

APUNTES DE ESTADISTICA .

Documento preparado por la División de Censos y Proyectos
Específicos - DICEPRO -

DIRECCION DE ANALISIS SOCIO ECONOMICO

Bogotá, Febrero, 1977

INDICE

Página

INTRODUCCION

CAPITULO I - CONCEPTOS GENERALES

1.	Breve reseña histórica	1
2.	Etimología de la palabra Estadística	4
3.	Algunas definiciones de la Estadística	6
4.	Aplicaciones de la Estadística	7
5.	Utilización de la Estadística	11

CAPITULO II - NOCIONES BASICAS DE ESTADISTICA

1.	Definiciones	15
1.1	Datos estadísticos	
1.2	Variables	
1.3	Universo o población	
1.4	Muestra	
2.	Etapas de una Investigación Estadística	18
2.1	Proyecto y preparación	
2.1.1	Definición de objetivos	
2.1.2	Definición o determinación de la población objeto de estudio	
2.1.3	Fuentes de información	
2.1.4	Conformación definitiva del equipo que va a realizar la investigación	
2.1.5	Costo de la investigación	
2.1.6	Propaganda	
2.2	Recolección	
2.3	Procesamiento de la información	
2.4	Exposición	
2.5	Análisis e interpretación	
2.6	Publicación	
3.	Series Estadísticas	25

- 3.1 Series cronológicas
- 3.2 Series espaciales
- 3.3 Series cualitativas
- 3.4 Series cuantitativas

4. / Tipos de Gráficos 32

- 4.1 Gráfica circular
- 4.2 Gráfica polar
- 4.3 Gráfica cuadrangular
- 4.4 Curva estadística
- 4.5 Gráficas de barras

5. Cuadros Estadísticos 41

- 5.1 Partes de que consta un cuadro estadístico
- 5.2 Clases de cuadros
 - 5.2.1 Simples
 - 5.2.2 De doble entrada
 - 5.2.3 Complejos

CAPITULO III - ESTADISTICA DE ATRIBUTOS

1. Nociones sobre teoría de conjuntos 48

- 1.1 Definiciones
- 1.2 Operaciones con conjuntos
- 1.3 Propiedades e implicaciones

2. Clasificación por Atributos 58

- 2.1 Terminología
- 2.2 Frecuencias y orden de las clases
- 2.3 Total de las frecuencias de clase

3. Relación de Atributos 78

- 3.1 Consistencia de Atributos
- 3.2 Criterios de independencia
- 3.3 Asociación de atributos
 - 3.3.1 Coeficiente de asociación

3.4	Asociaciones parciales	
3.4	Asociaciones ilusorias	
4.	Clasificación. Múltiple	123
4.1	Consideraciones generales	
4.2	Tablas de contingencia	
4.3	Asociación en las tablas de contingencia	
4.4	Interpretación del coeficiente de contingencia de PEARSON	
4.5	Coeficiente de TSCHUPROW	
4.6	Clasificaciones homogéneas y heterogéneas	
4.7	Clasificación múltiple como sucesión de clasificaciones dicotómicas	

CAPITULO IV - DISTRIBUCIONES DE FRECUENCIA

1./	Introducción	144
2.	Determinación de Intervalos y Grupos de Clase	146
3. /	Conceptos y gráficos de las Frecuencias	154
3.1	Frecuencias absolutas	
3.2	Frecuencias relativas	
3.3	Frecuencias absolutas acumuladas	
3.4	Frecuencias relativas acumuladas	
4.	Distribuciones de frecuencias para Variables Discretas	171
5.	Momentos estadísticos Unidimensionales	172
5.1	Momentos para datos originales	
5.1.1	Momentos con respecto a un origen cero (a_r)	
5.1.2	Momentos con respecto a la Media (m_r)	
5.2	Momentos para datos agrupados	
5.2.1	Momentos con respecto a un origen cero (a_r)	

- 5.2.2 Momentos con respecto a la Media Aritmética (m_r)
- 5.2.3 Momentos con respecto a un origen de trabajo (m'_r)
- 5.2.4 Momentos con respecto a un origen de trabajo medidos en unidades de intervalo (m''_r)

5.3 Relaciones entre momentos

6. Medidas de tendencia central - promedios 187

6.1 Matemáticos

6.1.1 Media aritmética

- 6.1.1.1 Media aritmética ponderada
- 6.1.1.2 Propiedades de la media aritmética
- 6.1.1.3 Procedimientos abreviados de cálculo
- 6.1.1.4 Ventajas y desventajas del promedio aritmético

6.1.2 Media geométrica

6.1.3 Media armónica

6.2 Promedios de posición

6.2.1 Mediana

- 6.2.1.1 Cálculo en datos originales
- 6.2.1.2 Cálculo en datos agrupados
- 6.2.1.3 Características de la mediana

6.2.2 Fractilas

6.2.3 Modo o moda

7. Medidas de Dispersión 243

7.1 Dispersión absoluta

7.1.1 Varianza

7.1.1.1 Procedimientos abreviados de cálculos

7.1.1.2 Propiedades de la varianza

7.1.2 Desviación Típica

7.1.3 Desviación Media

7.1.4 Desviación Mediana

7.1.5 Recorrido

7.1.6 Recorrido Intercuartílico

7.2 Dispersión Relativa - Coeficiente de Variación

8. Coeficiente de Deformación (asimetría) 275

9. Curtosis o estadígrafos de apuntamiento 278

CAPITULO V - PROBABILIDADES 282

ANEXO

BIBLIOGRAFIA

INTRODUCCION

El DANE, dentro de su política de mejorar y perfeccionar el nivel técnico y académico de las personas que se hallan vinculadas a la institución, ha querido iniciar una serie de publicaciones didácticas orientadas a refrescar y consolidar los conocimientos en la Universidad y para algunos iniciar con la presente publicación una formación estadística básica.

En desarrollo de esta idea, la primera publicación estará orientada al conocimiento general de los interesados por estas inquietudes del campo amplísimo de utilización estadística, comenzando por una breve reseña histórica y aspectos complementarios de generalidades

Se inicia el desarrollo del tema con un análisis bastante amplio sobre el tratamiento de los atributos y propiedades lógicas, los problemas y las aplicaciones dando al lector una idea general satisfactoria sobre estos tópicos. Viene luego la denominada estadística descriptiva, en la cual se hace un amplio análisis de los problemas que se presentan, la manera de tratarlos y la forma de llevarlos al campo de las conclusiones.

Como este aspecto es bastante elemental y de fácil comprensión, se comenzó el estudio de variables estadísticas teniendo en cuenta los

hechos que evidentemente se presentan, para luego pasar a los mo
delos teóricos, que serán materia de presentación futura.

Dentro de la estadística descriptiva se estudia lo relativo al análi
sis de datos dentro de las investigaciones corrientes y en particul
lar la construcción de distribuciones, su utilización para el cálcu
lo de las diferentes funciones, las aplicaciones pertinentes y en -
general el manejo de la información; también se desarrolla lo re-
lacionado con momentos de distinto tipo y uso.

