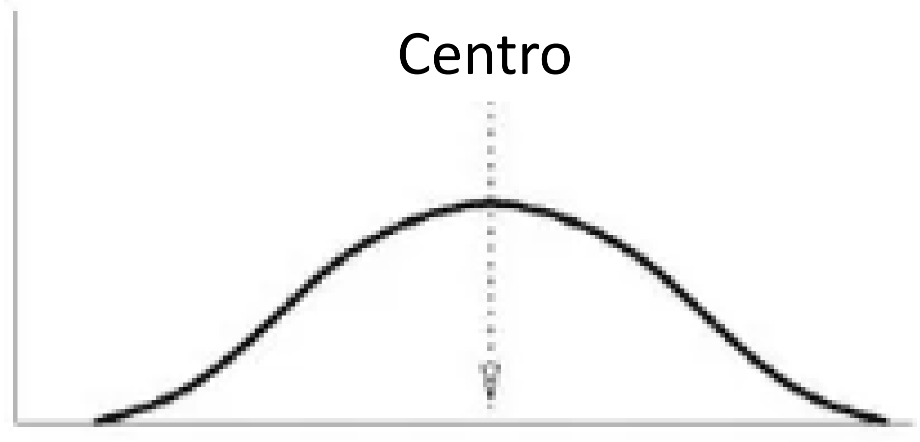
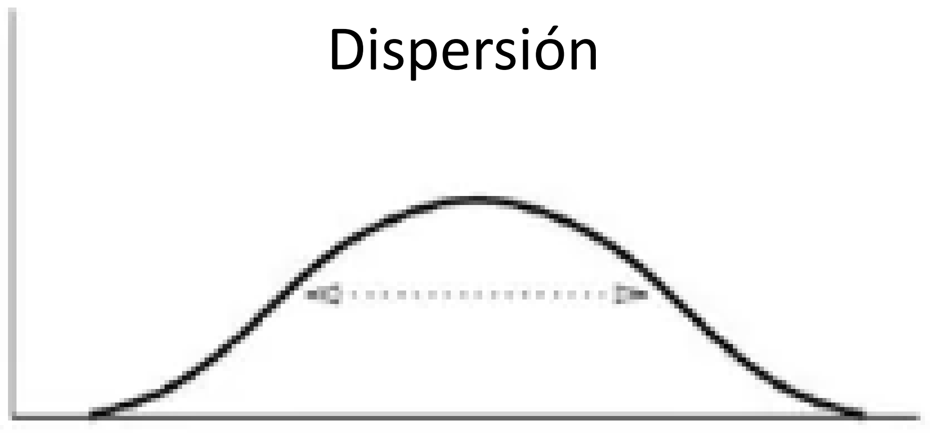
Medidas estadísticas



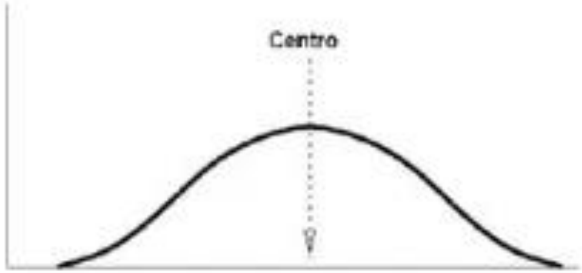


Medidas estadísticas

Tendencia central:

Indican valores con respecto a los que los datos parecen agruparse

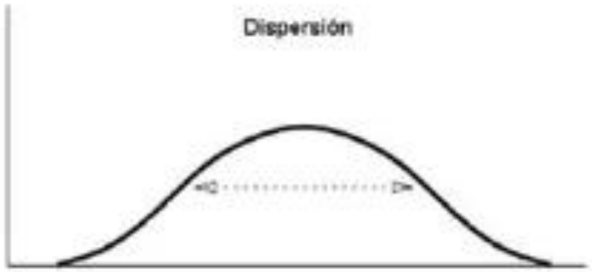
- Media
- Mediana
- Moda



Dispersión

Indican el nivel de concentración de los datos con respecto a la medida

- Rango
- Varianza
- Desv Estándar
- Coef. de Variación



Medidas de tendencia central

- Media
- Mediana
- Moda

Media

$$\bar{x}$$

Media

- Es el valor promedio del grupo de datos, es decir, la cifra que se obtiene al sumar todos los datos y dividir el resultado entre la cantidad de los mismos.

$$Media = \frac{Suma\ de\ valores}{Cantidad\ de\ valores}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \cdots\ x_n}{N}$$

Ejemplo. Se ha preguntado a un grupo de 5 estudiantes de estadística cuántas tazas de café beben a la semana. El resultado es: 21, 25, 10, 8 y 11 tazas. La media es, por tanto, 15.

1	2	3	4	5
21	25	10	8	11

$$\frac{21 + 25 + 10 + 8 + 11}{5} = 15$$

Cálculo de la media para datos agrupados

Supongamos que tenemos k clases diferentes en nuestra tabla de frecuencias, en donde para cada clase c_i, tenemos su media x_i y su frecuencia f_i correspondiente, entonces calculamos la media como:

$$\begin{aligned}\overline{X} &= \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \cdots + x_K \cdot f_K}{f_1 + f_2 + \cdots + f_K} \\ &= \frac{\sum_{i=1}^{K} x_i \cdot f_i}{\sum_{j=1}^{K} f_j} \\ &= \frac{\sum_{i=1}^{K} x_i \cdot f_i}{N}\end{aligned}$$

Debemos observar que ahora N es la suma de las frecuencias de cada clase, esto es

$$\mathrm{N} = \sum_{j=1}^{K} f_j$$

Ejemplo. En un test realizado a un grupo de 42 personas se han obtenido las puntuaciones que muestra la tabla. Calcula la puntuación media.

Intervalo	x_i	f_i
[10,20)	15	1
[20,30)	25	8
[30,40)	35	10
[40,50)	45	9
[50,60)	55	8
[60,70)	65	4
[70,80)	75	2
		42

Intervalo	x_i	f_i	$x_i * f_i$
[10,20)	15	1	15
[20,30)	25	8	200
[30,40)	35	10	350
[40,50)	45	9	405
[50,60)	55	8	440
[60,70)	65	4	260
[70,80)	75	2	150
		42	1820

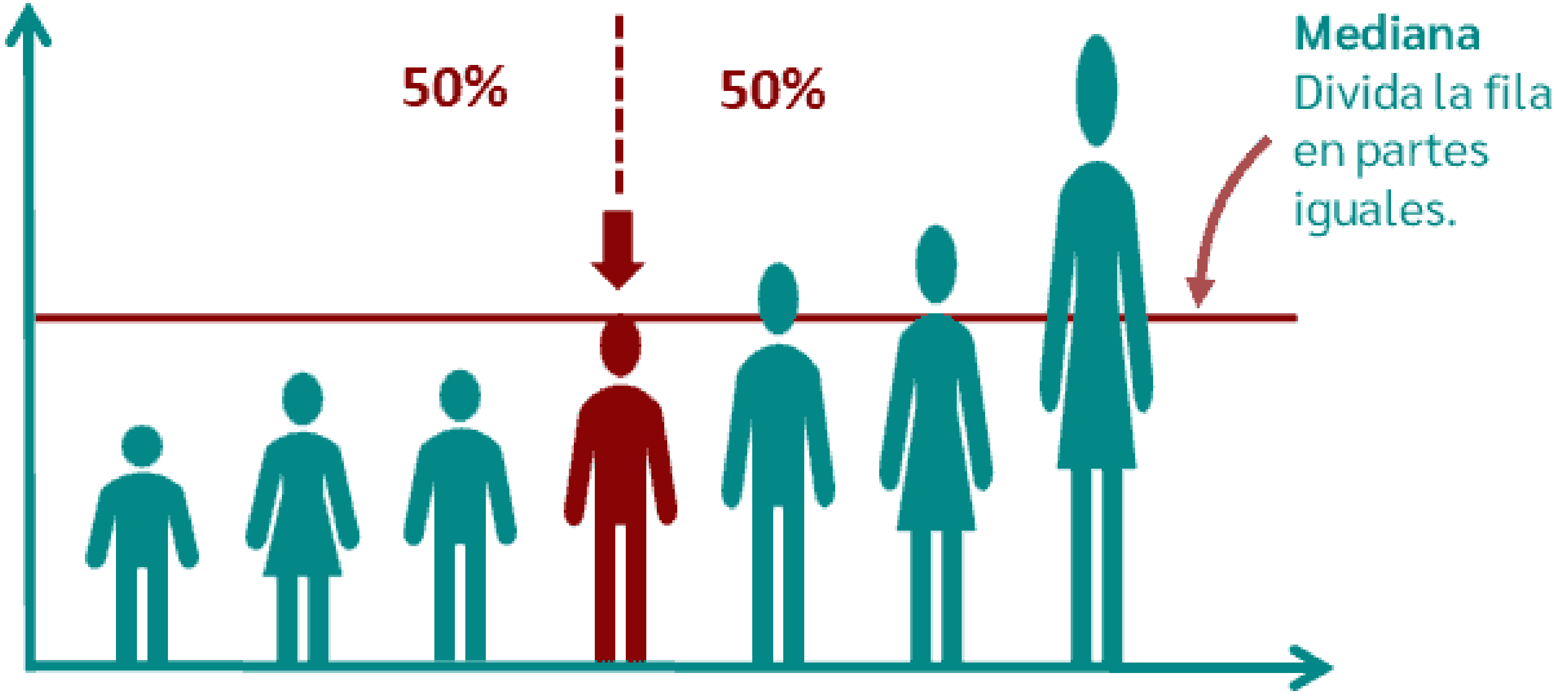
$$\begin{aligned}\overline{X} &= \frac{\sum_{i=1}^{7} x_i \cdot f_i}{\sum_{j=1}^{7} f_j} \\ &= \frac{\sum_{i=1}^{7} x_i \cdot f_i}{N} \\ &= \frac{1820}{42} \\ &= 43.33\end{aligned}$$

