INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

ÍNDICE

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

T3. LAS MEDIDAS DE POSICIÓN EN DISTRIBUCIONES UNIDIMENSIONALES

3.1. Introducción

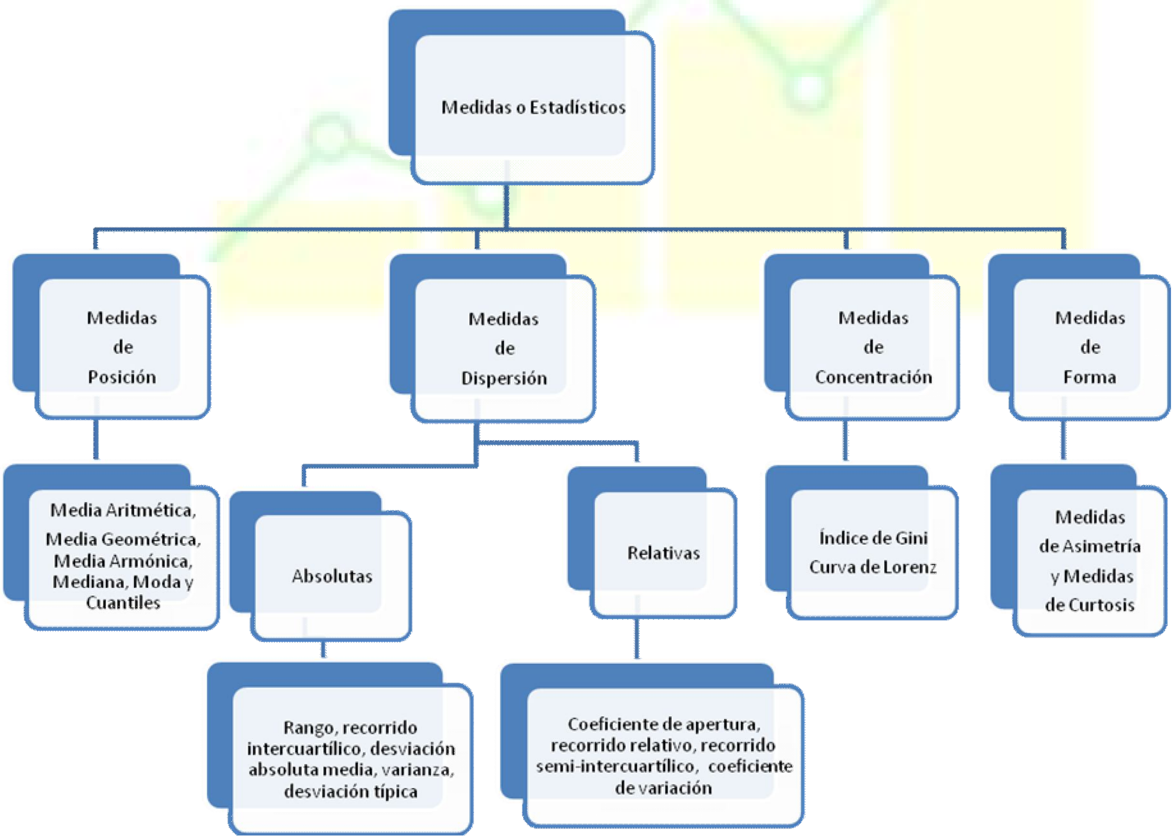
Las distribuciones de frecuencias de una variable estadística pueden estudiarse a través de unas medidas, que se conocen con el nombre genérico de estadísticos y que, analizadas conjuntamente, nos dan un panorama sobre las características de la distribución.

Son números que, al igual que las tablas o las gráficas vistas anteriormente, pretenden resumir la información recogida a la vez que ponen de manifiesto las principales características de una distribución.

