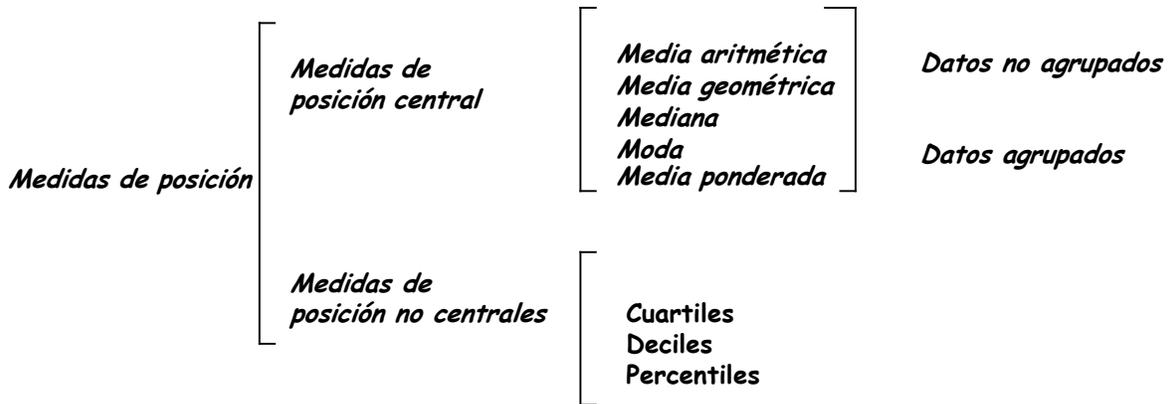


## Medidas de posición

Las medidas de posición nos facilitan información sobre la serie de datos que estamos analizando. Estas medidas permiten conocer diversas características de esta serie de datos.

Las **medidas de posición** son de dos tipos:

- a) **Medidas de posición central:** informan sobre los valores medios de la serie de datos. Con ellas se localiza el dato central de un conjunto de datos, el dato al alrededor del cual giran todos los demás, el dato que equilibra el momento de los datos más pequeños con el momento de los datos mayores.
- b) **Medidas de posición no centrales:** informan como se distribuye el resto de los valores de la serie.



## Medidas de variabilidad (dispersión)

Al igual que sucede con cualquier conjunto de datos, la media, la mediana y la moda sólo nos revelan una parte de la información que necesitamos acerca de las características de los datos. Para aumentar nuestro entendimiento del patrón de los datos, debemos medir también su dispersión, extensión o variabilidad. Son métodos para conocer que tan separados se encuentran los datos con respecto a la media aritmética que es el punto de equilibrio del grupo. Los valores de estas medidas de dispersión serán mayores cuando los datos estén muy disgregados, y serán menores cuando los datos estén más cercanamente agrupados. Se necesitan métodos para medir el grado en que los datos se dispersan o se diseminan, y las medidas estadísticas que proporcionan esta información son las medidas de dispersión o variación.

La dispersión es importante porque:

1. Proporciona información adicional que permite juzgar la confiabilidad de la medida de tendencia central. Si los datos se encuentran ampliamente dispersos, la posición central es menos representativa de los datos.
2. Ya que existen problemas característicos para datos ampliamente dispersos, debemos ser capaces de distinguir que presentan esa dispersión antes de abordar esos problemas.

3. Quizá se desee comparar las dispersiones de diferentes muestras. Si no se desea tener una amplia dispersión de valores con respecto al centro de distribución o esto presenta riesgos inaceptables, necesitamos tener habilidad de reconocerlo y evitar escoger distribuciones que tengan las dispersiones más grandes.

