



COLEGIO DE BACHILLERES

ESTADÍSTICA DESCRIPTIVA E INFERENCIAL I

FASCÍCULO 2. MEDIDAS DE TENDENCIA CENTRAL

Autores: Juan Matus Parra



Colaboradores

Asesoría Pedagógica

Revisión de Contenido

Diseño Editorial

Leonel Bello Cuevas
Javier Darío Cruz Ortiz

ÍNDICE

INTRODUCCIÓN	5
PROPÓSITO	7
CUESTIONAMIENTO GUÍA	9
PLANTEAMIENTO DEL PROBLEMA	11
CAPÍTULO 1. MEDIDAS DE TENDENCIA CENTRAL	12
1.1 MEDIDAS DESCRIPTIVAS, MEDIDAS DE TENDENCIA CENTRAL O DE POSICIÓN	12
1.1.1 Moda	13
1.1.2 La Mediana	16
1.1.3 La Media	19
1.1.4 La Moda de Datos Agrupados	22
1.1.5 Cálculo de la Mediana de Datos Agrupados	25
1.1.6 La Media de Datos Agrupados	27
1.1.7 Rango	27
1.1.8 La Desviación Media	28
1.1.9 La varianza	41
1.1.10 Cálculo de la Varianza de Datos Agrupados	44
1.1.11 Desviación Estándar	45

1.2 MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN EN CURVAS DE FRECUENCIAS	51
1.2.1 Curvas Simétricas	51
1.2.2 Momentos	53
1.2.3 Curvas Asimétricas	56
1.2.4 Sesgo	57
RECAPITULACIÓN	61
ACTIVIDADES DE CONSOLIDACIÓN	62
AUTOEVALUACIÓN	63
BIBLIOGRAFÍA CONSULTADA	66

INTRODUCCIÓN

La estadística como una de las ramas de la Matemática, nos ayuda a organizar la información obtenida en la investigación de campo de una población o muestra.

Una manera de organizar esta información son las tablas de frecuencias. Las tablas de frecuencia son una síntesis de la información que nos facilita su representación gráfica en un histograma o polígono de frecuencias.

La gráfica nos facilita el análisis e interpretación de la información, la relación de los parámetros de tendencia central: Moda (M_o), Mediana (M_e) y Media (\bar{x}) y la relación de los parámetros de tendencia central con los de dispersión: Desviación media (DM); Rango, Varianza (S^2) y desviación estándar (S).

La gráfica también nos muestra con claridad, la íntima relación que hay entre los parámetros de dispersión y cómo influyen éstos en su forma, dando como resultado gráficas:

- a) Simétricas, o
- b) Asimétricas.

La forma e inclinación de la gráfica conocida como curtosis nos permite visualizar hacia donde se dispersan los datos.

En síntesis: la gráfica nos facilita el análisis y la interpretación de la información al representarla en forma sencilla y compacta.

En este fascículo aprenderás a cuantificar e interpretar la información de cualquier problema de investigación.

PROPÓSITO

La investigación científica y tecnológica ha dejado de ser una actividad de laboratorio. En la actualidad toda actividad o actitud humana, todo funcionamiento de una máquina debe estudiarse y analizarse con la finalidad de determinar si el funcionamiento de la máquina es óptimo, si la actuación o características de las personas son las deseadas. El conocimiento que se deduce de cada estudio, nos permite hacer correcciones que nos garanticen el logro de las metas fijadas en las políticas de trabajo de una empresa, de un laboratorio, etcétera.

Este fascículo pretende darte los elementos metodológicos fundamentales que te permitirán abordar el importante estudio de cualquier fenómeno que te interese conocer con el fin de buscar su óptima realización; también el aprendizaje de conceptos estadísticos importantes para el análisis de cualquier fenómeno que quieras incursionar para conocerlo y tomar una decisión. Estos conceptos que aparecen en el índice te permitirán una aventura agradable e interesante, pasando momentos satisfactorios como los que vive todo científico cuando encuentra la comprobación de su hipótesis.

CUESTIONAMIENTO GUÍA

Seguramente has escuchado en los medios de información, noticias como las siguientes:

- El promedio de las edades de la selección de fútbol es 18 años.
- La temperatura promedio del día es 24 grados.
- El sueldo promedio de los trabajadores de una empresa está por arriba del salario mínimo.

Esta información nos hace reflexionar y deducir algún conocimiento:

- En el primer ejemplo se deduce que la selección es un equipo de jóvenes.
- En el segundo ejemplo, que el clima es agradable.
- En el tercero, que la empresa remunera muy bien a sus trabajadores.

Sin embargo nuestra deducción no es siempre correcta, analicemos el último ejemplo:

La información que nos proporcionó la empresa es que tiene 15 trabajadores y su nómina diaria es de \$ 360.00 y el promedio del salario es de \$ 24.00. Este resultado nos hace pensar que los sueldos de la empresa de sus trabajadores es casi el doble del sueldo mínimo del cual se deduce que la remuneración a sus trabajadores es buena, sin embargo no es así porque en la nómina esta incluido el sueldo del gerente y esta cantidad disfraza el sueldo real de cada uno. Si el gerente percibe \$ 150.00 diarios, entonces ¿Cuál es el sueldo real de los otros trabajadores?.

Como habrás notado en este ejemplo, conocer los promedios no es suficiente, es necesario saber cómo están dispersos los datos. En este fascículo aprenderás a calcular, analizar e interpretar otros parámetros que describen con precisión la información obtenida de una muestra de una población en estudio.

PLANTEAMIENTO DEL PROBLEMA

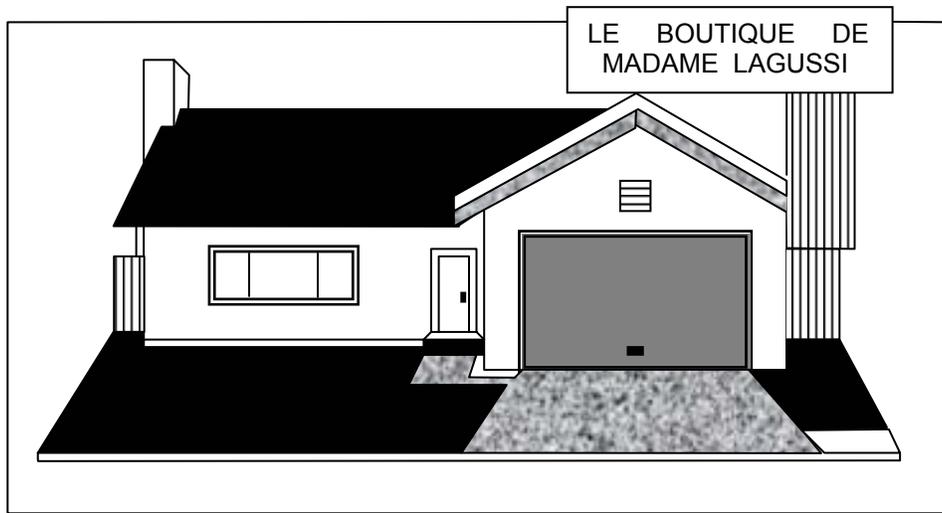


Figura 1

El sexo femenino suele vestir muy a la moda y esto se refleja en las ventas del mes de mayo en la cada de modas de Madame Lagussi.

M. Lagussi al hacer un balance obtuvo los siguientes resultados:

VENTAS DEL MES DE MAYO

Tipo de vestido	No. de vestidos vendidos
Suare	6
Corte de Princesa	8
Traje de Novia	6
Short	2
Maternidad	1
Bermudas	5
Minifaldas	10
Traje sastre	4
Vestido de cocktail	3

De acuerdo con esta información, ¿Qué tipo de ropa crees que esté de moda?

En toda actividad humana se presentan muchos problemas que se pueden analizar y estudiar en forma sistemática y lógica.

El estudio de estos problemas nos permiten tomar decisiones acertadas que nos conducen a lograr metas importantes de progreso.

La información que se obtiene de un determinado problema en estudio como el del ejemplo anterior, se resume mediante adecuadas descripciones estadísticas que llamamos:

CAPÍTULO 1

MEDIDAS DE TENDENCIA CENTRAL

1.1 MEDIDAS DESCRIPTIVAS, MEDIDAS DE TENDENCIA CENTRAL O DE POSICIÓN

Las medidas descriptivas que estudiaremos en este fascículo son de dos tipos:

- a) Medidas de Tendencia Central o de Posición y
- b) Medidas de Dispersión o de Variación.

Medidas de Tendencia Central o de Posición.

Las medidas de tendencia central describen las características básicas de un conjunto de datos. Son medidas representativas del conjunto y generalmente se resume mediante un valor numérico que indica la variación entre éstos.

Las medidas de tendencia central son:

- a) La Moda (M_o)
- b) La Mediana (M_e)
- c) La Media aritmética (\bar{X})

En esta unidad aprenderás el cálculo y aplicación de estas medidas. En estadística se usa el concepto de Moda (M_o) y representa lo mismo que dedujiste en el problema anterior.

En este ejemplo, el tipo de ropa con mayor frecuencia de venta es la minifalda, porque la frecuencia $f = 10$ es la mayor. En este caso denotamos $M_o = \text{minifalda}$.

Esta variable aleatoria describe una cualidad como en nuestro ejemplo (minifalda); pero también describe una cantidad representada por un valor numérico, como ya veremos.

1.1.1 Moda

La Moda (M_o) de un conjunto de datos es el valor (o cualidad) de la variable aleatoria que aparece con mayor frecuencia.

Ejemplos:

1. En la conferencia que se llevó a cabo en el hotel María Isabel Sheraton para altos ejecutivos, cuyo tema fue: "El tratado de libre comercio"; la asistencia fue la siguiente:

Día	No. de personas
Lunes	20
Martes	15
Miércoles	16
Jueves	16
Viernes	18
Sábado	16
Domingo	25

- a) ¿Cuál es la moda de la asistencia?
- b) ¿Qué tipo de variable es esta moda?
- c) ¿Qué día fue la moda de asistencia?
- d) ¿Qué tipo de variable aleatoria es el día de moda de asistencia?

R.

- a) La moda de la asistencia es $M_o = 16$.
- b) El tipo de variable es de cantidad.
- c) El día con mayor frecuencia en asistencia fue $M_o = \text{domingo}$.
- d) El tipo de variable es de cualidad.

2. En el entrenamiento de un equipo de básquetbol, los 5 jugadores estrella, dentro de los ejercicios de entrenamiento diario, deberán correr diariamente 5 minutos a toda su velocidad, tratando de rebasar diariamente su marca. Veamos el kilometraje del lunes:

Núm. de jugador	Recorrido en km.
1	8
2	5
3	7
4	3
5	6

¿Cuál es la moda del recorrido?

R.

Como pudiste notar en este ejemplo no hay moda porque ninguna distancia recorrida se repite. Esto siempre ocurre; es decir, si tienes un conjunto de datos en el cual cada dato aparece una sola vez, se dice que no hay moda.

3. El número de libros vendidos en la librería Acuario en la primera semana de clases del semestre 93B, registró las siguientes cantidades:

Lunes	Martes	Miércoles	Jueves	Viernes	Sábado	Domingo
150	200	180	200	275	200	300

- a) ¿Qué día fue la moda en ventas?
 b) ¿Cuál es la moda en ventas?

R.

- a) El día de moda en ventas fue el domingo porque la frecuencia en ventas es la mayor de todas, $M_o = \text{domingo}$.
 b) La moda en ventas es $M_o = 200$ porque es la cantidad que aparece con mayor frecuencia.

Como ya quedó definido, la variable aleatoria puede ser de cualidad o de cantidad. En este ejemplo lo hacemos notar nuevamente: fijaremos la variable dependiendo del interés que tengamos.

4. En un equipo de fútbol las estaturas del cuadro campeón son:

Delanteros	1.70	1.80	1.85	1.75 mts.
Medios	1.70			1.75 mts.
Defensas	1.75	1.70	1.75	1.70 mts.
Portero		1.88 mts.		

¿Cuál es la moda de las estaturas?

R.

Analizamos la frecuencia de cada medida y obtenemos:

Mts.	f.
1.70	4
1.75	4
1.80	1
1.85	1
1.88	1

La moda de las estaturas es:

Mo = 1.70 mts.

Mo = 1.75 mts.

De acuerdo a la definición de moda vemos que esta variable aleatoria tiene dos modas porque dos estaturas diferentes tienen la misma frecuencia y ésta es la mayor. En estos casos la variable se llama modal.

La variable aleatoria que tiene moda se llama unimodal.

La variable aleatoria que tiene dos modas se llama bimodal.

