

Calidad_Vino					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	Alta	855	53,5	53,5	53,5
	Baja	744	46,5	46,5	100,0
	Total	1599	100,0	100,0	

Por lo tanto, la probabilidad de que el vino rojo sea de alta calidad es de 53.5%.

b. ¿Cuál es la probabilidad conjunta de que un vino tenga bajo alcohol y baja calidad?

Para encontrar la probabilidad conjunta se utilizará una tabla cruzada, por lo que se da clic en Analizar > Estadísticos Descriptivos > Tablas cruzadas se selecciona como filas Nivel_Alcohol y como columnas Calidad_Vino, en casillas, se activa el porcentaje para filas y columnas y clic en aceptar.



Se obtiene la siguiente información:

Tabla cruzada Nivel_Alcohol*Calidad_Vino

			Calidad_Vino		Total
			Alta	Baja	
Nivel_Alcohol	Bajo	Recuento	238	509	747
		% dentro de Nivel_Alcohol	31,9%	68,1%	100,0%
		% dentro de Calidad_Vino	27,8%	68,4%	46,7%
	Alto	Recuento	617	235	852
		% dentro de Nivel_Alcohol	72,4%	27,6%	100,0%
		% dentro de Calidad_Vino	72,2%	31,6%	53,3%
Total	Recuento		855	744	1599
	% dentro de Nivel_Alcohol		53,5%	46,5%	100,0%
	% dentro de Calidad_Vino		100,0%	100,0%	100,0%

Por lo tanto, la probabilidad de que el vino sea de baja calidad y de un nivel de alcohol bajo es de 68.1%.

c. Dado que un vino tiene alto alcohol, ¿cuál es la probabilidad de que sea de alta calidad?

Para encontrar esta probabilidad se utilizará la misma tabla cruzada del literal anterior:

Tabla cruzada Nivel_Alcohol*Calidad_Vino

			Calidad_Vino		Total
			Alta	Baja	
Nivel_Alcohol	Bajo	Recuento	238	509	747
		% dentro de Nivel_Alcohol	31,9%	68,1%	100,0%
		% dentro de Calidad_Vino	27,8%	68,4%	46,7%
	Alto	Recuento	617	235	852
		% dentro de Nivel_Alcohol	72,4%	27,6%	100,0%
		% dentro de Calidad_Vino	72,2%	31,6%	53,3%
Total	Recuento		855	744	1599
	% dentro de Nivel_Alcohol		53,5%	46,5%	100,0%
	% dentro de Calidad_Vino		100,0%	100,0%	100,0%

Por lo tanto, la probabilidad de que el vino sea de alta calidad dado que tiene un nivel alto de alcohol es de 72.4%.

Autoevaluación 5.2.

1. Explica con tus propias palabras la Regla de la Adición para dos eventos; y escribe la diferencia entre eventos mutuamente excluyentes y no mutuamente excluyentes.
2. En un grupo de 1000 estudiantes universitarios, 450 están tomando un curso de Estadística, 600 están tomando un curso de Programación y 200 están tomando ambos cursos. Si se selecciona un estudiante al azar, ¿cuál es la probabilidad de que esté tomando solo el curso de programación.

5.5.6. Ejercicios propuestos

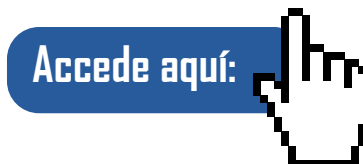
11. En la feria del pueblo hay un concurso donde se dará un premio si dos personas logran decir al mismo tiempo, el mismo número dentro del 1 al 20. Si Yuly está pensando en: 1, 8, 3, 4, 7, 6 y Julia en: 8, 5, 6, 14, 15, 1. ¿Cuál es la probabilidad que lo logren?
12. Si en un circuito en parejas hay 6 huecos donde cada persona deberá lanzar canicas hacia un hueco, cada persona por sorteo recibe diferente cantidad de canicas. Si Camilo recibió 3 canicas y María 4 canicas. ¿Cuál es la probabilidad de que Camilo y María metan todas las canicas en los huecos, sabiendo que existe un 50% de acertar?
13. En un juego de mesa llamado “Parchis” necesito que me salga un 4 en un dado y un 4 o 6 en el otro para Ganar el juego. ¿Cuál es la probabilidad de lograrlo?
14. En un juego por computadora, si necesito vencer al menos 4 de 8 enemigos en un desafío y a su vez lograr ver 3 de 9 códigos para conseguir pasar el nivel. ¿Cuál es la probabilidad de que lo logre?
15. En el aula de clases del bloque A de la Universidad hay 32 estudiantes, de los cuales 23 estudian solo matemática básica y 9 estudian solo logística; ¿cuál es la probabilidad de que, al preguntar a un estudiante, este estudie las 2 materias?
16. En una feria de universidades se les pregunta a 100 jóvenes que carrera les gustaría estudiar, si una carrera técnica o una carrera administrativa, todos los jóvenes desean estudiar cualquiera de las dos carreras. Pero, 82 responden que una carrera técnica y 44 responde que una carrera administrativa. ¿Cuál es la probabilidad de que al preguntarle a un estudiante les gusten los 2 tipos de carreras?
17. A un grupo de 300 personas se les pregunta qué genero de música prefieren; 170 contestan Reguetón y 190 Baladas. ¿Cuál es la probabilidad de que al preguntarle a una persona le gusten los dos géneros?
18. Al encuestar a 100 personas sobre sus deportes favoritos a 48 le gusta solo básquet, a 30 solo fútbol ¿Cuál es la probabilidad de que al preguntarle a una persona le guste los 2 deportes, se sabe que a todas las personas encuestadas les gusta alguno de los dos deportes?

19. Al consultar a 50 personas sobre su preferencia de películas a 30 personas les gusta las de acción, a 28 les gusta películas de terror y a 8 les gustan las de acción y terror. Si se le pregunta a una persona al azar, ¿cuál es la probabilidad de que les guste solo las de acción o solo las de terror?
20. Si se lanza dos dados al aire, ¿cuál es la probabilidad que el resultado de la suma de sus caras superiores sea 6 o un número primo?
21. Al 35% de los estudiantes de un salón de clases les gusta la torta de fresa y chocolate, mientras que al 80% le gusta la torta de chocolate. ¿Cuál es la probabilidad de que a un estudiante le guste solo la torta de fresa?
22. En una ciudad el 60% de la población tiene cabello castaño, el 35% tiene ojos castaños y el 25% tiene cabello y ojos castaños. Si se selecciona una persona al azar y esta tiene ojos castaños ¿Cuál es la probabilidad de que tenga cabello castaño?
23. El 40% de las estudiantes de un colegio practican fútbol, el 50% practican baloncesto y el 15% practica ambos deportes, si se selecciona un estudiante al azar y este practica baloncesto. ¿Cuál es la probabilidad que practique también fútbol?
24. En la Universidad de las Fuerzas Armadas el 80% de los estudiantes aprueba Matemática, el 75% aprueba Estadística, de los estudiantes que aprueban matemáticas el 90% aprueban Estadística. Si se selecciona un estudiante al azar y este aprobó estadística. ¿Cuál es la probabilidad de que haya aprobado matemáticas?
25. En una clínica veterinaria el 60% de los animales que atienden son perros, el 30% son de color blanco y de los perros que atienden el 25% son blancos; cada mes se selecciona un animalito para darlo en adopción Si se sabe que el animal es de color blanco, ¿Cuál es la probabilidad de que sea un perro?
26. Dos niños juegan a lanzar una moneda; si sale cara los niños van a la caja azul que tiene 4 bolas rojas y 5 bolas verdes. Si sale cruz, los niños van a la caja amarilla que tiene 6 bolas rojas y 3 bolas verdes. ¿Cuál es la probabilidad de sacar una bola roja?
27. En el aula de clases, hay 15 hombres y 8 mujeres. Además, se sabe que 6 hombres y 5 mujeres estudian física. Se selecciona un estudiante al azar, ¿cuál es la probabilidad de que sea mujer o estudie física?

28. En el juego de black Jack, se extraen dos cartas sin reemplazo. ¿Cuál es la probabilidad de sacar un As y un Rey, por lo tanto, ganar el juego?
29. En una caja de bingo hay 10 bolas rosadas y 12 bolas verdes. Si se extraen dos bolas sin reemplazo, ¿Qué probabilidad existe que la segunda bola sea verde, dado que la primera fue rosada?
30. En época de invierno en la ciudad de Guayaquil una enfermedad muy común es el dengue. Para realizar una investigación al respecto, se clasificó a la población de pacientes en tres grupos según su riesgo de contraer la enfermedad y su respectivo porcentaje: el 28% pertenece al grupo de alto riesgo, el 47% pertenece al grupo de riesgo medio, el 25% pertenece al grupo de bajo riesgo. A todos los individuos se les aplicó Test, obteniéndose los siguientes porcentajes de resultados positivos en cada grupo: Un 80% de positivos en el grupo de alto riesgo, un 65% de positivos en el grupo de riesgo medio y un 22% de positivos en el grupo de bajo riesgo. Con base en esta información, ¿cuál es la probabilidad total de que un paciente, seleccionado al azar, obtenga un resultado positivo en la prueba?

Ejercicio propuesto (con base de datos)

Utilizando la base de datos Default of Credit Card Clients que se encuentra en el siguiente enlace:



Información adicional:

Definición de las variables utilizadas en la base de datos y una breve descripción de lo que representan:

X1: Importe del crédito otorgado (dólares de Taiwán): incluye tanto el crédito al consumo individual como el crédito familiar (complementario).

X2: Género (1 = masculino; 2 = femenino).

X3: Educación (1 = posgrado; 2 = universitario; 3 = bachillerato; 4 = otros).

X4: Estado civil (1 = casado; 2 = soltero; 3 = otros).

X5: Edad (años).

X6-X11: Historial de pagos anteriores. Se registraron los pagos mensuales anteriores (de abril a septiembre de 2005) de la siguiente manera: X6 = estado de pago en septiembre de 2005; X7 = estado de pago en agosto de 2005; X11 = estado de pago en abril de 2005. La escala de medición para el estado de pago es: -1 = pago al día; 1 = retraso de un mes; 2 = retraso de dos meses; 8 = retraso de ocho meses; 9 = retraso de nueve meses o más.

X12-X17: Importe del extracto de factura (dólares taiwaneses). X12 = importe del extracto de factura en septiembre de 2005; X13 = importe del extracto de factura en agosto de 2005; X17 = importe del extracto de factura en abril de 2005.

X18-X23: Importe del pago anterior (dólares taiwaneses). X18 = importe pagado en septiembre de 2005; X19 = importe pagado en agosto de 2005; X23 = importe pagado en abril de 2005.

Utilizando SPSS, calcula las siguientes probabilidades:

- ¿Cuál es la probabilidad marginal de que en el próximo mes haya pagos incumplidos?
- ¿Cuál es la probabilidad que los pagos en el último mes estén al día?
- ¿Cuál es la probabilidad que un cliente incumpla el pago en el próximo mes y sea de sexo femenino?
- ¿Cuál es la probabilidad de que un cliente incumpla el pago en el próximo mes, dado que es de sexo femenino y pertenece al grupo etario adulto (30 a 49 años)?

5.6. Teorema de Bayes, Diagrama de Árbol y Tablas de Contingencia

5.6.1. Teorema de Bayes

El Teorema de Bayes describe cómo actualizar las probabilidades de una hipótesis, dado que se tiene nueva evidencia. Este teorema se utiliza para encontrar la probabilidad de un evento basado en conocimientos previos de condiciones relacionadas. (Lind et al., 2012), es muy útil en situaciones donde queremos tomar decisiones o inferir algo a partir de datos inciertos o incompletos.

La fórmula del teorema de Bayes es la siguiente:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Donde:

- $P(A | B)$ probabilidad condicional, probabilidad de que ocurra el evento A, dado que ha ocurrido el evento B.
- $P(B | A)$ probabilidad condicional de B dado A, probabilidad de que ocurra el evento B, dado que ha ocurrido el evento A.
- $P(A)$ probabilidad de que ocurra el evento A
- $P(B)$ probabilidad de que ocurra el evento B.

Ejemplo

Para detectar una enfermedad rara que el 2% de la población la padece, existe una prueba que la detecta con una precisión del 95% para personas que la padecen y el 90% para personas que no la padecen. Si una persona da positivo en la prueba, ¿cuál es la probabilidad de que realmente tenga la enfermedad?

1. Prepare los datos:

Asignar el nombre a cada uno de los eventos que suceden en el experimento

Evento A: “Padecer la enfermedad”

Evento B: “Resultado de la prueba sea positivo”

2. Análisis:

- a. Se identifica cada una de las probabilidades de los diferentes eventos.

$$P(A) = \frac{2}{100} = 0,02 \text{ (personas que padecen la enfermedad)}$$

$$P(B|A) = \frac{95}{100} = 0,95 \text{ (positiva la prueba sabiendo que padecen la enfermedad)}$$

$$P(B|A^c) = \frac{90}{100} = 0,90 \text{ (positiva la prueba sabiendo que no padecen la enfermedad)}$$

$$P(A^c) = 1 - P(A) = 1 - 0,02 = 0,98 \text{ (personas que no padecen la enfermedad)}$$

$$P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c) = 0,95 * 0,02 + 0,10 * 0,98 = 0,117$$

- b. Se plantea la formula del Teorema de Bayes; lo que se desea saber es la probabilidad de que una persona tenga la enfermedad dado que ha obtenido un resultado positivo en la prueba $P(A | B)$, por lo tanto:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{0.95 * 0.02}{0.117} = 0,162$$

3. Interpretación:

La probabilidad de que una persona tenga la enfermedad, dado que ha dado positivo en la prueba, es aproximadamente 16,2%.

5.6.2. Diagramas de árbol

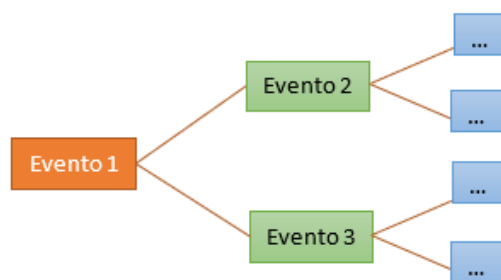
Los diagramas de árbol consisten en representar gráficamente cada uno de los eventos que intervienen en un experimento aleatorio; permite identificar de manera más fácil las probabilidades de cada evento y reconocer cuál es dependientes e independiente (Ross, 2019).

Como su nombre lo indica, tienen la forma de un árbol, donde desde el punto inicial (o punto de partida), salen las ramas que representan los posibles resultados de ese evento, al final de cada rama se coloca un nuevo punto y desde allí salen las nuevas ramas que se conectan a los eventos de ser el caso. Si existen más eventos se repite el proceso para los eventos posteriores, continuando hasta que se haya representado todos los eventos del experimento (Ross, 2019).

A lo largo de estas ramas se registran las probabilidades condicionadas correspondientes a cada evento. La probabilidad de seguir una trayectoria específica, desde la raíz inicial hasta cualquier nodo final, se calcula mediante el producto de las probabilidades de todas las ramas que componen ese camino. Se debe tener en cuenta que la suma de las probabilidades que llegan a un mismo nodo es de 1 o del 100%.

Figura 45

Representación de un diagrama de árbol

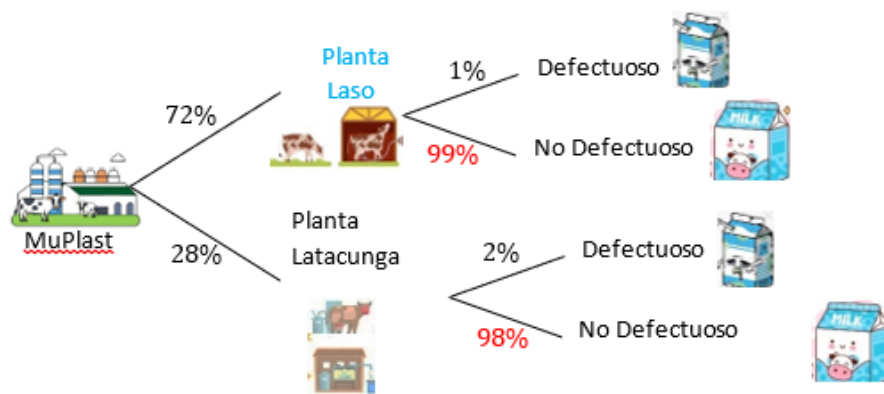


Ejemplo

Una importante empresa que fabrica envases plásticos para yogurt casero, tiene dos plantas, una en Laso y otra en Latacunga. La planta de Laso produce el 72% de los envases y la planta de Latacunga el 28%. Los envases defectuosos en la fabricación son del 1% y 2% respectivamente de cada planta. Se desea conocer la probabilidad de que un producto defectuoso provenga de la planta de Laso.

1. Prepare los datos:

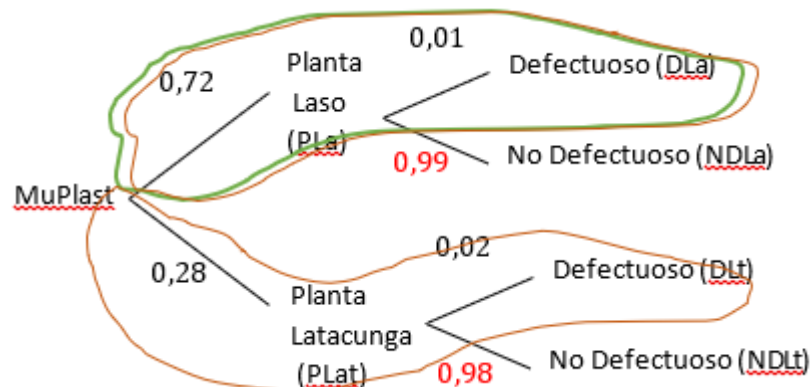
Se dibuja el diagrama de árbol representando las plantas y los envases con defecto y sin defecto con sus respectivas probabilidades.



Una vez que se realiza el diagrama de árbol, se desea conocer la probabilidad de que un producto defectuoso sea de la planta de Laso, esto es la probabilidad condicionada, pero también se observa que hay dos ramas con productos defectuosos la de la planta de Laso y la de la planta de Latacunga, así:

2. Análisis:

- Representar la información en un diagrama de árbol



- b. Se identifica y calcula las probabilidades de cada evento y las condiciones:

$$P(PLa|DLa) = 0,01$$

$$P(PLa) = 0,72$$

$$P(PLt|DLt) = 0,02$$

$$P(PLt) = 0,28$$

$$P(PLa) = P(PLa|DLa) * P(DLa) + P(PLa|DLt) * P(DLt)$$

$$P(PLa) = (0,01 * 0,72) + (0,02 * 0,28)$$

$$P(PLa) = 0,0128$$

- c. Se aplica la fórmula del Teorema de Bayes:

$$P(DLa|PLa) = \frac{P(PLa|DLa) * P(DLa)}{P(PLa)}$$

$$P(DLa|PLa) = \frac{0,72 * 0,01}{0,0128} = 0,5625$$

3. Interpretación:

Por lo tanto la probabilidad de que un envase plástico defectuoso haya sido fabricado en la planta de Laso es 56,25%

5.6.3. Tablas de contingencia

Son una herramienta estadística muy útil para organizar y analizar datos categóricos, permitiendo estudiar la relación entre dos o más variables. Estas tablas muestran la frecuencia con la que ocurren combinaciones específicas de categorías de las variables, ayudando a identificar posibles dependencias o correlaciones entre ellas (Ross, 2019).

Estructura de una tabla de contingencia

- Una tabla de contingencia se organiza en filas y columnas, donde cada fila representa la categoría de una variable y cada columna representa la categoría de otra variable.

- Las celdas contienen la frecuencia (o el conteo) de las observaciones que corresponden a la combinación de las categorías de ambas variables.
- La tabla de contingencia nos permite observar la relación entre las dos variables.
- La información de las frecuencias en la tabla se puede transformar en probabilidades.

Tabla 16

Representación de la tabla de Contingencia

Categoría Variable 2	Categoría Variable 1		TOTAL
	V11	V12	
V21			
V22			
TOTAL			

Ejemplo

En la tabla siguiente se recopila la información sobre los resultados de una prueba para detectar el COVID sobre una población de 1000 personas de la provincia de Latacunga. Encontrar la probabilidad de que:

- Las personas estén enfermas
- El resultado sea negativo
- Estén enfermas sabiendo que el resultado de la prueba fue negativo

	Resultados		
	Positivo	Negativo	TOTAL
Enfermo	95	5	100
No Enfermo	90	810	900
TOTAL	185	815	1000

1. Prepare los datos:

Los datos se organizan en Tablas de Contingencia y se procede a realizar los respectivos cálculos

2. Análisis:

- a. Para contestar el literal a), “las personas estén enfermas” aplicamos la fórmula básica de probabilidad, sucesos favorables sobre posibles:

$$P(enferma) = \frac{100}{1000} = \frac{1}{10} = 0,1$$

Por lo tanto, las personas enfermas son un 10%

- b. Para contestar el literal b), “el resultado sea negativo” se lo realiza como el literal a).

$$P(RNegativo) = \frac{815}{1000} = \frac{163}{200} = 0,815$$

Por lo tanto, las personas con un resultado negativo en sus pruebas es 81,5%

- c. Para responder el literal c), “La persona está enferma sabiendo que el resultado de la prueba fue negativo” aquí tenemos una probabilidad condicional, es decir, sabemos que el resultado de la prueba es negativo y además debe estar enferma; para resolver este apartado, se analiza la tabla, se aplica la fórmula básica de probabilidades y se tiene:

$$P(Enferma|RNegativo) = \frac{5}{815} = \frac{1}{163} = 0,0061$$

Es decir, en el numerador se ubica la condición “enfermo y resultado negativo” y en el denominador lo que ya sabe con seguridad “total resultado negativo de la prueba”.

Finalmente, se concluye que el 0,61% de las personas están enfermas, sabiendo que el resultado de la prueba fue negativo.

Autoevaluación 5.3.

1. Una fábrica tiene tres máquinas (A, B y C) que producen tapas para botella. La máquina A produce el 28%, la B el 47% y la C el 25% de la producción total. El porcentaje de tapas defectuosas que produce cada máquina es del 1,5%, 3% y 0,8% respectivamente.
 - a. Construye una tabla de contingencia hipotética que represente este escenario para una producción total de 15000 tapas.
 - b. Elabora un diagrama de árbol con la información dada de la fábrica
 - c. Si se selecciona una tapa al azar de la producción total y resulta ser defectuosa, ¿cuál es la probabilidad de que haya sido producido por la máquina C?

