

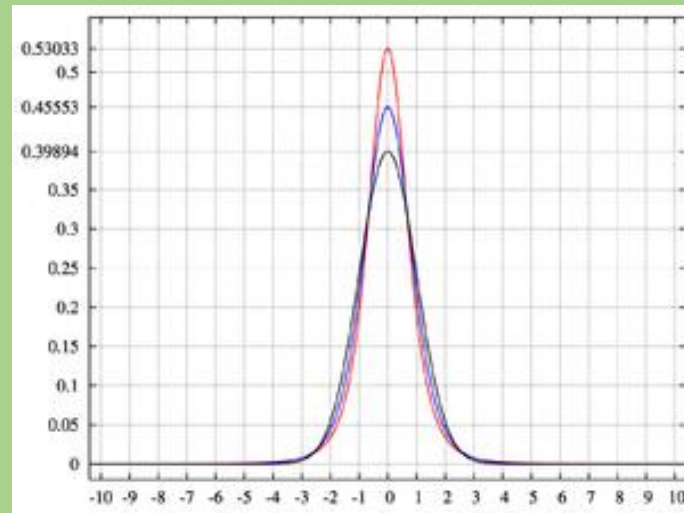
Capítulo 3

Caracterización y resumen de los datos

Continuar

Parámetro y estadístico

Un parámetro es la medida numérica que se calcula a partir de los datos observados en una población. Un estadístico es la medida numérica que se calcula a partir de los datos observados en una muestra.



Media ponderada

La media ponderada es una medida de interés, inicialmente el cálculo de la media considera que cada valor de la variable es de igual importancia. Sin embargo, en diferentes situaciones es de interés darle mayor peso a algunos de los valores observados.

$$MP = \frac{p_1 X_1 + p_2 X_2 + \dots + p_N X_N}{p_1 + p_2 + \dots + p_N}$$

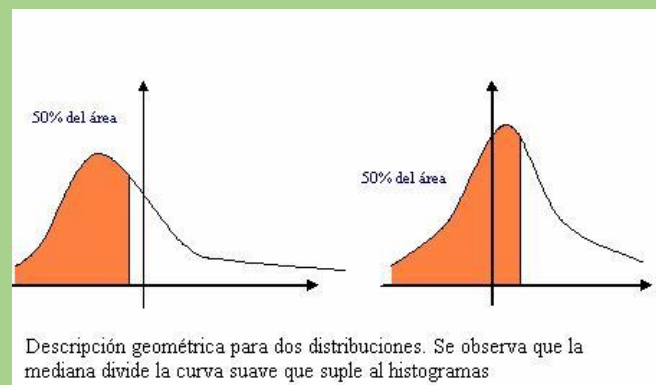
siendo (X_1, X_2, \dots, X_N) el conjunto de datos y (p_1, p_2, \dots, p_N) los pesos

Mediana

En general no se da una expresión para la mediana de una población, aquí la consideramos con el fin de tener presente la idea de la población y muestra.

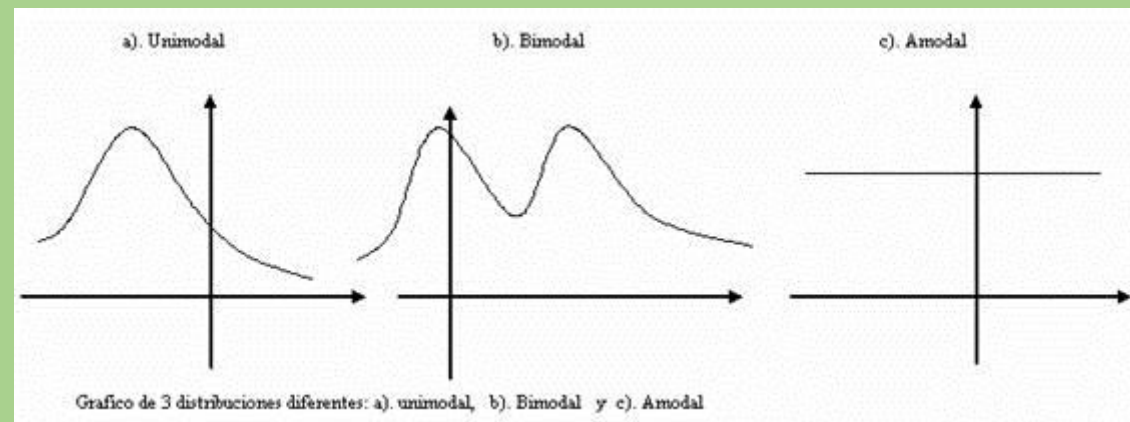
Sus características son:

- Tiene uso frecuente.
- Siempre existe y no toma en cuenta cada valor.
- No se ve afectada por los valores extremos, y es apropiada en estos casos.



