

UNIDAD TEMATICA 5 : “ Medidas de Posición”

Objetivo: Calcular medidas analíticas para describir las principales propiedades de los datos, interpretando correctamente los resultados de un análisis estadístico.

5.1 Introducción

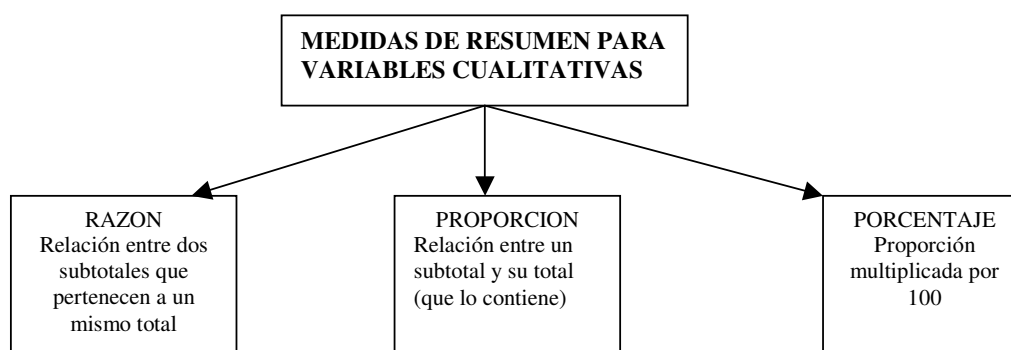
Cuando hemos reunido un conjunto de datos, es necesario condensarlos para que aparezcan claramente las características principales de dicho conjunto. Si debemos comparar dos o más conjuntos, la condensación es aún más necesaria. Por eso es importante que dispongamos de medidas cuantitativas para representar adecuadamente los datos y poder compararlos más fácilmente. Estas medidas nos proporcionan, en forma concisa, un resumen de la información obtenida.

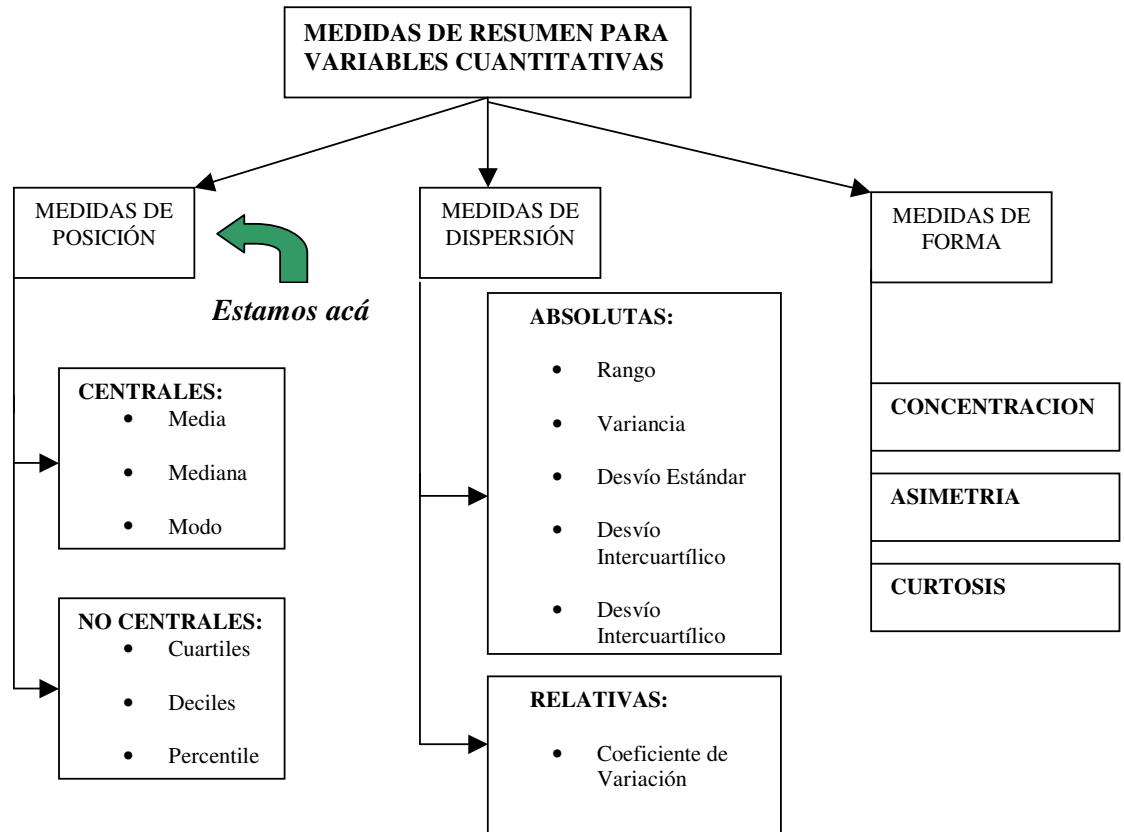
Como lo hemos demostrado, las tablas y las gráficas bien hechas, pueden ser de ayuda considerable al hacer que grandes cantidades de datos, de otra forma confusos, den a conocer los secretos que ocultan cuando están desorganizados. Sin embargo, un método aún más radical para condensar conjuntos de datos es el *Cálculo de Medidas aritméticas de Resumen* diseñadas para expresar las características más notables de los conjuntos de datos en la forma más compacta imaginable.

La Estadística Descriptiva es válida para cualquier tipo de información: la que proviene de una población o la que proviene de una muestra. Las medidas que se pueden obtener dependen del tipo de variables que se hayan manejado.

Los datos sobre variables cuantitativas se pueden resumir con nitidez en tres formas: Medidas de Posición o ubicación, Medidas de dispersión y Medidas de Forma.

Los datos sobre variables cualitativas se pueden resumir en tres medidas, que también pueden utilizarse para variables cuantitativas y que son: Tasa, Razón y Proporción.





5.2 Medidas de Posición: Centralización o Tendencia Central

Las medidas de centralización, nos sirven para representar con un solo número todo un conjunto de datos. Dicho número es fácil de comprender y comunicar. Estas medidas nos simplifican mucho la tarea de establecer conclusiones.

