

Tema 13: ESTADÍSTICA

La estadística es una ciencia que se encarga de recopilar y ordenar datos referidos a diversos fenómenos para su posterior análisis e interpretación.

1) DEFINICIONES:

Población: Conjunto de todos los elementos objeto de estudio.

Muestra: Parte de la población que se elige para hacer el estudio.

Individuo: Cada uno de los elementos que forman la población o muestra.

Carácter o variable estadística: Cualidad a estudiar en la muestra.

Puede ser: { Cuantitativo (mensurable) { Discreto (valores aislados)
Continuo (todos los valores de un intervalo)
Cualitativo (cualidad o modalidad no mensurable)

Serie estadística: Conjunto de datos que se obtienen al observar una variable en una población.

2) PASOS EN UN ESTUDIO ESTADÍSTICO:

1. Recogida de datos:

Una vez elegida la muestra de la población, se procede a la toma de datos. Llamaremos x_i ($i = 1, 2, \dots, n$) a cada valor numérico o modalidad de la variable estadística obtenidos.

2. Organización de los datos en tablas de frecuencia:

La información se ordena y resume en tablas en las que a cada valor o modalidad se le asigna su:

- **Frecuencia absoluta (f_i):** Número de veces que cada valor o modalidad de la variable estadística se repite.

A cada x_i le corresponde una f_i ($i = 1, 2, \dots, n$)

- **Frecuencia absoluta acumulada (F_i):** Suma de las frecuencias absolutas de los valores menores o iguales a x_i : $F_i = f_1 + f_2 + f_3 + \dots + f_i$

- **Frecuencia relativa (h_i):** Cociente entre la frecuencia absoluta y el número total de datos (N):

A cada x_i le corresponde una $h_i = \frac{f_i}{N}$

- **Frecuencia relativa acumulada (H_i):** Suma de las frecuencias relativas de los valores menores o iguales a x_i : $H_i = h_1 + h_2 + h_3 + \dots + h_i =$

$$\frac{f_1}{N} + \frac{f_2}{N} + \frac{f_3}{N} + \dots + \frac{f_i}{N}$$

- **Porcentaje:** El asignado a cada x_i será: $100 \cdot h_i$

Valores x_i	Frecuen- cias ab- solutas f_i	Frecuencias absolutas acumuladas F_i	Frecuencias relativas h_i	Frecuencias relativas acumuladas H_i	Porcentajes
x_1	f_1	f_1	h_1	h_1	$100 \cdot h_1$
x_2	f_2	$f_1 + f_2$	h_2	$h_1 + h_2$	$100 \cdot h_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	f_i	$f_1 + f_2 + f_3 + \dots + f_i$	h_i	$h_1 + h_2 + h_3 + \dots + h_i$	$100 \cdot h_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	f_n	N	h_n	1	$100 \cdot h_n$
Totales:	N		1		100%

3. Elaboración de gráficos estadísticos:

- 1) **Diagrama de barras:** Gráfico formado por barras separadas de altura proporcional a la frecuencia de cada valor o modalidad estadística (x_i):
 - Eje de abscisas: x_i
 - Eje de ordenadas: frecuencias.
 Se utiliza para variables cualitativas y cuantitativas discretas.
- 2) **Polígono de frecuencias:** Se obtiene uniendo los extremos de las barras en un diagrama de barras.
- 3) **Diagrama de sectores:** Gráfico que consiste en un círculo dividido en sectores de amplitud proporcional a la frecuencia de cada valor o modalidad de la variable estadística. El ángulo de cada sector es proporcional a la frecuencia de cada x_i :

$$\frac{360^\circ}{N} = \frac{\alpha_i}{f_i}$$

Gráfico de barras

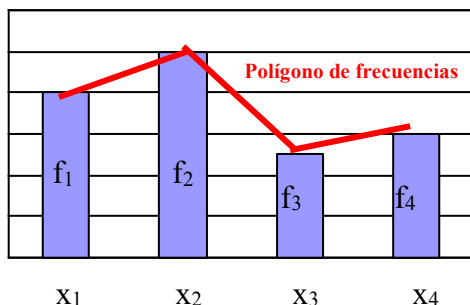
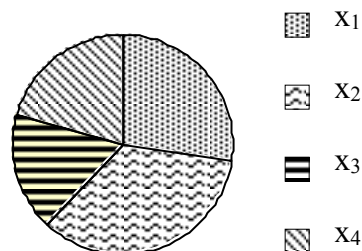


