



Universidad Nacional de Mar del Plata

Facultad de Ingeniería



Estadística Descriptiva

Repaso de los principales conceptos
teóricos:

¿Qué es la estadística?

“El contenido de la estadística moderna incluye la recopilación, presentación y caracterización de la información con el fin de auxiliar tanto en el análisis de datos como en el proceso de toma de decisiones”

Berenson y Levine, Estadística Básica en administración. (1992)

Algunos usos de la estadística...

Si nos ocupa el análisis de uno o varios conjuntos de datos de la **misma variable**, contamos con métodos gráficos y numéricos que reúne la **Estadística descriptiva**.

Resumiendo: La estadística descriptiva se ocupa de la organización, presentación y análisis de la información.

Algunas definiciones

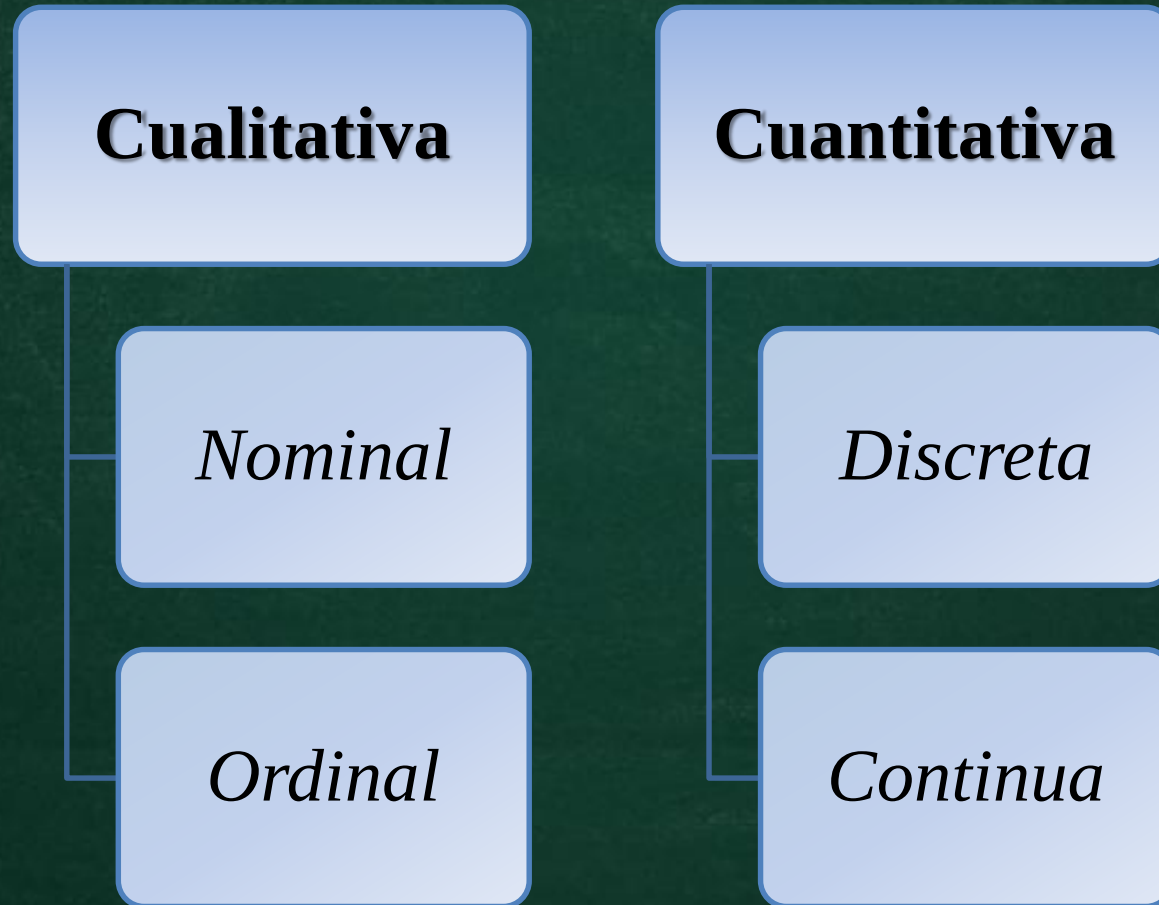
- ▣ *Población*: es el conjunto sobre el que se centra el objetivo de un análisis o investigación estadística. Esta compuesta por *unidades elementales* con características comunes observables.
- ▣ Una *unidad elemental* es cada objeto o sujeto que observamos de la población.
- ▣ Una *muestra* es un subconjunto «representativo» de unidades elementales tomadas de la población.
- ▣ Aquella característica que se observa o se mide sobre las unidades elementales, se denomina *variable estadística*.

Tipos de Variables

Variables: *Cualitativas* y *Cuantitativas*.

- ▣ Las variables *cualitativas* son aquellas que permiten la expresión de una característica, una categoría, un atributo o una cualidad de los elementos de estudio.
- ▣ Las variables *cuantitativas* son aquellas cuyos datos son de tipo numérico.

Clasificación de las Variables



¿Cómo organizar los datos?

Series Simples

Series de Frecuencias

Intervalos de Clases

Gráficos



¿Cómo organizar los datos?

Variables Cualitativas

- Barras Simples
- Barras Proporcionales
- Barras Agrupadas
- Diagramas Sectoriales

Variables Cuantitativas Discretas

- Bastones

Variables Cuantitativas Continuas

- Histograma
- Polígono de Frecuencias Simples
- Polígono de Frecuencias Acumuladas

Estadísticos

En todo análisis y/o interpretación de datos es necesario disponer de «valores» numéricos para extraer y resumir las principales características de los mismos.

Existen diversas medidas descriptivas que representan las propiedades de **tendencia central**, **dispersión** y **forma**.

Estadísticos

▣ Centralización

- ✓ Indican valores con respecto a los que los datos parecen agruparse.
 - ✓ Media, mediana y moda

▣ Posición

- ✓ Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.
 - ✓ Cuartiles, deciles, percentiles

▣ Dispersión

- ✓ Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización.
 - ✓ Rango, Varianza, Desviación Estándar, Coeficiente de Variación.

Medidas de Centralización

En la mayoría de los casos, **el conjunto de datos** obtenidos, ya sea de una muestra o de una población, **tienden a reunirse alrededor de un valor central**. De esta manera, es posible obtener un valor típico o representativo de todo el conjunto de datos, **el cual se denomina medida de tendencia central**.

Las medidas de tendencia central más representativas son:

- ✓ Media aritmética,
- ✓ Mediana,
- ✓ Moda.

Cálculo analítico para determinar las medidas de tendencia central más representativas:

- ✓ Media aritmética.
- ✓ Mediana.
- ✓ Moda.

Series Simples

Series de Frecuencias

Intervalos de Clases

Datos organizados:

Series Simples

Media Aritmética

- La **media aritmética** es la medida de centralización más común de un grupo de datos.

Serie Simple:

Si las observaciones de una muestra de tamaño n son $x_1, x_2, \dots, x_n$ entonces la media muestral se define como:

$$\overline{X} = \frac{\sum_{i=1}^n x_i}{n}$$

donde :

$\overline{X}$: media aritmética de la muestra

n : total de datos de la muestra

x_i : dato de la variable

$\sum_i x_i$: suma de todos los valores de la muestra

Datos organizados:

Series de Frecuencias

Media Aritmética

Cuando se agrupan los datos en una *Tabla de Frecuencias*, sin construir intervalos, se calcula la media aritmética mediante la siguiente formula:

X	f_i
x_1	f_1
x_2	f_2
x_3	f_3
.	.
.	.
x_i	f_i
.	.
.	.
x_k	f_k

n

Media Aritmética

Cuando se agrupan los datos en una *Tabla de Frecuencias*, sin construir intervalos, se calcula la media aritmética mediante la siguiente formula:

$$\overline{X} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

donde :

$\overline{X}$: media aritmética de la muestra

n : total de datos de la muestra

x_i : dato de la variable

f_i : frecuencia absoluta para cada valor de la variable

X	f _i
x ₁	f ₁
x ₂	f ₂
x ₃	f ₃
.	.
.	.
x _i	f _i
.	.
.	.
x _k	f _k

n

Datos organizados:

Intervalos de Clases

Media Aritmética

Cuando se agrupan los datos en *Intervalos de clases*, se calcula la media aritmética mediante la siguiente formula:

$$\bar{X} = \frac{\sum_{i=1}^k x_{m_i} f_i}{n}$$

donde :

$\bar{X}$: media aritmética de la muestra

n : total de datos de la muestra

x_{m_i} : marca de clase de i -ésimo intervalo. Corresponde al punto medio del intervalo

f_i : frecuencia absoluta para cada valor de la variable

Intervalos	x_{mi}	f_i
$[a_1, a_2)$	X_{m1}	f_1
$[a_2, a_3)$	X_{m2}	f_2
.	.	.
.	.	.
$[a_{k-1}, a_k)$	x_{mk}	f_k
		$\frac{\quad}{n}$

Mediana

La mediana es el valor de variable donde la muestra se divide en dos partes iguales, es decir, es aquel valor que deja el 50% de las observaciones por debajo de él y el otro 50% por encima de él.

Serie Simple:

Si las observaciones de una muestra de tamaño n son $x_1, x_2, \dots, x_n$ entonces la media muestral se define como:

$$Me = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Mediana

$$Me = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Ejemplo n (cantidad de datos) impar

$$x_1, x_2, x_3, x_4, x_5$$

Mediana

$$Me = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Ejemplo n (cantidad de datos) impar

$$x_1, x_2, x_3, x_4, x_5 \Rightarrow M_e = x_3$$

Mediana

$$Me = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Ejemplo n (cantidad de datos) impar

$$x_1, x_2, x_3, x_4, x_5 \Rightarrow M_e = x_3$$

Ejemplo n (cantidad de datos) par

$$x_1, x_2, x_3, x_4, x_5, x_6 \Rightarrow M_e = \frac{x_3 + x_4}{2}$$

Mediana

$$Me = \begin{cases} x_{(n+1)/2} & \text{si } n \text{ es impar} \\ \frac{x_{n/2} + x_{(n/2)+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

Ejemplo n (cantidad de datos) impar

$$x_1, x_2, x_3, x_4, x_5 \Rightarrow M_e = x_3$$

Ejemplo n (cantidad de datos) par

$$x_1, x_2, x_3, x_4, x_5, x_6 \Rightarrow M_e = \frac{x_3 + x_4}{2}$$

Mediana

Cálculo de la Mediana para datos agrupados en una Tabla de Frecuencias:

La mediana es aquel valor de la variable cuya Frecuencia Absoluta Acumulada (F_a) es inmediatamente mayor a la mitad de las observaciones.

El cálculo de la mediana para datos agrupados en una tabla de frecuencias se realiza mediante el siguiente procedimiento:

1. Hallar $n/2$, donde n es la cantidad de datos.
2. Ubicar el valor observado cuya frecuencia absoluta acumulada contiene a $n/2$.

Mediana

Cálculo de la Mediana para datos agrupados en una Tabla de Frecuencias:

Ejemplo correspondiente a las edades de los alumnos del segundo cuatrimestre de 2021 que cursan estadística básica.

Edad	f_i	Fa
18	3	3
19	3	6
20	8	14
21	5	19
22	4	23
23	2	25
24	3	28
26	1	29
27	1	30

Me →



$$\frac{n}{2} = \frac{30}{2} = 15$$

El 50 % de los alumnos del 2do cuatrimestre de 2021, tienen a lo sumo 21 años.

Mediana

Cálculo de la Mediana para datos agrupados en Intervalos de Clase:

$$\frac{n}{2} = \frac{30}{2} = 15$$

Intervalos de Edad	f_i	F_a
[18 – 20)	6	6
[20 – 22)	13	19
[22 – 24)	6	25
[24 – 26)	3	28
[26 – 28)	2	30
	$n = 30$	

$$Me = 20 + \frac{\frac{30}{2} - 6}{13} * 2 = 21,3846$$



frecuencia absoluta acumulada inmediatamente mayor a la mitad de las observaciones

donde:

L_{inf} = Límite inferior del primer intervalo cuya F_a es mayor a $n/2$.