La variable aleatoria que tiene tres modas se llama trimodal, etc.

Por lo tanto en el ejemplo anterior la variable aleatoria es bimodal por tener dos modas.

Multimodal es la variable aleatoria que tiene más de una moda. Es unimodal si tiene solamente una.

ACTIVIDAD DE REGULACIÓN

Para cada uno de los ejercicios siguientes contesta las preguntas que se indican:

- ¿Cuál es la variable aleatoria?
- La moda es $Mo =$
- ¿La moda es de cualidad o cantidad? ¿Por qué?
- ¿La variable aleatoria es unimodal o multimodal? ¿Por qué?

1. Las calificaciones en estadística descriptiva obtenidas en un grupo de 20 alumnos son; 7, 9, 10, 8, 6, 5, 4, 7, 9, 8, 5, 10, 9, 6, 7, 5, 4, 8, 9, 7.
2. En la fiesta de graduación de la generación 90-93 del Colegio de Bachilleres Plantel No. 11, se hizo un análisis sobre el color del vestido, tomando como muestra al azar, a las primeras 15 damas. Se obtuvieron los siguientes resultados: rojo, negro, azul, blanco, rojo, verde, blanco, azul, amarillo, verde, gris, rojo, negro, azul, negro.

La moda es una medida de tendencia central porque su valor encuentra en la proximidad de otros valores que también se localizan en la parte central de los valores que toma la variable aleatoria. Estas son: la mediana (Me) y la media aritmética (\bar{X}), de las cuales nos ocuparemos en las secciones siguientes.

1.1.2 La Mediana

GRADUACIÓN DE LA GENERACIÓN 2002 DEL COLEGIO DE BACHILLERES



Fig. 2

En la fiesta de graduación de la Generación 2002 del Colegio de Bachilleres se observó el color del vestido de las primeras 25 damas que llegaron al salón de fiestas, obteniéndose los resultados que se muestran en la siguiente tabla:

Color del vestido	F _i
Blanco	2
Negro	5
Rojo	7
Azul	6
Verde	3
Gris	1
Amarillo	1
total	25

Al ordenar los datos de la tabla en sentido creciente obtenemos:

1, 1, 2, 3, 5, 6, 7

De este ordenamiento puedes observar que el valor que está en el centro de todos los datos es el número 3; este valor recibe el nombre de mediana. En este caso denotamos $Me = 3$. En nuestro ejemplo corresponde al color verde del vestido.

5. Para el programa “Apoyo a la Universidad”, la Sria. de Hacienda y Crédito público, publicó una lista de las empresas y el número de universidades del país que recibieron recursos de aquellas. Dicha lista es:

Empresa	A	B	C	D	E	F	G	H	I	J	K	L	M
Universidad	6	19	11	10	11	13	12	9	16	11	16	6	8

Al ordenar en sentido creciente el número de aportaciones por empresa, obtenemos: 6, 6, 8, 9, 10, 11, 11, 11, 12, 13, 16, 16, 19.

De este conjunto ordenado podemos ver que el número que se encuentra en el centro de los datos es el 11 y por lo tanto $Me = 11$.

MEDIANA es el valor de la variable aleatoria que se encuentra en el centro de un conjunto ordenado de datos.

Para determinar el valor de la mediana (Me) de un conjunto con (n) datos, si n es impar, entonces aplicamos la fórmula:

$$\frac{n+1}{2} \dots\dots\dots (1)$$

En el ejemplo de las donaciones, $n = 13$, sustituyendo en (1) obtenemos:

$$\frac{13+1}{2} = 7$$

El 7 indica el número de orden del elemento que representa la mediana que es $Me = 11$.

Si el número de datos (n) es un número par, entonces la mediana recae en los dos valores centrales.

6. Para $n = 50$, sustituyendo en (1) obtenemos:

$$\frac{50 + 1}{2} = \frac{51}{2} = 25.5$$

Este valor nos indica que la mediana se obtiene por la semisuma de los elementos que ocupan los lugares 25° (vigésimoquinto) y 26° (vigésimo sexto).

7. El número de transacciones en dólares, realizadas por el Banco Nacional de México durante los primeros días del mes de julio de 1993, fueron los siguientes: el 1^{er}. día 9, el 2^{do}. día 13 y los siguientes días 12, 8, 3, 7, 14, 16, 18 y 17 respectivamente.

Determina la mediana (Me) de la variable aleatoria:

Al ordenar los datos en sentido creciente obtenemos 3, 7, 8, 9, 12, 13, 14, 16, 17, 18.

Aplicando la fórmula (1) para $n = 10$ obtenemos:

$$\frac{10 + 1}{2} = \frac{11}{2} = 5.5$$

El valor obtenido nos indica que la mediana se obtiene por la semisuma de los elementos que ocupan los 5° y los 6° Lugares o sea los valores 12 y 13.

$$Me = \frac{12 + 13}{2} = \frac{25}{2} = 12.5$$

$$Me = 12.5$$

Este valor de la mediana nos indica que el valor central de los datos corresponde a 12.5 y como ya se dijo, si el número de datos es par, para determinar el valor de la mediana se toma la semisuma de los dos valores centrales. También esto ilustra que la mediana no es necesariamente un elemento del conjunto en cuestión.

ACTIVIDAD DE REGULACIÓN

En cada uno de los problemas siguientes determina la moda y la mediana de cada conjunto de datos. Analiza los resultados e indica tus observaciones.

1. La oficina de correos de la Ciudad de México envió en el mes de julio, a diferentes estados de la República, el siguiente número de paquetes: 78, 38, 47, 84, 49, 55, 42, 32, 66, 60, 94, 67, 6, 68, 70.
2. Las tallas más comunes de los vestidos que se vendieron en la boutique de Madame Lagusy, en el mes de julio son: 7, 10, 14, 9, 14, 9, 18, 9, 16, 12, 14, 11, 14.

1.1.3 La Media

LAS DIEZ PRIMERAS FINALISTAS DEL CONCURSO MISS UNIVERSO



Fig 3

Las edades de las 10 primeras finalistas del Concurso de Belleza, son: 18, 19, 25, 19, 20, 21, 20, 22, 18 y 18 años, respectivamente. ¿Cuál es la edad del promedio (\bar{X}) de las diez finalistas?

R.

Para poder determinar la edad promedio, hacemos una suma de las edades y el resultado lo dividimos entre el número de finalistas.

$$\bar{X} = \frac{18 + 19 + 25 + 19 + 20 + 21 + 20 + 22 + 18 + 18}{10} = 20$$

$$\bar{X} = 20$$

Este resultado nos dice que la edad promedio de las 10 finalistas del concurso Miss Universo es de 20 años.

8. En la temporada pasada los equipos de fútbol que más goles anotaron fueron los siguientes: El Atlante 47, el Pachuca 50, el América 45, el Guadalajara 39, el Cruz Azul 38, el Necaxa 29, el Universidad 39, la U. de G. 42, el Toluca 44, el Querétaro 43 y el Puebla 46.

¿Cuál fue el promedio de goleo de la temporada pasada?

Para determinar la media aritmética (\bar{X}) o promedio de n datos, se suman y el resultado se divide entre n.

De nuestro ejemplo obtenemos $n = 11$

$$\bar{X} = \frac{47 + 50 + 45 + 39 + 38 + 29 + 39 + 42 + 44 + 43 + 46}{11} = 42 \text{ goles}$$

$\bar{X} = 42$ goles: Promedio de goleo de la temporada pasada.

Si los valores de la variable los representamos con X_1, X_2, \dots, X_n entonces la representación simbólica de la media es:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \dots \dots \dots (2)$$

Esta expresión se puede representar en su forma más compacta si usamos el símbolo (Σ sigma), para expresar la suma de los valores de la variable, obteniéndose:

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \dots \dots \dots (3)$$

Esta expresión indica que el numerador del segundo miembro existe una suma de la variable X, el subíndice (i) indica que el valor de X es la variable y la anotación abajo y arriba de Σ indica que el subíndice (i) toma valores desde $i = 1$ hasta $i = n$, es decir; la suma se hace desde $X = X_1$ hasta $X = X_n$.

9. En el informe de la sala de Urgencias del Hospital General sobre los pacientes atendidos en la primera semana de julio, se obtuvieron los siguientes datos: Lunes 25, martes 24, miércoles 20, jueves 30, viernes 26, sábado 35 y domingo 29.

Determinar el promedio de pacientes atendidos en la semana.

Para obtener este promedio sustituimos los datos en la fórmula (3).

$$\bar{X} = \frac{\sum_{i=1}^7 X_i}{7} = \frac{25 + 24 + 20 + 30 + 27 + 35 + 28}{7} = 27$$

$X_1 = 25$
 $X_2 = 24$
 $X_3 = 20$
 $X_4 = 30$
 $X_5 = 27$
 $X_6 = 35$
 $X_7 = 28$
 $n = 7$

$$\bar{X} = 27$$

El promedio semanal de atención de pacientes en la sala de urgencias es de 27 pacientes.

ACTIVIDAD DE REGULACIÓN

En cada uno de los ejercicios siguientes, determina la M_o , M_e y \bar{X} , analiza tus resultados, compáralos y expresa tus observaciones.

1. En el departamento de control de calidad se tomó una muestra al azar de 10 focos para determinar el número de horas de vida de cada foco, obteniéndose los siguientes datos:

De muestra	1	2	3	4	5	6	7	8	9	10
De horas	865	850	841	850	820	843	830	848	840	838

2. La producción de tornillos especiales elaborados por un empleado de la fábrica Mecanican Falk en una de las semanas que se toma de muestra, fue la siguiente:

Día de la semana	lunes	martes	miércoles	jueves	viernes	Sábado
No. de tornillos	240	225	215	208	295	230

Hasta este momento, hemos calculado las medidas de tendencia central (M_o = moda, M_e = mediana \bar{X} = media) de datos no agrupados y esto se puede hacer debido a que las muestras tomadas de la población son pequeñas. Si la muestra es grande entonces es necesario agrupar los datos en tablas de frecuencias como ya lo estudiaste en el fascículo anterior.

Ahora estudiaremos cómo determinar la moda, la mediana y la media para datos agrupados.

1.1.4 La Moda De Datos Agrupados.

Para determinar la moda se construye una tabla de frecuencia como el siguiente ejemplo:

10. Un grupo de 40 alumnos, acumuló en un mes 43 retardos reportados en minutos en la siguiente tabla:

Clase en minutos	Frecuencia de retardos
1 – 3	3
4 – 6	5
7 – 9	8
10 – 12	12
13 – 15	6
16 – 18	5
19 – 21	4
Total	43

Con estos datos aplicamos la siguiente ecuación:

$$M_o = L_r + A \left(\frac{d_1}{d_1 + d_2} \right) \dots\dots\dots (4)$$

A = Intervalo de clase.

L_r = Límite real inferior de la clase modal.

d_1 = frecuencia de la clase modal menos la frecuencia de clase anterior a la modal.

d_2 = frecuencia de la clase modal menos la frecuencia de la clase siguiente.

Recordarás que la moda es el valor que se repite con mayor frecuencia. Para los datos agrupados en clases, la moda es el valor que se encuentra en la clase de mayor frecuencia y a esta clase se le llama clase modal.

Para aplicar la ecuación (4), el procedimiento es el siguiente:

1) En nuestra tabla de distribución de frecuencias visualizamos la clase modal y la marcamos con doble raya, para no equivocarnos en la toma de datos, toda vez que para aplicar la ecuación (4) toda la información está basada en esta clase.