Mediana

$$M_e$$

Mediana

- La mediana es el valor medio de una secuencia ordenada de datos. Si no hay empates, la mitad de las observaciones serán menores y la otra mitad serán mayores.
    - Calcular la mediana es mucho más fácil porque es justo el valor central, es decir, el que se encuentra en la mitad de la lista.



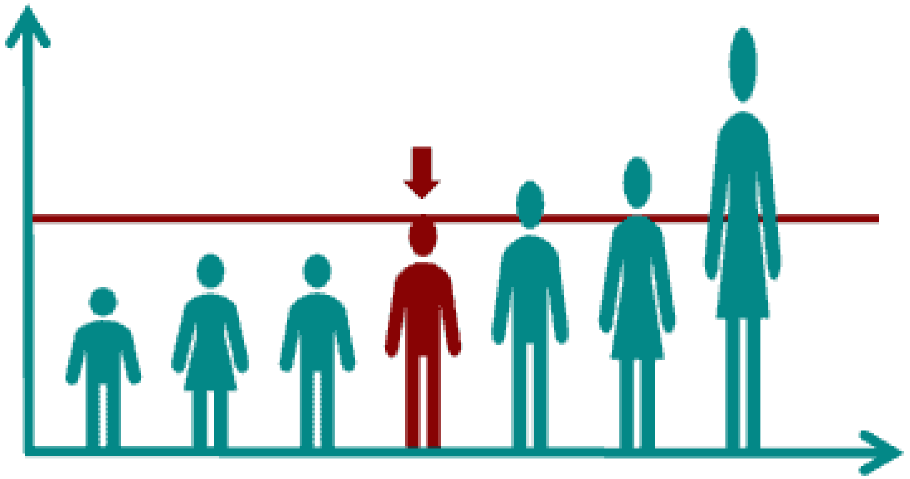
Procedimiento para calcular la mediana

1. Ordene los datos
2. Utiliza la expresión:

$$\tilde{X} = \begin{cases} X_{(n+1)/2} \; si\ n\ es\ impar \\ \frac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right) si\ n\ es\ par \end{cases}$$

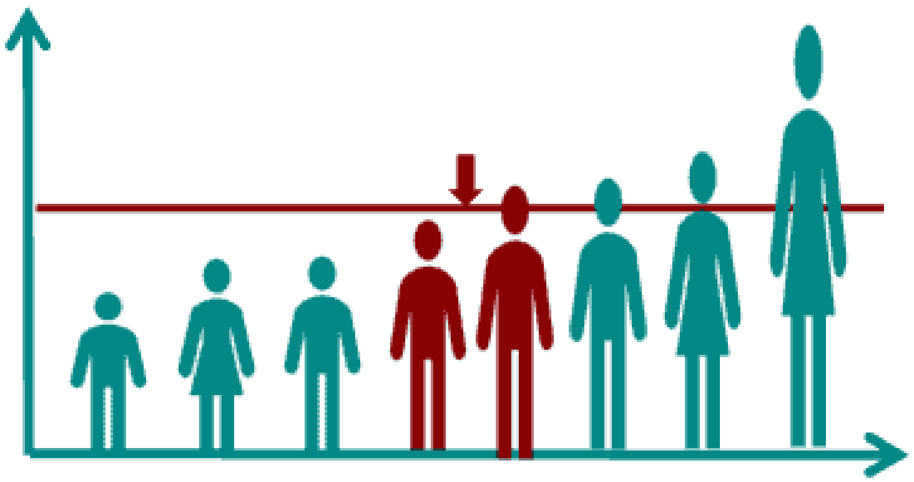
Número impar

La mediana es un valor que se da realmente.



Número par

Es la media de los dos valores medios.



Ejemplo. Tenemos el siguiente conjunto de datos:

2.0, 2.5, 5.3, 4.2, 3.8

Calcular la mediana.

1. Ordenamos los datos: 2.0, 2.5, 3.8, 4.2, 5.3
2. *N* es impar, *N* =5.
3. Utilizamos la expresión: $\tilde{x} = x_{(n+1)/2} = x_{(5+1)/2} = x_3$
4. La mediana es el número ubicado en la tercera posición: 3.8

Ejemplo. Calcular la mediana del siguiente conjunto de datos:

480, 485, 485, 490, 490, 495, 495, 500, 500, 505

$$480, 485, 485, 490, \boxed{490, 495,} 495, 500, 500, 505$$

1. Los datos están ordenados.
2. *N* es par, *N* =10.
3. Utilizamos la expresión: $\tilde{x} = \frac{1}{2}\left(x_{n/2} + x_{(n/2)+1}\right)$

$$\tilde{x} = \frac{1}{2}(x_5 + x_6)$$

$$Mediana = \frac{1}{2}(490 + 495) = 492.5$$

Ejemplo. Calcular la mediana del siguiente conjunto de datos:

460, 460, 470, 500, 500, 500, 500, 500, 500, 500

460, 460, 470, 500, 500, 500, 500, 500, 500, 500

1. Los datos están ordenados.
2. *N* es par, *N* =10.
3. Utilizamos la expresión: $$\tilde{x} = \frac{1}{2}\left(x_{n/2} + x_{(n/2)+1}\right)$$

$$\tilde{x} = \frac{1}{2}(x_5 + x_6)$$

$$Mediana = \frac{1}{2}(500 + 500) = 500$$

Ejemplo. Una toma de datos de un experimento arroja los siguientes resultados. Calcular la mediana.

x_i	*t(s) ±0.1*
x_1	0.48
x_2	0.50
x_3	0.52
x_4	**0.55**
x_5	0.56
x_6	0.58
x_7	0.60
x_8	0.63
x_9	0.66
x_{10}	0.69

