

Date:

1. ¿Qué es la Estadística?

La **estadística** es la ciencia que permite utilizar eficazmente un gran número de datos relativos a grupos de individuos o experimentos. La estadística trabaja con datos y nos enseña cómo recogerlos, organizarlos, analizarlos, resumirlos, interpretarlos y presentarlos.

La estadística recoge los datos, los procesa y los utiliza para responder a una pregunta, pero no responde cualquier pregunta.

Necesita una **pregunta estadística**, en la que espera obtener una variedad de respuestas que queramos ver como se organizan y a que tienden

Ejemplo:

¿Qué edad tienen los 100 alumnos del primer curso del IES Bernardo Balbuena?

12. Definiciones básicas

2.1 Población, muestra e individuo

Población : conjunto de elementos que vas a estudiar.

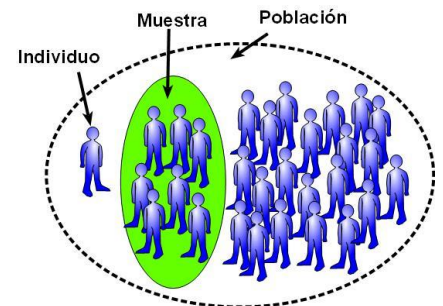
Ejemplo: Los 100 alumnos del primer curso

Muestra : un subconjunto representativo de la población .

Ejemplo: Por ejemplo, 10 estudiantes de cada clase en el primer curso.

Individuo: cada unidad de la población.

Ejemplo: cada uno de los alumnos del primer curso.



2.2 Variables estadísticas.

Una **variable estadística** es cada una de las características o cualidades que poseen los individuos de una población.

Ejemplo: edad, peso, cantante preferido,...

Date: _____

Los **valores estadísticos** son los diferentes resultados que se pueden obtener en un estudio estadístico.

Ejemplo: 12 años, 13 años, 14 años,...

Según como sean los valores que toman podemos hablar de :

- **variables cuantitativas** (numéricas)
Los datos cuantitativos pueden ser
 - **discretos** (limitados a valores concretos)
 - **continuos** (cualquier valor dentro de un rango).
- **Variabes cualitativas** (no numéricas).

Ejemplos:

- *El color favorito... Variable cualitativa (los valores posibles son azul, verde, rojo,...*
- *Cuántos años tiene... Variable cuantitativa discreta (los valores posibles son 12, 13, 14..)*
- *Cómo son las variables continuas cuantitativas (los valores posibles son cualquier número decimal entre 1 m y 2 m, por ejemplo)*

3. Organización de los datos. Tablas de frecuencia.

Una **tabla de frecuencias** (o distribución de frecuencias) es una organización en forma de tabla de los datos estadísticos, en la que a cada valor le corresponde una frecuencia (las veces que aparece)

Una tabla de frecuencias de valores puede utilizarse para resumir una lista grande.

Es aconsejable seguir estos pasos para hacer una tabla de frecuencias:

- Recoger datos.
- Organizar los datos y mostrarlos en una tabla de frecuencias.
- Dibuja un gráfico.

| Day | Number of customers | Frequency |
|-----------|---------------------|-----------|
| Monday | | 18 |
| Tuesday | | 13 |
| Wednesday | | 20 |
| Thursday | | 14 |
| Friday | | 21 |
| Saturday | | 27 |
| Sunday | | 26 |

Date: _____

3.1 Tipos de frecuencia

Podemos añadir muchas columnas en una tabla de frecuencias. Este curso veremos solo frecuencias absolutas y la relativas.

- **Frecuencia absoluta** (o sólo frecuencia): es el número de veces que aparece un determinado valor en un estudio estadístico. Se denota para n_i .

La suma de las frecuencias absolutas es igual al **número total de datos**, que se denomina **N**.

- **Frecuencia relativa**: es el cociente entre la frecuencia absoluta de un determinado valor y el número total de datos multiplicado por 100 (porque se suele expresar en forma de porcentaje) . Se denota por f_i .

La suma de las frecuencias relativas es siempre igual a 1.