2) Determinamos el límite real inferior de la clase modal $L_r = 9.5$

3) Calculamos $d_1 = 12 - 18 = 4$

4) Calculamos $d_2 = 12 - 6 = 6$

5) Calculamos A. Para el cálculo de A se resta al límite superior de cualquier clase, el límite inferior y al resultado se le suma la unidad. Para la clase modal de nuestro ejemplo A:

$$A = 12 - 10 + 1 = 2 + 1 = 3$$

Se sustituyen los datos en la fórmula y obtenemos:

$$M_o = 9.5 + 3 \left(\frac{4}{10} \right) = 9.5 + 1.2 = 10.7$$

La fórmula que hemos aplicado tiene su fundamento en el histograma que se obtiene al graficar los datos; analicemos la siguiente gráfica:

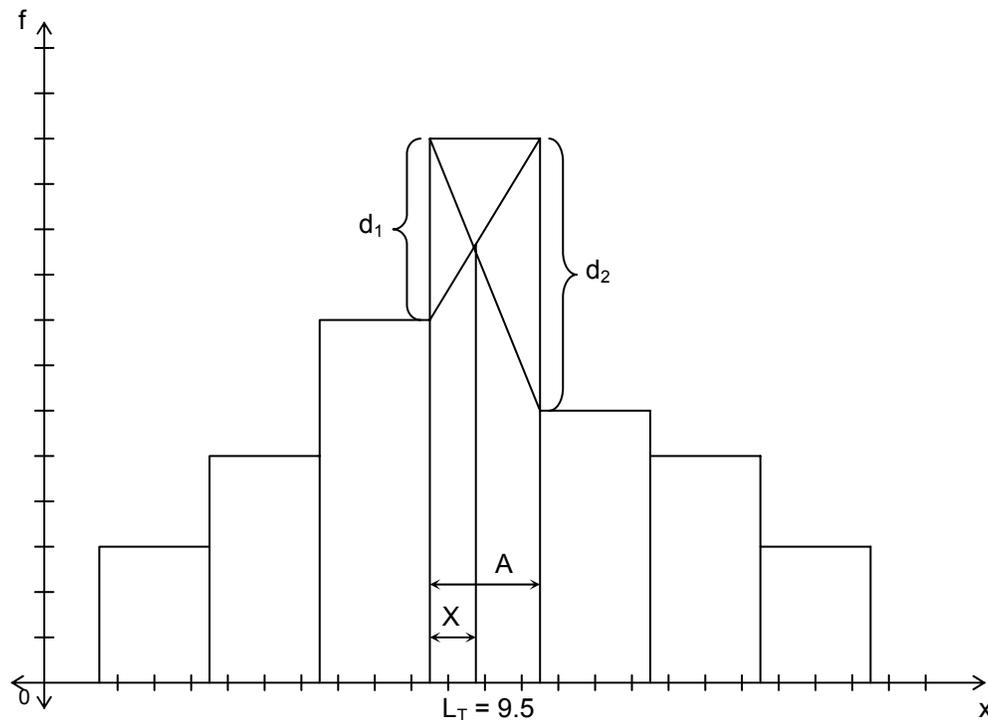


Fig 4

Analizando la gráfica vemos que el rectángulo mayor del histograma se localiza en la clase modal cuyo límite real inferior de clase es $Lr = 9.5$

d_1 es la distancia entre la frecuencia modal y la frecuencia de la clase anterior.

d_2 es la distancia entre la frecuencia mayor y la frecuencia de la clase siguiente a la modal.

A es intervalo de clase o amplitud de clase.

X es la distancia del límite real de clase al punto donde se localiza la moda; como estas distancias son proporcionales, podemos establecer la siguiente proporción:

$$\frac{X}{A} = \frac{d_1}{d_1 + d_2} \quad X = A \left[\frac{d_1}{d_1 + d_2} \right] \dots\dots\dots (a)$$

La moda es igual a la suma de las siguientes distancias:

$$Mo = Lr + X \dots\dots\dots (b)$$

Sustituyendo (a) en (b) obtenemos:

$$Mo = Lr + A \left[\frac{d_1}{d_1 + d_2} \right] \dots\dots\dots (c)$$

Del ejemplo anterior podemos concluir que la moda de datos agrupados se puede determinar de dos formas:

1. Aplicando la ecuación (4)
2. Cálculo gráfico.

Para el cálculo gráfico, una vez graficado el histograma, se trazan las diagonales uniendo los puntos superior derecho de los rectángulos de la clase modal y la clase anterior y los puntos superior izquierdo de la clase modal y la clase siguiente a ésta. En el punto donde se cortan las diagonales se traza una perpendicular al eje "X". El punto donde se cortan la perpendicular y el eje "X" es la moda de los datos agrupados.

ACTIVIDAD DE REGULACIÓN

En cada uno de los siguientes ejercicios, determina la moda (Mo) mediante la gráfica y comprueba el resultado, aplicando la fórmula.

1. En un grupo de 60 alumnos se obtuvieron los siguientes resultados en su examen final que contenía 20 reactivos. Los datos reportados son aciertos.

15,12,18,13,20,14,16,14,18,19
8,15,9,12,17,19,14,13,12,20
17,13,15,18,20,12,15,19,10,0
9,12,20,10,12,13,12,18,14,11
20,11,10,9,13,18,15,17,19,12
11,19,17,20,8,15,18,14,17,19

2. Las llamadas telefónicas diarias que una empresa hizo a sus clientes durante los dos primeros meses del año fueron:

6,12,7,15,13,18,16,20,25,12,10,8
13,15,6,9,18,20,25,12,7,10,11,13
9,12,15,18,20,13,17,23,25,14,18,6
14,16,9,6,10,12,20,13,17,14,25,7

1.1.5 Cálculo de la Mediana de Datos Agrupados.

La mediana (Me) de un conjunto de observaciones, agrupados en una tabla de distribución de frecuencias, se puede determinar aplicando la ecuación:

$$Me = L_i + A \left(\frac{\frac{n}{2} - F}{f} \right) \dots \dots \dots (5)$$

- L = límite inferior de la clase modal
- A = amplitud del intervalo de clase
- n = número de observaciones de la muestra
- F = frecuencia acumulada hasta la clase anterior a la clase modal
- f = frecuencia absoluta de la clase modal

Para aplicar esta ecuación, veamos el siguiente ejemplo:

11. El peso de 50 trabajadores de una empresa se representan en la siguiente tabla de distribución de frecuencias.

Intervalo de clase Kg.	Frecuencia absoluta	Frecuencia acumulada
53-57	2	2
58-62	7	9
63-67	10	19
68-72	12	31
73-77	9	40
78-82	6	46
83-87	4	50
total	50	

Para aplicar la ecuación, procedemos de la siguiente forma:

Sabemos que la mediana se encuentra en el centro de las observaciones, por lo tanto dividimos los n datos entre dos:

$$\frac{n}{2} = \frac{50}{2} = 25$$

De este resultado se deduce que la mediana ocupa el 25o. lugar, por lo que vemos en la frecuencia acumulada en qué clase se ubica, y ésta corresponde a la clase 68-72. Marcamos esta clase con doble raya para fácil referencia. Para determinar la mediana es necesario interpretar los datos, dado que no tenemos la mediana a la vista en la tabla sino que se encuentra entre los 12 valores que están incluidos en la clase 68-72.

Determinamos el valor de los parámetros observando los datos en la tabla:

$$\frac{n}{2} = 25$$

Ecuación:

$$Me = L_i + A \left(\frac{\frac{n}{2} - F}{f} \right) \dots \dots \dots (5)$$

$$L_i = 68$$

$$A = 77 - 73 + A = 4 + 1 = 5$$

$$F = 19$$

$$f = 12$$

Sustituyendo en la fórmula obtenemos

$$Me = 68 + 5 \left(\frac{25 - 19}{12} \right) = 68 + 5 \left(\frac{6}{12} \right) =$$

$$Me = 68 + 2.5 = 70.5 \text{ Kg.}$$

El valor de Me = 70.5 Kg. representa la mediana de los pesos de los 50 trabajadores de la empresa.

ACTIVIDAD DE REGULACIÓN

Determina la mediana en los problemas 1 y 2 del ejercicio anterior aplicando la fórmula (5).

1.1.6 La Media De Datos Agrupados

En tu fascículo anterior ya realizaste tablas de distribución de frecuencia para organizar los datos obtenidos de un problema. Los datos los agrupaste en clases. Las clases tienen el mismo intervalo de clase y el punto medio del intervalo de clase se llama marca de clase (M). Para determinar la marca de clase, se suman los extremos de la clase y el resultado se divide entre 2.

$$M_i = \frac{X_n + X_{n-1}}{2}$$

M_i = Marca de clase. Punto medio de clase.

La marca de clase es la mediana de las observaciones comprendidas en cada clase y ésta se obtiene solamente de dos valores como ya se indicó en la fórmula (6). La marca de clase no es útil porque para obtener el valor de las observaciones comprendidas en una clase, multiplicamos la frecuencias por la M. Este producto define todas las observaciones comprendidas en una clase. Para el cálculo de la media de datos agrupados, sumamos estos productos y los dividimos entre el número de observaciones.

La ecuación es la siguiente:

$$\bar{X} = \frac{M_1f_1 + M_2f_2 + \dots + M_n f_n}{n} = \frac{\sum_{i=1}^n M_i f_i}{n}$$
$$\bar{X} = \frac{\sum_{i=1}^n M_i f_i}{n} \dots \dots \dots (7)$$

12. Se desea saber la edad promedio de 50 ancianos de un asilo y para ello con la información se elabora la siguiente distribución de frecuencias:

Clase años	M _i	F _i	M _i f _i
70-72	71	15	1065
73-75	74	11	814
76-78	77	8	616
79-81	80	7	560
82-84	83	4	332
85-87	86	3	258
88-90	89	2	178
total		50	Σ = 3823

Sustituyendo los datos en la ecuación (7) obtenemos

$$\bar{X} = \frac{\sum_{i=1}^n M_i f_i}{n} = \frac{3823}{50} = 76.46 \text{ años es la edad promedio en el asilo de ancianos.}$$

Existe otro método para calcular la media de datos agrupados, lo llamaremos "Método abreviado para el cálculo de la media".

Este método consiste en fijar un origen muy aproximado a la media (\bar{X}) de los datos. Este valor que sirve de origen se resta a cada uno de los valores individuales obtenidos como diferencia, valores negativos y positivos cuya suma es cero; pero el producto de cada diferencia por la frecuencia nos da valores cuya suma es diferente de cero, con estas cantidades aplicamos la siguiente ecuación:

$$\bar{X} = X' + \frac{\sum_{i=1}^n \bar{X}' f_i}{n} \dots \dots \dots (8)$$

X' Es el valor que se toma como origen.

Se recomienda el valor que se tome como origen (X') sea un valor próximo a la mediana. El valor escogido de esta forma, facilita el cálculo y el valor obtenido de la media más exacta.

Veamos el siguiente ejemplo:

13. Las calificaciones obtenidas en la escala de 0 a 10 de un grupo de 50 alumnos se muestran en la siguiente tabla:

	Calificación	Calif.- origen $X_i - X' = \bar{X}'$	f_i	$\bar{X}' f_i$
	0	$0 - 5 = -5$	0	0
	1	$1 - 5 = -4$	0	0
	2	$2 - 5 = -3$	0	0
	3	$3 - 5 = -2$	2	-4
	4	$4 - 5 = -1$	3	-3
X'	(5)	$5 - 5 = 0$	4	0
	6	$6 - 5 = 1$	14	14
	7	$7 - 5 = 2$	12	24
	8	$8 - 5 = 3$	7	21
	9	$9 - 5 = 4$	4	16
	10	$10 - 5 = 5$	4	20
	Total		50	$\Sigma = 88$

$$X' = 5$$

$$n = 50$$

Sustituyendo en la ecuación obtenemos

$$\sum_{i=1}^n \bar{X}' f_i = 88$$

$$\bar{X} = X' + \frac{\sum_{i=1}^n \bar{X}' f_i}{n} = 5 + \frac{88}{50} = 5 + 1.76$$

$\bar{X} = 6.76$ promedio de calificaciones del grupo.

Cuando los datos están agrupados por clases en una tabla de frecuencias, se toma como punto de origen (X') un valor intermedio de las marcas de clase.

Se calculan las diferencias $Mc - X' = \bar{X}'$ y con este resultado se calculan los productos en $\bar{X}' f_i$ completando las columnas necesarias de la tabla de frecuencias; estos valores los sustituimos en la fórmula (8).

14. Aplicando esta fórmula calculemos la edad promedio de los 50 ancianos del problema (12), y comparemos los resultados obtenidos con ambos métodos.

	Clase años	M_i	F_i	$M_i - X' = \bar{X}'$	$\bar{X}' f_i$
	70 – 72	71	15	$71 - 77 = -6$	-90
	73 – 75	74	11	$74 - 77 = -3$	-33
$X' = 77$	76 – 78	77	8	$77 - 77 = 0$	0
	79 – 81	80	7	$80 - 77 = 3$	21
	82 – 84	83	4	$83 - 77 = 6$	24
	85 – 87	86	3	$86 - 77 = 9$	27
	88 – 90	89	2	$89 - 77 = 12$	24
	Total		50		$\Sigma = -27$

Sustituyendo valores en (8) , obtenemos:

$$\bar{X} = 77 + \frac{-27}{50} = 77 - 0.54 = 76.46$$

$\bar{X} = 76.46$ años es la edad promedio en el asilo de ancianos.

Al comparar este valor con el calculado por el otro método, podemos observar que el valor es el mismo.

ACTIVIDAD DE REGULACIÓN

PROBLEMA.

