Medidas de Tendencia Central

Media, mediana, moda, etc.

Instructor: Juan Luis Palacios Soto

palacios.s.j.l@gmail.com

Contenido

1. Media
2. Mediana
3. Moda
4. Media geométrica
5. Media armónica
6. Media ponderada

- Un automóvill va de la ciudad A a ciudad B que se encuentran separados a 100 km de distancia. Si en la ida va a 130 km/h y de regreso va a 90 km/h, calcule la velocidad media del vehículo.
- Una empresa tuvo los siguientes crecimientos de ventas en el semestre anterior de: 10 %, 32 %, 15 %, 8 %, 21 % y 29 %. Calcule el crecimiento promedio en el semestre anterior.
- Un alumno tiene las siguientes 5 calificaciones: 8,10,7,6.5 y 6. Las ponderaciones correspondientes a cada calificación son: 15 %, 25 %, 10 %, 30 %, 20 %, respectivamente. Calcule su promedio general.
- Dos esferas están unidas por medio de una varilla metálica y separados sus centros a una distancia de 10 metros. Calcule el punto de equilibrio entre las dos esferas.

$V=\frac{d}{t} \Rightarrow t_1=\frac{d_1}{v_1}=\frac{100\,km}{130}$

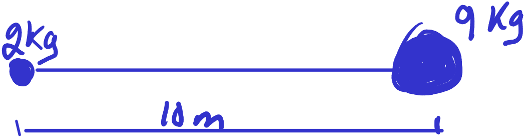
$t_1 \approx 46.15$ min

$t_2=\frac{d_2}{v_2}=\frac{100}{90}$

$t_2 \approx 66.66$ min

$V_m = ?$

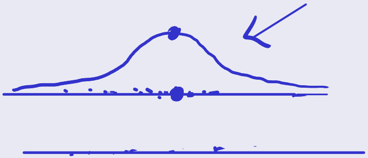
$T_t = 112h\,39s$



Medidas de tendencia central

1. Medidas de promedio
   - Media aritmética
   - Media Geométrica
   - Media Armónica
2. Medidas de particiones
   - Mediana
   - Cuantiles: Cuartiles, deciles, percentiles
3. Medidas de repetición
   - Moda
   - Bimodal
   - Multimodal

$f(x) = e^{-x^2}$



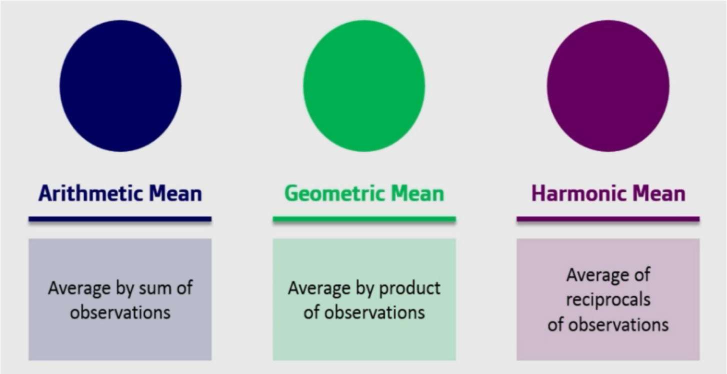
En Estadística, todos los datos están distribuidos a través de varios puntos, que es lo que se conoce como la distribución de los datos.

A partir de la distribución de los datos es muy complicado llegar a conclusiones, pues muchas veces corresponden a distribuciones demasiado complejas. Sin embargo, muchas veces existe una tendencia de los datos a *juntarse* al rededor de un cierto valor. Este valor es lo que se conoce como una tendencia central.

De esta manera, la tendencia central es un representante global de los datos. En particular,representa las características generales del conjunto de datos. Las medidas de tendencia central se clasifican en tres formas: valores promedios, valores de partición y valores repetidos.

Medidas de promedio

Las Medidas de promedio más empleadas son: media aritmética, media geométrica y media armónica.



Media aritmética

En general, es considerada como el valor promedio de las observaciones. Se trata de sumar todos los valores y dividir entre el número de observaciones. Una característica importante es que la suma de todas las desviaciones de cada valor respecto de su media aritmética siempre es 0. Además, se ve afectada por cambios en las escalas, así como si se suma un valor constante a cada uno de los datos.

