INFERENCIA ESTADISTICA

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

Tema 2

- ✓ Análisis exploratorio de datos.

2.1 Tablas de frecuencias.

Es La tabla de frecuencias (o distribución de frecuencias) es una tabla que muestra la distribución de los datos mediante sus frecuencias. Se utiliza para variables cuantitativas o cualitativas ordinales.
La tabla de frecuencias es una herramienta que permite ordenar los datos de manera que se presentan numéricamente las características de la distribución de un conjunto de datos o muestra.

Construcción de la tabla de frecuencias.

1. El número de intervalos de clase, k se puede elegir con Criterio de Sturges: $k = 1 + 3.3\log(n)$ ó $k = \sqrt{n}$
2. Calculamos el Ancho de Clase $A.C. = \frac{Rango}{k} = \frac{Obs.Máx - Obs.Min}{k}$

2.1 Tablas de frecuencias.

3. Una vez teniendo los datos anteriores, se calcula la Marca de Clase "X_i", es el promedio de los intervalos.
4. Frecuencia Absoluta, es el número de veces que un dato se repite o se encuentra dentro de un conjunto de datos. Se representa como f_i, donde la «i» corresponde al número de dato.
5. Frecuencia Absoluta Acumulada, es la suma de las frecuencias absolutas que se va acumulando hasta ese dato, es decir, la frecuencia absoluta acumulada de un dato en concreto se obtiene sumando su frecuencia absoluta a las frecuencias absolutas de los datos que son menores que él.
6. Frecuencia Relativa, es el número que se repite ese dato en relación al número total de datos, o en otras palabras, es la proporción de veces que aparece ese dato con respecto al total. $f_i = {}^{n_i}/_{N}$

2.1 Tablas de frecuencias.

7. Frecuencia Relativa Acumulada, es el mismo concepto que para la frecuencia absoluta acumulada.
8. Frecuencia Relativa Porcentual, es el mismo concepto que para la Frecuencia Relativa solo que con el porcentaje.
9. Frecuencia Relativa Porcentual, es el mismo concepto que para la Frecuencia Relativa Acumulada solo que con el porcentaje.

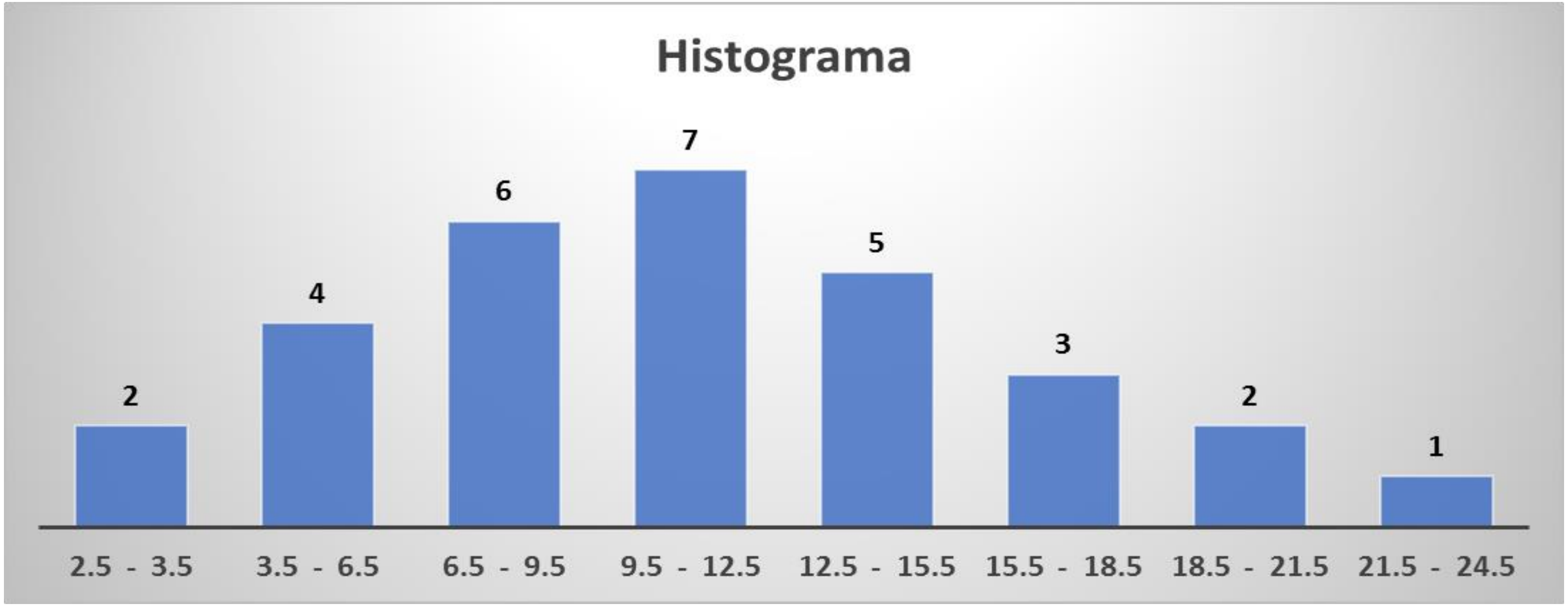
2.1 Tablas de frecuencias.

