

<https://www.youtube.com/watch?v=hODwSUX0kT4>

## 1. POBLACIÓN E INDIVIDUO. MUESTRAS

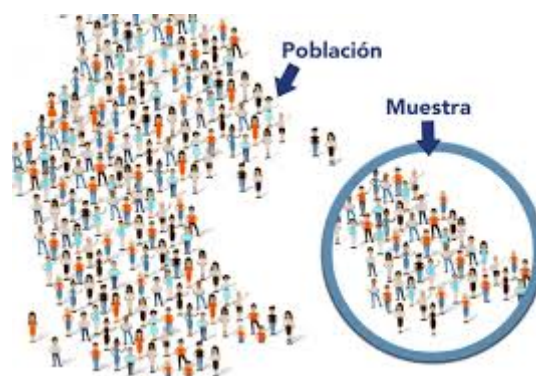
La **Estadística** estudia los métodos científicos para:

- Recoger, organizar, resumir y analizar los datos.
- Obtener conclusiones y tomar decisiones a partir del análisis de los datos.

La **Estadística descriptiva** se encarga de la primera fase del proceso y la **Estadística inferencial** de la segunda.

Llamamos **población**,  $N$ , al conjunto de elementos u objetos sobre los que hacemos el estudio estadístico. A cada elemento de la población se le denomina individuo.

Una **muestra** es un subconjunto de la población. El tamaño de la muestra es el número de elementos que contiene y se designa con la letra  $n$ . Una muestra es válida si es representativa de la población y para ello debe ser elegida de forma aleatoria.



## 2. VARIABLES ESTADÍSTICAS: CUALITATIVAS Y CUANTITATIVAS.

La variable estadística es una propiedad que permite clasificar a los individuos de una población. Pueden ser:

- **Variable estadística cualitativa:** son aquellas que no se pueden medir o expresar numéricamente (estado civil). Llamamos modalidad a cada una de las distintas situaciones que pueden aparecer (casado, soltero, viudo, divorciado,...). Si la variable admite una ordenación se denomina **cualitativa ordinal**. En caso contrario, se denomina **cualitativa nominal**.
- **Variable estadística cuantitativa:** son aquellas que se pueden medir o expresar numéricamente. Puede ser:
  - **Discreta:** si toma valores aislados, es decir, entre dos valores consecutivos no puede tomar valores intermedios (número de hermanos)
  - **Continua:** si entre dos valores cualesquiera se puede tomar otro valor intermedio (altura de los andaluces)



### 3. FRECUENCIAS ABSOLUTAS Y RELATIVAS. TABLAS DE FRECUENCIA.

Cuando el número de datos ( $x_i$ ) es muy grande o hay valores que se repiten conviene organizarlos en **tablas estadísticas o de frecuencias**.

- La **frecuencia absoluta** ( $f_i$ ) de un valor  $x_i$  es el número de veces que se repite ese valor.
- La **frecuencia absoluta acumulada** ( $F_i$ ) de un valor  $x_i$  es la suma de todas las frecuencias absolutas de los valores menores o iguales que  $x_i$ .
- La **frecuencia relativa** ( $h_i$ ) de un valor  $x_i$  es el cociente entre la frecuencia absoluta de ese valor y el número total de datos.

$$h_i = \frac{f_i}{N}$$

- La **frecuencia relativa acumulada** ( $H_i$ ) de un valor  $x_i$  es el cociente entre la frecuencia absoluta acumulada de ese valor y el número total de datos.

$$H_i = \frac{F_i}{N}$$

- Los **porcentajes** de cada valor  $x_i$  se hallan multiplicando por 100 las frecuencias relativas de ese valor.

$$p_i = 100 \cdot f_i$$

Las tablas de frecuencias tienen, por tanto, las siguientes columnas:

Variable $x_i$	Frecuencia absoluta $f_i$	Frecuencia absoluta acumulada $F_i$	Frecuencia relativa $h_i$	Frecuencia relativa acumulada $H_i$	Porcentajes $p_i$
-------------------	------------------------------	--	------------------------------	--	----------------------

Cuando la variable estadística tiene un número elevado de datos conviene agruparlos en intervalos o clases. En estos casos se tomará como representante de la variable la **marca de clase**,  $c_i$ , que es el valor medio de los extremos del intervalo.