Las medidas de tendencia central son valores alrededor de los que las observaciones tienden a agruparse, y permiten ubicar lo que en algún sentido pudiera llamarse el “centro” de un conjunto de datos. La tendencia central se refiere al punto medio de una distribución. Las medidas de tendencia central se conocen como medidas de posición.

No hay duda que la medida de Tendencia Central que se encuentra con más frecuencia, y es la más conocida, es la *media* llamada con más propiedad *Media Aritmética* (en la práctica cotidiana se conoce simplemente como el *promedio*). También se emplean con frecuencia la Mediana y la Moda.

El cálculo de estas medidas difiere de acuerdo al tipo de variables con que se trabaja y presenta pequeñas modificaciones según se disponga de datos agrupados o no agrupados.

• 5.3 Media

La media aritmética de un conjunto de observaciones es una medida de posición que se conoce comúnmente como promedio.

Si se ha realizado una investigación donde se han medido n unidades experimentales respecto a una característica determinada, **la media aritmética se calcula como la suma de todos los valores que toma la característica en estudio dividida por el número total de unidades experimentales observadas.**

Si disponemos de datos sin agrupar la definimos como:

$$\overline{X} = \frac{(X_1 + X_2 + X_3 + \dots + X_n)}{n}$$

donde:

$(X_1, X_2, \dots, X_n)$ son los n datos que tenemos recogidos de la variable en cuestión

n : tamaño de la muestra.

En símbolos:

$$Media(X) = \frac{\sum_{j=1}^n X_j}{n}$$

Este símbolo extraño que aparece por primera vez en esta Unidad Temática corresponde a la letra griega sigma mayúscula (Σ) y simplemente significa la suma de las observaciones indicadas.

La igualdad $j=1$ indica que los valores (datos) deben ser sumados empezando por el primero y terminando por el enésimo como nos indica la igualdad $j=n$, que está escrita encima del signo Σ .

El subíndice j en la expresión X_j indica los distintos valores que va tomando la variable en estudio. Por ejemplo si se han observado 3 unidades experimentales, la variable medida puede tomar únicamente 3 valores que se indican por : X_1, X_2, X_3 .

En este caso, el subíndice $j = 1, 2, 3$

En general, el símbolo X_j denota la observación correspondiente a la j – ésima unidad experimental de la población o de la muestra que se está estudiando.

Como ejemplo, consideremos 10 pacientes de edades 21 años, 32, 15, 59, 60, 61, 64, 60, 71, y 80.

$$\bar{X} = \frac{21 + 32 + 15 + 59 + 60 + 61 + 64 + 60 + 71 + 80}{10} = 52.3 \text{ años}$$

Generalizando sobre el ejemplo podemos decir que la media de esta muestra es igual a:

$$\text{Media}(X) = \frac{\sum_{j=1}^n X_j}{n} = 52,3$$

Interpretación: La edad media de estos pacientes es de: 52,3 años.

Puede ocurrir que cada uno de los valores de la variable se presente más de una vez o que se dispongan de una gran cantidad de datos, entonces, en estos casos, es preferible, usar datos agrupados para agilizar los cálculos. El cálculo de la media para datos agrupados se puede lograr sólo con suponer, quizás de manera incorrecta, que las observaciones que caigan en una clase, están igualmente espaciadas dentro de ellas y se ubican por lo tanto, en promedio, igual en el punto central del intervalo de clase. Cada punto medio se multiplica entonces por la frecuencia absoluta de clase, y la suma de estos productos se divide entre el valor de la población N o de la muestra n. Si cada punto medio está representado por X y la frecuencia absoluta de la clase correspondiente por f , el procedimiento se puede escribir de la siguiente manera:

$$\bar{X} = \frac{\sum X_j f_j}{n}$$

donde $n = \sum f_j$ es el total de datos, m el número total de clase y f_j es la frecuencia de datos.

Ejemplo:

Un investigador social está interesado en conocer el número promedio de hijos en una muestra de 10 familias entrevistadas para una encuesta en particular. Luego de efectuar el trabajo de recolección de datos, el listado de las familias con su correspondiente número de hijos se formó la siguiente tabla:

TABLA N° 5 -1: Número de hijos por familia

Familia N°	Número de Hijos
1	2
2	4
3	4
4	3
5	4
6	3
7	3
8	3
9	6
10	3

Con esta información se puede construir la Tabla de Frecuencias de la siguiente manera:

TABLA N° 5 -2: Número de hijos por familia

Número de Hijos (X_j)	Frecuencia (f_j)
2	1
3	5
4	3
6	1
TOTAL	10

La fórmula apropiada para calcular la media aritmética en el caso de variables discretas agrupadas en una tabla de frecuencias es:

$$\bar{X} = \frac{\sum X_j f_j}{n}$$

donde X_j representa los distintos valores que toma la variable y f_j su correspondiente frecuencia absoluta. Entonces, para calcular la media aritmética se agregará a la Tabla de frecuencias anterior una columna auxiliar que facilitará el cómputo de este promedio.

TABLA N° 5 -3: Número de hijos por familia

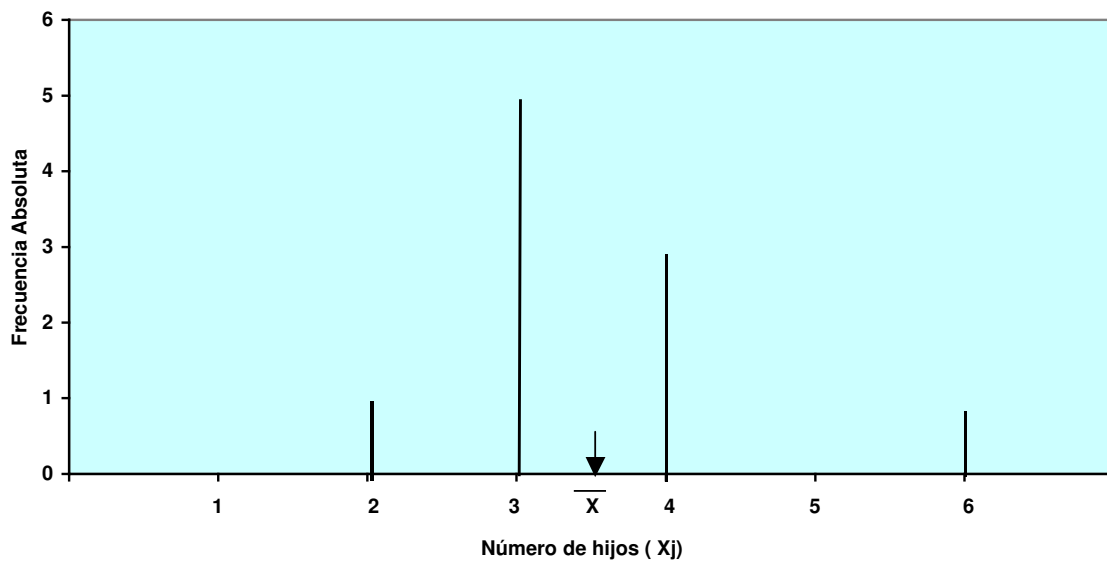
Número de Hijos (X_j)	Frecuencia (f_j)	$X_j f_j$
2	1	2
3	5	15
4	3	12
6	1	6
TOTAL	10	35

