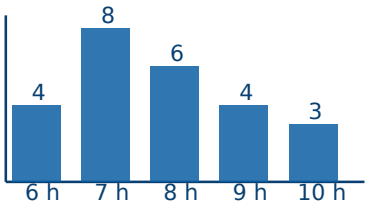
Apuntes - Estadística

Curso: 3 Eso

En este tema aprenderás: población, muestra y variables, tablas de frecuencias, gráficos estadísticos, medidas de centralización, medidas de posición y dispersión y cómo interpretar los resultados. Cada idea va en un cuadro de color, con ejemplos resueltos paso a paso.

Actividad de introducción: ¿cuántas horas duermen los alumnos?

En una clase de 3.º de ESO de Toledo se pregunta a 25 alumnos cuántas horas durmieron anoche. Hay respuestas de 6, 7, 8 y hasta 10 horas. ¿Cómo se resume toda esa información en un solo número? Se puede calcular la media, pero si un alumno durmió solo 3 horas, la media baja y no representa a la mayoría. Quizá sea mejor mirar la mediana o la moda. ¿Y cómo saber si las respuestas son parecidas entre sí o muy dispersas? Para eso existen la desviación típica y los cuartiles. La estadística consiste en recoger datos, organizarlos en tablas y gráficos, resumirlos con medidas y, sobre todo, interpretarlos con sentido crítico, igual que hacen los medios de comunicación con las encuestas y las noticias.



Población, muestra y variables

Definición: estadística

La **estadística** estudia cómo recoger, organizar e interpretar datos. La **población** es el conjunto de todos los individuos que se quiere estudiar y la **muestra** es una parte de ella. Para que las conclusiones sean fiables, la muestra debe ser **representativa** y elegirse al azar; si se elige de forma que algunos grupos quedan favorecidos, hay un **sesgo**. Una encuesta sobre el deporte favorito hecha solo a la puerta de un gimnasio estaría sesgada.

Definición: variable estadística

Una **variable estadística** es la característica que se estudia. Puede ser:

- **Cualitativa:** no es numérica (color del pelo, deporte favorito).
- **Cuantitativa discreta:** toma valores aislados, normalmente enteros (número de hermanos).
- **Cuantitativa continua:** puede tomar cualquier valor de un intervalo (estatura,

peso, tiempo).

Tablas de frecuencias

Propiedad: tablas de frecuencias

Para organizar los datos se construye una tabla con:

- **Frecuencia absoluta** f_i: número de veces que aparece el valor. La suma es N.
- **Frecuencia relativa** $h_i = \frac{f_i}{N}$; la suma es 1. Multiplicada por 100 da el porcentaje.
- **Frecuencias acumuladas** F_i: suma de las frecuencias hasta ese valor.

Propiedad

Si hay muchos valores distintos, se agrupan en **intervalos** y se usa la **marca de clase**, que es el punto medio del intervalo, $x_i = \frac{\text{extremo inferior} + \text{extremo superior}}{2}$.

Gráficos estadísticos

Propiedad

El **diagrama de barras** se usa para variables cualitativas o discretas, el **histograma** para datos agrupados en intervalos, el **polígono de frecuencias** une los puntos medios de las barras y el **diagrama de sectores** reparte los 360° según la frecuencia: el ángulo de cada sector es $360^\circ \cdot h_i$. Si el 40 % practica fútbol, su sector mide 144°.

Medidas de centralización

Definición: medidas de centralización

- **Media aritmética:** $\bar{x} = \frac{\sum x_i \cdot f_i}{N}$. Para datos agrupados se usan las marcas de clase.
- **Mediana:** valor central cuando los datos están ordenados. Si hay un número par de datos, es la media de los dos centrales.
- **Moda:** valor con mayor frecuencia (puede haber más de una).

Observación

La media se ve muy afectada por valores extremos. Si cinco empleados cobran 1 200, 1 300, 1 250, 1 350 y 3 900 €, la media es 1 800 € y la mediana es 1 300 €: la mediana representa mejor al grupo.

Cómo se hace: calcular media, mediana y moda

1. Ordena los datos de menor a mayor.
2. Media: suma todos los datos (o los productos $x_i \cdot f_i$) y divide entre N.
3. Mediana: toma el dato central; si hay un número par de datos, la media de los dos centrales.
4. Moda: busca el valor con mayor frecuencia.

Medidas de posición y dispersión

Definición: cuartiles

Los **cuartiles** Q_1, Q_2 (mediana) y Q_3 dividen los datos ordenados en cuatro partes con el mismo número de datos. El **rango** es la diferencia entre el mayor y el menor dato. La **varianza** es la media de los cuadrados de las desviaciones respecto a la media:

Propiedad

$$\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N} = \frac{\sum f_i x_i^2}{N} - \bar{x}^2$$

Definición: desviación típica

La **desviación típica** es $\sigma = \sqrt{\sigma^2}$ y tiene las mismas unidades que los datos. Cuanto menor es, más agrupados están los datos en torno a la media. El **coeficiente de variación** $CV = \frac{\sigma}{\bar{x}}$ permite comparar la dispersión de dos conjuntos con distinta media o distintas unidades.

Cómo se hace: calcular la desviación típica

1. Calcula la media $\bar{x}$.
2. Halla las desviaciones $x_i - \bar{x}$ y elévalas al cuadrado.
3. Multiplica por las frecuencias, suma y divide entre N: es la varianza.
4. Extrae la raíz cuadrada: es la desviación típica σ.