<i>Medidas de dispersión</i>	<i>Rango</i>	<i>Datos no agrupados</i>
	<i>Varianza</i>	<i>Datos agrupados</i>
	<i>Desviación estándar</i>	

## **Medidas de posición y variabilidad para datos no agrupados**

**Media:** es el valor medio de la serie de datos. Se pueden calcular diversos tipos de media, siendo las más utilizadas la media aritmética y la geométrica. Lo más positivo de la media es que en su cálculo se utilizan todos los valores de la serie, por lo que no se pierde ninguna información. Sin embargo, presenta el problema de que su valor (tanto en el caso de la media aritmética como geométrica) se puede ver muy influido por valores extremos, que se aparten en exceso del resto de la serie. Estos valores anómalos podrían condicionar en gran medida el valor de la media, perdiendo ésta representatividad. Según el tipo de datos que se analice será más apropiado utilizar la media aritmética o la media geométrica. En todo caso, la media aritmética es la medida de posición central más utilizada.

### **Media (media aritmética) ( $\bar{X}$ )**

Es la suma de todos los datos y dividido por el total de datos de la muestra: Es la medida de localización usada con mayor frecuencia se conoce simplemente como la media.

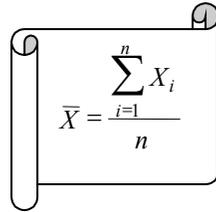
#### Ventajas:

1. Se trata de un concepto familiar para la mayoría de las personas y es intuitivamente claro.
2. Cada conjunto de datos tiene una media, es una medida que puede calcularse y es única debido a que cada conjunto de datos posee una y sólo una media.
3. Es útil para llevar a cabo procedimientos estadísticos como la comparación de medias de varios conjuntos de datos.

#### Desventajas:

1. Puede verse afectada por valores extremos que no son representativos del resto de los datos.
2. Resulta tedioso calcular la media debido a que utilizamos cada uno de los puntos de dato de nuestro cálculo.

3. *Somos incapaces de calcular la media para un conjunto de datos que tiene clases de extremo abierto, ya sea en el inferior o en el superior de la escala.*



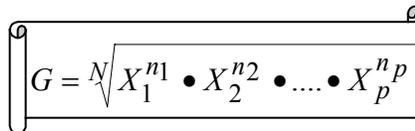
$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

datos

Donde:  $X_i$  representa cada uno de los  
 $n$  es el número total de datos

## Media geométrica (G)

La media geométrica se suele utilizar en series de datos como tipos de interés anuales, inflación, etc., donde el valor de cada año tiene un efecto multiplicativo sobre el de los años anteriores. Se eleva cada valor al número de veces que se ha repetido. Se multiplican todos estos resultados y al producto final se le calcula la raíz "n" (siendo "N" el total de datos de la muestra).



$$G = \sqrt[N]{X_1^{n_1} \cdot X_2^{n_2} \cdot \dots \cdot X_p^{n_p}}$$

## Mediana ( $\tilde{m}$ )

La mediana es el valor central de los datos, es decir, supuesta la muestra ordenada en orden creciente o decreciente, el valor que divide en dos partes la muestra. Es el valor de la serie de datos que se sitúa justamente en el centro de la muestra (un 50% de valores son inferiores y otro 50% son superiores).

No presentan el problema de estar influido por los valores extremos, pero en cambio no utiliza en su cálculo toda la información de la serie de datos (no pondera cada valor por el número de veces que se ha repetido). Es mejor usar la mediana que la media cuando se trata de un conjunto de datos en el cual existen datos con valores extremos.

### Ventajas y desventajas de la mediana:

Los valores extremos no afectan a la mediana tan intensamente como a la media. La mediana es fácil de entender y se puede calcular a partir de cualquier tipo de datos incluso a partir de datos agrupados con clases de extremo abierto a menos que la mediana entre en una clase de extremo abierto.

Podemos encontrar la mediana incluso cuando nuestros datos son descripciones cualitativas, en lugar de números.

Ciertos procedimientos estadísticos que utilizan la mediana son más complejos que aquellos que utilizan la media e implica consumo de tiempo para cualquier conjunto de datos que contenga un gran número de elementos. Por consiguiente, si deseamos utilizar una estadística de muestra para estimar un parámetro de población, la media es más fácil de usar que la mediana.

Si el tamaño de la muestra es  $n$  y todos los datos se ordenan en forma ascendente

Si  $n$  es **Impar**, hay un término central y la mediana es el valor del elemento intermedio, es decir el término  $X_{\frac{n+1}{2}}$  será el valor de la mediana.

