Medidas Descriptivas Numéricas

Frecuentemente una colección de datos se puede reducir a una o unas cuantas medidas numéricas sencillas que resumen al conjunto total. Tales medidas son más fáciles de comprender que el conjunto de datos originales o ya agrupados. Tres características importantes de los datos que las medidas numéricas ponen de manifiesto son:

1. El valor central o típico de los datos
2. La dispersión de los datos
3. La forma de la distribución de los datos

Medidas de Posición o Localización (tendencia central)

Las medidas de posición se utilizan para indicar un valor que tiende a tipificar o a ser el más representativo de un conjunto de datos. Las tres medidas que más comúnmente se emplean son la media, la mediana y la moda.

1. **Media**

a. ***Media Aritmética***

La media aritmética es lo que viene a la mente de la mayoría de las personas cuando se menciona la palabra "promedio". Como este término tiene ciertas propiedades matemáticas deseables, es la más importante de las medidas de tendencia central. La media aritmética se calcula al sumar los datos y al dividir este resultado entre el número de valores.

Ejemplo:
Si un granjero quiere conocer el peso promedio de sus ocho cerdos cuyos pesos en kilogramos son: 172, 177, 178, 173, 177, 174, 176, 173; realizará el siguiente cálculo:

$$\frac{172+177+178+173+177+174+176+173}{8}=\frac{1400}{8}=175$$

Es decir, el peso promedio de esos cerdos es 175 Kg.

Dada una colección de datos representada por $x_1, x_2, \ldots ,x_n$, la media aritmética de una muestra se denotará por el símbolo $\overline{x}$ (que se lee "equis barra"), y su calculo se puede expresar matemáticamente como:

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

El procedimiento para calcular la media aritmética es el mismo, independientemente si un conjunto de datos se refiere a las observaciones de la muestra o a todos lo valores de la población. Sin embargo, se utiliza el símbolo μ para la media de una población y N para el número de elementos en la misma:

$$\mu=\frac{\sum_{i=1}^{N}x_i}{N}$$

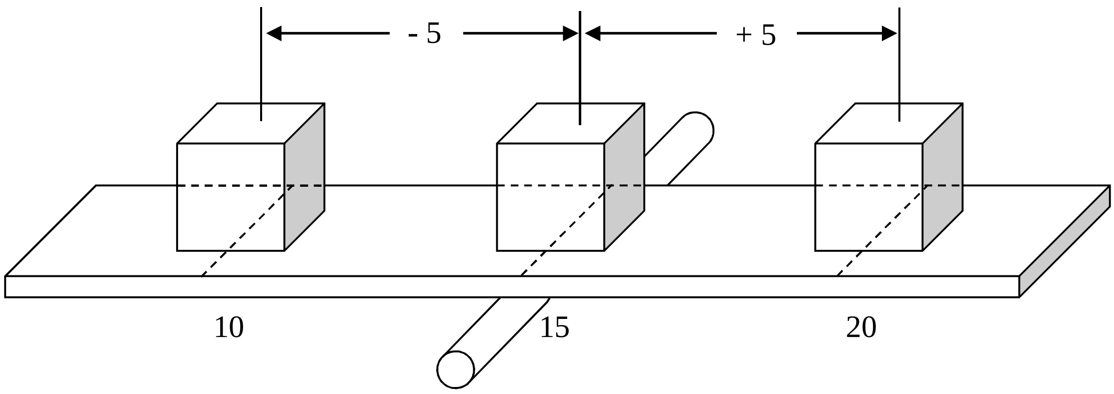
Nota

i. La media aritmética viene expresada en las mismas unidades que los datos originales.

ii. La media aritmética no tiene que coincidir con alguno de los datos de la colección. Como se observa en el último ejemplo el valor $\overline{x}$ =175 Kg. no aparece en los pesos del grupo de cerdos.

iii. Quizás la manera más adecuada de interpretar la media aritmética sea la que se hace desde el punto de vista de la física, en el sentido de que la media de una serie de datos representa el *centro de gravedad* o *punto de equilibrio* de esos datos.

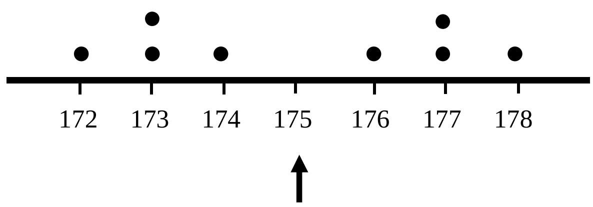
Una representación física de la media es imaginar una barra con un punto de apoyo central que sostiene pesos iguales en sitios correspondientes a los valores de un conjunto. La media de los números 10, 15 y 20 se puede ilustrar como se observa en la siguiente figura:



Nótese como la media es el *punto de equilibrio* de la tabla; las diferencias positivas y negativas se contrabalancean entre sí.

