

Conceptos básicos de estadística

Profesor: Miguel Jiménez

Material de los cursos:

<https://migueljimenezg.github.io/cursos/>

Distribuciones de frecuencias

Son una forma de organizar los datos para entenderlos más fácil. Particularmente en gran cantidad de datos.

Los datos se agrupan en clases que abarcan valores semejantes dentro de un rango.

Esta agrupación de clases se hace por medio de intervalos.

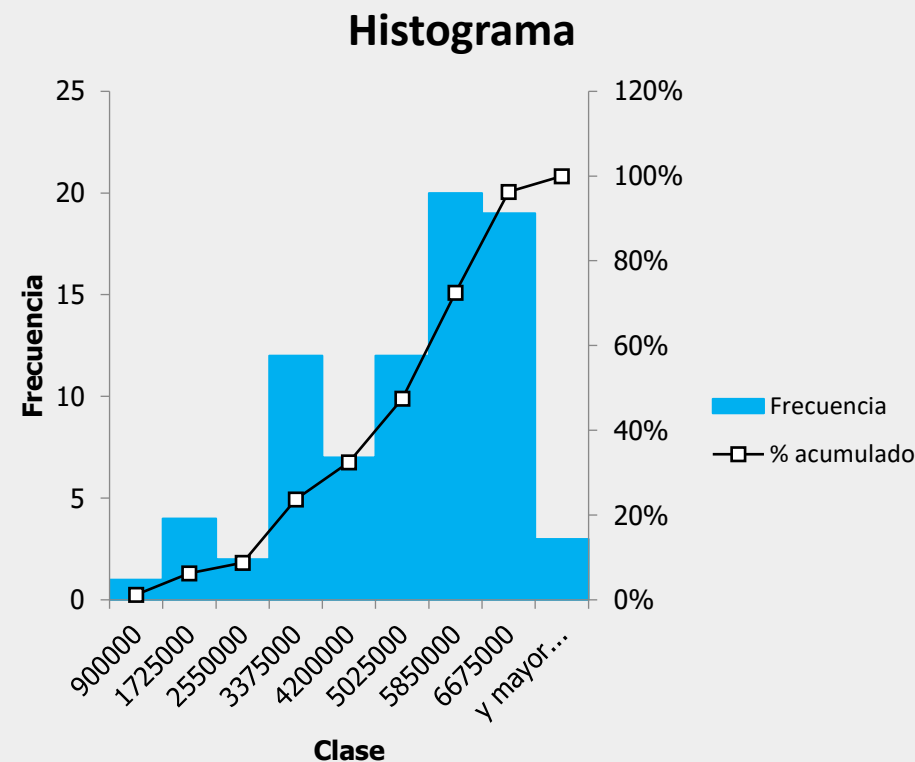
Estos intervalos deben tener la misma amplitud y todos los datos deben pertenecer a un intervalo y no estar en dos intervalos.

Se puede usar cualquier cantidad de clases o se puede utilizar las siguientes ecuaciones:

Cantidad de intervalos = $\sqrt{\text{cantidad de datos}}$ Aproximar al entero más cercano.

Amplitud de cada intervalo = $\frac{\text{valor máximo} - \text{valor mínimo}}{\text{Cantidad de intervalos}}$

Con esto se puede hacer tablas de frecuencias y la representación gráfica se llama histograma.



Probabilidades

Probabilidades individuales:

Es la probabilidad de ocurrencia de un valor en específico.

¿Cuál es la probabilidad de ocurrencia de obtener el valor de x ?

$$P(x) = \frac{\text{cantidad de valores iguales a } x}{\text{Total de datos}}$$

En Excel:

$$P(x) = \frac{\text{CONTAR.SI}(\text{datos}; "=x")}{\text{CONTAR}(\text{datos})}$$

Probabilidades acumuladas:

