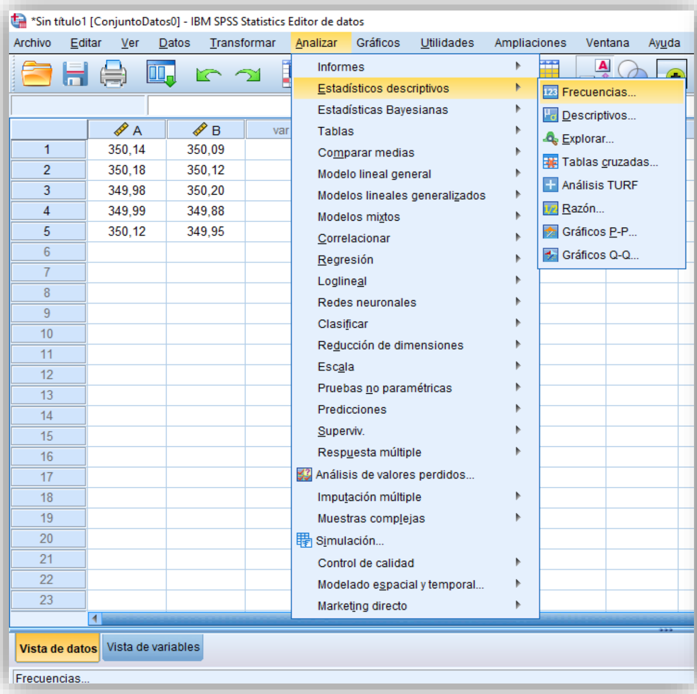
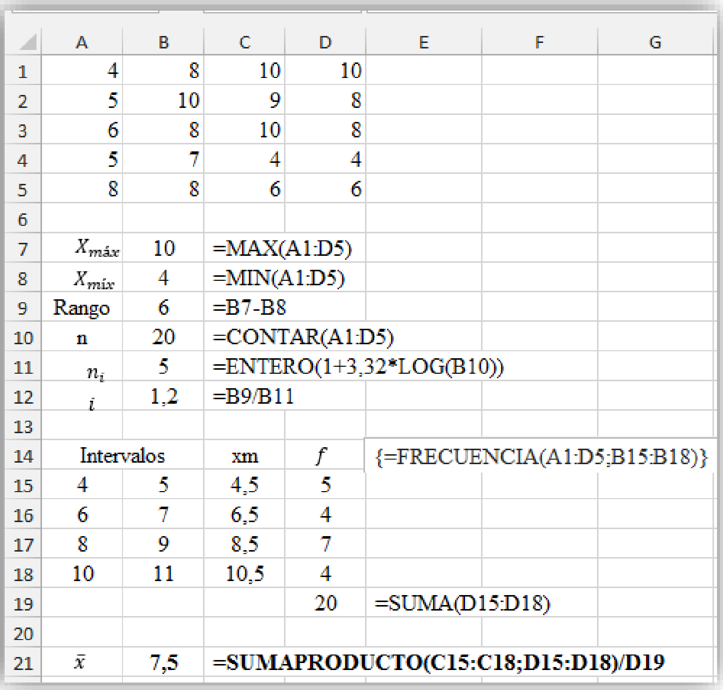
ESTADÍSTICA DESCRIPTIVA PARA TODOS: FUNDAMENTOS Y APLICACIONES - VOLUMEN II





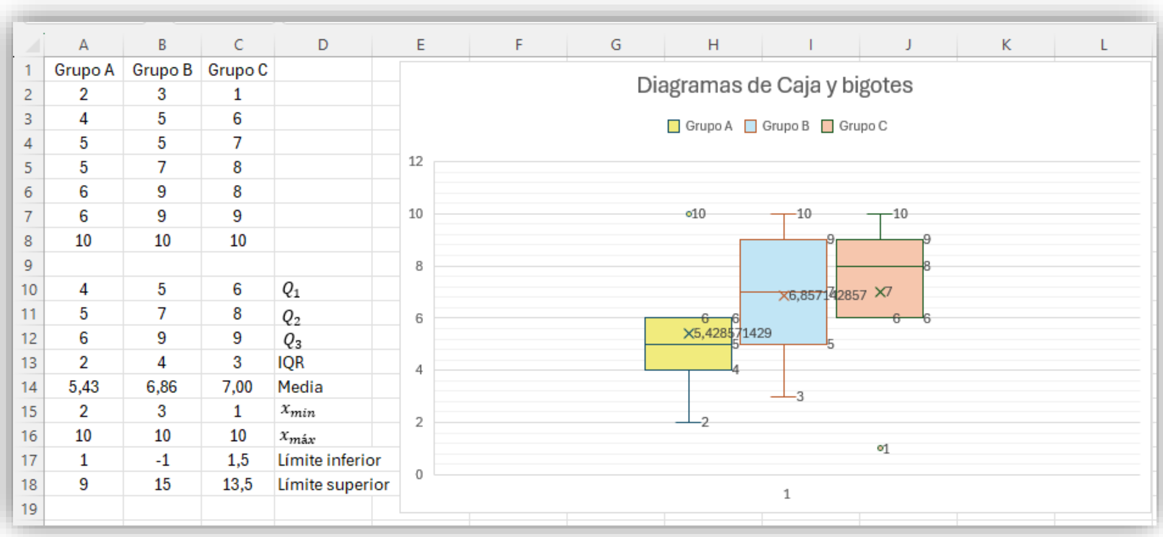
	A	B	C	D	E	F	G
1	4	8	10	10			
2	5	10	9	8			
3	6	8	10	8			
4	5	7	4	4			
5	8	8	6	6			
6							
7	$X_{máx}$	10	=MAX(A1:D5)				
8	$X_{mín}$	4	=MIN(A1:D5)				
9	Rango	6	=B7-B8				
10	n	20	=CONTAR(A1:D5)				
11	n_i	5	=ENTERO(1+3,32*LOG(B10))				
12	i	1,2	=B9/B11				
13							
14	Intervalos		xm	f	{=FRECUENCIA(A1:D5;B15:B18)}		
15	4	5	4,5	5			
16	6	7	6,5	4			
17	8	9	8,5	7			
18	10	11	10,5	4			
19				20	=SUMA(D15:D18)		
20							
21	$\bar{x}$	7,5	=SUMAPRODUCTO(C15:C18;D15:D18)/D19				

Autor

Mario Orlando Suárez Ibujés

IMBABURA - ECUADOR

ESTADÍSTICA DESCRIPTIVA PARA TODOS: FUNDAMENTOS Y APLICACIONES - VOLUMEN II



	A	B	C	D
1	Grupo A	Grupo B	Grupo C	
2	2	3	1	
3	4	5	6	
4	5	5	7	
5	5	7	8	
6	6	9	8	
7	6	9	9	
8	10	10	10	
9				
10	4	5	6	Q_1
11	5	7	8	Q_2
12	6	9	9	Q_3
13	2	4	3	IQR
14	5,43	6,86	7,00	Media
15	2	3	1	x_{min}
16	10	10	10	$x_{máx}$
17	1	-1	1,5	Límite inferior
18	9	15	13,5	Límite superior
19				

IMBABURA – ECUADOR

2025

AUTOR

Mgs. Mario Orlando Suárez Ibujés
Docente ocasional en la Facultad de Posgrado de la UTN
Correos electrónicos: mosuarez@utn.edu.ec; mariosuarezibujes@gmail.com
ORCID: https://orcid.org/0000-0002-3962-5433

PARES REVISORES

Mgs. Liliana Chamorro Hernández
Docente en la Universidad Politécnica Estatal del Carchi
Correo electrónico: lilianam.chamorro@upec.edu.ec

Mgs. Víctor Mario García Mora
Docente en la Escuela Superior Politécnica de Chimborazo
Correo electrónico: victor.garcia@espoch.edu.ec

Mgs. Gandy Patricio Rivadeneira Martínez
Asesor Educativo en la Coordinación Zonal 1- Educación
Correo electrónico: gandy.rivadeneira@educacion.gob.ec

PhD. Cristian Eduardo Terán Montalvo
Funcionario en la Coordinación Zonal 1- Educación
Correo electrónico: cristiane.teran@educacion.gob.ec

ISBN

Primera Edición 2025

Con el Auspicio

ISBN: 978-9942-51-446-2

9 789942 514462

Asesoría Educativa CZ1

COLPASE
Colegio Profesional de Asesores Educativos del Ecuador

Centro Cultural Antonio Ante

Excelencia Educativa
Asociación de maestros

DEDICATORIA

Con amor infinito en expansión,

para mi esposa, Dyana Rivera, el amor de mi vida y de todas mis existencias;

para mis hijos, Emily Monserrath y Mathías Josué, la continuación de mi ser,

mi inspiración más profunda y el sueño más anhelado hecho realidad.

A mis padres, Bertha Ibujés y Segundo Suárez,

por su ejemplo incansable de perseverancia y amor.

AGRADECIMIENTO

Agradezco profundamente a los docentes pares revisores por sus valiosas aportaciones y sugerencias, indispensables para el perfeccionamiento de esta obra. Cada observación fue recibida con gratitud y consideración, demostrando que el conocimiento se construye en comunidad.

De igual manera, expreso mi profundo agradecimiento a las instituciones que respaldaron esta obra y permitieron incluir su logo institucional como muestra de su apoyo: Equipo de Asesores Educativos de la Coordinación Zonal 1-Educación, Colegio Profesional de Asesores Educativos del Ecuador, Centro Cultural Antonio Ante, y Asociación de Maestros de Excelencia Educativa, instituciones de las cuales tengo el honor de ser integrante.

CONTENIDOS

PRESENTACIÓN

En un mundo cada vez más orientado a los datos, el dominio de la estadística descriptiva se ha convertido en una habilidad esencial para estudiantes, profesionales y cualquier persona interesada en el análisis de información. Este segundo volumen de *Estadística Descriptiva para Todos: Fundamentos y Aplicaciones* está diseñado para profundizar en los conceptos fundamentales de las medidas estadísticas, ofreciendo una explicación clara, ejemplos prácticos y ejercicios aplicados que facilitan el aprendizaje.

En este libro, el lector encontrará un recorrido detallado por:

- **Medidas de Tendencia Central** (media, mediana y moda), con métodos de cálculo para datos no agrupados, tablas de frecuencias y datos agrupados en intervalos.
- **Medidas de Posición** (cuartiles, quintiles, deciles, percentiles) y su representación gráfica mediante el *Diagrama de Caja y Bigotes (Box Plot)*.
- **Medidas de Dispersión** (rango, desviación media, varianza, desviación estándar y coeficiente de variación), fundamentales para entender la variabilidad de los datos.

Lo que hace especial a esta obra es su equilibrio entre rigor académico y accesibilidad. Cada concepto teórico va acompañado de ejemplos prácticos de aplicación resueltos paso a paso, muchos de ellos utilizando herramientas comunes como Excel y SPSS. Además, incluye una evaluación diagnóstica al inicio del libro, y actividades prácticas al final de cada capítulo con sus respuestas que permiten al lector verificar su comprensión. También, al finalizar el libro se encuentra el solucionario de la evaluación diagnóstica y un formulario resumen para una rápida consulta.

Este libro está dirigido a estudiantes de ciencias sociales, administración, ingeniería y otras áreas que requieran el manejo de datos, así como a docentes que busquen un material didáctico y accesible, y también para toda persona interesada en desarrollar competencias estadísticas básicas. Con un enfoque práctico y pedagógico, *Estadística Descriptiva para Todos: Fundamentos y Aplicaciones – Volumen II* busca eliminar el miedo a los números y demostrar que, con las herramientas adecuadas, todos pueden dominar la estadística.

Al concluir este libro, el lector no solo habrá adquirido conocimientos estadísticos, sino también la confianza para aplicarlos. En un mundo donde los datos son el nuevo lenguaje universal, esta obra busca fomentar una cultura estadística para tomar decisiones más informadas.

¡Bienvenidos a un viaje de descubrimiento numérico, donde los datos cobran sentido y la estadística se vuelve comprensible para todos!

El autor

EVALUACIÓN DIAGNÓSTICA

OBJETIVO: Verificar los resultados de aprendizaje previos adquiridos por los lectores a través del presente cuestionario para emitir juicios de valor y tomar decisiones.

INSTRUCCIONES:

Estimado lector:

- ✓ La evaluación tiene una duración de 2 horas.
- ✓ Cada pregunta tiene una valoración de dos puntos.
- ✓ No se ortorgará valoración a una respuesta correcta que no esté acompañada de un proceso de solución escrito.
- ✓ Emplee hojas adicionales y un esferográfico para resolver el presente cuestionario.
- ✓ Lea cuidadosamente el cuestionario y conteste empleando sus conocimientos previos.

¡Éxito!

CUESTIONARIO

1) Una Parroquia de una determinada ciudad está compuesta de 3 barrios, el barrio A de 2000 habitantes, el barrio B de 2500 habitantes y el barrio C de 1000 habitantes. Se va a realizar un estudio de mercado. Realice los cálculos en forma manual y con Excel. ¿Cuántas encuestas se debe aplicar si se emplea el cálculo del tamaño de la muestra al 95% de confianza con un error de muestreo del 5%? ¿Cuántas encuentras se deben aplicar en cada barrio? Realice la interpretación de los resultados.

Con los siguientes datos sobre las calificaciones trimestrales en la asignatura de Matemática evaluadas sobre 10 obtenidas de 45 estudiantes:

8,83	8,85	8,74	7,70	9,08
4,48	8,40	8,70	8,99	9,09
6,99	8,79	7,90	7,69	8,96
7,00	8,70	8,93	9,11	9,01
8,84	9,05	8,85	9,83	9,01
9,83	8,33	8,88	8,90	7,33
8,78	9,01	8,96	8,38	8,26
8,71	8,85	8,55	8,11	9,23
8,84	6,24	8,99	4,00	9,31

2) Calcule la media aritmética. Realice la interpretación.

3) Calcule la moda. Realice la interpretación.

4) Calcule los cuartiles. Realice las interpretaciones.

5) Calcule el rango. Realice la interpretación.

6) Calcule la desviación estándar de la muestra. Realice la interpretación.

7) Calcule el coeficiente de variación. Realice la interpretación.

8) Calcule el error estándar de la media. Realice la interpretación.

9) Elabore un diagrama de barras, un histograma y un diagrama de sectores considerando los intervalos empleados en el sistema educativo del Ministerio de Educación del Ecuador. Realice la interpretación.

10) Elabore un diagrama de caja y bigotes. Realice la interpretación.

CAPÍTULO I
MEDIDAS DE TENDENCIA CENTRAL

1.1) DEFINICIÓN

Las medidas de tendencia central son valores numéricos que representan el centro o punto de equilibrio de un conjunto de datos. Su objetivo es resumir la información en un solo valor que sea representativo de toda la distribución, facilitando su interpretación y comparación.

1.2) MEDIA ARITMÉTICA SIMPLE

A) DEFINICIÓN

La media aritmética simple (o simplemente media) es un promedio que representa el valor central de un conjunto de datos numéricos. Se calcula sumando todos los valores y dividiendo el resultado entre el número total de datos.
La media aritmética de la población se simboliza con la letra griega μ (mi/mu).
La media aritmética de muestra se simboliza con la letra $\bar{x}$ (x barra)

Características Clave
Medida de tendencia central: Resume todo el conjunto en un solo valor.
Sensible a valores extremos (outliers): Ejemplo: En 3, 5, 7, 9, 100, la media es 24,8 (distorsionada por el 100). No es robusta ante outliers o valores atípicos (un valor extremo puede distorsionarla).
Uso común: Calcular promedios de notas, ingresos, temperaturas, etc. No aplica para datos cualitativos (solo funciona con variables cuantitativas).

Alternativas si hay outliers:
Mediana: Valor central al ordenar los datos.
Moda: Valor más frecuente.

B) MÉTODOS DE CÁLCULO

1) Para datos no agrupados

Ejemplo de aplicación:

Calcular la media aritmética de las siguientes calificaciones de Estadística tomadas de una muestra:

4, 8, 10, 10, 5, 10, 9, 8, 6, 8, 10, 8, 5, 7, 4, 4, 8, 8, 6 y 6

Solución:

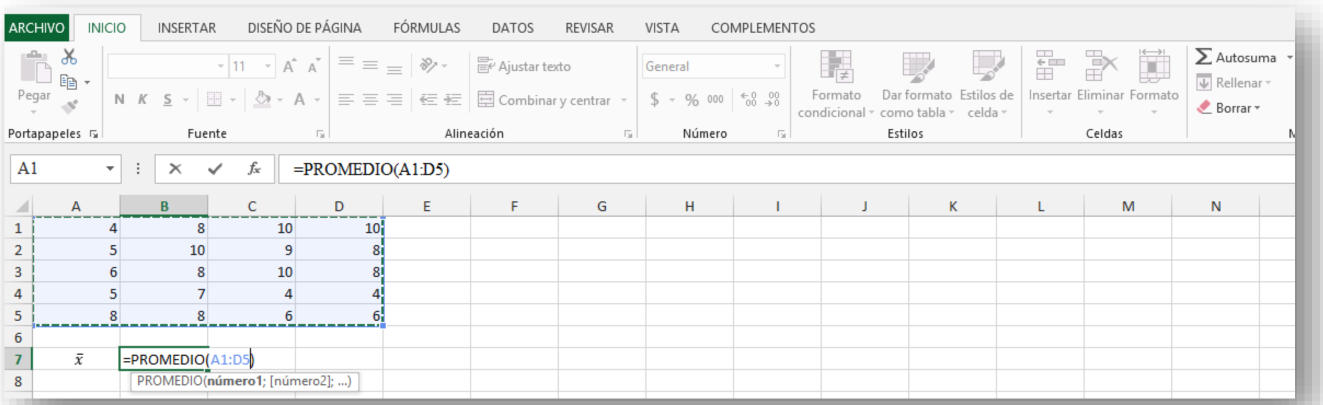
$$\bar{x} = \frac{\sum x_i}{n}$$

$$\bar{x} = \frac{4+8+10+10+5+10+9+8+6+8+10+8+5+7+4+4+8+8+6+6}{20}$$

$$\bar{x} = \frac{144}{20} = 7{,}2$$

Empleando Excel:

Se escriben los números, clic en Autosuma. Clic en Promedio. Seleccione los datos (Rango A1:D5).



Enter

	A	B	C	D
1	4	8	10	10
2	5	10	9	8
3	6	8	10	8
4	5	7	4	4
5	8	8	6	6
6				
7	$\bar{x}$	7,2	=PROMEDIO(A1:D5)	

Interpretación:

La media **7,2** representa el **promedio exacto** de las calificaciones.

Ventaja: Precisión al usar todos los datos individuales.

Limitación: Sensible a valores extremos (aunque en este caso no hay outliers).

2) Para datos en tablas de frecuencias

Ejemplo de aplicación:

Calcular la media aritmética de las siguientes calificaciones de Estadística tomadas de una muestra.

x	f
4	3
5	2
6	3
7	1
8	6
9	1
10	4
Total	20

Solución:

Se emplea la siguiente fórmula:

$$\bar{x} = \frac{\sum f x}{n}$$

$$\bar{x} = \frac{3 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 + 1 \cdot 7 + 6 \cdot 8 + 1 \cdot 9 + 4 \cdot 10}{3 + 2 + 3 + 1 + 6 + 1 + 4} = \frac{144}{20} = 7{,}2$$

Empleando Excel:

Se calcula la frecuencia absoluta. Luego se inserta la función SUMAPRODUCTO. Enter

B15 =SUMAPRODUCTO(A8:A14;B8:B14)/B15

	A	B	C	D	E	F	G	H
1	4	8	10	10				
2	5	10	9	8				
3	6	8	10	8				
4	5	7	4	4				
5	8	8	6	6				
6								
7	x	f						
8	4	3	=CONTAR.SI($A$1:$D$5;A8)					
9	5	2	=CONTAR.SI($A$1:$D$5;A9)					
10	6	3	=CONTAR.SI($A$1:$D$5;A10)					
11	7	1	=CONTAR.SI($A$1:$D$5;A11)					
12	8	6	=CONTAR.SI($A$1:$D$5;A12)					
13	9	1	=CONTAR.SI($A$1:$D$5;A13)					
14	10	4	=CONTAR.SI($A$1:$D$5;A14)					
15	Total	20	=SUMA(B8:B14)					
16	$\bar{x}$	4)/B15						

	A	B	C	D	E	F
1	4	8	10	10		
2	5	10	9	8		
3	6	8	10	8		
4	5	7	4	4		
5	8	8	6	6		
6						
7	x	f				
8	4	3	=CONTAR.SI($A$1:$D$5;A8)			
9	5	2	=CONTAR.SI($A$1:$D$5;A9)			
10	6	3	=CONTAR.SI($A$1:$D$5;A10)			
11	7	1	=CONTAR.SI($A$1:$D$5;A11)			
12	8	6	=CONTAR.SI($A$1:$D$5;A12)			
13	9	1	=CONTAR.SI($A$1:$D$5;A13)			
14	10	4	=CONTAR.SI($A$1:$D$5;A14)			
15	Total	20	=SUMA(B8:B14)			
16	$\bar{x}$	7,2	**=SUMAPRODUCTO(A8:A14;B8:B14)/B15**			

Interpretación:

La media **7.2** representa el promedio exacto de las calificaciones.

La nota **8** es la más frecuente (6 veces) y contribuye significativamente al promedio.

3) Para datos agrupados en intervalos

Ejemplo de aplicación:

Calcular la media aritmética con los datos de la siguiente tabla:

Intervalos	f	xm
4- 5	5	4,5
6 -7	4	6,5
8- 9	7	8,5
10-11	4	10,5

Solución:

Se emplea la siguiente formula:

$$\bar{x} = \frac{\sum f \cdot xm}{n}$$

$$\bar{x} = \frac{5 \cdot 4{,}5 + 4 \cdot 6{,}5 + 7 \cdot 8{,}5 + 4 \cdot 10{,}5}{5 + 4 + 7 + 4} = \frac{150}{20} = 7{,}5$$

Nota: En datos agrupados, la marca de clase aproxima el valor real, introduciendo un pequeño margen de error

Empleando Excel:

Se calcula el valor máximo($X_{máx}$), el valor mínimo($X_{mín}$), el Rango(R), el número de intervalos (n_i), el ancho de los intervalos (A_i) la marca del clase (xm), la frecuencia absoluta (f) y el número total de datos (n). Luego se inserta la función: SUMAPRODUCTO como se muestra en la siguiente imagen:

	A	B	C	D	E	F	G
1	4	8	10	10			
2	5	10	9	8			
3	6	8	10	8			
4	5	7	4	4			
5	8	8	6	6			
6							
7	$X_{máx}$	10	=MAX(A1:D5)				
8	X_{mix}	4	=MIN(A1:D5)				
9	Rango	6	=B7-B8				
10	n	20	=CONTAR(A1:D5)				
11	n_i	5	=ENTERO(1+3,32*LOG(B10))				
12	i	1,2	=B9/B11				
13							
14	Intervalos		xm	*f*	{=FRECUENCIA(A1:D5;B15:B18)}		
15	4	5	4,5	5			
16	6	7	6,5	4			
17	8	9	8,5	7			
18	10	11	10,5	4			
19				20	=SUMA(D15:D18)		
20							
21	$\bar{x}$	**7,5**	**=SUMAPRODUCTO(C15:C18;D15:D18)/D19**				

Interpretación:

La media agrupada (7,5) sobreestima ligeramente la media real (7,2) debido a la aproximación por marcas de clase.

Comparación de Métodos

Método	Media	Precisión	Ventajas	Limitaciones
Sin agrupar	7,2	Alta	Exactitud con datos individuales.	Tedioso para muestras grandes.
Tabla de frecuencias	7,2	Alta	Organiza datos y muestra patrones.	Requiere conteo.
Agrupado	7,5	Baja	Útil para grandes conjuntos.	Introduce error de aproximación.

1.3) MEDIA ARITMÉTICA PONDERADA

A) DEFINICION

Es una variante de la media simple donde cada valor tiene un peso que refleja su importancia relativa dentro del conjunto de datos. Se utiliza cuando no todas las observaciones contribuyen igualmente al promedio.

B) MÉTODO DE CÁLCULO

Cuando los números $x_1, x_2, x_3, \dots x_k$ se les asocian ciertos factores peso (o pesos) $p_1, p_2, p_3, \dots p_k$, dependientes de la relevancia asignada a cada número, en tal caso se requiere calcular la media aritmética ponderada, la cual se calcula así:

$$\bar{x} = \frac{p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 + \cdots p_k \cdot x_k}{p_1 + p_2 + p_3 + \cdots p_k} = \frac{\sum p \cdot x_i}{\sum p}$$

C) EJEMPLOS DE APLICACIÓN

1) Un estudiante de maestría en la asignatura de Estadística Aplicada obtiene 10 en foros, 9,5 en trabajos autónomos, 8,8 en deberes y 7,5 en pruebas. Los foros aportan con un punto al promedio, los trabajos autónomos, deberes y pruebas con 3 cada una. Calcule el promedio.

Solución:

Insumos	Calificación (x_i)	Ponderación o peso (p)	$p \cdot x_i$
Foros	10	1	10
Trabajos autónomos	9,5	3	28,5
Deberes	8,8	3	26,4
Pruebas	7,5	3	22,5
Total		10	87,4

$$\bar{x} = \frac{\sum p \cdot x_i}{\sum p} = \frac{87{,}4}{10} = 8{,}74$$

Empleando Excel:

	A	B	C	D	E
1	**Insumos**	**Calificaciones**	**Ponderación o peso**		
2	Foros	10	1		
3	Trabajos autónomos	9,5	3		
4	Deberes	8,8	3		
5	Pruebas	7,5	3		
6					
7	$\bar{x}$	8,74	=SUMAPRODUCTO(C2:C5;B2:B5)/SUMA(C2:C5)		
8		Mario Suárez			

Interpretación:

Promedio Final:

El estudiante obtiene un 8,74/10 en la asignatura, lo que refleja un buen rendimiento global, cercano al nivel de excelencia (9-10).

Contribución por Actividad:

Foros (10): Aunque tiene la calificación más alta, su bajo peso (1) limita su impacto en el promedio.

Trabajos autónomos (9,5): La mejor combinación de alta calificación y peso significativo (3), aportando 28,5/87,4 (32,6%) del total.

Deberes (8,8): Similar a los trabajos autónomos, pero con un ligero descenso en la nota.

Pruebas (7,5): El punto más débil, con la menor calificación y un peso alto (3), lo que reduce el promedio.

2) Emplear la media aritmética simple para el cálculo de la Exoneración en la Evaluación Final de Bachillerato en Ecuador y la media aritmética ponderada para la nota final de título de Bachiller. Aplique el Art. 37 y Art. 151 del Reglamento de la Ley Orgánica de Educación Intercultural (LOEI) sobre la evaluación final de bachillerato.

Solución:

El Artículo 37 del Reglamento de la LOEI en su parte pertinente sobre la exoneración menciona: La exoneración de la Evaluación Final de Bachillerato es un beneficio opcional para los estudiantes que dominen los aprendizajes requeridos; es decir aquellos que obtengan un promedio simple entre nueve (9) y diez (10) puntos en la trayectoria educativa en Educación General Básica Media y Superior, así como en Bachillerato y en el Programa de Participación Estudiantil.

El Artículo 151 de la LOEI en su parte pertinente sobre la obtención del título de bachiller menciona: Para obtener el título de bachiller en todas las modalidades educativas, el estudiante deberá conseguir una nota de grado mínima de siete (7) sobre diez (10), la misma que será el promedio ponderado mínimo de las notas correspondientes a:

a. Trayectoria educativa en Educación General Básica Media: Promedio obtenido en los tres años correspondientes al subnivel de Educación General Básica Media, equivalente al veinte por ciento (20%).
b. Trayectoria educativa en Educación General Básica Superior: Promedio obtenido en los tres (3) años correspondientes al subnivel de Educación General Básica Superior, equivalente al treinta por ciento (30%).
c. Trayectoria educativa en Bachillerato: Promedio obtenido en los tres años correspondientes al nivel de Bachillerato, equivalente al treinta por ciento (30%).
d. Participación estudiantil: Nota obtenida en el programa de participación estudiantil, equivalente al diez por ciento (10%).
e. Evaluación Final de Bachillerato: Nota obtenida de la evaluación que el estudiante desarrolla en el tercer año de Bachillerato, que valorará los logros establecidos en los estándares de aprendizaje equivalente al diez por ciento (10%).

Cumpliendo con esta normativa se realiza los cálculos en Excel

	A	B	C	D	E
1		Cálculo de la exoneración de la Evaluación Final de Bachillerato			
2		Componentes	Calificaciones		
3	1	Promedio EBG Media	9,55		
4	2	Promedio EGB Superior	8,75	¿Se exonera?	
5	3	Promedio Bachillerato	8,50		
6	4	Participación Estudiantil	9,85		
7		Promedio simple	9,16	SI	
8					
9		Cálculo de la nota final del título de bachiller			
10		a) Estudiante opta por la exoneración			
11			Calificaciones	Ponderación	Calificación ponderada
12	1	Promedio EBG Media	9,55	20%	1,91
13	2	Promedio EGB Superior	8,75	30%	2,63
14	3	Promedio Bachillerato	8,50	30%	2,55
15	4	Participación Estudiantil	9,85	10%	0,99
16	5	Evaluación Final de Bachillera	9,16	10%	0,92
17		Promedio ponderado			**8,99**
18					**Nota final del título de bachiller**
19	b) Estudiante NO opta por la exoneración (ejemplo con 9,5 en la evaluación Final de Bachillerato)				
20			Calificaciones	Ponderación	Calificación ponderada
21	1	Promedio EGB Superior	9,55	20%	1,91
22	2	Promedio EGB Superior	8,75	30%	2,63
23	3	Promedio Bachillerato	8,50	30%	2,55
24	4	Participación Estudiantil	9,85	10%	0,99
25	5	Evaluación Final de Bachillera	9,50	10%	0,95
26		Promedio ponderado			**9,02**
27					**Nota final del título de bachiller**

Interpretación: Para este ejemplo el estudiante se exonera con 9,16. Si el estudiante opta por la exoneración, el estudiante obtiene una nota final del título de bachiller de 8,99. Si el estudiante no opta por la exoneración y en la evaluación final de Bachillerato obtiene 9,50, el estudiante obtendría una nota final del título de bachiller de 9,02.

1.4) LA MEDIANA

A) DEFINICIÓN

La mediana es un valor central que divide un conjunto de datos ordenados en dos partes iguales:

50% de los datos son menores o iguales a la mediana.

50% de los datos son mayores o iguales a la mediana.

Características clave:

Medida robusta que No se ve afectada por valores extremos (outliers), a diferencia de la media.

Su aplicación es ideal en datos ordinales.

Útil en distribuciones asimétricas (Ejemplo: ingresos, precios de viviendas).

B) MÉTODOS DE CÁLCULO

1) Para datos no agrupados

Ejemplos de aplicación:

1) Calcular la media aritmética y la mediana de las siguientes calificaciones del curso de Estadística evaluadas en la escala de 0 a 10: 10, 8, 6, 4, 9, 7, 10, 9 y 6

Solución:

Se ordena los datos de menor a mayor:

4	6	6	7	8	9	9	10	10
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9

Se aplica la fórmula para encontrar la ubicación:

$$Me = x_{\frac{n+1}{2}}$$

$Me = x_{\frac{9+1}{2}} = Me = x_5$

Por lo tanto, Me = 8

En Excel:

	A	B	C	D	E	F	G	H	I
1	4	6	6	7	8	9	9	10	10
2									
3	Me	8	=MEDIANA(A1:I1)						
4	$\bar{x}$	7,67	=PROMEDIO(A1:I1)						

Interpretación:

Mediana (Me) = 8:
El 50% de los estudiantes obtuvo una calificación igual o inferior a 8, y el otro 50% igual o superior a 8.

Es una medida robusta que no se ve afectada por valores extremos (Ejemplo: las notas 4 y 10 no distorsionan el resultado).

Uso preferente:

En este caso, la mediana es más representativa que la media si hubiera habido más notas extremas (Ejemplo: varios 4 o 10).

La mediana (8) es ligeramente mayor que la media (7,67), lo que sugiere una ligera asimetría hacia notas bajas (pero mínima, dado que 8 y 7,67 están cerca).

Conclusión:

La mediana (8) es más representativa que la media en este caso porque:

La presencia del 4 (aunque no extremo) baja la media, subestimando el rendimiento general.

La distribución es casi simétrica, pero la mediana resiste mejor pequeñas asimetrías.

El 8 refleja mejor el "centro real" de las notas, donde la mayoría está entre 7 y 10.

Recomendación:

Usar ambas medidas para análisis complementarios:

Mediana: Para describir el rendimiento típico.

Media: Para cálculos de promedios grupales o sumatorias.

2) Calcular la mediana de las siguientes calificaciones del curso de Matemática evaluadas de 0 a 10. Las calificaciones son: 10, 10, 10, 6, 8, 8, 8, 8, 10 y 9. También calcule la media aritmética.

Solución:

Se ordena los datos de menor a mayor:

6	8	8	8	8	9	10	10	10	10
$\boldsymbol{x_1}$	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}

Se aplica la ecuación

$$Me = x_{\frac{n+1}{2}}$$

$$Me = x_{\frac{10+1}{2}} = x_{\frac{11}{2}} = x_{5,5}$$

Promedio entre x_5 y x_6

$$Me = \frac{x_5 + x_6}{2} = \frac{8+9}{2} = 8{,}5$$

Empleando Excel

	A	B	C	D	E	F	G	H	I	J
1	6	8	8	8	8	9	10	10	10	10
2										
3	Me	8,5	=MEDIANA(A1:J1)							
4	$\bar{x}$	8,70	=PROMEDIO(A1:J1)							

Interpretación:

Mediana (Me) = 8,5:

El 50% de los estudiantes obtuvo 8,5 o menos, y el otro 50% 8,5 o más.

La mediana es robusta ante valores extremos (como el 6 o los múltiples 10).

Media = 8,70:

El promedio del grupo es 8,70, ligeramente superior a la mediana (8,5).

Esto indica que los valores altos (10) tienen más peso que el valor bajo (6).

Conclusión:

La mediana (8,5) es más representativa en este caso porque:

Resume mejor el rendimiento típico: La mayoría de los estudiantes obtuvieron 8 o 9.

Es robusta: No se ve afectada por el 6 (única nota baja) ni por los múltiples 10.

Refleja la distribución real: Cae en la zona de mayor concentración de datos.

Recomendación:

Usar la mediana (8,5) para describir el rendimiento típico, ya que resiste mejor la influencia del 6.

Usar la media (8,70) para cálculos que requieran incluir todas las notas (Ejemplo: promedio grupal).

2) Para datos en tablas de frecuencia

Para calcular la posición de la mediana se aplica la siguiente fórmula:

$$\boldsymbol{Me = x_{\frac{n+1}{2}}}$$

Ejemplo de aplicación:

Calcular la mediana agrupando los datos en tabla de frecuencia de los siguientes 20 números: 1, 3, 3, 5, 5, 5, 5, 2, 2, 2, 6, 6, 4, 4, 4 ,4, 5, 5, 5, 5.

Solución:

x	f
1	1
2	3
3	2
4	4
5	8
6	2
Total	20

Calculando la posición de la mediana se obtiene:

$$Me = x_{\frac{20+1}{2}} = x_{\frac{21}{2}} = x_{10,5} \rightarrow Me = \frac{x_{10} + x_{11}}{2}$$

Para observar cuáles son x_{10} y x_{11} se calcula la frecuencia acumulada (F)

x_i	f	F	Ubicación
1	1	1	x_1
2	3	4	x_2 $hasta\ el$ x_4
3	2	6	x_5 $hasta\ el$ x_6
4	4	10	x_7 $hasta\ el$ x_{10}
5	8	18	x_{11} $hasta\ el$ x_{18}
6	2	20	x_{19} $hasta\ el$ x_{20}
Total	20		

Reemplazando valores se obtiene:

$$Me = \frac{x_{10} + x_{11}}{2} = \frac{4+5}{2} = 4{,}5$$

Interpretación:

Mediana = 4,5:

El 50% de los valores son ≤ 4,5, y el otro 50% son ≥ 4,5.

Esto significa que la mitad de los datos están en el rango 1-4, y la otra mitad en 5-6.

3) Para datos agrupados en intervalos

Se emplea la ecuación

$$Me = L_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i}\right) \cdot A_i$$

Donde:
$L_i =$ Límite inferior del intervalo de clase de la mediana
$n =$ Número total de datos
F_{i-1} =Frecuencia acumulada del intervalo de clase que antecede al intervalo de la mediana.
f_i = Frecuencia absoluta del intervalo de la mediana
A_i = Amplitud o Ancho del intervalo

Ejemplo de aplicación:

Calcular la mediana de los siguientes datos3:

Intervalos	*f*
[10-20)	4
[20-30)	8
[30-40)	6
[40-50)	2

Solución:

Intervalos	*f*	*F*	Ubicación
[10-20)	4	4	x_1 *hasta el* x_4
[20-30)	8	12	x_5 *hasta el* x_{12}
[30-40)	6	18	x_{13} *hasta el* x_{18}
[40-50)	2	20	x_{19} *hasta el* x_{20}

Se calcula la posición de la mediana de la siguiente manera:

$$Me = x_{\frac{n}{2}} = x_{\frac{20}{2}} = x_{10}$$

Por lo tanto, el intervalo donde se ubica la mediana es:
[20-30)

Al aplicar la ecuación respectiva se obtiene:

$$Me = L_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i}\right) \cdot A_i \Rightarrow Md = 20 + \left(\frac{\frac{20}{2} - 4}{8}\right) \cdot 10 = 20 + \left(\frac{6}{8}\right) \cdot 10 = 20 + \frac{60}{8} = 27{,}5$$

Interpretación:

Significado:

El valor 27,5 divide la distribución en dos partes iguales:

50% de los datos son menores o iguales a 27,5.

