RESUMEN DE CONTENIDOS

cREAgal

Introducción

Este resumen contiene los contenidos de 2º de ESO del tema de estadística.

Si necesitas los conceptos estudiados en 1º de ESO, accede al "Repositorio de contidos educativos", y bájalos del REA del curso pasado "Ocho apelli2 gallegos".

3. Elige tu propio camino

3.1. Mejor en grupo

En Estadística, cuando tenemos mucha información, también es conveniente hacer grupos para resumirla.
Las **clases** son los intervalos en que se agrupan los datos de una variable estadística, cuando el número de datos es grande y ninguno de los datos se repite. Esto ocurre con las variables cuantitativas continuas.

Variable cuantitativa continua, es aquella que en un determinado intervalo puede tomar cualquier valor. En teoría podría tomar infinitos valores.

La **marca de clase**, x_i , es el punto medio de la clase. Se obtiene sumando sus extremos y dividiendo entre 2.

La **frecuencia absoluta** es el número de valores que están en cada uno de los intervalos.

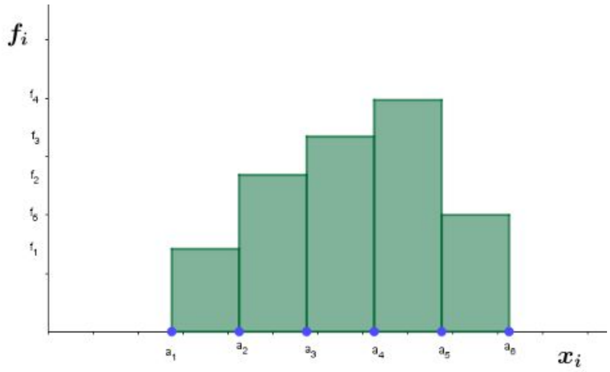
Las clases, marcas de clase y la frecuencia absoluta suelen expresarse mediante una **tabla de frecuencias**. En ella se ordenan todos los datos.

La tabla quedará como sigue:

Intervalo	**Marca de clase** x_i	**Frecuencias absolutas** f_i
$[(a_1, a_2)$	x_1	f_1
$[(a_2, a_3)$	x_2	f_2
...	...	...
$[a_n, a_{n+1}]$	x_n	f_n
		$N=\sum f_i$

3.2. Consultando el mapa

El **histograma** es un gráfico estadístico que se utiliza para representar datos agrupados en clases o intervalos, normalmente valores de variables cuantitativas continuas.



En el eje horizontal, eje de abscisas, se representan las clases, y en el eje vertical, eje de ordenadas, las frecuencias absolutas.

Finalmente, se dibujan los rectángulos que tienen de base los intervalos y de altura la frecuencia absoluta correspondiente.

3.3. El camino de invierno

Para resumir los datos de una variable estadística cuantitativa se utilizan distintos cálculos. Son los parámetros estadísticos.

Los que nos dan una idea de valor alrededor del cual están agrupados se llaman **parámetros de centralización**.

La **media** es un parámetro de centralización cuya fórmula es:

$\bar{x}=\frac{\sum x_i \cdot f_i}{N}$, donde N es el número de datos, $N=\sum f_i$.

Para hacer este cálculo de forma más sencilla se suele añadir una nueva columna en la tabla del apartado anterior con los productos $x_i \cdot f_i$.

La tabla quedará como sigue:

Intervalo	**Marca de clase** x_i	**Frecuencias absolutas** f_i	$x_i \cdot f_i$
$[(a_1, a_2)$	x_1	f_1	$x_1 \cdot f_1$
$[(a_2, a_3)$	x_2	f_2	$x_2 \cdot f_2$
...	...	...	...
$[a_n, a_{n+1}]$	x_n	f_n	$x_n \cdot f_n$
		$N=\sum f_i$	$\sum x_i \cdot f_i$

3.4. Cercanos pero distintos

Las **medidas de dispersión** nos dan información sobre cómo de dispersos están los datos en una variable estadística.

Las medidas de dispersión más usadas son el **rango o recorrido**, la **desviación media**, la **varianza**, la **desviación típica** y el **coeficiente de variación**.

El **rango o recorrido, R,** mide la diferencia entre el valor más grande de una variable y el más pequeño.

