INTRODUCCIÓN

La demostración de las matemáticas como una de las ciencias exactas permitió la identificación de la estadística como la metodología de esa importante ciencia, de tal manera que en la actualidad se le considera como el método cuantitativo aplicable prácticamente a todas las ramas del saber científico.

Como una consecuencia de la generalización anterior, la ciencia económica se ha beneficiado con la aplicación de este valioso método numérico que facilitó la prueba de hipótesis a partir de las cuales se formularon algunas teorías, que una vez demostradas, permitieron arribar a la determinación de ciertas leyes que le dieron categoría de ciencia a la economía.

En efecto, la estadística como brazo operativo de las matemáticas se revela como una disciplina de gran ayuda para la configuración, análisis e interpretación de cualquiera de los fenómenos económicos conocidos o por identificar.

La importancia de esta disciplina en el análisis económico determinó la conveniencia de escribir esta obra. Se estima que resultara de gran ayuda al poder aplicar los principales métodos numéricos en el análisis de la economía y los negocios.

Aun cuando existe una amplia bibliografía sobre el tema, dentro de la cual, se deduce, que existen libros de excelente calidad en el país, creo que esta obra tiene cualidades que le dan originalidad y la ubican como un libro de texto de introducción a la estadística que ya hacia falta para llenar el hueco surgido por la inexistencia del análisis estadístico aplicado a la economía mexicana. En otras palabras, esta obra es original porque la presentación de su contenido se caracteriza por; primero, la exposición del método, sus características y alcance, fenómenos factibles de analizar y, finalmente, se aplica con el análisis e interpretación correspondientes. Con ello se hace una aportación en la nueva presentación del conocimiento, cuya transmisión resulta rápida y atractiva; en ocasiones se ratifican o rectifican algunas interpretaciones superficiales o radicales en cuanto a la bondad del método estadístico aplicado a la empresa y la economía en general.

Con base en lo anterior, la presentación de la obra se caracteriza por el desarrollo siguiente:

En el capitulo I se establece la relación entre la Estadística y la Economía, así como la función especifica que tiene la primera como instrumento de análisis de la segunda.

En el capitulo II se define y caracteriza a la ESTADÍSTICA DESCRIPTIVA al igual que se identifica la importancia que tiene en la tipificación de los fenómenos bajo estudio por el investigador.

Aquí se presentan los métodos necesarios para identificar, obtener, clasificar, procesar, analizar e interpretar la información de un fenómeno de interés para el investigador, hombre de negocios, estudiante o analista.

En los capítulos III a VIII se establece la relación que existe entre la información posible y la información probable de una variable, en el ámbito de la predicción que muchas veces es necesario hacer aún cuando halla riesgo e incertidumbre. Se hace una introducción a la teoría de la probabilidad, la identificación de los resultados posibles que genera un experimento realizado en determinadas condiciones, el tipo y la caracterización de esos resultados, mismos que se analizan en el marco de una distribución probabilística que sienta las bases para introducirnos al muestreo, que a su vez es el fundamento para realizar investigaciones de campo, con muestras probabilísticas, asi como para la estimación de parámetros y pruebas de hipótesis.

Finalmente, deseo expresar mi agradecimiento a los profesores J. Ekatherina Peregrina Guerreroy J. Alberto Reyes de la Rosa, sin cuya ayuda habría sido dificil actualizar los contenidos temàticos, mejorar sustancialmente los diseños gráficos, incorporar los anexos bibliografico y estadístico, así como su grabación electrónica para que las autoridades de la Facultad de Economía, con el apoyo que tradicionalmente me han brindado para la modernización de la cátedra y la difusión del conocimiento, promuevan estos materiales en discos compactos como libros electrónicos al igual que su inserción en la red de Intranet, pàgina web de la Facultad de Economía.

I. GENERALIDADES

Desde que el hombre tuvo conocimiento de su existencia, buscó expresar sus pensamientos, sus actos en forma tal que éstos le permitieran valorarse en su interrelación con el grupo social a que pertenecía.

La matemática surgió como una experiencia en la mente humana, ella refleja la voluntad activa, su objeto es precisar en forma sistematizada el mundo interno y externo en que se desenvuelve el individuo.

Observan los estudiosos de esta ciencia que sus elementos básicos son: Lógica e intuición, análisis y construcción, generalidad y particularidad. Advierten que diversas actividades han destacado sus enfoques diferentes, y que es únicamente el juego de estas fuerzas opuestas y la lucha por su síntesis lo que constituye su teoría, su utilidad y el supremo valor de la ciencia matemática.

Sin duda, todo el desarrollo matemático ha tenido sus raíces psicológicas en necesidades más o menos prácticas. Pero una vez en marcha, bajo la presión de las aplicaciones necesarias, dicho desarrollo gana impulso por sí mismo y trasciende los confines de una utilidad inmediata. Esta tendencia de la ciencia aplicada dio origen a la Estadística.

Con base en lo aquí dicho, podremos decir que la Teoría de la Estadística es una rama de las matemática pura, conocida con el nombre de Probabilidad.

La Economía es una ciencia cuyos límites rayan en la familiaridad con otras (Historia, Estadística, Sociología, etc.), por lo regular se encuentran interrelacionadas y se complementan unas con otras, para el enfoque teórico práctico, y las soluciones que se obtienen son resultado de sus acciones conjuntas.

Sigificado de estadística

Su significado[21] emana del vocablo "estado" y en general es sinónimo de ***datos.***

Lo anterior se debe basicamente a que cuando el hombre se organiza en sociedad y aparece el Estado, entonces es cuando el gobernante se empieza a preocupar por la recabación de datos relativos a la población y a la riqueza, para fines guerreros y de administración pública. Con el transcurso del tiempo la sociedad se fué desarrollando y con ella se fueron obteniendo datos de carácter más variado para uso general de los gobiernos, datos que vinieron a constituir lo que hoy conocemos como ESTADÍSTICA DESCRIPTIVA, en virtud de que sirven para describir las características de una variable determinada.

El origen de la teoría estadística inferencial, basada en muestras, puede atribuirse a las personas que estudiaron los juegos de azar, los seguros de vida y principalmente ciertas zonas de la experimentación biológica, llegando a un punto en su desarrollo, que a fin de continuar profundizando se hizo necesaria la aplicación de los métodos estadísticos. Consideran algunos autores que fueron precisamente los biólogos los que desarrollaron la teoría de esta ciencia.

A fin de dar una mayor referencia de la estadística, como preámbulo diré que los señores F. E. Croxton y D. J. Cowden la consideran no como ciencia sino como un método científico, es en esto último donde está de acuerdo el Profesor Gilberto Loyo quien en cierta ocasión me indicó que precisamente la estadística es un conjunto de métodos.

¿Pero ha existido siempre un criterio uniforme a través del tiempo sobre el concepto de estadística?

Es obvio que no, ya que lo que es ahora estadística, es completamente distinto a lo que se creía hace medio siglo, y aún hace un siglo.

Como se indicó previamente, es sinónimo de "dato", ya que por ejemplo cuando hacemos mención a las *estadísticas* de alumnos, de su matrícula, de su número, de sus calificaciones, el semestre que cursan.etcc, nos referimos a sus datos.

Sin embargo dicha acepción no corresponde, no es congruente con la función que desempeña como disciplina dentro del método científico, ni con las actividades que desempeñan en la actrualidad los expertos en estadística, puesto que no son meros "recolectores y tabuladores de datos numéricos[20] .

Pensando que es conveniente desglosar y explicar los diferentes criterios que han existido, expondré varias definiciones sobre la materia y se observará como han variado a travéz del tiempo.

1.1 DEFINICIONES DE ESTADÍSTICA[1]

Como todas las ciencias la Estadística ha sido considerada, por los teóricos dedicados a ella, según el grado de desarrollo en que se encuentra su teoría y su aplicación.

Al dar a conocer las definiciones que sobre ella existen, las estaremos interpretando como la expresión de lo que se consideró en una fecha dada; lo que era y para qué servía.

Enumerándolas conforme al tiempo en que fueron elaboradas, obtendremos el orden siguiente:

Achenwall (1748).- "La Estadística tiene por objeto el conocimiento de las cosas públicas, y enseña los medios para percibir las relaciones que hay entre ellas, siempre que sean dignas de notarse en cada República".

Achenwall (1749).-"La Estadística es la ciencia del Estado que se ocupa de la riqueza y contiene el conocimiento básico de las verdaderas posibilidades de una sociedad burguesa".

Achenwall (1749).-"La Estadística es la ciencia del Estado que se ocupa de determinar la riqueza individual".

Bielfield (1770).- "La Estadística es aquella rama del conocimiento político cuyo objeto de estudio es el poder real y relativo de los diversos estados modernos, el poder emanado de sus ventajas naturales, la industria y la civilización de sus habitantes y la sabiduría de sus gobiernos".

A. F. Luder (1792).- "La Estadística describe la situación de un estado en la actualidad o como era en una época determinada.

Meusel (1794).- "La Estadística es una exposición científica ordenada de la constitución y actual organización política de los Estados".

Sociedad Estadística de Londres (1838).- "La Estadística es la investigación y coordinación de aquellos hechos que son calculados para ilustrar las condiciones y posibilidades de una Sociedad".

D. E. Worl (1840).- "La finalidad de la Estadística consiste en la legitimidad de las diversas relaciones, en la detección de lo más posiblemente absoluto de los fenómenos relativos, de lo constante obtenido de la variable y en sacar de lo nuevo las leyes relativas".

Joe Fallati (1843).- "La noción de lo real es el punto medio de la Estadística, la realidad se encuentra, en parte, en los hechos, en parte en las leyes de los fenómenos".

Noreau de Jonneis (1847).- "La Estadística es la ciencia de los hechos sociales, expresados en términos numéricos".

Romelín (1863).-"La Estadística describe las características de la sociedad humana a base de observaciones metodológicas y de enumeraciones de fenómenos similares".

Levasser (1889).- "La Estadística es el estudio numérico de los hechos Sociales".

Arturo Bowley (1901).- "La Estadística es la ciencia de los promedios, la ciencia de los grandes números".

W. F. Willcox (1934).- "La Estadística es el estudio numérico de grupos o masas a través del estudio de las unidades que las componen, ya sea que estas unidades sean humanas o subhumanas, animadas o inanimadas".

McFarlane Mood (4) (1950).-"Estadística es la tecnología del método científico; proporciona instrumentos y técnicas para los investigadores, y estos instrumentos pueden ser de aplicación complementaria general y útiles en cualquier campo de la ciencia".

Significado profano de la Estadística.- Algunos la consideran como dato, otros dicen que comprende la recolección de grandes masas de datos y la presentación de éstos en tablas o gráficas; suele incluir también el cálculo de totales, promedios, porcentajes, etc. Este significado, según Mood tiene 30 años de retraso, porque estas operaciones más o menos rutinarias constituyen solamente parte inicial de la estadística de hoy.

Claudio Napoleoni (5) (1960).-La Estadística económica es aquella rama de la estadística aplicada que utiliza los métodos estadísticos para el estudio de los fenómenos económicos, en cuanto sean susceptibles de expresión numérica".

Wilburg Jimenez Castro.-(6) (1963).-La define como "método científico o ciencia de previsión de hechos futuros con base en el conocimiento de datos pasados y presentes."

Croxton y Cowden (7)(1965).-"Estadística es la recopilación, presentación, análisis e interpretación de los datos numéricos".

Ana María Flores (1983).-"La Estadística es la ciencia de medir variaciones".

Para obtener un resumen de las definiciones anteriores, puede decidirse que cada una de ellas refleja lo que se entendía por dicha ciencia, esto es, son viva expresión del campo en que se le aplicaba y expresan la tendencia a que se quería llegar una vez obtenidas sus conclusiones. En otras palabras, estas definiciones indican para quien se investigaba y qué es lo que interesaba saber (alimentación, riqueza, número de hombres disponibles para el trabajo, producción y otros).

Así vemos que las definiciones que abarcan toda la segunda mitad del siglo XVIII están enfocados a hacer de la Estadística una ciencia de información acorde con el industrialismo que ya se gestaba en Inglaterra, y a la consolidación de los Estados Europeos.

Con base en la doctrina del liberalismo y el surgimiento de nacionalismo en la Europa continental, se fortalece el Estado cuyo poder se encuentra en manos de esa clase social dinámica en sus orígenes llamada BURGUESIA, la que diera impulso en general al estudio de las ciencias entre las cuales contamos la Estadística.

