

ESTADÍSTICA

1. Población, muestra e individuo

Las características de una distribución se pueden estudiar directamente sobre la población o se pueden inferir a partir de l estudio de una muestra.

Población estadística es el conjunto de todos los elementos objeto de un estudio estadístico.

Por ejemplo: el conjunto de todos los tornillos producidos diariamente por un fabricante.

Muestra es un subconjunto, extraído de la población, cuyo estudio sirve para inferir características de toda la población.

Por ejemplo: 1000 de esos tornillos cuyas características (longitud, correcto/defectuoso, nº de pasos de rosca, etc.) van a ser estudiadas para inferir características del total de los producidos.

Individuo es cada uno de los elementos que forman la población o la muestra.

Por ejemplo: cada uno de los tornillos producidos es un individuo.

2. Variables estadísticas

Son los caracteres objeto del estudio estadístico.

Las variables pueden ser:

- **Cuantitativas:** numéricas
 - **Discretas:** sólo pueden tomar varios valores.
Por ejemplo: nº de pasos de rosca de un tornillo, nº de hermanos, etc.
 - **Continuas:** pueden tomar todos los valores en un intervalo.
Por ejemplo: longitud de un tornillo, estatura de una persona, etc.
- **Cualitativas:** no numéricas.
Por ejemplo: tornillos correctos o defectuosos, estudios universitarios que pueden realizarse, etc.

3. El proceso que se sigue en estadística

1º Determinar lo que se quiere estudiar.

2º Seleccionar y acotar las variables que se van a analizar.

3º Recolectar los datos.

4º Organizar los datos.

5º Elaborar tablas y gráficas.

6º Obtener parámetros.

4. Tablas de frecuencias

Una vez recogidos los datos, hay que tabularlos; es decir hay que elaborar una tabla en la que aparezcan bien organizados los valores de la variable que se esta analizando y el número de individuos que toma cada valor o cada intervalo de valores. Es lo que se llama una **tabla de frecuencias**.

Frecuencia absoluta (frecuencia) es el número de individuos correspondiente a cada valor o a cada intervalo de valores de la variable.

5. Gráficos estadísticos

Son representaciones que permiten visualizar fácilmente la información estadística recogida.

Los gráficos de uso más frecuente son:

Diagrama de barras se utiliza para representar tablas de frecuencias correspondientes a variables cualitativas o cuantitativas discretas.

Las barras son estrechas y se sitúan sobre los valores puntuales de la variable.

Histograma se utiliza para representar tablas de frecuencias correspondientes a variables cuantitativas continuas y también para las cuantitativas discretas cuando toman gran número de valores y se tabulan agrupándolos en intervalos.

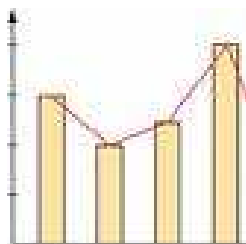
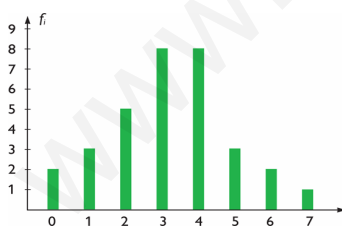
Polígono de frecuencias se obtiene al unir los puntos medios de los lados superiores de las barras (en un diagrama de barras) o de los lados superiores de los rectángulos (en un histograma).

- También pueden representarse diagramas de barras, histogramas y polígonos de frecuencias acumuladas.

Diagrama de sectores se utiliza para representar tablas de frecuencias correspondientes a cualquier tipo de variable.

El ángulo de cada sector es proporcional a la frecuencia.

Diagrama de barras y polígono de frecuencias



Histograma y polígono de frecuencias

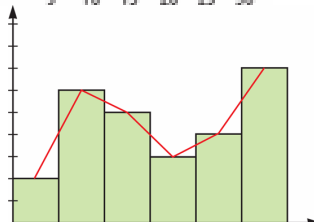
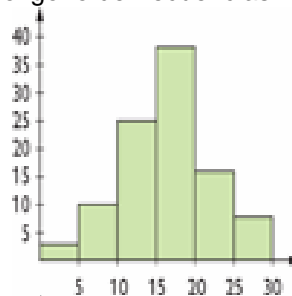


Diagrama de sectores

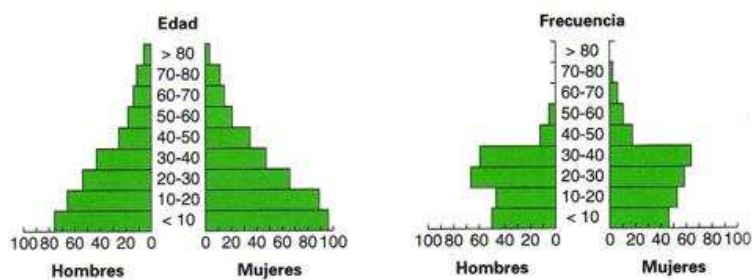


Otras representaciones gráficas:

Pirámides de población

Cuando se realizan representaciones correspondientes a edades de población, cambiamos el eje Y por el eje X para obtener las llamadas pirámides de población, que no son más que 2 histogramas a izquierda y derecha, para hombres y mujeres.