Luego:

$$\bar{x} = \frac{\sum X_j f_{j-}}{n} = \frac{35}{10} = 3,5$$

Interpretación: La familia promedio proporcionada por la encuesta es aquella que presenta entre 3 y 4 hijos; el valor 3,5 es el resultado matemático del cálculo de la media aritmética pero no es un valor posible de la variable por su propia definición.

Si se representaran los datos por medio de un gráfico a bastones se tiene:



En el caso de datos numéricos continuos agrupados en intervalos de clase, el cálculo de la media aritmética es similar al caso anterior, es decir :

$$\bar{Y} = \frac{\sum Y_i f_i}{n}$$

donde las i varían de 1 a k , siendo k la cantidad de intervalos.

Simbolizamos ahora la variable por Y_i simplemente para marcar la diferencia entre las variables discretas y variables continuas agrupadas en intervalos de clase. Ahora bien ¿Qué valor tomará Y_i en la fórmula de cálculo de la media aritmética?.

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

Ejemplo:

Calcular la media aritmética de la longitud de 100 tornillos fabricados por una máquina.

En una Tabla de frecuencias agregamos ahora dos columnas. Una simbolizada por Y_i que contiene los puntos medios de los intervalos de clase y la otra que expresa el producto de cada uno de ellos por su correspondiente frecuencia absoluta. Esto da lugar a la **TABLA 5-4**

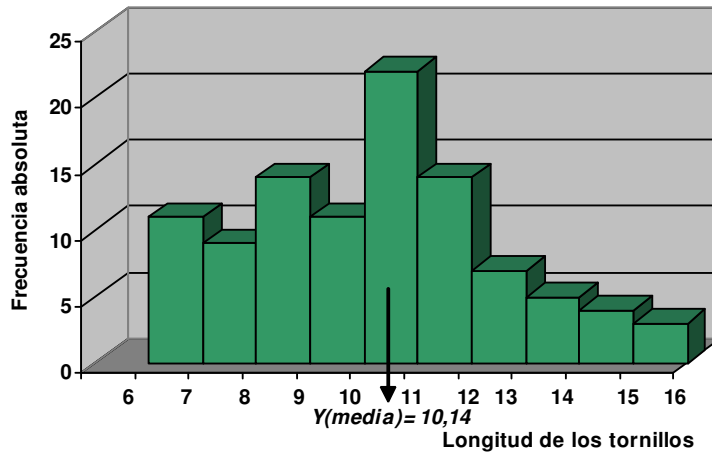
Intervalos de Clase	Punto Medio (Y_i)	Frecuencia (f_i)	$Y_i * f_i$
6 – 7	6,5	11	71,5
7 – 8	7,5	9	67,5
8 – 9	8,5	14	119,0
9 – 10	9,5	11	104,5
10 – 11	10,5	22	231,0
11 – 12	11,5	14	161,0
12 – 13	12,5	7	87,5
13 – 14	13,5	5	67,5
14 – 15	14,5	4	58,0
15 – 16	15,5	3	46,5
TOTAL		100	1.014,0

Luego:

$$\bar{Y} = \frac{\sum Y_i f_i}{n} = \frac{1014,0}{100} = 10,14 \text{ mm}$$

Interpretación: *En promedio el proceso productivo fabrica tornillos de 10,14 mm de longitud.*

Si hubiésemos calculado este promedio tomando los 100 datos originales sin agrupar, el resultado hubiese sido de Y (media) = 10,169 mm. La diferencia se debe a que cuando se agrupan los datos en intervalos de clase, se pierde precisión en el cálculo de las medidas descriptivas.



Representando a la Media Aritmética en el Histograma de Frecuencias correspondiente también se puede observar cómo esta medida posiciona la distribución. El valor de la media aritmética se encuentra a la izquierda de la distribución denotando esta situación una asimetría hacia la derecha a causa de que los últimos intervalos presentan frecuencias menores.

Ventajas

Es de fácil cálculo e interpretación sencilla. Es la más utilizada y es útil en muchos desarrollos matemáticos.

Desventajas

La principal desventaja se presenta cuando alguno o los dos valores extremos de la muestra son desproporcionados respecto al resto de los datos, sobre todo cuando éstos son poco numerosos. En este caso la media se aleja de la realidad, es decir, deja de ser representativa de los datos.

• 5.4 Mediana

La mediana de un conjunto finito de valores es aquel valor que divide al conjunto en dos partes iguales, de forma que el número de valores mayor o igual a la mediana es igual al número de valores menores o igual a estos. Su aplicación se ve limitada ya que solo considera el orden jerárquico de los datos y no alguna propiedad propia de los datos, como en el caso de la media.

A continuación se muestran los criterios para construir la mediana. Se puede construir los siguientes criterios:

- Lo primero que se requiere es ordenar los datos en forma ascendente o descendente, cualquiera de los dos criterios conduce al mismo resultado.

Sean ordenados los datos en orden ascendente $x_1, x_2, x_3, \dots, x_n$

- Si n (tamaño de la muestra) es impar, entonces, la mediana coincide con el valor medio, el cual corresponde al dato $X_{n/2}$.
- Si n (tamaño de la muestra) es par, no existe un solo valor medio, si no que existen dos valores medios, en tal caso, la mediana es el promedio de esos valores, es decir, los sumamos y luego los dividimos por dos.

$$Md = \frac{\frac{x_n}{2} + \frac{x_{n+1}}{2}}{2}$$

Podemos describir algunas propiedades para la mediana:

1.- Es única.

