

# **Laboratorio de Datos**

## **Estadística descriptiva**

Primer Cuatrimestre 2024  
Turnos tarde y noche

Facultad de Ciencias Exactas y Naturales, UBA

# Variables Categóricas vs Numéricas

## Variables Categóricas

Son variables que representan diferentes categorías o grupos. Las variables categóricas se pueden dividir en dos tipos: nominal y ordinal.

- **Nominal:** No tienen un orden intrínseco.
- **Ordinal:** Tienen un orden intrínseco.

- **Variable Categórica:**
  - Nominal: Color de los ojos
    - Azul
    - Marrón
    - Verde
    - Gris
  - Ordinal: Nivel de educación
    - Primaria
    - Secundaria
    - Universitaria
    - Posgrado

# Variables Categóricas vs Numéricas

## Variables Numéricas

Son variables que representan cantidades numéricas. Las variables numéricas se pueden dividir en dos tipos: discreta y continua.

- **Discreta:** Representa valores contables y no puede tomar valores intermedios.
- **Continua:** Representa medidas y puede tomar cualquier valor dentro de un rango.

- **Variable Numérica:**

- Discreta: Número de hijos

- 0
- 1
- 2
- 3

- Continua: Altura (en metros)

- 1.60
- 1.75
- 1.82
- 1.68

# Medidas de tendencia central: media y mediana

## Media

La media es el promedio de un conjunto de datos y se calcula sumando todos los valores y dividiéndolos por el número total de valores.

$$\text{Media} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

## Mediana

La mediana es el valor del medio cuando un conjunto de datos se ordena de menor a mayor. Si hay una cantidad par de datos, se toma el promedio entre los dos valores del medio.

# Ejemplos Numéricos

Conjunto de datos 1: 2, 6, 8, 10, 9

- Media:

$$\frac{2 + 6 + 8 + 10 + 9}{5} = \frac{35}{5} = 7$$

- Mediana: ordenamos de menor a mayor 2, 6, 8, 9, 10  
La mediana es 8.

Conjunto de datos 2: 2, 6, 8, 50, 9

- Media:

$$\frac{2 + 6 + 8 + 50 + 9}{5} = \frac{75}{5} = 15$$

- Mediana: ordenamos de menor a mayor 2, 6, 8, 9, 50  
La mediana es 8.

# Diferencia entre Media y Mediana: Outliers

Un outlier o valor atípico es una observación dentro de una muestra que no es consistente con el resto.

## Media

La media es sensible a los valores atípicos (outliers), ya que se calcula sumando todos los valores y dividiéndolos por el número total de valores.

- Si hay outliers, la media puede sesgarse significativamente hacia los valores extremos.

## Mediana

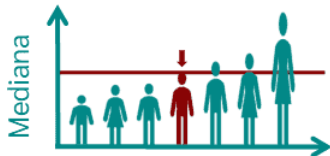
La mediana no se ve afectada por los valores atípicos, ya que es el valor central de un conjunto de datos ordenados.

- Es más robusta frente a valores extremos que la media.

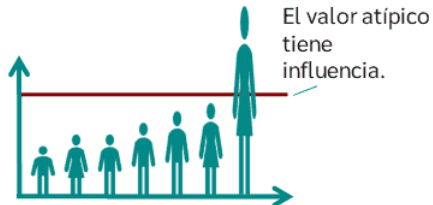
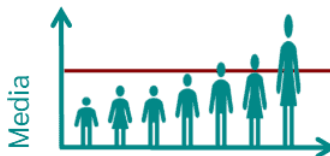
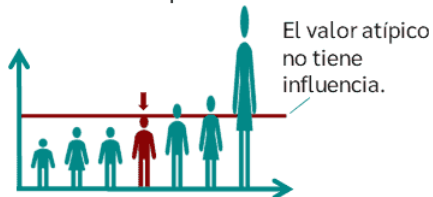


# Media y mediana

Sin valores atípicos



Con valores atípicos



# Varianza y desvío Estándar

## Definición

La varianza y el desvío estándar son *medidas de dispersión* que indican cuánto varían los valores de un conjunto de datos respecto a la media.

## Fórmula

La varianza se calcula mediante la fórmula:

$$\text{varianza} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n},$$

donde  $\mu$  es la media de los datos.

El desvío se calcula como la raíz cuadrada de la varianza:

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

# Interpretación

- Un desvío estándar pequeño indica que los datos están cercanos a la media.
- Un desvío estándar grande indica que los datos están dispersos y alejados de la media.

## Ejemplo

Consideramos el siguiente conjunto de datos: 1, 2, 3, 4, 5

- Media ( $\mu$ ): 3
- Varianza:  $\frac{2^2+1^2+0+1^2+2^2}{5} = 2$
- Desvío Estándar ( $\sigma$ ):  $\sqrt{2} = 1.41$