

# ESTADÍSTICA DESCRIPTIVA.

**1.- Distribución de frecuencias.** Agrupación de los datos en clases mutuamente excluyentes mostrando el número de observaciones en cada una.

**Intervalo total.**  $IT = Max - Min + UM$

Naturaleza del dato	UM	h
Entero	1	0.5
1 decimal	0.1	0.05
2 decimales	0.01	0.005

**Número de clases.**

$k = 1 + 3.322 \log n$  (Regla de Sturges)

$2^k < n \Rightarrow k > \frac{\log n}{\log 2}$  (Regla de 2 a la k)

$k = \sqrt{n}$  ( $n > 400$ )

Aproximar por exceso a un entero.

**Intervalo de clases.**  $IC = \frac{IT}{k}$

Aproximar por exceso al mismo número de decimales que los datos.

**Clases. Límites aparentes.**

Primera clase:  $L_{i,1} = Min$ ,  $L_{s,1} = Min + IC - UM$

Segunda clase:  $L_{i,2} = L_{i,1} + IC$ ,  $L_{s,2} = L_{s,1} + IC$

k-ésima clase:  $L_{i,k} = L_{i,1} + (k-1)IC$ ,  $L_{s,k} = L_{s,1} + (k-1)IC$

**Clases. Límites reales.**

Primera clase:  $L_{ri,1} = Min - h$ ,  $L_{rs,1} = Min + IC - UM + h$

Segunda clase:  $L_{ri,2} = L_{ri,1} + IC$ ,  $L_{rs,2} = L_{rs,1} + IC$

k-ésima clase:  $L_{ri,k} = L_{ri,1} + (k-1)IC$ ,  $L_{rs,k} = L_{rs,1} + (k-1)IC$

**Tabla de distribución de frecuencias.**

i	Clases		Frecuencias absolutas			Frecuencias acumuladas		
	$L_{ir} - L_{sr}$	$L_i - L_s$	$f_i$	$f_{ri}$	$\% f_{ri}$	$F_i$	$F_{ri}$	$\% F_{ri}$
1	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	-	-	-	-	-	n	1.00	100.00
			n	1.00	100.00			

**Notas:**

- Si existe la necesidad de anexas otra clase, se puede agregar, pero conservando la amplitud.

- Todas las clases deben ser del mismo ancho; es decir, la diferencia entre el límite superior e inferior debe ser igual para todas las clases.

- Las clases deben ser continuas a lo largo de la distribución, no se puede eliminar una determinada clase aunque la frecuencia en esa clase sea cero.

**Frecuencia relativa.**

$$f_{ri} = \frac{f_i}{n}$$

**Frecuencia relativa porcentual.**

$$\% f_{ri} = f_{ri} \times 100$$

**Frecuencia acumulada.**

Primera clase:  $F_i = f_i$

2<sup>da</sup> clase en adelante:  $F_i = F_{i-1} + f_i$

**Frecuencia acumulada relativa.**

$$F_{ri} = \frac{F_i}{n}$$

**Frecuencia acumulada relativa porcentual.**

$$\% F_{ri} = F_{ri} \times 100$$

**Marca de clase.**  $X_i = \frac{L_i + L_s}{2}$

**2.- Gráficos.**

**Diagrama de barras.** Es el procedimiento gráfico empleado para representar datos en escalas nominales y ordinales, y se obtiene dibujando en el eje de la abscisa (x) barras separadas por cada categoría, y con una altura asociada a la frecuencia absoluta ( $f_i$ ) por categoría, que se indica en el eje vertical (y).

**Gráficas circulares o de pastel.** Se emplean generalmente para representar distribuciones de razones. Su nombre se deriva de la semejanza de sus porciones a trozos de un pastel. El círculo representa la suma porcentual del conjunto de la distribución de razones (100%). Cada porción indica una razón en la serie. Los diagramas de barras y de pastel se usan para representar datos no continuos.

**Histograma.** Gráfica en la que las clases reales se marcan en el eje horizontal y las frecuencias de clases en el eje vertical. Las frecuencias de clases se representan mediante la altura de las barras y éstas últimas se dibujan una junto a otra.