2.- Es simple.

3.- Los valores extremos no tienen efectos importantes sobre la mediana, lo que si ocurre con la media.

La notación mas usual que se utiliza para representar a la mediana es Md o Me

La mediana para datos no agrupados.

Ejemplo 1:

Dados los siguientes datos: 1, 2, 3, 4, 0, 1, 4, 3, 1, 1, 1, 1, 2, 1, 3 correspondientes al número de hijos de 15 empleados de una empresa. Para la obtención de la mediana se deberán de ordenar. Tomemos el criterio de orden ascendente con lo que, tendremos:

0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 3, 3, 3, 4, 4

Por otro lado el número de datos $n = 15$, siendo el número de datos impar se elige el dato que se encuentra a la mitad, una vez ordenados los datos, en este caso es 1.

0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 3, 3, 3, 4, 4

 Mediana

Interpretación: *El número mediano de hijos para estos empleados es 1.*

Ejemplo 2:

Las calderas de una planta de energía de vapor a alta presión tuvieron las siguientes eficiencias en porcentajes:

90,3 - 91,6 - 90,9 - 90,4 - 90,3 - 91,0 - 87,9 - 89,4

El tamaño de la muestra, $n=8$, número par. Luego los ordenamos y la mediana es la semisuma de los valores centrales o sea el promedio de esos valores.

87,9 - 89,4 - 90,3 - 90,3 - 90,4 - 90,9 - 91,0 - 91,6


 (los dos valores centrales)

$$\text{Mediana} = \frac{90,3 + 90,4}{2} = 90,35$$

Interpretación: *El número mediano de eficiencia en porcentaje de las calderas de una planta de energía es de 90,35 % aunque el mismo no sea un valor posible de la variable.*

La mediana para datos agrupados.

Si tenemos datos agrupados en tablas simples de frecuencia, procedemos de la siguiente manera:

- Antes de obtener la Mediana, vamos a calcular el lugar que ocupa. Lo llamaremos **orden de la mediana**, cuya fórmula es:

$$\text{Orden} = \frac{n+1}{2}$$

Ejemplo:

Supongamos que el gerente de personal de una empresa obtuvo los siguientes datos, correspondientes al número de días que 19 de sus empleados faltan por enfermedad en un año.

Luego :

$$\text{Orden} = \frac{19+1}{2} = 10$$

Los datos se presentan en la siguiente **TABLA 5 - 4**.

Nº de días	f _j	F _j
5	3	3
6	3	6
7	1	7
8	3	10
10	4	14
12	1	15
15	1	16
16	2	18
17	1	19
TOTAL	19	

Orden de la Mediana

Seguimos con el mismo ejemplo para ver como se calcula la Mediana.

- Buscamos en la tabla, en la columna de frecuencias acumuladas (F_j), el orden que hallamos.

Observando la tabla deducimos que el valor correspondiente a la Mediana es 8, pues ocupa el décimo lugar. Se cree necesario aclarar algo sobre esta tabla para comprender lo anterior. Podemos interpretarla así:

1°	2°	3°	4°	5°	6°	7°	8°	9°	10°	11°	12°	13°	14°	15°	16°	17°	18°	19°
5	5	5	6	6	6	7	8	8	8	10	10	10	10	12	15	16	16	17

donde la primera fila son los lugares que ocupan los datos y la segunda son los valores de los datos.

Es por esta razón que en lugar de hacer lo anterior para cada tabla, directamente se mira la columna de frecuencias acumuladas.

El valor 8 lo podemos pensar así:

Interpretación: El 50 % de los 19 empleados faltan menos de 8 días y el 50% restante más de 8 días.

¿Qué pasa se el orden de la mediana no coincide con uno de los valores de la frecuencia acumulada?

Veamos el siguiente ejemplo: Supongamos que la siguiente tabla corresponde a la vida útil en horas de 100 válvulas. Vemos el resultado en la **TABLA N° 5 -5**

X (N° de horas)	f_i	F_i
34	2	2
35	3	5
36	5	10
37	12	22
38	15	37
39	20	57
40	22	79
41	15	94
42	4	98
43	2	100
TOTAL	100	

→ 50,5: Orden

$$\text{Orden} = \frac{100 + 1}{2} = \frac{101}{2} = 50,5$$

Esto nos indica que la Mediana se encuentra entre el lugar 50 y el lugar 51. Pero, qué valores ocupan esos lugares?.

Por lo explicado anteriormente, desde el lugar 38 y hasta el lugar 57, hay valores 39. Luego el valor número 50 y el valor número 51 son 39. Entonces:

$$\text{Mediana} = \frac{39 + 39}{2} = 39$$

- Si los datos están agrupados en intervalo de clase, veamos cómo se calcula la mediana.

Ejemplo: Tenemos los siguientes datos agrupados en una Tabla de Frecuencia que representan los montos de 40 préstamos personales, en dólares, en una compañía financiera de consumidores.

Tabla N° 5-6

Intervalo de Clase	f_i	F_i
299 - 685	13	13
685 - 1071	11	24
1071 - 1547	6	30
1457 - 1843	5	35
1843 - 2229	3	38
2229 - 2615	1	39
2615 - 3001	1	40

► 20,5 = Orden

Siguiendo el razonamiento anterior, desde el lugar 14 hasta el 24 están en la clase 685 – 1071. Luego la Mediana pertenece a dicha clase.

La calculamos mediante la siguiente fórmula:

$$\text{Mediana} = a + x$$

Donde:

a : extremo inferior de la clase a la cual pertenece la mediana.

$$x = \frac{(\text{Orden} - F_a) \cdot h}{f_c}$$

Donde:

F_a : Frecuencia acumulada correspondiente a la clase anterior a la que contiene la Mediana. En caso que esta última sea la del primer intervalo, $F_a = 0$

h: amplitud de la clase

f_c : Frecuencia absoluta de la clase a la que pertenece la mediana.