F_{aa} = Frecuencia acumulada del intervalo anterior al primer intervalo cuya F_a es mayor a $n/2$.

f_i = Frecuencia absoluta del primer intervalo cuya F_a es mayor a $n/2$.

a = Amplitud de los intervalos

$$Me = L_{inf} + \frac{\frac{n}{2} - F_{aa}}{f_i} \times a$$

Moda

Se denomina moda de un conjunto de datos al valor que más se presenta, es decir, el atributo o el valor de mayor frecuencia. La moda se representa por Mo.

Serie Simple:

Ejemplo 1: 1, 3, 3, 3, 3, 4, 5, 6, 6, 6, 8, 9, 10

La moda es 3.

Ejemplo 2: 1, 3, 3, 3, 3, 3, 4, 4, 5, 5, 6, 6, 6, 6, 6, 8, 9, 10

Las modas son 3 y 6.

Ejemplo 3: 1, 2, 3, 4, 7, 8, 9, 19

En este caso, no hay moda.

Moda

Se denomina moda de un conjunto de datos al valor que más se presenta, es decir, el atributo o el valor de mayor frecuencia. La moda se representa por M_o .

Cálculo de la Media Aritmética para datos agrupados en una Tabla de Frecuencias:

Ejemplo correspondiente a las edades de los alumnos del primer cuatrimestre de 2018 que cursan estadística básica.

$M_o = 20$ años

20 años es la edad más frecuente entre los alumnos del 2do cuatrimestre de 2021.

Edad	f_i
18	3
19	3
20	8
21	5
22	4
23	2
24	3



Moda

Cálculo de la Mediana para datos agrupados en Intervalos de Clase:

$$Mo = L_{\text{inf}} + \frac{d_1}{d_1 + d_2} \times a$$

donde:

L_{inf} = Límite inferior del intervalo que tiene mayor frecuencia absoluta (intervalo modal).

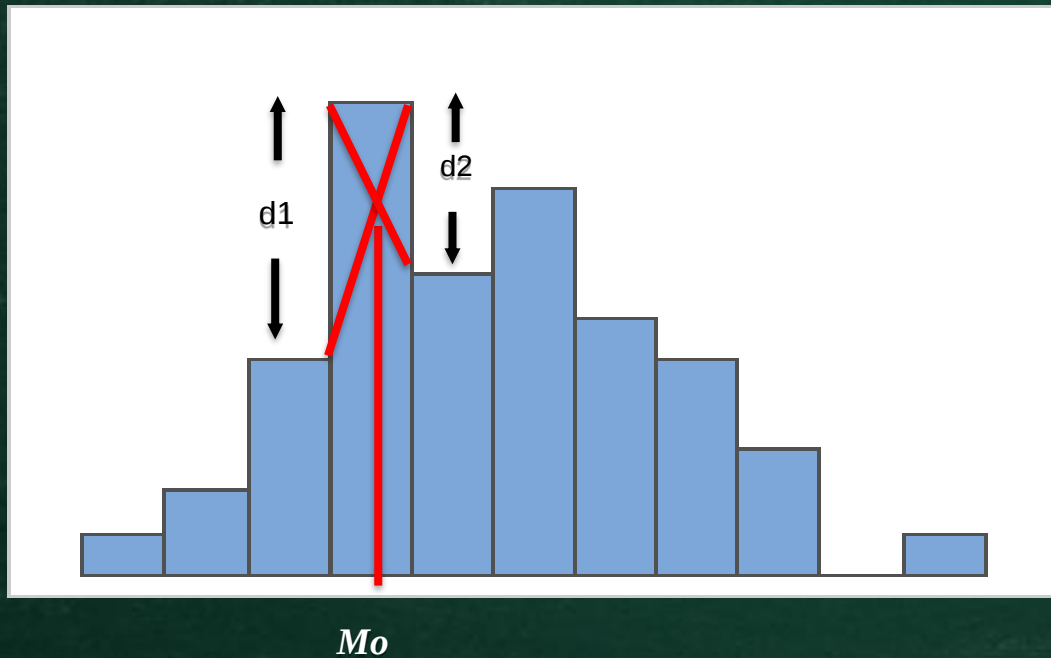
d_1 = Diferencia entre las frecuencias absolutas del intervalo modal y el intervalo pre-modal.

d_2 = Diferencia entre las frecuencias absolutas del intervalo modal y el intervalo post-modal.

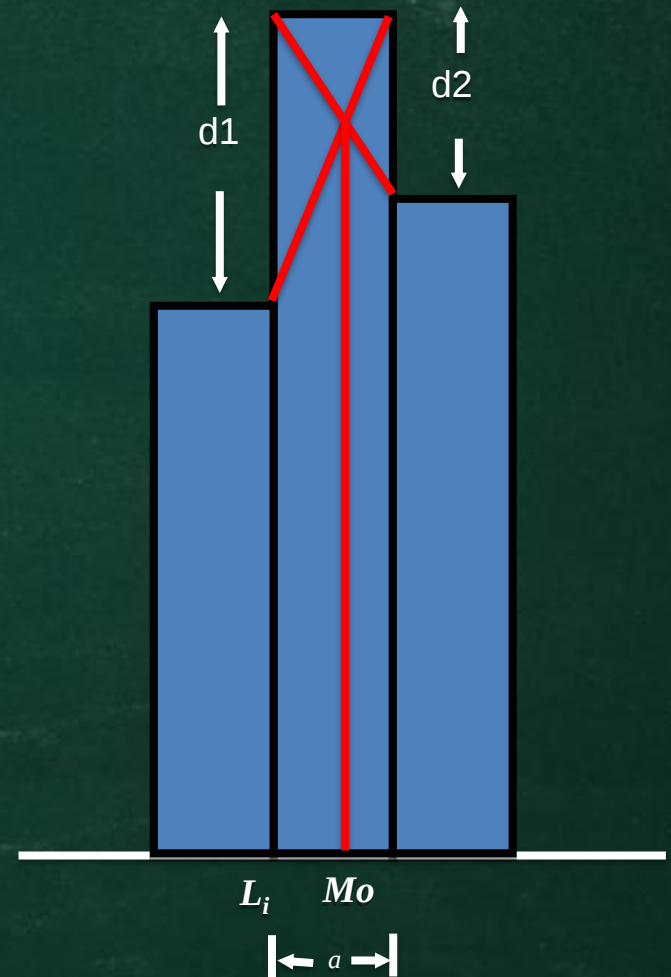
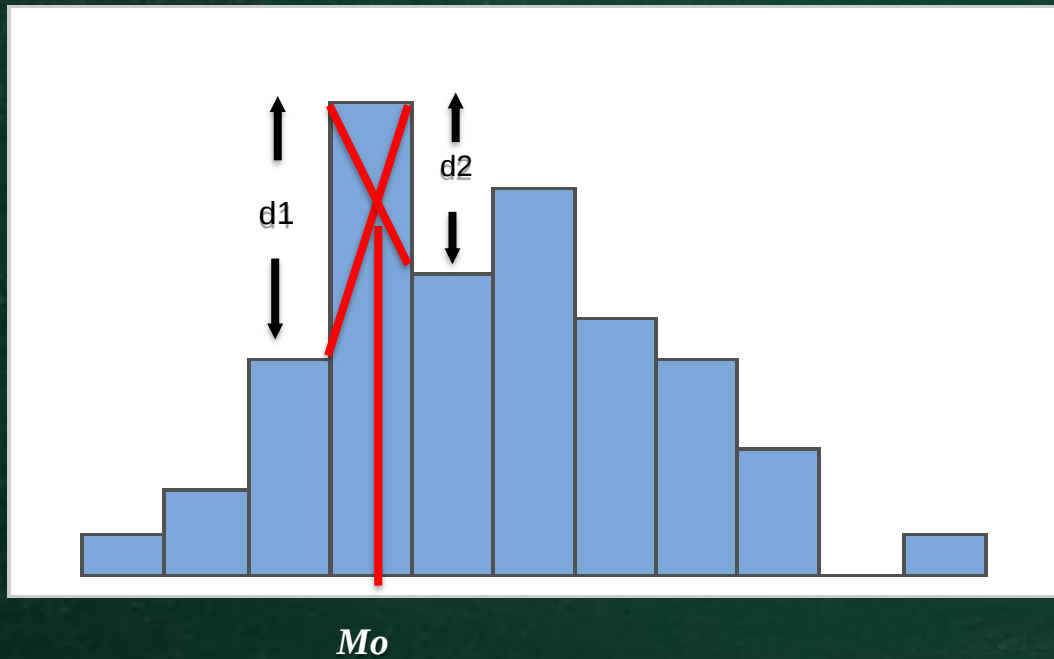
a = Amplitud de los intervalos

Representación Gráfica de la Moda

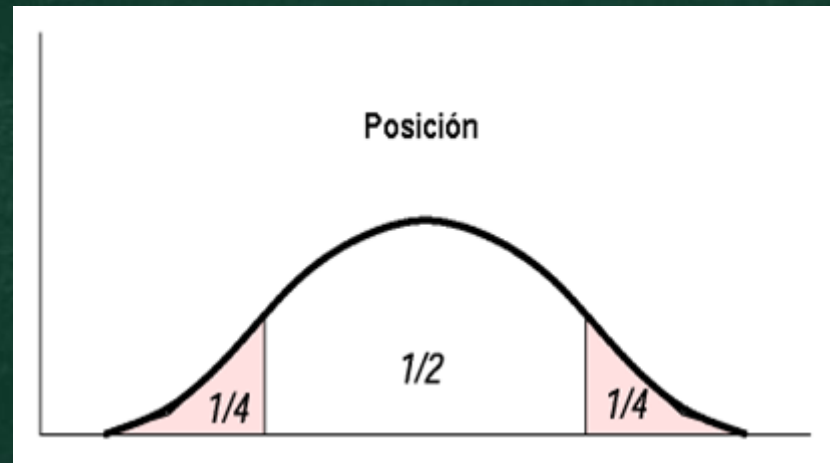
La siguiente figura muestra de forma simplificada cómo determinar gráficamente la moda para datos agrupados por intervalos.