En el último ejemplo también podemos observar visualmente el punto de equilibrio o centro de gravedad de esos datos:

$\overline{x}$ =175 Kg. constituye el punto en donde se logra el equilibrio.



Nota

No debe interpretarse la media como punto medio de los datos. La media representa el punto de equilibrio de las observaciones, el cual no tiene que ser igual al punto medio. En el gráfico anterior el punto de equilibrio coincide con el punto medio debido a que esos datos se distribuyen *simétricamente*.

Ejercicio:
Para los datos no agrupados, de estudio en clase, calcule la media aritmética para las variables peso, número de hermanos, visitas a la discoteca, visitas al cine, estatura e ingreso mensual del hogar.

b. *Media Ponderada*
La fórmula de la *media aritmética* supone que cada observación es de igual importancia. Habitualmente, suele suceder así, sin embargo, existen algunas excepciones. Por ejemplo, un profesor informa a su clase que efectuará cuatro parciales. Estos, con respecto a la calificación final del curso equivalen a:
Parcial 1: 20%, Parcial 2: 30%, Parcial 3: 20% y Parcial 4: 30%
El cálculo de la media deberá considerar las diferentes *ponderaciones* de los exámenes. Se conoce como *peso o ponderación* a los factores cuantitativos que modifican a cada uno de los datos.
La media ponderada de una colección de datos $x_1, x_2, \ldots, x_n$, cuyas respectivas ponderaciones son $w_1, w_2, \ldots, w_n$ se define como:

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

Así un alumno que logre las siguientes calificaciones:

Evaluación	Calificación	Ponderación
1	15	0,30
2	12	0,20
3	19	0,20
4	12	0,30

$$\bar{x} = \frac{0{,}30(15) + 0{,}20(12) + 0{,}20(19) + 0{,}30(12)}{0{,}30 + 0{,}20 + 0{,}20 + 0{,}30} = 14{,}3$$

Obtendrá un promedio de 14,3 puntos. Si todas las evaluaciones poseen la misma importancia, entonces el promedio sería 14,5 puntos. ¿Por qué?

Ejemplo:
Supóngase que el semestre anterior un estudiante cursó Matemática I, Inglés, Métodos Estadísticos I y Sociología, obteniendo las siguientes calificaciones:

Materia	Unidades Crédito	Calificación
Matemáticas I	6	10
Sociología	3	16
Métodos Estadísticos I	5	13
Inglés	3	20

Así el promedio ponderado del estudiante fue de:

$$\overline{x}_p = \frac{6(10) + 3(16) + 5(13) + 3(20)}{6+3+5+3} = 13{,}71 \text{ puntos}$$

y su promedio aritmético simple:

$$\overline{x} = \frac{10+16+13+20}{4} = 14{,}75 \text{ puntos}$$

¿A qué se debe que los dos promedios anteriores sean distintos?

c. ***Media aritmética para datos agrupados en distribuciones de frecuencias***
Es posible utilizar una variante de la fórmula para calcular la media ponderada, a fin de obtener la media de una distribución de frecuencias. Las ponderaciones son sustituidas por las frecuencias absolutas simples y la fórmula se convierte en:

$$\overline{x} = \frac{\sum_{i=1}^{n} f_i m_i}{n}$$

Ejercicio:
Calcular la media aritmética para las distribuciones de frecuencias de las variables peso, visitas a la discoteca, estatura, número de hermanos, visitas al cine e ingreso mensual del hogar.

Nota:
En el caso de una *distribución de frecuencias para valores individuales de la variable*, mediante la fórmula se obtendrá la misma respuesta como si se trabajara con datos originales. Si las clases de la distribución de frecuencias son intervalos, el agrupamiento hace que se pierda información y por tanto la media resultante es una aproximación. El uso de los puntos medios de clase (marcas de clase) los considera como promedios de clase, que representan a la clase respectiva, lo cual no siempre se cumple. Sin embargo, si no se dispone de datos originales, no existe otra alternativa razonable. Además la aproximación de esta fórmula a la verdadera media es generalmente buena.

Propiedades de la media aritmética

La media aritmética presenta ciertas propiedades útiles e interesantes, que explican por qué es la medida de tendencia central que se utiliza más ampliamente.
Sea $x_1, x_2, \ldots, x_n$, una colección de datos cuya media aritmética es $\overline{x}$, entonces se cumple que:

i. La suma de las desviaciones o diferencias de cada uno de los datos con respecto a su media, es cero:

$$\sum_{i=1}^{n} \left(x_i - \overline{x} \right) = 0$$

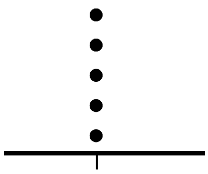
Ejemplo:
En el ejemplo de los pesos de los cerdos se obtuvo que la media aritmética es 175 Kg. Ahora calculando las desviaciones con respecto a $\overline{x}$ =175 se tiene que:

172 - 175 =	- 3
177 - 175 =	+2
178 - 175 =	+3
173 - 175 =	-2
177 - 175 =	+2
174 - 175 =	-1
176 - 175 =	+1
173 - 175 =	-2
	0

ii. $\sum_{i=1}^{n}\left(x_i-\overline{x}\right)^2$ es un valor mínimo.