5.6.4. Ejercicios propuestos

Para resolver los siguientes ejercicios, utilice las tablas de contingencia o diagramas de árbol, según sea el caso:

31. Una empresa recibe importaciones de materia prima de tres proveedores en proporciones del 50%, 30% y 20%. Se sabe que el 0,1% de los lotes del primer proveedor, el 0,5% del segundo, y el 1% del tercero son rechazados debido a fallas del material. Calcule la probabilidad de que un lote sea rechazado.
32. Don Zura panadería y pastelería tiene dos hornos para hornear pan, los mismos hornean el 46% y 54% del total de panes que venden en la panadería, respectivamente. El primer horno quema un 2% de panes y el segundo horno quema 2,5% de los panes. Se selecciona un pan al azar y esta quemado, ¿cuál es la probabilidad que haya sido horneado en el segundo horno?
33. En una Institución Educativa, los estudiantes pueden elegir entre estudiar inglés o mandarín como lengua extranjera. En un curso, el 80% de los estudiantes opta por inglés, mientras que el resto elige mandarín. El 40% de los estudiantes que estudian inglés son hombres, y de los que estudian mandarín, el 65% son mujeres. Si se elige un estudiante al azar, ¿cuál es la probabilidad de que sea una mujer y que estudia mandarín?
34. En el programa navideño “Luz de Noche” se realiza un sorteo, se sortearán 30 cocinas y 20 lavadoras; el 40% de las cocinas son marca Franke

y el resto Electrolux; mientras que el 35% de las lavadoras son marca Electrolux y el resto Franke.

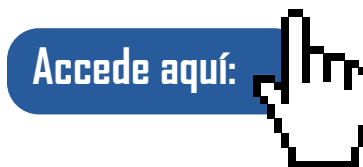
- ¿Cuántas cocinas y cuantas lavadoras existen de cada marca?
- ¿Se realiza el sorteo al azar, encuentre la probabilidad que el ganador se lleve una lavadora Electrolux?

35. Tenemos dos urnas: la urna 1 contiene 5 bolas azules y 3 roja, mientras que la urna 2 tiene 4 bolas azules y 4 rojas. Se extrae una bola de la urna 1 y se coloca en la urna 2. Luego se extrae una bola de la urna 2, resultando ser azul. ¿Cuál es la probabilidad de que la bola transferida de la urna 1 a la 2 haya sido azul?
36. Una importante fábrica de juguetes tiene dos plantas de producción una en Ambato y otra en Latacunga, la producción de las plantas es 235 a 265 respectivamente. El 14% de los juguetes de la planta de Ambato y el 16% de la planta de Latacunga son defectuosos. Si se elige un juguete al azar de toda la producción, ¿cuál es la probabilidad de que sea defectuoso y provenga de la planta de Latacunga?
37. Se realiza una encuesta sobre las preferencias de comida a los estudiantes de una universidad. Los encuestados se dividen en tres grupos: vegetarianos, veganos y omnívoros. La probabilidad de que un encuestado sea vegetariano es 0,45, vegano es 0,3, y omnívoro es 0,25. La probabilidad de que un vegetariano prefiera ensaladas es 0,60, un vegano prefiera frutas es 0,53, y un omnívoro prefiera carnes es 0,78. ¿Cuál es la probabilidad de que un encuestado prefiera ensaladas o frutas o carnes?
38. En un casino “Royal”, hay tres tipos de juegos: keno, dados y blackjack. La probabilidad de que un jugador elija keno es 43%, dados es 35%, y blackjack es 22%. La probabilidad de ganar en keno es 38%, en dados es 48%, y en blackjack es 46%. ¿Cuál es la probabilidad de que un jugador elegido al azar gane en el juego que elija?
39. En almacenes POINT se venden tres tipos de electrodomésticos: televisores, lavadoras y frigoríficos. La probabilidad de que un electrodoméstico vendido sea un televisor es 0,35, una lavadora es 0,42, y un frigorífico es 0,23. La probabilidad de que un televisor sea defectuoso es 0,056; una lavadora sea defectuosa es 0,12; y un frigorífico sea defectuoso es 0,084. Si se sabe que un electrodoméstico vendido al azar es defectuoso, ¿cuál es la probabilidad de que sea frigorífico?

40. Se realizó una encuesta a los invitados a una fiesta sobre las preferencias que tienen de sus bebidas, las tres más elegidas son: agua, tequila y ron. La probabilidad de que un encuestado prefiera agua es 0,32, tequila es 0,38, y ron es 0,30. La probabilidad de que un amante del agua prefiera la marca Absolut es 0,40, un amante del tequila prefiera la marca Absolut es 0,36, y un amante del ron prefiera la marca Absolut es 0,52. ¿Cuál es la probabilidad de que un encuestado prefiera la bebida de otras marcas?

Ejercicio propuesto (con base de datos)

La siguiente base de datos contiene información sobre el rendimiento estudiantil en educación secundaria de dos escuelas portuguesas student-por.csv, se la puede encontrar en el siguiente link:



Utilizando SPSS, se desea analizar la probabilidad de que un estudiante tenga un alto consumo de alcohol los fines de semana, dado que proviene de una familia donde al menos uno de los padres tiene un nivel educativo alto.

5.7. Distribución de probabilidades

Los posibles resultados de un experimento y la probabilidad de ocurrencia de cada uno se considera una distribución de probabilidades.

Variable aleatoria

Representa el valor numérico de cada posible resultado de un experimento aleatorio (Ross, 2019).

Los posibles resultados de lanzar dos monedas al aire son:



Eventos:

$$E=\{(c,c);(c,s);(s,c);(s,s)\}$$

Las variables aleatorias pueden ser Discretas y continuas

Variable aleatoria discreta

Una variable aleatoria discreta es la que puede tomar un conjunto finito de valores, es decir toma valores enteros. Como ejemplo se puede mencionar el número de llamadas que recibe un centro de atención al cliente en una hora o el número de veces que una moneda cae en “cara” después de lanzarla varias veces (Lind et al., 2012).

Variable Aleatoria

Discreta

$X =$



Función densidad de probabilidad discreta o masa de probabilidad

Para cada valor posible de x , siendo x la variable aleatoria, la función de densidad de probabilidad es la probabilidad de obtener dicho valor cuando se realiza un experimento dado.

$$p(x) = P(X = x)$$

Para que sea una función de densidad debe cumplir las siguientes condiciones:

1. $p(x) \geq 0$
2. $\sum_{i=1}^n p(x) = 1$

Función de densidad acumulada

Para una variable aleatoria x con función de densidad de probabilidad discreta $p(x)$, para cada valor de x es:

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y)$$

$F(x)$ representa la probabilidad de que el valor observado X tomara un valor máximo de x .

Media o esperanza matemática

De una variable aleatoria x es un número real igual a los datos por su respectiva frecuencia: (Galindo, 2011).

$$E(x) = \sum_{i=1}^n x_i * p_i$$

Donde:

- n es el número de datos
- x_i es la variable aleatoria (dato)
- p_i la probabilidad de ocurrencia de la variable aleatoria

Varianza

Representa como se distribuyen los datos alrededor de la media y es el cuadrado de la diferencia entre el dato de variable aleatoria y la esperanza matemática multiplicado por la probabilidad de ocurrencia: (Galindo, 2011).

$$Var(x) = \sigma^2 = \sum_{i=1}^n (x_i - E(x))^2 * p_i$$

A mayor varianza los datos se encuentran más alejados de la media; mientras que a menor varianza los valores están más concentrados alrededor de la media.

Desviación estándar

Es la raíz de la varianza:

$$\sigma = \sqrt{Var(x)}$$

Ejemplo

En una fábrica de peluches se sabe que la probabilidad de obtener un peluche defectuoso es:

Número de peluches defectuosos x	0	1	2	3	4	5
Probabilidad $P(x)$	0,4	0,2	0,1	0,1	0,15	0,05

- Encontrar el valor esperado de los peluches defectuosos
- La varianza y la desviación típica de un peluche defectuoso

1. Prepare los datos:

Realizar los cálculos necesarios para aplicar las fórmulas y organizarlos en una tabla.

Número de peluches defectuosos x	Probabilidad $p(x)$	$x * p(x)$	$(x - E(x))^2$	$(x - E(x))^2 * p(x)$
0	0,40	0,00	2,40	0,96
1	0,20	0,20	0,30	0,06
2	0,10	0,20	0,20	0,02
3	0,10	0,30	2,10	0,21
4	0,15	0,60	6,00	0,90
5	0,05	0,25	11,90	0,60
		1,55		2,75

2. Análisis:

Cálculo de la media o esperanza matemática:

El resultado del cálculo de la esperanza matemática se encuentra en la tercera columna de la tabla, ya que esta columna tiene los cálculos de cada producto de x por $p(x)$, en la última celda se muestra el resultado de la suma de todos estos productos.

$$E(x) = \sum_{i=1}^n x_i * p_i = \sum_{i=1}^6 x_i * p_i$$

$$E(x) = 1,55$$

Al igual que en la esperanza matemática, se realizan los cálculos para encontrar la varianza.

$$Var(x) = \sigma^2 = \sum_{i=1}^6 (x_i - E(x))^2 * p_i$$

$$Var(x) = \sigma^2 = 2,75$$

Cálculo de la desviación típica:

$$\sigma = \sqrt{Var(x)} = \sqrt{2,75}$$

$$\sigma = 1,66$$

5.7.1. Distribución de probabilidades discretas

Distribución binomial

Como su palabra lo indica, tiene dos posibles resultados: éxito o fracaso; es decir indica el número de éxitos en una serie de ensayos independientes (Ross, 2019).

Características de la distribución binomial

Para que una variable siga una distribución binomial, debe cumplir con las siguientes condiciones:

- Número fijo de intentos o ensayos (n): Hay un número determinado de intentos que no pueden cambiar mientras se desarrolla el experimento.
- Dos posibles resultados en cada ensayo: Éxito o fracaso.
- Independencia de los ensayos: El resultado de cada ensayo no afecta a los otros resultados.

La probabilidad de obtener exactamente k éxitos en n ensayos está dada por la función de probabilidad binomial:

$$P(x = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Donde:

- x número de éxitos de la variable aleatoria.
- n es el número total de ensayos o intentos.
- k es el número de éxitos deseados.
- p es la probabilidad de éxito en cada ensayo.

- $(1 - p)$ es la probabilidad de fracaso (q)
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ es el coeficiente binomial, que cuenta las posibles combinaciones de k éxitos en n intentos.

Ejemplo práctico N°1

Si se lanza una moneda al aire 15 veces, se desea conocer la probabilidad de obtener 6 veces cara.

1. Prepare los datos:

A. Identificar los datos

$S = \{\text{Cara}, \text{Sello}\} = 2$ Espacio muestral o suceso posibles de una moneda

$p = \frac{1}{2} = 0,5$ probabilidad de obtener cara en una moneda

$n = 15$ número total de lanzamientos

$k = 6$ éxitos, lo que se desea obtener

2. Análisis:

$$P(x = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$P(x = 6) = \binom{15}{6} 0,5^6 (1 - 0,5)^{15-6}$$

$$P(x = 6) = \frac{15!}{6! (15 - 6)!} 0,5^6 (1 - 0,5)^{15-6}$$

$$P(x = 6) = \frac{5005}{32768} = 0,1527$$

Construcción en SPSS:

1. Preparación de los datos

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana "Vista de variables" cree las variables k , n y p con los siguientes parámetros:

- **Nombre:** k
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** éxitos
 - **Medida:** Escala
-
- **Nombre:** n
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** número de ensayos
 - **Medida:** Escala
-
- **Nombre:** p
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** probabilidad de éxitos
 - **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	k	Numérico	8	0	Éxitos	Ninguno	Ninguno	4	Centro	Escala	Entrada
2	n	Numérico	8	0	Número de ensayos	Ninguno	Ninguno	4	Centro	Escala	Entrada
3	p	Numérico	8	4	Probabilidad de éxito	Ninguno	Ninguno	4	Centro	Escala	Entrada
4	Prob_6_caras	Numérico	8	4	Probabilidad k éxitos	Ninguno	Ninguno	14	Centro	Escala	Entrada

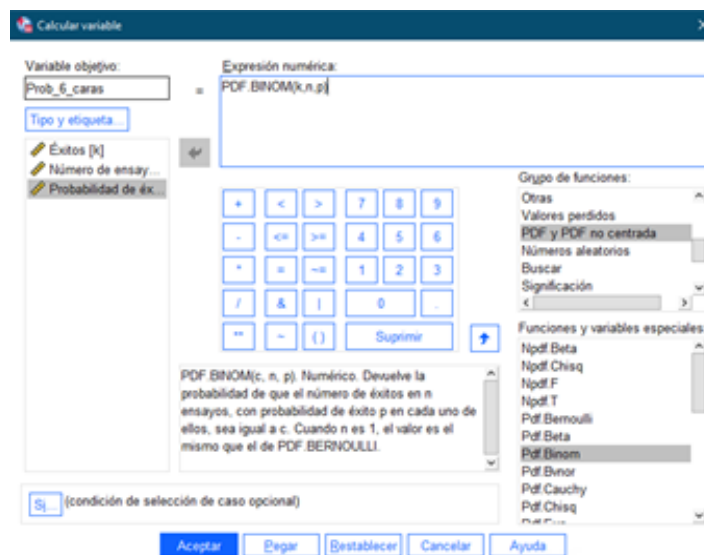
- c. En la pestaña “Vista de datos” ingrese los valores de las variables éxitos, número de intentos y probabilidad de éxito.

Archivo	Editar	Ver	Datos	Tr
14 :				
1	k	n	p	
6	15	,5000		

2. Análisis probabilístico:

En el menú principal seleccione: Transformar / Calcular variable y de despliega otra ventana

- Nombre de la variable
- En grupo de funciones elegimos PDF y PDF no centrada
- En la sección de Funciones y variables especiales seleccionamos Pdf: Binom
- En la sección expresión numérica aparece la función Binomial, y completamos con los parámetros de las variables k , n y p ; cuyas variables se encuentran en la sección Tipo y etiqueta.
- Finalmente, clic en Aceptar



3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la probabilidad a encontrar.

Archivo Editar Ver Datos Transformar Analizar				
3 : n				
	k	n	p	Prob_6_caras
1	6	15	.5000	.1527

4. Interpretación

La probabilidad de obtener exactamente 6 caras al lanzar una moneda 15 veces es 15.27%

Esperanza y Varianza de la Distribución Binomial

Para una variable aleatoria X que sigue una distribución binomial $X \sim \text{Binomial}(n,p)$: (Ross, 2019)

Esperanza matemática (media): $E(x)=n \cdot p$

Varianza: $\text{Var}(x)=n \cdot p \cdot (1-p)$

Ejemplo

Calcule la esperanza matemática, la varianza y la desviación típica del lanzamiento de una moneda al aire 15 veces con la probabilidad de obtener 6 caras

1. Prepare los datos:

A. Identificar los datos

$S = \{\text{Cara}, \text{Sello}\} = 2$ Espacio muestral o suceso posibles de una moneda

$p = \frac{1}{2} = 0,5$ probabilidad de obtener cara en una moneda

$n = 15$ número total de lanzamientos

$k = 6$ éxitos, lo que se desea obtener

2. Análisis probabilístico

A. Esperanza matemática

$$E(x) = n \cdot p$$

$$E(x) = 15 \cdot 0,5$$

$$E(x) = 7,5$$

B. Varianza y Desviación Típica

$$\text{Var}(x) = n \cdot p \cdot (1 - p)$$

$$\text{Var}(x) = 15(0,5) \cdot (1 - 0,5)$$

$$\text{Var}(x) = 3,75$$

$$\sigma = \sqrt{\text{Var}(x)} = \sqrt{3,75}$$

$$\sigma = 1,94$$

Distribución Hipergeométrica

Esta distribución se fundamenta en la distribución binomial, con la diferencia que los ensayos o intentos no son independientes entre sí (Ross, 2019).

Características de la Distribución Hipergeométrica

- Población finita y fija de tamaño N
- Dos tipos de elementos en la población éxito o fracaso.
- Muestra sin reemplazo

La probabilidad de obtener exactamente k éxitos en una muestra de tamaño n se calcula con la ecuación:

$$P(x = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Donde:

- x número de éxitos de la variable aleatoria en la muestra.
- N es el tamaño total de la población.
- n es el tamaño de la muestra.
- K es el número total de éxitos en la población.
- k es el número total de éxitos en la muestra.
- $\binom{K}{k} = \frac{K!}{k!(K-k)!}$ es el número de maneras de elegir k éxitos de K éxitos disponibles.
- $\binom{N-K}{n-k} = \frac{(N-K)!}{(n-k)!(N-K-(n-k))!}$ es el número de maneras de elegir $(n-k)$ fracasos de $(N-K)$ fracasos disponibles.
- $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ es el número total de maneras de elegir n elementos de N elementos totales

Ejemplo práctico N°2

En una caja con estuches para celulares hay 200 estuches, de los cuales 80 tienen falla y 120 no presentan fallas. Si se seleccionan 10 estuches al azar sin

reemplazo, ¿cuál es la probabilidad de que exactamente 4 de los estuches seleccionados tengan fallas?

1. *Prepare los datos:*

$N = 200$	total, estuches
$K = 80$	estuches con fallas
$n = 10$	muestra
$k = 4$	número de estuches con fallas en la muestra
$N - K = 120$	número de estuches sin falla
$n - k = 10 - 4 = 6$	número de estuches sin falla de la muestra

2. *Análisis:*

$$P(x = k) = \frac{\binom{K}{k} \binom{N - K}{n - k}}{\binom{N}{n}}$$

$$P(x = 4) = \frac{\binom{80}{4} \binom{120}{6}}{\binom{200}{10}}$$

$$P(x = 4) = 0,2573$$

3. *Interpretación:*

La probabilidad de seleccionar exactamente 4 defectuosos de la muestra es 25,73%

Construcción en SPSS:

1. *Preparación de los datos*

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las variables N , K , k y n con los siguientes parámetros:

- **Nombre:** N
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** Población
- **Medida:** Escala

- **Nombre:** K
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** éxitos población
- **Medida:** Escala

- **Nombre:** k
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** éxitos muestra
- **Medida:** Escala

- **Nombre:** n
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** número de ensayos muestra
- **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	N	Numérico	8	2	Población	Ninguno	Ninguno	8	Centro	Desconocido	Entrada
2	K	Numérico	8	2	éxitos población	Ninguno	Ninguno	8	Centro	Desconocido	Entrada
3	n_muestra	Numérico	8	2	número de ensayos muestra	Ninguno	Ninguno	8	Centro	Desconocido	Entrada
4	k_muestra	Numérico	8	2	éxitos muestra	Ninguno	Ninguno	8	Centro	Desconocido	Entrada

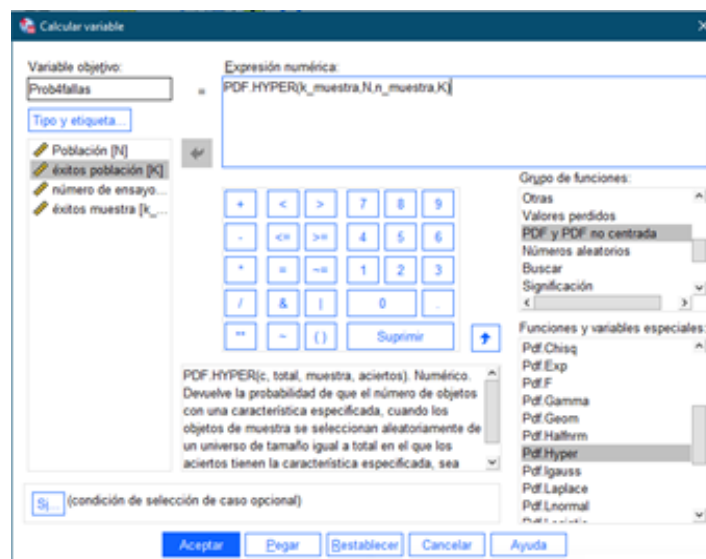
- c. En la pestaña “Vista de datos” ingrese los valores de las variables Población, éxitos población, éxitos muestra y número de ensayos muestra.

Archivo Editar Ver Datos Transformar Analizar Gráficos Utili				
1 :				
	N	K	n_muestra	k_muestra
1	200,00	80,00	10,00	4,00

2. Análisis probabilístico:

En el menú principal seleccione: **Transformar / Calcular variable** y se despliega otra ventana

- Nombre de la variable
- En grupo de funciones elegimos **PDF y PDF no centrada**
- En la sección de Funciones y variables especiales seleccionamos **PDF: Hyper**
- En la sección expresión numérica aparece la función Hipergeométrica, y completamos con los parámetros de las variables **N, K, k** y **n**; cuyas variables se encuentran en la sección Tipo y etiqueta.
- Finalmente, clic en **Aceptar**



3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la probabilidad a encontrar.

	N	K	n_muestra	k_muestra	Prob4fallas
1	200,00	80,00	10,00	4,00	.2573

4. Interpretación

La probabilidad de seleccionar exactamente 4 estuches defectuosos de la muestra es 25.73%.

Esperanza, Varianza y desviación Típica de la Distribución Hipergeométrica.

Para una variable aleatoria X que sigue una distribución hipergeométrica representada por $H(N, K, n)$, la esperanza y la varianza es igual a:

Esperanza matemática (media): $E(x) = n * \frac{K}{N}$

Varianza: $Var(x) = n * \frac{K}{N} * \frac{N-K}{N} * \frac{N-n}{N-1}$

Ejemplo

Calcule la esperanza matemática, la varianza y la desviación típica del Ejemplo Practico anterior

1. Prepare los datos:

$N = 200$	total, estuches
$K = 80$	estuches con fallas
$n = 10$	muestra
$k = 4$	número de estuches con fallas en la muestra
$N - K = 120$	número de estuches sin falla
$n - k = 10 - 4 = 6$	número de estuches sin falla de la muestra

2. Análisis:

$$E(x) = n * \frac{K}{N} = 10 * \frac{80}{200}$$

$$E(x) = 4$$

$$Var(x) = n * \frac{K}{N} * \frac{N-K}{N} * \frac{N-n}{N-1} = 10 * \frac{80}{200} * \frac{200-80}{200} * \frac{200-10}{200-1} = 10 * \frac{80}{200} * \frac{120}{200} * \frac{190}{199}$$

$$Var(x) = 2,29$$

$$\sigma = \sqrt{Var(x)} = \sqrt{2,29}$$

$$\sigma = 1,51$$

Distribución Geométrica

Al igual que en la distribución binomial, cada ensayo tiene dos posibles resultados, éxitos y fracasos y la probabilidad de éxito es constante en cada intento. Presenta el número de ensayos o intentos necesarios hasta obtener el primer éxito en una serie de ensayos independientes. (Ross, 2019)

Características de la Distribución Geométrica

- Dos posibles resultados, éxito o fracaso.
- El resultado de cada ensayo no influye en los otros.
- La probabilidad de éxito (p) no cambia en cada ensayo.

La probabilidad de que el primer éxito ocurra en el intento k es igual a:

$$P(x = k) = p * (1 - p)^{k-1}$$

Donde:

- x número de ensayos hasta el primer éxito.
- k es el ensayo en el cual ocurre el primer éxito.
- p es la probabilidad de éxito en cada ensayo.
- $1-p$ es la probabilidad de fracaso en cada ensayo.

Ejemplo práctico N°3

La probabilidad de que una persona compre un producto de oferta en Temu es del 30%. Se desea saber la probabilidad de que el primer cliente que compre la oferta sea el tercero en entrar a la aplicación de Temu.

1. *Prepare los datos:*

$p = 30\% = 0,30$ *probabilidad de comprar un producto en oferta*

$k = 3$ *ubicación del primer éxito*

2. *Análisis:*

$$P(x = k) = p * (1 - p)^{k-1}$$

$$P(x = 3) = 0,30 * (1 - 0,30)^{3-1}$$

$$P(x = 3) = 0,30 * (0,70)^2$$

$$P(x = 3) = 0,147$$

3. *Interpretación:*

La probabilidad de que el primer cliente que compre sea el quinto en entrar a la tienda es 14,7%.