En nuestro ejemplo:

$$x = \frac{(20,5 - 13) \cdot 386}{11} = 263,18$$

$$\text{Mediana} = 685 + 263,18 = 948,18$$

Ventajas

La Mediana no se ve afectada por los valores extremos, por lo que la podemos utilizar en aquellos casos en que la media aritmética no es útil.

Desventajas

La más importante, es que no podemos hacer cálculos adicionales con la mediana. Además, no utiliza mucha información de un conjunto de datos.

Finalmente, al menos que dispongamos de una computadora o unos esclavos, no es fácil ordenar un conjunto grande de números. En este caso, la mediana no es fácil de calcular.

- **5.5 Moda**

La moda es el valor que aparece con mayor frecuencia, es decir, el que ocurre más frecuentemente. La moda es la medida que se relaciona con la frecuencia con que se presenta el dato o los datos con mayor incidencia, con lo que se considera la posibilidad de que exista más de una moda para un conjunto de datos. La notación mas frecuente es la siguiente: Mo y $\hat{x}$. Esta medida se puede aparecer tanto para datos cualitativos como cuantitativos. Se dice que cuando un conjunto de datos tiene una moda la muestra es unimodal, cuando tiene dos modas bimodal, cuando la muestra contiene mas de un dato repetido se dice que es multimodal y un último caso es cuando ningún dato tiene una frecuencia, en dicho caso se dice que la muestra es amodal.

Moda para datos no agrupados

Si tenemos datos sin agrupar, la encontramos fácilmente observando cuál es el valor que más se repite.

Ejemplos:

1.- Determinar la moda del siguiente conjunto de datos:

a).- 1, 2, 3, 3, 4, 5, 6, 7, 7, 3, 1, 9, 3

Respuesta: La moda de este conjunto de datos es igual a 3 y si considera unimodal.

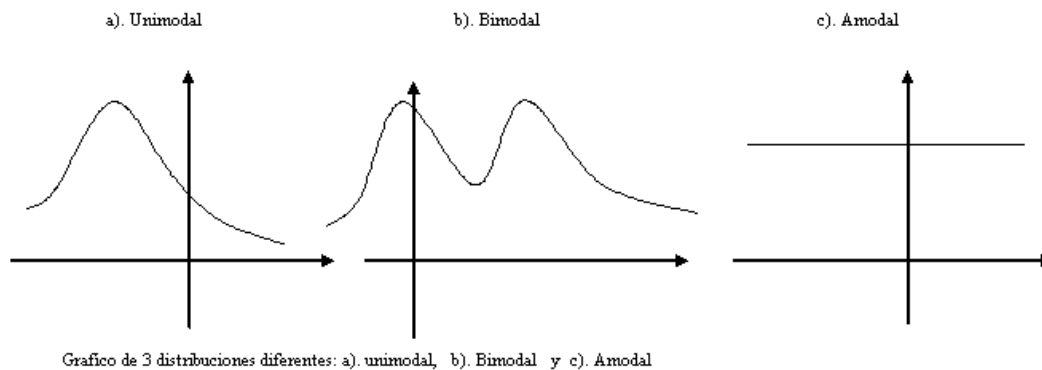
b).- 1, 2, 3, 4, 4, 5, 2, 1, 3, 4, 2, -3, 4, 6, 3, 3

Respuesta: Las modas de este conjunto de datos son 3 y 4 ya que ambas tienen la mas alta frecuencia, por lo que la muestra es bimodal

c).- 1, 2, 3, 4, 5, 6, 7, 8, 9

Respuesta: La muestra no contiene ningún dato repetido por lo que se considera que la muestra es amodal.

Gráficamente eso se puede reflejar mediante el análisis de un histograma de frecuencias.



Moda para datos agrupados

En datos agrupados en tablas simples de frecuencias, nos fijamos que valor corresponde a la mayor frecuencia absoluta. En la **TABLA 5-4**

Número de Días	f_i	F_i
5	3	3
6	3	6
7	1	7
8	3	10
10	4	14
12	1	15
15	1	16
16	2	18
17	1	19
TOTAL	19	

En este ejemplo, la mayor frecuencia absoluta es 4, que corresponde al valor 10. Luego la Moda es 10.

Interpretación: *La cantidad de días más frecuente que los empleados faltan por enfermedad es 10.*

En datos agrupados en intervalos de clases, existen varios métodos para calcular la Moda. Cada método puede darnos un valor diferente, pero aproximado, para un mismo conjunto de datos.

Un método poco usado, es tomar como moda a la marca de la **clase modal** que es aquella que contiene a la moda.

Otro método, más usado para determinar la moda de datos agrupados en clases de igual tamaño su cálculo se puede realizar de la siguiente forma:

$$\text{Moda} = a + \frac{f_2}{f_1 + f_2} \cdot h$$

donde:

a: Extremo inferior de la clase modal

f_1 : frecuencia absoluta de la clase anterior a la modal.

f_2 : frecuencia absoluta de la clase posterior a la modal.

h: amplitud de la clase

En la **TABLA 5 – 6**, LA CLASE MODAL ES LA 299 – 685 Y LA FRECUENCIA ABSOLUTA 13. Luego:

$$\text{Moda} = 299 + \frac{11}{0 + 11} \cdot 386 = 685$$

Interpretación: *El monto de préstamos personales en dólares más frecuente otorgados por una compañía financiera de consumidores es de 685 dólares*

Ventajas: Es la que más fácilmente se determina, puesto que la obtenemos por inspección y no por cómputo.

- **5.6 Posiciones relativas de la Media, la Mediana y la Moda.**

Si la Media, Mediana y Modo se localizan en el centro y son siempre iguales, la distribución es simétrica. Ello significa que si se doblara por la mitad al polígono de frecuencia, ambos lados tienen la misma forma.

El punto más alto de la curva, corresponde al modo. Como la curva es simétrica, la mediana corresponde al punto en que la distribución se parte a la mitad. Las frecuencias más altas se compensan con la más baja y así la Media, Mediana y Modo coinciden, lo que significa que cualquiera de las tres medidas *es adecuada* para representar una distribución.

