

ESTADÍSTICA DESCRIPTIVA

A. MEDIDAS DE TENDENCIA CENTRAL

B. MEDIDAS DE VARIABILIDAD

C. MEDIDAS DE FORMA

RESUMEN:

A. MEDIDAS DE TENDENCIA CENTRAL

Son estadígrafos de posición que son interpretados como valores que permiten resumir a un conjunto de datos dispersos, podría asumirse que estas medidas equivalen a un centro de gravedad que adoptan un valor representativo para todo un conjunto de datos predeterminados. Estas medidas son:

1. Promedio Aritmético (Media o simplemente promedio)
2. Mediana
3. Moda
4. Promedio Geométrico
5. Promedio Ponderado
6. Promedio Total
7. Media Armónica

Otras medidas de posición son: Cuartiles, Deciles y Percentiles

B. MEDIDAS DE VARIABILIDAD

Son estadígrafos de dispersión que permiten evaluar el grado de homogeneidad, dispersión o variabilidad de un conjunto de datos. Estas medidas son:

1. Amplitud o Rango
2. Variancia
3. Desviación Estándar
4. Coeficiente de Variabilidad

C. MEDIDAS DE FORMA

Evalúa la forma que adopta la distribución de frecuencias respecto al grado de distorsión (inclinación) que registra respecto a valor promedio tomado como centro de gravedad, el grado de apuntamiento (elevamiento) de la distribución de frecuencias. A mayor elevamiento de la distribución de frecuencia significará mayor concentración de los datos en torno al promedio, por tanto, una menor dispersión de los datos. Estas medidas son:

1. Asimetría o Sesgo
2. Curtosis

Los Gráficos de Cajas como indicadores de forma

A. MEDIDAS DE TENDENCIA CENTRAL

1. LA MEDIA ARITMETICA

- **Para Datos No Agrupados.**

El promedio aritmético de un conjunto de valores (x_1 x_2 x_3 x_n) es:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n}$$

Ejemplo: Durante los últimos 32 días el valor de las compras en periódicos fue:

{ 5.2, 10.2, 7.0, 7.1, 10.2, 8.3, 9.4, 9.2, 6.5, 7.1, 6.6, 7.8, 6.8, 7.2, 8.4, 9.6, 8.5, 5.7, 6.4, 10.1, 8.2, 9.0, 7.8, 8.2, 5.3, 6.2, 9.1, 8.6, 7.0, 7.7, 8.3, 7.5 }

El promedio aritmético del valor de las compras de periódicos es:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{250.2}{32} = 7.82$$

- **Para Datos Agrupados.**

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

Donde: f_i = Frecuencia en la clase k-ésima

X_i = Marca de clase en la intervalo k-ésimo

Ejemplo: Para los gastos diarios en periódicos del hotel agrupados en una tabla de frecuencia:

Intervalo	X_i	f_i	h_i	F_i	H_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7.0	6.55	5	0.156	8	0.250
7.0 - 7.9	7.45	9	0.281	17	0.531
7.9 - 8.8	8.35	7	0.219	24	0.750
8.8 - 9.7	9.25	5	0.156	29	0.906
9.7 - 10.6	10.15	3	0.094	32	1.000
TOTAL		32	1.000		

7.87

El promedio aritmético es:

$$\bar{x} = \frac{\sum_{i=1}^k f_i X_i}{n} = \frac{3(5.65) + 5(6.55) + 9(7.45) + 7(8.35) + 5(9.25) + 3(10.15)}{32} = \frac{251.9}{32} = 7.87$$

Durante los 32 días el hotel tuvo un gasto promedio en periódicos de 7.87 soles

2. LA MEDIANA

Es el valor que ocupa la posición central de un conjunto de observaciones ordenadas. El 50% de las observaciones son mayores que este valor y el otro 50% son menores.

