ESTADÍSITICA DESCRIPTIVA

Las distribuciones de frecuencias aportan información preliminar de interés. acerca de una variable investigada en una población o en una muestra. Sin embargo, a menudo necesitamos una medida que caracterice o distinga a la distribución, en los siguientes aspectos:

- En cuanto a su posición
- En cuanto a su dispersión
- En cuanto a su forma

Como ya dijimos, cuando a estos valores se los obtiene a partir de toda la población se denominan parámetros y si son calculados con los datos de una muestra se denominan estadísticos.

a) las **medidas de posición**: nos indican la "posición", que ocupa la distribución sobre el eje de las abscisas. También se las denomina de tendencia central porque muchas de ellas tienden a ubicarse en el centro de la distribución.

b) las **medidas de dispersión**: nos reflejan cómo se distribuyen los valores de la variable a lo largo del eje de las abscisas.

c) Entre las **medidas de forma**, tenemos:

- las de asimetría: que indican la deformación horizontal que tiene la distribución. Una distribución perfectamente simétrica es aquella en la que las frecuencias equidistantes de la frecuencia máxima son iguales.

- las de puntiagudez: expresan la altura relativa de la distribución.

MEDIDAS DE POSICIÓN

Se definen varios tipos de medidas de posición o tendencia central, siendo las más comunes:

Medidas de Posición
- media aritmética
- moda
- mediana
- cuartiles
- deciles
- percentiles

Medidas de orden: mediana, cuartiles, deciles, percentiles

Cada una tiene ventajas y desventajas, según los datos y el objetivo perseguido.

El cálculo de las mismas difiere de acuerdo al tipo de variable con que se trabaja y presenta pequeñas modificaciones según se disponga de datos agrupados o no.

MEDIA ARITMÉTICA

O simplemente media, o promedio, es el más conocido y quizás el más usado de los parámetros de posición por las ventajas que ofrece en algunos aspectos. Se denota con $\bar{x}$, (léase "x barra") y para un conjunto n de números $x_1 , x_2 , x_3 \ldots \ x_n$. Se define por:

$$\bar{x} = \frac{x_1 , x_2 , x_3 \ldots \ x_n}{n} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Si los valores de la variable x se repiten, es decir si tienen frecuencias, se debe hacer la suma de los productos de cada valor de la variable con su frecuencia y dividir por n, es decir:

$$\bar{x} = \frac{x_1 . f_1 + x_2 . f_2 + \ x_3 . f_3 + \cdots + x_n . f_n}{n} = \sum_{i=1}^{n} \frac{x_i . f_i}{n}$$

Donde N es la frecuencia total, es decir, la cantidad total de casos, que en la distribución de frecuencias, está dado por la sumatoria de los f_i

En el caso de datos numéricos agrupados en intervalos de clase, el cálculo de la media aritmética es similar al caso anterior, es decir: $\bar{x} = \sum_{i=1}^{n} \frac{x_i . f_i}{n}$

Ahora bien, qué valor tomará x_i en la fórmula de cálculo de la media aritmética.

Cuando se agrupan datos continuos en intervalos de clase, se pierde información original. Luego, para solucionar este problema, x_i se calcula como el promedio entre los extremos de cada intervalo, es decir, x_i representa el punto medio, (o marca de clase), de cada intervalo de clase.

Calculemos la media aritmética en la siguiente tabla de distribución de frecuencias que corresponden a la cantidad de Kilogramos de carga que han transportado 27 vehículos.

Kg transportados (xi)	Cant. de Vehículos (fi)	xi	xi. fi
50 – 100	2	75	150
100 – 150	6	125	750
150 – 200	11	175	1925
200 – 250	5	225	1125
250 – 300	3	275	825
	n =27		Total : 4775

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i . f_i}{n} = \frac{4775}{27} = 176{,}85$$

O sea que en promedio, cada vehículo transportó una carga de 176,85kg.

