

Tema 4

Medidas de dispersión

Contenido

4.1. Introducción	1
4.2. Medidas de dispersión absolutas	1
4.3. Medidas de dispersión relativas	3

4.1. Introducción

Las **medidas de dispersión** están encaminadas a cuantificar lo próximos o alejados que están los datos de la muestra de un punto central.

Estas medidas indicarán por un lado el **grado de variabilidad** que hay en la muestra y, por otro, la **representatividad** de dicho **punto central**, ya que si se obtiene un valor pequeño, eso significará que los valores se concentran entorno a ese centro (por lo que habrá poca variabilidad y el centro representará bien a todos). En cambio, si se obtiene un valor grande, significará que los valores no están concentrados, sino dispersos (por lo que habrá mucha variabilidad y el centro no será muy representativo).

4.2. Medidas de dispersión absolutas

Se tomará como punto central de referencia la media aritmética, aunque de forma análoga podría utilizarse la mediana (posteriormente, se introducirán algunas medidas de dispersión relacionadas con la mediana, ver Tema 7).

En primer lugar se buscará una forma de medir la dispersión en términos absolutos (i.e., dependiendo de las unidades en las que se mida la variable).

Matemáticamente la distancia de un valor a otro es la diferencia entre esos valores sin signo, es decir, en valor absoluto. Sin embargo, el valor absoluto tiene muy malas propiedades matemáticas. En cambio, si se elevan las distancias al cuadrado, también “se quita” el signo y eso sí que tiene buenas propiedades matemáticas. Así surge la **varianza**, que sirve igualmente para medir la dispersión conjunta en torno a la media, pero es más operativa.

El resultado se mide en unidades al cuadrado y es difícil de interpretar, por lo que lo habitual es calcular su raíz cuadrada (y surge así lo que se llama **desviación típica**).

Matemáticamente, para calcular la varianza y la desviación típica en general, que se denotan respectivamente S_x^2 y S_x , se puede operar como sigue a partir de la tabla de frecuencias de las distancias al cuadrado:

Distancias ²	n_i	Distancias ² $\times$ n_i
$(x_1 - \bar{x})^2$	n_1	$(x_1 - \bar{x})^2 \times n_1$
$\vdots$	$\vdots$	$\vdots$
$(x_k - \bar{x})^2$	n_k	$(x_k - \bar{x})^2 \times n_k$
Total	N	$\sum_{i=1}^k (x_i - \bar{x})^2 n_i$

La varianza es la media de las distancias al cuadrado y la desviación es su raíz, es decir:

$$S_x^2 = \frac{\text{suma de todas las distancias}^2}{\text{número de individuos}} = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 n_i}{N}.$$

$$S_x = \sqrt{S_x^2}.$$

Puede comprobarse que la varianza se puede calcular alternativamente como la media de los valores de la variable al cuadrado menos la media al cuadrado de la variable, es decir, $S_x^2 = \overline{x^2} - \bar{x}^2$. Esta segunda fórmula suele resultar más cómoda a la hora de hacer los cálculos.

Problemas propuestos: Problemas 4.1 y 4.2.

4.3. Medidas de dispersión relativas

A la hora de establecer comparaciones no es aconsejable manejar magnitudes absolutas, ya que las unidades no son siempre comparables. Cuando se pretende comparar la dispersión de variables medidas en distintas unidades o variables con distinto orden de magnitud, es necesario relativizar.

Una forma de relativizar es considerar la dispersión en relación al valor absoluto de la media, consiguiendo así el **coeficiente de variación**, que suele ser interpretado en términos de proporción o porcentaje:

$$CV_x = \frac{S_x}{|\bar{x}|}.$$

El coeficiente de variación se suele utilizar con variables que toman valores positivos (precios, rentas, etc.) y cuya media no está próxima al 0, ya que en caso contrario al dividir por un valor muy cercano a 0, el resultado puede desvirtuarse.

Tampoco se debe utilizar la desviación para comparar la dispersión de dos variables medidas en las mismas unidades (a no ser que el valor de la media sea el mismo), porque, por ejemplo, no es lo mismo 1 cm de dispersión en relación a una media de 170 cm que 1 cm de dispersión en relación a una media de 2 cm.

Como la dispersión también se utiliza para medir la representatividad de la media, el CV también se puede utilizar para *comparar* la representatividad de dos medias.

Por otro lado, como el valor de CV es una cantidad relativa, ayuda a *valorar* la dispersión/representatividad de la media, en el sentido de establecer si es *grande* o *pequeña*. Valores bajos del CV indicarán poca dispersión/mucha representatividad y valores altos indicarán mucha dispersión/poca representatividad.

No hay criterios universales para decir que un valor del CV es “bajo” o “alto”, aunque en la práctica se suelen considerar bajos los valores inferiores al 30 o 40 %, moderados entre esas cantidades y aproximadamente el 80 % y cuando se superan el 120 o 140 % ya se considera que la dispersión es bastante elevada.

En resumen, se utiliza el CV para variables positivas cuya media no está próxima a 0 con el objetivo de medir la dispersión entorno a la media y/o su representatividad cuando se deben establecer comparaciones o valoraciones.

Problemas propuestos: Problemas 4.3 y 4.4.