MEDIDAS DESCRIPTIVAS

Las medidas descriptivas son valores numéricos calculados a partir de la población o de una muestra y que nos resumen la información contenida en ella.

MEDIDAS DESCRIPTIVAS

CENTRALIZADAS

Indica valores con respecto a los que los datos parecen agruparse:
Media, mediana y moda.

POSICIÓN

Dividen un conjunto ordenado de datos en grupos con la misma cantidad de individuos:
Percentiles, cuartiles, deciles.

DISPERSIÓN

Indica la mayor o menor concentración de los datos con respecto a las medidas de centralización:
Varianza, desviación típica o estándar, coeficiente de variación, rango, desviación media.

FORMA

Indica hacia donde se encuentra la mayor o menor concentración de los datos con respecto a la media:
Asimetría, curtosis.

Elaborado por: Ing. Joan Betancourt

MEDIA ARITMÉTICA

La media aritmética o media es la medida de tendencia central que frecuentemente llamamos promedio, consiste en la suma de los valores del grupo de datos dividida entre la cantidad de valores. La media aritmética de una población se representa con el símbolo µ (mu), y la media aritmética de una muestra se representa con el símbolo $\overline{X}$ (equis barra):

$$\mu = \frac{\sum X}{N} \qquad \overline{X} = \frac{\sum X}{n}$$

$\sum X$: es la sumatoria de todos los datos
N: es la cantidad de datos de la población
n: es la cantidad de datos de la muestra

Cuando es para datos agrupados:

$$\overline{X} = \frac{\sum fX}{n}$$

X: son los puntos medios de cada intervalo
f: es la frecuencia absoluta de cada intervalo
n: es la cantidad de datos
fX: es la multiplicación de la frecuencia por el punto medio de cada intervalo
$\sum$fX: es la sumatoria del producto de la frecuencia por el punto medio de cada intervalo

Propiedades de la Media Aritmética:

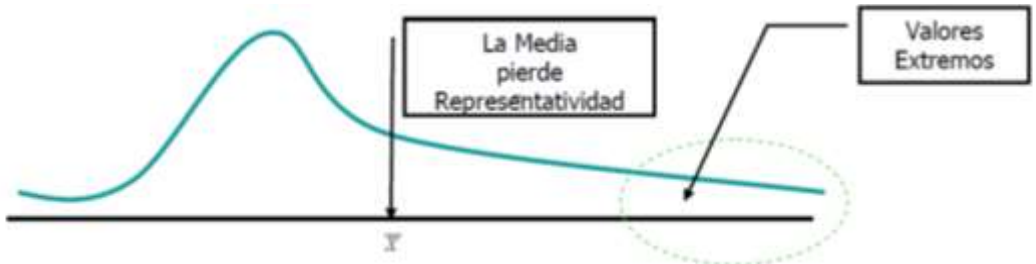
- ✓ Para calcular la media se toman todos los valores
- ✓ Un conjunto de datos sólo tiene una media. La media es única
- ✓ La media es una medida útil para comparar dos o más poblaciones
- ✓ La media se puede hallar sólo para variables cuantitativas.
- ✓ La media es independiente de las amplitudes de los intervalos.
- ✓ La media aritmética es la única medida de posición en la que las suma de las desviaciones de los valores de la media es siempre cero:

$$\sum(X_i - \bar{X}) = 0$$

Ejemplo:
La media de los números 3, 8 y 4 es 5

$$\sum(X - \overline{X}) = (3–5) + (8 –5) + (4 –5) = –2 + 3+ –1 = 0$$

- ✓ La media no se puede calcular si hay un intervalo o clase abierto (con una amplitud indeterminada).
- ✓ La media es un estadístico “suficiente” porque usa toda la información de la muestra.
- ✓ La media es muy sensible a las puntuaciones extremas:



EJEMPLOS:

1.- Durante cada hora de trabajo de un día una cooperativa produce las siguientes cantidades de artículos de limpieza: 14, 19, 20, 15, 12, 18, 16, 10 ¿Cuál es el número medio de unidades producidas?

$$\mu = \frac{\sum X}{N} = \frac{14+19+20+15+12+18+16+10}{8} = \frac{124}{8} = 15{,}50$$

2.- Un pediatra obtuvo la siguiente tabla sobre los meses de edad de 50 niños de su consulta en el momento de empezar a caminar. Desea saber a cuantos meses en promedio los niños empiezan a caminar.

Meses (X)	Niños (f)	X * f
9	1	9
10	4	40
11	9	99
12	16	192
13	11	143
14	8	112
15	1	15
∑ =	**50**	**610**

$$\overline{X} = \frac{\sum fX}{n}$$

$\overline{X}$ = 610 / 50

$\overline{X}$ = 12,2 meses

3.- ¿Cuál es la media del precio de venta de los vehículos del plan Venezuela móvil de acuerdo a la siguiente tabla de datos recolectados?