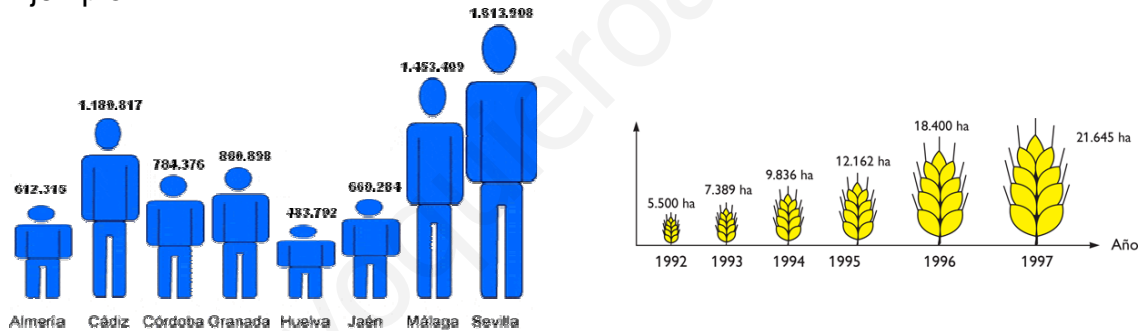
Ejemplo:



Pictogramas

Son gráficos con dibujos alusivos al carácter que se está estudiando y cuyo tamaño es proporcional a la frecuencia que representan; dicha frecuencia se suele representar.

Ejemplo:

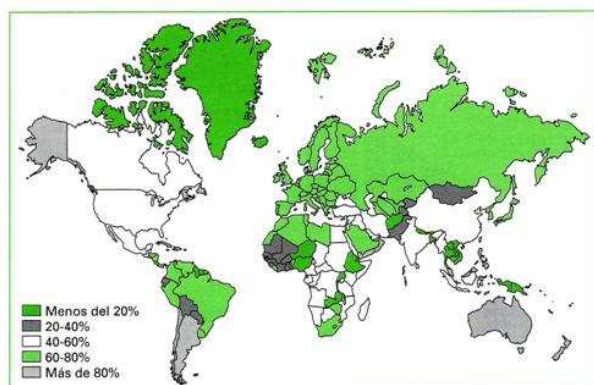


Padrón Municipal de Habitantes a 1 de Enero de 2005.
Habitantes de cada una de las 8 provincias de Andalucía.

Cartogramas

Son gráficos realizados sobre mapas, en los que aparecen indicados sobre las distintas zonas cantidades o colores de acuerdo con el carácter que representan.

Ejemplo



Urbanización en el mundo atendiendo a la industrialización

6. Parámetros estadísticos

Los parámetros estadísticos sirven para sintetizar la información y permiten apreciar con rapidez y eficacia las características más relevantes de una determinada distribución.

Parámetros o medidas de centralización

Indican en torno a que valor (centro) se distribuyen los datos.

- **Media** o promedio, es el promedio de los datos de la distribución.

Se calcula así: $\bar{x} = \frac{\sum x_i}{N}$ y cuando los datos están en tablas de frecuencias: $\bar{x} = \frac{\sum x_i f_i}{\sum f_i}$

- **Moda**: valor con mayor frecuencia.
- **Mediana**: valor que ocupa el lugar central.

Parámetros o medidas de dispersión

Informan sobre cuánto se alejan del centro los valores de la distribución, es decir sirven para medir cómo de dispersos están los datos. En todos ellos la idea clave es medir el grado de separación de los datos de la media.

- **Desviación media**: es el promedio de las distancias de los datos a la media.

Se calcula así: $D.M. = \frac{\sum |x_i - \bar{x}|}{N}$, $D.M. = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i}$
↑ en tablas de frecuencias

- **Varianza**: es el promedio de los cuadrados de las distancias de los datos a la media.

Se calcula así:
$$\left\{ \begin{array}{l} V = \frac{\sum (x_i - \bar{x})^2}{N} \\ V = \frac{\sum x_i^2}{N} - \bar{x}^2 \end{array} \right. \quad \left\{ \begin{array}{l} V = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} \\ V = \frac{\sum x_i^2 f_i}{\sum f_i} - \bar{x}^2 \end{array} \right.$$

↑ en tablas de frecuencias

- **Desviación típica**: es la raíz cuadrada de la varianza. $\sigma = \sqrt{V}$
- **Rango o recorrido**: Diferencia entre el valor mayor y el menor. Es decir, la longitud del tramo donde están los datos.

Coefficiente de variación

Sirve para comparar la dispersión de dos poblaciones heterogéneas.

Se calcula así: $C.V. = \frac{\sigma}{\bar{x}}$

Al dividir la desviación típica entre la media se está relativizando la dispersión.

El resultado se da, a veces en tantos por ciento.