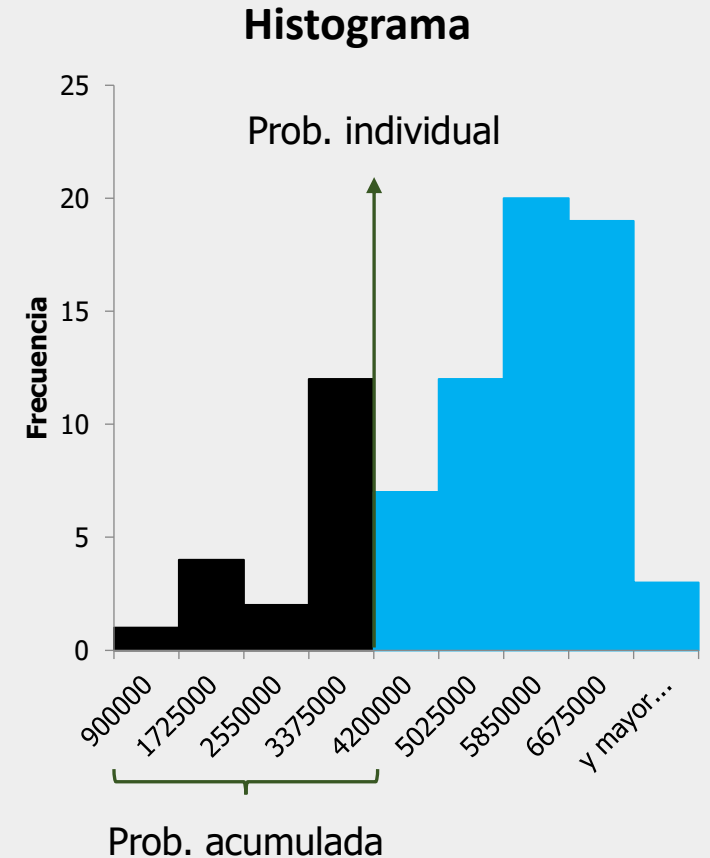
Es la probabilidad acumulada de obtener resultados iguales o menores que un valor en específico.

¿Cuál es la probabilidad acumulada x ?

$$P(\leq x) = \frac{\text{cantidad de valores iguales o menores a } x}{\text{Total de datos}}$$

En Excel:

$$P(x) = \frac{\text{CONTAR.SI}(\text{datos}; "<=x")}{\text{CONTAR}(\text{datos})}$$



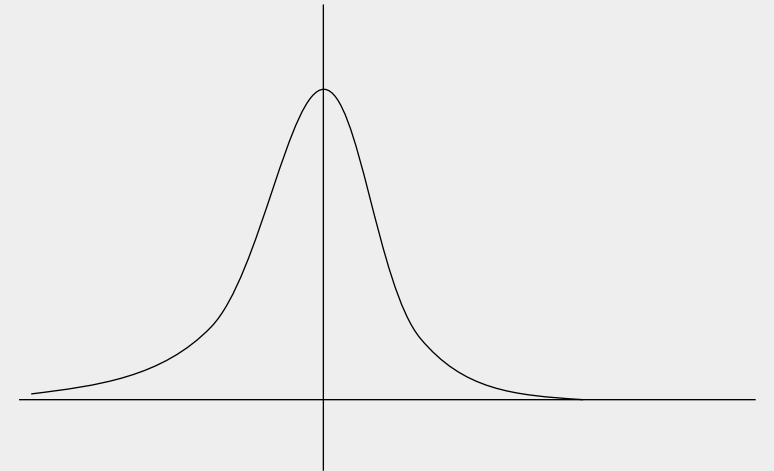
Estadística descriptiva

Medidas de tendencia o posición:

- Media aritmética.
- Media ponderada.
- Mediana.
- Moda.

Medidas de dispersión:

- Rango.
- Percentiles.
- Varianza.
- Desviación estándar.
- Coeficiente de variación.
- Coeficiente de asimetría o sesgo.
- Curtosis.



Medidas de tendencia o posición

Media aritmética: promedio simple.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

En Excel: =PROMEDIO(datos)

Media ponderada: promedio ponderado.

$$\bar{x} = \frac{\sum_{i=1}^n x_i \times w_i}{\sum_{i=1}^n w_i}$$

En Excel: =SUMAPRODUCTO(datos; pesos)

Mediana: divide el conjunto de datos en dos partes iguales. Misma cantidad de observaciones hacia la derecha y hacia la izquierda.

En Excel: =MEDIANA(datos)

Moda: observación que aparece con más frecuencia.

En Excel: =MODA(datos)

x_i : valor de la i -ésima observación.

n : número de observaciones.

w_i : peso o proporción de la i -ésima observación.

\bar{x} : media.

\tilde{x} : mediana.

Medidas de dispersión

Mide el grado de concentración de los datos en la distribución.

Rango: Este intervalo calcula la distancia entre los valores mínimo y máximo. Extremos de los datos.

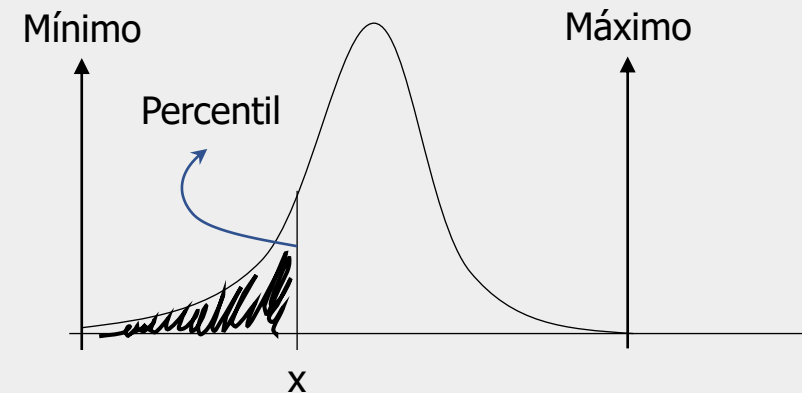
Rango = máximo valor – mínimo valor

En Excel: = MAX(datos) – MIN(datos)

Problemas cuando existen valores atípicos.

