2. **MEDIDAS DE DISPERSIÓN.** La medida de dispersión o de variabilidad mide qué tan diferentes o distantes son las observaciones de una medida de tendencia central (generalmente la media aritmética)

   a. Rango o Amplitud. Se expresa como la diferencia entre el mayor y el menor valor observado en un conjunto de datos, obtiene así:

   Rango = $X_{max} – X_{min}$

   Se tiene los datos: 490, 540, 560, 590, 600, 610, 620, 630, 660, 670, 700

   Rango = $X_{max} – X_{min}$

   Rango = 700 – 490 Rango = 210

   b. Desviación Media. Conocida como desviación absoluta promedio, se define como la media aritmética de los valores absolutos de las desviaciones con respecto a la media aritmética.

Población	Muestra
$DM = \frac{\sum \lvert X-\mu \rvert}{N}$	$DM = \frac{\sum \lvert X-\bar{X} \rvert}{n}$

Se tiene los siguientes datos:

	Observación X	Media	Desviación	Desviación Media
1	3.600	5.820	-2.220	2.220
2	4.200	5.820	-1.620	1.620
3	4.700	5.820	-1.120	1.120
4	4.900	5.820	-920	920
5	5.300	5.820	-520	520
6	5.700	5.820	-120	120
7	6.700	5.820	880	880
8	7.300	5.820	1.480	1.480
9	7.700	5.820	1.880	1.880
10	8.100	5.820	2.280	2.280
Σ	58.200		0	13.040

DM = $\frac{\sum |X-\mu|}{N}$ DM = $\frac{13.040}{10}$ DM = 1304

La desviación media nos indica la distancia a la que se encuentra las observaciones de la media del conjunto de datos. No es muy usada.

c. Varianza. Cada población tiene una varianza, que se simboliza con σ^2 (sigma cuadrada). Para calcular la varianza de una población se divide la suma de las distancias al cuadrado entre cada elemento de la población y la media, entre el número total de observaciones de dicha población. La varianza de la población se define por:

Población	Muestra
$\sigma^2 = \frac{\sum (X-\mu)^2}{N}$ O $\sigma^2 = \frac{\sum X^2}{N} - \mu^2$	$S^2 = \frac{\sum (X-\bar{X})^2}{n-1}$

d. Desviación Estándar. La desviación estándar de un conjunto de datos se define como la raíz cuadrada del valor de la varianza.

La desviación estándar de la población se define por:

Población	Muestra
$\sigma = \sqrt{\frac{\sum (X-\mu)^2}{N}}$	$S = \sqrt{\frac{\sum (X-\bar{X})^2}{n-1}}$

La tabla siguiente presenta las observaciones de un determinado evento. Halle su varianza y la desviación estándar

	Observación X	Media	Desviación	Desviación al Cuadrado	Observación al Cuadrado
1	3.600	5.820	-2.220	4928400	12960000
2	4.200	5.820	-1.620	2624400	17640000
3	4.700	5.820	-1.120	1254400	22090000
4	4.900	5.820	-920	846400	24010000
5	5.300	5.820	-520	270400	28090000
6	5.700	5.820	-120	14400	32490000
7	6.700	5.820	880	774400	44890000
8	7.300	5.820	1.480	2190400	53290000
9	7.700	5.820	1.880	3534400	59290000
10	8.100	5.820	2.280	5198400	65610000
Σ	58.200		0	21636000	360360000

La varianza es:

$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$; $\sigma^2 = \frac{21\,636\,000}{10}$; $\sigma^2 = 2\,163\,600$, se puede utilizar

$\sigma^2 = \frac{\sum X^2}{N} - \mu^2$; $\sigma^2 = \frac{360\,360\,000}{10} - (5.820)^2$; $\sigma^2 = 2\,163\,600$

La desviación estándar es:

$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}}$; $\sigma = \sqrt{\frac{21\,636\,000}{10}}$; $\sigma =$ 1.470, 91808

e. Coeficiente de Variación El coeficiente de variación mide la variabilidad porcentual de datos respecto a su media, se define:

$$C.V = \frac{Desviación\ Estándar}{Media}$$

Población	Muestra
$C.V(\%) = \frac{\sigma}{\mu}.100$	$C.V(\%) = \frac{S}{\overline{X}}.100$

El coeficiente de variación sirve para comparar la variabilidad de diferentes variables (no tiene unidades de medida), y se utiliza cuando:

a) Los datos están en unidades diferentes. Se desea comparar la dispersión de las distribuciones de los pesos, en kg, y las alturas, en cm, de cierta población.

b) Los datos están en las mismas unidades, pero las medidas son muy diferentes.