- **Para Datos No agrupados.**

La ubicación de la mediana de n datos ordenados se determina por : $\frac{(n+1)}{2}$. Ejemplos:

En los 7 datos ordenados: {4, 5, 5, 6, 7, 8, 9 }

La ubicación de la mediana es: $\frac{(7+1)}{2} = 4$ Luego el valor de la mediana es: $Me=6$

En los 8 datos ordenados: {3, 4, 5, 5, 6, 7, 8, 9}

La mediana se ubica en el lugar $\frac{(8+1)}{2} = 4.5$ Luego el valor de la mediana es $Me = \frac{5+6}{2} = 5.5$

- **Para Datos Agrupados.**

$$Me = L_i + \frac{c \left(\frac{n}{2} - F_{i-1} \right)}{f_i} = L_i + \frac{c(0.50 - H_{i-1})}{h_i}$$

Donde:

L_i = Límite Inferior del intervalo que contiene a la Mediana

F_{i-1} = Frecuencia Acumulada en la clase anterior i -ésima

f_i = Frecuencia en la clase que contiene a la mediana

H_{i-1} = Frecuencia Relativa Acumulada en la clase anterior i -ésima

h_i = Frecuencia Relativa en la clase que contiene a la mediana

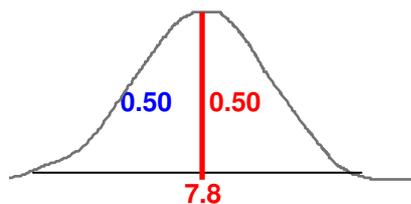
c = Tamaño del intervalo de clase.

Ejemplo: Para los gastos diarios en periódicos del hotel en una tabla de frecuencia:

Intervalo	X_i	f_i	h_i	F_i	H_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7.0	6.55	5	0.156	8	0.250
7.0 - 7.9	7.45	9	0.281	17	0.531
7.9 - 8.8	8.35	7	0.219	24	0.750
8.8 - 9.7	9.25	5	0.156	29	0.906
9.7 - 10.6	10.15	3	0.094	32	1.000
TOTAL		32	1.000		

La Mediana es: $Me = 7.0 + \frac{0.9 \left(\frac{32}{2} - 8 \right)}{9} = 7.0 + \frac{0.9(0.5 - 0.25)}{0.281} = 7.8$

El 50% de los días el hotel gastó menos de 7.8 soles en la compra de periódicos



3. LA MODA

Es el valor, clase o categoría que ocurre con mayor frecuencia y sus características son:

- Puede no existir o existir más de una moda
- Su valor no se ve afectado por los valores extremos en los datos
- Se utiliza para analizar tanto la información cualitativa como la cuantitativa
- Es una medida "inestable" cuando en número de datos es reducido.

- **Para Datos No Agrupados.**

Por ejemplo, durante los últimos 32 días el valor de las compras en periódicos fue:

{ 5.2, 10.2, 7.0, 7.1, 10.2, 8.3, 9.4, 9.2, 6.5, 7.1, 6.6, 7.8, 6.8, 7.1, 8.4, 9.6, 8.5, 5.7, 6.4, 10.1, 8.2, 9.0, 7.8, 8.2, 5.3, 6.2, 9.1, 8.6, 7.0, 7.7, 8.3, 7.5 }

Moda = $M_o = 7.1$; Es el valor más frecuente, ocurre 3 veces.

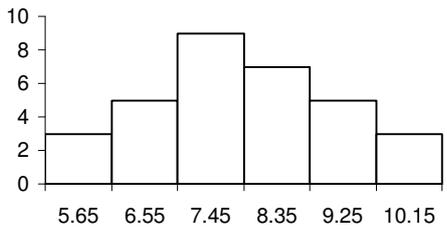
- **Para Datos Agrupados.**

$$M_o = L_i + c \left[\frac{d_1}{d_1 + d_2} \right]$$

Donde: $d_1 = (f_i - f_{i-1})$ y $d_2 = (f_i - f_{i+1})$ f_i = Valor de la mayor frecuencia

Ejemplo: El gasto diario en periódicos del hotel "AAA" agrupados en una tabla de frecuencia:

Intervalo	X_i	f_i	h_i	F_i	H_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7.0	6.55	5	0.156	8	0.250
7.0 - 7.9	7.45	9	0.281	17	0.531
7.9 - 8.8	8.35	7	0.219	24	0.750
8.8 - 9.7	9.25	5	0.156	29	0.906
9.7 - 10.6	10.15	3	0.094	32	1.000
TOTAL		32	1.000		



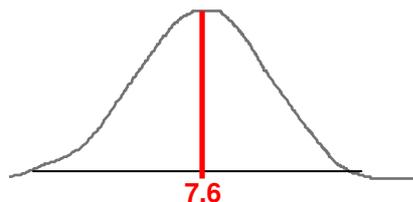
$d_1 = 9 - 5 = 4$ $d_2 = 9 - 7 = 2$ $c = 0.9 =$ Tamaño de Intervalo de Clase