Construcción en SPSS:

1. *Preparación de los datos*

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree las variables p y k con los siguientes parámetros:

- **Nombre:** p
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** Probabilidad de éxito
- **Medida:** Escala
- **Nombre:** k
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** ensayo ocurre primer éxito
- **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	p	Númerico	8	2	Probabilidad de éxito en cada ensayo	Ninguno	Ninguno	8	Centro	Escala	Entrada
2	k	Númerico	8	2	Ensayo ocurre el primer éxito	Ninguno	Ninguno	8	Centro	Escala	Entrada

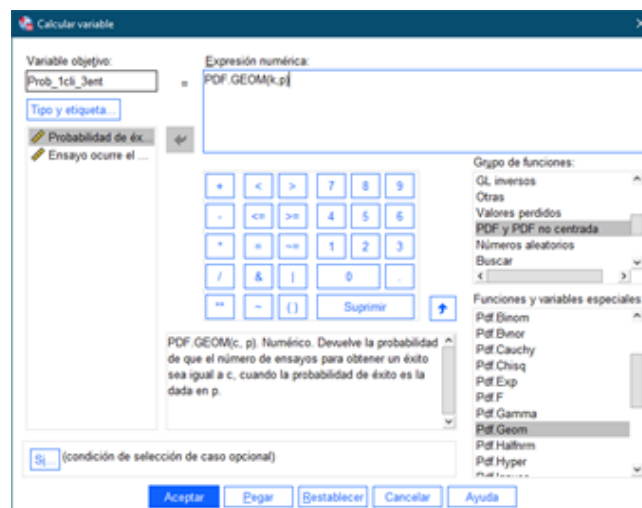
- c. En la pestaña “Vista de datos” ingrese los valores de las variables Población y el valor del ensayo donde ocurre el primer éxito.

Archivo	Editar	Ver	Datos	Transformar
7 :				
1	p	k		
	,30	3,00		

2. Análisis probabilístico:

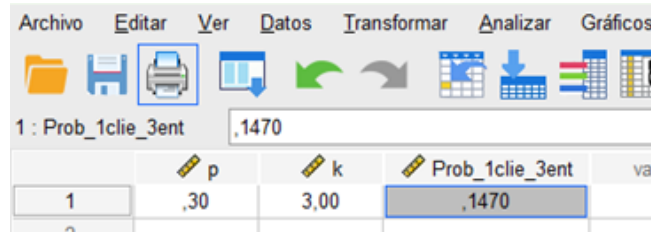
En el menú principal seleccione: **Transformar / Calcular variable** y se despliega otra ventana

- Nombre de la variable
- En grupo de funciones elegimos **PDF y PDF no centrada**
- En la sección de Funciones y variables especiales seleccionamos **Pdf: Geom**
- En la sección expresión numérica aparece la función Geométrica, y completamos con los parámetros de las variables **k** y **p**; cuyas variables se encuentran en la sección Tipo y etiqueta.
- Finalmente, clic en **Aceptar**



3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la probabilidad a encontrar.



	p	k	Prob_1clie_3ent	var
1	.30	3.00	.1470	

4. Interpretación

La probabilidad de que el primer cliente que compre sea el quinto en entrar a la tienda es 14,7%.

Esperanza, varianza y desviación típica de la distribución geométrica

Para una variable aleatoria X que sigue una distribución geométrica $X \sim \text{Geom}(p)$, la esperanza y la varianza son:

Esperanza matemática (media): $E(x) = \frac{1}{p}$

Varianza: $\text{Var}(x) = \frac{1-p}{p^2}$

Desviación Típica: $\sigma = \sqrt{\text{Var}(x)}$

Ejemplo

Calcule la esperanza matemática, la varianza y la desviación típica del ejemplo Práctico anterior

1. Prepare los datos:

$$p = 30\% = 0,30$$

probabilidad de comprar un producto en oferta

$$k = 3$$

ubicación del primer éxito

2. Análisis:

a. Cálculo de la esperanza matemática:

$$E(x) = \frac{1}{p}$$

$$E(x) = \frac{1}{0,30} = 3,33$$

b. Cálculo de la varianza y la desviación típica

$$Var(x) = \frac{1-p}{p^2}$$

$$Var(x) = \frac{1-0,30}{0,30^2} = 7,78$$

$$\sigma = \sqrt{Var(x)} = \sqrt{7,78} = 2,79$$

Distribución de Poisson

Esta distribución describe las veces que ocurre un evento en un intervalo fijo de tiempo, espacio o área, cuando estos eventos ocurren de forma independiente y con una tasa promedio constante (Lind et al., 2012).

Características de la distribución de Poisson

- Número de eventos en un intervalo fijo de tiempo, área o espacio.
- Eventos independientes uno de otro.
- La tasa promedio de los eventos en el intervalo debe ser constante.

La probabilidad de que ocurra exactamente k eventos en un intervalo dado, con una tasa promedio λ , se calcula con la función de probabilidad de Poisson:

$$P(x = k) = \frac{\lambda^k * e^{-\lambda}}{k!}$$

Donde:

- x número de eventos en el intervalo.
- k es el número específico de eventos que queremos que ocurra.
- λ es la tasa promedio de ocurrencia de eventos en el intervalo.
- e constante de Euler

Ejemplo Práctico N°4

El centro de atención al cliente de CNT ubicado en el Mall el Jardín recibe en promedio 17 llamadas por minuto. Se desea saber la probabilidad de que en un minuto se reciban exactamente 10 llamadas.

1. Prepare los datos:

$\lambda = 17$	promedio de llamadas por minuto
$k = 10$	número de llamadas que se desea en un minuto

2. Análisis:

$$P(x = k) = \frac{\lambda^k * e^{-\lambda}}{k!}$$

$$P(x = 5) = \frac{17^{10} * e^{-17}}{10!}$$

$$P(x = 3) = 0,023$$

3. Interpretación:

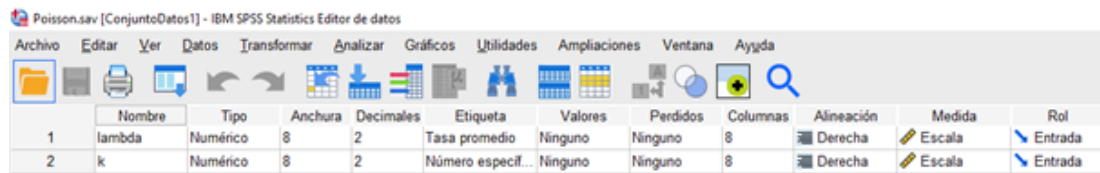
La probabilidad de que en un minuto se reciban exactamente 10 llamadas es 2,3% un porcentaje mínimo.

Construcción en SPSS:

1. Preparación de los datos

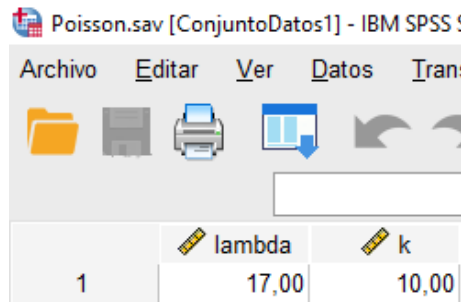
- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las variables lambda y k con los siguientes parámetros:

- **Nombre:** Lambda
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** Tasa promedio
 - **Medida:** Escala
-
- **Nombre:** k
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** Número específico de eventos
 - **Medida:** Escala



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	lambda	Númerico	8	2	Tasa promedio	Ninguno	Ninguno	8	Derecha	Escala	Entrada
2	k	Númerico	8	2	Número especif...	Ninguno	Ninguno	8	Derecha	Escala	Entrada

- c. En la pestaña “Vista de datos” ingrese los valores de las variables landa y número específico de eventos.

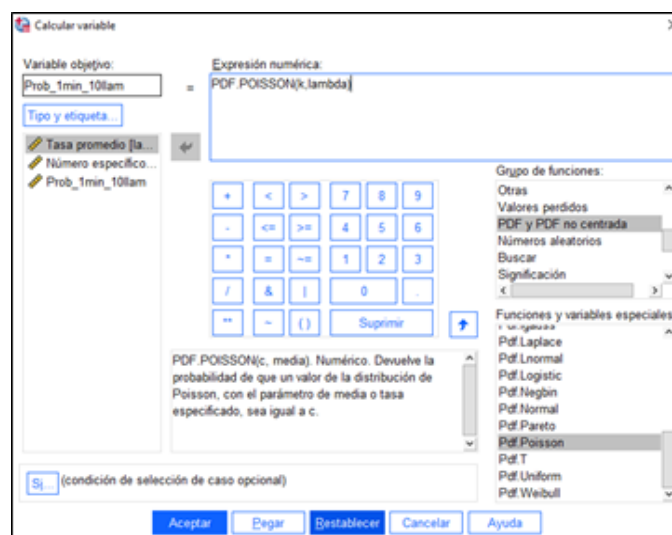


	lambda	k
1	17,00	10,00

2. Análisis probabilístico:

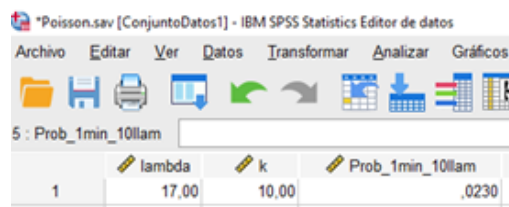
En el menú principal seleccione: **Transformar / Calcular variable** y se despliega otra ventana

- Nombre de la variable
- En grupo de funciones elegimos **PDF y PDF no centrada**
- En la sección de Funciones y variables especiales seleccionamos **PDF: Poisson**
- En la sección expresión numérica aparece la función Poisson, y completamos los parámetros de las variables λ y k ; cuyas variables se encuentran en la sección Tipo y etiqueta.
- Finalmente, clic en **Aceptar**



3. Resultados obtenidos:

En la Ventana de Resultados, encontrará el resultado de la probabilidad a encontrar.



	lambda	k	Prob_1min_10llam
1	17,00	10,00	,0230

4. Interpretación:

La probabilidad de que en un minuto se reciban exactamente 10 llamadas es 2.3% un porcentaje mínimo.

Esperanza y Varianza de la Distribución de Poisson

Esperanza matemática (media): $E(x)=\lambda$

Varianza: $Var(x)=\lambda$

Desviación Típica: $\sigma = \sqrt{Var(x)}$

Ejemplo

Calcule la esperanza matemática, la varianza y la desviación típica del ejemplo práctico anterior

1. Prepare los datos:

$\lambda = 17$ promedio de llamadas por minuto

$k = 10$ número de llamadas que se desea en un minuto

2. Análisis:

a. Cálculo de la esperanza o media

$$E(x) = \lambda = 17$$

b. Cálculo de la varianza y la desviación típica

$$Var(x) = 17$$

$$\sigma = \sqrt{Var(x)} = \sqrt{17} = 4,12$$

Ejercicio práctico (con base de datos)

Se tiene una base de datos de mediciones horarias de calidad del aire Air Quality, disponible en el siguiente enlace:

Accede aquí:



Donde una de las variables clave para de la medición es CO(GT) (Monóxido de Carbono en mg/m^3). Tomar en cuenta que:

- Nivel alto de Concentración de CO: $\text{CO}(\text{GT}) \geq 2.5$
- Evento raro de benceno $\text{C}_6\text{H}_6(\text{GT}) > 20 \mu\text{g}/\text{m}^3$.
- Nivel alto de óxidos de nitrógeno totales $\text{NO}_x(\text{GT}) \geq 150 \text{ ppb}$.

Con lo antes expuesto, se pide calcular las siguientes probabilidades:

- Suponiendo que en un día se registran 28 mediciones horarias de CO(-GT) y que la probabilidad de que una hora tenga un nivel alto de CO es del 30%, calcular la probabilidad de que al menos 4 horas tengan niveles altos.
- Si el número promedio de horas con eventos raros de benceno es de 3 por día, calcula la probabilidad de que en un día no ocurra ningún evento raro.
- En un lote de 48 mediciones de NO_x , se sabe que 12 tienen niveles altos. Si se seleccionan 6 mediciones al azar sin reemplazo, calcular a probabilidad de que a lo mucho una sea alta.
- Si la probabilidad de que una hora tenga un nivel crítico de CO ($\text{CO}(\text{GT}) \geq 5$) es del 15%, calcular la probabilidad de que la primera hora crítica ocurra en la cuarta medición.

Solución:

- Suponiendo que en un día se registran 28 mediciones horarias de CO(-GT) y que la probabilidad de que una hora tenga un nivel alto de CO es del 30%, calcular la probabilidad de que al menos 4 horas tengan niveles altos.

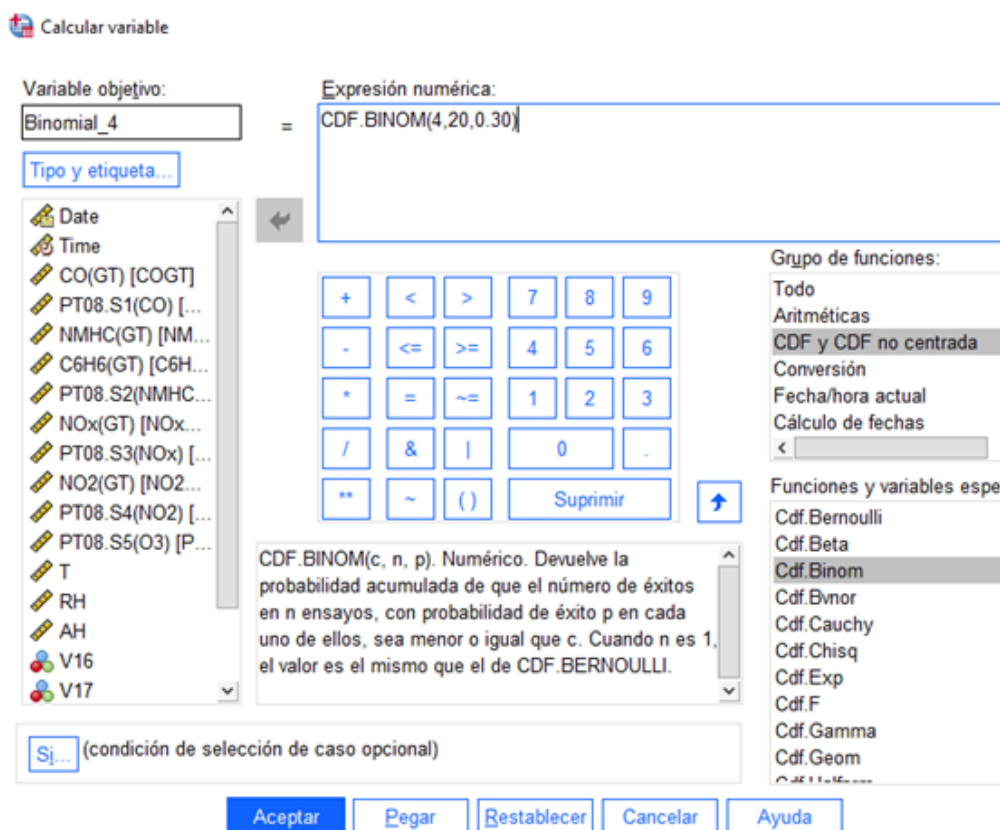
1. Eliminar valores negativos de CO, se supondrá que los negativos son valores que no se leyeron correctamente, por lo que dar clic en Datos > Seleccionar Casos > Si se satisface la condición e ingresar $CO(GT) \geq 0$.



2. Definir “éxito” como $CO(GT) \geq 3$ (nivel alto de CO), dar clic en transformar > Calcular Variable y escribir la condición:



Ahora se definen los datos para poder aplicar la distribución binomial: $n=28$, $p=0.30$ y $k \leq 4$; estos valores se ingresan en el SPSS:



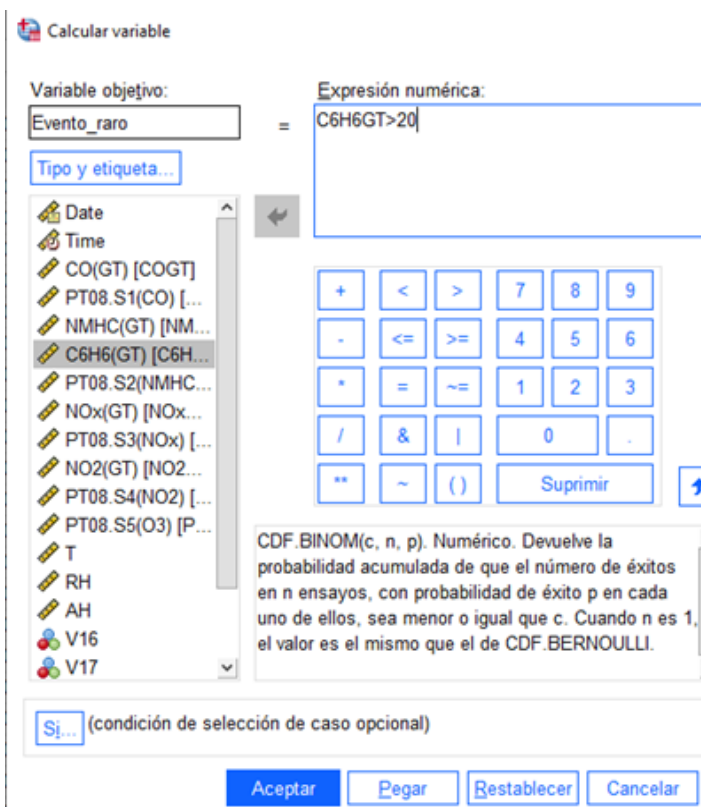
Obteniendo el siguiente resultado:

Binomial_4
,24

Por lo tanto, la probabilidad de que al menos 4 horas tengan niveles altos de CO es de 24%.

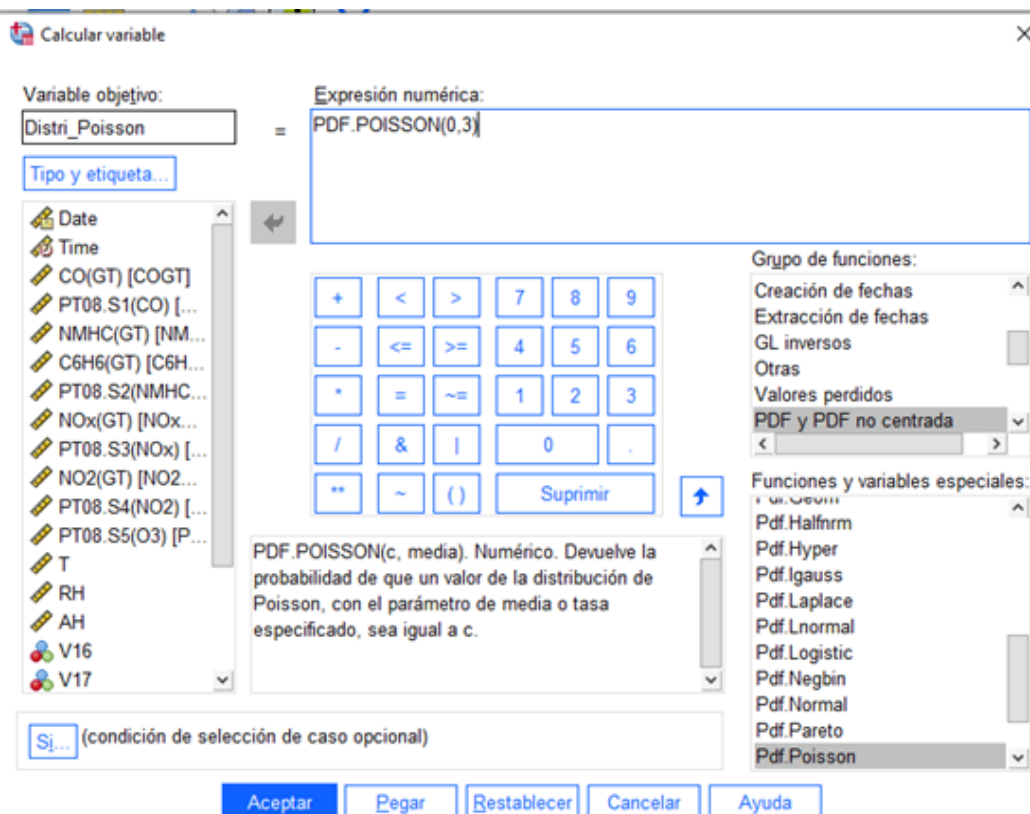
- b. Si el número promedio de horas con eventos raros de benceno es de 3 por día, calcula la probabilidad de que en un día no ocurra ningún evento raro.

Definir el evento raro de Benceno $> 20 \mu\text{g}/\text{m}^3$, dar clic en Transformar > Calcular variable:



Como ya se tiene todos los datos $\lambda=3, k=0$.

Se ingresa la información en el SPSS y se calcula la probabilidad deseada:



Obteniendo un valor de:

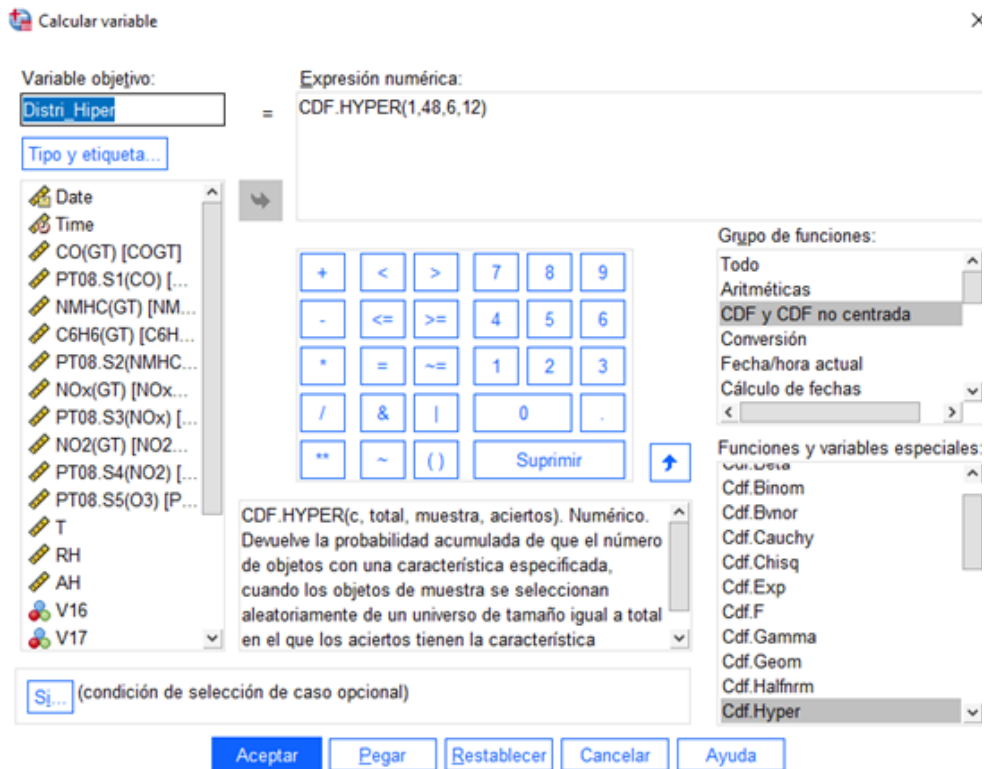
 Distri_Poisson
.05

Por lo tanto, la probabilidad de que en un día no ocurra ningún evento raro de Benceno es de 5%.