Aunque la clasificación no es definitiva ni excluyente veremos:

- Medidas de posición
- Medidas de dispersión
- Momentos
- Medidas de forma
- Medidas de concentración

Los más habituales son:



INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

3.2. La media aritmética

3.2.1. La media aritmética simple

Suele denotarse por:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{N} = \frac{\sum_{i=1}^{N} Xi}{N}$$

3.2.2. La media aritmética ponderada por las frecuencias

En las distribuciones de tipo II o de tipo III es necesario utilizar las frecuencias para obtener la media aritmética.

$$\bar{x} = \frac{\sum_{i=1}^{N} Xi \cdot n_i}{N} = \sum_{i=1}^{N} Xi \cdot f_i$$

En el caso de que los datos estén agrupados en clases, es decir, en las distribuciones de tipo III, se opera igual, tomando la marca de clase m_i como x_i.

$$\bar{x} = \frac{\sum_{i=1}^{N} n_i m_i}{N}$$

3.2.3. La media aritmética ponderada por coeficientes

En ocasiones resulta conveniente introducir un coeficiente de ponderación que de mayor peso a algunos valores de la variable. Estos coeficientes o pesos suelen denominarse w_i

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i w_i}{\sum_{i=1}^{N} w_i}$$

Para las distribuciones de tipo II y III sería:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i n_i w_i}{\sum_{i=1}^{N} n_i w_i}$$

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

3.2.4. Propiedades de la media aritmética

Propiedades:

- Si a todos los valores observados les sumamos una constante K (cambio de origen), la media de los nuevos valores se obtiene sumando a la media de los valores originales esta constante K.

$$X_A = \{x_1, x_2, \ldots, x_n\} \Rightarrow \overline{X}_A$$

$$X_B = \{x_1 + k, x_2 + k, \ldots, x_n + k\} \Rightarrow \overline{X}_B = \overline{X}_A + k$$

- Si todas las observaciones se multiplican o se dividen por una constante K (cambio de escala), la nueva media queda multiplicada o dividida por dicho número K.

Como consecuencia de estas dos últimas propiedades, sí a la variable estadística x_i la sometemos al mismo tiempo a un *cambio de origen* O_t y a un *cambio de escala C* mediante la transformación $y_1 = \frac{x_1 - O_1}{c}$, (siendo O_t y C constantes), resulta que: $\bar{x} = C_{\bar{y}} + O_1$

Esta propiedad es bastante utilizada para la simplificación de los cálculos cuando los valores observados son muy elevados y tienen un máximo común divisor.

- La media de una constante es la misma constante. Es centro de gravedad de la distribución. La suma de las desviaciones de todos los valores respecto a su media aritmética es cero

$$\sum_{i=1}^{r} \left(Xi - \overline{X}\right) = 0.$$

- La media es la cantidad que hace mínima la suma de los cuadrados de las desviaciones respecto a un valor.

$\sum_{i=1}^{n} \left(Xi - \overline{X}\right)^2$: esta diferencia es mínima cuando la constante es la media aritmética.

3.2.5 Ventajas e inconvenientes de la media aritmética

Ventajas

- Se trata de un concepto familiar para la mayoría de las personas y es intuitivamente claro.
- Es calculable en todas las variables, es decir siempre que nuestras observaciones sean cuantitativas.
- Para su cálculo se utilizan todos los valores de la distribución.
- Es única para cada distribución de frecuencias.
- Tiene un claro significado, ya que al ser el centro de gravedad de la distribución representa todos los valores observados.
- Es útil para llevar a cabo procedimientos estadísticos como la comparación de medias de varios conjuntos de datos.

Inconvenientes

- Que es un valor muy sensible a los valores extremos, con lo que n las distribuciones con gran dispersión de datos puede llegar a perder totalmente su significado.
- Que no es calculable cuando los parámetros son cualitativos.
- Podemos tener dificultades para su cálculo en distribuciones de tipo III con intervalos abiertos; en estos casos es necesario estimar una marca de clase para poder calcular la media y ésta nos varía sí cambiamos la marca de clase.

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

1.2. Media geométrica

Es la raíz de índice N del producto de las observaciones elevado a sus respectivas frecuencias.

