PRIMERA PARTE:

ANALISIS DE UNA VARIABLE

Por:

Ing. Roberto Piol Puppio
CIV: 32.290
SOITAVE: 260
E-Mail: rpiol@yahoo.com
Website: www.rpiol.com

CONTENIDO

I INTRODUCCIÓN

1.0 La estadística o métodos estadísticos, como a veces se llama, este desempeñando un importante papel ascendente en casi todas las facetas del progreso humano. Anteriormente, solo era aplicada a los asuntos del Estado, de donde deriva su nombre; pero hoy en día la influencia de la estadística se extiende a muchos campos de la ciencia e ingeniería.

2.0 El propósito de este curso es presentar una introducción a los principios generales de la estadística que sea de utilidad para todos, muy en especial en el campo de la valuación de los bienes muebles e inmuebles, tema el cuál es de nuestro especial interés.

3.0 Cada tema se inicia con una exposición de las definiciones pertinentes, teoremas y principios básicos con ejemplos ilustrativos.

4.0 <u>Definición de Estadística:</u> La estadística está ligada con los métodos científicos en la toma, organización, recopilación, presentado y análisis de datos, tanto para la deducción de conclusiones como para tomar decisiones razonables de acuerdo con tales análisis.

En un sentido más estricto, el término se utiliza para denotar los mismos datos o números, que se derivan de ellos, como por ejemplo, se habla de Estadísticas de Empleos, Estadísticas de Accidentes, etc.

5.0 <u>Población y Muestra. Estadística Descriptiva e Inductiva:</u> En un; colección de datos que atañen a las características de un grupo de individuos u objetos, tal como las alturas. y pesos de los estudiantes de un curso o el Precio Unitario de apartamentos en una urbanización; es imposible o poco práctico observar la totalidad de los individuos, sobre todo si estos son muchos o no hay registros completos. En lugar de examinar el grupo entero, llamado “Población o Universo”, se examina una pequeña parte del grupo llamada “Muestra”.

Una población puede ser Finita o Infinita, por ejemplo las ventas de Casas en una ciudad en un año determinado es Finita; mientras que la población formada por todos los posibles sucesos (Caras, Sellos) en tiradas sucesivas de una moneda es Infinita.

Si una muestra es representativa de una población, se pueden deducir importantes conclusiones acerca de esta a partir del análisis de la misma. La parte de la estadística que trata de las condiciones bajos las cuáles tales inferencias son válidas se llama Estadística Inductiva o Estadística Inferencial.

Al no poder estar absolutamente ciertos de la veracidad de tales inferencias, se ha de utilizar con frecuencia en estas condiciones el término “Probabilidad”.

La parte de la estadística que trata solamente de describir y analizar un grupo dado sin sacar conclusiones o inferencias de un grupo mayor se llama: “Estadística Descriptiva o Estadística Deductiva”.

¿Qué es la Estadística, Profesor? Preguntó la distraída alumna. Lentamente el viejo profesor dirigió mirada a su discípula acotando: - La Estadística señorita, es como una muchacha en bikini.. .

Todo el salón quedando perplejo ante la respuesta del venerable anciano prestaron su máxima atención al profesor que continuó diciendo: -...No lo muestra todo, sin embargo da una idea muy aproximada de la situación...

II DISTRIBUCIONES DE FRECUENCIA.

1.0 Levantamiento de los Datos

1.1 DEFINICIÓN: Es la obtención de una colección de datos que no han sido ordenados numéricamente.

1.2 ORDENACIÓN: Una ordenación es la colocación de datos numéricos tomados, en orden Creciente o Decreciente de magnitud.

La diferencia entre el Mayor y el Menor de los números se llama “Recorrido o Rango” de los Datos.

EJEMPLO

Si el área mayor de una muestra de 20 apartamentos es de 100 M2 y la menor de 43 M2
El RANGO o RECORRIDO será:
100 M2 – 43 M2 = 57 M2

1.3 LA TABLA DE DISTRIBUCION DE FRECUENCIA: Cuando se dispone de gran número de datos, es útil el distribuirlos en “Clases o Categorías” y determinar el número de individuos pertenecientes a cada Clase, que es lo que se denomina la “Frecuencia de Clase”.

Una ordenación tabular de los datos en Clases y con las Frecuencias correspondientes a cada una, se conoce como una “Distribución de Frecuencia” o “Tablas de Frecuencia”.

Los datos ordenados y resumidos en una “Tabla de Distribución de Frecuencia”, se suelen llamar “Datos Agrupados”.

EJEMPLO

La siguiente es una "Distribución de Frecuencia de Áreas" (expresadas en M2) de una muestra de Cien (100) apartamentos tomada en la Oficina de Registro Inmobiliario, correspondiente a la Parroquia "La Candelaria" de la ciudad de Caracas:

TABLA 1

AREA DE APARTAMENTOS	**NUMERO DE APARTAMENTOS**
60 - 62	5
63 - 65	18
66 - 68	42
69 - 71	27
72 - 74	8
TOTAL:	**100**

La Primera Clase o Categoría, comprende las áreas de apartamentos entre 60 y 62 M2 y viene indicada por el símbolo "60-62". Puesto que 5 apartamentos tienen un área perteneciente a esta clase, la correspondiente Frecuencia de esta "Clase" es 5.

La Segunda Clase o Categoría, comprende las áreas de apartamentos entre 63 y 65 M2 y viene indicada por el símbolo "63-65". Puesto que 18 apartamentos tienen un área perteneciente a esta clase, la correspondiente Frecuencia de esta "Clase" es 18.

1.4 INTERVALOS DE CLASES Y LIMITES DE CLASES: Un símbolo que define a una Clase, tal como "60-62" del ejemplo anterior, se conoce como "Intervalo de Clase".

Los números extremos 60 y 62, son los "Limites de Clases": El número menor "60" es el "Límite Inferior" de la Clase y el mayor "62" es el "Límite Superior" de la Clase.

Los términos. "Clase" e "Intervalo de Clase", se utilizan a menudo indistintamente, aunque el "Intervalo de Clase" es realmente un símbolo alfanumérico para identificar una "Clase".

Un "Intervalo de Clase", que al menos teóricamente, NO TIENE LIMITE (superior o inferior), se conoce como "Intervalo de Clase Abierto".

EJEMPLO

Al referirse a las áreas de grupos de apartamentos, el Intervalo de Clase: "> 65", es un Intervalo de Clase Abierto que agrupa a todos los apartamento cuyas áreas sean mayores a 65 M2.

1.5 LIMITES REALES DE CLASES: Si las áreas de apartamentos se registran con aproximación de M2, el Intervalo de Clase 60-62 teóricamente incluye todas las áreas desde 59.50 M2 a 62.50 M2. Estos números se conocen como Limites Reales de Clases o Limites Verdaderos de Clases.

El menor de ellos 59.50 es el “Limite Real Inferior” y el Mayor de ellos, 62.50 es el Limite Real Superior.

Prácticamente, Los Limites Reales de Clase se obtienen sumando al Limite Superior de un Intervalo de Clase el Límite Inferior del Intervalo de Clase contiguo superior y dividiendo entre 2.

A veces los Límites Reales de Clases se utilizan para simbolizar las Clases, por ejemplo las diferentes Clases de la 1ra. columna de la Tabla 1 podrían indicarse como:

EJEMPLO

TABLA 2

LIMITE REAL INFERIOR	LIMITE REAL SUPERIOR
59,5	62,5
62,5	65,5
65,5	68,5
68,5	71,5
71,5	74,5

Nótese, que la notación parece ser una ambigüedad, pues los Límites reales de Clase no coincidirían con las observaciones reales. Si una observación fuese 62.5 no sería posible discernir si pertenece al Intervalo de Clase “59.50 - 62.50” o al Intervalo de Clase “62.50 - 65.50. Por esto, es de especial cuidado predefinir el criterio: "Incluido a la derecha" .o viceversa, el cual deberá ser aplicado a todas las clases”.