Se puede afirmar que el temario de las presentes conferencias es
un vehículo que permitirá a los lectores conocer inicialmente la
estadística en sus aspectos preliminares y ~~te~~ inducirá a lo~~gr~~ar -
una más sólida formación. Para ello, se procurará en el futuro
desarrollar los aspectos complementarios de distribuciones de -
más de una variable, cálculo de probabilidades, sus distribuciones
y algunas nociones de muestreo.

Por otra parte parece ser importante estudiar también lo relacio
nado con el análisis de series cronológicas, índices y la teoría de
la regresión. Como los temas expuestos son extensos y es la intenu
ción desarrollarlos por áreas más o menos compatibles, éstas se
realizarán con cierta holgura en el futuro.

Será satisfactorio que estas conferencias contribuyan al mejoramien
to académico e intelectual de quienes las estudien.

CAPITULO I

CONCEPTOS GENERALES

1. BREVE RESEÑA HISTORICA

La estadística surgió como una necesidad para el análisis de sucesos y obtención de resultados y conclusiones, así como también, la observación de fenómenos parecidos o aún iguales en tiempos o espacios diferentes o simultáneos. Las primeras investigaciones de recopilación datan de tiempos anteriores a la era cristiana, cuando se hicieron los primeros censos de población. Naturalmente, estos censos se referían exclusivamente a contar el número de individuos, e incluso se llegaba a obligar a las personas a radicarse temporalmente en su lugar de origen. Esta circunstancia traía como consecuencia, un error de datos de población de una provincia o lugar, ya que artificialmente se hacían desplazamientos masivos para estos eventos. Posteriormente y aplicada a la astronomía, se observó que los fenómenos terrestres en el transcurso del año, presentaban determinadas características más o menos parecidas; esto hizo pensar a los científicos de la época, la necesidad de aplicar ciertos sistemas de medición a las experiencias que ellos observaban.

En épocas posteriores al Imperio Romano, cuando la humanidad empezaba a tener un mayor desenvolvimiento económico y social, se -

hizo necesario tener informaciones de hechos mediatos y futuros. Se empezó a efectuar recopilación de informaciones sobre algunos productos agropecuarios tratando de sistematizarlos en forma tal que sirviera a los gobiernos como pauta para previsiones futuras; se observó que en muchos de ellos se marcaba una tendencia definida, lo cual incrementó este tipo de investigaciones. En esta forma las recopilaciones de datos y agrupaciones fueron cada vez más importantes, con el fin de producir una serie de informaciones más completas en el futuro y por ende, trabajar con menos incertidumbre.

En la Edad Media se encuentran estadísticas más sistematizadas. El clero se dedicó a la recopilación, ordenación y estudio de datos. En América es posible encontrar algunos antecedentes estadísticos de la época precolombina. El Imperio Incaico llevaba registros de su población y otros recuentos. Las tribus nómadas que lucharon para establecerse en México, efectuaron catastros de contribuyentes y sabían estimar cuantitativamente sus pertenencias; posteriormente, Felipe II dispuso una investigación general de los territorios colonizados.

La estadística pasó a ser una organización estatal durante el despotismo ilustrado de Federico Guillermo I y de Federico el Grande de Prusia. Se dió especial importancia a la demografía, que hasta entonces había sido iniciativa particular de la Iglesia, que a partir del Concilio de Trento había hecho obligatoria la inscripción de los nacimientos, matrimonios y defunciones.

En Alemania se creó la primera Cátedra y cursos de Estadística, que dió origen a la Escuela Universitaria Administrativa, que la consideraba como una descripción de fenómenos de interés para el Estado. Hermann Conring empezó a dictar un curso de Estadística en la Universidad de Helmstadt, en el que pasaba revista a los casos más notables del Estado; su discípulo Godofredo de Achenwall, le dió el nombre de Estadística y la separó de la Sociología.

En Inglaterra, por esta época se desarrollaba la escuela demográfica, o de los aritméticos políticos quienes se proponían fijar en forma cuantitativa las leyes empíricas que regían los fenómenos políticos y sociales. Casi contemporánea fue la Escuela Probabilística originada en Francia, con los problemas de los juegos de azar, y cuya solución motivó el auge del cálculo de probabilidades,

en el cual se interesaron entre otros: Fermat, Laplace, Poisson, los hermanos Bernowilli, Gauss y de Moivre. El desarrollo de esta escuela ha continuado hasta nuestros días, en los cuales el cálculo de probabilidades es la base de la teoría del muestreo.

2. ETIMOLOGIA DE LA PALABRA ESTADISTICA

La etimología de la palabra Estadística es dudosa; para unos procede del Latín, Status: Situación de personas o cosas; mientras que otros la hacen derivar de la voz griega STATERA: Balanza, fundándose en que la Estadística mide o pesa los hechos sometidos a su conocimiento, no faltando quienes la hacen derivar del Alemán, Staat: Estado, como expresión de la unidad política superior.

La etimología Latina es la que más explica actualmente la finalidad de la ciencia Estadística, porque sus resultados contribuyen de manera decisiva al mejoramiento de la sociedad. La ciencia es un producto del esfuerzo colectivo de hombres y mujeres a través de una multitud de generaciones; de manera que sus conquistas y resultados pertenecen en común a la humanidad entera. Los investigadores científicos ejecutan un gran número de operaciones mentales y manuales, mediante las cuales consiguen descubrir la existencia de nuevos obje

tos, conocer sus distintos aspectos, determinar sus vínculos, com
probar las conclusiones establecidas y encontrar las maneras de -
utilizar los resultados. Así, en cualquier momento el conjunto de
todos los conocimientos elaborados se encuentra a disposición de
quienes tienen necesidad de aprender o de beneficiarse con sus aplii
caciones. La aplicación científica del conocimiento distingue las
fases observadas en el desarrollo de los procesos; por eso es que
los conocimientos científicos pueden ser comprobados por cualquier
persona y su verificación se puede practicar en todo momento. De
esta manera se pone en claro que la existencia de los objetos es la
fuente del conocimiento científico y constituye, a la vez, base inelu
dible para su comprobación. El método científico se funda estricta
mente en las técnicas experimentales, las operaciones lógicas y la
imaginación racional, para servir así como instrumento de adquisii
ción del conocimiento científico. Se desarrolla a través del esfuerzo
conjunto de los pensadores y los experimentadores. En cada inve
stigación que se realiza además de adquirir un nuevo conocimiento
también se obtiene una comprobación concreta de la eficacia del
método empleado.

Teniendo en cuenta lo anterior se puede afirmar que la etimología Alemana se aproxima a la misma finalidad en cuanto inicialmente la ciencia y el Método Estadístico fueron empleados por los Estadistas o Gobernante para obtener información actualizada, especialmente sobre los Recursos Humanos.

3. ALGUNAS DEFINICIONES DE LA ESTADISTICA

En cuanto hace relación a la definición propia de estadística, aunque los autores tienen conceptos que difieren entre sí, sus propósitos son los mismos.

En general la ESTADÍSTICA es la ciencia que trata de analizar un determinado conglomerado con características definidas, para obtener información básica o fundamental, es decir, que trata de definir un resultado y que éste sirva de base a futuras experiencias.