- Para distribuciones unitarias o distribuciones de tipo I:

$$G = \sqrt[N]{x_1 x_2 \dots x_N} = \sqrt[N]{\prod_{i=1}^{N} x_i}$$

- En distribuciones no unitarias:

$$G = \sqrt[N]{x_1^{n_1} x_2^{n_2} \dots x_r^{n_r}} = \sqrt[N]{\prod_{i=1}^{N} x_i^{n_i}}$$

Sólo se puede calcular si no hay observaciones nulas, también puede no tener sentido su cálculo cuando algún valor es negativo, ya que podemos obtener números irracionales.

Debe emplearse cuando los valores e la variable no son de naturaleza aditiva (tasas, tipos de interés, porcentajes, números índices, etc.)

Ventajas

- En su determinación intervienen todos los valores de la distribución.
- Es menos sensible que la media aritmética cuando la distribución tiene valores extremos.
- Es más representativa que la media aritmética cuando la distribución evoluciona de forma acumulativa o con efectos multiplicativos.
- Cuando la distribución no tiene valores nulos, su valor está definido de forma objetiva y es único.

Inconvenientes

- Su significado es menos intuitivo que la media aritmética.
- La mayor complicación de los cálculos.
- Su indefinición (da números con naturaleza imaginaria) cuando tiene valores negativos y su valor nulo cuando una observación toma este valor.

Exigirá normalmente la utilización de logaritmos o de programas informáticos.

$$\log G = \frac{1}{N} \sum_{i=1}^{r} n_i \log x_i$$

Y en neperianos:

$$\ln G = \frac{1}{N} \sum_{i=1}^{r} n_i \ln x_i$$

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

1.3. Media armónica

La media armónica de N observaciones es la inversa de la media de las inversas de las observaciones; suele denotarse con la letra H.

Para distribuciones unitarias o de tipo I:

$$H = \frac{N}{\sum_{i=1}^{n} \frac{1}{X_i}}$$

Para distribuciones de tipo II:

$$H = \frac{N}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \cdots + \frac{n_r}{x_r}} = \frac{N}{\sum_{i=1}^{r} \frac{n_i}{x_i}}$$

Su utilización es bastante poco frecuente y sólo debe emplearse cuando la variable está medida en unidades relativas, por ejemplo, Km/h., es decir, para promediar velocidades, tiempos, rendimientos, etc.

Ventajas

- Está definida de forma objetiva y es única.
- Para su cálculo tiene en cuenta todos los valores de la distribución.
- Es más representativa que otras medidas en los casos de obtener promedios de velocidades, rendimientos, productividades, etc.
- Los valores extremos tienen una menor influencia que en la media aritmética.

Inconvenientes

- Sólo se puede calcular si no hay observaciones iguales a cero.
- Cuando la variable toma algunos valores muy pequeños puede carecer de significado.

3.5 Relación entre las medias armónica, geométrica y aritmética

$$H \leq G \leq \overline{X}$$

3.6 La mediana

Las medias estudiadas hasta ahora son medidas que tratan de equilibrar los valores de una distribución compensando los más grandes con los más pequeños para buscar su centro de gravedad o posicionamiento central; estos estadísticos tienen 2 problemas:

- Son muy sensibles a los valores extremos de las distribuciones de forma que cuando existe mucha dispersión los hacen poco representativos. Cálculo de la **mediana**.

- No es posible calcularlos en las distribuciones cualitativas, para solucionarlo se calcula la **moda**.

La **mediana** de una distribución de frecuencias, previamente ordenada en orden creciente o decreciente, se define como el valor central de la variable que divide la distribución en dos partes iguales, es decir, es el valor que deja el mismo número de observaciones o de frecuencias a su izquierda que a su derecha.

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

- Cálculo de la mediana en el caso de distribuciones de tipo I

Es el valor que quedaría en medio de la distribución si ordenásemos de menor a mayor todas las observaciones. La mediana deja un 50% de las observaciones a cada lado.