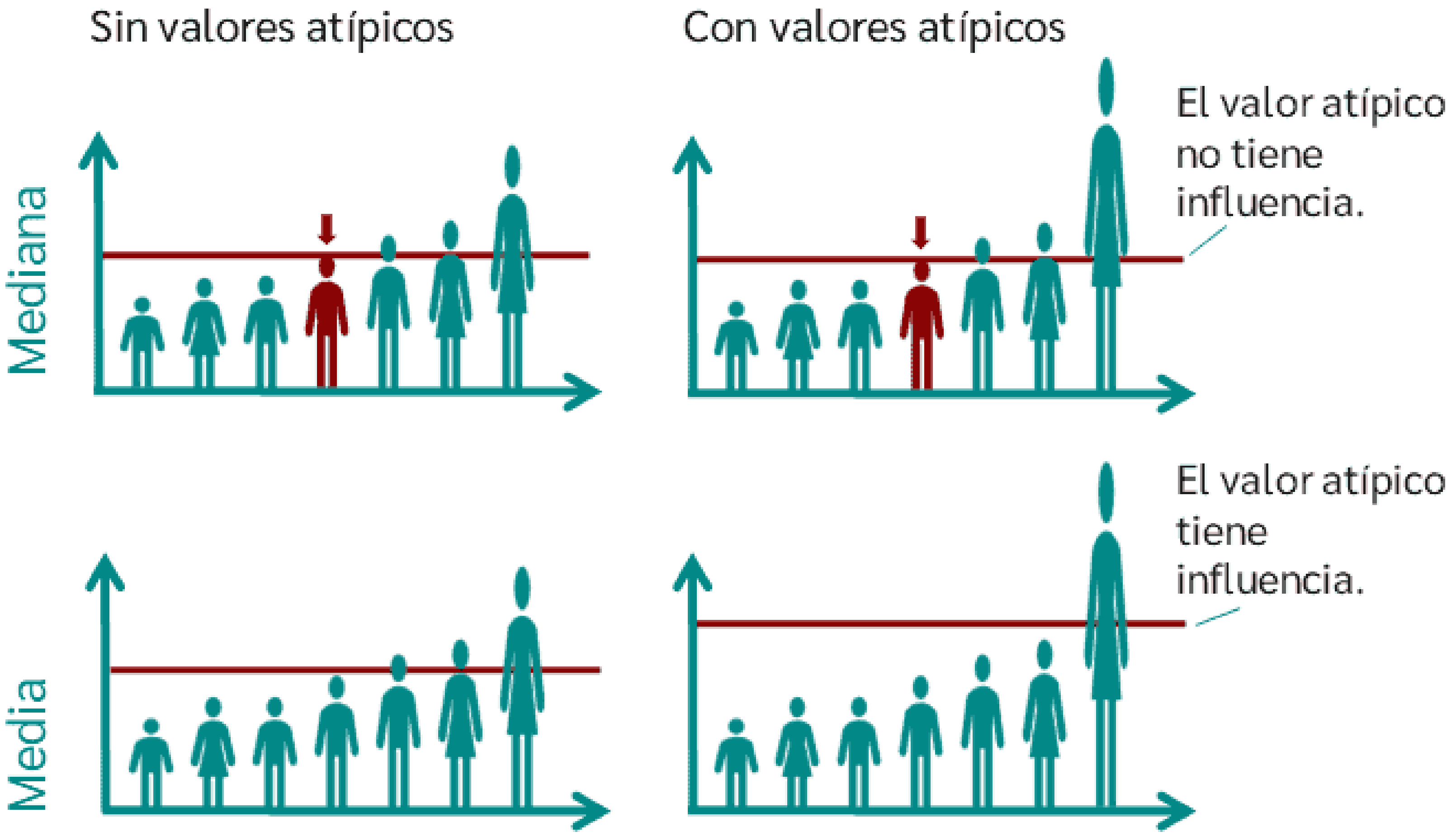
x_i	$t(s)\ \pm 0.1$
x_1	0.48
x_2	0.50
x_3	0.52
x_4	0.55
x_5	0.56
x_6	0.58
x_7	0.60
x_8	0.63
x_9	0.66
x_{10}	0.69

1. Los datos están ordenados.
2. *N* es par, *N* = 10.
3. Utilizamos la expresión:

$$\frac{1}{2}\left(X_{\frac{n}{2}} + X_{\frac{n}{2}+1}\right) = \frac{1}{2}\left(X_{\frac{10}{2}} + X_{\frac{10}{2}+1}\right) = \frac{1}{2}(X_5 + X_6) = 0.57$$

Media frente a mediana

- En comparación con la media, la mediana es mucho más robusta frente a la dispersión.
- Un valor atípico no suele influir en la mediana, pero tiene una influencia más o menos grande en la media.

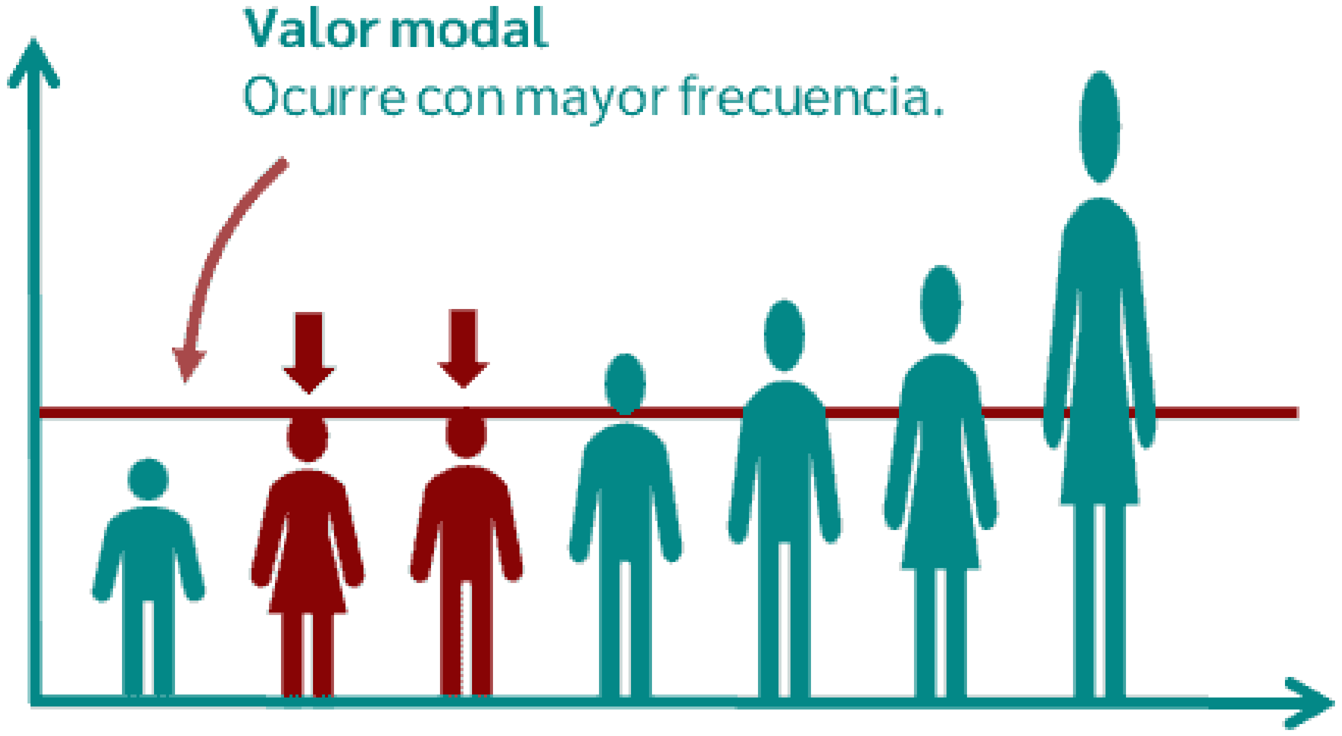


Moda

$$M_o$$

Moda

- La moda de un conjunto de datos es el dato que más veces se repite, es decir, aquel que tiene mayor frecuencia absoluta.



Moda

- En caso de existir dos valores de la variable que tengan la mayor frecuencia absoluta, habría dos modas.

- Si no se repite ningún valor, no existe moda.

- A diferencia de la media aritmética, la moda estadística no se ve afectada por la ocurrencia de los valores extremos.

Ejemplo: En una muestra de 70 directivos de Berlín, 20 conducen un Daimler, 25 un BMW, 10 un VW y 15 un Audi. La marca de coches BMW es la más común. Por tanto, la Moda es "BMW".

