

Tema 3

Medidas de tendencia central

Contenido

3.1. Introducción	1
3.2. Media aritmética	2
3.3. Media ponderada	3
3.4. Media geométrica	4
3.5. Mediana	5
3.5.1. Cálculo de la mediana para datos agrupados	5
3.6. Moda	6

3.1. Introducción

En ocasiones es conveniente resumir la información de una muestra (que se representa mediante las distribuciones de frecuencias vistas anteriormente) en un solo valor para obtener indicadores del comportamiento de la variable en diferentes sentidos, como punto alrededor del que toma valores, variabilidad, etc.

Resumir la información mediante un solo número es interesante para **comprender** mejor cómo se comporta la variable y para poder **realizar comparaciones**.

En este capítulo se considerarán las **medidas de tendencia central** más habituales. La idea de centro de una distribución no es única, aunque en términos generales se puede decir que se trata de encontrar un **punto alrededor del cuál tome valores la variable**.

3.2. Media aritmética

En general, la **media aritmética** (o simplemente media si no ha lugar a confusión) consiste en repartir o dividir la suma de todos datos entre el número de individuos y se denotará por $\bar{x}$. Matemáticamente, si se tiene una variable genérica X con distribución de frecuencias

x_i	n_i
x_1	n_1
$\vdots$	$\vdots$
x_k	n_k
Total	N

y se denota por $x_T = \sum_{i=1}^k x_i n_i$ el total de la variable en la muestra, la media aritmética $\bar{x}$ sería

$$\bar{x} = \frac{\text{suma de todos los valores}}{\text{número de valores}} = \frac{\text{suma de modalidades por número de repeticiones}}{\text{número de valores}}$$

$$= \frac{x_1 n_1 + \dots + x_k n_k}{N} = \frac{\sum_{i=1}^k x_i n_i}{N} = \frac{x_T}{N}.$$

La media se mide en las mismas unidades que la variable (artículos, euros, kg, ...). Es un valor que representa el centro de la distribución, luego tiene que estar entre el valor mínimo y el máximo. Además, no tiene por qué coincidir con ningún valor exacto de la variable, ya que simplemente representa un punto alrededor del cuál se mueve esa variable (**nunca se debe redondear** para conseguir un valor de la variable).

Problema propuesto: Problema 3.1.

Si se dispone de **datos agrupados**, la media aritmética se aproxima resumiendo la información de cada clase a través de las respectivas marcas y aplicando la fórmula anterior sobre dichas marcas.

Por otro lado, si se conocen únicamente las **frecuencias relativas** (y no se conoce el tamaño de la muestra para poder calcular las absolutas) se puede calcular la media teniendo en cuenta que, dado que $f_i = n_i/N$ y $\bar{x} = \sum_{i=1}^k x_i n_i/N$, se tiene que $\bar{x} = \sum_{i=1}^k x_i f_i$.

La media aritmética es la medida de tendencia central **más utilizada** cuando la variable es de tipo **cardinal**. Es la más **operativa**, porque es muy sencilla de calcular y tiene muy buenas propiedades matemáticas. Es muy **informativa**, porque tiene en cuenta la magnitud de todos los valores de la muestra. Sin embargo, presenta dos inconvenientes fundamentales. El primero es que **no se puede calcular** para datos agrupados **cuando no se conoce el límite del último intervalo**. El segundo es que **está muy influenciada por los posibles valores extremos**.

En general, cuando la distribución tenga **datos extremos**, **no** se utiliza la **media** como medida de tendencia central. Posteriormente se verán medidas más adecuadas (*medidas robustas*, ver Tema 7).

Problema propuesto: Problema 3.2.

Hay situaciones en las que se conoce la media (o el total) en dos muestras y, sin embargo, no se dispone de todos los datos. En estas ocasiones se debe prestar atención a la hora de calcular la media conjunta, cómo se ilustra en el siguiente ejemplo.

En general, la fórmula para la media conjunta es

$$\bar{x} = \frac{\bar{x}_a \times N_a + \bar{x}_b \times N_b}{N_a + N_b}.$$

Problemas propuestos: Problemas 3.3, 3.4 y 3.5.

3.3. Media ponderada

La media ponderada es análoga a la media aritmética, pero se utiliza cuando los **valores tienen una importancia que no tiene que ver con sus frecuencias**.

En general, las importancias suelen denotarse como w_i y se denominan **ponderaciones**.

La fórmula de la **media ponderada** es, en general:

$$\bar{x}_w = \frac{\text{suma de todos los valores por sus ponderaciones}}{\text{suma de las ponderaciones}} = \frac{\sum_{i=1}^k x_i w_i}{\sum_{i=1}^k w_i}.$$

Problema propuesto: Problema 3.6.