Ejemplo:

Invervalo	Marca de Clase (Xi)	Frecuencia Absoluta (n_i)	Frecuencia Absoluta Acumulda (N_i)	Frecuencia Relativa ($fi=n_i/N$)	Frecuencia Relativa Acumulada ($fi=n_i/N$)	Frecuencia Relativa ($fi=n_i/N$) en %	Frecuencia Relativa Acumulada ($fi=n_i/N$)
2.5 - 3.5	3.0	2	2	0.07	0.07	7%	7%
3.5 - 6.5	5.0	4	6	0.13	0.20	13%	20%
6.5 - 9.5	8.0	6	12	0.20	0.40	20%	40%
9.5 - 12.5	11.0	7	19	0.23	0.63	23%	63%
12.5 - 15.5	14.0	5	24	0.17	0.80	17%	80%
15.5 - 18.5	17.0	3	27	0.10	0.90	10%	90%
18.5 - 21.5	20.0	2	29	0.07	0.97	7%	97%
21.5 - 24.5	23.0	1	30	0.03	1.00	3%	100%
		30		**1**			

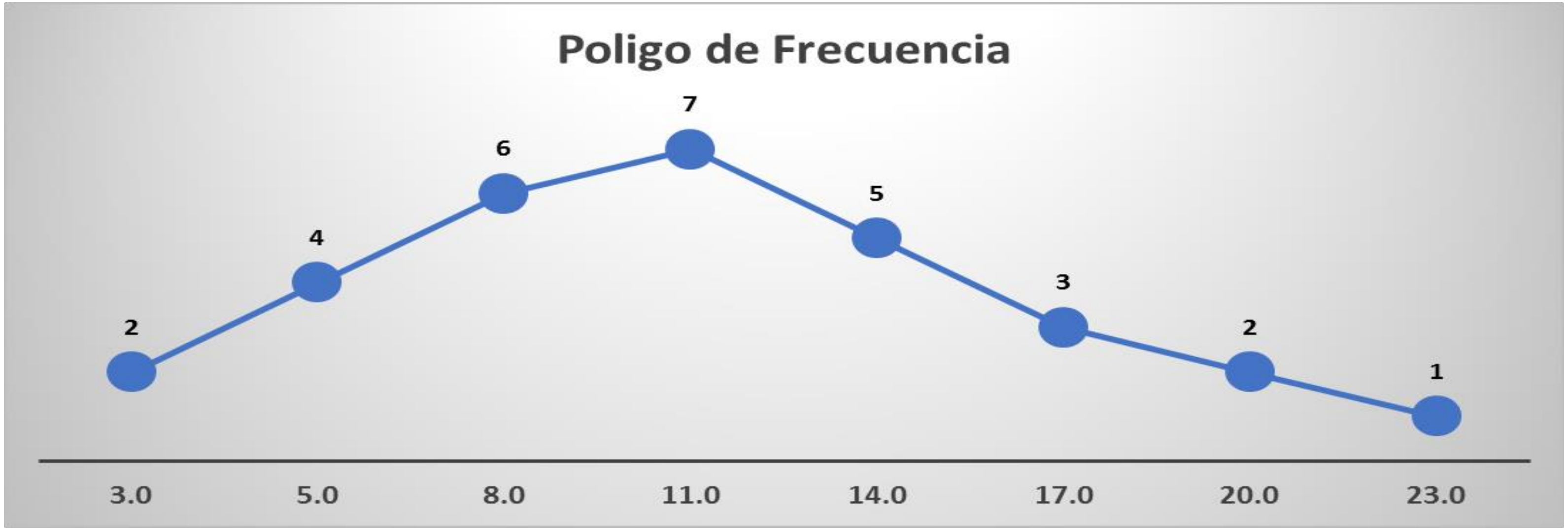
2.2 Métodos gráficos.