Si  $N$  es **Par**, hay dos términos centrales, la mediana es el valor promedio de los dos elementos intermedios,  $X_{\frac{n}{2}}$ ,  $X_{\frac{n}{2}+1}$  es decir la mediana será la media de esos dos valores

*Ejemplo*

<i>n es par</i>	<i>n es impar</i>
1,4,6,7,8,9,12,16,20, 24,25,27	1,4,6,7,8,9,12,16,20, 24,25,27,30
$N=12$	$N=13$
Términos Centrales el 6° y 7° 9 y 12	Término Central el 7°, 12
$\bar{m} = \frac{9+12}{2} = 10.5$	$\bar{m} = 12$



**Moda (Mo)**

La moda es el valor de la variable que tenga mayor frecuencia absoluta, la que más se repite, es la única medida de centralización que tiene sentido estudiar en una variable cualitativa, pues no precisa la realización de ningún cálculo.

Si bien a simple vista no se ve la centralidad de la moda, debemos indicar que en un grupo normal, de tendencia a un grupo central, la mayoría de los datos se encuentran cercanos a este punto central, por lo que se presume que el dato que más se repite estará cercano al punto central. No existe moda si todos los valores son diferentes o si se presentan el mismo número de veces. Si dos o más ocurren el mismo número de veces (con mayor frecuencia que las demás observaciones) entonces existe más de una moda y la muestra se puede considerar como

polimodal). Si el dato que más se repite no es uno sino dos datos, se dice que el espacio es bimodal, si tres trimodal,... o en general el espacio es polimodal

#### Ventajas y desventajas de la moda:

La moda, al igual que la mediana, se puede utilizar como una posición central para datos tanto cualitativos como cuantitativos.

También, al igual que la mediana, la moda no se ve mayormente afectada por los valores extremos. Incluso si los valores extremos son muy altos o muy bajos, nosotros escogemos el valor más frecuente del conjunto de datos como el valor modal. Podemos utilizar la moda sin importar qué tan grandes o qué tan pequeños sean los valores del conjunto de datos, e independientemente de cuál sea su dispersión.

La podemos utilizar aun cuando una o más clases sean de extremo abierto.

Muy a menudo, no existe un valor modal debido a que el conjunto de datos no contiene valores que se presenten más de una vez. En otras ocasiones, cada valor es la moda, pues cada uno de ellos se presenta el mismo número de veces. Otra desventaja consiste en que cuando los datos contienen dos, tres o más modas, resultan difíciles de interpretar y comparar.

#### **Comparación entre la media, la mediana y la moda.**

Cuando trabajamos un problema de estadística, debemos decidir si vamos a utilizar la media, la mediana o la moda como medidas de tendencia central. Las distribuciones simétricas que sólo contienen una moda, siempre tienen el mismo valor para la media, la mediana y la moda. En tales casos, no es necesario escoger la medida de tendencia central, pues ya está hecha la selección.

En una distribución positivamente sesgada (es decir, sesgada hacia la derecha), la moda todavía se encuentra en el punto más alto de la distribución, la mediana está hacia la derecha de la moda y la media se encuentra todavía más a la derecha de la moda y la mediana.

En una distribución negativamente sesgada, la moda sigue siendo el punto más alto de la distribución, la mediana está hacia la izquierda de ella y la media se encuentra todavía más a la izquierda de la moda y la mediana.

Cuando la población está sesgada negativa o positivamente, con frecuencia la mediana resulta ser la mejor medida de posición, debido a que siempre está entre la moda y la media. La mediana no se ve altamente influida por la frecuencia de aparición de un solo valor como es el caso de la moda, ni se distorsiona con la presencia de valores extremos como la media.