Características de la media aritmética:

- ✓ Como dijimos anteriormente es un valor comprendido entre el mínimo y el máximo valor de la variable en estudio.
- ✓ Posee la misma unidad de medida que la variable considerada.
- ✓ En su cálculo intervienen todos los valores de la variable estudiada. Esto se presenta como una ventaja ya que permite el tratamiento algebraico de la misma.
- ✓ Otra ventaja es que resulta de fácil cálculo e interpretación.
- ✓ No se la puede calcular cuando los datos están agrupados en una tabla de distribución de frecuencias con intervalos abiertos, (porque de los mismos no se puede obtener el punto medio). Obviamente esto es una desventaja.
- ✓ Se ve afectada o arrastrada por los valores extremos, lo que la hace poco significativa cuando éstos existen. Por lo tanto, no se aconseja su cálculo en estos casos.

Propiedades de la media aritmética

1. "*La suma de los desvíos de cada valor de la variable con respecto a la media aritmética es siempre igual a cero*". En símbolos:

$$\sum_{i=1}^{n} x_i - \bar{x} = 0$$

En general, entendemos por desvió, a la diferencia entre los valores de la variable y un valor fijo cualquiera. Cuando ese valor fijo es la media aritmética tendremos desvíos con respecto a ella.

2.- "*La suma de los cuadrados de los desvíos con respecto a la media aritmética, da un mínimo*". Es decir que cuando los desvíos son con respecto a la media, la suma de los cuadrados nos da un valor que será siempre menor que el valor que se obtiene cuando los desvíos y sus cuadrados se calculan respecto de cualquier otra constante distinta a la media.

MEDIA ARMÓNICA (O PROMEDIO ARMÓNICO)

Es una medida de posición central de la estadística descriptiva. La media armónica se calcula dividiendo el número total de datos estadísticos entre la suma de los inversos de cada valor.

La media armónica sirve para calcular promedios de velocidades, tiempos o hacer cálculos en electrónica. Esta característica diferencia la media armónica de los otros tipos de medias, que se utilizan frecuentemente en el cálculo medio de precios o porcentajes.

La media armónica suele representarse con una H mayúscula:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Para calcular la media armónica se deben hacer los siguientes pasos:

1. Calcular el inverso de cada dato estadístico de la muestra
2. Sumar todos los inversos calculados.
3. Dividir el número total de datos entre la suma hallada en el paso anterior.
4. El resultado obtenido es la media armónica de la muestra estadística

Vamos a ver cómo hallar la media armónica de un conjunto de datos resolviendo un ejemplo paso a paso del precio de unas acciones.

Una persona compra acciones de una empresa cada año durante 5 años seguidos. Durante este periodo el precio de las acciones ha fluctuado bastante: el primero año valía 7€ cada acción, el segundo año 10€, el tercer año 15€, el cuarto año la empresa tuvo grandes pérdidas económicas y el precio bajó hasta los 6€ por acción y, finalmente, el quinto año la compañía hizo una fuerte inversión que provocó un aumento del precio hasta los 11€. ¿Cuál ha sido el precio medio de compra de las acciones?

Una opción sería hacer el cálculo de la media aritmética, es decir, sumar todos los precios y dividirlos entre cinco. Sin embargo, como las compras se han realizado en diferentes años, sacar la media aritmética sería un error. Por tanto, debemos hallar la media armónica de todos los precios.

Entonces, aplicamos la fórmula de la media armónica:

$$H = \frac{5}{\frac{1}{7} + \frac{1}{10} + \frac{1}{15} + \frac{1}{6} + \frac{1}{11}} = 8,82$$

De modo que el precio medio de las acciones que ha comprado durante este periodo es de 8,82 euros por acción.

Para datos agrupados se usa la fórmula:

$$H = \frac{n}{\frac{f_1.1}{x_1} + \frac{f_2.1}{x_2} + \cdots + \frac{f_n.1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$

Propiedades de la media armónica

La media armónica tiene las siguientes características:

- ✓ Los valores grandes afectan poco a la media armónica de un conjunto, es decir, un valor muy grande en comparación con los otros datos no supondrá un cambio notable en la media armónica.
- ✓ En cambio, un valor pequeño sí que influye mucho en la media armónica de un conjunto, haciendo reducir considerablemente su valor. Esto es debido a que entonces los recíprocos del denominador de la fórmula toman valores muy grandes.
- ✓ La media armónica no se puede calcular si alguno de los datos es cero, ya que se obtendría una indeterminación en la fórmula. En tal caso, se dice que la media armónica no está definida.
- ✓ La inversa de la media armónica es equivalente a la media aritmética de los inversos de las observaciones.
- ✓ Para un mismo grupo de datos, la media armónica será menor o igual a la media aritmética.