Ejemplo: Se ha preguntado a 30 alumnos por el número de hermanos que tienen en su familia. Estas son las respuestas:

1. Recoge los datos: 1, 2, 1, 3, 6, 3, 2, 1, 1, 2, 2, 3, 2, 3, 2, 2, 4, 2, 3, 3, 2, 3, 2, 3, 4, 2, 2, 3, 1 y 2.

2. Organiza los datos en una tabla de frecuencias:
Hay 30 datos. $N=30$

| Número de herman@s: x_i | Recuento | Frecuencia absoluta: n_i | Frecuencia relativa : f_i |
|---------------------------|---------------|----------------------------|----------------------------------|
| 1 | IIII I | 6 | $\frac{6}{30} \cdot 100 = 20 \%$ |
| 2 | IIII IIII III | 13 | $13/30 \cdot 100 = 43,3 \%$ |
| 3 | IIII III | 8 | $8/30 \cdot 100 = 26,7 \%$ |
| 4 | II | 2 | $2/30 \cdot 100 = 6,7 \%$ |
| 5 | | 0 | $0/30 \cdot 100 = 0 \%$ |
| 6 | I | 1 | $1/30 \cdot 100 = 3,3 \%$ |
| Total | | 30 | 100% |

Date: _____

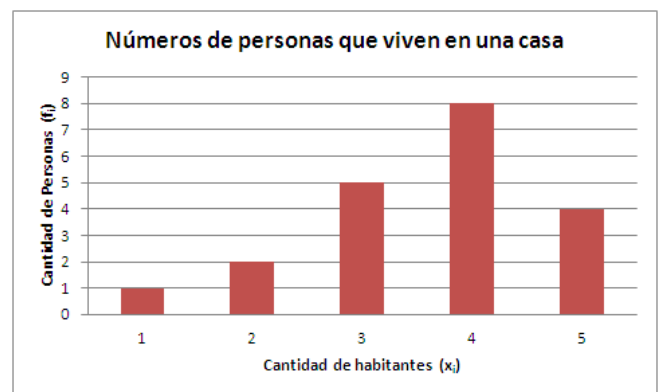
4. Gráficos estadísticos.

Los datos registrados en los experimentos estadísticos o en las encuestas se muestran mediante un **gráfico estadístico**. Cada tipo de gráfico tiene sus ventajas e inconvenientes y, según el tipo de datos, podemos elegir uno u otro.

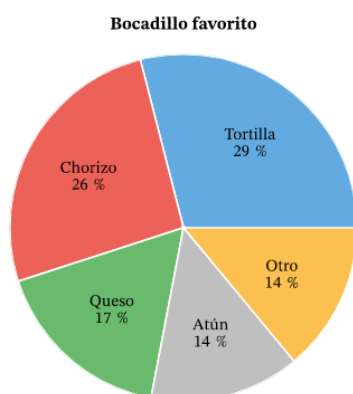
4.1 Diagrama de barras.

La información se presenta en un eje de coordenadas. Los valores de la variable se representan en el eje horizontal y las frecuencias absolutas, relativas o acumuladas se representan en el eje vertical.

Los datos se representan mediante barras cuya altura es proporcional a la frecuencia.



4.2 .Diagrama de sectores



Los datos se representan en un círculo y el ángulo de cada sector circular es proporcional a la frecuencia absoluta correspondiente.

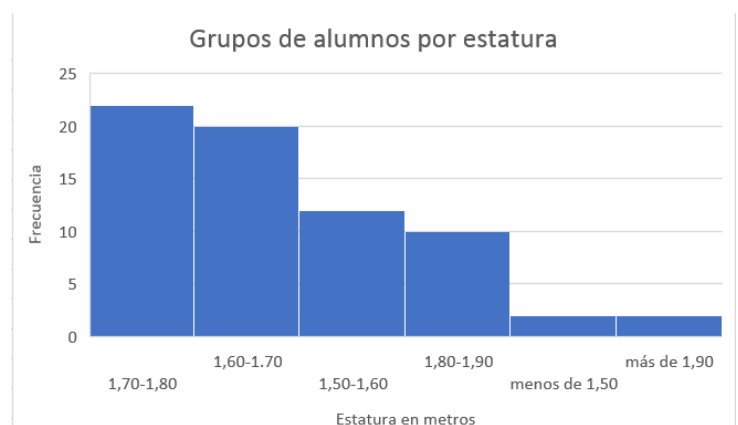
El gráfico de sectores se tiene que construir con la ayuda de un transportador .