- c. En un lote de 48 mediciones de NO_x, se sabe que 12 tienen niveles altos. Si se seleccionan 6 mediciones al azar sin reemplazo, calcular la probabilidad de que a lo mucho una sea alta.

Se definen los datos necesarios para el cálculo: N=48, K=12, n=6, k≤1

Se ingresa los valores al SPSS:



Calcular variable

Variable objetivo: **Distri_Hiper**

Expresión numérica: **CDF.HYPER(1,48,6,12)**

Grupo de funciones: **CDF y CDF no centrada**

Funciones y variables especiales: **Cdf.Hyper**

Botones: Aceptar, Pegar, Restablecer, Cancelar, Ayuda

Obteniendo el siguiente resultado:

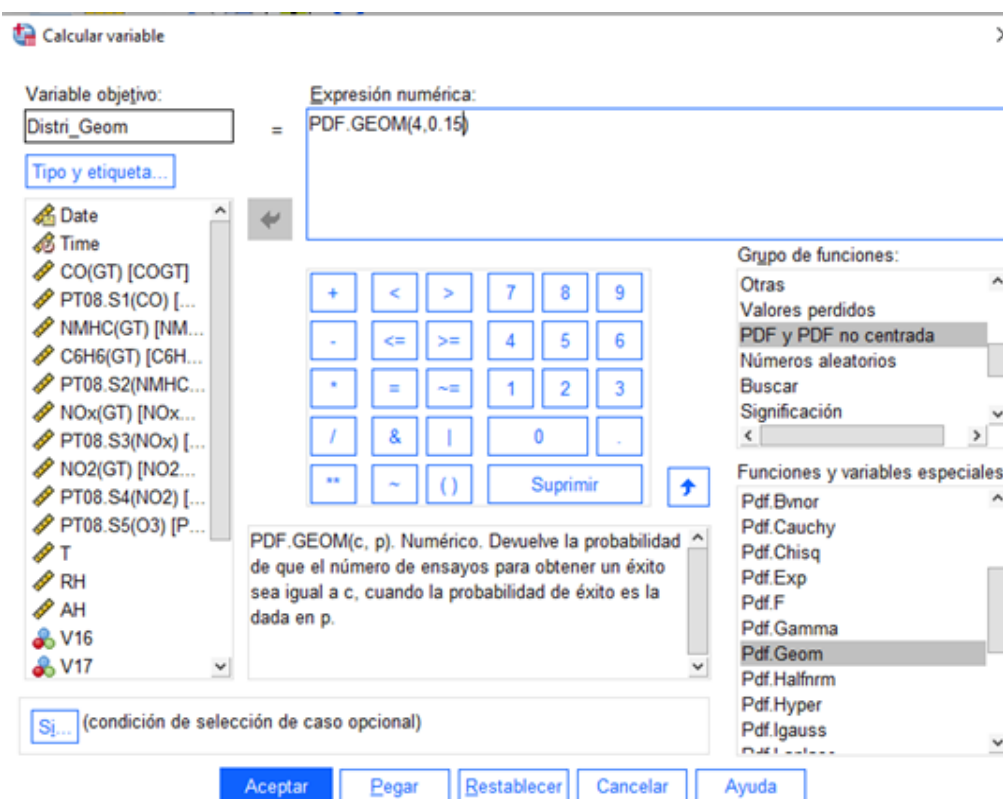
 Distri_Hiper
.53

Por lo tanto, la probabilidad de que a lo mucho una medición de óxido de nitrógeno sea alta es de 53%.

- d. Si la probabilidad de que una hora tenga un nivel crítico de CO (CO(GT) ≥ 5) es del 15%, calcular la probabilidad de que la primera hora crítica ocurra en la cuarta medición.

Se definen los datos: $k=4$, $p=0.15\%$

Estos datos se ingresan en el SPSS:



Obteniendo el siguiente resultado:

Distri_Geom	
	.09

La probabilidad de que la primera hora crítica ocurra en la cuarta medición es de 9%.

Autoevaluación 5.4.

1. Menciona un ejemplo real donde se puede aplicar la distribución binomial.
2. ¿Escribe una diferencia entre la distribución hipergeométrica y la binomial?
3. ¿Cuál es el propósito de la distribución de Poisson?
4. Un jugador de baloncesto de la NBA tiene una probabilidad de 85% de encestar un lanzamiento de tiros libres:
 - a. Calcula la probabilidad de que enceste exactamente 8 de 10 lanzamientos.
 - b. Si en un partido intenta 30 lanzamientos de tiro libre, ¿cuál es la probabilidad de que enceste al menos 25?

Ejercicios propuestos

41. En uno de los centros de salud del Puyo existe un 62% de probabilidad de ocurrir una emergencia grave cada vez que alguien entra a este centro de salud. ¿Cuál es la probabilidad de que dicha emergencia grave suceda entre 3 de 8 personas ingresadas el día de hoy?
42. El equipo de baloncesto femenino U15 del Ecuador tiene una probabilidad del 40% de ganar el campeonato sudamericano desarrollado en Brasil. Si juegan 6 partidos, calcula la probabilidad de que ganen:
 - 3 partidos
 - 0 partidos
 - Todos los partidos
43. De tres a cinco de la tarde los padres de Samantha una niña de 12 años le permiten usar su computadora para realizar tareas en internet; el software de supervisión le admite una de cada cuatro páginas que ella busca. ¿Cuál es la probabilidad de que, cuando se busquen 8 páginas elegidas al azar, el software solo autorice dos?
44. Un agente de seguros vende pólizas a diez personas de entre 50 y 60 años de edad, que son muy saludables. Según las tablas proporcionadas por el INEC, la probabilidad de que una persona entre 50 y 60 años

saludable viva 25 años o más es $3/4$. Hállese la probabilidad de que, transcurridos 25 años, vivan:

- Las diez personas
 - Al menos dos personas
 - Exactamente tres personas
45. En una fábrica de tuercas se sabe que el 2,34% son defectuosos; las tuercas son empaquetadas en cajas de 100 tuercas. Calcular la probabilidad de que en una caja haya:
- Ninguna tuerca defectuosa
 - 5 tuercas defectuosas
 - Mas de dos tuercas defectuosas
46. Un lote destinado a Ferrisariato contiene 1000 piezas de tubería de baño de un proveedor extranjero y 2000 unidades de tubería de baño de un proveedor nacional. Si se seleccionan 5 piezas al azar y sin reemplazo, calcular la probabilidad de que al menos una pieza de la muestra sea del proveedor nacional.
47. Se sabe que 60 de los 900 clavos para madera que se producen en un día no pasan el control de calidad del cliente. Se seleccionan del lote 8 clavos al azar y sin reemplazo. Calcular la probabilidad de que a lo mucho 4 clavos no cumplan con los requerimientos del cliente.
48. En un recinto militar se dispone de armas de fuego, dentro de este se tiene una caja con 100 balas, 22 de ellas son de punta blindada, 28 son de punta de plástico. Se selecciona al azar 8 balas. ¿Cuál es la probabilidad de que se seleccione exactamente 5 balas de plástico?
49. Para un campeonato de bolas de acero, el organizador tiene una bolsa que contiene 25 bolas de acero, de las cuales 18 son nuevas y 7 están oxidadas. Si se extraen 4 bolas al azar y sin reemplazo, ¿Cuál es la probabilidad de que exactamente 2 estén oxidadas?
50. En la sala de espera de un Hospital se encuentran 200 personas. Se les toma una muestra de sangre y 70 son de tipo O-. Si se eligen 15 personas al azar, ¿Cuál es la probabilidad de que 9 tengan O-?
51. De un conjunto de 100 microcontroladores, cada uno se considera aceptable si presenta como máximo cinco defectos. El método de evaluación

- del lote consiste en seleccionar al azar 4 microcontroladores y rechazar el lote si alguno de ellos es defectuoso. ¿Qué probabilidad existe de que, en la muestra, exactamente uno de los microcontroladores seleccionados sea defectuoso, suponiendo que en todo el lote hay cinco defectuosos?
52. Un viajero intenta pasar por la aduana escondiendo 10 paquetes de droga dentro de una maleta que también contiene 20 paquetes de ropa. En el aeropuerto, un oficial de control de aduanas selecciona aleatoriamente cinco paquetes de la maleta para analizarlos. Según la normativa vigente en el aeropuerto, el viajero solo será arrestado si en el control aparecen dos o más paquetes que contengan droga. ¿Cuál es la probabilidad de que el viajero no sea arrestado por tráfico de drogas?
 53. El Megapopular tiene 75 bombillos navideños, de los cuales 20 son rojos. Si se extraen aleatoriamente y sin sustitución 50 bombillos. ¿Cuál es la probabilidad de que cinco de ellos sean rojos?
 54. El hospital del IESS de Quito recibe un promedio de 40 llamadas de emergencia por hora. Encuentra la probabilidad de que reciban exactamente 15 llamadas en una hora.
 55. En una fábrica de balones, se producen en promedio 5 balones defectuosos por cada 100 balones fabricados. ¿Cuál es la probabilidad de encontrar 3 defectos en una muestra de 6 productos?
 56. El promedio de accidentes de tránsito en el cruce las avenidas Bolivariana y Víctor Hugo en la ciudad de Ambato es de 5 por mes. Calcula la probabilidad de que haya 3 accidentes en el próximo mes.
 57. El centro de Atención Ciudadana de Latacunga recibe en promedio 20 llamadas por hora. ¿Cuál es la probabilidad de que reciban exactamente 5 llamadas en la próxima hora?
 58. En un restaurante “Entre Fuegos”, el número promedio de clientes que pide postre de chocolate es de 6 por hora. Encuentra la probabilidad de que exactamente 10 clientes pidan postre en la próxima hora.
 59. BioImágenes, laboratorio médico, recibe en promedio 20 muestras de orina para análisis por día. Calcula la probabilidad de que reciban exactamente 30 muestras en un día dado.
 60. Aseguradora del Sur recibe un promedio de 5 reclamaciones sobre seguros vehiculares por semana. ¿Cuál es la probabilidad de no recibir reclamaciones en una semana?

Ejercicio práctico (con base de datos)

Con la información de la base de datos sobre estudiantes disponible en el siguiente link:

Accede aquí:



Realiza un análisis de distribuciones de probabilidad discretas utilizando el SPSS:

- Calcula la probabilidad de que exactamente 200 estudiantes se gradúen, asumiendo una probabilidad de éxito del 55% en una muestra de 1000 estudiantes seleccionados al azar.
- En un grupo de 1500 estudiantes seleccionados del total de estudiantes, se tiene que 300 son becados, calcula la probabilidad de que al menos 150 tengan beca.
- Calcula la probabilidad de que el primer abandono ocurra en el segundo semestre, con una probabilidad de abandono del 25% por semestre.
- Calcula la probabilidad de que un estudiante apruebe exactamente 4 unidades en un semestre, asumiendo una tasa media de 6 aprobaciones por semestre.

5.7.2. Distribución de probabilidades continuas

Variable aleatoria continua

Una variable aleatoria continua es la que puede tomar cualquier valor dentro de un rango continuo (por ejemplo, la altura, el peso o el tiempo) (Lind et al., 2012).



A diferencia de las variables discretas, no se asigna probabilidad a valores individuales, sino, se calcula la probabilidad de que la variable caiga dentro de un intervalo (Lind et al., 2012).

Función de densidad de probabilidad

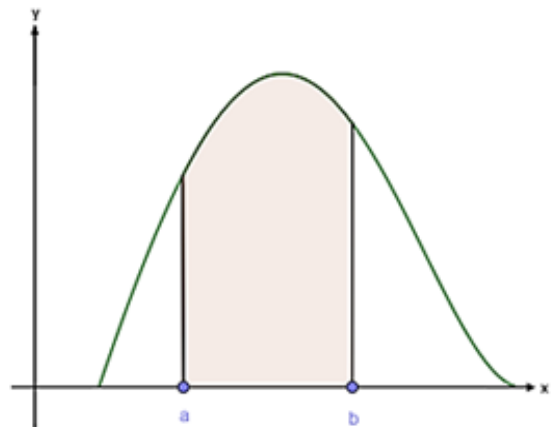
La función de densidad de probabilidad continua, $f(x)$ para un intervalo de valores entre a y b con $a \leq b$, se define como:

$$P(a \leq x \leq b) = \int_a^b f(x)dx$$

La probabilidad de que X asuma un valor en el intervalo $[a, b]$ es el área bajo la curva entre este intervalo como se observa en la figura.

Figura 46

Representación de la función de densidad de probabilidad continua



Una función de densidad de probabilidad, debe satisfacer dos condiciones:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

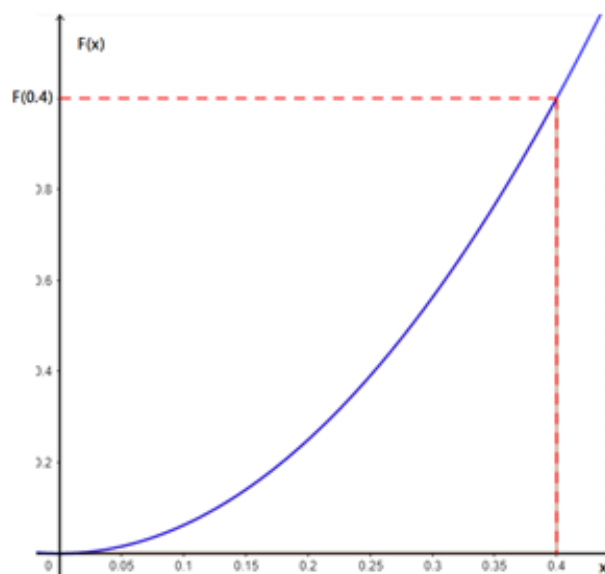
Función de distribución acumulada

La función de distribución acumulada de una variable aleatoria continua X se define como:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y)dy$$

Figura 47

Gráfica función de distribución acumulada



Media o esperanza matemática

De una variable aleatoria x es un número real igual a la integral del producto del dato por su respectiva función de densidad de probabilidad, entre los límites de la función:

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

Donde:

x es la variable aleatoria

$f(x)$ función de densidad de probabilidad

Varianza

Es un número no negativo igual a: (Galindo, 2011)

$$Var(x) = \int_{-\infty}^{\infty} [x - E(x)]^2 * f(x) dx$$

$$Var(x) = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(x))^2$$

Desviación típica

Es la raíz de la varianza:

$$\sigma = \sqrt{Var(x)}$$

Ejemplo

El tiempo de espera de un pasajero en una parada de autobús para dirigirse a su casa, en minutos, está dado por la siguiente función de densidad de probabilidad:

$$f(x) = \begin{cases} \frac{1}{25} + \frac{8}{125}x, & 0 \leq x \leq 5 \\ 0, & \text{en cualquier otro caso} \end{cases}$$

- Determine que corresponde a una función de densidad continua
- Cuál es la probabilidad de que el pasajero espera en la parada de autobús entre 1 y 3 minutos.
- Calcule la función de distribución acumulada
- Calcular la media, la varianza y la desviación típica

1. *Prepare los datos:*

- Identificar los datos:

$$f(x) = \frac{1}{25} + \frac{8}{125}x$$

$$a = 0$$

$$b = 5$$

2. *Resolución y análisis:*

- Para el cálculo del literal a), debe cumplir con las dos condiciones para que sea una función de densidad de probabilidad:

$$1. f(x) \geq 0$$

En este caso si cumple con la primera condición, ya que la función $f(x) = \frac{1}{25} + \frac{8}{125}x$ esta entre 0 y 5

$$2. \int_{-\infty}^{\infty} f(x)dx = 1$$

Para esta segunda condición, se debe evaluar la integral entre 0 y 5 y debe cumplirse la igualdad:

$$\int_0^5 \left(\frac{1}{25} + \frac{8}{125}x \right) dx = 1$$

$$\left(\frac{1}{25}x + \frac{8x^2}{250} \right) \Big|_0^5 = 1$$

$$\frac{1(5)}{25} + \frac{8(5)^2}{250} - \frac{1(0)}{25} - \frac{8(0)^2}{250} = 1$$

$$\frac{5}{25} + \frac{200}{250} = 1$$

$$1 = 1$$

Se comprobó que cumple con la segunda condición

Por lo tanto, si es una función de densidad de probabilidad continua.

b. Para el cálculo del literal b)

$$P(a \leq x \leq b) = \int_a^b f(x) dx$$

$$P(1 \leq x \leq 3) = \int_1^3 \left(\frac{1}{25} + \frac{8}{125}x \right) dx = \left(\frac{1}{25}x + \frac{8x^2}{250} \right) \Big|_1^3$$

$$P(1 \leq x \leq 3) = \frac{1(3)}{25} + \frac{8(3)^2}{250} - \frac{1(1)}{25} - \frac{8(1)^2}{250} = \frac{3}{25} + \frac{72}{250} - \frac{1}{25} - \frac{8}{250}$$

$$P(1 \leq x \leq 3) = 0,336$$

c. Para el literal c):

$$F(x) = \int_{-\infty}^x f(y) dy$$

$$F(x) = \int_0^x \left(\frac{1}{25} + \frac{8}{125}y \right) dy = \left(\frac{1}{25}y + \frac{8y^2}{250} \right) \Big|_0^x = \frac{1(y)}{25} + \frac{8(y)^2}{250} - \frac{1(0)}{25} - \frac{8(0)^2}{250}$$

$$F(x) = \frac{x}{25} + \frac{8x^2}{250}$$

d. Para el literal d):

Cálculo de la esperanza matemática:

$$\blacktriangle E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

$$E(x) = \int_0^5 x * \left(\frac{1}{25} + \frac{8}{125} x \right) dx = \frac{1}{25} \int_0^5 x dx + \frac{8}{125} \int_0^5 x^2 dx$$

$$E(x) = \left(\frac{1}{25} * \frac{x^2}{2} + \frac{8}{125} * \frac{x^3}{3} \right) \Big|_0^5 = \frac{1}{25} * \frac{5^2}{2} + \frac{8}{125} * \frac{5^3}{3} - \frac{1}{25} * \frac{0^2}{2} - \frac{8}{125} * \frac{0^3}{3}$$

$$E(x) = \frac{1}{2} + \frac{8}{3} = \frac{19}{6}$$

$$E(x) = 3,17$$

Cálculo de la varianza y desviación típica:

$$Var(x) = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(x))^2$$

$$Var(x) = \int_0^5 x^2 \left(\frac{1}{25} + \frac{8}{125} x \right) dx - \left(\frac{19}{6} \right)^2$$

$$Var(x) = \frac{1}{25} \int_0^5 x^2 dx + \frac{8}{125} \int_0^5 x^3 dx - \frac{361}{36}$$

$$Var(x) = \left(\frac{1}{25} * \frac{x^3}{3} + \frac{8}{125} * \frac{x^4}{4} \right) \Big|_0^5 - \frac{361}{36} = \frac{1}{25} * \frac{5^3}{3} + \frac{8}{125} * \frac{5^4}{4} - \frac{1}{25} * \frac{0^3}{3} - \frac{8}{125} * \frac{0^4}{4} - \frac{361}{36}$$

$$Var(x) = \frac{5}{3} + 10 - \frac{361}{36} = \frac{59}{36}$$

$$Var(x) = 1,6389$$

Cálculo de la desviación típica:

$$\sigma = \sqrt{Var(x)}$$

$$\sigma = \sqrt{1,6389} = 1,28$$

Distribución uniforme

Todos los valores dentro de un intervalo $[a, b]$ son igualmente probables de ser elegidos; se utiliza cuando no hay una preferencia por ningún valor dentro del intervalo. (Ross, 2019)

Características de la distribución uniforme

- Probabilidad constante, es decir que cada resultado dentro del intervalo definido tiene la misma posibilidad de ser seleccionado.

- La distribución es simétrica, es decir no tiene sesgos hacia un lado u otro del rango.
- No tiene moda única, porque todos los valores tienen la misma probabilidad
- La media de una distribución uniforme es el punto medio del intervalo, y la varianza se calcula a partir de la longitud del intervalo.

Función de densidad de probabilidad

De una variable aleatoria continua X con distribución uniforme está dada por: (Lind et al., 2012)

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{si no} \end{cases}$$

Donde:

- a es el límite inferior del intervalo.
- b es el límite superior del intervalo.

Función de distribución acumulada

Es la probabilidad acumulada hasta un punto x , y está definida como: (Lind et al., 2012).

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$$

Media de una distribución uniforme:

La media, $E(X)$, se encuentra en el centro del intervalo:

$$E(x) = \frac{a+b}{2}$$

Desviación estándar de una distribución uniforme

La desviación estándar (σ) mide la dispersión de los valores y está en función de los extremos del intervalo:

$$\sigma = \sqrt{\frac{(a+b)^2}{12}}$$

Para poder calcular la probabilidad de que X que sigue una distribución uniforme entre un subintervalo $[c,d]$ se calcula como el área bajo la curva entre esos puntos:

$$P(c \leq x \leq d) = \int_c^d f(x)dx = \frac{d-c}{b-a}$$

Ejemplo práctico N°1

La temperatura de la Mitad del Mundo varía uniformemente entre 20°C y 32°C durante el día. ¿Cuál es la probabilidad de que la temperatura de la Mitad del Mundo esté entre 25°C y 30°C?

1. *Prepare los datos:*

- $a=20^\circ\text{C}$ límite inferior del intervalo
- $b=35^\circ\text{C}$ límite superior del intervalo
- $c=25^\circ\text{C}$ límite inferior del subintervalo
- $d=30^\circ\text{C}$ límite superior del subintervalo

2. *Análisis:*

a. Encontrar la función de densidad de probabilidad:

$$f(x) = \frac{1}{b-a} = \frac{1}{35-20} = \frac{1}{15}$$

b. Calcular la probabilidad de que la temperatura este entre 25°C y 30°C:

$$P(25 \leq x \leq 30) = \int_{25}^{30} \frac{1}{15} dx = \frac{x}{15} \Big|_{25}^{30} = \frac{30-25}{15} = \frac{5}{15} = \frac{1}{3} = 0,333$$

Construcción en SPSS:

1. *Preparación de los datos*

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** a
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** Límite inferior del intervalo
 - **Medida:** Escala
- **Nombre:** b
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** Límite superior del intervalo
 - **Medida:** Escala
- **Nombre:** c
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** Límite inferior del subintervalo
 - **Medida:** Escala
- **Nombre:** d
 - **Tipo:** Numérico
 - **Decimales:** cuatro
 - **Etiqueta:** Límite superior del subintervalo
 - **Medida:** Escala

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda											
	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	a	Numérico	8	2	Límite inferior del intervalo	Ninguno	Ninguno	4	Centro	Escala	Entrada
2	b	Numérico	8	2	Límite superior del intervalo	Ninguno	Ninguno	4	Centro	Escala	Entrada
3	c	Numérico	8	2	Límite inferior del subintervalo	Ninguno	Ninguno	4	Centro	Escala	Entrada
4	d	Numérico	8	2	Límite superior del subintervalo	Ninguno	Ninguno	4	Centro	Escala	Entrada

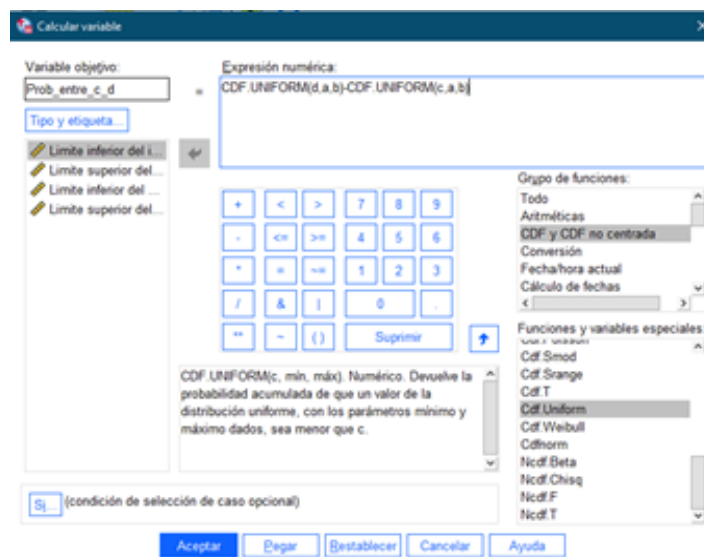
- c. En la pestaña “Vista de datos” ingrese los valores de las variables dato, límite superior y límite inferior.