El Histograma, es un gráfico de la representación de distribuciones de frecuencias, en el que se emplean rectángulos dentro de unas coordenadas



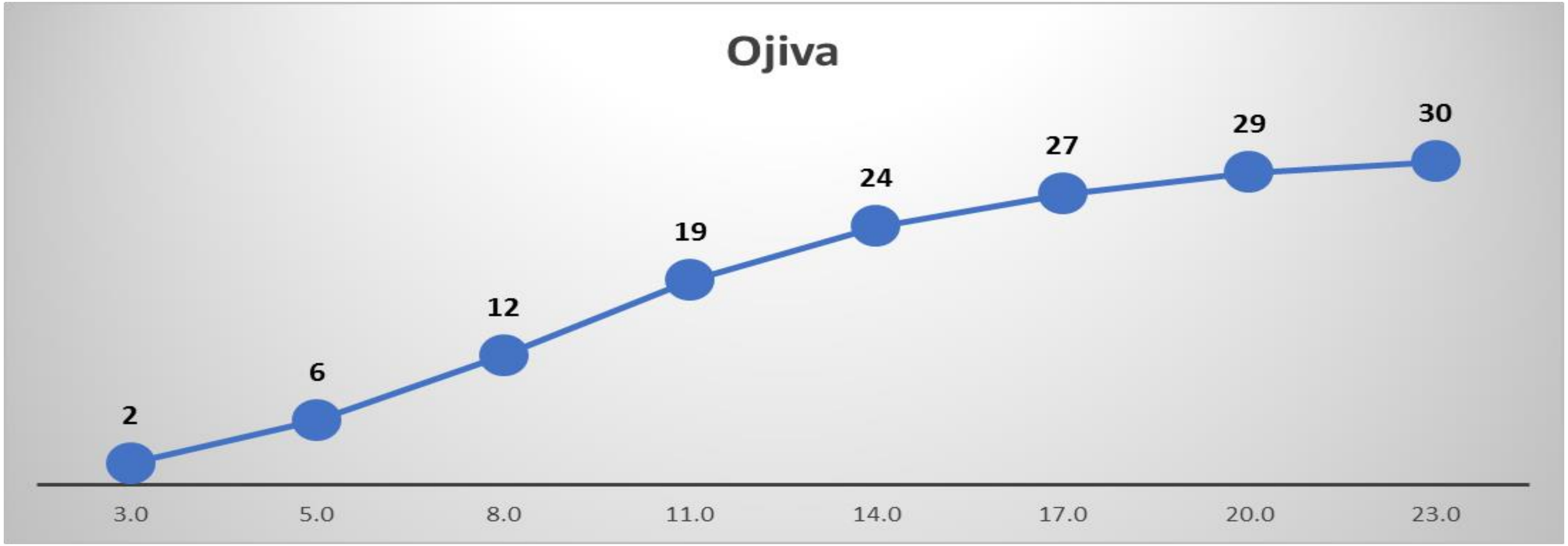
2.2 Métodos gráficos.

El polígono de frecuencias es un grafico que permite la rápida visualización de las frecuencias de cada una de las categorías del estudio. Normalmente se utiliza el polígono de frecuencias con frecuencia absoluta, pero también se utiliza con frecuencia relativas.