Precio en millones de Bs.	f	X	f * X
18 a 23	25	20,5	512,5
23 a 28	28	25,5	714
28 a 33	26	30,5	793
33 a 38	17	35,5	603,5
38 a 42	13	40,5	526,5
∑ =	109		3149,5

$$\overline{X} = \frac{\sum fX}{n}$$

$\overline{X}$ = 610 / 50

$\overline{X}$ = 12,2 meses

MEDIA PONDERADA

La media ponderada o promedio ponderado es una media aritmética en al que cada uno de los valores se le pondera de acuerdo a su importancia con el grupo general.

$$\mu_w \text{ ó } \overline{X}_w = \frac{\sum(wX)}{\sum w}$$

$\overline{X}$w: Media Ponderada
X: Observación individual
w: Peso o ponderación asignada a cada observación

Cuando calculamos la media aritmética no sale a discusión si cada uno de los datos tiene igual importancia, sin embargo en ciertos casos puede ocurrir que determinados datos tengan más valor que otro de su mismo conjunto.

EJEMPLO:

Un estudiante obtuvo las siguientes calificaciones en su curso de estadística I: 19, 20, 18 y 16. ¿Cuál es la media ponderada? A continuación se reflejan los datos en la siguiente tabla:

Calificaciones (X)	Ponderación (w)	X * w
9	10	90
10	15	150
8	12	96
6	20	120
	57	456

$$\mu_w \text{ ó } \overline{X}_w = \frac{\sum(wX)}{\sum w}$$

$\overline{X}$ w = 456 / 57

$\overline{X}$ w = 8

MEDIA GEOMÉTRICA

La media geométrica es útil para encontrar el promedio de porcentajes, proporciones, índices o tasas de crecimiento. Tiene mucha aplicación en el comercio y en la economía debido a que nos interesa encontrar el porcentaje de cambio en ventas, salarios o cualquier otro dato económico, también es apropiada para promediar porcentajes o medidas que aumentan a una razón constante. La media de un conjunto n de números positivos se define como la n-ésima raíz del producto de los n valores. La fórmula de la media geométrica se escribe así:

$$MG = \sqrt[n]{(x_1)(x_2)\ldots(x_n)}$$

La media geométrica será siempre menor o igual a la media aritmética, pero nunca mayor.
Antes de escribir las tasas o porcentajes dentro del signo radical, se le debe aplicar a cada tasa el ***factor de crecimiento***: 1 + (tasa / 100). Por ejemplo, si una tasa es de 35%, para escribirla dentro del signo radical, se le aplica el factor de crecimiento: 1 + (35 / 100), el resultado seria 1,35 (este es el valor que se escribiría dentro del signo radical). Si la tasa es negativa (indica una perdida), por ejemplo una tasa de 25% de perdida, el cual se escribiría -35%, al aplicarle el factor de crecimiento, quedaría de la siguiente manera: 1 + (-35 / 100), el resultado seria 0,65.

EJEMPLO:
La recuperación de una inversión que realizó la compañía Alfa durante cuatro años consecutivos fue de 30%, 20%, - 40% y 200%. ¿Cuál es la media geométrica de la recuperación de la inversión?
El número 1.3 representa 30% de la recuperación de la inversión, que es la inversión original de 1.0 más la recuperación de 0.3. El número 0.6 representa la pérdida de 40%, que es la inversión original de 1.0 menos la pérdida de 0.4. Este cálculo supone que el total de la inversión de cada periodo se reinvierte o se convierte en la base de la siguiente. En otras palabras, la base del segundo periodo es 1.3 y la base del tercer periodo es (1.3)(1.2) y así sucesivamente. En consecuencia, la media geométrica de la tasa de recuperación es de 29.4%, que se determina por medio del siguiente cálculo:
Al aplicar el factor de crecimiento a cada una de las tasa, los valores que van dentro del signo radical quedaron de la siguiente manera:

$$MG = \sqrt[4]{(1{,}3)(1{,}2)(0{,}6)(3)} = \sqrt[4]{2{,}808} = 1{,}294$$

A este resultado se le resta 1 y se divide entre 100, quedando en 29,4%

Otro modelo de aplicación de la media geométrica se relaciona con la determinación de un cambio porcentual promedio durante cierto periodo. Por ejemplo, si usted ganó $30 000 en 2000 y $50 000 en 2010, ¿cuál es la tasa anual de incremento durante el periodo? Ésta es de 5.24%. La tasa de incremento se determina a partir de la siguiente fórmula:

$$MG = \sqrt[n]{\frac{valor\ al\ final\ delm\ periodo}{valor\ al\ inicio\ del\ periodo}} - 1$$

Donde “n” es el número de periodos.

EJEMPLO
Durante la década de los noventa y hasta los primeros años de 2000, Las Vegas, Nevada, fue la ciudad de mayor crecimiento en Estados Unidos. La población se incrementó de 258 295 en 1990 a 607 876 en 2009. Es un incremento de 349 581 personas o 135.3% durante el periodo. ¿Cuál es el incremento anual promedio?