**Polígono de frecuencia.** Consiste en segmentos de línea que conectan los puntos formados por las intersecciones de los puntos medios de clase y las frecuencias de clase. El punto medio de cada clase se representa en una escala en el eje x y las frecuencias de clase en el eje y. Para anclar el polígono es necesario incluir dos puntos adicionales, que se calculan al sustraer el intervalo de clase de la marca de clase más baja y sumando el intervalo de clase de la marca de clase más alta. En ambos casos, la ordenada del punto es cero.

**Polígono de frecuencias acumuladas.** Consiste en segmentos de línea que conectan los puntos formados por los límites reales superiores de cada clase y las frecuencias acumuladas de clase. El límite real superior se representa en una escala en el eje x y las frecuencias acumuladas de clase en el eje y. Para anclar el polígono es necesario incluir un punto adicional, que corresponde al límite real inferior de la primera clase con ordenada cero. La forma de esta curva es sigmoide (parecido a una S).

**3.- Medidas de tendencia central.**

**3.1.- Media aritmética.**

**Datos no agrupados.**

**Población**

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

**Muestra**

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

**Propiedades de la media aritmética.**

- La suma de las desviaciones de los valores con respecto a su media es igual a cero:  $\sum_{i=1}^n (X_i - \bar{X}) = 0$

- Si todos los valores de una serie de datos se aumentan o disminuyen en una cierta cantidad k, la media aritmética queda aumentada o

disminuida en esa cantidad:  $\bar{X}' = \frac{\sum_{i=1}^n (X_i + k)}{n} = \bar{X} + k$

- Si las observaciones son multiplicadas por una constante, la media aritmética queda multiplicada por dicha constante:

$$\bar{X}' = \frac{\sum_{i=1}^n (X_i \times k)}{n} = \bar{X} \times k$$

**Datos agrupados.**

**Población**

$$\mu = \frac{\sum_{i=1}^k f_i X_i}{n}$$

**Muestra**

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

Para determinar la media aritmética, completar la tabla de distribución de frecuencias con una columna que contenga el producto  $f_i X_i$ .

Sumar los elementos de dicha columna y dividir entre el número de datos.

**3.2.- Media geométrica.**

**Datos no agrupados.**

$$G = \sqrt[n]{\prod_{i=1}^n X_i}$$

$$\ln G = \frac{\sum_{i=1}^n \ln X_i}{n}$$

**Datos agrupados.**

$$G = \sqrt[n]{\prod_{i=1}^k X_i^{f_i}}$$

$$\ln G = \frac{\sum_{i=1}^k f_i \ln X_i}{n}$$

Para determinar la media geométrica, completar la tabla de distribución de frecuencias con una columna que contenga  $X_i^{f_i}$ ,

multiplicar los elementos de dicha columna y extraer la raíz n-ésima. Una forma alternativa para determinar la media geométrica es completar la tabla de distribución de frecuencias con una columna que contenga  $f_i \ln X_i$ , sumar los elementos de dicha columna, dividir entre el número de datos y determinar el exponencial.

**3.3.- Media armónica.**

**Datos no agrupados.**

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

**Datos agrupados.**

$$H = \frac{n}{\sum_{i=1}^k \frac{f_i}{X_i}}$$

Para determinar la media armónica, completar la tabla de distribución de frecuencias con una columna que contenga el cociente  $\frac{f_i}{X_i}$ . Sumar

los elementos de dicha columna y aplicar la ecuación para determinar H.

**Relaciones importantes.**  $\bar{X} > G > H$

$$- G = \sqrt[n]{\bar{X}H}$$

**3.4.- Mediana.**

El punto medio de los valores después de que se ordenan desde el más bajo hasta el más alto o desde el más alto hasta el más bajo. Es una medida descriptiva, indica un punto sobre o bajo el cual se encuentran el 50% de los datos, es decir, es un valor que divide a un conjunto de datos en dos partes iguales.

**Datos no agrupados.**

Es necesario ordenar los datos en forma creciente ó decreciente.