Así pues el siglo XlX, es un período en que se fortalece la idea de aplicar los métodos estadísticos al análisis general de las ciencias sociales (Véase definiciones de Ravaseer y Romelin).

Nuevas definiciones (Véase definiciones de Arturo Bowley y W.F.Willcox) habían de formularse en torno al inicio de un siglo cuya primera mitad se caracteriza por cambios profundos e imprevistos. -Estos hechos hicieron una necesidad la existencia de datos estadísticos que sirvieron para la formulación de planes bélicos o científicos. Esto fue un primer paso hacia la programación adecuada porque se basaba en datos estadísticos.

Una vez terminadas las dos guerras mundiales, viene una paz que hace posible que se logre un gran avance en la técnica de producción, en donde una vez más surge la competencia entre las grandes corporaciones (monopolios), la que da lugar a la búsqueda de nuevos métodos estadísticos que garanticen la producción en masa y con el mínimo de defectivos (control estadístico de calidad). Hay otra característica importante en esta segunda mitad del siglo XX; la liberación de una gran cantidad de países que antes de la segunda guerra mundial eran "colonias", y que, ahora como países independientes elaboran sus planes de desarrollo con un conocimiento aceptable de la realidad en que se desenvuelven, gracias a la aplicación de los métodos estadísticos en el estudio de sus economías.

Considero que las definiciones de los señores Claudio Napoleoni, Mood y los autores que les siguen conforme al orden establecido, corresponden al significado que tiene actualmente la estadística. Por su permanencia en el tiempo, la definición de la maestra Ana María Flores es la más conveniente para el concepto general de la Estadística.

Como mi objetivo es presentar y exponer el uso de los Métodos Estadísticos, aplicados a la economía he considerado convenientemente ajustarme a la definición dada por el señor Claudio Napoleoni ya que ésta es la más idónea para los propósitos del economista.

1.2. DIFERENTES CLASES DE ESTADÍSTICAS

1. Estadística Inductiva o Inferencia Estadística. Incluye los métodos para obtener inferencias a partir de datos muestrales. Para ser específicos, inducción estadística incluye los métodos de generalización, estimación ó

predicción de las características de una población ó universo basados en una muestra.

2. Estadística Moderna. Se conoce como una gama de conocimientos y métodos de muestreo. Tiene una identificación con la estadística inductiva, al usar métodos probabilísticos. El uso de la probabilidad matemática de evaluar la confiabilidad de las conclusiones e inferencias basadas en los datos, es la contribución de la estadística al problema de la inferencia inductiva_ el problema de pasar de datos específicos a conclusiones generales _ . Por esta razón la teoría de la probabilidad desempeña un papel fubdamental en la teoría y aplicación de la estadística moderna [(20)].
3. Estadística Tradicional. Es la ciencia de los grandes números y de los grandes promedios.
4. Estadística de Variables. Es aquella cuyo universo de datos se expresa en magnitudes, ó sea que se puede medir.
5. Estadísticas de Atributos. Es la que se relaciona con fenómenos cuyo universo de datos está expresado de manera cualitativa ejemplo sexo, religión, filiación política, etc.
6. Estadística Descriptiva. Incluye los métodos de recopilación, organización presentación, análisis e interpretación de un grupo de datos, ya sean datos de muestreo o información completa sin ningún intento por hacer una predicción basada sobre los datos.
7. Estadística Paramétrica. Se sustenta en el supuesto de la normalidad de un fenómeno.
8. Estadística no Paramétrica. Es aquella que no se basa en el comportamiento normal del fenómeno bajo estudio.

II. LA ESTADÍSTICA COMO MÉTODO DE ANÁLISIS ECONÓMICO

Como se ha explicado en párrafos anteriores, la Economía es una ciencia compleja y amplia; para poder penetrar y caracterizar a los fenómenos que en ella se experimentan se vale de sus propios métodos y a la vez recurre a los pertenecientes a otras disciplinas.

La estadística es un conjunto de métodos que usamos para análisis de las variables económicas medibles directa e indirectamente. Conscientes de que la intuición no es una base general aceptable para estructurar el criterio sobre asuntos económicos, los investigadores de esta rama siguiendo el camino de los de las ciencias físicas, trabajan denodadamente con métodos de estudio que se basan en la observación y análisis de hechos, cuando las observaciones son de carácter cuantitativo, se requieren métodos apropiados para organizarlas e interpretarlas;

después la estadística nos dá los instrumentos necesarios para efectuar la combinación y análisis de tales observaciones.

Cabe mencionar que aunque los métodos estadísticos en general son prácticamente universales en su aplicación, siempre aparecen problemas especiales en cualquier campo de la investigación, esto es valioso para la economía, donde surgen dificultades peculiares y problemas característicos.

Los métodos que en cierto modo están dedicados a responder a estos requisitos, han experimentado amplio desarrollo, siendo confiable su aplicación a la ciencia económica; ya que como afirma J. M. Keynes "este método se basa en la observación cuantitativa de agregados, en el estudio de ellos y encaminado a descubrir uniformidades y constancias entre los elementos que los constituyen. Se funda en la observación porque considera directamente los hechos, y los reúne, selecciona y clasifica; se asienta en la observación cuantitativa porque sólo opera con hechos que son medibles: Y se ocupa en la observación de agregados, porque aún cuando para llegar al análisis de ellos hayan de pasar antes por el de los individuos o cosas que los forman, su verdadero campo de aplicación es el estudio de los conjuntos, no el de los elementos que lo forman".

Si recordamos que las leyes económicas son la expresión de sucesos que se repiten uniformemente en fenómenos globales. entonces la estadística es un buen instrumento para su análisis.

El Profesor Alonso Aguilar Monteverde afirma "las leyes económicas son estadísticas en virtud de que requieren de la repetición para poder configurarse". Con base a ello se puede afirmar que son hechos repetidos en sucesos masivos dentro del sistema económico.

Por consiguiente el método científico, que es la estadística, nos sirve para el análisis de los fenómenos económicos dentro de sus múltiples manifestaciones. Nos permite evaluar, hasta donde es posible la magnitud y el impacto que tiene el acto del hecho económico dentro de la sociedad, si se puede preveer o proyectar; en otras palabras, nos permite cuantificar las diferentes actividades que ejecutan los individuos dentro del sistema económico en que se desarrollan.

Así, si queremos saber la producción de bienes y servicios en un período determinado; si deseamos conocer las características de la población económicamente activa, su aportación al Producto Interno Bruto; de que manera contribuye a su formación, con tendencias al fortalecimiento del mercado interno, de consumo y de capitales; si nuestro interés es saber cuántos empleos deben crearse cada fin de año a fin de que el nivel de ingresos vaya acorde con el crecimiento demográfico, entonces la estadística es el método que ayuda a contestar estas interrogantes.

En conclusión si deseamos verificar cuál es la fuerza de trabajo en un momento dado, cuánta de ella es activa, cómo está distribuída y cuál es su posición con respecto a la tasa de desarrollo de un país, y muchas cosas más, no nos queda otro camino que recurrir a la ciencia estadística.

Este breve análisis nos permite observar la estrecha relación que hay entre ambas disciplinas y a la vez la importancia que tiene la estadística dentro de la economía. Esto no debe llevarnos al extremo de pensar que la economía vale por la estadística, o que trabaja a expensas de ella; hubo algunos estudios de la economía como el profesor Moore, quien dijera, "nada se sabe en tanto que no pueda medirse ".

Esta concepción es un error derivado posiblemente de la falta de profundidad en el conocimiento de la ciencia económica, ya que como afirma Marshall "semejante opinión es exagerada e inexacta".

Se acepta como complementaria más no como sustituto de la economía; para fundamentar su razonamiento establece que:

a) Todo estudio cuantitativo exige una selección y organización de los datos numéricos, o sea, la existencia previa de una teoría. Antes de cuantificar el consumo es necesario definirlo como una categoría económica particular.
b) Una serie de datos numéricos, un cuadro estadístico, con un estudio sólo cuantitativo, carece en sí de interés, si no se le somete a un trabajo cualitativo de interpretación.

A esta fundamentación debemos sumar la de Samuelson: "el razonamiento lógico es la clave del éxito para dominar los principios fundamentales (teoría económica), mientras que la ponderación sagaz de los datos empíricos es la llave para dominar las aplicaciones económicas".

Resumiendo diremos que la aplicación de la estadística en los fenómenos económicos es conveniente dentro de ciertos límites y tomando en cuenta las características del fenómeno en estudio, esto es, ver si es posible aplicarle determinado método que favorezca la obtención de resultados deseados, a la vez tomar en cuenta si es de interés social la realización del trabajo que con ella se logre.

II.1. APLICACIÓN DEL MÉTODO ESTADÍSTICO A LA ECONOMÍA

Con el objeto de describir los métodos estadísticos que se aplican con máxima frecuencia al análisis del sistema económico, se ha considerado necesario hacer una exposición en forma detallada de los mismos a fin de demostrar su uso, y con ello tratar de hacer clara su aplicación en el desarrollo del curso que se pretende dar al alumnado.

Para iniciar dicha exposición se ha juzgado conveniente comenzar a tratar sobre el significado de la terminología del mencionado método científico.

Empezaremos por definir lo que se entiende "población" en su acepción estadística.

Llamamos población o Universo a todo grupo o conjunto de elementos con ciertos atributos que lo caracterizan, como por ejemplo: tenemos los habitantes de la República Mexicana con un determinado grado de enseñanza primaria para el año 2002. Como se ve, este es un grupo con una característica, que es la instrucción primaria para el año 2002.

Las poblaciones pueden ser finitas o infinitas. Decimos que una población es finita cuando está compuesta por un número finito de elementos.
Ejemplo de ellos puede ser los habitantes de una localidad que tienen agua potable en sus hogares; será infinita una población, aquella cuyos términos sean inconmensurables, por ejemplo, la población de moscas en todo el mundo.

SERIE ESTADISTICA.- Es la descripción de la relación que se gesta entre dos variables, donde una de ellas es el tiempo, en función del cual se observa la evolución de la otra variable.

AÑO	PRODUCCIÓN	KILOS
1	10.334	"
2	9.756	"
3	9.339	"
4	8.357	"
5	7.364	"
6	7.401	"
7	6.531	"

Fuente: Investigación directa

Como puede observarse, la producción de oro del año 1 al 5 constituye una serie, ya que su volumen se modifica al variar el tiempo.

II.2. DISTRIBUCIÓN DE FRECUENCIAS

Sabemos que en la práctica los datos de un fenómeno bajo estudio se encuetran dispersos y es necesario organizarlos y distribuirlos a fin de que tengan utilidad y podamos analizarlos e interpretarlos.

Cuando hacemos esta operación de agrupamiento, decimos que estamos elaborando una distribución de frecuencias. Sea la edad en años de 30 instituciones bancarias establecidas en el país: 10, 7, 6, 5, 8, 9, 10, 11, 6, 7, 7, 12, 9, 6, 5, 9, 8, 13, 11, 12, 10, 9, 6, 7, 6, 6, 6, 6, 7, 8.

Cada número expresa la edad de cada banco. Cuando los datos están presentados en esta forma es difícil hacer observaciones; y se dice que estos datos no están agrupados, que son simples o que pertenecen a una serie simple.

Sin embargo nosotros podemos proceder a ordenarlos en forma ascendente o descendente conforme a sus respectivos valores y obtenemos la siguiente tabla:

13	10	9	7	6	6
12	10	9	7	6	6
12	10	8	7	6	6
11	9	8	7	6	5
11	9	8	7	6	5

Una vez ordenados los datos en forma decreciente, podemos hacer análisis y conocer los límites entre los cuales varían las edades de los bancos, ellas son: de 5 años son 2; de 6 años son 8; de 7 años son 5; etc.; lo que nos da la siguiente tabla:

EDADES		NÚMERO DE BANCOS
13	años	1
12	años	2
11	años	2
10	años	3
9	años	4

8	años	3
7	años	5
6	años	8
5	años	2
TOTAL		30

En virtud de que la suma nos da treinta, significa que fueron concentrados en forma correcta las edades de los bancos, ya que efectivamente son treinta los que tienen edad entre 5 y 13 años.

La columna cuyo encabezado dice "número de bancos", suele llamarse ***"Columna de frecuencia"***, por lo que una frecuencia se define como el número de veces que un términos se repite o existe en una serie; así podemos decir que los bancos cuya edad es de siete años, tiene una frecuencia de cinco, o lo que equivale a decir que hay cinco bancos cuya edad es de siete años.