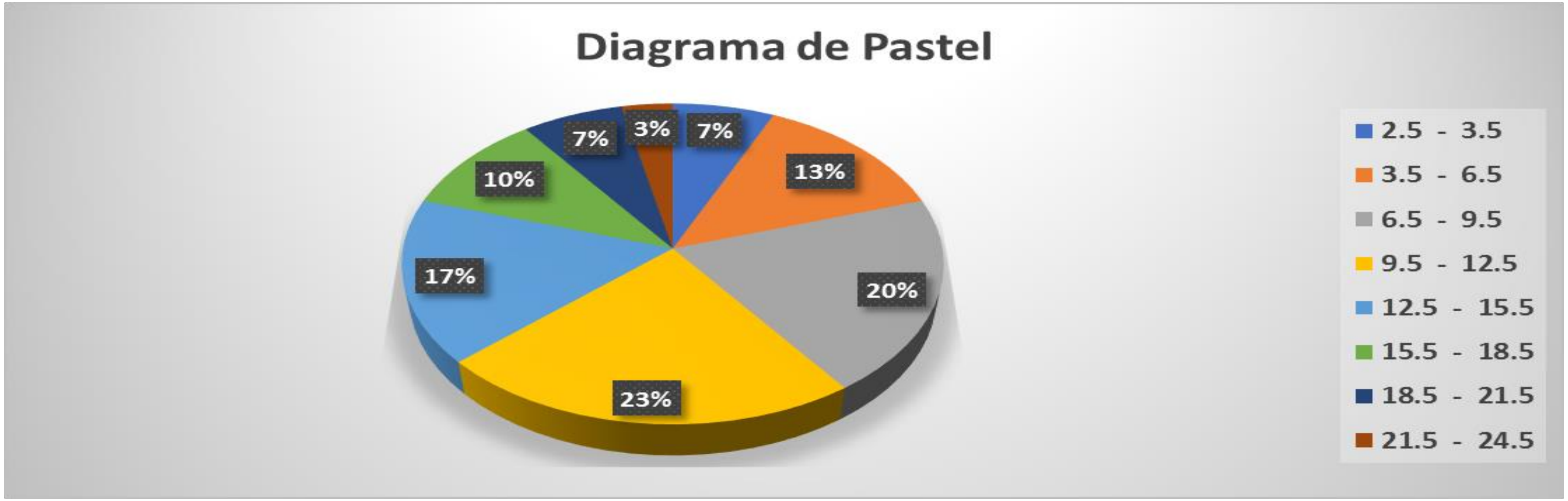
2.2 Métodos gráficos.

La ojiva es el polígono frecuencial acumulado, es decir, que permite ver cuántas observaciones se encuentran por encima de ciertos valores, en lugar de solo exhibir los números asignados a cada intervalo.



2.2 Métodos gráficos.

Un gráfico circular o gráfica circular, también llamado "gráfico de pastel", "gráfico de tarta", "gráfico de torta" o "gráfica de 360 grados", es un recurso estadístico que se utiliza para representar porcentajes y proporciones. El número de elementos comparados dentro de una gráfica circular suele ser de más de cuatro.



2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

Son estadígrafos de posición que son interpretados como valores que 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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE VARIABILIADAD

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MEDIA ARITMETICA. Para Datos No Agrupados.

El promedio aritmético de un conjunto de valores $(x_1, x_2, \dots, x_n)$ es:

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

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:

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MEDIA ARITMETICA. Para Datos Agrupados.

El promedio aritmético es:

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i X_i}{n} = \frac{3(5.65) + 5(6.55) + \cdots + 3(10.15)}{32} = \frac{251.9}{32} = 7.87$$

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

Intervalo	Xi	f_i	r_i	F_i	R_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7	6.55	5	0.156	8	0.250
7 - 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
		32	**1**		

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MEDIANA. 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: $\frac{(7+1)}{2} = 4$, por lo tanto la $M_e = 6$

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 $M_e = \frac{(5+6)}{2} = 5.5$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MEDIANA. Para Datos Agrupados.

$$M_e = L_i + \frac{c\left(\frac{n}{2} - F_{i-1}\right)}{f_i} = L_i + \frac{c(0.5 - R_{i-1})}{r_i}$$

Donde:

Li = Límite Inferior del intervalo que contiene a la Mediana

Fi-1 = Frecuencia Acumulada en la clase anterior i-ésima

fi = Frecuencia en la clase que contiene a la mediana

Ri-1 = Frecuencia Relativa Acumulada en la clase anterior i-ésima

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

c =Tamaño del intervalo de clase.