La moda estimada utilizando estos datos agrupados es:

$$M_o = 7.0 + (0.9) \left[\frac{4}{4 + 2} \right] = 7.0 + 0.6 = 7.6$$

Utilizando las frecuencias relativas, la moda estimada es:

$$M_o = 7.0 + (0.9) \left[\frac{0.125}{0.125 + 0.062} \right] = 7.0 + 0.6 = 7.6$$



El gasto diario en periódicos más frecuente es 7.6 soles

4. MEDIA GEOMÉTRICA

Corresponde al valor representativo central de observaciones secuenciales y estrechamente relacionadas entre sí tales como tasas de: interés, inflación, devaluación, variación, crecimiento, disminución. El promedio geométrico de los valores: $(X_1 X_2 \dots X_t)$ es:

$$\bar{X}_G = \sqrt[t]{FC_1 FC_2 \dots FC_t} \quad \text{ó} \quad \bar{X}_G = \sqrt[t]{\frac{X_f}{X_i}} \quad \text{Donde } X_f = \text{Valor final y } X_i = \text{Valor inicial}$$

Ejemplo: La tasa de interés mensual que se pagó por un préstamo recibido por 3 meses fue cambiando mes a mes; en el primer mes se pagó un interés de 15%, en el segundo mes 10% y en el tercer mes 16%. La tasa de interés promedio mensual que se pagó es:

Mes	1	2	3
Tasa	0.15	0.10	0.16
Factor	1.15	1.10	1.16

$$\bar{X}_G = \sqrt[3]{(1.15)(1.10)(1.16)} = \sqrt[3]{1.4674} = 1.136 \quad (13.6\% \text{ mensual})$$

Ejemplo: El Producto Bruto Interno de un país durante los últimos cinco años tuvo la evolución siguiente: Año1: +5%. Año 2: 0% Año3: - 1% Año 4: +2% y Año5: + 4%. La tasa de crecimiento anual promedio del PBI sería:

$$\bar{X}_G = \sqrt[5]{(1.05)(1.00)(0.99)(1.02)(1.04)} = 1.0197 \quad (1.97\% \text{ anual})$$

Ejemplo: Se recibió un préstamo de 1000 soles por 3 meses y al final del período se pagó un total 1467.40 soles; ¿Cuál fue la tasa promedio de interés mensual que se pagó?

Mes	0	Mes 1	Mes 2	Mes 3
Saldo	1000			1467.40

$$\bar{X}_G = \sqrt[3]{\frac{1467.40}{1000}} = 1.136 \quad (13.6\% \text{ mensual})$$

5. PROMEDIO PONDERADO

Cuando se desea encontrar el promedio de valores $(X_1 X_2 \dots X_k)$ que ocurren con frecuencias $(f_1 f_2 \dots f_k)$ diferentes se deberán ponderar los valores observados con pesos diferentes:

$$\bar{x} = \sum_{i=1}^K W_i X_i$$

Donde los valores $W_i = f_i/n$ se denominan “ponderaciones o pesos”

Ejemplo: En una agencia de viajes se han vendido 200 pasajes a los precios siguientes:

Precio de Venta (soles) X_i	Número de pasajes f_i	Ponderación W_i
12	60	0.30
14	100	0.50
16	40	0.20
Total	200	1.00

El precio promedio de venta de los 200 pasajes: $\bar{x} = 0.30(12) + 0.50(14) + 0.20(16) = 13.8$

6. PROMEDIO TOTAL

Corresponde al valor promedio representativo de grupos de observaciones separadas o diferentes y que podrían estar consolidadas en tablas de frecuencia independientes, por tanto:

$$\bar{X}_T = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots + n_k \bar{X}_k}{n_1 + n_2 + \dots + n_k}$$

n_i : Número de observaciones en el grupo i -ésimo.