1.6 TAMAÑO O ANCHO DE UN INTERVALO DE CLASE: El tamaño o ancho de un Intervalo de Clase es la diferencia entre los límites reales de clase que lo forman y se conoce como “Ancho de Clase”, “Tamaño de Clase” o “Longitu|d de Clase”.

Si todos los intervalos de Clase de una “Distribución de Frecuencia” tienen igual Ancho; esta Anchura común se representa como “C”.

En tal caso, C es igual a la diferencia entre Dos (2) sucesivos Límites de Clase.

EJEMPLO

Para los datos de la Primera Clase de la “TABLA 2” el Intervalo de Clase sería:

C = 62.5 - 59.5 = 3

El Profesor H.A. Sturges, estableció un algoritmo genérico para el cálculo del tamaño o ancho de un Intervalo de Clase en una serie.

Este modelo matemático es muy útil cuando se utilizan métodos automatizados de cálculo como pueden serlo un simple programa en Basic o una Hoja de Cálculo:

$$C = \frac{(RANGO)}{(1 + 3.322 * \log N)}$$

Donde N es el número total de los datos utilizados.

1.7 ELEMENTOS QUE CONFORMAN UNA TABLA DE DISTRIBUCION DE FRECUENCIAS: Además de los Límites Superior e Inferior de la Clase antes nombrados, los otros componentes de una “Tabla de Distribución de Frecuencia” son:

1.7.1 Marca de Clase (Xi):

Se define como el punto medio del Intervalo de Clase y se obtiene sumando los Límites Inferior y Superior de la Clase y a este resultado dividirlo entre Dos (2).

A la “Marca de Clase” (Xi), también se le conoce como “Punto Medio de la Clase” (o “Midpoint” en los textos anglosajones).

$$Xi = \frac{L_i + L_{i+1}}{2}$$

1.7.2 Frecuencia Absoluta (fi):

Se define como Frecuencia Absoluta o Frecuencia Ordinaria de una Clase (fi), como el número de datos de la serie en estudio, que aparece ente el Límite Inferior y el Límite Superior de dicha Clase.

1.7.3 Frecuencia Acumulada (Fi)

Se define como Frecuencia Acumulada de una Clase (Fi), como la sumatoria de las “Frecuencias Absolutas” (fi) de todas las “Clases” inferiores o igual a la “Clase” en estudio. En algunos textos también se conoce como “Frecuencia Acumulada Absoluta”.

1.7.4 Representación Gráfica de la Tabla de Distribución de Frecuencia

Generalmente, las representaciones gráficas de las “Tablas de Distribución de Frecuencia son:

a) El Histograma

b) El Polígono de Frecuencias

a) El Histograma, consiste en una serie de rectángulos o gráfico de barras, que tienen las siguientes características:

i.- Su base sobre el eje horizontal (eje X) con centros en las "Marcas de Clase" (Xi) y anchos iguales al ancho de los intervalos de clase (C).

ii.- Superficies proporcionales a las "Frecuencias Absolutas" de cada clase.

NOTA: Si los intervalos de clase tienen todos iguales tamaños, las alturas de los rectángulos son proporcionales a las frecuencias de clase y se acostumbra en tal caso a tomar las alturas numéricamente iguales a las frecuencias absolutas (fi) de cada clase.

b) El Polígono de Frecuencia, es un "gráfico de línea", trazado sobre las "Marcas de Clase" (Xi). Se construye, uniendo los puntos medios de los techos de los rectángulos en el "Histograma".

NOTA: El punto inicial de un "Polígono de Frecuencias" será el "Límite Inferior de la Serie" y el punto final, el "Límite Superior de la Serie".

EJEMPLO

Sea la siguiente Serie, que representa Once (11) datos de Valores Unitarios de Parcelas de Terreno en la Urbanización "Playa del Angel" de la ciudad de Pampatar (Isla Margarita, Venezuela):