Car brand	Daimler	BMW	VW	Audi
Frequency	20	25	10	15

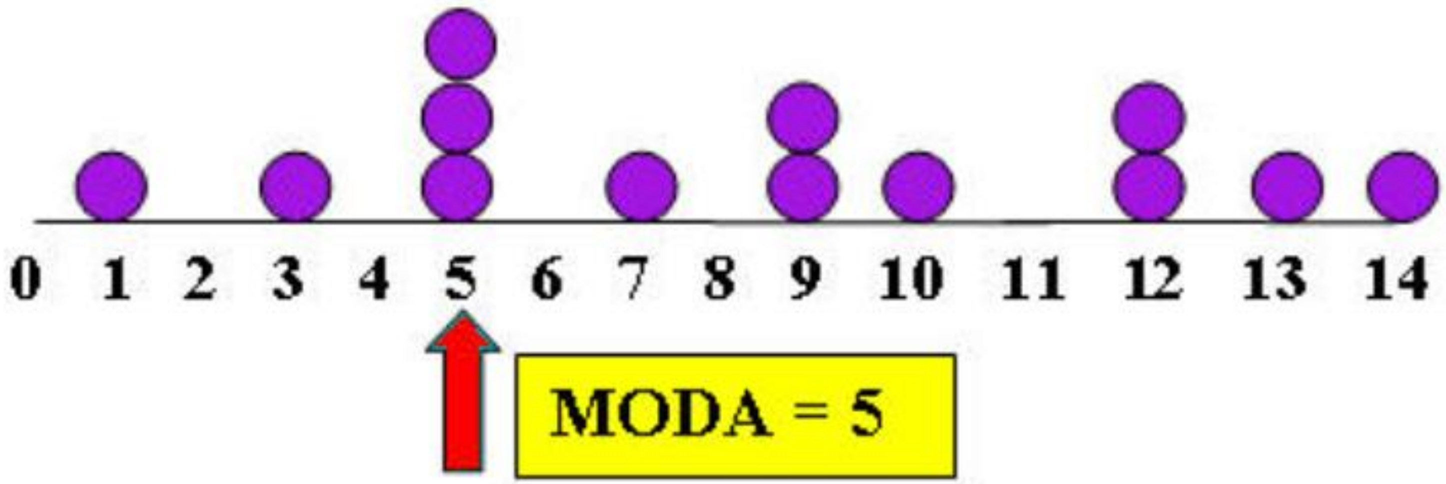
Ejemplos. Hallar la moda de las siguientes distribuciones:

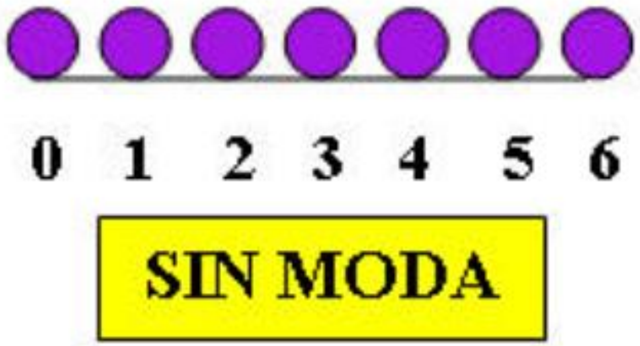
$2, 3, 3, 4, 4, 4, 5, 5$ $\quad M_o = 4$

$1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9$ $\quad M_o = 1, 5, 9$

$2, 2, 3, 3, 6, 6, 9, 9$ $\quad$ No hay moda

Ejemplos.





Cálculo de la moda para datos agrupados

Caso 1: Cuando todos los intervalos tienen la misma amplitud.

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

L_i es el límite inferior de la clase modal

f_i es la frecuencia absoluta de la clase modal

f_{i-1} es la frecuencia absoluta inmediatamente inferior a la clase modal

f_{i+1} es la frecuencia absoluta inmediatamente posterior a la clase modal

a_i es la amplitud de la clase

- También se utiliza otra fórmula de la moda que da un valor aproximado de ésta:

$$M_o = L_i + \frac{f_{i+1}}{f_{i-1} + f_{i+1}} \cdot a_i$$

Ejemplo:

Calcular la moda de una distribución estadística que viene dada por la siguiente tabla:

Intervalo	*Frecuencia absoluta* f_i
(60,63)	5
(63,66)	18
(66,69)	42
(69,72)	27
(72,75)	8
	100

- Los intervalos tienen la misma amplitud: $a_i = 3$

$$a_i = Límite\ superior\ de\ la\ clase - Límite\ inferior\ de\ la\ clase$$

Intervalo	*Frecuencia absoluta f_i*
(60,63)	5
(63,66)	18
(66,69)	42
(69,72)	27
(72,75)	8
	100

1. Buscamos el intervalo donde se encuentra la moda, que será el intervalo que tenga la mayor frecuencia absoluta f_i.

   La **clase modal** es: (66,69)

Aplicaremos la fórmula para el cálculo de la moda para datos agrupados:

$$M_o = L_i + \frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \cdot a_i$$

Intervalo	*Frecuencia absoluta f_i*
(60,63)	5
(63,66)	18
(66,69)	42
(69,72)	27
(72,75)	8
	100

$Límite\ inferior = 66$

$f_i = 42$

$f_{i-1} = 18$

$f_{i+1} = 27$

$a_i = 3$

$$M_o = 66 + \frac{(42-18)}{(42-18)+(42-27)} \cdot 3 = 67.846$$

$$M_o = 66 + \frac{27}{18+27} \cdot 3 = 67.8$$

Cálculo de la moda para datos agrupados

Caso 2: Cuando los intervalos tienen amplitudes distintas.

1. En primer lugar tenemos que hallar las alturas. $h_i = \dfrac{f_i}{a_i}$

2. La clase modal es la que tiene mayor altura.

$$M_o = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

3. La fórmula de la moda aproximada cuando existen distintas amplitudes es:

$$M_o = L_i + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} \cdot a_i$$

Ejemplo:

En la siguiente tabla se muestra las calificaciones obtenidas por un grupo de 50 alumnos. Calcular la moda.

Intervalo	*Frecuencia absoluta* f_i
[0,5)	15
[5,7)	20
[7,9)	12
[9,10)	3

0,1,2,3,4 Amplitud: 5

5,6 Amplitud: 2

7,8 Amplitud: 2

9 Amplitud: 1

- Creamos una nueva columna con las alturas, dividiendo las frecuencias absolutas entre las amplitudes de los intervalos correspondientes:

$$h_1 = \frac{15}{5} = 3$$

$$h_1 = \frac{20}{2} = 10$$

$$h_3 = \frac{12}{2} = 6$$

$$h_4 = \frac{3}{1} = 3$$

	Frecuencia absoluta f_i	*Altura* h_i
[0,5)	15	3
[5,7)	20	10
[7,9)	12	6
[9,10)	3	3
	50	

- La **clase modal** es [5,7) porque es la que tiene mayor altura

Aplicaremos la fórmula para el cálculo de la moda para datos agrupados con intervalos de amplitudes distintas:

$$M_o = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

	Frecuencia absoluta f_i	*Altura h_i*
[0,5)	15	3
[5,7)	20	10
[7,9)	12	6
[9,10)	3	3
	50	