\bar{X}_i : Promedio correspondiente el grupo i -ésimo

Grupo A		
Nota	X_i	F_i
5-10	7.5	4
10-15	12.5	16
15-20	17.5	5
Total		25

Promedio del grupo A:

$$\bar{x}_A = \frac{4(7.5) + 16(12.5) + 5(17.5)}{25} = 12.7$$

Grupo B		
Nota	X_i	f_i
0-5	2.5	8
5-10	7.5	10
10-15	12.5	16
15-20	17.5	6
Total		40

Promedio del grupo B:

$$\bar{x}_B = \frac{8(2.5) + 10(7.5) + 16(12.5) + 6(17.5)}{40} = 10$$

Promedio Total	Grupo	\bar{X}_i	f_i
	A	12.7	25
	B	10.0	40
	Total		65

$$\bar{x}_T = \frac{25(12.7) + 40(10.0)}{65} = 11.04$$

7. MEDIA ARMÓNICA

El promedio armónico de los valores: (X_1, X_2, \dots, X_n) donde ninguno toma el valor "cero" es:

$$\bar{X}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}$$

Este promedio se utiliza para que los valores "extremos" no afecten al valor del promedio. Los valores extremos sí afectan cuando se usa el promedio aritmético o el promedio geométrico.

Ejemplo: Calcular el rendimiento promedio para el caso de tres automóviles que recorrieron 500 kilómetros y cada auto tuvo el rendimiento siguiente:

Auto	A	B	C
Rendimiento (Km/galón)	50	62.4	77.6

$$\bar{X}_H = \frac{3}{\frac{1}{50} + \frac{1}{62.4} + \frac{1}{77.6}} = \frac{3}{0.0489121} = 61.334 \frac{\text{Kilómetros(CONSTANTE)}}{\text{galón}}$$

Verificación:

Auto	Km	Rendimiento	Total galones
A	500	50	10
B	500	62.4	8.0128
C	500	77.6	6.4433
Total	1500		24.4561

$$\bar{X}_H = \frac{1500}{24.4561} = 61.334$$

PERCENTILES, CUARTILES Y DECILES

- **Para Datos Agrupados**

Percentiles: Son 99 valores que dividen a un conjunto de datos en 100 partes iguales

$$P_k = L_i + \frac{c \left(\frac{kn}{100} - F_{i-1} \right)}{f_i}$$

- L_i = Límite Inferior del intervalo que contiene al Percentil
- F_{i-1} = Frecuencia Acumulada en la clase anterior k-ésima
- f_i = Frecuencia en la clase que contiene al Percentil
- c = Tamaño del intervalo de clase.
- k = 1%, 2%, 3%, ... , 97%, 98%, 99% Percentiles

Intervalo De Clase	Marca de Clase X_i	Frecuencia Absoluta f_i	Frecuencia Relativa h_i	Frec.Acum. Absoluta F_i	Frec. Acum. Relativa H_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7.0	6.55	5	0.156	8	0.250
7.0 - 7.9	7.45	9	0.281	17	0.531
7.9 - 8.8	8.35	7	0.219	24	0.750
8.8 - 9.7	9.25	5	0.156	29	0.906
9.7 - 10.6	10.15	3	0.094	32	1.000
TOTAL		32	1.000		

Ejemplo: El Percentil 80% de los gastos diarios en periódicos estará en intervalo 5

$$P_{80\%} = L_i + \frac{c(80n/100 - F_{i-1})}{f_i} = 8.8 + \frac{0.9(25.6-24)}{5} = 9.088$$

El 80% de los datos analizados serán menores a 9.088 y el 20% restante serán superiores

Cuartiles: Son 3 valores Q_1 ; Q_2 y Q_3 que dividen a los datos en 4 partes iguales

El Cuartil 3 (Percentil 75%) se ubicará en el cuarto intervalo

$$P_{75\%} = L_i + \frac{c(75n/100 - F_{i-1})}{f_i} = 7.9 + \frac{0.9(24-17)}{7} = 8.8$$

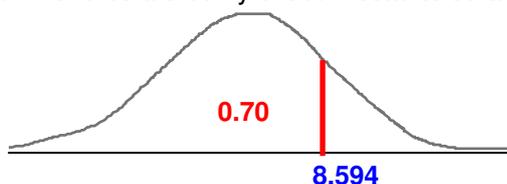
75% de los datos serán menores a 8.8 y el 25% de los datos restantes serán superiores

Deciles: Son 9 valores D_1 , D_2 ; D_3 ; D_4 ; D_5 ; D_6 ; D_7 ; D_8 y D_9 que dividen a un conjunto de datos en 10 partes iguales.