TABLA 3

SERIE PRIMITIVA

DATOS	PRECIO UNITARIO Bs./M2
1	155
2	210
3	280
4	157
5	210
6	150
7	165
8	210
9	180
10	250
11	155

TABLA 4

SERIE ORDENADA

REFERENCIALES	PRECIO UNITARIO Bs./M2
1	150
2	155
3	155
4	157
5	165
6	180
7	210
8	210
9	210
10	250
11	280

LIMITE SUPERIOR DE LA SERIE: ***280***

LIMITE INFERIOR DE LA SERIE: ***150***

NUMERO DE DATOS (n): ***11***

RANGO (R): ***130***

ANCHO DE LA CLASE (C): ***29,15*** $C = \frac{(RANGO)}{(1+3.322*\log N)}$

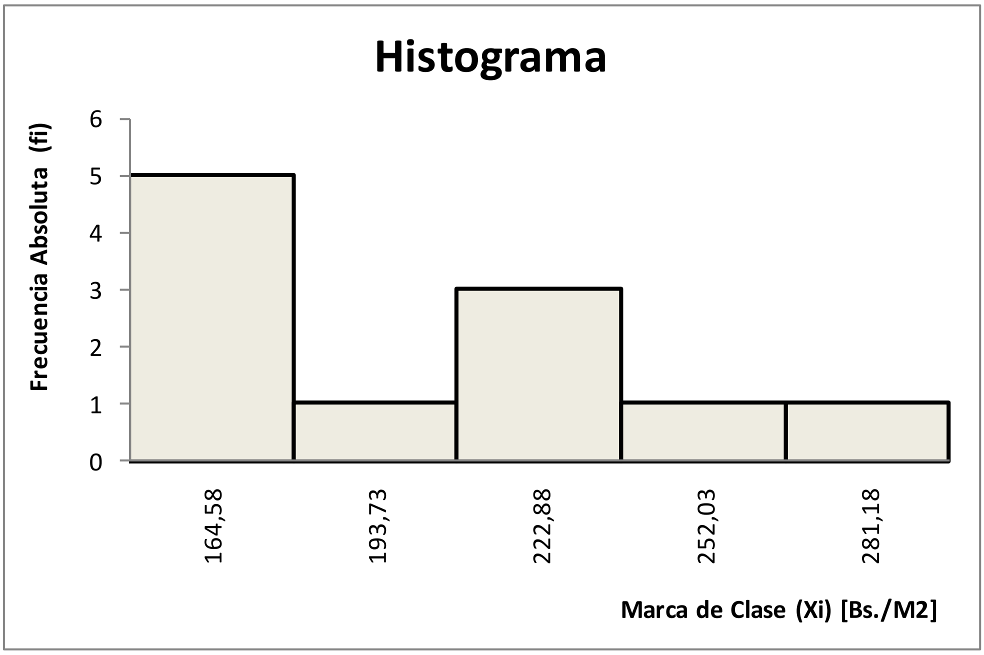
CONSTRUCCION DE LA TABLA DE DISTRIBUCION DE FRECUENCIAS

TABLA 5

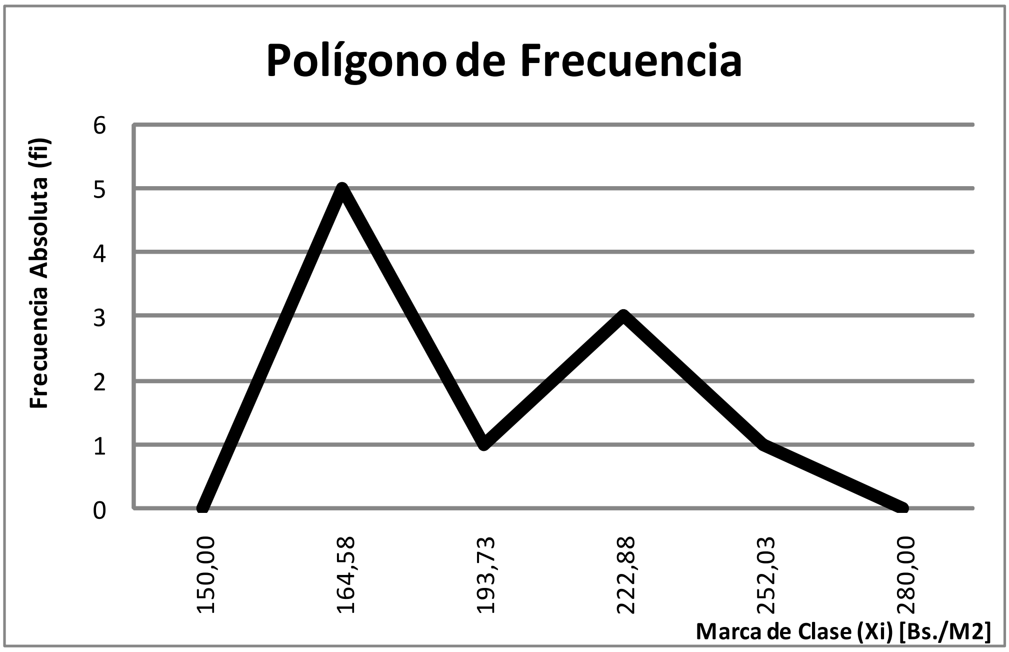
CLASE	LIMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i
1ra.	**150,00**	**179,15**	**164,58**	**5**	**5**
2da.	**179,15**	**208,30**	**193,73**	**1**	**6**
3ra.	**208,30**	**237,45**	**222,88**	**3**	**9**
4ta.	**237,45**	**266,60**	**252,03**	**1**	**10**
5ta.	**266,60**	**295,76**	**281,18**	**1**	**11**
			Σ	**11**	

REPRESENTACION GRAFICA DE LA TABLA DE DISTRIBUCION DE FRECUENCIA

A) HISTOGRAMA



B) POLIGONO DE FRECUENCIAS



III MEDIDAS DE TENDENCIA CENTRAL.

1.0 Tipos de Estimadores de la Tendencia Centra:

1.1 Las medidas o estimadores de la tendencia central, se dividen en dos grupos:

- PROMEDIOS MATEMATICOS
- PROMEDIO NO MATEMATICOS O POSICIONALES

1.2 Entre los estimadores de los Estimadores Matemático de la Tendencia Central para datos agrupados en Clases, los más comunes son:

- MEDIA ARITMETICA (También conocido como “Momento”)
- MEDIA GEOMETRICA
- MEDIA ARMONICA

1.3 Entre los estimadores de los Estimadores “No Matemático” o “Posicionales” de la Tendencia Central para datos agrupados en Clases, los más comunes son:

- MEDIANA
- MODO

2.0 La Media Aritmética: La media aritmética o media de un conjunto de n datos x1, x2, x3,…, xn ordenados en “Clases”, se define “Media Aritmética” ($\bar{x}$) como:

$$\bar{x} = \frac{f1 * X1 \; + \; f2 * X2 \; + \; \dots\dots..+fn * Xn}{f1 + f2 + \cdots + fn}$$

O sea:

$$\bar{x} = \sum \frac{fi * Xi}{n}$$

Donde: n es la frecuencia total, es decir, el número total de datos
fi es la “Frecuncia Absoluta” de cada Clase
Xi es la “Marca de Clase” de cada una de ellas

EJEMPLO

Siguiendo el ejemplo anterior, se procederá a calcular la "Media Aritmética"

TABLA 6

CLASE	LIMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	$X_i * f_i$
1ra.	150,00	179,15	164,58	5	5	822,88
2da.	179,15	208,30	193,73	1	6	193,73
3ra.	208,30	237,45	222,88	3	9	668,63
4ta.	237,45	266,60	252,03	1	10	252,03
5ta.	266,60	295,76	281,18	1	11	281,18
			Σ	11		2.218,45

El cálculo de la "Media Aritmética", será:

$$\bar{x} = \frac{2.218{,}45}{11} = 201{,}68\ [\frac{Bs}{M2}]$$

3.0 <u>La Media Geométrica de una serie de "n" datos x1, x2, x3,…, xn:</u> Es la raíz enésima del producto de las "Frecuencias Absolutas" y las "Marcas de Clases" de una "Tabla de Distribución de Frecuencias"; y su cálculo tiene sentido solo en series clasificadas con escalas de razones.

Para datos agrupados en "Clases", se define la "Media Geométrica" como:

$$\bar{G} = \sqrt[n]{\prod fi * Xi}$$

La "Media Geométrica", rara vez tiene uso en la valuación inmobiliaria.

4.0 <u>La Media Armónica de una serie de "n" datos x1, x2, x3,…, xn</u>: Es la reciproca de la media aritmética de los inversos de los mismos.

Para datos agrupados en "Clases", se define la "Media Armónica" como:

$$\overline{H} = \frac{n}{\sum \frac{1}{f_i * X_i}}$$

La "Media Armónica", rara vez tiene uso en la valuación inmobiliaria. Su uso fundamental se presenta en el caso de que existan dos o más modas en una serie, y para estos caso, se utiliza la versión de la "Media Armónica para Datos No Agrupados en Clases", que es la siguiente:

$$\overline{H} = \frac{n}{\sum \frac{1}{x_i}}$$

EJEMPLO

Suponiendo que el Diagrama de Dispersión siguiente, se tratase de una "Distribución Bimodal":

Donde el cálculo de las Dos (2) Modas de la Distribución sería:

Moda 1 =	160	Bs/M2
Moda 2 =	280	Bs/M2

Para poder calcular el Valor Unitario del inmueble representado, se utilizará el "Promedio Armónico", ya que un inmueble solo puede tener un solo valor:

$$\overline{H} = \frac{n}{\sum \frac{1}{x_i}} \implies \overline{H} = \frac{2}{\frac{1}{160} + \frac{1}{280}} \implies \overline{H} = 203{,}64\ [\frac{Bs.}{M^2}]$$

Nótese, que este sería un valor más sincero que el de simplemente aplicar un promedio a las Dos (2) modas, ya que el mismo tenderá al mayor de los valores.

5.0 <u>LA MEDIANA</u>: Se define a la Mediana de un conjunto de datos, como aquel valor que divide al conjunto en dos partes iguales, de forma que el número de valores mayor o igual a la mediana es igual al número de valores menores o igual a estos.

También puede definirse, como el término de la serie supera y a la vez es superado a lo sumo por la mitad de los datos.

5.1 Para datos No Agrupados en Clases, el cálculo de la Mediana es la siguiente:

A) Si la Serie tiene un número par de datos, como por ejemplo la siguiente serie:

1, 3, 4, 5, 6 y 9

$$Me = \frac{x_i + x_{i+1}}{2} \implies Me = \frac{4+5}{2} \implies Me = 4.5$$

B) Si la Serie tiene un número impar de datos, como por ejemplo la siguiente serie:

1, 2, 3, 4 y 5

$$Me = Toma\ el\ Valor\ Central\ de\ la\ Serie \implies Me = 3$$

5.2 Para datos Agrupados en Clases, el cálculo de la Mediana es el siguiente:

$$Me = Lm + \frac{\frac{n}{2} - Fm}{fm} * C$$

Donde:

Me:	*Mediana*
Lm:	*Límite Inferior de la Clase Medianal*
n:	*Número de Datos de la Serie*
Fm	*Frecuencia Acumulada anterior a la Clase Medianal*
fm:	*Frecuencia Absoluta de la Clase Medianal*
C:	*Ancho de la Clase*

NOTA: Se define como "Clase Medianal", a la Clase dónde la "Frecuencia Acumulada (Fi)" contiene por lo menos a la mitad de los datos (n/2).