Conforme la distribución se hace menos simétrica o sesgada, la relación entre los tres promedios cambia. En una distribución positivamente sesgada, la Media aritmética es el mayor de los tres promedios. ¿Por qué? Porque la media es más influida que la Moda o la Mediana por valores extremadamente altos. La Mediana suele ser el siguiente promedio en una distribución de frecuencias positivamente sesgada y el Modo el menor de los tres. Si la distribución es muy sesgada no sería bueno emplear a la Media como promedio. La Mediana y el Modo serían más representativos.

Inversamente, en una distribución negativamente sesgada, la Media aritmética es el menor de los tres promedios. Es evidente que la Media se ve influida por unas cuantas observaciones extremadamente bajas. La Mediana es mayor que la Media Aritmética y el Modo es el mayor de los tres promedios. Aquí también si la distribución es muy sesgada, no se debe emplear la Media para representar a los datos.

Medidas de posición no centrales.

- **5.7 Cuartiles**

Los cuarteles de una distribución, como si nombre lo indica, son valores de la variable que dividen al conjunto de datos (ordenados de menor a mayor) en cuatro subconjuntos que contienen la misma cantidad de datos.

Para calcular los cuartiles de una distribución de frecuencias se procede del mismo modo que en el caso de la mediana, salvo que ahora dividiremos a la distribución de la variable en cuatro partes iguales en lugar de dos.

A partir de esta definición es evidente que la mediana coincide con el segundo cuartil. Los cuarteles se simbolizan con la letra Q.

Para ubicar los valores de la variable que se encuentran en cada cuarto de la distribución, veremos un ejemplo sencillo para el caso de datos no agrupados.

Ejemplo:

Supongamos que un veterinario ha registrado los pesos de 8 pollos de seis semanas de vida y ordenó de menor a mayor, obteniendo:

150 - 151 - 152 - 154 - 155 - 156 - 157 - 159 gramos.

La mediana de este conjunto de datos estará posicionada entre el 4° y 5° valor de la serie, siendo:

Mediana = Q_2 = 154,5 gramos

El primer cuartil Q_1 , debe dividir a la primera mitad de la serie en dos partes iguales, por lo cual Q_1 se ubicará entre el 2° y el 3° valor de la serie.

Luego:

Q_1 = 151,5 gramos

Del mismo modo Q_3 , el tercer cuartil, divide a la segunda mitad de la serie en dos partes iguales.

Es decir:

Q_3 = 156,5 gramos

1° Cuarto	2° Cuarto	3° Cuarto	4° Cuarto
150 - 151	152 - 154	155 - 156	157 - 159
↓	↓	↓	
Q_1 = 151,5 gramos	Q_2 = 154,5 gramos	Q_3 = 156,5 gramos	

Interpretación:

Si $Q_1 = 151,5$ gramos significa que el 25 % de los pollos tendrán un peso inferior a 151,5 gramos y el 75 % un peso superior a ese valor.

Si $Q_2 = 154,5$ gramos significa que el 50 % de los pollos tendrán un peso inferior a 154,5 gramos y el 50% restante superior a ese peso.

Si $Q_3 = 156,5$ gramos significa que el 75 % de los pollos tendrán un peso inferior a 156,5 y un 25% será superior a ese peso.

En el caso de trabajar con datos agrupados y la variable es discreta, se procederá de la misma manera que con la mediana. La única diferencia que existe es que, en el caso de ubicar el orden para Q_1 se debe considerar $n/4$ y en el caso de Q_3 se debe utilizar $3n/4$.

Ejemplo:

Número de hijos por familia. **TABLA 5 – 3**

	Número de Hijos	f_j	F_j	
	0	5	5	F_{j-1}
$Q_1 \leftarrow$	1	11	16	F_j
$Q_3 \leftarrow$	2	35	51	F_{j+1}
	3	2	53	
	4	2	55	
	25	1	56	
	TOTAL	56		

Para calcular el primer cuartil Q_1 en el caso de familias urbanas, se establece primero el orden $n/4$.

$$\text{Orden } n/4 = 56/4 = 14$$

$$F_{j-1} = 5 ; \text{ por lo cual } F_{j-1} < n/4 = 14$$

$$Q_1 = X_j = 1$$

Para calcular el cuartil 3, Q_3 , el orden se busca haciendo:

$$3n/4 = 42$$

$$\text{Ahora } F_{j-1} = 16$$

$$Q_3 = X_j = 2$$

Interpretación: En el caso de la variable cantidad de hijos en familias urbanas se tiene:

$Q_1 = 1$ hijo, es decir que el 25% de la población tiene un solo hijo y, cuando $Q_3 = 2$ significa que el 75% de la población tiene 2 hijos.

El problema se vuelve a presentar cuando se deben calcular los cuartiles para datos agrupados continuos donde también se aplican fórmulas de interpolación.

En este caso las fórmulas son:

$$1^{\circ} \text{ Cuartil} = X_{j-1} + c \frac{(n/4 - F_{j-1})}{F_j - F_{j-1}}$$

Donde:

$n/4$: es el número total de observaciones dividido por 4.

F_{j-1} : es la mayor de las frecuencias acumuladas que no supera a $n/4$

F_j : es la frecuencia acumulada que le sigue a F_{j-1}

X_{j-1} : es el extremo inferior del intervalo que tiene como frecuencia acumulada F .

h : amplitud de dicho intervalo

$$3^{\circ} \text{ Cuartil} = X_{j-1} + h \frac{(3n/4 - F_{j-1})}{F_j - F_{j-1}}$$

Calcularemos el primer y tercer cuartil en el ejemplo de la longitud de los tornillos :

Como el primer cuartil Q_1 está posicionado en el primer cuarto de la distribución, el orden surge de dividir al número total de observaciones por 4.

$$\text{Orden: } n/4 = 100/4 = 25$$

A continuación se debe buscar en la columna de la frecuencia absoluta acumulada a la mayor de dichas frecuencias que no supera al valor $n/4 = 25$

Esta frecuencia se simboliza como F_{j-1} .

En nuestro ejemplo : $F_{j-1} = 20$

Por extensión, $F = 34$.

El extremo inferior del intervalo que contiene a la frecuencia F_j se simboliza con X_{j-1} y en este caso es igual a 8.

La amplitud de dicho intervalo es $h=1$.