Para datos no agrupados: Si tenemos datos no agrupados, $x_1, x_2, ..., x_n$, la media aritmética, denotada por $\bar{x}$, se define como

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Para datos agrupados: Para datos agrupados se define como

$$\Rightarrow \bar{x} = \frac{1}{N}\sum_{i=1}^{k} x_i f_i,$$

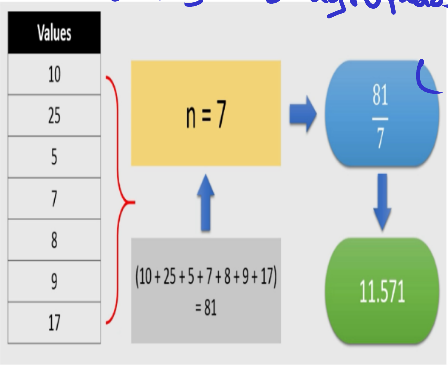
donde x_i es la marca de clase, f_i la frecuencia de la clase $i = 1, 2, ..., k$ y k es el total de clases.

	clase	x_i	f_i	F_i
1	[0,2)	1	9	9
2	[2,4)	3	8	17
⋮				
K	[,)			

Definición (Media aritmética en Excel)

En Excel, utilizamos las funciones PROMEDIO y PROMEDIO.SI para calcular la media aritmética. Su sintaxis es idéntica a la de las funciones SUMA y SUMA.SI, respectivamente.

Datos no agrupados



x_i Datos agrupados.

$C_1 = \{1, 3, 5\}$ $\bar{x} = 3$ $C_3 = \{-1, 3, 7\}$

$C_2 = \{0, 1, 2, 3, 4, 5, 6\}$ $\bar{x} = 3$ $C_4 = \{2, 3, 4\}$

Definición (Media Geométrica)

Se trata de multiplicar todos los valores y calcular a ese producto la raíz n-ésima donde n es la cantidad de observaciones. Puede aplicarse en Finanzas, ya que está íntimamente relacionada con la tasa de crecimiento anual compuesto.

Para datos no agrupados: *Si tenemos datos no agrupados, $x_1, x_2, ..., x_n$, la media geométrica, se define como*

$$MG = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = (x_1 \cdot x_2 \cdots x_n)^{1/n}$$

Para datos agrupados: *Para datos agrupados se define como*

$$MG = \sqrt[N]{x_1^{f_1} \cdot x_2^{f_2} \cdots x_k^{f_k}}$$

donde x_i es la marca de clase, f_i la frecuencia de la clase $i = 1, 2, ..., k$ y k es el total de clases.

$$C_1 = (C_0 + C_0 i) = C_0(1+i)$$

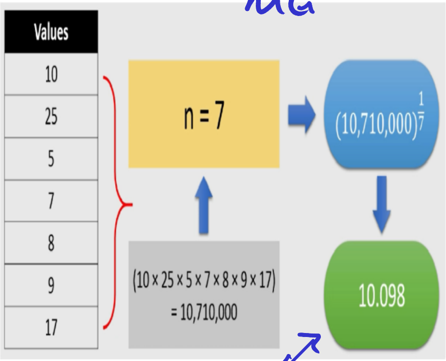
$$C_2 = C_1 + C_1 i = C_1(1+i) = C_0(1+i)(1+i)$$

$$= C_0(1+i)^2 \Rightarrow C_n = C_0(1+i)^n.$$

Definición (Media geométrica en Excel)

En Excel, utilizamos la función MEDIA.GEOM (GEOMEAN) para calcular la media geométrica. Su sintaxis es idéntica a la de la función SUMA.

MG



$\bar{x} = 11.571$

$\bar{x} = 6.307$

$MG \leq MA.$

Definición (Media Armónica)

Se trata de dividir el total de elementos entre la suma de los inversos multiplicativos de cada dato. Se aplica en situaciones que involucran tasas y razones ya que la media armónica proporciona el promedio correcto. Ejemplos en velocidad, densidad, electricidad (series en paralelo), óptica, finanzas (precio-beneficio), geometría y en general en situaciones en que los datos son del tipo fracción.

