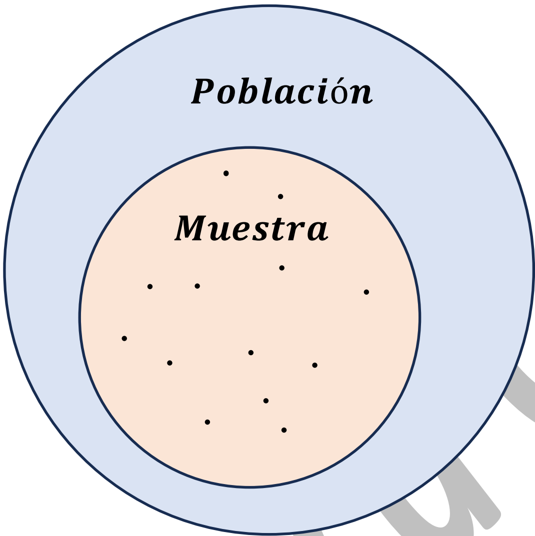
# 10. Estadística

## 10.1. Conceptos básicos

La Estadística es la rama de las matemáticas que desarrolla unas técnicas para el conocimiento numérico de un conjunto de datos.




- **Población**: El conjunto de todos los elementos de los que nos interesa hacer el estudio.
- **Muestra**: Es un subconjunto de la población. Se estudia la muestra y luego se infiere a toda la población.
- **Individuo**: Es cada uno de los elementos que forman la población. Indicado con puntos dentro del conjunto muestra.
- **Caracteres**: Son los aspectos que queremos estudiar en los individuos de una población.
- **Variable Estadística**: Es aquella que recorre todos lo valores de un cierto carácter. Y pueden ser:
    - **Cuantitativa**: Si tiene valores numéricos.
        - **Discreta**: Solo toma valores aislados.
        - **Continua**: Puede tener cualquier valor de un intervalo.
    - **Cualitativa**: Si no toma valores numéricos.

La **estadística descriptiva** expone y analiza algunos caracteres de los individuos de un grupo dado (población) sin extraer conclusiones para un grupo más grande.

La **estadística inferencial** trabaja con muestras y pretende, a partir de ellas, "inferir" características de toda la población.

## 10.2. Tabla de frecuencias

$$x_i \rightarrow variable\ estadística$$

$$f_i \rightarrow frecuencia\ absoluta$$

$$F_i \rightarrow frecuencia\ absoluta\ acumulada$$

$$f_{r_i} = \frac{f_i}{N} \rightarrow frecuencia\ relativa$$

$$F_{r_i} \rightarrow frecuencia\ relativa\ acumulada$$

$$\% = f_{r_i} \cdot 100 \rightarrow Porcentaje$$

$$o = \frac{\% \cdot 360º}{100} \rightarrow grados$$

$$N \rightarrow número\ total\ de\ datos$$

Ejemplo: variable estadística discreta

| $x_i$ | $f_i$ | $F_i$ | $f_{r_i}$ | $F_{r_i}$ | % | $o$ |
|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 0,15 | 0,15 | 15 | 54 |
| 2 | 2 | 5 | 0,10 | 0,25 | 10 | 36 |
| 3 | 1 | 6 | 0,05 | 0,30 | 5 | 18 |
| 4 | 2 | 8 | 0,10 | 0,40 | 10 | 36 |
| 5 | 4 | 12 | 0,20 | 0,60 | 20 | 72 |
| 6 | 5 | 17 | 0,25 | 0,85 | 25 | 90 |
| 7 | 3 | 20 | 0,15 | 1 | 15 | 54 |
| | $N = 20$ | | + 1 | | + 100 | + 360 |