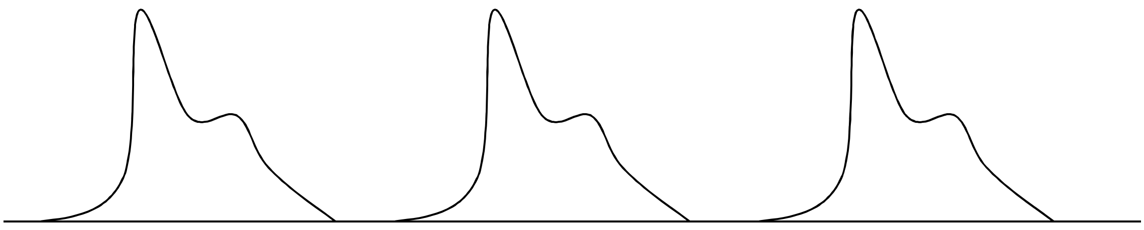
Si se calcula la expresión anterior sustituyendo $\overline{x}$ por cualquier otro valor arbitrario que se nos ocurra, se obtiene un valor mayor al que se consigue utilizando $\overline{x}$.

iii. Si todos los datos son iguales a un mismo valor fijo o constante *c*, entonces la media de esos datos también es igual a *c*:



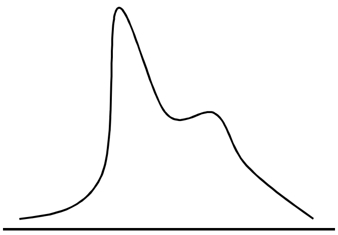
iv. Si a cada uno de los datos originales se le suma un mismo número real *c*, entonces se tiene una nueva colección de datos $x_1 + c$, $x_2 + c$, ... ,$x_n + c$, cuya media viene dada por $\overline{x} + c$.

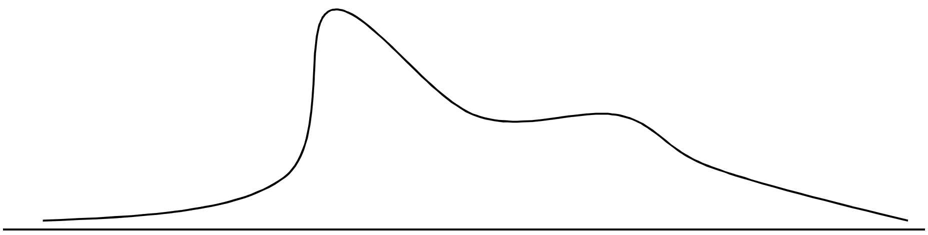
Esta situación se puede visualizar gráficamente de la siguiente manera:



Al sumar la misma constante a cada uno de los datos, realmente lo que estamos haciendo es desplazar sobre el eje horizontal los datos hacia la derecha si la constante es positiva o hacia la izquierda si la constante es negativa. Entonces la media aritmética se "corre" con los datos.

v. Si cada uno de los datos originales se multiplica por un mismo número real c, entonces se genera una nueva colección de datos $x_1 * c, x_2 * c, \ldots, x_n * c$, cuya media viene dada por $\overline{x} * c$.
En la siguiente ilustración se puede observar como se ensancha la distribución de los datos originales cuando estos han sido modificados al multiplicar cada uno por una constante, con lo cual la media se ve afectada.





vi. Si se tienen m diferentes grupos de datos de distintos tamaños $n_1, n_2, \ldots , n_m$ respectivamente, entonces la media de todos esos datos juntos viene dada por:

$$\overline{\overline{x}} = \frac{\sum_{i=1}^{m} n_i \overline{x}_i}{\sum_{i=1}^{m} n_i}$$

***Nota*:**

Obsérvese que: $\overline{\overline{x}} \neq \dfrac{\overline{x}_1 + \overline{x}_2 + \cdots + \overline{x}_m}{m}$

Ejemplo:
Si en un semestre un estudiante aprobó sus cuatro materias con 15 puntos ¿Cuál fue su calificación promedio?
De acuerdo a la propiedad iii. la media aritmética de sus calificaciones fue de 15 puntos.
Ejemplo:

Haciendo referencia al ejemplo de los pesos de los cerdos, suponga que al granjero le han recomendado un nuevo alimento para cerdos que según parece los engorda 20 Kg. en quince días. ¿Cuál será el peso promedio de los cerdos dentro de quince días, luego de utilizar el nuevo alimento?
Nótese que todos los cerdos aumentan 20 Kg., así que a cada uno de los pesos originales se le debe sumar la constante $c = 20$. En consecuencia, de acuerdo a la propiedad iv. dentro de quince días el peso promedio de los cerdos debe ser 175+20 = 195 Kg.