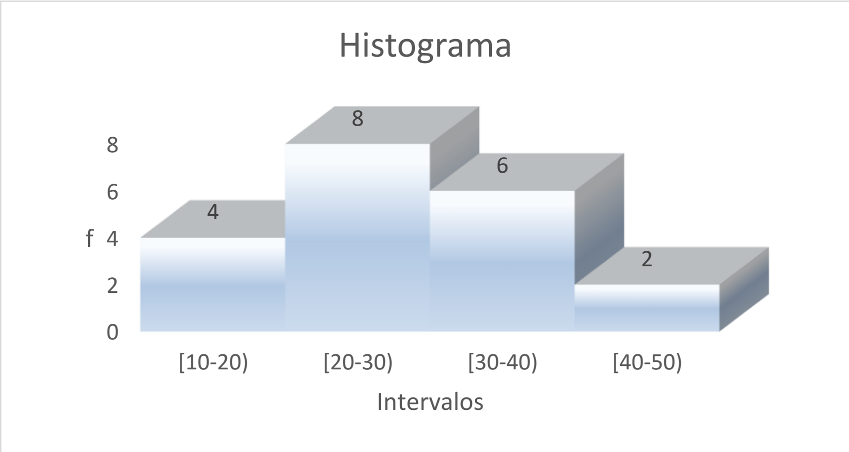
50% de los datos son mayores o iguales a 27,5.

Contexto:

La clase mediana [20, 30) contiene 8 observaciones (40% de los datos).

La mediana está más cerca del límite inferior (20) porque la posición x_{10} está en la primera mitad del intervalo (entre x_5 y x_{12}).

Visualización:



Si se trazara una línea vertical en 27,5 dividiría el gráfico en dos partes iguales

1.5) LA MODA

A) DEFINICIÓN

La moda es una medida de tendencia central que identifica el valor (o valores) que aparece con mayor frecuencia en un conjunto de datos. A diferencia de la media o la mediana, la moda puede aplicarse tanto a datos numéricos como categóricos (Ejemplo: calificaciones, edades, colores, categorías).

Aplicaciones Prácticas

Ventas: Identificar el producto más vendido.

Educación: Determinar la nota más común en un examen.

Sociología: Analizar la categoría más frecuente en encuestas (Ejemplo: estado civil).

B) MÉTODOS DE CÁLCULO

1) Para datos no agrupados

a) Unimodal: Un único valor con mayor frecuencia.

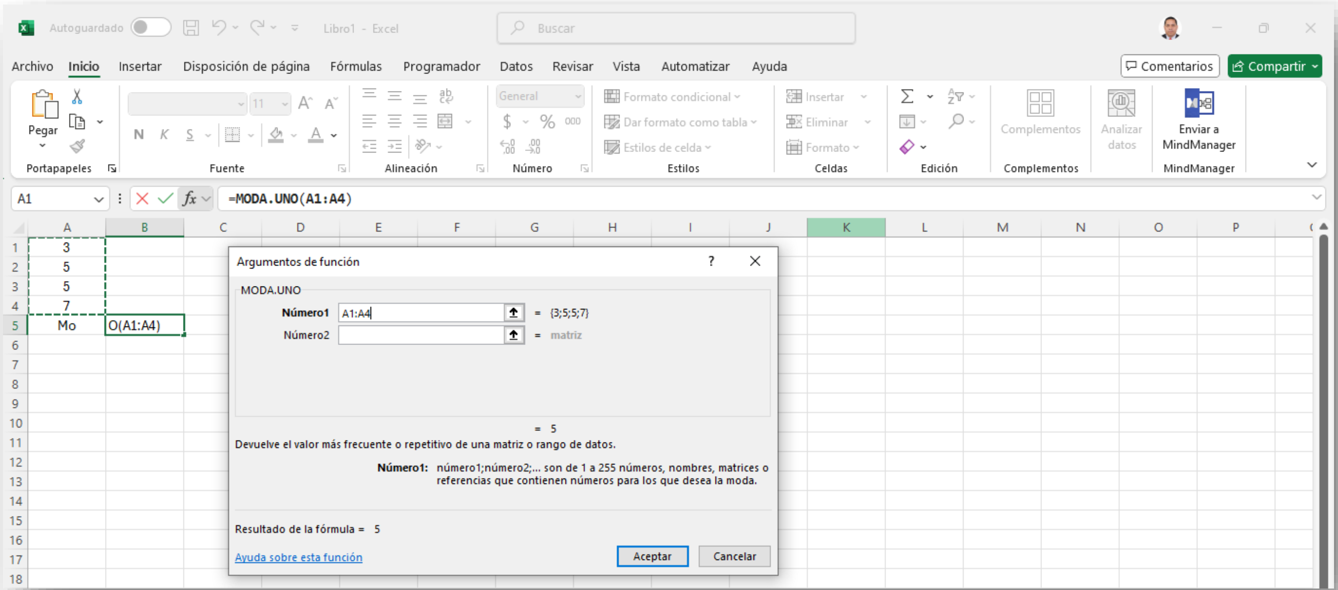
Ejemplo de aplicación: Calcule la moda de 3, 5, 5, 7

Solución:

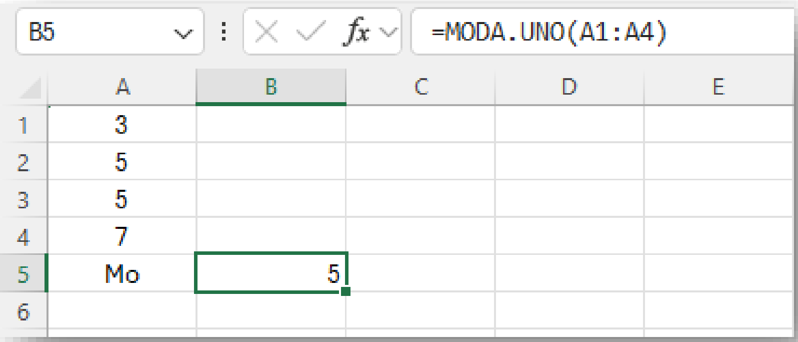
En [3, 5, 5, 7], la moda es **5**.

Empleando Excel:

Insertar función. Seleccionar MODA.UNO. Seleccionar los datos

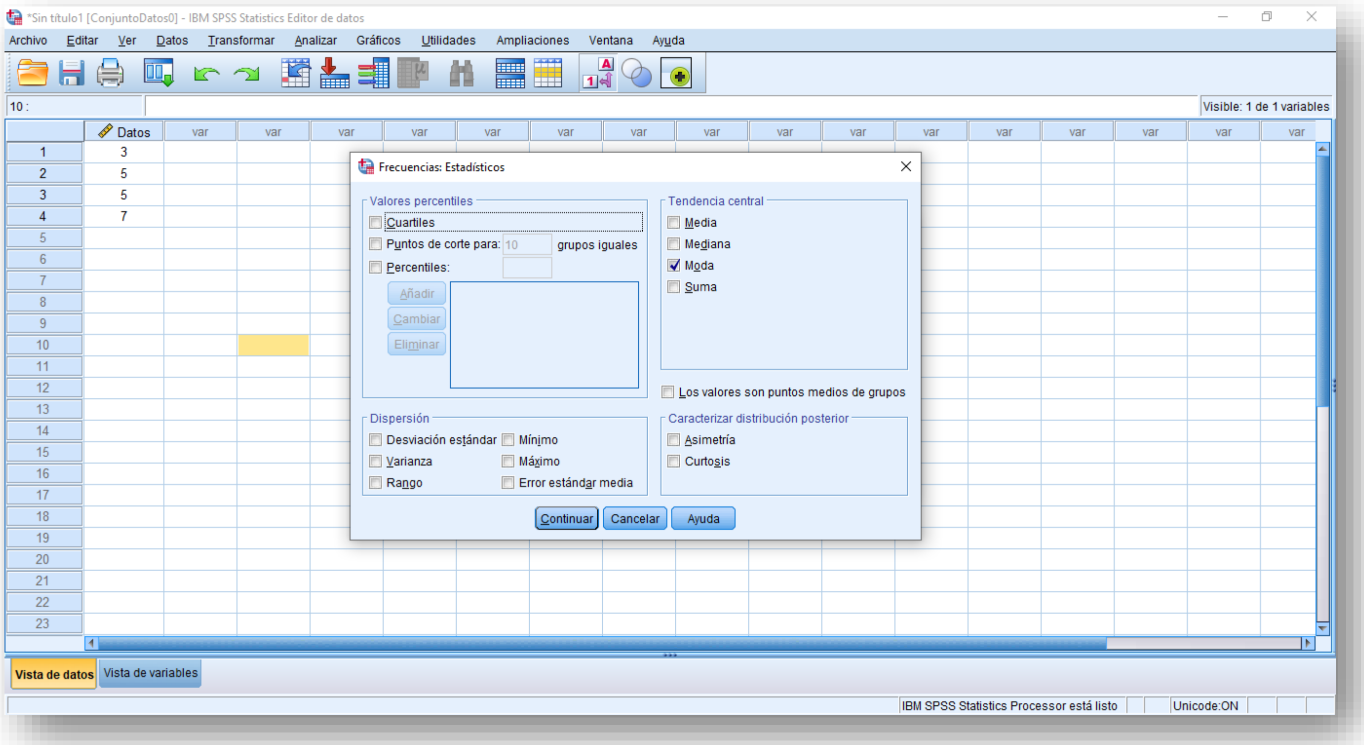


Clic en aceptar

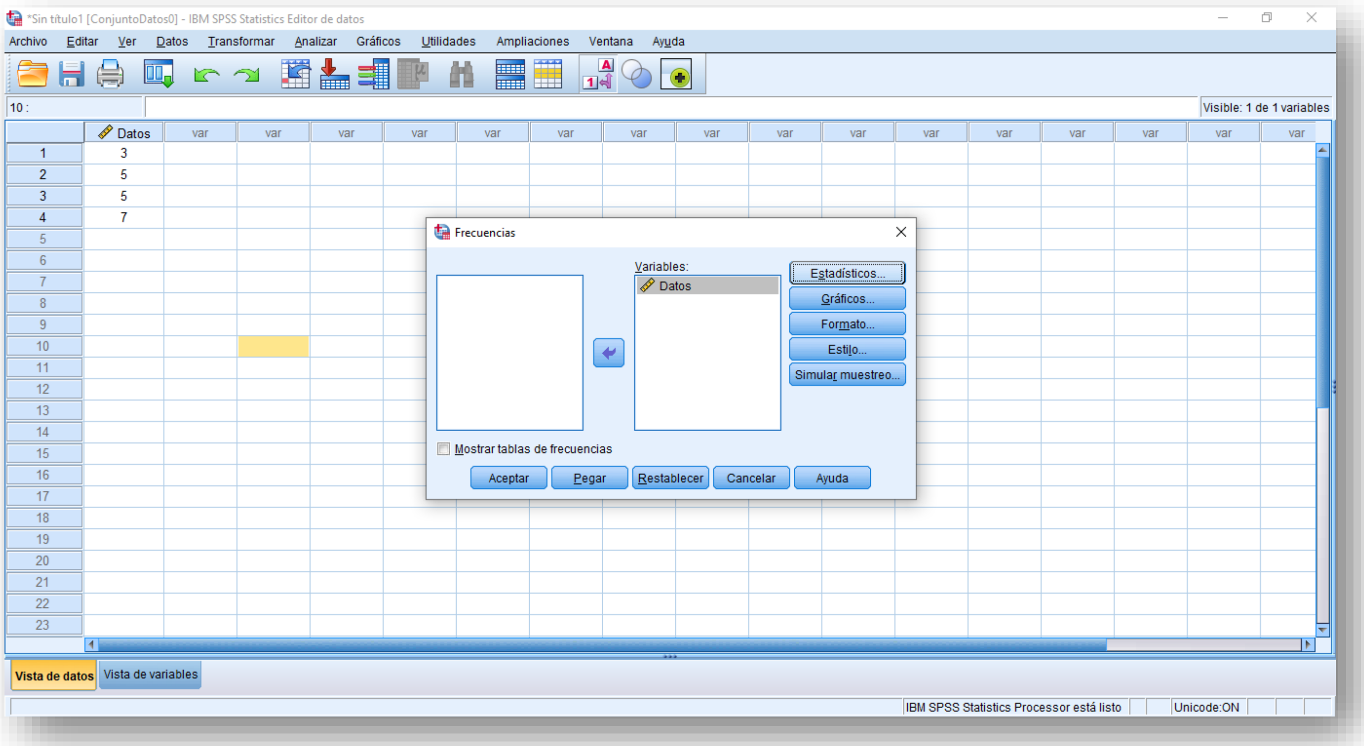


En SPSS

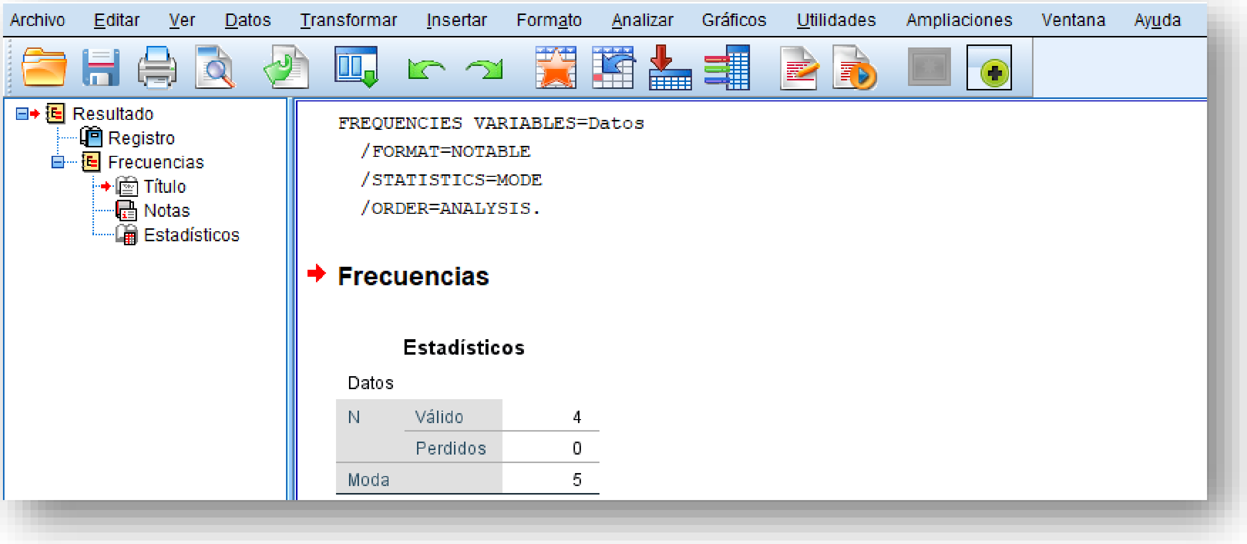
Clic en Analizar. Estadísticos Descriptivos. Frecuencias. Clic en la Flecha. Estadísticos. Moda



Clic en Continuar.



Clic en Aceptar



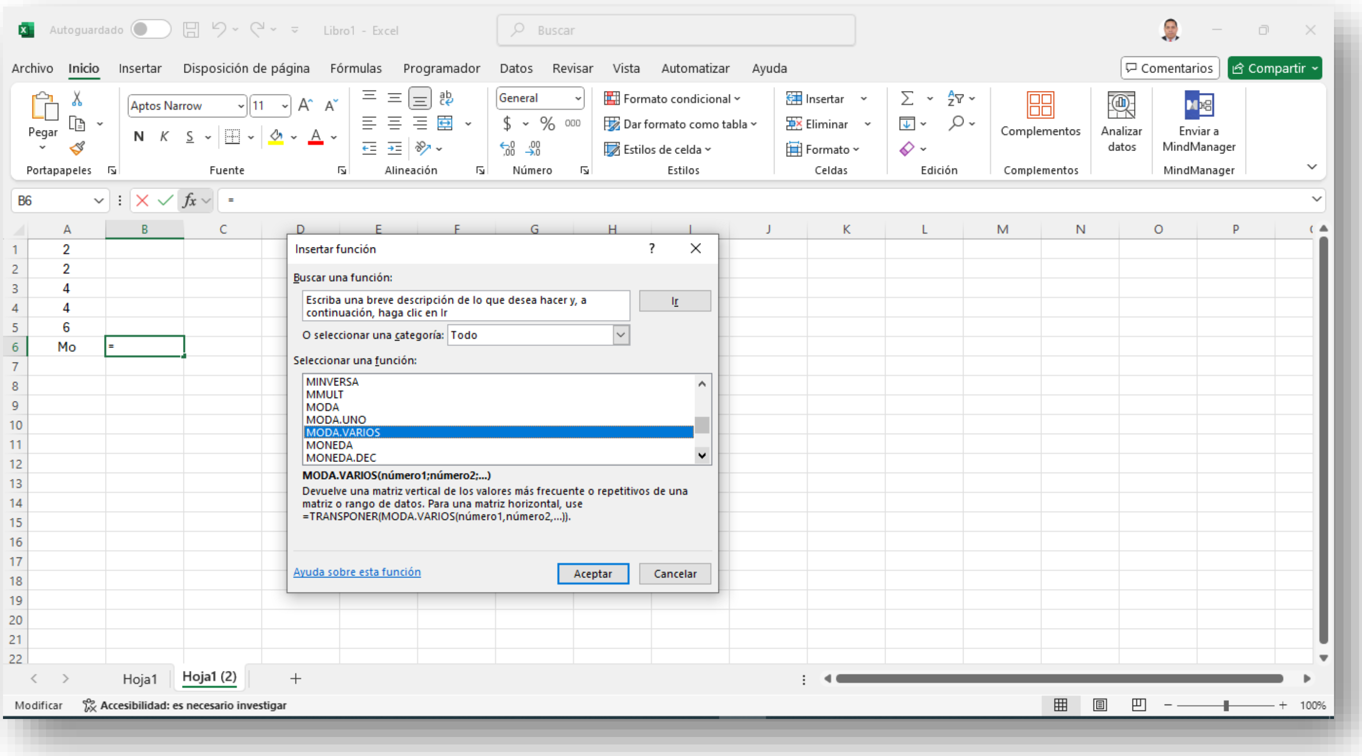
Interpretación: Una distribución es unimodal cuando un solo valor tiene la frecuencia máxima. En el ejemplo: El 5 aparece 2 veces, mientras que el 3 y el 7 aparecen solo 1 vez.

b) Bimodal: Dos valores con la misma frecuencia máxima.

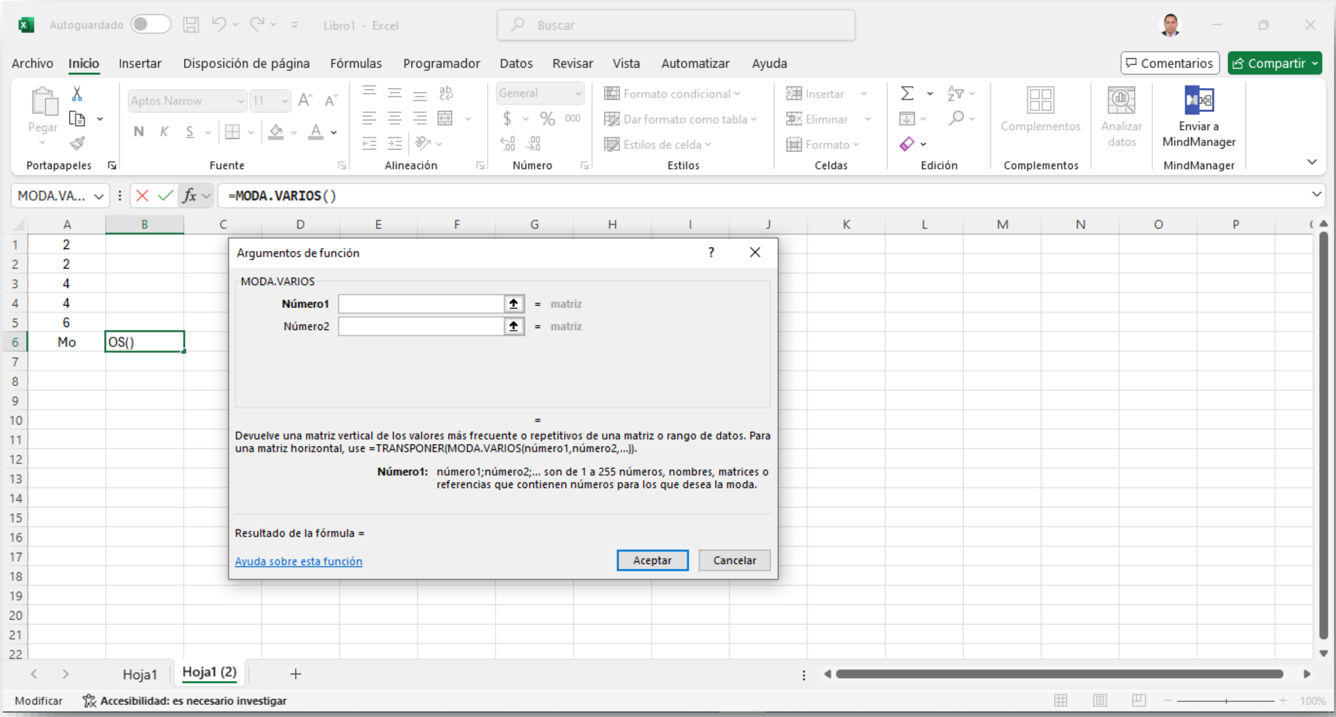
Ejemplo de aplicación: Calcule la moda de 2, 2, 4, 4, 6

Solución: En [2, 2, 4, 4, 6], las modas son **2 y 4**.

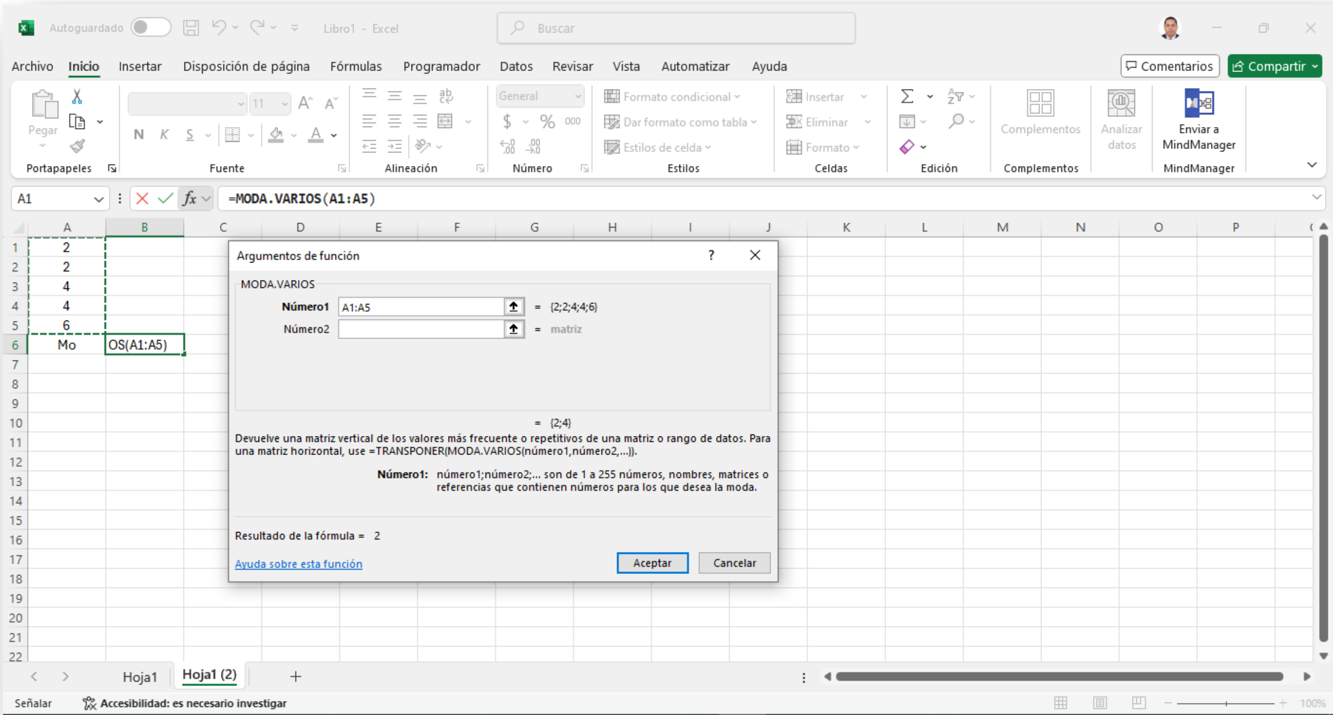
En Excel: Insertar función. Seleccionar MODA.VARIOS



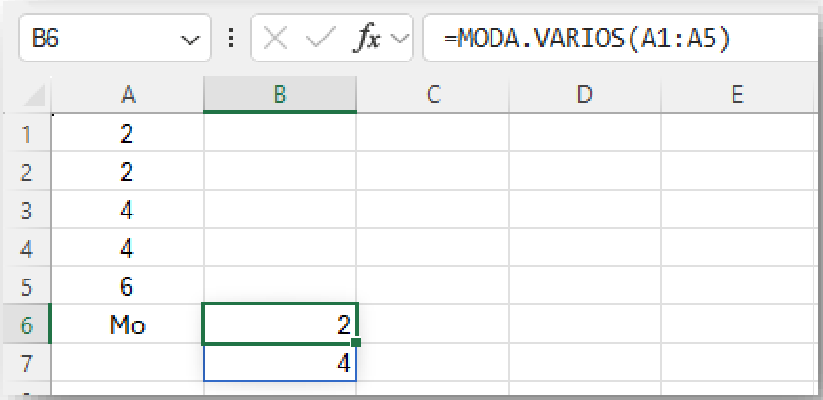
Clic en Aceptar



Seleccionar los datos



Ctrl + Shift + Enter



Interpretación:

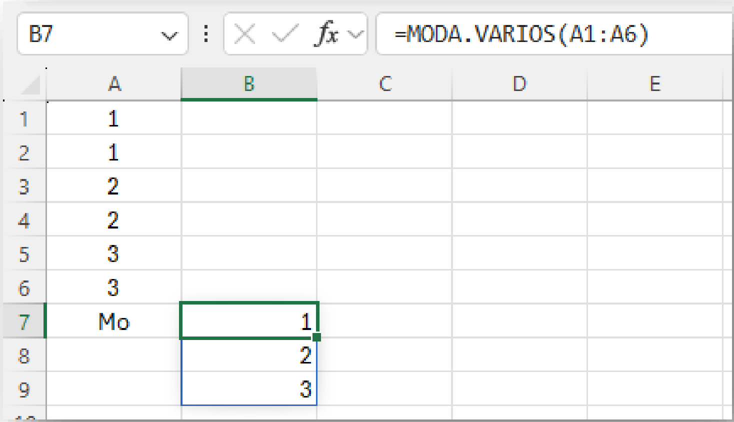
Una distribución es bimodal cuando dos valores distintos comparten la misma frecuencia máxima (es decir, ambos son los más repetidos). En el ejemplo: Tanto **2** como **4** aparecen 2 veces cada uno, mientras que el 6 aparece solo 1 vez.

En el ejemplo se revela que existen dos valores dominantes con igual importancia, es decir, la distribución no está centrada en un solo punto, sino que tiene dos picos.

c) Multimodal: Tres o más valores con frecuencia máxima.

Ejemplo de aplicación:

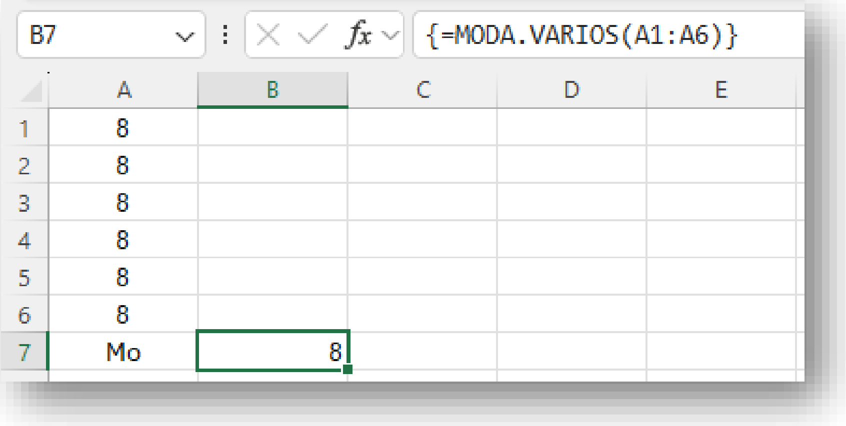
En [1, 1, 2, 2, 3, 3], las modas son **1, 2 y 3**.



c) Plana: Sin Moda. Cuando todos los valores tienen la misma frecuencia.

Ejemplo de aplicación:

En [8, 8, 8, 8, 8, 8], no hay moda (o todos los valores son moda).



2) Para datos en tablas de frecuencias

Ejemplo de aplicación:

En una encuesta se pregunta por el estado civil. Calcule la moda, si se obtuvieron los siguientes resultados:

Estado civil	*f*
Soltero/a	27
Casado/a	20
Divorciado/a	10
Viudo/a	7
Unión Libre	3

Solución:

Se observa que el estado civil Sotero/a tiene mayor frecuencia igual a 27, por lo tanto, Mo = Soltero/a

Interpretación:

La moda es el valor o categoría que aparece con mayor frecuencia en un conjunto de datos. En este caso, es una variable cualitativa (categórica).

Soltero/a es la categoría más frecuente, representando a 27 de 67 encuestados. Esto indica que, en la muestra estudiada, el estado civil predominante es Soltero/a.

3) Para datos agrupados en intervalos

Ejemplo de aplicación:

Calcule la moda de los siguientes datos:

Intervalo o Clase	*f*
$[10, 20)$	3
$[20, 30)$	7
$[30, 40)$	12
$[40, 50)$	15
$[50, 60)$	8

Solución:

El intervalo con mayor frecuencia (clase modal) es $[40, 50)$.

Se emplea la siguiente fórmula:

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) \cdot A_i$$

Donde:

$L_i =$ Límite inferior de la clase modal

$d_1 =$ Diferencia entre la frecuencia de la clase modal y la frecuencia del intervalo anterior.

$d_2 =$ Diferencia entre la frecuencia de la clase modal y la frecuencia del intervalo posterior.

$A_1 =$ Amplitud o ancho de la clase modal.

Aplicando la fórmula

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) \cdot A_i$$

Se obtiene:

$$Mo = 40 + \left(\frac{15 - 12}{(15 - 12) + (15 - 8)}\right) \cdot 10 = 40 + \left(\frac{3}{3 + 7}\right) \cdot 10 = 40 + \frac{30}{10} = 43$$

Interpretación:

El valor 43 es el punto dentro del intervalo [40, 50) donde se concentra la mayor frecuencia de datos. Esto indica que es el valor más común o repetido en la distribución.

La clase modal [40, 50) contiene 15 observaciones (la mayor cantidad).

La moda está más cerca del límite inferior (40) porque la frecuencia del intervalo anterior (12) es mayor que la del posterior (8), lo que "atrae" la moda hacia la izquierda.

1.6) ACTIVIDADES DE APLICACIÓN

1) Realice un organizador gráfico del Capítulo I

2) ¿Qué entiende por media aritmética simple y por media aritmética ponderada? Mediante un ejemplo ilustre su respuesta.

3) Con los siguiente datos 6, 9, 9, 12, 12, 15, 17, 18 calcule la media aritmética, la media y la moda. Interprete los resultados.

12,25; 12; Bimodal (9 y 12).

4) Un estudiante de maestría en la asignatura de Estadística obtiene 10 en foros, 9 en trabajos autónomos, 9,5 en deberes y 7,5 en pruebas. Los foros aportan con un punto al promedio, los trabajos autónomos, deberes y pruebas con 3 puntos cada uno. Empleando Excel calcule el promedio ponderado.

8,8

5) Cree y resuelva empleando Excel un ejercicio aplicando el Art 37 y Art 151 del Reglamento de la LOEI sobre la evaluación final de bachillerato en Ecuador.

6) Con los siguiente datos calcule la media aritmética, la media y la moda. Interprete los resultados.

x	***f***
6	1
9	2
12	3
15	1
17	1

11,5; 12; 12

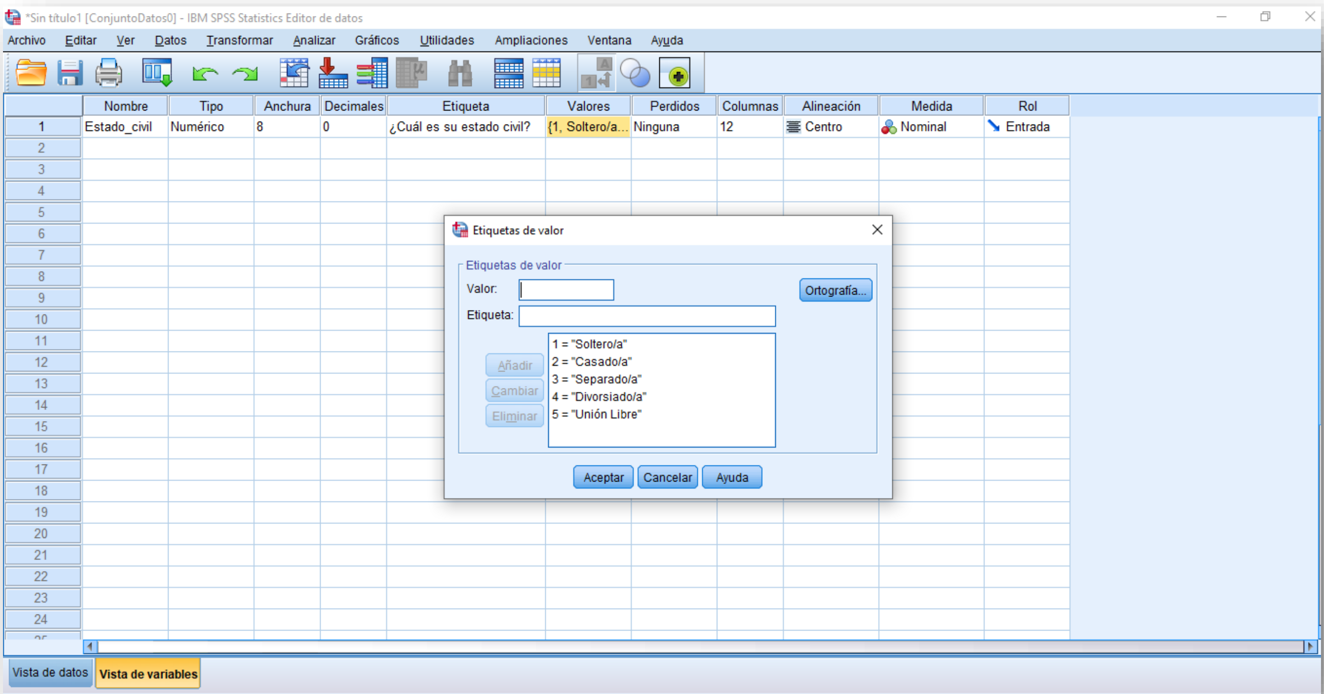
7) Con los siguiente datos sobre edades calcule la media aritmética, la media y la moda. Elabore un histograma. Interprete los resultados.