Ejemplo: variable estadística continua

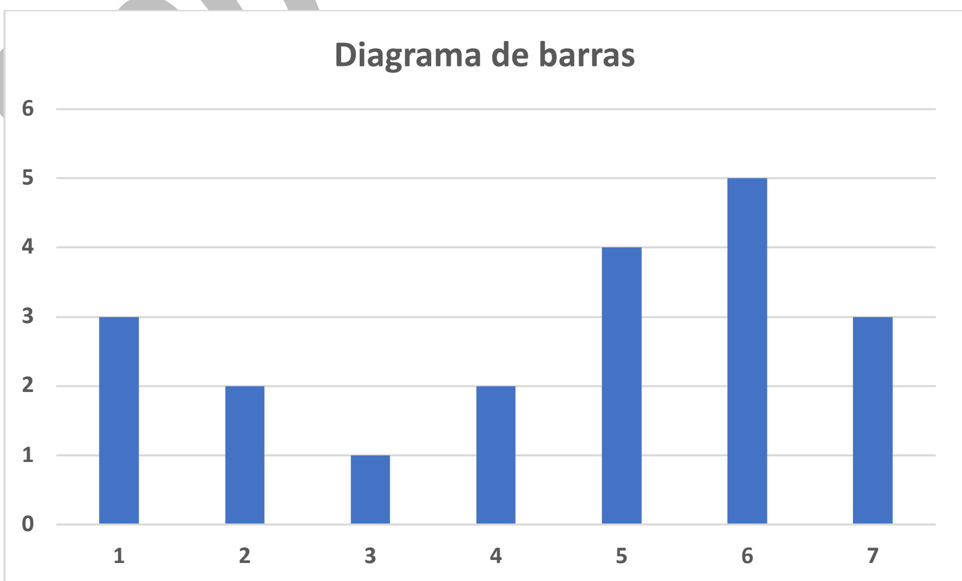
$$[a, b] \rightarrow Intervalo$$

$$x_i = \frac{a+b}{2} \rightarrow Marca\ de\ classe$$

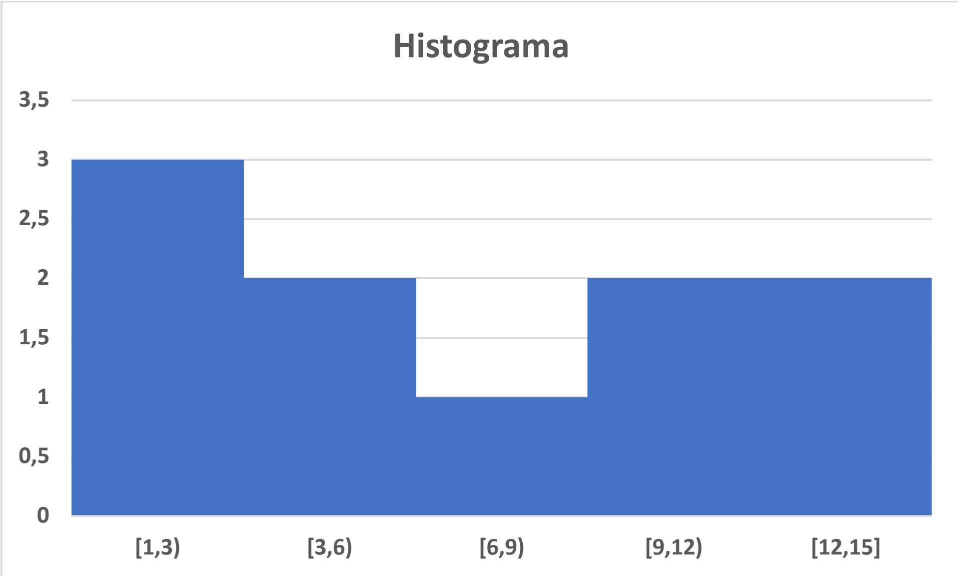
| $\boldsymbol{Intervalo}$ | $\boldsymbol{x_i}$ | $\boldsymbol{f_i}$ | $\boldsymbol{F_i}$ | $\boldsymbol{f_{r_i}}$ | $\boldsymbol{F_{r_i}}$ | $\%$ | $\boldsymbol{o}$ |
|---|---|---|---|---|---|---|---|
| $[1, 3[$ | 1,5 | 3 | 3 | 0,30 | 0,30 | 30 | 108 |
| $[3, 6[$ | 4,5 | 2 | 5 | 0,20 | 0,50 | 20 | 72 |
| $[6, 9[$ | 7,5 | 1 | 6 | 0,10 | 0,60 | 10 | 36 |
| $[9, 12[$ | 10,5 | 2 | 8 | 0,20 | 0,80 | 20 | 72 |
| $[12, 15]$ | 13,5 | 2 | 10 | 0,20 | 1 | 20 | 72 |
| | | $N = 10$ | | + 1 | | + 100 | + 360 |

