



Análisis e Interpretación de Datos

Dra. Mariana-Edith Miranda-Varela

30 junio 2025

Índice

1. Estructuras de datos en R
2. Medidas de tendencia central
3. Medidas de dispersión
4. Medidas de posición y forma
5. Práctica

Medidas de tendencia central

Medias

Estas medidas indican la distribución de los datos a partir del medio o mitad del conjunto.

- Media aritmética

$$\bar{x} = \frac{\sum x_i}{n}$$

- Es la medida más usada
- Es sensible a los valores atípicos

- Media aritmética para datos con frecuencia

$$\bar{x} = \frac{\sum n_i x_i}{\sum n_i}$$

- Media ponderada

$$\bar{x} = \frac{\sum w_i x_i}{\sum w_i}$$

- Se emplea cuando unos datos tienen más importancia que otros

Medias

- Media armónica
 - Promedia variaciones con respecto a la misma variable
- Media geométrica
 - Promedio que se usa para calcular el rendimiento medio de porcentajes o tasas de crecimiento
 - Se usa cuando los datos cambian de forma multiplicativa y no aditiva
- Media cuadrática
 - Raíz de la media cuadrática
 - Se emplea para dar importancia a valores pequeños de una variable

Mediana (Me)

- Es más robusta que la media cuando se tienen valores *outliers*
- Información más fiable en grandes cantidades de datos
- Observación central del conjunto de datos
 - Número impar – valor central
 - Número par – promedio de las observaciones centrales
- Distribución de datos se encuentra en intervalos

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} a_i$$

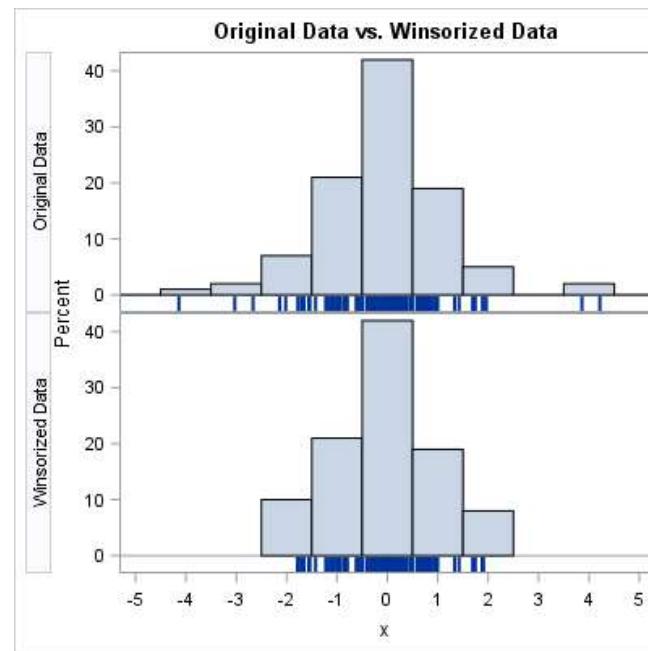
Moda y otras medias

- **Moda**
 - Es el valor más frecuente de los datos
 - Si se tienen 2 datos con la misma frecuencia es **bimodal**
- **Media recortada**
 - Calcula la media aritmética a un subconjunto central del conjunto de datos
 - Media recortada al y%
 - Media recortada al 0% = media aritmética
 - Media recortada al 25% se conoce como "**centrimedia**"
- **Media winsorizada**
 - La diferencia con la media recortada consiste en que se sustituyen el menor y mayor valor (tras el proceso de eliminación) en las posiciones eliminadas.

Otras medias

Media recortada

~~1 5 32 35 38 43 45 49 52 56 80 98~~



FUENTE: <https://blogs.sas.com/content/iml/2017/02/08/winsorization-good-bad-and-ugly.html>

Medidas de dispersión

Estas medidas indican cuánto se desvían los datos (distribución)

- Rango

$$rango = x_{max} - x_{min}$$

- Varianza

$$\text{Varianza} = s^2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

- Desviación estándar (o típica)

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}}$$

- Promedio de las desviaciones de datos con respecto a su media
- Varianza poblacional σ^2

- Depende de la escala de los datos, por lo que no se puede usar para comparar datos

Medidas de posición y forma

Sirven para saber si un valor esta alejado de su media

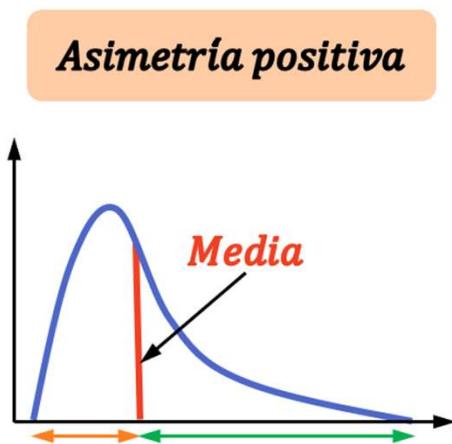
- Cuartil
 - Dividir en cuartos iguales los datos
 - Q1 – 25% de los datos son inferiores a él
 - Q2 – 50% de las observaciones, coincide con la mediana
 - Q3 – 75% de las observaciones a la izquierda
 - Percentiles – Dejan un % de datos a su izquierda

$$Q_k = L_{i-1} + \frac{\frac{kN}{4} - N_{i-1}}{n_i} a_i$$

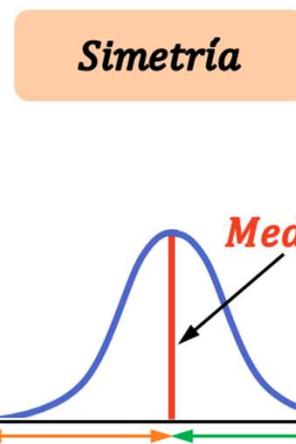
Distribución de datos

- Engloban la simetría y apuntamiento
- Distribución

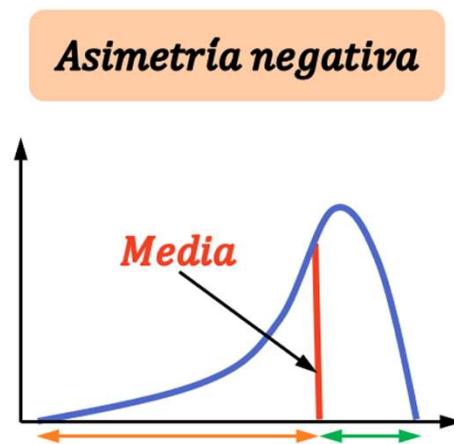
Indican sesgos



Media > moda



Mediana=Moda=Media

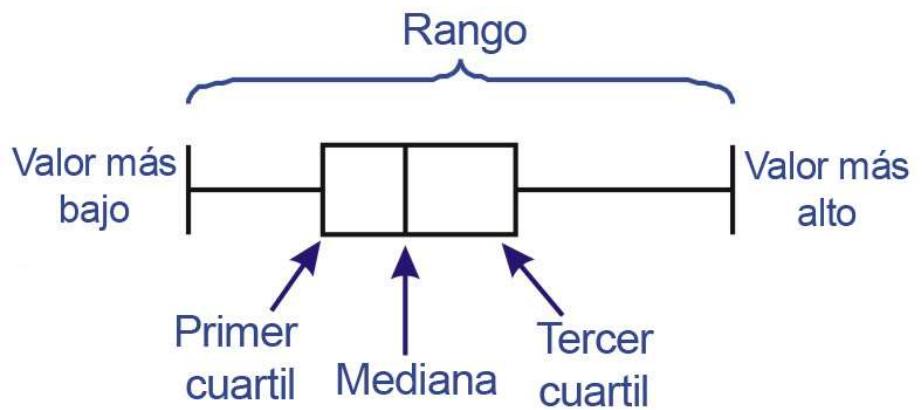


Moda > media

FUENTE: <https://www.probabilidadyestadistica.net/tipos-de-asimetria/>

Gráficos de caja y bigotes

- Resumen la información de los datos
- Medidas que se representan
 - Mínimo
 - Q1 (límite de la caja)
 - Mediana (divide la caja en 2)
 - Q3 (límite de la caja)
 - Máximo
- Rango intercuartílico



FUENTE: <https://flexbooks.ck12.org/cbook/ck-12-%C3%A1lgebra-i-en-espa%C3%B1ol/section/11.8/primary/section/diagramas-de-caja-y-bigotes/>



Práctica



| Próxima sesión

Semana 5
7 al 11 de julio

Muchas gracias.