Archivo Editar Ver Datos Transforn				
4 :				
	a	b	c	d
1	20,00	35,00	25,00	30,00

2. Análisis probabilístico:

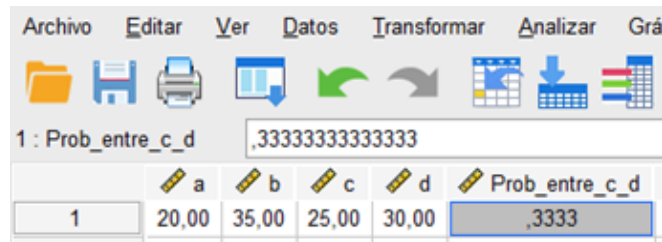
En el menú principal seleccione: Transformar / Calcular variable y se despliega otra ventana, en la cual se completa lo siguientes:

- Nombre de la variable
- En grupo de funciones elegimos **CDF y CDF no centrada**
- En la sección de Funciones y variables especiales seleccionamos **Cdf: Uniform**
- En la sección expresión numérica aparece la función Uniforme, y completamos los parámetros de las variables **a, b, c y d**; cuyas variables se encuentran en la sección Tipo y etiqueta. En la expresión numérica se realiza una diferencia de la probabilidad uniforme para el límite superior del subintervalo menos el límite inferior del subintervalo.
- Finalmente, clic en **Aceptar**



3. Resultados obtenidos:

En la Ventana de Resultados, encontrará el resultado de la probabilidad a encontrar.



	a	b	c	d	Prob_entre_c_d
1	20,00	35,00	25,00	30,00	,3333

4. Interpretación:

Se puede concluir, que existe un 33,33% de probabilidad de que la temperatura en la Mitad del Mundo está entre 25°C y 30°C.

Distribución normal

Llamada también distribución Gaussiana, muy importante en la estadística y la probabilidad debido a su aparición en muchos fenómenos naturales y científicos (Ross, 2019).

La distribución normal describe datos que tienden a agruparse alrededor de un promedio, con una mayor frecuencia cerca del centro y una menor frecuencia a medida que se alejan del centro (Lind et al., 2012).

Función de densidad de probabilidad

De una variable aleatoria X con distribución normal está dada por la siguiente función de distribución de probabilidades:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde:

μ : Media de la distribución ubicada en el centro de la curva

σ : Desviación típica

σ^2 : Varianza de la distribución.

x : Variable aleatoria.

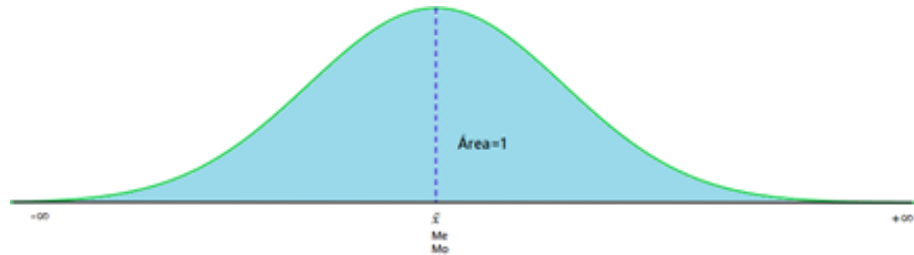
Características de una distribución normal:

- Campana simétrica: La curva es simétrica alrededor de la media, mediana y moda.
- El área total bajo la curva es igual a 1 que representa la probabilidad total.

- Los extremos de la curva, se las conoce como colas de la distribución y se extienden hacia el infinito en direcciones opuestas (Ross, 2019).
- A la distribución normal se la representa como: $N(\mu, \sigma)$

Figura 48

Representación de una curva normal



Distribución normal estándar

Por la dificultad de cálculo de la distribución normal y para hacer uso de tabla de distribución, se utiliza la estandarización de la distribución normal. Es decir, se convierte una distribución normal con cualquier media y desviación típica $N(\mu, \sigma)$, a una distribución normal estándar, de media de 0 (cero) y una desviación típica de 1 $N(0,1)$.

Función de densidad de probabilidad

De la distribución normal estándar está dada por:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Donde:

x es el valor estándar o puntuación estándar.

Se utiliza la distribución normal estándar por facilidad de cálculo, ya que se puede usar una única tabla para calcular probabilidades, sin importar las características originales de la distribución de datos. Además, la estandarización permite utilizar una misma escala para comparar datos de diferentes distribuciones. Finalmente, estandarizar reduce la complejidad de trabajar con diferentes parámetros de media y desviación estándar (Ross, 2019).

Cualquier distribución de probabilidad normal puede convertirse en una distribución de probabilidad normal estándar si se resta la media de cada observación y se divide esta diferencia entre la desviación estándar, como se indica (Lind et al., 2012):

$$z = \frac{x - \mu}{\sigma}$$

Donde:

z: valor estándar o tipificado

x: variable aleatoria original

μ : Media de la distribución

σ : desviación estándar de la distribución

Esta transformación convierte a la variable aleatoria (x) a una escala universal donde:

- Un z positivo indica que X está por encima de la media.
- Un z negativo indica que X está por debajo de la media.
- El valor de z representa el área bajo la curva a la izquierda del valor z, como se observa en la figura siguiente.

Figura 49

Representación de una curva normal estándar



Con la estandarización se puede hacer uso de tablas Normalizadas para facilitar el cálculo de probabilidades.

Ejemplo práctico N°2

En la fábrica de azúcar Valdez, una máquina empaquetadora llena las bolsas pequeñas de azúcar con un peso promedio de 1kg y desviación estándar 0,025kg. ¿Cuál es la probabilidad de que una bolsa pese menos de 0,95kg?

1. Prepare los datos:

$$\mu=1\text{kg}$$

$$\sigma=0,025\text{kg}$$

$$x=0,95\text{kg}$$

2. Análisis:

- a. Estandarizar la variable aleatoria x al valor tipificado z

$$z = \frac{x - \mu}{\sigma} = \frac{0,95 - 1}{0,025} = \frac{-0,05}{0,025}$$

$$z = -2$$

- b. Se plantea la probabilidad solicitada, se realiza una operación adicional ya que la tabla de valores z solo da la parte izquierda del valor límite z .

$$P(x \leq 0,95) = P(z \leq -2)$$

$$P(x \leq 0,95) = P(z > 2)$$

$$P(x \leq 0,95) = 1 - P(z \leq 2)$$

- c. Se busca en la tabla el valor de $z=2,00$, para buscarlo en la tabla, las filas representan el valor de z y el primer decimal, mientras que las columnas es el segundo decimal, para el ejemplo fila 2,0 y columna ,00 se tiene:

1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

- d. Por lo tanto, la probabilidad o área bajo la curva es la intersección de los valores, lo ubicamos para los cálculos respectivos:

$$P(x \leq 0,95) = 1 - 0,9772$$

$$P(x \leq 0,95) = 0,0228$$

3. Interpretación:

En conclusión, la probabilidad de que la bolsa pequeña de azúcar pese menos de 0,95Kg es de 2,28%

Construcción en SPSS:

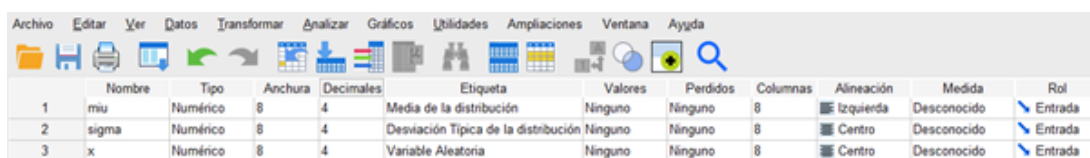
1. Preparación de los datos

- a. Abra SPSS y cree un nuevo archivo de datos
- b. En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** miu
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** media de la variable
- **Medida:** Escala

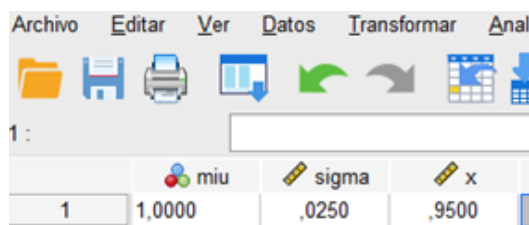
- **Nombre:** sigma
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** desviación típica
- **Medida:** Escala

- **Nombre:** x
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** variable aleatoria
- **Medida:** Escala



	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	miu	Numérico	8	4	Media de la distribución	Ninguno	Ninguno	8	Izquierda	Desconocido	Entrada
2	sigma	Numérico	8	4	Desviación Típica de la distribución	Ninguno	Ninguno	8	Centro	Desconocido	Entrada
3	x	Numérico	8	4	Variable Aleatoria	Ninguno	Ninguno	8	Centro	Desconocido	Entrada

- c. En la pestaña “Vista de datos” ingrese los valores de las variables miu, sigma y x:



	miu	sigma	x
1	1,0000	,0250	,9500

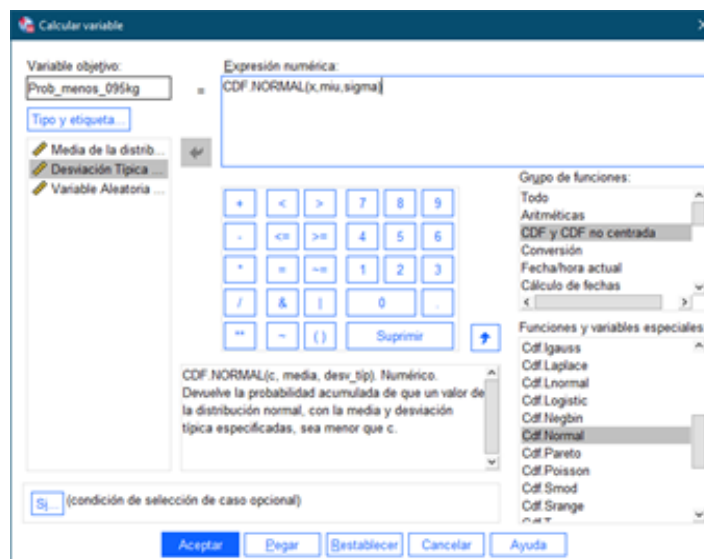
2. Análisis probabilístico:

En el menú principal seleccione: Transformar / Calcular variable y de despliega otra ventana, en la cual se completa lo siguientes:

- Nombre de la variable
- En grupo de funciones elegimos CDF y CDF no centrada
- En la sección de Funciones y variables especiales seleccionamos Cdf: Normal
- En la sección expresión numérica aparece la función Normal, y completamos los parámetros de las variables x, miu, sigma; cuyas variables se encuentran en la sección Tipo y etiqueta.

La función CDF devuelve la el valor normal menor al valor de comparación x, por lo que no se realiza ninguna operación adicional.

- Finalmente, clic en Aceptar



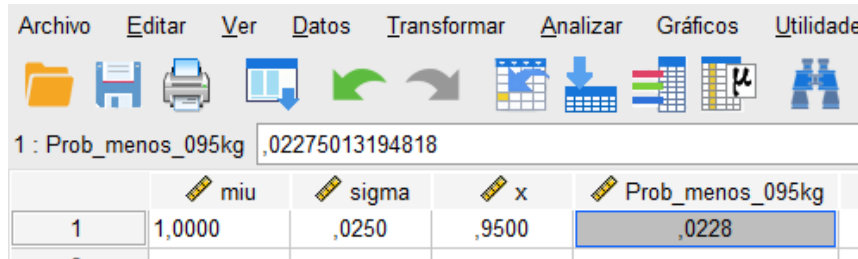
3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la probabilidad a encontrar.

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidade				
1 : Prob_menos_095kg ,02275013194818				
	miu	sigma	x	Prob_menos_095kg
1	1,0000	,0250	,9500	,0228

3. Resultados obtenidos

En la Ventana de Resultados, encontrará el resultado de la probabilidad a encontrar.



	miu	sigma	x	Prob_menos_095kg
1	1,0000	,0250	,9500	,0228

4. Interpretación:

En conclusión, la probabilidad de que la bolsa pequeña de azúcar pese menos de 0,95 kg es de 2,28%

Distribución exponencial

La distribución exponencial se utiliza para modelar tiempos de espera hasta un evento, como tiempos de vida útil o tiempos entre fallos. También puede aplicarse a distancias, volúmenes o intervalos entre eventos. Si un evento ocurre al azar varias veces, los tiempos entre eventos son independientes y siguen una distribución exponencial (Ross, 2019).

Permite identificar el tiempo entre eventos o intentos en una distribución de Poisson, donde los eventos ocurren de manera independiente y a una tasa constante; además mide la probabilidad de que en un determinado tiempo suceda un evento, o el tiempo entre un evento y el siguiente (Lind et al., 2012).

Función de densidad de probabilidad

La función de densidad de probabilidad de la distribución exponencial es:

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Donde:

$\lambda > 0$, inverso de la media

x , variable aleatoria continua

Función de distribución acumulativa

La función de distribución acumulativa es:

$$F(x, \lambda) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

La función de distribución acumulada es también la probabilidad de que un valor de X sea menor que x .

Para calcular la probabilidad de X sea mayor que x se utiliza: $P(X > x) = e^{(-\lambda x)}$

Finalmente, para calcular la probabilidad que X se encuentre entre dos valores a y b se utiliza: $P(a < X < b) = e^{(-\lambda a)} - e^{(-\lambda b)}$

Media y desviación típica de una distribución exponencial

$$\text{Media: } \mu = \frac{1}{\lambda}$$

$$\text{Desviación típica: } \sigma = \sqrt{\frac{1}{\lambda^2}}$$

Ejemplo práctico N°3

Supongamos que los tiempos de espera entre llamadas en el centro de atención telefónica de Claro ubicado en el Mall de los Andes, siguen una distribución exponencial con una tasa promedio de 3 llamadas por minuto.

- a. ¿Qué probabilidad existe que el próximo cliente llame en menos de 35 segundos?
- b. ¿Qué probabilidad existe que el tiempo de espera sea mayor a un minuto y medio?

1. Prepare los datos:

$$\lambda = 3 \text{ llamadas}$$

$$x_1 = 35 \text{ segundos} = 0,583 \text{ minutos}$$

$$x_2 = 1,5 \text{ minutos}$$

2. Análisis:

- a. Para el literal a), se calcula la probabilidad de que existe que el próximo cliente llame en menos de 35 segundos.

$$F(x, \lambda) = 1 - e^{-\lambda x}$$

$$F(0.583, 3) = 1 - e^{-3 \cdot 0.583} = 0,826$$

- b. Para el literal b), se encuentra la probabilidad que el tiempo de espera sea mayor a un minuto y medio.

$$F(x > 1,5) = 1 - P(x < 1,5) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x} = e^{-3 \cdot 1,5} =$$

$$F(x > 1,5) = 0,011$$

3. Interpretación:

La probabilidad de que el próximo cliente llame en menos de 35 segundos es 82,6%

La probabilidad de que el tiempo de espera sea mayor a un 1,5 minuto es 1,1%

Construcción en SPSS:

1. Preparación de los datos

- Abra SPSS y cree un nuevo archivo de datos
- En la ventana “Vista de variables” cree las siguientes variables:

- **Nombre:** lambda
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** Tasa Promedio
- **Medida:** Escala
- **Nombre:** x1
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** Variable aleatoria literal a
- **Medida:** Escala

- **Nombre:** x2
- **Tipo:** Numérico
- **Decimales:** cuatro
- **Etiqueta:** Variable aleatoria literal b
- **Medida:** Escala

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	lambda	Númerico	8	2	Tasa promedio	Ninguno	Ninguno	8	Centro	Escala	Entrada
2	x1	Númerico	8	4	Variable aleatoria a	Ninguno	Ninguno	8	Centro	Escala	Entrada
3	x2	Númerico	8	4	Variable aleatoria b	Ninguno	Ninguno	8	Centro	Escala	Entrada

- c. En la ventana “Vista de datos” ingrese los valores de las variables lambda, x1 y x2:

	lambda	x1	x2
1	3,00	35,0000	1,5000

2. *Análisis probabilístico:*

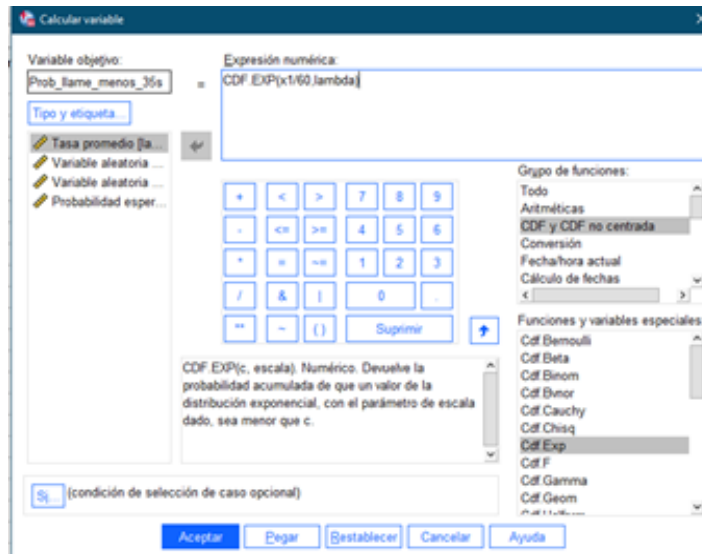
Para el literal a)

En el menú principal seleccione: **Transformar / Calcular variable** y de despliega otra ventana, en la cual se completa lo siguientes:

- Nombre de la variable
- En grupo de funciones elegimos **CDF y CDF no centrada**
- En la sección de Funciones y variables especiales seleccionamos **Cdf: Exp**
- En la sección expresión numérica aparece la función exponencial, y completamos los parámetros de las variables **x1 y lambda**; cuyas variables se encuentran en la sección Tipo y etiqueta.

X1 se divide para 60 ya que el cálculo se debe realizar en las mismas unidades, en este caso en minutos.

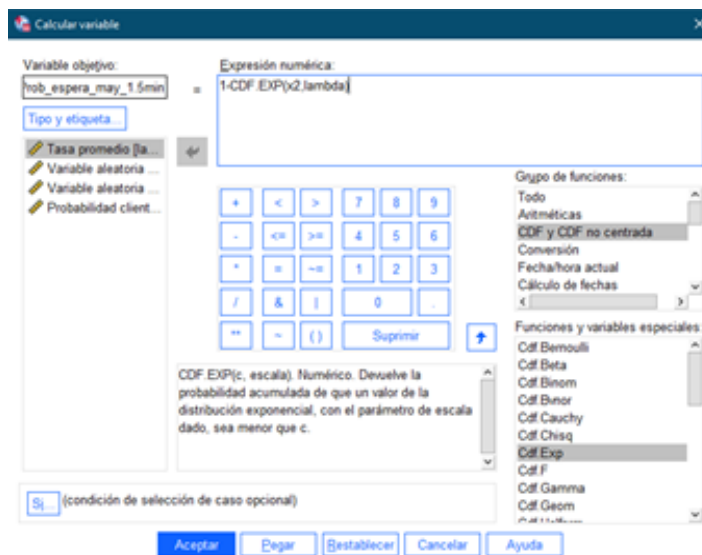
e. Finalmente, clic en **Aceptar**



Para el literal b)

En el menú principal seleccione: Transformar / Calcular variable y de despliega otra ventana, en la cual se completa lo siguientes:

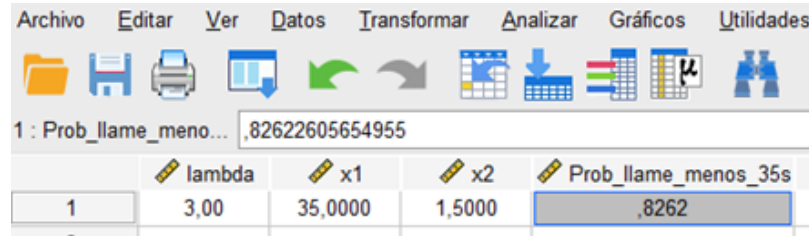
- Nombre de la variable
- En grupo de funciones elegimos CDF y CDF no centrada
- En la sección de Funciones y variables especiales seleccionamos Cdf: Exp
- En la sección expresión numérica aparece la función exponencial, y completamos los parámetros de las variables **x2** y **lambda**; cuyas variables se encuentran en la sección Tipo y etiqueta.
- Finalmente, clic en **Aceptar**



3. Resultados obtenidos

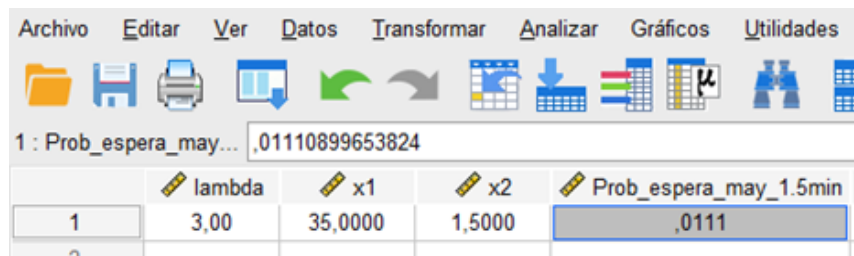
El resultado de la variable encontrada se muestra en la vista de datos

a. Para el literal a)



	lambda	x1	x2	Prob_llame_menos_35s
1	3,00	35,0000	1,5000	,8262

b. Para el literal b)



	lambda	x1	x2	Prob_espera_may_1.5min
1	3,00	35,0000	1,5000	,0111

4. Interpretación

La probabilidad de que el próximo cliente llame en menos de 35 segundos es 82,6%.

La probabilidad de que el tiempo de espera sea mayor a un 1,5 minuto es 1,1%

Ejercicio práctico (con base de datos)

Utilizando la base de datos proporcionada sobre el consumo de energía eléctrica en un hogar (household_power_consumption.txt) disponible en el enlace:

Accede aquí:



Realiza un análisis estadístico para determinar:

- Si los valores de voltaje están distribuidos uniformemente entre un valor mínimo y un valor máximo, calcular la probabilidad de que el voltaje esté entre 220 V y 245 V.