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MEDIANA. Para Datos Agrupados.

$$M_e = 7.0 + \frac{0.9\left(\frac{32}{2} - 8\right)}{9} = 7.0 + \frac{0.9(16 - 8)}{9} = 7.8$$

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

Intervalo	Xi	f_i	r_i	F_i	R_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7	6.55	5	0.156	8	0.250
7 - 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
		32	**1**		

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MODA. Para Datos No Agrupados.

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.

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 = Mo = 7.1; Es el valor más frecuente, ocurre 3 veces.

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MODA. 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 $d_2 = (f_i - f_{i+1})$ $\quad f_i = Valor\ de\ la\ mayor\ frecuencia$

Intervalo	Xi	f_i	r_i	F_i	R_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7	6.55	5	0.156	8	0.250
7 - 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
		32	**1**		

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

LA MODA. Para Datos Agrupados.

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

La moda estimada utilizando estos datos agrupados es:

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

Utilizando las frecuencias relativas, la moda estimada es:

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

El gasto diario en periódicos más frecuente es 7.6 pesos.

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

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: $(x_1, x_2, \dots, x_n)$ es:

$$\bar{X}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1, x_2, \dots, x_n} \quad \text{ó} \quad \bar{X}_G = \sqrt[n]{\frac{X_f}{X_i}}\ Donde\ X_f = Valor\ final\ y\ X_i = Valor\ inicial$$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

MEDIA GEOMÉTRICA.

Ejemplo 1: 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

$$\bar{X}_G = \sqrt[3]{(1.15)(1.10)(1.16)} = \sqrt[3]{1.4674} = 1.136 \qquad (13.6\%\, mensuales)$$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

MEDIA GEOMÉTRICA.

Ejemplo 2: 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:

$$\bar{X}_G = \sqrt[5]{(1.05)(1.00)(.99)(1.02)(1.04)} = 1.0197 \qquad (1.97\%\ anual)$$

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

Mes	0	Mes 1	Mes 2	Mes 3
Saldo	1000			1467.40

$$\bar{X}_G = \sqrt[3]{\frac{1467.40}{1000}} = 1.136$$

$$(13.6\%\ mensual)$$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

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".

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

Precio de Venta 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: $\bar{x} = .30(12) + .50(14) + .20(16) = 13.8$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

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:

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

n_i: Número de observaciones en el grupo i-ésimo.
$\bar{X}_i$: Promedio correspondiente el grupo i-ésimo

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

PROMEDIO TOTAL

Ejemplo:

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:

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

PROMEDIO TOTAL

Sigue Ejemplo:

Promedio Total		
Grupo	**X_i**	**fi**
A	12.5	25
B	10	40
Total		**65**

$$\bar{X}_T = \frac{25(12.7) + 40(10.0)}{65} = \frac{717.5}{65} = 11.04$$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

MEDIA ARMÓNICA

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

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

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.

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

MEDIA ARMÓNICA

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	9.98	13.5	10.45

$$\bar{X}_H = \frac{3}{\frac{1}{9.8} + \frac{1}{13.5} + \frac{1}{10.45}} = \frac{3}{0.26997} = 11.11242$$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

PERCENTILES, CUARTILES Y DECILES. Para Datos Agrupados.

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

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

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

Intervalo	Xi	f_i	r_i	F_i	R_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7	6.55	5	0.156	8	0.250
7 - 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
		32	**1**		

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

PERCENTILES, CUARTILES Y DECILES. Para Datos Agrupados.