EJEMPLO

Siguiendo el ejemplo anterior, se procederá a calcular la "Mediana"

TABLA 7

CLASE	LÍMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	
1ra.	150,00	179,15	164,58	5	5	
2da.	179,15	208,30	193,73	1	6	CLASE MEDIANAL
3ra.	208,30	237,45	222,88	3	9	
4ta.	237,45	266,60	252,03	1	10	
5ta.	266,60	295,76	281,18	1	11	
			Σ	11		

CLASE MEDIANAL:	2da. Clase
n:	11
Lm:	179,15
Fm:	5
fm:	1
C:	29,15

$$Me = Lm + \frac{\frac{n}{2} - Fm}{fm} * C \implies Me = 179{,}15 + \frac{\frac{11}{2} - 5}{1} * 29{,}15 \implies Me = 193{,}73$$

6.0 LA MODA: Se define a la Moda como el valor de mayor frecuencia en una serie de datos.

6.1 Para datos No Agrupados en Clases, el cálculo de la Moda es el siguiente:

$$Mo = Max\,\{\,x_i\,\}$$

EJEMPLO

Cálculo de la Moda para una Serie No Agrupada en Clases

TABLA 8

SERIE PRIMITIVA	SERIE ORDENADA
Bs/M2	Bs/M2
80	48
70	54
125	54
114	55
114	55
114	70
55	77
77	80
55	**114**
48	**114**
54	**114**
54	125

Mo = 114 Bs/M2

6.2 Para datos Agrupados en Clases, el cálculo de la Moda es el siguiente:

$$Mo = Li + \frac{D1}{D1 + D2} * C$$

Donde:

Mo = Moda

Li = Límite inferior de la Clase Modal

D1 - Diferencia entre la Frecuencia Ordinaria Absoluta de la Clase Modal y la Frecuencia Ordinaria Absoluta de la Clase Anterior.

D2 = Diferencia entre la Frecuencia Ordinaria Absoluta de la Clase Modal y la Frecuencia Ordinaria Absoluta de la Clase Siguiente.

C = Ancho de la Clase.

NOTA: Se define como “Clase Modal”, a la Clase de mayor “Frecuencia Absoluta (fi)”.

EJEMPLO

Siguiendo el ejemplo anterior, se procederá a calcular la "Moda"

TABLA 8

CLASE	LIMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	
1ra.	150,00	179,15	164,58	5	5	CLASE MODA
2da.	179,15	208,30	193,73	1	6	
3ra.	208,30	237,45	222,88	3	9	
4ta.	237,45	266,60	252,03	1	10	
5ta.	266,60	295,76	281,18	1	11	
			Σ	11		

D1 D2

CLASE MODAL:	1ra. Clase
Li:	150,00
D1:	5
D2:	4
C:	29,15

$$Mo = Li + \frac{D1}{D1 + D2} * C \quad \Rightarrow \quad Mo = 150{,}00 + \frac{(5-0)}{(5-0)+\ (5-1)} * 29{,}15 \quad \Rightarrow \quad Mo = 166{,}20\ [\frac{Bs.}{M2}]$$

III MEDIDAS DE DISPERSIÓN

1.0 El conocimiento de la tendencia central de una serie de datos no es suficiente para tener una idea clara de su distribución.

Por ejemplo: Si recogemos datos referenciales de parcelas de terreno ubicadas en Dos (2) urbanizaciones similares, con similar zonificación y uso, con similares áreas y distanciadas una de la otra en 10 Km y las Dos (2) urbanizaciones tienen todos los servicios y están en el mismo municipio…..

¿Se podría afirmar que también los Precios Unitarios de esas Dos (2) urbanizaciones, también serán similares?

Datos Referenciales de las Dos (2) Urbanizaciones:

URBANIZACION "A"	URBANIZACION "B"
Bs/M2	Bs/M2
60	120
80	280
140	310
210	380
250	400
1.200	450

Si calculamos el Promedio Aritmético de los Referenciales de cada una de las urbanizaciones se obtendrían los siguientes resultados:

PROMEDIO ARITMETICO

URBANIZACION "A"	URBANIZACION "B"
Bs/M2	Bs/M2
323,33	323,33

Ambos grupos tienen la misma media de 323,33 Bs/M2 pero las distribuciones no son parecidas.

En la "Urbanización B", los Precios Unitarios están más uniformemente distribuidos, lo que resulta en mayor concentración de los datos en torno a la media.

Mientras que en la "Urbanización A", el mayor de los Precios Unitarios difiere del "Promedio Aritmético":

1.200,00 – 323,33 = 876,67 Bs/M2

Mientras que en la "Urbanización B" solo difiere en:

450,00 – 323,33 = 126,67.Bs/M2.

Para un mejor conocimiento de la naturaleza de una distribución, además de una medida de la tendencia central, es necesaria otra clase de estimador, que exprese el grado de representatividad de la primera, una de ellas se refieren al agrupamiento de los datos en torno a la tendencia central. Estas nuevas medidas se denominan "Estadísticos de Dispersión" o "Dispersores.

Los "Estimadores de la Dispersión" son de dos clases: las absolutas, que se expresan en las unidades de la variable: Kg., M2, Número de hijos, Bs/M2, etc., y las relativas, que se expresan en números adimensionales, producto de una relación: tanto por ciento, tanto por uno, tanto por mil, etc.

Entre las Medidas Absolutas de Dispersión, las más usuales son:

- El Recorrido o Rango
- La Desviación Media
- La Desviación Mediana
- La Desviación Típica o Desviación Estándar
- La Varianza

Entre las Medidas Relativas de Dispersión, la más conocida es:

- El Coeficiente de Variación

MUY IMPORTANTE: *En el capítulo anterior, se estudió con detalle los estimadores del "Término Central" de una serie:*

  - *MEDIA ARITMETICA*
  - *MEDIANA*
  - *MODO*

Ahora bien, los Tres (3) estimadores anteriores son efectivamente medidores del "Término Central", los Tres (3) son precisos, los Tres (3) son matemáticamente correctos; pero solo uno de ellos será ciertamente el "Término Central" de esa serie.

La selección de uno de ellos no es una cuestión del azar, ni tampoco del "criterio" del profesional tasador, ni tampoco los mismos se prestan para una "suerte de ponderación", ya que los mismos son de conformación heterogénea.

Efectivamente; la "Media Aritmética" procede del resultado de una expresión matemática, mientras los otros dos (Mediana y Moda), dependen exclusivamente de la "Posición del Término Central" en la serie y solamente en esa serie.

Por lo tanto, para la determinación del exacto "Término Central" de una serie, se deberá primero prestar atención a los estimadores de la dispersión, antes de que a los propios medidores del término central de la misma.

De manera que, primero se seleccionará al estimador de la dispersión de la serie y una vez de haberlo comprobado, se seleccionará al "Término Central" que lo acompañe.

ESTIMADOR DE LA DISPERSION	TERMINO CENTRAL QUE LO ACOMPAÑA	RELACIONADO CON
Desviación Media D_x	Media Aritmética $\bar{x}$	La Simetría del Diagrama de Dispersión
Desviación Mediana D_m	Mediana M_e	Cantidad de datos que se agrupan alrededor del Término Central
Desviación Estándar σ	Media Aritmética $\bar{x}$	Distribución Normal del Diagrama de Dispersión

2.0 <u>RECORRIDO O RANGO:</u> El Recorrido de una Distribución, que también llamaremos rango, y denotaremos por "R" y es simplemente la distancia entre sus valores extremos.

Como medida de dispersión tiene poco interés porque sólo suministra información relativa en base al análisis de los valores extremos de una serie, no obstante, se emplea en muestras pequeñas (de 10 o menos datos) cuando sus extremos no están muy distanciados.

EJEMPLO

En el caso del análisis de los Precios Unitarios de las Dos (2) Urbanizaciones antes citadas, al calcular el Estadístico "RECORRIDO o RANGO", de cada una de ellas se obtendrá:

URBANIZACION "A"

Bs/M2
60
80
140
210
250
1.200

URBANIZACION "B"

Bs/M2
120
280
310
380
400
450

Maximo:	1.200	**Maximo:**	450
Minimo:	60	**Minimo:**	120
RANGO "A":	**1.140**	**RANGO "B":**	**330**

Demostrando evidentemente, que se trata de Dos (2) Distribuciones diferentes.

3.0 DESVIACIÓN MEDIA: La Desviación Media de una distribución es la media aritmética de los valores absolutos de las desviaciones de los datos respecto a la medida de la tendencia central.