**Numero par de datos:** La mediana es el promedio de los dos datos centrales.  $X_{Med} = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$

**Número impar de datos:** La mediana es el dato central.  $X_{Med} = X_{\frac{n+1}{2}}$

**Datos agrupados.**

La clase donde se encuentre la frecuencia acumulada inmediatamente mayor a  $p = n/2$  es la clase mediana, y la que debe tomarse para la determinación de la mediana.

$$X_{Med} = L_{ir} + \frac{\frac{n}{2} - F_{i,ant}}{f_i} \times IC$$

### 3.5.- Moda.

**Datos no agrupados.** El valor de la observación que aparece con mayor frecuencia.

**Datos agrupados.** 
$$X_{Mod} = L_{ir} + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times IC$$

- La clase donde se encuentre la mayor frecuencia simple es la clase modal, y la que debe tomarse para la determinación de la moda.

- Si la moda se encuentra en la  $k$ -ésima clase, entonces:  $\Delta_1 = f_k - f_{k-1}$ ,

$$\Delta_2 = f_k - f_{k+1}$$

Es posible que una distribución de datos tenga más de una moda.

### 3.6.- Criterios de simetría.

Asimetría positiva (+):  $X_{Mod} < X_{Med} < \bar{X}$

Asimetría negativa (-):  $\bar{X} < X_{Med} < X_{Mod}$

Curva simétrica:  $\bar{X} = X_{Med} = X_{Mod}$

**3.7.- Cuartiles.** Son aquellos que dividen una serie de datos en cuatro porciones iguales en términos de la proporción de observaciones de cada una de ellas. Los puntos de división ( $Q_1, Q_2, Q_3$ ) se denominan cuartiles.

**Datos no agrupados.** Es necesario ordenar los datos en forma creciente.

**Número par de datos:** La ubicación del  $k$ -ésimo cuartil es  $\frac{kn}{4}$

**Numero impar de datos:** La ubicación del  $k$ -ésimo cuartil es  $\frac{k(n+1)}{4}$

**Datos agrupados.** La clase donde se encuentre la frecuencia acumulada inmediatamente mayor a  $p = \frac{kn}{4}$  es la clase que debe tomarse para la determinación del cuartil correspondiente.

$$Q_k = L_{ir} + \frac{\frac{kn}{4} - F_{i,ant}}{f_i} \times IC \quad (k = 1, 2, 3)$$

**3.8.- Deciles.** Son los valores que dividen al conjunto de datos en diez (10) partes iguales.

**Datos no agrupados.** Es necesario ordenar los datos en forma creciente.

**Número par de datos:** La ubicación del  $k$ -ésimo decil es  $\frac{kn}{10}$

**Numero impar de datos:** La ubicación del  $k$ -ésimo decil es  $\frac{k(n+1)}{10}$

**Datos agrupados.** La clase donde se encuentre la frecuencia acumulada inmediatamente mayor a  $p = \frac{kn}{10}$  es la clase que debe tomarse para la determinación del decil correspondiente.

$$D_k = L_{ir} + \frac{\frac{kn}{10} - F_{i,ant}}{f_i} \times IC \quad (k = 1, 2, 3, \dots, 10)$$

**3.9.- Percentiles.** Son los valores que dividen al conjunto de datos en cien (100) partes iguales.

**Datos no agrupados.** Es necesario ordenar los datos en forma creciente.

**Número par de datos:** La ubicación del  $k$ -ésimo percentil es  $\frac{kn}{100}$

**Numero impar de datos:** La ubicación del  $k$ -ésimo percentil es  $\frac{k(n+1)}{100}$

**Datos agrupados.** La clase donde se encuentre la frecuencia acumulada inmediatamente mayor a  $p = \frac{kn}{100}$  es la clase que debe tomarse para la determinación del percentil correspondiente.

$$P_k = L_{ir} + \frac{\frac{kn}{100} - F_{i,ant}}{f_i} \times IC \quad (k = 1, 2, 3, \dots, 100)$$

