La segunda manera de organizar los datos es mediante elaboración de intervalos que se conoce como datos agrupados

DATOS AGRUPADOS.

Se realiza el siguiente algoritmo:

1. Se debe calcular el Número de Clases (NC), para elaborar una tabla de distribución de frecuencias

   NC= $\sqrt{n}$, se utiliza cuando n $\leq$ 100

   NC= $2{,}5 * \sqrt[4]{n}$

   NC= 1 + 3,32*log(n), es la Ecuación de Sturges

   Mendenhall (1990), afirma que, es mejor utilizar de 5 a 20 clases.

   Otras ecuaciones para calcular el número de clases son:

   NC = 5* log(n) si n>100

   Scott (1979) basado en la normalidad de los datos. Establece que NC= $\frac{A * n^{1/3}}{3{,}49 * s}$;

   Donde A: amplitud, n: tamaño de la muestra y s: desviación estándar

2. Determine la Amplitud total o rango = $X_{max} - X_{min}$
3. Calcule la amplitud o intervalos de clase (IC)

   IC= $\frac{\text{Rango}}{\text{Nº de Clases}}$

   Punto Medio o conocido como marca de clase (Xi) es el punto medio del intervalo de clase y se obtiene

   Xi = $\frac{\text{Lim Sup} + \text{Lim Inf}}{2}$

1. MEDIDAS DE TENDENCIA CENTRAL

a. Media Aritmética.

$$\bar{X} = \frac{\sum_{i=1}^{n} x_i f_i}{N}$$

Xi: Punto medio o marca de clase

fi: Frecuencia de cada clase

N: El número total de frecuencias

b. La Mediana. La mediana para datos agrupados se obtiene a partir de las frecuencias absolutas acumuladas, mediante interpolación lineal (teorema de Thales),

$$Me = Li + \frac{\frac{N}{2} - Fi{-}1}{fi} * C$$

Identifique donde está la mitad de los datos con N/2

Li: Límite Inferior de la clase medianal

Fi-1: Frecuencia Acumulada antes de la frecuencia acumulada que contiene la mitad de los datos

fi: Frecuencia de la clase medianal

N: Total de frecuencias.

C: Amplitud del intervalo

c. Moda

$$Mo = Li + \frac{\Delta 1}{\Delta 1 + \Delta 2} * C$$

Li: Límite Inferior

Δ1: Diferencia entre la frecuencia de la clase modal y la clase inmediata anterior

Δ2: Diferencia entre la frecuencia de la clase modal y la clase inmediata posterior.

C: Intervalo de clase

Clase modal. Es la clase que presenta la mayor frecuencia

2. MEDIDAS DE DISPERSIÓN.

a. Rango (Recorrido) o Amplitud Total = $X_{max} - X_{min}$

b. Desviación Media.

$$D.M = \frac{\sum * |Xi - \vec{X}| * fi}{N}$$

c. Varianza.

$$\sigma^2 = \frac{\sum fi*(Xi-\mu)^2}{N} \qquad s^2 = \frac{\sum fi*(Xi-X)^2}{N-1}$$

d. Desviación Estándar.

$$\sigma = \sqrt{\frac{\sum fi*(X-\mu)^2}{N}}; \qquad s = \sqrt{\frac{\sum fi*(Xi-X)^2}{N-1}}$$

e. Coeficiente de Variación

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

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

3. MEDIDAS DE POSICIÓN RELATIVA.

Primero busque la posición y luego observe en la frecuencia acumulada

a. Cuartiles.

$$QK = Li + \frac{\frac{k.n}{4} - F_{i-1}}{fi} * C; \qquad k = 1, 2, 3$$

b. Deciles.

$$DK = Li + \frac{\frac{k.n}{10} - F_{i-1}}{fi} * C; \qquad k = 1, 2, 3, \ldots 8, 9$$

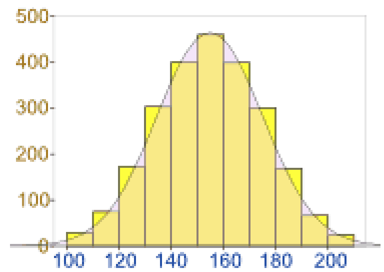
c. Percentiles.

$$PK = Li + \frac{\frac{k.n}{100} - F_{i-1}}{fi} * C; \qquad k = 1, 2, 3, \ldots 8, 9, 10, 11, \ldots 97, 98, 99$$

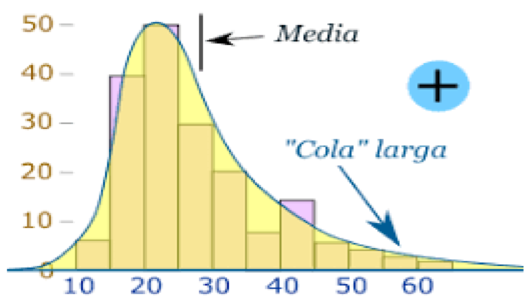
4. MEDIDAS DE FORMA O APUNTALAMIENTO

Coeficiente de asimetría de Fisher	**Coeficiente de asimetría de Pearson**
$As = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^3 \ast fi}{n\cdot s^3}$	$AS= \frac{\vec{X}-Mo}{s}$