Percentiles:

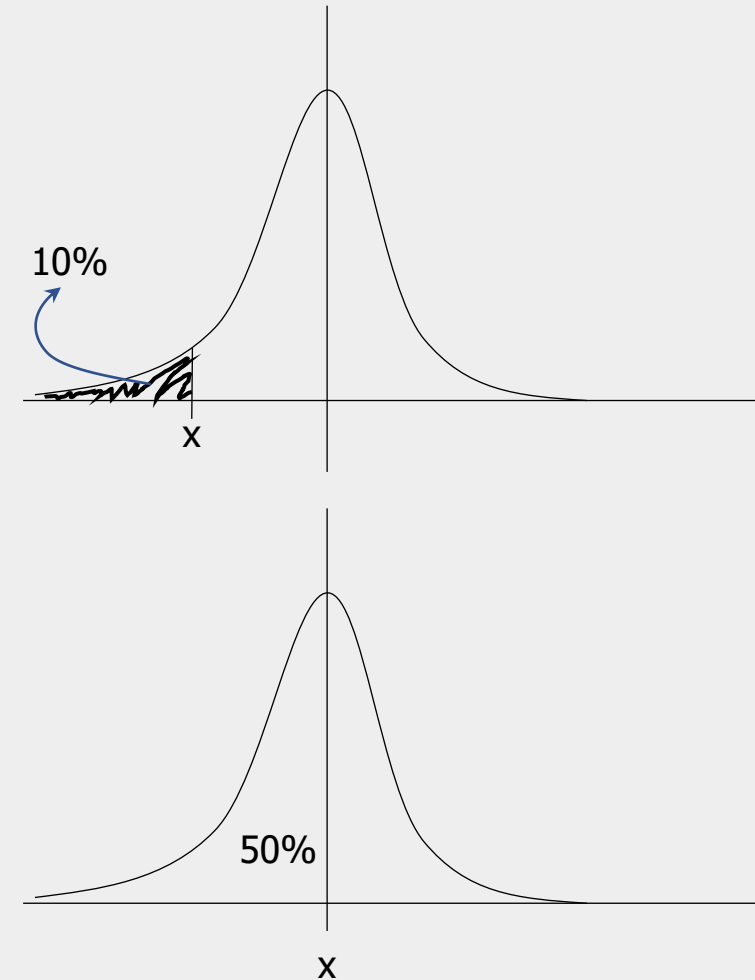
Se construyen con las frecuencias relativas acumuladas de los datos ordenados de forma ascendente y que tomarán valores entre 0% y 100%.



Medidas de dispersión

Percentiles: son las probabilidades acumuladas.

- Percentil del 0% (es el mínimo valor): valor mínimo.
- Percentil del 100% (es el máximo valor): acumula todos los valores.
- Percentil del 10% (primer decil): valor que acumula el 10% de los valores.
- Percentil del 20% (segundo decil): valor que acumula el 20% de los valores.
- Percentil del 27,5%: valor que acumula el 27,5% de los valores.
- Percentil del 50%: es la mediana. Acumula la mitad de los valores.
- Cuartiles: divide los datos en cuatro partes iguales.
 - Q1 (primer cuartil): 25%.
 - Q2 (segundo cuartil): 50% (mediana).
 - Q3 (tercer cuartil): 75%.



Medidas de dispersión

Percentiles:

En Excel:

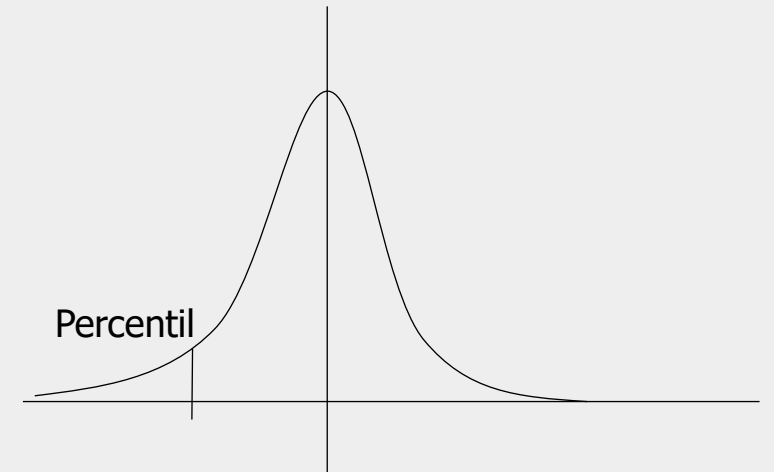
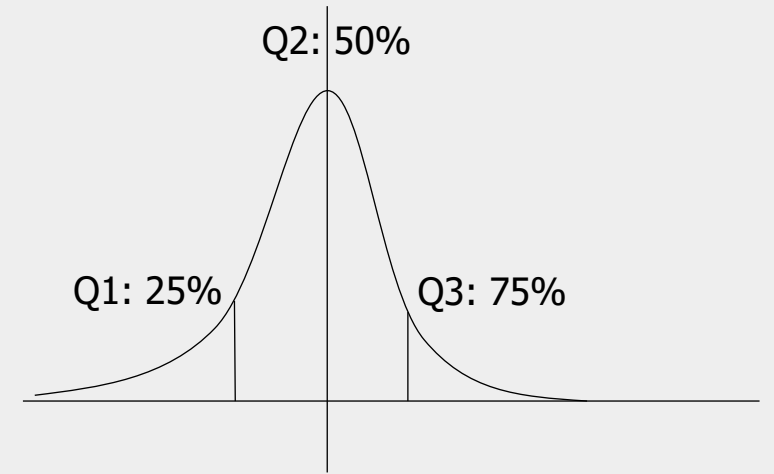
=PERCENTIL(datos, percentil)

percentil: valores entre 0 y 1. Es el porcentaje.

=CUARTIL(datos, cuartil)

cuartil: valores de 0, 1, 2, 3 o 4.

Con la función de Excel =PERCENTIL() también se pueden calcular los cuartiles.



Medidas de dispersión

Varianza: mide la variabilidad de los datos con respecto al promedio.

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

En Excel: =VAR(datos)

Desviación estándar: es la raíz cuadrada de la varianza.

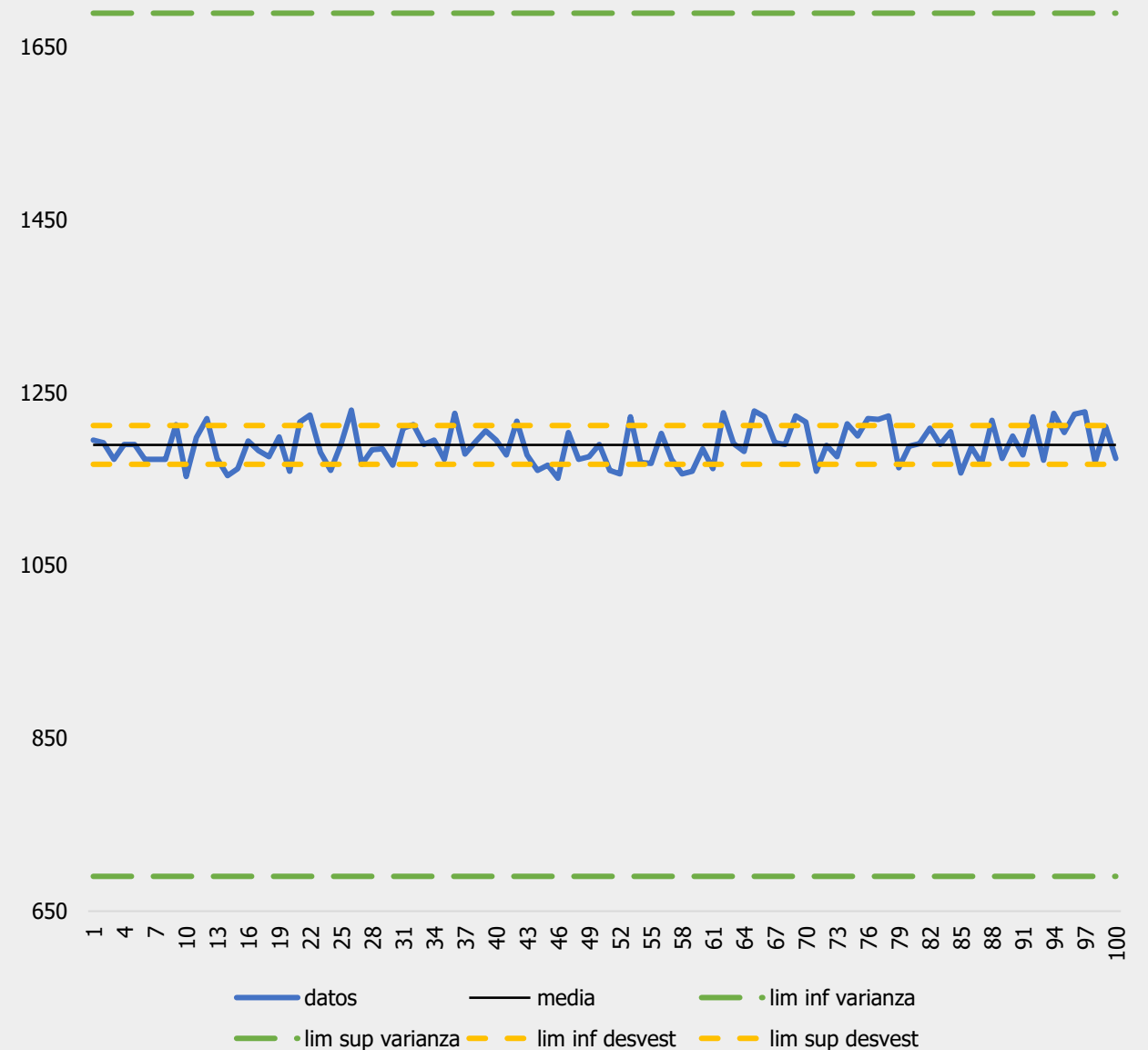
$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