Calcula la media (\bar{X}) tomando como punto origen $X' = 80$ y repite los cálculos para $X' = 8$ compara tus resultados con el anterior y expresa tus conclusiones.

En el siguiente problema, aplica el método abreviado para calcular la media.

1. Una zapatería pone en remate 200 pares de zapatos cuyos precios y cantidades de existencia se indican en la siguiente tabla de frecuencias:

PRECIO EN \$	f_i
10 – 19	19
20 – 29	46
30 – 39	69
40 – 49	35
50 – 59	22
60 – 69	9
total	200

2. La producción de pantalones de una fábrica de ropa en el mes de julio y agosto fue la siguiente:

142, 163, 108, 157, 160, 124, 140, 132, 135,
 128, 135, 150, 149, 120, 126, 128, 145, 150,
 139, 149, 144, 129, 131, 137, 142, 163, 108,
 133, 146, 137, 149, 142, 163, 165, 158, 146
 120, 126, 150, 159, 160, 148, 120, 108, 110, 115

Calcula la media de la producción aplicando las ecuaciones (7) y (8) y compara los resultados. Explica cuál es el más recomendable usar y por qué.

La media de una muestra se simboliza como ya se dijo \bar{X} (equis testada) y la media de una población se simboliza como μ (letra griega mu).

El nombre que los estadísticos dan para diferenciarlas es:

\bar{X} (estadística de la muestra)
 μ (parámetro de la población)

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \dots \dots \dots (9)$$

Recuerda que *población* es todo el conjunto de observaciones obtenidas de un cierto fenómeno, y *muestra* es una parte representativa de la población.

El total de observaciones de la población se simboliza con (N) y los elementos de la muestra se simbolizan con (n).

También recordarás que, para que la muestra sea representativa de la población, esta debe ser una “muestra aleatoria”.

Hasta el momento hemos calculado medidas representativas de un conjunto de datos concentrados en un valor numérico que describe a todo el conjunto y cuyo valor es característico del mismo. A estas medidas les hemos llamado *medidas de tendencia central* porque su valor se encuentra cerca del centro de observaciones obtenidas al analizar un determinado problema y éstas son:

- a) La Moda (Mo)
- b) La Mediana (Me)
- c) La Media (\bar{X})

Hay otras características de las observaciones que son importantes conocer, por ejemplo: ¿Cuánto varían los datos entre sí? ¿Cuál es el grado de dispersión de los datos?. Esta información no la proporcionan las medidas de tendencia central, por lo que a continuación analizaremos cómo determinar estos nuevos valores.

Analícemos el siguiente ejemplo:

15. Las ventas realizadas en 10 días de uno de los meses por tres vendedores de una compañía fueron las siguientes:

Ventas de Pedro: 5, 18, 8, 12, 17, 19, 25, 17, 17, 20

Ventas de Andrés: 5, 17, 17, 17, 17, 17, 17, 17, 17, 17

Venta de Carlos: 5, 6, 10, 20, 21, 20, 19, 18, 19, 20

Ordena los datos y calcula las medidas de tendencia central de las ventas de cada uno.

Ventas de Pedro:	Mo	_____	Me	_____	\bar{X}	_____
Ventas de Andrés:	Mo	_____	Me	_____	\bar{X}	_____
Ventas de Carlos:	Mo	_____	Me	_____	\bar{X}	_____

De acuerdo a tus cálculos, la media de las ventas que obtuviste fue la misma para los tres:

$$\bar{X} = 15.8$$

¿Qué podrías deducir de estos resultados?

Expresa brevemente tu deducción: _____

De acuerdo a los valores de tendencia central de Andrés y Pedro, ¿Qué puedes decir de estos vendedores? _____

¿Cuál de los tres consideras que es más consistente en sus ventas? _____

¿Considera a Pedro y Carlos ¿Quién consideras que es más consistente en sus ventas?

Comenta tus resultados con tu asesor de contenido.

Los resultados que debiste obtener son:

Ventas de Pedro:	Mo	17;	Me	17;	\bar{X}	15.8
Ventas de Andrés:	Mo	17;	Me	17;	\bar{X}	15.8
Ventas de Carlos:	Mo	20;	Me	19;	\bar{X}	15.8

Esta información podría hacernos pensar que todo está bien, sin embargo no lo es, por lo que es necesario tomar alguna medida para mejorar las ventas.

Si observas los resultados anteriores, podrás notar que el valor de la media es el mismo para los tres, por lo que no nos indica en donde las ventas andan mal.

Un cálculo rápido y sencillo consiste en restar al valor mayor de los datos el valor menor, este resultado nos indica el rango en que están dispersos los datos.

Calcula el rango de dispersión de cada uno de los vendedores.

Los resultados que debiste obtener son:

Rango de ventas de Pedro $25 - 5 = 20$

Rango de ventas de Andrés $17 - 5 = 12$

Rango de ventas de Carlos $21 - 5 = 16$

Las ventas de Andrés son muy consistentes porque el rango de dispersión es de 12.

Las ventas de Carlos también son consistentes aunque menor que las de Andrés; la dispersión de los datos es 16.

Las ventas de Pedro registran el mayor rango de dispersión, por lo tanto nos indica que sus ventas no son consistentes.

Con este análisis debemos recomendar que se haga una investigación de campo para determinar las causas de la irregularidad de Pedro con el fin de dar una solución para mejorar.

La variabilidad y dispersión de los datos la proporcionan otras medidas que son:

1. El rango
2. La desviación media
3. La desviación estándar
4. La varianza

1.1.7 El Rango:

Una forma rápida de conocer la dispersión de los datos determinando el rango o recorrido de la variable aleatoria.

El rango (R) es la diferencia entre el valor mayor y el menor que toma la variable.

R = rango

$$R = X_n - X_1 \dots \dots \dots (10)$$

X_1 = el menor valor

X_n = el mayor valor

En el ejemplo anterior obtuvimos los rangos de dispersión de las ventas aplicando la ecuación (10).

Para determinar el rango de datos agrupados X_n es el límite superior del último intervalo de clase y X_1 es el límite inferior del primer intervalo de clase.

ACTIVIDAD DE REGULACIÓN

En cada uno de los ejercicios determina el rango, analízalo y da tus conclusiones.

1. El departamento de control de calidad tomó una muestra de 10 flechas al azar, para medir su diámetro y conocer si la tolerancia está dentro de las especificaciones. Se obtuvieron los siguientes datos (y las unidades son cm.):

12, 12.05, 12.03, 11.95, 12.01, 12.04, 12, 11.98, 11.99 y 12.03

2. El departamento de control de calidad de una fábrica de baterías tomó una muestra aleatoria de 10 pilas con el fin de determinar su vida útil. Se obtuvieron los siguientes resultados:

6.2 hrs., 6.8 hrs., 6.0 hrs., 6.4 hrs., 5.9 hrs., 6.6 hrs.,
5.8 hrs., 6.3 hrs., 6.1 hrs., 6.2 hrs., 6.8 hrs.

El rango es muy útil en el control de calidad porque en esa área manejan diferencias muy pequeñas (tolerancias). Sin embargo, la media y el rango no permite conocer con precisión cómo están dispersos los datos, por lo tanto es necesario introducir otra variable que permita conocer la dispersión y ésta es la desviación media que introduciremos en la siguiente sección.

1.1.8 La Desviación Media

La desviación (d_i) que hay de cada observación (X_i) con respecto a la media (\bar{X}) se obtiene mediante la siguiente ecuación:

$$d_i = X_i - \bar{X} \dots\dots\dots (11)$$

d_i = Desviación de cada observación con respecto a la media

La desviación es negativa para valores $X_i < \bar{X}$ y positiva para valores $X_i > \bar{X}$. La suma de todas las desviaciones es igual a cero y se simboliza mediante la siguiente ecuación:

$$D = \sum_{i=1}^n (X_i - \bar{X}) = 0 \dots\dots\dots (12)$$

Este valor $D = 0$ no nos ayuda en el cálculo; para evitar que la suma sea igual a cero, se toma el valor absoluto de cada desviación y la ecuación (12) se transforma en:

$$D = \sum_{i=1}^n |(X_i - \bar{X})| = \sum_{i=1}^n |d_i| \dots\dots\dots (13)$$

La desviación media (DM) es la suma de los valores absolutos de todas las desviaciones dividido entre el número de datos u observaciones.

Su definición matemática es:

Para una muestra:

$$DM = \frac{\sum_{i=1}^n |(X_i - \bar{X})|}{n} \dots\dots\dots (14)$$

Para una población:

$$DM = \frac{\sum_{i=1}^N |(X_i - \mu)|}{N} \dots\dots\dots (15)$$

- DM = desviación media
- X_i = 1-esimo dato u observación
- μ = media población
- \bar{X} = media muestral
- N = número de observaciones de la población
- n = número de observaciones de la muestra
- Σ = signo de sumatoria
- $| |$ = signo del valor absoluto

16. El gerente de personal entrevistó a 15 personas para su contratación; el tiempo que duró la entrevista de cada aspirante fue:

37 min., 30 min., 23 min., 46 min., 18 min., 40 min., 58 min.,
 43 min., 39 min., 55 min., 64 min., 42 min., 28 min., 20 min.,
 57 min.

Determina la desviación media de las entrevistas.

Determinamos la media de los datos:

n	X_i min.	$X_i - \bar{X}$	$ X_i - \bar{X} $
1	18	$18 - 40 = -22$	22
2	20	$20 - 40 = -20$	20
3	23	$23 - 40 = -17$	17
4	28	$28 - 40 = -12$	12
5	30	$30 - 40 = -10$	10
6	37	$37 - 40 = -3$	3
7	39	$39 - 40 = -1$	1
8	40	$40 - 40 = 0$	0
9	42	$42 - 40 = 2$	2
10	43	$43 - 40 = 3$	3
11	46	$46 - 40 = 6$	6
12	55	$55 - 40 = 15$	15
13	57	$57 - 40 = 17$	17
14	58	$58 - 40 = 18$	18
15	64	$64 - 40 = 24$	24
	Total		$\Sigma = 170$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \dots\dots\dots (3)$$

$$\bar{X} = \frac{600}{15} = 40$$

$$DM = \frac{\sum_{i=1}^n |X_i - \bar{X}|}{n} = \frac{170}{15} = 11.3 \dots\dots\dots (14)$$

DM = 11.3 minutos.

El valor de la desviación media de este ejemplo se debe interpretar que la desviación de los tiempos de cada entrevista es de 11.3 min. mayor o menor que la media, es decir el tiempo de cada entrevista se expresa por medio de la media en

$$\bar{X} \pm DM$$

Esto significa que los datos se distribuyen en torno a la media en:

$$\bar{X} \pm 11.3 \text{ minutos (Es una distancia promedio a la media).}$$

Si los datos están agrupados en una tabla de distribución de frecuencias, entonces para determinar la desviación media DM es necesario multiplicar cada desviación por su frecuencia y de esto se obtienen otra ecuación de la DM para datos agrupados, la cual es:

$$DM = \frac{\sum_{i=1}^n f_i |(X_i - \bar{X})|}{n} \dots \dots \dots (16)$$

f_i = frecuencia absoluta

17. La siguiente distribución de frecuencias corresponde al número de aviones que despegan diariamente en el aeropuerto capitalino tomando como muestra aleatoria los primeros 20 días de un mes.

X_i	f_i	$f_i X_i$	$X_i - \bar{X}$	$ X_i - \bar{X} $	$f_i X_i - \bar{X} $
41	1	41	41 - 46 = -5	5	5
42	2	84	42 - 46 = -4	4	8
44	4	176	44 - 46 = -2	2	8
46	6				
48	4				
50	2				
51	1				
	$\Sigma 20$	Σ	$\Sigma = 0$	Σ	Σ

- a) Completa la tabla de frecuencias.
- b) Determinamos la media de los datos.

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{920}{20} = 46$$

$$\bar{X} = 46$$

c) Determinamos la diferencia de cada observación menos la media. $X_i - \bar{X}$

d) Determinamos el valor absoluto de las desviaciones.

e) Realizamos el producto $f_i |X_i - \bar{X}|$ y sumamos los productos.

f) Sustituimos en la fórmula:

$$DM = \frac{\sum_{i=1}^n f_i |X_i - \bar{X}|}{n} = \frac{42}{20} = 2.1$$

DM = 2 El valor de la desviación media lo redondeamos al valor entero porque de acuerdo al problema, no tiene sentido hablar de fracciones de avión que despegan del aeropuerto.

Este resultado nos indica que los datos están distribuidos en 46 ± 2 aviones a partir de la media, que despegan diariamente.