Simétrica: $\vec{X}$ = Mo = Me

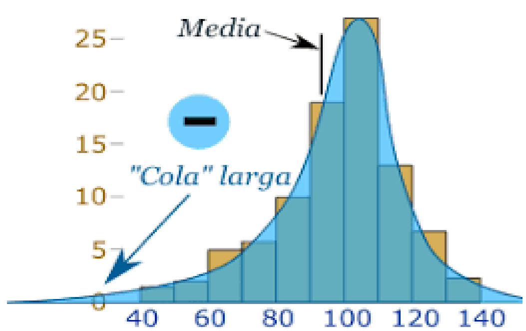


Asimétrica Positiva: $\vec{X}$ se desplaza a la derecha de la moda, el valor $\vec{X}$ - Mo > 0.



Mo < Me < $\vec{X}$

Asimétrica Negativa: $\vec{X}$ se ubica por debajo de la moda, el valor $\vec{X}$ - Mo < 0.



$\vec{X}$ < Me < Mo

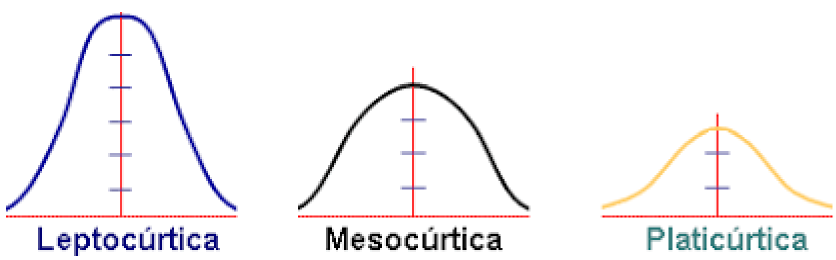
5. CURTOSIS Las medidas de curtosis estudian la distribución de frecuencias en la zona central de la misma, se las conoce como medidas de apuntamiento o concentración central. Para estudiar la curtosis de una distribución se toma como modelo la distribución Normal, que corresponde a fenómenos muy corrientes en la naturaleza, y cuya representación gráfica es una campana de Gauss. Dependiendo del valor del coeficiente de curtosis (K), una distribución es:

$$K = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4 * fi}{n \cdot s^4} - 3$$

a) Leptocúrtica (apuntamiento mayor que el de la normal) sí K > 0,

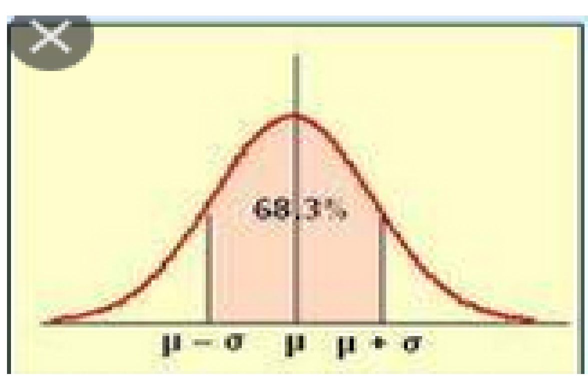
b) Mesocúrtica (apuntamiento igual al de la normal) cuando K = 0,

c) Platicúrtica (apuntamiento menor que el de la normal) sí K < 0.

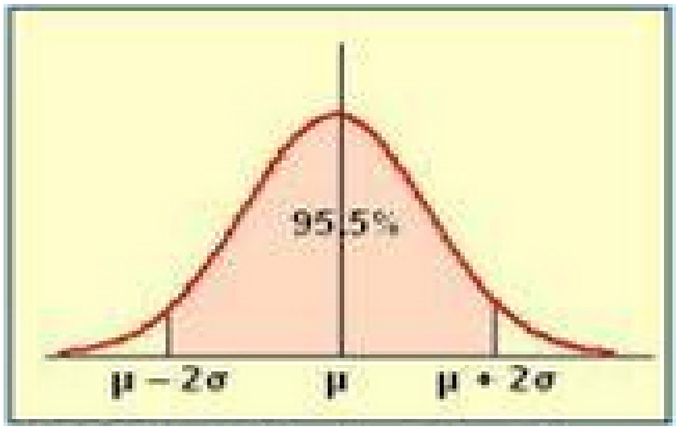


6. **TEOREMA DE TCHEBYSHEV**. El matemático ruso P.L. Tchebyshev (1821−1894), establece este teorema que se conoce como Regla Empírica, afirma que no importa la forma tenga la distribución de los datos se tiene:

1. Aproximadamente el 68% de las observaciones están comprendidas entre 1 desviación estándar de la media, esto es, entre $\mu - \sigma$ y $\mu + \sigma$.



2. Aproximadamente el 95% de las observaciones están comprendidas entre 2 desviación estándar de la media, esto es, entre $\mu - 2\sigma$ y $\mu + 2\sigma$.



3. Aproximadamente 99,7% de las observaciones están comprendidas entre 3 desviación estándar de la media, esto es, entre $\mu - 3\sigma$ y $\mu + 3\sigma$.