En Excel: =DESVEST(datos)

σ^2 : varianza muestral.

σ : Desviación estándar.

Estas dos medias son en términos absolutos.



Medidas de dispersión

Coeficiente de variación:

Medida de dispersión en términos relativos.

Útil para comparar distribuciones.

El resultado está dado entre el 0% y el 100%.

$$CV = \frac{\sigma}{\bar{x}}$$

Coeficiente de asimetría o sesgo (skewness):

Mide el grado de asimetría de una distribución con respecto a su media.

$$Sk = \frac{n}{(n-1) \times (n-2)} \times \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$

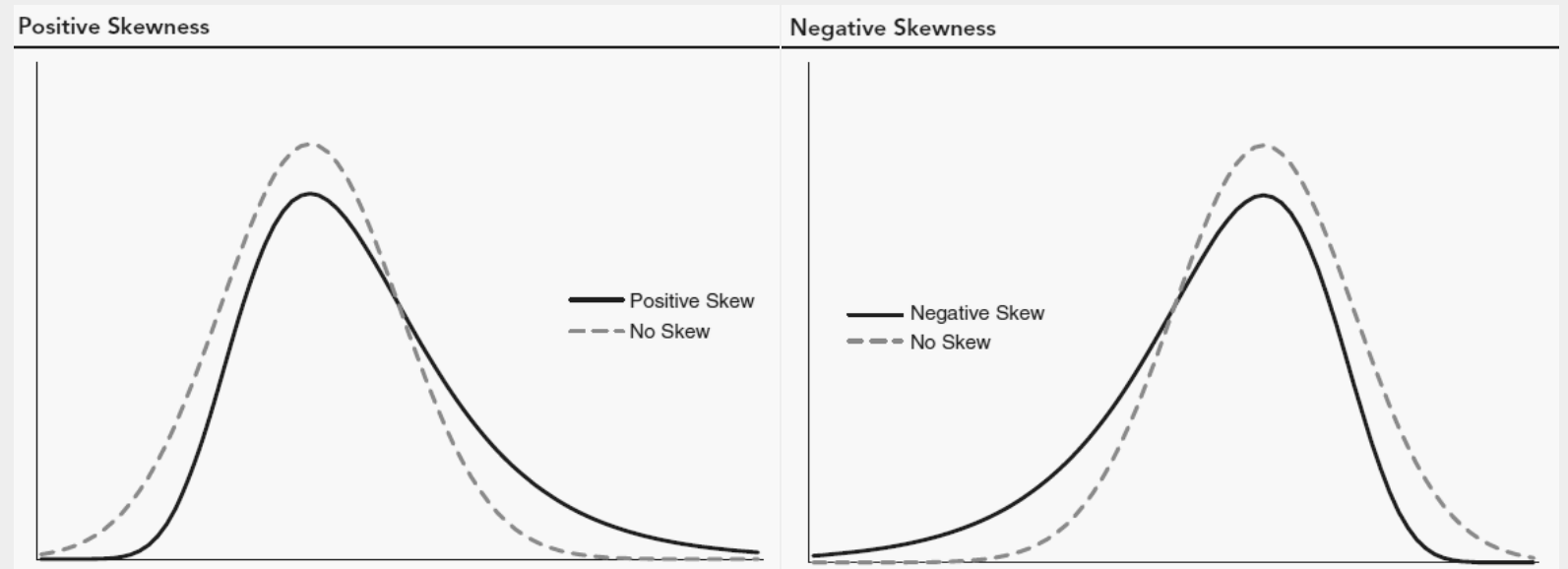
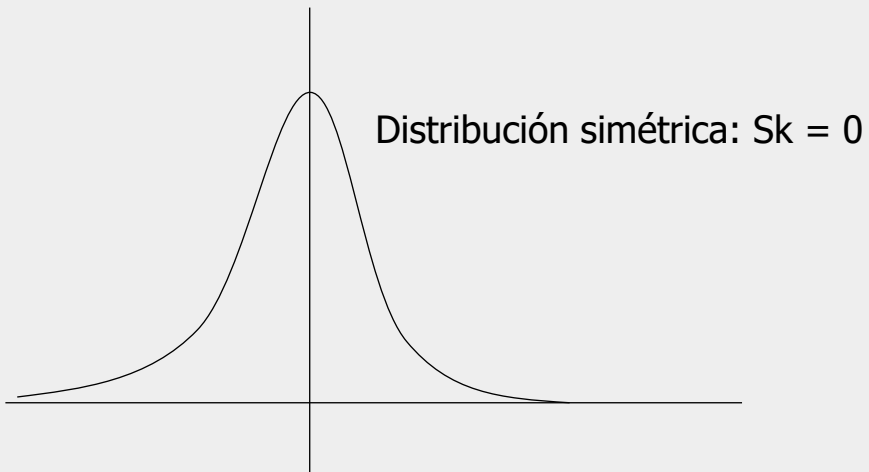
En Excel: =COEFICIENTE.ASIMETRIA(datos)