- Si N es par la mediana se calcula como la media aritmética de los dos valores centrales. La media aritmética de los valores N/2 y (N/2)+1

$$Me = \frac{x\left(\frac{N}{2}\right) + x\left(\frac{N}{2}+1\right)}{2}$$

Ejemplo: $X = \{1,2,5,7,9,10,13,14\}$

N=8 (par) $\Rightarrow$ N/2=4 (número que ocupa la 4ª posición, en nuestro ej. es el 7).
(N/2) + 1=5 (número que ocupa la 5ª posición, en nuestro ej. es el 9).

Me = (7+9)/2 = 8.

- **Si N es impar** la mediana es la posición (N+1)/2.

$$Me = x\left(\frac{n+1}{2}\right)$$

Ejemplo: $X=\{1,2,5,7,9,10,13\}$

N = 7 (impar) $\Rightarrow$ (N+1)/2 = 4 (valor que ocupa la 4ª posición, en nuestro ej. es el 7).
Me = 7.

- Cálculo de la mediana en las distribuciones de tipo II.

Es preciso ordenar los valores y trabajar con la frecuencia absoluta acumulada Ni, obteniendo en concreto el valor N/2.

Se distinguen 2 casos:

1. Que exista un N_i igual a N/2: En este caso la mediana es la media aritmética de X_i y del siguiente X_{i+1}, si la variable no admite decimales, la mediana serían los 2 valores, conjuntamente.

2. Cuando no existe un N_i que iguale a N/2, la mediana corresponde al primer X_i cuyo valor supere al de N/2.

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

- Cálculo de la mediana en las distribuciones de tipo III o agrupadas por intervalos.

1. Si existe N_i que es igual a N/2, la mediana por convenio, es el límite superior del intervalo mediano o intervalo en el que N_i = N/2

2. Si no existe un N_i = N/2, la mediana está en el siguiente intervalo, es decir en el primer intervalo cuya N_i supero a N/2; diremos que dicho intervalo es el intervalo mediano.

$$Me = L_{i-1} + \frac{(N/2) - N_{i-1}}{n_i} c_i$$

L_{i-1} : límite inferior del intervalo mediano.

N_{i-1} : frecuencia acumulada hasta el intervalo inmediatamente anterior al intervalo mediano.

n_i : frecuencia absoluta del intervalo mediano.

c_i : amplitud del intervalo mediano.

Ventajas

- Es la medida más representativa en el caso de las variables cualitativas o atributos.
- Su cálculo es sencillo.
- Tiene una fácil interpretación
- No es sensible a los valores extremos de la distribución.

Inconvenientes

- En su determinación no se tienen en cuenta todos los valores de la variable; puede constituir una ventaja, ya que es posible su cálculo cuando no se conocen los valores extremos pero sí su frecuencia.

3.7. La moda

La moda es el valor de la variable que se repite más veces; suele designarse por M_o y se define como el valor de la variable que presenta mayor frecuencia absoluta.

Cuando existan varios valores en esta situación se dice que la distribución es bimodal, trimodal o multimodal.
Se diferencian entre moda o modas absolutas y moda o modas relativas.

Un valor de una variable constituye una moda relativa cuando su frecuencia absoluta no es superada por la de sus valores contiguos.

Obtención de la moda en los 3 tipos de distribuciones:

- En las distribuciones de frecuencias de tipo I no tiene sentido hablar de moda, ya que las frecuencias absolutas son todas unitarias.

- Para obtener la moda de las distribuciones de tipo II basta con observar la columna de las n_i.

- En las distribuciones de tipo III, es decir, cuando los datos están agrupados en clases o intervalos pueden darse 2 supuestos:

  1. Que los intervalos sean de igual amplitud, en este caso la moda absoluta se situará en el intervalo que presente mayor frecuencia absoluta y las modas relativas en el intervalo o intervalos que superen la frecuencia absoluta de los intervalos contiguos. Para determinar el valor exacto de la moda podríamos optar por considerar la marca de clase del intervalo o proceder a prorratear un valor dentro del intervalo.

$$Mo = L_{i-1} + \frac{n_{i+1}}{n_{i+1} + n_{i-1}} \cdot c_i$$

  2. Que los intervalos tengan distinta amplitud: Para este caso es necesario obtener un ratio de densidad de frecuencia (frecuencia absoluta dividida por amplitud del intervalo); el intervalo con mayor valor en este ratio constituirá el intervalo modal. Es imprescindible para evitar situaciones de desequilibrio que pudieran desvirtuar el propio concepto, interés e interpretación de la moda.

