



## Estadística con Excel

Héctor Manuel Garduño Castañeda

Octubre, 2021



# Contenido

## Medidas de dispersión

### Medidas de dispersión absoluta

- El rango

- Desviación media absoluta

- Desviación intercuartil

- Desviación estándar





## Conceptos

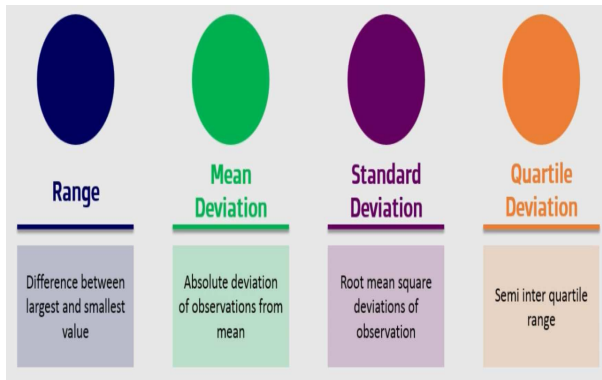
Las dispersiones o desviaciones son una manera de medir *qué tan alejados están los puntos de alguna medida de tendencia central*. Se pueden calcular para cualquiera de las medidas de tendencia central que estudiamos en el capítulo anterior.

Básicamente, son medidas que utilizamos para conocer que tan dispersos se encuentran nuestros datos de algún valor central. Esto significa que si hay una desviación alta, entonces nuestros datos se alejan mucho de la medida de tendencia central y por lo tanto esta última no es un buen representante de nuestros datos.

Existen dos tipos de medidas de dispersión: absolutas y relativas.



Son cuatro: el rango, desviación absoluta media, la desviación intercuartil y la desviación estándar.



Entendamos qué es la **rango**.

**Concepto.** Se define simplemente como la diferencia entre el valor más alto y el valor más pequeño. Por lo tanto, se ve fuertemente afectado por los valores atípicos. Es por esta razón que no se trata de la mejor medida de dispersión, además de que su cálculo no involucra ninguna medida de tendencia central.

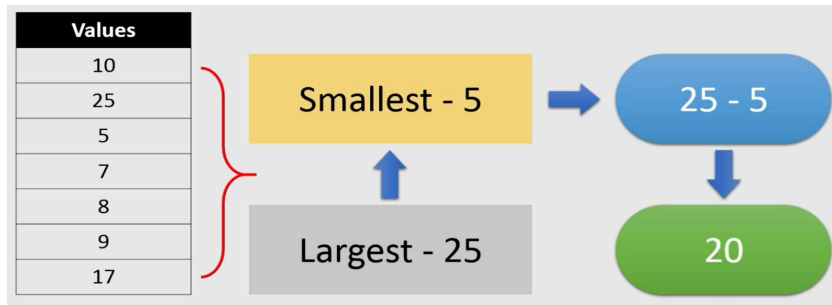
**Descripción matemática** Como la misma definición de rango lo indica, simplemente tenemos

$$Rango = val_{\max} - val_{\min}$$



## Rango en Excel

No existe en Excel una manera de calcular el rango de los datos directamente; sin embargo, recuerda que Excel tiene las funciones MIN y MAX.



Entendamos qué es la **desviación media absoluta**.

**Concepto.** Considera directamente todos los valores de los datos uno por uno. Se trata de medir cómo se aleja, en promedio, la ”‘distancia’” entre cada dato y alguna medida de tendencia central. Generalmente esta última se toma como la media aritmética o la mediana.

**Descripción matemática** Supongamos que tenemos los  $n$  datos  $x_1, x_2, \dots, x_n$  (no necesariamente están ordenados). Sea  $A$  alguna medida de tendencia central. Entonces la desviación media absoluta se calcula como

$$MD = \frac{1}{n} \sum |x_i - A|$$



## Desviación media absoluta en Excel

En Excel, cuando la medida de tendencia central es la media aritmética (es decir,  $A = MA$  en la fórmula de  $MD$ ), se utiliza la función **DESVPROM**.

$x$	$ x - \bar{x} $
4	8
10	2
15	3
12	0
13	1
18	6

$$(4 + 10 + 15 + 12 + 13 + 18) = 72$$

$$\frac{72}{6} = 12$$

$$\frac{20}{6}$$



Entendamos qué es la **Desviación intercuartil**.