Habiendo calculado todos estos datos, ya se puede aplicar la fórmula de interpolación que permitirá obtener el de Q_1 .

$$Q_1 = 8 + 1 \frac{25 - 20}{34 - 20} = 8,36 \text{ mm}$$

Interpretación: Este valor indica que el 25% de los tornillos miden menos de 8,36 mm mientras que el 75% restante mide más de 8,36mm.

TABLA 5-4

Intervalos de Clase	n_j	N_j
6 – 7	11	11
7 – 8	9	20
8 – 9	14	34
9 – 10	11	45
10 – 11	22	67
11 – 12	14	81
12 – 13	7	88
13 – 14	5	93
14 – 15	4	97
15 – 16	3	100
TOTAL	100	

Diagrama de anotaciones en la tabla:

- En la fila 8-9, el valor 8 en el intervalo "8 – 9" está circulado y etiquetado como X_{j-1} con una flecha que apunta a la izquierda.
- En la fila 8-9, el valor 34 en la columna N_j está circulado y etiquetado como F_j con una flecha que apunta a la derecha.
- En la fila 10-11, el valor 67 en la columna N_j está en un recuadro y etiquetado como F_{j-1} con una flecha que apunta a la derecha.
- En la fila 11-12, el valor 11 en el intervalo "11 – 12" está en un recuadro y etiquetado como X_{j-1} con una flecha que apunta a la izquierda.
- En la fila 11-12, el valor 81 en la columna N_j está en un recuadro y etiquetado como F_j con una flecha que apunta a la derecha.

Como el tercer cuartil Q_3 está posicionado en el tercer cuarto de la distribución, el orden surge de multiplicar por 3 al número total de observaciones y dividir por 4.

Orden: $3n/4 = 3 \cdot 100/4 = 75$

A continuación, se debe buscar en la columna de la frecuencia absoluta acumulada a la mayor de dichas frecuencias que no supera el valor $3n/4 = 75$.

Esta frecuencia se simboliza con F_{j-1}

En nuestro ejemplo : $F_{j-1} = 67$

Por extensión: $F_j = 81$

El extremo inferior del intervalo que contiene a la frecuencia F_j se simboliza con X_{j-1} y en este caso es igual a 11.

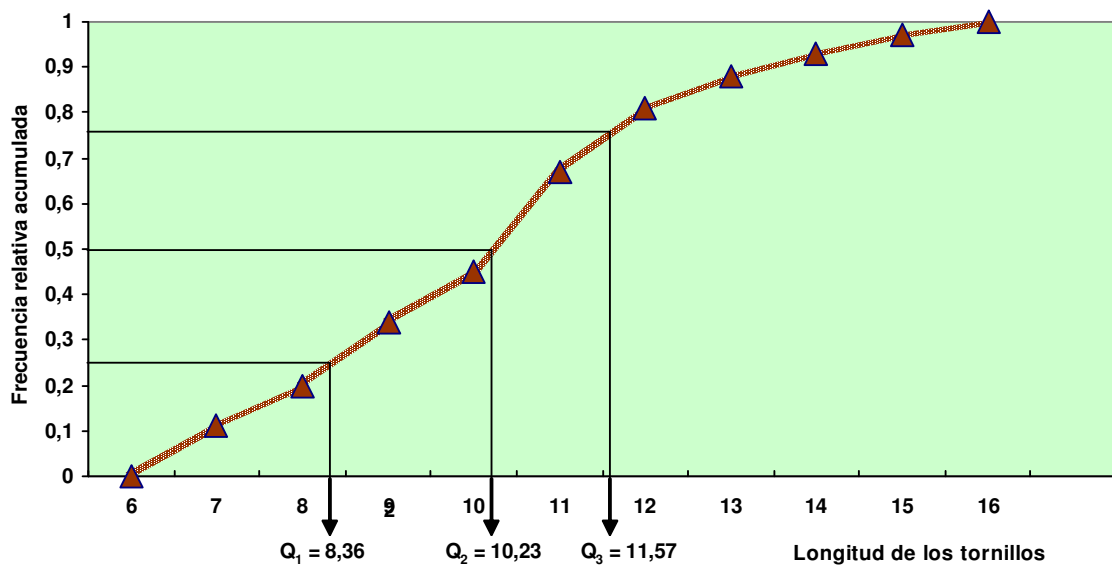
La amplitud de dicho intervalo es $h = 1$

Habiendo calculado todos estos datos, ya se puede aplicar la fórmula de interpolación que permitirá obtener el valor de Q_3 .

$$Q_1 = 11 + 1 \cdot \frac{75 - 67}{81 - 67} = 11,57 \text{ mm}$$

Interpretación: Este valor indica que el 75% de los tornillos miden menos de 11,57 mm mientras que el 25% restante mide más de 11,57mm.

Podemos resumir los resultados obtenidos mediante la utilización de un polígono de frecuencias relativas a los efectos de visualizar la información aportada por los cuartiles.



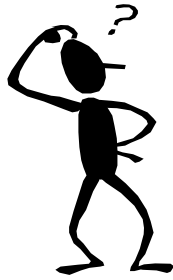
• 5.8 Percentiles

Los percentiles de una distribución, como su nombre lo indica, son valores de la variable, que dividen al conjunto de datos (ordenados de menor a mayor) en cien partes iguales. Los percentiles tienen el mismo significado y la misma forma de cálculo que los cuartiles. Así, cuando se habla del percentil 15 se quiere expresar que es el valor de la variable que deja el 15% de los datos a su izquierda y el 85 % de los mismos a su derecha o lo que es lo mismo decir que es el valor de la variable que deja al 15 % de los datos por debajo de él y el 85% por encima.

Así, por ejemplo, el percentil 80 es el valor de la variable que es igual o deja por debajo de sí, al 80% del total de los datos.

Importante recordar

$$P_{50} = Q_2 = \text{Mediana.}$$



5.9. Vamos a Resolver ejercicios.....

Ejercicio 5 – 1 : En una sección de una fábrica se obtuvieron las siguientes cantidades que representan el número de unidades producidas por día por cada uno de los 10 obreros de esta sección: 5 - 4 - 6 - 5 - 5 - 4 - 6 30 - 30 - 5

- 1 Calcular la Media aritmética, la mediana y la moda.
2. ¿Cuál de las tres medidas es menos representativa de los datos? Y Por qué?.

