Apuntes - Estadística

Curso: 2 Eso

La estadística organiza datos para sacar conclusiones. En este curso trabajas con datos agrupados, medidas de dispersión y gráficos.

Actividad de introducción: la encuesta del recreo

En el recreo, Irene y Hugo preguntan a 20 compañeros cuántas horas duermen cada noche. Anotan las respuestas, pero una lista de 20 números no dice nada a simple vista. ¿Cuántas horas duermen la mayoría? ¿Cuál es el valor típico? ¿Hay mucha diferencia entre unos y otros? Para responder, organizan los datos en una tabla, calculan medidas como la media y la mediana y dibujan un gráfico. Eso es la estadística: recoger datos, organizarlos, resumirlos y sacar conclusiones. En este tema aprenderás a hacerlo y a interpretar el resultado con espíritu crítico.



Diagrar
de barr

Población, muestra y variables

Definición: población y muestra

La **población** es el conjunto de todos los individuos de un estudio y la **muestra** un subconjunto de ella. Una variable es **cualitativa** o **cuantitativa** (discreta o continua). Una muestra debe ser **representativa**.

Tablas de frecuencias

Definición: frecuencias

La **frecuencia absoluta** f_i cuenta las veces que aparece un valor; la **relativa** $h_i = \frac{f_i}{N}$; las acumuladas suman las anteriores. Con datos continuos se agrupan en **intervalos** y se usa la **marca de clase** (punto medio).

Medidas de centralización

Regla: media, mediana y moda

Media: $\bar{x} = \frac{\sum x_i \cdot f_i}{N}$. **Mediana**: valor central de los datos ordenados. **Moda**: valor de

mayor frecuencia.

Ejemplo 1

Calcula media, mediana y moda de 2, 3, 3, 4, 4, 4, 5, 9.

Media $= \frac{34}{8} = 4{,}25$. Mediana: centrales 4 y 4, vale 4. Moda: 4.

Medidas de dispersión

Definición: rango y desviación típica

El **rango** es $\max - \min$. La **varianza** es $\sigma^2 = \frac{\sum f_i (x_i - \bar{x})^2}{N}$ y la **desviación típica** $\sigma = \sqrt{\sigma^2}$. Cuanto menor es, más agrupados están los datos.

Ejemplo 2

Calcula la desviación típica de 2, 4, 6.

Media 4. Varianza $= \frac{4+0+4}{3} = \frac{8}{3}$. $\sigma = \sqrt{\frac{8}{3}} \approx 1{,}63$.

Gráficos estadísticos

Propiedad: gráficos

Diagrama de barras (discretas y cualitativas), histograma (continuas agrupadas), polígono de frecuencias y diagrama de sectores, en el que el ángulo de cada sector es $360^\circ \cdot h_i$.

Ejemplo resuelto con datos agrupados

Ejemplo 3

Las alturas (cm) de 20 alumnos están en los intervalos [150,160): 4 alumnos, [160,170): 10, [170,180): 6. Calcula la media.

Marcas: 155, 165, 175.

$$\bar{x} = \frac{155 \cdot 4 + 165 \cdot 10 + 175 \cdot 6}{20} = \frac{620 + 1650 + 1050}{20} = \frac{3320}{20} = 166 \text{ cm}.$$

Ejemplo 4

Un diagrama de sectores muestra que el 35 % practica fútbol. ¿Qué ángulo tiene el sector?

$360^\circ \cdot 0{,}35 = 126^\circ$.

Método: estudio estadístico

1. Define población, muestra y variable.
2. Recoge y ordena los datos.
3. Construye la tabla y el gráfico.
4. Calcula media, mediana, moda y desviación típica.
5. Interpreta los resultados.

¡Atención!

Con datos agrupados usa siempre la marca de clase, no los extremos. Y la media se deja influir por los valores extremos, la mediana no.

Frecuencias y porcentajes

Definición: tipos de frecuencia

La **frecuencia absoluta** f_i es el número de veces que se repite un valor. La **frecuencia relativa** es $h_i = \dfrac{f_i}{n}$, donde n es el total de datos; multiplicada por 100 da el porcentaje. La **frecuencia acumulada** F_i suma las frecuencias hasta ese valor, y la última siempre es n. Estas columnas permiten comprobar los cálculos: las relativas suman 1 y los porcentajes suman 100.

Ejemplo 5

En una clase de 25 alumnos, 10 van en autobús, 5 en coche y 10 andando. Calcula las frecuencias relativas y los porcentajes.

$h_{\text{autobús}} = \dfrac{10}{25} = 0{,}4$ (40 %); $h_{\text{coche}} = \dfrac{5}{25} = 0{,}2$ (20 %); $h_{\text{andando}} = \dfrac{10}{25} = 0{,}4$ (40 %).
Suman $0{,}4 + 0{,}2 + 0{,}4 = 1$.

Elegir y comparar medidas de centralización

Propiedad: qué medida usar

La **media** usa todos los datos y es la mejor cuando los datos son simétricos, pero se ve muy afectada por los valores extremos. La **mediana** es el valor central y no se altera por los valores extremos, por lo que es mejor cuando hay datos atípicos (como sueldos o precios de viviendas). La **moda** es el único resumen posible para variables

cualitativas y representa el valor más frecuente. A veces hay más de una moda o ninguna.

Ejemplo 6

Los precios de seis pisos (en miles de €) son 120, 130, 125, 135, 128 y 600. ¿Qué medida resume mejor el precio típico?

