Apuntes - Estadística

Curso: 1 Eso

La estadística recoge, organiza y analiza datos para sacar conclusiones. Verás cómo construir tablas y gráficos y calcular las medidas más usadas.

Actividad de introducción: la encuesta de la clase

La profesora de Educación Física quiere organizar el torneo de recreo y pregunta a los 20 alumnos de 1.º A qué deporte prefieren: fútbol (8), baloncesto (5), balonmano (4) y voleibol (3). Para decidir, anota los resultados en una tabla y dibuja un diagrama de barras. Así ve a simple vista que el fútbol es lo más votado. Esta organización de los datos, su representación gráfica y los números que los resumen son lo que estudia la estadística. En este tema aprenderás a hacerlo con cualquier conjunto de datos.

Conceptos básicos de estadística

Definición: población y muestra

La **población** es el conjunto de individuos que se estudian; la **muestra** es un subconjunto representativo. Una **variable** es la característica estudiada: **cualitativa** (color, deporte) o **cuantitativa**, que puede ser discreta (nº de hermanos) o continua (altura).

Tablas de frecuencias

Definición: frecuencias

La **frecuencia absoluta** f_i es el número de veces que aparece un valor. La **frecuencia relativa** es $h_i = \frac{f_i}{N}$, con N el total de datos. La **frecuencia acumulada** suma las frecuencias hasta ese valor. La relativa multiplicada por 100 es el porcentaje.

Ejemplo 1

Se pregunta el número de hermanos a 20 alumnos: 0 (4), 1 (8), 2 (6), 3 (2). Halla las frecuencias relativas.

$h = \frac{4}{20} = 0{,}20$, $\frac{8}{20} = 0{,}40$, $\frac{6}{20} = 0{,}30$, $\frac{2}{20} = 0{,}10$. Suman 1.

Gráficos estadísticos

Propiedad: tipos de gráficos

El **diagrama de barras** se usa con variables discretas o cualitativas; el **polígono de frecuencias** une los extremos de las barras; el **histograma** se usa con datos agrupados en intervalos; el **diagrama de sectores** reparte los 360° en proporción a la frecuencia: ángulo $= 360^\circ \cdot h_i$.

Ejemplo 2

En un sector que representa el 25 %, ¿qué ángulo corresponde?

$360^\circ \cdot 0{,}25 = 90^\circ$.

Medidas de centralización y de dispersión

Definición: media, mediana y moda

La **media** $\bar{x}$ es la suma de los datos dividida entre el número de datos. La **mediana** es el valor central cuando los datos están ordenados (si hay dos centrales, su media). La **moda** es el valor con mayor frecuencia. El **rango** es la diferencia entre el mayor y el menor dato y mide la dispersión.

Ejemplo 3

Para los datos 3, 5, 5, 6, 8, 9 calcula media, mediana, moda y rango.

Media: $\frac{36}{6} = 6$. Mediana: centrales 5 y 6, $\frac{5+6}{2} = 5{,}5$. Moda: 5. Rango: $9 - 3 = 6$.

Ejemplo 4

Con la tabla de hermanos del ejemplo anterior, ¿cuál es la media?

$\bar{x} = \frac{0 \cdot 4 + 1 \cdot 8 + 2 \cdot 6 + 3 \cdot 2}{20} = \frac{26}{20} = 1{,}3$ hermanos.

Ejemplo 5

Las notas de Ana son 6, 7 y 8. ¿Qué necesita sacar para tener 7,5 de media en cuatro exámenes?

Suma necesaria: $4 \cdot 7{,}5 = 30$. Ya tiene 21, así que necesita un 9.

Método: hacer un estudio estadístico

1. Define la población y la variable.
2. Recoge los datos y haz el recuento.
3. Construye la tabla de frecuencias.
4. Representa con el gráfico adecuado.
5. Calcula las medidas y escribe una conclusión.

¡Atención!

La moda es un valor, no una frecuencia. Y la mediana exige ordenar antes los datos. La media se ve muy afectada por un valor extremo.

Cómo organizar y resumir datos

Método: elaborar una tabla de frecuencias

1. Escribe los valores distintos de la variable ordenados de menor a mayor.
2. Cuenta cuántas veces aparece cada valor: es la **frecuencia absoluta**.
3. Calcula la **frecuencia relativa** dividiendo cada frecuencia absoluta entre el total de datos.
4. Multiplica por 100 para obtener el porcentaje.
5. Comprueba que la suma de frecuencias absolutas es el total y que la suma de relativas es 1.

Ejemplo 5

Se han anotado las notas de 10 alumnos: 5, 7, 7, 6, 8, 5, 7, 9, 6, 7. Construye la tabla y halla la media, la mediana y la moda.

Valores: 5 (2 veces), 6 (2), 7 (4), 8 (1), 9 (1). Media = (10 + 12 + 28 + 8 + 9) : 10 = 6,7. Ordenados: 5, 5, 6, 6, 7, 7, 7, 7, 8, 9; mediana = (7 + 7) : 2 = 7. Moda = 7.