$$[a, b) \rightarrow c_i = \frac{a + b}{2}$$

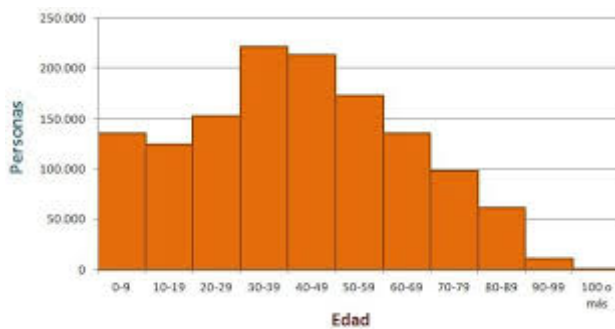
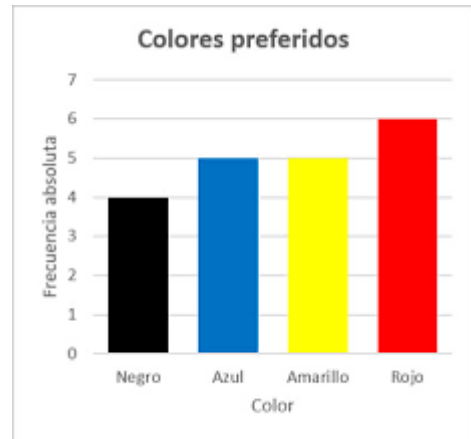
Los intervalos serán cerrados por la izquierda y abiertos por la derecha, coincidiendo éste extremo con el extremo izquierdo del siguiente intervalo. La diferencia entre los dos extremos del intervalo se denomina **amplitud del intervalo**.

#### 4. GRÁFICOS ESTADÍSTICOS: DIAGRAMAS DE BARRAS Y SECTORES. POLÍGONO DE FRECUENCIA.

Los gráficos estadísticos muestran la información de forma sencilla y clara y dan una visión global de los datos. Los gráficos más usuales son:

- **Diagrama de barras.** Es una representación de los datos mediante rectángulos cuyas bases son los valores que toma la variable y las alturas son las frecuencias absolutas de dichos valores. Este tipo de representación se utiliza para variables cuantitativas discretas o cualitativas.

Si unimos los puntos medios de los extremos superiores de las barras mediante una línea poligonal se obtiene el **polígono de frecuencias**.



- **Histograma.** Es una representación similar al diagrama de barras con las diferencias que los rectángulos son adosados y las bases son los intervalos en los que se ha agrupado la variable. Se utiliza para variables cuantitativas continuas.

- **Diagrama de sectores.** Es una representación con forma de círculo que se divide en tantos sectores como valores toma la variable. La amplitud de cada sector se obtiene multiplicando la frecuencia relativa de cada valor por 360°. Este gráfico se utiliza principalmente en variables cualitativas.



## 5. PARÁMETROS DE CENTRALIZACIÓN: MEDIA, MODA Y MEDIANA.

Los parámetros de centralización son datos que representan de forma global a toda la población. Por ejemplo, si hacemos un examen en la clase y queremos tener una idea global del resultado de dicho examen, ¿cómo lo podríamos hacer? Parece lógico que sumando todas las notas y dividiendo el resultado por el número de alumnos, es decir, lo que todos conocemos como calculando la media.

Los parámetros de centralización son tres: media, moda y mediana.

- **Media:** es el valor obtenido al sumar todos los datos ( $x_i$ ) y dividir el resultado entre el número total de datos ( $N$ ). Se representa por  $\bar{x}$  y se puede hallar sólo para variables cuantitativas.

Si los datos vienen agrupados en una tabla de frecuencias, la fórmula que permite calcular la media es:

$$\bar{x} = \frac{\sum x_i \cdot f_i}{N}$$

- **Moda:** es el valor que tiene mayor frecuencia absoluta. Se representa por **Mo** y se puede hallar para variables cualitativas y cuantitativas.

Si en un grupo hay dos o varias puntuaciones con la misma frecuencia y esa frecuencia es la máxima, la distribución es bimodal o multimodal, es decir, tiene varias modas. (1, 1, 1, 4, 4, 5, 5, 5, 7, 8, 9, 9, 9 Mo= 1, 5, 9)

Cuando todas las puntuaciones de un grupo tienen la misma frecuencia, no hay moda.

- **Mediana:** es el valor que ocupa el lugar central de todos los datos cuando éstos están ordenados de menor a mayor. Se representa por **Me** y se puede hallar sólo para variables cuantitativas.

Para calcular la mediana cuando la variable estadística es **cuantitativa discreta**, primero ordenamos los datos de menor a mayor.

- Si la serie tiene un número impar de medidas la mediana es la puntuación central de la misma.

$$2, 3, 4, 4, 5, 5, 5, 6, 6 \rightarrow Me = 5$$

- Si la serie tiene un número par de puntuaciones la mediana es la media entre las dos puntuaciones centrales.

$$7, 8, 9, 10, 11, 12 \rightarrow Me = 9.5$$

Cuando la variable estadística es **cuantitativa continua**, la mediana se encuentra en el intervalo donde la frecuencia acumulada llega hasta la mitad de la suma de las frecuencias absolutas.

## 6. PARÁMETROS DE DISPERSIÓN: RANGO

Las medidas de dispersión nos informan sobre cuánto se alejan del centro los valores de la distribución.

Uno de los parámetros de dispersión es el **rango** que se calcula hallando la diferencia entre el mayor y el menor de los datos de una distribución estadística. Sólo se puede calcular en el caso de variables estadísticas cuantitativas.