Ejemplo:
Suponga ahora que todos los cerdos del granjero se enferman a causa de un virus y se detecta cinco días después que todos estos animales han disminuido exactamente 10 Kg. ¿cuál es ahora el peso promedio de los cerdos?

Ejemplo:
Las secciones 03 y 05 de la asignatura Estadística I, tienen 66 y 73 alumnos respectivamente. Se realiza la primera evaluación y se obtienen las siguientes notas promedio por sección: $\overline{x}_1 = 15$ y $\overline{x}_2 = 12$. Entonces la nota promedio del primer parcial para las dos secciones juntas es:

$$\overline{\overline{x}} = \frac{66*15+73*12}{139} = 13,42 \text{ puntos}$$

¡OJO es falso que:! $\overline{\overline{x}} = \frac{12+15}{2} = 13,5$

Ejemplo:
Si en el ejemplo de los cerdos, se incluye otro de esos animales cuyo peso es de 490 Kg., Calcule la media aritmética.

$$\overline{x} = \frac{172+177+178+173+177+174+176+173+490}{9} = \frac{1890}{9} = 210$$

Este valor 210 Kg. transmite una idea equivocada de la realidad en cuanto al peso de la mayoría en ese grupo de cerdos. ¿Qué es lo que provoca que $\overline{x}$ no sea representativa de los pesos de los cerdos?
En estos casos no debe utilizarse la media para calcular el peso promedio, sino que se recomiendan otras medidas de tendencia central.

Desventajas de la media aritmética

- No puede calcularse cuando los datos están agrupados en distribuciones de frecuencias que tienen un intervalo de clase abierto.
- La principal desventaja es que se ve afectada por la presencia de *valores extremos o atípicos* en los datos.

Ventajas de la media aritmética

- Es un promedio que toma en cuenta todos los valores de una colección de datos.
- Es fácil de calcular y se presta a operaciones algebraicas, lo que la convierte en la medida de tendencia central más utilizada tanto en estudios descriptivos como para realizar inferencias.
- En general, para una serie dada de datos existe una buena aproximación entre el valor de la media para los datos no agrupados y la media de los datos agrupados.

2. **Mediana**

La mediana de una colección de datos, que previamente han sido *ordenados*, es aquél valor más central o que está más en medio en el conjunto de datos. En otras palabras, la mediana es *mayor* que aproximadamente la mitad de los datos y *menor* que (aproximadamente) la otra mitad. Así se tiene que aproximadamente 50% de las observaciones se encuentran por arriba y 50% (aproximadamente) por debajo de ella. La mediana se denota *Md* (también algunos autores la denotan como $\tilde{X}$.

Ejemplo:
Los tiempos de los miembros de un equipo de atletismo en una carrera de 1,6 Km están dados en la siguiente tabla, calcule la mediana.

Miembro	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Tiempo (en minutos)	4.2	9.0	4.7	5.0	4.3	5.1	4.8

En primer lugar se deben ordenar los datos: 4.2 4.3 4.7 4.8 5.0 5.1 9.0.

Mediana

Md = 4.8 minutos, es el valor que está en el centro de los datos.

Ejemplo:
Calcule la mediana para el número de pacientes tratados en la sala de emergencias de un hospital durante ocho días consecutivos:

Día	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
No. de pacientes	49	52	86	30	35	31	43	11

Los datos ordenados son: 86 52 49 43 35 31 30 11

Centro de los datos

La mediana en este caso puede ser 43 ó 35, o también cualquier valor entre 43 y 35. Para evitar esta imprecisión, se acepta tomar como mediana la suma de los dos valores centrales y se dividen entre dos: $Md = \frac{43+35}{2} = 39$.

Nota:
Si se tienen n observaciones ordenadas, la mediana es la observación que ocupa la posición $\frac{n+1}{2}$ cuando n es impar y la media de las observaciones que ocupan las posiciones $\frac{n}{2}$ y $\frac{n+2}{2}$ cuando n es par.

Ejemplo:
Regresando al ejemplo de los tiempos del equipo de atletismo, se pide calcular la media y comparar este resultado con el de la mediana ya obtenida.

Entonces, se obtiene que $\bar{x}$ = 5.3 minutos y antes se obtuvo que *Md = 4.8* minutos. Nótese que en esos datos existe un *valor atípico*: 9.0 minutos. Por tanto, la media aritmética $\bar{x}$ se distorsiona. La mediana, en cambio, *no se ve distorsionada* por la presencia del valor 9.0. Este valor pudo haber sido 15.0 o incluso 45.0 y la mediana ¡seguirá siendo la misma!