Representación Gráfica de la Moda



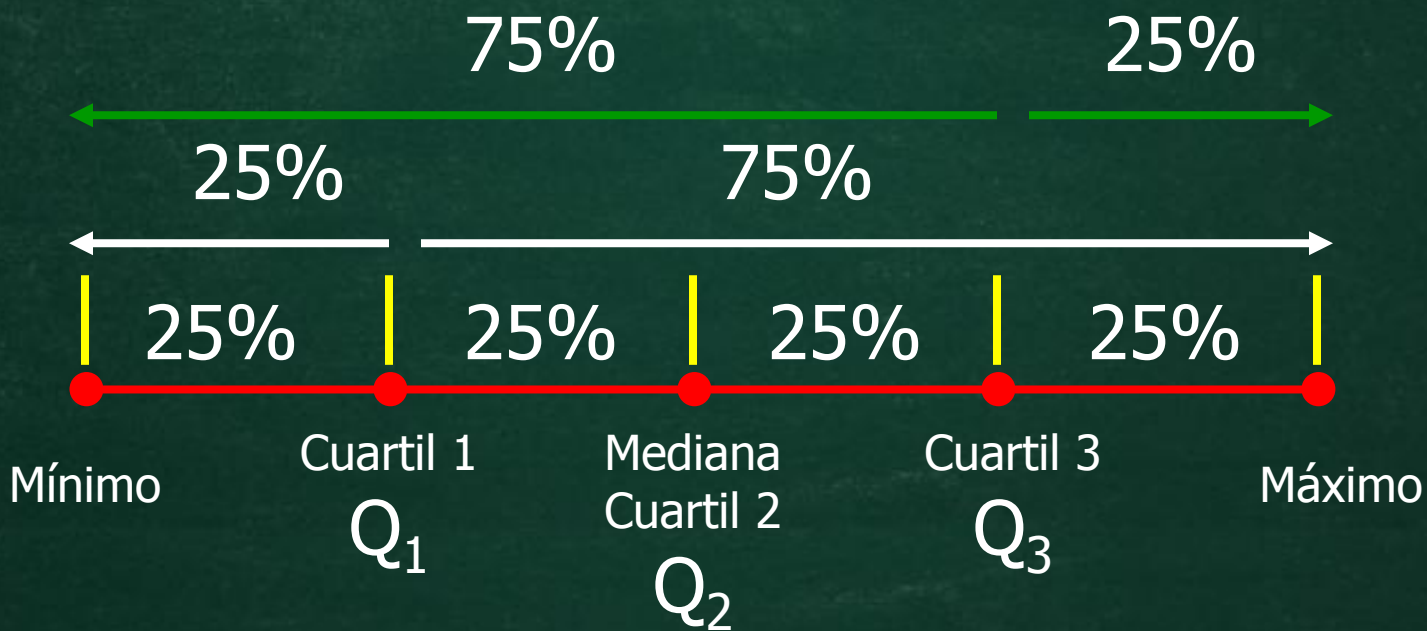
Medidas de Posición



- ✓ Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.
 - ✓ Cuartiles, deciles, percentiles

Cuartiles

Los cuartiles (Q_k) son valores que fraccionan la distribución de los datos en cuatro partes iguales. Existen tres cuartiles y cada una de las partes representa un 25% de los datos.



Cuartiles

Los cuartiles (Q_k) son valores que fraccionan la distribución de los datos en cuatro partes iguales. Existen tres cuartiles y cada una de las partes representa un 25% de los datos.

Serie Simple: Si las observaciones de una muestra de tamaño n son $x_1, x_2, \dots, x_n$ entonces el cuartil j (para $j = 1, 2, 3$) se define como:

$$Q_j = \begin{cases} x_{(n+1)j/4} & \text{si } n \text{ es impar} \\ \frac{x_{nj/4} + x_{(nj/4)+1}}{2} & \text{si } n \text{ es par} \end{cases}$$

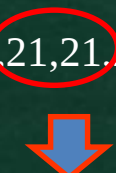
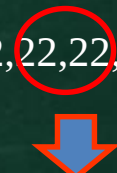
Cuartiles

Cálculo de los Cuartiles para datos dados por una Serie Simple:

Ejemplo correspondiente a las edades de los alumnos del segundo cuatrimestre de 2021 que cursan estadística básica.

24	19	26	19	22	21	19	24	20	23
22	21	22	23	20	20	18	24	20	18
22	18	20	21	20	20	21	20	21	27

18,18,18,19,19,19,20,20,20,20,20,20,20,20,20,21,21,21,21,21,22,22,22,22,23,23,24,24,24,26,27

$q_1=20$
 $q_2=21$
 $q_3=22$

Cuartiles

Cálculo de los Cuartiles para datos agrupados en Serie de Frecuencias:

Edad	f_i	Fa
18	3	3
19	3	6
20	8	14
21	5	19
22	4	23
23	2	25
24	3	28
26	1	29
27	1	30

$q_1 = 20$ años

$q_3 = 22$ años

$$j \cdot \frac{n}{4} = j \cdot \frac{30}{4}$$

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

$$3 \cdot \frac{n}{4} = 3 \cdot \frac{30}{4} = 22,5$$

Cuartiles

Cálculo de los Cuartiles para datos agrupados en Intervalos de Clase:

Intervalos de Edad	fi	Fa
[18 – 20)	6	6
[20 – 22)	13	19
[22 – 24)	6	25
[24 – 26)	3	28
[26 – 28)	2	30
	<i>n</i> = 30	

$$1 \times \frac{n}{4} = \frac{30}{4} = 7,5$$

$$3 \times \frac{n}{4} = 3 \times \frac{30}{4} = 22,5$$

$$Q_j = L_{\text{inf}} + \frac{j \times \frac{n}{4} - F_{aa}}{f_i} \times a$$

$$q_1 = 20 + \frac{\frac{30}{4} - 6}{13} \times 2 = 20,2307$$

$$q_3 = 22 + \frac{3 \cdot \frac{30}{4} - 19}{6} \times 2 = 23,166$$

L_{inf} = Límite inferior del primer intervalo cuya F_a es mayor a $j \cdot n/4$

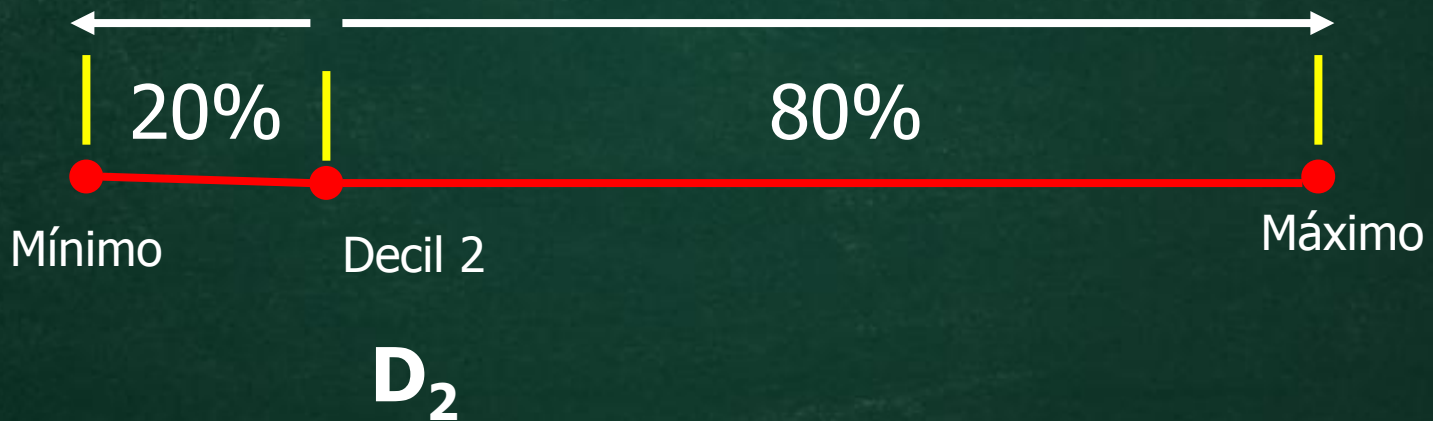
F_{aa} = Frecuencia acumulada del intervalo anterior al primer intervalo cuya F_a es mayor a $j \cdot n/4$.

f_i = Frecuencia absoluta del primer intervalo cuya F_a es mayor a $j \cdot n/4$

a = Amplitud de los intervalos.

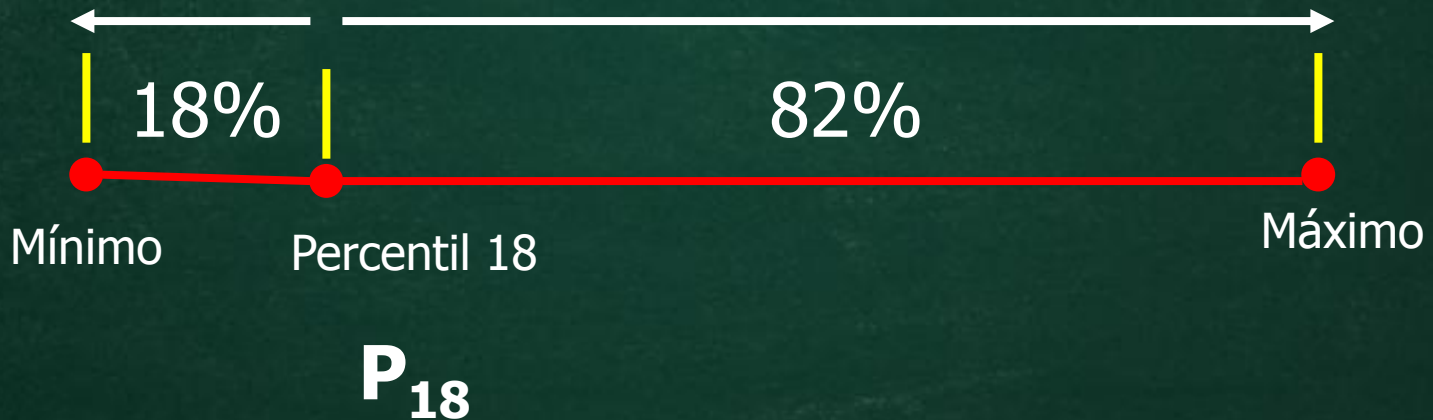
Deciles

Cuando se divide un conjunto ordenado de datos en diez partes iguales, los puntos de división se conocen como deciles.

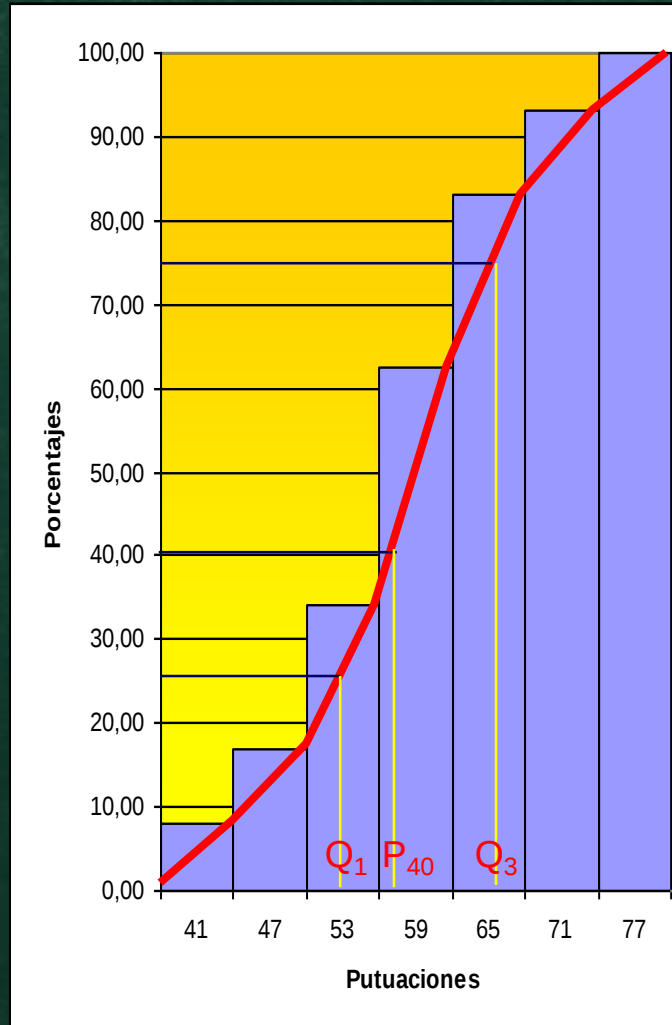


Percentiles

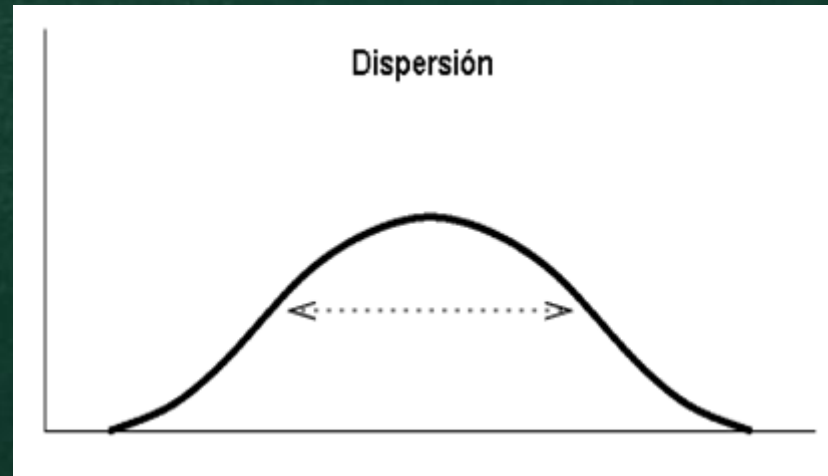
Cuando se divide un conjunto ordenado de datos en cien partes iguales, los puntos de división se conocen como percentiles.