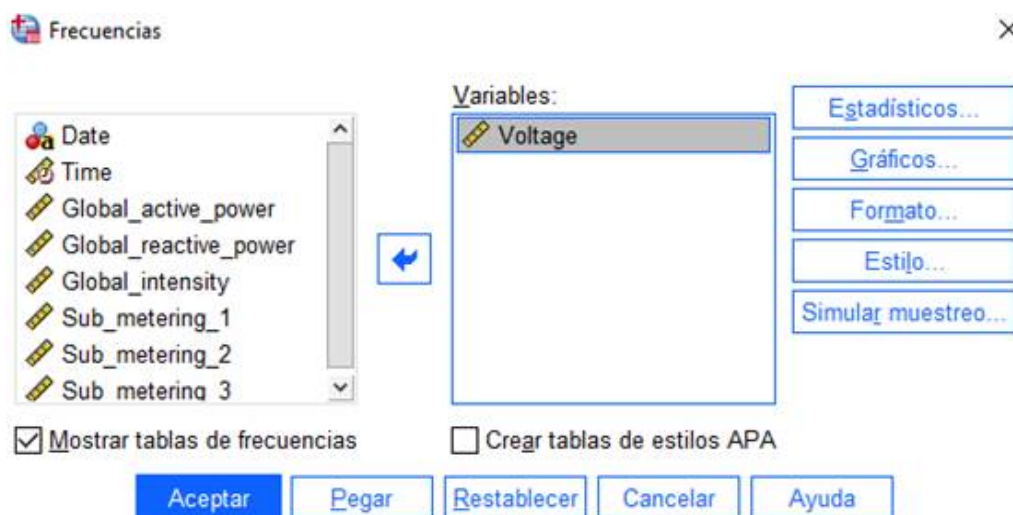
- b. Se supone que la variable Potencia activa (Global_active_power) sigue una distribución normal con media μ y desviación estándar σ calculadas a partir de la información encontrada en la base de datos, determina la probabilidad de que la potencia activa sea mayor a 5 kW.
- c. Se supone que el tiempo entre consumos significativos sigue una distribución exponencial con parámetro $\lambda=1/\text{media}(\text{Sub_metering_1})$. Encontrar la probabilidad que el consumo sea menor que 18 Wh.

Solución:

Cargar la base de datos household_power_consumption.txt en el spss.

a) Suponer que los valores de voltaje están distribuidos uniformemente entre un valor mínimo y un valor máximo, calcular la probabilidad de que el voltaje esté entre 230V y 240V.

1. Cargar la base de datos household_power_consumption.txt en el SPSS.
2. Encontrar los valores de a y b que son los voltajes mínimos y máximos del consumo de energía; para ello se da clic en análisis -> Estadísticos Descriptivos -> Frecuencia selecciona la variable voltaje y clic en aceptar:



3. En la tabla de resultados como esta ordenado elegimos el primer valor que es el mínimo y el ultimo valor que es el máximo:

$$a = 223.200$$

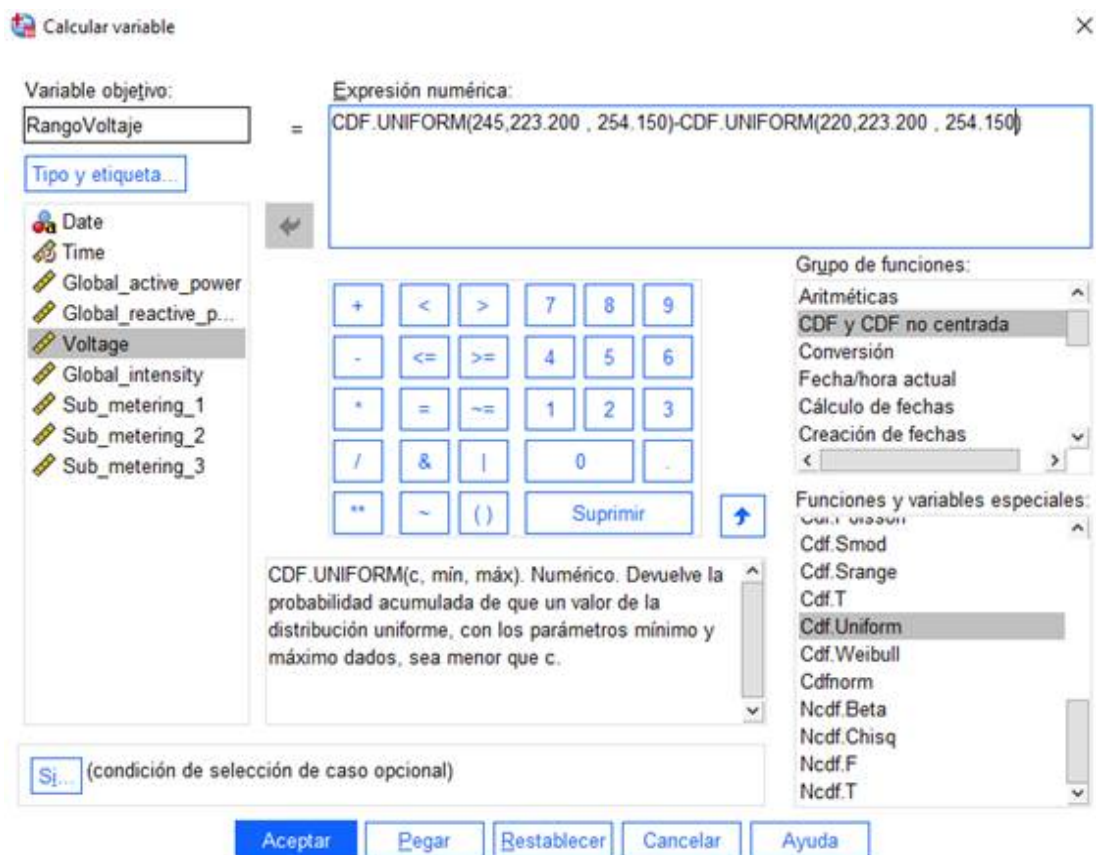
$$b = 254.150$$

4. Se anota el rango de valores de c y d:

$$c = 220$$

$$d = 245$$

5. Se realiza el cálculo en el SPSS, Transformar -> Calcular Variable, se elige la función Cdf. Uniform y se completan los datos necesarios:



6. Finalmente se obtiene el resultado:

RangoVoltaje
70

La probabilidad de que el voltaje este entre 220 y 245 voltios es de 70 %

b) Se supone que la variable Potencia activa (*Global_active_power*) sigue una distribución normal con media μ y desviación estándar σ calculadas a partir de la información encontrada en la base de datos, determina las probabilidades de que la potencia activa sea mayor a 5 kW.

1. Encontrar la media y la desviación típica de la distribución. Clic en Analizar -> Estadísticos descriptivos -> Frecuencias, se elige la variable *Global_active_power*, en estadísticos marcar media y desviación típica, clic en continuar y aceptar:

Frecuencias: Estadísticos

Valores percentiles

☐ Cuantiles

☐ Puntos de corte para: 10 grupos iguales

☐ Percentiles:

Añadir

Cambiar

Eliminar

Tendencia central

☒ Media

☐ Mediana

☐ Modo

☐ Suma

☐ Los valores son puntos medios de grupos

Dispersión

☒ Desv. estándar

☐ Varianza

☐ Rango

☐ Mínimo

☐ Máximo

☐ Error estándar media

Distribución

☐ Asimetría

☐ Curtosis

Continuar Cancelar Ayuda

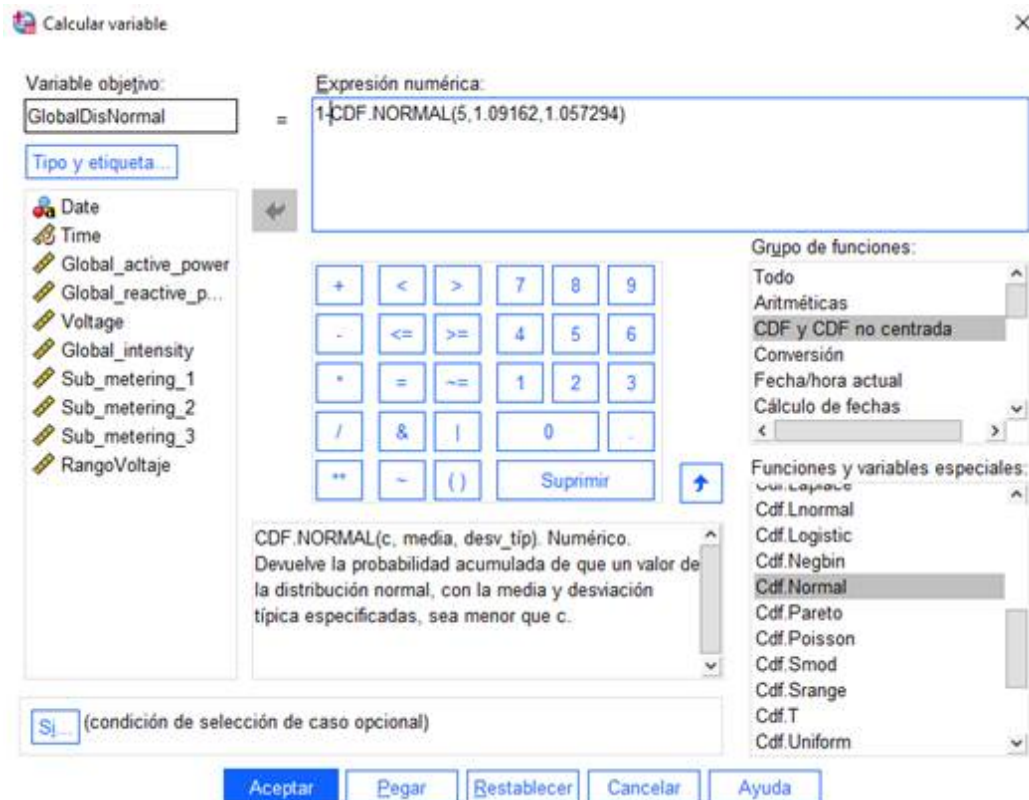
En los resultados se observa los valores calculados:

Estadísticos		
Global_active_power		
N	Válido	2049280
	Perdidos	25979
Media		1,09162
Desv. Desviación		1,057294

$$\mu = 1.09162$$

$$\sigma = 1.057294$$

- Calcular la distribución normal, clic en transformar -> Calcular Variable, seleccionar la función Cdf.Normal, escribir los datos necesarios para el cálculo, como son media, desviación, valor límite y clic en aceptar:



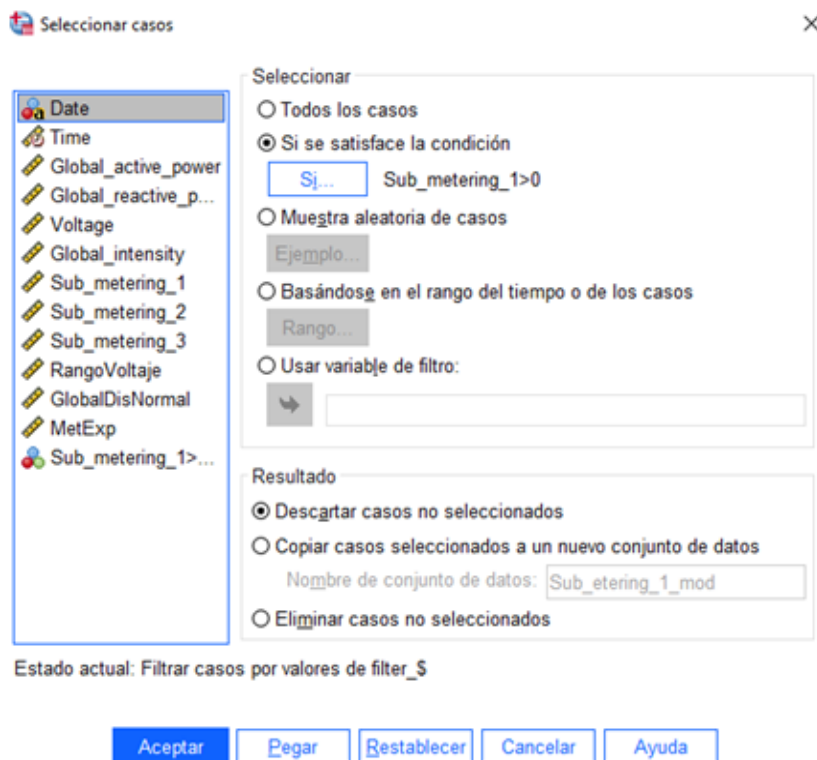
- Se obtiene el valor de la probabilidad normal:

GlobalDisNormal
 ,000109258

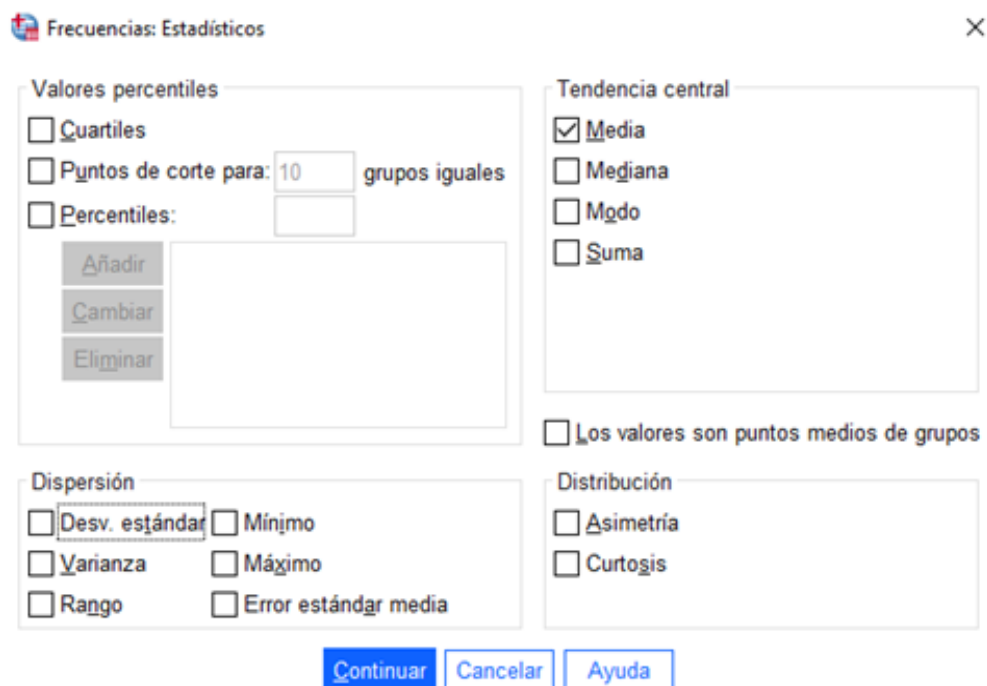
La probabilidad de que la potencia activa sea mayor a 5 kWh es 0.011 % es decir nula.

c) Se supone que el tiempo entre consumos significativos (Sub_metering_1) sigue una distribución exponencial con parámetro λ . Encontrar la probabilidad que el consumo sea menor que 18 Wh.

- Filtrar los valores de la variable mayores a 0, para no tener problemas con el cálculo, por lo que clic en Datos > Seleccionar casos, escribe la condición a cumplir $\text{Sub_metering_1} >, 0$ clic en continuar y aceptar.



2. Una vez filtrado los 0, se calcula la media de Sub_metering para poder así calcular el λ ; para ello clic en Analizar -> Estadísticos descriptivos -> Frecuencias, se elige la variable Sub_metering_1, en estadísticos marcar media, clic en continuar y aceptar:



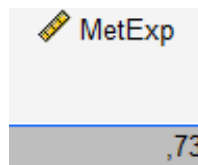
Obteniendo así el valor de la media $\mu=13.5959$

Estadísticos		
Sub_metering_1		
N	Válido	169105
	Perdidos	0
Media		13,59590

3. Con el valor de la media se encuentra el valor de λ :

$$\lambda = \frac{1}{\mu} = \frac{1}{13.5959} = 0.073552$$

4. Aplicar la función para calcular la distribución exponencial, Transformar > Calcular variable, Cdf.Exp y colocamos la información necesaria.
5. Se obtiene el valor deseado:



Por lo tanto, la probabilidad de que el consumo entre tiempos significativos sea menor de 18 Wh es 73 %.

Autoevaluación 5.5.

- Describe las características de una distribución uniforme.
- La distribución normal es fundamental en estadística. Explica el papel de sus dos parámetros, μ y σ .
- La distribución exponencial modela el tiempo entre eventos. ¿Qué relación tiene con la distribución de Poisson?
- Las estaturas de un grupo de personas siguen una distribución aproximadamente normal con una media de 170 cm y desviación estándar de 8 cm.
 - ¿Qué porcentaje de personas tienen una estatura entre 160 cm y 180 cm?
 - Se selecciona una persona al azar, ¿cuál es la probabilidad de que mida más de 180 cm?

Ejercicios propuestos

61. La cantidad de tiempo que un cliente espera en una fila está uniformemente distribuida entre 3 y 7 minutos. ¿Cuál es la probabilidad de que espere entre 4 y 5,5 minutos?
62. El tiempo de carga de una batería sigue una distribución uniforme continua entre 1 y 3 horas. ¿Cuál es la probabilidad de que tarde más de 2 horas y media?
63. Una máquina expendedora entrega un producto en un tiempo uniformemente distribuido entre 1 y 5 segundos. ¿Cuál es la probabilidad de que tarde entre 1 y 2 segundos?
64. El tiempo de espera para un autobús sigue una distribución uniforme entre 0 y 15 minutos. ¿Cuál es la probabilidad de esperar entre 5 y 10 minutos?
65. Alex un empleado de Plasticaucho tarda entre 10 y 30 minutos en completar una tarea diaria. ¿El jefe inmediato desea saber la probabilidad que Alex tarde en realizar la tarea asignada en menos de 15 minutos?
66. En la pasteurizadora Quito, el tiempo de pasteurizar la leche entera sigue una distribución uniforme que tarda entre 20 y 40 minutos. ¿Se desea conocer la probabilidad que la leche entera tarde entre 25 y 36 minutos en pasteurizarse?
67. La preparación de un almuerzo ejecutivo en el restaurant Ramada sigue una distribución uniforme y tarda aproximadamente entre 25 y 40 minutos. ¿Llega un cliente al restaurante en una hora determinada, cuál será la probabilidad que reciba su almuerzo en menos de 30 minutos?
68. Los tiempos de entrega de los paquetes de Servientrega siguen una distribución normal con una media de 7 días y una desviación estándar de 2 días para entregas interprovinciales. ¿Un cliente desea saber la probabilidad de que su paquete llegue entre 5 y 9 días?
69. El peso del grupo de estudiantes mujeres de una Institución Educativa sigue una distribución normal, de media 74 kg y desviación típica 4 kg, determinar:
 - La probabilidad de que el peso de las estudiantes mujeres esté comprendido entre los 70 y 80 kg.
 - ¿Cuál es la probabilidad de que pese más de 80 kg?

70. Las estaturas de las mujeres en el Ecuador en los últimos años siguen una distribución normal con media 165 cm y desviación estándar 4 cm.
- ¿Si se selecciona una mujer al azar, cual es la probabilidad que mida más de 173 cm?
 - Un dato adicional es que el 7% de las mujeres se consideran más altas, ¿a partir de qué altura se encuentra este porcentaje de mujeres?
71. Sylvania empresa de fabricación de luminarias led, en sus especificaciones dice que el tiempo de vida de una luminaria led sigue una distribución normal con media 38000 horas y desviación estándar 10000 horas.
- ¿Cuál es la probabilidad de que una luminaria led dure menos de 20000 horas?
 - ¿Si los fabricantes de luminarias led desean que solo el 1% de las luminarias fallen antes del tiempo medio de vida, ¿Cuál debe ser la vida útil mínima de la luminaria Led, para garantizar lo que los fabricantes desean?
72. Los tarros de café de Productos Minerva Cia. Ltda. siguen una distribución normal con un peso promedio de 1 kg, y una desviación estándar de 150 g.
- ¿Qué porcentaje de los tarros de café pesa entre 900 g y 1,5 Kg?
 - Si se tiene un 15% de la producción, ¿cuál es el peso máximo de esos tarros de café?
73. PsicoActiva realiza un test de IQ (coeficiente intelectual) a un grupo de niños de una escuela rural, los mismos se distribuyen normalmente con media 100 y desviación estándar 12.
- ¿Se desea conocer el porcentaje de niños de esa escuela que tienen un IQ mayor a 120?
 - ¿Cuál es el rango de IQ donde se encuentra el 75% de la población?
74. El tiempo de espera de un usuario del cajero automático de una Entidad Bancaria sigue una distribución exponencial, el cual tiene una tasa de servicio de 0,4 clientes por minuto.
- ¿Calcule la probabilidad de que un usuario espere más de 8 minutos?
 - ¿Qué probabilidad tiene un usuario de esperar entre 3 y 8 minutos?

- ¿Qué tiempo promedio espera un usuario?
75. El tiempo duración en promedio de un Arduino Nano según el fabricante sigue una distribución exponencial con una media de 75 horas con una batería de 1200 mAh.
- ¿Cuál es la probabilidad de que un Arduino Nano se apague antes de las 60 horas?
 - ¿Cuál es la probabilidad de que un Arduino Nano dure más de 80 horas?
76. El departamento de TICs de la Universidad ha descubierto que el servidor de internet presenta fallas con una media de 600 horas y de manera exponencial.
- ¿Cuál es la tasa de fallos del servidor de internet?
 - ¿Qué probabilidad tiene el servidor de fallar antes de las 500 horas?
77. Los estudiantes de la Espe esperan al autobús para trasladarse al campus Belisario Quevedo en promedio unos 20 minutos entre un bus y otro.
- ¿Qué tasa de llegadas tienen los autobuses?
 - ¿Cuál es la probabilidad de que el próximo autobús llegue en menos de 15 minutos?
 - ¿Cuál es la probabilidad de que el próximo autobús se tarde más de 30 minutos?
78. Camila llama a su madre para contarle cómo van sus estudios en Argentina; el tiempo que tardan las llamadas siguen una distribución exponencial con una tasa de 0,2 llamadas por minuto.
- ¿Qué probabilidad existe en que la duración de una llamada sea mayor de 20 minutos?
 - ¿Cuál es el tiempo promedio de llamadas entre Camila y su madre?
79. El volcán Cotopaxi es un volcán activo y las emisiones de ceniza en el 2023 siguieron una distribución exponencial con una media de una emisión de ceniza al día.
- ¿Cuál es la tasa de emisiones de ceniza?
 - ¿Cuál es la probabilidad de que el volcán Cotopaxi no emita ceniza en más de una semana?

80. La tasa de ocurrencia de accidentes de tránsito en el Ecuador de enero a octubre del 2024 es de 0,0183 accidentes por día.

- ¿Cuál es la probabilidad de que ocurra un accidente de tránsito en el Ecuador en menos de 8 días?
- ¿Cuál es la probabilidad de que pasen más de 20 días sin un accidente de tránsito?