Probablemente las definiciones en general coinciden con el PROFESOR ROS: Estadística es la ciencia que estudia los fenómenos colectivos, mediante la observación numérica, el análisis matemático y la interpretación lógica, investigando especialmente sus causas y sus leyes empíricas.

No conviene confundir la Estadística con las estadísticas: ya que estadísticas o series estadísticas, no son otra cosa que colecciones metódicas de datos numéricos, comunmente llamadas cementerio de cifras.

A. M. MOOD: Define la estadística como la tecnología del método científico que tiene como aspectos principales, los siguientes:

- 1) Realización de experimentos y observaciones,
- 2) Obtención de conclusiones o proposiciones objetivas, a partir de los resultados de dicho experimento y observación.
- 3) Formulación de leyes que simplifiquen la descripción de un - gran número de experiencias y observaciones.

CRAMER: Define por su parte la estadística, como la ciencia que se ocupa de obtener inferencias válidas a partir de los datos estadísticos y de la construcción de métodos para realizar dichas inferencias.

4. APLICACIONES DE LA ESTADISTICA

El campo de la investigación estadística es en la actualidad muy amplio, se puede decir que en todas las actividades humanas tiene aplicación; por lo tanto, el primer paso para sistematizar las informa-

ciones, será el de elegir la variable o la función a investigar; la función a investigar podrá ser en su forma más amplia por ejemplo: una población. Desde el punto de vista estadístico, ésta puede estar constituida por personas o por cosas, si dichas cosas son materia de investigación; por ejemplo, un lote de escritorios que tiene características comunes y no comunes y se quieren elegir unos, según su superficie; el total de escritorios se llamará población o conglomerado. Elegida la población se tiene el campo de actividad a investigar y bastará elegir las características fundamentales de dicho conglomerado, mediante el análisis previo a la investigación con una serie de preguntas sobre las características, clase y condición de los factores que se requieren para tratar de obtener resultados de esta experiencia; por lo tanto se habrá definido la magnitud de las observaciones necesarias para obtener resultados satisfactorios; dicha magnitud estará de acuerdo con las características del conglomerado si por ejemplo, la característica a investigar es la estatura de una serie de personas, y se ve que existe una cierta homogeneidad, se puede extraer un porcentaje de dicho grupo y analizarlo por separado; se obtendría un resultado y se generalizaría a todo el grupo. Esta selección tiene el in-

conveniente de asumir errores en su aplicación. Pero si observamos que el grupo a investigar posee por ejemplo, tres características en cuanto a su estatura predominante y éstas se presentan en la misma proporción, se efectuará un análisis mediante la observación detenida, de las características de tres grupos pequeños del mismo conglomerado; los resultados obtenidos se generalizan en la misma proporción para el conglomerado general. Este procedimiento se basa en la observación espontánea y sin relación científica; es un proceso de observación subjetiva que puede conllevar a graves errores en la conclusión de los hechos.

En medicina la Estadística ha encontrado un amplio campo de acción, ya que en los fenómenos biológicos, así como en las diferentes razas, poblaciones y las circunstancias ambientales de los individuos, se presentan aspectos que deben ser tratados por la Estadística; será necesario diferenciar peso, estatura o determinadas características cualitativas según la región, o por ejemplo, el número de glóbulos rojos, que es diferente según el sitio o región donde viva el individuo. Estas informaciones han servido de base como prevención y tratamiento de enfermedades y mejoramiento de drogas.

La Estadística se puede extender a todo el reino animal desarrollándose lo que se denomina BIOESTADISTICA.

En el campo de la Física se emplea al estudiar una densidad, un índice de refracción, que después de una serie de experimentos lleva al conocimiento de la verdad.

En la Química para experimentos tales como el análisis de una sustancia, índice de solubilidad, etc., la Estadística permite llegar al conocimiento definitivo de los resultados.

En aspectos Técnicos la aplicación de la Estadística no es menos importante, es común observar el empleo de métodos estadísticos en la Administración en general, en la Agricultura, en la Milicia, en la Astronomía, Control de Calidad, Demografía, Econometría, Meteorología, etc.

La Programación Lineal, es una de las más modernas aplicaciones de la Estadística, pues solamente hace diez años que en Estados Unidos se inició su estudio. Trata de resolver el problema que se presenta a todo empresario que ha de tomar decisiones, a fin de elegir el camino óptimo que maximice ciertas magnitudes deseables (ganancia, calidad, cantidad, etc.) y minimice otras no deseables (costos, pérdida de tiempo, etc.).

Matemáticamente es problema análogo el de calcular los máximos y mínimos de una función de varias variables, cuando estas variables están sujetas a limitaciones expresadas mediante inecuaciones.

El Profesor Cansado la define diciendo: Programación Lineal es - un problema que consiste en hallar el máximo (o el mínimo) de una función lineal de varias variables, cuando estas variables están sujetas a limitaciones expresadas mediante desigualdades o inecuaciones lineales. El adjetivo lineal, lo emplea en el sentido de que la función, o las desigualdades, son de primer grado respecto a todas las variables consideradas.

5. UTILIZACION DE LA ESTADISTICA

Existen dos tipos extremos de actitudes respecto a la estadística: una aceptación indiscriminada, que conduce al engaño, o una desconfianza sistemática, por la cual se ignoran conscientemente muchos hechos. Una actitud intermedia, en cambio, permite interpretar adecuadamente las cifras estadísticas. Pese a que no es necesario el conocimiento estadístico, éste ayuda bastante; muy importante son, un pensamiento claro y un buen uso del sentido común, con un enfoque analítico crítico.

Ningún procedimiento estadístico, por sí mismo, puede conducir directamente a resultados correctos; el camino está lleno de errores, de inexactitudes y de razonamientos falsos o incorrectos, y los procedimientos que se utilizan sólo son herramientas para llegar al fin buscado. Su adecuada utilización depende únicamente de la habilidad y experiencia de quien las emplea.

Las actividades de toda persona o institución están influenciadas, en mayor o menor grado, por decisiones basadas en antecedentes de tipo cuantitativo. Los modernos métodos estadísticos practicamente constituyen elementos indispensables en toda ciencia. El hombre de negocios utiliza encuestas estadísticas para determinar la reacción del público a sus nuevos productos; el astrónomo confecciona tablas estadísticas para determinar la posición actual y futura de los astros; el auditor de una empresa no pierde tiempo en examinar todos los elementos que ésta posee, sino que mediante métodos adecuados selecciona algunas muestras de los inventarios, máquinas de escribir o documentos en cobro; la semejanza entre los resultados observados y esperados en una experiencia genética se determina estadísticamente; el ingeniero industrial, que no puede examinar cada ampolleta o fósforo fabricados, sin

incurrir en la destrucción del producto, controla la cantidad mediante muestras tomadas al azar; el actuário determina las primas de seguro de vida o de incendio mediante tablas estadísticas de siniestros; el economista emplea una amplia gama de estadísticas para estudiar los planes de los consumidores y efectuar pronósticos de la tendencia de las actividades económicas; el gerente de una central eléctrica proporciona un buen servicio a la comunidad mediante la inspección de las variaciones estacionales de las necesidades de carga; el sociólogo trata de auscultar la opinión pública mediante encuestas, para determinar su preferencia por un candidato presidencial o su posición frente a determinados problemas económicos, políticos o sociales; los investigadores determinan la significación de ensayos agrícolas mediante consideraciones estadísticas; el biólogo aplica las leyes de Mendel esencialmente estadísticas; el geólogo también utiliza métodos estadísticos para determinar las edades de las rocas; el filólogo emplea otros métodos cuantitativos para determinar la época de separación de dialectos y lenguas del tronco común; el criterio literario considera la frecuencia del uso de diversas partes de la oración o de determinadas palabras o la longitud de las frases para estudiar el estilo de diversos autores, etc.