Por tanto, el rango nos dice cuál ha sido la variación máxima en la variable aleatoria.

$$R = x_{max} - x_{min}$$

Si los datos están **agrupados** se calcula restando el extremo superior del último intervalo menos el extremo inferior del primero, es decir:

$$R = L_{max} - L_{min}$$

La **desviación media, D_m,** de un conjunto de datos de una variable es una medida de dispersión que consiste en el promedio de las distancias de cada dato a la media. Viene dada por la siguiente fórmula:

$$D_m = \frac{\sum |x_i - \overline{x}| \cdot f_i}{N}$$

Para facilitar su cálculo se suele añadir dos columnas en la tabla del apartado anterior una con $|x_i - \overline{x}|$, y otra con $|x_i - \overline{x}| \cdot f_i$.

La tabla quedará como sigue:

Intervalo	**Marca de clase** x_i	**Frecuencias absolutas** f_i	$x_i \cdot f_i$	$\lvert x_i - \overline{x} \rvert$	$\lvert x_i - \overline{x} \rvert \cdot f_i$
$[(a_1, a_2)$	x_1	f_1	$x_1 \cdot f_1$	$\lvert x_1 - \overline{x} \rvert$	$\lvert x_1 - \overline{x} \rvert \cdot f_1$
$[(a_2, a_3)$	x_2	f_2	$x_2 \cdot f_2$	$\lvert x_2 - \overline{x} \rvert$	$\lvert x_2 - \overline{x} \rvert \cdot f_2$
...	...	...	...	...	...
$[a_n, a_{n+1}]$	x_n	f_n	$x_n \cdot f_n$	$\lvert x_n - \overline{x} \rvert$	$\lvert x_n - \overline{x} \rvert \cdot f_n$
		$N = \sum f_i$	$\sum x_i \cdot f_i$		$\sum \lvert x_i - \overline{x} \rvert \cdot f_i$

La **varianza**, que se representar por σ^2 o Var, es una medida de dispersión que consiste en calcular la media de las desviaciones de cada uno de los datos con respecto a la media elevada al cuadrado. Es decir: $\sigma^2 = Var = \frac{\sum (x_i - \overline{x})^2 \cdot f_i}{N}$

Para facilitar su cálculo se suele añadir una nueva columna con $(x_i - \bar{x})^2$.

Las unidades de la **varianza** no coinciden con las unidades de la variable que estamos estudiando, por lo que su interpretación es más difícil, para evitar esto se utiliza la **desviación típica**. La tabla quedará como sigue:

Intervalo	**Marca de clase** x_i	**Frecuencias absolutas** f_i	$x_i \cdot f_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
$[(a_1, a_2)$	x_1	f_1	$x_1 \cdot f_1$	$(x_1 - \bar{x})^2$	$(x_1 - \bar{x})^2 \cdot f_1$
$[(a_2, a_3)$	x_2	f_2	$x_2 \cdot f_2$	$(x_2 - \bar{x})^2$	$(x_2 - \bar{x})^2 \cdot f_2$
...	...	...	...	...	...
$[a_n, a_{n+1}]$	x_n	f_n	$x_n \cdot f_n$	$(x_n - \bar{x})^2$	$(x_n - \bar{x})^2 \cdot f_n$
		$N = \sum f_i$	$\sum x_i \cdot f_i$		$\sum (x_i - \bar{x})^2 \cdot f_i$

La **desviación típica**, (también conocida como desviación estándar) se representa con la letra griega σ y es la raíz cuadrada de la varianza. Tendremos entonces:

$$\sigma = \sqrt{Var}$$

Como todas las medidas de dispersión nos indican como de separados están los datos con respecto a la media.

- Si la desviación típica es **alta**, los datos están más dispersos en torno a la media.
- Si la desviación típica es **baja**, los datos están más concentrados en torno a la media.

El **coeficiente de variación**, que representaremos como CV, es una medida de dispersión que consiste dividir la desviación típica entre la media de un conjunto de valores. Normalmente se expresa en tanto por ciento, es decir:

$$CV = \frac{\sigma}{\bar{x}} \cdot 100$$

El coeficiente de variación se suele utilizar para comparar dos o más conjuntos de datos.

Proporciona una visión más completa y precisa de la dispersión relativa en cada conjunto de datos, permitiendo una comparación adecuada, que no es posible utilizando solo la desviación típica.

Atribución de los recursos incorporados al documento

Las imágenes que figuran en este documento son de elaboración propia (proxecto cREAgal).