Pero si estos datos deseamos clasificarlos en grupos o clases, (también se conocen como ***intervalos de clases***), hacemos el siguiente procedimiento:

1. Buscamos el valor más pequeño (mínimo) y el más grande (máximo), que son respectivamente el 5 y el 13.
2. Procedemos a calcular la amplitud.
Amplitud = valor máximo - valor mínimo
Substituyendo: amplitud = 13 - 5 = 8
3. Calculamos la amplitud de la clase o grupo

$$Amplitud\ \ de\ \ la\ \ clase = \frac{Amplitud}{Número\ \ de\ \ clases\ \ que\ \ se\ \ desean} = \frac{8}{4} = 2$$

El número cuatro indica que se agruparon los datos en cuatro clases o grupos, y el número dos expresa que cada clase tendrá una amplitud de dos unidades.

Con estos resultados procedemos a elaborar la siguiente tabla:

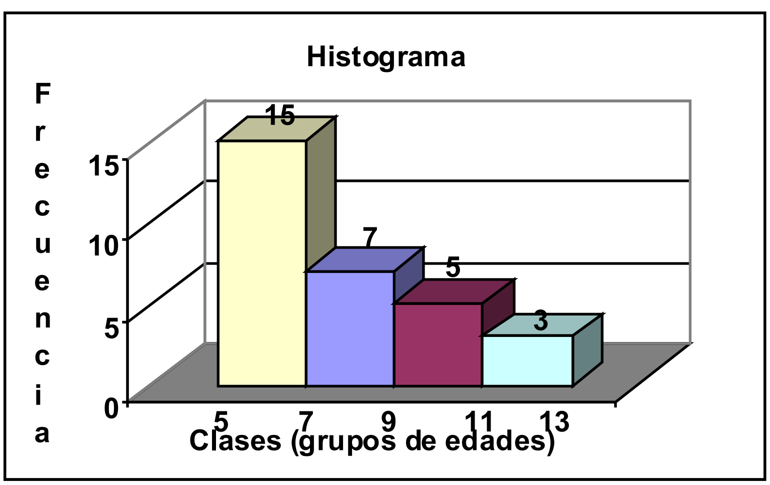
CLASES
(Grupos de edades)

	F
De 5 a 7	15
De más de 7 a 9	7
De más de 9 a 11	5
De más de 11 a 13	3
Total	30

F=Frecuencia

A esta forma de agrupar los datos le llamamos serie de clases y frecuencias, en donde cada **clase tiene su límite inferior y superior.**

Si queremos ver *gráficamente* cómo están distribuidas las edades de los bancos; basta hacer uso de los ejes cartesianos, usando el primer cuadrante y poniendo en el eje de las "Y" las frecuencias, en el de las " X " los grupos de edades, tendremos:

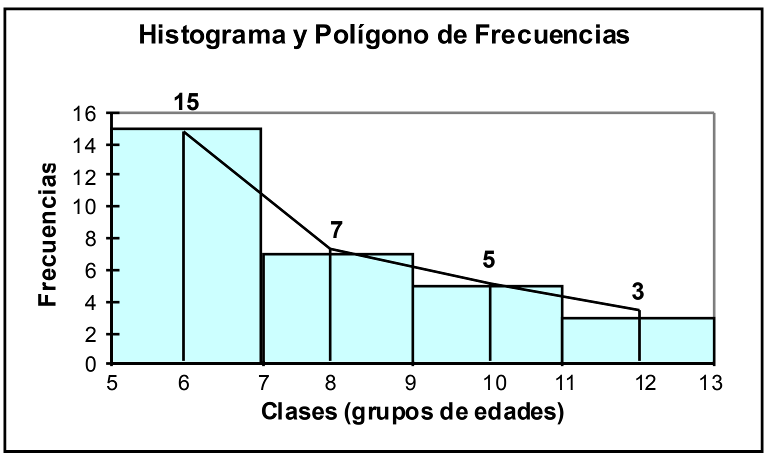


A esta representación gráfica suele llamársele "histograma". Si calculamos los puntos medios de los intervalos de clase, obtenemos una nueva tabla* .

Intervalos de clase	F	X_i
De 5 a 7	15	6
De más de 7 a 9	7	8
De más de 9 a 11	5	10
De más de 11 a 13	3	12
Total	30	

$_i$ = 1, 2, 3...n
X_i = marca de clase o punto medio.

Gráficamente tendremos



Marca de clases o puntos medios (X_i)

A esta segunda curva se le llama "polígono de frecuencia".

Es interesante decir que autores como el Dr. Raúl Rojas Soriano(19), Croxton & Cowden(2), entre otros, no cierran el Polígono de Frecuencias. Sin embargo, Yu

* Obsérvese que un punto medio es la suma de los límites inferior y superior de cada clase, la cual se divide entre dos. También se le llama"punto medio de la clase o marca de clase", que en esencia es el valor representativo de cada clase.

Lun Chou[21] comenta que: " Aunque el histograma es una presentación gráfica eficaz y vívida de distribuciones de frecuencias, el polígono no representa muy bien los datos básicos. La diferencia más notable del polígono es que las áreas situadas debajo de él generalmente no son proporcionales a las frecuencias. Un remedio es cerrar el polígono en la base prolongando ambos extremos de la curva hasta los puntos medios de dos clases hipotéticas situadas en los extremos de las distribuciones que tienen cero frecuencias."

Resumiendo podemos decir, en función de la forma en que están ordenados los datos, que hay tres tipos de series:

1. Serie simple
2. Serie de frecuencia
3. Serie de clases y frecuencias

Por lo que respecta a la representación gráfica de los datos, se puede afirmar que es más útil usar el polígono de frecuencias, porque se presta más para que puedan trazarse dos o más curvas en los mismos ejes para fines de comparación.

Con objeto de reafirmar la forma como se constituye una serie de clases y frecuencias, a continuación se presenta:

EJEMPLO # 2 CÁLCULO DE UNA SERIE DE CLASES Y FRECUENCIAS

Sean los datos:

96,500;	93,590;	93,268;	92,807;	90,196;
88,500;	87,500;	85,453;	84;925;	82,579;
80,813;	79,947;	79,504;	77,867;	74,635;
69,800;	69,145;	66,500;	66,317;	66,260;
60,310;	59,500;	57,486;	55,861;	53,500;
51,580;	46,963;	45,509;	44,148;	42,000;
41,558;	40,648;	39,729;	39,499;	39,000;

38,498;	37,719;	34,333;	33,055;	31,635;
80,852;	89,690;	28,710;	24,948;	20,500;
18,819;	14,500;	14,004;	13,681;	11,169.

PASOS A SEGUIR

1. Se identifican el valor más grande y el más pequeño, que son: 96,500 - 11,169
2. Se calcula la amplitud: AMPLITUD = VALOR MAXIMO - VALOR MINIMO

A = 96,500 - 11,169 = 85,331

3. Calculamos la amplitud de la clase o grupo.

$$Amplitud \ \ de \ \ la \ \ clase = \frac{85{,}331}{5} = 17{,}066$$

Intervalos de clase	F	PUNTOS MEDIOS PM
de 11,169 a 28,235	12	19.702
de más de 28,235 a 45,301	15	36.768
de más de 45,301 a 62,367	8	53.834
de más de 62,367 a 79,433	8	70.900
de más de 79,433 a 96,500	7	87.966
Total	50	

En la práctica el procedimiento anterior de agrupar los datos resulta ser el más usual; de hecho es el más conveniente porque parte del conocimiento del fenómeno y de los objetivos que se persiguen con la investigación. Con ello se evita el manejo de tablas extensas que hacen más complejo el análisis del fenómeno en estudio, o por el contrario tablas con unas cuantas clases que ocultan las características o detalles relevantes de la distribución.

Particularidades de una distribución de datos continuos

Como estos datos, a diferencia de los discretos o discontinuos, son fraccionables o divisibles, surge en problema del redondeo que se resuelve calculando las **fronteras** de clase, mismas que se definen como el punto preciso que separa una clase de otra, cuyo valor no está incluido en ninguna de las dos clases. Una frontera de clase es un punto medio entre el límite superior de una clase y el límite inferior de la que le sigue en la distribución de datos. Así en el siguiente ejemplo correspondientes a becas que reciben 50 estudiantes expresadas en pesos, tenemos:

Límites de clase en $	Fronteras de clase en $	Frecuencia	Punto medio de clase	Amplitud de la clase
160- 162	159.5- 162.5	1	161	3
163- 165	162.5- 165.5	2	164	3
166- 168	165.5-168.5	13	167	3
169-171	168.5-171.5	20	170	3
172-174	171.5-174.5	11	173	3
175-177	174.5-177.5	3	176	3
		50		

Observamos que a).- a diferencia de una frontera de clase, el límite superior de una clase no es el límite inferior de la siguiente clase; b).- la frecuencia o agrupamiento de datos se hace con base en los límites y no de las fronteras de clase. Así, digamos el valor 165.5, que es el valor de una frontera de clase, por el **criterio de redondeo,** se registra en 166, límite inferior de la siguiente clase; c).- la amplitud del intervalo de una clase se determina sustrayendo el valor de la frontera inferior del superior. En la clase 166-168, su amplitud = 168.5-165.5= 3. También se puede obtener sustrayendo el valor de su frontera inferior de la frontera inferior de la clase siguiente; o el límite inferior de la clase del límite inferior de la clase siguiente; también, el límite superior de la clase del límite superior de la clase que le sigue.

Es incorrecto obtener la amplitud del intervalo de la clase sustrayendo el límite inferior del límite superior.

Sin embargo, con el propósito de partir de una base matemática y no empírica en la construcción de las tablas de frecuencias, **H.A.Sturges** sugirió un procedimiento basado en la siguiente fórmula.

Donde: $i = \dfrac{Oscilación}{1 + 3.32 \quad \log(n)}$ i : Intervalo de clase

n: Número de términos de la distribución

Esto supone que una vez conocida la amplitud de la clase o intervalo de clase denotado por (i), la tabla de clases y frecuencias puede calcularse de inmediato. Como podrá intuirse una vez conocido su valor se determina automáticamente el número de grupos o clases de la distribución.

Aplicación del **Método de Sturges** al ejemplo # 2, redondeando cifras:

$$i = \frac{Oscilación}{1 + 3.32 \quad \log(n)}$$

OSCILACIÓN: Es la diferencia absoluta que existe entre el dato de menor valor y el de valor más elevado

$$i = \frac{85{,}331}{1 + 3.32 \quad \log(50)}$$ Oscilación = 96,500 - 11,169 = 85,331

$$i = \frac{85{,}331}{1 + 3.32 \quad \log(50)} = \frac{85{,}331}{6.6} = 12{,}929 \approx 13, \text{ redondeando a miles}$$

Serie de clases y frecuencias	x_i	F	PUNTOS MEDIOS PM
de	11 a 24	6	17.5
de más de	24 a 37	7	30.5
de más de	37 a 50	11	43.5
de más de	50 a 63	6	56.5
de más de	63 a 76	6	69.5
de más de	76 a 89	8	82.5
de más de	89 a 102	6	95.5
Total		50	

Acumulación de frecuencias u ojiva

	X_i	Frecuencia Acumulada (fa)
menos de	24	6
menos de	37	13
menos de	50	24
menos de	63	30
menos de	76	36
menos de	89	44
menos de	102	50

Al graficar esta serie se observa a través del histograma o del polígono de frecuencias la distribución que tienen los datos; algunas veces hay más de ellos a la izquierda, otras veces a la derecha.

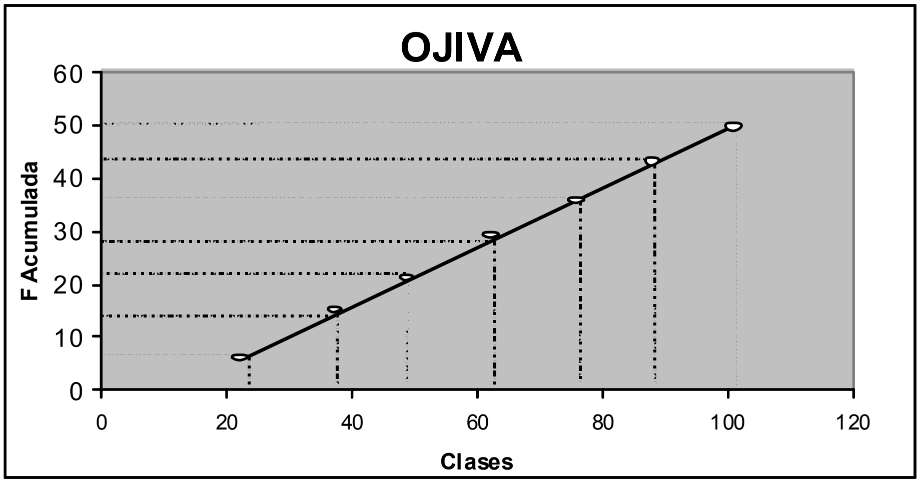
Derivado de lo anterior puede decirse que gráficamente la forma de las curvas o representaciones de una distribución de frecuencias, puede describirse de dos maneras: en términos de asimetría que se conoce como dispersión y en términos de su picudez que se conoce como Kurtosis.