3.4. Media geométrica

Para calcular la media aritmética se suman todos los datos de la muestra, lo que presupone un carácter aditivo de la variable (por ejemplo, 3 artículos + 5 artículos hacen un total de 8 artículos). Sin embargo, hay **variables que presentan variaciones acumulativas**, por lo que ni la suma ni la media tienen un sentido real (por ejemplo, una rebaja del 50 % sobre otra rebaja del 50 % no hacen en total una rebaja del 100 %). En estos casos se utilizará la media geométrica.

La media geométrica representa la **tasa** (también llamada razón o proporción) **media de variación**. Si se tienen los valores absolutos (con unidades, como pueden ser los precios), la **tasa o razón de variación en un momento t** se calcula como el cociente (ver el Apéndice 1 del Tema 1)

$$\frac{\text{valor absoluto después de } t}{\text{valor absoluto antes de } t}.$$

Dependiendo del problema, se conocen las tasas directamente o se obtienen a partir de los datos.

El **cálculo** de la **media geométrica** se puede hacer también utilizando tablas. Únicamente hay que tener en cuenta que para hallar el total aquí se *multiplican* los efectos, en vez de sumarlos, que era lo que se hacía con la media aritmética, luego habrá que corregir los papeles que juegan las distintas operaciones matemáticas en las tablas.

En primer lugar, se debe **determinar** a partir de los datos del problema **todas las tasas** o razones de crecimiento entre dichos datos (ordenados) y representarlos en una tabla, junto con sus frecuencias. Se denotarán genéricamente las tasas por t_i . A continuación se deben multiplicar todas las razones de crecimiento, por lo que, de forma análoga al caso de la media aritmética, se debe crear la columna $\overbrace{t_i \dots t_i}^{n_i} = t_i^{n_i}$, es decir,

Tasas de variación	n_i	$t_i^{n_i}$
t_1	n_1	$t_1^{n_1}$
$\vdots$	$\vdots$	$\vdots$
t_k	n_k	$t_k^{n_k}$
Total	N	$t_1^{n_1} \dots t_k^{n_k}$

$$\text{Tasa de variación total} = t_1^{n_1} \dots t_k^{n_k}$$

$$\bar{x}_G = \text{media geométrica} = \text{tasa media de variación}$$

$$= \sqrt[N]{\text{Tasa de variación total}} = \sqrt[N]{t_1^{n_1} \cdots t_k^{n_k}}.$$

Problemas propuestos: Problemas 3.7, 3.8 y 3.9.

3.5. Mediana

La **mediana** es un punto que divide la muestra ordenada (de menor a mayor) en dos grupos iguales (deja el 50 % de los valores por debajo y el otro 50 % por encima) y se suele denotar por Me. **Es un punto que está justo en mitad de la muestra**, y no tiene por qué coincidir con la media.

El *método de cálculo* general sería entonces el siguiente; se busca el valor 0,5 en la columna de frecuencias relativas acumuladas:

- si 0,5 no está en la tabla, la mediana Me es el primer valor para el que F_i sobrepasa el 0,5.
- si 0,5 está en la tabla, la mediana Me es cualquier valor entre el x_i correspondiente y el siguiente. Por ejemplo, se puede tomar el punto medio.

La mediana se puede calcular tanto para variables cardinales como ordinales, y es **la que más se utiliza** para variables **ordinales**. Es **menos operativa que la media** desde el punto de vista matemático. No es tan informativa como la media, ya que en su cálculo interviene sólo el orden de los valores y no su magnitud. Sin embargo, por este mismo motivo **no está influenciada por la magnitud de los valores extremos** (es una *medida robusta*) y por eso, se puede utilizar cuando la media no sea representativa (ver Tema 7).

Problema propuesto: Problema 3.10.

3.5.1. Cálculo de la mediana para datos agrupados

Si se dispone de una tabla de frecuencias con datos agrupados, existen dos formas de calcular la mediana de forma aproximada:

- Resumir cada intervalo por la marca de clase y aplicar sobre ellas el método anterior para datos aislados.
- Suponer que los individuos están repartidos uniformemente en cada clase e interpolar utilizando triángulos semejantes.

El segundo método suele ser más exacto que el primero, por lo que será el que se utilice en general.

En general, el intervalo que contiene a la mediana es el primero para el que F_i supere el 50 %. Si aparece $F_i = 0,5$ en la tabla, es más sencillo, ya que al coincidir el extremo superior de ese intervalo con el inferior del siguiente, sólo podría ser que la mediana estuviese justo en ese número.

Después de localizar el intervalo que contiene a la mediana, se interpola, utilizando triángulos semejantes, para hallar el punto que correspondería con $F_i = 0,5$ *si los datos estuviesen uniformemente repartidos en ese intervalo*.

Problemas propuestos: Problemas 3.12 y 3.13.

3.6. Moda

La **moda** representa el valor (o valores) más frecuente de la distribución. Es decir, el que tenga un n_i o un f_i más alto. Se suele denotar por Mo y se puede calcular para todo tipo de distribuciones, pero como no resulta muy informativa, sólo se utilizará cuando no sea posible usar la media ni la mediana (por ejemplo, para variables nominales).

Problema propuesto: Problema 3.14.