Para datos no agrupados: *Si tenemos datos no agrupados, $x_1, x_2, ..., x_n$, la media harmónica se define como*

$$\frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$

$7 \to \frac{1}{7}$

Para datos agrupados: *Para datos agrupados se define como*

$-3 \to -\frac{1}{3}$

$$\frac{N}{\sum_{i=1}^{k} \frac{f_i}{x_i}} = \frac{N}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \cdots + \frac{f_k}{x_k}}$$

$\frac{1}{5} \Rightarrow 5$

donde x_i es la marca de clase, f_i la frecuencia de la clase $i = 1, 2, ..., k$ y k es el total de clases.

$V = \frac{d}{t}$

$V_1 = 10 Km/h$

$V_2 = 11 Km/h$

$V_3 = 8 Km/h$

$V_4 = 15 Km/h$

Definición (Media armónica en Excel)

En Excel, utilizamos la función MEDIA.ARMO (HARMEAN) para calcular la media armónica. Su sintaxis es idéntica a la de la función SUMA.

Values	Reciprocals	Calculated
10	1/10	0.1
25	1/25	0.04
5	1/5	0.2
7	1/7	0.142857
8	1/8	0.125
9	1/9	0.111111
17	1/17	0.058824

(0.1 + 0.04 + 0.2 + 0.142857 + 0.125 + 0.11111 + 0.058824) = 0.777791783

n = 7

$\dfrac{7}{0.777791783}$ 8.999

Class (x)	Frequency (f)	f/x
2	2	2/2 = 1
4	3	3/4 = 0.75
6	2	2/6 = 0.3333
8	3	3/8 = 0.375
10	3	3/10 = 0.3

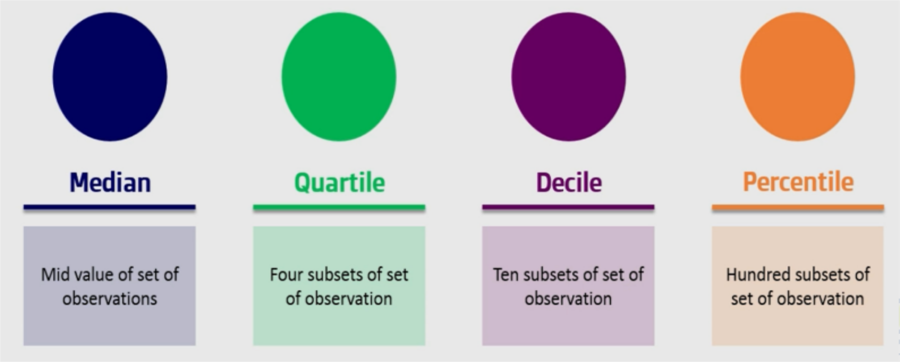
(1 + 0.75 + 0.3333 + 0.375 + 0.3) = 2.7583

N = 13

$\dfrac{13}{2.7583}$ 4.713

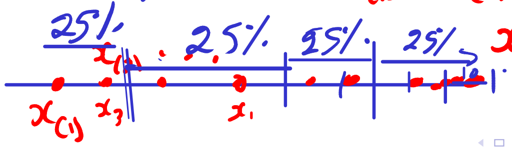
Med Arm ≤ Med Geo ≤ Med Art.

La función más importante de una medida de partición es dividir la colección de datos en varios subconjuntos. Tenemos aquí la mediana, los cuartiles, los deciles y los percentiles, entre otros.



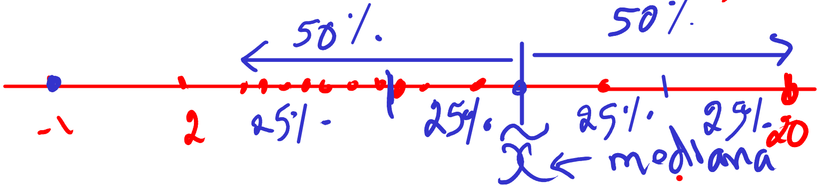
$x_1, x_2, \ldots, x_n$ $\qquad x_{(1)}, x_{(2)}, \ldots, x_{(n)}$