En capítulos posteriores se ilustrará la metodología usada para medir tanto la asimetría como la Kurtosis de las curvas de frecuencias.

Resumiendo puede decirse que los datos pueden agruparse o clasificarse en:

1. Series simples
2. Series de frecuencias
3. Series de clases y frecuencias

En lo que se refiere a su representación gráfica, éstos pueden representarse a través del histograma, del polígono de frecuencias ó de la ojiva.



Hasta el momento hemos estudiado y usado clases cerradas para constituir series de clases y frecuencias; sin embargo esto no siempre es así, ya que el investigador puede decidir trabajar clases abiertas, es decir, que alguna clase no tenga un límite inferior o el superior; ejemplo:

SERIE DE CLASES ABIERTA

Clases			F	PM
1	a	2	5	1.5
3	a	4	0	3.5
5	a	6	2	5.5
7	a	9	5	8.0
más	de	9	2	no tiene

SERIE DE CLASES ABIERTAS Y AMPLITUD VARIABLE

La Secretaría de hacienda y Crédito Público para calcular el impuesto del año 2001, elaboró y dio a conocer a los contribuyentes la siguiente tabla:

Tarifa actualizada del impuesto correspondiente al ejercicio de 2001

Límite inferior $	Límite superior $	Cuota fija $	% para aplicarse sobre el excedente del límite inferior.
0.01	5,153.22	0.00	3.00
5,153.23	43,739.22	154.56	10.00
43,739.23	76,867.80	4,013.10	17.00
76,867.81	106,982.82	12,767.04	32.00
106,982.83	215,769.06	18,407.70	33.00
215,769.07	629,030.10	54,307.20	34.00
629,030.11	1,887, 090.18	194,815.74	35.00

1,887,090.19	2,516,120.46	635, 136.96	37.50
2,516,120.47	en adelante	871,023.24	40.00

En general, en la práctica se acostumbra ordenar o agrupar los datos en la forma anterior.

Ejemplos adicionales de Distribuciones o Series Estadísticas de Datos

Serie Simple

3	13	8	5	14	10	5	6	14	18	6
1	10	13	14	2	10	11	6	19	9	3
10	9	2	9	6	14	10	10	6	5	11
17	6	17	13	8	18	19	9	8	17	5
11	9	11	13	9	8					

Estos datos como aparecen en desorden no pueden analizarse ni interpretarse, para ello es recomendable ordenarlos en forma creciente, dando origen a una serie de frecuencias.

La series es de frecuencias como aparece en las dos primeras columnas, que puede convertirse en una serie de clases y frecuencias, con:

X	f(X)	
1	1	
2	2	5
3	2	
5	4	10
6	6	
8	4	
9	6	20
10	6	
11	4	
13	4	8

Ls - Li = 19 - 1

19 - 1 = 18

18/5 = 3.6

14	4	
17	3	
18	2	7
19	2	
	50	

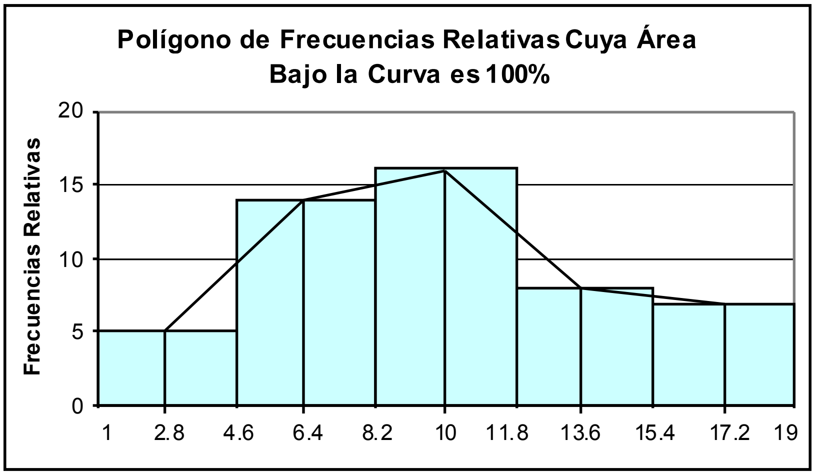
$$\sum fx = 50$$

Con ello se puede construir una:

SERIE DE CLASES Y FRECUENCIAS RELATIVAS
(GRUPOS PEQUEÑOS)

Clases	F	F Relativa (%)	PM
de 1 a 4.6	5	5/50*100 = 10	2.8
de más de 4.6 a 8.2	14	14/50*100 = 28	6.4
de más de 8.2 a 11.8	16	16/50*100 = 32	10.0
de más de 11.8 a 15.4	8	8/50*100 = 16	13.6
de más de 15.4 a 19	7	7/50*100 = 14	17.2
	50	50/50*100 = 100	

Las frecuencias relativas son muy importantes en economía por que permiten conocer la ponderación o importancia de los datos comprendidos en cada clase, además de que constituye **la base o introducción de la probabilidad en el análisis económico.**



Observación: La curva tiende a ser simétrica o normal aún con pocos datos. La identificación de esta característica es muy importante, ya que permite aplicar más métodos estadísticos al análisis de un mismo fenómeno económico, como se verá posteriormente, haciendo o corroborando que la estadística es un apoyo significativo para estudiar el comportamiento y caracterización de los fenómenos económicos.

II.3 ANÁLISIS DE LAS DISTRIBUCIONES DE FRECUENCIAS [2]

El ordenamiento o clasificación de las edades de los niños en una tabla de frecuencias así como su correspondiente representación gráfica permiten deducir ciertas características de la distribución, dentro de las cuales destacan las siguientes: Los términos (en este caso las edades) difieren, esto es, son diferentes y su grado de dispersión o variación quedó de manifiesto cuando se calculó la amplitud de la distribución. Por otra parte, al elaborar la serie de clases y frecuencias se conoció la clase con mayor número de frecuencias, es decir la que comprende el mayor número de edades de los niños.

Estas características son comunes a todas las distribuciones, no importa el área de investigación de donde provengan, siempre habrá una concentración máxima de términos, y éstas habrán de mostrar variaciones, algunas veces pequeñas y otras veces variaciones significativas.

Para la cuantificación de estas características y distinguir unas distribuciones de otras, existen ciertos métodos estadísticos que permiten analizar con una base científica el comportamiento de los términos en la distribución.

Los principales métodos usados para tal propósito son: Las de tendencia central para medir la acumulación o concentración alrededor de cierto valor, y las medidas de dispersión que sirven para medir la variación de los términos con respecto a una medida de la tenencia central.

II.3.1 MEDIDAS DE TENDENCIA CENTRAL

Sabiendo que un cuadro estadístico nos representa cómo están clasificados los elementos de una población y que a su representación gráfica le llamamos polígono de frecuencias, comprobamos que hay valores que se presentan más seguido y otros que ocurren con menos frecuencia, entonces los valores más característicos o de máxima frecuencia están por lo general en la parte central de las distribuciones. Es por ello que se dicen medidas de tendencia central, para identificar estos valores que pueden calcularse, con el fin de caracterizar la distribución de frecuencias.

Entre las medidas de tendencia central encontramos que las más usadas son la media aritmética, la media armónica, la media geométrica, la mediana y la moda.

II.3.1.1 LA MEDIA ARITMÉTICA ($\overline{X}$)[3]

Se define como un valor medio tal, que si a cada término se le da ese valor, resulta una suma igual a la de los valores de los términos de la sucesión dada.

Ejemplo (13) Sean los términos 1,5,2,9,7,8,5,3; como puede observarse son términos no agrupados.

Por definición, la media o promedio aritmético, es igual a 5 comprobando podemos obtener las siguientes expresiones:

$1+5+2+9+7+8+5+3 = 40$

$5+5+5+5+5+5+5+5 = 40$

Más generalmente:

Sea la sucesión cuyos términos son:
$x_1, x_2, x_3 ...... x_n$.

Designando con $\overline{X}$ a la media aritmética obtenemos:

$$\overline{x} + \overline{x} + \overline{x} + \overline{x} + \overline{x} + ... + \overline{x} = x_1 + x_2 + x_3 + ... x_n$$

Por lo tanto: para una serie simple:

$$n\overline{x} = \overline{x}_1 + \overline{x}_2 + ... + \overline{x}_n$$

.

Despejando la $\overline{x}$, queda de la siguiente forma: $\overline{x} = \dfrac{\sum_{i=1}^{n} xi}{n}$

$$\overline{x} = \frac{\overline{x}_1 + \overline{x}_2 + \overline{x}_3 + ... + \overline{x}_n}{n}$$ donde i= 1,2,3,...,n

Conforme a esto podemos definir también a la media aritmética de una sucesión como el valor medio que resulta sumando los términos de la sucesión y dividiéndolos entre el número de ellos. Con base a esta definición podemos obtener las medias aritméticas de una sucesión de frecuencias clasificadas.

Para una sucesión de frecuencias la media aritmética está dada por:

$$\overline{x} = \frac{\sum_{i=1}^{n} xiF}{N}$$

Para una serie de clases y frecuencias:

$$\boldsymbol{\overline{x} = \frac{\sum_{i=1}^{n} (P.M.)F}{N}}$$

donde P. M. nos indica el punto medio del intervalo de la clase.

PROPIEDADES DE LA MEDIA ARITMÉTICA.

1. La suma algebraica de las desviaciones de un conjunto de términos con respecto a su media aritmética es igual a cero.

Demostración

$$\sum (xi - \bar{x}) = \sum xi - \sum \bar{x} = \sum xi - n\bar{x} = n\bar{x} - n\bar{x} = 0$$

sean los términos 8, 3, 5, 10, 12.

Calculando su media $\bar{X} = \frac{\sum xi}{n} = \frac{38}{5} = 7.6$

Desviaciones = (8-7.6) + (3-7.6) + (5-7.6) + (10-7.6) + (12-7.6)= 0
= 0.4-4.6-2.6+2.4+4.4= 0

2. La suma de los cuadrados de las desviaciones de un conjunto de términos x_i de cualquier número A, es un mínimo si y sólo si

$$A = \bar{x}.$$

Demostración.

$$\sum (xi - A)^2 = Q(X_1, X_2, \ldots X_n; A)$$

Tomando la derivada parcial de Q respecto a A se tiene:

$$\frac{\partial Q}{\partial A} = -2\sum [(xi - A)] = -2\left[\sum xi - nA\right];$$

$$\sum xi - nA = 0 \rightarrow A = \frac{1}{n}\sum xi = \bar{x}$$

Esta expresión tiene un mínimo $\Leftrightarrow \boldsymbol{A = \bar{x}}$

Ejemplo: sean los términos 3, 4, 6, 8, 7.

$$\bar{x} = \frac{28}{5} \therefore \bar{x} = 5.6$$

Cuando $A < \bar{x}$ A=5

Tenemos $(3-5)^2 + (4-5)^2 + (6-5)^2 + (8-5)^2 + (7-5)^2$

= 4 + 1 + 1 + 9 + 4 = 19

Cuando $A > \bar{x}$ A = 6

Tenemos $(3 - 6)^2 + (4 - 6)^2 + (6 - 6)^2 + (7 - 6)^2$

= 9 + 4 + 0+ 1 = 18

Cuando $A = \bar{x}$ A = 5.6

$(3-5.6)^2 + (4-5.6)^2 + (6-5.6)^2 + (8-5.6)^2 + (7-5.6)^2 =$

$(-2.6)^2 + (-1.6)^2 + (0.4)^2 + (2.4)^2 + (1.4)^2 =$

6.76 + 2.56 + 0.16 + 5.76 + 1.96 = 17.20 l.q.q.d.

3. El promedio aritmético por el número de términos es igual a la suma de los valores de los términos.

Demostración.

$$\text{Como } \bar{x} = \frac{1}{n}\sum xi \rightarrow n\bar{x} = \sum xi$$

$$n\bar{x} = x_1 + x_2 + \ldots + x_n$$

Si los términos de la serie son: 3, 5, 6, 4, 2.