**Concepto.** Representa qué tanto se desvía el 50% de los datos ”‘de enmedio’” respecto de la mediana, por lo que se calcula utilizando  $Q1$  y  $Q3$ . Como no toma en cuenta valores debajo de  $Q1$  ni arriba de  $Q3$ , no se ve afectado por los valores extremos. Tampoco se ve afectado si se suma una constante a todos los datos, pero sí se afecta si se multiplican todos los datos por cualquier número.

**Descripción matemática** Se calcula directamente como

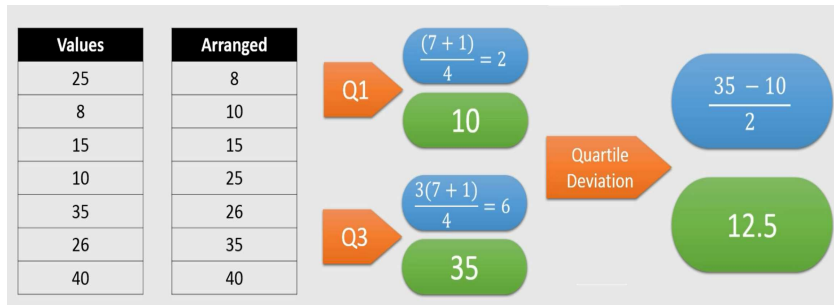
$$DQ = \frac{Q3 - Q1}{2}.$$

Al numerador de la fórmula anterior se le conoce como **rango intercuartil** y se representa como  $RQ$ .



## Desviación intercuartil en Excel

No existe en Excel una función directa para calcularlo, pero recordemos que tenemos la función CUARTIL.EXC.



Entendamos qué es la **Desviación estándar**.

**Concepto.** Es considerada como la mejor medida de dispersión. Se utiliza, junto con la media aritmética, para dar reglas empíricas acerca de los valores poco comunes (como la regla de las desviaciones para distribuciones gaussianas). Al igual que la anterior, no se ve afectada por si sumamos un valor a todos los datos; sin embargo, se afecta (y en gran medida) si todos los datos se multiplican por un mismo número.

**Descripción matemática** Si consideramos una muestra  $x_1, x_2, \dots, x_n$  (los datos no están necesariamente ordenados), se calcula como

$$SD = \sqrt{\frac{1}{n-1} \sum (x_i - MA)^2}.$$



## Desviación estándar en Excel

En Excel existen dos funciones para calcular desviación estándar:

**DESVEST.M** y **DESVEST.P**. La diferencia entre ellas es que la primera calcula la desviación estándar MUESTRAL (y por lo tanto aplica la fórmula de la diapositiva anterior) y la segunda calcula la desviación estándar POBLACIONAL (que en lugar dividir por  $n - 1$  en la fórmula anterior, divide por el tamaño de la población).

Class (x)	$x - \bar{x}$	$(x - \bar{x})^2$
25	10	100
10	- 5	25
15	0	0
12	- 3	9
13	- 2	4

$$(25 + 10 + 15 + 12 + 13) = 75$$

$$\frac{75}{5} = 15$$

$$(100 + 25 + 0 + 9 + 4) = 138$$

$$\sqrt{\frac{138}{4}}$$

$$5.25$$

Entendamos qué es la **Varianza**.

**Concepto.** Básicamente se trata de la desviación estándar elevada al cuadrado

**Descripción matemática** Si consideramos una muestra  $x_1, x_2, \dots, x_n$  (los datos no están necesariamente ordenados), se calcula como

$$Var = \frac{1}{n-1} \sum (x_i - MA)^2.$$



## Varianza en Excel

Nuevamente, en Excel existen dos funciones para calcular la varianza: **VAR.S** y **VAR.P**. La diferencia entre ellas es que la primera calcula la varianza MUESTRAL (sample, en inglés) y la segunda calcula la varianza POBLACIONAL.

Class (x)	$x - \bar{x}$	$(x - \bar{x})^2$
30	-11	121
22	-19	361
78	37	1369
20	-21	441
55	14	196

$(121 + 361 + 1369 + 441 + 196) = 2488$

$$\frac{2488}{4}$$

$$622$$

$$(30 + 17 + 78 + 20 + 55) = 205$$

$$\frac{205}{5} = 41$$