Se puede utilizar una fórmula: $Angle = \frac{frecuencia \cdot 360}{Total\ data}$

4.3 Histograma.

Es similar a un diagrama de barras, pero un histograma agrupa los números en **intervalos** La altura de cada barra muestra cuántos valores entran en cada intervalo.

Tú decides qué intervalos utilizar.



Date:

5. 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

5.1 Medidas de centralización:

Una medida de centralización es un número que se utiliza para representar la media de un conjunto de valores de datos. La moda, la mediana y la media son tres tipos diferentes de **media**.

Media

La media de un conjunto de valores se calcula sumando todos los valores y dividiendo por el número de valores de la misma.

Ejemplo.-Hallar la media de 10, 11, 7 y 8 a la centésima más cercana.

$$\text{Mean} = \bar{x} = \frac{10+11+7+8}{4} = 9$$

Mediana

La mediana de un conjunto de números es el número central cuando los números se escriben en orden .

- Si la lista tiene un número impar de valores, el número central en este orden es la mediana.
- Si hay un número par de valores, la mediana es la suma de los dos números del medio, dividida por 2.

Observa que siempre hay tantos números mayores o iguales a la mediana en la lista como menores o iguales a la mediana en la lista.

Ejemplo: Mis notas en matemáticas fueron 7, 10, 8, 10 y 7 el año pasado. Encuentra la mediana. Colocadas en orden, las notas fueron 7, 7, 8, 10, 10. El número 8 es la mediana.

Date: _____

Imagina que mis notas eran 7, 10, 8, 10, 7 y 7. Colocadas en orden, las marcas eran 7, 7, 7, 8, 10, 10. Ahora los números del medio son 7 y 8, el tercero y el cuarto

La mediana es la media de estas dos cifras: $Me = \frac{7+8}{2} = 7.5$

Moda

La moda de un conjunto de números es el número (o números) que se dan con más frecuencia.

Ejemplo: Los alumnos de la clase de Julián tienen las siguientes edades: 5, 7, 5, 3, 4, 6, 6, 7, 3. Encuentra la moda de sus edades. El número que más veces aparece en la lista es el 6, que aparece tres veces. Ningún otro número aparece tantas veces. La moda de sus edades es el 6.

5.2 Medidas de la dispersión

Una medida de dispersión es un parámetro que indica el grado de dispersión de los valores de los datos. El recorrido y la desviación estándar (ne se ve este cursi) son medidas de dispersión.

Recorrido .

El recorrido de un conjunto de números es la diferencia entre el número mayor y el menor.

Para calcular el recorrido:

- Poner los números en orden: 3 6 6 6 7 9 11 11 13
- Resta el número menor al mayor: $13 - 3 = 10$

El rango de este conjunto de números es 10.

Date: _____

EXTRA:

A veces es útil utilizar grupos (**intervalo**) de valores para que los datos sean más fáciles de entender. El punto medio de cada intervalo se llama **marca de clase**

En este caso:

- Encuentra el dato mayor y el dato menor valor.
- Decide cuántos intervalos quieres (entre 4 y 10)
- Resta el menor dato al mayor dato.
- Divide esta cantidad por el número de intervalos que has elegido .
- Encuentra un número natural que se aproxime a este cociente. Será la amplitud de tus intervalos

Ejemplo: Agrupar estos datos en una tabla de frecuencias:

35 47 63 25 31 8 19 55 47 14 24 36 56 61 15 43 22 5 10 36 45 66 18 20 53 31 40 60
44 47

- El mayor valor es 66 y el menor es 5:
- $66-5=61$.
- Decidimos 5 intervalos de clase.
- $61/5=12.3$.

| x_i (data) | Tally | f_i (frequency) |
|--------------|-------|-------------------|
| [3,16) | | 5 |
| [16,29) | | 6 |
| [29,42) | | 6 |
| [42,55) | | 7 |
| [55,68) | | 6 |

- La amplitud es 13. Así que las clases son [3,16), [16,29), [29,42), [42,55) y [55,68)

Polígono de frecuencia

Se forma uniendo la marca de clase de los intervalos mediante segmentos de línea .