Cuartiles: Son 3 valores Q1, Q2 y Q3 que dividen a los datos en 4 partes iguales. El Cuartil 3 (Percentil 75%) se ubicará en el cuarto

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

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

Intervalo	Xi	f_i	r_i	F_i	R_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7	6.55	5	0.156	8	0.250
7 - 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
		32	**1**		

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

PERCENTILES, CUARTILES Y DECILES. Para Datos Agrupados.

Deciles: Son 9 valores D1, D2,...,D9 que dividen a un conjunto de datos en 10 partes iguales. El Decil 7(Percentil 70%) se ubicará en el cuarto intervalo

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

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

Intervalo	Xi	f_i	r_i	F_i	R_i
5.2 - 6.1	5.65	3	0.094	3	0.094
6.1 - 7	6.55	5	0.156	8	0.250
7 - 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
		32	**1**		

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

PERCENTILES, CUARTILES Y DECILES. 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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE TENDENCIA CENTRAL

PERCENTILES, CUARTILES Y DECILES. Para Datos No Agrupados.

Ejemplo: Para la representación tallo hoja de los gastos en periódicos del hotel: Determine los 3 cuartiles correspondientes a los 32 datos ordenados:

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

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE VARIABILIADAD

AMPLITUD O RANGO

Sean los valores:$(X_1, X_2, \ldots, X_k)$. La amplitud o rango de estos dato es $A = (X_{max} - X_{min})$.

La Varianza o Variancia

Es una medida de la dispersión de una variable aleatoria (valores que se obtienen de manera aleatoria). La varianza, junto con la desviación estándar, son medidas de dispersión de datos u observaciones. La dispersión de estos datos indica la variedad que estos presentan, es decir, si todos los valores en un conjunto de datos son iguales, entonces no hay dispersión, pero en cambio, si no todos son iguales entonces hay dispersión.

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE VARIABILIADAD

VARIANCIA Para Datos No Agrupados.

La variancia de los datos de esta muestra $(X_1, X_2, \ldots, X_n)$:

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

Ejemplo Calcular la varianza de los 4 datos siguientes:

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

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}{n-1} = \frac{3^2+4^2+6^2+7^2-4(5)^2}{4-1} = \frac{10}{3} = 3.33$$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE VARIABILIADAD

VARIANCIA Para Datos Agrupados.

La varian[...] frecuencias $(f_1, f_2, \ldots, f_k)$ es:

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

Intervalo	Xi	fi	fiXi	fiX2i
5.2 – 6.1	5.65	3	16.95	95.77
6.1 – 7	6.55	5	32.75	214.51
7 – 7.9	7.45	9	67.05	499.52
7.9 – 8.8	8.35	7	58.45	488.06
8.8 – 9.7	9.25	5	46.25	427.81
9.7 – 10.6	10.15	3	30.45	309.07
		32	251.90	2034.74

Ejemplo: Utilizando la tabla de datos de los gastos de diarios. Los cálculos necesarios para determinar la variancia de los gastos diarios son:

$$S^2 = \frac{\sum_{i=1}^{n} f_i X_i^2 - n\bar{X}^2}{n-1} = \frac{2{,}034.74 - 32(7.8719)^2}{32-1} = 1.671$$

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE VARIABILIADAD

DESVIACIÓN ESTÁNDAR.

La desviación estándar es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media. Mientras mayor sea la desviación estándar, mayor será la dispersión de los datos.

$$s = \sqrt{s^2} = \sqrt{1.671} = 1.293\ pesos$$

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:

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDAS DE VARIABILIADAD

COEFICIENTE DE VARIABILIDAD.