La población de estudiantes se presenta en la siguiente tabla, con la media y la varianza de los pesos y alturas. ¿Cuál de las dos variables tiene mayor dispersión relativa?

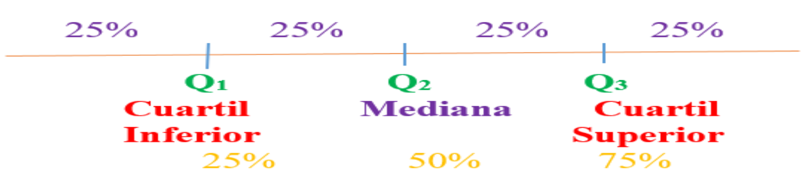
ATRIBUTO	MEDIA	VARIANZA	Desviación Estándar
Altura	167 cm	324 cm^2	18 cm
Peso	60 Kg.	121 Kg^2	11 Kg.

Coeficiente de variación: Altura	Coeficiente de variación: Peso
C.V = $\frac{\sigma}{\mu}$. C.V = $\frac{18\ cm}{167\ cm}$. C.V = 10,778%	C.V = $\frac{\sigma}{\mu}$. C.V = $\frac{11\ \text{Kg}}{60\ \text{Kg}}$. C.V = 18,333%

La variable con mayor dispersión relativa es el Peso con 18,333%

3. MEDIDAS DE LOCALIZACIÓN O POSICIÓN RELATIVA La desviación estándar es la medida de dispersión que más se utiliza. Los cuantiles son los valores de la variable que dividen a la distribución en un cierto número de partes con igual número de elementos, los cuantiles más utilizados son: los cuartiles, deciles y percentiles o centiles

a. Cuartiles**.** Son los más utilizados y divide la distribución en cuatro partes, existen tres cortes que se denota con: Q_1, Q_2 y Q_3



Q_1: Deja por debajo el 25% de los datos

Q_2: Deja por debajo el 50% de los datos es igual a la MEDIANA

Q_3: Deja por debajo el 75% de los datos

Cálculo de los cuartiles. Realice los siguientes pasos:

1). Ordene los datos

2). Encuentre la Posición (P_K) del cuartil que desea utilizando la siguiente fórmula

$P_k = \frac{K(n+1)}{4}$, donde K= 1, 2, 3   y n: es el total de datos

3). Interpole si es necesario

Rango Intercuartílico (r). Se define como: $r = Q_3 - Q_1$

La desviación cuarílica. Se define como: $Q = \frac{Q_3 - Q_1}{2}$

Considere los siguientes datos, hallar los tres cuartiles y la desviación cuartílica

Pi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Xi	15,	45,	47,	53,	58,	58,	60,	62,	67,	74,	75,	78,	80,	80,	81,	85,	85,	85,	90,	92

Cálculo del Q_1, es decir K=1

$P_1 = \frac{1(20+1)}{4}$ $P_1 = 5{,}25$

$Q_1 = 58+(58 - 58)(0{,}25)$ $Q_1 = 58$

Cálculo del Q_2, es decir K=2

$P_2 = \frac{2(20+1)}{4}$ $P_2 = 10{,}\ 5$

$Q_2 = 74+(75 - 74)(0{,}50)$ $Q_2 = 74{,}5$ Es igual a la Mediana

Cálculo del Q_3, es decir K=3

$P_3 = \frac{3(20+1)}{4}$ $P_3 = 15{,}75$

$Q_3 = 81+(85 - 81)(0{,}75)$ $Q_3 = 84$

Cálculo de la desviación cuartílica

$Q = \frac{Q_3 - Q_1}{2}$ $Q = \frac{84-58}{2}$ $Q = 13$ puntos

b. Deciles. Dividen un conjunto de observaciones en 10 partes iguales. El decil quinto es igual a la Mediana. Se calcula así:

1). Ordene los datos
2). Encuentre la Posición (P_K) del decil que desea utilizando la siguiente fórmula:
$P_k = \frac{K(n+1)}{10}$, donde K= 1, 2, 3,4, 5, 6, 7, 8, 9   y n: es el total de datos
3). Interpole si es necesario

Considere los siguientes datos, hallar el decil 1, decil 5 y decil 7

Pi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Xi	15,	45,	47,	53,	58,	58,	60,	62,	67,	74,	75,	78,	80,	80,	81,	85,	85,	85,	90,	92