Hay 19 años entre 1990 y 2009, así que n = 19.

$$MG = \sqrt[19]{\frac{6007876}{258295}} - 1 \ = 1{,}0461 - 1 = 0{,}0461$$

Luego el valor obtenido se multiplica por 100:

0.0461 x 100 = 4,61%

La población de Las Vegas creció a una tasa de **4.61%** por año de 1990 a 2009.

EJEMPLO

Una inversión de $100 crece a razón del 5% anual durante 7 años, de manera que su valor al final de cada uno de estos 7 años seria:

Fin del año	Valor	
1	100(1,05) =	105
2	105(1,05) =	110,25
3	110,25(1,05) =	115,76
4	115,76(1,05) =	121,55
5	121,55(1,05) =	127,63
6	127,63(1,05) =	134,01
7	134,01(1,05) =	140,71
	Total =	854,91

Al aplicar el factor de crecimiento a 5%, da como resultado **1,05**
La media geométrica seria:

$$MG = \sqrt[7]{(105)(110{,}25)(115{,}76)(121{,}55)(127{,}63)(134{,}01)(140{,}71)} = 121{,}55$$

EJEMPLO
El CEO (director ejecutivo) de White-Knuckle Airlines desea determinar la tasa de crecimiento promedio en los ingresos con base en las cifras dadas en la tabla. Si la tasa de crecimiento promedio es menor que el promedio industrial del 10%, se asumirá una nueva campaña publicitaria.

Ingresos para White-Knuckle Airlines		
Año	**Ingreso ($)**	**Porcentaje del año anterior**
1992	50000	
1993	55000	55/50 = 1,10
1994	66000	66/55 = 1,20
1995	60000	60/66 = 0,91
1996	78000	78/60 = 1,30

Primero es necesario determinar el porcentaje que los ingresos de cada año representan respecto de los obtenidos el año anterior. En otras palabras, ¿qué porcentaje del ingreso de 1992 es el ingreso en 1993? Esto se encuentra dividiendo los ingresos de 1992 entre los de 1993. El resultado, 1.10 revela que los ingresos de 1993 son 110% de los ingresos de 1992. También se calculan los porcentajes para los tres años restantes. Tomando la Media Geométrica (MG) de estos porcentajes da:

$$MG = \sqrt[4]{(1{,}10)(1{,}20)(0{,}91)(1{,}30)} = 1{,}1179$$

Al restarle 1 y multiplicarlo por 100, la tasa de crecimiento queda en 11,79%. En conclusión como la tasa resultante es mayor que el 10% del promedio industrial, la nueva campaña publicitaria no se llevara a cabo.

MEDIANA

Se define como "aquel valor de la variable que supera a no más de la mitad de las observaciones, al mismo tiempo, es superado por no más de la mitad de las observaciones", en otras palabras se puede definir como el "valor central".
También se puede decir que es el valor que divide al conjunto en dos conjuntos de igual tamaño, o bien, es el promedio de los dos valores centrales.

Cálculo en datos sin agrupar
a.- Número impar de observaciones: lo primero que se debe hacer es ordenarlos de menor a mayor o de mayor a menor. La mediana se obtiene con la siguiente ecuación:

$$Me = (n + 1) / 2$$

EJEMPLO:
Calculemos la mediana de los kilos (ordenados de forma ascendente) de materia prima utilizadas durante esta semana: 33, 36, 40, 45, 57,60 y 68.
Me = (7 + 1) / 2 = 4
La mediana es el valor que está en la posición ***4***: 33, 36, 40, **45**, 57,60 y 68.

b.- Número par de observaciones: una vez ordenado los datos, encontramos dos valores en el centro de la serie, por tal razón, la mediana deberá ser el promedio de ellos. Aplicamos la misma ecuación del caso anterior.

$$Me = (n + 1) / 2$$

EJEMPLO

Datos: 10, 15, 18, 25, 31, 36, 45, 60, 77, 80. Calcular la mediana.

Me = (10 + 1) / 2 = 5,5

El punto 5,5 estaría entre los valores de las posiciones 5 y 6, por lo buscamos ambos valores y los promediamos 10, 15, 18, 25, **32, 36**, 45, 60, 77, 80

Se calcula la media entre los dos números:

(32 + 36) / 2 = 86/2 = 43

Cálculo en datos agrupados

$$Me = Li + \frac{N/2 - fa_{i-1}}{f} * a$$

L_i: límite inferior del intervalo de la clase modal

N: es la cantidad de datos

f: frecuencia absoluta del intervalo de la clase modal

fa_{i-1}: es la frecuencia absoluta acumulada que esta antes de la frecuencia absoluta acumulada de la clase modal

a: es la amplitud del intervalo

Clase modal: para hallarla, ubicar en la columna de la frecuencia absoluta (fa), empezando desde la frecuencia absoluta menor, el primer número mayor o igual que el resultado de dividir N/2. Una vez ubicado ese número, seleccionar la fila donde está el número y es ahí donde ubicaremos la clase modal.

