



Estadística con Excel

Héctor Manuel Garduño Castañeda

Octubre, 2021



Contenido

Medidas de tendencia central

Medidas de promedio

Media aritmética

Media Geométrica

Media Armónica

Medidas de particiones

Mediana

Cuartiles

Percentiles

Deciles

Medidas de repetición



- ○
○○
○○
○○○

- ○○
○○
○○
○○

-

Conceptos de tendencia central

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

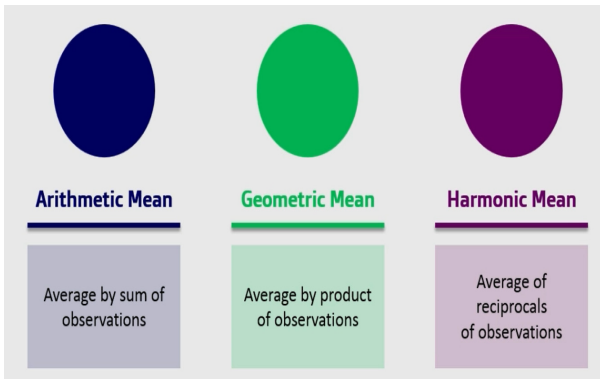
A partir de la distribución de los datos es muy complicado llegar a conclusiones, pues muchas veces corresponden a distribuciones demasiado complejas. Sin embargo, muchas veces existe una tendencia de los datos a *juntarse* al rededor de un cierto valor.

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.



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



Entendamos qué es la **Media aritmética**.

Concepto. 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.

Descripción matemática Si tenemos datos no agrupados, la media aritmética viene dada por $\frac{1}{n} \sum x_i$ donde x_i es el valor de la observación i -ésima y n es el número de observaciones.

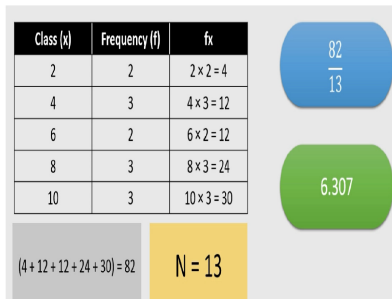
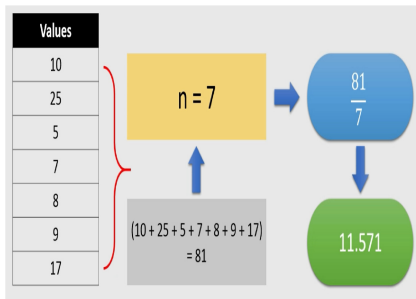
Para datos agrupados se define como

$\frac{1}{N} \sum f_i x_i = \sum f_i x_i / \sum f_i$ donde f_i es la frecuencia en que el valor del grupo i , x_i , se repite.



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.



Media Geométrica

Entendamos qué es la **Media geométrica**.

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

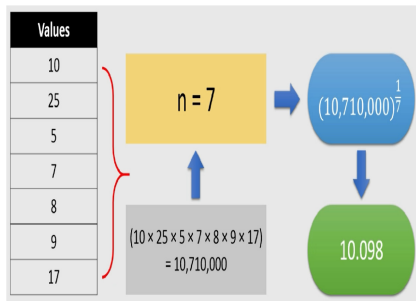
Descripción matemática Si tenemos datos no agrupados, la media geométrica viene dada por $\sqrt[n]{x_1 \cdot x_2 \cdots x_n}$ donde x_i es el valor de la observación i -ésima y n es el número de observaciones.

Para datos agrupados se define como $\sqrt[N]{x_1^{f_1} x_2^{f_2} \cdots x_N^{f_N}}$ donde f_i es la frecuencia en que el valor del grupo i , x_i , se repite.



Media Geométrica en Excel

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



Class (x)	Frequency (f)	x^f
2	2	$2^2 = 4$
4	3	$4^3 = 64$
6	2	$6^2 = 36$
8	3	$8^3 = 512$
10	3	$10^3 = 1000$

$(4 \times 64 \times 36 \times 512 \times 1000) = 4,718,592,000$
N = 13
 $(4,718,592,000)^{\frac{1}{13}}$
5.548

Media Geométrica

Entendamos qué es la **Media armónica**.

Concepto. Se trata de dividir el total de elementos entre la suma de los inversos multiplicativos de cada dato. Se suele utilizar cuando los datos son de tipo fracción. Por ejemplo, goles de equipos por partido de futbol.

Descripción matemática Si tenemos datos no agrupados, la media geométrica viene dada por $n / \sum \frac{1}{x_i}$ donde x_i es el valor de la observación i -ésima y n es el número de observaciones.

Para datos agrupados se define como $N / \sum \frac{f_i}{x_i}$ donde f_i es la frecuencia en que el valor del grupo i , x_i , se repite.