Cuando los datos están agrupados en clase, para obtener la diferencia de cada observación con la media, se toma la M_i , esto es:

$$M_i - \bar{X} \dots \dots \dots (17)$$

con esto obtenemos otra ecuación para la desviación media:

$$DM = \frac{\sum_{i=1}^n f_i |M_i - \bar{X}|}{\sum_{i=1}^n f_i} \dots \dots \dots (18)$$

18. Las estaturas de la compañía de fusileros del batallón de guardias presidenciales se registran en la siguiente tabla de distribución de frecuencias, calcula la DM de las estaturas.

Intervalo de clase	M_i	f_i	$f_i M_i$	$ f_i - \bar{X} $	$f_i M_i - \bar{X} $
1.54 – 1.56	1.55	5	7.75	$1.55 - 1.62 = 0.07$	0.35
1.56 – 1.58	1.57	6	9.42	$1.57 - 1.62 = 0.5$	0.30
1.58 - 1.60	1.59	8	12.72	$1.59 - 1.62 = .03$	0.24
1.60 – 1.62	1.61	20	32.20	$1.61 - 1.62 = .01$	0.20
1.62 – 1.64	1.63	25	40.75	$1.63 - 1.62 = .01$	0.25
1.64 – 1.66	1.65	16			
1.66 – 1.68	1.67	9			
1.68 – 1.70	1.69	7			
1.70 – 1.72	1.71	3			
1.72 – 1.74	1.73	1			
Total		$\Sigma 100$	Σ		Σ

- a) Completa la tabla de frecuencias anterior.
b) Calculamos la media

$$\bar{X} = \frac{162.96}{100} = 1.62 \text{ metros}$$

- c) Calculamos las otras columnas de la tabla.
d) Aplicamos la ecuación (18)

$$DM = \frac{3.14}{100} = 0.0314$$

Este resultado dice que las estaturas se distribuyen en promedio en 0.0314 mts., a uno y otro lado de la media. Como éste es un valor muy pequeño, se interpreta que los datos están agrupados muy próximos a la media.

ACTIVIDAD DE REGULACIÓN

1. En una fábrica no se ha realizado debidamente el programa de mantenimiento por lo que una máquina está fuera de operación debido a las fallas que a menudo se presentan. El tiempo de inactividad consecutivos en minutos durante un año se registran en la siguiente tabla.

Esta información incluye el tiempo inactividad de algunas máquinas cuya falla ocurre más de una vez al día.

32	63	52	62	35	34	74	53	64	80
51	55	48	60	76	51	35	44	45	54
33	45	61	53	21	68	85	60	77	61
42	67	34	53	45	47	52	68	52	69
73	61	55	65	62	54	41	59	53	50
26	58	82	74	41	70	38	50	47	35
36	67	43	28	56	79	84	49	36	65
82	22	62	55	72	68	40	37	78	43
65	73	57	39	46	57	56	60	50	88
45	56	75	40	51	70	74	76	48	59

- Agrupar los datos en una distribución de frecuencias que contenga 17 clases.
- Elaborar un histograma.
- Calcular la M_o .
- Calcular la M_e aplicando la ecuación y gráficamente.
- Calcular la media aritmética \bar{X} .
- Calcular el rango.
- Calcular la DM.

Otras medidas que describen la dispersión de los datos son la:

- 1) Varianza S^2 ó σ^2 y la
- 2) desviación estándar S o σ

S es la desviación estándar de la muestra.
 σ es la desviación estándar de la población.

La desviación estándar es la medida más útil de la variación de los datos. En los ejemplos hemos podido observar que la dispersión de un conjunto de datos, es pequeña si los datos se reúnen muy cerca de la media y es grande si estos se dispersan ampliamente en torno de la media.

Vimos que la suma de las desviaciones es igual a 0 y como este valor no nos sirve, para el cálculo de la desviación media (DM) de un conjunto de datos, introducimos el valor absoluto de las desviaciones

$$X_i - \bar{X}$$

para eliminar el signo y de esta forma tener siempre un valor positivo en la suma de las desviaciones.

otra forma de eliminar el signo de las desviaciones es elevando al cuadrado de cada una de éstas, es decir;

$$(X_i - \bar{X})^2 = (X_i - \bar{X})^2 \dots \dots \dots (19)$$

Recuerda que todo número elevado al cuadrado siempre es positivo.

La ecuación (19) es el cuadrado de las desviaciones, estos valores nos permiten el cálculo de la varianza de la muestra (S^2) cuya definición matemática se da a continuación.

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \dots\dots\dots (20)$$

X_i = i-ésimo elemento de la muestra

\bar{X} = media de las observaciones

n = número de elementos de la muestra

$\sum_{i=1}^n (X_i - \bar{X})^2$ sumatoria del cuadrado de las desviaciones, desde la primer desviación i = 1 hasta la enésima i = n

El cálculo de la varianza (σ) de una población se determina mediante la siguiente ecuación:

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{N} \dots\dots\dots (21)$$

1.1.9 La Varianza

Es una medida de dispersión, en cuyo cálculo interviene el cuadrado de las desviaciones de cada puntuación. Para obtener su valor, calculamos la sumatoria de los cuadrados de las desviaciones y el resultado lo dividimos entre n-1 o N.

19. En un almacén de fábrica de jeringas, al estibar las cajas en forma incorrecta, se detectó que algunas se habían roto, por lo que se hicieron una inspección. Se tomaron 10 cajas al azar para su revisión habiéndose obtenido la siguiente información:

De la primera caja 2 jeringas rotas, de las siguientes: 3,1,0,4,2,1,3,0,2.

Determina la varianza de las jeringas rotas.

El procedimiento es el siguiente:

a) Organizamos los datos en una tabla:

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
0	0-1.8=-1.8	3.24
0	0-1.8=-1.8	3.24
1	1-1.8=-0.8	0.64
1	1-1.8=-0.8	0.64
2	2-1.8=0.2	0.04
2	2-1.8=0.2	0.04
2	2-1.8=0.2	0.04
3	3-1.8=1.2	1.44
3	3-1.8=1.2	1.44
4	4-1.8=2.2	4.84
18	$\Sigma = 0$	$\Sigma = 15.60$

b) Calculamos la media $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{18}{10} = 1.8$

$$\bar{X} = 1.8$$

c) Calculamos las desviaciones y las registramos en la tabla.

d) Calculamos el cuadrado de las desviaciones.

e) Con los datos de la tabla calculamos la varianza sustituyendo la ecuación:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{15.60}{10-1} = \frac{15.60}{9} = 1.73$$

La varianza de las jeringas rotas es de $S^2 = 2$ toda vez que no podemos hablar de fracciones de jeringa.

20. Regresemos al problema de los agentes de ventas (ejemplo 14) y calculemos la varianza de cada uno:

Pedro

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
5	-10.8	116.64
8	-7.8	60.84
12	-3.8	14.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
18	2.2	4.84
19	3.2	10.24
20	4.2	17.64
25	9.2	84.64

$$\bar{X} = \frac{158}{10} = 15.8$$

Andrés

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
5	-10.8	116.64
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44
17	1.2	1.44

$$\bar{X} = \frac{158}{10} = 15.8$$

Carlos

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
5	-10.8	116.64
6	-9.8	96.04
10	-5.8	33.64
18	2.2	4.84
19	3.2	10.24
19	3.2	10.24
20	4.2	17.64
20	4.2	17.64
20	4.2	17.64
21	5.2	27.04

$$\bar{X} = \frac{158}{10} = 15.8$$

Sustituyendo los datos en la ecuación (19), obtenemos:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \dots\dots\dots (20)$$

$$S_A^2 = \frac{129.6}{9} = 14.4$$

$$S_P^2 = \frac{313.6}{10-1} = \frac{313.6}{9} = 34.84$$

$$S_C^2 = \frac{351.6}{9} = 39.06 = \underline{351.6} = 39.06$$

$$S_A^2 = 14$$

$$S_P^2 = 35$$

$$S_C^2 = 39$$

Estos valores se han redondeado al entero más próximo porque no podemos hablar de ventas fraccionarias.

El mayor valor de la varianza nos dice que hay mayor dispersión de los datos y esto lo podemos constatar en las tablas.

La varianza de las ventas de Andrés es menor que las otras dos y de hecho, sólo un valor se aleja de la media, los demás están simétricamente agrupados alrededor de ésta.

La relación de la varianza de los tres vendedores es:

$$S_A^2 < S_P^2 < S_C^2$$

a) ¿Cuál de los tres tiene mayor varianza? _____

b) ¿Qué podemos inferir de este resultado? _____

c) Si tú fueras el gerente de ventas, ¿Qué medidas dictaminarías? _____

1.1.10 Cálculo de la Varianza con Datos Agrupados

Si los datos están agrupados en una tabla de frecuencia, entonces el cuadrado de cada desviación se multiplica por sus frecuencias obteniéndose la siguiente ecuación:

$$S^2 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{n - 1} \dots\dots\dots (22)$$

21. Tomamos los datos del ejemplo de las jeringas (19), los agrupamos en una tabla de frecuencias y obtenemos:

X_i	f_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$
0	2	-1.8	3.24	6.48
1	2	-1.8	0.64	1.28
2	3	0.2	0.04	0.12
3	2	1.2	1.44	2.88
4	1	2.2	4.84	4.84

Con los datos de la tabla calculamos S^2 aplicando la ecuación (21):

$$S^2 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{n - 1} = \frac{15.6}{10 - 1} = \frac{15.6}{9} = 1.73$$

$S^2 = 2$ este valor significa que la varianza de las jeringas rotas es 2, toda vez que no podemos hablar de fracciones de jeringa.

Compara los dos cálculos del problema y explica tus conclusiones.

Si los datos están agrupados en clases, entonces la puntuación que se toma para determinar la desviación de cada uno con respecto a la media, es la marca de clase M_i , obteniéndose la siguiente ecuación:

$$S^2 = \frac{\sum_{i=1}^n f_i (M_i - \bar{X})^2}{n - 1} \dots\dots\dots (23)$$

22. Tomemos los datos del ejercicio VIII referente a la falla de la máquina que constantemente está fuera de servicio. En este ejercicio debiste haber seguido este procedimiento de cálculo:

a) Ordenamos los datos de menor a mayor

- b) Los distribuimos en siete clases como se indicó
- c) Determinamos la marca de clase M_i
- d) Determinamos la frecuencia
- e) Calculamos la media de los datos multiplicando la M_i por la frecuencia y los sumamos
- f) Calculamos la desviación de cada dato con respecto a la media
- g) Elevamos al cuadrado las desviaciones
- h) Multiplicamos el cuadrado de las desviaciones por su frecuencia
- i) Sumamos este producto

Con esta información tomamos los valores de la tabla y los sustituimos en la ecuación (22) para el cálculo de S^2 .

Intervalo de clase	M_i	f_i	$f_i M_i$	$M_i - \bar{X}$	$(M_i - \bar{X})^2$	$f_i (M_i - \bar{X})^2$
20-29	24.5	4	98.0	-31.2	973.44	3893.76
30-39	34.5	12	414	-21.2	449.44	5393.28
40-49	44.5	18	801	-11.2	125.44	2257.92
50-59	54.5	26	1417	-1.2	1.44	37.44
60-69	64.5	20	1290	8.8	77.44	1548.80
70-79	74.5	14	1043	18.8	353.44	4948.16
80-89	84.5	6	507	28.8	829.44	4976.64
		$\Sigma 100$	$\Sigma 5570$		$\Sigma 2810.08$	$\Sigma 23056$

Cálculo de la media

$$\bar{X} = \frac{\sum_{i=1}^n f_i M_i}{\sum_{i=1}^n f_i} = \frac{5570}{100} = 55.7$$

$$S^2 = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{n - 1} = \frac{23056}{100 - 1} = \frac{23056}{99} = 232.88$$

$$S^2 = 233$$

La varianza de los tiempos en que la máquina está fuera de servicio por falla es de $S^2 = 233 \text{ min}^2$.

Como puedes notar en el resultado, los tiempos están en unidades cuadradas por lo que no nos da una idea precisa de los tiempos que la máquina está fuera de servicio toda vez que las unidades de la varianza no coinciden con las unidades de los datos, por lo que no podemos saber si la dispersión con respecto a la media es muy grande o pequeña.

Para poder determinar la dispersión de las puntuaciones con la misma unidad de la variable aleatoria, se usa otra medida de dispersión que se llama desviación estándar (S) o (σ).

S es la desviación estándar de la muestra y
 σ es la desviación estándar de la población.

1.1.11 Desviación Estándar

Es una medida de dispersión que es igual a la raíz cuadrada de la varianza.

Su definición matemática es:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}} \dots\dots\dots (24)$$

Datos no agrupados de una muestra aleatoria.