$\bar{x} = \frac{20}{5} = 4$ n=5

Tendremos: 4(5) = 3 + 5 +6 + 4 + 2

20 = 20 1.q.q.d.

Ventajas de la media aritmética.

1. Su cálculo es sencillo.
2. Con su valor y el número de términos se puede calcular la suma de los valores de los términos.
3. Puede calcularse conociendo solamente la suma y el número de los términos de la serie.

Desventajas de la media aritmética.

1. Hay ocasiones en que un término que mide una modalidad anormal del fenómeno influye en el promedio y puede ser que este no refleje la realidad.

2. Puede darse el caso de que el valor de la media para dos más series distintas sea el mismo.

Comparación de los métodos de calculo abreviado y largo de la media.

Serie Simple

Con el método abreviado

$$\bar{x} = A + \frac{\sum (xi - A)}{N}$$

Con el método largo

$$\bar{x} = \frac{\sum xi}{N}$$

Donde: A = media supuesta arbitrariamente
xi-A = desviación de cada término respecto a A
N = número de términos
xi = término i-ésimo

Ejemplo numérico de una serie simple

xi	xi-A
3	-1
5	1
6	2
7	5
21	5

Si A=4 entonces $\bar{x} = A + \frac{\sum (xi - A)}{N}$

$$= 4 + \frac{5}{4}$$

$$= 4 + 1.25$$

$$= 5.25$$

Método directo o largo: $\bar{x} = \frac{\sum xi}{N} = \frac{21}{4} = 5.25 = \bar{x}$

Serie de Frecuencias

Xi	Fi	Xi-A	Fi(Xi-A)	Xifi
3	8	0	0	24
5	6	2	12	30
6	2	3	6	12
7	3	4	12	21
	19	9	30	87

Método abreviado

$$\overline{X} = A + \frac{\sum Fi(Xi - A)}{\sum fi}$$

Método directo o largo

$$\bar{x} = \frac{\sum Xifi}{\sum fi}$$

Con A = 3

$$\overline{X} = 3 + \frac{30}{19} = 4.8$$

$$\overline{X} = \frac{87}{19} = 4.57$$

Serie de Clases y Frecuencias

Clases	Fi	PMi	PMi-A	Fi (PMi-A)	FiPMi
0 y menos de 2	1	1	-4	-4	1
2 y menos de 4	3	3	-2	-6	9
4 y menos de 6	5	5	0	0	25
6 y menos de 8	7	7	2	14	49
	16	16	-4	4	84

Método directo ó largo

$$\overline{X} = A + \frac{\sum Fi(P * Mi - A)}{\sum fi}$$

$$\bar{x} = \frac{\sum (fi * PMi)}{\sum fi}$$

Con A = 5

$$\overline{X} = 5 + \frac{4}{16} = 5.25$$

$$\overline{X} = \frac{84}{16} = 5.25$$

II.3.1.2 MEDIANA (Md)

Se define como el valor central que divide una distribución de modo que quede de cada lado de ella un número igual de términos ordenados, éstos en orden creciente ó decreciente.

Para obtener la Md. a partir de una serie de clases y frecuencias se usa la fórmula:

$$Md = Li + \frac{\frac{N}{2} - C}{Fi}(i)$$

donde:

Md.=Mediana

Li = Límite inferior de la clase que contiene a la mediana;
N = Número de términos ó suma de las frecuencias
C = Frecuencia acumulada de la clase anterior a la que contiene la mediana.
Fi = Frecuencia de la clase que tiene a la mediana.
i = Amplitud del intervalo de la clase que contiene la Md.

En el caso de una serie simple, si los datos son:
1, 2, 3, 4, 5

Md. = 3

En el caso de una serie de frecuencias

Xi	Fi	Fi

		Acumulada
2	5	5
3	6	11
4	3	14
5	3	17
Total	17	

No. de orden o terminos = $\frac{\sum fi + 1}{2}$

No= $\frac{17+1}{2} = \frac{18}{2} = 9$

Indicando que el termino noveno es el que contiene a la mediana. Para conocer el valor de la mediana se acumularán las frecuencias hasta encontrar el número 9 que corresponde al término 3, por consiguiente la Md = 3

Lo anterior se puede comprobar abriendo la serie de frecuencias en una serie simple, esto es:

2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5

Por definición Md. = 3, puesto que es el término que parte o divide a una serie o distribución en dos partes iguales.

EJEMPLO SERIE DE CLASES Y FRECUENCIAS

CLASES	Pm	F	Pm*F	F-Media	Fac.
1 y menos 3	2	5	10	0.07	5
3 y menos 5	4	2	8	-3.93	7
5 y menos 7	6	4	24	-1.93	11
7 y menos 9	8	4	32	-1.93	15

Total	15	74

$$\bar{x} = \frac{74}{15} = 4.93$$

Así, No= $\frac{\sum F + 1}{2} = \frac{15+1}{2} = 8$

Acumulado Fi se observa que Md. por consiguiente será

$$Md. = 5 + \frac{\frac{15}{2} - 7}{4}(2)$$

$$Md. = 5 + \frac{7.5 - 7}{4}(2)$$

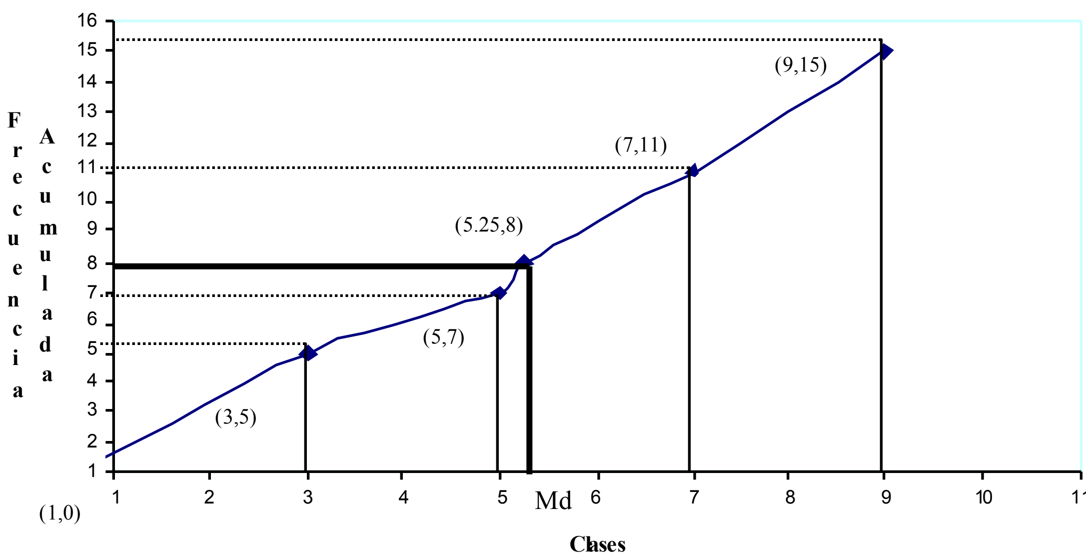
Md = 5 + 0.25(2) = 5 + .25 = 5.25
Md = 5.25

Obtención de la Md. por el método gráfico.

Se obtiene a partir de la ojiva, ordenando los datos en base a "menor de" "mayor de" o "más de". Si ordenamos los datos con base a "menor de" obtenemos lo siguiente:

Clase	F Acumulado
Menos de 1	0
Menos de 3	5
Menos de 5	7
Menos de 7	11
Menos de 9	15

Ojiva



Md = 5.25

8 es el término que contiene a la mediana, en torno al cual bajamos una línea que cruza el eje de las abscisas en el punto Md = 5.25.

II.3.1.3 MODA (Mo)

Se define como el valor de máxima frecuencia o dicho en otras palabras, la moda es el término que más aparece o se repite en una distribución.
Ejemplo:
En una serie Simple :
Datos: 1, 2, 2, 2, 3, 4.
Mo = 2 porque es el término que más se repite.

Serie de frecuencias

X	F
10	6
11	40
12	2
13	1

Mo=11 porque es el término que más aparece, en este caso 40 veces.

Serie de Clase y Frecuencias :

CLASES	F
1 y menos de 3	5
3 y menos de 5	2
5 y menos de 7	4
7 y menos de 9	4

Total	15

Partiendo de la definición de Mo, se observa que Mo está contenida en la clase "1 y menos de 3". Su valor exacto se determina con la fórmula de interpolación.

$$\boldsymbol{Mo.} = \boldsymbol{Li} + \frac{\boldsymbol{d_1}}{\boldsymbol{d_1} + \boldsymbol{d_2}}(\boldsymbol{i})$$

donde:
Li = Límite inferior de la clase que contiene a Mo.
d_1= fm - f_1
d_2= fm - f_2
fm= frecuencia de la clase que contiene a Mo.
f_1= frecuencia de la clase anterior que contiene a la Mo.
f_2= frecuencia de la clase posterior que contiene a la Mo.
i= Amplitud del intervalo.
Luego:

$$\boldsymbol{Mo.} = 1 + \frac{5}{5+3}(2) \qquad d_1=5-0=5$$

$$\boldsymbol{Mo.} = 1 + \frac{10}{8} = 2.25 \qquad d_2=5-2=3$$

Mo.=2.25

Uno de los procedimientos alternos ó métodos para sacar las Modas es ver que la frecuencia que le anteceden sea menor y la que le siga también. Este procedimiento se aplica cuando el investigador desea identificar los valores más representativos de un arreglo numérico.

Ejemplo:

Xi	Fi
2	1
3	10

4	3
5	4
6	7
7	2
8	3
9	1

Mo. = 3 Mo. = 6 Mo. = 8

En este caso se obtiene una situación multimodal, que, se reitera, en algunas situaciones es útil conocerla.

RELACIÓN ENTRE LA MEDIA ARITMÉTICA,LA MODA Y LA MEDIANA (7)

Expresaremos esta relación en función de su calidad como estimadores de los datos, observaciones o mediciones de una distribución determinada, la cual se gesta y expresa deacuerdo con los criterios matemático y empírico que utilicemos. Primero describiremos el criterio matemáticoy la forma en que es satisfecho por cada una de estas tres medidas de tendencia central y, posteriormente,veremos cómo sus valores difieren sistemáticamente entre si debido a varios tipos de distribuciones de mediciones.

Así, decimos que el criterio matemático de un "buen promedio " que satisface la **moda** se expresa como **Ne= mín, el que podemos interpretar** así: cuando usamos la moda como el mejor estimador del valor de cada medición en una distribución de mediciones, el número (N) de errores (e) es un mínimo. En otras palabras, se dice que la medida de tendencia central que produce el menor número de errores, cuando se usa como el mejor estimador de cada medición en un grupo o distribución de mediciones u observaciones, es la moda.

Por otra parte, si tomamos en cuenta **la magnitud** de cada error dentro del criterio matemático, si denominamos a "e" como la cantidad del error sin considerar su dirección o signo aritmético, y si deseamos **minimizar la suma de errores** en que se incurre al estimar el valor de cada medición u observación, el criterio matemático se expresa como **Σe= mín,** que **solo la mediana lo satisface**. Lo anteror significa que si usaramos otra medida de tendencia central para estimar cada estimación, la suma aritmética de los valores absolutos de los errores sería mayor que la suma de los errores obtenidos cuando usamos la mediana como estimador.

Ahora bien el tercer gran criterio que deben satisfacer los estimadores al utilizarlos, es que la **suma de errores al cuadrado sea un mínimo, que por cierto, como ya vimos en páginas anteriores, solo lo satisface la media aritmética; se representa por Σe^2 = mínimo,** que es muy importante en el análisis estadístico, en especial en el análisis de regresión (relación de asociación o de causalidad), campo en el que se le conoce como el criterio de **"mínimos cuadrados"** .

Al utilizar el **criterio empírico,** se observa que la media aritmética es la más afectada por la adición de datos en cualquier extremo de la distribución, que ya describimos como una desventaja en páginas anteriores. Le sigue la mediana pero en menor cuantía y, como también vimos, la moda no es afactada por la incorporación de mas datos a la serie o distribución estadística.

Derivado de lo anterior podemos decir que habrà distribuciones simétricas de datos, de sesgo positivo y de sesgo negativo, según hacia donde se distribuyan la mayoría de los datos, que se verà màs adelante con las medidas de asimetría y kúrtosis. Por el momento digamos que en una distribución simétrica se obtiene $\bar{x} = Md. = Mo$.