EJEMPLO

Calcular la mediana en la siguiente tabla de distribución de frecuencia:

Calcular N al sumar todas las frecuencias absolutas: N = 50

Dividimos N entre 2: 50/2 = 25

Ubicamos el número mayor o igual a 25 en la columna de fa: es 35

Li	Ls	f	fa
45	55	6	6
55	65	10	16
65	**75**	19	35
75	85	11	46
85	95	4	50

Límite inferior del intervalo (Li)

fa_{i-1}

Primer Número encontrado mayor que 25

Clase modal

La amplitud del intervalo es de 75 – 65 = 5

$$Me = Li + \frac{{}^{N}/_{2} - fa_{i-1}}{f} * a$$

Me = 65 + [(25 – 16)/19] * 5 = **67,35**

MODA

Es la medida de tendencia central que se define como el valor que se presenta con mayor frecuencia, es decir el más común.

La moda para datos no agrupados presenta los siguientes casos:

Caso 1:

2, 3, 3, 4, 4, 4, 5, 7, 8, 9. Moda = 4.

Caso 2:

2, 5, 5, 6, 6, 7, 9, 16. Moda = 6, 5.

$$Mo = Li + \frac{f - f_{i-1}}{(f - f_{i-1}) + (f - f_{i+1})} * a$$

L_i: límite inferior del intervalo de la clase modal

f: frecuencia absoluta del intervalo de la clase modal

f_{i-1}: es la frecuencia absoluta que esta ante de la de la clase modal

f_{i+1}: es la frecuencia absoluta que esta después de la de la clase modal

a: es la amplitud del intervalo

Clase modal: es el intervalo que tiene la mayor frecuencia absoluta

EJEMPLO

Calcular la moda en la siguiente tabla de distribución de frecuencia:

Ubicamos la clase modal, que es la fila que tiene la mayor frecuencia.

Li	Ls	f	fa
45	55	6	6
55	65	10	16
65	75	19	35
75	85	11	46
85	95	4	50

Límite inferior del intervalo (Li)

Clase modal

f_{i-1}

f

f_{i+1}

Mo = $Li + \frac{f - f_{i-1}}{(f - f_{i-1}) + (f - f_{i+1})} * a$

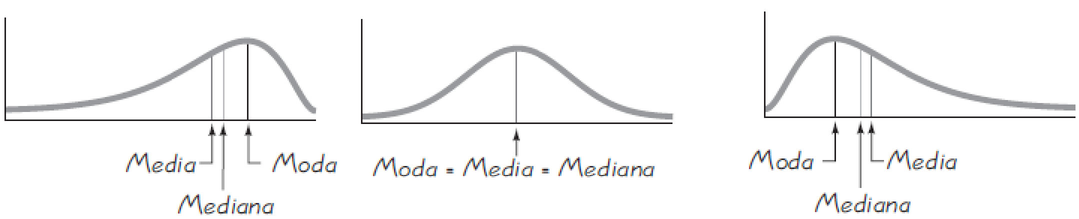
Mo = $65 + \frac{19 - 10}{(19 - 10) + (19 - 11)} * 5$ = **67,65**

RELACIÓN ENTRE LAS MEDIDAS DE TENDENCIA CENTRAL O PROMEDIOS.

$$\bar{X} - \text{Moda} \approx 3\,(\bar{X} - \text{Mediana})$$

SESGO

Una distribución de datos está sesgada si no es simétrica y se extiende más hacia un lado que hacia el otro. (Una distribución de datos es simétrica si la mitad izquierda de su histograma es aproximadamente una imagen en espejo de su mitad derecha).



a) Sesgada a la izquierda (sesgo negativo): la media y la mediana están a la izquierda de la moda.

b) Simétrica (sesgo cero): la media, la mediana y la moda son iguales.

c) Sesgada la derecha (sesgo positivo): la media y la mediana están a la derecha de la moda.

Los datos **sesgados a la izquierda** (lo que también se conoce como *sesgo negativo*) poseen una cola izquierda más larga, y la media y la mediana se encuentran a la izquierda de la moda. [Aunque no siempre es posible predecirlo, los datos sesgados a la izquierda suelen tener una media menor a la mediana]. Los datos **sesgados a la derecha** (lo que también se denomina *sesgo positivo*) poseen una cola derecha más larga, y la media y la mediana se encuentran a la derecha de la moda. [Una vez más, aunque no siempre es posible predecirlo, en los datos sesgados a la derecha la media suele estar a la derecha de la mediana].

En la práctica, muchas distribuciones de datos son simétricas y carecen de sesgo. Las distribuciones sesgadas hacia la derecha son más comunes que las sesgadas hacia la izquierda, ya que con frecuencia es más fácil obtener valores excepcionalmente grandes que valores excepcionalmente pequeños. Por ejemplo, en el caso de los ingresos anuales, es imposible obtener valores por debajo del límite inferior de cero, pero hay algunas personas que ganan millones de dólares en un año. Por lo tanto, el ingreso anual tiende a mostrar un sesgo hacia la derecha.

MEDIDAS DE LOCALIZACIÓN: CUARTILES, DECILES Y PERCENTILES.

Las medidas de localización dividen la distribución en partes iguales, sirven para clasificar a un individuo o elemento dentro de una determinada población o muestra. Para calcular los cuartiles y percentiles se ordena los datos de menor a mayor:

CUARTILES

Medida de localización que divide la población o muestra en cuatro partes iguales.