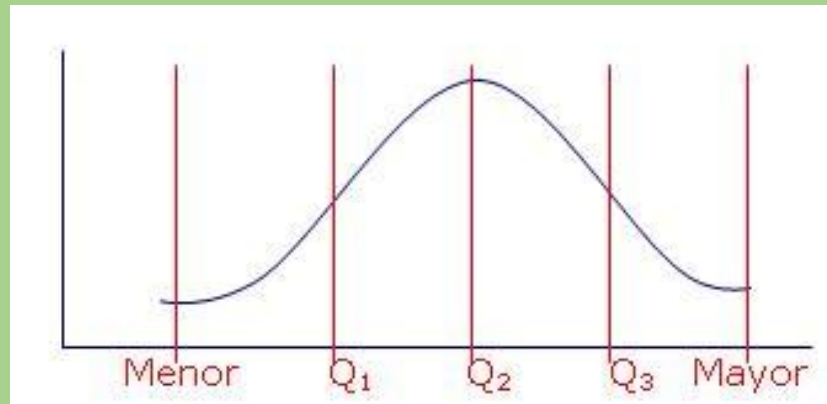
Moda

La moda es otra medida de tendencia central y ésta se define como el valor de la observación que más veces se repite en la muestra (o en la población); es decir, el de mayor frecuencia.



Cuartiles

Los cuartiles se conocen como medidas de posición relativa, dan información sobre la posición de una observación en la muestra. Éstos se consideran medidas adicionales a la media y mediana. Es un indicador que permite completar un resumen estadístico de los datos.



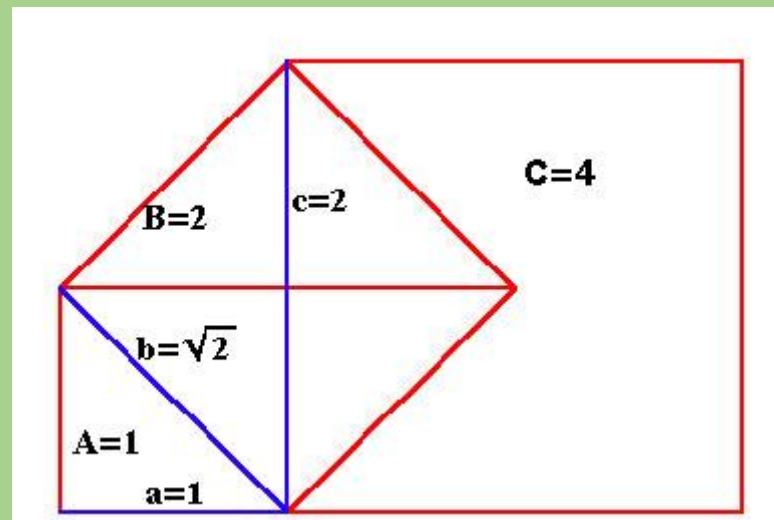
Media armónica

La media armónica de un conjunto de datos x_1, x_2, \dots, x_n es el recíproco de la media aritmética del recíproco de esos datos. Se emplea cuando se desea promediar velocidades, tiempo, rendimiento, etc., Es decir, cuando influyen los valores pequeños.

$$H = \frac{3}{\frac{1}{3} + \frac{1}{2} + \frac{1}{6}} = 3$$

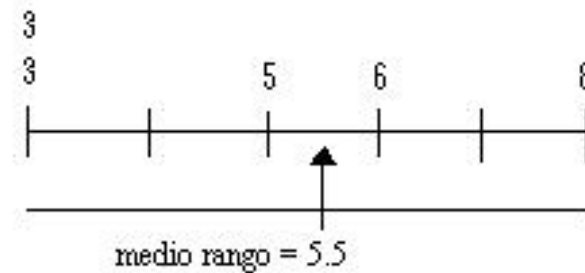
Media geométrica

Si algunos valores son muy grandes en magnitud y otros son pequeños, entonces la media geométrica es una medida que representa los datos mejor que la media.



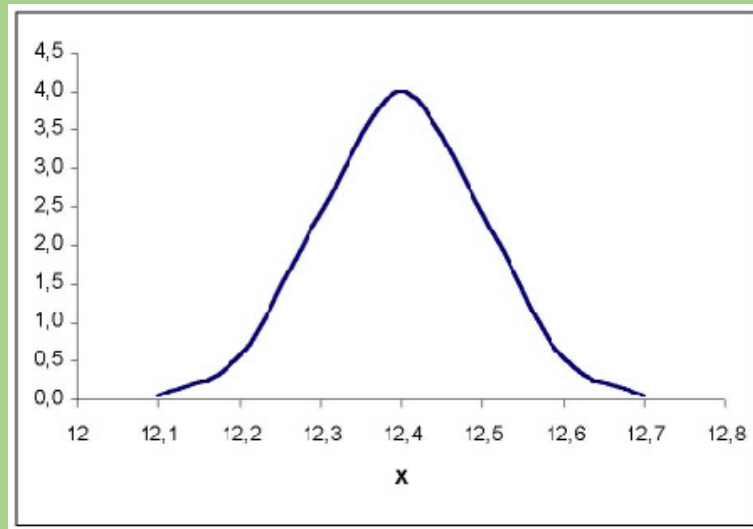
Rango

El rango es una de las medidas más sencillas para expresar la dispersión de los datos. Tan solo se requiere considerar el valor máximo y mínimo de las observaciones.



Rango intercuartil

El rango intercuartil es otra medida de dispersión. A menudo, en los datos aparecen observaciones con valores muy pequeños o demasiado grandes. En el contexto estadístico a estos valores se les llaman datos anómalos, y el rango es una medida sensible a este tipo de datos.



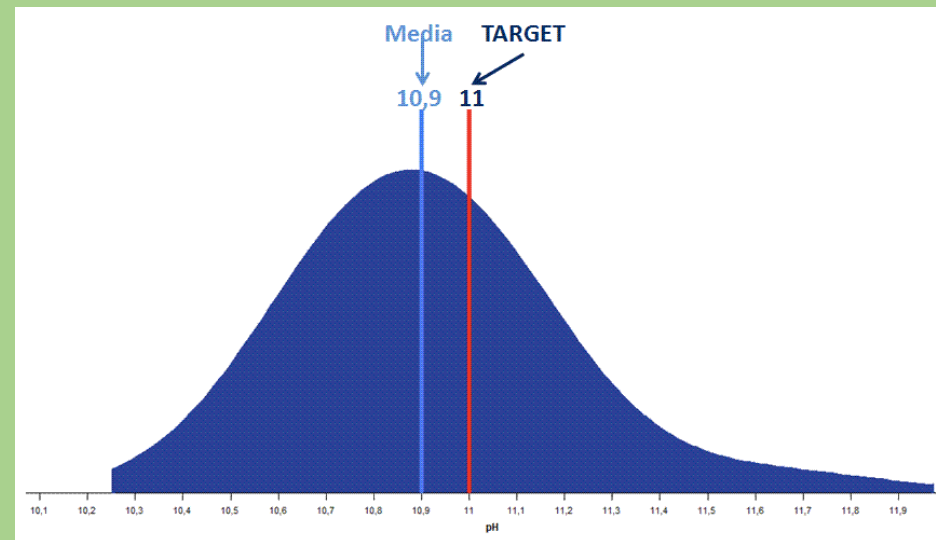
Desviación media muestral

La desviación media muestral es la media del valor absoluto de las desviaciones de cada observación con respecto a la media muestral.

$$MD = \frac{\sum X - X}{n}$$

Medidas estadísticas de datos agrupados

Las fórmulas y procedimientos indicados para calcular tanto las medidas de tendencia central como las de dispersión se aplican a cualquier conjunto de datos donde la variable sea cuantitativa. Si el número de datos es grande, el esfuerzo se presenta en la captura de éstos.



Coeficiente de variación

Otra medida apropiada para medir la variación de un conjunto de datos es el llamado Coeficiente de variación. Esta medida es útil para comparar las variabilidades de dos conjuntos de datos cuando existe una clara diferencia en la magnitud tanto en la media como en la desviación estándar, es independiente de las unidades de medición.

$$C.V. = \frac{S}{\bar{X}} * 100$$

Diagrama de caja

El diagrama de caja es una técnica de graficación relativamente moderna y sencilla, que permite apreciar las características principales de los datos y con ello tener una idea aproximada de la distribución. Asimismo, nos permite en una sola gráfica comparar la distribución de varios conjuntos de datos, ello con el objetivo de determinar las diferencias principales de los datos.