Mediana y Cuartiles representados en el polígono de frecuencias acumuladas



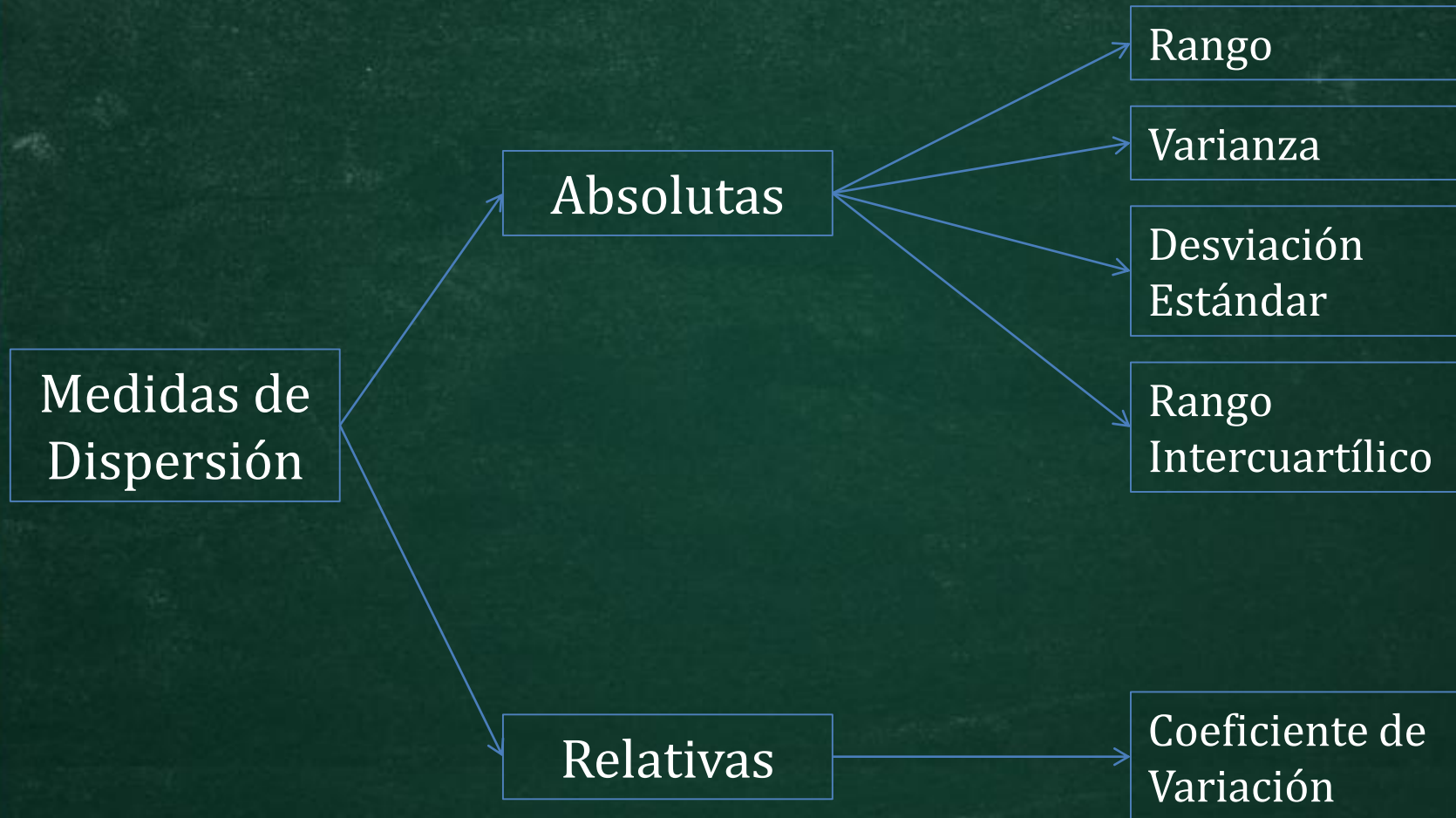
Medidas de Dispersión



Medidas de Dispersión

Además de las medidas de tendencia central que posibilitan la representación del conjunto de datos por medio de un valor, es necesario conocer la variabilidad o la dispersión que los datos pueden tener en relación a una medida de tendencia central.

Medidas de Dispersión



Rango

El rango se define como la diferencia entre la observación más grande y la más pequeña :

$$r = x_{\max} - x_{\min}$$

Rango Intercuartílico (RIC)

$$RIC = Q_3 - Q_1$$

Los valores extremos NO influyen en el conjunto de datos.

Varianza

Para el conjunto de datos $x_1, x_2, \dots, x_n$ de una población de tamaño N . **Las diferencias de cada dato y la media, determinan los desvíos o desviaciones.** Dado que la suma de estas desviaciones es cero, se utiliza como medida de variabilidad el promedio de los cuadrados de tales desvíos.

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}$$

(1)

$$\sigma^2 = \frac{\sum_{i=1}^k (x_i - \bar{X})^2 f_i}{N}$$

(2)

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}$$

(3)

$$s^2 = \frac{\sum_{i=1}^k (x_i - \bar{X})^2 \cdot f_i}{n-1}$$

(4)

Varianza Poblacional

siendo N el tamaño de la población.

Para datos sin agrupar (1) y agrupados (2)

Varianza muestral

siendo n el tamaño de la muestra.

Para datos sin agrupar (3) y agrupados (4)

Si los datos se agrupan por intervalos, usamos x_{mi} en lugar de x_i

Desvío estándar muestral

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



Para datos sin agrupar

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



Para datos agrupados
por frecuencias

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^k (x_{mi} - \bar{X})^2 f_i}$$



Para datos agrupados
por Intervalos

Coeficiente de variación

El **coeficiente de variación** (CV) es una medida que relaciona la desviación estándar con la media aritmética para determinar qué tan homogénea o dispersa es la información.

$$CV = \frac{S}{\bar{X}}$$

- Mide el grado de variabilidad en una muestra o población.
- Está desprovisto de unidades. Permite comparar la variabilidad entre distintas variables y poblaciones.
- El valor expresado en términos porcentuales, se llama **coeficiente de variación porcentual**.

$$CV\% = \frac{S}{\bar{X}} \times 100\%$$

Consideraremos poca variabilidad, si el $CV\%$ es a lo sumo del 30 %

Coeficiente de variación: Comparación entre distintas variables y poblaciones.

Ejemplo: Si tenemos el peso de 5 pacientes (70, 60, 56, 83 y 79 Kg) cuya media es de 69,6 kg. y su desviación estándar (s) = 10,44 y la presión arterial de los mismos (150, 170, 135, 180 y 195 mmHg) cuya media es de 166 mmHg y su desviación estándar de 21,3. La pregunta sería: ¿qué distribución es más dispersa, el peso o la presión arterial? Si comparamos las desviaciones estándar observamos que la desviación estándar de la presión arterial es mucho mayor; sin embargo, no podemos comparar dos variables que tienen escalas de medidas diferentes, por lo que calculamos los coeficientes de variación:

$$CV \text{ de la variable peso} = \frac{10,44}{69,6} = 15\%$$

$$CV \text{ de la variable presión} = \frac{21,30}{166} = 12,8\%$$

Cuestionario

Actividades clase Nº 2

1. ¿QUÉ ES, CÓMO se calcula (según estén los datos agrupados o no), QUÉ SIGNIFICADO TIENE en el contexto del problema, la media, la mediana y la moda?
2. Aprenda el manejo de su calculadora en el modo estadístico (SD) para todos los cálculos.
3. ¿Cuál es el gráfico adecuado en cada caso para ubicar la media, la mediana y la moda?
4. ¿QUÉ SON, CÓMO se calculan (según estén los datos agrupados o no) y CÓMO SE INTERPRETAN los cuartiles? ¿Y los percentiles?
5. ¿Cómo se construye un gráfico de caja y cómo se interpreta? ¿Cuáles son las dos formas de construcción de los gráficos de caja? ¿Cuál brinda mayor información?
6. ¿Cómo analizar la simetría de una distribución de datos en forma gráfica y analítica?
7. ¿Cuáles son las medidas de dispersión y cómo se calculan, según estén los datos agrupados o no?
8. ¿Cuándo puede afirmarse que una medida (como la media o la mediana) resume los datos?

Ejemplos

La siguiente tabla contiene la información de una consulta que se le hizo a 500 alumnos de la facultad de ingeniería para conocer la frecuencia con la que acudieron a la biblioteca, durante el último cuatrimestre, para acceder a material bibliográfico.

Número de visitas	Frecuencia (Cantidad de alumnos)
0	210
1	178
2	68
3	24
4	14
5	6

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

a) En primer lugar debemos definir la variable estadística bajo estudio.

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

a) En primer lugar debemos definir la variable estadística bajo estudio.

En este ejemplo la variable bajo estudio es “*número de visitas a la biblioteca*”

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

a) En primer lugar debemos definir la variable estadística bajo estudio.

En este ejemplo la variable bajo estudio es “*número de visitas a la biblioteca*”

b) En segundo lugar debemos clasificarla:

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

a) En primer lugar debemos definir la variable estadística bajo estudio.

En este ejemplo la variable bajo estudio es “*número de visitas a la biblioteca*”

b) En segundo lugar debemos clasificarla:

La variable bajo estudio es una variable cuantitativa discreta

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

a) En primer lugar debemos definir la variable estadística bajo estudio.

En este ejemplo la variable bajo estudio es “*número de visitas a la biblioteca*”

b) En segundo lugar debemos clasificarla:

La variable bajo estudio es una variable cuantitativa discreta

Finalmente determinamos los valores numéricos solicitados en el ejercicio.

Mínimo: 0, **Máximo:** 5. La cantidad de visitas que hicieron los alumnos encuestados oscilan entre 0 y 5.

A los valores numéricos debemos darle su interpretación en función de la variable estadística.

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

Media Aritmética:

Número de visitas	Frecuencia (Cantidad de alumnos)
0	210
1	178
2	68
3	24
4	14
5	6

Media Aritmética

Cuando se agrupan los datos en una *Tabla de Frecuencias*, sin construir intervalos, se calcula la media aritmética mediante la siguiente formula:

$$\overline{X} = \frac{\sum_{i=1}^k x_i f_i}{n}$$

donde :

$\overline{X}$: media aritmética de la muestra

n : total de datos de la muestra

x_i : dato de la variable

f_i : frecuencia absoluta para cada valor de la variable

X	f _i
x ₁	f ₁
x ₂	f ₂
x ₃	f ₃
.	.
.	.
x _i	f _i
.	.
.	.
x _n	f _n

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q1 y q3.

Solución/

Media Aritmética:

$$\overline{X} = \frac{\sum_{i=1}^k x_i f_i}{n} = 0.94$$

Número de visitas	Frecuencia (Cantidad de alumnos)
0	210
1	178
2	68
3	24
4	14
5	6

Interpretación: Si se suman todas las visitas efectuadas durante el cuatrimestre a la biblioteca y se reparten equitativamente entre los estudiantes entrevistados, daría como resultado casi una visita por persona.

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q_1 y q_3 .

Solución/

Mediana:

Número de visitas	Frecuencia (Cantidad de alumnos)
0	210
1	178
2	68
3	24
4	14
5	6

Mediana

Cálculo de la Mediana para datos agrupados en una Tabla de Frecuencias:

La mediana es aquel valor de la variable cuya Frecuencia Absoluta Acumulada (F_a) es inmediatamente mayor a la mitad de las observaciones.