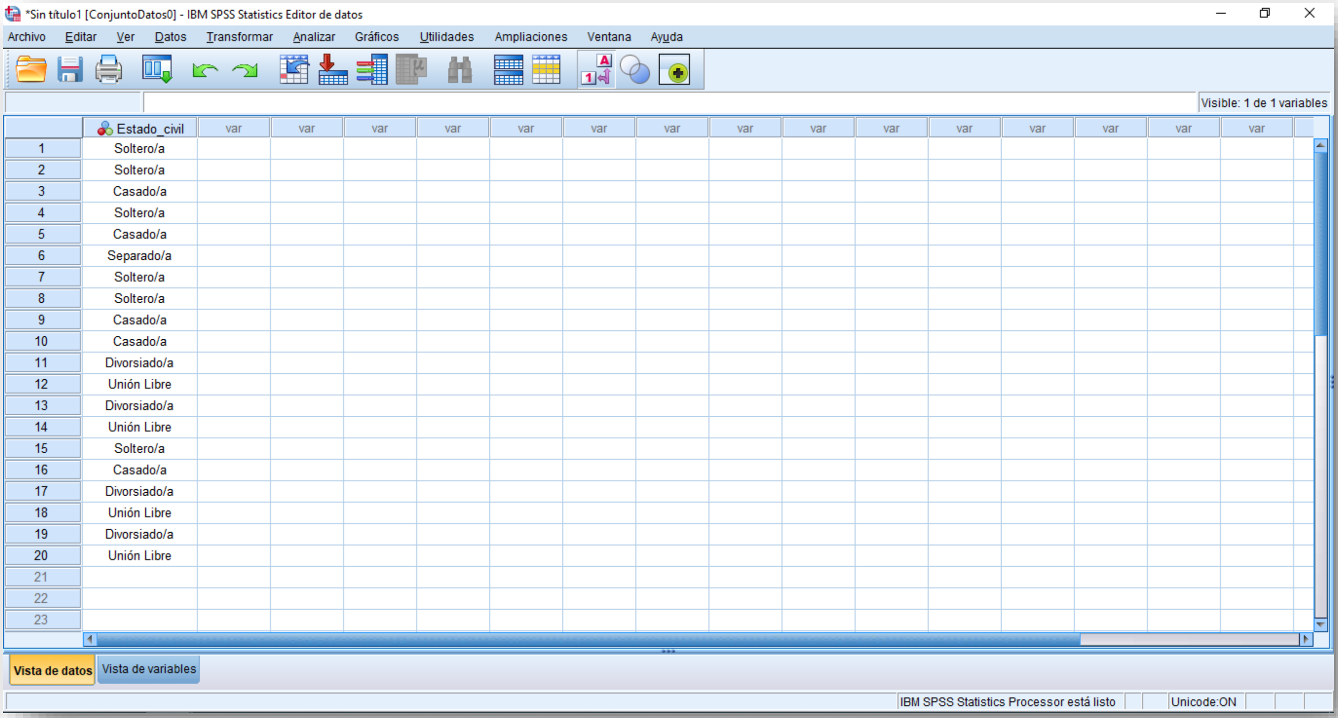
Intervalos	*f*	*xm*
[45-55)	6	50
[55-65)	10	60
[65-75)	19	70
[75-85)	11	80
[85-95)	4	90

69,4; 69,74; 70,29

8) Cree y resuelva un ejercicio similar al anterior

9) Empleando SPSS, calcule las frecuencias, elabore un diagrama de barras y calcule la moda de los siguientes datos.





10) Cree y resuelva un ejercicio similar al anterior

CAPÍTULO II
MEDIDAS DE POSICIÓN

2.1) DEFINICIÓN

Las medidas de posición (también llamadas cuantiles) son valores que dividen un conjunto de datos ordenados en partes iguales, permitiendo analizar la distribución de los datos en segmentos porcentuales.

Son puntos de corte que dividen una distribución de datos en intervalos con la misma proporción de observaciones. Se usan para entender:

- Dónde se concentran los datos (Ejemplo: percentil 50 = mediana).
- La dispersión (Ejemplo: rango intercuartílico).
- Valores atípicos (datos fuera de los percentiles 5 o 95).

A) VENTAJAS DE LOS CUANTILES

No requieren supuestos distribucionales: Funcionan bien con distribuciones no normales (asimétricas o con outliers). No asumen que los datos siguen una distribución específica.

Robustez ante valores extremos: A diferencia de la media o la desviación estándar, no se ven afectados por outliers. Ejemplo: El percentil 50 (mediana) es más resistente a datos atípicos que la media.

Útiles para comparar distribuciones: Permiten comparar la posición relativa de un dato en diferentes conjuntos (Ejemplo: percentiles en pruebas estandarizadas).

Flexibilidad en el análisis: Se pueden calcular en datos ordinales, y de razón. Útiles en datos discretos y continuos.

Interpretación intuitiva: Ejemplo: *"El percentil 90 indica que el 90% de los datos están por debajo de ese valor"*.

B) DESVENTAJAS DE LOS CUANTILES

Pérdida de información en datos agrupados: Si los datos están agrupados en intervalos, los cuantiles pueden ser menos precisos.

No aprovechan toda la información de los datos: Solo consideran la posición relativa, no la magnitud de las diferencias entre valores.

Cálculo menos eficiente en grandes conjuntos de datos: Requieren ordenar los datos, lo que puede ser computacionalmente costoso en Big Data.

No son útiles para operaciones algebraicas: No se pueden sumar, restar o promediar cuantiles directamente.

C) USOS PRINCIPALES DE LOS CUANTILES

Estadística descriptiva y reportes:

Resumir datos con mediana (P_{50}), rango intercuartílico ($IQR = Q_3 - Q_1$).

Ejemplo: El salario mediano en la empresa es 500; $IQR = Q_3 - Q_1$= $1,500 - $470

Detección de outliers:

Regla del $1,5 \cdot IQR$

Ejemplo: Datos fuera de $Q_1 - 1,5 \cdot IQR$; $Q_3 + 1,5 \cdot IQR$ se consideran atípicos

Medicina y ciencias de la salud:

Establecer rangos de referencia (Ejemplo: percentiles de presión arterial por edad).

Ejemplo de aplicación: Cuartiles en Salarios

Datos ordenados (salarios mensuales en $): 1200,1500,1800,2000,2200,2500,3000,3500,4000,5000

Q_1 (Percentil 25): 1800

Q_2 (Mediana, Percentil 50): 2350

Q_3 (Percentil 75): 3500

IQR (Rango intercuartílico): 3500−1800=3500−1800=1700

Interpretación:

El 25% de los empleados gana menos de $1800.

El 50% gana menos de $2350.

El 25% más alto gana más de $3500.

La mayoría de los salarios están entre $1800 *y* $3500 (IQR).

2.2) CLASIFICACIÓN

A continuación, se presenta una tabla sobre las medidas de posición.

Medida	Símbolo	División	Ejemplo de Uso
Cuartiles	Q_1 , Q_2, Q_3	Divide los datos en 4 partes iguales (25%, 50%, 75%).	$Q_1 = P_{25}$ $Q_2 =$ Mediana (50% de los datos) $Q_3 = P_{75}$
Deciles	D_1 a D_9	Divide los datos en 10 partes iguales (10%, 20%, ..., 90%).	$D_5 =$ Mediana (50% de los datos)
Percentiles	P_1 a P_{99}	Divide los datos en 100 partes iguales (1%, 2%, ..., 99%).	$P_{90} =$ Umbral alto (solo el 10% supera este valor)
Quintiles	K_1 a K_4	Divide los datos en 5 partes iguales (20%, 40%, 60%, 80%).	Usado en economía para analizar desigualdad de ingresos.

Nota:

$D_1 = P_{10}$; $D_2 = P_{20} = K_1$; $D_3 = P_{30}$; $D_4 = P_{40} = K_2$

$Md = Q_2 = D_5 = P_{50}$

$D_6 = P_{60} = K_3$; $D_7 = P_{70}$; $D_8 = P_{80} = K_4$; $D_9 = P_{90}$

2.3) MÉTODOS DE CÁLCULO

A) PARA DATOS NO AGRUPADOS

Ordenar de menor a mayor y aplicar la fórmula de posición

$$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$$

Donde:

$x = x_i$ = posición del dato (ejemplo: x_2)

k = Número del cuantil (Ejemplo: 1 para Q_1, 25 para P_{25})

n = Número total de datos

m = Número de divisiones (4 para cuartiles, 5 para quintiles, 10 para deciles, 100 para percentiles)

Ejemplo de aplicación: Calcule los percentiles 25, 50 y 75 de la siguiente distribución: 6, 9, 9, 12, 12, 15, 17 y 18

Solución: Se ordena los datos de menor a mayor

6	9	9	12	12	15	17	18
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

Percentil 25 (P_{25}):

$$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$$

$$P_k = P_{25} = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{25(n+1)}{100}\right]} = x_{\left[\frac{n+1}{4}\right]} = x_{\left[\frac{8+1}{4}\right]} = x_{\left[\frac{9}{4}\right]} = x_{2,25}$$

Interpolamos entre x_2 y x_3

$$P_{25} = 9 + 0{,}25 \cdot (9 - 9) = 9$$

Interpretación:

El 25% de los datos son ≤ 9.

Esto significa que el 25% de las notas más bajas están en el valor **9** o por debajo (pero el mínimo es 6, lo que indica que el 25% inferior incluye 6 y parte de los 9).

Percentil 50 (P_{50}):

$$P_k = P_{50} = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{50(n+1)}{100}\right]} = x_{\left[\frac{n+1}{2}\right]} = x_{\left[\frac{8+1}{2}\right]} = x_{\left[\frac{9}{2}\right]} = x_{4,5}$$

Promedio entre x_4 y x_5

$$P_{50} = \frac{x_4 + x_5}{2} = \frac{12 + 12}{2} = 12$$

Interpretación:

El 50% de los datos son ≤ 12.

Este es también el valor de la **mediana**, indicando que la mitad de las notas son 12 o menos, y la otra mitad son 12 o más.

Percentil 75 (P_{75}):

$$P_k = P_{75} = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{75(n+1)}{100}\right]} = x_{\left[\frac{75(8+1)}{100}\right]} = x_{\left[\frac{27}{4}\right]} = x_{6,75}$$

Interpolamos entre x_6 y x_7

$$P_{75} = 15 + 0,75 \cdot (17 - 15) = 16,5$$

Interpretación:

El 75% de los datos son ≤ 16,5.

Esto implica que el 25% de las notas más altas están entre 16,5 y 18.

Empleando Excel

	A	B	C	D	E	F	G	H
1	6	9	9	12	12	15	17	18
2								
3	P_{25}	9	=CUARTIL.EXC(A1:H1;1)					
4	P_{50}	12	=CUARTIL.EXC(A1:H1;2)					
5	P_{75}	16,5	=CUARTIL.EXC(A1:H1;3)					

Análisis de la Distribución

1. Rango Intercuartílico (IQR):

$$IQR = Q_3 - Q_1 = P_{75} - P_{25} = 16,5 - 9 = 7,5$$

El 50% central de los datos está entre 9 y 16,5, con una dispersión moderada.

2. Valores Atípicos:

Límite inferior:

$Q_1 - 1,5 \cdot IQR = 9 - 1,5 \cdot 7,5 = -2,25$→ Ningún dato es atípico.

Límite superior:

$Q_3 + 1,5 \cdot IQR = 16,5 + 1,5 \cdot 7,5 = 27,75$ → Ningún dato es atípico.

B) PARA DATOS EN TABLAS DE FRECUENCIAS

Ejemplo de aplicación: Calcular el decil 5 empleando los siguientes datos:

x	f
6	1
9	2
12	3
15	1
17	1

Solución:

Aplicando la primera ecuación para el cuartil dos se obtiene:

Decil 5 (D_5):

$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$

$$D_k = D_5 = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{5(n+1)}{10}\right]} = x_{\left[\frac{n+1}{2}\right]} = x_{\left[\frac{8+1}{2}\right]} = x_{\left[\frac{9}{2}\right]} = x_{4,5}$$

Como la posición del decil 5 es 4,5, su valor es el promedio de los datos cuarto y quinto

Para observar con claridad cuáles son los datos cuarto y quinto se calcula la frecuencia acumulada (F)

x	f	F	Ubicación
6	1	1	x_1
9	2	3	x_2 $hasta$ x_3
12	3	6	x_4 $hasta$ x_6
15	1	7	x_7
17	1	8	x_8

Se observa que el cuarto dato es 12 y el quinto dato es 12, por lo tanto

$$D_5 = x_{4,5=}\frac{x_4 + x_5}{2} = \frac{12 + 12}{2} = 12$$

Interpretación:

$D_5 = 12$ significa que:

El 50% de los datos son ≤ 12.

Este valor coincide con la mediana de la distribución, indicando que la mitad de los datos son ≤ 12, y la otra mitad ≥12.

C) PARA DATOS AGRUPADOS EN INTERVALOS

$$Cuantil = L_i + \left(\frac{\frac{nk}{m} - F_{i-1}}{f_i} \right) \cdot A_i$$

Donde:

$Cuantil$: $Cuartiles\ Q_k, Quintiles\ K_k, Deciles\ D_k, Percentiles\ P_k$
L_i = Límite inferior del intervalo del cuantil
n = Número total de datos
k = Número del cuantil (Ejemplo: 1 para Q_1, 25 para P_{25})
m = Número de divisiones (4 para cuartiles, 5 para quintiles, 10 para deciles, 100 para percentiles)
F_{i-1} = Frecuencia acumulada del intervalo que antecede al intervalo del cuantil
f_i = Frecuencia absoluta del intervalo del cuantil
A_i = Amplitud o Ancho del intervalo

Ejemplo de aplicación: Calcular el percentil 30 de los siguientes datos:

Intervalos	*f*	*F*
[10-20)	4	4
[20-30)	8	12
[30-40)	6	18

Donde:
$F = frecuencia\ acumulada$

Solución:

Se calcula la ubicación del percentil 30

$$\frac{nk}{m} = \frac{18 \cdot 30}{100} = 5{,}4$$

Por lo tanto, el intervalo de donde está ubicado el percentil 30 es [20-30). Al aplicar la ecuación se obtiene:

$$P_{30} = Cuantil = L_i + \left(\frac{\frac{nk}{m} - F_{i-1}}{f_i} \right) \cdot A_i = 20 + \left(\frac{\frac{18 \cdot 30}{100} - 4}{8} \right) \cdot 10 = 21{,}75$$

Interpretación:

$P_{30} = 21{,}75$ indica que:

El 30% de los datos son ≤ 21,75.

En contexto, el 30% de las observaciones (como calificaciones, ingresos, etc.) están por debajo de 21,75 dentro del rango [20-30).

2.3) DIAGRAMA DE CAJA Y BIGOTES (BOX PLOT)

El diagrama de caja y bigotes (o *box plot*) es una representación gráfica que resume la distribución de un conjunto de datos numéricos a través de sus cuartiles, mostrando la mediana, la dispersión y posibles valores atípicos (*outliers*).

Fue introducido por el matemático John Tukey en 1977 y es ampliamente utilizado en estadística descriptiva.

A) CAJA

Caja (Box): Representa el rango intercuartílico ($IQR = Q_3 - Q_1$), que abarca desde el primer cuartil (Q_1, percentil 25%) hasta el tercer cuartil (Q_3, percentil 75%).

Dentro de la caja, una línea marca la mediana (Q_2, percentil 50%).

B) BIGOTES

Bigotes (Whiskers): Líneas que se extienden desde la caja hasta los valores mínimo y máximo dentro de un rango aceptable (generalmente $1{,}5 \cdot IQR$ por encima de Q_3 o por debajo de Q_1).

Límite inferior: $Q_1 - 1{,}5 \cdot IQR$

Límite superior: $Q_3 + 1{,}5 \cdot IQR$

C) VALORES ATÍPICOS

Son valores discordantes o raros llamados outliers (aislados). Son puntos que quedan fuera de los bigotes, indicando datos inusuales.

D) EJEMPLOS DE APLICACIÓN

1) Elaborar un Diagrama de Caja y Bigotes con los siguientes datos: 12, 15, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 36

Solución: Ordenando los datos y codificando

12	15	17	18	19	20	21	22	23	24	25	30	36
x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}

Cuartil 1 (Q_1):

$$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$$

$$Q_k = Q_1 = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{1(n+1)}{4}\right]} = x_{\left[\frac{n+1}{4}\right]} = x_{\left[\frac{13+1}{4}\right]} = x_{\left[\frac{14}{4}\right]} = x_{3,5}$$

Promedio entre x_3 y x_4

$$Q_1 = \frac{x_3 + x_4}{2} = \frac{17 + 18}{2} = 17{,}5$$

Cuartil 2 (Q_2):

$$Q_k = Q_2 = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{2(n+1)}{4}\right]} = x_{\left[\frac{n+1}{2}\right]} = x_{\left[\frac{13+1}{2}\right]} = x_{\left[\frac{14}{2}\right]} = x_7$$

$$Q_2 = x_7 = 21$$

Cuartil 3 (Q_3):

$$Q_k = Q_3 = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{3(n+1)}{4}\right]} = x_{\left[\frac{3(13+1)}{4}\right]} = x_{\left[\frac{42}{4}\right]} = x_{10,5}$$

Promedio entre x_{10} y x_{11}

$$Q_3 = \frac{x_{10} + x_{11}}{2} = \frac{24 + 25}{2} = 24,5$$

Calculando el Rango Intercuartílico:

$$IQR = Q_3 - Q_1 = 24,5 - 17,5 = 7$$

Límites para Bigotes:

Límite inferior: $Q_1 - 1,5 \cdot IQR = 17,5 - 1,5 \cdot 7 = 7$

Valor dentro del rango: **12** (mayor que 7; el bigote llega a 12)

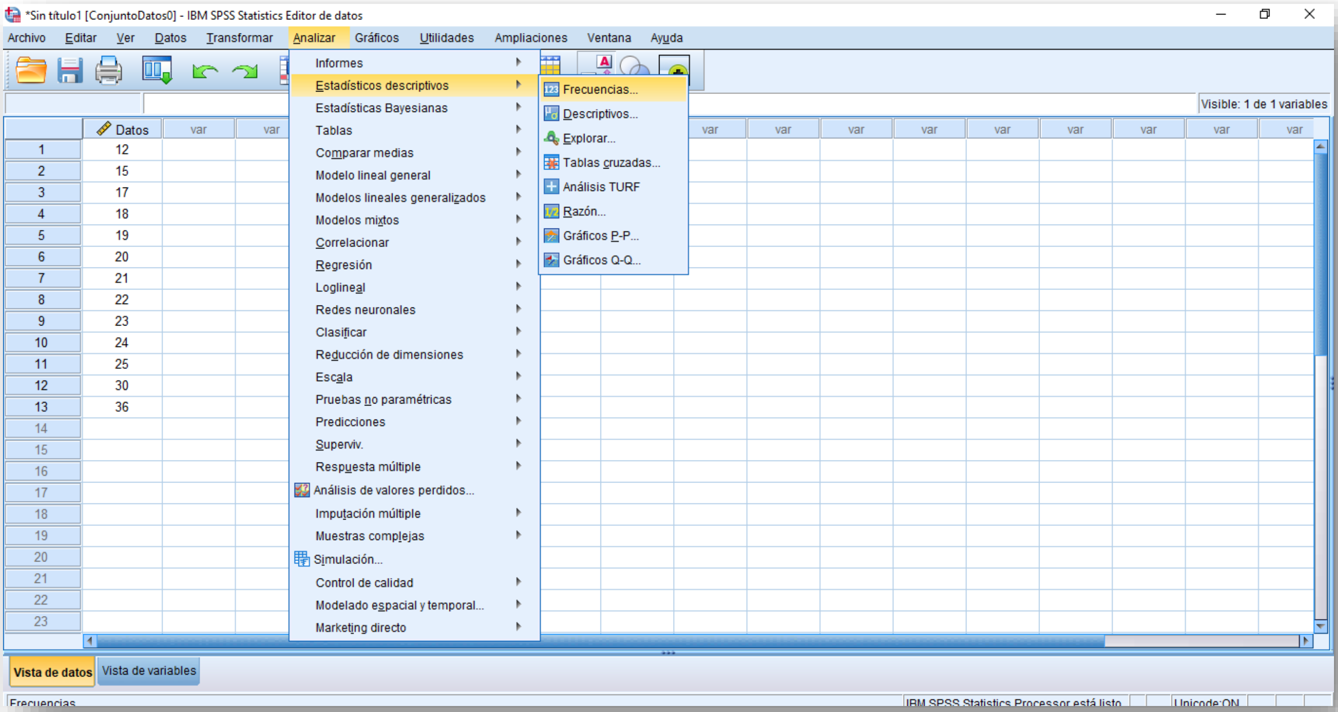
Límite superior: $Q_3 + 1,5 \cdot IQR = 24,5 + 1,5 \cdot 7 = 35$

Valor máximo dentro del rango: **30** (menor que 35; el bigote llega a 30)

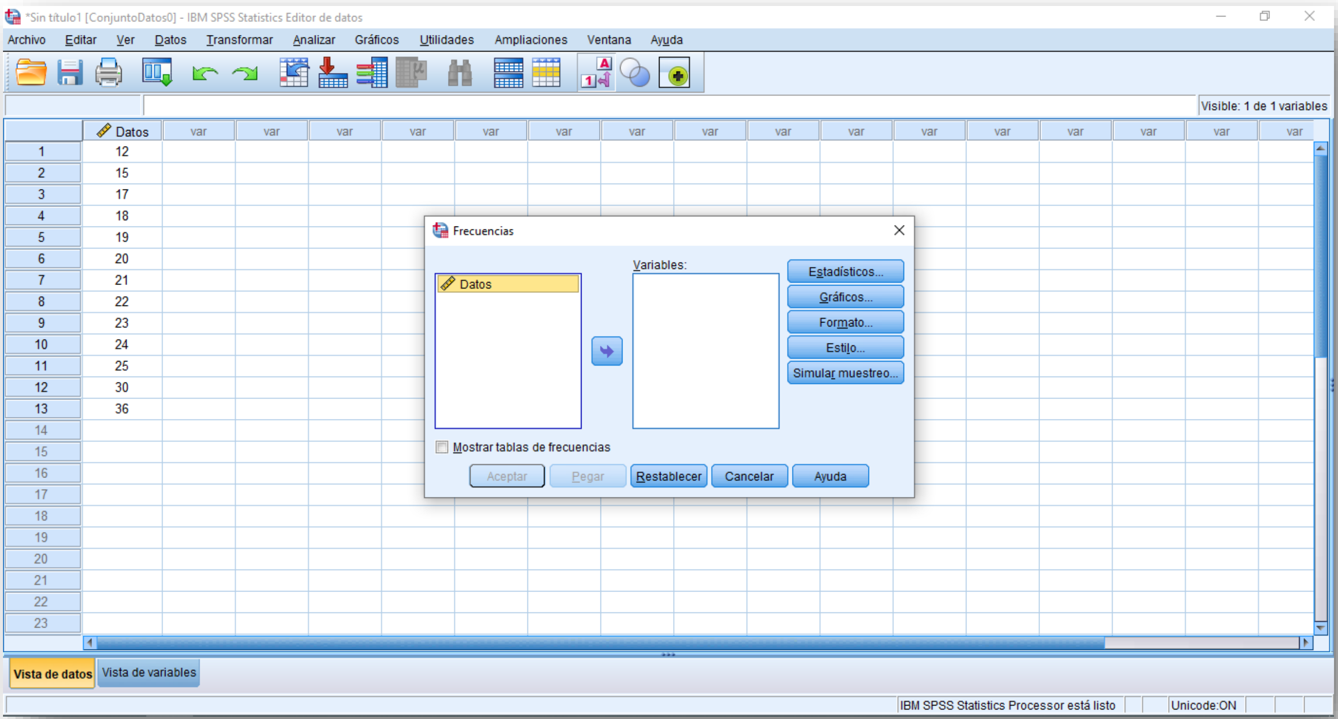
Outliers: 36 (supera el límite superior de 35, es decir, marcado como punto fuera del bigote superior).

En SPSS

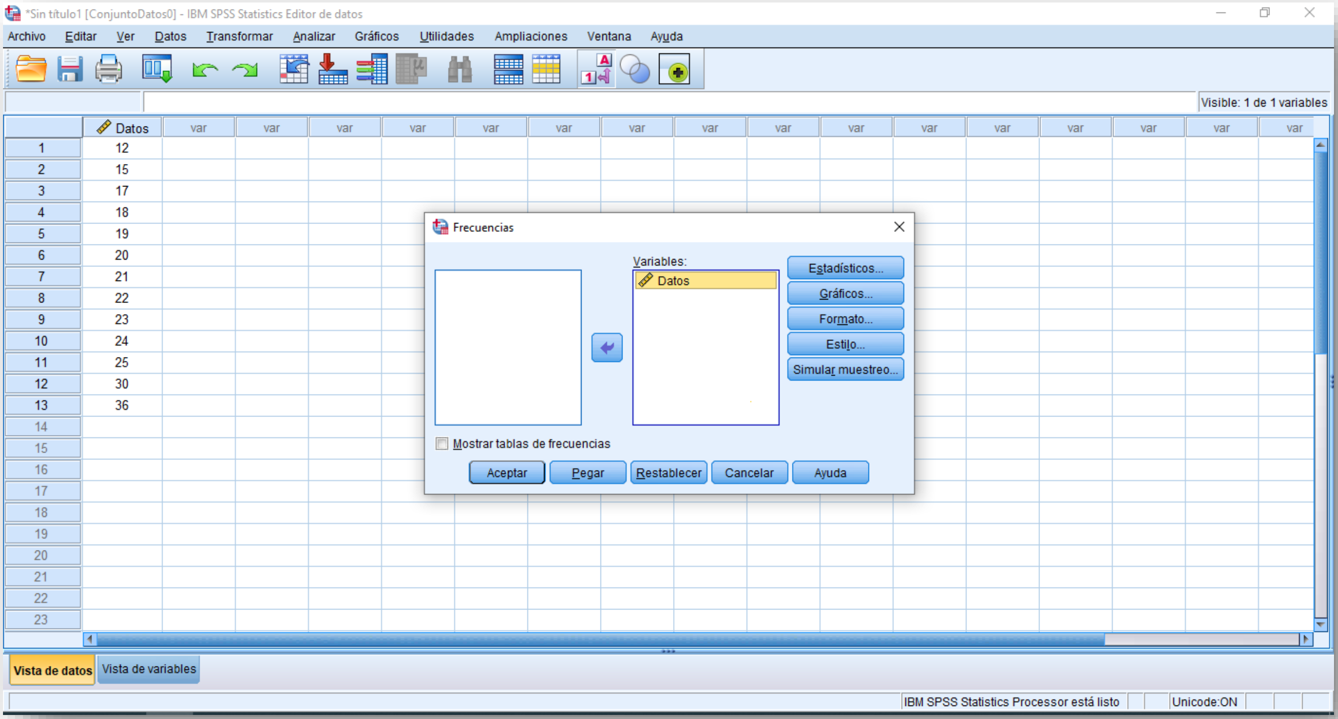
Para calcular los cuartiles. Clic en Analizar. Frecuencias.



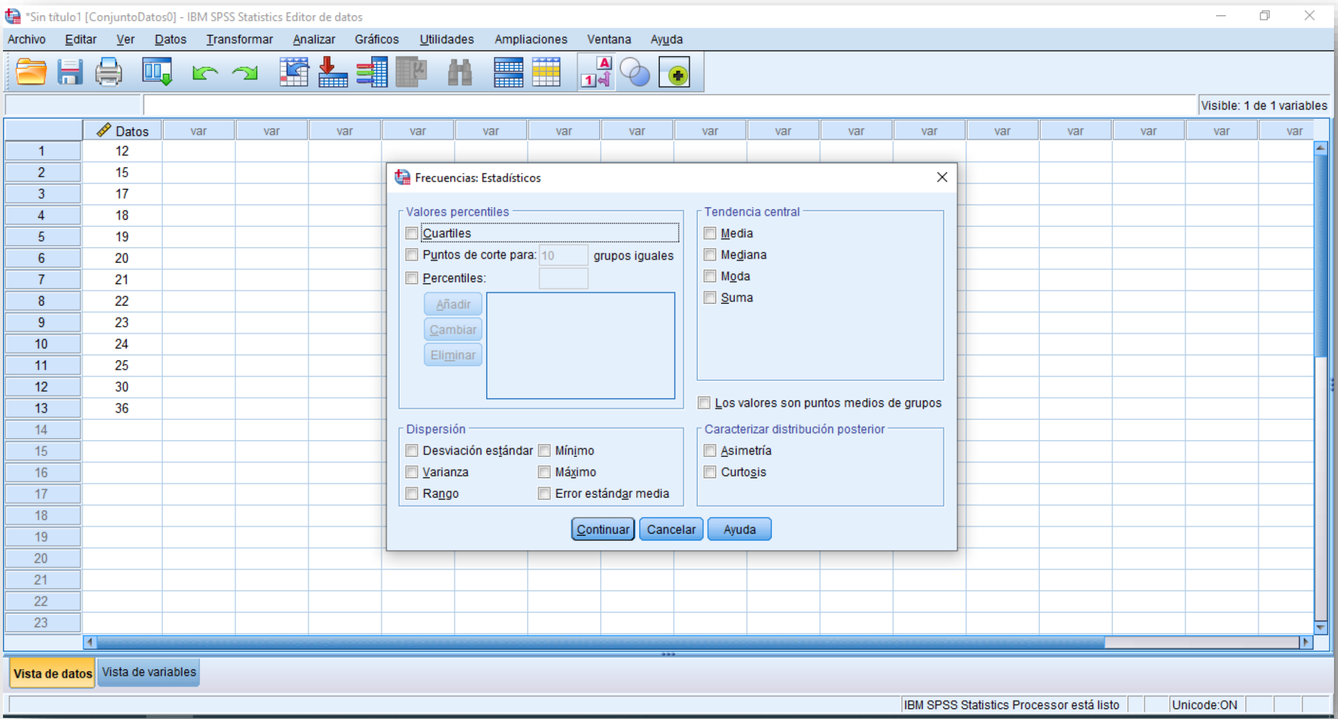
Clic en Frecuencias



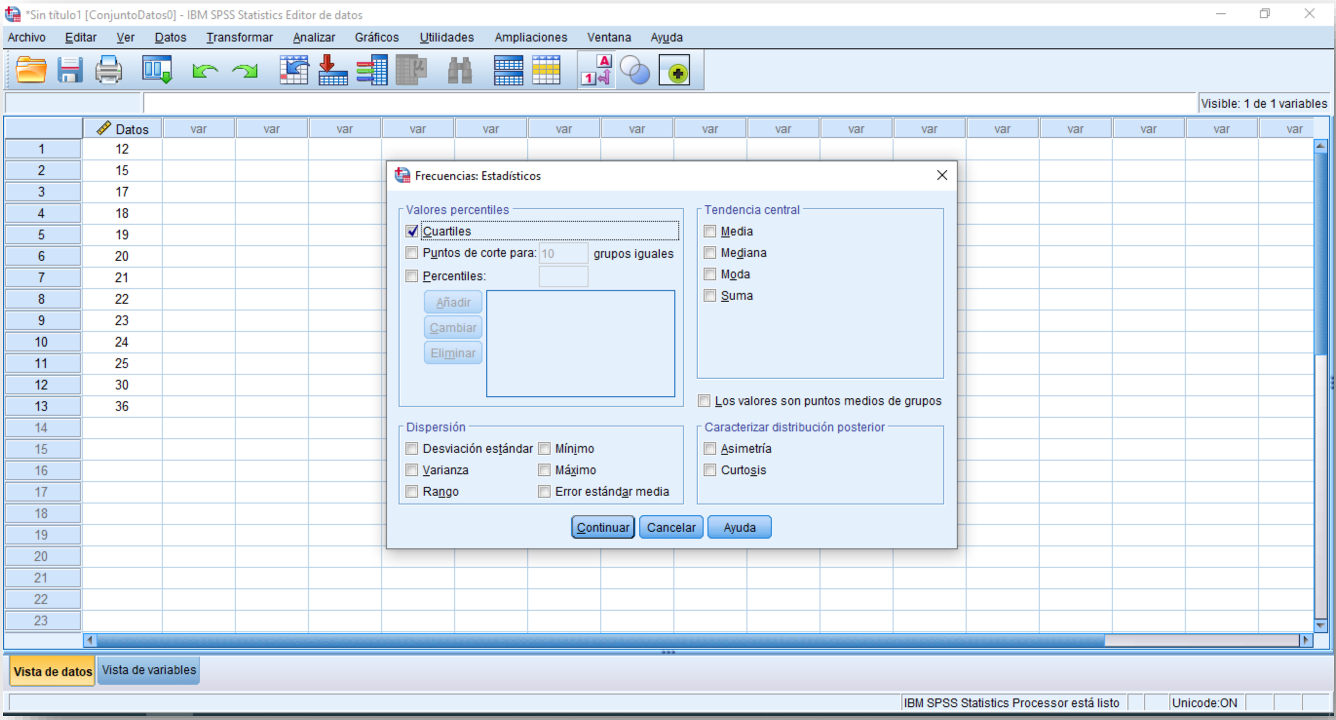
Clic en la fecha.



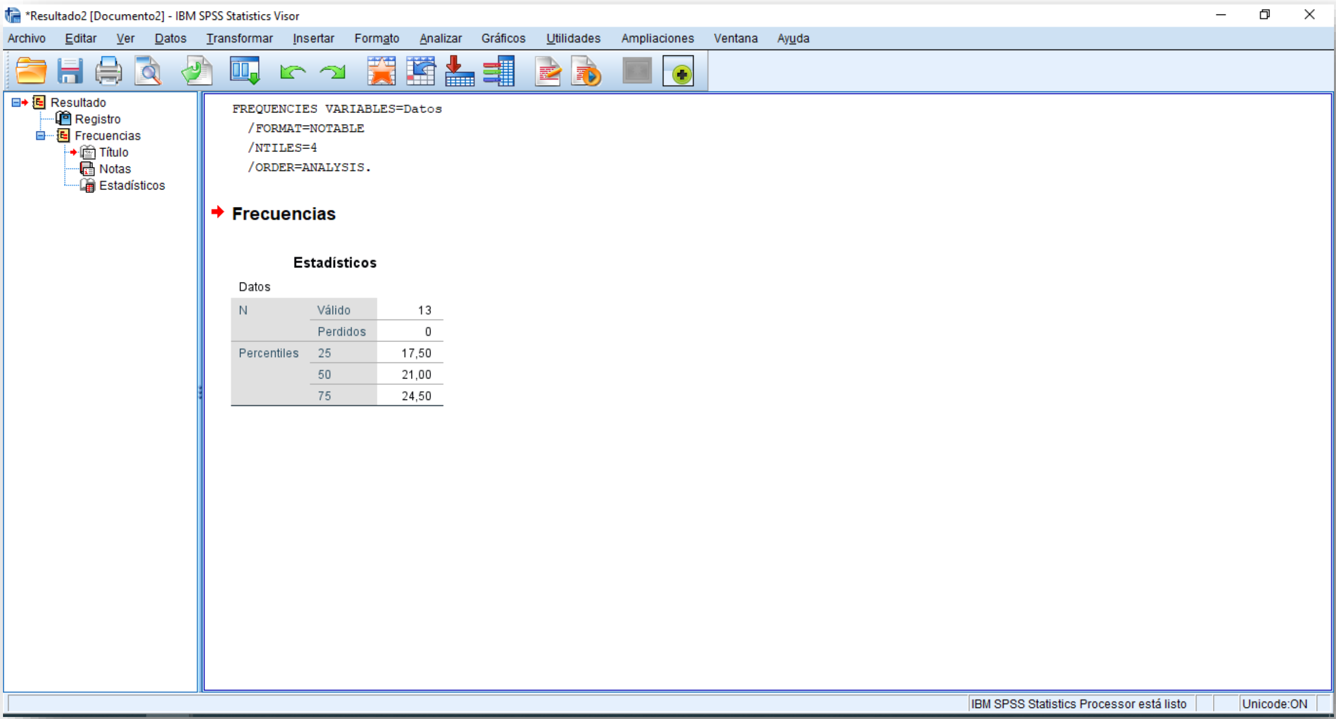
Clic en Estadísticos.