MEDIA GEOMÉTRICA O PROMEDIO GEOMÉTRICO

La media geométrica de un conjunto de datos estadísticos es igual a la raíz n-ésima del producto de todos los valores de la variable.
La media geométrica se utiliza en las finanzas de las empresas para calcular tasas de retorno, promedios sobre porcentajes, e intereses compuestos.

$$MG = \sqrt[n]{x_1 \,.\, x_2 \,.\, x_3 \,.\, \ldots \;\; .\, x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

La media geométrica solo se puede calcular cuando todos los datos de la muestra son positivos. Porque si algún valor es negativo la raíz tendría solución negativa o no tendría solución, por otro lado, si algún dato es nulo entonces la multiplicación de los datos daría cero y, en consecuencia, la media geométrica sería igual a 0.
La principal diferencia entre la media geométrica y la media aritmética es que la media geométrica es menos sensible a los valores extremos que la media aritmética. Además, la media aritmética se puede calcular con valores negativos y nulos, en cambio, la media geométrica solamente se puede calcular con valores positivos.
Asimismo, la media geométrica será, en general, menor que la media aritmética para un mismo conjunto de datos.
Para calcular la media geométrica se deben hacer los siguientes pasos:

1. Calcular el producto de todos los datos estadísticos de la muestra.

2. Hallar la raíz n-ésima del producto calculado.
3. El resultado obtenido es la media geométrica de la muestra estadística.

Veamos un ejemplo para de cómo obtener la media geométrica.
Se conocen los resultados económicos de un empresa de los últimos cinco años. El primer año la empresa generó un 10% de rentabilidad económica, el segundo año el beneficio llegó al 23%, en el tercer año el dinero ganado fue del 16%, el cuarto año logró un 7% de rentabilidad económica y la inversión del quintó año supuso una rentabilidad del 20%. Se pide calcular el promedio de todos los porcentajes.

$$MG = \sqrt[5]{1,10 \cdot 1,23 \cdot 1,16 \cdot 1,07 \cdot 1,20} = 1,15 \longrightarrow 15\%$$

El resultado numérico de la media geométrica es 1,15, lo que significa que la empresa ha crecido económicamente un 15% de media cada año.
Ten en cuenta que hemos podido sacar la media geométrica porque todos los valores eran positivos, pero si algún porcentaje hubiese sido negativo tendríamos que haber puesto el dato en la fórmula en forma de decimal positivo con la parte entera igual a cero. Por ejemplo, un crecimiento del -30% se debe expresar en la fórmula como 0,70 (1-0,3=0,7).

Para datos agrupados se usa la fórmula:

$$MG = \sqrt[n]{{x_1}^{f_1} . {x_2}^{f_2} . \ \dots \ . {x_n}^{f_n}} = \sqrt[n]{\prod_{i=1}^{n} {x_i}^{f_i}}$$

MODO O MODA

Es el valor de la variable que se repite la mayor cantidad de veces, es decir, al que le corresponde la máxima frecuencia.

En símbolos: $\hat{x}$

Si tenemos datos sin agrupar, bastará con identificar cuál es el valor de la variable que más se repite.

Podremos tener series, con un modo es decir unimodales, con más de un modo, o sea bimodales, o que no tengan modo, llamadas amodales.

Por ejemplo, para las siguientes series de datos el modo seria:

a) 2 3 5 7 2 $\hat{x} = 2$

b) 10 14 10 12 10 20 14 45 14 $\hat{x} = 10\ y\ 14$

c) 23 24 25 30 45 54 Amodal

En el caso de datos provenientes de una variable discreta con datos agrupados, es posible determinar inmediatamente el valor modal. Bastará con identificar al valor de la variable al que le corresponde la mayor frecuencia.

