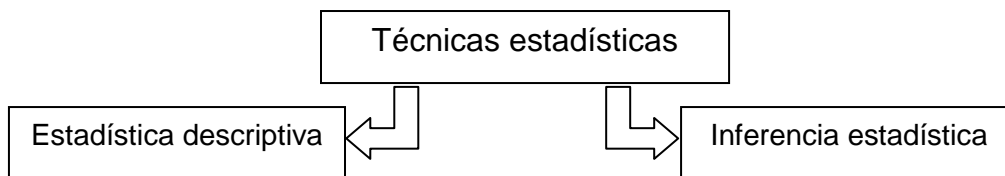


INTRODUCCIÓN

- **Fenómeno determinista:** al repetirlo en idénticas condiciones se obtiene el mismo resultado. (Ejemplo: kilómetros recorridos en un intervalo de tiempo a una velocidad determinada)
- **Fenómeno aleatorio:** no es posible predecir el resultado. (Ejemplo: resultado al lanzar un dado)

La **estadística** se ocupa de aquellos fenómenos no deterministas donde es imposible predecir los resultados. Es una rama de las matemáticas que trata de la recopilación, el análisis, la interpretación y la representación de una gran cantidad de datos numéricos.



La **estadística descriptiva** se ocupa del problema del manejo de ingentes cantidades de datos. Busca, fundamentalmente, el resumen de los mismos, presentándolos de forma más manejable con el objetivo de facilitar su interpretación. En pocos valores se refleja la máxima información posible.

Conceptos:

- **Población:** conjunto objeto de estudio.
- **Individuo** (o unidad): cada uno de los elementos de la población.
- **Muestra:** Subconjunto de la población que se estudia.
- **Tamaño de la población:** número de individuos de la población (M).
- **Tamaño de la muestra:** número de individuos de la muestra (n).
- **Carácter o variable:** propiedad que se desea observar sobre los elementos de la población.
- **Modalidad:** cada uno de los posibles valores observables de una característica. Las modalidades han de ser incompatibles y exhaustivas.

Tipos de variables: $\left\{ \begin{array}{l} \text{Cualitativas} \\ \text{Cuantitativas} \end{array} \right. \left\{ \begin{array}{l} \text{Ordinales} \\ \text{Nominales} \\ \text{Discretas} \\ \text{Continuas} \end{array} \right.$

Variable cualitativa: Las modalidades de la variable son características no numéricas. (Ejemplo: color de pelo, nivel de estudios)

Variable cuantitativa: Las modalidades de la variable son características numéricas. (Ejemplo: ingresos mensuales, edad)

- **Discreta:** El paso de un valor de la variable al siguiente representa un salto. (Ejemplo: número de empleados en una empresa)
- **Continua:** Dados dos valores de la variable siempre existen infinitos valores posibles entre ellos dos. (Ejemplo: tiempo que transcurre entre dos averías de una máquina)

TABLAS DE FRECUENCIAS

$n_i \rightarrow$ **Frecuencia absoluta** de la modalidad M_i : número de individuos de la población que presentan esta modalidad (número de veces que aparece esta modalidad en la población)

$f_i \rightarrow$ **Frecuencia relativa** de la modalidad M_i : proporción de individuos de la población que presentan esta modalidad. Se obtiene dividiendo la frecuencia absoluta entre el número de elementos de la población (N).

En variables
cuantitativas
u ordinales

$N_i \rightarrow$ **Frecuencia absoluta acumulada** de la modalidad M_i : suma de las frecuencias absolutas hasta la i -ésima modalidad.

$F_i \rightarrow$ **Frecuencia relativa acumulada** de la modalidad M_i : suma de las frecuencias relativas hasta la de la i -ésima modalidad.

Modalidades	Frecuencias absolutas n_i	Frecuencias relativas f_i	Frecuencias absolutas acumuladas N_i	Frecuencias relativas acumuladas F_i
M_1	n_1	f_1	N_1	F_1
M_2	n_2	f_2	N_2	F_2
...
M_k	n_k	f_k	$N_k=N$	$F_k=1$
Suma (\sum)	N	1		

Cuando las modalidades vienen dadas en intervalos, la tabla de frecuencias se completa con las columnas