Clic en Cuartiles

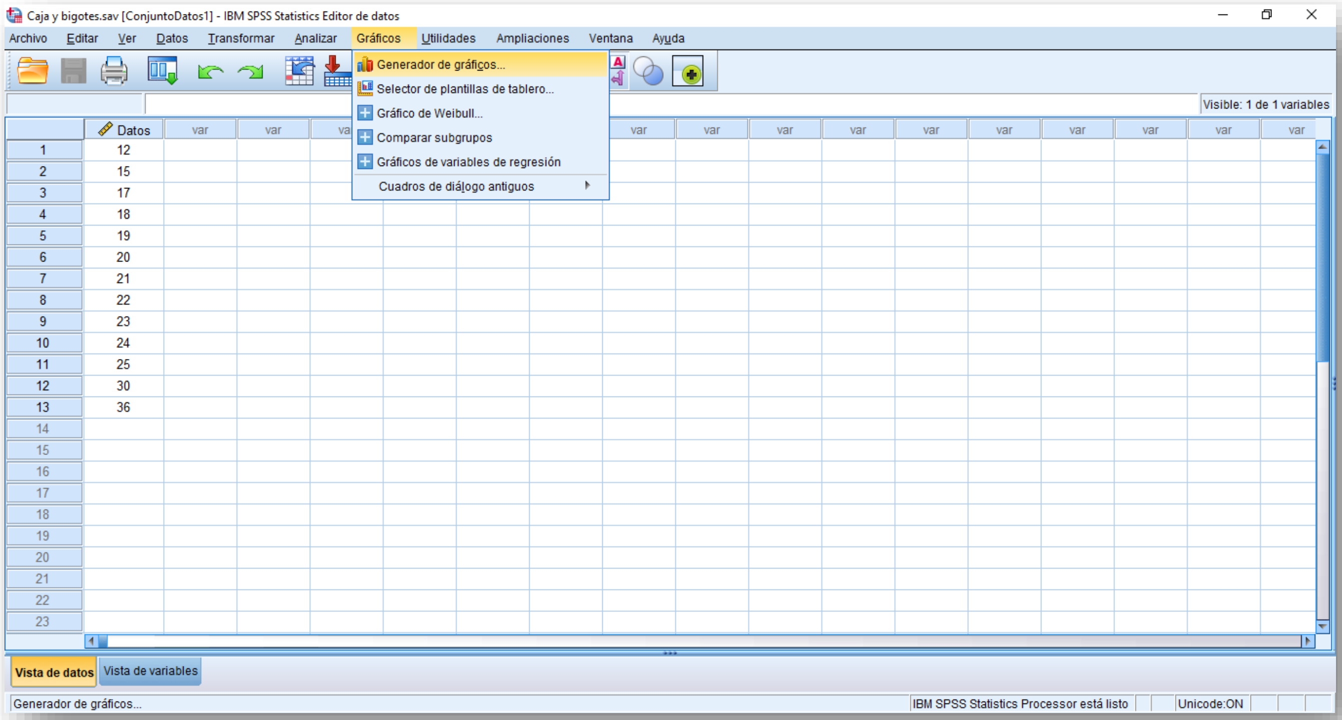


Clic en Continuar. Clic en Aceptar

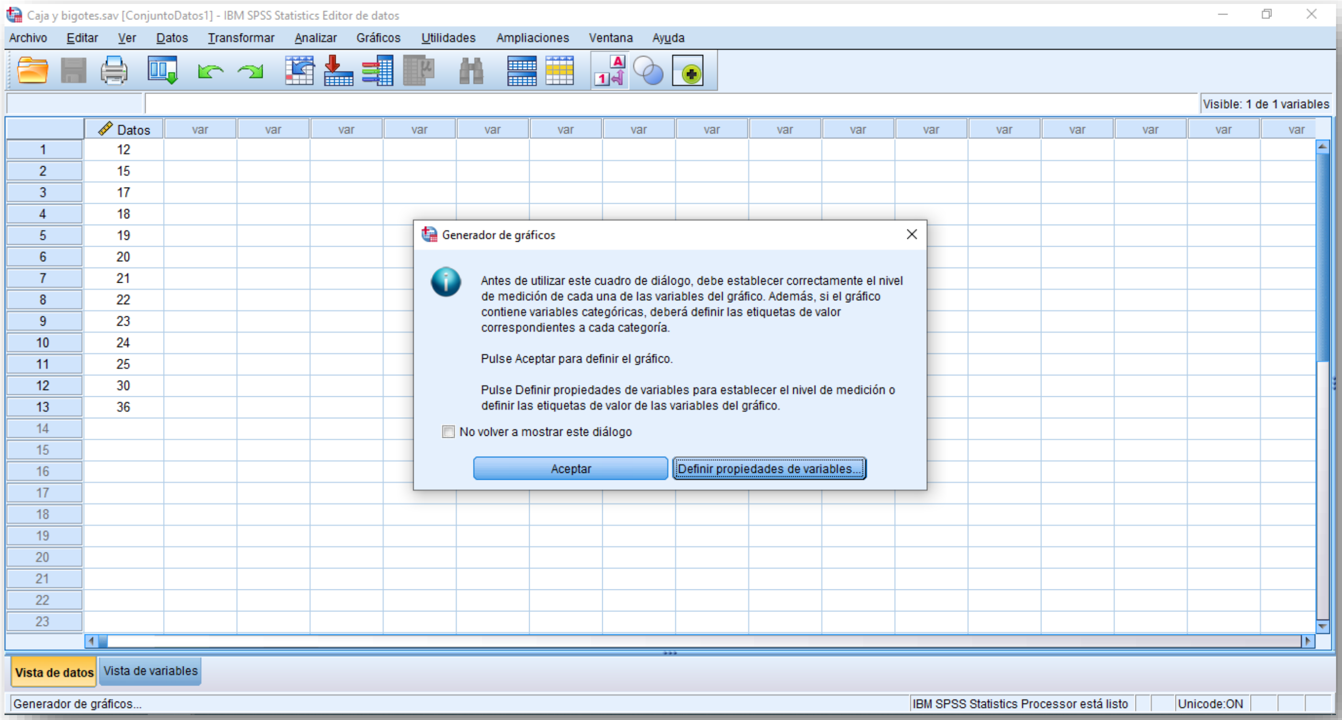


Para graficar el diagrama de Caja y Bigotes

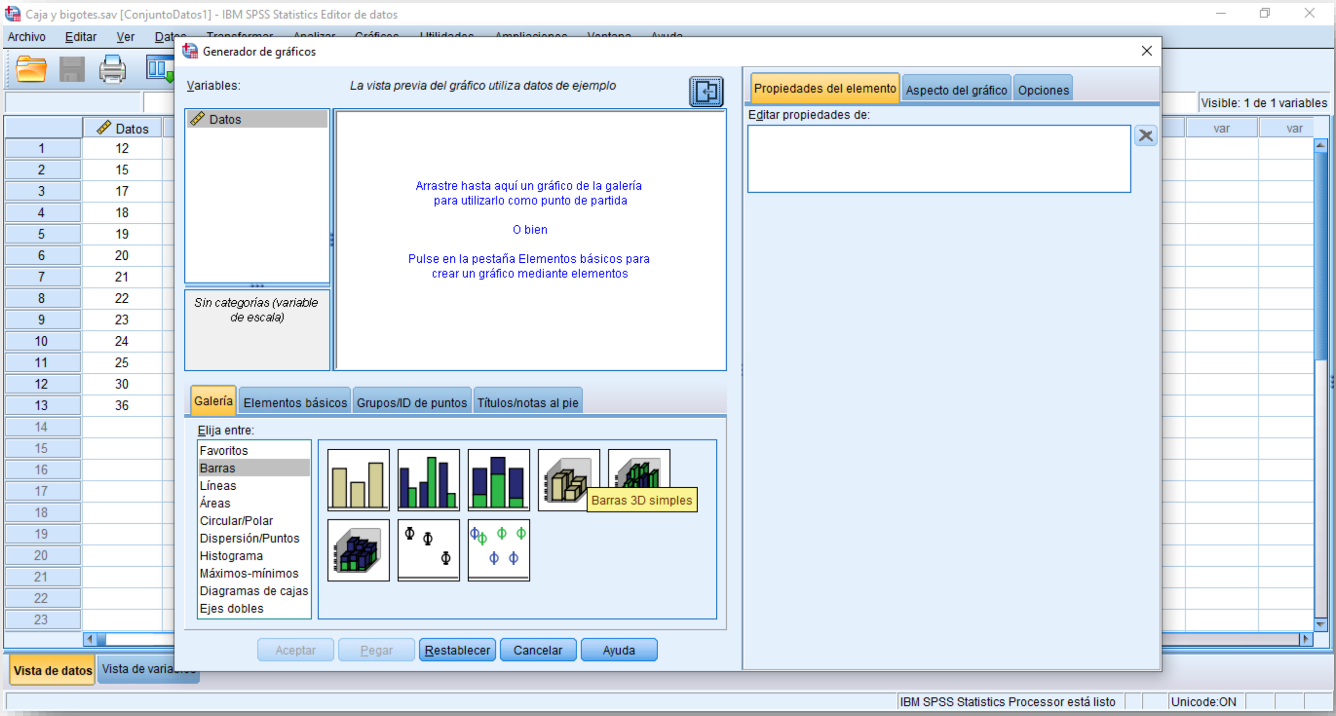
Clic en Gráficos. Generador de gráficos.



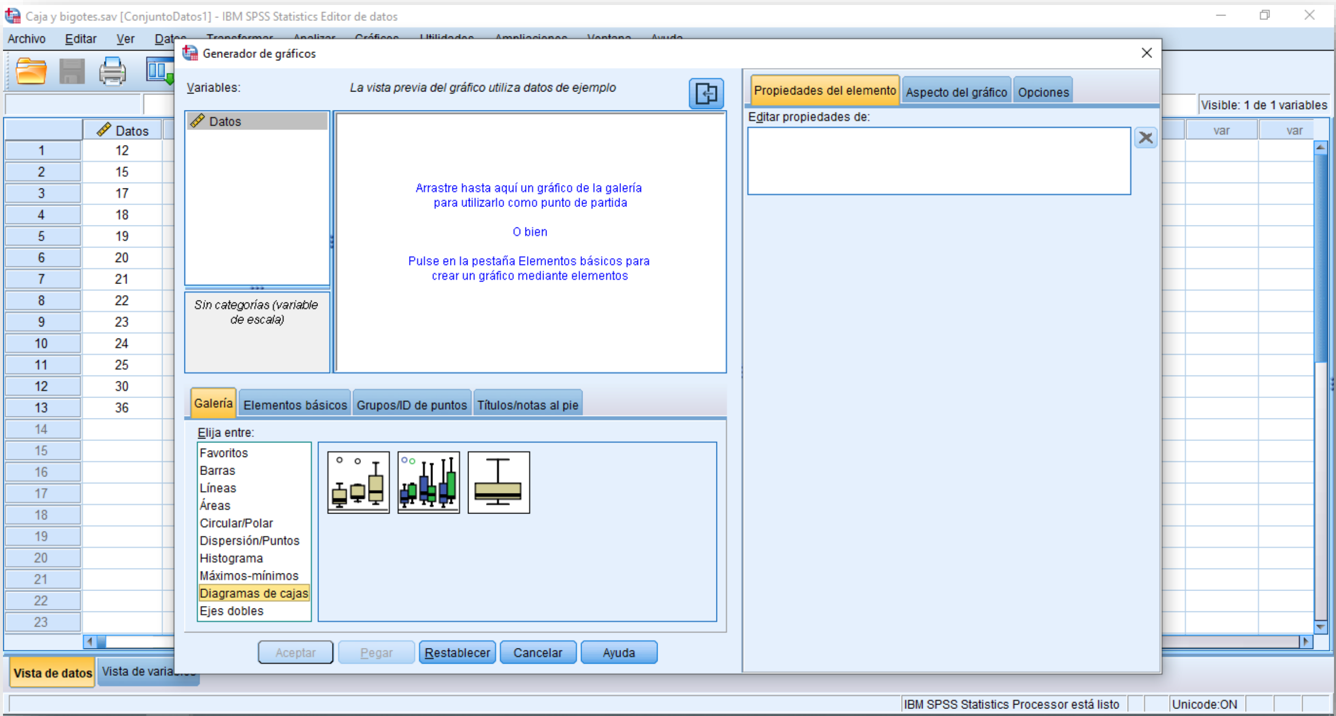
Clic en Generador de gráficos.



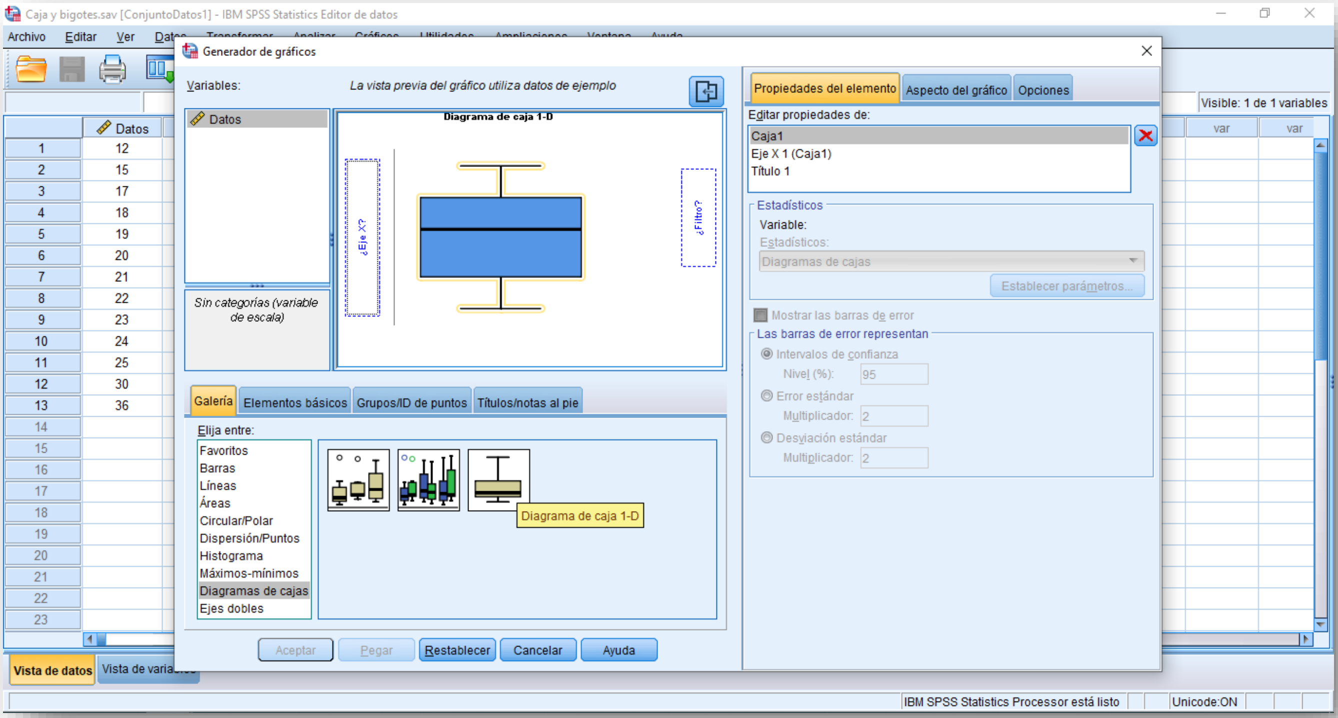
Clic en Aceptar



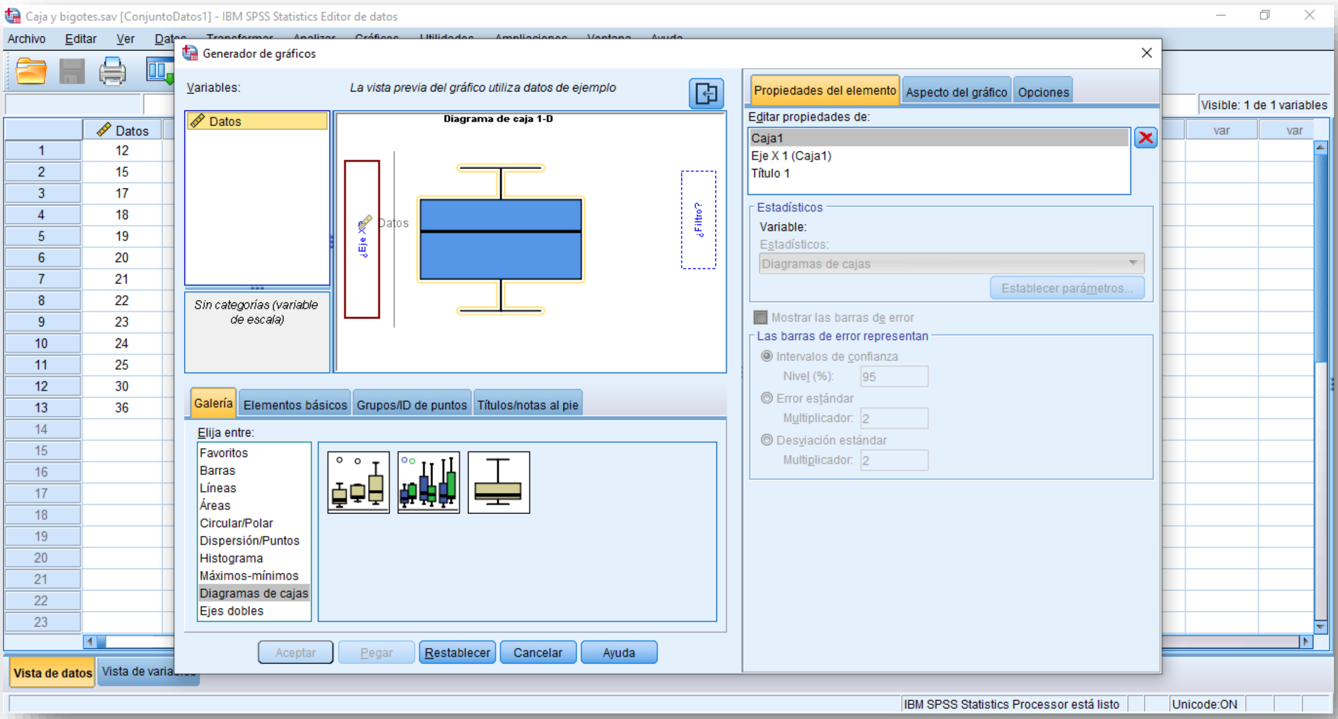
Clic en Diagrama de cajas



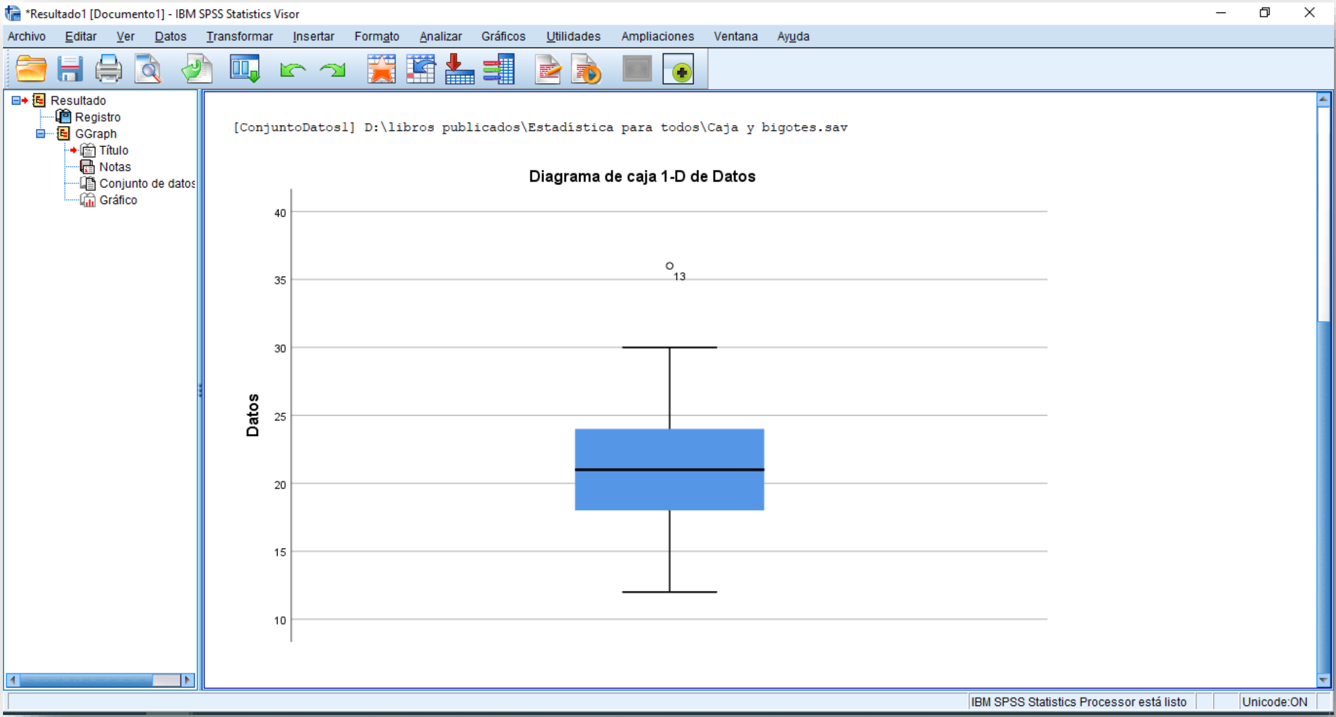
Arrastre el gráfico Diagrama de caja 1-D a la parte superior



Arrastre la variable Datos al Eje del gráfico



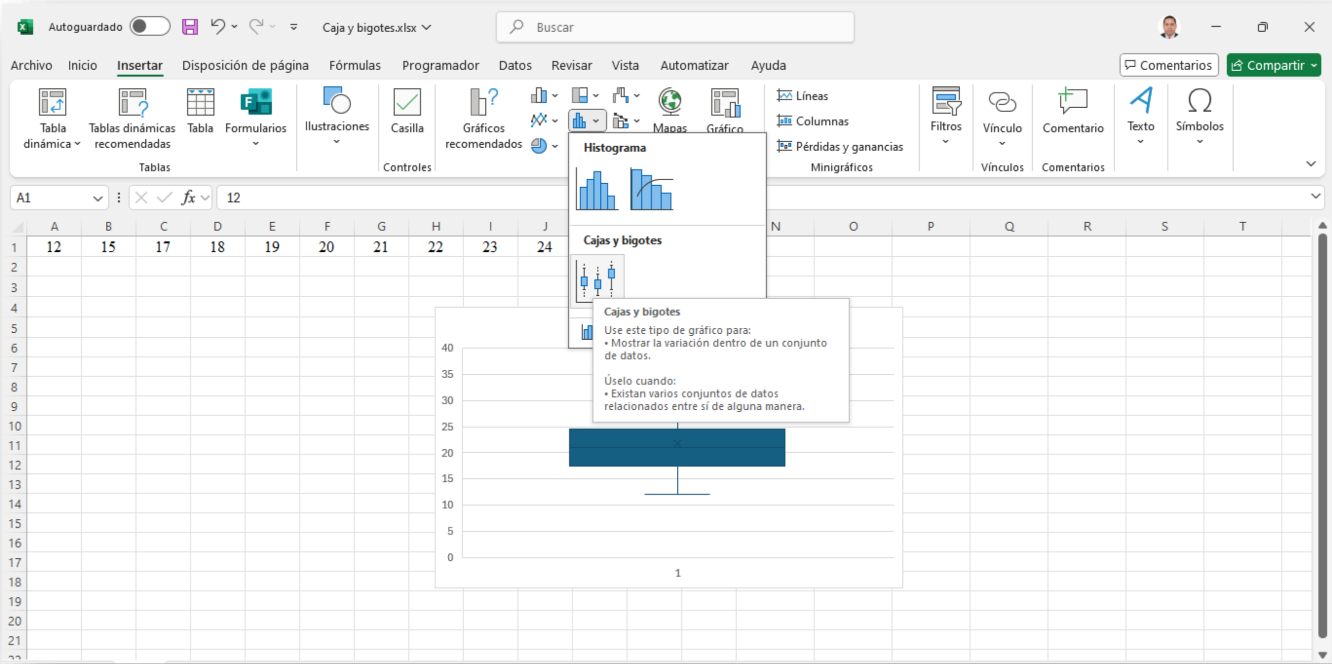
Clic en Aceptar



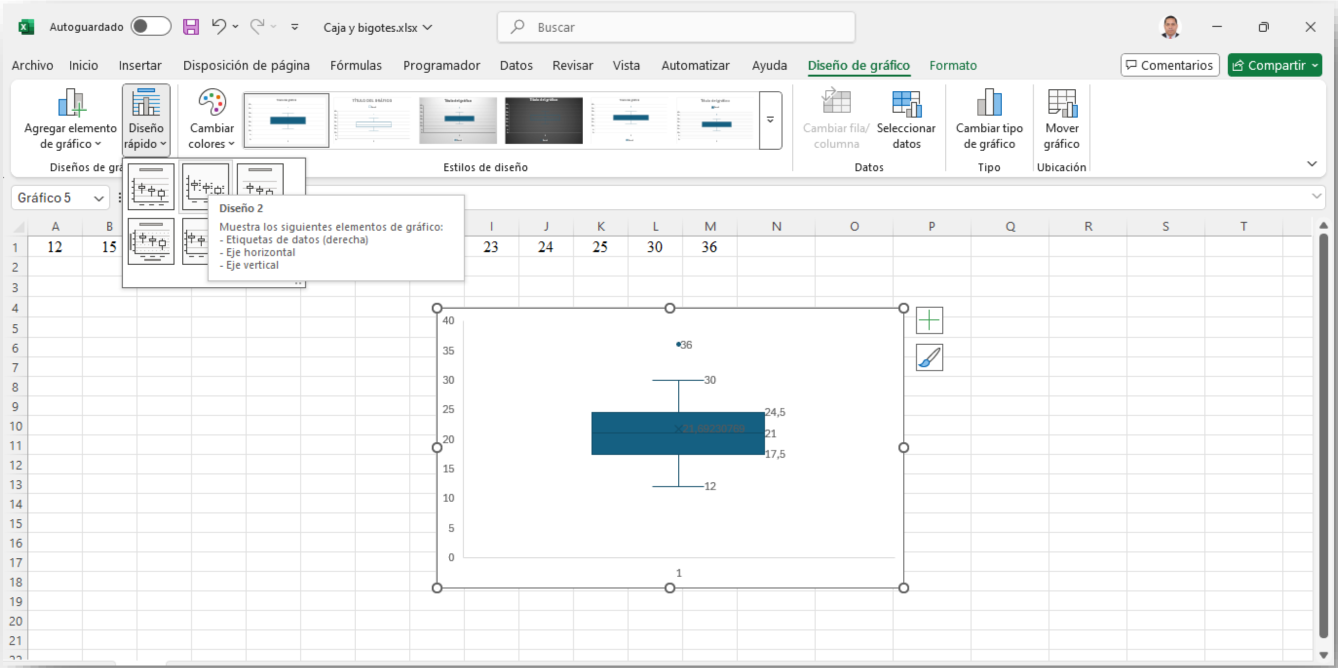
Interpretación: El dato N° 13 es un valor atípico, que este ejemplo corresponde al valor de 36

En Excel:

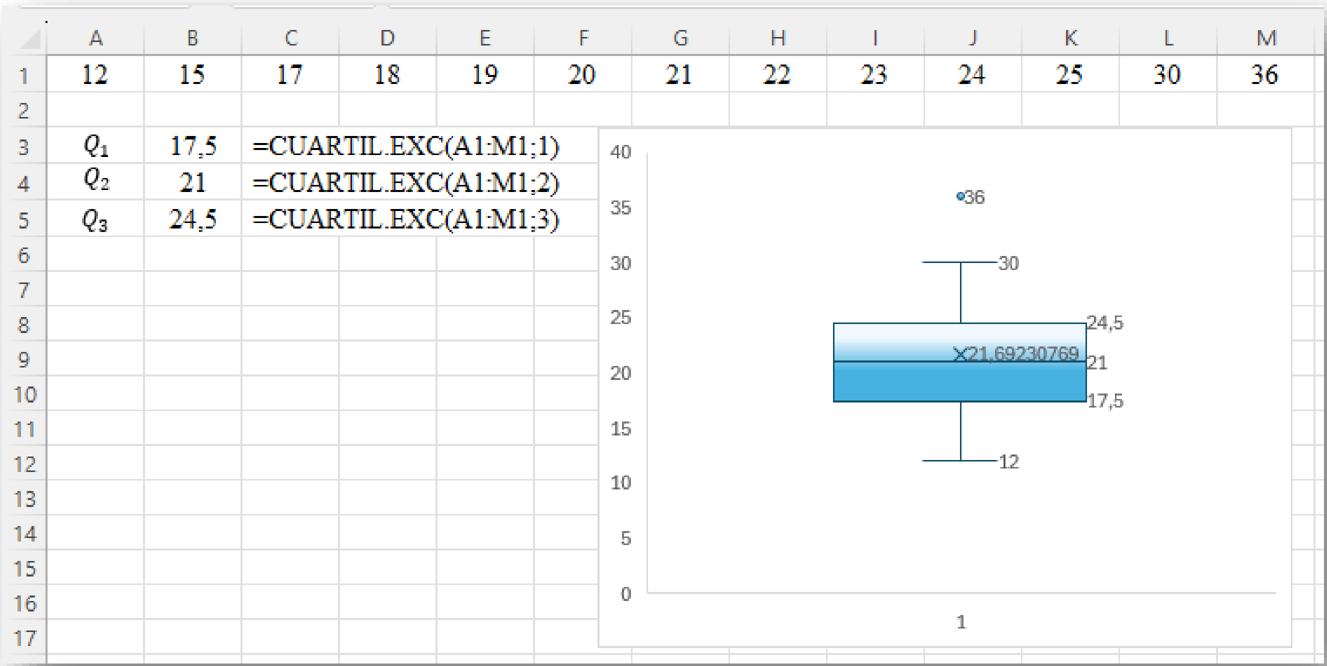
Seleccione los datos. Clic en Insertar. Clic en Cajas y bigotes.



Clic en Diseño rápido. Seleccione Diseño 2



Seguir realizando la edición del gráfico. Calcule los cuartiles.



Interpretación:

Dispersión: El IQR (7) indica que el 50% central de los datos está entre 17,5 y 24,5.
Los bigotes abarcan desde 12 hasta 30, mostrando una dispersión moderada.

Outlier: El valor 36 es un dato atípico, posiblemente debido a un error de medición, una anomalía o un evento excepcional.

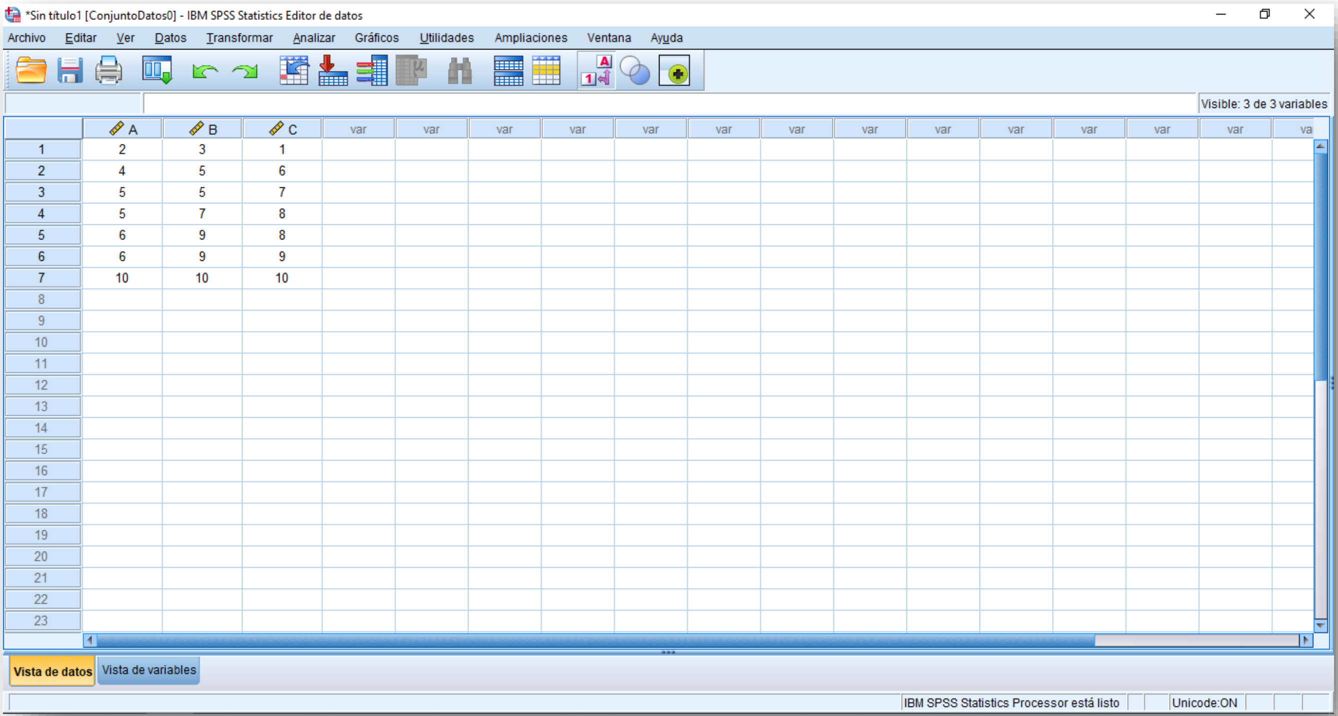
2) Elaborar un diagrama de caja y bigotes con los siguientes datos que corresponde a calificaciones (escala de 0 a 10) de Matemática en 3 grupos. Las calificaciones son de un mismo examen. No hubo condiciones diferentes entre grupos.

Grupo A	Grupo B	Grupo C
2	3	1
4	5	6
5	5	7
5	7	8
6	9	8
6	9	9
10	10	10

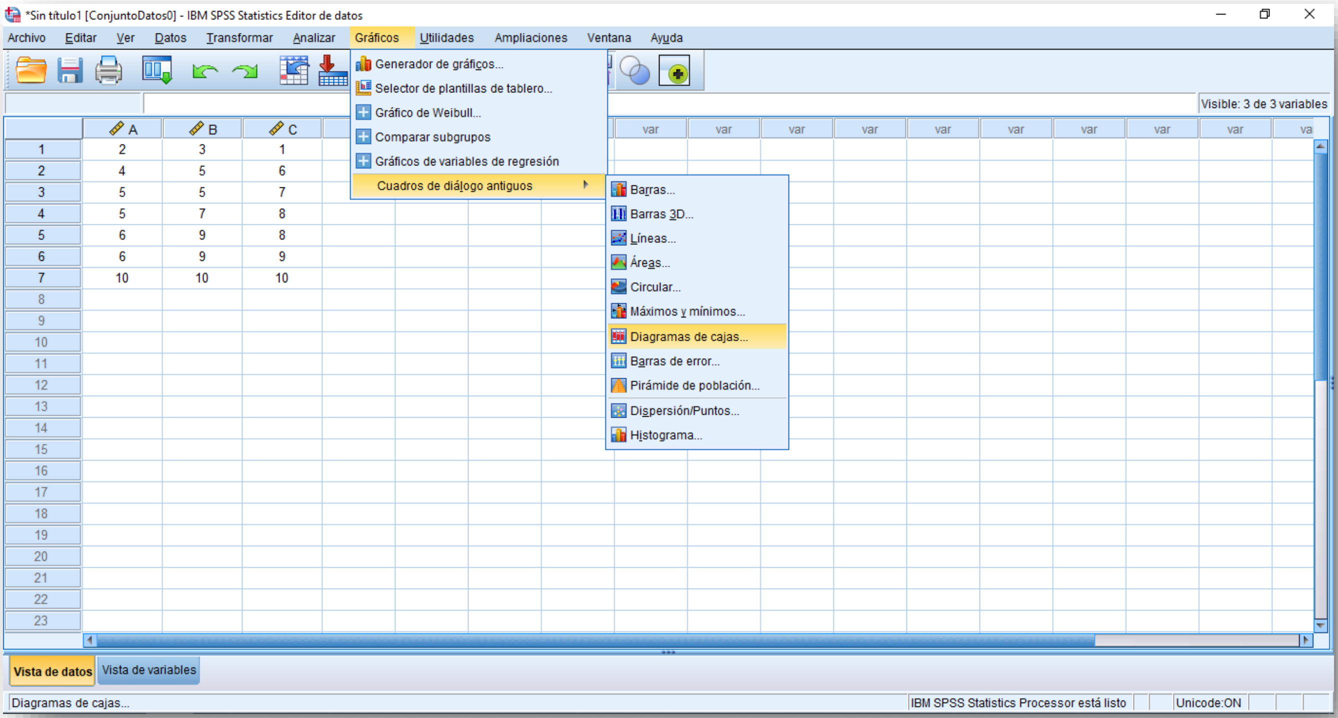
Solución:

En SPSS

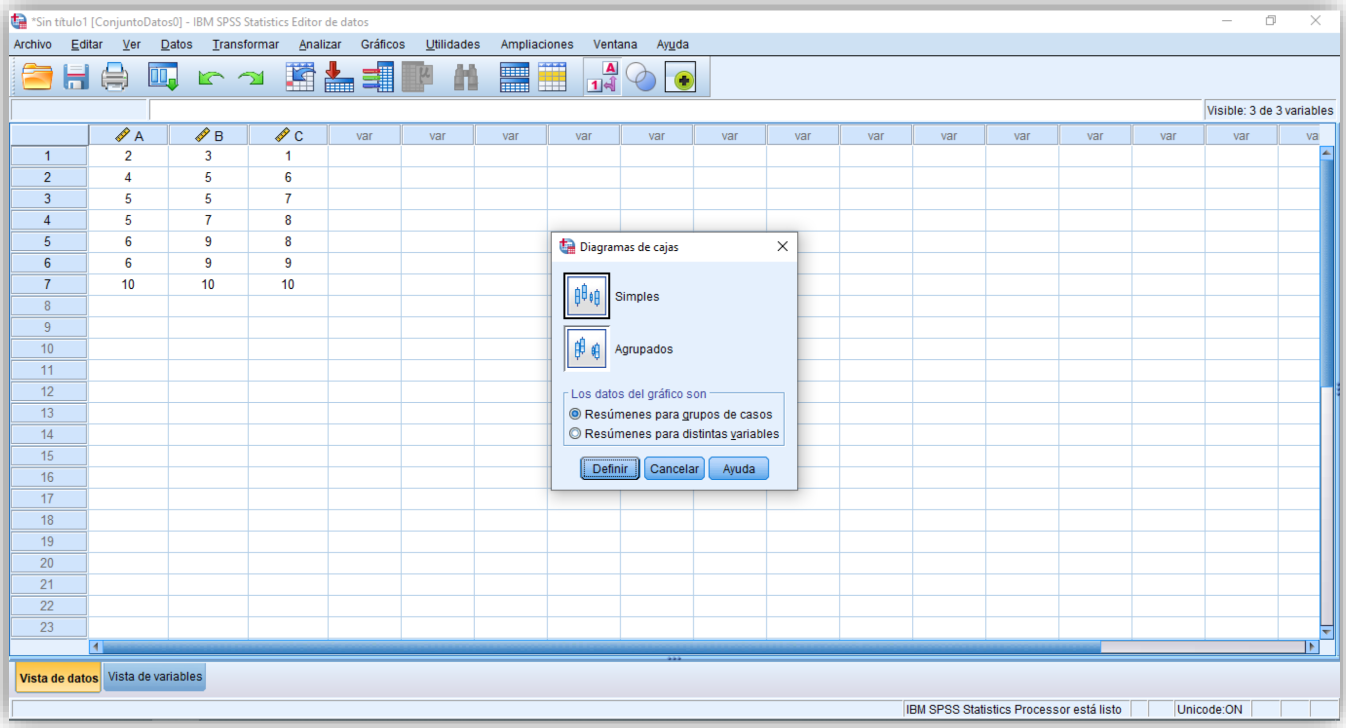
Ingresar los datos



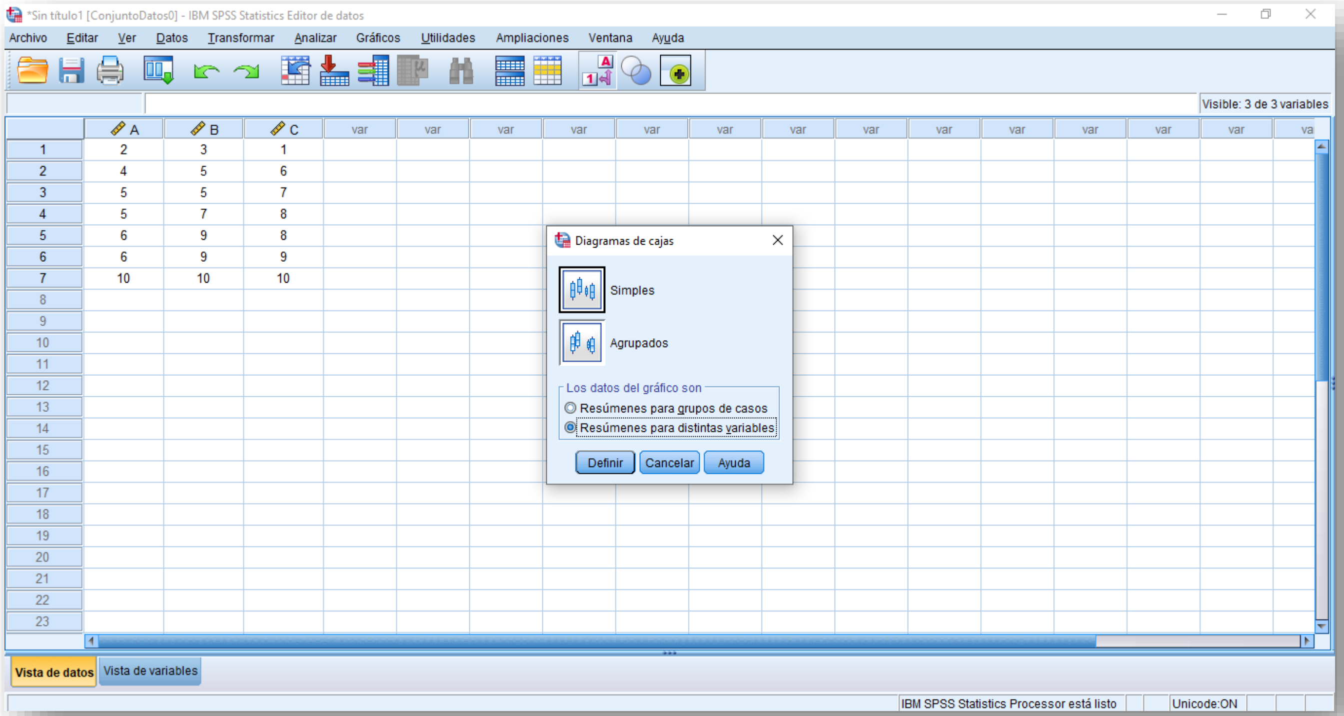
Clic en Gráficos, Cuadros de diálogo antiguos, Diagramas de cajas



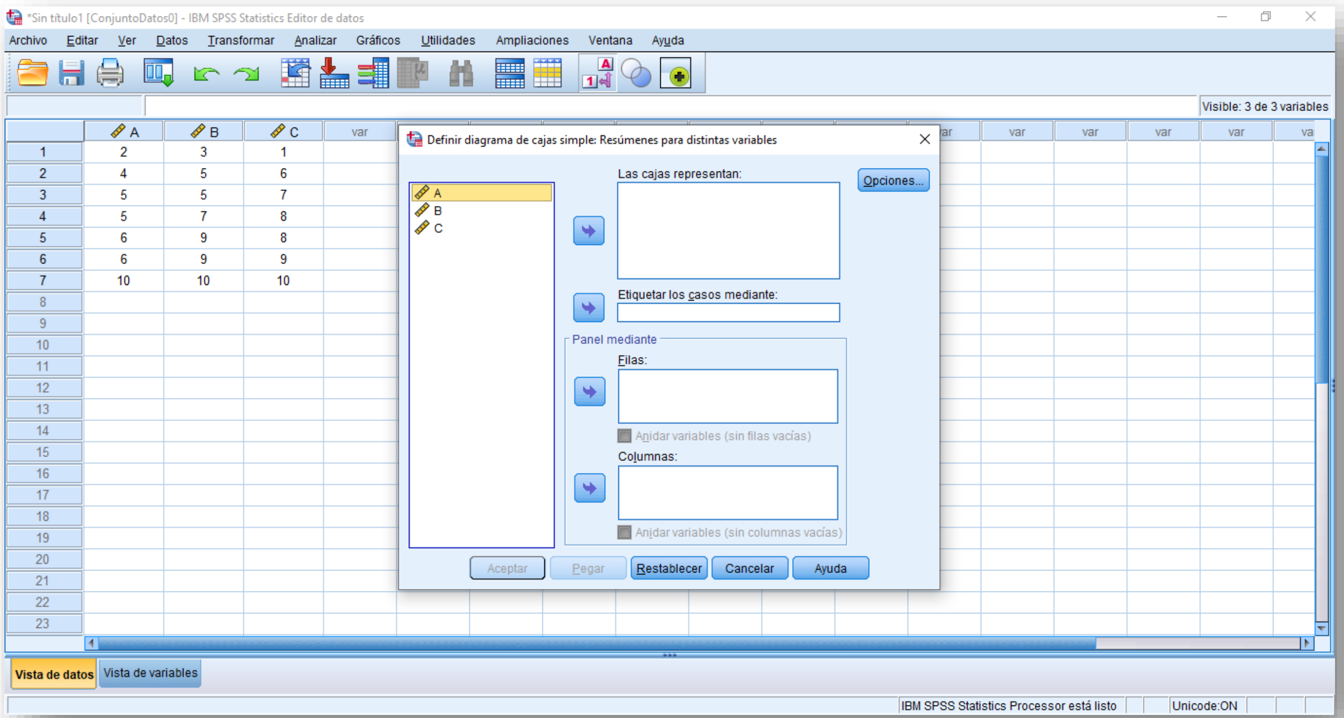
Clic en Diagramas de cajas



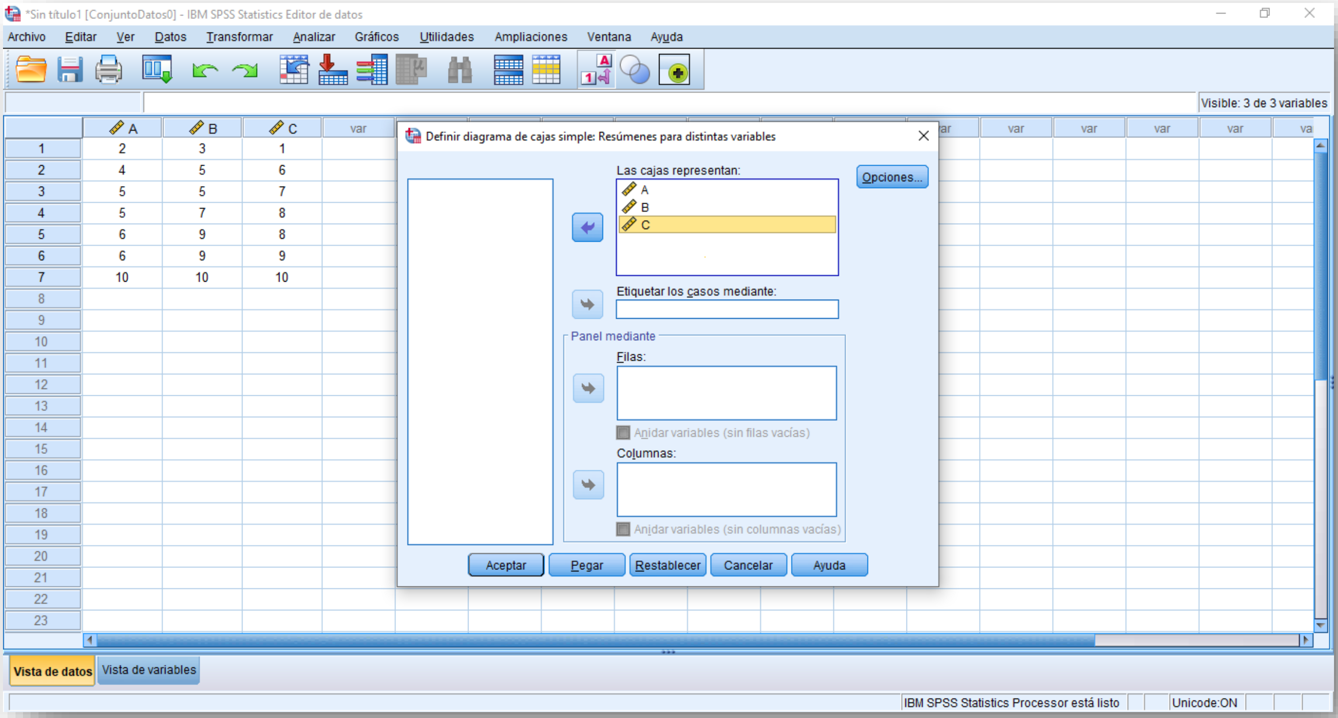
Clic en Resúmenes para distintas variables



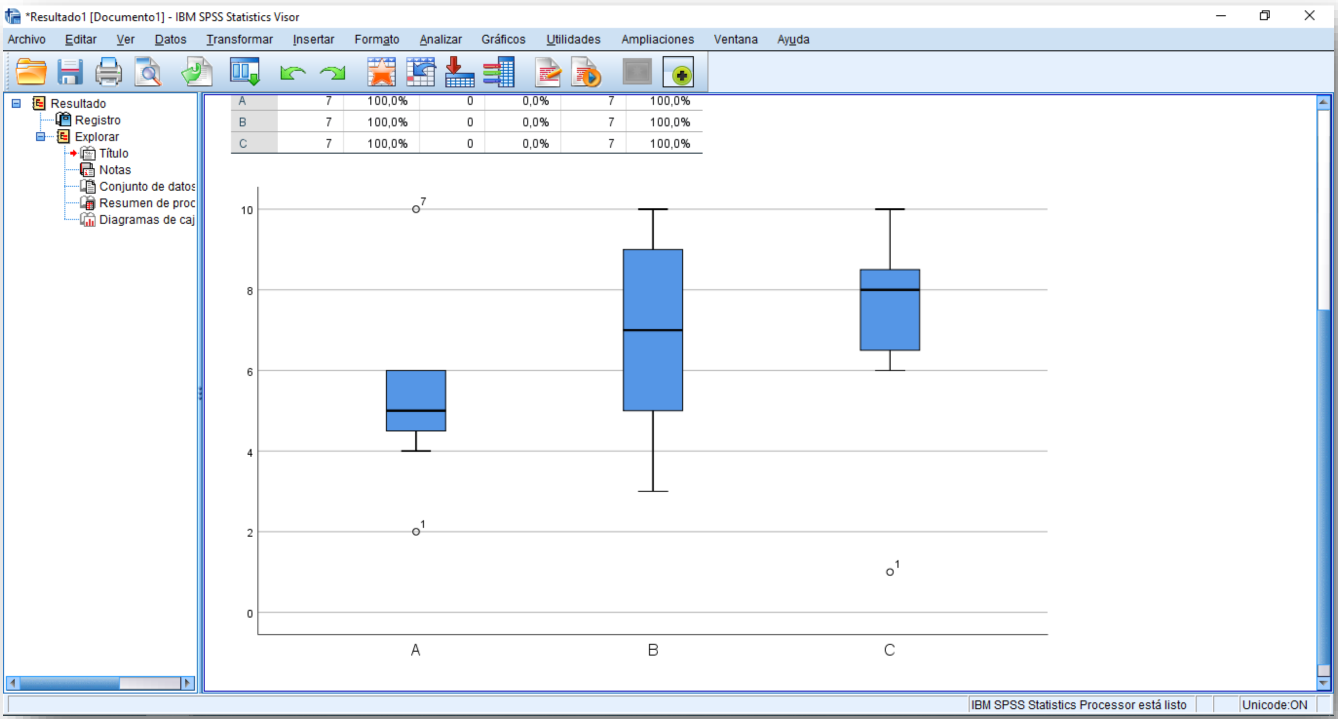
Clic en Definir



Seleccionar A y clic en la fecha a la derecha en Las cajas representan. Lo mismo con B y C.



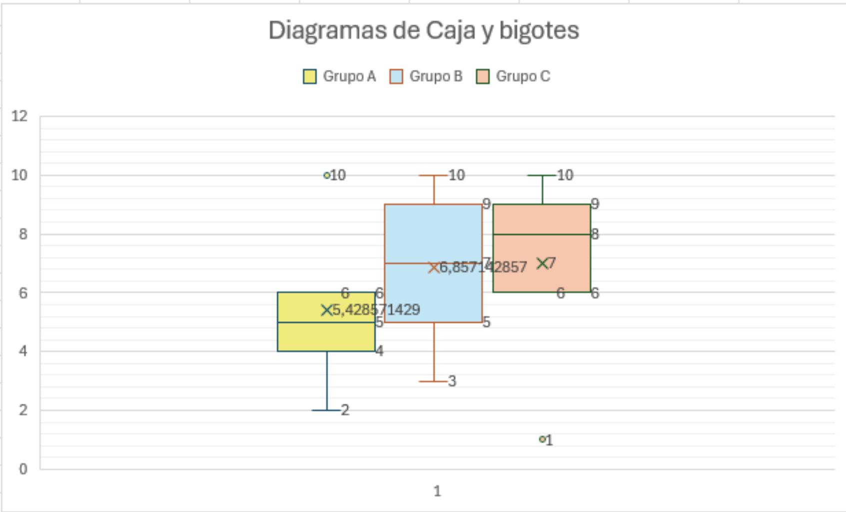
Clic en Aceptar



En Excel:

Seleccione las 3 columnas. Inserte Gráfico. Realice la ediciones y cálculos respectivos.

	A	B	C	D
1	Grupo A	Grupo B	Grupo C	
2	2	3	1	
3	4	5	6	
4	5	5	7	
5	5	7	8	
6	6	9	8	
7	6	9	9	
8	10	10	10	
9				
10	4	5	6	Q_1
11	5	7	8	Q_2
12	6	9	9	Q_3
13	2	4	3	IQR
14	5,43	6,86	7,00	Media
15	2	3	1	x_{min}
16	10	10	10	$x_{máx}$
17	1	-1	1,5	Límite inferior
18	9	15	13,5	Límite superior
19				



Interpretación:

Estadístico	Grupo A	Grupo B	Grupo C	Observación
Q_1 (25%)	4,5	5	6,5	Grupo C tiene el Q_1 más alto
Mediana (Q_2)	5,5	7	8	Grupo C destaca con mejor desempeño
Q_3 (75%)	6,75	9	9	Grupos B y C comparten Q_3
IQR	2,25	4	2,5	Grupo B es el más disperso
Media	5,43	6,86	7,0	Coincide con la mediana en Grupo C
Mínimo	2	3	1	Outlier en Grupo C (1)
Máximo	10	10	10	Máximo común en todos
Outliers	10*	Ninguno	1	*10 en A está en el límite

Hallazgos clave:

El Grupo C muestra el mejor rendimiento general (mediana=8) pero con un caso extremo preocupante (nota 1)

El Grupo B tiene la distribución más equilibrada (mediana=7) pero con mayor variabilidad

El Grupo A presenta el rendimiento más bajo (mediana=5,5) con posible caso de dificultad extrema (nota 2)

Análisis detallado por grupo:

Grupo A (Rendimiento bajo)

50% de estudiantes obtuvo entre 4,5 y 6,75 (IQR=2,25)

Nota mínima: 2 (posible indicador de dificultades graves)

Nota máxima: 10 (único caso destacado)

Acción recomendada:

Revisión pedagógica urgente para estudiantes bajo 6

Investigar caso de la nota 2 (¿necesidad de apoyo especial?)

Grupo B (Rendimiento intermedio)

Distribución más equilibrada (mediana=7, media=6,86)

Mayor variabilidad (IQR=4)

25% de estudiantes entre 9-10 (alto rendimiento)

25% de estudiantes entre 3-5 (bajo rendimiento)

Acción recomendada:

Tutorías diferenciadas

Capitalizar estrategias de los estudiantes con notas altas

Grupo C (Mejor rendimiento con anomalía)

75% de estudiantes obtuvo entre 6,5-9

Mediana de 8 (mejor desempeño global)

Outlier claro: nota 1 (inconsistente con el resto del grupo)

Acción recomendada:

Verificación inmediata del caso de nota 1 (¿error de registro?). Si es válido, requiere intervención individualizada

2.4) ACTIVIDADES DE APLICACIÓN

1) Realice un organizador gráfico del capítulo II

2) Calcular los cuartiles de los siguientes datos sobre calificaciones en Matemática. Realice las interpretaciones

8	9	10	8	9	7	6	8	9	8
8	10	7	6	6	8	7	9	10	8

7; 8; 9

3) Calcular los cuartiles de los siguientes datos sobre calificaciones en Matemática. Realice las interpretaciones

x_i	f
6	4
7	8
8	10
9	4
10	14

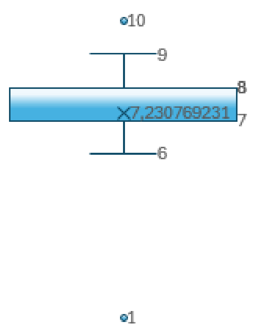
7; 8; 10

4) Con los siguientes datos calcule el quintil 4. Realice la interpretación.

Intervalos	f
$[10, 20)$	10
$[20, 30)$	14
$[30, 40)$	16
$[40, 50)$	20
$[50, 60)$	40

55

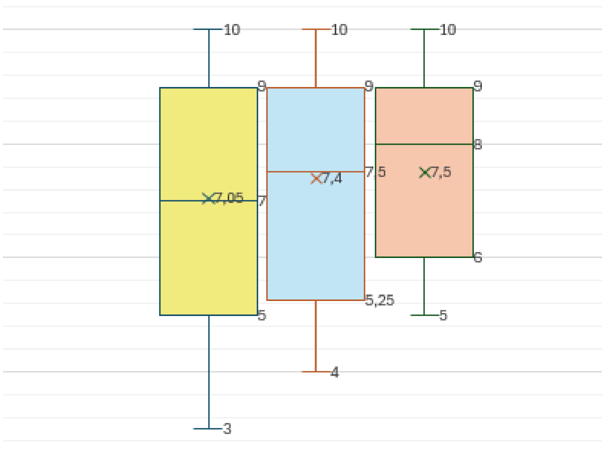
5) Calcule los cuartiles, el IQR, la media, el valor máximo, el valor mínimo, los límites inferior y superior de los bigotes de 1, 6, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 10. Elabore un diagrama de Caja y Bigotes empleando Excel. Interprete los resultados.



6) Elabore un diagrama de Caja y Bigotes empleando SPSS con los datos del ejercicio anterior.

7) Calcule los cuartiles, el IQR, la media, el valor máximo, el valor mínimo, los límites inferior y superior de los siguientes grupos. Elabore los diagramas de caja y bigotes empleando Excel. Interprete los resultados

Grupo A	Grupo B	Grupo C
3	4	5
4	5	5
4	5	5
6	5	6
5	5	6
5	6	6
5	6	7
6	7	7
6	7	7
7	7	8
7	8	8
8	8	8
8	9	8
9	9	8
9	9	9
9	9	9
10	9	9
10	10	9
10	10	10
10	10	10



8) Cree y resuelve un ejercicio similar al anterior. Elabore el diagrama de caja y bigotes con SPSS.

CAPÍTULO III
MEDIDAS DE DISPERSIÓN

3.1) DEFINICIÓN

Son medidas estadísticas que cuantifican la variabilidad o dispersión de un conjunto de datos. Indican la distancia entre los valores individuales y una medida de tendencia central, que puede ser la media, la mediana o la moda. Un nivel bajo de dispersión indica que los datos tienden a agruparse alrededor del valor central, mientras que un nivel alto indica que los datos están más dispersos.

3.2) CLASIFICACIÓN

Clasificación de las Medidas de Dispersión

Tipo de Medida	Nombre	Fórmula	Características
Medidas Absolutas	Rango (Amplitud Total)	$Rango = x_{máx} - x_{mín}$	- Simple, pero sensible a valores extremos.
	Desviación Media (DM)	$\mathrm{DM} = \frac{\sum\lvert \mathrm{x} - \bar{\mathrm{x}}\rvert}{\mathrm{n}}$	- Usa valores absolutos para evitar cancelaciones.
	Varianza *Población*: σ^2 *Muestra*: s^2	$\sigma^2 = \frac{\sum(\mathrm{x_i} - \mu)^2}{\mathrm{N}}$ $\mathrm{s}^2 = \frac{\sum(\mathrm{x_i} - \bar{\mathrm{x}})^2}{\mathrm{n} - 1}$	- Mide dispersión en unidades al cuadrado. - Sensible a outliers.
	Desviación Estándar *Población*: σ *Muestra*: s	$\sigma = \sqrt{\frac{\sum f(x_i - \mu)^2}{N}}$ $s = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{n - 1}}$	- Interpretación en unidades originales. - Más usada que la varianza.
Medidas Relativas	Coeficiente de Variación (CV)	$CV = \frac{\sigma}{\mu} \cdot 100\%$ $CV = \frac{s}{\bar{x}} \cdot 100\%$	- Compara dispersión entre conjuntos con distintas escalas. - Expresado en %.
	Rango Intercuartílico (IQR)	$IQR = Q_3 - Q_1$	- Elimina el 25% superior e inferior (robusto a outliers).
	Coeficiente de Dispersión (CD)	$CD = \frac{Rango}{\bar{x}}$	- Versión relativa del rango. - Poco usado en estadística moderna.

Cuando existe una dispersión pequeña se dice que los datos están dispersos o acumulados cercanamente respecto a un valor central, en este caso el dato central es un valor muy representativo. En el caso que la dispersión sea grande el valor central no es muy confiable. Cuando una distribución de datos tiene poca dispersión toma el nombre de distribución *homogénea* y si su dispersión es alta se llama *heterogénea*.

3.3) RANGO, AMPLITUD TOTAL O RECORRIDO

A) DEFINICIÓN

Dada una serie de valores $x_1, x_2, \dots, x_n$, su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores.

Características:

- Simple de calcular, pero muy sensible a valores atípicos (*outliers*).
- No proporciona información sobre la distribución interna de los datos (solo depende de dos valores).
- Se usa como una primera aproximación a la variabilidad de los datos.

B) MÉTODO DE CÁLCULO

$$Rango = x_{máx} - x_{mín}$$

Ejemplo de aplicación:

Calcula el rango de las siguientes distribuciones:

1) 4, 6, 8, 10, 12, 14, 16
2) 5, 10, 13, 13, 14, 15, 17

Solución:

$Rango_1 = x_{máx1} - x_{mín1} = 16 - 4 = 12$

$Rango_2 = x_{máx2} - x_{mín2} = 17 - 5 = 12$

Empleando Excel

Insertar la función MAX(Celdas) – MIN(Celdas) como muestra la siguiente figura:

	A	B	C	D	E	F
1	4			5		
2	6			10		
3	8			13		
4	10			13		
5	12			14		
6	14			15		
7	16			17		
8						
9	R	12	=MAX(A1:A7)-MIN(A1:A7)	R	12	=MAX(D1:D7)-MIN(D1:D7)

Interpretación:

Ambas series tienen rango 12, pero están desigualmente distribuidas, pues mientras la primera se distribuye uniformemente a lo largo de todo el recorrido, la segunda tiene una mayor concentración en el centro.

Nota:

La amplitud es una medida de dispersión cuya ventaja es la facilidad con que se calcula. Tiene en cambio las siguientes desventajas:

- En su cálculo sólo intervienen dos elementos del conjunto.

- Al aumentar el número de observaciones, puede esperarse que aumente la variabilidad. Puesto que la amplitud no tiene en cuenta el tamaño del conjunto, no es una medida adecuada para comparar la variabilidad de dos grupos de observaciones, a menos que éstos sean del mismo tamaño.

3.4) DESVIACIÓN MEDIA

A) DEFINICIÓN

La desviación media (DM) es una medida de dispersión que indica cuánto se alejan, en promedio, los datos de un conjunto con respecto a su media aritmética. A diferencia de la desviación estándar, aquí se usan valores absolutos para evitar que las diferencias positivas y negativas se cancelen.

Ventajas:

Fácil de interpretar: Mide la dispersión en las mismas unidades que los datos originales.

Considera todas las observaciones: A diferencia del rango, usa todos los datos del conjunto.

Evita cancelación de desviaciones: Al usar valores absolutos, no hay compensación entre diferencias positivas y negativas.

Robustez ante outliers (en comparación con la varianza): Al no elevar al cuadrado las desviaciones, es menos sensible a valores extremos que la desviación estándar.

Desventajas:

Menos eficiente en inferencia estadística: No se usa en pruebas paramétricas (como ANOVA o regresión), donde la varianza/desviación estándar son preferidas.

Poca relación matemática: Cuando mayor sea el valor de la desviación media, mayor es la dispersión de los datos. Sin embargo, no proporciona una relación matemática precisa entre su magnitud y la posición de un dato dentro de una distribución. La desviación media al tomar los valores absolutos mide una observación sin mostrar si la misma está por encima o por debajo de la media aritmética.

Menos precisa para distribuciones complejas: En distribuciones con múltiples modas o asimetrías, puede no reflejar adecuadamente la dispersión.

Usos Principales:

Análisis descriptivo: Ideal para reportar dispersión en informes no técnicos (Ejemplo: economía, ciencias sociales).

Enseñanza de estadística: Ayuda a entender el concepto de dispersión sin la complejidad de cuadrados (varianza).

Control de calidad: En procesos industriales donde se requiere medir desviaciones promedio de un valor objetivo.

Comparación de consistencia: Por ejemplo, evaluar cuán estables son los rendimientos de dos inversiones.

B) MÉTODOS DE CÁLCULO

1) Para datos no agrupados

$$DM = \frac{\sum|x_i - \bar{x}|}{n}$$

Donde:

x_i = cada valor del conjunto de datos.

$\bar{x}$ = media aritmética.

n = número total de datos.

Ejemplo de aplicación:

Calcule la desviación media de las siguientes calificaciones 5,7,8,9,10

Solución:

Calculando la media aritmética

$$\bar{x} = \frac{\sum x_i}{n} = \frac{5 + 7 + 8 + 9 + 10}{5} = \frac{39}{5} = 7{,}8$$

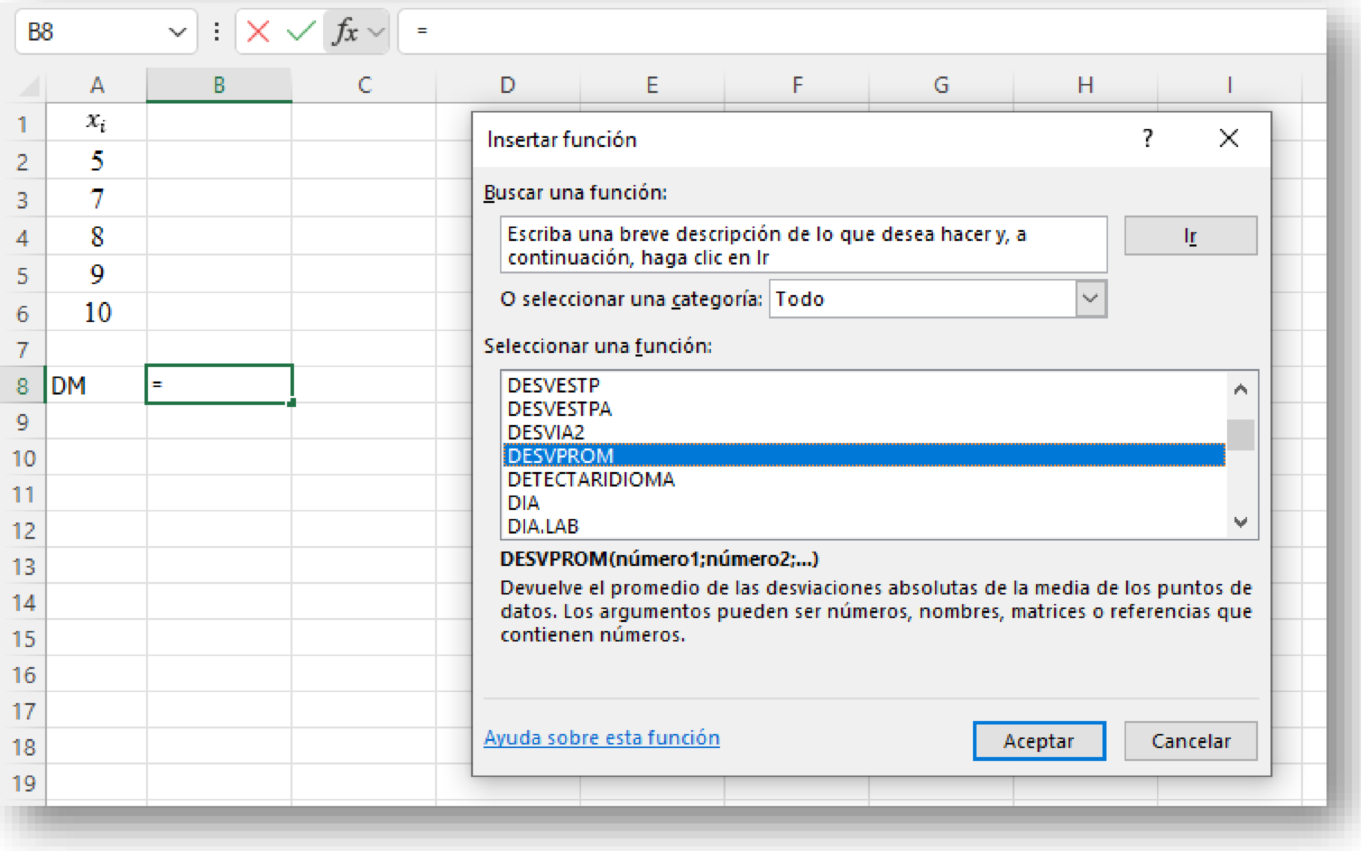
Calculando las desviaciones absolutas $|x_i - \bar{x}|$. Para cada dato, restamos la media y tomamos el valor absoluto

$$DM = \frac{\sum|x_i - \bar{x}|}{n} = \frac{|5 - 7{,}8| + |7 - 7{,}8| + |8 - 7{,}8| + |9 - 7{,}8| + |10 - 7{,}8|}{5}$$

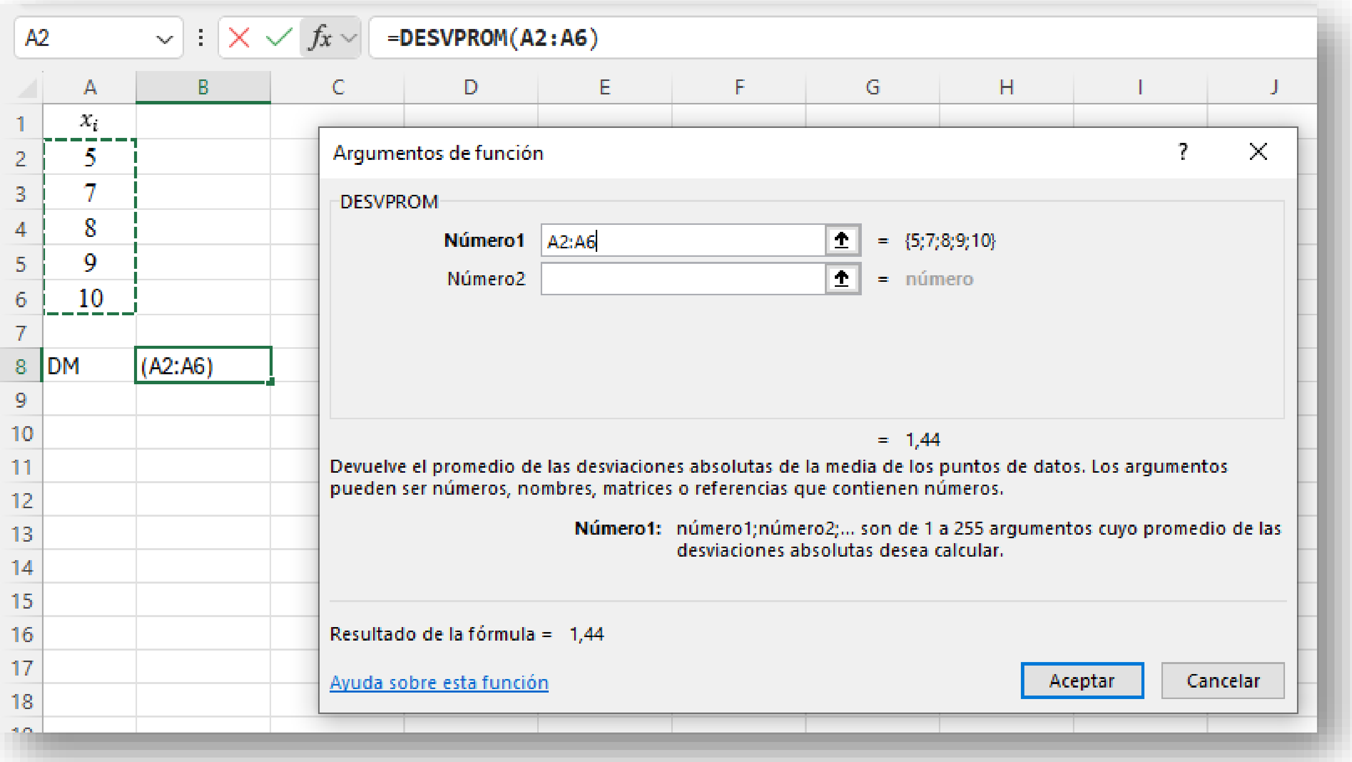
$$DM = \frac{2{,}8 + 0{,}8 + 0{,}2 + 1{,}2 + 2{,}2}{5} = 1{,}44$$

Empleando Excel

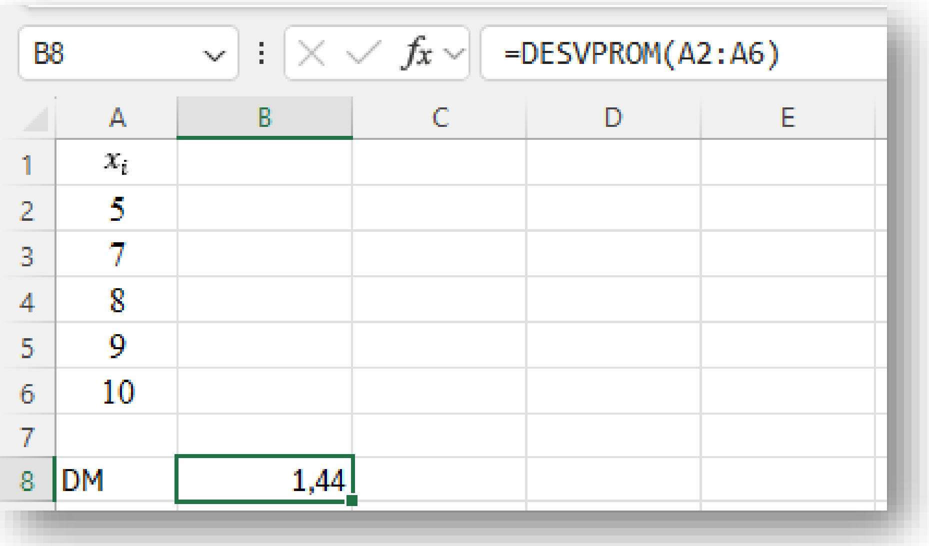
Se inserta la función DESVPROM como se muestra en la siguiente imagen:



Clic en Aceptar y seleccionar las celdas.



Clic en Aceptar



Interpretación:

Las calificaciones se desvían, en promedio, 1,44 puntos respecto a la media (7,8). Esto significa que, en general, las notas están dispersas alrededor de 7,8, con una variación promedio de ±1,44 puntos.

2) Para datos en tablas de frecuencia

$$DM = \frac{\sum f|x_i - \bar{x}|}{n}$$

Donde:

f = frecuencia absoluta

x_i = cada valor del conjunto de datos.

$\bar{x}$ = media aritmética.

n = número total de datos.

Ejemplo de aplicación:

Calcular la desviación media en base a la siguiente tabla sobre las calificaciones de un estudiante en 12 asignaturas evaluadas sobre 10**.**

Calificación	Cantidad de asignaturas
6	4
7	2
8	3
9	2
10	1
Total	12

Solución:

Se calcula la media aritmética.

$$\bar{x} = \frac{\sum f x_i}{n} = \frac{4 \cdot 6 + 2 \cdot 7 + 3 \cdot 8 + 2 \cdot 9 + 1 \cdot 10}{12} = \frac{24 + 14 + 24 + 18 + 10}{12} = \frac{90}{12} = 7,5$$

Se llena la siguiente tabla:

x_i	f	$\lvert x_i - \bar{x}\rvert$	$f\lvert x_i - \bar{x}\rvert$
6	4	1,5	6
7	2	0,5	1
8	3	0,5	1,5
9	2	1,5	3
10	1	2,5	2,5
Total	12		14

Se emplea la ecuación de la desviación media.

$$DM = \frac{\sum f\lvert x_i - \bar{x}\rvert}{n} = \frac{14}{12} = 1,167$$

Interpretación:

Variabilidad Moderada: Una DM de 1,167 indica que, en general, las calificaciones tienden a estar ±1,167 puntos alrededor del promedio (7,5).