Cálculo del D_1, es decir K=1

$P_1 = \frac{1(20+1)}{10}$ $P_1 = 2{,}1$

$D_1 = 45+(47 - 45)(0{,}10)$ $D_1 = 45{,}2$

Cálculo del D_5, es decir K=5

$P_5 = \frac{5(20+1)}{10}$ $P_5 = 10{,}\ 5$

$D_5 = 74+(75 - 74)(0{,}50)$ $D_5 = 74{,}5$ Es igual a la Mediana

Cálculo del D_7, es decir K=7

$P_7 = \frac{7(20+1)}{10}$ $P_7 = 14{,}7$

$D_7 = 80+(81 - 80)(0{,}70)$ $D_7 = 80{,}7$

c. Percentiles. Dividen un conjunto de observaciones en 100 partes iguales el Percentil 25 es igual a Q_1, el percentil cincuenta es igual a la medina (Q_2) y el percentil 75 es igual a Q_3. Se calcula así:

1). Ordene los datos
2). Encuentre la Posición (P_K) del decil que desea utilizando la siguiente fórmula:
$P_k = \frac{K(n+1)}{100}$, donde K= 1, 2, 3, …,95, 96, 97, 98, 99 y n: es el total de datos
3). Interpole si es necesario

Considere los siguientes datos, hallar el percentil 25, el percentil 50 y percentil 75

Pi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Xi	15,	45,	47,	53,	58,	58,	60,	62,	67,	74,	75,	78,	80,	80,	81,	85,	85,	85,	90,	92

Cálculo del P_{25}, es decir K=25

$P_{25} = \frac{25(20+1)}{100}$ $P_{25} = 5{,}25$

$P_{25} = 58+(58-58)(0{,}25)$ $P_{25} = 58$ Es igual a Q_1

Cálculo del P_{50}, es decir K=50

$P_{50} = \frac{50(20+1)}{100}$ $P_{50} = 10{,}\ 5$

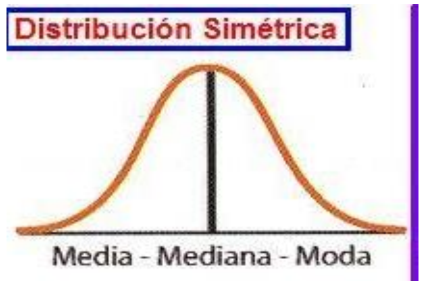
$P_{50} = 74+(75 - 74)(0{,}50)$ $P_{50} = 74{,}5$ Es igual a la Mediana y a Q_2

Cálculo del P_{75}, es decir K=75

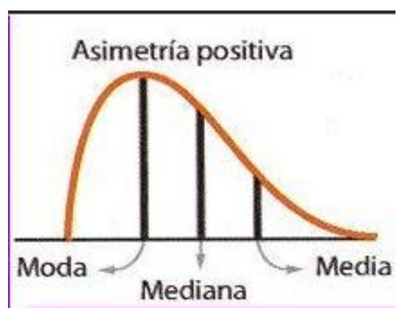
$P_{75} = \frac{75(20+1)}{100}$ $P_{75} = 15{,}75$

$P_{75} = 81+(85 - 81)(0{,}75)$ $P_{75} = 84$ Es igual a Q_3

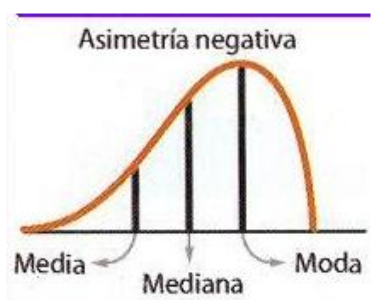
4. **MEDIDAS DE FORMA**. Las distribuciones se caracterizan porque el mayor número de observaciones se agrupan en valores de la variable más o menos centrales, siendo poco común los valores extremos. las distribuciones pueden tener forma: Simétrica y asimétrica (sesgo a la derecha o sesgo a la izquierda)
   a. Distribución Simétrica (sesgo Cero). En cualquier distribución simétrica, la moda, la mediana y la media siempre son iguales



   b. Distribución sesgada a la derecha (sesgo Positivo). La media aritmética es la mayor de las tres medidas. Por lo general, la mediana es la siguiente medida más grande en una distribución de frecuencias con sesgo positivo y la moda es la menor de las tres medidas.



   c. Distribución sesgada a la izquierda (sesgo Negativo). La media es la menor medida de las tres generalmente.



d. Medidas de Asimetría

Coeficiente de asimetría de Fisher	Coeficiente de asimetría de Pearson
$As = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n \cdot s^3}$	$As = \dfrac{\bar{X} - Me}{S}$

As = 0; la distribución es simétrica alrededor de la media.