Ejemplo:

N° de obreros	Cant. de establecimientos
10	44
11	84
12	73
13	**189**
14	32

Mirando la tabla, directamente podemos determinar el modo o valor modal fijándonos cual es la máxima frecuencia absoluta, (189 en este caso), y luego a qué valor de la variable le corresponde, (13, en este ejemplo). Por lo tanto:

$\hat{x}$ =13 obreros

Esta medida de tendencia central está indicando que lo más frecuente es que los establecimientos observados, tengan 13 obreros.

En una tabla con intervalos de clase el modo se puede obtener gráfica y analíticamente.

El siguiente es un ejemplo de una distribución de frecuencias referida a la producción de un grupo de establecimientos, (en kg).

Peso (en kg)	xi	fi	Fi
50 – 100	75	1	1
100 – 150	125	3	4
150 – 200	175	9	13
200 – 250	225	30	43
250 – 300	275	60	103
300 – 350	325	52	155

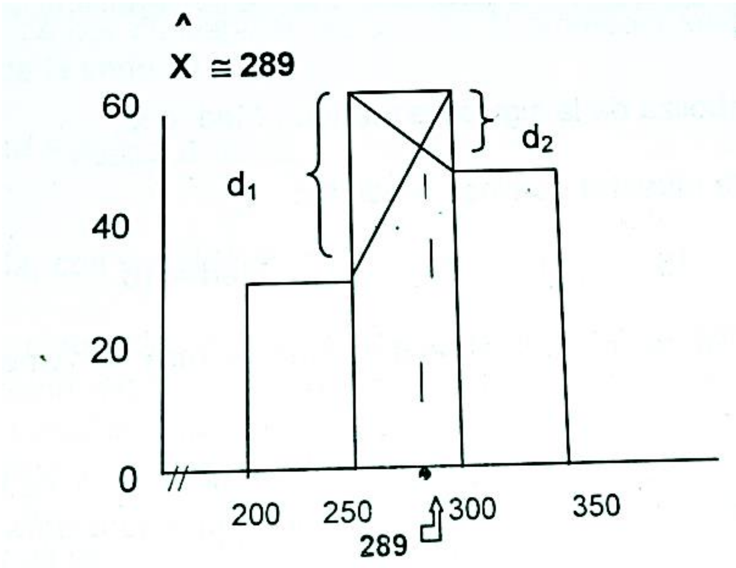
Total: 155

En primer término, vamos a determinar el intervalo con mayor frecuencia absoluta: es 250-300. Dicho intervalo se denomina intervalo de clase modal.

Determinación gráfica

1. Se confecciona el histograma con la barra de mayor frecuencia y las adyacentes.
2. Se trazan dos diagonales en el interior de la barra del intervalo modal, partiendo de los vértices de la barra hasta los vértices de las adyacentes.
3. Se traza luego una línea perpendicular desde la intersección de las dos diagonales hasta el eje de las x, (escala horizontal). El punto donde se cortan será el valor de la variable al que le corresponde la máxima frecuencia, o sea el modo.

En el ejemplo:



Determinación analítica

Para datos agrupados en intervalo se utiliza la siguiente fórmula:

$$\hat{x} = L_i + \frac{d_1}{d_1 + d_2} \cdot h$$

Donde:

$L_i = límite\ inferior\ del\ intervalo\ modal$

$d_1 = f_i - f_{i-1}$, $es\ decir, la\ diferencia\ entre\ la\ frecuencia\ absoluta\ del\ intervalo\ modal,$ $menos\ la\ inmediata\ anterior.$

$d_2 = f_i - f_{i+1}$, $es\ decir, la\ diferencia\ entre\ la\ frecuencia\ absoluta\ del\ intervalo\ modal,$ $menos\ la\ inmediata\ posterior.$

$h = amplitud\ del\ intervalo$

MEDIANA

Se define como el valor de la variable, (en una serie ordenada), que divide al conjunto de datos en dos subconjuntos con igual número de elementos.
Se simboliza de la siguiente manera: $\tilde{x}$

En la siguiente muestra de cinco medidas:
14 15 16 19 23 $\tilde{x} = 16$
ya que la tercera observación tiene el mismo número de observaciones a ambos lados.
Podemos hacer visible la mediana fácilmente si pensamos en una ordenación de menor a mayor. Por ejemplo, una fila de personas, alineadas por sus estaturas.
La "persona mediana" será el/la que tiene igual número de personas a su derecha y a su izquierda. Su altura será la altura mediana de la muestra considerada. Esta cantidad se calcula fácilmente en una muestra ordenada de un número impar de individuos.