El Decil 7(Percentil 70%) se ubicará en el cuarto intervalo

$$P_{70\%} = L_i + \frac{c(70n/100 - F_{i-1})}{f_i} = 7.9 + \frac{0.9(22.4-17)}{7} = 8.594$$

70% de los datos serán menores a 8.594 y el 30% restante serán superiores a 8.594.



- **Para Datos No Agrupados**

El lugar o posición donde se encuentran los cuartiles para n datos ordenados es:

Cuartel	$Q_1 = P_{25\%}$	$Q_2 = P_{50\%}$	$Q_3 = P_{75\%}$
Posición	$\frac{25(n+1)}{100}$	$\frac{50(n+1)}{100}$	$\frac{75(n+1)}{100}$

Ejemplo: Determine los cuartiles y el decil 8 de los 13 datos ordenados siguientes:

10	11	11	12	12	13	13	13	14	15	17	18	20
----	----	----	----	----	----	----	----	----	----	----	----	----

Percentil	Posición	Valor del Cuartel
$Q_1 = P_{25}$	$0.25(13+1)=3.5$	$Q_1 = 11 + (12-11)0.5 = 11.5$
$Q_2 = P_{50}$	$0.50(13+1)=7$	$Q_2 = 13$
$Q_3 = P_{75}$	$0.75(13+1)=10.5$	$Q_3 = 15 + (17-15)0.5 = 16$
$D_8 = P_{80}$	$0.80(13+1)=11.2$	$P_{80} = 17 + (18-17)0.2 = 17.2$

Ejemplo: Para la representación tallo hoja de los gastos en periódicos del hotel:

	Tallo	Hojas
3	5	2 3 7
8	6	2 4 5 6 8
(9)	7	0 0 1 1 2 5 7 8 8
15	8	2 2 3 3 4 5 6
8	9	0 1 2 4 6
3	10	1 2 2

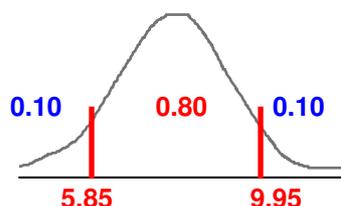
Determine los 3 cuartiles correspondientes a los 32 datos ordenados:

Cuartil	Posición	Valor
$Q_1 = P_{25\%}$	$\frac{25(32+1)}{100} = 8.25$	$Q_1 = 6.8 + (7.0-6.8)0.25 = 6.85$
$Q_2 = P_{50\%}$	$\frac{50(32+1)}{100} = 16.5$	$Q_2 = 7.8 + (7.8-7.8)0.50 = 7.80$
$Q_3 = P_{75\%}$	$\frac{75(32+1)}{100} = 24.75$	$Q_3 = 8.6 + (9.0-8.6)0.75 = 8.90$

¿Entre qué valores está el 80% central de los gastos diarios en periódicos?

Percentil	Posición	Valor
P_{10}	$\frac{10(32+1)}{100} = 3.3$	$P_{10\%} = 5.7 + (6.2-5.7)0.3 = 5.85$
P_{90}	$\frac{90(32+1)}{100} = 29.7$	$P_{90\%} = 9.6 + (10.1-9.6)0.7 = 9.95$

El 80% de los gastos diarios en periódicos está definido entre los 5.85 y 9.95 soles



B. MEDIDAS DE VARIABILIDAD

1. AMPLITUD O RANGO

Sean los valores: $(x_1 x_2 x_3 \dots x_n)$. La amplitud o rango de estos dato es $A=(X_{\max}-X_{\min})$

2. VARIANCIA

- **Para Datos No Agrupados**

La variancia de los datos de esta muestra $(x_1 x_2 x_3 \dots x_n)$:

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

Ejemplo: Calcular la variancia de los cuatro datos siguientes (X_i : 3, 4, 6 y 7)

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n} = \frac{3+4+6+7}{4} = \frac{20}{4} = 5$$

$$S^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n-1} = \frac{3^2 + 4^2 + 6^2 + 7^2 - 4(5)^2}{4-1} = \frac{10}{3} = 3.333$$