El estudiante mantiene un desempeño regular (media 7,5), con algunas asignaturas destacadas (9-10) y otras por mejorar (6).

La DM sugiere que sus calificaciones no son extremadamente variables, pero hay margen para reducir la dispersión (Ejemplo: subir las notas de 6 a 7).

3) Para datos agrupados en intervalos

$$DM = \frac{\sum f\,|xm - \bar{x}|}{n}$$

Donde:

f = frecuencia absoluta.

xm = marca de clase.

$\bar{x}$ = media aritmética.

n = número total de datos.

Ejemplo de aplicación:

Calcular la desviación media de un curso de 40 estudiantes en la asignatura de Estadística en base a la siguiente tabla:

Calificación	Cantidad de estudiantes
2-4	6
4-6	8
6-8	16
8-10	10
Total	40

Solución:

Se calcula la media aritmética llenando la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$
2-4	6	3	18
4-6	8	5	40
6-8	16	7	112
8-10	10	9	90
Total	40		260

Calculando la media aritmética se obtiene:

$$\bar{x} = \frac{\sum f \cdot x\,m}{n} = \frac{260}{40} = 6{,}5$$

Para calcular la desviación media se llena la siguiente tabla:

Intervalo	f	xm	$\lvert xm-\bar{x}\rvert$	$f\lvert xm-\bar{x}\rvert$
2-4	6	3	3,5	21
4-6	8	5	1,5	12
6-8	16	7	0,5	8
8-10	10	9	2,5	25
Total	40			66

$$DM = \frac{66}{40} = 1{,}65$$

Interpretación:

Una DM de 1,65 indica que las notas están dispersas alrededor del promedio (6,5), pero sin extremos excesivos. El curso tiene un rendimiento medianamente homogéneo

3.5) VARIANZA

A) DEFINICIÓN

La varianza es una medida de dispersión que cuantifica cuánto se alejan los valores de un conjunto de datos de su media aritmética. Matemáticamente, se define como el promedio de los cuadrados de las desviaciones de cada dato respecto a la media.

B) MÉTODOS DE CÁLCULO

1) Para datos no agrupados

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum(x_i-\mu)^2}{N}$$

Donde:

$x_i = observaciones\ individuales\ de\ la\ población$

μ = media aritmética poblacional

$N = número\ de\ observaciones\ de\ la\ población$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

Donde:

$x_i = observaciones\ individuales\ de\ la\ muestra$

$\bar{x} = media\ aritmética\ de\ la\ muestra$

$n = número\ de\ obsevaciones\ de\ la\ muestra$

2) Para datos en tablas de frecuencia

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(x_i - \mu)^2}{N}$$

Donde:

f = frecuencia absoluta.

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(x_i - \bar{x})^2}{n-1}$$

3) Para datos agrupados en intervalos

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(xm - \mu)^2}{N}$$

Donde:
$f = frecuencia\ absoluta$
$xm = marca\ de\ clase$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(xm - \bar{x})^2}{n-1}$$

3.6) DESVIACIÓN ESTÁNDAR

A) DEFINICIÓN

La desviación estándar (o *desviación típica*) es una medida de dispersión que indica cuánto se alejan, en promedio, los valores de un conjunto de datos respecto a su media aritmética. Es la raíz cuadrada de la varianza y se expresa en las mismas unidades que los datos originales, lo que facilita su interpretación.

Características clave:

Unidades: A diferencia de la varianza (que está al cuadrado), la desviación estándar usa las mismas unidades que los datos originales (ejemplo: metros, dólares, etc.).

Relación con la varianza: Es su raíz cuadrada $\left(\sigma = \sqrt{\sigma^2} \;\; o \;\; s = \sqrt{s^2}\right)$

Interpretación:

La siguiente tabla muestra el nivel de dispersión de los datos según la desviación estándar

Rango (desviación estándar)	Interpretación	Ejemplo de Aplicación
Baja	Datos muy concentrados cerca de la media.	Control de calidad: piezas con medidas casi idénticas.
Moderada	Dispersión media. Variabilidad esperada en muchos contextos.	Notas de estudiantes en un examen bien balanceado.
Alta	Datos muy dispersos o con *outliers*.	Ingresos en una población con alta desigualdad económica.

B) MÉTODOS DE CÁLCULO

1) Para datos no agrupados

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

La desviación estándar de una muestra se calculó con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

Nota:

La división por $n-1$ corrige el sesgo introducido al estimar μ con $\bar{x}$, asegurado que s^2 sea un reflejo fiel de σ^2.

Además, se divide por $n-1$ debido a que una muestra generalmente está un poco menos dispersa que la población de la cual se tomó, es decir, al dividir para $n-1$ se cumple con la tendencia y sentido lógico de que la desviación estándar de la muestra debe tener un valor más pequeño que la desviación estándar de la población.

En la práctica, los parámetros poblacionales μ (media) y N (tamaño de la población) suelen ser desconocidos. Por ello, se recurre a una muestra para estimarlos, ya que, en contextos reales, rara vez se calcula la desviación estándar poblacional (σ) de manera directa. Esto se debe a limitaciones de tiempo, costos y recursos, lo que hace más eficiente trabajar con datos muestrales que representen adecuadamente a la población bajo estudio.

Ejemplo de aplicación:

Dos empresas, A y B, venden sobres de café instantáneo de 350 gramos. Se seleccionaron al azar en los mercados cinco sobres de cada una de las compañías y se pesaron cuidadosamente sus contenidos. Los resultados fueron los siguientes:

A	B
350,14	350,09
350,18	350,12
349,98	350,20
349,99	349,88
350,12	349,95

1) ¿Qué empresa proporciona más café en sus sobres?

2) ¿Qué empresa llena sus sobres de manera más consistente?

Solución:

1) Se calcula las medias aritméticas.

$$\bar{x}_A = \frac{\sum x_i}{n} = \frac{350{,}14 + 350{,}18 + 349{,}98 + 349{,}99 + 350{,}12}{5}$$

$$\bar{x}_A = \frac{1750{,}41}{5} = 350{,}082$$

$$\bar{x}_B = \frac{\sum x_i}{n} = \frac{350{,}09 + 350{,}12 + 350{,}20 + 349{,}88 + 349{,}95}{5} = \frac{1750{,}24}{5} = 350{,}048$$

$$\bar{x}_B = \frac{1750{,}24}{5} = 350{,}048$$

Interpretación:

Como la media aritmética de la empresa A es mayor que la media aritmética de la empresa B, por lo tanto, la empresa A proporciona más café en sus sobres.

2) Se calcula las desviaciones estándar.

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

A		
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
350,14	0,058	0,003364
350,18	0,098	0,009604
349,98	-0,102	0,010404
349,99	-0,092	0,008464
350,12	0,038	0,001444
Total		0,03328

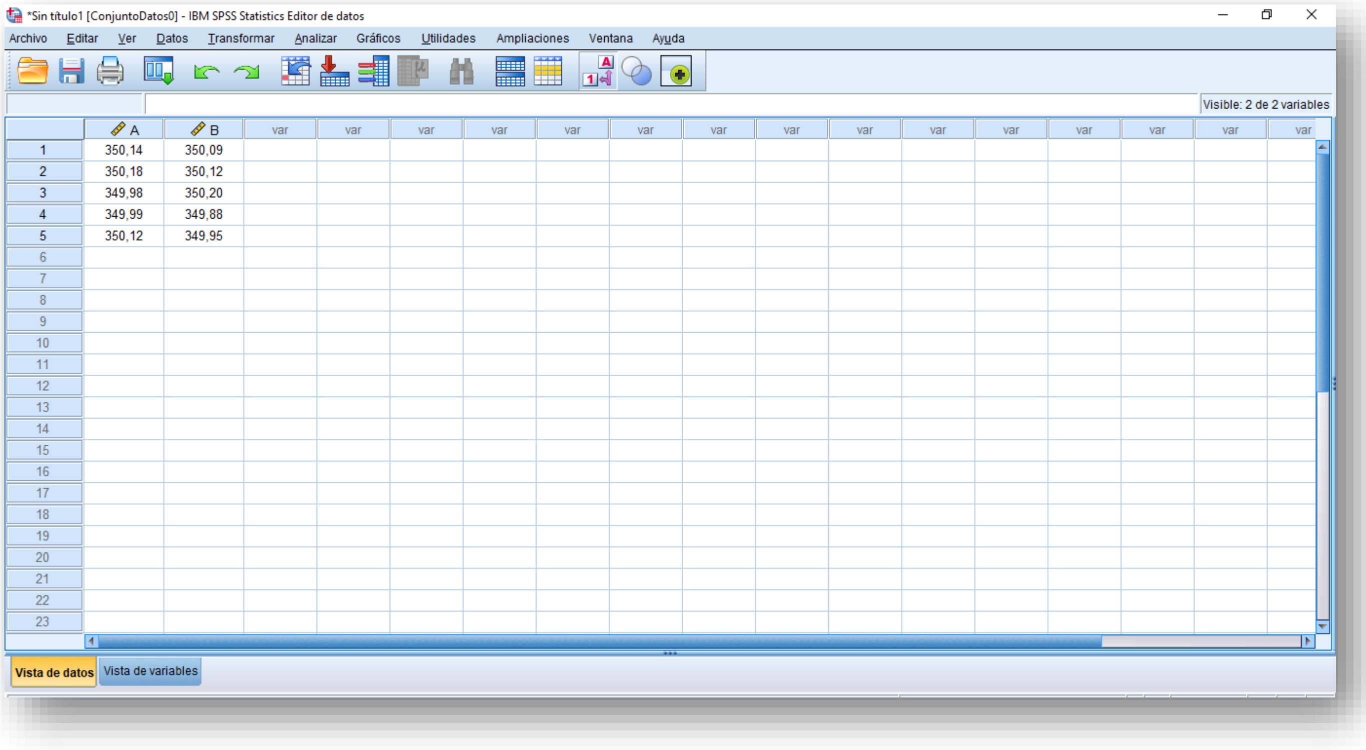
B		
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
350,09	0,042	0,001764
350,12	0,072	0,005184
350,20	0,152	0,023104
349,88	-0,168	0,028224
349,95	-0,098	0,009604
Total		0,06788

$$s_A = \sqrt{\frac{0{,}03328}{5-1}} = \sqrt{\frac{0{,}03328}{4}} = 0{,}0912$$

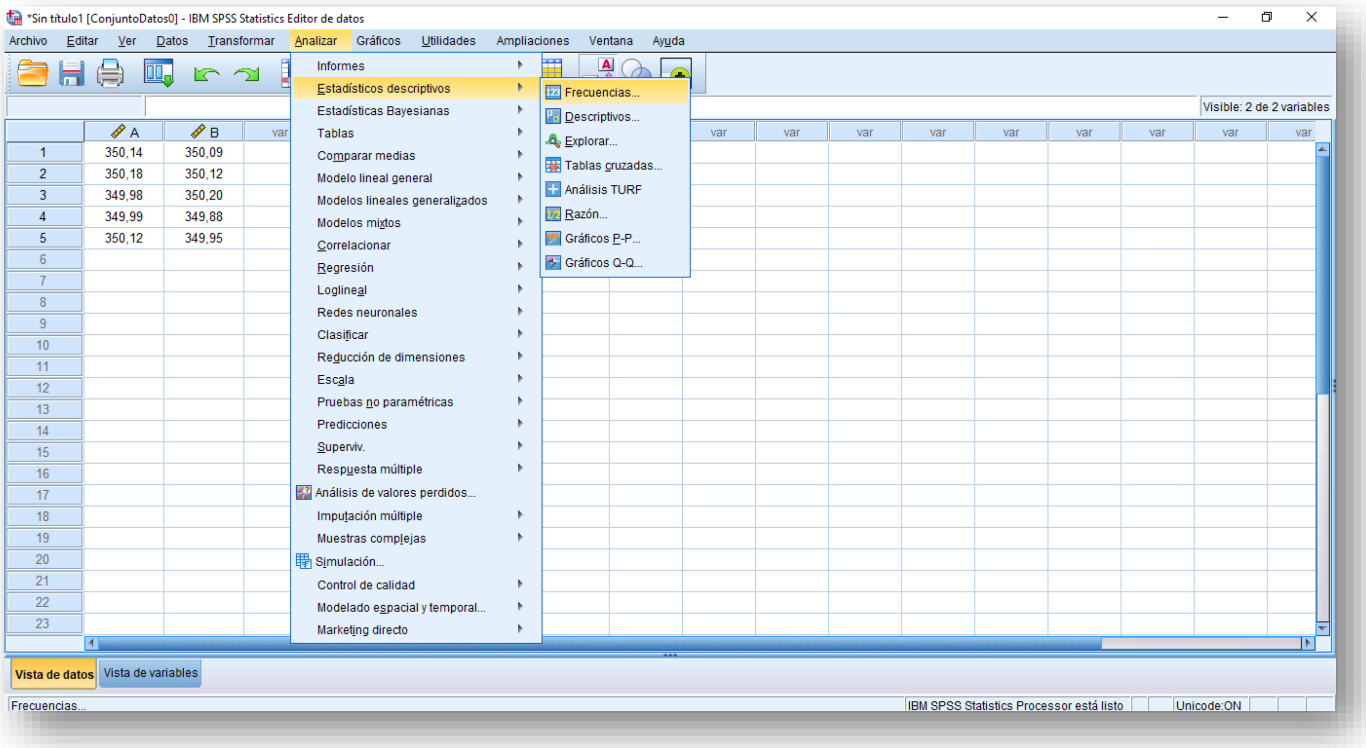
$$s_B = \sqrt{\frac{0{,}06788}{5-1}} = \sqrt{\frac{0{,}06788}{4}} = 0{,}13$$

Empleando SPSS

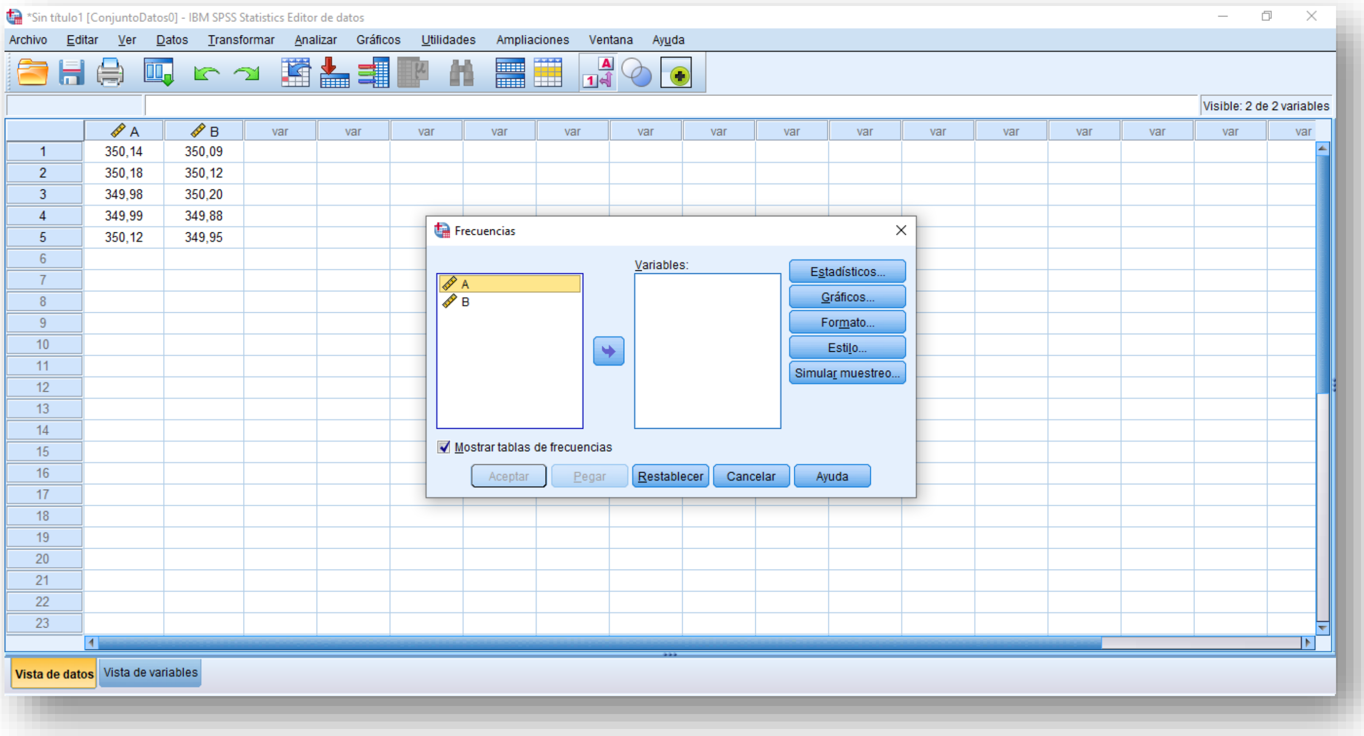
Ingresar los datos



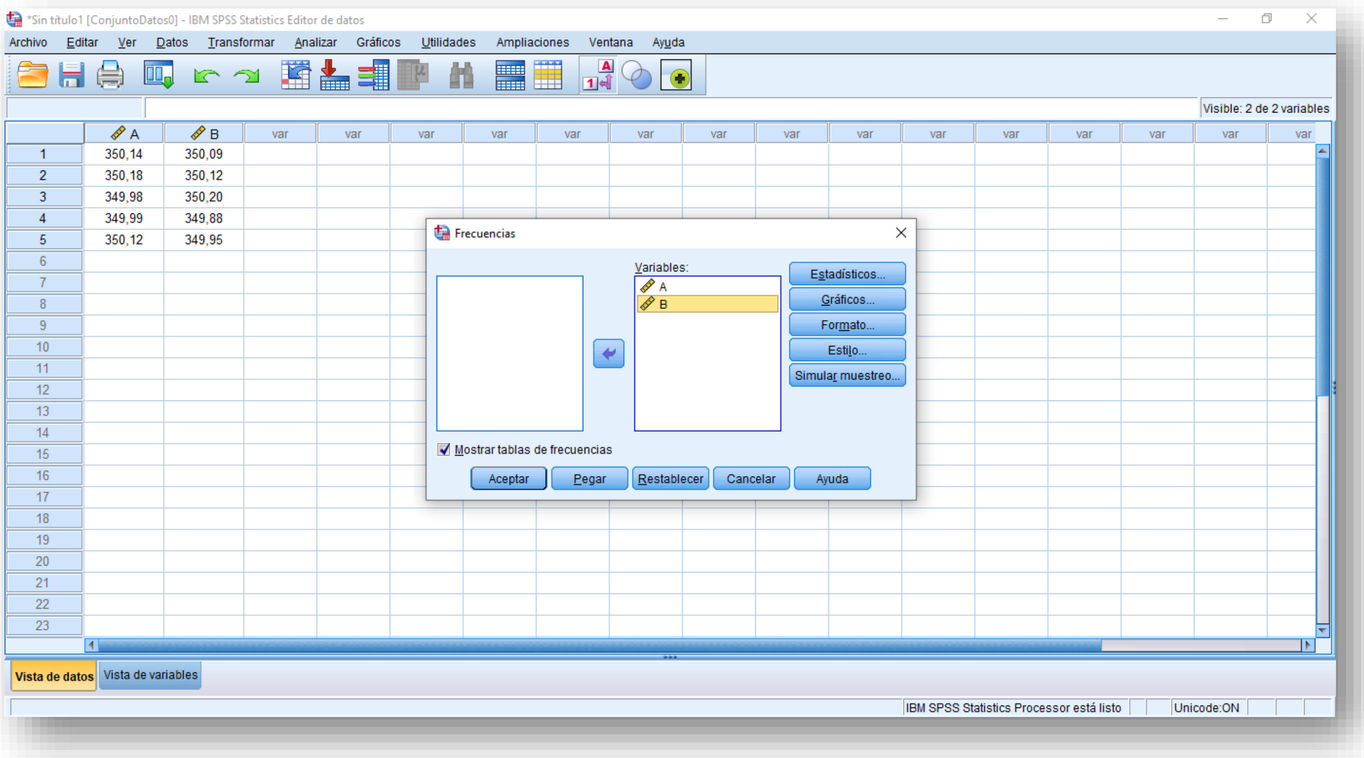
Clic en Analizar, Estadísticos descriptivos, Frecuencias



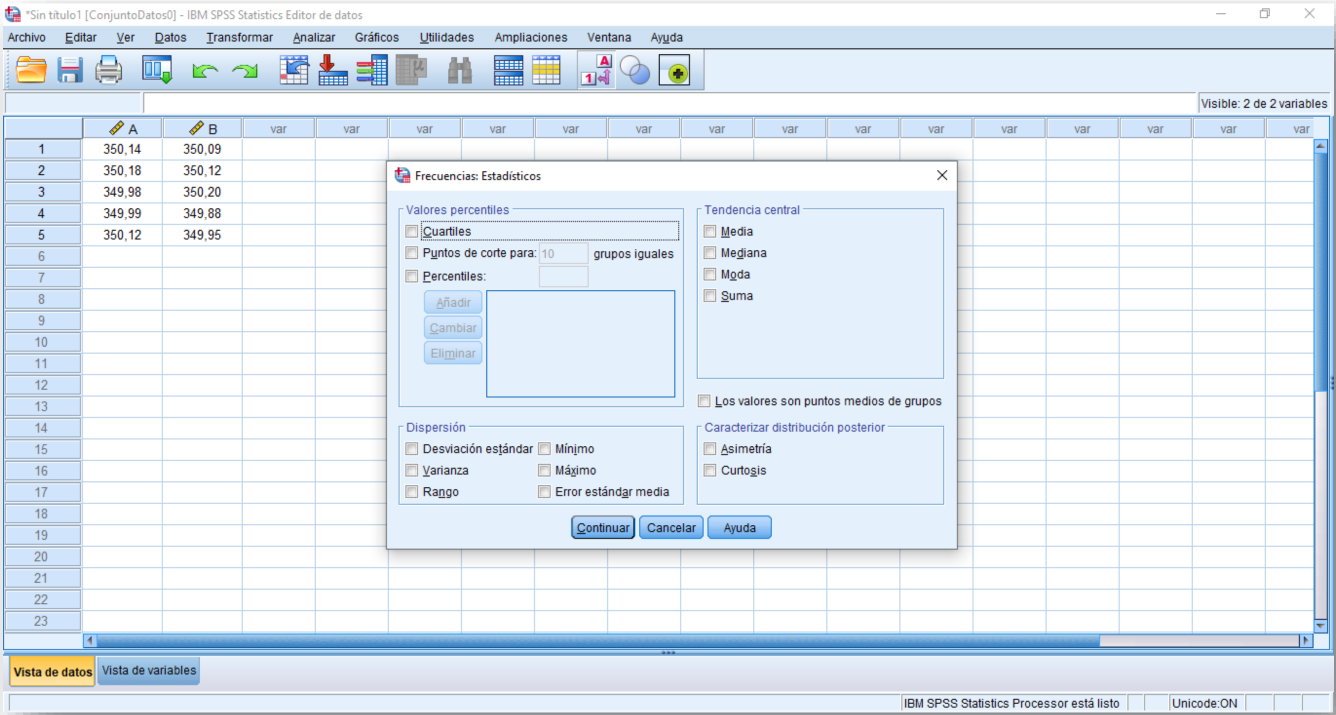
Clic en Frecuencias



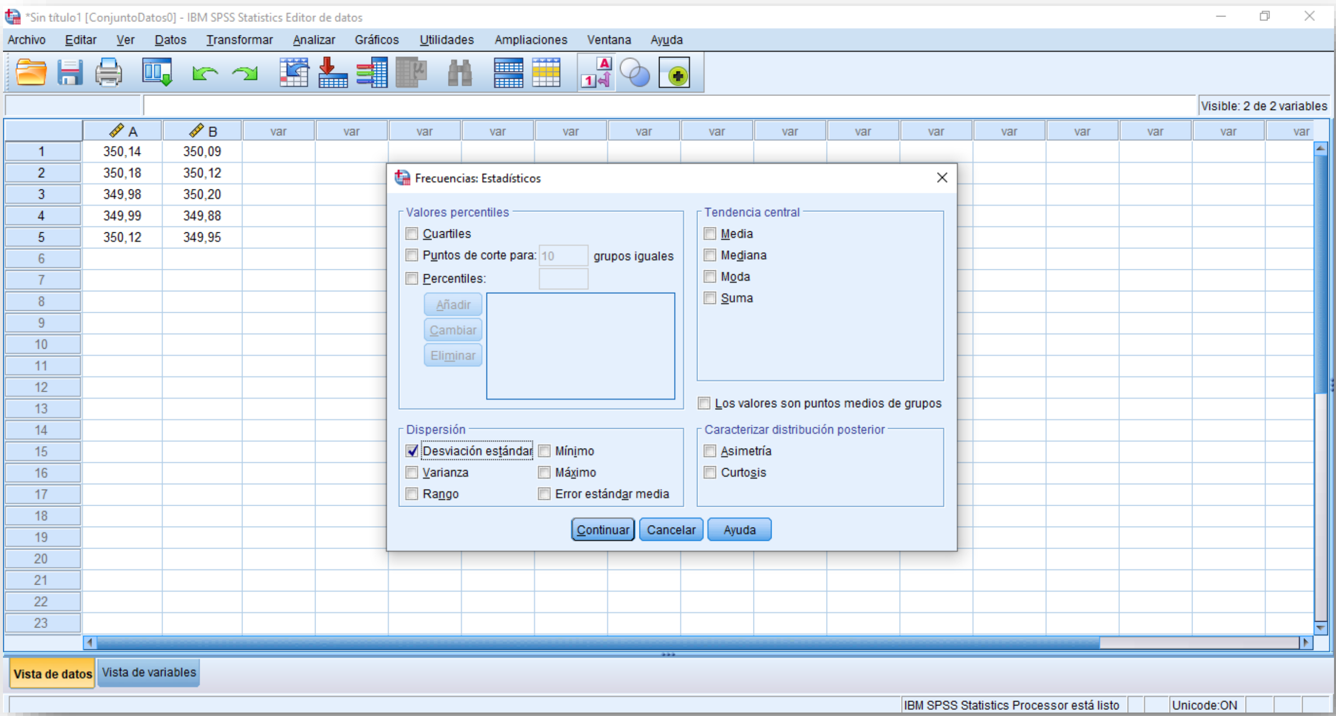
Seleccionar A y luego clic en la flecha. Seleccionar B y luego clic en la flecha.



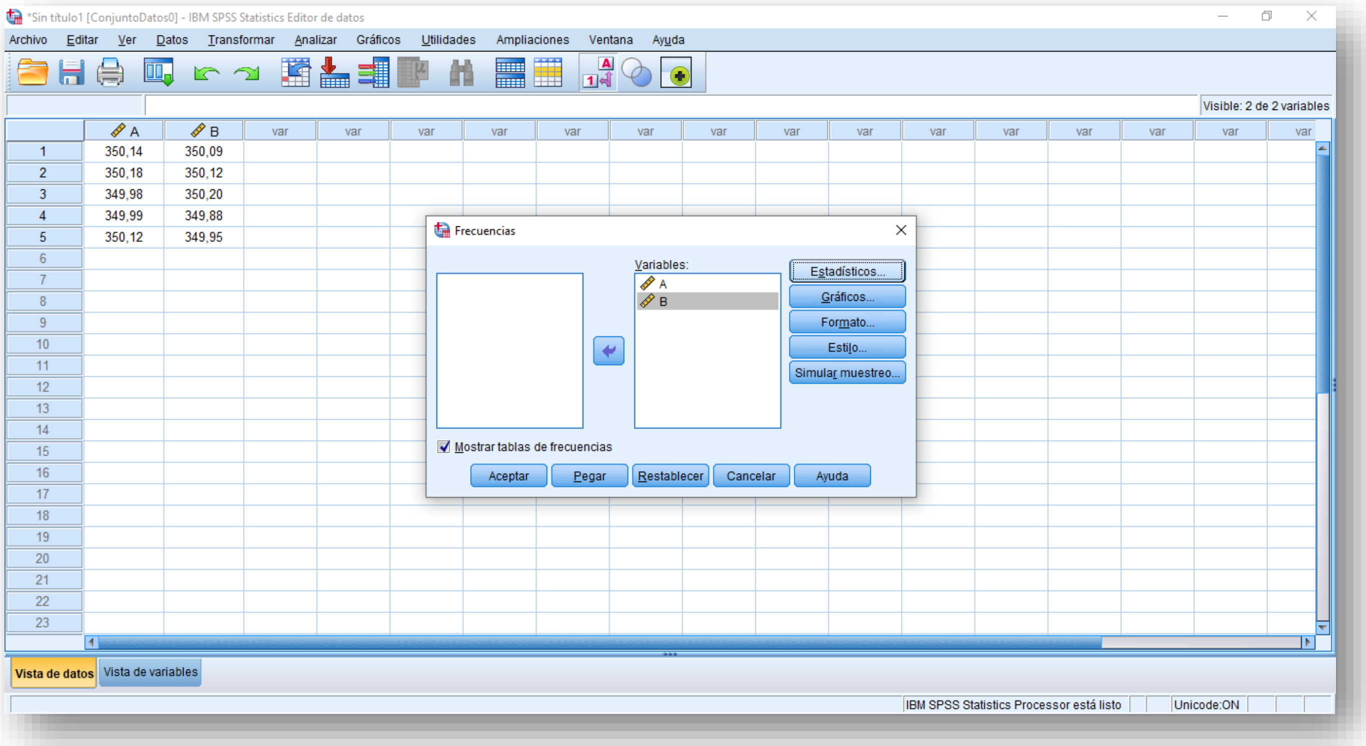
Clic en Estadísticos.



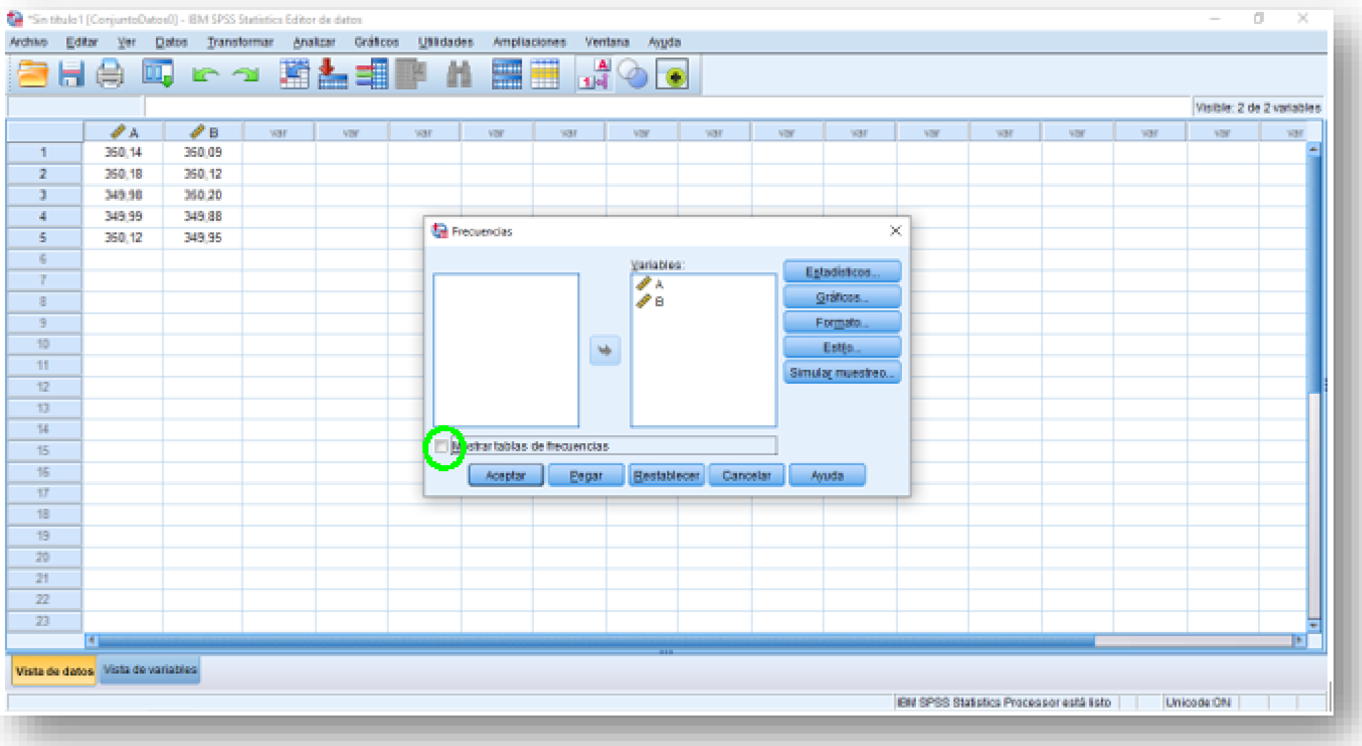
Clic en desviación estándar



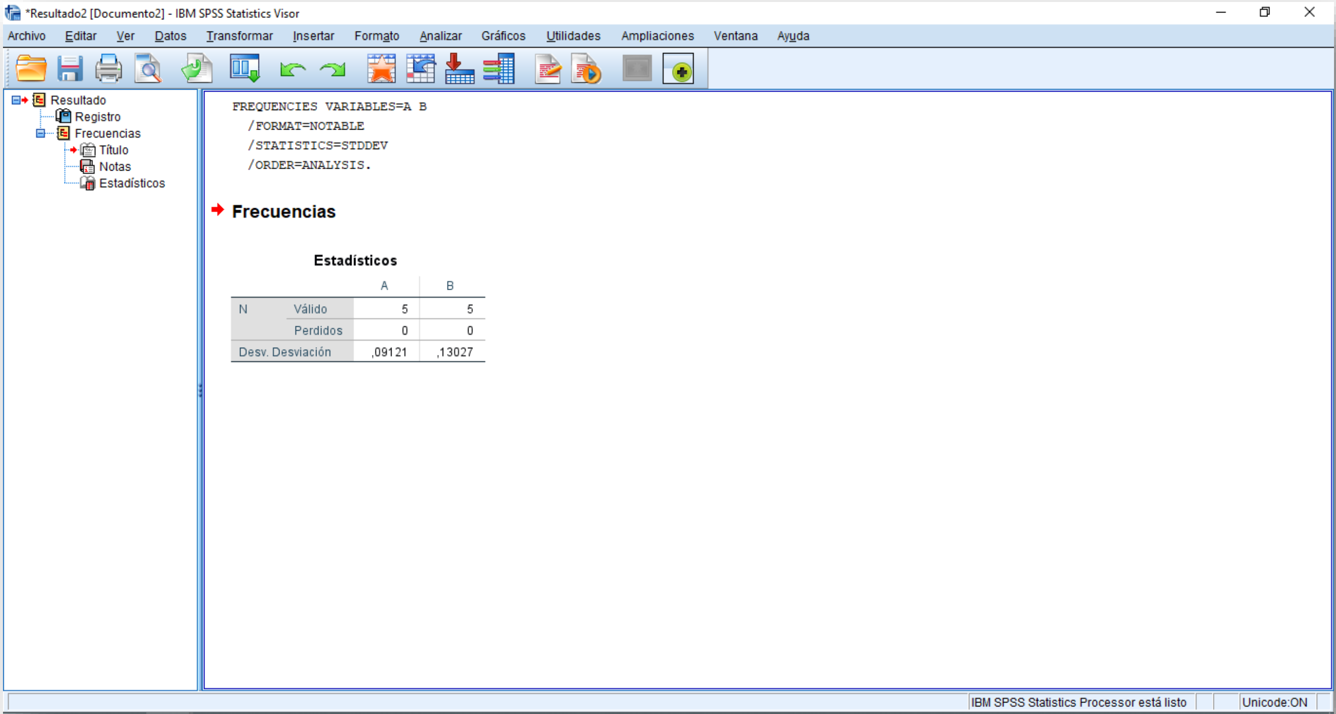
Clic en Continuar



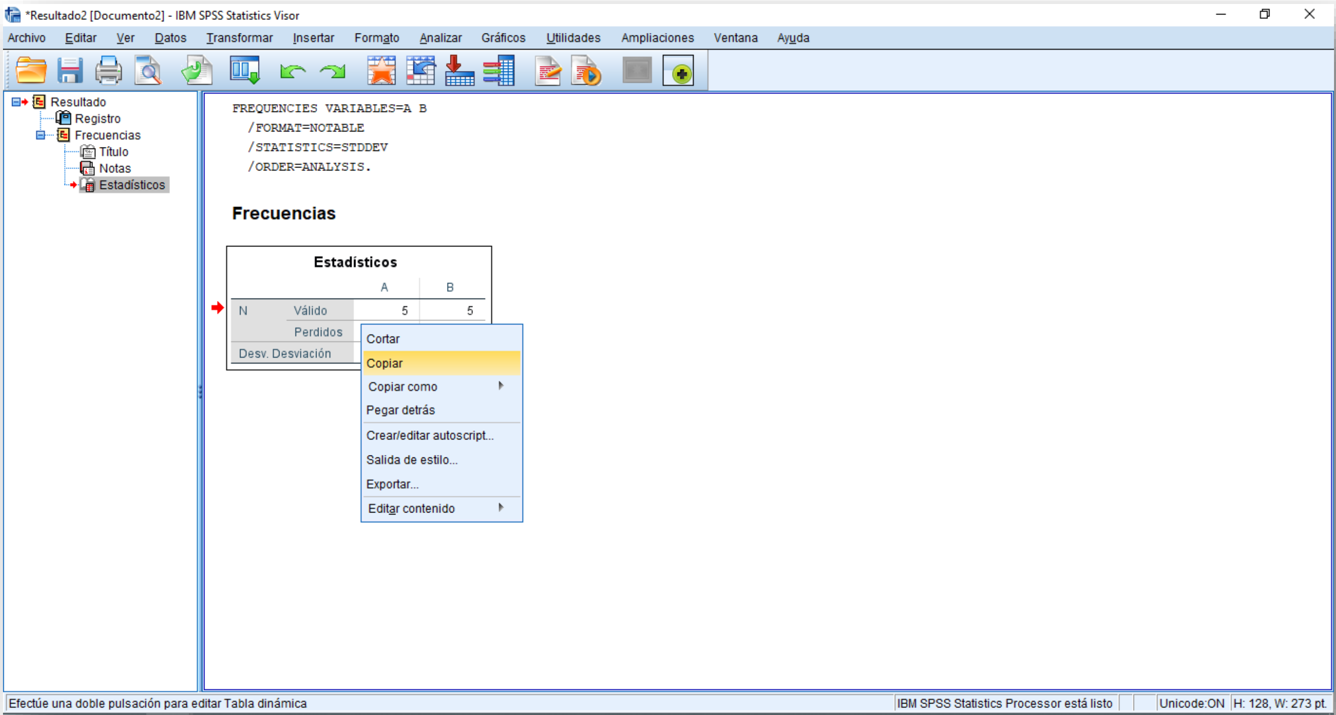
Clic Mostrar tablas de frecuencias (se borra el visto).