Ejercicio propuesto (con base de datos)

Utilizar la base de datos ObesityDataSet_raw_and_data_synthetic.csv disponible en el siguiente link:



Calcular las siguientes probabilidades:

- Se conoce que los pesos corporales de las personas siguen una distribución uniforme distribuida entre un máximo y un mínimo, calcular la probabilidad que el peso corporal de las personas:
 - Este entre 65 y 82 Kg.
 - Sea mayor a 95 Kg.
 - Sea menor a 50 Kg.
- Se conoce que las alturas de las personas siguen una distribución normal, con media μ y desviación estándar σ , calculadas con la información de la base de datos; calcular la probabilidad que la altura:
 - Este entre 1.50 y 1.75 Kg.
 - Sea mayor a 1.80 Kg.
 - Sea menor a 1.65 Kg.
- La edad de las personas se distribuye exponencialmente con una tasa $\lambda=1/\mu$, calcular la probabilidad que la edad de las personas:
 - Este entre 25 y 55 Kg.
 - Sea mayor a 45 Kg.
 - Sea menor a 30 Kg.

ANEXOS

Respuestas a los ejercicios
propuestos

Capítulo II

1. a.

Nº <i>hermanos</i>	Frecuencia absoluta	Frecuencia relativa	Frecuencia absoluta acumulada	Frecuencia relativa acumulada
0	3	7.5	3	7.5
1	11	27.5	14	35.0
2	10	25.0	24	60.0
3	10	25.0	34	85.0
4	4	10.0	38	95.0
5	2	5.0	40	100.0
Total	40	100.0		

b. 3 estudiantes

c. 52.5%

d. 8 estudiantes

2. a.

<i>Diámetro del perno roscable (mm)</i>				
	Frecuencia absoluta	Frecuencia Relativa (%)	Frecuencia absoluta acumulada	Porcentaje acumulado (%)
Válido 5.0	5	25.0	5	25
5.1	5	25.0	10	50.0
5.2	5	25.0	15	75.0
5.3	3	15.0	18	90.0
5.4	2	10.0	20	100.0
Total	20	100.0		

b. 5.0, 5.1 y 5.2 mm

c. 75%

3. a.

<i>Emisiones de CO2 de un motor diésel</i>				
	Frecuencia absoluta	Frecuencia relativa (%)	Frecuencia absoluta acumulada	Frecuencia relativa acumulado
Válido 120	3	20.0	3	20.0
125	4	26.7	7	46.7
130	4	26.7	11	73.3
135	3	20.0	14	93.3
140	1	6.7	15	100.0
Total	15			

b. 125 g/km y 130 g/km

c. 73.3%

4. a.

Error de precisión de brazos robóticos

		Frecuencia absoluta	Frecuencia Relativa (%)	Frecuencia absoluta acumulada	Frecuencia Relativa acumulada (%)
Válido	0.5	4	33.3	4	33.3
	0.6	3	25.0	7	58.3
	0.7	3	25.0	10	83.3
	0.8	1	8.3	11	91.7
	0.9	1	8.3	12	100.0
	Total	12	100.0		

b. 33.3%.

c. 83.3%

d. 1 brazo robótico

e. 7

5. a.

Gastos mensuales de una empresa

		Frecuencia absoluta	Frecuencia relativa (%)	Frecuencia absoluta acumulada	Frecuencia relativa acumulada (%)
Válido	1200	3	30.0	3	30.0
	1300	2	20.0	5	50.0
	1500	3	30.0	8	80.0
	1600	1	10.0	9	90.0
	1800	1	10.0	10	100.0
	Total	10	100.0		

b. 1200 dólares y 1500 dólares

c. 80%

d. 1800 dólares

e. 80%

6. a.

Edad (en años completos) de sus empleados (Agrupada)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	<= 22	3	6.0	6.0	6.0
	23 - 29	7	14.0	14.0	20.0
	30 - 36	7	14.0	14.0	34.0
	37 - 43	17	34.0	34.0	68.0
	44 - 50	9	18.0	18.0	86.0
	51 - 57	4	8.0	8.0	94.0
	58 - 64	3	6.0	6.0	100.0
	Total	50	100.0	100.0	

- b. 17 empleados
- c. 32%
- d. 0.34
- e. 13 empleados

7. a.

		Consumo de combustible (Agrupada)			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	<= 11.7	4	13.3	13.3	13.3
	11.8 - 12.5	5	16.7	16.7	30.0
	12.6 - 13.3	5	16.7	16.7	46.7
	13.4 - 14.1	6	20.0	20.0	66.7
	14.2 - 14.9	4	13.3	13.3	80.0
	15.0 - 15.7	5	16.7	16.7	96.7
	15.8+	1	3.3	3.3	100.0
	Total	30	100.0	100.0	

- b. 20%
- c. 14 vehículos
- d. 16 vehículos

8. a.

		inversión anual, en miles de dólares (Agrupada)			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	<= 17	4	10.0	10.0	10.0
	18 - 23	7	17.5	17.5	27.5
	24 - 29	14	35.0	35.0	62.5
	30 - 35	9	22.5	22.5	85.0
	36 - 41	4	10.0	10.0	95.0
	42 - 47	2	5.0	5.0	100.0
	Total	40	100.0	100.0	

- b. clase 3
- c. 37.5%
- d. 27.5%

9. a.

Tiempo (min)	Frecuencia Absoluta	Frecuencia Relativa
0 – 3	32	0.2
3 – 6	48	0.30
6 – 9	56	0.35
9 – 12	8	0.05
12 – 15	16	0.10
Total	160	1

b. 160

c. 6 – 9 minutos

d. 50%

e. 24 personas

10. a.

Gastos operativos mensuales (en dólares) (Agrupada)

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	<= 237	9	15.0	15.0	15.0
	238 - 356	7	11.7	11.7	26.7
	357 - 475	10	16.7	16.7	43.3
	476 - 594	10	16.7	16.7	60.0
	595 - 713	9	15.0	15.0	75.0
	714 - 832	8	13.3	13.3	88.3
	833 - 951	7	11.7	11.7	100.0
	Total	60	100.0	100.0	

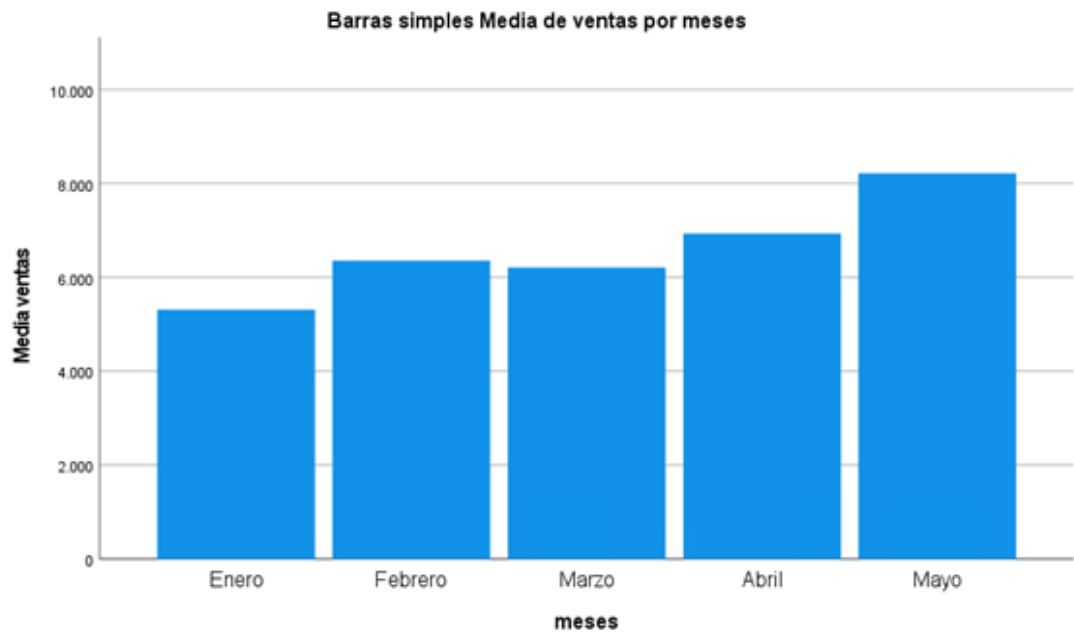
b. USD356.5 – USD475.5 y USD475.5 – USD594.5

c. 7 empresas

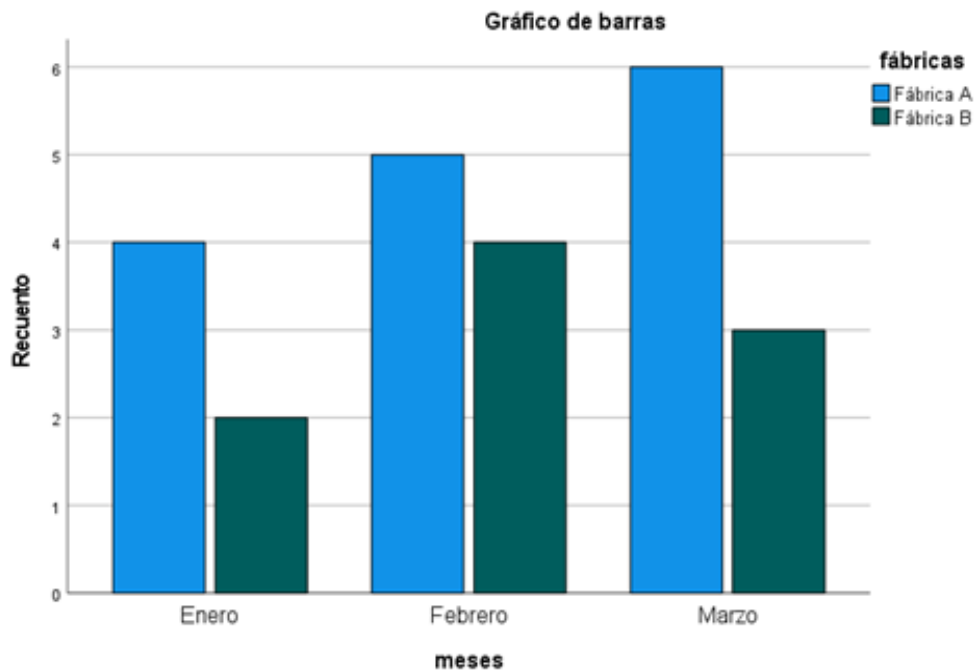
d. 36 empresas

Capítulo III

1.

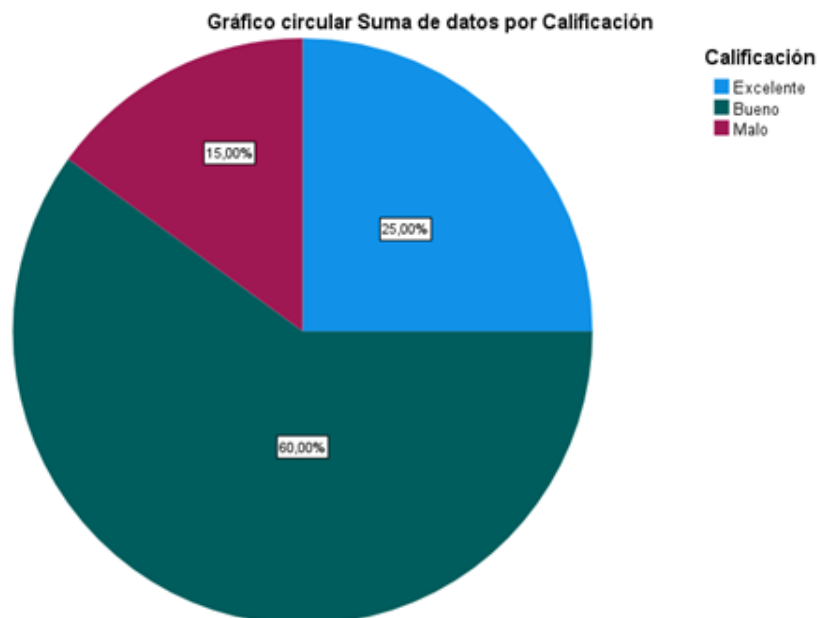


2.

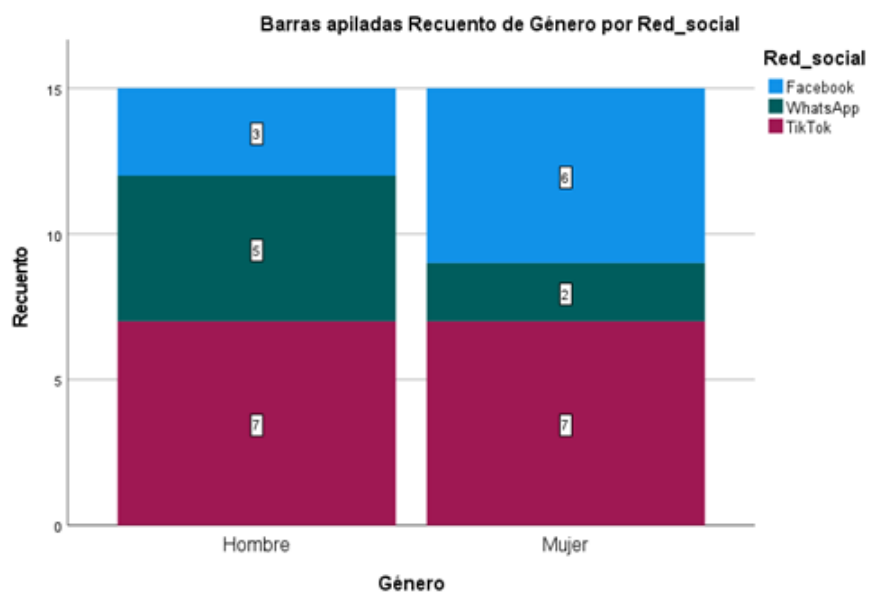


3. 45

4.



5.

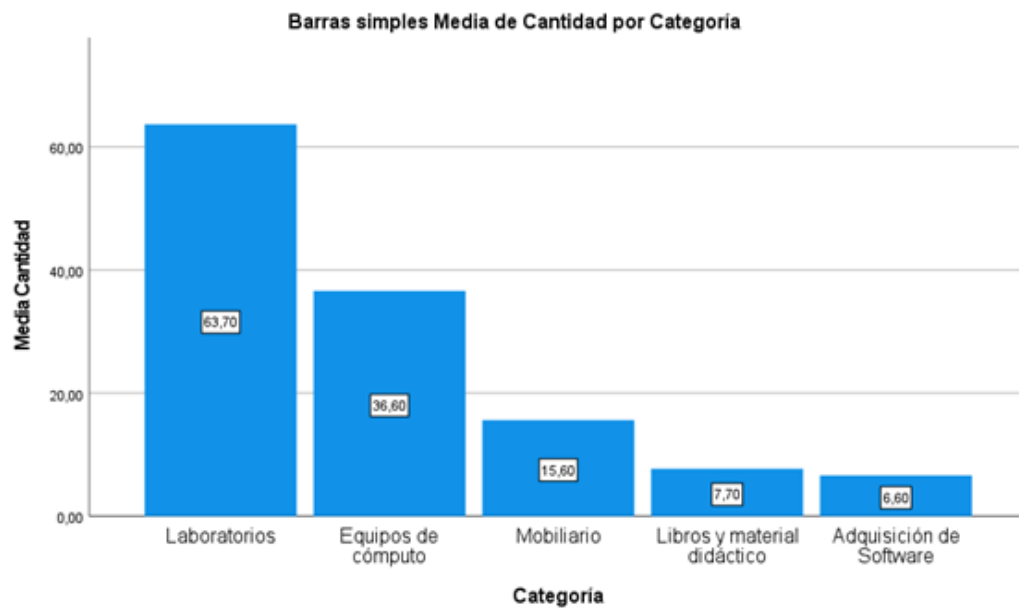


TikTok es la red social más usada en ambos géneros (7 hombres y 7 mujeres).

WhatsApp es la red social más utilizada por los hombres (7) que por las mujeres (2)

Después del TikTok la red social más utilizada por las mujeres es el Facebook (6) respecto a los hombres (3)

6. a.



b. 0.33

c. 0.77

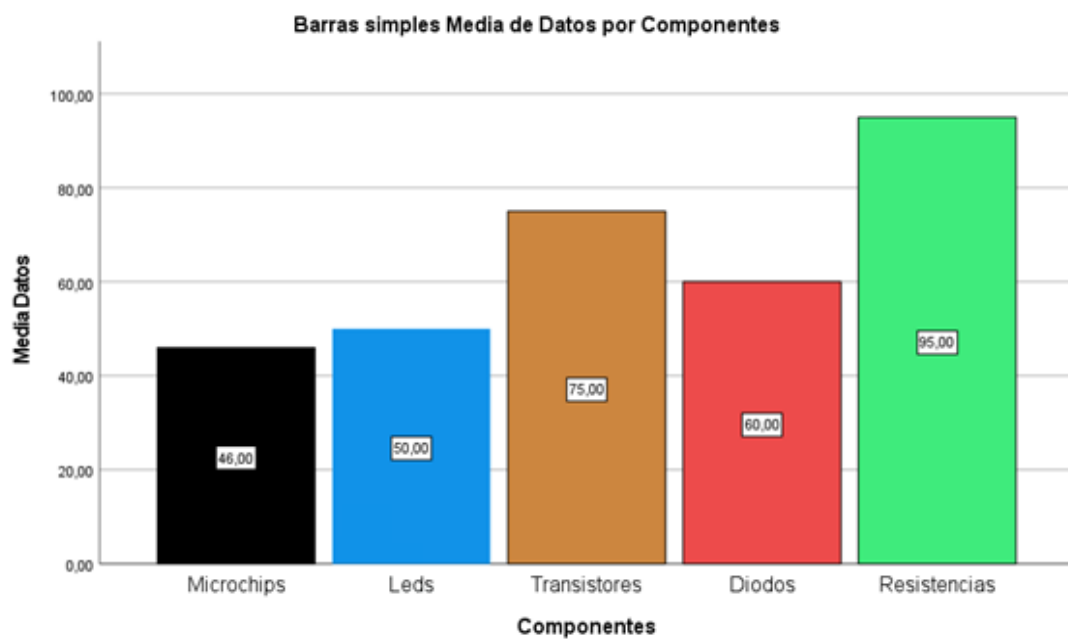
7.

a. Si: 216°

b. No: 108°

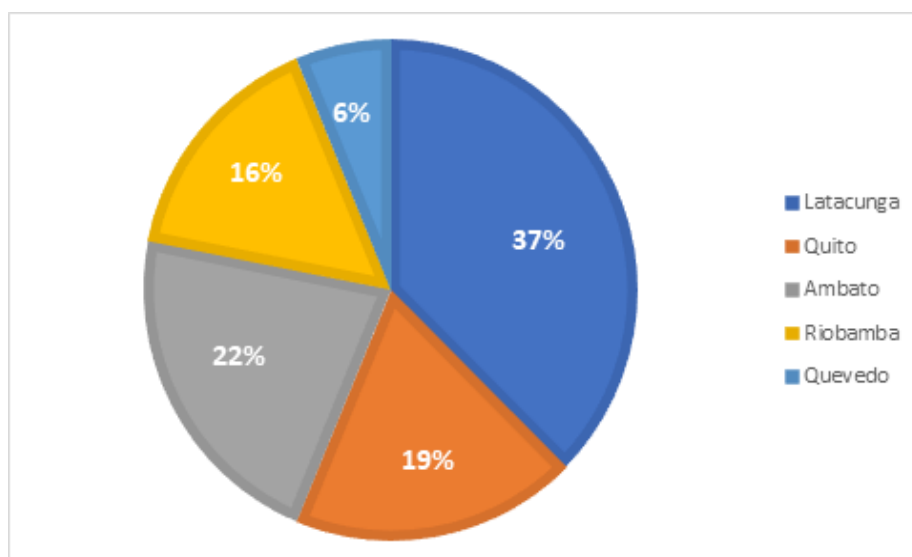
c. No contestan: 36°

8.



9. a. Java
b. HTML

10.



11.



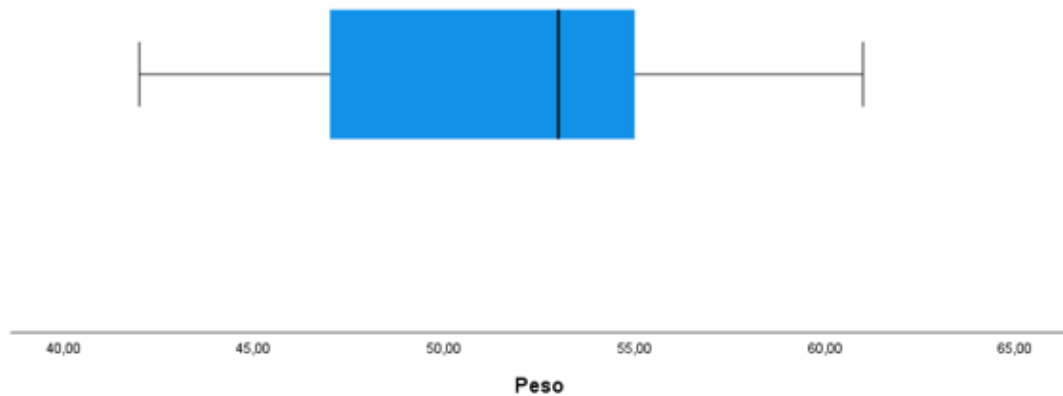
12. a. 30
b. 5 (0-5)
c. 8

13. 26

14. Cálculos previos:

$V_{\min}$	42
Q_1	47
Q_2	53
Q_3	55
$V_{\max}$	61

Diagrama de caja 1-D de Peso



No existen la presencia de valores atípicos evidentes.

15.