Cálculo de la mediana para datos agrupados en distribuciones de frecuencias

i. Cuando las clases son intervalos

- Se ubica la *clase medianal*, la cual viene dada por aquella clase que contiene a la frecuencia acumulada $n/2$ o equivalentemente a la frecuencia relativa acumulada 0,5.
- Luego de ubicada la clase medianal, el cálculo de la mediana se hace mediante un proceso de interpolación el cual conduce a la siguiente fórmula:

$$Md = LI_m + \left(\frac{\frac{n}{2} - F_{am}}{f_m} \right) * C_m$$

en donde,
LI_m: Límite inferior de la clase medianal
n: No. total de observaciones o datos
F_{am}: Frecuencia acumulada anterior a la clase medianal
f_m: Frecuencia absoluta de la clase medianal
c_m: Amplitud de la clase medianal

Ejemplo: Calcular la mediana para la distribución de frecuencias de la variable peso.

En primer lugar se debe ubicar la clase medianal, para esto se debe calcular:

$$\frac{n}{2} = \frac{43}{2} = 21{,}5$$

Ahora se ubica la frecuencia acumulada que contiene a 21,5:

Clase medianal

Clases	mi	fi	fri	Fi	Fri
[40-49)	44,5	4	0,0930	4	0,0930
[49-58)	53,5	10	0,2326	14	0,3256
[58-67)	62,5	15	0,3488	29	0,6744
[67-76)	71,5	7	0,1628	36	0,8372
[76-85)	80,5	5	0,1163	41	0,9535
[85-94)	89,5	1	0,0233	42	0,9767
[94-103)	98,5	1	0,0233	43	1
Totales		**43**	**1**		

Frecuencia acumulada que contiene a 21,5

También se puede ubicar la clase medianal encontrando la frecuencia relativa acumulada que contiene a 0,5000.

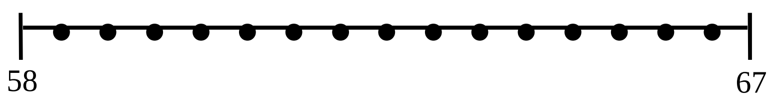
Entonces, se tiene que:

$$Md = 58 + \left(\frac{21{,}5 - 14}{15}\right) * 9$$

$$Md = 62{,}5$$

De esta manera, Md = 62,5 Kg. representa el valor central de los pesos. Es decir, aproximadamente la mitad de los estudiantes de Métodos Estadísticos I tienen un peso inferior a 62,5 Kg. y aproximadamente la otra mitad pesa más de 62,5 Kg.

Nota:
En la fórmula de la mediana se está suponiendo que los valores en el intervalo de clase que contiene la mediana están *uniformemente espaciados (o equidistantes)*. Entonces, en el ejemplo anterior se está suponiendo que los 15 valores que contiene la clase medianal están uniformemente espaciados en [58 -67):



<u>Ejercicio</u>:
Calcule la mediana para las distribuciones de frecuencias correspondientes a las variables estatura, índice académico e ingreso mensual del hogar.

ii. Cuando las clases son valores individuales

- Se calcula $n/2$ (o se considera el valor 50% de las observaciones)
- Si el valor $n/2$ **NO APARECE** en la columna de la F_i, entonces se ubica aquella frecuencia acumulada que lo contiene y la mediana será el valor de la variable correspondiente a esa frecuencia acumulada.

- Una forma equivalente de hacer lo anterior es la siguiente, si el valor 50% no aparece en la columna de las Fr_i * 100 entonces se ubica aquella frecuencia que lo contenga y la mediana será el valor de la variable correspondiente a esa clase.
- Si el valor $n/2$ **APARECE** en la columna de las F_i , es decir que coincide con la frecuencia acumulada de alguna clase, entonces la mediana viene dada por la media aritmética de ese valor de la variable y el siguiente valor.
- También, si el valor 50% coincide con alguna de las Fr_i * 100, entonces la mediana viene dada por el promedio de los valores de la variable correspondiente a esa clase y a la siguiente.

Ejemplo:
La siguiente distribución de frecuencias corresponde al número de materias que cursan 112 estudiantes de la carrera de Contaduría Pública. Calcule la mediana.

Inicialmente se debe calcular $\frac{n}{2} = 56$. Entonces 56 no aparece en la columna de las F_i.

Por tanto, Md = 4 materias.

Número de materias	***fi***	***fri***	***Fi***	***Fri***
1	1	0,0089	1	0,0089
2	1	0,0089	2	0,0179
$\boxed{3}$	12	0,1071	14	0,1250
4	56	0,5000	70	0,6250
5	40	0,3571	110	0,9821
6	2	0,0179	112	1,0000
Totales	**112**	**1,0000**		

Mediana

Frecuencia acumulada que contiene a $\frac{n}{2} = 56$

Ejemplo:
Calcule la mediana para la siguiente distribución de frecuencias, en donde $\frac{n}{2} = 30$. Es decir, $n/2$ aparece en la columna de las frecuencias acumuladas:

Clase	$\mathbf{f_i}$	$\mathbf{F_i}$	$\mathbf{Fr_i}$
5	8	8	0,1333
6	9	17	0,2833
7	13	30	0,5000
8	10	40	0,6667
9	6	46	0,7667
10	14	60	1

n = 60

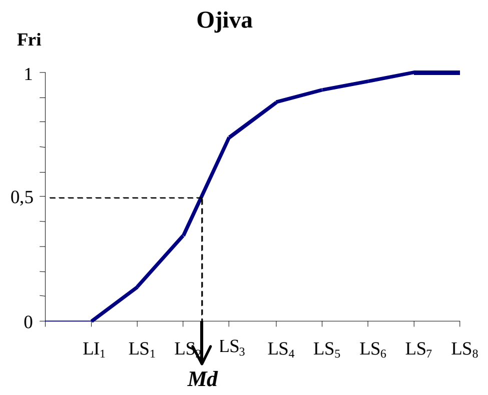
Mediana

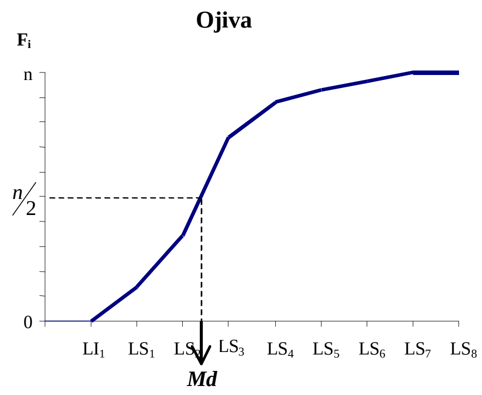
Frecuencia acumulada que coincide con $\frac{n}{2} = 30$

Entonces la mediana viene dada por: $Md = \dfrac{7+8}{2} = 7,5$

La mediana gráficamente

Mediante la ojiva y a través del método de interpolación visto en esa sección se puede obtener de manera gráfica el valor de la mediana de una colección de datos agrupados en una distribución de frecuencias cuyas clases son intervalos. Si se usa la ojiva construida con la frecuencia acumulada Fi la mediana será aquél valor en el eje horizontal cuya ordenada sea $n/2$. En el caso de usar la ojiva construida con la frecuencia relativa acumulada Fri (o Fri*100), la mediana vendrá dada por el valor en el eje de las abscisas que corresponda a la ordenada 0,5 (o 50%).





Así aplicando el método de interpolación visto antes se obtiene la fórmula del cálculo de la mediana:

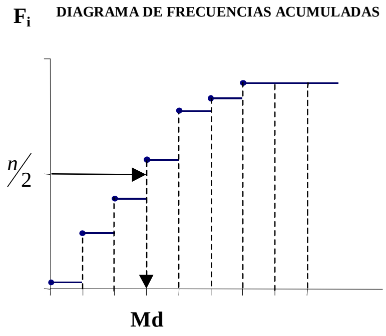
$$Md = LI_m + \left(\frac{\frac{n}{2} - F_{am}}{f_m} \right) * C_m$$

<u>Ejercicio</u>:

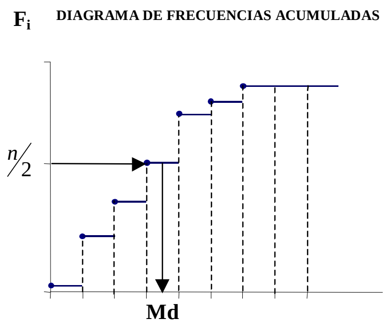
Obtenga gráficamente la fórmula anterior para el cálculo de la mediana.

Para el caso de distribuciones de frecuencias cuyas clases son valores individuales de la variable, se puede hallar gráficamente la mediana por medio del diagrama de frecuencias acumuladas. El procedimiento es similar que cuando se usa la ojiva. Se ubica en el eje vertical $n/2$ (o 50% si se usó Fri*100) y se traza una línea paralela al eje horizontal, así se presentan las dos situaciones siguientes:

- Si la línea intercepta el gráfico, entonces la mediana viene dada por el valor en el eje de las abscisas que corresponde a la ordenada $n/2$ (o 50%).



- Si la línea coincide con uno de los escalones del gráfico, la mediana vendrá dada por el punto medio de ese escalón.



Propiedades de la mediana

i. La mediana es una medida de tendencia central de fácil comprensión pero que solamente toma en cuenta la posición que ocupan las observaciones y no el valor en sí de las mismas. Esto hace que la mediana no sea susceptible de operaciones algebraicas y en consecuencia limita su utilidad, por ejemplo para fines de inferencia estadística.

ii. Puede calcularse en el caso de distribuciones de frecuencias con clases abiertas siempre y cuando se disponga de la información correspondiente a la clase medianal.

iii. No se ve afectada ante la presencia de unos pocos *valores atípicos* y es por ello que se recomienda su uso en el caso de distribuciones marcadamente asimétricas.