De esta manera hay una desviación media respecto a la media aritmética, que se denotará por Dx y vendrá expresada como:

$$D_x = \sum \frac{|X_i - \bar{x}| * f_i}{n}$$

Donde:

Dx:	Desviación Media
Xi:	Marca de Clase
$\bar{x}$:	Media Aritmética
fi:	Frecuencia Absoluta
n:	Número de Datos

Es necesario tomar el valor absoluto de las diferencias porque, de lo contrario al ir precedidas del. signo correspondiente, las Desviaciones respecto a la Media, su suma podría ser nula, ya que esta es una propiedad de la Media Aritmética.

EJEMPLO

Siguiendo el mismo ejemplo que se ha desarrollado a lo largo de esta monografía , se procederá a calcular la "Desviación Media"

TABLA 9

CLASE	LIMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	$X_i * f_i$	$\lvert X_i - \bar{x}\rvert$	$\lvert X_i - \bar{x}\rvert * f_i$
1ra.	150,00	179,15	164,58	5	5	822,88	37,10	185,52
2da.	179,15	208,30	193,73	1	6	193,73	7,95	7,95
3ra.	208,30	237,45	222,88	3	9	668,63	21,20	63,59
4ta.	237,45	266,60	252,03	1	10	252,03	50,35	50,35
5ta.	266,60	295,76	281,18	1	11	281,18	79,50	79,50
			Σ	11		2.218,45		386,92

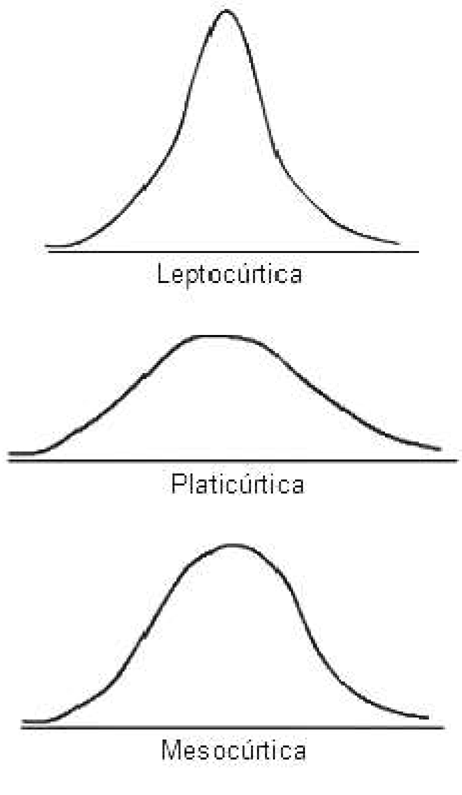
$$\bar{x} = \frac{2.218,45}{11} = 201,68\,[\frac{Bs}{M2}]$$

$$D_x = \sum \frac{|X_i - \bar{x}| * f_i}{n} \Rightarrow D_x = \frac{386,92}{11} \Rightarrow D_x = 35,17\,[\frac{Bs.}{M2}]$$

2.1 La Desviación Media como Indicador de la Simetría del Diagrama de Dispersión

El "Diagrama de Dispersión", es la representación gráfica de una "Tabla de Distribución de Frecuencia".

Debido que al asumir un "Ancho de Clase" dado, al construir una "Tabla de Distribución de Frecuencia"; el Diagrama de Dispersión podría variar en su representación gráfica, tal como se observa a continuación:



A veces, haciendo imposible determinar gráficamente, si la distribución es o no simétrica.

El dispersor "Desviación Media", está íntimamente ligado al grado de simetría de un "Diagrama de Dispersión"; por lo tanto a través de la "Desviación Media", se puede determinar en forma analítica, la simetría o no de la curva.

Un “diagrama de Dispersión”, puede ser simétrico o no de acuerdo a las dispersión de los propios datos con los que fue construida la “Tabla de Distribución de Frecuencia”; tal como se observan en los siguientes gráficos:

Diagrama de Dispersión Simétrico

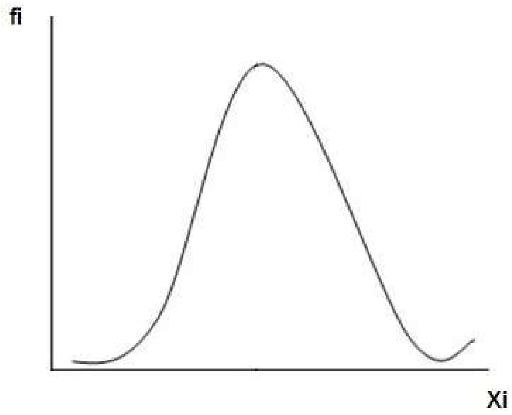
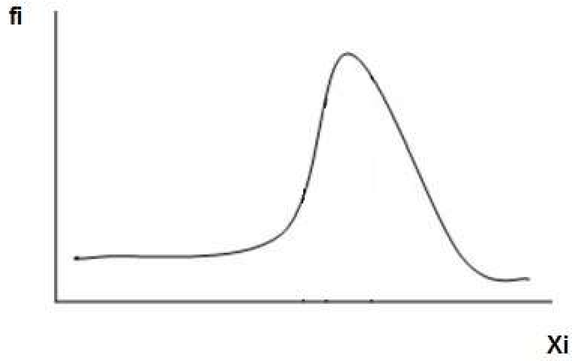


Diagrama de Dispersión No Simétrico



2.2 <u>El Concepto de Simetría</u>

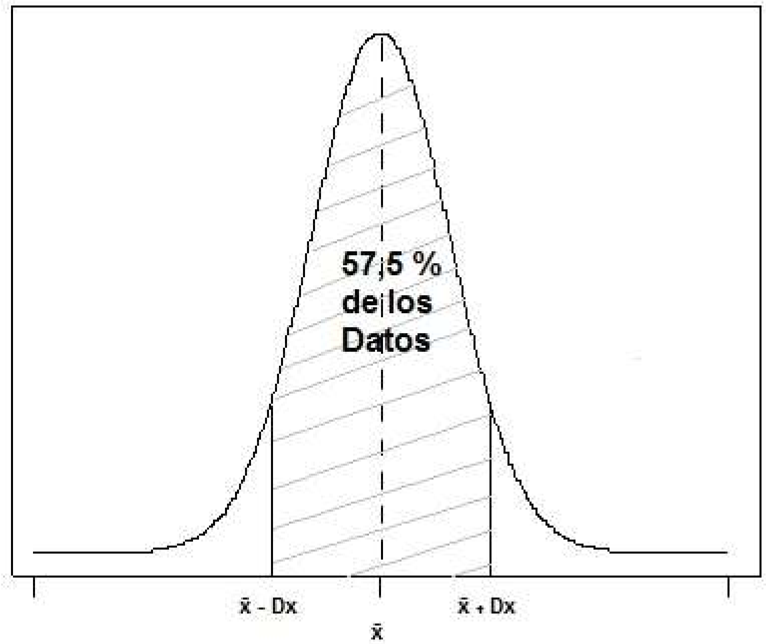
Para que un “Diagrama de Dispersión”, se le considere simétrico, debe cumplir que por lo menos el 57,5% de la data está comprendida debajo de la campana conformada por intervalo que comienza en $\bar{x} - Dx$ y finaliza en $\bar{x} + Dx$.

$$x_i \in [(\bar{x} - D_x) \ ; \ (\bar{x} + D_x)]$$

O sea:

$$|\bar{x} \pm D_x| \geq 57.5\%$$

Tal como se observa en la siguiente gráfica:



2.3 <u>Relación entre la “Media Aritmética” y el “Diagrama de Dispersión”</u>

Como se puede observar en la ilustración anterior, el estimador “Media Aritmética” ($\bar{x}$) es el único eje de simetría de esa curva.