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n f_i (X_i - \bar{X})^2}{n - 1}} \dots\dots\dots (25)$$

Datos agrupados en frecuencias de una muestra aleatoria.

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n f_i (M_i - \bar{X})^2}{n - 1}} \dots\dots\dots (26)$$

Datos agrupados en clases de una muestra aleatoria.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^N f_i (M_i - \mu)^2}{N - 1}} \dots\dots\dots (27)$$

Datos agrupados en clases de una muestra población.

23. Calcula la desviación estándar del ejemplo anterior correspondiente a los tiempos fuera de servicio de la máquina por descompostura.

R.-

Como ya calculamos la varianza basta con extraerle la raíz cuadrada, obteniéndose:
 $S^2 = 233 \text{ min}^2$.

$$S = \sqrt{S^2} = \sqrt{233} = 15.26$$

$$S = 15.3$$

Esta información es más comprensible y debemos entender que los tiempos de descompostura de la máquina se distribuyen a $X \pm 15.3 \text{ min}$.

ACTIVIDAD DE REGULACIÓN

En cada uno de los siguientes problemas, calcula la varianza y la desviación estándar. Interpreta tus resultados y expresa tus deducciones.

1. La oficina de quejas de los microbuses del D.F. recibieron en una semana el siguiente número de quejas: Lunes 16, martes 10, miércoles 8, jueves 12, viernes 15, sábado 18 y domingo 20.
2. El bufete jurídico de una empresa envió cartas de cobranza a los clientes morosos del mes de Junio. El primer día del mes envió 64, el segundo día 62 y los siguientes días 51, 58, 83, 54, 57, 51, 42, 54, 74, 62, 47, 59, 51, 67, 53, 45, 58, 78, 69, 51, 72, 69, 78, 45, 64, 67.
3. La siguiente tabla muestra la distribución de frecuencias de las horas/hombre que requiere una compañía de pintura para pintar 100 casas clasificadas por el grado de deterioro.

Horas	No. casas
0 – 19	4
20 – 39	5
40 – 59	13
60 – 79	17
80 – 99	24
100 – 119	11
120 – 139	10
140 – 159	7
160 – 179	5
180 – 199	4

Existe una forma simplificada para calcular la desviación estándar y para ello se aplican las siguientes ecuaciones:

Para datos no agrupados

$$S = \sqrt{\frac{\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n \frac{X_i}{n}\right)^2}{n-1}} \dots\dots\dots (28)$$

$$S = \sqrt{\sum_{i=1}^n \frac{f_i X_i^2}{f_i} - \left(\sum_{i=1}^n \frac{f_i X_i}{n}\right)^2} \dots\dots\dots (29)$$

$$S = \sqrt{\sum_{i=1}^n \frac{f_i M_i^2}{f_i} - \left(\sum_{i=1}^n \frac{f_i M_i}{n}\right)^2} \dots\dots\dots (30)$$

En el siguiente problema, calcula la desviación estándar aplicando las dos ecuaciones. Compara los procedimientos y resultados.

24. Una compañía de computadoras recibió un pedido urgente de computadoras domésticas para todas las ciudades del país que deberá entregar en un periodo de 6 semanas. De acuerdo con el contrato firmado con los distribuidores, la entrega diaria es para el primer día 22, para los siguientes días 65, 77, 79,83, 65, 50, 75, 73, 60, 33, 66, 65, 30, 63, 41, 55, 65, 57, 62, 45, 49, 75, 59, 55, 54, 51, 28, 39, 25, 50, 48, 68, 55, 87, 35, 65, 65, 79, 61, 45, 53.

R.-

Secuencia de operaciones:

- a) Organizamos los datos en clases.
- b) Determinamos la marca de clase M_i
- c) Determina la frecuencia.
- d) Calculamos el producto $f_i M_i$
- e) Determinamos la media \bar{X} .
- f) Calculamos el producto $f_i M_i^2$
- g) Calculamos el producto $f_i (X_i - \bar{X})^2$
- h) Calculamos M_i^2
- i) Calculamos el producto $f_i M_i^2$

Aplicamos las fórmulas para datos agrupados, la general y la simplificada.

Fórmula general

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n f_i (M_i - \bar{X})^2}{\sum_{i=1}^n f_i}} \dots \dots \dots (26)$$

Fórmula simplificada

$$S = \sqrt{\frac{\sum_{i=1}^n f_i M_i^2}{\sum_{i=1}^n f_i} - \left(\frac{\sum_{i=1}^n f_i M_i}{n}\right)^2} \dots \dots \dots (30)$$

1	2	3	4	5	6	7	8	9
Clase	M_i	f_i	$f_i M_i$	$M_i - \bar{X}$	$(M_i - \bar{X})^2$	$f_i (M_i - \bar{X})^2$	M_i^2	$f_i M_i^2$
20-29	24.5	3	73.5	-31.9	1017.61	3052.83	600.25	1800.75
30-39	34.5	4	138.0	-21.9	479.61	1918.44	1190.25	4761.00
40-49	44.5	5	222.5	-11.9	141.6	708.05	1980.25	9901.25
50-59	54.5	10	544.0	- 2.0	4.0	40.0	2959.36	29593.60
60-69	64.5	12	774.0	8.1	65.61	787.32	4160.25	49923.00
70-79	74.5	5	447.0	18.1	327.61	1965.66	5550.25	33301.50
80-89	84.5	2	169.0	28.1	789.61	1579.22	7140.25	14280.50
		$\Sigma 42$	$\Sigma 2368$			10051.62		143561.6

- Cálculo de la media

$$\bar{X} = \frac{\sum_{i=1}^n f_i M_i}{\sum_{i=1}^n f_i} = \frac{2368}{42} = 56.4 \quad \therefore \quad \bar{X} = 56.4$$

- Cálculo de la varianza

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n f_i (M_i - \bar{X})^2}{\sum_{i=1}^n f_i}} \dots \dots \dots (26)$$

Tomados los valores de la columna 3 y 7, obtenemos:

$$S = \sqrt{\frac{10051.52}{42}} = \sqrt{239.32} = 15.46$$

Si $S = 15.46$, entonces $S^2 = 239.32$

Si analizamos las columnas de la tabla de frecuencias necesarias para cada ecuación, vemos que para la ecuación normal debemos elaborar de la primera a la séptima columna.

– Cálculo con la fórmula simplificada

$$S = \sqrt{\frac{\sum_{i=1}^n f_i M_i^2}{\sum_{i=1}^n f_i} - \left(\frac{\sum_{i=1}^n f_i M_i}{\sum_{i=1}^n f_i} \right)^2} \dots\dots\dots (30)$$

Tomados los valores de la columna 4 y 9 obtenemos

$$S = \sqrt{\frac{143561.6}{42} - \left(\frac{2368}{42} \right)^2} = \sqrt{3418.13 - (56.38)^2} = \sqrt{239.32} = 15.46$$

Si $S = 15.46$, entonces $S^2 = 239.32$

La desviación estándar del problema es $S = 15$ y este valor significa que el número de computadoras que se entregan diariamente se dispersa en ± 15 computadoras alrededor de la media (\bar{X}), es decir $X \pm 15$ es la entrega diaria de computadoras.

El valor decimal se ha eliminado toda vez que no podemos pensar que estamos entregando fracciones de computadora.

Para el cálculo de la desviación estándar con la ecuación simplificada, las columnas que deben elaborarse son las tres primeras, la 8 y la 9.

Los resultados obtenidos en ambas ecuaciones, son exactamente los mismos por lo tanto puedes usar para tus cálculos la que consideres que es más sencilla.

ACTIVIDAD DE REGULACIÓN

Para familiarizarte con la ecuación simplificada, repite los cálculos del ejercicio (IX) y compara tus resultados.

Las medidas de posición y las medidas de variación son descripciones estadísticas muy importantes, porque nos describen cómo se están agrupando o alejando los datos de una distribución con respecto a las medidas de tendencia central. Existen otras maneras de describir los datos estadísticos en forma más ilustrativa que un simple número que describe todo un proceso y que finalmente su interpretación se hace más compleja. La otra forma de analizar e interpretar los datos, es mediante su representación gráfica que a continuación estudiaremos.

1.2 MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN EN CURVAS DE FRECUENCIAS.

La representación gráfica de la distribución de frecuencias toma diferentes formas, más ilustrativas y fácilmente identificables, que nos permite conocer y comparar los estadísticos o parámetros visualmente.

Las formas en que se presentan las gráficas de una distribución de frecuencias se pueden generalizar en dos tipos:

- a) Curvas simétricas y
- b) Curvas asimétricas.

1.2.1 Curvas Simétricas

Curva simétrica es la representación gráfica de una distribución de frecuencias cuyo eje de simetría es la media \bar{X} o μ .

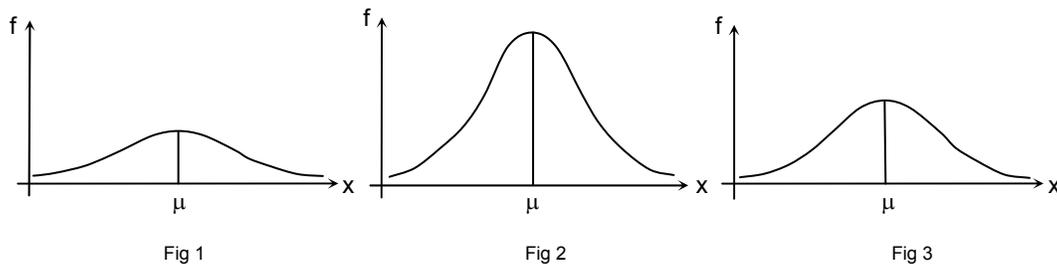
Las curvas simétricas se caracterizan por su curtosis.

Curtosis es la forma característica de la puntiagudez que presenta la parte superior de la gráfica.

Por su curtosis las curvas simétricas se clasifican en:

1. Particúrtica (fig. 1)
2. Lepticúrtica (fig. 2)
3. Mesocúrtica (fig. 3)

Las siguientes gráficas muestran la forma característica de cada una de ellas:



Analiza cada una de las gráficas y contesta las siguientes preguntas:

1. Expresa las características particulares de cada gráfica.
2. ¿Qué características son comunes a las tres gráficas?

R.-

1. Fig. 1. La Particúrtica presenta una zona casi horizontal en su punto máximo, su puntiagudez es casi nula.
Fig. 2. La Lepticúrtica presenta un pico muy agudo.
Fig. 3. La Mesocúrtica es semejante a la curva normal de Gauss.

2. Como podemos ver, las tres gráficas son simétricas con respecto a la media y tienen forma de campana.

Como ya se dijo, la curtosis es el grado de “puntiagudez” de la gráfica de una distribución de frecuencias y ésta se simboliza con la letra K.

K = Grado de apuntamiento o puntiagudez de la gráfica de la distribución de frecuencias.

El apuntamiento (K) es el cociente del promedio de las desviaciones con respecto a la media elevadas a la 4ª. Potencia entre el cuadrado de la varianza, disminuido en tres unidades.

De la definición de apuntamiento se desprenden nuevos estadígrafos que llamamos momentos (M).

1.2.2 Momentos

Momentos son los promedios de las series de potencias de las desviaciones de la variable con respecto a la media.

La definición matemática de cada uno de los momentos es:

1. Momento de primer orden o media aritmética:

$$M_1 = \frac{\sum_{i=1}^n X_i - \bar{X}}{n} = \frac{\sum X_i}{n} \dots\dots\dots (31)$$

2. momento de segundo orden o varianza:

$$M_2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum X_i^2}{n} = S^2 \dots\dots\dots (32)$$

3. Momento de tercer orden:

$$M_3 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n} = \frac{\sum X_i^3}{n} \dots\dots\dots (33)$$

4. Momento de cuarto orden:

$$M_4 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n} = \frac{\sum X_i^4}{n} \dots\dots\dots (34)$$

5. Momento de orden m:

$$M_5 = \frac{\sum_{i=1}^n (X_i - \bar{X})^5}{n} = \frac{\sum X_i^5}{n} \dots\dots\dots (35)$$

6. Momento de orden m para datos agrupados:

$$M_m = \frac{\sum_{i=1}^n f_i (X_i - \bar{X})^m}{n} \dots \dots \dots (36)$$

Basándose en los momentos, la definición matemática del apuntamiento (K) es:

$$K = \frac{M_4}{(M_2)^2} - 3 \dots \dots \dots (37)$$

M_4 = Cuarto momento

M_2 = Varianza

En el siguiente problema calcula los momentos sucesivos hasta el de cuarto orden y determina el apuntamiento (K).