Diagrama de sectores



4. Resumen de la información con las medidas de centralización:

Las medidas de centralización son valores en torno a los que se agrupan los datos.

- **Media aritmética ($\bar{x}$):** (sólo para variables estadísticas cuantitativas)

Es el valor medio de los datos. Para calcularla:

1. Se añade en la tabla de frecuencias la columna $x_i \cdot f_i$
2. Se calcula el total de dicha columna como suma de sus valores.
3. Se divide el valor anterior entre N

- **Mediana (M_e):** (sólo para variables estadísticas cuantitativas)

Es el valor central del conjunto de datos ordenados: hay tantos datos mayores que él como menores.

Para calcularla, se utiliza la columna de frecuencias absolutas acumuladas.

- **Moda (M_o):**

Valor que tiene mayor frecuencia.

Si existe más de un valor con la frecuencia mayor, la serie es **bimodal** (dos modas) o **multimodal** (más de dos modas).

5. Resumen de la información con las medidas de dispersión.

Las medidas de dispersión indican el grado de agrupación de los datos en torno a la media, la mediana y la moda.

- **Recorrido o rango:** (sólo para variables estadísticas cuantitativas)

Es la diferencia entre el mayor y el menor de los datos.

-. Cuanto mayor es el rango, mayor es la dispersión.

- **Varianza (s^2):** (sólo para variables estadísticas cuantitativas)

-. La Varianza de un dato es el cuadrado de la diferencia entre el dato y la media.

-. La Varianza de una serie de datos es la media aritmética de los valores absolutos de las varianzas de cada dato.

$$s^2 = \frac{\sum_{i=1}^n f_i \cdot (x_i - \bar{x})^2}{N} = \frac{\sum_{i=1}^n f_i \cdot x_i^2}{N} - \bar{x}^2$$

-. Para calcularla, se añaden dos columnas a la tabla de frecuencias: una en la que se calcula x_i^2 y otra en la que aparezca el producto $f_i \cdot x_i^2$

- **Desviación típica (s):** (sólo para variables estadísticas cuantitativas)

-. La Desviación Típica de un dato es la diferencia entre el dato y la media.

-. La Desviación Típica de una serie de datos es la media aritmética de los valores absolutos de las desviaciones de cada dato. En la práctica, es la raíz cuadrada de la varianza.

Coefficiente de variación (CV):

El Coeficiente de Variación de una distribución es el cociente entre la desviación típica y la media aritmética:

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

El resultado se suele expresar en forma de porcentaje. Se utiliza para comparar la dispersión relativa de distintas series de datos.

3) PARA VARIABLES CUANTITATIVAS CONTINUAS:

Los pasos en los que se desarrolla el estudio estadístico son los mismos, las únicas diferencias a tener en cuenta aquí son:

- Los datos se agrupan en intervalos semiabiertos, todos de la misma amplitud.
- A cada intervalo anterior se le asocia una **marca de clase, x_i** ; que coincide con el valor medio del intervalo.
- El **Histograma** sustituye al diagrama de barras, las barras siguen siendo del mismo grosor pero no hay espacio entre ellas, comienzan en el extremo inferior del intervalo y acaban en el superior, siendo su altura proporcional a la f_i de cada intervalo.

Ejemplo:

Intervalo	x_i	f_i	F_i
[50, 60)	55	8	8
[60, 70)	65	10	18
[70, 80)	75	16	34
[80, 90)	85	14	48
[90, 100)	95	10	58
[100, 110)	105	5	63
[110, 120)	115	2	65