- Q_1= Valor de la variable que deja a la izquierda el 25% de la distribución.
- Q_2= Valor de la variable que deja a la izquierda el 50% de la distribución = mediana.
- Q_3= Valor de la variable que deja a la izquierda el 75% de la distribución.

Al igual que ocurre con el cálculo de la mediana, el cálculo de estos estadísticos, depende del tipo de variable.

$$Q_k = Li + \frac{\frac{K * N}{4} - Fa_{i-1}}{fi} * A$$

DECILES

Medida de localización que divide la población o muestra en 10 partes iguales
$d_{k\,=}$ Decil k-simo es aquel valor de la variable que deja a su izquierda el k·10 % de la distribución.

$$D_k = Li + \frac{\frac{K * N}{10} - Fa_{i-1}}{fi} * A$$

PERCENTILES:

Medida de localización que divide la población o muestra en 100 partes iguales
pk = Percentil k-simo es aquel valor de la variable que deja a su izquierda el k % de la distribución.

$$P_k = Li + \frac{\frac{K * N}{100} - Fa_{i-1}}{fi} * A$$

EJEMPLO:

Como se puede observar la forma de calcular estas medidas es muy similar a la del cálculo de la mediana.

Veamos el cálculo de algunas de estas medidas en el ejemplo que estamos estudiando.

Vamos a calcular Q1,Q3, d3, y p45

Li	Ls	fi	Fa
45	55	6	6
55	65	10	16
65	75	19	35
75	85	11	46
85	95	4	50

Cálculo de Q_1: Buscamos en la columna de las frecuencias Acumuladas (Fa) el valor que supere al 25% de N=50, usando (K*N)/4, donde K = 1 porque estamos buscando Q_1

. (1*50)/4=12,5 por lo tanto corresponde al 2º intervalo, ya que si no encontramos el valor calculado (12,5) usamos el valor inmediatamente superior, el cual es 16 y así sabemos cuál es el intervalo que se va a usar.

Li: es el límite inferior del intervalo

fi: es la frecuencia absoluta

Fa_{i-1}: es la frecuencia acumulada anterior

A: es la amplitud del intervalo

$$Q_1 = 55 + \frac{\frac{1*50}{4} - 6}{10} * 10 = 55 + \frac{12{,}5 - 6}{10} * 10 = 55 + \frac{6{,}5}{10} * 10$$

$$= 55 + 0{,}65 * 10 = 55 + 6{,}5 = 61{,}5$$

Análogamente calculemos Q_3, Buscamos ahora en la misma columna el correspondiente al 75 %de N que en este caso es el 4º intervalo (3*50)/4=37.5

$$Q_3 = 75 + \frac{\frac{3*50}{4} - 35}{11} * 10 \qquad = 75 + \frac{37{,}5 - 35}{11} * 10 = 75 + \frac{2{,}5}{11} * 10$$

$$= 75 + 0{,}23 * 10 = 75 + 2{,}3 = 77{,}3$$

Veamos ahora el decil 3º. corresponde al 30 %, (3 * 50) / 10 = 15 sería el 2º intervalo.

$$D_3 = 55 + \frac{\frac{3*50}{10} - 6}{10} * 10 = 55 + \frac{15 - 6}{10} * 10 = 55 + \frac{9}{10} * 10$$

$$= 55 + 0{,}9 * 10 = 55 + 9 = 64$$

Por último veamos el percentil 45 (45*50)/100 = 22.5 Corresponde al intervalo 3º.

$$P_{45} = 65 + \frac{\frac{45*50}{100} - 6}{10} * 10 \qquad = 65 + \frac{22{,}5 - 16}{19} * 10 = 65 + \frac{6{,}5}{19} * 10$$

$$= 65 + 0{,}34 * 10 = 65 + 3{,}4 = 68{,}4$$

DESVIACIÓN MEDIA

Media aritmética de los valores absolutos de las desviaciones con respecto a la media aritmética. Mide la cantidad media respecto de la cual los valores de una población o muestra varían.

$DM = \sum|X_i - \overline{X}| / n$

X_i: es el valor de cada observación
$\overline{X}$: es la media aritmética de los valores
| |: valor absoluto
n: es el número de datos de la muestra

Para datos agrupados hay que agregarle la frecuencia absoluta (f):

$DM = (\sum|X_i - \overline{X}| * f_i) / n$
X_i: son los puntos medios de cada intervalo
X: es la media aritmética
f_i: es la frecuencia absoluta de cada intervalo
n: es la cantidad de datos
$\sum(X_i - \overline{X})$: es la sumatoria de las diferencias entre los puntos medios y la media aritmética

EJEMPLO

Calcule la desviación media de la siguiente tabla de datos.