Medidas de dispersión

$Sk = 0$: Distribución simétrica.

$Sk > 0$: Los datos descienden más lentamente por la derecha. Asimetría positiva o hacia la derecha.

$Sk < 0$: Los datos descienden más lentamente por la izquierda. Asimetría negativa o hacia la izquierda.



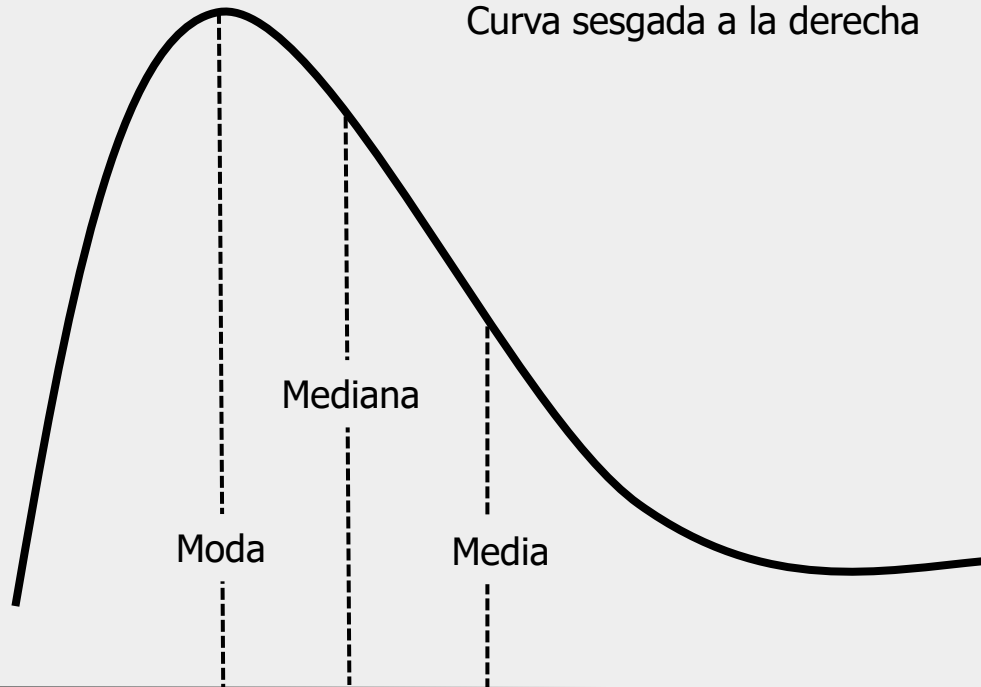
Medidas de dispersión

$S_k = 0$: moda = media = mediana

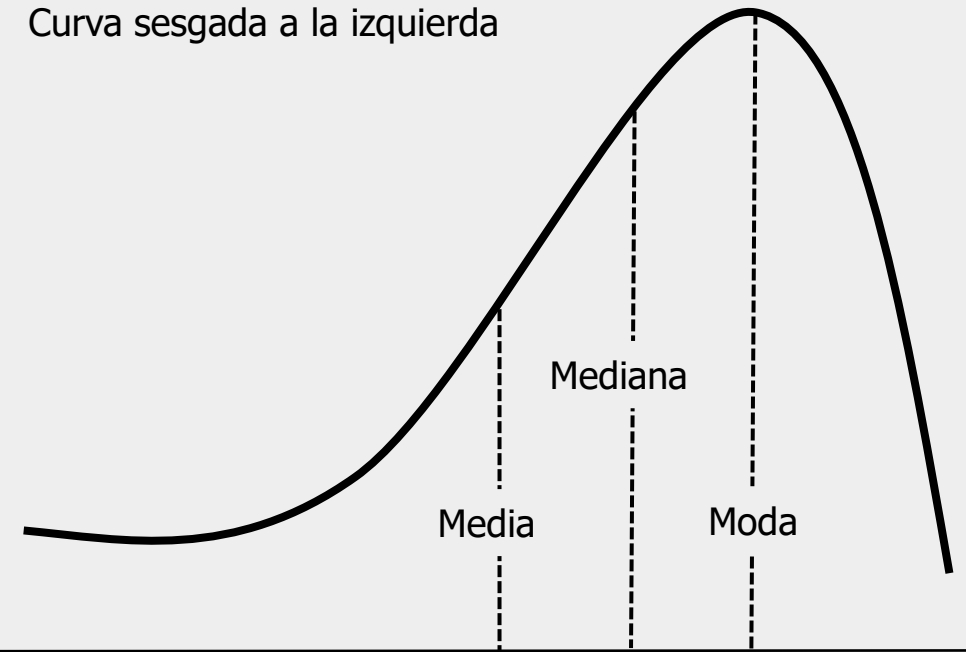
$S_k > 0$: moda > media y mediana

$S_k < 0$: moda < media y mediana