También puede anotarse como una utilización de la estadística las siguientes:

- 1) Descripción de grandes colecciones de datos empíricos, reduciéndolos a un pequeño número de características que concen tren la parte importante de la información suministrada por los datos. A esta finalidad se refiere la parte general de la Estadística que se conoce como "REDUCCION DE DATOS".
- 2) Análisis científico de datos experimentales y de los fenómenos observados. Esta parte de la estadística recibe el nombre de "ESTIMACION Y PRUEBA DE HIPOTESIS".
- 3) Predicción del futuro con base en las tendencias expresadas, por los fenómenos a través de datos preexistentes que consti tuyen una herramienta de mucha utilidad en los actuales mo - mentos de la vida moderna, en los cuales el análisis y estudio de tendencias es tan importante.

CAPITULO II

NOCIONES BASICAS DE LA ESTADISTICA

1. DEFINICIONES

1.1 Datos Estadísticos. Fundamentalmente se distinguen dos tipos de datos:

- a) Datos absolutos
- b) Datos relativos

Los datos absolutos son aquellos que resultan directamente del conteo de los elementos o características investigadas. Ejemplo 1: Número de habitantes del país, número de billetes emitidos por el Banco de la República, ingresos, etc.

Los datos relativos son los que se obtienen mediante la comparación de dos cifras absolutas entre sí. Ejemplo 2: El 49% de la población de un país son hombres, el 37% de la población económicamente activa gana menos de 800 unidades monetarias al mes, en un país X, el 12% de los trabajadores de una ciudad ganan en promedio 1.200 unidades monetarias.

1.2 Variables. Son las expresiones cuantitativas de las características objeto de estudio. Ejemplo 3: Ingreso, consumo, - número de personas en el hogar, número de perceptores de ingreso, etc. Las unidades en que se miden algunas características solo admiten ciertos valores dentro de un intervalo considerado, de tal manera, que la unidad de medida no admite subdivisión sin el riesgo a perder su naturaleza; son éstas las variables discretas. Ejemplo 4: Número de viviendas en una ciudad, habitantes de un país, número de fincas de un municipio, etc. Variable continua es aquella que puede tomar cualesquiera de los valores especificados dentro de un intervalo, en otras palabras son aquellas cuya unidad se puede subdividir sin que se pierda su naturaleza. Ejemplo 5: Ingreso, consumo, peso o estatura de las personas. etc.

Las variables continuas serán además derivables en un recorrido dado y se pueden definir sus funciones como una integral,

$$F(x) = \int_a^b f(x) dx$$

mientras que una variable discreta no es derivable y se define como una sumatoria.

$$F(x) = \sum_{x=a}^b f(x)$$

1.3 Universo o Población. Cualquier conjunto de elementos o individuos. Ejemplo 6: Al investigar "La Distribución de Ingresos de la Población" podríamos tomar como elemento, objeto de análisis, o sujeto de estudio, bien al hogar, o bien al individuo.

Para definir el conjunto que será objeto de análisis, se debe conocer el ámbito de la investigación (marco de la investigación), que viene a constituir el marco o espacio universal de todos los elementos, objeto de análisis.

Algunos autores distinguen entre universo y población de la siguiente manera:

Universo: Conjunto de elementos, personas, u objetos, que determinan la investigación.

Población: Conjunto de valores obtenidos midiendo o contando determinada característica de los elementos del universo; de tal manera que se pueden obtener varias poblaciones de un solo universo. Ejemplo 7: En el caso anterior del universo de hogares o personas podríamos obtener:

- Población de ingresos
- Población de gastos, etc.

1.4 Muestra. Resulta de tomar del conjunto universal un subconjunto, cuyos elementos van a ser objeto del estudio. Es así una parte del universo, o subconjunto de valores de una característica; obtenidos con el fin de investigar las propiedades de la población o conjunto de procedencia. Siempre se busca que la muestra sea representativa y que proporcione la información que interesa, sobre la población. Sobre este particular, que en la actualidad es un campo de amplia investigación, se tratará lo pertinente, con suficiente detenimiento en el desarrollo de técnicas e investigaciones por muestreo (otro volumen).

2. ETAPAS DE UNA INVESTIGACION ESTADISTICA

Para una investigación de tipo estadístico es necesario tener en cuenta las siguientes etapas:

2.1 Proyecto y Preparación

2.2 Recolección

2.3 Procesamiento

2.4 Exposición

2.5 Análisis e Interpretación

2.6 Publicación

2.1 Proyecto y Preparación. Al iniciar una investigación hay que definir exactamente el tema objeto de la investigación y de acuerdo a ello, preparar el plan de trabajo, para lo cual se debe tener en cuenta:

2.1.1 Definición de Objetivos:

- a) Generales: que vienen a ser el enunciado global de la investigación;
- b) Específicos: son propiamente el desarrollo de la investigación, o sea los capítulos que componen el estudio. Estos últimos van a condicionar el que la investigación sea de tipo explicativo, es decir, que trate de verificar o rechazar hipótesis. O bien, - que sea de tipo descriptivo, o sea, analizar cómo actúan una serie de elementos sobre una población determinada.

2.1.2 "Definición o Determinación de la Población Objeto de Estudio"; aquella a la cual se dirige la encuesta.

2.1.3 "Fuentes de Información" a las cuales se va a recurrir. Pueden ser: directas o indirectas. Directas, cuando se acude a obtener la información en el mismo fenómeno que se desea investigar; indirectamente cuando se va a tomar la información de investigaciones realizadas anteriormente o de publicaciones previas, sobre el tema investigado. Si son directas se debe definir si se va a realizar una investigación censal sobre la población o por el contrario, una investigación por muestreo.

2.1.4 Conformación Definitiva del Equipo que va a Realizar la Investigación. El jefe e inmediatos subalternos, así como el personal de campo o de implementación en la investigación.

2.1.5 Costo de la Investigación. Una vez cumplidas las etapas anteriores se pasa a traducir a unidades monetarias el trabajo del equipo y demás gastos que ocasione el estudio.

2.1.6 Propaganda. Se debe dar a conocer a los interesados (y personas que cubra la investigación, si es el caso) directa o indirectamente la importancia que tiene la investigación y los beneficios que se pueden lograr si - ésta se lleva a feliz término.