$x_{(3)} \leq x_{(4)}$



$x_1, x_2, x_3, \ldots, x_n \rightarrow 7, -1, 2, \ldots, 5$

ord. $x_{(1)}, x_{(2)}, \ldots, x_{(n)} \Rightarrow -1, 2, \ldots, 20$



$1, 7, 9$ cant. impar

$1, 7, 9, 21$ cant. par

8

$1, 7, 21$

$1 \quad \frac{1}{7} \quad \frac{1}{21}$

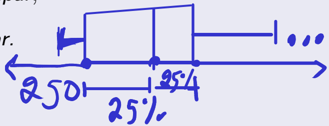
$$\frac{3}{1 + \frac{1}{7} + \frac{1}{21}}$$

Definición (Mediana)

La **mediana** *divide las observaciones en el 50 % más alto y el 50 % más bajo. Se le considera también una medida de la posición. Además, es muy importante en los llamados gráficos de bigote.*

Para datos no agrupados: *Sean $x_1, x_2, ..., x_n$ un conjunto de datos numérico. Para determinar la mediana en este conjunto de datos debemos primero ordenar de menor a mayor estos datos, los cuales denotaremos por, $x_{(1)}, x_{(2)}, ..., x_{(n)}$. La mediana en un conjunto de datos pequeño, dependerá de si n es par, $n = 2k$, o si es impar, $n = 2k - 1$ con $k \in \mathbb{N}$. Denotaremos la mediana como $\tilde{x}$, calculada por*

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{si } n \text{ es impar,} \\ \dfrac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{si } n \text{ es par.} \end{cases}$$



Para datos agrupados: *Para datos agrupados se define como*

$$M_e = L_i + \left(\frac{\frac{N}{2} - F_{i-1}}{f_i} \right) \Delta x$$

donde Δx es la amplitud de las clases, f_i la frecuencia de la clase $i = 1, 2, ..., k$, L_i es el límite inferior de la clase donde se encuentra el dato medio $\frac{N}{2}$ y F_{i-1} es la frecuencia acumulada en la clase anterior.

$\Delta x = 2 \rightarrow [0, 2)$ ——— F_i

$[2, 4) \leftarrow$

Definición (Mediana en Excel)

En Excel, utilizamos la función MEDIANA para calcular la mediana. Su sintaxis es idéntica a la de la función SUMA.

Values	Arranged
25	8
8	10
15	13
10	15
13	25

13

Values	Arranged
25	8
8	10
15	13
10	15
35	25
26	26
40	35
13	40

$$\frac{15+25}{2}$$

20

Definición (Mediana para datos no agrupados)

$$M_e = L_i + \left(\frac{\frac{N}{2} - F_{i-1}}{f_i}\right)\Delta x$$

donde Δx *es la amplitud de las clases,* f_i *la frecuencia de la clase* $i = 1, 2, ..., k$*,* L_i *es el límite inferior de la clase donde se encuentra el dato medio* $\frac{N}{2}$ *y* F_{i-1} *es la frecuencia acumulada en la clase anterior.*

Datos agrupados

Edad (años)	Marca de clase x_i	Frecuencia f_i	$x_i \cdot f_i$
[0 - 2)	1	7	7
[2 - 4)	3	8	24
[4 - 6)	5	8	40
[6 - 8]	7	7	49
	Σ	30	120

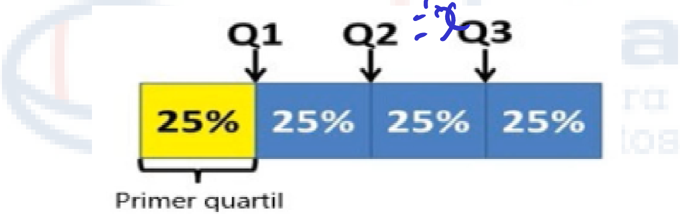
$$M_e = 2 + \left(\frac{\frac{30}{2} - 7}{8}\right)2 = 4$$

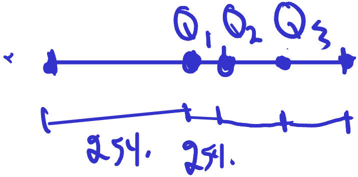
Definición (Cuartiles)

Se utiliza para dividir las observaciones en cuatro partes iguales. Por lo tanto existen tres ellos: Q_1, Q_2 y Q_3, donde Q_1 es la partición hasta el 25 % de los datos; Q_2 es la partición hasta el 50 % de los datos, y por lo tanto Q_2 = mediana; y Q_3 es la partición hasta el 75 % de los datos.