23. Se tomaron al azar 20 alumnos para determinar su aptitud en la lectura; para ello se les dio a leer un texto de 100 palabras, habiéndose obtenido los siguientes tiempos de lectura en segundos: 70, 78, 60, 58, 80, 70, 75, 78, 60, 65, 70, 85, 78, 80, 75, 60, 55, 58, 78.

R.

Elaboramos la siguiente tabla:

	1	2	3	4	5	6	7	8	9	10
X_i	f_i	$f_i X_i$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$f_i (X_i - \bar{X})^2$	$(X_i - \bar{X})^3$	$f_i (X_i - \bar{X})^3$	$(X_i - \bar{X})^4$	$f_i (X_i - \bar{X})^4$	
55	2	110	-14.4	207.36	414.72	2985.98	5971.96	42998.17	85996.34	
58	2	116	-11.4	129.96	259.92	1481.54	2963.08	16889.60	33779.20	
60	3	180	-9.4	88.36	265.08	830.58	2491.74	7807.48	23422.46	
65	1	65	-4.4	19.36	19.36	85.18	85.18	374.80	374.80	
70	3	210	0.6	0.36	1.08	0.22	0.66	0.13	0.39	
75	2	150	5.6	31.36	62.72	175.61	351.22	983.45	1966.90	
78	4	312	8.6	73.96	295.84	636.06	2544.24	5470.08	21880.32	
80	2	160	10.6	112.36	224.72	1191.02	2382.02	12624.77	25249.54	
85	1	85	15.6	243.36	243.36	3796.42	3796.42	59224.09	59224.09	
	Σ	1388			$\Sigma 1786.8$		$\Sigma 20596.52$		$\Sigma 251894.04$	

$$\bar{X} = \frac{\sum_{i=1}^n f_i X_i}{n} = \frac{1388}{20} = 69.4$$

$$M_1 = 69.4$$

$$M_2 = \frac{1786.8}{20} = 89.34$$

$$M_3 = \frac{205960.52}{20} = 10298$$

$$M_4 = \frac{551894}{20} = 27594.7$$

Con estos valores sustituyendo en la ecuación (36) podemos calcular el grado de apuntamiento (K).

$$K = \frac{M_4}{(M_2)^2} - 3 \quad \dots \quad (37) \quad M_4 = \text{Cuarto momento}$$

$$M_2 = \text{Varianza}$$

$$M_4 = 27594.7$$

$$M_2 = 89.34$$

$$K = \frac{27594.7}{(89.34)^2} - 3 = 3.57 - 3$$

$$K = - 1.4$$

El valor de K define la curtosis de la gráfica, tal que:

$$E \begin{cases} = 0 & \text{La curva es mesocúrtica (curva normal)} \\ > 0 & \text{La curva es leptocúrtica} \\ < 0 & \text{La curva es platicúrtica} \end{cases}$$

En nuestro ejemplo donde $K = -0.21$ significa que $K < 0$ por lo tanto si graficamos la tabla de frecuencia, obtendremos una gráfica similar al de la figura 1 o sea una gráfica platicúrtica.

ACTIVIDAD DE REGULACIÓN

Del siguiente problema calcula el valor de K e indica qué tipo de gráfica tiene. Interpreta el resultado y expresa tus conclusiones.

1. De una guardería infantil se toma una muestra de 36 niños para determinar su altura. Los resultados obtenidos en centímetros son:

63, 64, 64, 65, 65, 66, 66, 66, 67, 67, 67, 67, 67, 68, 68, 68, 69, 69, 69, 69, 69, 70, 70, 70, 71, 72, 72, 72, 72, 73, 73, 74, 74, 76, 76, 77.

1.2.3 Curvas Asimétricas

Ya se dijo que las medidas de dispersión solamente señalan la magnitud de las variaciones, pero no dan ninguna información acerca de la dirección hacia la cual se dispersan.

Las curvas asimétricas son las que nos indican hacia donde se inclina la dispersión de los datos.

Curva Asimétrica es la representación gráfica de una distribución de frecuencias, la cual no es simétrica.

Estas gráficas se caracterizan por la posición que guardan las medidas de tendencia central. Podemos generalizar su presentación mediante las siguientes gráficas:

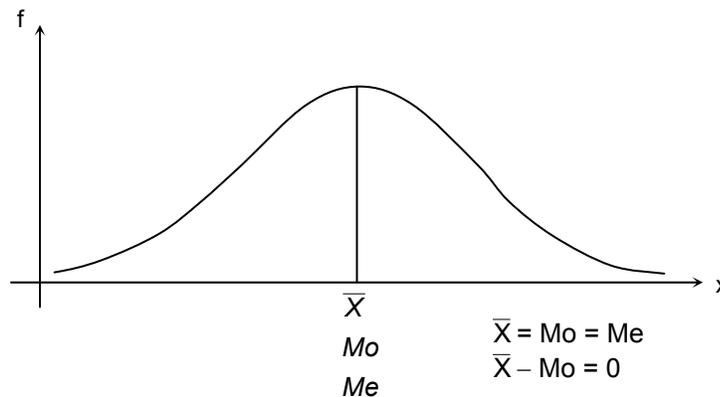


Fig. 4

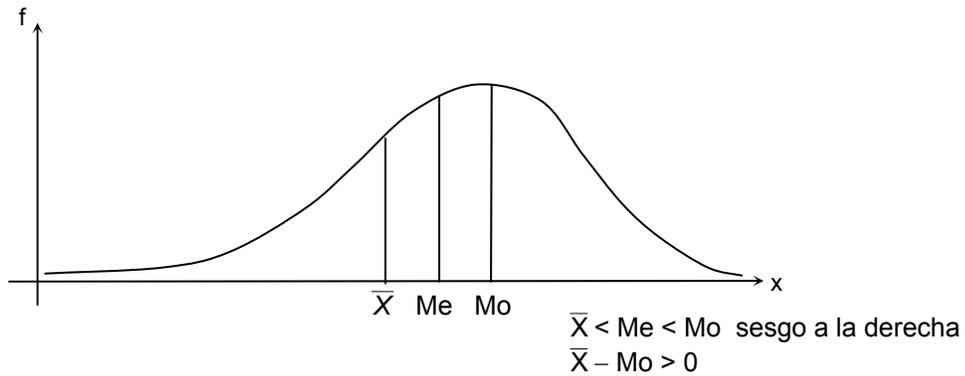


Fig. 5

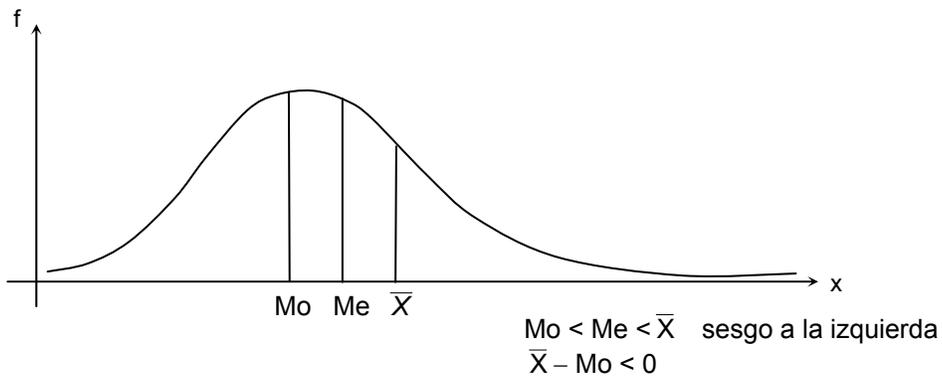


Fig. 6

Analiza cada una de las gráficas anteriores y contesta las siguientes preguntas:

a) ¿Qué características tiene cada una? _____

b) ¿Qué características tienen en común? _____

La asimetría de la gráfica se mide mediante el sesgo.

1.2.4 Sesgo

Sesgo es la medida de la asimetría de la gráfica y se calcula por el cociente de la media menos la moda entre la desviación estándar.

Su definición matemática es:

$$SG = \frac{\bar{X} - Mo}{S} \dots \dots \dots (37)$$

SG = Sesgo de la gráfica.

De acuerdo con esta ecuación y analizando el numerador podemos definir el sesgo de la gráfica de un conjunto de datos de la siguiente forma:

$$\frac{\bar{X} - Mo}{S} = 0 \quad \text{Implica que la curva está centrada con respecto a las medidas de tendencia central (curva simétrica)}$$

$$\frac{\bar{X} - Mo}{S} > 0 \quad \text{Implica que la gráfica está sesgada a la derecha.}$$

$$\frac{\bar{X} - Mo}{S} < 0 \quad \text{Implica que la gráfica está sesgada a la izquierda.}$$

Del siguiente problema calcula el apuntamiento (K) y el sesgo (SG), analiza tus resultados e indica cómo es la gráfica y cómo se distribuyen los datos del problema.

26. El departamento del Distrito Federal recibió varias quejas del público sobre la contaminación que produce la Planta de óxido de sulfuro. El Departamento envió un equipo de expertos para investigar el problema.

Para hacer el estudio de la contaminación, los investigadores se establecieron dentro de la Planta y tomaron una muestra diaria durante 40 días de los humos que ésta despiden. Los resultados obtenidos de la emisión promedio de óxido de sulfuro en kg. Por día, son los que se indican a continuación:

17, 15, 20, 29, 19, 18, 22, 25, 29, 9, 24, 20, 17, 16, 24,
14, 15, 23, 24, 26, 19, 23, 28, 19, 16, 22, 24, 17, 20, 13,
19, 10, 23, 18, 31, 13, 20, 17, 24, 14.

R.-

Secuencias de operaciones:

a) Ordenamos los datos en sentido creciente.

9, 10, 13, 13, 14, 14, 15, 15, 16, 16, 17, 17, 17, 17, 18,
18, 19, 19, 19, 19, 20, 20, 20, 20, 22, 22, 23, 23, 23, 24,
24, 24, 24, 24, 25, 26, 27, 28, 29, 31.

b) Los agrupamos en cinco clases y completamos las columnas necesarias para el cálculo.

	1	2	3	4	5	6	7	8	9
CLASE	f_i	M_i	$M_i f_i$	$M_i - \bar{X}$	$(M_i - \bar{X})^2$	$f_i (M_i - \bar{X})^2$	$(M_i - \bar{X})^4$	$f_i (M_i - \bar{X})^4$	
9-13	4	11	44	-9.125	83.3	333.2	6938.9	27755.6	
14-18	12	16	192	-4.125	17.0	204.0	289.0	3468.0	
19-23	13	21	273	0.875	0.76	9.95	0.58	7.6	
24-28	9	26	234	5.875	34.5	310.5	1190.3	10712.3	
29-33	2	31	62	10.875	118.3	236.6	13994.9	27989.8	
Total	40		Σ			$\Sigma 1094.3$		$\Sigma 69933.3$	

c) Cálculo de la media (\bar{X}). Sustituimos valores de la tabla en la ecuación (7) y obtenemos

$$\bar{X} = \frac{805}{40} = 20.125$$

d) Calculamos los momentos de segundo y cuarto orden, para ello sustituimos los datos de la tabla en las ecuaciones (33) y (35) obtenemos:

$$M_2 = \frac{1094.3}{40} = 27.36$$

$$M_4 = \frac{69933.3}{40} = 1748.3$$

e) Calculamos (K), para ello sustituimos los datos de la tabla en la ecuación (36) y obtenemos:

$$K = \frac{1748.3}{(27.36)^2} - 3 = 2.3 - 3 = -0.66$$

$$K = -0.7$$

Este resultado nos dice que $k < 0$, por lo tanto la gráfica de la distribución es de tipo platicúrtida.

f) Calculamos el sesgo (SG). Para este cálculo primero determinamos:

1. La Moda.

De la tabla obtenemos que la Moda $M_o = 21$ porque es el valor de mayor frecuencia.

2. Calculamos la desviación estándar.

Sustituimos los valores de la tabla en la ecuación (25) y obtenemos:

$$S = \sqrt{\frac{1094.3}{40}} = \sqrt{27.34} = 5.23$$

$$S = 5.23$$

Para calcular el sesgo (SG) sustituimos estos valores en la ecuación (37) y obtenemos:

$$SG = \frac{20.125 - 21}{5.23} = -0.167$$

$$SG = -0.2$$

Este resultado nos dice que $SG < 0$ lo cual implica que la gráfica esta cargada a la izquierda, por lo tanto los datos del problema están dispersos hacia la izquierda de la media.