2.2 Recolección. Para la recolección es de suma importancia la elaboración del formulario; debe en términos generales, estar diseñado teniendo en cuenta entre otros los siguientes aspectos:

- a) Nombre de la entidad encargada de la investigación.
- b) Número del formulario.
- c) Nombre de la investigación.
- d) Cabeza del formulario.
- e) Cuerpo del formulario.
- f) Observaciones.
- g) Fecha.
- h) Firma del recolector responsable.
- i) Firma de la persona que suministra la información, etc.

NOMBRE DE LA EMPRESA

Investigación . . .

CONCEPTO	Ventas a Crédito		Ventas al Contado		Total de Ventas	
	No. Unidades	Valor	No. Unidades	Valor	No. Unidades	Valor

Firma y Fecha.

En la elaboración del formulario es de primordial importancia te
ner en cuenta que todos los datos incluídos sean posteriormente
objeto de estudio, es decir su elaboración se debe hacer de acuero
do a la lista de objetivos; las preguntas deben seguir una secuencia
lógica; el lenguaje técnico debe ser generalmente desechado y no
deben existir términos vagos, debe estar al nivel intelectual de -
quien suministra la información.

La recolección de la información conduce a una clasificación de los
datos de acuerdo a:

- a) Investigaciones que obedecen a datos preliminares o a datos definitivos.
- b) Representativas y no representativas.
- c) Ocasionales, periódicas o continuas.
- d) Incompletas o completas.
- e) Públicas y privadas.
- f) Universales, nacionales, seccionales, regionales y locales.

Sigue la selección e instrucción de entrevistadores y la entrevista propiamente dicha. El control de la entrevista es definitivo, sin descuidar naturalmente todas y cada una de las etapas, es una labor de equipo, en el que la falta de un elemento es generalmente irreparable.

2.3 Procesamiento de la Información; comprende:

- a) Codificación de cada pregunta y respuesta.
- b) Crítica; se determina si hay o no lógica en las diferentes respuestas del formulario diligenciado y entre los cruces o pruebas que se hayan programado de ellas (correlación de las respuestas).

c) Tabulación de los resultados; puede ser manual, mecánica o electrónica. Para estas dos últimas se utilizan tarjetas en las cuales se encuentran codificadas todas las respuestas para procesarlas mecánicamente o pasarlas por el computador y producir las tablas finales, o resultados finales debidamente tabulados.

2.4 Exposición. Consiste en la adecuada presentación de los datos obtenidos; se debe hacer en forma literal, tabular, gráfica o mediante una combinación de las anteriores.

2.5 Análisis e Interpretación. Es la última fase de la metodología y se puede dividir:

- a) Análisis y evaluación estadístico-matemática de los resultados.
- b) Análisis y evaluación técnica de acuerdo a la materia investigada.

Esto permitirá determinar el grado de consistencia y confianza de los diversos resultados de la investigación.

2.6 Publicación. Se debe seguir un plan previo según que tenga orientación:

- a) Técnica.
- b) Periodística (para el público en general).
- c) Mixta.

3. SERIES ESTADISTICAS

El estudio de un fenómeno conlleva a la determinación de éste en el tiempo y espacio; así se define como momento estadístico el período de tiempo al cual se refiere la investigación. En general los datos se pueden clasificar inicialmente de acuerdo a las características a que las series se refieran:

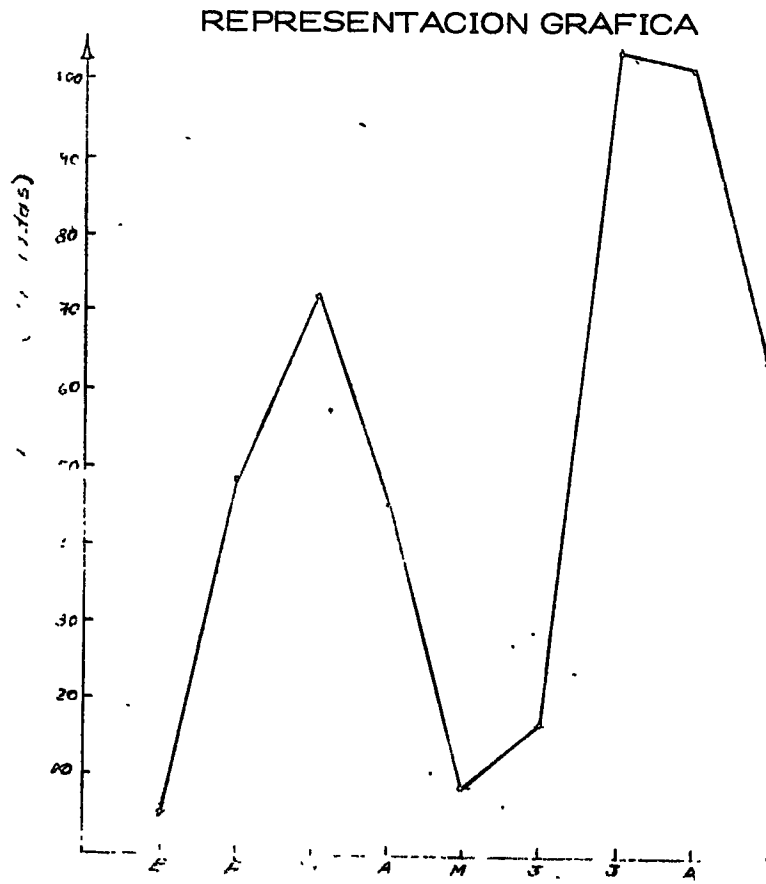
- Cuantitativas es decir medibles (contables); a su vez de acuerdo al tipo de variable puede ser discreta o continua. Ejemplo 8: Número de personas, viviendas, etc.; o estatura, ingreso, etc.
- Cualitativas referentes a cualidades y que en estadística son objeto del estudio de atributos. Ejemplo 9: Profesiones, marcas de carros, sexo, etc.

Serie estadística es una sucesión de datos sobre un fenómeno determinado, o mejor aún sobre determinada o determinadas características de los elementos objeto de estudio. Se acostumbra clasificar estas series así:

3.1 Series Cronológicas. Es una sucesión de datos que se refieren a una misma característica, considerada en un mismo espacio, pero en tiempos o épocas diferentes. Ejemplo 10:

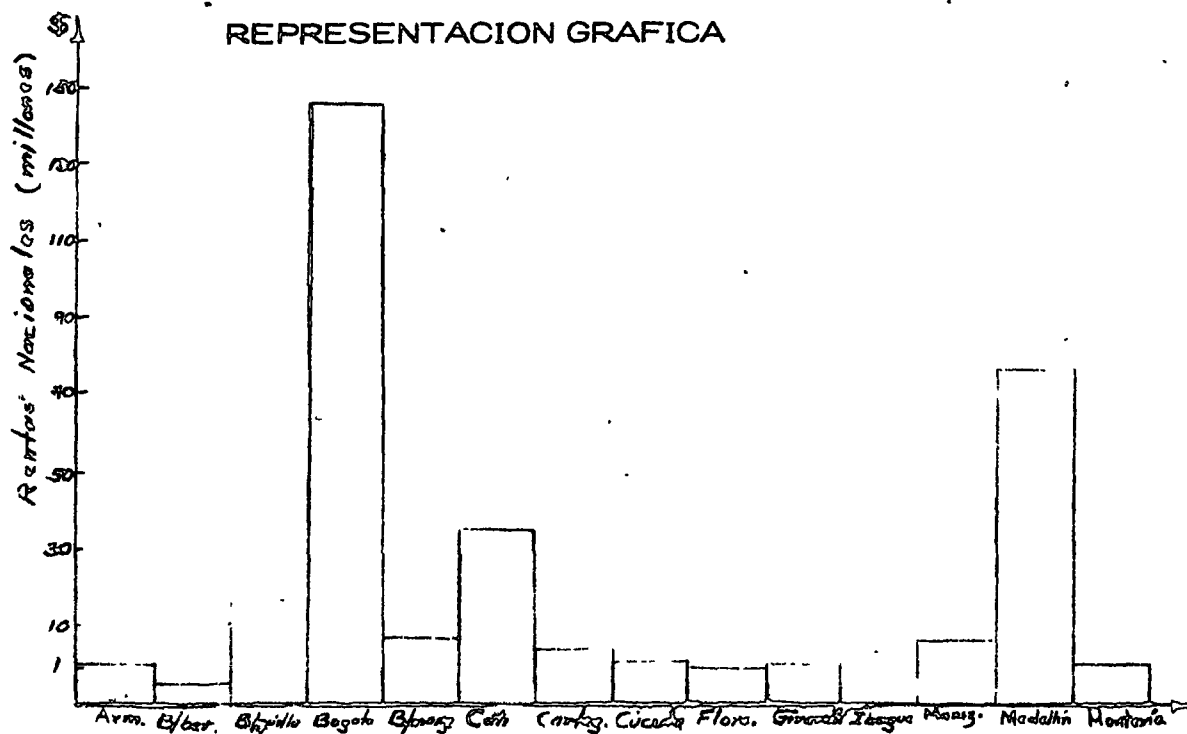
PRODUCCION DE SAL PARA 19X1

Enero	5.028	toneladas
Febrero	48.344	toneladas
Marzo	71.454	toneladas
Abril	45.450	toneladas
Mayo	8.807	toneladas
Junio	16.345	toneladas
Julio	103.708	toneladas
Agosto	101.400	toneladas
Septiembre	64.318	toneladas



3.2 Series Espaciales. Es una sucesión de datos que se refieren a una misma característica, considerada en un momento determinado del tiempo, pero clasificada de acuerdo a su situación en el espacio (o área geográfica). Ejemplo 11: Rentas Nacionales. Recaudos por Administración en las diferentes ciudades en el año 19X2.

ADMINISTRACIONES	AÑO 19X2
Armenia	1.475.367
Barrancabermeja	449.623
Barranquilla	13.613.529
Bogotá	145.900.141
Bucaramanga	6.875.800
Cali	33.727.638
Cartagena	4.441.418
Cúcuta	1.873.144
Florencia	1.163.825
Girardot	1.233.063
Ibagué	3.591.175
Manizales	5.323.904
Medellín	75.583.337
Montería	1.053.093

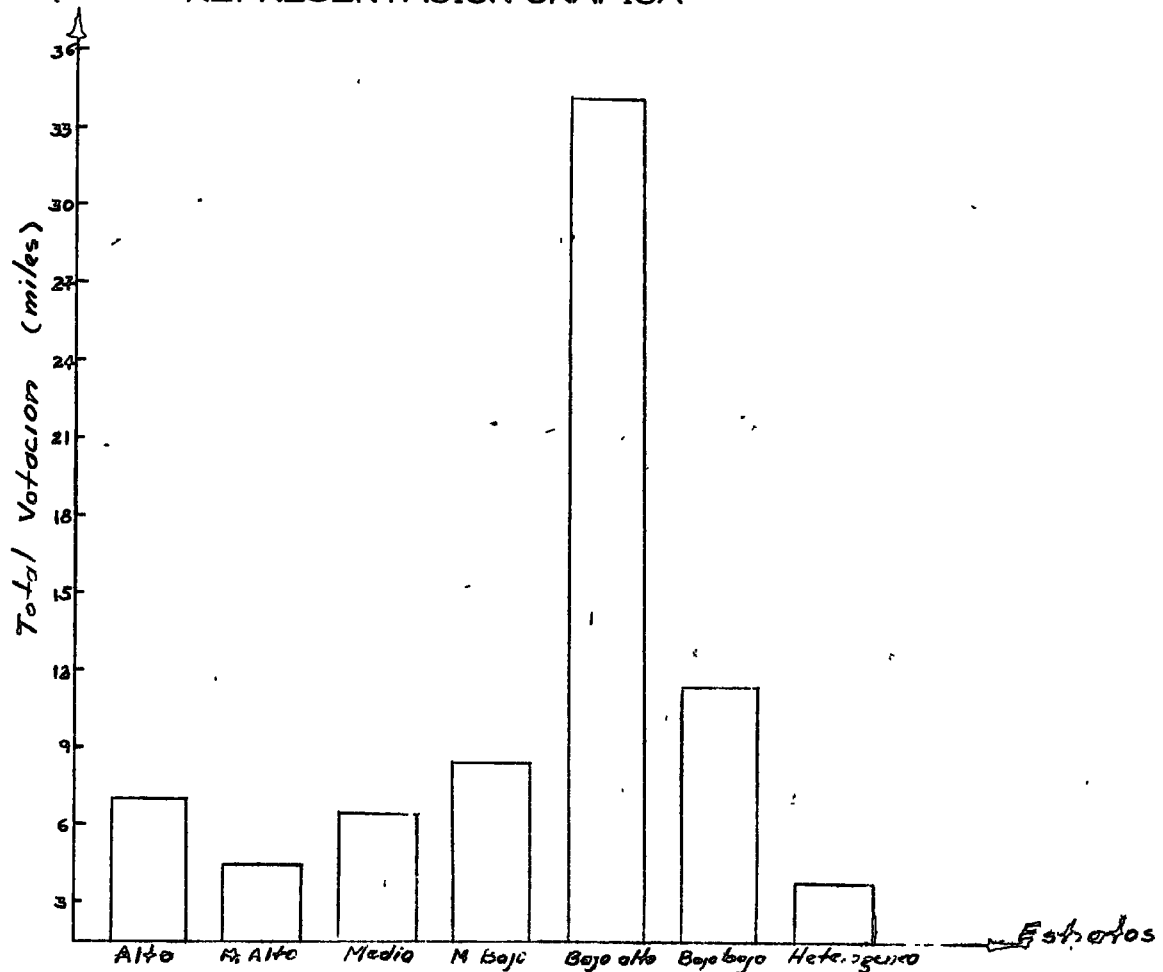


Cuando se hacen comparaciones de dos series espaciales, particularmente las geográficas en diferentes épocas o tiempos, se deben tener en cuenta los cambios de extensión que hayan podido tener las unidades geográficas a través del tiempo; de lo contrario lo único que se puede obtener es una tergiversación de la realidad.

3.3 Series Cualitativas. Una sucesión de datos en un momento de terminado del tiempo y del espacio, clasificada de acuerdo a diferentes características objeto de estudio. Ejemplo 12: Votación estratificada para 19X2 en la ciudad de Cali (Colombia).