Si tenemos los valores $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ entonces

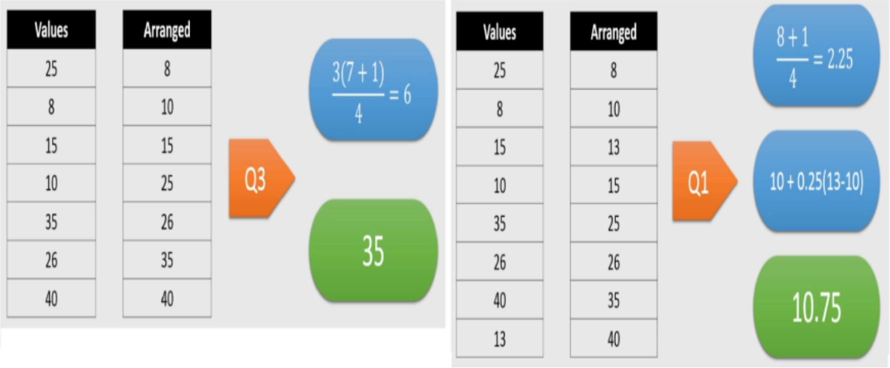
$$Q_1 \sim x_{(n+1)/4}, \quad Q_2 \sim x_{(n+1)/2}, \quad Q_3 \sim x_{3(n+1)/4}$$





Definición (Cuartiles en Excel)

En Excel, utilizamos las funciones CUARTIL.INC y CUARTIL.EXC para calcular los cuartiles. La sintaxis de CUARTIL.INC es CUARTIL.INC(rango, número de cuartil) al igual que la de CUARTIL.EXC. La diferencia entre ambas es que INC incluye los valores extremos para realizar el cálculo en tanto que EXC no.

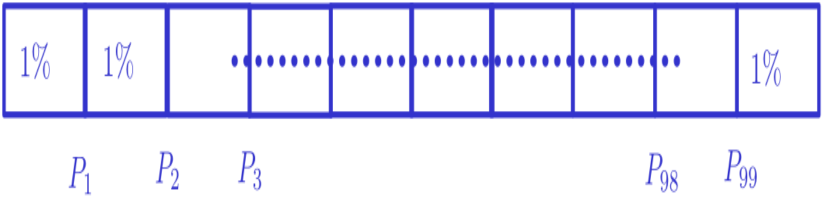


Definición (Percentiles)

Se utiliza para dividir las observaciones en cien partes iguales. Por lo tanto existen 99 de ellos: $P_1, P_2, ..., P_{99}$ *, donde* P_1 *es la partición hasta el 1 % de los datos;* P_2 *es la partición hasta el 2 % de los datos,..., y* P_{99} *es la partición hasta el 99 % de los datos..*

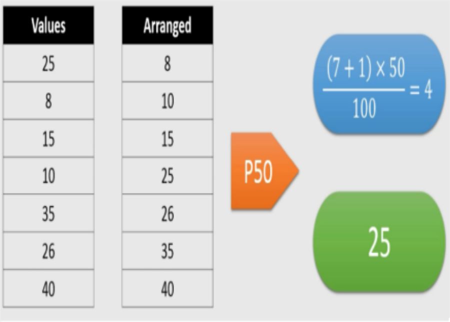
Si tenemos los valores $x_1, x_2, ..., x_n$ *entonces*

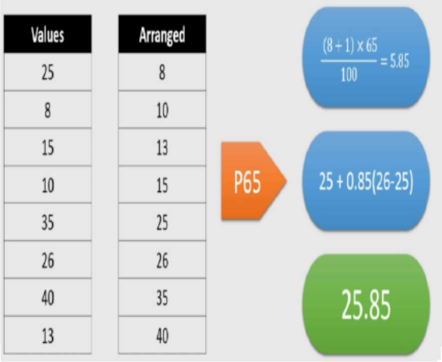
$$P_i \sim x_{(n+1)} \times i/100$$



Definición (Percentiles en Excel)

En Excel, utilizamos las funciones PERCENTIL.INC y PERCENTIL.EXC para calcular los ~~cuartiles~~ percentiles. La sintaxis son idénticas a las de CUARTIL.