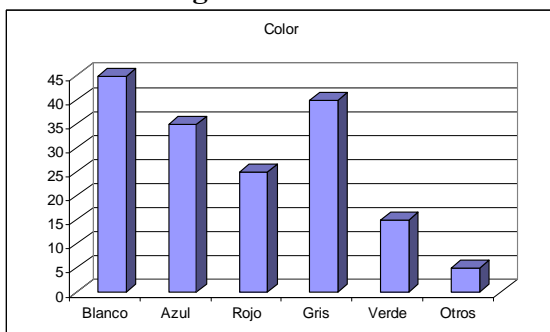
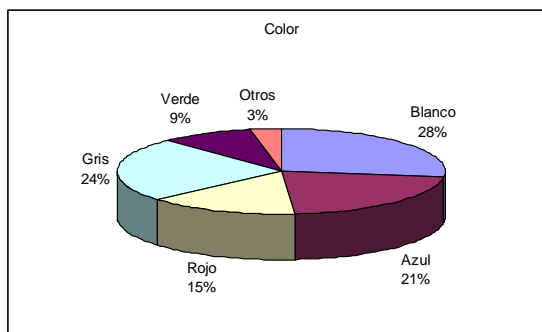
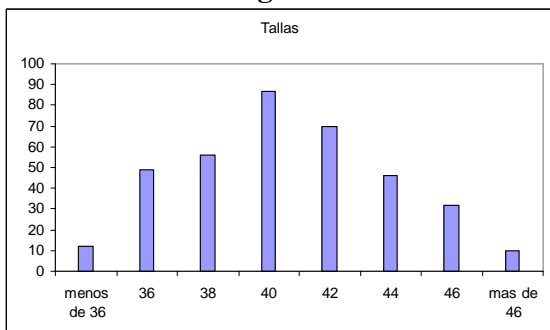
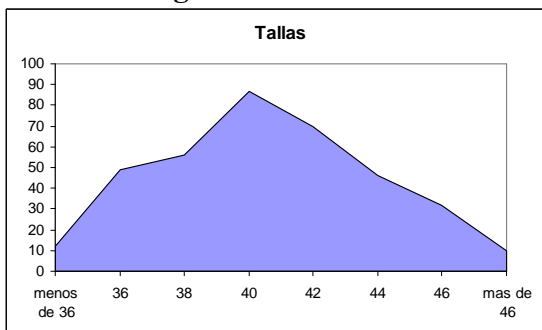
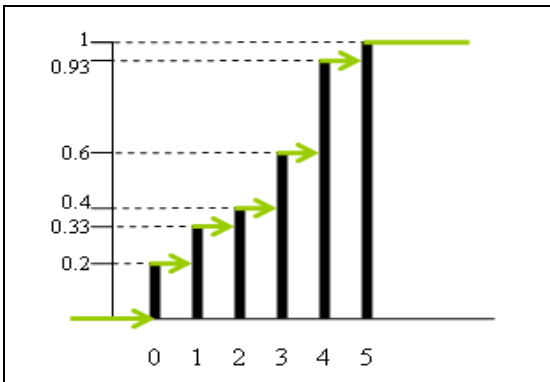
$(I_{i-1}, I_i] \rightarrow$ **Intervalos**.

$x_i \rightarrow$ **Marca de clase**: punto medio del intervalo $(I_{i-1}, I_i]$.

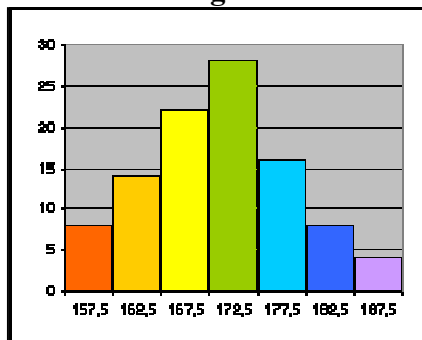
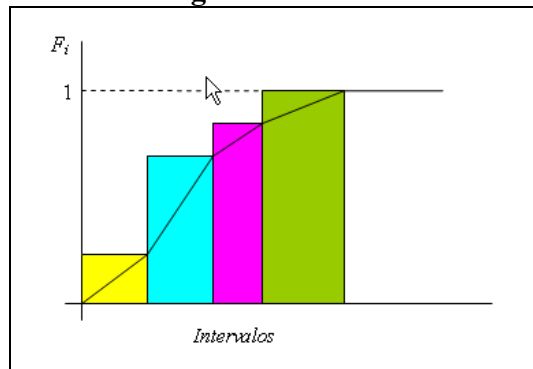
$a_i \rightarrow$ **Amplitud** del intervalo: Límite superior menos límite inferior del intervalo.

$h_i \rightarrow$ **Densidad de frecuencia**: Frecuencia absoluta del intervalo entre la amplitud del mismo.

Intervalos	Marca de clase x_i	Amplitud a_i	Densidad de frecuencia h_i	n_i	f_i	N_i	F_i
$(I_0, I_1]$	x_1	a_1	h_1	n_1	f_1	N_1	F_1
$(I_1, I_2]$	x_2	a_2	h_2	n_2	f_2	N_2	F_2
...
$(I_{k-1}, I_k]$	x_k	a_k	h_k	n_k	f_k	$N_k=N$	$F_k=1$
Suma				N	1		

REPRESENTACIONES GRÁFICAS**VARIABLES CUALITATIVAS****Diagrama de barras****Gráfico de sectores****VARIABLES CUANTITATIVAS DISCRETAS****Gráfico o diagrama de barras****Polígono de frecuencias****Curva acumulativa o de distribución****Diagrama de tallo y hojas:**