ESTRATO	TOTAL VOTACION
Alto	7.018
Medio Alto	4.867
Medio	6.292
Medio Bajo	8.639
Bajo Alto	34.502
Bajo Bajo	11.802
Heterogéneo	3.400

REPRESENTACION GRAFICA



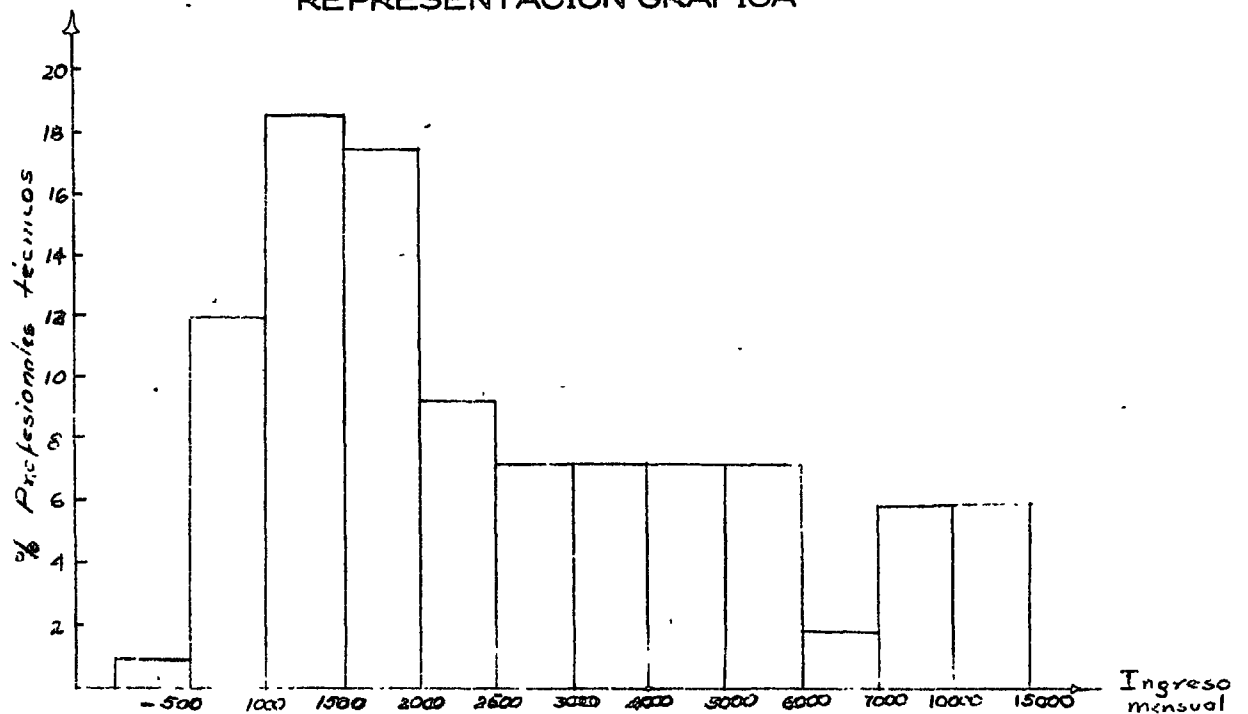
3.4 Serie Cuantitativa. Es una sucesión de datos estadísticos, referentes a una misma característica considerada en un mismo tiempo y espacio y clasificada de acuerdo al tamaño de la característica en mención. Ejemplo 13:

INGRESO MENSUAL

% PROFESIONALES TECNICOS

Menos de 500	.92
De 501 a 1.000	11.93
De 1.001 a 1.500	18.35
De 1.501 a 2.000	17.43
De 2.001 a 2.500	9.17
De 2.501 a 3.000	7.34
De 3.001 a 4.000	7.34
De 4.001 a 5.000	7.34
De 5.001 a 6.000	7.34
De 6.001 a 7.000	1.84
De 7.001 a 10.000	5.50
De 10.001 a 15.000	5.50
Más de 15.000	0.00

REPRESENTACION GRAFICA



4. TIPOS DE GRAFICOS

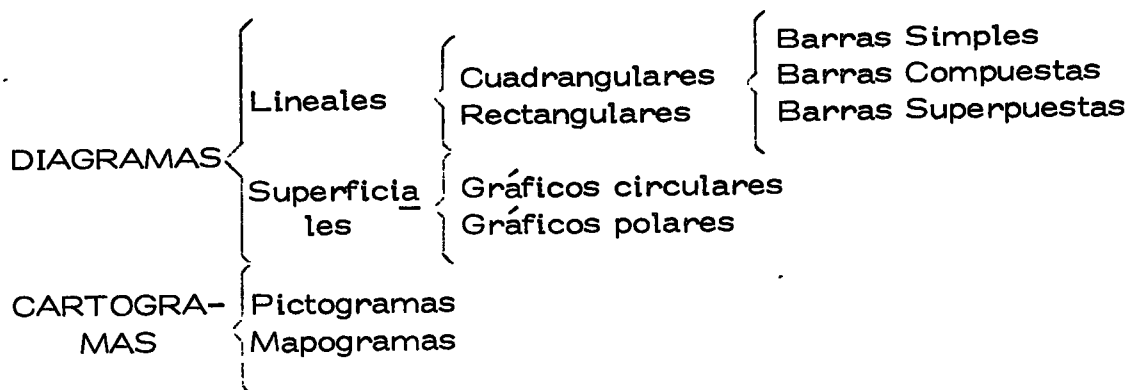
Como complemento del tipo de agrupación suele emplearse una serie de modelos de tipo gráfico que sirvan de complemento al análisis propiamente numérico ya que la gráfica en forma sencilla, puede dar idea del comportamiento de las observaciones originales. Para cada caso en particular se tendrá un modelo gráfico distinto pero que se acomode a la realidad del fenómeno.

Las más comunes se construyen teniendo como base los ejes de L coordenadas cartesianas rectangulares en la cual podemos ubicar sobre el eje de las abscisas los diferentes valores que tome la función analizada o lo que es lo mismo, los diferentes valores que tome la variable.

En la ordenada se anota el número de veces que dichos valores se presentan, por lo tanto para cada punto en el plano les debe corrésponder dos valores: uno relativo a la variable y otro al número de veces que se repite dicha variable.

Esta gráfica puede ser del tipo de barras o rectángulos sucesivos o separados, de línea continua, discontinua, o las modalidades que

pueda surgir de los modelos anteriormente descritos; cuando por circunstancias especiales los valores de las observaciones se incrementen en una forma de tipo geométrico, la gráfica más aconsejable para este fenómeno sería la que se denominaría "Semilogarítmica", que consiste básicamente en medir en unidades naturales los valores de la variable y en función de logaritmos los valores de las funciones, es decir, que tendríamos una escala logarítmica en el eje de las ordenadas. Si los dos valores de las dos coordenadas fluctúan en función geométrica, su representación gráfica más aconsejable será la de tipo logarítmico, en el cual tanto las abscisas como las ordenadas se medirán en escala logarítmica. Otro tipo de gráfica es el de coordenadas polares en el cual la importancia del fenómeno se mide bajo dos conceptos: un segmento rectilíneo y un determinado ángulo.