Por básica geometría, se concluiría que: ***Para que la “Media Aritmética” ($\bar{x}$), sea el indicador del “Término Central de una Serie”; su “Diagrama de Dispersión” debe ser simétrico.***

En caso contrario, indicaría que la Distribución es Asimétrica y no se deberá usar a la “Media Aritmética” ($\bar{x}$), como medida de la Tendencia Central. Deberá usarse entonces la Mediana o la Moda como “Término Central” de esa Serie.

EJEMPLO

Retomando el ejemplo anterior, donde se ha determinado la "Media Aritmética" y la "Desviación Media" de la data.

$$\bar{x} = \frac{2.218,45}{11} = 201,68\ [\frac{Bs}{M2}]$$

$$D_x = \sum \frac{|X_i - \bar{x}| * f_i}{n} \Rightarrow D_x = \frac{386,92}{11} \Rightarrow D_x = 35,17\ [\frac{Bs.}{M2}]$$

Se procederá a determinar si el "Diagrama de Dispersión" es simétrico, y por lo consiguiente a conocer si la "Media Aritmética," es el Término Central de dicha Serie

Se procederá a calcular el Intervalo $|x_i \pm D_x| \geq 57.5\%$ *para determinar cuantos datos están incluidos en el mismo:*

El Intervalo:

$$\left.\begin{array}{l} x_i - D_x = 201,68 - 35,17 = 166,50 \\ \\ x_i + D_x = 201,68 + 35,17 = 236,85 \end{array}\right\}$$

Debe contener Por lo menos el 57.5% de la Data, para ser Considerada una "Curva Simétrica"

TABLA 10

SERIE ORDENADA

REFERENCIALES	PRECIO UNITARIO Bs./M2
1	150
2	155
3	155
4	157
5	165
6	**180**
7	**210**
8	**210**
9	**210**
10	250
11	280

Solamente Cuatro (4) Datos están dentro del Intervalo **36,4%**

Por lo tanto se concluye que:

a) Solo el 36,4 % de la Data está comprendida debajo de la campana conformada por intervalo que comienza en $\bar{x} - Dx$ *y finaliza en* $\bar{x} + Dx$.

b) Para que el "Diagrama de Dispersión" fuera simétrico, deberían estar comprendidos en el intervalo por lo menos Siete (7) datos; y solo están incluidos Cuatro (4)

c) Se define al "Diagrama de Dispersión" como "No Simétrico"

d) La "Media Aritmética" ($\bar{x}$ *), no será el "Término Central de la Serie"*

4.0 DESVIACIÓN MEDIANA: La Desviación Mediana de una distribución es la media aritmética de los valores absolutos de las desviaciones de los datos respecto a la medida de la tendencia central. En otras palabras; la Desviación Mediana" es la media aritmética de los valores de las observaciones con respecto a la Mediana

De esta manera, hay una desviación media respecto a la MEDIANA, que se denotará por Dm y vendrá expresada como:

$$D_m = \sum \frac{|X_i - Me| * f_i}{n}$$

El criterio del uso de la Mediana como "Término Central" de una serie de datos, radica en la magnitud de diferencias de cada dato respecto al "Término Central" de la misma.

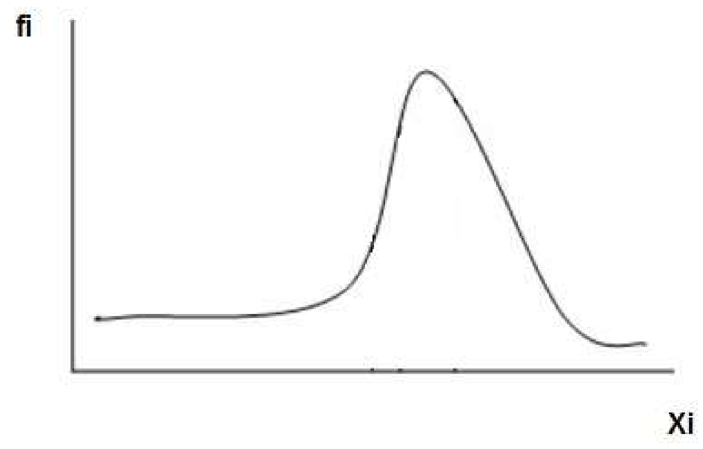
Pero el sólo uso de las diferencias no garantiza que se pueda medir discrepancias porque algunas (prácticamente la mitad) serán menores que la mediana, con diferencias negativas, y el resto mayores que la mediana, con diferencias positivas, y al sumar dichos valores habría compensaciones entre valores negativos y positivos.

Por lo tanto, una salida a esta dificultad es considerar el valor absoluto de las diferencias calculadas y promediarlos.

Esto conduce a la definición siguiente: Para que el "Término Central" de una serie de datos sea la "Mediana" se deberá cumplir que:

A) El “Diagrama de Dispersión” de la serie no sea simétrico

Diagrama de Dispersión No Simétrico

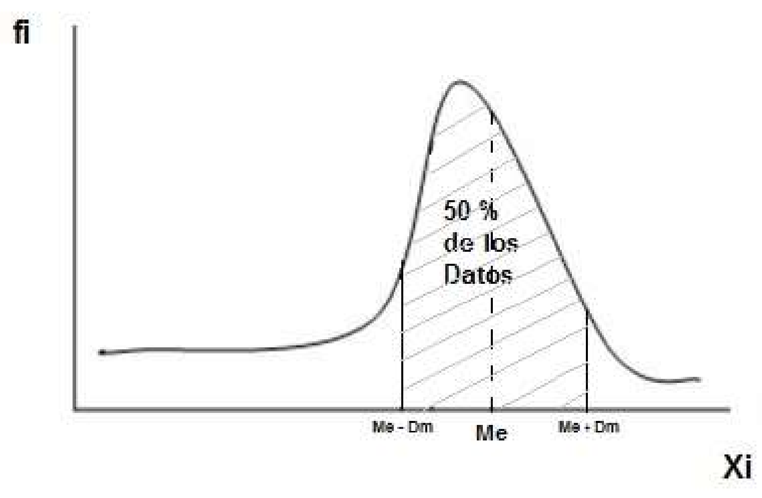


B) Que por lo menos el 50% de la data esté incluida en el intervalo:

$$xi \in [\ (Me - Dm)\ ,\ (Me + Dm)\ ]$$

O sea,

$$|M_e \pm D_m| \geq 50\%$$



EJEMPLO

Debido a que en el ejemplo anterior, donde se ha determinado la "Media Aritmética" No Es el "Término Central" de la misma; se procederá a determinar si la Mediana sería el "Término Central" de dicha serie

TABLA 11

CLASE	LIMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	$\lvert X_i - Me \rvert$	$\lvert X_i - Me \rvert * f_i$
1ra.	150,00	179,15	164,58	5	5	29,15	145,77
2da.	179,15	208,30	193,73	1	6	0,00	0,00
3ra.	208,30	237,45	222,88	3	9	29,15	87,44
4ta.	237,45	266,60	252,03	1	10	58,30	58,30
5ta.	266,60	295,76	281,18	1	11	87,45	87,45
			Σ	11			378,97

$$Me = Lm + \frac{\frac{n}{2} - Fm}{fm} * C \implies Me = 179,15 + \frac{\frac{11}{2} - 5}{1} * 29,15 \implies Me = 193,73$$

$$D_m = \sum \frac{|X_i - Me| * f_i}{n} \implies D_m = \sum \frac{378,97}{11} \implies D_m = 34,45 \ [\frac{Bs}{M2}]$$

Ya se ha demostrado que el "Diagrama de Dispersión" NO es simétrico, y por lo consiguiente la "Media Aritmética," NO es el "Término Central" de dicha Serie, por lo tanto se intentará conocer si la "Mediana" lo sería.