## **Media ponderada ó pesada ( $\bar{x}_w$ )**

Es un promedio de un conjunto de datos, pero donde te permitís definir el grado de importancia al aporte de ese promedio. La media pesada nos permite calcular, el promedio que toma en cuenta la importancia de cada valor con respecto al total. Los promedios pesados toman en

$$\bar{x}_w = \frac{\sum (w \times x)}{\sum w}$$

*cuenta las diferentes cantidades de cada nivel de trabajo que se utiliza en la elaboración de los productos. Además toma en cuenta no nada más el valor de cada de dato, sino también con qué frecuencia éste se presenta en la población.*

*Donde:  $w$  es el peso asignado a cada observación*

*$\sum(w \times x)$  es la suma de los productos del peso de cada elemento por el elemento correspondiente*

*$\sum w$  es la suma de todos los pesos*

### *Ejemplo*

*Para calcular el promedio final de los exámenes de un alumno, son considerados de la siguiente manera, cada examen final es cuatro veces mas importante que los parciales, así si un alumno obtuvo: 69 puntos, 75 puntos, 62 puntos y 73 puntos en parciales mas 78 puntos en final, su promedio ponderado será:  $(69.1 + 75.1 + 62.1 + 73.1 + 78.4) / 5$ .*



### *Ejemplo*

*A continuación se mencionan las materias que Luis Pérez llevó en el primer semestre de Ingeniería Química, el número de créditos y la calificación obtenida. Determina la calificación promedio que obtuvo Luis Pérez en su primer semestre*



<i>MATERIA</i>	<i>NUMERO CREDITOS</i>	<i>DE CALIFICACIÓN</i>
<i>Metodología de la investigación</i>	<i>8</i>	<i>90.5</i>
<i>Matemáticas I</i>	<i>10</i>	<i>100.0</i>
<i>Programación</i>	<i>8</i>	<i>81.0</i>
<i>Química</i>	<i>10</i>	<i>78.0</i>
<i>Dibujo</i>	<i>4</i>	<i>100.0</i>
<i>Economía</i>	<i>8</i>	<i>84.0</i>

$$\tilde{x}_w = \frac{(8)(90.5) + (10)(100) + (8)(81) + (10)(78) + (4)(100) + (8)(84)}{8+10+8+10+4+8} = 88$$

*Nota: Si comparamos este promedio con el que se obtiene usando simplemente la media aritmética, que es un 88.91, nos damos cuenta de que este último es mayor, por no tomar en cuenta el peso o número de créditos que aporta cada materia a la carrera que se estudia, el promedio de esta persona es menor al de la media aritmética debido a que obtiene una calificación baja es Química que es una de las materias que aporta más créditos.*

## Rango (R) [o recorrido, también se le conoce como amplitud o intervalo]

*Mide la amplitud de los valores de la muestra y se calcula por diferencia entre el valor más elevado y el valor más bajo.*

*También llamado desviación total, es la diferencia entre los valores mayor y menor de una muestra. Casi nunca se usa porque se basa sólo en dos de los elementos y, por consiguiente, está muy influido por los valores extremos de los datos.*

$$R = D_M - D_m$$

Donde:  $D_M$  es el dato mayor  
 $D_m$  es el dato menor

## Varianza (s<sup>2</sup>)

*Mide la distancia existente entre los valores de la serie y la media. Se calcula como sumatoria de las diferencias al cuadrado entre cada valor y la media, multiplicadas por el número de veces que se ha repetido cada valor. la sumatoria obtenido se divide por el tamaño de la muestra.*

*Es una medida de la dispersión que emplea todos los valores de los datos. Se basa en la diferencia entre cada valor y la media. La varianza es útil para comparar la dispersión, o variabilidad de dos conjuntos de datos. Se define como el cuadrado de la desviación típica.*

$$s^2 = \frac{\sum_{i=1}^k f_i (M_i - \bar{x})^2}{n - 1}$$

Muestra:

$$\sigma^2 = \frac{\sum_{i=1}^k f_i (M_i - \mu)^2}{N}$$

Población:

La varianza siempre será mayor que cero. Mientras más se aproxima a cero, más concentrados están los valores de la serie alrededor de la media. Por el contrario, mientras mayor sea la varianza, más dispersos están.