Media: $\frac{120+130+125+135+128+600}{6} = \frac{1\,238}{6} \approx 206{,}3$ mil €. Mediana: ordenados 120, 125, 128, 130, 135, 600; $\frac{128+130}{2} = 129$ mil €. La mediana describe mejor el precio típico, porque el piso de 600 mil € eleva mucho la media.

Desviación típica paso a paso

Método: calcular la desviación típica

1. Calcula la media $\bar{x}$.
2. Resta la media a cada dato: $x_i - \bar{x}$.
3. Eleva al cuadrado cada diferencia.
4. Suma los cuadrados y divide entre el número de datos: es la varianza σ^2.
5. Haz la raíz cuadrada de la varianza: es la desviación típica σ.

¡Atención!

- Calcular la mediana sin ordenar antes los datos.
- Olvidar que con un número par de datos la mediana es la media de los dos centrales.
- Dividir entre 5 en vez de entre el total de datos al calcular una media con frecuencias: hay que dividir entre $n = \sum f_i$.
- Dar la varianza como si fuera la desviación típica: la desviación se obtiene después de hacer la raíz cuadrada.
- Concluir que dos grupos son «iguales» solo porque tienen la misma media.

Datos agrupados en intervalos

Método: media con datos agrupados

1. Calcula la marca de clase de cada intervalo: el punto medio $\frac{\text{extremo inferior} + \text{extremo superior}}{2}$.
2. Multiplica cada marca por su frecuencia absoluta.
3. Suma los productos y divide entre el total de datos.

4. El intervalo modal es el de mayor frecuencia (si todos tienen la misma amplitud).
5. La mediana está en el primer intervalo cuya frecuencia acumulada llega a $\frac{n}{2}$.

Ejemplo 7

Se agrupan los tiempos de 20 corredores: [10, 20): 4; [20, 30): 10; [30, 40): 6 minutos. Calcula la media aproximada.

Marcas de clase: 15, 25 y 35. Suma: $15 \cdot 4 + 25 \cdot 10 + 35 \cdot 6 = 60 + 250 + 210 = 520$. Media: $520 : 20 = 26$ minutos.

¿Sabías que...?

La palabra «estadística» viene de «estado»: los gobiernos recogían datos de población y cosechas para organizar impuestos y ejércitos. En España se hace un censo de población cada diez años, y el Instituto Nacional de Estadística publica datos sobre empleo, precios o población. Hoy los datos están en todas partes: en las redes, en los deportes y en las encuestas, y saber interpretarlos es una forma de pensamiento crítico.

Elegir el gráfico adecuado

Regla: gráfico según el tipo de dato

Para variables cualitativas o cuantitativas discretas con pocos valores se usan diagramas de barras o de sectores. El diagrama de sectores es útil cuando interesa comparar cada parte con el total, porque el ángulo de cada sector es $360^\circ \cdot h_i$. Para datos continuos agrupados en intervalos se usa el histograma, con barras pegadas, y para ver una evolución en el tiempo (la temperatura mes a mes) conviene un polígono o una línea. Elegir bien el gráfico hace que la información se entienda de un vistazo.

Ejemplo 8

En una clase de 30 alumnos, el 40 % practica fútbol, el 30 % baloncesto y el resto natación. ¿Qué ángulos tiene el diagrama de sectores?

Fútbol: $0{,}40 \cdot 360^\circ = 144^\circ$. Baloncesto: $0{,}30 \cdot 360^\circ = 108^\circ$. Natación: $30\,\%$ restante, 108°. Suman $144 + 108 + 108 = 360^\circ$. En número de alumnos son 12, 9 y 9.

Gráficos que engañan

Un diagrama de barras con el eje vertical cortado o con una escala distinta puede exagerar diferencias pequeñas. Antes de fiarte de un gráfico, mira siempre el origen y la escala de los ejes, la fuente de los datos y si la muestra es suficientemente grande y representativa. Esa actitud crítica es parte del trabajo del estadístico.

Comparar dos grupos de datos

Muchas veces la estadística se usa para comparar: las notas de dos clases, las temperaturas de dos ciudades o los tiempos de dos atletas. Para comparar bien no basta con mirar la media; también hay que fijarse en la dispersión, porque dos grupos pueden tener la misma media y comportarse de forma muy diferente.

Ejemplo 9

En dos clases las notas de un examen han sido: 2.º A: 4, 5, 5, 6, 10; 2.º B: 5, 6, 6, 6, 7. Compara los dos grupos.

Las medias son iguales: $30 : 5 = 6$ en los dos casos. Sin embargo, en 2.º A el rango es $10 - 4 = 6$ y en 2.º B es $7 - 5 = 2$. Las desviaciones típicas confirman la diferencia: en 2.º A los cuadrados de las desviaciones suman $4 + 1 + 1 + 0 + 16 = 22$ y $\sigma = \sqrt{4{,}4} \approx 2{,}10$; en 2.º B suman $1 + 0 + 0 + 0 + 1 = 2$ y $\sigma = \sqrt{0{,}4} \approx 0{,}63$.

Conclusión: 2.º B es un grupo mucho más homogéneo, mientras que en 2.º A hay notas muy distintas entre sí.

Observación: el coeficiente de variación

Para comparar la dispersión de datos con medias muy distintas se divide la desviación típica entre la media. Ese cociente, el coeficiente de variación, indica qué parte de la media representa la dispersión y permite comparar, por ejemplo, pesos de elefantes con pesos de ratones.

El consejo de Otto

Compara siempre dos conjuntos de datos con la media y la desviación típica: la media dice dónde está el centro y la desviación cuánto se separan los datos de él.