Análisis Hombres:

Cálculos previos:

$V_{\min}$	1100
Q_1	1400
Q_2	1500
Q_3	1700
$V_{\max}$	3000

Valores atípicos:

Li=950

Ls=2150

Existe la presencia de valores atípicos, son los valores mayores a 2150 como son 2500 (2 personas) y un valor muy extremo que es el 3000 (1 persona)

Análisis Mujeres:

Cálculos previos:

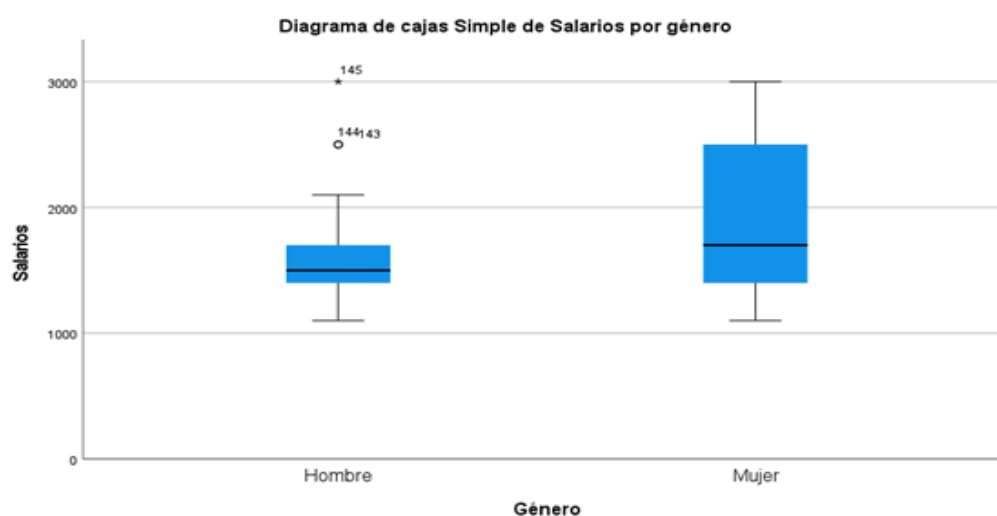
$V_{\min}$	1100
Q_1	1400
Q_2	1700
Q_3	2500
$V_{\max}$	3000

Valores atípicos

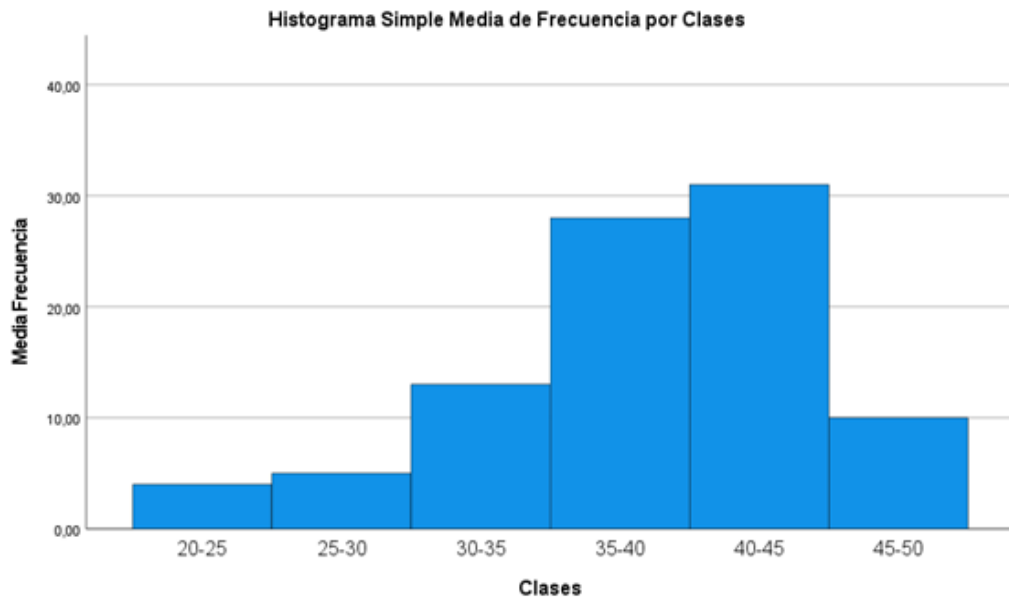
$Li=-250$

$Ls=4150$

No existe la presencia de valores atípicos



16.



17. 46.15%

18. a. 3620.50

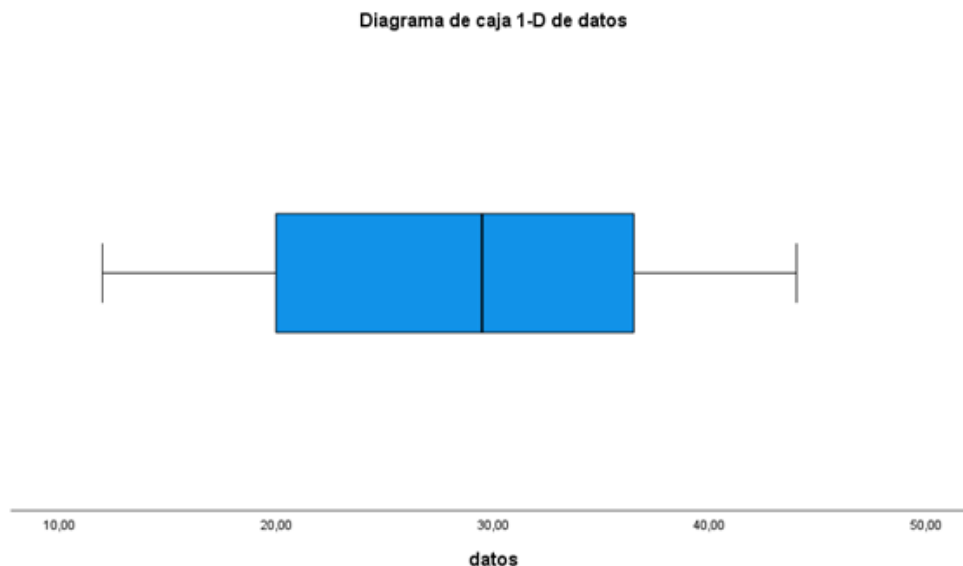
b. $Q1=3067.25$ y $Q3=4699.50$

c. 1632.25

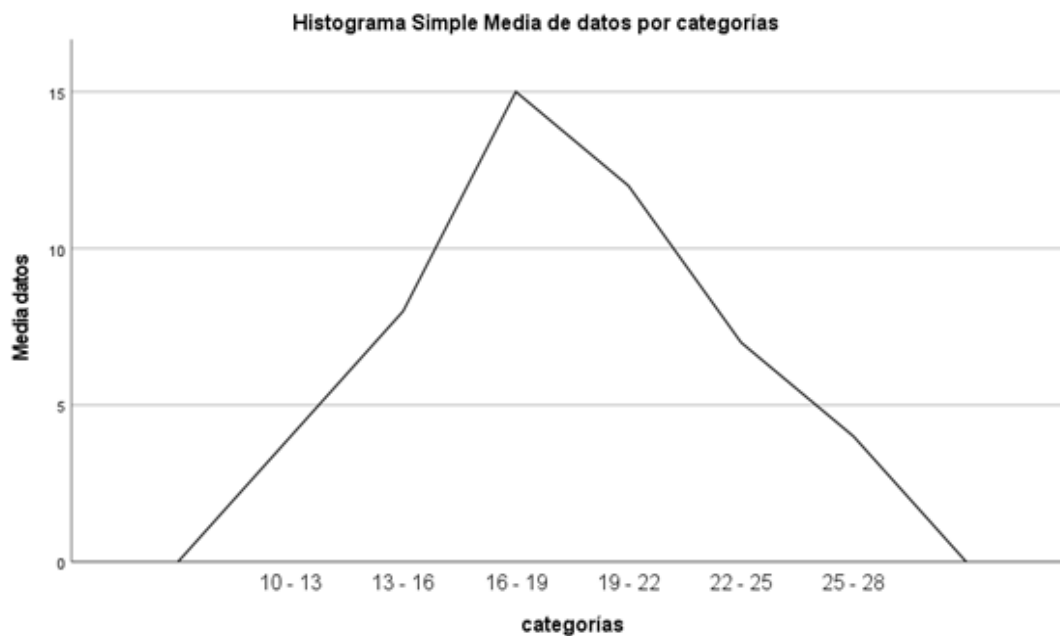
d. Menos de 2253.00 o más de 7200.00. Si existe un valor atípico

e. La distribución es asimétrica con sesgo positivo

19. La distribución tiene un ligero sesgo a la izquierda



20.



Capítulo IV

1. Media=20.25
Mediana=21
Moda=22
2. Media=14.11
Mediana=14.13
Moda=11.12
3. 476.92
4. a. Media=16.89
Mediana=17
Moda=15

b. Enfocarse en aquellos que obtuvieron notas de 15 debido a que esta nota está por debajo de la media y la mediana.
5. a. La media se incrementa en 5 unidades
b. La mediana se mantiene inalterable
6. 1196.41 dólares

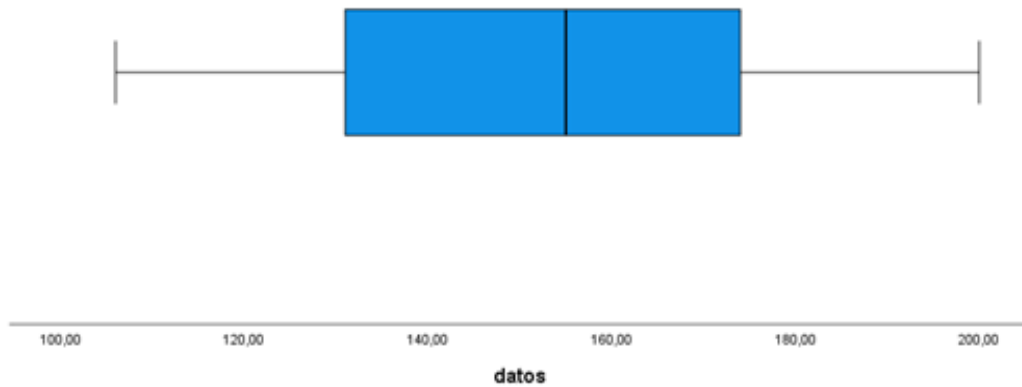
7. a. Media=153.44

Mediana=155

Moda=117

b.

Diagrama de caja 1-D de datos



c.

Clase	Intervalo	Frecuencia
1	[104 – 117]	9
2	[118 – 131]	4
3	[132 – 145]	6
4	[146 – 159]	8
5	[160 – 173]	10
6	[174 – 187]	8
7	[188 – 202]	5
		50

d. Media=152.16

Mediana=156.5

Moda=167

8.

La mediana y la moda con USD550 no se ven afectadas por valores extremos con el sueldo del gerente

La media es USD1163 está distorsionada por el sueldo del gerente

9. a. 13.87

b. 13.53

c. 13.16

d. Media armónica es menor que la media geométrica y ésta es menor que la media aritmética

10. a. 12.81

b. moda=13, mediana=12.5

11. mediana=10, Q1=7, Q3=12.75

12. a.

Clase	Intervalo	Frecuencia
1	[9 - 22[	5
2	[22 - 35[	7
3	[35 - 48[	7
4	[48 - 61[	10
5	[61 - 74[	7
6	[74 - 87[	6
7	[87 - 100[	8

b. 16% es 27.57

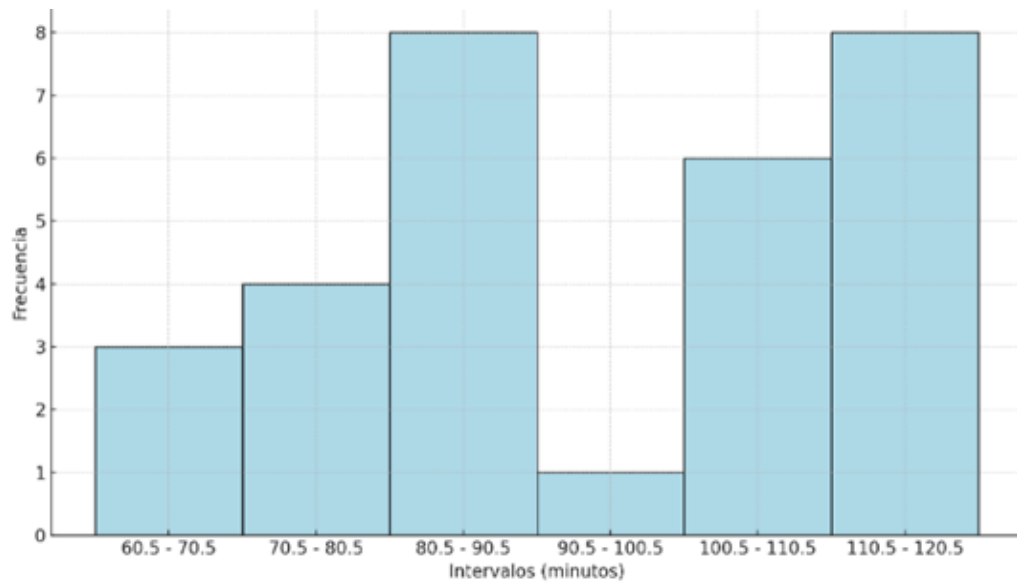
45% es 52.55

85% es 87.81

13. a.

Clase	Intervalos	Frecuencia
1	[60,5 - 70,5[	3
2	[70,5 - 80,5[	4
3	[80,5 - 90,5[	8
4	[90,5 - 100,5[	1
5	[100,5 - 110,5[	6
6	[110,5 - 120,5[	8
		30

b.



c. promedio=94.50

mediana=90.5

d. D3=83

e. P80=113

14. a. P70=16.8

b. P25=13

c. P50=15.33

15. a.

Clases	Frecuencia
[28 – 33[	7
[33 – 38[	9
[38 – 43[	10
[43 – 48[	12
[48 – 53[	6
[53 – 58[	9
[58 – 63[	3

b. Velocidad mínima=30km/h

Velocidad máxima=61km/h

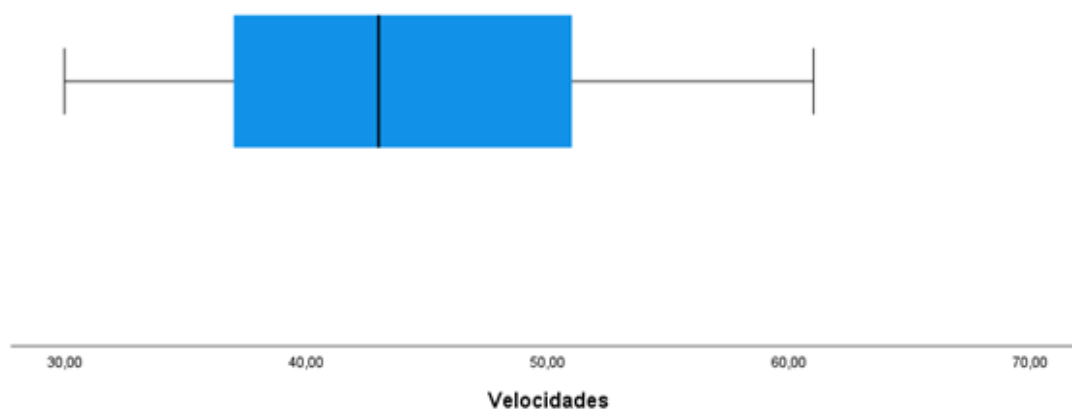
c. 44.07km/h

d. $Q1=36.89$, $Q3=51.33$

e. 14.44

f.

Diagrama de caja 1-D de datos



16. a. 14 puntos ($P40$)

b. 19 puntos ($P80$)

c. 15.5 puntos ($P50$)

d. Deben recibir Tutorías aquellos estudiantes con notas inferiores a 14 puntos y reconocimiento con notas mayores o iguales a 19 puntos

17. a. 34 minutos

b. 20,6 minutos

c. 26.2 minutos

18. 277.78 dólares

19. $Q_1=31.66$

$$Q_2=42.5$$

$$Q_3=50$$

20. $D_7=48.88$

$$P_{30}=32$$

21. a. 18

b. $\sigma^2=27.96$ $\sigma=5.29$

c. 17,51%

22. 2.08

23. 0.013

24. a. 64.16

b. 9.82%

25. $\sigma_Alfa= 0.89$

$$\sigma_Beta=3.74$$

26. $R=4;$

$$\sigma^2=1.90;$$

$$CV=7.55\%$$

27. $\sigma= 1.95;$

$$CV=59.10\%$$

28. $\sigma_{\text{Andres}}=0.13$

$\sigma_{\text{Camila}}=0.33$

$\sigma_{\text{Esteban}}=0.64$

$\sigma_{\text{Maria}}=0.29$

29. $\sigma^2=1.53$

$\sigma=1.24$

30. $CV_{\text{SMD}}=1.58\%$

$CV_{\text{AXC}}=1.90\%$

31. $-0,04$

32. $As=-0.081$

$K=-0.51$

33. 1.11

34. $As=-0.16$

$K=-0.63$

35. $As=-1.42$

$K=1.23$

36. $As=-0.18$

$K=-0.92$

37. $As=-0.13$

$K=-0.77$

38. $A_s = -0,00$

$K = -0,85$

39. 0.77

40. Asimetría negativa

Capítulo V

1. 0.8125

2. 0.7222

3. 0.083

4. 0.1111

5. 0,1111

6. 0.9343

7. 0.934

8. 0.3750

9. 0.375

10. 0.2380

11. 0.15

12. 0.0078

13. 0.0566

14. 0.1667

15. 0.327

16. 0.26

17. 0.23

18. 0.22

19. 0.8

20. 0.6389

21. 0.20

22. 0.7143

23. 0.30

24. 0.96

25. 0.83

26. 0.5556

27. 0.826

28. 0.0121

29. 0.62

30. 0.5845

31. 0.004

32. 0.5947

33. 0.13

34.

a. Cocinas: 12 Franke y 18 Electrolux y Lavadoras: 13 Franke y 7 Electrolux.

b. 0.14

35. 0.6757

36. 0.5631

37. 0.624

38. 0.4326

39. 0.2163

40. 0.5972

41. 0.1058

42. a. 0.2765

b. 0.0467

c. 0.0410

43. 0.3115

44. a. 0.0560

b. 0.9999

- c. 0.0031
- 45. a. 0.0947
 - b. 0.0557
 - c. 0.2245
- 46. 0.9959
- 47. 1
- 48. 0.0315
- 49. 0.2540
- 50. 0.0265
- 51. 0.1765
- 52. 0.4990
- 53. 0,000008623
- 54. 0.000003488
- 55. 0.0012
- 56. 0.1404
- 57. 0.000055
- 58. 0.0413
- 59. 0.0083
- 60. 0.006737
- 61. 0.375
- 62. 0.25
- 63. 0.25
- 64. 0.3333
- 65. 0.25
- 66. 0.55
- 67. 0.3333
- 68. 0.6826
- 69. a. 0.7745
 - b. 0.0668

70. a. 0.0228
b. 170.92 cm
71. a. 0.0359
b. 14700
72. a. 0.7482
b. 844 Kg
73. a. 0.0475
b. entre 86,2 y 113,8
74. a. 0.2019
b. 0.2604
c. 2.5 minutos
75. a. 0.5507
b. 0.3442
76. a. tasa de 0,0017
b. 0.0263
77. a. 0.05
b. 0.5276
c. 0.2231
78. a. 0.0183
b. Promedio de 5 llamadas
79. a. 1
b. 0.0009119
80. a. 0.1362
b. 0.6935

Referencias

- Castañeda, M. B., Cabrera, A. F., Navarro, Y., & de Vries, W. (2010). *Procesamiento de datos y análisis estadísticos utilizando SPSS: Un libro práctico para investigadores y administradores educativos* (1st ed.). EDIPUCRS.
- Galindo, E. (2011). *Estadística metodos y aplicaciones*. Prociencia.
- Galindo, E. (2010). *Problemas y ejercicios de probabilidad y estadística* (3.ª ed.). Prociencia.
- Gorgas, G. J., Cardiel L. N., & Zamorano C. J. (2011). *Estadística básica para estudiantes de ciencias*. Universidad Complutense de Madrid.
- IBM. (2025). Obtención de ayuda. IBM Knowledge Center. <https://www.ibm.com/docs/es/spss-statistics/27.0.0?topic=system-getting-help>
- Jay, D. (2005). *Probabilidad y estadística para ingeniería y ciencias*. Thomson Editores, México.
- Johnson, N. L., Kemp, A. W., & Kotz, S. (2005). *Univariate discrete distributions* (3rd ed). Wiley.
- J. R., Rodríguez, A. I. P., & Jiménez, E. C. R. (2017). *Estadística para administración*. Grupo Editorial Patria.
- Lind, D. A., Marchal, W. G., & Wathen, S. A. (2012). *Statistical techniques in business & economics* (15th ed). McGraw-Hill/Irwin.
- Martínez Bencardino, C. (2012). *Estadística y muestreo* (13a ed.). Bogotá, D.C: Ecoe
- Martín Marín, Jorge, Portillo Poblador, Nuria, Vidal Puig, Santiago, Montesinos Guillot, Amparo, & Villa Juliá, María Fulgencia. (2013). *DISTRIBUCIONES DE PROBABILIDAD. PROBLEMAS Y RECURSOS PRÁCTICOS* (Primera). Universitat Politècnica de València.
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1973). *Introduction to the theory of statistics* (3d ed). McGraw-Hill.
- Probabilidad y Estadística. (n.d.). *Frecuencia estadística. Probabilidad y Estadística*. <https://www.probabilidadyestadistica.net/frecuencia-estadistica/>
- Ochoa Torrez, R. R. (2014). *Análisis estadístico con el SPSS (IBM SPSS Statistics 22)*. La Paz, Bolivia.
- Olvera, B. G. (2013). *Estadística y probabilidad*. Pearson Educación.
- Probabilidad y Estadística básica para Ingenieros – RODRIGUEZ.pdf. (s. f.).
- Quezada Lucio, N. (2014). *Estadística con SPSS 22* (1ª ed.). Macro E.I.R.L.
- Romero Villafranca, R., & Zúnica Ramajo, L. R. (2013). *Métodos estadísticos para ingenieros* (Primera). Universitat Politècnica de Valencia.

- ROSS, S. M. (s. f.). *PROBABILIDAD Y ESTADISTICA PARA INGENIERIA TY CIENCIA*.
- Ross, S. M. (2019). *Introduction to probability models (Twelfth edition)*. Elsevier, Acad. Press. <https://doi.org/10.1016/C2017-0-01324-1>
- Solano, H. L., & Álvarez, C. R. (2005). *Estadística descriptiva y distribuciones de probabilidad*. Ediciones Uninorte. <https://books.google.com.ec/books?id=3Tkb8HJ5toUC>
- Spiegel, M. R., & Stephens, L. J. (2008). *Schaum's outline of theory and problems of statistics* (4. ed). McGraw-Hill.
- Valeria Bizet, Elena Molina Portillo, & José Miguel Contreras. (2023). Situaciones-problemas sobre variable aleatoria y sus aplicaciones en distribuciones de probabilidad según libros de texto chilenos. 27(2), 32. <https://doi.org/10.30827/profesorado.v27i2.21345>
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2012). *Probabilidad y estadística para ingeniería y ciencias* (Novena). Pearson. https://bibliotecavirtualaserena.wordpress.com/wp-content/uploads/2017/05/libro_probabilidad-y-estadistica-para-ingenerc3ada-y-ciencias-ronald-e-walpole-mayers.pdf

Estadística descriptiva con aplicaciones en SPSS

La estadística es una herramienta esencial para comprender la realidad y tomar decisiones acertadas en cualquier ámbito profesional. Este libro ha sido diseñado pensando en estudiantes de ingenierías, licenciaturas y tecnologías que buscan desarrollar una base sólida en el análisis estadístico.

A través de explicaciones claras y un enfoque práctico, se introducen los fundamentos del SPSS —desde su instalación y configuración— junto con los conceptos esenciales de la estadística descriptiva, los gráficos estadísticos, las medidas descriptivas y la teoría de probabilidades.

Con el apoyo de bases de datos reales, el lector encontrará ejercicios resueltos y propuestos que permiten reforzar los conocimientos y aplicar cada tema utilizando el software estadístico SPSS. En un entorno académico y laboral en el que el manejo de información es cada vez más importante, el uso de herramientas informáticas se convierte en un recurso indispensable para optimizar el tiempo, mejorar la eficiencia y garantizar análisis confiables.

Los autores, con amplia experiencia en la enseñanza universitaria, investigación y en el campo de la estadística, han identificado la necesidad de contar con un material claro, didáctico e interactivo que facilite el aprendizaje. Este libro nace con el propósito de guiar al estudiante paso a paso, motivándolo a descubrir la relevancia de la estadística como ciencia fundamental para la toma de decisiones.

ISBN: 978-9942-652-37-9



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA