

# ESTADÍSTICA

Camerina Laura Ramírez Gallegos

# DEFINICIÓN

- La **Estadística** es una disciplina que utiliza recursos matemáticos para organizar y resumir una gran cantidad de datos obtenidos de la realidad, e inferir conclusiones respecto de ellos.



# CLASIFICACIÓN

- Hay una estadística **descriptiva** y una estadística **inferencial**.
- La primera se ocupa de describir la muestra, y la segunda infiere conclusiones a partir de los datos que describen la muestra (por ejemplo con respecto a la población).



# ASPECTOS A REPRESENTAR

## Medidas descriptivas

- **Posición**
  - Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.
    - Cuantiles, percentiles, cuartiles, deciles,...
- **Centralización**
  - Indican valores con respecto a los que los datos parecen agruparse.
    - Media, mediana y moda
- **Dispersión**
  - Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización.
    - Varianza, desviación típica, coeficiente de variación, rango
- **Forma**
  - Asimetría
  - Apuntamiento o curtosis



# MEDIDAS DE TENDENCIA CENTRAL

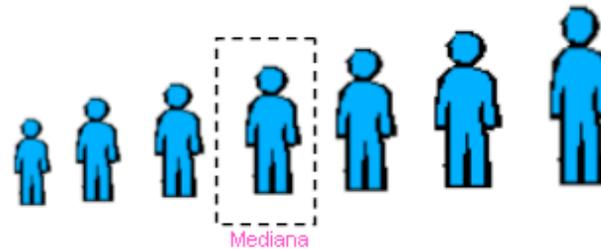
○ Entre las más comunes se encuentran:

○ Promedio o media

$$\frac{X_1 + X_2 + X_3 + X_4 + \dots + X_n}{n}$$

○ Mediana

○ Moda



Unimodal



Bimodal



Multimodal





## MEDIDAS DE UBICACIÓN

- La descripción de un conjunto de datos, incluye como un elemento de importancia la ubicación de éstos dentro de un contexto de valores posibles. Entre las más comunes se encuentran:
- **Percentiles** (son 99 valores que dividen en cien partes iguales el conjunto de datos ordenados. Ejemplo, el percentil de orden 15 deja por debajo al 15% de las observaciones, y por encima queda el 85% )
- **Cuartiles** (son los tres valores que dividen al conjunto de datos ordenados en cuatro partes iguales)
- **Deciles** (son los nueve valores que dividen al conjunto de datos ordenados en diez partes iguales)



# MEDIDAS DE DISPERSION

- Al analizar un conjunto de datos, resulta a menudo conveniente expresar numéricamente la variabilidad que existe entre ellos. .
- Entre las más comunes se encuentran:
  - Rango
  - Desviación media
  - Varianza



# MEDIDAS DE DISPERSION

- **Rango**, Es la diferencia entre el valor de las observaciones mayor y el menor ( $R = \mathbf{x}_{\max} - \mathbf{x}_{\min}$ )

- **Varianza**, es el promedio del cuadrado de las distancias entre cada observación y la media aritmética del conjunto de observaciones.

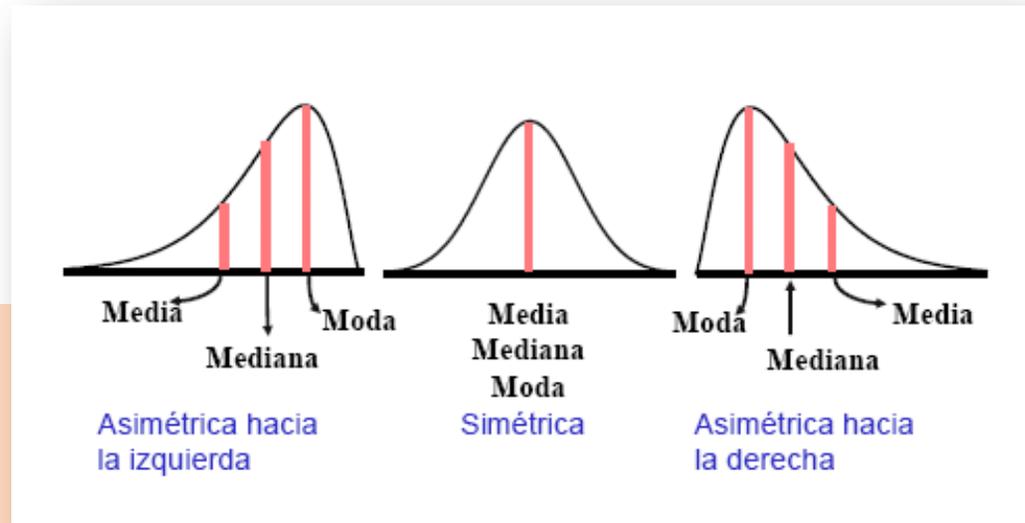
$$s^2 = \frac{\sum_i x_i^2 n_i}{n} - \bar{x}^2$$

- **Desviación media**, La varianza viene dada por las mismas unidades que la variable pero al cuadrado, para evitar este problema podemos usar como medida de dispersión la desviación típica que se define como la raíz cuadrada positiva de la varianza

$$s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2 n_i}{n - 1}}$$



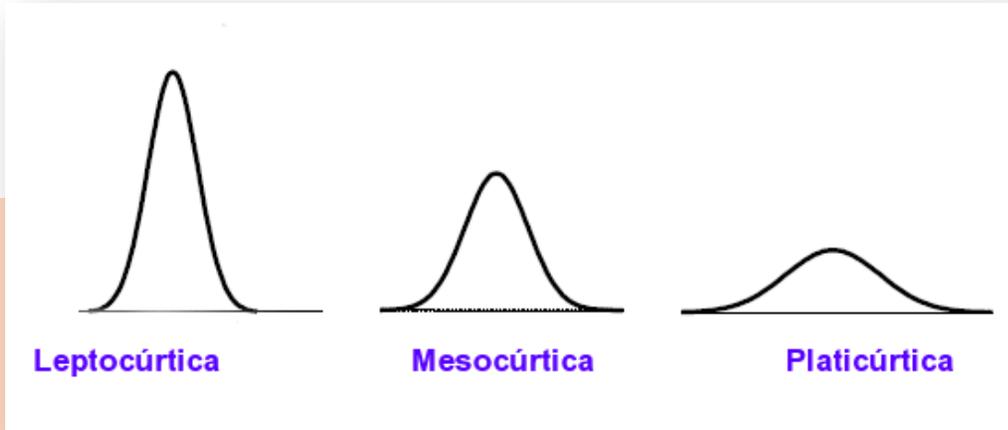
# MEDIDAS DE FORMA



- Diremos que una distribución es **simétrica** cuando su mediana, su moda y su media aritmética coinciden.
- Diremos que una distribución es **asimétrica a la derecha** si las frecuencias (absolutas o relativas) descienden más lentamente por la derecha que por la izquierda.
- Si las frecuencias descienden más lentamente por la izquierda que por la derecha diremos que la distribución es **asimétrica a la izquierda**.



## MEDIDAS DE FORMA



- **Curtosis** miden la mayor o menor cantidad de datos que se agrupan en torno a la moda. Se definen 3 tipos de distribuciones según su grado de curtosis:
- Distribución **mesocúrtica**: presenta un grado de concentración medio alrededor de los valores centrales de la variable (el mismo que presenta una distribución normal).
- Distribución **leptocúrtica**: presenta un elevado grado de concentración alrededor de los valores centrales de la variable.
- Distribución **platicúrtica**: presenta un reducido grado de concentración alrededor de los valores centrales de la variable.



# ASPECTOS A REPRESENTAR

## Medidas descriptivas

- **Posición**
  - Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos.
    - Cuantiles, percentiles, cuartiles, deciles,...
- **Centralización**
  - Indican valores con respecto a los que los datos parecen agruparse.
    - Media, mediana y moda
- **Dispersión**
  - Indican la mayor o menor concentración de los datos con respecto a las medidas de centralización.
    - Varianza, desviación típica, coeficiente de variación, rango
- **Forma**
  - Asimetría
  - Apuntamiento o curtosis



<b>Centralización</b>	{	<b>Media aritmética:</b>	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
		<b>Mediana:</b>	En el conjunto de datos ordenado, valor que ocupa la posición central.
		<b>Moda:</b>	Es el valor más frecuente.

### Poblaciones.

### Inferencia de poblaciones (muestra)

<b>Dispersión</b>	{	<b>Varianza:</b>	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
		<b>Desviación típica:</b>	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$
		<b>Coef. Variación:</b>	$CV = \frac{s}{\bar{x}}$	
		<b>Rango:</b>	diferencia entre el valor de las observaciones mayor y el menor	

Para las medidas de las poblaciones se suele utilizar letras griegas ( $\mu, \sigma$ )  
y para las de muestras letras latinas ( $\bar{x}, s$ )