21	0 11111222222333444444
43	0 555555566778888888889
50	1 0233444
50	1 5555667777789
37	2 1133
33	2 556789
27	3 111224
21	3 55666789
13	4 1222333
6	4 88899
1	5
1	5
1	6 4

VARIABLES CUANTITATIVAS CONTINUAS**Histograma****Histograma acumulado**

1. MEDIDAS DE POSICIÓN

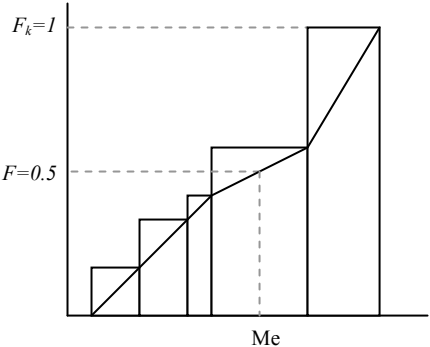
Medidas de posición central

Media aritmética: Es la suma, ponderada por sus frecuencias relativas, de los valores de la variable.

Caso discreto	Caso continuo
$\bar{x} = \sum_{i=1}^k x_i f_i = \frac{1}{N} \sum_{i=1}^k x_i n_i$	Sustituimos los intervalos por sus marcas de clase $\bar{x} = \sum_{i=1}^k x_i f_i = \frac{1}{N} \sum_{i=1}^k x_i n_i$

Mediana: Valor de la variable que divide a los individuos de la población en dos partes iguales, supuestos ordenados los datos.

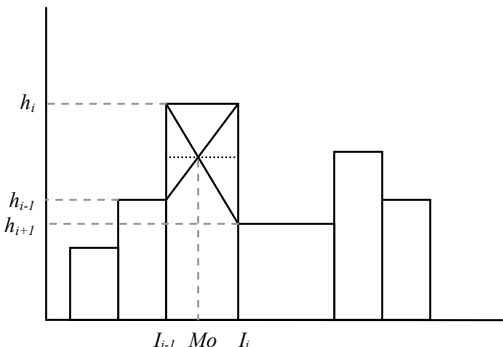
Caso discreto
<ul style="list-style-type: none"> · Si existe un valor para el cual $F_i = 0.5$, se toma como valor mediano el punto medio entre x_i y x_{i+1}. · En caso contrario, la mediana es el primer valor de la variable cuya frecuencia relativa acumulada F_i supere el valor 0.5.
Caso continuo
<ul style="list-style-type: none"> · Si existe algún intervalo para el cual $F_i = 0.5$, la mediana es el extremo superior de ese intervalo. · En caso contrario, la mediana es un valor entre los extremos del intervalo para el cual F_i es mayor que 0.5 por primera vez.



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

Moda: Es el valor más frecuente de la distribución. (No tiene por qué ser única)

Caso discreto
La moda es el valor de la variable que corresponde a la máxima frecuencia. En el gráfico de barras, es la modalidad a la que corresponde la barra más alta.
Caso continuo
El intervalo al que pertenece la moda es el que tiene la base del rectángulo más alto en el histograma



$$Mo = I_{i-1} + \frac{h_i - h_{i-1}}{(h_i - h_{i-1}) + (h_i - h_{i+1})} a_i$$

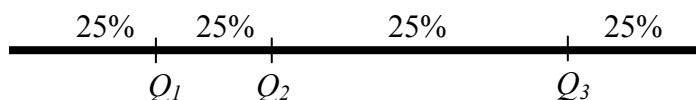
Medidas de posición no central

Cuantiles: El cuantil de orden α (α es un valor entre 0 y 1) es el valor de la variable que acumula el 100α % de la distribución.

$$\text{Variable continua: } C(\alpha) = I_{i-1} + \frac{\alpha \cdot N - N_{i-1}}{n_i} a_i$$

- **Cuartiles** (Q_i): Son puntos de la distribución que la dividen en cuatro partes, cada una de las cuales engloba el 25% de los datos.

$\alpha = 0.25, 0.5, 0.75$



- **Deciles** (D_i): Son puntos que dividen a la distribución en diez partes, cada una de las cuales engloba el 10% de los datos.

$\alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

- **Percentiles** (P_i): Son puntos que dividen a la distribución en cien partes, cada una de las cuales engloba el 1% de los datos.

$\alpha = 0.01, 0.02, 0.03, \dots, 0.99$

- Rango intercuartílico (**RIQ**): Es la diferencia entre Q_3 y Q_1

$$RIQ = Q_3 - Q_1$$

2. MEDIDAS DE DISPERSIÓN: cuantificamos la variabilidad de forma que sepamos si las medidas de posición central son o no representativas del conjunto de datos.

Varianza

$$\sigma_X^2 = Var(X) = \frac{1}{N} \sum_{i=1}^k (x_i - \bar{x})^2 n_i = \frac{1}{N} \sum_{i=1}^k x_i^2 n_i - \bar{x}^2$$

Desviación típica

$$\sigma_X = \sqrt{Var(X)}$$

Coefficiente de variación

$$CV = \frac{\sigma}{\bar{x}}$$