Intervalos		f
45	55	6
55	65	10
65	75	19
75	85	11
85	95	4

Se calcula la media aritmética:

$\overline{X} = \frac{\sum Xi * f}{N}$ $\overline{X}$ = 3470 / 50 = 69,4

Intervalos		f	Xi	Xi * f	Xi - $\overline{X}$	|Xi - $\overline{X}$|	|Xi - $\overline{X}$| * f
45	55	6	50	300	-19,4	19,4	116,4
55	65	10	60	600	-9,4	9,4	94
65	75	19	70	1330	0,6	0,6	11,4
75	85	11	80	880	10,6	10,6	116,6
85	95	4	90	360	20,6	20,6	82,4
		50		**3470**			420,8

$DM = (\sum |X_i - \overline{X}| * f_i) / n$

DM = 420,8 / 50 = **8,42**

VARIANZA

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.

$$\sigma^2 = S^2 = \frac{\sum (X_i - \overline{X})^2 * f}{N} = \frac{\sum X_i^2}{N} - \overline{X}^2$$

EJEMPLO

Calcule la varianza de la siguiente tabla de datos.

Intervalos		f
45	55	6
55	65	10
65	75	19
75	85	11
85	95	4

$$\sigma^2 = \frac{\sum (X_i - \overline{X})^2 * f}{N}$$

Calculo de la media aritmética.

$\overline{X} = \frac{\sum Xi * f}{N}$ $\overline{X}$ = 3470 / 50 = **69,4**

Intervalos		f	Xi	Xi * f	Xi - $\overline{X}$	(Xi - $\overline{X}$)^2	(Xi - $\overline{X}$)^2 * f
45	55	6	50	300	-19,4	376,36	2258,16
55	65	10	60	600	-9,4	88,36	883,6
65	75	19	70	1330	0,6	0,36	6,84
75	85	11	80	880	10,6	112,36	1235,96
85	95	4	90	360	20,6	424,36	1697,44
		50		**3470**			**6082**

σ^2 = 6082 / 50 = **121,64**

DESVIACIÓN ESTÁNDAR O TÍPICA

La desviación estándar, también llamada desviación típica, es una medida de dispersión usada en estadística que nos dice cuánto tienden a alejarse los valores concretos del promedio en una distribución. De hecho, específicamente, el cuadrado de la desviación estándar es "el promedio del cuadrado de la distancia de cada punto respecto del promedio". Se suele representar por una S o con la letra sigma σ.
La desviación estándar de un conjunto de datos es una medida de cuánto se desvían los datos de su media. Esta medida es más estable que el recorrido y toma en consideración el valor de cada dato. Es la raíz cuadrada de la varianza:

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

EJEMPLO

Calcule la desviación estandar de la siguiente tabla de datos.

Intervalos		f
45	55	6
55	65	10
65	75	19
75	85	11
85	95	4

Una vez calculada la varianza:

σ^2 = 6082 / 50 = 121,64

Se calcula la raíz cuadrada:

$$\boldsymbol{\sigma} = \sqrt{\boldsymbol{\sigma}^2} \qquad \boldsymbol{\sigma} = \sqrt{121,64} \quad \mathbf{= 11,03}$$

COEFICIENTE DE VARIACIÓN:

Es un estadístico de dispersión que tiene la ventaja de que no lleva asociada ninguna unidad, por lo que nos permitirá decir entre dos muestras, cual es la que presenta mayor dispersión. La denotaremos por **C.V.**

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 unidas que los datos de la serie. Se interpreta como el número de veces que la media está contenida en la desviación típica. Suele darse su valor en tanto por ciento, multiplicando el resultado anterior por 100. De este modo se obtiene un porcentaje de la variabilidad.
Es importante que todos los valores sean positivos y su media dé, por tanto, un valor positivo. A mayor valor del coeficiente de variación mayor heterogeneidad de los valores de la variable; y a menor C.V., mayor homogeneidad en los valores de la variable. Suele representarse por medio de las siglas C.V.

$$C.V. = \frac{\sigma_x}{\bar{x}} \bullet 100\,\%$$

Propiedades y aplicaciones:

- El coeficiente de variación no posee unidades.
- El coeficiente de variación es típicamente menor que uno. Sin embargo, en ciertas distribuciones de probabilidad puede ser 1 o mayor que 1.
- Para su mejor interpretación se expresa como porcentaje.
- Depende de la desviación típica, también llamada "desviación estándar", y en mayor medida de la media aritmética, dado que cuando ésta es 0 o muy próxima a este valor el C.V. pierde

significado, ya que puede dar valores muy grandes, que no necesariamente implican dispersión de datos.

- El coeficiente de variación es común en varios campos de la probabilidad aplicada, como teoría de renovación y teoría de colas. En estos campos la distribución exponencial es a menudo más importante que la distribución normal. La desviación típica de una distribución exponencial es igual a su media, por lo que su coeficiente de variación es 1. La distribuciones con un C.V. menor que uno, como la distribución de Erlang se consideran de "baja varianza", mientras que aquellas con un C.V. mayor que uno, como la distribución hiperexponencial se consideran de "alta varianza". Algunas fórmulas en estos campos se expresan usando el cuadrado del coeficiente de variación, abreviado como S.C.V. (por sus siglas en inglés).