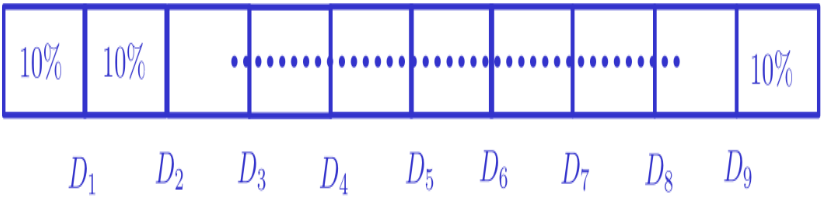


Definición (Deciles)

Se utiliza para dividir las observaciones en cien partes Se utiliza para dividir las observaciones en 10 partes iguales. Por lo tanto existen 9 de ellos:$D_1, D_2, ..., D_9$*, donde* D_1 *es la partición hasta el 10 % de los datos;* D_2 *es la partición hasta el 20 % de los datos,..., y* D_9 *es la partición hasta el 90 % de los datos.*

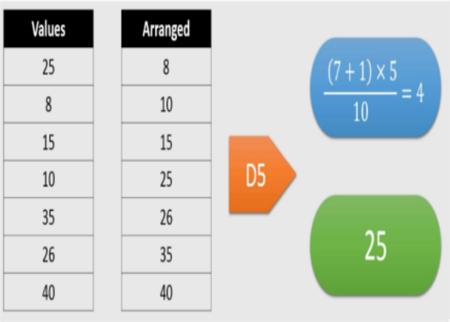
Si tenemos los valores $x_1, x_2, ..., x_n$ *entonces*

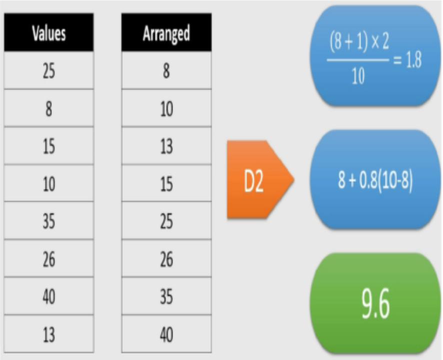
$$D_i \sim x_{(n+1)} \times i/10$$



Definición (Deciles en Excel)

No hay una función para deciles en Excel; sin embargo utilizamos las funciones PERCENTIL.INC y PERCENTIL.EXC como Di = PERCENTIL.EXC(rango,$10 \times i$).





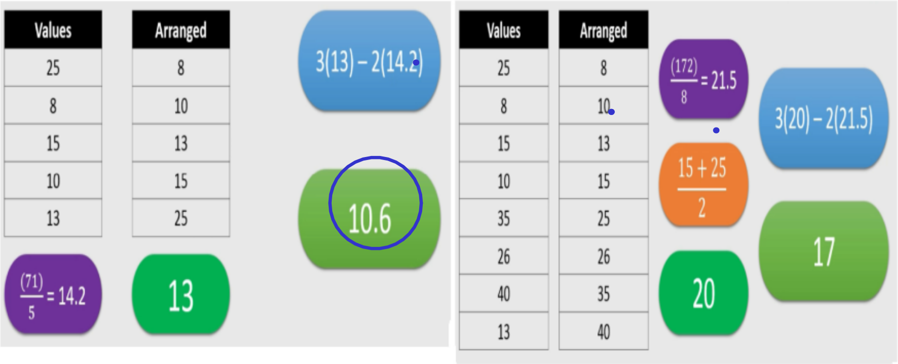
Definición (Moda)

Sólo existe una medida de repetición: la moda. Esta se refiere al valor que más veces se repite. En la práctica es la menos usada de todas las medidas de tendencia central ya que no provee suficiente claridad acerca de las características de un conjunto de datos, aunque suele usarse para datos nominales.

En el caso en que no hay valores repetidos, se calcula como Moda $= 3\times$ *Mediana* $-2\times$ *Media Aritmética. En el caso en que hay valores repetidos, es el que más se repite.*

Definición (Moda en Excel)

En Excel utilizamos la función MODA.UNO en el caso en que hay valores repetidos. Su sintaxis es idéntica a la de la función SUMA.



Values	Arranged
25	8
8	10
15	13
10	15
13	25

Values	Arranged
25	8
8	10
15	13
10	15
35	25
26	26
40	35
13	40