3. **Moda**

La moda es el valor que más se repite, es decir el que aparece con mayor frecuencia. En otras palabras la moda es el valor más común de los datos, se denota por *Mo* y viene expresada en las mismas unidades que los datos.

Ejemplo:
Calcule la moda de los siguientes datos: 5, 3, 6, 5, 4, 5, 2, 4.
En este caso el valor que más se repite es el 5, por tanto *Mo* = 5.

Ejemplo:
Calcule la moda de los siguientes datos: 5, 3, 6, 5, 4, 5, 2, 4, 4.
En este conjunto de datos existen dos valores que se repiten con la misma frecuencia: 4 y 5. Así, se tienen dos modas: $Mo_1 = 4$ y $Mo_2 = 5$.

Ejemplo:
Calcule la moda de los siguientes datos: 5, 3, 3, 5, 6, 2, 6, 4, 2, 4.
En este caso no existe la moda dado que no hay datos que se repitan más que otros.

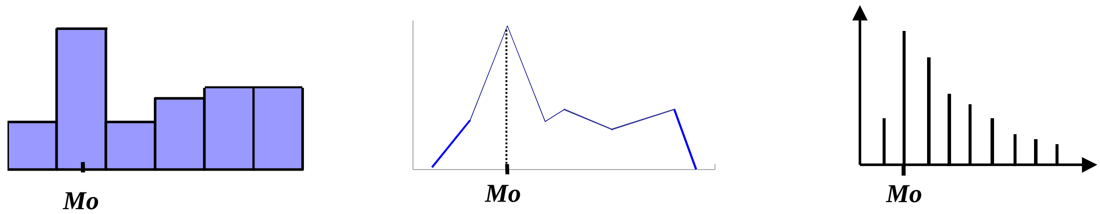
En conclusión, una colección de datos puede que no tenga moda o puede ser que posea una o más modas.

Nota:
Cuando hay una sola moda la distribución de datos se llama *unimodal*, con dos modas *bimodal*, con tres modas *trimodal* y con 4 o más modas se llama *polimodal* o *multimodal*. Si todos los valores se presentan la misma cantidad de veces, la distribución se llama *amodal*.

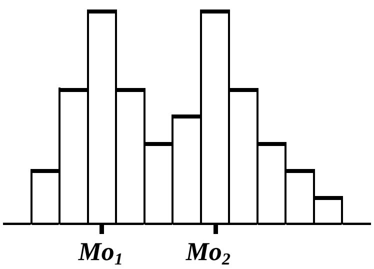
Cuando los datos están agrupados en distribuciones de frecuencias cuyas clases presenten igual amplitud, se toma el *punto medio* de la clase con mayor frecuencia absoluta como la moda.

Representación gráfica de la moda:

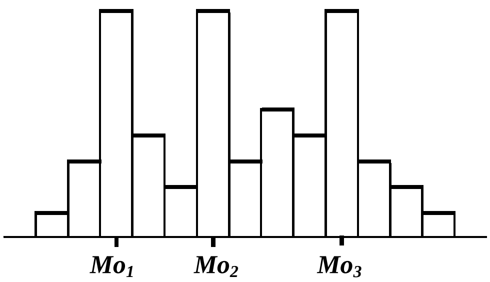
Distribuciones unimodales:



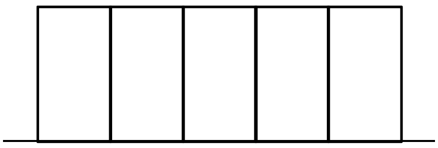
Distribución bimodal:



Distribución trimodal:



Distribución amodal:



Ejercicio:
Calcular la moda para las distribuciones de frecuencias correspondientes a las variables peso, número de hermanos, estatura, ingreso mensual del hogar, número de veces que visita la discoteca e índice académico.

Propiedades de la Moda:

i. La moda en realidad no es una medida de tendencia central, sino más bien indica punto(s) de concentración de datos.
ii. No es susceptible de operaciones algebraicas y de allí que su uso es limitado.
iii. Es la única de las medidas descriptivas que puede utilizarse para datos cualitativos de cualquier tipo.
iv. Es posible su cálculo en algunos casos de distribuciones de frecuencias con intervalos de clase abiertos.
v. Es una medida muy imprecisa e inestable. En una distribución de frecuencias depende de la forma en como se construyen las clases.

Ejemplo:
Considere la siguiente distribución de frecuencias:

Clases	f_i
[0 - 5)	3
[5 - 10)	5
[10 - 15)	6
[15 - 20)	6
[20 - 25)	4
[25 - 30)	7
[30 - 35)	2
Total	**33**

La clase modal es [25 - 30) y la moda es *Mo* = 27,5

Si se introduce una pequeña modificación en las clases, por ejemplo agrupando las dos primeras, se tiene:

Clases	f_i
[0 - 10)	8
[10 - 15)	6
[15 - 20)	6
[20 - 25)	4
[25 - 30)	7
[30 - 35)	2
Total	**33**

La clase modal pasa a ser [0 - 10) y Mo = 5. Obsérvese el cambio tan grande que se produce en la moda ya que pasa de 27,5 a 5.