El cálculo de la mediana para datos agrupados en una tabla de frecuencias se realiza mediante el siguiente procedimiento:

1. Hallar $n/2$.
2. Ubicar el valor observado cuya frecuencia absoluta acumulada contiene a $n/2$.

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q1 y q3.

Solución/

Mediana:

Me →

Número de visitas	f_i	f_a
0	210	210
1	178	388
2	68	456
3	24	480
4	14	494
5	6	500



$$\frac{n}{2} = \frac{500}{2} = 250$$

Interpretación: El 50% (la mitad) de los alumnos consultados visitaron a lo sumo 1 vez la biblioteca en ese cuatrimestre.

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q1 y q3.

Solución/

Q1: 0. El 25% (la cuarta parte) de los estudiantes consultados no hicieron ninguna visita a la biblioteca en ese cuatrimestre.

q1 →

Número de visitas	f_i	f_a
0	210	210
1	178	388
2	68	456
3	24	480
4	14	494
5	6	500

← $1 \times \frac{n}{4} = \frac{500}{4} = 125$

- Hallar las medidas descriptivas e interpretar cada una: mínimo, máximo, media aritmética, mediana, moda, q1 y q3.

Solución/

Q1: 0. El 25% (la cuarta parte) de los estudiantes consultados no hicieron ninguna visita a la biblioteca en ese cuatrimestre.

Q3: 1. El 75% (las tres cuartas partes) de los estudiantes consultados visitaron a lo sumo 1 vez la biblioteca en ese cuatrimestre.

Número de visitas	f_i	f_a
0	210	210
1	178	388
2	68	456
3	24	480
4	14	494
5	6	500

q3 →

← $3 \times \frac{n}{4} = 3 \times \frac{500}{4} = 375$

Ejemplos

La siguiente tabla contiene los valores del pH sanguíneo de 32 individuos

a) Agrupar los datos en 5 intervalos y confeccionar la tabla de frecuencias.

Solución/

7,33	7,31	7,26	7,33	7,37	7,27	7,30	7,33
7,33	7,32	7,35	7,39	7,33	7,38	7,33	7,31
7,37	7,35	7,34	7,32	7,29	7,35	7,38	7,32
7,32	7,33	7,32	7,40	7,33	7,32	7,34	7,33

La siguiente tabla contiene los valores del pH sanguíneo de 32 individuos

a) Agrupar los datos en 5 intervalos y confeccionar la tabla de frecuencias.

Solución/

7,33	7,31	7,26	7,33	7,37	7,27	7,30	7,33
7,33	7,32	7,35	7,39	7,33	7,38	7,33	7,31
7,37	7,35	7,34	7,32	7,29	7,35	7,38	7,32
7,32	7,33	7,32	7,40	7,33	7,32	7,34	7,33

Cantidad de intervalos: 5

Amplitud: $(7,4 - 7,26) : 5 = 0,028$. Vamos a usar una amplitud de 0.03.

La siguiente tabla contiene los valores del pH sanguíneo de 32 individuos

a) Agrupar los datos en 5 intervalos y confeccionar la tabla de frecuencias.

Solución/

Cantidad de intervalos: 5

Amplitud: $(7,4 - 7,26) : 5 = 0,028$. Vamos a usar una amplitud de 0.03.

Intervalos de Clase	X_{mi}	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
		f_i	$fr = f_i/n$	$(100 \cdot fr) \%$	Fa	Fa %
[7,26; 7,29)	7,275	2	0.0625	6.250	2	6.250
[7,29; 7,32)	7,305	4	0.125	12.500	6	18.750
[7,32; 7,35)	7,335	17	0.53125	53.125	23	71.875
[7,35; 7,38)	7,365	5	0.15625	15.625	28	87.500
[7,38; 7,41)	7,395	4	0.125	12.500	32	100.000

Media Aritmética

Cuando se agrupan los datos en *Intervalos de clases*, se calcula la media aritmética mediante la siguiente formula:

$$\bar{X} = \frac{\sum_{i=1}^k x_{m_i} f_i}{n}$$

donde :

$\bar{X}$: media aritmética de la muestra

n : total de datos de la muestra

x_{m_i} : marca de clase de i -ésimo intervalo

f_i : frecuencia absoluta para cada valor de la variable

Intervalos	x_{mi}	f_i
$[a_1, a_2)$	X_{m1}	f_1
$[a_2, a_3)$	X_{m2}	f_2
.	.	.
.	.	.
$[a_{k-1}, a_k)$	x_{mk}	f_k
		$n = 30$

b) Calcular la media aritmética, la moda y la mediana.

Solución/

Intervalos de Clase	X_{mi}	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
		f_i	$fr = f_i/n$	$(100 \cdot fr) \%$	Fa	$Fa \%$
[7,26;7,29)	7,275	2	0.0625	6.250	2	6.250
[7,29;7,32)	7,305	4	0.125	12.500	6	18.750
[7,32;7,35)	7,335	17	0.53125	53.125	23	71.875
[7,35;7,38)	7,365	5	0.15625	15.625	28	87.500
[7,38;7,41)	7,395	4	0.125	12.500	32	100.000

$$\bar{X} = \frac{\sum_{i=1}^k x_{m_i} f_i}{n} = 7,34$$

Mediana

Cálculo de la Mediana para datos agrupados en Intervalos de Clase:

El cálculo de la mediana para datos agrupados en intervalos de clase se realiza mediante el siguiente procedimiento:

1. Hallar $n/2$.
2. Ubicar el valor observado cuya frecuencia absoluta acumulada contiene a $n/2$.
3. Calcular la mediana mediante la fórmula:

$$Me = L_{inf} + \frac{\frac{n}{2} - F_{aa}}{f_i} \times a$$

donde:

L_{inf} = Límite inferior del primer intervalo cuya F_a es mayor a $n/2$.

F_{aa} = Frecuencia acumulada del intervalo anterior al primer intervalo cuya Fa es mayor a $n/2$.

f_i = Frecuencia absoluta del primer intervalo cuya Fa es mayor a $n/2$.

a = Amplitud de los intervalos

c) Graficar el histograma y polígono de frecuencias relativas.

Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	f_i	$fr = f_i/n$	$(100 \cdot fr) \%$	F_a	$F_a \%$
$[7,26; 7,29)$	2	0.0625	6.250	2	6.250
$[7,29; 7,32)$	4	0.125	12.500	6	18.750
$[7,32; 7,35)$	17	0.53125	53.125	23	71.875
$[7,35; 7,38)$	5	0.15625	15.625	28	87.500
$[7,38; 7,41)$	4	0.125	12.500	32	100.000

$$Me = L_{\text{inf}} + \frac{\frac{n}{2} - F_{aa}}{f_i} \times a = 7,32 + \frac{\frac{32}{2} - 6}{17} \times 0,03 \cong 7,3476$$

b) Calcular la media aritmética, la moda y la mediana.

Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	f_i	$fr = f_i/n$	$(100 \cdot fr) \%$	F_a	$F_a \%$
[7,26;7,29)	2	0.0625	6.250	2	6.250
[7,29;7,32)	4	0.125	12.500	6	18.750
[7,32;7,35)	17	0.53125	53.125	23	71.875
[7,35;7,38)	5	0.15625	15.625	28	87.500
[7,38;7,41)	4	0.125	12.500	32	100.000

Cálculo de la Moda para datos agrupados en Intervalos de Clase:

Moda

Cálculo de la Moda para datos agrupados en Intervalos de Clase:

$$Mo = L_{\text{inf}} + \frac{d_1}{d_1 + d_2} \times a$$

donde:

L_{inf} = Límite inferior del intervalo que tiene mayor frecuencia absoluta (intervalo modal).

d_1 = Diferencia entre las frecuencias absolutas del intervalo modal y el intervalo pre-modal.

d_2 = Diferencia entre las frecuencias absolutas del intervalo modal y el intervalo post-modal.

a = Amplitud de los intervalos

b) Calcular la media aritmética, la moda y la mediana.

Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	fi	fr = fi/n	(100.fr) %	Fa	Fa %
[7,26;7,29)	2	0.0625	6.250	2	6.250
[7,29;7,32)	4	0.125	12.500	6	18.750
[7,32;7,35)	17	0.53125	53.125	23	71.875
[7,35;7,38)	5	0.15625	15.625	28	87.500
[7,38;7,41)	4	0.125	12.500	32	100.000

$$Mo = L_{\text{inf}} + \frac{d_1}{d_1 + d_2} \times a = 7,32 + \frac{13}{13 + 12} \times 0,03 \cong 7,3356$$

c) Graficar el histograma y polígono de frecuencias relativas.

Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	f_i	$fr = f_i/n$	$(100.fr) \%$	F_a	$F_a \%$
[7,26;7,29)	2	0.0625	6.250	2	6.250
[7,29;7,32)	4	0.125	12.500	6	18.750
[7,32;7,35)	17	0.53125	53.125	23	71.875
[7,35;7,38)	5	0.15625	15.625	28	87.500
[7,38;7,41)	4	0.125	12.500	32	100.000

d) Hallar el coeficiente de variabilidad. ¿Qué se puede concluir a partir del valor hallado?

Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	f_i	$fr = f_i/n$	$(100.fr) \%$	F_a	$F_a \%$
[7,26;7,29)	2	0.0625	6.250	2	6.250
[7,29;7,32)	4	0.125	12.500	6	18.750
[7,32;7,35)	17	0.53125	53.125	23	71.875
[7,35;7,38)	5	0.15625	15.625	28	87.500
[7,38;7,41)	4	0.125	12.500	32	100.000

Coeficiente de variación

El **coeficiente de variación** (CV) es una medida que relaciona la desviación estándar con la media aritmética para determinar qué tan homogénea o dispersa es la información.

$$CV = \frac{S}{\bar{X}}$$

- Mide el grado de variabilidad en una muestra o población.
- Está desprovisto de unidades. Permite comparar la variabilidad entre distintas variables y poblaciones.
- El valor expresado en términos porcentuales, se llama **coeficiente de variación porcentual**.

$$CV\% = \frac{S}{\bar{X}} \times 100\%$$

Consideraremos poca variabilidad, si el $CV\%$ es a lo sumo del 30 %

Desvío estandar muestral

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



Para datos sin agrupar

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



Para datos agrupados
por frecuencias

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^k (x_{mi} - \bar{X})^2 f_i}$$



Para datos agrupados
por Intervalos

d) Hallar el coeficiente de variabilidad. ¿Qué se puede concluir a partir del valor hallado?

Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	f_i	$fr = f_i/n$	$(100.fr) \%$	Fa	$Fa \%$
[7,26;7,29)	2	0.0625	6.250	2	6.250
[7,29;7,32)	4	0.125	12.500	6	18.750
[7,32;7,35)	17	0.53125	53.125	23	71.875
[7,35;7,38)	5	0.15625	15.625	28	87.500
[7,38;7,41)	4	0.125	12.500	32	100.000

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^k (x_{mi} - \bar{X})^2 f_i}$$

$$\bar{X} = 7,34$$

$$CV\% = \frac{S}{\bar{X}} \times 100\% = 0,42\%$$

Consideraremos poca variabilidad, si el CV% es a lo sumo del 30 %

e) Hallar el tercer decil. Interpretar su significado en el contexto del problema.

Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	f_i	$fr = f_i/n$	$(100 \cdot fr) \%$	F_a	$F_a \%$
[7,26; 7,29)	2	0.0625	6.250	2	6.250
[7,29; 7,32)	4	0.125	12.500	6	18.750
[7,32; 7,35)	17	0.53125	53.125	23	71.875
[7,35; 7,38)	5	0.15625	15.625	28	87.500
[7,38; 7,41)	4	0.125	12.500	32	100.000

Deciles

Los Deciles (D_k) son valores que fraccionan la distribución de los datos en diez partes iguales.

$$D_j = L_{\text{inf}} + \frac{j \times \frac{n}{10} - F_{aa}}{f_i} \times a$$

e) Hallar el tercer decil. Interpretar su significado en el contexto del problema.