Demostración: sea la siguiente distribución de datos:

CLASES	Fi	PM	PM-A	F(PM-A)
2 y menos de 4	2	3	-4	-8
4 y menos de 6	3	5	-2	-6
6 y menos de 8	5	7	0	0
8 y menos de 10	3	9	2	6
10 y menos de 12	2	11	4	8
Total	15			0

Verificación:

Si A = 7

$$\bar{X} = A + \frac{\sum fi * (P.M. - A)}{\sum fi}$$

$$\bar{X} = A + \frac{0}{15} = 7; \bar{X} = 7$$

$$\boldsymbol{Md}.= \boldsymbol{L} + \frac{\frac{\boldsymbol{N}}{2} - \boldsymbol{C}}{\boldsymbol{fi}}(\boldsymbol{i})$$

$$\boldsymbol{Md}.= 6 + \frac{\frac{15}{2} - 5}{5}(2) = 7$$

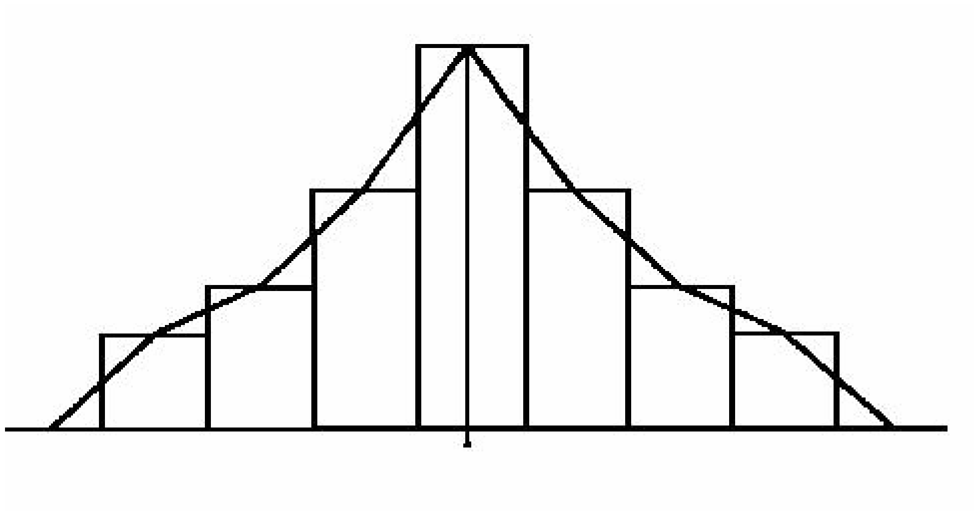
Md.=7

$$\boldsymbol{Mo}.= \boldsymbol{Li} + \frac{\boldsymbol{d}_1}{\boldsymbol{d}_1 + \boldsymbol{d}_2}(\boldsymbol{i})$$

$$\boldsymbol{Mo}.= 6 + \frac{2}{2+2}(2) = 7$$

$$\boldsymbol{Mo}.= \boldsymbol{Md}.= \bar{\boldsymbol{x}}$$

Histograma y Poligono de frecuencias de una distribución simétrica donde: $\overline{X} = Md. = Mo.$



histograma
$\overline{X} = Md. = Mo. = 7$

Podemos concluir diciendo que **la relación** entre estas tres medidas de tendencia central **es indicativa** de la dirección y extensión del alejamiento de los datos de la distribución, **de la simetría**.

Con base en lo anterior podemos preguntarnos, entonces ¿ cúal de las tres representa el mejor promedio? La respuesta dependerá de si, o no, la distribución está sesgada, así como del uso que pretendamos darle al promedio.

II.3.1.4 MEDIA GEOMETRICA (Mg)

Definición: Es un valor tal, que multiplicado ese valor tantas veces como el número de términos, resulta un producto igual al producto de los valores de los términos de la serie dada.

Si los términos de la serie dada son: $x_1, x_2, \ldots x_n$, aplicando la propiedad apuntada antes, resulta la expresión:

$$Mg.Mg.Mg.Mg.\ldots Mg = x_1, x_2, \ldots, x_n$$

Luego $Mg^n = x_1, x_2, \ldots, x_n$

Despejando $Mg = \sqrt[n]{x_1 * x_2 * x_3 * \ldots * x_n}$

Igualdad que nos dice: el promedio geométrico de una serie simple es igual a la raíz que tiene como índice el número de términos, del producto de los valores de los términos de la serie.

Sea el ejemplo: 3 * 6 * 12 * 24 * 48 =248,832
como $Mg = \sqrt[5]{248832}$
Mg= 12
Por definición : 12 * 12 * 12* 12 * 12 = 248,832
También 3 * 6 *12 * 24 * 48 = 248.832 = 12^5 = Mg^n.
Ahora bien, si sabemos que:
$Mg = \sqrt[n]{x_1 * x_2 * x_3 * \ldots * x_n}$
elevando a la potencia " n " ambos miembros
$Mg^n = x_1, x_2, \ldots, x_n$
Tomando logaritmo $n * \log(Mg) = \log x_1 + \log x_2 + \ldots + \log x_n$

luego: $\log(Mg) = \dfrac{\sum \log xi}{n}$,fórmula usual para el cálculo del promedio geométrico.

Ejemplo de su cálculo en una serie simple

$$\log(Mg) = \frac{\sum \log xi}{n}$$

SERIE SIMPLE

Xi	LogXi
1	0.00000
20	1.30103
7	0.84510
30	1.47712
18	1.25527
	4.87852

Sustituyendo
Log Mg = 4.87852/5 = 0.975704
Log Mg = 0.975704
Antilog de 0.975704 = 9.0303
Luego Mg. = 9.0303

Generalizando podemos decir que su cálculo es el siguiente:

Para una serie de frecuencias.

$$\log \boldsymbol{Mg} = \frac{\sum \boldsymbol{fi} * \log \boldsymbol{xi}}{\sum \boldsymbol{fi}}$$

Para una serie de clases y frecuencias :

$$\log \boldsymbol{Mg} = \frac{\sum \boldsymbol{fi} * \log \boldsymbol{P.M.}}{\sum \boldsymbol{fi}}$$

Para fines prácticos es preferible calcular el logaritmo de la media geométrica y luego el antilogaritmo de ésta.

Cálculo de Mg

En una SERIE DE FRECUENCIAS.

Xi	Fi	Log Xi	Fi LogXi
12	3	1.0792	3.2375
10	6	1.0000	6.0000
15	9	1.1761	10.5848
20	12	1.3010	15.6124
22	7	1.3424	9.3970
	37		44.8317

$$\log \boldsymbol{Mg} = \frac{\sum \boldsymbol{fi} * \log \boldsymbol{xi}}{\sum \boldsymbol{fi}}$$

$$\log \boldsymbol{Mg} = \frac{44.8313}{37} = 1.2116567$$

Antilog.de1.2116567=16.28

Mg=16.28

En una serie de clases y frecuencias

CLASES	Fi	PM	Log PM	Fi LogPM
de 10 a menos de 20	5	15	1.1761	5.8805
de 20 a menos de 30	6	25	1.3979	8.3876
de 30 a menos de 40	7	35	1.5441	10.8085
de 40 a menos de 50	8	45	1.6532	13.2257
Total	26			38.3023

$$\log \boldsymbol{Mg} = \frac{\sum \boldsymbol{fi} * \log \boldsymbol{P.M.}}{\sum \boldsymbol{fi}}$$

$$\log \boldsymbol{Mg} = \frac{38.3022}{26} = 1.4731615$$

Antilogaritmo de 1.4731615 = 29.72

Mg.= 29.72

Nota : Mg. es conveniente utilizarla cuando se manejan grandes números.

II.3.1.5 MEDIA ARMONICA (Ma).

Se define como el recíproco de la media aritmética de los valores recíprocos de la variable.

Es un número tal que la suma de las relaciones entre él y cada término es igual al número de términos de la serie.*

La Media Armónica es igual al número de sus términos divididos entre la sumas de los recíprocos de ellos.

$$Ma = \frac{n}{\sum \frac{1}{xi}}$$

donde:

n = número de observaciones
*x*i = Observaciones i-ésima .

Por consiguiente su cálculo se efectúa de la siguiente manera:

SERIE SIMPLE

Sean los cinco términos de la serie : 1, 2, 3, 4, 5

$$Ma = \frac{n}{\sum \frac{1}{xi}} = \frac{5}{2.28} = 2.192$$

Xi	$\frac{1}{Xi}$
1 2 3 4 5	1.0 0.5 0.3 0.3 0.2
15	2.28

* Aplicando la propiedad tenemos :

$$\frac{2.192}{1} = 2.19$$
$$\frac{2.192}{2} = 1.092$$
$$\frac{2.192}{3} = 0.730$$
$$\frac{2.192}{4} = 0.548$$
$$\frac{2.192}{5} = 0.434$$

5=5 términos

SERIE DE FRECUENCIAS

Xi	Fi	1/Xi	Fi*(1/Xi)
3	1	0.33	0.33
4	2	0.25	0.50
2	3	0.50	1.50
5	4	0.20	0.80
Total	10		3.13

$$\boldsymbol{Ma} = \frac{\sum \boldsymbol{fi}}{\sum \frac{1}{\boldsymbol{xi}} \boldsymbol{f}} = \frac{\sum \boldsymbol{fi}}{\sum \frac{\boldsymbol{fi}}{\boldsymbol{xi}}}$$
$$\boldsymbol{Ma} = \frac{10}{3.13} = 3.1948$$

SERIE DE CLASES Y FRECUENCIAS

CLASES	Fi	PM	Fi/Pm

1 a 2	2	1.5	1.33
de más de 2 a 4	3	3	1.00
de más de 4 a 6	4	5	0.80
de más de 6 a 8	1	7	0.14
	10		3.27

$$Ma = \frac{\sum fi}{\sum \frac{fi}{P.M.}} \; o \; \frac{\sum fi}{\sum \frac{1}{P.M.} fi}$$

$$Ma = \frac{10}{3.24} = 3.08$$

NOTA: Ma. es conveniente aplicarlo en tasas de crecimiento, cuando manejamos fenómenos como la velocidad, es decir con un crecimiento gradual.

II.3.1.6. RELACIÓN ENTRE LAS MEDIDAS DE TENDENCIA CENTRAL[4]

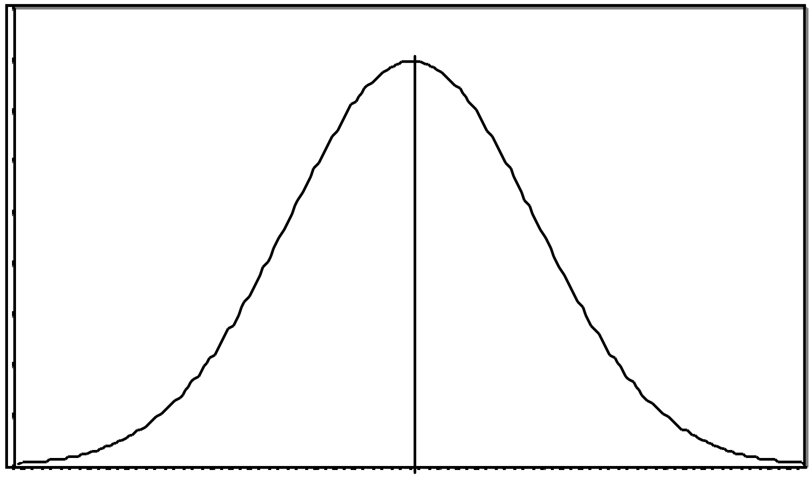
Representan valores centrales que se les designa como promedios y son de extraordinaria utilidad tanto en el análisis de una distribución como en la comparación entre distribuciones.

Una vez conseguida la clasificación de los datos originales cuyas características más esenciales se destacan, será preciso calcular en conjunto de indicadores que caractericen en forma algo más precisa la distribución que se está efectuando.

Se necesita disponer de medidas que representen valores centrales en torno de los cuales se agrupan las observaciones, en general se les designa como promedios.