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 vienes 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 va a comparar la dispersión en los precios anuales de las acciones que se venden a menos de $10 (dólares) y la dispersión en los precios de aquellas que se venden por arriba de $60. El precio medio de las acciones que se venden a menos de $10 es 5,25 y la desviación estándar es $1,52. El precio medio de las acciones que se negocian a más de $60 es $92,50 y su desviación estándar es $5,28.

Calcule los coeficientes de variación. ¿Cuál es su conclusión?

Acciones menores a $10

CV = (1,52 / 5,25) * 100 = 28,95%

Arriba de $60

CV = (5,28 / 92,5) * 100 = 5,70%

Se observa que las acciones a menos de $10 tienen una dispersión mayor relativa, en comparación con las que se venden por arriba de los $60.

ASIMETRÍA

Es una medida de forma de una distribución que permite identificar y describir la manera como los datos tiende a reunirse de acuerdo con la frecuencia con que se hallen dentro de la distribución. Permite identificar las características de la distribución de datos sin necesidad de generar el gráfico.

TIPOS DE ASIMETRÍA

La asimetría presenta las siguientes formas:

Asimetría Negativa o a la Izquierda.- Se da cuando en una distribución la minoría de los datos está en la parte izquierda de la media. Este tipo de distribución presenta un alargamiento o sesgo hacia la izquierda, es decir, la distribución de los datos tiene a la izquierda una cola más larga que a la derecha. As < 0.

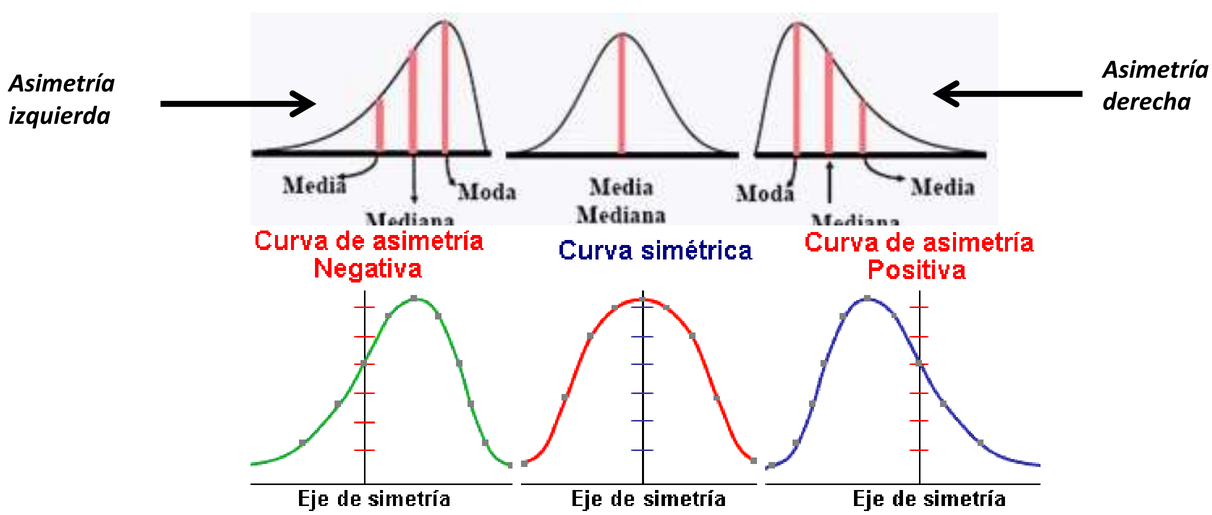
También se dice que una distribución es simétrica a la izquierda o tiene sesgo negativo cuando el valor de la media aritmética es menor que la mediana y éste valor de la mediana a su vez es menor que la moda, en símbolos $\bar{x} < Md < Mo$

Nota: Sesgo es el grado de asimetría de una distribución, es decir, cuánto se aparta de la simetría.

Simétrica.- Se da cuando en una distribución se distribuyen aproximadamente la misma cantidad de los datos a ambos lados de la media aritmética. No tiene alargamiento o sesgo. Se representa por una curva normal en forma de campana llamada campana de Gauss (matemático Alemán 1777-1855) o también conocida como de Laplace (1749-1827).También se dice que una distribución es simétrica cuando su media aritmética, su mediana y su moda son iguales, en símbolos $\bar{x}$Md=Mo

Asimetría Positiva o a la Derecha.- Se da cuando en una distribución la minoría de los datos está en la parte derecha de la media aritmética. Este tipo de distribución presenta un alargamiento o sesgo hacia la derecha, es decir, la distribución de los datos tiene a la derecha una cola más larga que a la izquierda. As > 0.