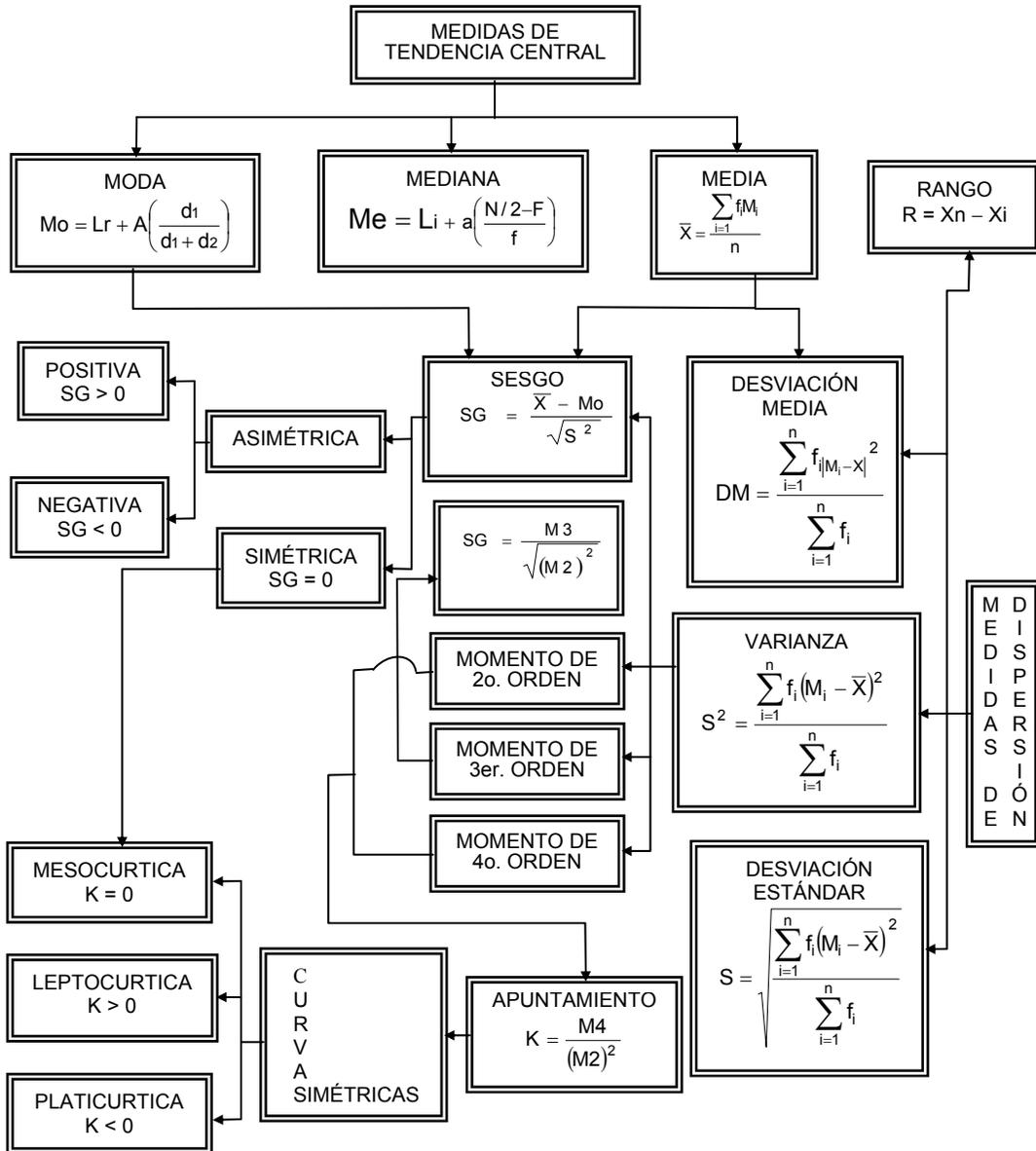
ACTIVIDAD DE REGULACIÓN

Para el siguiente problema calcula el apuntamiento (K) y el sesgo (SG). Analiza tus resultados e indica qué tipo de gráfica tiene el problema y cómo están dispersos los datos.

1. En el expendio "El huevo de oro" de la Central de Abastos de Iztapalapa, se vendieron durante los primeros 40 días del año, las siguientes cajas de huevo:

13, 19, 22, 14, 13, 16, 19, 21, 23, 11, 27, 25, 17, 17, 13, 20
23, 17, 26, 20, 24, 15, 20, 21, 23, 17, 29, 17, 19, 14, 20, 20
10, 22, 18, 25, 16, 23, 19, 20

RECAPITULACIÓN



ACTIVIDADES DE CONSOLIDACIÓN

Con el fin de afirmar los conceptos que estudiaste en este fascículo, del siguiente problema realiza lo que se indica:

1. Ordena los datos y agrúpalos en 9 clases.
2. Elabora la tabla de frecuencias.
3. Determina la moda de la muestra.
4. Calcula la mediana.
5. Calcula la media.
6. Traza el polígono de frecuencia.
7. Determina el rango de datos.
8. Calcula la desviación media.
9. Calcula el polígono de frecuencias.
10. Calcula la desviación estándar.
11. Calcula el 2° y 4° momentos.
12. Calcula el valor de K.
13. Calcula el sesgo.
14. Indica el tipo de gráfica que obtuviste.
15. Analiza tus cálculos y expresa tus conclusiones.

PROBLEMA.- Se aplicó una prueba de aptitud a 100 aspirantes de obrero para la planta de VW de Puebla, habiendo obtenido en los resultado las siguientes puntuaciones:

60, 25, 47, 61, 54, 78, 36, 67, 40, 51, 49, 53, 44, 94, 51, 65, 55, 61, 48, 70,
42, 58, 69, 40, 51, 88, 54, 83, 79, 14, 13, 72, 57, 27, 46, 62, 43, 51, 82, 45,
64, 52, 71, 82, 53, 41, 65, 62, 75, 60, 49, 64, 40, 61, 73, 80, 71, 53, 36, 90,
60, 59, 41, 29, 86, 57, 61, 85, 44, 92, 27, 56, 39, 43, 54, 35, 59, 59, 89, 60,
61, 71, 53, 58, 26, 77, 68, 62, 62, 57, 48, 69, 76, 52, 49, 45, 54, 41, 77, 85.

Comenta tus resultados con tu asesor o profesor.

AUTOEVALUACIÓN

A continuación se presentan las respuestas del ejercicio que realizaste en las actividades de consolidación, compáralos con tus resultados y si encuentras alguna diferencia entonces revisa tus procedimientos de cálculo y corrige tus resultados.

Si tienes alguna duda, consulta a tu profesor o asesor.

1. Datos ordenados en 9 clases.

13, 14, 25, 26, 27, 27, 29, 35, 36, 36, 39, 40, 40, 40, 41, 41, 41, 42, 43, 43, 44, 44, 45, 45, 46, 47, 48, 48, 49, 49, 49, 51, 51, 51, 51, 51, 52, 52, 53, 53, 53, 53, 54, 54, 54, 54, 55, 56, 57, 57, 57, 58, 58, 59, 59, 59, 60, 60, 60, 61, 61, 61, 61, 62, 62, 62, 62, 64, 65, 65, 67, 68, 69, 69, 70, 71, 71, 71, 72, 73, 75, 76, 77, 77, 78, 79, 80, 82, 82, 83, 85, 85, 86, 88, 89, 90, 92, 94.

Clases: (10-19), (20-29), (30-39), (40-49), (50-59), (60-69), (70-79), (80-89), (90-99).

2. Tabla de frecuencias.

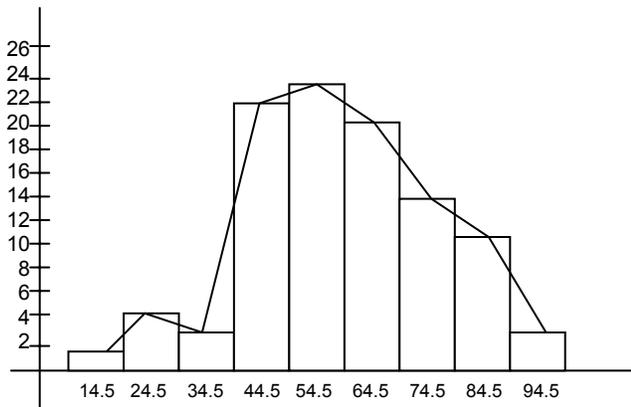
	1	2	3	4	5	6	7	8	9
CLASE	f_i	M_i	$M_i f_i$	$M_i - \bar{X}$	$(M_i - \bar{X})^2$	$(M_i - \bar{X})^4$	$ f_i (M_i - \bar{X}) $	$f_i (M_i - \bar{X})^2$	
10-19	2	14.5	29	-43	1849	3418801	86	3698	
20-29	5	24.5	123	-33	1096	1200361	165	5445	
30-39	4	34.5	138	-23	529	279841	92	2116	
40-49	20	44.5	890	-13	169	28561	260	3380	
50-59	26	54.5	1417	3	9	81	78	234	
60-69	19	64.5	1226	7	49	2401	133	931	
70-79	12	74.5	894	17	289	83521	204	3468	
80-89	9	84.5	760	27	729	531441	243	6561	
90-99	3	94.5	283	37	1369	1874169	111	4107	
Total	100		$\Sigma 5760$		6088	7419169	$\Sigma 1372$	$\Sigma 29940$	

3. Cálculo de la moda: $Mo = 49.5 + 10 (6/10) = 54.1$

4. Cálculo de la mediana: $Me = 50 + 10 \left(\frac{50 - 31}{26} \right) = 57.3$

5. Cálculo de la media: $\bar{X} = \frac{5760}{100} = 57.6$

6. Traza el polígono de frecuencias:



7. Cálculo del rango: $R = 94 - 13 = 81$

8. Cálculo de la desviación media: $DM = \frac{1372}{100} = 13.7$

9. Cálculo de la varianza: $S^2 = \frac{29940}{100} = 299.4 u^2$

10. Cálculo de la desviación estándar: $S = \sqrt{299.4} = 17.3$

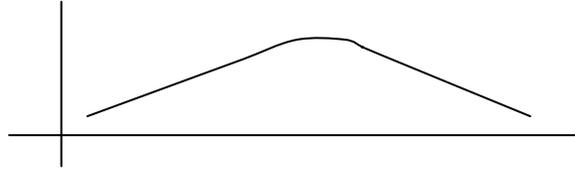
11. Cálculo del 2o. y 4o. momentos:

$M_2 = \frac{29940}{100} = 299.4$ $M_4 = \frac{7419169}{100} = 74191.69$

12. Cálculo del apuntamiento: $K = \frac{74191.69}{(299.4)^2} - 3 = -2.1$

13. Cálculo del sesgo: $SG = \frac{57.6 - 54.5}{17.4} = 0.2$

14. Tipo de gráfica:



BIBLIOGRAFÍA CONSULTADA

- FREUD** Williams, Perles. Estadística para la Administración. Editorial Prentis HALL Hispanoamérica, S.A., Ed. Méx. 1990.
- G. HOEL**, Paul. Estadística Elemental. Editorial CECSA. 2a. Impresión, México, 1976.
- KREYSZING**, Erwin. Introducción a la Estadística Matemática. Editorial Limusa, México, 1981.
- LEVIN**, Jack. Fundamentos de Estadística en la Investigación Social. Editorial Harla. 2a. Ed. México, 1979.
- LINCOLN** L. Chao. Introducción a la Estadística. Editorial CECSA. México, 1985.
- LINCOYAN** P., Govinden. Curso práctico de Estadística. Editorial McGraw Hill. México, 1985.
- LIZÁRRAGA**, G., Ignacio M. Estadística. Editorial McGraw Hill. México, 1986.
- MENDEHALL**, William. Introducción a la Probabilidad y la Estadística. Traductor Segami Carlos. Grupo Editorial Iberoamericana. México, 1989.
- MILLS** R., Richard. Estadística para Economía y Administración. Editorial McGraw Hill. Colombia, 1980.
- PERSEN**, Emmanuel. Teoría Moderna de Probabilidad y sus Aplicaciones. Editorial Limusa. México, 1973.
- STEVENSON**, William J. Estadística para Administración y Economía. Editorial Harla. México, 1981.
- SEYMOUR** Lipschutz, Ph. D. Teoría y problemas de probabilidad. Editorial McGraw Hill. México, 1973.
- TARO**, Yamane. Estadística. Editorial Harla, 3a. ed. México, 1979.
- TURNER**, J.C. Matemática Moderna Aplicada. Probabilidades, Estadística e Investigación Operativa. Alianza Editorial, S.A. Madrid, 1974.
- WILLOUGHBY**, Stephens. Probabilidad y Estadística. Publicaciones Culturales, S.A. México, D.F., 1983.