Ejemplos resueltos

Ejemplo 1

Datos

2, 3, 3, 4, 4, 4, 5, 5, 6, 8. Suma = 44, $N = 10$, media 4,4. Moda 4. Mediana: los centrales son 4 y 4, luego 4. Rango $8 - 2 = 6$.

Ejemplo 2

Datos 2, 4, 4, 4, 5, 5, 7, 9

media $\frac{40}{8} = 5$. Cuadrados de las desviaciones: 9, 1, 1, 1, 0, 0, 4, 16, suma 32. Varianza $\frac{32}{8} = 4$ y $\sigma = 2$.

Ejemplo 3

En una tabla de intervalos [0,10), [10,20), [20,30) con frecuencias 5, 10 y 5, las marcas son 5, 15 y 25 y la media es $\frac{25+150+125}{20} = 15$.

Cómo interpretar los resultados

Observación

Calcular es solo la mitad del trabajo: hay que saber qué dicen los números. Una media de 6,2 en un examen indica el nivel general de la clase, pero si la desviación típica es grande significa que hay alumnos con notas muy altas y otros con notas muy bajas. Si dos clases tienen la misma media, la que tiene menor desviación típica es más homogénea. La mediana informa de que la mitad de los datos está por debajo y la otra mitad por encima, y la moda indica el valor más habitual. Para describir un conjunto de datos conviene dar una medida de centralización junto con una de dispersión, y acompañar los resultados de un gráfico adecuado y de una frase con las unidades.

- Calcular la mediana sin ordenar los datos.
- Calcular la media de los valores sin multiplicar por sus frecuencias.
- Olvidar la raíz cuadrada al pasar de varianza a desviación típica.
- Usar los extremos del intervalo en lugar de la marca de clase.
- Pensar que una media mayor significa siempre mayor dispersión (son medidas distintas).

Comparar dos distribuciones

Método: calcular la desviación típica

1. Calcula la media $\bar{x}$.
2. Resta la media a cada dato y eleva al cuadrado esas diferencias.

3. Suma los cuadrados (multiplicando por f_i si hay frecuencias) y divide entre el número de datos: es la varianza σ^2.
4. Calcula la raíz cuadrada de la varianza: es la desviación típica σ.

Ejemplo 4

Calcula la desviación típica de 3, 5, 5, 7.

Media $= \frac{20}{4} = 5$. Desviaciones al cuadrado: $4,\ 0,\ 0,\ 4$. Varianza $= \frac{8}{4} = 2$. Desviación típica $= \sqrt{2} \approx 1{,}41$.

Propiedad: elegir la medida adecuada

La media usa todos los datos, pero se ve muy afectada por los valores extremos. La mediana divide los datos ordenados en dos mitades y resiste mejor a los extremos. La moda es la única medida válida para variables cualitativas. Para comparar la dispersión de dos grupos con medias muy distintas se usa el coeficiente de variación $\frac{\sigma}{\bar{x}}$, que no depende de las unidades.

Ejemplo 5

Dos grupos tienen media 10 y 100, y desviaciones típicas 2 y 10. ¿Cuál es más disperso en términos relativos?

Coeficientes de variación: $\frac{2}{10} = 0{,}20$ y $\frac{10}{100} = 0{,}10$. El primer grupo es relativamente más disperso, aunque su desviación típica sea menor.

- Calcular la mediana sin ordenar antes los datos.
- Olvidar multiplicar por la frecuencia al calcular la media de una tabla.
- Pensar que la media es siempre un dato que existe en la muestra: puede no coincidir con ninguno.
- Sacar conclusiones de una muestra que no es representativa o es demasiado pequeña.

¿Sabías que?

La palabra «estadística» viene de «estado»: los gobiernos antiguos contaban a la población y sus bienes para cobrar impuestos y organizar ejércitos. Hoy, el Instituto Nacional de Estadística realiza el censo y las encuestas de población activa en España, y las empresas de internet analizan miles de millones de datos para recomendar canciones, series o productos.

Interpretar los resultados

Ejemplo 6

Dos grupos de 3.º de ESO tienen la misma nota media, 6. En el grupo A la desviación típica es 0,8 y en el grupo B es 2,5. ¿Qué podemos decir de cada grupo?

La media es igual, así que «en conjunto» los dos grupos han rendido lo mismo. Sin embargo, la desviación típica nos dice cómo se reparten las notas alrededor de esa media. En el grupo A casi todos los alumnos tienen notas cercanas a 6, entre 5 y 7 aproximadamente: es un grupo homogéneo. En el grupo B hay notas muy altas y muy bajas: es un grupo heterogéneo, con alumnos que dominan el tema y otros que necesitan refuerzo. Por eso un informe estadístico completo siempre da una medida de centralización y otra de dispersión.

¡Atención!

Antes de calcular la mediana hay que ordenar los datos. Un error muy frecuente es tomar el valor central de la lista tal como aparece en el enunciado. Recuerda también que, con datos agrupados en intervalos, la media obtenida con las marcas de clase es solo una aproximación.

El consejo de Otto

Después de calcular la media, comprueba que está entre el dato menor y el mayor. Si no es así, hay un error de cálculo.