Se procederá a calcular el Intervalo $|Me \pm Dm| \geq 50\%$ *para determinar cuantos datos están incluidos en el mismo:*

$$\left.\begin{array}{l} Me - Dm = 193,73 - 34,45 = 159,28 \\ \\ Me + Dm = 193,73 + 34,45 = 228,18 \end{array}\right\}$$

Debe contener por lo menos el 50% de la Data, para que se considere a la Mediana como Término Central de la serie

TABLA 12
SERIE ORDENADA

REFERENCIALES	PRECIO UNITARIO Bs./M2
1	150
2	155
3	155
4	157
5	**165**
6	**180**
7	**210**
8	**210**
9	**210**
10	250
11	280

Solamente Cinco (5) Datos están dentro del Intervalo **45,5%**

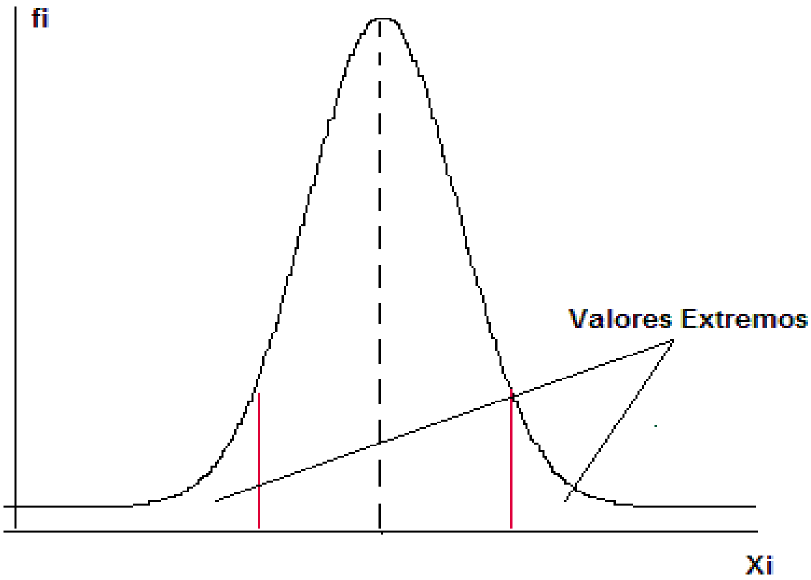
Por lo tanto se concluye que "La Mediana", tampoco es el "Término Central" de esta serie. Entonces "La Moda", será el "TérminoCentral" de la misma:

$$Mo = Li + \frac{D1}{D1 + D2} * C \implies Mo = 150{,}00 + \frac{(5-0)}{(5-0) + (5-1)} * 29{,}15 \implies Mo = 166{,}20 \ [\frac{Bs.}{M2}]$$

5.0 DESVIACIÓN TIPICA O ESTANDAR: La Desviación Estándar, indica sí una distribución simétrica es normal o no.

Esto es importante de conocer, porque si la distribución es "Normal", se podrán calcular cuales son los valores de la serie que se encuentran fuera de la campana.

Los valores que se encuentran fuera de la campana (o debajo de las colas), se consideran "VALORES EXTREMOS" y se definirán como aquellos valores que afectan notablemente a la "Media Aritmética".



Se define como “Valores Extremos”, aquellos datos de la serie que se encuentran fuera del intervalo:

$$[\,(\bar{x} - \sigma)\,;\,(\bar{x} + \sigma)\,]$$

Debido a que estos “Valores Extremos”, afectan a la “Media Aritmética”, los mismos pueden ser eliminados y el término central de la serie será la “Media Aritmética” de los datos que se encuentran dentro del intervalo arriba indicado.

5.1 Se define a la DESVIACIÓN ESTÁNDAR o DESVIACION TIPICA: Como la medida absoluta de dispersión de una Distribución Simétrica con respecto a la “Media Aritmética”.

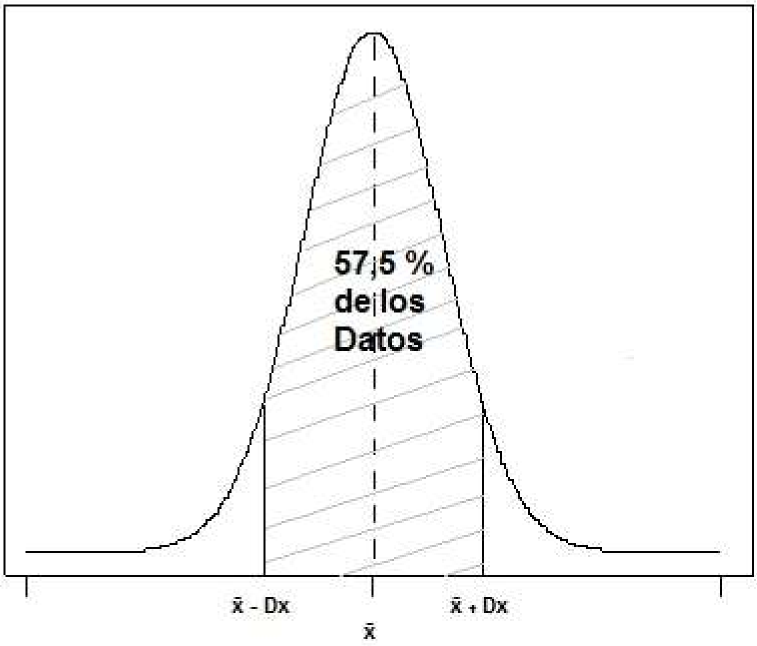
A mayor una mayor dispersión de la datar mayor será el valor de la Desviación Estándar.

Para el cálculo de la “Desviación Estándar”, se utiliza el siguiente modelo:

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{x})^2 * f_i}{n}}$$

5.2 Para que un “Diagrama de Dispersión”, sea considerado como una “Distribución Normal”, se deben cumplir las siguientes condiciones:

A) El “Diagrama de Dispersión” de la serie debe ser simétrico



Debiendo cumplir que por lo menos el 57,5% de la data está comprendida debajo de la campana, conformada por intervalo que comienza en $\bar{x} - Dx$ y finaliza en $\bar{x} + Dx$.

$$x_i \in [(\bar{x} - D_x) \; ; \; (\bar{x} + D_x)]$$

O sea:

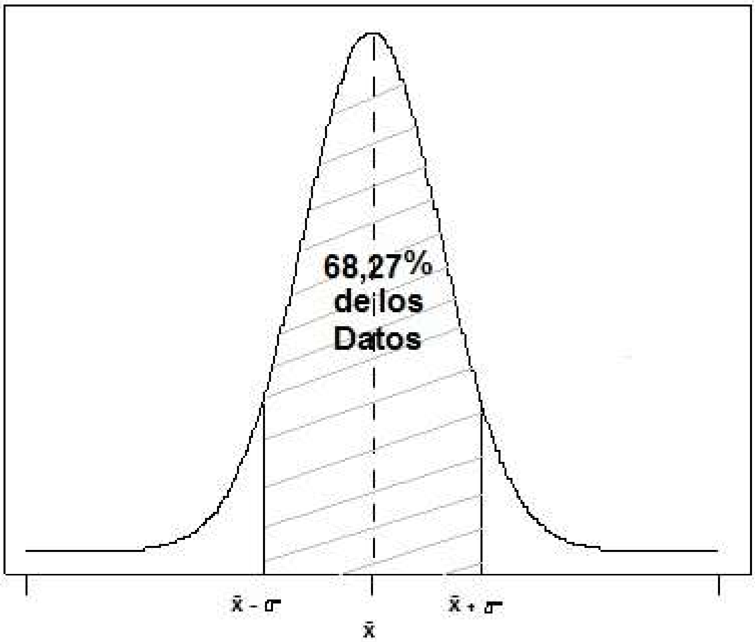
$$|\bar{x} \pm D_x| \geq 57.5\%$$

B) Que por lo menos el 68,27% de la data esté incluida en el intervalo:

$$x_i \in [(\bar{x} - \sigma) ; (\bar{x} + \sigma)]$$

O, sea:

$$|\bar{x} \pm \sigma| \geq 68{,}27\,\%$$



EJEMPLO

Se quiere determinar el Valor de Una Parcela de Terreno con un área de 250 M2 ubicada en las afueras de la ciudad de Maracaibo.