### 3.10.- Determinación de un percentil dado un $\bar{X}$ .

$$\% = \left[ \frac{f_i(P_k - L_{ir})}{IC} + F_{i,ant} \right] \frac{100}{n}$$

### Relaciones importantes.

$$X_{Med} = Q_2 = D_5 = P_{50}$$

$$Q_j = P_{25j}, \quad j = 1, 2, 3$$

$$D_j = P_{10j}, \quad j = 1, 2, 3, \dots, 8, 9, 10$$

### Rango.

$$R = L_{s,k} - L_{i,1}$$

### Rango intercuartílico.

$$RI = Q_3 - Q_1$$

### Rango semi-intercuartílico.

$$RSI = \frac{Q_3 - Q_1}{2}$$

### 4.- Medidas de dispersión.

#### 4.1.- Desviación media.

**Datos no agrupados.**

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

**Datos agrupados.**

$$DM = \frac{\sum_{i=1}^k f_i |X_i - \bar{X}|}{n}$$

#### 4.2.- Varianza.

La media aritmética de las desviaciones cuadradas de la media. La varianza implica homogeneidad o heterogeneidad de los datos en el grupo. A mayor heterogeneidad en el grupo, mayor desviación con respecto a la media.

**Datos no agrupados.**

**Población.**

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^n X_i^2}{n} - (\mu)^2$$

**Muestra.**

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$s^2 = \frac{\sum_{i=1}^n X_i^2 - n(\bar{X})^2}{n-1}$$

**Datos agrupados.**

**Población.**

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

$$\sigma^2 = \frac{\sum_{i=1}^k f_i X_i^2}{n} - \mu^2$$

**Muestra.**

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

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

**4.3.- Desviación estándar.** La raíz cuadrada de la varianza.

**Población:**  $\sigma = \sqrt{\sigma^2}$

**Muestra:**  $s = \sqrt{s^2}$

**4.4.- Teorema de Chebyshev.** Para cualquier grupo de observaciones (muestra o población), la proporción de los valores que se encuentra dentro de  $k$  desviaciones estándar de la media es por lo menos  $1 - 1/k^2$ , donde  $k$  es cualquier constante mayor que 1.

**4.5.- Regla empírica.** Para una distribución de frecuencia simétrica en forma de campana, aproximadamente 68% de las observaciones estarán entre más y menos una desviación estándar de la media; alrededor de 95% de las observaciones se encontrarán entre más y menos dos desviaciones estándar de la media, y prácticamente todas (99.7%) estarán entre más y menos de tres desviaciones estándar de la media.

### 5.- Medidas de dispersión relativa.

#### 5.1.- Coeficiente de variación.

La razón de la desviación estándar con la media aritmética, expresada como porcentaje:  $CV = \frac{s}{\bar{X}} \times 100$

#### 5.2.- Coeficiente de asimetría de Pearson.

$$A_p = \frac{\bar{X} - X_{Mod}}{s}$$

$$A_p = \frac{3(\bar{X} - X_{Med})}{s}$$

#### 5.3.- Coeficiente de asimetría de Fisher.

$$\gamma_1 = \frac{\mu_3}{\sigma^3}, \quad \mu_3 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n}, \quad \mu_3 = \frac{\sum_{i=1}^k f (X_i - \bar{X})^3}{n}$$

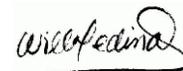
Coefficiente de sesgo positivo: Distribución sesgada a la derecha.  
 Coeficiente de sesgo negativo: Distribución sesgada a la izquierda.  
 Coeficiente de sesgo nulo: Distribución sin sesgo, o insesgada.

#### 5.4.- Kurtosis.

$$K = \frac{\mu_4}{\sigma^4}, \quad \mu_4 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n}, \quad \mu_4 = \frac{\sum_{i=1}^k f (X_i - \bar{X})^4}{n}$$

$K = 3$ : Distribución mesocúrtica (normal).  
 $K < 3$ : Distribución platicúrtica (curva achatada)  
 $K > 3$ : Distribución leptocúrtica (curva aguda)

Autor: **MSc. Ing. Willians Medina.**  
 Teléfono / WhatsApp: **+58-424-9744352**  
 e-mail: **medinawj@gmail.com**  
 Twitter: **@medinawj**



El presente formulario está disponible en formato digital en la siguiente dirección:  
<https://www.tutoruniversitario.com/>  
 Puerto La Cruz, abril de 2025.