Ejemplo: Utilizando la tabla de datos de los gastos de diarios de las ventas de periódicos.

$$C.V = \left(\frac{s}{\bar{x}}\right) * 100\% = \left(\frac{1.293}{7.87}\right) * 100\% = 16.4\%$$

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

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDA DE FORMA

ASIMETRIA O SESGO

Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética). La asimetría presenta tres estados diferentes, cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría.

$$A_k = \frac{3(\bar{X} - M_e)}{S} \text{ (Coeficiente de asimetría de Pearson)}$$

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

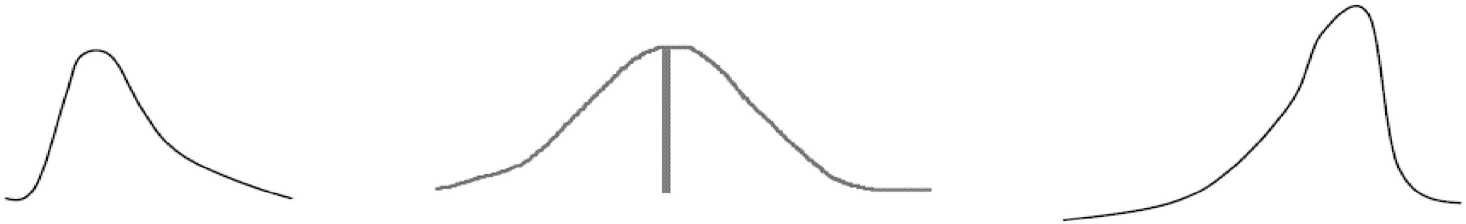
2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDA DE FORMA

ASIMETRIA O SESGO

Ejemplo: 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 +0.1624

$$A_k = \frac{3(7.87 - 7.80)}{1.293} = 0.1624$$



Asimetría Positiva (Promedio>Mediana) | Simétrica Promedio=Mediana | Asimetría Negativa Promedio<Mediana

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDA DE FORMA

CURTOSIS

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del Coeficiente de Curtosis, podemos identificar si existe una gran concentración de valores (Leptocúrtica), una concentración normal (Mesocúrtica) ó una baja concentración (Platicúrtica).

$$K_U = \frac{P_{75} - P_{25}}{2(P_{90} - P_{10})}$$

Grado de Apuntamiento	*Valor de la Curtosis*
Mesocurtica (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

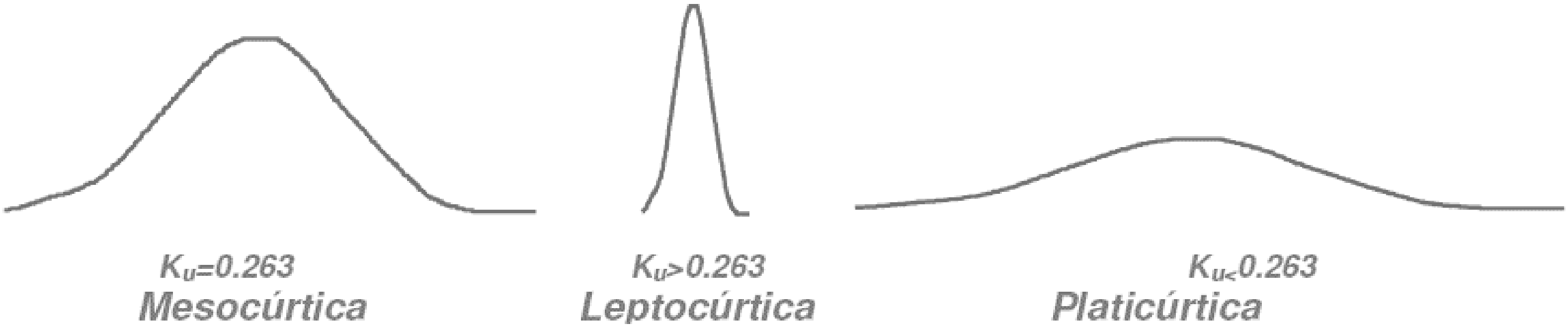
2.3 Medidas descriptivas para datos sin agrupar y agrupados.