Ejercicio 5 – 2: Los siguientes datos nos muestran el número de piezas producidas por una máquina diariamente:

34 - 42 - 46 - 36 - 60 - 44 - 43 - 30 - 50 - 62 - 48 - 50 - 45 - 36 - 43 - 32 - 68 - 51 - 35 - 34 - 50 - 60 - 53 - 30 - 50

- 1) Construya una Tabla con 6 intervalos de clase.
- 2) Determine la media aritmética, la mediana y la moda.

Ejercicio 5 – 3: Se registró la estatura de cien estudiantes de un colegio secundario y luego se agruparon los datos de la siguiente manera en la siguiente Tabla de distribución de Frecuencias:

Intervalos de clase	Cantidad de estudiantes
1,20 - 1,30	5
1,30 - 1,40	18
1,40 - 1,50	42
1,50 - 1,60	27
1,60 - 1,70	8

- a) Calcular la Media, Mediana y Moda.
- b) Calcular Q_1 y Q_3



Clave de respuestas

Ejercicio 5 - 1

1. Cálculo de la Media:

$$\text{Media } \frac{5 + 4 + 6 + 5 + 5 + 4 + 6 + 30 + 30 + 5}{10} = \frac{100}{10} = 10$$

Cálculo de la Mediana:

Los ordenamos de menor a mayor:

4 4 5 5 5 5 6 6 30 30
 ↓ ↓

Como son 10 valores, es decir n es par, la Mediana es la semisuma de los valores centrales.
Luego:

$$\text{Mediana} = \frac{5 + 5}{2} = 5$$

Moda: es el valor 5 pues es el que mayor número de veces aparece.

2. La Media aritmética es la menos representativa, pues está afectada por los valores extremos.

Ejercicio 5 - 2

1) $\text{Rango} = 68 - 30 = 38$

$$\text{Ancho} = \frac{38}{6} = 6,33 = 7$$

$$\text{Ancho Real} = 6 \cdot 7 = 42$$

Sobran 4, pues $42 - 38 = 4$ (los repartimos entre el 1° y el último intervalo). Luego, el primer intervalo tendrá como extremo inferior $30 - 2 = 28$, le sumo la amplitud 7 y obtengo 35 y así sucesivamente.

Intervalo de clase	Conteo	f	f _r	F	X _m
28 – 35	////	5	0,20	5	31,5
35 – 42	///	3	0,12	8	38,5
42 – 49	////////	7	0,28	15	45,5
49 – 56	////////	6	0,24	21	52,5
56 – 63	///	3	0,12	24	59,5
63 – 70	/	1	0,04	25	66,5
Total		25	1		

2) Media, Mediana y Moda

$$\text{Media} = \frac{31,5 \cdot 5 + 38,5 \cdot 3 + 45,5 \cdot 7 + 52,5 \cdot 6 + 59,5 \cdot 3 + 66,5 \cdot 1}{25} = \frac{1151,5}{25} = 46,06$$

$$\text{Mediana: Para la mediana: Orden} = \frac{n+1}{2} = \frac{25+1}{2} = 13$$

Ubicándolo en la columna de F, la mediana pertenece a la clase **42 – 49**

a = Extremo inferior de la clase que contiene la mediana = 42

F_a = Frecuencia acumulada correspondiente a la clase anterior a la que contiene la Mediana = 8

h = Amplitud del intervalo de clase = 7

f_c = Frecuencia absoluta de la clase a la que pertenece la mediana = 7.

$$\text{Luego: Mediana} = 42 + \frac{(13 - 8) \cdot 7}{7} = 42 + \frac{5 \cdot 7}{7}$$

$$\text{Mediana} = 42 + 5 = \mathbf{47}$$

Moda: Para la Moda la clase modal es 42 – 49

a = Extremo inferior de la clase modal = 42

h = Amplitud del intervalo de clase = 7

f₁ = Frecuencia absoluta de la clase anterior a la Modal = 3

f₂ = Frecuencia absoluta de la clase posterior a la Modal = 6

$$\text{Luego: Moda} = 42 + \frac{6 \cdot 7}{(3 + 6)} = 42 + \frac{42}{9} = 42 + 4,66 = 46,66$$

Interpretación: La distribución de frecuencias está negativamente sesgada, ya que la media es la menor de los tres promedios. No obstante, los valores de ellos son cercanos, por lo que la Media es adecuada para representar los datos.

Ejercicio 5 - 3

Intervalo de Clase	Frecuencia absoluta f	Marcas de clase o punto medio X_m	f . X_m	Frecuencia Absoluta Acumulada F	Frecuencia relativa f_r	Frecuencia Relativa Acumulada F_r
1,20 – 1,30	5	1,25	6,25	5	0,05	0,05
1,30 - 1,40	18	1,35	24,30	23	0,18	0,23
1,40 – 1,50	42	1,45	60,90	65	0,42	0,65
1,50 – 1,60	27	1,55	41,85	92	0,27	0,92
1,60 – 1,70	8	1,65	13,20	100	0,08	1,00
TOTAL	100		146,50			

$$\text{Media} = \frac{146,5}{100} = 1,465 \text{ m}$$

$$\text{Mediana} = 1,40 + 0,10 \frac{(50 - 23)}{(65 - 23)} = 1,40 + 0,064 = 1,464 \text{ m}$$

$$\text{Modo} = 1,40 + 0,10 \frac{(42 - 18)}{(42 - 18) + (42 - 27)} = 1,40 + 0,061 = 1,461 \text{ m}$$

$$Q_1 = 1,40 + 0,10 \frac{(25 - 23)}{(65 - 23)} = 1,40 + 0,005 = 1,405 \text{ m}$$

$$Q_3 = 1,50 + 0,10 \frac{(75 - 65)}{(92 - 65)} = 1,40 + 0,037 = 1,537 \text{ m}$$

Interpretación: La distribución de frecuencias está positivamente sesgada, ya que la media es el mayor de los tres promedios. No obstante, los valores de ellos son cercanos, por lo que la Media es adecuada para representar los datos.

El cuartil N° 1 expresa que el 25% de las observaciones se encuentran por debajo del valor 1,405m y el 75 % por encima de ese valor.

El cuartil N° 3 expresa que el 75% de las observaciones se encuentran por debajo del valor 1,537m y el 25 % por encima de ese valor.