Para hallar el punto modal exacto lo más habitual es operar con la siguiente expresión:

$$Mo = L_{i-1} + \frac{h_{i+1}}{h_{i+1} + h_{i-1}} \cdot c_i$$

Ventajas:

- Puede obtenerse en todas las distribuciones (cuantitativas como cualitativas), ya que siempre es posible determinar el valor, la categoría o la modalidad que más se repite.
- Su cálculo es sencillo.
- Fácil interpretación estadística, a que nos da el valor o modalidad que más se repite.

Inconveniente:

En su determinación no intervienen todos los valores de la distribución, centrándonos sólo en la mayor frecuencia absoluta de un determinado valor de la variable o de la modalidad de los atributos.

EJEMPLO:

L_{i-1} - L_i	n_i	N_i	C_i	X_i	h_i
0-25	20	20	25	12.5	20/25=0.8
25-50	140	160	25	37.5	140/25=5.6
50-100	180	340	50	75	180/50=3.6
100-150	40	380	50	125	40/50=0.8
150-200	20	400	50	175	20/50=0.4

1.- Media: $\overline{X} = \frac{12.5 \cdot 20 + 37.5 \cdot 140 + 75 \cdot 180 + 125 \cdot 40 + 175 \cdot 20}{400} = 68.75$

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

2.- Mediana: N/2 = 400/2 = 200.

Si vamos a la tabla, el intervalo que acumula 200 observaciones es el intervalo 50-100, por tanto, éste será el intervalo mediano.

$$Me = 50 + \frac{200 - 160}{180} \cdot 50 = 61.1$$

3.- Moda.

El intervalo modal es aquel que tiene una mayor densidad de frecuencia, en nuestro caso es el intervalo 25-50. Como los intervalos son de diferente amplitud, la moda será:

$$Mo = 25 + \frac{3.6}{0.8 + 3.6} \cdot 25 = 45.45$$

3.8. MEDIDAS DE POSICION NO CENTRALES: LOS CUANTILES

Los cuantiles son los valores de la variable que dividen una distribución de frecuencias en partes iguales.

Los más habituales son:

- Cuartiles, son 3 valores que dividen la serie de datos en cuatro partes iguales. La mediana coincide con el segundo cuartil divide la distribución en dos partes iguales.
- Quintiles, son 4 valores que dividen la distribución en 5 partes iguales.
- Deciles, son nueve valores que dividen la distribución en 10 partes iguales.
- Percentiles, que son 99 valores que dividen la distribución en cien partes iguales.

Considerando N el número de datos de la distribución, o frecuencia absoluta acumulada, con carácter general los cuantiles se obtienen con la expresión:

$$\frac{r \cdot N}{q}$$

en la que r indica el cuantil correspondiente (r = 1, primer cuantil, r = 2, segundo cuantil, etc.) y q el número de intervalos con iguales frecuencias en los que se pretende dividir la distribución (si q = 4 hablamos de cuartiles, si q = 10 de percentiles, etc.).

Para distribuciones agrupadas en intervalos utilizamos la siguiente expresión:

$$Q_q^r = L_{i-1} + \frac{\frac{rN}{q} - N_{i-1}}{n} \cdot c_i$$

Algunos programas informáticos no utilizan los mismos criterios o algoritmos indicados con anterioridad; en concreto, la Excel considera a todas las distribuciones como si fueran continuas y sitúa los cuartiles no el valor de la variable cuya frecuencia absoluta acumulada supera al establecido por el cuartil, sino en un punto intermedio que obtiene mediante un algoritmo particular.