MEDIDA DE FORMA

CURTOSIS

Ejemplo: En el ejemplo de los gastos diarios en periódicos como $Q_3=P_{75}=8.9$, $Q_1=P_{25}=6.85$, $P_{90}=9.7$ y $P_{10}=6.1$ la curtosis de la distribución es 0.2847, por tanto, la distribución es ligeramente Leptocútrica.

$$K_U=\frac{P_{75}-P_{25}}{2(P_{90}-P_{10})}=\frac{8.9-6.85}{2(9.7-6.1)}=0.2847$$



2.3 Medidas descriptivas para datos sin agrupar y agrupados.

DIAGRAMA DE CAJA (BOX PLOT)

El gráfico de caja ("box-plot" en inglés) es una forma de presentación estadística destinada, fundamentalmente, a resaltar aspectos de la distribución de las observaciones en una o más series de datos cuantitativos. Reemplaza, en consecuencia, al histograma y a la curva de distribución de frecuencias sobre los que tiene ventajas en cuanto a la información que brinda y a la apreciación global que surge de la lectura. Fue ideado por John Tukey, de la Universidad de Princeton (U.S.A.) en 1977.

Procedimiento para construir una gráfica de caja

1. Elabore el resumen de los 5 números consistente en el valor mínimo, Q_1, la mediana, Q_2, y el valor máximo.
2. Construya una escala con valores que incluyan el valor mínimo y el valor máximo.
3. Construya una caja (un rectángulo) que se extienda desde Q_1 hasta Q_3, y dibuje una línea en la caja, en el valor de la mediana.
4. Dibuje líneas que se extiendan hacia fuera de la caja hasta los valores mínimo y máximo.

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

Ejemplo: Se obtuvo una muestra aleatoria simple de las páginas del *Merriam-Webster's Collegiate Dictionary, decimoprimera edición*. A continuación se presentan, ordenados, los números de palabras definidas en estas páginas. Construya una gráfica de caja que incluya los valores del resumen de los 5 números.

34, 36, 39, 43, 51, 53, 62, 63, 73 y 79

Dato y Cuartil	Posición	Valor
D. Menor		34
Q_1	$Q_1 = 0.25(n+1) = 0.25(10+1) = 2.75$	$Q_1 = (36+39)/2 = 37.5$
Q_2	$Q_2 = 0.5(n+1) = 0.5(10+1) = 5.5$	$Q_2 = (51+53)/2 = 52$
Q_3	$Q_3 = 0.75(n+1) = 0.75(10+1) = 8.25$	$Q_3 = (63+73)/2 = 68$
D. Mayor		79

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

Ejercicio

Los datos que se dan a continuación son los recorridos de unidades correspondientes, a las operaciones de una línea de tráiler:

Fecha	2 IMPO/EXPO MTY	2 IMPO/EXPO SLT		Fecha	2 IMPO/EXPO MTY	2 IMPO/EXPO SLT
ene-00	2.91	2.89		oct-00	2.83	2.68
feb-00	3.04	3.08		nov-00	3.11	3.73
mar-00	2.83	2.98		dic-00	2.66	2.53
abr-00	2.92	3.43		ene-01	2.75	2.96
may-00	2.63	3.02		feb-01	3.07	2.83
jun-00	2.78	2.66		mar-01	2.79	2.79
jul-00	2.88	2.29		abr-01	2.94	3.02
ago-00	2.94	2.65		may-01	2.98	3.21
sep-00	2.40	2.91		jun-01	2.86	2.63

2.3 Medidas descriptivas para datos sin agrupar y agrupados.

Participación en Clase

a) Calcule las medidas de tendencia y variabilidad
b) Construya Histograma, Ojiva, Poligo y Pastel
c) Construya un diagrama de caja y explique su uniformidad
d) Calcular la asimetría y la curtosis
e) Calcular los deciles 2, 6 y 9.
f) Calcular los Percentiles 12%, 35%, 87% y 98%.

Utilizando la información de lo puntos anteriores, y sabiendo que los KM recorridos en promedio, a que conclusión podemos llegar:

Promedio de KM Recorridos por Mes	
2 IMPO/EXPO MTY	2 IMPO/EXPO SLT
61,852	74,920

Bibliografía.

Badii, M.H., J. Castillo, J.N. Barragán & A.E. Flores. 2007d. Análisis discriminante. Pp. 119-136. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Martínez Garza, A. 1988. Diseños Experimentales. Editorial Trillas. México.

Montgomery, D.C. 2001. Design of Análisis of Experiments. 15th Edition, Wiley. New Cork.

TRIOLA, M. F. (2004). Estadística. México: PEARSON EDUCACIÓN.