Cuando el número es par, la mediana se calcula convencionalmente como el punto medio entre los valores que ocupan el lugar: $\frac{n}{2}$ y $\frac{n}{2}+1$. En una muestra de cuatro medidas:
20 21 22 26

la mediana será el punto medio entre la segunda y la tercera medida, o sea: 21,5.

También podemos definir a la mediana como aquel valor de la variable que cumple con la condición de superar a no más de la mitad de las observaciones y ser superado por no más de la mitad de las observaciones.

Simbólicamente para calcular la mediana hay que distinguir distintas situaciones:
a) Cuando la serie es simple y la cantidad de observaciones es un número impar, es decir, n número impar. Sea la serie simple: $x_1, x_2, x_3 \ldots\ x_n$
En éste caso hay que ordenar las observaciones de menor a mayor y luego localizar la observación central que será aquella que ocupe el lugar $\frac{n+1}{2}$. Es decir $\tilde{x} = x_{\frac{n+1}{2}}$

b) Cuando la serie es simple y la cantidad de observaciones es par.
En ese caso hay dos valores centrales: los que ocupan la posición $\frac{n}{2}$ y $\frac{n}{2}$ +1. Como dijimos por convención se adopta el promedio simple de ambos comoúnica mediana de la serie. O sea:

$$\tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$$

c) Serie agrupada, con variable discreta:

El procedimiento de cálculo resulta de practicar el análisis anterior para serie simple, pero teniendo en cuenta las ponderaciones que ahora aparecen.

Hay que calcular el valor de $n/_2$ y las frecuencias absolutas acumuladas. Luego se relaciona el valor n/2 con las frecuencias absolutas acumuladas para encontrar dos de estos valores entre los que esté comprendido el mismo. Supongamos que ese par de valores sean F_{j-1} y F_j y que satisface que: $F_{j-1} < \frac{n}{2} < F_j$

Ejemplo:

xi	fi	Fi
7	32	32
8	40	72
9	12	84
10	10	94
11	22	116

Total: 116

58

$$\frac{n}{2} = 58$$

$$32 < \frac{n}{2} < 72$$

$$\tilde{x} = 8$$

d) Distribución de variable continua

Ejemplo:

intervalos	fi	Fi
20 – 40	2	2
40 – 60	6	8
60 – 80	11	19
80 – 100	6	25
100 – 110	5	30

Total: 30

n/2 = 15

Los pasos para obtener el valor mediana, son:

Agregar la columna de frecuencias acumuladas (Fi). Obtener el punto medio de la distribución mediante la siguiente operación: n / 2 = 30/2 = 15

Ubicar en la columna de Fi (frecuencias acumuladas), los valores entre los cuales se encuentra el n/2, (15 para nosotros), o sea entre 8 y 19, quiere decir que la mediana, en éste caso, se ubicará en una posición mayor que 8 y menor que 19. Por lo tanto ya sabemos que será un valor entre 60 y 80.-

Una vez determinado el intervalo en el cual cae la mediana, usaremos la siguiente fórmula:

$$\tilde{x} = L_i + \frac{\frac{n}{2} - F_{i-1}}{fi} \cdot h$$

Donde:

$L_i = límite\ inferior\ del\ intervalo\ que\ contiene\ la\ mediana$

$F_{i-1} = frecuencia\ acumulada\ \ anterior\ a\ la\ que\ contiene\ la\ mediana$

$f_i = \ frecuencia\ absoluta\ del\ intervalo\ que\ contiene\ la\ mediana$

$h = amplitud\ del\ intervalo$