- **Para Datos Agrupados**

La variancia de los valores: $(x_1 x_2 \dots x_k)$ que ocurren con las frecuencias $(f_1 f_2 \dots f_k)$ es:

$$S^2 = \frac{\sum_{i=1}^n f_i X_i^2 - n\bar{X}^2}{n-1}$$

Ejemplo: Los gastos diarios en periódicos del hotel agrupados en la tabla de frecuencia:
Los cálculos necesarios para determinar la variancia de los gastos diarios son:

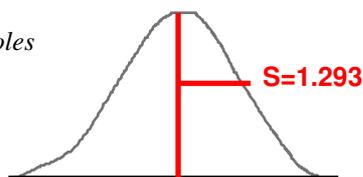
Intervalo	X_i	f_i	$f_i X_i$	$f_i X_i^2$
5.2 - 6.1	5.65	3	16.95	95.7675
6.1 - 7.0	6.55	5	32.75	214.5125
7.0 - 7.9	7.45	9	67.05	499.5225
7.9 - 8.8	8.35	7	58.45	488.0575
8.8 - 9.7	9.25	5	46.25	427.8125
9.7 - 10.6	10.15	3	30.45	309.0675
TOTAL		32	251.9	2034.74

$$S^2 = \frac{\sum_{i=1}^n f_i X_i^2 - n\bar{X}^2}{n-1} = \frac{2034.74 - 32(7.8719)^2}{32-1} = 1.671$$

3. DESVIACIÓN ESTÁNDAR

Es una medida de variabilidad que corresponde a la raíz cuadrada de la variancia. Este indicador tiene la misma unidad de medida en la que se expresa el promedio.

$$S = \sqrt{S^2} = \sqrt{1.671} = 1.293 \text{ soles}$$



4. COEFICIENTE DE VARIABILIDAD

Es una medida de variabilidad de los datos que se expresa en porcentaje en la cual se compara la desviación estándar con el respectivo valor del promedio de los datos:

$$C.V. = \left(\frac{S}{\bar{x}} \right) \times 100$$

Grado de variabilidad de los datos	Coefficiente de variabilidad
Con variabilidad baja	Menos de 10%
Con variabilidad moderada	De 10% a 30%
Con alta variabilidad	Más de 30%

En el ejemplo anterior el coeficiente de variabilidad es:

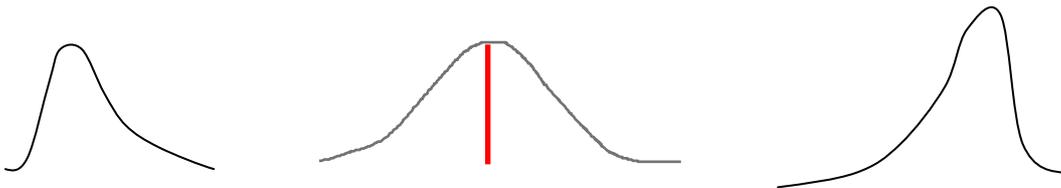
$$C.V. = \left(\frac{1.293}{7.87} \right) \times 100 = 16.4\%$$

C. MEDIDA DE FORMA: ASIMETRÍA O SESGO

Evalúa el grado de distorsión o inclinación que adopta la distribución de los datos respecto a su valor promedio tomado como centro de gravedad. El coeficiente de asimetría de Pearson es:

$$A_k = \frac{3(\bar{X} - M_e)}{S}$$

Grado de Asimetría	Valor del Sesgo
Simetría Perfecta	Cero. El promedio es igual a la mediana
Sesgo Positivo	Positivo. Promedio mayor que la mediana
Sesgo Negativo	Negativo. Promedio menor que mediana



Asimetría Positiva
(Promedio > Mediana)

Simétrica
Promedio = Mediana

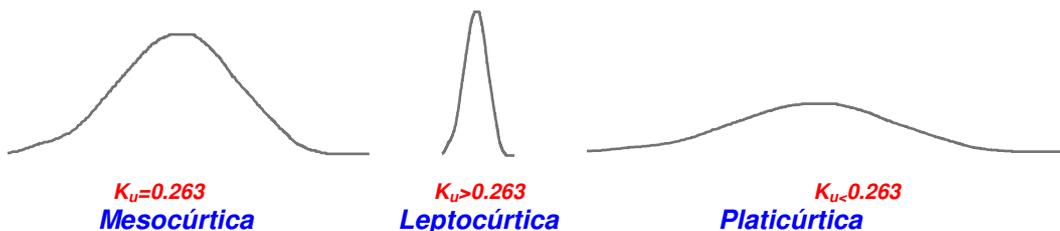
Asimetría Negativa
Promedio < Mediana

En el ejemplo sobre los gastos diarios en periódicos el Promedio es 7.87 le Mediana es 7.80 y la desviación estándar 1.293, por tanto el sesgo es ligeramente positivo +0.16