## 10.3. Representación gráfica

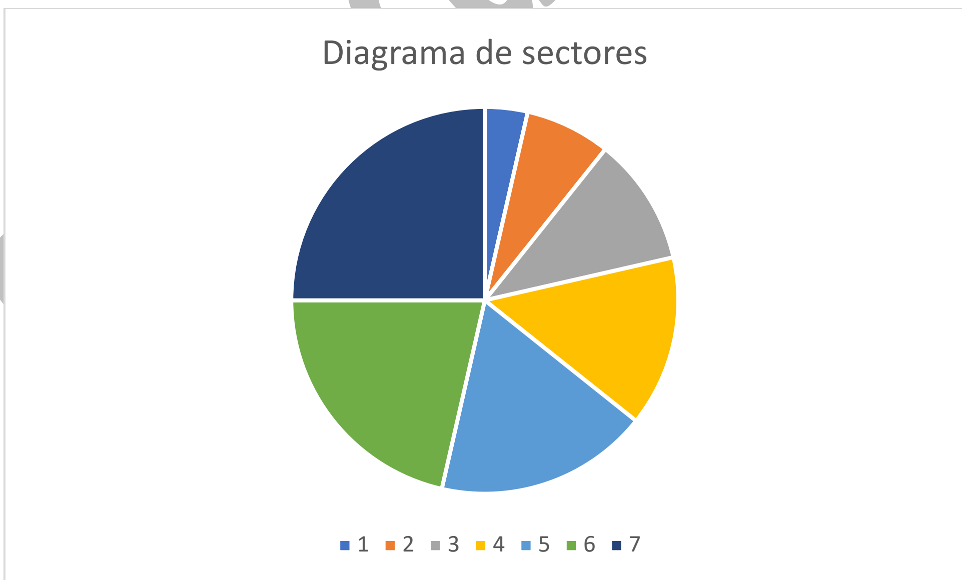
En el ejemplo de variable estadística discreta se representa mediante **diagrama de barras**, el eje X seria $x_i$ y el eje Y $f_i$.



En el ejemplo de variable estadística continua se representa mediante un **histograma**, el eje X seria $x_i$ y el eje Y los $intervalos$.




En los dos tipos de variables, tanto discreta como continua, se puede realizar un **diagrama de sectores**.



## 10.4. Medidas de centralización

Las medidas de centralización en un estudio estadístico estudian dónde se encuentra el valor central del estudio. Estudiaremos la moda, la mediana y la media aritmética.

### 10.4.1. Moda ($Mo$)

Es el valor que más se repite en el estudio, el que tiene la $f_i$ mayor.

En el ejemplo de la variable discreta: $Mo = 5$ ya que es el valor de $x_i$ que tiene el mayor valor de $f_i = 4$

### 10.4.2. Mediana ($Me$)

Es el valor central de los datos del estudio. Para llegar a él se divide el número total de datos entre 2 y se busca ese valor en la $F_i$, el que esté más próximo por encima, y entonces su $x_i$ correspondiente seria la mediana.

En el ejemplo de la variable discreta:

$$\frac{N}{2} = \frac{20}{2} = 10 \rightarrow buscamos\ ese\ valor\ en\ F_i = 12, que\ corresponde\ al\ x_i = 5$$

$$Me = 5$$

### 10.4.3. Media aritmética ($\overline{x}$)

Es el valor promedio del conjunto de datos del estudio estadístico. Para su cálculo debemos añadir una columna más a la tabla de frecuencias, $x_i \cdot f_i$.