Media Armónica en Excel

En Excel, utilizamos la función **MEDIA.ARMO** 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

$$(0.1 + 0.04 + 0.2 + 0.142857 + 0.125 + 0.111111 + 0.058824) = 0.777791783$$

$$n = 7$$

$$\frac{7}{0.777791783}$$

$$8.999$$

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

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

$$N = 13$$

$$\frac{13}{2.7583}$$

$$4.713$$

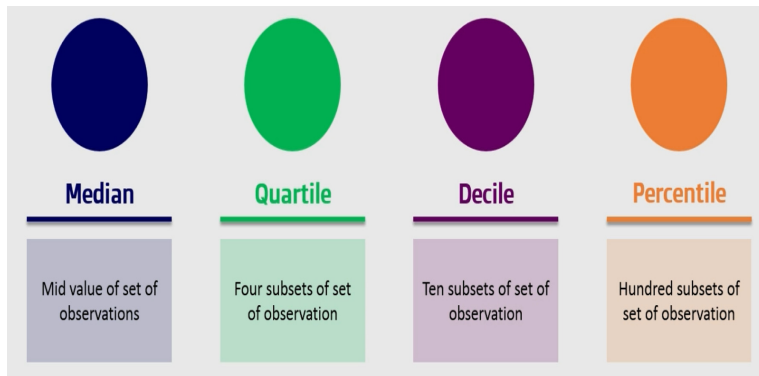
Lectura complementaria

Para mas información, echar un vistazo a la siguiente liga de lectura complementaria:

<https://towardsdatascience.com/on-average-youre-using-the-wrong-average-geometric-harmonic-means-in-data-analysis-2a703e21ea0>



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.



Mediana

Entendamos qué es la **Mediana**.

Concepto. Se utiliza para dividir 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.

Descripción matemática Si tenemos las observaciones x_1, x_2, \dots, x_n , entonces:

► Si n es impar, se tiene $mediana = \frac{x_{n+1}}{2}$

► Si n es par, se tiene $mediana = \frac{x_{n/2} + x_{n/2+1}}{2}$



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	13
25	8	
8	10	
15	13	
10	15	
13	25	

Values	Arranged	$\frac{15 + 25}{2}$
25	8	
8	10	
15	13	
10	15	
35	25	
26	26	20
40	35	
13	40	

Cuartiles

Entendamos qué son los **cuartiles**.

Concepto. 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 $Q_2 = \text{mediana}$; y Q_3 es la partición hasta el 75% de los datos.

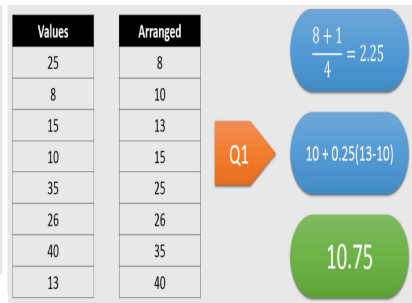
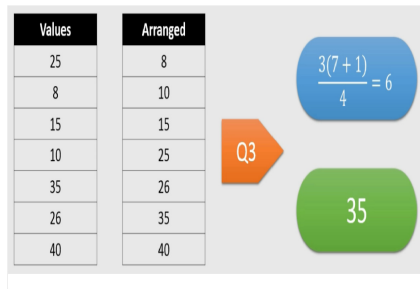
Descripción matemática Si tenemos los valores x_1, x_2, \dots, x_n entonces:

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



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.



Percentiles

Entendamos qué son los **percentiles**.

Concepto. Se utiliza para dividir las observaciones en cien partes iguales. Por lo tanto existen 99 de ellos: P_1, P_2, \dots, 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.

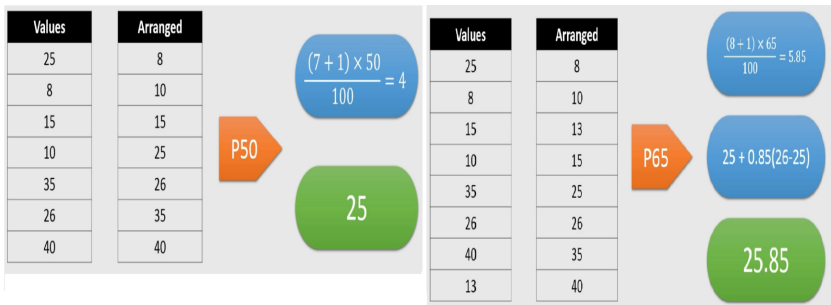
Descripción matemática Si tenemos los valores x_1, x_2, \dots, x_n entonces:

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



Percentiles en Excel

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



Deciles

Entendamos qué son los **deciles**.

Concepto. Se utiliza para dividir las observaciones en 10 partes iguales. Por lo tanto existen 9 de ellos: D_1, D_2, \dots, 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.

Descripción matemática Si tenemos los valores x_1, x_2, \dots, x_n entonces:

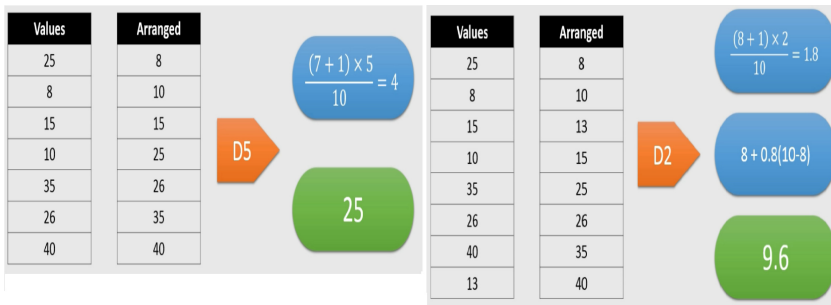
$$D_i \sim (n + 1) \times \frac{i}{10}$$



Deciles en Excel

No hay una función para deciles en Excel; sin embargo utilizamos las funciones **PERCENTIL.INC** y **PERCENTIL.EXC** como

$D_i = \text{PERCENTIL.EXC}(\text{rango}, 10 \times i)$.



Moda

Entendamos qué es la **Moda**.

Concepto. Solo 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.

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



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.