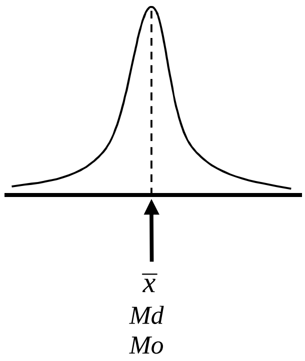
vi. La moda es de utilidad en aquellos casos donde la naturaleza de los datos así lo indique.

Ejemplo:
Para una fábrica de zapatos, el interés está en conocer la o las tallas más frecuentes en la población.

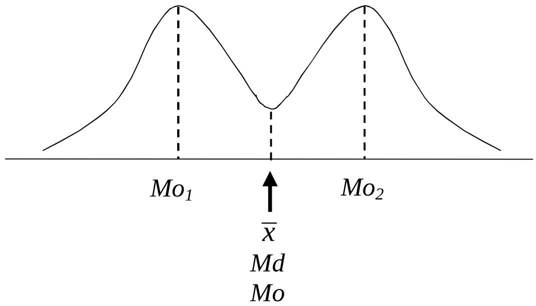
Relación entre la Media Aritmética, la Mediana y la Moda

En función de la simetría de una distribución se presentan las siguientes relaciones entre esas tres medidas:

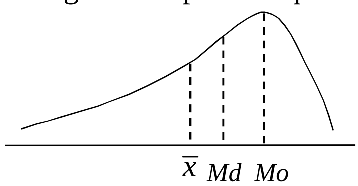
1. En distribuciones simétricas unimodales la media, la mediana y moda coinciden:



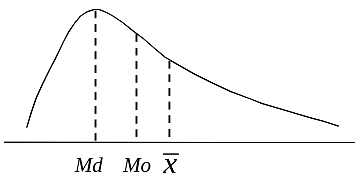
2. En distribuciones simétricas bimodales, la media y la mediana son iguales pero no coinciden con las modas.



3. En distribuciones asimétricas negativas o por la izquierda, se cumple que $\overline{x} < Md < Mo$



4. En distribuciones asimétricas positivas o por la derecha, se cumple que $\overline{x} > Md > Mo$



Selección de la Medida de Tendencia Central adecuada
Los siguientes factores deben tomarse en cuenta en el momento de la selección de la medida numérica apropiada para describir la posición o tendencia central de los datos:

1. De acuerdo al tipo de dato se puede utilizar una u otra medida de tendencia central. Las medidas que pueden aplicarse con cada tipo de dato son las siguientes:
   i. Datos Nominales: Moda
   ii. Datos Ordinales: Moda y Mediana
   iii. Datos Discretos: Todas
   iv. Datos Continuos: Todas

2. Teniendo en cuenta lo anterior se recomienda tener presente los siguientes aspectos:
   a. ***La naturaleza de la distribución de los datos***. Gráficamente se puede observar la forma general en que se distribuyen los datos. Esto es determinante en la selección del promedio adecuado.
      - Si se trata de una *distribución simétrica* o aproximadamente simétrica, se sabe que la media, la mediana y la moda coinciden y en consecuencia se puede utilizar cualquiera de ellas.
      - Si la *distribución es asimétrica*, la media aritmética no va a ser adecuada y es preferible inclinarse por la moda o la mediana.

   b. ***El concepto de tendencia central o de posición que interese reflejar en una situación dada.***
      - Si interesa conocer el valor más común de una serie de datos como por ejemplo la estatura típica de las personas que ingresan al ejército, es necesario usar *la moda*.
      - Si se desea ubicar a una persona en cuanto a su salario anual diciendo que gana por encima o por debajo de lo que gana la mitad de los trabajadores del país, entonces habrá que usar *la mediana***.**
      - Cuando interesa el total de datos o reflejar el punto de equilibrio de los mismos se utiliza *la media aritmética*.

   c. ***Riesgos que se corren ante la presencia de valores atípicos.***
      Si existen valores atípicos, hay que verificar si se incurrió en algún error en la recolección de la información o puede ser el alerta de alguna situación no esperada por el investigador. En todo caso hay que tener presente que la media aritmética se ve seriamente afectada ante la presencia de valores atípicos y será necesario recurrir a alguna de las otras medidas conocidas.

   d. ***Posibilidad de realizar inferencia estadística***
      Cuando el análisis estadístico se realiza sobre una muestra de la población con la intención de generalizar a la totalidad, lo que se conoce como inferencia estadística, prácticamente la única medida de tendencia central utilizada hasta ahora satisfactoriamente es la media aritmética y esto se debe a que existe un fundamento teórico bien fundamentado que la respalda.