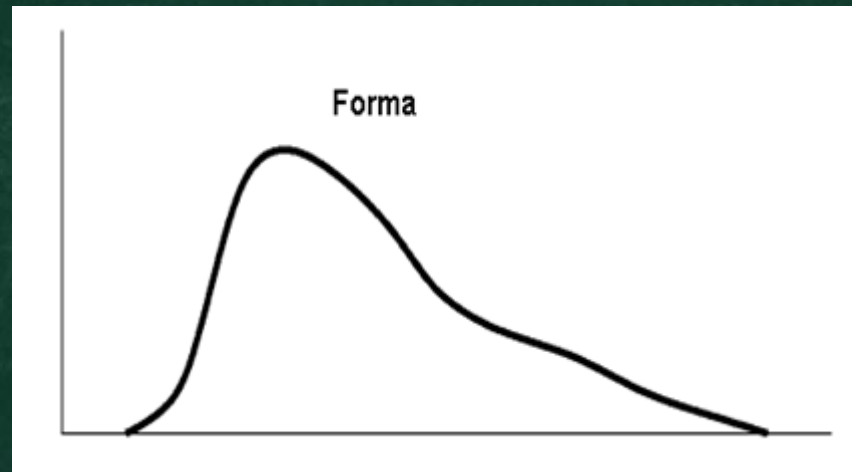
Solución/

Intervalos de Clase	Fr. Absoluta	Fr. Relativa	Fr. Relativa Porcentual	Fr. Acumulada	Fr. Acum. Porcentual
	f_i	$fr = f_i/n$	$(100 \cdot fr) \%$	F_a	$F_a \%$
[7,26; 7,29)	2	0.0625	6.250	2	6.250
[7,29; 7,32)	4	0.125	12.500	6	18.750
[7,32; 7,35)	17	0.53125	53.125	23	71.875
[7,35; 7,38)	5	0.15625	15.625	28	87.500
[7,38; 7,41)	4	0.125	12.500	32	100.000

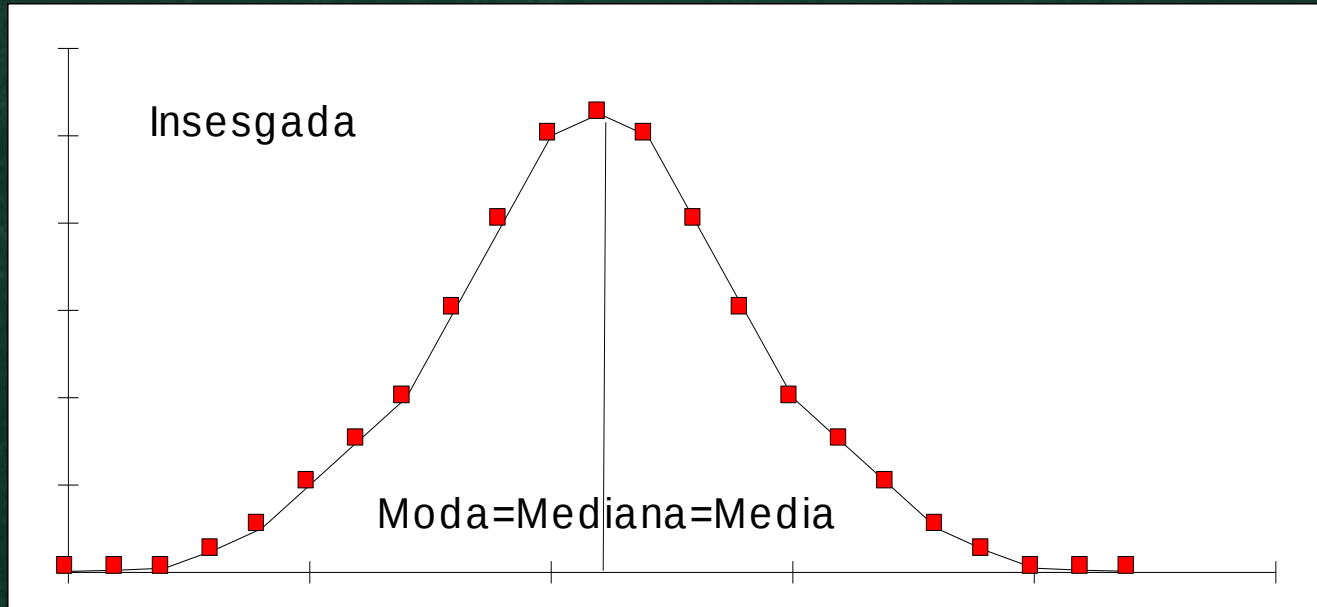
$$D_j = L_{\text{inf}} + \frac{j \times \frac{n}{10} - F_{aa}}{f_i} \times a = 7.32 + \frac{3 \times \frac{32}{10} - 6}{17} * 0,03 \cong 7,3264$$

El 30% de los individuos tiene un PH de 7,3264 como máximo.

Análisis de la Forma



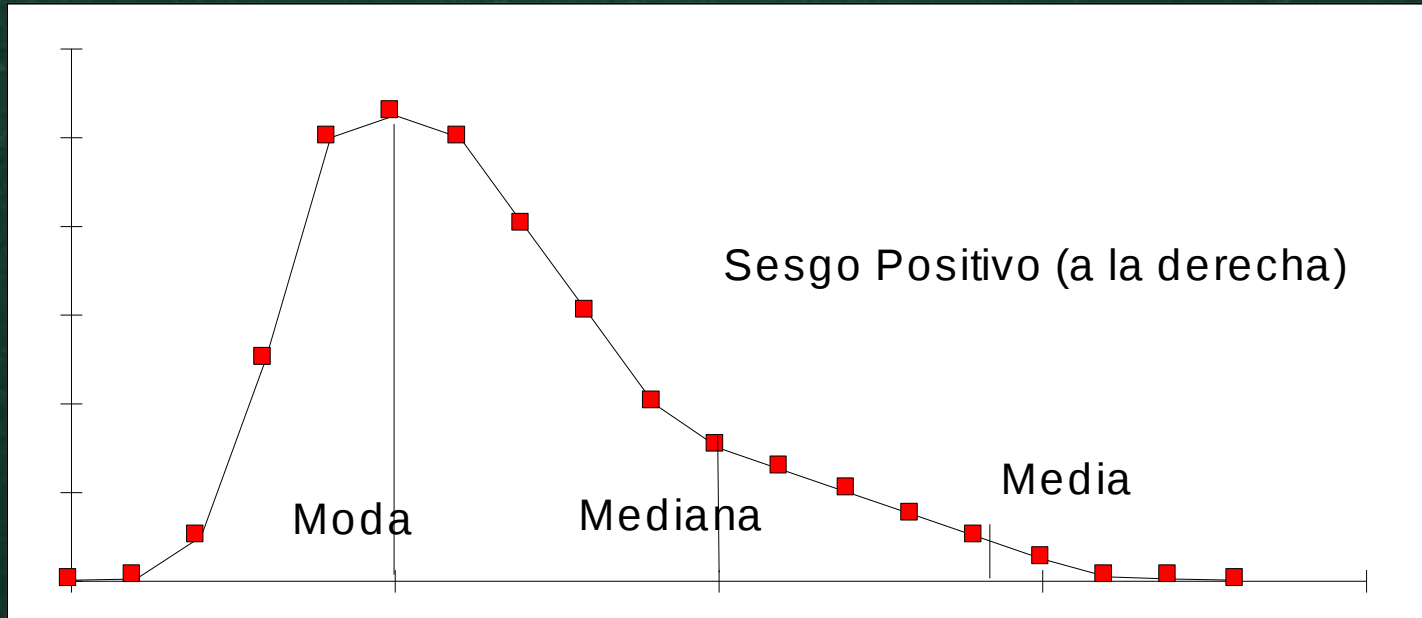
Distribuciones Simétricas



$$\overline{X} = Me = Mo$$

La distribución de los datos es simétrica

Distribuciones Asimétricas



Si $Mo < Me < \bar{X}$: Asimétrica Positiva

Si la distribución es asimétrica positiva, la media no representa al conjunto de datos.

Gráficos

JAN
2020

FACEBOOK'S ADVERTISING AUDIENCE

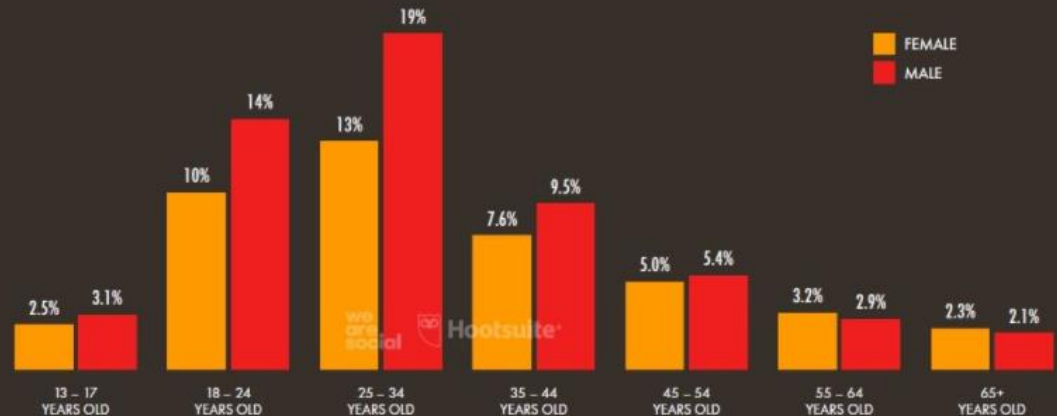
POTENTIAL REACH OF ADVERTISING ON FACEBOOK BY AGE GROUP AND BY GENDER *

AGE	TOTAL AUDIENCE	FEMALE TOTAL	FEMALE SHARE	MALE TOTAL	MALE SHARE
13-17	109,000,000	48,700,000	2.5%	60,300,000	3.1%
18-24	474,800,000	196,300,000	10%	278,500,000	14%
25-34	627,000,000	253,400,000	13%	373,600,000	19%
35-44	332,500,000	148,200,000	7.6%	184,300,000	9.5%
45-54	201,500,000	97,200,000	5.0%	104,400,000	5.4%
55-64	119,000,000	63,000,000	3.2%	56,000,000	2.9%
65+	85,100,000	45,100,000	2.3%	40,100,000	2.1%
TOTAL	1,949,000,000	851,900,000	43.7%	1,097,100,000	56.3%