As > 0; la distribución es asimétrica a la derecha

As < 0; la distribución es asimétrica a la izquierda

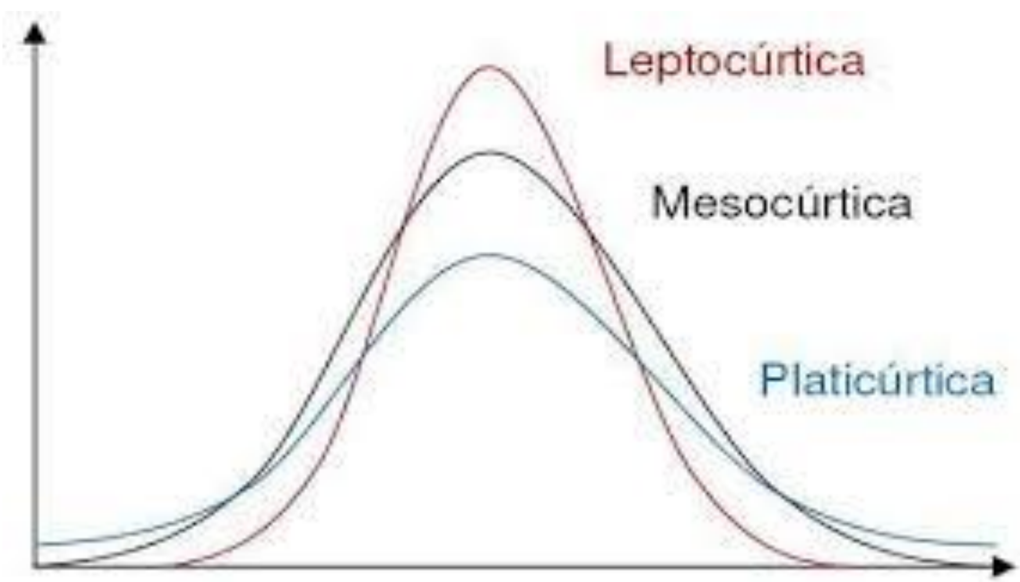
e. Medidas de Curtosis. La curtosis analiza la deformación, en sentido vertical, de una distribución con respecto a la Normal.

$$K = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n \cdot s^4}$$

K > 3; la distribución tiene mayor grado de apuntamiento que la distribución normal y se conoce como LEPTOCÚRTICA (mayor concentración de los datos en torno a la media).

K = 3; la distribución es normal se conoce como MESOCÚRTICA

K < 3; la distribución tiene menor grado de apuntamiento que la distribución normal y se conoce como PLATICÚRTICA.



5. REPRESENTACIÓN GRÁFICA PARA DATOS SIN AGRUPAR.

a. Diagrama de tallo y hojas. Es una técnica estadística para presentar un conjunto de datos ordenados de una variable cuantitativa. Cada valor numérico se divide en dos partes. El dígito principal se convierte en el tallo y los dígitos secundarios en las hojas. El tallo se localiza a lo largo del eje vertical y los valores de las hojas se apilan unos contra otros a lo largo del eje horizontal y estará dada por un solo dígito.

En la siguiente tabla, se presentan los pesos (Kg) de 27 alumnos de la carrera de Contabilidad y Auditoría. Elabore una de tallo y hojas

45	52	53	56	57	58	60	65	67
66	75	53	55	55	58	64	65	66
4 3	72	75	102	106	112	117	124	139

43	45	52	53	53	55	55	56	57
58	58	60	64	65	65	66	66	67
72	75	75	102	106	112	117	124	139

Tallo	Hoja
4	3 5
5	2 3 3 5 5 6 7 8 8
6	0 4 5 5 6 6 7
7	2 5 5
10	2 6
11	2 7
12	4
13	9

b. Diagramas de puntos. Son más útiles en el caso de conjuntos de datos pequeños, mientras que los histogramas lo son para conjuntos grandes de datos.

c. Diagrama de caja de dispersión. Es una representación gráfica, basada en cuartiles, que ayuda a presentar un conjunto de datos. Para construir un diagrama de caja, sólo necesita cinco estadísticos: el valor mínimo, Q_1 (primer cuartil), la mediana (Q_2), Q_3 (tercer cuartil) y el valor máximo.

d. Otros. Tenemos el diagrama de barras, diagrama de barras compuestas, El diagrama de pastel (sectores o pie), gráfico de barras y líneas, gráfico de líneas, gráfico de anillo o tipo dona entre otros.

DISTRIBUCIÓN DE FRECUENCIAS. Es la agrupación de datos cuantitativos en clases mutuamente excluyentes, que muestra el número de observaciones que hay en cada clase. Estos datos cuantitativos se representan en un histograma, polígono de frecuencias, diagrama de cajas, lineales, de dispersión entre otros.