1.- Distribución Simétrica

$\overline{X}$ = Mo = Md

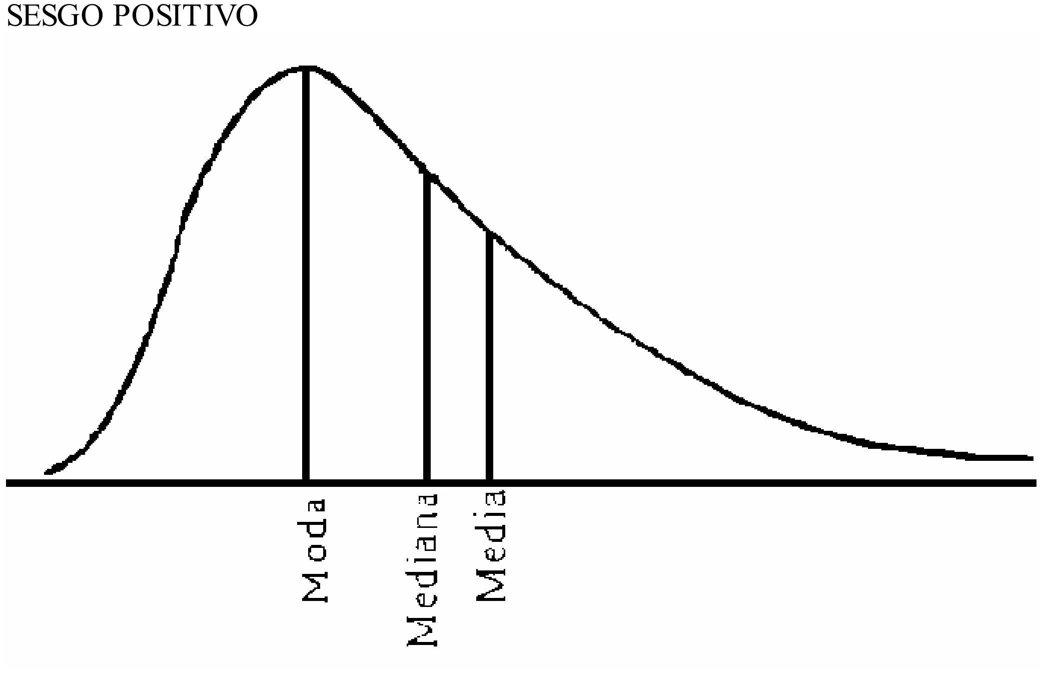


$\overline{X} = Mo = Md$

2.- Distribución Asimétrica

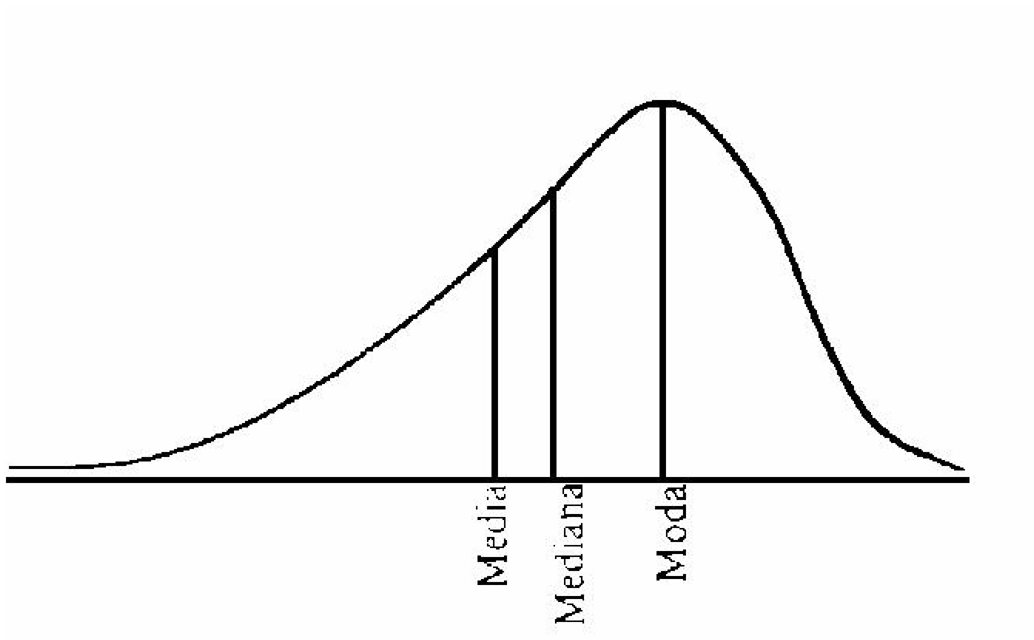
$Mo < Md < \overline{X}$

SESGO POSITIVO

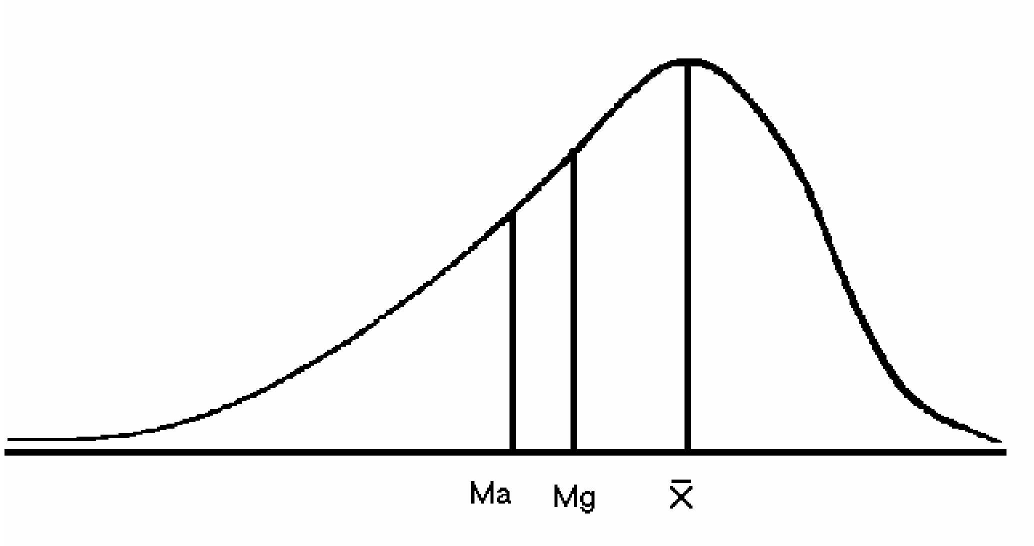


$\overline{X} < Md < Mo$

SESGO NEGATIVO



3.-Otra relación muy importante es:
$\overline{X} > Mg > Ma$.



II.3.I.7 PRACTICAS NÚMERO I,II y III

PRÁCTICA I

Nombre:_______________________________________Grupo:_________

Problema 1. Construya usted una serie simple con los siguientes datos, que representan la estatura de 20 estudiantes de la Facultad de Economía.

1.67, 1.72, 1.54, 1.57, 1.61, 1.61, 1.67, 1.54, 1.57, 1.72
1.85, 1.81, 1.54, 1.61, 1.81, 1.67, 1.81, 1.67, 1.61, 1.67

Problema 2. Con los datos siguientes que representan el número de hijos de 60 familias campesinas, construya una serie de frecuencias.

4	5	8	7	6	7	8	9	14	15
7	9	10	8	6	11	0	10	1	3
12	16	12	1	15	15	4	13	0	6
2	3	4	5	16	5	11	6	9	12
9	13	6	10	18	4	14	8	9	13
11	6	8	12	4	20	17	10	7	6

a) En base a estos datos, diga usted cuál es el número de hijos que se presentan con mayor frecuencia en las familias campesinas y emita su opinión al respecto.

b) Si uno de los objetivos del pasado régimen era el control de la natalidad, y se pensaba que como resultado de esa campaña, el promedio de hijos entre las familias sería menos de 5, en base a los datos anteriores que, porcentaje de ellas no cumplieron con el objetivo y explique a qué conclusiones les lleva este análisis.

PRÁCTICA II

Problema 1. Los accidentes de trabajo ocurridos en 60 fábricas de la zona industrial de Tlalnepantla en 1998, están dados en el siguiente cuadro.

No. de Accidentes	No. Fabricas
0 a 4	3
5 a 9	6
10 a 14	15

15 a 19	12
20 a 24	10
25 a 29	9
30 a 34	5
	60

a) Calcule usted la media aritmética, la media geométrica y la media armónica e interprete cada uno de estos resultados, asimismo explique la relación que existe entre ellas.

b) Si en la zona industrial de Tlalnepantla existen 1350 fábricas cuantos accidentes ocurrieron allí durante 1998.

c) Si tomamos esta distribución como un fiel reflejo de la situación que impera en el país en la actualidad, en la gran mayoría de las industrias, ¿cual debería de ser la política del Estado en este renglón y porqué?

Problema 2.- Para poder garantizar la duración de una determinada marca de llantas, se realizó una investigación en 100 llantas, con los kilómetros recorridos y se obtuvieron los siguientes datos.

Miles de kilómetros	No. de llantas
de + 25 a 30	18
de + 30 a 35	12
de + 35 a 40	35
de + 40 a 45	20
de + 45 a 50	15
	100

a) Determine gráficamente la mediana, por medio de método gráfico de la ojiva
b) Calcule el valor de la mediana y la moda y explique sus resultados.
c) Si el lema de la marca llantera era garantizarlas por mas de 40,000 km, que porcentaje de la producción no cumple ese requisito.

PRÁCTICA III.

Problema 1.- Las estaturas de un grupo de 40 estudiantes de una escuela secundaria fueron las siguientes.

1.38	1.64	1.50	1.32	1.44	1.25	1.49	1.57
1.46	1.58	1.40	1.47	1.36	1.48	1.52	1.44
1.68	1.26	1.38	1.76	1.63	1.19	1.54	1.65
1.46	1.73	1.42	1.47	1.35	1.53	1.40	1.35
1.61	1.45	1.35	1.42	1.50	1.56	1.45	1.28

PREGUNTAS

a) Ordene los datos anteriores en una serie de clases y frecuencias, de acuerdo al método de Sturges.

b) Construya usted el histograma y el polígono de frecuencias correspondientes.

c) Calcule la media aritmética, la mediana y la moda, explicando la relación que existe entre estos valores.

PROBLEMA 2 .- De un estudio realizado por la Secretaría de la Reforma Agraria, se obtuvieron los siguientes datos, relacionados con el número de hectáreas que concentra cada agricultor en una zona del país.

Hectáreas	No. de Agricultores
'0-2	6
3-5	10
6-8	14
9-11	6
12-14	4
15-17	2
18-20	8
	50

PREGUNTAS.

a) En base a estos datos, explique y compruebe la primera y la segunda propiedad de la media aritmética.

b) Determine gráficamente y numéricamente si esta es una distribución simétrica, calculando su media aritmética y trazando el polígono de frecuencias.

II.3.2 MEDIDAS DE DISPERSIÓN[4]

Varias distribuciones pueden tener iguales medidas de posición pero con diferentes grados de dispersión.

Como por ejemplo: en las distribuciones siguientes todas tienen la media y la mediana igual a 5, pero con diferentes composiciones.

A: 0, 0, 1, 2, 5, 8, 9, 10, 10

B: 2, 3, 3, 4, 5, 5, 5, 6, 7, 7, 8

C: 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5

Como se ve la distribución A presenta un grado más alto de variaciones, mientras la C es la más homogénea por no haber variabilidad de los datos. Para medir esta dispersión de los términos con respecto a una medida de tendencia central (que generalmente es $\bar{x}$) se aplican las medidas de dispersión que son las siguientes:

II.3.2.1 RANGO: La medida de dispersión más sencilla es el campo de variación del Rango, definida como la diferencia entre el mayor y el menor de los valores observados

El Rango también lo podemos encontrar con el nombre de Recorrido u otro nombre, según el autor.

El Rango o Recorrido no refleja en modo alguno la forma de la distribución.

Ejemplo: datos: 2, 3, 4, 5, 6, 7, 8, 9

Rango = Valor máximo - valor mínimo de la serie.

Rango = 9 - 2 = 7

II.3.2.2 DESVIACION MEDIA: Se define como la suma de las desviaciones en términos absolutos de los datos que integran la serie, respecto a la media, entre el número de términos de la serie.

$$D.M. = \frac{\sum |xi - \bar{x}|}{n}$$ Serie simple

$$D.M.=\frac{\sum fi|xi-\bar{x}|}{\sum fi}$$ Serie de frecuencias

$$D.M.=\frac{\sum fi|P.M.-\bar{x}|}{\sum fi}$$ Serie de clases y frecuencias

Ejemplo: la temperatura en ciertos días del mes de mayo, expresada en grados centígrados fué:

X_1	$\lvert X_1 - \bar{X} \rvert$
22	2
23	1
23	1
24	0
25	1
26	2
27	3
170	10

i = 22, 23, ..., 27

$$\bar{x}=\frac{170}{7}=24^\circ$$

$$D.M.=\frac{\sum|x_1-\bar{x}|}{n}$$

$$D.M.=\frac{10}{7}=1.42^\circ$$

Primero se calcula $\overline{X}$ para después desviar con respecto a ella el valor de los términos de la serie, en este caso las temperaturas observadas durante el mes de mayo.