$Límite\ inferior = 5$

$h_i = 10$

$f_{i-1} = 3$

$f_{i+1} = 6$

$a_i = 2$

$$M_o = L_i + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} \cdot a_i$$

$$M_o = 5 + \frac{(10-3)}{(10-3)+(10-6)} \cdot 2 = 6.27$$

$$M_o = 5 + \frac{6}{3+6} \cdot 2 = 6.33$$

Ventajas y desventajas de la media, la mediana y la moda

… la cuestión es …

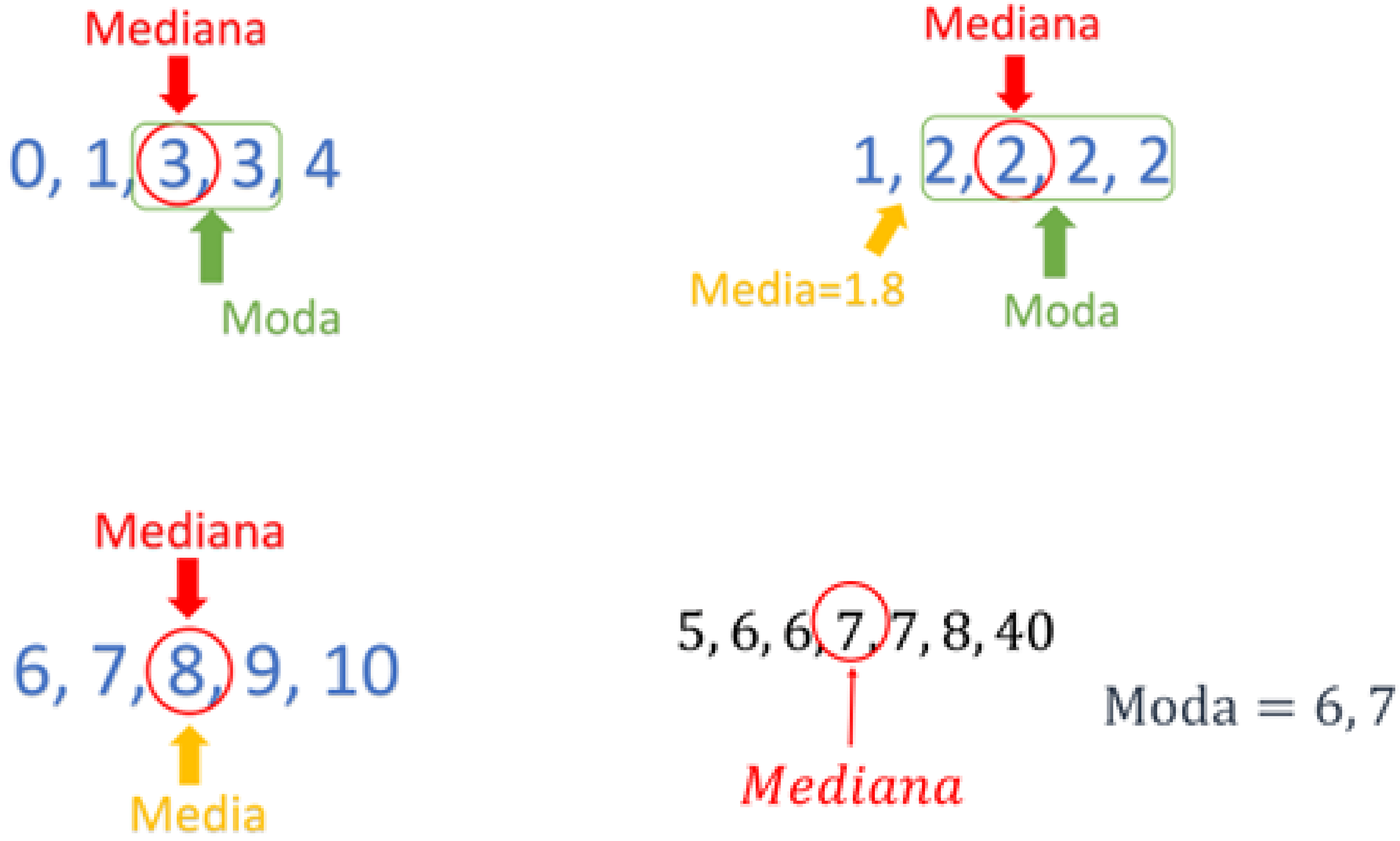
¿cuál de las medidas de tendencia central hay que utilizar?

- Si la distribución es simétrica, la media y la mediana son iguales, y si la distribución es simétrica y unimodal, las tres medidas son iguales.

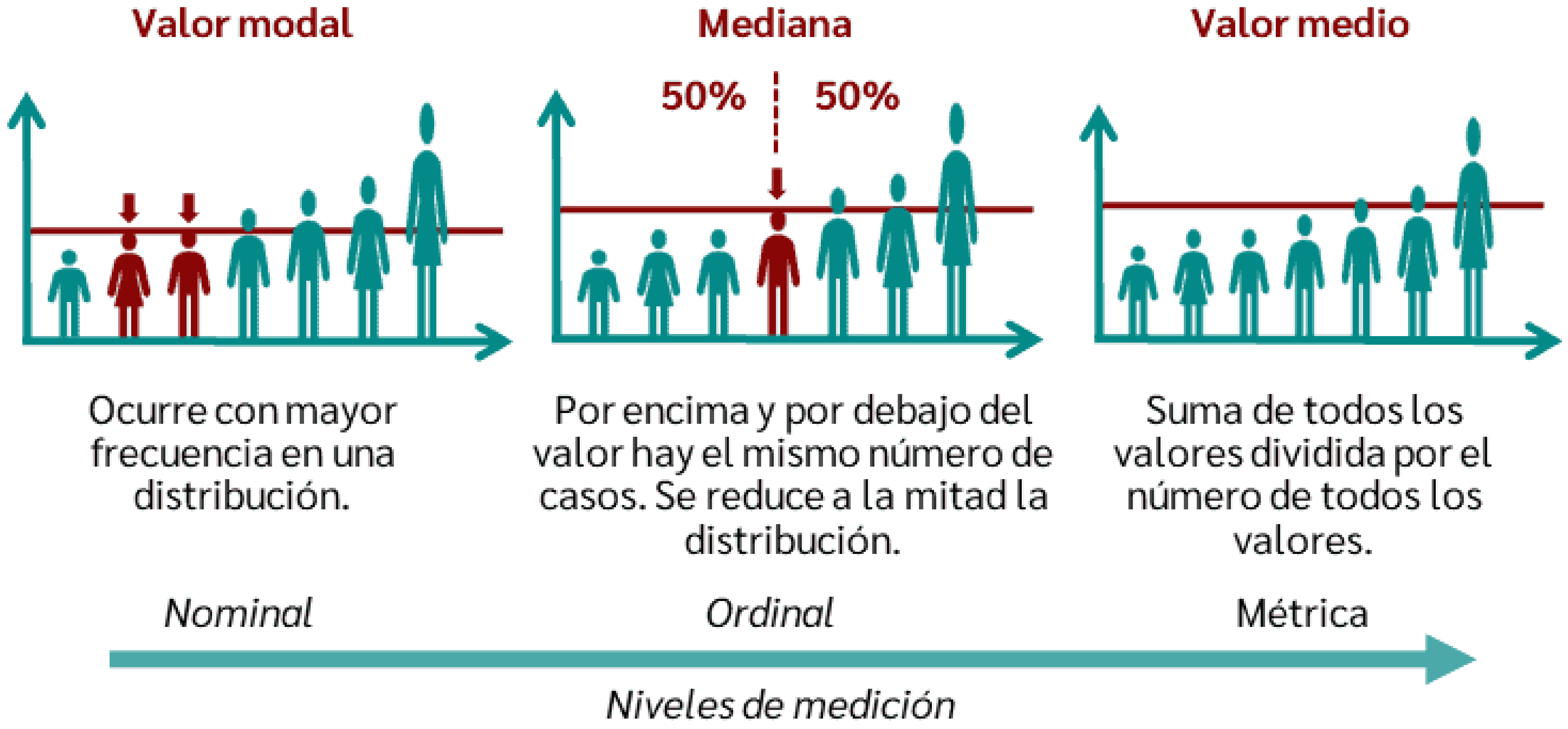
- **Media**: El valor medio es, con mucho, el más utilizado. Las desventajas de la media son que es sensible a los valores atípicos. **Los datos deben tener un nivel de escala métrica**.

- **Mediana**: La gran ventaja de la mediana es que es muy robusta frente a los valores atípicos y que **los datos sólo tienen que tener una escala ordinaria**.

- **Moda**: La moda es el valor que se da con mayor frecuencia, lo que tiene la ventaja de que el valor se da realmente. Además, la moda también puede calcularse para los datos que no pueden ordenarse y, por tanto, tienen un nivel de **escala nominal**. La desventaja es que la moda no tiene en cuenta los demás datos existentes.



Mediana
0, 1, 3, 3, 4
Moda
Mediana
1, 2, 2, 2, 2
Media=1.8
Moda
Mediana
6, 7, 8, 9, 10
Media
5, 6, 6, 7, 7, 8, 40
Moda = 6, 7
Mediana



Valor modal
Mediana
Valor medio
50%
50%
Ocurre con mayor frecuencia en una distribución.
Por encima y por debajo del valor hay el mismo número de casos. Se reduce a la mitad la distribución.
Suma de todos los valores dividida por el número de todos los valores.
Nominal
Ordinal
Métrica
Niveles de medición

Medidas de dispersión

- Rango
- Desviación media
- Varianza
- Desviación típica

Medidas de dispersión

- Las medidas de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.