D. MEDIDA DE FORMA: CURTOSIS

Evalúa el grado de apuntamiento de la distribución, el coeficiente es: $K_U = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}$

Grado de Apuntamiento	Valor de la Curtosis
Mesocúrtica (Distribución normal)	0.263
Leptocúrtica (Elevada)	Mayor a 0.263 ó se aproxima a 0.5
Platicúrtica (Aplanada)	Menor a 0.263 ó se aproxima a 0



$K_U = 0.263$
Mesocúrtica

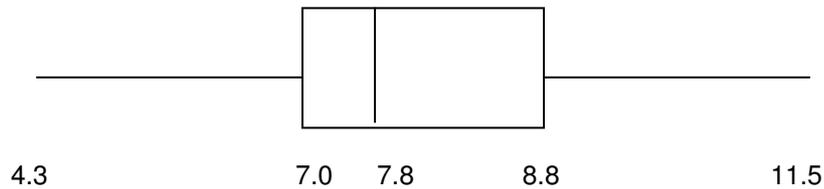
$K_U > 0.263$
Leptocúrtica

$K_U < 0.263$
Platicúrtica

En el ejemplo de los gastos diarios en periódicos como $Q_3=8.8$; $Q_2=7.0$; $P_{90}=9.7$ y $P_{10}=6.1$ la curtosis de la distribución es **0.25**; por tanto, la distribución es ligeramente *platicúrtica*.

GRÁFICOS DE CAJAS

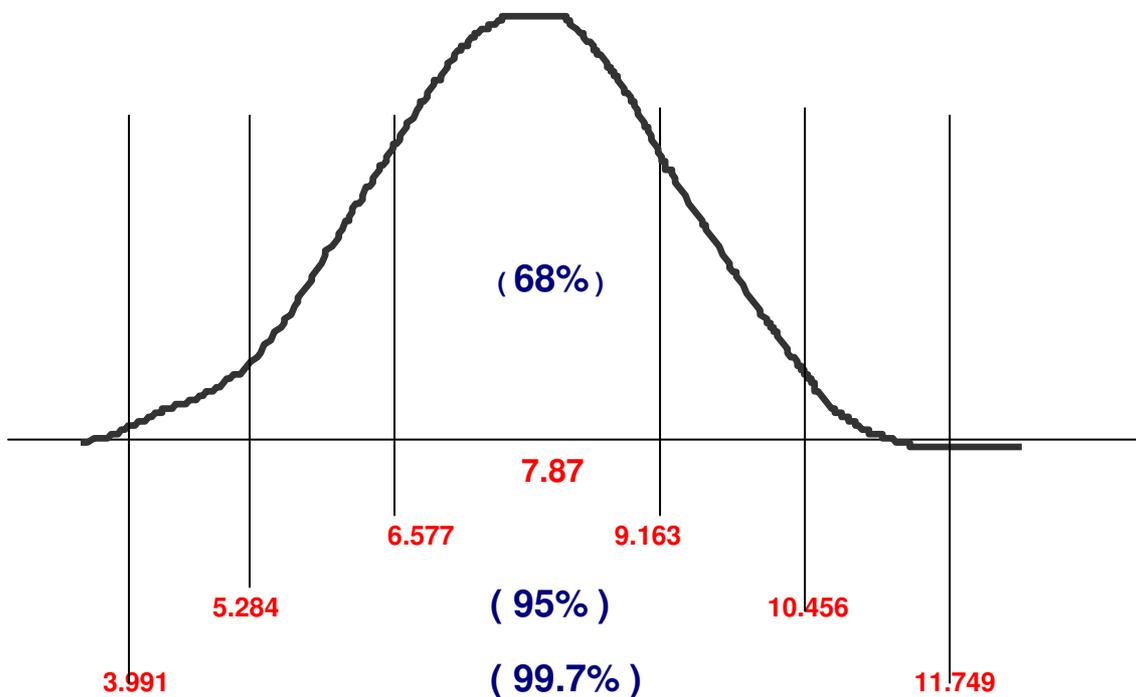
- Tercer Cuartil: $Q_3 = 8.8$
- Segundo Cuartil: $Q_2 = 7.8$
- Primer Cuartil: $Q_1 = 7.0$
- Rango Intercuatílico: $IQR = Q_3 - Q_1 = 8.8 - 7.0 = 1.8$
- Límite inferior: $Q_1 - 1.5(IQR) = 7.0 - 1.5(1.8) = 4.3$
- Límite Superior: $Q_3 + 1.5(IQR) = 8.8 + 1.5(1.8) = 11.5$