También se dice que una distribución es simétrica a la derecha o tiene sesgo positivo cuando el valor de la media aritmética es mayor que la mediana y éste a valor de la mediana a su vez es mayor que la moda, en símbolos $\bar{x} > Md > Mo$



$$As = \frac{\frac{\sum (X_i - \overline{X})^3 * f_i}{N}}{\sigma^3}$$

X$_i$: son los puntos medios de cada intervalo
X: es la media aritmética
f$_i$: es la frecuencia absoluta de cada intervalo
N: es la cantidad de datos
σ^3: es la desviación estándar elevada a la 3

$\sum(X_i - \overline{X})^3$: es la sumatoria de las diferencias elevada al cubo entre los puntos medios y la media aritmética

EJEMPLO

Calcule el coeficiente de asimetría de la siguiente tabla de datos.

Intervalos		f
45	55	6
55	65	10
65	75	19
75	85	11
85	95	4

Se calcula la media aritmética:

$\overline{X} = \frac{\sum Xi * f}{N}$ $\overline{X}$ = 3470 / 50 = **69,4**

Intervalos		f	Xi	Xi * f	Xi - $\overline{X}$	(Xi - $\overline{X}$)^3	(Xi - $\overline{X}$)^3 *f
45	55	6	50	300	-19,4	-7301,384	-43808,304
55	65	10	60	600	-9,4	-830,584	-8305,84
65	75	19	70	1330	0,6	0,216	4,104
75	85	11	80	880	10,6	1191,016	13101,176
85	95	4	90	360	20,6	8741,816	34967,264
		50		**3470**			-4041,6

$$As = \frac{\frac{\sum(X_i - \overline{X})^3 * f_i}{N}}{\sigma^3}$$

$\boldsymbol{\sigma} = \sqrt{121,64}$ **= 11,03** $\sigma^3 = 11,03^3$ = 1341,92

As = (-4041,6 / 50) / 1341,92 = **-0,06**

CURTOSIS O APUNTAMIENTO

La curtosis mide el grado de agudeza o achatamiento de una distribución con relación a la distribución normal, es decir, mide cuán puntiaguda es una distribución.

Tratan de estudiar la proporción de la varianza que se explica por la combinación de datos extremos respecto a la media en contraposición con datos poco alejados de la misma. Una mayor curtosis implica una mayor concentración de datos muy cerca de la media de la distribución coexistiendo al mismo tiempo con una relativamente elevada frecuencia de datos muy alejados de la misma. Esto explica una forma de la distribución de frecuencias con colas muy elevadas y un con un centro muy apuntado.

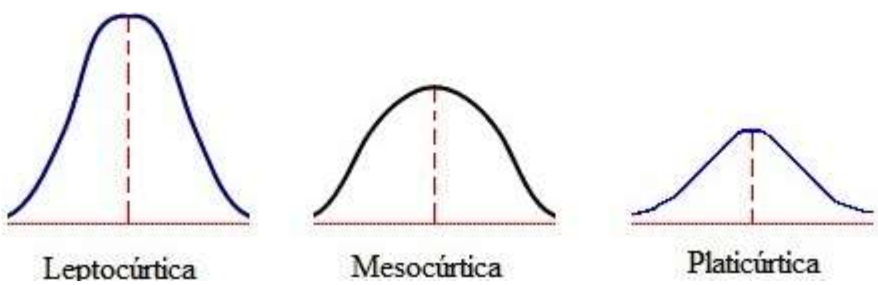
TIPOS DE CURTOSIS

La curtosis determina el grado de concentración que presentan los valores en la región central de la distribución. Así puede ser:

Leptocúrtica.- Existe una gran concentración, K > 0.

Mesocúrtica.- Existe una concentración normal, K = 0.

Platicúrtica.- Existe una baja concentración, K < 0.



$$K = \frac{\frac{\sum(X_i-\overline{X})^4 * f_i}{N}}{\sigma^4} - 3$$

X_i: son los puntos medios de cada intervalo
X: es la media aritmética
f_i: es la frecuencia absoluta de cada intervalo
N: es la cantidad de datos
σ^4: es la desviación estándar elevada a la 4
$\sum(X_i - X)^4$: es la sumatoria de las diferencias elevada a la cuatro entre los puntos medios y la media aritmética

EJEMPLO

Calcule el coeficiente de asimetría de la siguiente tabla de datos.

Intervalos		f
45	55	6
55	65	10
65	75	19

75	85	11
85	95	4

Se calcula la media aritmética:

$\overline{X} = \frac{\sum Xi * f}{N}$ $\overline{X}$ = 3470 / 50 = **69,4**

Intervalos		f	Xi	Xi * f	Xi - X	(Xi - X)^4	(Xi - X)^4 * f
45	55	6	50	300	-19,4	141646,8	849881,098
55	65	10	60	600	-9,4	7807,49	78074,896
65	75	19	70	1330	0,6	0,1296	2,4624
75	85	11	80	880	10,6	12624,77	138872,466
85	95	4	90	360	20,6	180081,4	720325,638
		50		**3470**			1787156,56

$$K = \frac{\frac{\sum (X_i - \overline{X})^4 * f_i}{N}}{\sigma^4} - 3$$

$\boldsymbol{\sigma} = \sqrt{121{,}64}$ **= 11,03** $\sigma^4 = 11{,}03^4$ = 14801,37

K = [(1787156,56 / 50) / 14801,37] – 3 = 2,42 – 3 = **-0,58**