Clic en Aceptar



Clic derecho en la tabla



Clic en Copiar. Luego Pegar

Estadísticos

		A	B
N	Válido	5	5
	Perdidos	0	0
Desv. Desviación		,09121	,13027

Empleando Excel

Se emplea la función DESVEST.M

	A	B	C
1	A	B	
2	350,14	350,09	
3	350,18	350,12	
4	349,98	350,2	
5	349,99	349,88	
6	350,12	349,95	
7			
8	S_A	0,091214	=DESVEST.M(A2:A6)
9	S_B	0,130269	=DESVEST.M(B2:B6)

Interpretación:

Como la desviación estándar de la empresa A (0,091) es menor a la desviación estándar de la empresa B (0,13), por lo tanto, la empresa A es más consistente al llenar los sobres de café.

2) Para datos en tablas de frecuencia

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(x_i - \mu)^2}{N}}$$

La desviación estándar de una muestra se calcula con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{n - 1}}$$

Ejemplo de aplicación:

Calcular la desviación estándar de los siguientes datos correspondientes a una muestra.

Calificaciones	f
4	3
5	6
6	4
7	13
8	7
10	6
Total	39

Solución:

Se llena la siguiente tabla:

Calificaciones (x_i)	f	$f \cdot x_i$
4	3	12
5	6	30
6	4	24
7	13	91
8	7	56
10	6	60
Total	39	273

Se calcula la media aritmética.

$$\bar{x} = \frac{\sum f \cdot x_i}{n} = \frac{273}{39} = 7$$

Se llena la siguiente tabla:

x_i	f	$f \cdot x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f \cdot (x_i - \bar{x})^2$
4	3	12	-3	9	27
5	6	30	-2	4	24
6	4	24	-1	1	4
7	13	91	0	0	0
8	7	56	1	1	7
10	6	60	3	9	54
Total	39	273			116

Se calcula la desviación estándar.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{116}{39-1}} = \sqrt{\frac{116}{38}} = \sqrt{3{,}0526} = 1{,}747$$

Interpretación:

La desviación estándar es 1,747, lo que significa que, en promedio, las calificaciones se desvían ±1,747 puntos respecto a la media (7).

Una desviación estándar de 1,747 indica que la mayoría (aproximadamente el 68% de los datos si la distribución es normal) de las calificaciones están en el intervalo de variación típico ($\bar{x} \pm s$)

$\bar{x} \pm s = 7 \pm 1{,}747$

$7 - 1{,}747 = 5{,}25 \ ; \ 7 + 1{,}747 = 8{,}75$

Esto sugiere que la mayoría de los estudiantes obtuvieron calificaciones entre 5,25 y 8,75.

3) Para datos agrupados en intervalos

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(xm - \mu)^2}{N}}$$

Donde:

$f = frecuencia\ absoluta$

$xm = marca\ de\ clase$

La desviación estándar de una muestra se calcula con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(xm - \bar{x})^2}{n-1}}$$

Ejemplo de aplicación:

Calcular la desviación estándar de los siguientes datos correspondientes a una muestra.

Intervalo	*f*
60-65	5
65-70	20
70-75	40
80-85	27
85-90	8
Total	100

Solución:

Se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$
60-65	5	62,5	312,5
65-70	20	67,5	1350
70-75	40	72,5	2900
80-85	27	82,5	2227,5
85-90	8	87,5	700
Total	100		7490

Se calcula la media aritmética.

$$\bar{x} = \frac{\sum f \cdot xm}{n} = \frac{7490}{100} = 74{,}9$$

Se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$	$xm - \bar{x}$	$(xm - \bar{x})^2$	$f \cdot (xm - \bar{x})^2$
60-65	5	62,5	312,5	-12,4	153,76	768,8
65-70	20	67,5	1350	-7,4	54,76	1095,2
70-75	40	72,5	2900	-2,4	5,76	230,4
80-85	27	82,5	2227,5	7,6	57,76	1559,52
85-90	8	87,5	700	12,6	158,76	1270,08
Total	100		7490			4924

Se calcula la desviación estándar.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f\,(xm - \bar{x})^2}{n-1}} = \sqrt{\frac{4924}{100-1}} = \sqrt{\frac{4924}{99}} = \sqrt{49{,}737} = 7{,}052$$

Interpretación:

La desviación estándar es 7,052, lo que indica que, en promedio, los datos se desvían ±7,052 unidades respecto a la media (74,9).

Intervalo de variación típico (considerando ±1 desviación estándar):

$\bar{x} \pm s = 74{,}9 \pm 9{,}052$

$74{,}9 - 9{,}052 = 67{,}848$; $74{,}9 + 9{,}052 = 81{,}952$

La mayoría de los datos (alrededor del 68% si la distribución es normal) se encuentran en el rango 67,848 y 81,952.

C) RELACIÓN EMPÍRICA ENTRE DESVIACIÓN ESTÁNDAR Y RANGO

En distribuciones aproximadamente normales (en forma de campana), se cumple la siguiente regla práctica:

$$s \approx \frac{Rango}{4}$$

Esto se debe a que, en una distribución normal:

Casi el 100% de los datos está dentro de ±3 desviaciones estándar de la media ($\bar{x} \pm 3s$).

Por tanto, el rango total suele ser cercano a $6s$ (de $-3s$ a $+3s$), lo que implica:

$$s \approx \frac{Rango}{6}$$

Sin embargo, en distribuciones no perfectamente normales o con muestras pequeñas, esta relación puede variar. En la práctica, para una estimación más conservadora, se usa:

$$s \approx \frac{Rango}{4}$$

En nuestro ejemplo ilustrativo

$s = 7{,}052$

$Rango = x_{máx} - x_{mín} = 90 - 60 = 30$

Si aplicamos la relación empírica:

$$\frac{Rango}{4} = \frac{30}{4} = 7{,}5$$

El valor obtenido (7,5) es muy cercano a la desviación estándar calculada (7,052), lo que confirma que la dispersión es moderada y coherente con lo esperado en una distribución no extrema. Es decir, la relación con el rango valida que la dispersión es equilibrada o moderada.

A continuación, se presenta una tabla sobre la relación empírica entre desviación estándar y el rango

Relación	Interpretación	Implicación en los Datos
$s \ll \frac{Rango}{4}$	**Dispersión baja** (datos muy concentrados alrededor de la media).	- Poca variabilidad. - La mayoría de los valores están cerca del promedio.
$s \approx \frac{Rango}{4}$	**Dispersión moderada** (variabilidad esperada, sin valores extremos dominantes).	- Distribución equilibrada. Ejemplo: $s = 7{,}052$ $\frac{Rango}{4} = 7{,}5$
$s \gg \frac{Rango}{4}$	**Dispersión alta** (datos muy dispersos o con *outliers*).	- Presencia de valores atípicos. - Distribución posiblemente sesgada o con colas pesadas.

Nota: ≪ se lee "mucho menor que" ; ≈ se lee "casi igual a" ; ≫ se lee "mucho mayor que"

3.7) COEFICIENTE DE VARIACIÓN

A) DEFINICIÓN

El Coeficiente de Variación (CV) es una medida de dispersión relativa que expresa la desviación estándar como porcentaje de la media. Es útil para comparar la variabilidad entre conjuntos de datos con diferentes unidades o escalas.

En la siguiente tabla se muestra la interpretación del CV según su valor

Valor del CV (%)	Interpretación	Implicaciones
CV < 15%	Dispersión baja (datos homogéneos).	- Los datos están muy concentrados alrededor de la media. - Poca variabilidad. Ideal en controles de calidad. Ejemplo: Tiempos de producción en una fábrica eficiente (CV=10%).
15% ≤ CV ≤ 30%	Dispersión moderada (datos relativamente homogéneos).	- Variabilidad aceptable. - Común en estudios sociales o biológicos. Ejemplo: Edades en un grupo de adultos (CV=20%).
CV > 30%	Dispersión alta (datos heterogéneos o con outliers).	- Alta variabilidad. - Puede indicar inconsistencia o grupos mixtos (ej.: ingresos económicos muy desiguales). Ejemplo: Ventas mensuales de productos con algunos éxitos y muchos fracasos (CV=50%).

Posibles ejemplos de uso

- Comparar riesgo/rendimiento en finanzas (Ejemplo: volatilidad de acciones).
- Analizar calidad en producción (Ejemplo: variabilidad en diámetros de tornillos).
- Estudiar datos biológicos (Ejemplo: variación en crecimiento de plantas bajo distintos tratamientos).

Propiedades Principales

Adimensional (No tiene unidades)

Al ser una razón entre la desviación estándar y la media, el CV es independiente de las unidades de medida. Esto permite comparar la dispersión de variables con escalas distintas (Ejemplo: peso en kg vs. altura en cm).

Sensible a cambios en la media

Si la media es cercana a cero, el CV puede tender a infinito, lo que lo hace no interpretable. No se recomienda para datos con media próxima a cero. Si la media es muy pequeña, el CV amplifica artificialmente la dispersión

Útil para comparar variabilidad entre grupos

Permite determinar qué conjunto de datos tiene mayor dispersión relativa:

Ejemplo: Si el CV de los salarios en la empresa A es 15% y en la B es 25%, la empresa B tiene mayor variabilidad relativa.

No es robusto ante asimetría

En distribuciones muy asimétricas, el CV puede ser engañoso, ya que la media se ve afectada por valores extremos. En estos casos, es mejor usar otras medidas como el rango intercuartílico (IQR).

B) MÉTODOS DE CÁLCULO

Para una **población** se emplea la siguiente fórmula:

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Donde:

$CV = Coeficiente\ de\ variación$

$\sigma = desviación\ estándar\ de\ la\ población$

$\mu = media\ aritmética\ de\ la\ población$

Para una **muestra** se emplea la siguiente fórmula:

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Donde:

$CV = Coeficiente\ de\ variación$

$s = desviación\ estándar\ de\ la\ muestra$

$\bar{x} = media\ aritmética\ de\ la\ muestra$

C) EJEMPLOS DE APLICACIÓN

1) Comparar la variabilidad en los salarios de las empresas A y B según los siguientes datos

Empresa	Media Salarial ($\bar{x}$)	Desviación Estándar (s)
A	$3,000	$450
B	$5,000	$750

Solución:

Cálculo del CV

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Empresa A:

$$CV_A = \frac{450}{3000} \cdot 100\% = 15\%$$

Empresa B:

$$CV_B = \frac{750}{5000} \cdot 100\% = 15\%$$

Interpretación:

Ambas empresas tienen el mismo CV (15%), lo que indica que, proporcionalmente, la variabilidad de los salarios es idéntica en relación con sus medias. Además, la dispersión en las dos empresas es moderada (datos relativamente homogéneos).

Aunque la Empresa B tiene una desviación estándar mayor ($750 vs. $450), esto se compensa con su media salarial más alta ($5,000 vs. $3,000).

2) Comparar la variabilidad en tiempos de entrega de los repartidores X, Y empleando los siguientes datos:

Repartidor	Media de Tiempo ($\bar{x}$)	Desviación Estándar (s)
X	30 minutos	5 minutos
Y	60 minutos	20 minutos

Solución:

Cálculo del CV

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Repartidor X:

$$CV_X = \frac{5}{30} \cdot 100\% = 16{,}67\%$$

Repartidos Y:

$$CV_Y = \frac{20}{60} \cdot 100\% = 33{,}33\%$$

Interpretación:

El Repartidor X tiene menor dispersión relativa (16, 67%) frente al Y (33, 33%).

Conclusión: El Repartidor X es más consistente en sus tiempos de entrega, mientras que Y muestra el doble de variabilidad relativa.

3) Comparar el crecimiento de plantas (en cm) con dos fertilizantes.

Fertilizante	Media ($\bar{x}$)	Desviación Estándar (s)
Orgánico	50 cm	7 cm
Químico	80 cm	20 cm

Solución:

Cálculo del CV

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Fertilizante Orgánico:

$$CV_O = \frac{7}{50} \cdot 100\% = 14\%$$

Fertilizante Químico:

$$CV_Q = \frac{20}{80} \cdot 100\% = 25\%$$

Interpretación:

El fertilizante orgánico produce un crecimiento más homogéneo (CV = 14%).

El fertilizante químico tiene mayor variabilidad (CV = 25%), posiblemente por efectos desiguales en distintas plantas.

4) Realizar la comparación de la variabilidad en las notas de dos grupos de estudiantes en un examen de Matemáticas (escala 0-10) con los siguientes datos

Grupo	Media ($\bar{x}$)	Desviación Estándar (s)	Número de Alumnos
A	7,2	1,1	30
B	6,8	1,8	25

Solución:

Cálculo del CV

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Grupo A:

$$CV_A = \frac{1,1}{7,2} \cdot 100\% = 15,28\%$$

Grupo B:

$$CV_B = \frac{1,8}{6,8} \cdot 100\% = 26,47\%$$

Interpretación:

Grupo A (CV = 15,28%):

Baja variabilidad relativa.

Las notas están más concentradas alrededor de la media (7,2). *Ejemplo:* La mayoría de los estudiantes obtuvieron entre 6,1 y 8,3 puntos ($\bar{x} \pm s$).

Grupo B (CV = 26,47%):

Alta variabilidad relativa.

Notas más dispersas, con mayor diferencia entre estudiantes. *Ejemplo:* Notas típicas entre 5,0 y 8,6 puntos ($\bar{x} \pm s$), posiblemente con casos extremos (Ejemplo: 4 o 9).

Cálculo adicional: Error típico de la media

El error típico de la media (o error estándar de la media, SE por sus siglas en inglés) es una medida estadística que cuantifica la variabilidad esperada de la media muestral al repetir el muestreo de una población. Indica cuánto se alejaría, en promedio, la media calculada de una muestra de la media verdadera de la población si repitiéramos el estudio múltiples veces.

El error típico de la media **no es una medida de dispersión**, ya que no describe la variabilidad de los datos originales, sino la confiabilidad de la media. Se calcula empleando la siguiente fórmula:

$$SE = \frac{s}{\sqrt{n}}$$

Donde:
s = Desviación estándar muestral (mide la dispersión de los datos).
n = Tamaño de la muestra.

Ejemplo de aplicación:

Calcular el error típico de la media con los datos del ejemplo ilustrativo 4

Solución:

Cálculo del SE

$$SE = \frac{s}{\sqrt{n}}$$

Grupo A:

$$SE_A = \frac{1,1}{\sqrt{30}} = 0,20$$

Mide cuánto podría variar la media muestral (7,2) si repitiéramos el estudio.

El intervalo de confianza al 95% es $\bar{x} \pm 1,96 \cdot SE \rightarrow 7,2 \pm 1,96 \cdot 0,20$

$7,2 - 1,96 \cdot 0,20 = 6,8;\ 7,2 + 1,96 \cdot 0,20 = 7,6$

Interpretación:

Esto significa que podemos estar 95% seguros de que si repitiéramos el examen con otros 30 alumnos la media estaría entre 6,8 y 7,6.

Grupo B:

$$SE_B = \frac{1,8}{25} = 0,36$$

El intervalo de confianza al 95% es $\bar{x} \pm 1,96 \cdot SE \rightarrow 6,8 \pm 1,96 \cdot 0,36$

$6,8 - 1,96 \cdot 0,36 = 6,09;\ 6,8 + 1,96 \cdot 0,36 = 7,51$

Interpretación:

Esto significa que si repitiéramos el examen con otros 25 alumnos, el promedio estaría entre 6,09 y 7,51

5) Evaluar la variabilidad en resultados de glucosa en sangre (mg/dL) entre dos laboratorios clínicos empleando los siguientes datos.

Laboratorio	Media ($\bar{x}$)	Desviación Estándar (s)	Número de Muestras (n)
A	98 mg/dL	4,2 mg/dL	50
B	102 mg/dL	8,1 mg/dL	50

Nota: mg/dL significa miligramos por decilitro y es una unidad de concentración comúnmente utilizada en medicina para medir la glucosa en sangre.

Solución:

Cálculo del CV

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Laboratorio A:

$$CV_A = \frac{4,2}{98} \cdot 100\% = 4,29\%$$

Laboratorio B:

$$CV_B = \frac{8,1}{102} \cdot 100\% = 7,94\%$$

Interpretación Clínica:

Laboratorio A (CV = 4,29%):

Alta precisión analítica. Los resultados son muy consistentes.

Implicación: Confiable para diagnóstico de diabetes o hipoglucemia, donde pequeñas variaciones son críticas.

Laboratorio B (CV = 7.94%):

Precisión aceptable pero inferior. Mayor dispersión en los resultados.

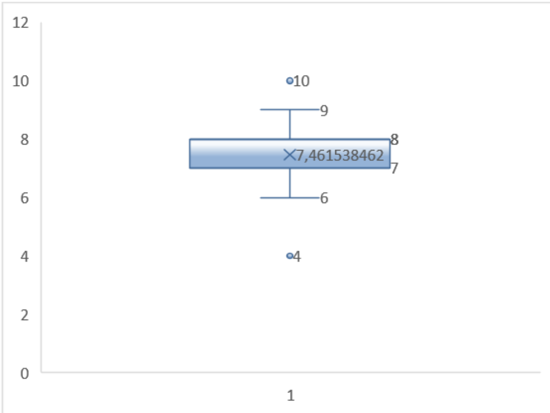
Implicación: Podría generar falsos positivos/negativos en casos límite (Ejemplo: glucosa en 100-110 mg/dL).

3.8) ACTIVIDADES DE APLICACIÓN

1) Realice un organizador gráfico del capítulo III

2) Con los siguientes datos sobre calificaciones de una muestra: 4, 6, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 10

a) Calcule los cuartiles, el IQR, el valor máximo, el valor mínimo, los límites inferior y superior de los bigotes, y valores atípicos. Elabore un diagrama de Caja y Bigotes empleando Excel. Interprete los resultados.



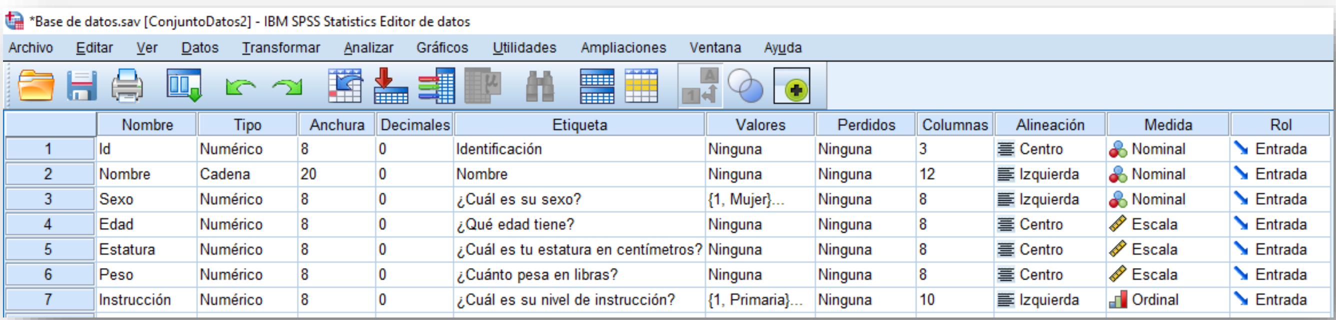
b) Calcule la media, moda, mediana, desviación media, desviación estándar, el coeficiente de variación y el error típico de la media. Interprete los resultados.

7,46; 8; 8; 1,04; 2,10; 28,18%; 0,40

c) Ordene los datos en una tabla de frecuencias y calcule media, moda, cuartiles, desviación media, desviación estándar, el coeficiente de variación y el error típico de la media. Interprete los resultados.

7,46; 8; 7; 8; 8; 1,04; 2,10; 28,18%; 0,40

3) Ingrese los siguientes datos en SPSS.


*Base de datos.sav [ConjuntoDatos2] - IBM SPSS Statistics Editor de datos

Archivo Editar Ver Datos Transformar Analizar Gráficos Utilidades Ampliaciones Ventana Ayuda

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	Id	Numérico	8	0	Identificación	Ninguna	Ninguna	3	Centro	Nominal	Entrada
2	Nombre	Cadena	20	0	Nombre	Ninguna	Ninguna	12	Izquierda	Nominal	Entrada
3	Sexo	Numérico	8	0	¿Cuál es su sexo?	{1, Mujer}...	Ninguna	8	Izquierda	Nominal	Entrada
4	Edad	Numérico	8	0	¿Qué edad tiene?	Ninguna	Ninguna	8	Centro	Escala	Entrada
5	Estatura	Numérico	8	0	¿Cuál es tu estatura en centímetros?	Ninguna	Ninguna	8	Centro	Escala	Entrada
6	Peso	Numérico	8	0	¿Cuánto pesa en libras?	Ninguna	Ninguna	8	Centro	Escala	Entrada
7	Instrucción	Numérico	8	0	¿Cuál es su nivel de instrucción?	{1, Primaria}...	Ninguna	10	Izquierda	Ordinal	Entrada

¿Cuál es su sexo? tiene los valores:

1=Mujer

2=Hombre

¿Cuál es su nivel de instrucción? tiene los valores:

1= Primaria

2= Secundaria

3= Universitario

4=Maestría

5=PhD

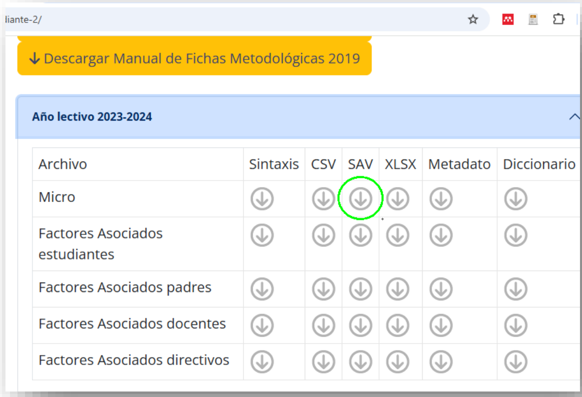
Y la vista de datos es:

	Id	Nombre	Sexo	Edad	Estatura	Peso	Instrucción
1	1	Mario Suárez	Hombre	47	168	170	Maestría
2	2	Mathías Suárez	Hombre	17	168	140	Primaria
3	3	Emyli Suárez	Mujer	12	130	100	Primaria
4	4	Dyana Rivera	Mujer	49	170	150	Universitario
5	5	Segundo Suárez	Hombre	76	172	160	Primaria
6	6	Bertha Ibujés	Mujer	77	156	170	Primaria
7	7	Sandra Suárez	Mujer	55	160	165	Primaria
8	8	Ximena Suárez	Mujer	41	162	175	Secundaria
9	9	Darío Suárez	Hombre	43	165	155	Primaria
10	10	Pablo Yépez	Hombre	54	160	180	Primaria

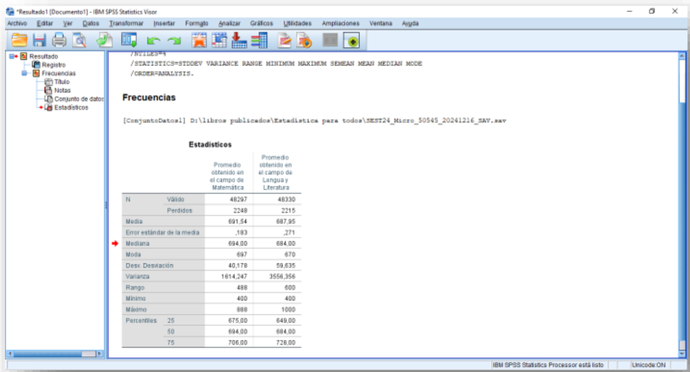
a) Elabore una tabla de frecuencias y una tabla cruzada de las variables que usted estime pertinente. Interprete los resultados.

b) Calcule la media, error estándar de la media, mediana, moda, cuartiles, varianza y desviación estándar de la variable estatura. Interprete los resultados.

4) Descargue la base de datos de la Prueba Ser Estudiante 2023-2024 (Micro SAV) disponible en el portal oficial: https://evaluaciones.evaluacion.gob.ec/BI/ser-estudiante-2/

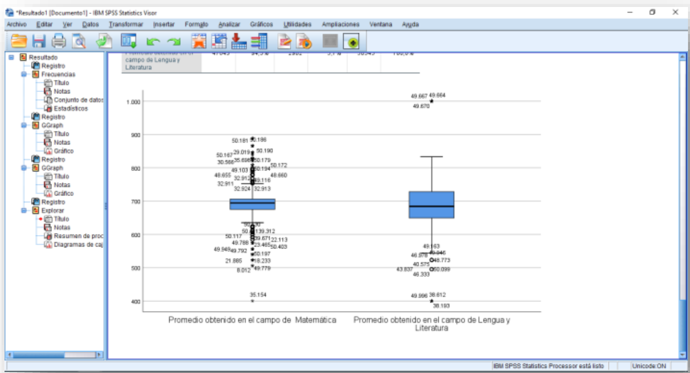


a) Empleando SPSS calcule la media, error estándar de la media, mediana, moda, desviación estándar, varianza, rango, mínimo, máximo y cuartiles del promedio obtenido en el campo de Matemática (imat) y del promedio obtenido en el campo de Lengua y Literatura (ilyl). Interprete los resultados.



		Promedio obtenido en el campo de Matemática	Promedio obtenido en el campo de Lengua y Literatura
N	Válido	48297	48330
	Perdidos	2248	2215
Media		691,54	687,95
Error estándar de la media		,183	,271
Mediana		694,00	684,00
Moda		697	670
Desv. Desviación		40,178	59,635
Varianza		1614,247	3556,356
Rango		488	600
Mínimo		400	400
Máximo		888	1000
Percentiles	25	675,00	649,00
	50	694,00	684,00
	75	706,00	728,00

b) Calcule el IQR, límites inferior y superior de los bigotes y los valores atípicos de imat y ilyl. Elabore un diagrama de Caja y Bigotes empleando SPSS. Interprete los resultados.



5) Cree y resuelva un ejercicio similar al anterior empleando la misma base de datos.

6) Con los datos sobre edades que representan a una muestra de personas, calcule e interprete los resultados.

Intervalos	***f***
50 – 59	2
60 – 69	4
70 – 79	5
80 – 89	6
90 – 99	3
Total	20

a) La media aritmética

76,5

b) El decil 5

78

c) La moda

82,5

d) La desviación estándar

12,4

e) El coeficiente de variación

16,2

7) Cree y resuelva un ejercicio similar al anterior.

FORMULARIO

Media aritmética simple

a) Para datos sin tabular

La media de una población es el parámetro μ (que se lee "miu"). Si hay N observaciones en el conjunto de datos de la población, la media se calcula así:

$$\mu = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \frac{\sum x_i}{N}$$

La media de una muestra es un estadístico $\bar{x}$ (que se lee "x barra"). Con n observaciones en el conjunto de datos de la muestra $(x_1, x_2, \ldots)$, la media se determina así:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum x_i}{n}$$

b) Para datos ordenados en tablas de frecuencias

$$\bar{x} = \frac{f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3 + \cdots + f_n \cdot x_n}{f_1 + f_2 + f_3 + \cdots f_n} = \frac{\sum f_i \cdot x_i}{\sum f} = \frac{\sum f x_i}{n}$$

Donde $n = \sum f$ es la frecuencia total (o sea, el número total de casos)

c) Para datos agrupados en intervalos

$$\bar{x} = \frac{f_1 \cdot xm_1 + f_2 \cdot xm_2 + f_3 \cdot xm_3 + \cdots f_n \cdot xm_n}{f_1 + f_2 + f_3 + \cdots f_n} = \frac{\sum f_i \cdot xm_i}{\sum f} = \frac{\sum f \cdot xm}{n}$$

Media aritmética ponderada

$$\bar{x} = \frac{p_1 \cdot x_1 + p_2 \cdot x_2 + p_3 \cdot x_3 + \cdots p_k \cdot x_k}{p_1 + p_2 + p_3 + \cdots p_k} = \frac{\sum p \cdot x}{\sum p}$$

La Mediana

a) Para datos no agrupados

Se aplica la fórmula para encontrar la ubicación:

$$Me = \boldsymbol{x}_{\frac{\boldsymbol{n+1}}{\boldsymbol{2}}}$$

b) Para datos en tablas de frecuencia

Se aplica la fórmula para encontrar la ubicación:

$$Me = \boldsymbol{x}_{\frac{\boldsymbol{n+1}}{\boldsymbol{2}}}$$

c) Para datos agrupados en intervalos

Se emplea la ecuación

$$Me = L_i + \left(\frac{\frac{n}{2} - F_{i-1}}{f_i}\right) \cdot A_i$$

Donde:

$L_i =$ Límite inferior del intervalo de clase de la mediana

$n =$ Número total de datos

F_{i-1} =Frecuencia acumulada del intervalo de clase que antecede al intervalo de la mediana.

$f_i =$ Frecuencia absoluta del intervalo de la mediana

$A_i =$ Amplitud o Ancho del intervalo

La Moda

Para datos agrupados en intervalos se emplea la siguiente fórmula:

$$Mo = L_i + \left(\frac{d_1}{d_1 + d_2}\right) \cdot A_i$$

Donde:

$L_i =$ Límite inferior de la clase modal

$d_1 =$ Diferencia entre la frecuencia de la clase modal y la frecuencia del intervalo anterior.

$d_2 =$ Diferencia entre la frecuencia de la clase modal y la frecuencia del intervalo posterior.

$A_1 =$ Amplitud o ancho de la clase modal.

Cuantiles (cuartiles, quintiles, deciles, percentiles)

a) Para datos no agrupados

Ordenar de menor a mayor y aplicar la fórmula de posición

$$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$$

Donde:

$x = x_i$ = posición del dato (ejemplo: x_2)

k = Número del cuantil (Ejemplo: 1 para Q_1, 25 para P_{25})

n = Número total de datos

m = Número de divisiones (4 para cuartiles, 5 para quintiles, 10 para deciles, 100 para percentiles)

b) Para datos en tablas de frecuencia

Ordenar de menor a mayor y aplicar la fórmula anterior de posición para datos no tabulados

$$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$$

$$Cuantil = L_i + \left(\frac{\frac{nk}{m} - F_{i-1}}{f_i}\right) \cdot A_i$$

Donde:

$Cuantil: Cuartiles\ Q_k\ , Quintiles\ K_k\ , Deciles\ D_k\ , Percentiles\ P_k$

L_i = Límite inferior del intervalo del cuantil

n = Número total de datos

k = Número del cuantil (Ejemplo: 1 para Q_1, 25 para P_{25})

m = Número de divisiones (4 para cuartiles, 5 para quintiles, 10 para deciles, 100 para percentiles)

F_{i-1} = Frecuencia acumulada del intervalo que antecede al intervalo del cuantil

f_i = Frecuencia absoluta del intervalo del cuantil

A_i = Amplitud o Ancho del intervalo

Desviación media o desviación promedio

Para datos no agrupados

$$DM = \frac{\sum|x - \bar{x}|}{n}$$

Para datos ordenados en tablas de frecuencia

$$DM = \frac{\sum f|x - \bar{x}|}{n}$$

Para datos agrupados en intervalos

$$DM = \frac{\sum f|xm - \bar{x}|}{n}$$

Donde xm es la marca de clase**.**

Varianza y desviación estándar

Para datos no agrupados

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

Donde:

$x_i = observaciones\ individuales\ de\ la\ población$

μ = media aritmética poblacional

$N = número\ de\ observaciones\ de\ la\ población$

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1}$$

Donde:

$x_i = observaciones\ individuales\ de\ la\ muestra$

$\bar{x} = media\ aritmética\ de\ la\ muestra$

$n = número\ de\ obsevaciones\ de\ la\ muestra$

La desviación estándar de una muestra:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Para datos ordenados en tablas de frecuencias

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(x_i - \mu)^2}{N}$$

Donde:

f = frecuencia absoluta.

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(x_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(x_i - \bar{x})^2}{n-1}$$

La desviación estándar de una muestra se calcula con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(x_i - \bar{x})^2}{n-1}}$$

Para datos agrupados en intervalos

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(xm_i - \mu)^2}{N}$$

Donde:

$f = frecuencia\ absoluta$; $xm = marca\ de\ clase$

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(xm_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(xm_i - \bar{x})^2}{n-1}$$

La desviación estándar de una muestra se calcula con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(xm_i - \bar{x})^2}{n-1}}$$

Relación empírica entre desviación estándar y rango

$$s \approx \frac{Rango}{4}$$

Error típico de la media

$$SE = \frac{s}{\sqrt{n}}$$

Donde:

s = Desviación estándar muestral (mide la dispersión de los datos).

n = Tamaño de la muestra.

Rango intercuartílico

Es la distancia entre el tercer cuartil Q_3 y el primer cuartil Q_1.

Amplitud intercuartílica o rango intercuartílico = tercer cuartil - primer cuartil

$IQR = Q_3 - Q_1$

Límites para bigotes

Límite inferior: $Q_1 - 1{,}5 \cdot IQR$

Límite superior: $Q_3 + 1{,}5 \cdot IQR$

Dispersión relativa o coeficiente de variación

Para una población

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Donde:

$CV = Coeficiente\ de\ variación;\ \sigma = desviación\ estándar\ de\ la\ población$

$\mu = media\ aritmética\ de\ la\ población$

Para una muestra

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Donde:

$CV = Coeficiente\ de\ variación; s = desviación\ estándar\ de\ la\ muestra$

$\bar{x} = media\ aritmética\ de\ la\ muestra$

SOLUCIONARIO DE LA EVALUACIÓN DIAGNÓSTICA

CUESTIONARIO

1) Una Parroquia de una determinada ciudad está compuesta de 3 barrios, el barrio A de 2000 habitantes, el barrio B de 2500 habitantes y el barrio C de 1000 habitantes. Se va a realizar un estudio de mercado. Realice los cálculos en forma manual y con Excel. ¿Cuántas encuestas se debe aplicar si se emplea el cálculo del tamaño de la muestra al 95% de confianza con un error de muestreo del 5%? ¿Cuántas encuentras se deben aplicar en cada barrio? Realice la interpretación de los resultados.

Solución:

Para determinar el tamaño de la muestra necesario para realizar el estudio de mercado en la parroquia compuesta por tres barrios (A, B y C), seguiremos los siguientes pasos:

Datos Poblacionales:

Barrio A: 2000 habitantes

Barrio B: 2500 habitantes

Barrio C: 1000 habitantes

Total de la población (N): 2000 + 2500 + 1000 = 5500 habitantes

Parámetros del Estudio:

Nivel de confianza (Z): 95% → Z = 1,96 (valor estándar para 95% de confianza)

Error de muestreo (e): 5% → e = 0,05

Proporción estimada (p): Como no se tiene información previa, se asume p = 0,5 (maximiza el tamaño muestral).

Fórmula para el Tamaño de la Muestra (Población Finita):

La fórmula para calcular el tamaño de la muestra es:

$$n = \frac{NpqZ^2}{(N-1)e^2 + pqZ^2}$$

Sustituyendo los valores:

$$n = \frac{NpqZ^2}{(N-1)e^2 + pqZ^2} = \frac{5500 \cdot 0{,}5 \cdot 0{,}5 \cdot 1{,}96^2}{(5500-1) \cdot 0{,}05^2 + 0{,}5 \cdot 0{,}5 \cdot 1{,}96^2} = 359{,}13 = 359$$

Como el tamaño de la muestra debe ser un número entero, redondeamos: n = 359 encuestas.