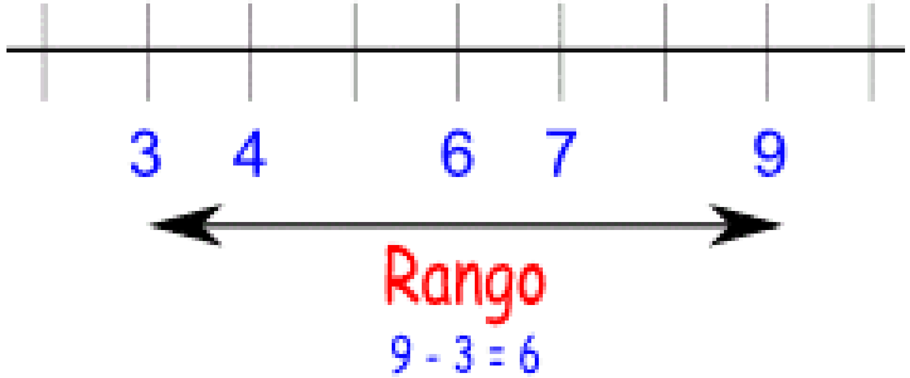
Rango

$$R = \text{Máx} - \text{Mín}$$

Rango o recorrido

- En estadística, el **rango** es la diferencia entre el mayor y el menor de los datos de una distribución estadística.

$$R = \text{Máx} - \text{Mín}$$



Ejemplo. Calcula el rango estadístico de las ventas mensuales de una empresa en el último año. Tienes registradas las ventas obtenidas de cada mes en la siguiente tabla:

Mes	Ventas
Enero	4785
Febrero	1803
Marzo	3636
Abril	1671
Mayo	1968
Junio	4694
Julio	5024
Agosto	4705
Setiembre	3494
Octubre	2461
Noviembre	2986
Diciembre	3145

$$R = \text{Máx} - \text{Mín}$$

$$\text{Máx} = 5024$$

$$\text{Mín} = 1671$$

$$R = 5024 - 1671 = 3353$$


Rango de las ventas de los últimos 12 meses
Máximo
Mínimo
RANGO
Ventas
Mes
6000
5000
4000
3000
2000
1000
0
1
2
3
4
5
6
7
8
9
10
11
12


PRODUCCIÓN DE MICROCHIPS
Microchips producidos
Mes
MÁXIMO
49.079
MÍNIMO
10.124
RANGO
60.000
50.000
40.000
30.000
20.000
10.000
0
0
10
20
30
40
50
60

Desviación media

$$D_{\bar{x}}$$

Desviación respecto a la media

- es la diferencia entre cada valor de la variable estadística y la media aritmética.

$$D_i = x_i - \bar{x}$$

Desviación media

- es la media aritmética de los valores absolutos de las desviaciones respecto a la media.

$$D_{\bar{x}} = \frac{|D_1|+|D_2|+\cdots+|D_n|}{N} = \frac{\sum_{i=1}^{n}|D_i|}{N}$$

$$D_{\bar{x}} = \frac{|x_1-\bar{x}|+|x_2-\bar{x}|+\cdots+|x_n-\bar{x}|}{N} = \frac{\sum_{i=1}^{n}|x_i-\bar{x}|}{N}$$

Ejemplo: Calcular la desviación media de la distribución:

$$9,\ 3,\ 8,\ 8,\ 9,\ 8,\ 9,\ 18$$

Ejemplo: Calcular la desviación media de la distribución:

$$9,\ 3,\ 8,\ 8,\ 9,\ 8,\ 9,\ 18$$

Primero calculamos la media:

$$\overline{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

Calculamos la desviación media:

$$\begin{aligned} D_{\overline{x}} &= \frac{|9-9|+|3-9|+|8-9|+|8-9|+|9-9|+|8-9|+|9-9|+|18-9|}{8} \\ &= 2.25 \end{aligned}$$

Desviación media para datos agrupados

- Si los datos vienen agrupados en una tabla de frecuencias, la expresión de la desviación media es:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| \cdot f_1 + |x_2 - \bar{x}| \cdot f_2 + \cdots + |x_n - \bar{x}| \cdot f_n}{N} = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| \cdot f_n}{N}$$

Ejemplo: Calcular la desviación media de la distribución:

Intervalo	x_i	f_i
[10,15)	12.5	3
[15,20)	17.5	5
[20,25)	22.5	7
[25,30)	27.5	4
[30,35)	32.5	2

x_i es la marca de clase

Intervalo	x_i	a_i	f_i	$x_i \cdot f_i$
[10,15)	12.5	5	3	37.5
[15,20)	17.5	5	5	87.5
[20,25)	22.5	5	7	157.5
[25,30)	27.5	5	4	110
[30,35)	32.5	5	2	65
			21	457.5

Calculamos la media:

$$\overline{x} = \frac{457.5}{21} = 21.786$$

Intervalo	x_i	a_i	f_i	$x_i \cdot f_i$	$\lvert x_i - \bar{x} \rvert$	$\lvert x_i - \bar{x} \rvert \cdot f_i$
[10,15)	12.5	5	3	37.5	9.286	27.858
[15,20)	17.5	5	5	87.5	4.286	21.43
[20,25)	22.5	5	7	157.5	0.714	4.998
[25,30)	27.5	5	4	110	5.714	22.856
[30,35)	32.5	5	2	65	10.174	21.428
			21	457.5		98.57

La desviación media es

$$D_{\overline{x}} = \frac{98.57}{21} = 4.69$$

Varianza

$$\sigma^2$$

Varianza

- es la media aritmética del cuadrado de las desviaciones respecto a la media de una distribución estadística.

$$\sigma^2 = \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2}{N}$$

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

Ejemplo: Calcular la varianza de la distribución:

$$9,\ 3,\ 8,\ 8,\ 9,\ 8,\ 9,\ 18$$

Primero calculamos la media:

$$\overline{x} = \frac{9+3+8+8+9+8+9+18}{8} = 9$$

Con el valor de la media, ya podemos encontrar la varianza:

$$\begin{aligned} \sigma^2 &= \frac{(9-9)^2+(3-9)^2+(8-9)^2+(8-9)^2+(9-9)^2+(8-9)^2+(9-9)^2+(18-9)^2}{8} \\ &= 15 \end{aligned}$$

Ejemplo: Calcular la varianza de la distribución:

$$12, 6, 7, 3, 15, 10, 18, 5.$$

$$12,\ 6,\ 7,\ 3,\ 15,\ 10,\ 18,\ 5.$$

Calculamos la media aritmética

$$\bar{x} = \frac{12+6+7+3+15+10+18+5}{8} = \frac{76}{8} = 9.5$$

Calculamos la varianza

$$\sigma^2 = \frac{(12-9.5)^2+(6-9.5)^2+(7-9.5)^2}{8} +$$

$$+ \frac{(3-9.5)^2+(15-9.5)^2+(10-9.5)^2}{8} +$$

$$+ \frac{(18-9.5)^2+(18-9.5)^2+(5-9.5)^2}{8} = 23.75$$

Ejemplo: Calcular la varianza de la distribución:

$$5,\ 5,\ 12,\ 13,\ 15,\ 15,\ 15,\ 20,\ 20,\ \ 23.$$

$$5,\ 5,\ 12,\ 13,\ 15,\ 15,\ 15,\ 20,\ 20,\ \ 23.$$

Usando las mismas etapas como en los ejemplos anteriores, primero calculamos la media aritmética y luego la varianza.

$$\bar{x} = \frac{5+5+12+13+15+15+15+20+20+23}{10} = \frac{143}{10} = 14.3$$

$$\sigma^2 = \frac{(5-14.3)^2 + (5-14.3)^2 + (12-14.3)^2}{10} +$$

$$+\frac{(13-14.3)^2 + (15-14.3)^2 + (15-14.3)^2}{10} + \quad +\frac{(13-14.3)^2 + (20-14.3)^2 + (20-14.3)^2}{10} +$$

$$+\frac{(23-14.3)^2}{10} = = \frac{86.49 + 86.49 + 5.29 + 1.69 + 0.49 + 0.49}{10} +$$

$$= \frac{0.49 + 32.49 + 32.49 + 75.69}{10} = \frac{322.1}{10} = 32.21$$

https://www.superprof.es/apuntes/escolar/matematicas/estadi
stica/descriptiva/varianza.html