## Desviación estándar (S)

La desviación estándar nos permite determinar, con un buen grado de precisión, dónde están localizados los valores de una distribución de frecuencias con relación a la media. El teorema de Chebyshev dice que no importa qué forma tenga la distribución, al menos 75% de los valores caen dentro de  $\pm 2$  desviaciones estándar a partir de la media de la distribución, y al menos 89% de los valores caen dentro de  $\pm 3$  desviaciones estándar a partir de la media.

Con más precisión:

- Aproximadamente 68% de los valores de la población cae dentro de  $\pm 1$  desviación estándar a partir de la media.
- Aproximadamente 95% de los valores estará dentro de  $\pm 2$  desviaciones estándar a partir de la media.
- Aproximadamente 99% de los valores estará en el intervalo que va desde tres desviaciones estándar por debajo de la media hasta tres desviaciones estándar por arriba de la media.

La desviación estándar se define como la raíz cuadrada positiva de la varianza. Nos indica el grado más o menos fuerte de dispersión respecto al valor central "media aritmética". Es más fácil interpretar la desviación estándar que la varianza, ya que está va a tener unidades al cuadrado, mientras la desviación estándar nos da las unidades tal como son.

$$s = \sqrt{s^2}$$

$$\sigma = \sqrt{\sigma^2}$$

*Ejemplo*

Halla la desviación estándar y la varianza de la siguiente serie de datos: 10, 18, 15, 12, 3, 6, 5, 7.



$$\bar{x} = 9.5$$

$$\sigma^2 = \frac{(10-9.5)^2 + (18-9.5)^2 + (15-9.5)^2 + (12-9.5)^2 + (3-9.5)^2 + (6-9.5)^2 + (5-9.5)^2 + (7-9.5)^2}{8} = 23.7169$$

$$\sigma = 4.87$$

*Ejemplo*

Vamos a utilizar la tabla de distribución de frecuencias con los datos de la estatura de los alumnos. Para calcular los valores de las distintas posiciones centrales



Variable (valor)	Frecuencia absoluta		Frecuencia relativa	
	Simple	acumulada	Simple	Acumulada
1.20	1	1	3.3%	3.3%
1.21	4	5	13.3%	16.6%
1.22	4	9	13.3%	30%
1.23	2	11	6.6%	36.6%
1.24	1	12	3.3%	40%
1.25	2	14	6.6%	46.6%
1.26	3	17	10%	56.6%
1.27	3	20	10%	66.6%
1.28	4	24	13.3%	80%
1.29	3	27	10%	90%
1.30	3	30	10%	100%

1.- *Media aritmética:*  $\bar{X} = \frac{1.20 \cdot 1 + 1.21 \cdot 4 + 1.22 \cdot 4 + \dots + 1.30 \cdot 3}{30} = 1.253$

2.- *Media geométrica:*  $G = \sqrt[30]{1.20^1 + 1.21^4 + 1.22^4 + \dots + 1.30^3}$

3.- *Mediana:*

*La mediana de esta muestra es 1,26 cm, ya que por debajo está el 50% de los valores y por arriba el otro 50%. Esto se puede ver al analizar la columna de frecuencias relativas acumuladas.*

*En este ejemplo, como el valor 1,26 se repite en 3 ocasiones, la media se situaría exactamente entre el primer y el segundo valor de este grupo, ya que entre estos dos valores se encuentra la división entre el 50% inferior y el 50% superior.*

4.- *Moda:*

*Hay 3 valores que se repiten en 4 ocasiones: el 1,21, el 1,22 y el 1,28, por lo tanto esta sería cuenta con 3 modas.*

5. *Rango:*  $R = 1.30 - 1.20 = 0.10$

6. *Varianza:*  $s^2 = \frac{(1.20 - 1.253)^2 \cdot 1 + (1.21 - 1.253)^2 \cdot 4 + (1.22 - 1.253)^2 \cdot 4 + \dots + (1.30 - 1.253)^2 \cdot 3}{29} = 0.0329$