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

EJEMPLO:

Xi	ni	Ni
1	10	10
2	15	25
5	7	32
8	10	42
9	20	62
20	40	102
	n = 102	

Podemos buscar, por ejemplo, el percentil 30 si queremos buscar aquel valor X (puede ser que no haya sido observado), tal que el 40% de las observaciones serían inferiores a él.

El percentil k-ésimo se calcula de la siguiente manera:

$$P_{k\%} = \frac{n}{100} \cdot k$$

Ej. El percentil 40% del ejemplo anterior sería:

$$P_{40\%} = \frac{102}{100} \cdot 40 = 40.8$$

Aproximamos a 41. El 41 nos indica la posición que ocupa el valor que buscamos. Si observamos en la tabla la columna de frecuencias acumuladas, vemos que el valor que ocupa esta posición es el 8. El 8 deja un 40% de las observaciones por debajo.

Observaciones:

Los cuartiles Q_1, Q_2, Q_3 son los percentiles 25, 50 y 75 respectivamente. Los deciles $D_1, D_2, \ldots\ldots, D_9$ se corresponden con los percentiles 10, 20, 30, 40, 50, 60, 70, 80 y 90.

P (Me) = 50%.

$$P_{50\%} = Me = Q_2 = D_5$$

3.9. Medidas de posición robustas

Tratan de paliar los problemas de estimación asociados a distribuciones anómalas, siendo estadísticos que funcionan bien para varios tipos distintos de distribuciones teóricas, aunque pueden no ser el mejor estimador para ningún tipo concreto de distribución siendo, por tanto, el mejor compromiso.

3.9.1. La media k-recortada

Es la media de los datos que quedan después de eliminar el k por ciento de los datos más grandes y k por ciento de los datos más pequeños. A la media recortada al 25% se le denomina **centrimedia**. La media recortada al 0% es igual a la media aritmética.

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

3.9.2. La media k-winsorizada

En lugar de prescindir de los k por ciento datos más grandes y más pequeños, se sustituyen por el valor mayor y menor de los datos restantes.

3.9.3. La trimedia

Es un índice de tendencia central que consiste en calcular una media aritmética ponderada de tres medidas, la Mediana (con peso doble).

$$Trimedia = \frac{Q_1 + 2Q_2 + Q_3}{4}$$

3.10 momentos de una distribución unidimensional de frecuencias

Los momentos son medidas que caracterizan a una distribución de frecuencias y que tienen como principal utilidad su condición de operadores para el cálculo simplificado de las medidas de posición, dispersión o forma de una distribución; también tienen una importante utilidad para efectuar las regresiones estadísticas.

Existen dos clases:

- Respecto al origen, que se representan con a_h

$$a_h = \sum_{i=1}^{r} x_i^h \frac{n_i}{N}$$

- Los momentos centrales o respecto a la media, se representan con m_h

$$m_h = \sum (x_i - \overline{x})^h \frac{n_i}{N}$$

Al momento de orden 2 respecto a la media m_2, se le denomina varianza, constituye la medida de dispersión más utilizada.

Puede demostrarse, utilizando el desarrollo del Binomio de Newton, que los momentos respecto a la media están relacionados con los momentos respectos al origen:

$$m_h = f(a_{h-j})$$

$$m_2 = a_2 - a_1^2$$

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871

EJEMPLO:

Dada la siguiente distribución de frecuencias:

Xi	ni	Ni
1	**3**	**3**
2	**4**	**7**
5	**9**	**16**
7	**17**	**33**
	$\sum$ = 33	

- Media: $\overline{X} = \dfrac{1\cdot 3 + 2\cdot 4 + 5\cdot 9 + 7\cdot 17}{33} = 5.30.$
- Mediana: n = 33 (impar) $\Rightarrow$ (n+1)/2 = 17. Vemos en la columna de frecuencia acumulada que el valor que ocupa la posición 17 es el 7. Me = 7.
- Moda: Mo = 7, ya que el 7 es el valor que más veces se repite.

INFO@ADEFACIL.COM
ELENA GONZALO NOGUÉS
671 227 871