Varianza para datos agrupados

$$\sigma^2 = \frac{(x_1-\overline{x})^2\cdot f_1+(x_2-\overline{x})^2\cdot f_2+\cdots+(x_n-\overline{x})^2\cdot f_n}{N}$$

$$= \frac{\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2\cdot f_i}{N}$$

Para simplificar el cálculo de la varianza vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma^2 = \frac{x_1^2+x_2^2+\cdots+x_n^2}{N}-\overline{x}^2$$

$$= \sum_{i=1}^{n}\frac{x_i^2}{N}-\overline{x}^2$$

Ejemplo: Calcular la **varianza** de la distribución de la tabla:

	x_i	f_i
[10,20)	15	1
[20,30)	25	8
[30,40)	35	10
[40,50)	45	9
[50,60)	55	8
[60,70)	65	4
[70,80)	75	2
		42

	x_i	f_i	$x_i \cdot f_i$	$x_i^2 \cdot f_i$
[10,20)	15	1	15	225
[20,30)	25	8	200	5000
[30,40)	35	10	350	12.250
[40,50)	45	9	405	18.225
[50,60)	55	8	440	24.200
[60,70)	65	4	260	16.900
[70,80)	75	2	150	11.250
		42	1.820	88.050

Calculamos la media

$$\overline{x} = \frac{1820}{42} = 43.33$$

La varianza es

$$\sigma^2 = \frac{88050}{42} - 43.33^2 = 218.94$$

Desviación típica

$$\sigma$$

Desviación típica

- es la raíz cuadrada de la varianza.

$$\sigma = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{N}}$$

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

Ejemplo. Calcular la desviación típica de la distribución:

$$9,3,8,8,9,8,9,18$$

Calculamos la media aritmética

$$=\frac{9+3+8+8+9+8+9+18}{8}=9$$

Sustituimos en la fórmula de la desviación típica

$$\sigma=\sqrt{\frac{(9-9)^2+(3-9)^2+(8-9)^2+(8-9)^2+(9-9)^2+(8-9)^2+(9-9)^2+(18-9)^2}{8}}$$

$$\sigma=3,87$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{(x_1-\overline{x})^2\cdot f_1+(x_2-\overline{x})^2\cdot f_2+\cdots+(x_n-\overline{x})^2\cdot f_n}{N}}$$

$$= \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2\cdot f_i}{N}}$$

Para simplificar el cálculo vamos o utilizar las siguientes expresiones que son equivalentes a las anteriores.

$$\sigma = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{N} - \overline{x}^2}$$

$$= \sqrt{\sum_{i=1}^{n} \frac{x_i^2}{N} - \overline{x}^2}$$

Desviación típica para datos agrupados

$$\sigma = \sqrt{\frac{x_1^2 \cdot f_1 + x_2^2 \cdot f_2 + \cdots + x_n^2 \cdot f_n}{N} - \overline{x}^2}$$

$$= \sqrt{\sum_{i=1}^{n} \frac{x_i^2 \cdot f_i}{N} - \overline{x}^2}$$

Ejemplo. Calcular la **desviación típica** de la distribución de la tabla:

	x_i	f_i
[10,20)	15	1
[20,30)	25	8
[30,40)	35	10
[40,50)	45	9
[50,60)	55	8
[60,70)	65	4
[70,80)	75	2

$$\sigma = \sqrt{\frac{88.050}{42} - 43,33^2} = 14,797$$

Coeficiente de variación

- es la relación entre la desviación típica de una muestra y su media.

$$\mathrm{C.V} = \frac{\sigma}{\overline{x}} \cdot 100$$

Puntuaciones típicas

- son el resultado de dividir las puntuaciones diferenciales entre la desviación típica.
- este proceso se llama tipificación.

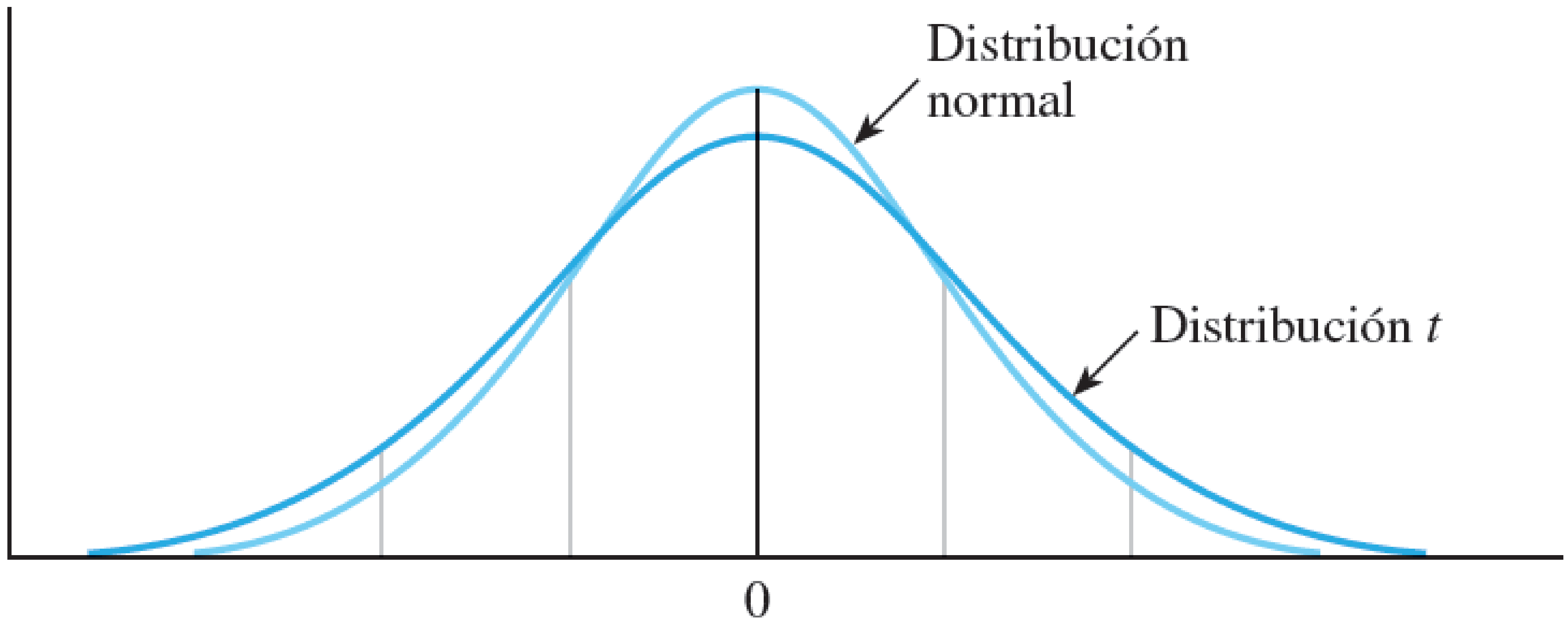
$$\mathrm{z} = \frac{x_i - \overline{x}}{\sigma}$$