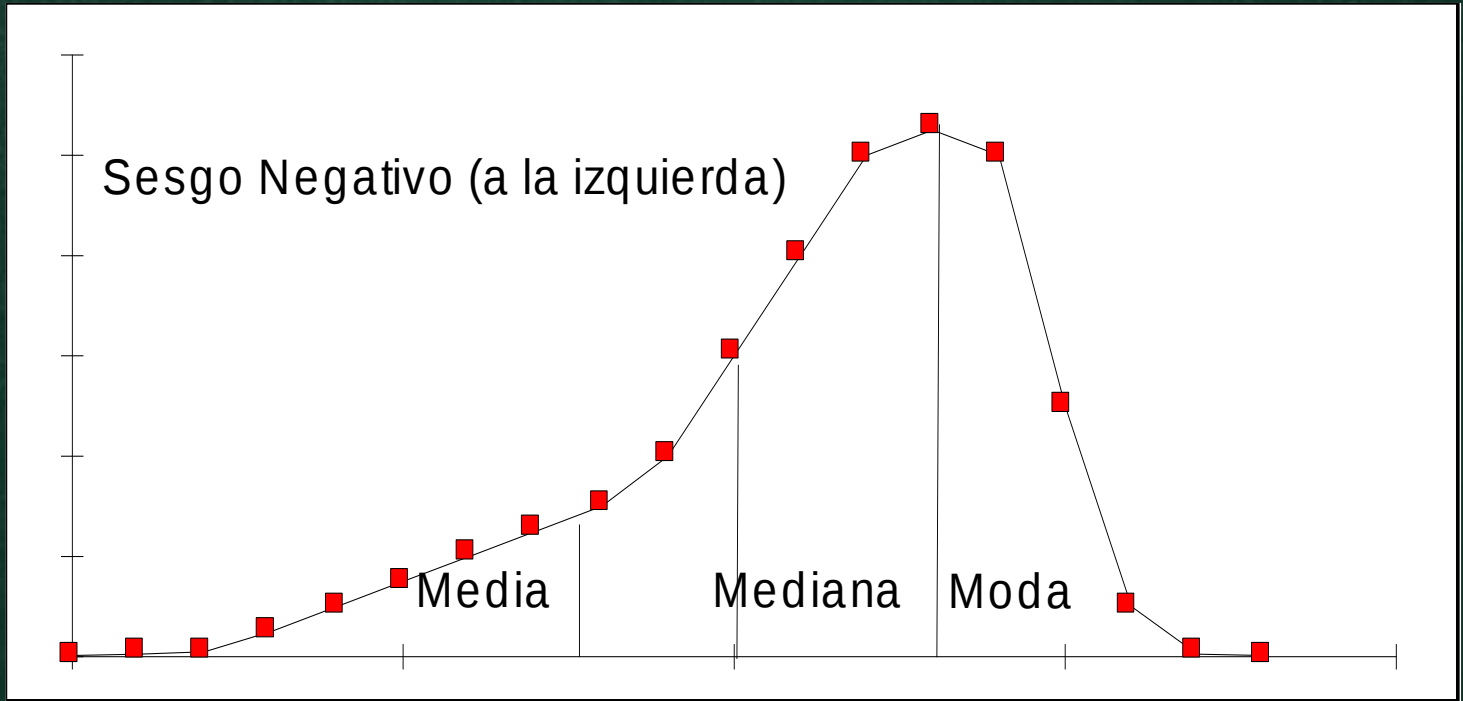
JAN
2020

PROFILE OF FACEBOOK'S ADVERTISING AUDIENCE

SHARE OF FACEBOOK'S GLOBAL ADVERTISING AUDIENCE* BY AGE GROUP AND GENDER *



Distribuciones Asimétricas



Si $\bar{X} < Me < Mo$: Asimétrica Negativa

Si la distribución es asimétrica negativa,
la media no representa al conjunto de datos.

Análisis de la simetría

Coefficiente
de asimetría

$$A_s = \frac{\bar{X} - M_e}{s}$$

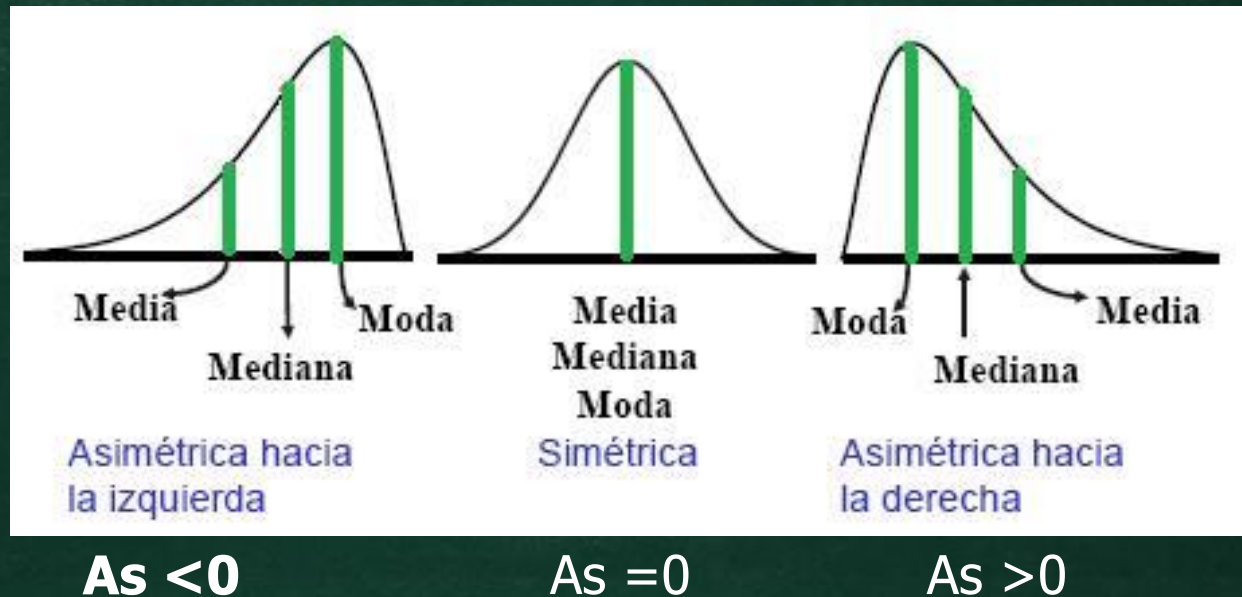


Gráfico de caja y bigotes (Box-Plot)

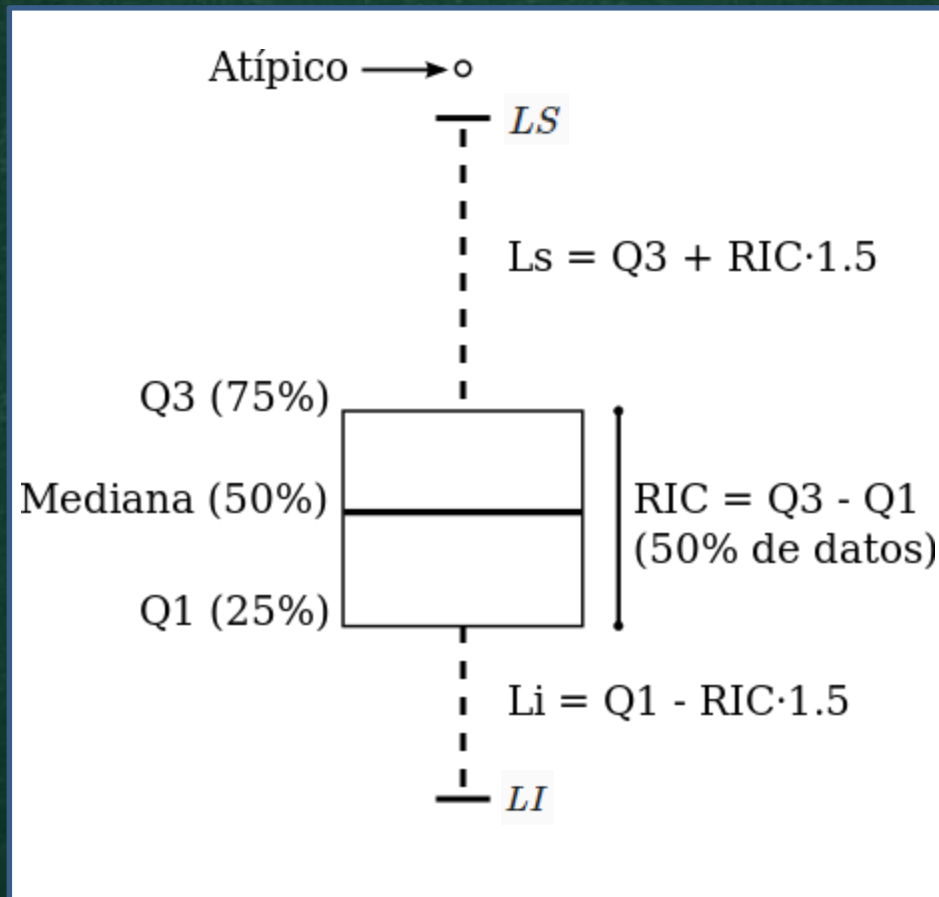
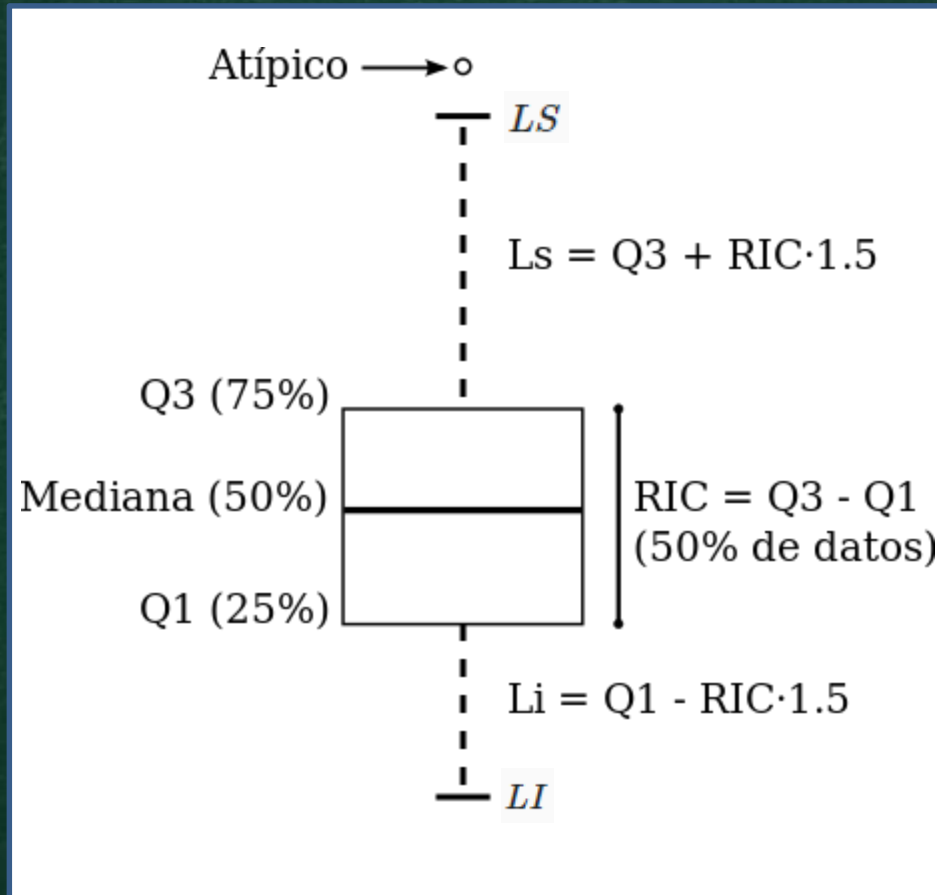


Gráfico de caja y bigotes (Box-Plot)



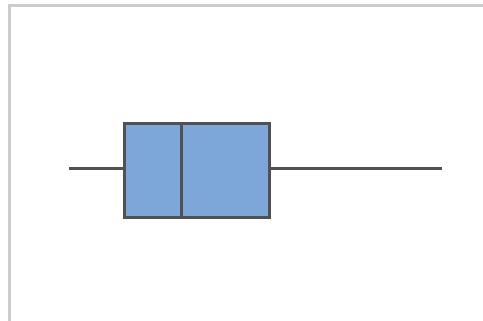
Para calcular el valor del bigote inferior, se compara al número Li con el valor más chico de todos los datos y el más grande entre ellos será el bigote inferior, LI . Para calcular el valor del bigote superior, se comparan al número Ls con el valor más grande de todos los datos y el más chico entre los dos números será el bigote superior LS .

Gráfico de caja y bigotes (Box-Plot)

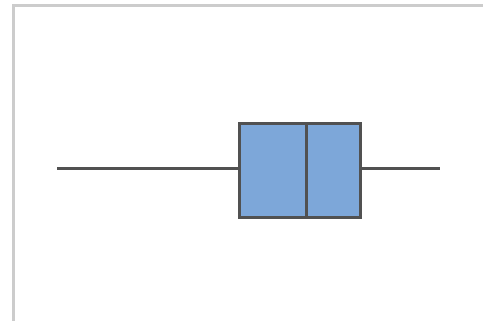
Interpretación.