7. *Desviación estándar:*  $s = 0.18149$

## Coeficiente de variación de Pearson: (V)

*Se calcula como cociente entre la desviación típica y la media. Una característica de la desviación estándar es que depende de las unidades de medición, por lo que se requiere de una medida de variación relativa, como el coeficiente de variación (V) que expresa la desviación estándar como un porcentaje de la media:*

*Al comparar los coeficientes de variación (que son porcentajes), podemos comparar las dispersiones de dos o más conjuntos de datos que pertenecen a diferentes clases de mediciones, por ejemplo: altura, peso, velocidad, edad, cantidades monetarias, etc...*

$$V = \frac{s}{\bar{x}} \cdot 100 \quad \text{ó} \quad V = \frac{\sigma}{\mu} \cdot 100$$

*El interés del coeficiente de variación es que al ser un porcentaje permite comparar el nivel de dispersión de dos muestras. Esto no ocurre con la desviación típica, ya que viene expresada en las mismas unidades que los datos de la serie.*

*Por ejemplo, para comparar el nivel de dispersión de una serie de datos de la altura de los alumnos de una clase y otra serie con el peso de dichos alumnos, no se puede utilizar las desviaciones típicas (una viene expresada en cm y la otra en kg). En cambio, sus coeficientes de variación son ambos porcentajes, por lo que sí se pueden comparar.*

**Ejemplo**

Se tiene una serie de datos en los que  $\bar{x} = 1.253$  y con  $\sigma = 0.0320$  entonces el Coeficiente de variación de Pearson: se calcula como cociente entre la desviación típica y la media de la muestra.  
 $V = 0.0320 / 1.253 = 0.0255$



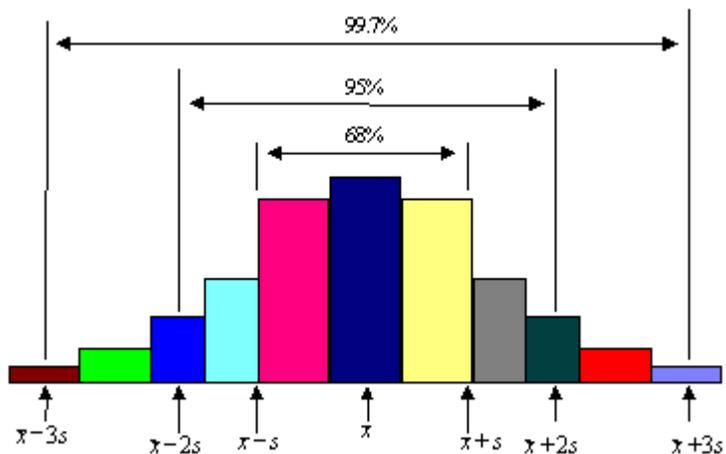
## Teorema de Chebyshev (Tchebycheff)

Para cualquier conjunto de datos (de una población o una muestra) y cualquier constante  $k$  mayor que 1, el porcentaje de los datos que debe caer dentro de  $k$ -veces la desviación típica de cualquier lado de la media es de por lo menos:  $1 - \frac{1}{k^2}$

El teorema de Chebyshev se aplica a cualquier tipo de datos, pero sólo nos indica “por lo menos que porcentaje” debe caer entre ciertos límites. Pero para casi todos los datos, el porcentaje real de datos que cae entre esos límites es bastante mayor que el que especifica el teorema de Chebyshev.

Para las distribuciones que tienen forma de campana puede hacerse una aseveración más fuerte:

- (1) alrededor del 68% de los valores caerán dentro de una desviación típica de la media esto es: entre  $\bar{x} - \sigma, \bar{x} + \sigma$  ;
- (2) aproximadamente el 95% de los valores caerán dentro de dos desviaciones típicas de la media, esto es :  $\bar{x} - 2\sigma, \bar{x} + 2\sigma$  ;



(3) *aproximadamente el 99,7% de los valores caerán dentro de dos desviaciones típicas de la media, esto es :  $\bar{X}-3\sigma$ ,  $\bar{X}+3\sigma$  ;*