Ejemplo 6

En un diagrama de sectores, 45 de 180 alumnos van en bici. ¿Qué ángulo corresponde?

Fracción: 45/180 = 1/4. Ángulo: 360° : 4 = 90°.

Datos agrupados en intervalos

Definición: intervalo y marca de clase

Cuando una variable continua toma muchos valores distintos (alturas, pesos, tiempos), se agrupan los datos en **intervalos** de igual amplitud, por ejemplo [150, 160). La **marca de clase** es el punto medio del intervalo: para [150, 160) es 155. Se

usa para calcular la media aproximada.

Ejemplo 7

Las alturas de 20 alumnos se agrupan así: [140, 150) 4 alumnos; [150, 160) 10 alumnos; [160, 170) 6 alumnos. Calcula la media aproximada.

Marcas de clase: 145, 155 y 165. Media = (4·145 + 10·155 + 6·165) : 20 = (580 + 1 550 + 990) : 20 = 3 120 : 20 = 156 cm. El intervalo modal es [150, 160).

Regla: elegir el gráfico adecuado

El **diagrama de barras** sirve para comparar frecuencias de variables cualitativas o discretas. El **diagrama de sectores** sirve para ver la proporción de cada parte respecto al total. El **gráfico de líneas** es ideal para ver cómo cambia una magnitud en el tiempo. El **histograma** se usa para datos agrupados en intervalos.

Propiedad: media, mediana y moda

La media usa todos los datos y se ve muy afectada por los valores extremos. La mediana no se deja influir por los valores extremos. La moda es el único estadístico que se puede calcular con datos cualitativos. Si hay un número par de datos, la mediana es la media de los dos centrales.

¡Atención! Errores frecuentes

- Calcular la mediana sin ordenar antes los datos.
- Dividir entre el número de valores distintos en vez de entre el total de datos al calcular la media.
- Confundir frecuencia absoluta con relativa.
- Dibujar un diagrama de barras con una escala que no empieza en cero y exagera las diferencias.

¿Sabías que...?

La palabra «estadística» viene de «estado»: los primeros datos que se recogieron fueron los censos de población que hacían los gobernantes para cobrar impuestos. La enfermera inglesa Florence Nightingale, pionera de la estadística, usó gráficos de sectores llamados «rosas» para convencer al gobierno de que mejoraran los hospitales militares. Hoy el Instituto Nacional de Estadística (INE) publica cada año datos sobre la población de España.

Interpretar y criticar la información estadística

La estadística no consiste solo en hacer tablas y calcular medias. En los periódicos, en la

televisión y en las redes sociales aparecen continuamente gráficos y porcentajes, y hay que saber leerlos con espíritu crítico. Un mismo conjunto de datos puede parecer muy distinto según cómo se presente.

Método: leer un gráfico con atención

1. Lee el título y averigua qué variable se representa y en qué población.
2. Mira las unidades de los ejes y comprueba si el eje vertical empieza en 0.
3. Busca el valor más alto y el más bajo y compáralos.
4. Comprueba si los porcentajes suman 100 % o si los ángulos suman 360°.
5. Escribe una o dos conclusiones con frases completas.

Ejemplo 8

Un anuncio dice: «Las ventas de nuestra bicicleta se han disparado», y muestra un gráfico de barras con un eje vertical que empieza en 950 unidades. Las barras indican 960 ventas en enero y 990 en febrero. ¿Es cierto que se han disparado?

Como el eje no empieza en 0, la barra de febrero parece cuatro veces más alta que la de enero. Sin embargo, el aumento real es de $990 - 960 = 30$ bicicletas, que respecto a 960 supone poco más de un 3 %. El gráfico exagera la subida. Conclusión: las ventas han aumentado ligeramente, no se han disparado.

Ejemplo 9

En una pequeña empresa los sueldos mensuales son 1 200, 1 200, 1 300, 1 400 y 5 900 euros. ¿Qué medida representa mejor el sueldo típico?

La media es $(1\,200 + 1\,200 + 1\,300 + 1\,400 + 5\,900) : 5 = 11\,000 : 5 = 2\,200$ €. Pero cuatro de los cinco trabajadores cobran mucho menos. La mediana, que es el valor central de los datos ordenados, vale 1 300 €, y la moda, 1 200 €. Un único valor muy alto, llamado valor atípico, hace subir la media. En este caso la mediana describe mejor el sueldo típico.

Observación: muestras representativas

Para que las conclusiones de una encuesta sean válidas, la muestra debe parecerse a la población. Si para conocer el deporte favorito de los alumnos de un instituto solo preguntas a los del equipo de baloncesto, el resultado estará sesgado. Lo correcto es elegir alumnos de todos los cursos y grupos, por ejemplo al azar.

El consejo de Otto

Para calcular la media con una tabla, añade una columna $x_i \cdot f_i$, suma y divide entre N. Comprueba siempre que las frecuencias relativas suman 1.