Distribución de las Encuestas por Barrio (Afijación Proporcional):

Para distribuir las 360 encuestas proporcionalmente al tamaño de cada barrio:

$$Encuestas\ en\ el\ Barrio\ A = \frac{2000}{5500} \cdot 359 = 131$$

$$Encuestas\ en\ el\ Barrio\ B = \frac{2500}{5500} \cdot 359 = 163$$

$$Encuestas\ en\ el\ Barrio\ C = \frac{1000}{5500} \cdot 359 = 65$$

Total: 131 + 163 + 65 = 359 encuestas (se ajusta correctamente).

Cálculo en Excel:

	A	B	C	D	E	F
1	N	5.500	=A16			
2	p	**0,5**				
3	q	0,5	=1-B2			
4	Confianza	**95**				
5	Área a la izquierda de -Z	0,025	=(100-B4)/200			
6	-Z	-1,96	=INV.NORM.ESTAND(B5)			
7	Z	1,959964	=B6*-1			
8	e = error de muestreo	**5**	%			
9	e	0,05	=B8/100			
10	$n = \frac{NpqZ^2}{(N-1)e^2 + pqZ^2}$	359,13	=(B1*B2*B3*B7^2)/((B1-1)*B9^2+B2*B3*B7^2)			
11		**359**	=REDONDEAR(B10;0)			
12	Habitantes	Barrios		**Encuestas**		
13	2000	A	130,55	=(A13/$A$16)*$B$11	**131**	=REDONDEAR(C13;0)
14	2500	B	163,18	=(A14/$A$16)*$B$11	**163**	=REDONDEAR(C14;0)
15	1000	C	65,27	=(A15/$A$16)*$B$11	**65**	=REDONDEAR(C15;0)
16	**5500**	=SUMA(A13:A15)			**359**	=SUMA(E13:E15)

Interpretación:

a) Se requiere un mínimo de 360 encuestas para toda la parroquia, con un nivel de confianza del 95% y un error máximo del 5%.

b) La distribución proporcional asegura que cada barrio esté representado de acuerdo con su tamaño poblacional:

Barrio A (2000 habitantes): 131 encuestas.

Barrio B (2500 habitantes): 163 encuestas.

Barrio C (1000 habitantes): 65 encuestas.

c) Esto garantiza que los resultados del estudio sean estadísticamente representativos de toda la parroquia.

Con los siguientes datos sobre las calificaciones trimestrales en la Asignatura de Matemática evaluadas sobre 10 obtenidas de 45 estudiantes:

8,83	8,85	8,74	7,70	9,08
4,48	8,40	8,70	8,99	9,09
6,99	8,79	7,90	7,69	8,96
7,00	8,70	8,93	9,11	9,01
8,84	9,05	8,85	9,83	9,01
9,83	8,33	8,88	8,90	7,33
8,78	9,01	8,96	8,38	8,26
8,71	8,85	8,55	8,11	9,23
8,84	6,24	8,99	4,00	9,31

2) Calcule la media aritmética. Realice las interpretación.

Solución:

$$\bar{x} = \frac{\sum \boldsymbol{x_i}}{\boldsymbol{n}} = \frac{8{,}83 + 8{,}85 + 8{,}74 + \cdots 9{,}31}{45} = \frac{379{,}01}{45} = 8{,}42$$

Interpretación:

La calificación promedio es 8,42/10, lo que indica un rendimiento general alto en la asignatura.

3) Calcule la moda. Realice la interpretación.

Solución:

Valor más frecuente: El número que más se repite.

En los datos 8,85 y el 9,01 aparecen 3 veces (máxima frecuencia).

Moda = 8,85; Moda = 9,01.

Interpretación:

Existe una tendencia notable hacia notas cercanas a 9,0, lo que refuerza el buen desempeño del grupo. Sin embargo, al haber múltiples modas (como 8,85 y 9,01), la distribución no es perfectamente unimodal.

4) Calcule los cuartiles. Realice las interpretaciones.

Solución:

Ordenando de menor a mayor:

4,00; 4,48; 6,24; 6,99; 7,00; 7,33; 7,69; 7,70; 7,90; 8,11; 8,26; 8,33; 8,38; 8,40; 8,55; 8,70; 8,70; 8,71; 8,74; 8,78; 8,79; 8,83; 8,84; 8,84; 8,85; 8,85; 8,85; 8,88; 8,90; 8,93; 8,96; 8,96; 8,99; 8,99; 9,01; 9,01; 9,01; 9,05; 9,08; 9,09; 9,11; 9,23; 9,31; 9,83; 9,83

Fórmula de posición

$Posición = p \cdot (n + 1)$

$Donde$:

$p = 0{,}25; 0{,}50; 0{,}75$

Q_1 (Percentil 25):

$Posición = p \cdot (n + 1) = 0{,}25 \cdot (45 + 1) = 11{,}5$

Otra fórmula de posición (fórmula general)

$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$

Donde:

$x = x_i$ = posición del dato (ejemplo: x_2)

k = Número del cuantil (Ejemplo: 1 para Q_1)

n = Número total de datos

m = Número de divisiones (4 para cuartiles, 100 para percentiles)

$Posición = x_{\left[\frac{k(n+1)}{m}\right]}$

$P_k = P_1 = x_{\left[\frac{k(n+1)}{m}\right]} = x_{\left[\frac{1(45+1)}{4}\right]} = x_{\left[\frac{46}{4}\right]} = x_{11,5}$

Por lo tanto, al ser la posición 11,5 (obtenida con las fórmulas) se debe calcular el promedio de las posiciones 11 y 12

$$Q_1 = \frac{8{,}26 + 8{,}33}{2} = 8{,}295$$

Interpretación:

El 25% de los estudiantes obtuvo una nota ≤ 8,295.

Q_2 (Mediana, Percentil 50):

$Posición = p \cdot (n+1) = 0{,}5 \cdot (45+1) = 23$

Por lo tanto, al ser la posición 23 es el valor en la posición 23

$Q_2 = 8{,}84$

Interpretación:

El 50% de las calificaciones son $\leq$ 8,84, muy cercana a la media, lo que sugiere una distribución equilibrada en el centro.

Q_3 (Percentil 75):

$Posición = p \cdot (n+1) = 0{,}75 \cdot (45+1) = 34{,}5$

Por lo tanto, al ser la posición 34,5 se debe calcular el promedio de las posiciones 34 y 35

$$Q_3 = \frac{8{,}99 + 9{,}01}{2} = 9{,}0$$

Interpretación:

El 75% de los estudiantes sacó $\leq$ 9,00.

5) Calcule el rango. Realice la interpretación.

Solución:

$$Rango = Valor\ máximo - Valor\ mínimo = x_{máx} - x_{mín} = 9{,}83 - 4{,}00 = 5{,}83$$

Interpretación:

La diferencia entre la nota máxima (9,83) y mínima (4,00) es amplia (5,83 puntos).

Implicación: Existe una gran variabilidad debido a valores atípicos bajos (Ejemplo 4,00; 4,48), que podrían distorsionar el análisis global.

6) Calcule la desviación estándar de la muestra. Realice la interpretación.

Solución:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{(4{,}00 - 8{,}42)^2 + (4{,}48 - 8{,}42)^2 + \cdots + (9{,}83 - 8{,}42)^2}{45-1}} = 1{,}15$$

Interpretación:

Las notas se desvían en promedio 1,15 puntos de la media.

Implicación: Una desviación estándar >1,0 en una escala de 10 indica dispersión moderada-alta. Esto se explica por la presencia de notas atípicas bajas y altas (Ejemplo 4,00 vs. 9,83).

7) Calcule el coeficiente de variación. Realice la interpretación.

Solución:

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

$$CV = \frac{1{,}15}{8{,}42} \cdot 100\% = 13{,}66\%$$

Interpretación:

Un CV del 13,66% indica que las calificaciones tienen una dispersión moderada-baja respecto al promedio. Esto sugiere que la mayoría de los estudiantes obtuvieron puntuaciones cercanas a 8,42, con una desviación estándar (s=1,15) relativamente pequeña.

8) Calcule el error estándar de la media. Realice la interpretación.

Solución:

Error estándar de la media (SEM)

$$SEM = \frac{s}{\sqrt{n}} = \frac{1{,}15}{\sqrt{45}} = 0{,}17$$

Interpretación:

Si se repitiera el estudio, la media oscilaría en ±0,17 puntos alrededor de 8,42.

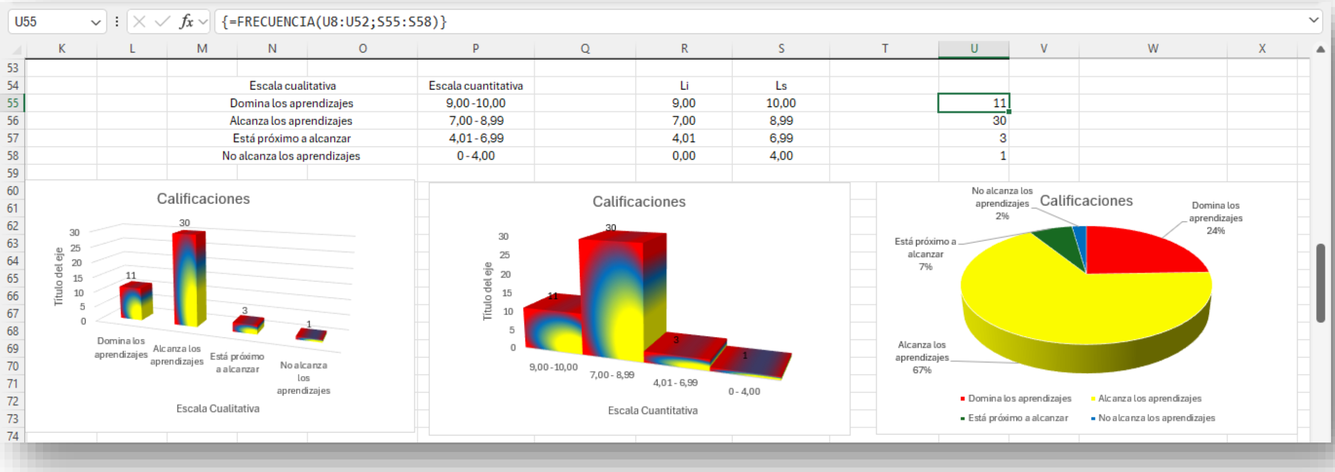
Resumen de Resultados

Estadístico	Valor	Interpretación
Media	8,42	Promedio de las calificaciones.
Moda	8,85 9,01	Nota más frecuente.
Mediana (Q_2)	8,84	50% de las notas son ≤ 8,84.
Q_1	8,30	25% de las notas son ≤ 8,30.
Q_3	9,00	75% de las notas son ≤ 9,00.
Rango	5,83	Diferencia entre la nota máxima (9,83) y mínima (4,00).
Desviación Estándar	1,15	Dispersión moderada alrededor de la media.
Coeficiente de variación	13,66%	Dispersión moderada-baja respecto al promedio
Error Estándar de la Media	0,17	Precisión de la media estimada.

9) Elabore un diagrama de barras, un histograma y un diagrama de sectores considerando los intervalos empleados en el sistema educativo del Ministerio de Educación del Ecuador. Realice la interpretación.

Solución:

Empleando Excel



Interpretación:

La Mayoría está en "Alcanza los aprendizajes" (30 estudiantes): La gran parte del grupo se ubica en un nivel satisfactorio (7,00 – 8,99), lo que indica un buen desempeño general.

Segundo lugar: "Domina los aprendizajes" (11 estudiantes): Un grupo menor pero significativo demuestra un excelente dominio (9,00 – 10,00), destacándose por su alto rendimiento.

Baja frecuencia en niveles inferiores: Solo 3 estudiantes están "próximos a alcanzar" (4,01 – 6,99), y 1 estudiante no logra los aprendizajes (0 – 4,00). Esto sugiere que la mayoría supera los estándares mínimos.

10) Elabore un diagrama de caja y bigotes. Realice la interpretación.

Solución:

Los cuartiles previamente calculados fueron

$Q_1 = 8{,}295 = 8{,}30$

$Q_2 = 8{,}84$

$Q_3 = 9{,}00$

Cálculo del Rango Intercuartílico (IQR)

$IQR = Q_3 - Q_1 = 9{,}00 - 8{,}30 = 0{,}70$

Límites para los Bigotes

Límite Inferior (LI):

$LI = Q_1 - 1{,}5 \cdot IQR = 8{,}30 - 1{,}5 \cdot 0{,}70 = 7{,}25$

Valor mínimo dentro del límite: Primer dato ≥ 7,25 → 7,33.

Límite Superior (LS):

$LS = Q_3 + 1{,}5 \cdot IQR = 9{,}00 + 1{,}5 \cdot 0{,}70 = 10{,}05$

Valor máximo dentro del límite: Último dato ≤ 10,05 → 9,83.

Valores Atípicos

Atípicos inferiores: Valores < 7,25 → 4,00; 4,48; 6,24; 6,99; 7,00.

Atípicos superiores: Valores >10,05 → Ninguno.

Nota: En algunos criterios, solo se marcan como atípicos los valores de

$Q_1 - 3 \cdot IQR$

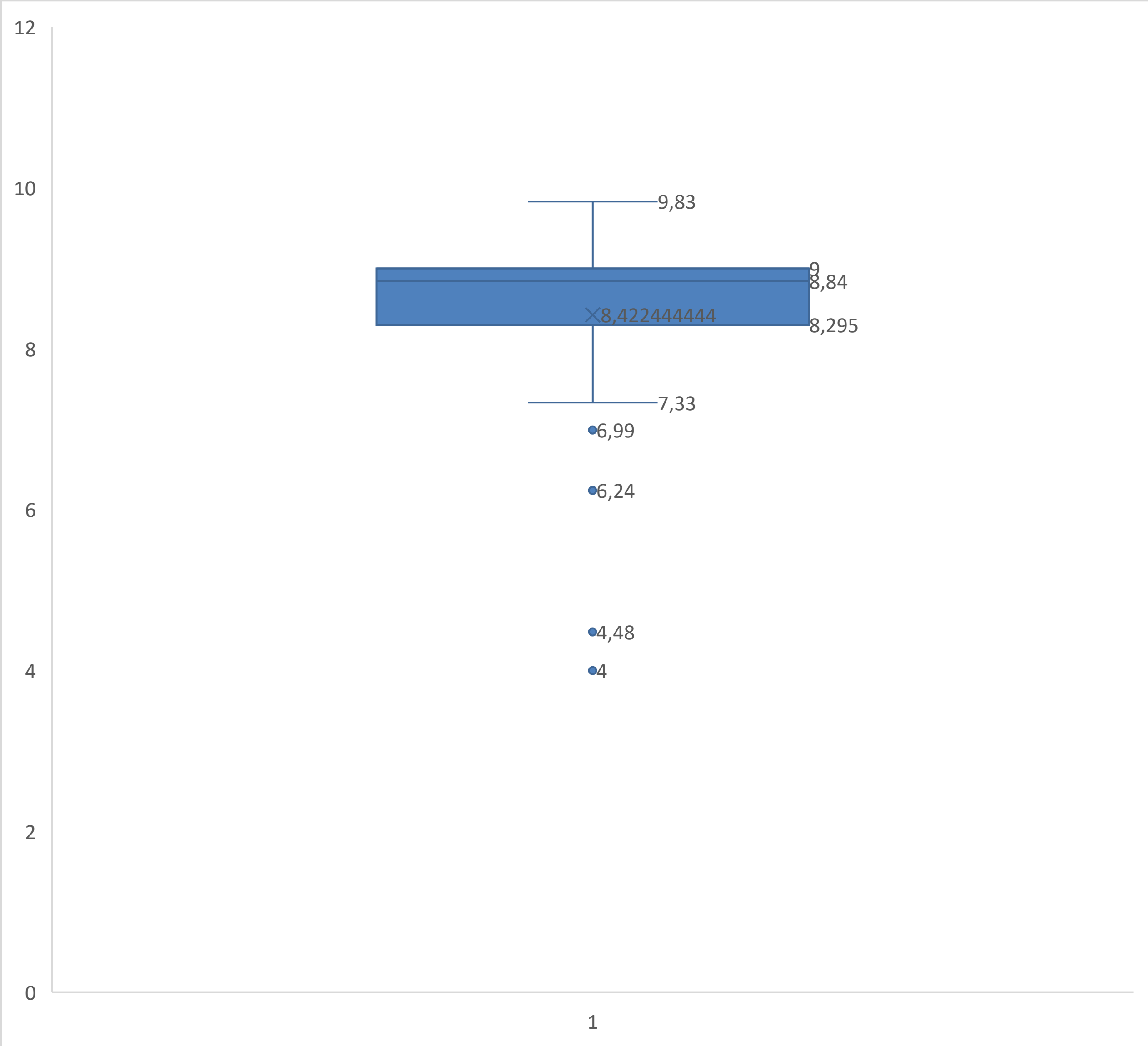
$Q_3 + 3 \cdot IQR$

Si aplicamos este criterio más estricto

$LI_{estricto} = Q_1 - 3 \cdot IQR = 8{,}30 - 3 \cdot 0{,}70 = 6{,}20$

Atípicos inferiores: Valores < 6,20 → 4,00; 4,48

Empleando Excel



Interpretación:

a) Tendencia Central

Mediana (8,84): El 50% de los estudiantes obtuvo una nota ≤ 8,84.

Caja (8,30–9,00): El 50% central de las notas está en este rango estrecho, lo que indica baja dispersión en la mayoría de los datos.

b) Dispersión y Variabilidad

IQR (0,70): Rango intercuartílico pequeño, confirmando concentración de datos.

Rango total (5,83): Amplio debido a valores atípicos bajos.

c) Valores Atípicos

4,00; 4,48; 6,24; 6,99; 7,00**:** Notas significativamente más bajas que el resto.

BILIOGRAFÍA

Cachuput, J., Suárez, M., Salguero, S., y Reyes, E., (2024). *Estrategias pedagógicas basadas en el enfoque constructivista para mejorar la comprensión de las matemáticas*. Reincisol, 3(6), pp. 4718-4742. https://doi.org/10.59282/reincisol.V3(6)4718-4742

Guangasi Gómez, E. M., Valencia Nuñez, E. R., Montoya Puglla, S. D., & Suárez Ibujés, M. O. (2024). *Exploring the impact of indebtedness on the financial profitability of the automotive sector: a statistical analysis and prediction of bankruptcy*. Runas. Journal of Education and Culture, 5(9), e240168. https://doi.org/10.46652/runas.v5i9.168

Ineval. (2024). *Prueba Ser Estudiante 2023-2024*. https://evaluaciones.evaluacion.gob.ec/BI/ser-estudiante-2/

Ministerio de Educación. (2023). *Reglamento General a la Ley Orgánica de Educación Intercultural*. Segundo suplemento N° 254-Registro oficial.

Moreira Parrales ML, Mejía Carrillo M de J, Suarez Ibujes MO, Torres Penafiel JS. *Gamification for learning mathematics in secondary school: Most effective gamified strategies to motivate students and improve their performance in mathematics*. Salud, Ciencia y Tecnología. 2024; 4:1016. https://doi.org/10.56294/saludcyt20241016

Olivo Solis, J. E. ., Murillo García, M. A., Suárez Ibujés, M. O., & Rizzo Orellana, E. B. (2025). *Una educación más innovadora y de mayor impacto a través de la inteligencia artificial, mediante el aprendizaje personalizado: transformando las estrategias de enseñanza en el nivel superior*. *Revista Social Fronteriza*, *5*(2), e–637. https://doi.org/10.59814/resofro.2025.5(2)637

Shao, S. (1980). *Estadística para economistas y administradores de empresas*. Herrero Hermanos.

Suárez, M. y Tapia, F. (2018). *Interaprendizaje de Estadística Básica*. http://repositorio.utn.edu.ec/handle/123456789/8696

Suárez-Ibujés, M. O., & Sánchez-Pozo, N. N. (2023). *Análisis de la influencia geográfica en la evaluación Ser Estudiante a través de estadística multivariante*. Prometeo Conocimiento Científico, 3(1), e27. https://doi.org/10.55204/pcc.v3i1.e27

Suárez-Ibujés, M. O., Hernández-Dávila, C. A., Peñafiel, E. J. A., & Villena-Atoche, C. A. (2024). *Utilización de juegos de razonamiento lógico para potenciar competencias matemáticas en estudiantes de bachillerato.* MQRInvestigar, 8(2), 2931–2950. https://doi.org/10.56048/MQR20225.8.2.2024.2931-2950

DATOS BIOGRÁFICOS DEL AUTOR

Mario Orlando Suárez Ibujés nació el 24 de marzo de 1978 en el barrio La Florida de Ibarra, Imbabura (Ecuador). Hijo de Segundo Suárez y Bertha Ibujés. Realizó sus estudios primarios en la Escuela Fiscal Mixta "Alejandro Pasquel Monge" de su localidad, donde destacó como Abanderado del Estandarte de la Escuela, Abanderado del Pabellón Nacional y Mejor Egresado. Posteriormente, cursó la educación secundaria en el Colegio "Teodoro Gómez de la Torre" de Ibarra, donde sobresalió académicamente: fue Mejor Alumno en Matemática durante sus tres últimos años, Abanderado del Estandarte del Colegio y Mejor Egresado. A nivel superior, se graduó como Licenciado en Ciencias de la Educación, con especialización en Física y Matemática, en la Universidad Técnica del Norte (UTN) de Ibarra, donde obtuvo el reconocimiento de Mejor Egresado.

Su formación de cuarto nivel incluye:

- Magíster en Gerencia de Proyectos Educativos y Sociales (UTN).
- Magíster en Estadística Aplicada (Universidad Politécnica Estatal del Carchi), graduado con Mención Honorífica.

Formación Doctoral en progreso:

- Doctorado PhD en Educación (Universidad Benito Juárez, México).
- Doctorado PhD en Educación e Innovación (Universidad de Investigación e Innovación de México).

Actualmente se desempeña como Asesor Educativo en la Coordinación Zonal 1-Educación de Ecuador. Además, es docente ocasional de posgrado en la UTN y miembro activo del Centro Cultural Antonio Ante, la Asociación de Maestros de Excelencia Educativa y el Colegio Profesional de Asesores Educativos del Ecuador. Tiene cuatro Doctorados Honoris Causa. Ha publicado 14 libros, 8 artículos en revistas indexadas, es autor de 5 rompecabezas llamados Poliprismas, un Juego Matemático en la Chakana y de 403 recursos educativos en repositorios digitales.

Trayectoria profesional:

- Docente de Matemática del Bachillerato Internacional (BI) y (EGB) en la Unidad Educativa. Teodoro Gómez de la Torre, en donde inicia su experiencia profesional docente a los 20 años de edad.
- Docente de Matemática en la Escuela Alejandro Pasquel Monge.
- Docente de Matemática de Educación General Básica (EGB) y Bachillerato General Unificado (BGU) en la Academia Militar San Diego.
- Docente de Matemática de EGB en la Unidad Educativa Mariano Suárez Veintimilla.

- Docente de Matemática de BGU y BI en la Unidad Educativa Ibarra.
- Director Distrital de Educación 10D02 Antonio Ante-Otavalo.
- Docente de la UTN en la Facultad de Ciencias Administrativas y Económicas (FACAE).
- Actual Docente ocasional en la Facultad de Posgrado de la UTN y Asesor Educativo en la Coordinación Zonal 1-Educación.

Producción Académica

Libros publicados (15 obras):

- Unidades para Producir Medios Instruccionales en Educación (coautor a los 24 años).
- Interaprendizaje Holístico de Matemática (autor).
- Hacia un Interaprendizaje Holístico de Álgebra y Geometría (autor).
- Matemática Recreativa (coautor).
- Interaprendizaje de Probabilidades y Estadística Inferencial Empleando Excel, Winstats y Graph (autor).
- Interaprendizaje de Estadística Básica (coautor).
- Probabilidades y Estadística empleando las TIC (autor).
- Matemática y sus aplicaciones empleando las TIC (coautor).
- Los Poliprismas y su aplicación en la enseñanza de la Matemática (autor).
- El PAPT en Cotacachi (coautor).
- Docentes en Iberoamérica: Reflexiones hacia la Excelencia Educativa (coautor).
- Hacia un interaprendizaje de Matemática Financiera (autor).
- Maestros de Excelencia Transformando la Educación en Iberoamérica Reflexiones y Desafíos (coautor).
- Estadística Descriptiva para Todos. Fundamentos y Aplicaciones-Volumen I (autor).
- Estadística Descriptiva para Todos. Fundamentos y Aplicaciones-Volumen II (autor).

https://isbnecuador.com/catalogo.php?mode=busqueda_menu&id_autor=6289

Artículos científicos (8 en revistas indexadas):

- Guía didáctica para el interaprendizaje de Trigonometría Básica empleando el Poliprisma. Revista el Investigador N° 4 de la Universidad Técnica del Norte.
  https://issuu.com/utnuniversity/docs/el-investigador-n04

- Análisis de la influencia geográfica en la evaluación Ser Estudiante a través de estadística multivariante. Prometeo Conocimiento Científico, 3(1), e27. https://doi.org/10.55204/pcc.v3i1.e27
- Gamification for learning mathematics in secondary school: Most effective gamified strategies to motivate students and improve their performance in mathematics. Salud, Ciencia y Tecnología. 2024; 4:1016. https://doi.org/10.56294/saludcyt20241016
- Exploring the impact of indebtedness on the financial profitability of the automotive sector: a statistical analysis and prediction of bankruptcy. Runas. Journal of Education and Culture, 5(9), e240168. https://doi.org/10.46652/runas.v5i9.168
- Utilización de juegos de razonamiento lógico para potenciar competencias matemáticas en estudiantes de bachillerato. MQRInvestigar, 8(2), 2931–2950. https://doi.org/10.56048/MQR20225.8.2.2024.2931-2950
- Estrategias pedagógicas basadas en el enfoque constructivista para mejorar la comprensión de las matemáticas. Reincisol, 3(6), pp. 4718-4742. https://doi.org/10.59282/reincisol.V3(6)4718-4742
- Una educación más innovadora y de mayor impacto a través de la inteligencia artificial, mediante el aprendizaje personalizado: transformando las estrategias de enseñanza en el nivel superior. *Revista Social Fronteriza*, *5*(2), e–637. https://doi.org/10.59814/resofro.2025.5(2)637
- Optimización del aprendizaje conceptual y práctico en matemáticas, física y química mediante la implementación de tecnologías digitales y estrategias de gamificación en la educación superior. Revista Social Fronteriza; 5(3): e707. https://doi.org/10.59814/resofro.2025.5(3)707

Obras literarias inéditas y material didáctico:

- Poliprisma 3.0, Poliprisma 4.0, Poliprisma 7.0, Poliprisma 9.0, Poliprisma 9.1.
- Juego Matemático en la Chakana.

Temas publicados en repositorios digitales

401 temas sobre temas educativos se encuentran publicados en:

https://www.researchgate.net/profile/Mario-Suarez-Ibujes

http://es.scribd.com/mariosuarezibujes

https://repositorio.utn.edu.ec/simple-search?filterquery=Su%C3%A1rez+Ibuj%C3%A9s%2C+Mario+Orlando&filtername=author&filtertype=equals

https://amee.ec/author/mario-amee/

Reconocimientos y Premios Nacionales e Internacionales

Doctorados Honoris Causa

- Doctor Honoris Causa por Remzion University of Jerusalem. Cartagena de Indias, Colombia, año 2025
  https://es.scribd.com/presentation/839031232/Doctor-Honoris-Causa-en-Remzion-University-of-Jerusalem
- Doctor Honoris Causa en Andragogía por parte de la Organización Internacional para la Inclusión y Calidad Educativa (OIICE). Perú-Cusco, año 2024.
  https://youtu.be/-KZ9pK44hBw
  https://www.scribd.com/document/839033529/Doctor-Honoris-Causa-en-Andragogia
- Doctor Honoris Causa otorgado por la Universidad del Norte de Tamaulipas. Perú-Lima, año 2024.
  https://es.scribd.com/document/708966545/Mario-Orlando-Suarez-Ibujes-Doctor-Honoris-Causa-UNT
  https://www.youtube.com/shorts/NA-hLFXaxqo
  https://www.youtube.com/watch?v=5cREEjBLBZw
- Doctor Honoris Causa por parte de la Organización Internacional para la Inclusión y Calidad Educativa OIICE. Perú- Cusco, año 2023.
  https://es.scribd.com/document/649433171/Galardon-a-La-Excelencia-Educativa-OIICE-Cusco-2023
  https://www.imbaburaenlinea.com/2023/06/08/docente-imbabureno-recibe-reconocimiento-internacional/?fbclid=IwAR0ls92m6pLJ_rPo6d5WQkibkLO7U65w87WJSn7gHKx-ZwaoMnJIKkLTRXo
  https://youtu.be/JJGzdiSlu6c

Premios Internacionales

- Premio a la Excelencia Gerencial otorgado por la Sociedad Internacional de Gerencia. Cartagena de Indias, Colombia, año 2025
  https://es.scribd.com/presentation/839031232/Doctor-Honoris-Causa-en-Remzion-University-of-Jerusalem
- Education Leadership Awards (Premio al Liderazgo en Educación) 2024 otorgado por parte de Christian Chambers Entrepreneurs. Perú-Lima, año 2024.
  https://youtu.be/Smsfu20MAoE

- Global Education Prize (Premio Global de Educación) 2024 otorgado por parte de la Asociación de Acreditación Internacional y Certificación de Entidades Privadas (AAICEP). Perú-Lima, año 2024.
  https://www.youtube.com/watch?v=B-qqUbNMCoQ
- Global Learning Awards (Premio Global de Aprendizaje) en la Categoría Educación Innovadora otorgado por la Cámara Iberoamericana de Educación (IBEROCAM). Lima-Perú, año 2024.
  https://www.youtube.com/watch?v=tenrlEKFpXY
- Premio Educa Latinoamérica 2024 en la categoría Educación de Excelencia otorgado por la Cámara Peruana de Desarrollo y Educación. Lima-Perú, año 2024.
  https://www.youtube.com/watch?v=Ok6alM5GXiU&t=15s
  https://www.scribd.com/document/717050277/Premio-Educa-Latinoamerica-2024
- Orden al Mérito Educativo y Cultural Magnus Docentis otorgado por parte de la Organización Internacional para la Inclusión y Calidad Educativa (OIICE). Perú-Cusco, año 2024.
  https://youtu.be/-KZ9pK44hBw
  https://www.scribd.com/document/839033529/Doctor-Honoris-Causa-en-Andragogia
- Galardón a la Excelencia Educativa Edición Cusco 2024 otorgado por parte de la Organización Internacional para la Inclusión y Calidad Educativa (OIICE). Perú-Cusco, año 2024.
  https://youtu.be/-KZ9pK44hBw
  https://www.scribd.com/document/839033529/Doctor-Honoris-Causa-en-Andragogia
- Colegiatura Oficial Internacional de Doctorado Honoris Causa otorgado por el Colegio Internacional de Doctores. Costa Rica-San José, año 2024
  https://es.scribd.com/document/708961975/Mario-Orlando-Suarez-Ibujes-Colegiatura-Internacional-de-Doctor-Honoris-Causa
- Educador de Eminencia otorgado por Universidad Ricardo Palma-Escuela de Marketing y Administración Comercial. Costa Rica-San José, año 2024
  https://es.scribd.com/document/708961975/Mario-Orlando-Suarez-Ibujes-Colegiatura-Internacional-de-Doctor-Honoris-Causa
- Orden Dorada Magisterial por parte de la Organización Internacional para la Inclusión y Calidad Educativa OIICE. Perú- Cusco, año 2023.
  https://es.scribd.com/document/649433171/Galardon-a-La-Excelencia-Educativa-OIICE-Cusco-2023
  https://youtu.be/JJGzdiSlu6c

- Galardón a la Excelencia Educativa Cusco 2023 por parte de la Organización Internacional para la Inclusión y Calidad Educativa OIICE. Perú- Cusco, año 2023.
  https://es.scribd.com/document/649433171/Galardon-a-La-Excelencia-Educativa-OIICE-Cusco-2023
  https://youtu.be/JJGzdiSlu6c

Distinciones Nacionales

- Mención Especial en Ciencias Básicas (Matemática), Premio Nacional otorgado por la VI Feria de Ciencia y Tecnología por haber triunfado con el Proyecto Multiprisma (Un rompecabezas tridimensional bicolor integrado por partes prismáticas). Ecuador-Quito, año 2001.
- Premio Nacional a la Excelencia Docente "Rita Lecumberri" en la categoría Educador Innovador otorgado por el Ministerio de Educación del Ecuador, año 2013.
  https://www.youtube.com/watch?v=fN614do_3lI
  https://es.scribd.com/doc/135847484/Premio-Rita-Lecumberri
- Premio Nacional Galardón Nacional Estatuilla "Nöus" por ser el ganador del VI Concurso Nacional de Excelencia Educativa, otorgando por la Fundación para la Integración y Desarrollo de América Latina (FIDAL) y la Revista Edu@news. Ecuador-Quito, año 2014. Se encuentra publicado en
  https://www.youtube.com/watch?v=hilX-jZUM8g
  https://www.youtube.com/watch?v=l-H_rkSZdbs
- Condecoración al Mérito Educativo "Alfredo Pérez Guerrero" otorgado por el Gobierno Autónomo Descentralizado Municipal de San Miguel de Ibarra. Ibarra-Ecuador, año 2025.
  https://youtu.be/KNLH6qQeYQ8
  https://www.facebook.com/share/195kUYAsPF/
- Condecoración "Medalla Julio Miguel Aguinaga" al mérito educativo otorgado por el Gobierno Autónomo Descentralizado Municipal de Antonio Ante. Atuntaqui-Ecuador, año 2024.
  https://www.youtube.com/watch?v=3VqYG7lLhjQ
  https://www.youtube.com/watch?v=99t2paNDI6I
  https://fb.watch/qA4Z5YGhnR/
- Diploma y Placa de Reconocimiento por trayectoria docente y aporte invaluable al Magisterio Fiscal por la Coordinación Zonal 1 del Ministerio de Educación del Ecuador. Antonio Ante-Ecuador, año 2024.
  https://es.scribd.com/document/839014751/Diploma-y-Placa-de-Reconocimiento-CZ1#logout

- Diploma de Reconocimiento por la destacada contribución a la innovación educativa otorgado por la Dirección Distrital 10D02 Antonio Ante-Otavalo-Educación. Otavalo-Ecuador, año 2024. https://www.facebook.com/share/v/RNti5UtrqkDQC8QC/?mibextid=WC7FNe
- Diploma de Reconocimiento al mérito profesional por el aporte a la educación otorgado por la Unidad Educativa "Sarance". Otavalo-Ecuador, año 2024. https://www.facebook.com/share/p/YTyLodqDJpfLHahv/?mibextid=xfxF2i
- Placa de reconocimiento por excelente trayectoria académica y profesional otorgado por el área de Matemática y Física de la Unidad Educativa Teodoro Gómez de la Torre. Ecuador-Ibarra, año 2024 https://es.scribd.com/document/839071664/Reconocimientos-Profesionales-Hasta-Marzo-de-2025
- Placa de Homenaje de Gratitud y Reconocimiento, otorgado por la 1ra Cohorte de la Maestría en Pedagogía Mención Currículo Modalidad en Línea de la Universidad Técnica del Norte. Ecuador-Ibarra, año 2023.
- Diploma de reconocimiento por el aporte a la investigación científica y tecnológica al haber contribuido con publicaciones científicas durante el año 2017. Universidad Técnica del Norte. Ecuador- Ibarra, año 2018.
- Diploma por el valioso aporte al cumplimiento de los objetivos institucionales, otorgado por la Unidad Educativa Ibarra. Ecuador- Ibarra, año 2014.
- Diploma y placa de reconocimiento por la excelente trayectoria como docente investigador y destacado profesor universitario. Universidad Técnica del Norte. Asociación de Profesores de la Facultad de Ciencias Administrativas y Económicas. Ecuador- Ibarra, año 2013.
- Estatuilla "El Pensador" al Mérito Académico. Asociación General de Profesores de la Universidad Técnica del Norte. Ecuador-Ibarra, año 2013.
- Diploma como Profesor tutor de estudiantes ganadores de Concursos Intercolegiales de Matemática. Academia Militar "San Diego". Ecuador-Ibarra, año 2008.
- Diploma de Honor por haber aportado positivamente al desarrollo académico de Academia Militar "San Diego". Academia Militar "San Diego". Ecuador-Ibarra, año 2005.
- Diploma como Asesor de proyectos ganadores en la Primera Feria Binacional de Ciencia y Tecnología Ecuador Colombia. Unidad Educativa Experimental "Teodoro Gómez de la Torre". Ecuador-Ibarra, año 2005.

- Mejor Trabajo de Investigación. Certificado de la UTN-Centro Universitario de Investigación Científica y Tecnológica, por haber presentado la Tesis "Interaprendizaje de poliedros irregulares de bases paralelas empleando al Multiprisma" en la Casa Abierta. Ecuador- Ibarra, año 2003.
- Diploma por haber asesorado satisfactoriamente en el VII Concurso Provincial de Matemática, otorgado por el Colegio Nacional Ibarra. Ecuador- Ibarra, año 2003 https://es.scribd.com/document/839071664/Reconocimientos-Profesionales-Hasta-Marzo-de-2025

MÉRITOS ESTUDIANTILES

PRODUCCIÓN ACADÉMICA

RECONOCIMIENTOS DESTACADOS