Distribución Normal y Distribuciones Muestrales

Distribución Normal y Distribuciones Muestrales

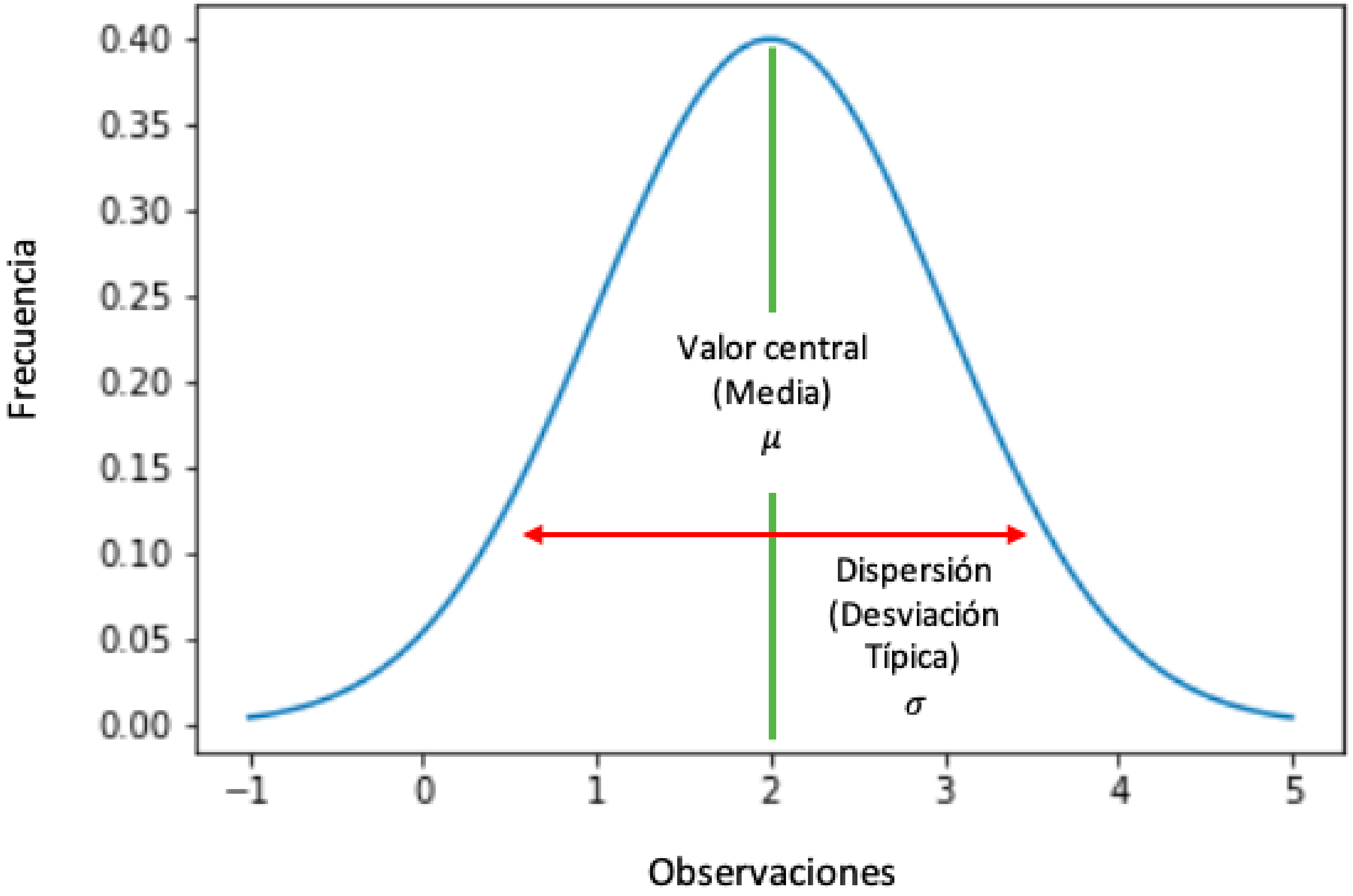
- La **distribución normal (o gaussiana)** se usa cuando la desviación estándar de la población es conocida y el tamaño de la muestra es grande (generalmente $n \geq 30$).

- La **distribución *t* de Student** es una versión más adaptable, usada cuando se trabaja con muestras pequeñas ($n < 30$) y se desconoce la variabilidad de la población.



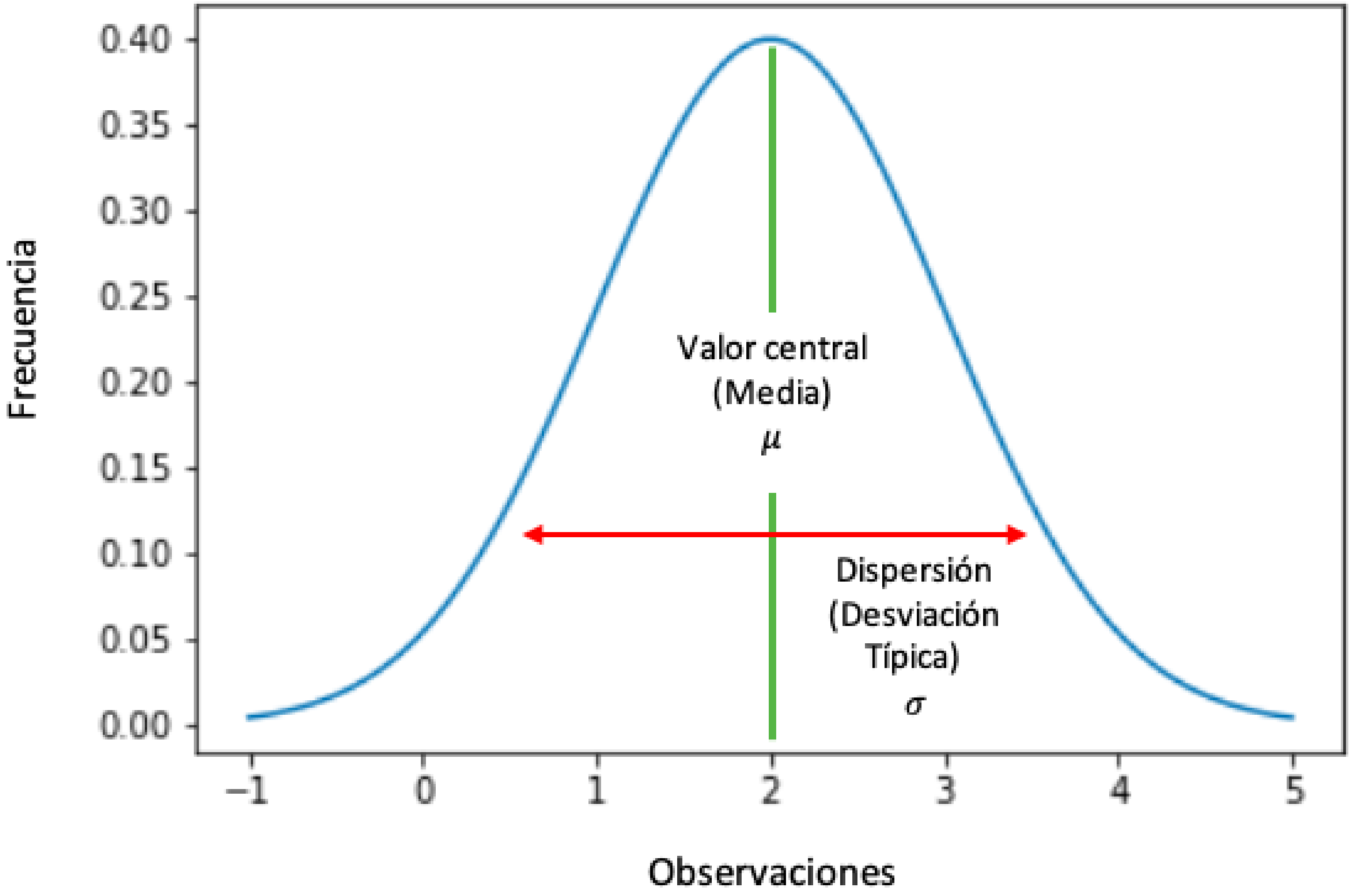
Distribución Normal

- la frecuencia de una variable aleatoria *x* puede representarse mediante una **distribución normal.**



Distribución Normal

- la frecuencia de una variable aleatoria *x* puede representarse mediante una **distribución normal.**



Distribución Normal

Propiedades

- Es una distribución simétrica. El valor de la media, la mediana y la moda coinciden.

$$\text{Media} = \text{Mediana} = \text{Moda}$$

- Distribución unimodal. Los valores que son más frecuentes o que tienen más probabilidad de aparecer están alrededor de la media. En otras palabras, cuando nos alejamos de la media, la probabilidad de aparición de los valores y su frecuencia descienden.

Ejemplo. Suponemos que queremos saber si los resultados de un examen pueden aproximarse satisfactoriamente a una distribución normal. Sabemos que en este examen participan 476 estudiantes. Calculamos la media y la desviación típica a partir de las observaciones (resultados del examen).

Resultados	Frecuencia
0	20
1	31
2	44
3	56
4	64
5	66
6	62
7	51
8	39
9	26
10	16
TOTAL	475

Calculamos la media y la desviación típica a partir de las observaciones (resultados del examen).

Media = 4.86 Desviación típica = 2.56

¿Los resultados del examen pueden aproximarse a una distribución normal?

Razones para considerar dicho caso son:

- Distribución simétrica. Es decir, existe el mismo número de observaciones tanto a la derecha como a la izquierda del valor central. También, que la media, la mediana y la moda tienen el mismo valor.

$$\text{Media} = \text{Mediana} = \text{Moda} = 5$$

- Las observaciones con más frecuencia o probabilidad están alrededor del valor central. En otras palabras, las observaciones con menos frecuencia o probabilidad se encuentran lejos del valor central.

¿Los resultados del examen pueden aproximarse a una distribución normal?

¿Los resultados del examen pueden aproximarse a una distribución normal?