Interpretación: de conformidad con el rango, la temperatura fluctúo entre 22 y 27 grados centígrados; con base en la desviación media, la temperatura media fué de 24 grados centígrados y tuvo una variación media de 1.42 grados durante el mes.

Serie de frecuencias

$F_i \lvert X_i - \bar{X} \rvert$

15
6
4
7
2
2
4
40

$\lvert X_1 - \bar{X} \rvert$	F_i	X_iF_i	$F_i \lvert X_1 - \bar{X} \rvert$
3	5	120	15
2	4	102	6
2	2	46	4
1	7	168	7
0	8	200	2
1	2	52	2
2	2	54	4
	30	742	40

$$\bar{x} = \frac{\sum x_i f_i}{\sum fi} \qquad D.M. = \frac{\sum fi \lvert x_1 - \bar{x} \rvert}{\sum fi}$$

$$\bar{x} = \frac{742}{30} = 25 \qquad D.M. = \frac{40}{30} = 1.33^{\circ}$$

Serie de clases y frecuencias

Cuando los datos aparecen ya ordenados o agrupados en una serie de clases y frecuencias, las fórmulas que se deben aplicar serán:

$$\bar{x} = \frac{\sum P.M. * f_i}{\sum fi}$$

$$D.M. = \frac{\sum fi \lvert P.M. - \bar{x} \rvert}{\sum fi}$$

Temperatura en Intervalos	Fi	PM	Fi PM	$\|PM - \bar{X}\|$
22 y menos de 24	5	23	115	-3
24 y menos de 26	9	25	225	-1
26 y menos de 28	10	27	270	1
28 y menos de 30	6	29	174	3
	30		784	0

$F_i \|PM - \bar{X}\|$
15
9
10
18
52

$$\bar{x} = \frac{784}{30} = 26^\circ \qquad \boldsymbol{D.M.} = \frac{52}{30} = 1.73^\circ$$

Interpretación: independientemente de que la información aparezca ordenada en una serie simple, de frecuencias o de clases y frecuencias, como los datos son los mismos, la desviación media nos permite verificar que la temperatura no varió mucho en el mes de mayo, ya que en promedio fué de 26 grados, cuando los datos provenían de una serie de clases y frecuencias y sin embargo, durante los 30 días del mes, en promedio se observo una variación o dispersión de 1.73 grados con respecto a los 26 grados centígrados.

Es muy importante recordar que la serie de clases y frecuencias da resultados de menor exactitud que la simple y la de frecuencias, ya que maneja los puntos medios como valores sustitutos de los valores originales de la serie.

II.3.2.3 DESVIACIÓN ESTÁNDAR (σ)

Es la raíz cuadrada positiva de las desviaciones al cuadrado de los valores observados, respecto a la media aritmética; indica el grado de desviación que tienen los términos de la serie con respecto a su media aritmética. Para una serie simple se calcula así:

X_i	$\delta_a = X_i - \bar{X}$	δ_a^2
1	-9	81
2	-8	64
5	-5	25
9	-1	1
11	1	1
13	3	9
14	4	16
25	15	225
80	0	422

$$\bar{X} = \frac{\sum x}{n} = \frac{80}{8} = 10$$

si δ: desviación con respecto ó $\bar{x}$, tenemos

$$\sigma = \sqrt{\frac{\delta_a^{\ 2}}{n}} = \sqrt{\frac{422}{8}} = \sqrt{52.75}, \sigma = 7.26$$

MÉTODO DIRECTO PARA UNA SERIE DE FRECUENCIAS.

X_i	F_i	Xf	δ_a	δ_a^2	$\delta_a^2 F_i$
1	9	9	-1.87	3.50	31.50
2	15	30	-0.87	0.76	11.40
3	29	87	0.13	0.02	0.58
4	10	40	1.13	1.28	12.80
5	7	35	2.13	4.54	31.78
	70	201			88.06

$$\overline{X} = \frac{\sum xifi}{\sum fi} = \frac{201}{70} = 2.87$$

$$\sigma = \sqrt{\frac{\sum \delta_a^2 fi}{\sum fi}}$$

$$\sigma = \sqrt{\frac{88.06}{70}} = 1.1$$

SERIE DE CLASES Y FRECUENCIAS

METODO ABREVIADO

Definiendo δ_a como desviación arbitraria *xi* como amplitud del intervalo

Intervalos de clase	F_i	δa	δaf	δa^2	$\delta a^2 F_i$
De 1.0 a 1.5 inclusive	2	-3	-6	9	18
de más de 1.5 a 2.0	5	-2	-10	4	20
de más de 2.0 a 2.5	12	-1	-12	1	12
de más de 2.5 a 3.0	28	0	0	0	0
de más de 3.0 a 3.5	20	1	20	1	20
de más de 3.5 a 4.0	14	2	28	4	56
de más de 4.0 a 4.5	3	3	9	9	27
de más de 4.5 a 5.0	1	4	4	16	16
	85		33		169

$$\sigma = i\sqrt{\frac{n(\sum \delta a^2 F) - (\sum \delta af)^2}{n(n-1)}} = 0.5\sqrt{\frac{85(169) - (33)^2}{85(85-1)}} = 0.5\sqrt{\frac{14365-1089}{85(84)}} = 0.5\sqrt{\frac{13276}{7140}} = 0.5 * 1.36 = 0.680$$

CON EL MÉTODO DIRECTO σ SE CALCULA ASÍ

Intervalos de Clase	F	PM	PMF	PM-$\bar{X}$	(PM-$\bar{X}$)^2	(PM-$\bar{X}$)^{2}F
De 1.0 a 1.5 inclusive	2	1.25	2.5	-1.69	2.86	5.71
de más de 1.5 a 2.0	5	1.75	8.75	-1.19	1.42	7.08
de más de 2.0 a 2.5	12	2.25	27	-0.69	0.48	5.71
de más de 2.5 a 3.0	28	2.75	77	-0.19	0.04	1.01
de más de 3.0 a 3.5	20	3.25	65	0.31	0.10	1.92
de más de 3.5 a 4.0	14	3.75	52.5	0.81	0.66	9.19
de más de 4.0 a 4.5	3	4.25	12.75	1.31	1.72	5.15
de más de 4.5 a 5.0	1	4.75	4.75	1.81	3.28	3.28
	85		250.25			39.05

Teniendo los datos agrupados en clases y frecuencias, se procede a obtener la medida aritmética de ellos.

Como se recordar, la fórmula de la media aritmética viene dada por

$$\bar{X} = \frac{\sum P.M.f}{\sum f}$$

$$\bar{X} = \frac{250.25}{85}$$

$$\bar{X} = 2.94$$

Después se procede a desviar el punto medio con respecto a la media $(PM - \bar{X})$.

Se eleva al cuadrado y multiplicamos por su frecuencia respectiva, llegando a la fórmula de la desviación estándar:

$$\sigma = \sqrt{\frac{\sum (P.M. - \bar{X})^2 fi}{\sum fi}} = \sqrt{\frac{39.38}{85}}$$

El resultado es el mismo que el obtenido con el método abreviado:

II.3.2.4 VARIANZA

Se define como el cuadrado de la desviación estándar

$$\sigma^2 = \frac{\sum (Xi - \bar{X})^2}{N} = \frac{\sum xi^2}{N} - \bar{X}^2 \text{ para una serie simple.}$$

$$\sigma^2 = \frac{\sum fi(Xi - \bar{X})^2}{\sum fi} \text{ para una serie de frecuencias.}$$

$$\sigma^2 = \frac{\sum fi(P.M. - \bar{X})^2}{\sum fi} \text{ para una serie de clases y de frecuencias.}$$

II.3.2.5 COEFICIENTE DE VARIABILIDAD (C.V.).

Se define como la razón porcentual entre la desviación estándar y la media aritmética, es decir.

$$C.V. = \frac{\sigma}{\bar{X}} * 100$$

La razón es conveniente multiplicarla por 100 para expresarla en términos de %.

Se dice que cuando C.V. toma valores alrededor del 10% se acepta $\bar{X}$ como medida central representativa de los datos de un fenómeno bajo estudio; cuando C.V. es mayor a 10%, se debe optar por otra medida de tendencia central para representar el promedio de los datos del fenómeno bajo estudio.

Ejercicios

Con los siguientes datos:

15,11, 10,18, 17, 14, 14,15, 16,13
12,12, 9,11,14,16,15,14,13, 10
13,12,10,12,14,16,15,17,13,10
14,11,11,13,14,16,15,17,13,10
15,11,14,14,14,16,15,17,13,10

obtenga:

1.- Una serie de clases y frecuencias con el métodode Sturges;

2.- la relación de la media aritmética con la media geométrica y la moda;
3.- Las dos propiedades de la media aritmética, es decir, "suma cero y mínimo ".
4.- la varianza y desviación estándar;
5.- El coeficiente de variación con su interpretación correspondiente.

Respuestas

1.- La fórmula de Sturges:

i = oscilación / 1+3.32 log n ; donde n = 50 y *oscilación* es la diferencia en valores absolutos que existe entre el dato de mayor valor y el dato de menos valor.
Así:
I = 18-9/1 + 3.32(1.6990) = 9/ 1+ 5.64 = 9/ 6.64 = 1.36 redondeando a 1.4

Clases	fi	PMi	fiPMi	logPMi	logPMifi
De 9 a 10.4	7	9.7	67.9	0.9542	6.6794
De + de 10.4 a 11.8	5	11.1	55.5	1.0453	5.2265
De + de 11.8 a 13.2	11	12.5	137.5	1.0969	12.0659
De + de 13.2 a 14.6	10	13.9	139.0	1.1430	11.4300
De + de 14.6 a 16.0	12	15.3	183.6	1.1847	14.2164
De + de 16.0 a 17.4	4	16.7	66.8	1.2227	4.8908
De + de 17.4 a 18.8	1	18.1	18.1	1.2577	1.2577
Total	50		668.4		55.7667

Respuesta 2

$\bar{X}$ = Σ fi(PMi) / Σ fi = 668.4/50 = 13.368, redondeado a **13.36**

logMg = Σ (logPMi)fi/ Σ fi = 55.7667/50 = 1.115334; su ***antilogaritmo* = 13.04**
luego Mg = 13.04
Moda = Li + $[d1/d1+d2]$(i)= 14.6 + $[12-10/2+(12-4)](1.4)$ =14.6+
$[2/2+8](1.4)$ = 14.6 + 2.8/10 = 14.88

Luego Moda = 14.88 mayor que $\bar{X}$ = 13.36 mayor que Mg = 13.04

Respuesta 3

3a. Primera propiedad: $\Sigma (PMi- \bar{x})fi = 0$

(PMi - $\bar{x}$)	**(PMi - $\bar{x}$)fi**	$(PMI - \bar{x})^2$	$(PMi - \bar{x})^2$ **fi**
9.7-13.36=-3.66	-3.66(7)= -25.62	13.3956	93.7692
11.1-13.36=-2.26	-2.26(5)=-11.30	5.1076	25.5380
12.5-13.36=-0.86	-0.86(11)=-9.46	0.7396	8.1323
13.9-13.36=0.54	0.54(10)= 5.40	0.2916	2.9160
15.3-13.36=1.94	1.94(12)=23.28	3.7636	45.1632
16.7-13.36=3.34	3.34(4)=13.36	11.1556	44.6224
18.1-13.36=4.74	4.74(1)=4.74	22.4676	22.4676
Total	0.00		242.6087

3b.- Segunda propiedad: $\Sigma (PMi- \bar{x})^2 fi$= MINIMO=242.6087

Respuesta 4ª: la varianza y desviación estándar

$\sigma^2 = \Sigma (PMi- \bar{x})^2 fi \;/\; \Sigma fi = 242.6087/50 = 4.852174$

$\sigma = \sqrt{4.852174} = 2.20$

Respuesta 5ª: Coeficiente de variación,CV.

CV= $\sigma / \bar{x}$ *100 = 2.20/13.36*100 = **16.5%**

Interpetación:hay una variación significativa que supera el 10% recomendable, entre los valores de los términos Xi, que se expresa en la alta proporción de la desviación estándar con respecto a la media aritmética; se recomienda cambiar de medida de tendencia central a otra, digamos, la mediana, la moda, etc.