Curva sesgada a la derecha



Curva sesgada a la izquierda



Medidas de dispersión

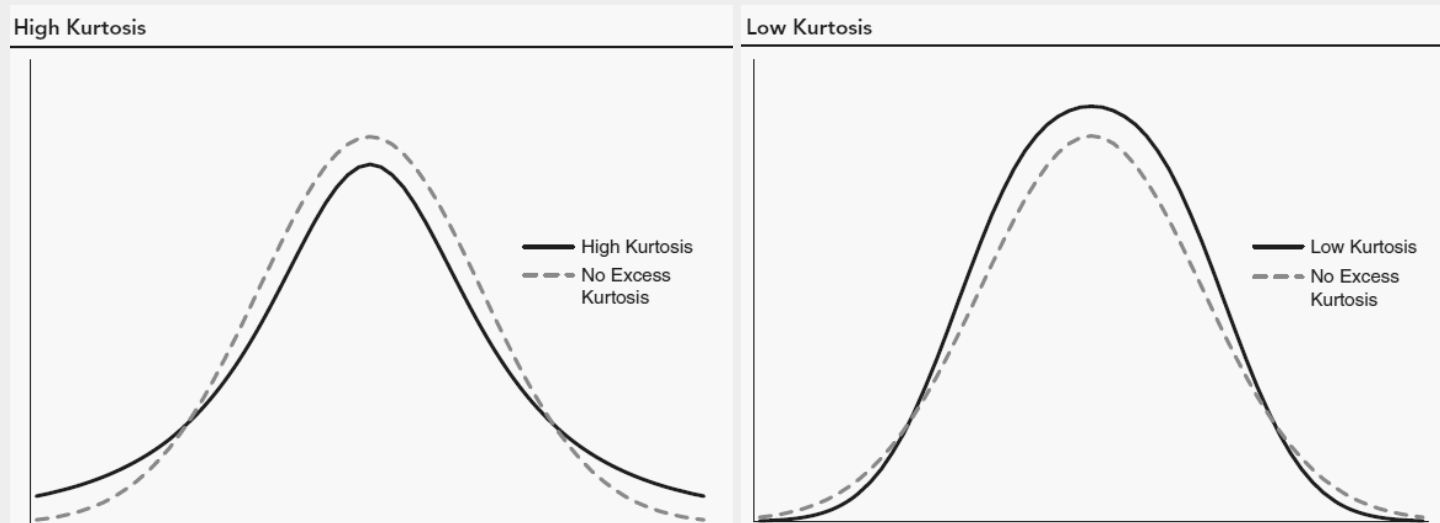
Curtosis (Kurtosis):

Mide el grado de concentración de los datos.

Curvas elevadas o curvas aplanadas.

$$K = \frac{n \times (n + 1)}{(n - 1) \times (n - 2) \times (n - 3)} \times \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 - \frac{3 \times (n - 1)^3}{(n - 2) \times (n - 3)}$$

En Excel: =CURTORIS(datos)