La fórmula de la media aritmética es:

$$\boldsymbol{\overline{x} = \frac{\sum_{i=1}^{i=N} x_i \cdot f_i}{N}}$$

En el ejemplo de la variable discreta:

| $x_i$ | $f_i$ | $F_i$ | $f_{r_i}$ | $F_{r_i}$ | % | o | $x_i \cdot f_i$ |
|---|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 0,15 | 0,15 | 15 | 54 | 3 |
| 2 | 2 | 5 | 0,10 | 0,25 | 10 | 36 | 4 |
| 3 | 1 | 6 | 0,05 | 0,30 | 5 | 18 | 3 |
| 4 | 2 | 8 | 0,10 | 0,40 | 10 | 36 | 8 |
| 5 | 4 | 12 | 0,20 | 0,60 | 20 | 72 | 20 |
| 6 | 5 | 17 | 0,25 | 0,85 | 25 | 90 | 30 |
| 7 | 3 | 20 | 0,15 | 1 | 15 | 54 | 21 |
| | $N = 20$ | | + 1 | | + 100 | + 360 | + 89 |

$$\bar{x} = \frac{89}{20} = 4{,}45$$

## 10.5. Medidas de dispersión

Las medidas de dispersión nos indican las desviaciones que sufren las medidas de centralización. Estudiaremos la varianza y la desviación típica.

### 10.5.1. Varianza ($V$)

Su fórmula es:

$$V = \frac{\sum x_i^2 \cdot f_i}{N} - \bar{x}^2$$

Debemos añadir en la tabla de frecuencias una columna más: $x_i^2 \cdot f_i$. El valor de la varianza no se puede interpretar.

En el ejemplo de la variable estadística discreta:

| $x_i$ | $f_i$ | $F_i$ | $f_{r_i}$ | $F_{r_i}$ | % | $o$ | $x_i \cdot f_i$ | $x_i^2 \cdot f_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 3 | 3 | 0,15 | 0,15 | 15 | 54 | 3 | 3 |
| 2 | 2 | 5 | 0,10 | 0,25 | 10 | 36 | 4 | 8 |
| 3 | 1 | 6 | 0,05 | 0,30 | 5 | 18 | 3 | 9 |
| 4 | 2 | 8 | 0,10 | 0,40 | 10 | 36 | 8 | 32 |
| 5 | 4 | 12 | 0,20 | 0,60 | 20 | 72 | 20 | 100 |
| 6 | 5 | 17 | 0,25 | 0,85 | 25 | 90 | 30 | 180 |
| 7 | 3 | 20 | 0,15 | 1 | 15 | 54 | 21 | 147 |
| | $N = 20$ | | + 1 | | + 100 | + 360 | + 89 | + 479 |

$$V = \frac{479}{20} - 4{,}45^2 = 4{,}1475$$

### 10.5.2. Desviación típica ($\sigma$)

La desviación típica expresa lo que se desvía la media de los datos. Se puede interpretar porque se encuentra en las mismas unidades que los datos. Su fórmula es:

$$\boldsymbol{\sigma = \sqrt{V}}$$

En el ejemplo de la variable estadística discreta:

$$\sigma = \sqrt{4{,}1475} = 2{,}04$$

Para su interpretación obtenemos el intervalo $(\bar{x} - \sigma, \bar{x} + \sigma)$.

En nuestro ejemplo: $(4{,}45 - 2{,}04, 4{,}45 + 2{,}04) = (2{,}41, 6{,}49)$

Lo que significa que la mayoría de los datos del estudio que se realiza se encuentran entre 2,41 y 6,49.

## 10.6. Medidas que relacionan las medias de centralización y de dispersión: coeficiente de variación ($C.V.$)

El coeficiente de variación, como relaciona la media aritmética y la desviación típica, nos indica la variación relativa.

$$\boldsymbol{C.V.=\frac{\sigma}{\bar{x}}}$$

Cuanto más bajo sea indica que más agrupados están los datos.

En el ejemplo de la variable estadística discreta: $C.V.=\frac{2{,}04}{4{,}45}=0{,}46$