Datos asimétricos

Cuando los datos son asimétricos, la mayoría de los datos se ubican en la parte superior o inferior de la gráfica.



Asimétrico hacia la derecha



Asimétrico hacia la izquierda

Gráfico de caja y bigotes (Box-Plot)

Interpretación.

Valores atípicos

Los valores atípicos, que son valores que están muy alejados de otros valores de datos, pueden afectar fuertemente sus resultados. Frecuentemente, es más fácil identificar los valores atípicos en una gráfica de caja.

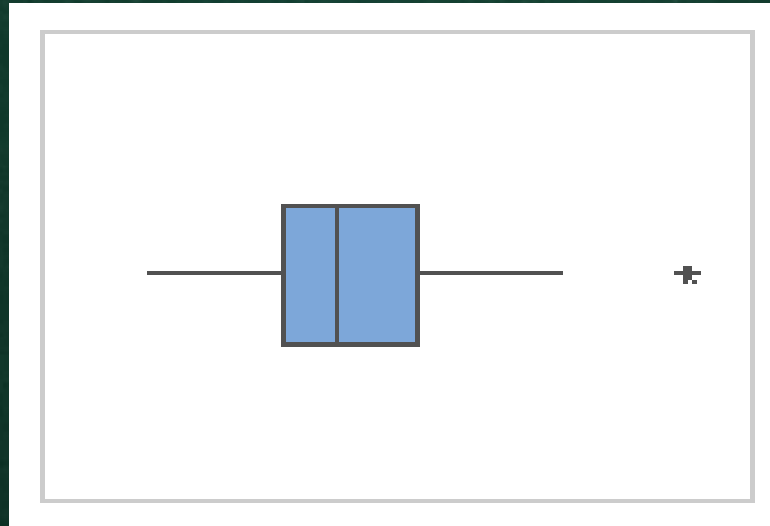
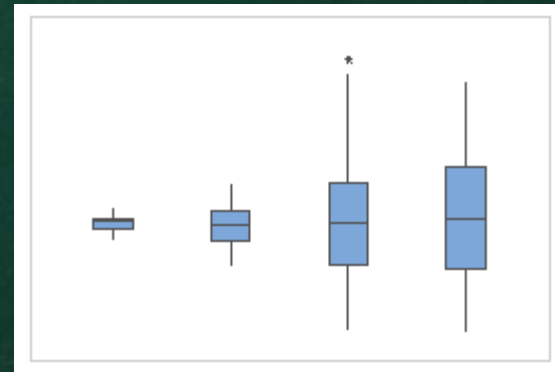
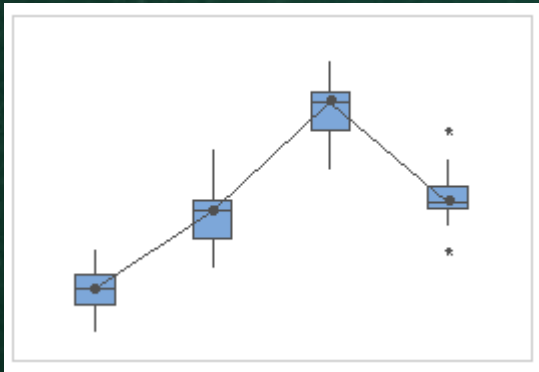


Gráfico de caja y bigotes (Box-Plot)

Interpretación.

Evaluar y comparar los grupos

Permiten evaluar y comparar el centro y la dispersión de distintos grupos.



La mediana de los grupos son similares, pero algunos de los grupos presentan mayor variabilidad.

Observaciones finales

- Comenzar por el estudio de la variabilidad de los datos, puede ahorrar pasos en el análisis.
- Si el CV es mayor que 30 %, ninguna medida resume los datos.
- Si existe poca variación en los datos, debemos analizar la forma. En ese caso, si los datos son simétricos, la media representa los mismos. Si son asimétricos, la medida que los representa es la mediana.

Situación problemática.

Un ingeniero es responsable de la producción de rulemanes y tiene dos máquinas para ello. Le interesa que los rulemanes producidos tengan diámetros similares, independientemente de la máquina que los produce, pero tiene sospechas de que existe una falta de calibración entre ellas. ¿Qué elementos le proporciona la estadística al ingeniero para poder concluir que los rulemanes tienen diámetros significativamente diferentes?

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

Diámetros producidos por la máquina A

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

Diámetros producidos por la máquina A

Diámetros producidos por la máquina B

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

Diámetros producidos por la máquina A

Clasificación de la Variable Estadística.

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

Diámetros producidos por la máquina A

Clasificación de la Variable Estadística.

Variable estadística cuantitativa continua.

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

Series Simples

Series de Frecuencias

Intervalos de Clases

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

Series Simples

Series de Frecuencias

Intervalos de Clases

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Organización de los datos:

¿Cómo saber cuántos intervalos considerar? ¿Cómo determinar su amplitud?

Primero debemos determinar el rango de los datos, que es la diferencia entre el mayor y el menor de los valores obtenidos.

$$\text{Rango} = X_{\text{máx}} - X_{\text{mín}}$$

Luego debemos establecer el número de intervalos (K) y determinar la amplitud (A) de los mismos.

$$K = 1 + 3,3 \cdot \log n \quad (\text{regla de } \mathbf{Sturges})$$

$$A = \text{Rango} / K$$

Organización de los datos:

¿Cómo saber cuántos intervalos considerar? ¿Cómo determinar su amplitud?

Primero debemos determinar el rango de los datos, que es la diferencia entre el mayor y el menor de los valores obtenidos.

$$\text{Rango} = X_{\text{máx}} - X_{\text{mín}}$$

Luego debemos establecer el número de intervalos (K) y determinar la amplitud (A) de los mismos.

$$K = 1 + 3,3 \cdot \log n \quad (\text{regla de } \mathbf{Sturges})$$

$$A = \text{Rango} / K$$

Organización de los datos:

¿Cómo saber cuántos intervalos considerar? ¿Cómo determinar su amplitud?

Primero debemos determinar el rango de los datos, que es la diferencia entre el mayor y el menor de los valores obtenidos.

$$\text{Rango} = X_{\text{máx}} - X_{\text{mín}}$$

Luego debemos establecer el número de intervalos (K) y determinar la amplitud (A) de los mismos.

$$K = 1 + 3,3 \cdot \log n \quad (\text{regla de } \mathbf{Sturges})$$

$$A = \text{Rango} / K$$

Organización de los datos:

¿Cómo saber cuántos intervalos considerar? ¿Cómo determinar su amplitud?

Primero debemos determinar el rango de los datos, que es la diferencia entre el mayor y el menor de los valores obtenidos.

$$\text{Rango} = X_{\text{máx}} - X_{\text{mín}}$$

Luego debemos establecer el número de intervalos (K) y determinar la amplitud (A) de los mismos.

$$K = 1 + 3,3 \cdot \log n \quad (\text{regla de Sturges})$$

$$A = \text{Rango} / K$$

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

$$\text{MaxA} = 5.47$$

$$\text{MinA} = 4.5$$

$$\text{rangoA} = 0.97$$

$$\text{Rango} = X_{\text{máx}} - X_{\text{mín}}$$

Luego debemos establecer el número de intervalos (K) y determinar la amplitud (A) de los mismos.

$$K = 1 + 3,3 \cdot \log n \quad (\text{regla de Sturges})$$

$$A = \text{Rango} / K$$

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

$$\text{MaxA} = 5.47$$

$$\text{MinA} = 4.5$$

$$\text{rangoA} = 0.97$$

$$\text{Rango} = X_{\text{máx}} - X_{\text{mín}}$$

$$K = 8$$

$$A = 0.13$$

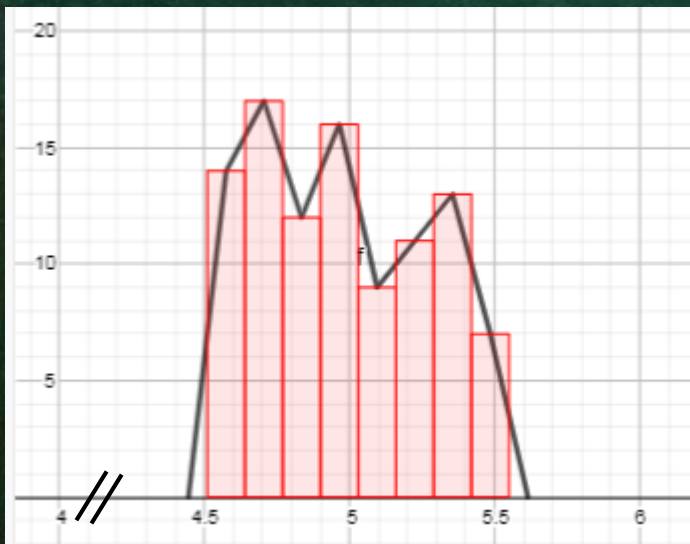
	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

Intervalo	Cuenta
[4.51 – 4.64)	14
[4.64 – 4.77)	16
[4.77 – 4.9)	11
[4.9 – 5.03)	12
[5.03 – 5.16)	12
[5.16 – 5.29)	17
[5.29 – 5.42)	10
[5.42 – 5.55)	6
n = 98	

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
•	•	•
•	•	•

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.

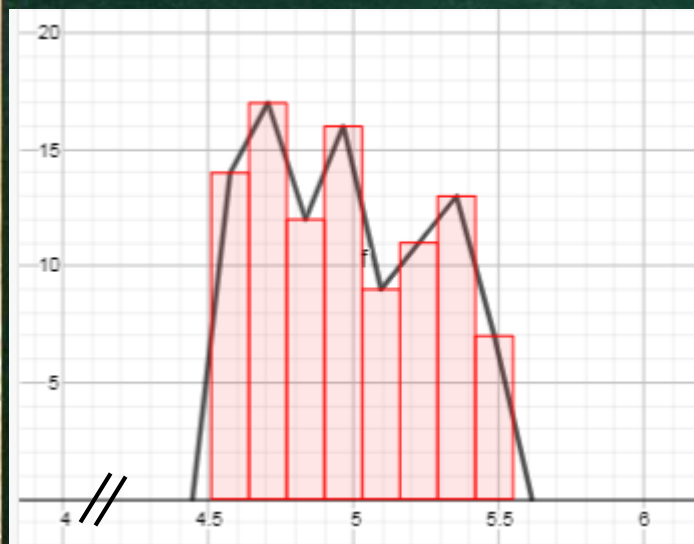


Intervalo	Cuenta
[4.51 – 4.64)	14
[4.64 – 4.77)	16
[4.77 – 4.9)	11
[4.9 – 5.03)	12
[5.03 – 5.16)	12
[5.16 – 5.29)	17
[5.29 – 5.42)	10
[5.42 – 5.55)	6

n = 98

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
⋮	⋮	⋮

Supongamos que el ingeniero para cada una de las máquinas toma al azar 98 rulemanes y registra sus diámetros.



Intervalo	Cuenta
[4.51 – 4.64)	14
[4.64 – 4.77)	16
[4.77 – 4.9)	11
[4.9 – 5.03)	12
[5.03 – 5.16)	12
[5.16 – 5.29)	17
[5.29 – 5.42)	10
[5.42 – 5.55)	6

n = 98

	A	B
1	5.41	5.3
2	4.53	5.32
3	5.35	5.06
4	4.66	5.43
5	5.24	5.08
6	5.08	4.58
7	4.78	5.28
8	5.18	4.93
9	5.31	5.32
10	4.87	5.27
11	5.02	4.58
12	4.76	5.45
13	5.48	4.91
14	4.64	4.98
15	5.11	4.88
16	5.09	5.02
17	5.39	4.66
18	5.36	4.54
19	5.02	5.39
⋮	⋮	⋮

El mismo análisis se puede hacer para la máquina B.