- La mitad (50%) de los datos son menores a 7.8
- La mitad (50%) de los datos toman valores entre 7.0 y 8.8
- La cuarta parte (25%) de los datos son menores a 7.0 (Antes de Primer Cuartil)
- La cuarta parte (25%) de los datos toman valores entre a 7.0 y 7.8
- La cuarta parte (25%) de los datos toman valores entre a 7.8 y 8.8
- La cuarta parte (25%) de los datos son mayores a 8.8 (Después del Tercer Cuartil)
- Los datos tienen mayor variabilidad entre 7.8 y 8.8.
- Los datos superiores a 11.5 y los datos inferiores a 4.3 se denominan **ATÍPICOS**

REGLA EMPÍRICA

Cuando la distribución de frecuencia es **simétrica**:



Si el Promedio es 7.87 y Desviación estándar 1.293 podremos afirmar que:

- 68% (22 datos) están entre: $[7.87 + 1(1.293)] = 9.163$ y entre $[7.87 - 1(1.293)] = 6.577$
- 95% (30 datos) están entre: $[7.87 + 2(1.293)] = 10.456$ y entre $[7.87 - 2(1.293)] = 5.284$
- 99.7% (32 datos) están entre: $[7.87 + 3(1.293)] = 11.749$ y entre $[7.87 - 3(1.293)] = 3.991$

TRANSFORMACIONES LINEALES DE VARIABLES

Si la variable X_i tiene promedio \bar{X} y variancia S_x^2 y sea la transformación lineal: $Y_i = aX_i + b$

- El promedio de la variables Y_i es : $\bar{Y} = a\bar{X}_i + b$
- La variancia de la variables Y_i es: $S_y^2 = a^2 S_x^2$
- La desviación estándar de la variables Y_i es: $S_y = a S_x$

Ejemplo: Las calificaciones de un examen de estadística son:

Nota	X_i	f_i	F_i	$f_i X_i$	$f_i X_i^2$
0-4	2	3	3	6	12
4-8	6	10	13	60	360
8-12	10	39	52	390	3900
12-16	14	38	90	532	7448
16-20	18	7	97	126	2268
Total		97		1114	13988

Promedio = 11.4845
 Mediana = 11.641
 Moda = 11.867
 Variancia = 12.44
 Desviación estándar = 3.53

Si el profesor decide transformar las calificaciones en la forma: $Y_i = 0.8X_i + 2$

- El promedio de la notas modificadas Y_i es : $\bar{Y} = 0.8(11.4845) + 2 = 11.1876$
- La mediana de la notas modificadas Y_i es : $M_e = 0.8(11.641) + 2 = 11.313$
- La moda de la notas modificadas Y_i es : $M_o = 0.8(11.867) + 2 = 11.493$
- La variancia de la variables Y_i es: $S_y^2 = 0.8^2(12.44) = 7.96$
- La desviación estándar de la variables Y_i es: $S_y = 0.8(3.53) = 2.82$

Verificación: Utilizando la tabla de frecuencia transformada donde $c=3.2$:

Nota	Y_i	f_i	F_i	$f_i Y_i$	$f_i Y_i^2$
2-5.2	3.6	3	3	10.8	38.88
5.2-8.4	6.8	10	13	68.0	462.40
8.4-11.6	10.0	39	52	390.0	3900.00
11.6-14.8	13.2	38	90	501.6	6621.12
14.8-18.0	16.4	7	97	114.8	1882.72
Total		97		1085.2	12905.12

Promedio = 11.1876
 Mediana = 11.313
 Moda = 11.493
 Variancia = 7.96
 Desviación estándar = 2.82