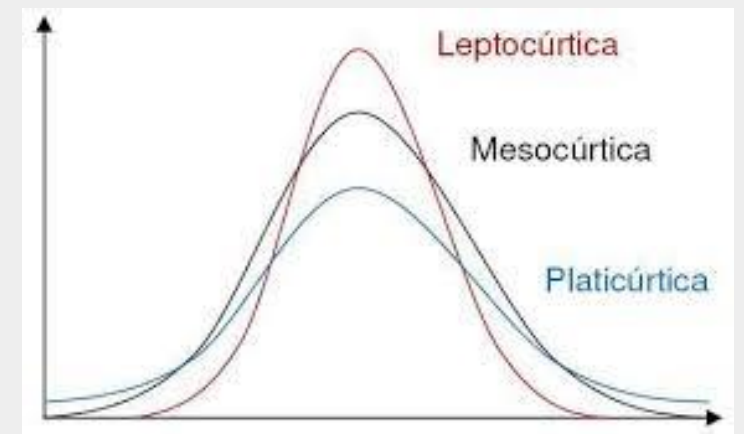
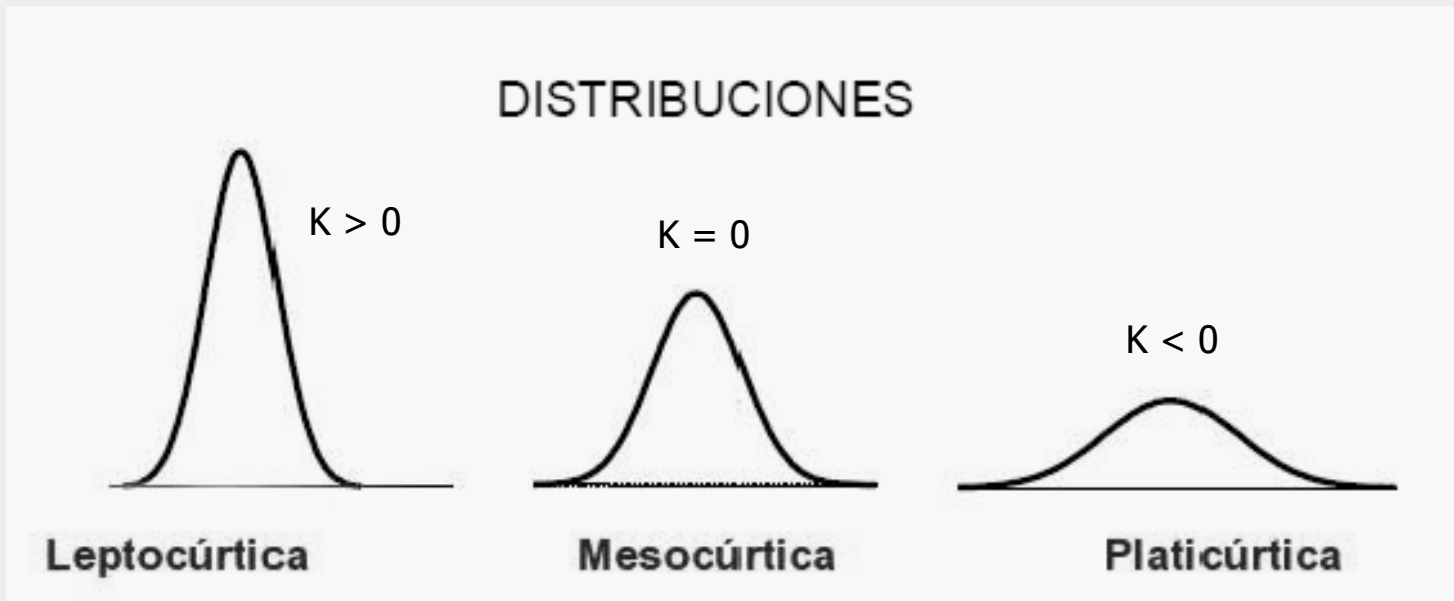
Medidas de dispersión

Curtosis (Kurtosis):

$K = 0$. Distribución mesocúrtica. Ej: distribución Normal.

$K > 0$: distribución leptocúrtica.

$K < 0$: distribución platicúrtica.



Resumen

Medida de tendencia	Fórmula	Excel
Media aritmética	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$	=PROMEDIO(datos)
Media ponderada	$\bar{x} = \frac{\sum_{i=1}^n x_i \times w_i}{\sum_{i=1}^n w_i}$	=SUMAPRODUCTO(datos; pesos)
Mediana		=MEDIANA(datos)
Moda		=MODA(datos)
Medida de dispersión		
Rango	Rango = máximo valor – mínimo valor	= MAX(datos) – MIN(datos)
Percentiles		=PERCENTIL(datos, percentil)
		=CUARTIL(datos, cuartil)
Varianza	$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$	=VAR(datos)
Desviación estándar	$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$	=DESVEST(datos)
Coeficiente de variación	$CV = \frac{\sigma}{\bar{x}}$	
Sesgo (skewness)	$Sk = \frac{n}{(n - 1) \times (n - 2)} \times \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma}\right)^3$	=COEFICIENTE.ASIMETRIA(datos)
Curtosis	$K = \frac{n \times (n + 1)}{(n - 1) \times (n - 2) \times (n - 3)} \times \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma}\right)^4 - \frac{3 \times (n - 1)^3}{(n - 2) \times (n - 3)}$	=CURTORIS(datos)