Los siguientes datos corresponden a todas las ventas de parcelas terreno, similares a la que se va a valuar:

TABLA 13

SERIE PRIMITIVA

REFEREN.	P.U.
1	250
2	400
3	700
4	375
5	280
6	200
7	485
8	430
9	375
10	280
11	130
12	580
13	475
14	455
15	430
16	300
17	250
18	90
19	375

SERIE ORDENADA

REFEREN.	P.U.
A	90
B	130
C	200
D	250
E	250
F	280
G	280
H	300
I	375
J	375
K	375
L	400
M	430
N	430
O	455
P	475
Q	485
S	580
R	700

LIMITE SUPERIOR DE LA SERIE:	*700*
LIMITE INFERIOR DE LA SERIE:	*90*
NUMERO DE DATOS (n):	*19*
RANGO (R):	*610*
ANCHO DE LA CLASE (C):	*116,23*

Se agrupan los datos en "Clases", de la siguiente manera:

TABLA 14

CLASE	LIMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	$X_i * f_i$	$\lvert X_i - \bar{x}\rvert$	$\lvert X_i - \bar{x}\rvert * f_i$	$(X_i - \bar{x})^2$	$(X_i - \bar{x})^2 * f_i$
1ra.	90,00	206,23	148,12	3	3	444,35	214,12	642,35	45.845,60	137.536,81
2da.	206,23	322,47	264,35	5	8	1.321,76	97,88	489,41	9.580,80	47.903,98
3ra.	322,47	438,70	380,59	6	14	2.283,51	18,35	110,12	336,82	2.020,95
4ta.	438,70	554,94	496,82	3	17	1.490,46	134,59	403,76	18.113,69	54.341,07
5ta.	554,94	671,17	613,05	1	18	613,05	250,82	250,82	62.911,39	62.911,39
6ta.	671,17	787,41	729,29	1	19	729,29	367,06	367,06	134.729,93	134.729,93
			$\sum$	19		6.882,43		2.263,51		439.444,13

$$\bar{x} = \sum \frac{fi * Xi}{n} \quad = \quad 362,23$$

$$D_{\bar{x}} = \sum \frac{|X_i - \bar{x}| * f_i}{n} \quad = \quad 119,13$$

$$|\bar{x} \pm D_{\bar{x}}| \geq 57.5\% \quad \left\{ \begin{array}{l} 243,10 \\ 481,37 \end{array} \right.$$

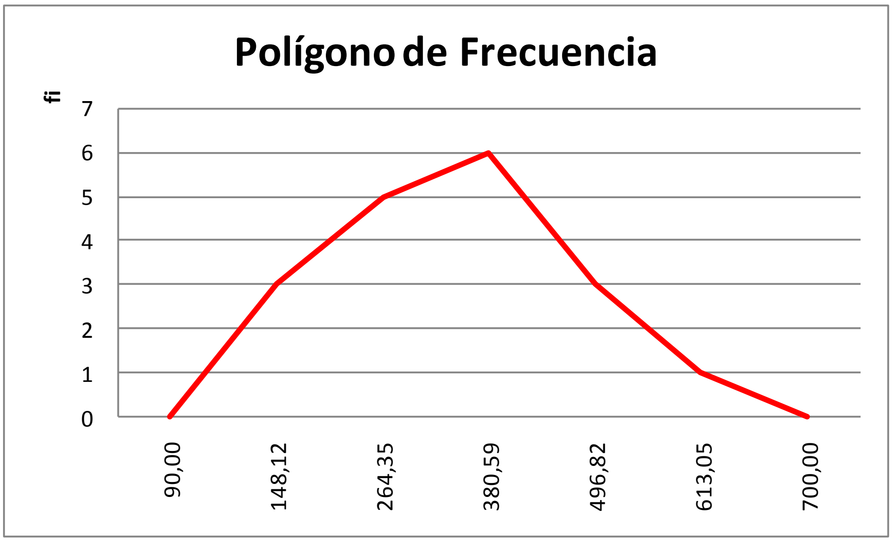
13 Datos 68,42% SIMETRICA

$$\sigma = \sqrt{\frac{\sum (X_i - \bar{x})^2 * f_i}{n}} = \quad 152,08$$

$$|\bar{x} \pm \sigma| \geq 68,27\% \quad \left\{ \begin{array}{l} 210,15 \\ 514,31 \end{array} \right.$$

14 Datos 73,68% NORMAL

Siendo su "Diagrama de Dispersión" el siguiente:



Debido a que el "Diagrama de Dispersión" es Simétrico y Normal; se podrán determinar los "Valores Extremos", eliminarlos y de esa forma depurar o afinar la Media Aritmética:

TABLA 15

REFEREN.	P.U.	
A	90	Valores
B	130	Extremos
C	200	
D	250	
E	250	
F	280	
G	280	
H	300	
I	375	
J	375	
K	375	
L	400	
M	430	
N	430	
O	455	
P	475	
Q	485	
S	680	Valores
R	700	Extremos

SERIE DEPURADA

REFEREN.	P.U.
D	250
E	250
F	280
G	280
H	300
I	375
J	375
K	375
L	400
M	430
N	430
O	455
P	475
Q	485

PARA LA NUEVA SERIE:

LIMITE SUPERIOR DE LA SERIE:	*485*
LIMITE INFERIOR DE LA SERIE:	*250*
NUMERO DE DATOS (n):	*14*
RANGO (R):	*235*
ANCHO DE LA CLASE (C):	*48,88*

$$C = \frac{(RANGO)}{(1 + 3.322 * \log N)}$$

Se calcula la "Media Aritmética" de la Serie depurada:

TABLA 16

CLASE	LIMITE INFERIOR L_i	LIMITE SUPERIOR $L_{(i+1)}$	MARCA DE CLASE X_i	FRECUENCIA ABSOLUTA f_i	FRECUENCIA ACUMULADA F_i	$X_i * f_i$
1ra.	250,00	298,88	274,44	4	4	1.097,77
2da.	298,88	347,77	323,32	1	5	323,32
3ra.	347,77	396,65	372,21	3	8	1.116,62
4ta.	396,65	445,53	421,09	3	11	1.263,27
5ta.	445,53	494,41	469,97	3	14	1.409,91
			Σ	14		5.210,89

$$\bar{x} = \sum \frac{fi * Xi}{n} \qquad = \qquad 372,21$$

Termino Central de la Serie: 372,21

Por lo tanto el Valor de la Parcela de Terreno a valuar será el siguiente:

AREA:	250,00 M2
TERMINO CENTRAL DE LA SERIE:	372,21 Bs./M2
VALOR DE LA PARCELA DE TERRENO:	93.051,62 Bs.

BIBLIOGRAFIA

ABAD, F.; VARGAS, M.; Estadística. (Volumen 1). Ed. Júfer. Granada. 1992.

ALONSO, F.; GARCÍA, P.; OLLERO, J.; Estadística para Ingenieros. Colección Escuelas. Ed. Colegio de I. C. C. P. Madrid. 1996.

CALOT, G.; Curso de Estadística Descriptiva. Ed. Paraninfo. Madrid. 1974.

FREEMAN, D.; PISANI, R.; PURVES, R; ADHIKARI, A.; Estadística. Editorial W.W. Noton & Company, 1991.

MILLER, I.R.; FREUND, J.E.; JOHNSON, R.; Probabilidad y Estadística para Ingenieros. Editorial Prentice-Hall Hispanoamericana, S. A., 1992.

HERMOSO, J; HERNÁNDEZ, A.; Curso básico de Estadística Descriptiva y Probabilidad. Ed. Némesis. Granada. 1997.

PIOL, R.; Estadística I. Curso de la Sociedad de Ingeniería de Tasación de Venezuela (SOITAVE). Caracas. 1999

SCHEFLER, W; BIOESTADÍSTICA. Fondo Educativo Interamericano, S.A. ,1981.

WALPOLE, R.; MYRES, R.H.; Probabilidad y Estadística. Mcgraw-Hill / Interamericana de Mexico, S.A. de C.V.