4.1 Explicación de la Gráfica Circular

Ejemplo 14: Ingresos totales de:

Cine: \$250.000 Toros: \$85.000

Foot-Ball: \$175.000 Otros: \$20.000

$$\text{Superficie circunferencia} = S = \pi r^2 \quad r = \sqrt{\frac{S}{\pi}}$$

S = Suma de todos los Ingresos

$$S = 530.000 \quad r = \sqrt{\frac{530.000}{3,1416}} \quad r = 168.704$$

$$r = 410.74$$

Escala: $\frac{1}{150}$ Por cada cms. se emplean 50 unidades.

$$\frac{410,74}{150} = 2,74 \text{ cms. de radio.}$$

$$\text{Ahora: } 360^\circ = 530.000 \quad \angle X = \frac{250.000 \times 360}{530.000} = 169^\circ 48' 41''$$
$$X = 250.000$$

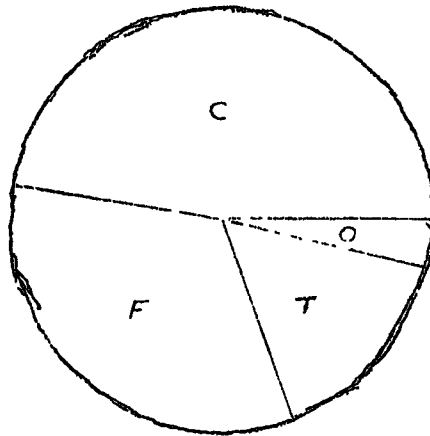
Esta gráfica se usa para explicar fenómenos agropecuarios,
financieros, producción industrial, etc.

./,

$$\frac{530.000}{175.000} \times \frac{360^\circ}{X} = \frac{175.000 \times 360}{530.000} = \frac{175 \times 36}{53} = 118^\circ 52' 4''$$

$$\frac{530.000}{85.000} \times \frac{360^\circ}{X} = \frac{85.000 \times 360}{530.000} = \frac{85 \times 36}{53} = 57^\circ 44' 3''$$

$$\frac{530.000}{20.000} \times \frac{360^\circ}{X} = \frac{20.000 \times 360}{530.000} = \frac{20 \times 36}{53} = 13^\circ 35' 3''$$



4.2 Gráfica Polar. Se usa para representar fenómenos que presentan máximos y mínimos. Ejemplo 15:

Lunes	150;	Miércoles	175;	Viernes	180
Martes	170;	Jueves	155;	Sábado	130

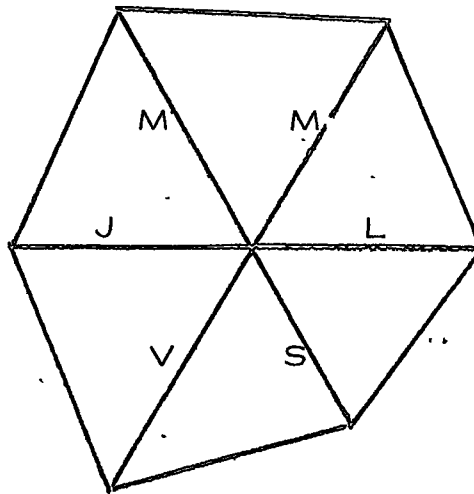
Como hay 6 normas debemos dividir la circunferencia en 6 - partes iguales: $360^\circ \div 6 = 60^\circ$

Convenciones

Esc. 1/25 1 cm. por cada 25 unidades

$$L = \frac{150}{25} = 6 \text{ cm.} \quad M = \frac{170}{25} = 6,8 \text{ cm.} \quad M = \frac{175}{25} = 7 \text{ cm.}$$

$$J = \frac{155}{25} = 6,2 \text{ cm.} \quad V = \frac{180}{25} = 7,2 \text{ cm.} \quad S = \frac{130}{25} = 5,2 \text{ cm.}$$



Nota: La escala se redujó a 1/50.

4.3 Gráfico Cuadrangular. Es utilizado para representar fenómenos que hacen relación a áreas o superficies. Ejemplo 16:

Area Construida.

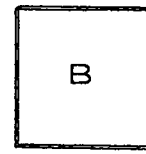
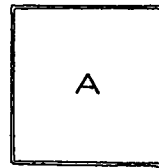
Ciudad A = 10.000 Mt^2 Superficie del cuadrado = $S=L^2$

Ciudad B = 8.000 Mt^2 $L = \sqrt{\frac{A}{S}}$ A = 10.000 Mt^2

Ciudad C = 5.000 Mt^2 $L = \sqrt{10.000 \text{ } Mt^2}$ $L = 100 \text{ Mt}$

Convenciones

Esc. 1/50



$$A = \frac{100 \text{ Mt}}{50} = 2 \text{ cm.} = \text{lado del gráfico A}$$

$$B = 8.000 \text{ mt}^2 \quad L = \sqrt{8.000 \text{ Mt}^2} = 89.4 \text{ Mt.}$$

$$B = \frac{89.4 \text{ mt}}{50} = 1,8 \text{ cm.} = \text{lado del gráfico B ; etc.}$$

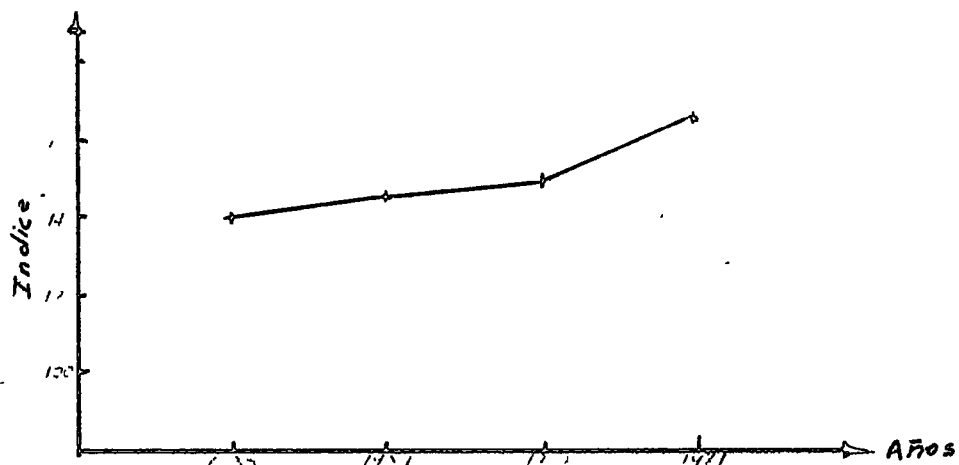
4.4 Curva Estadística: Ejemplo 17: Índice Producción Nacional de un Artículo.

19X8	:	140
19X9	:	145
19X0	:	148
19X1	:	165

Convenciones

Esc. 1/2.

NOTA: Si la escala es 1/2 se va tomando de 2 en 2 cms., para cada 20 Unidades. Ejemplo 18: Empieza en 100 luego 120 - 140 etc.

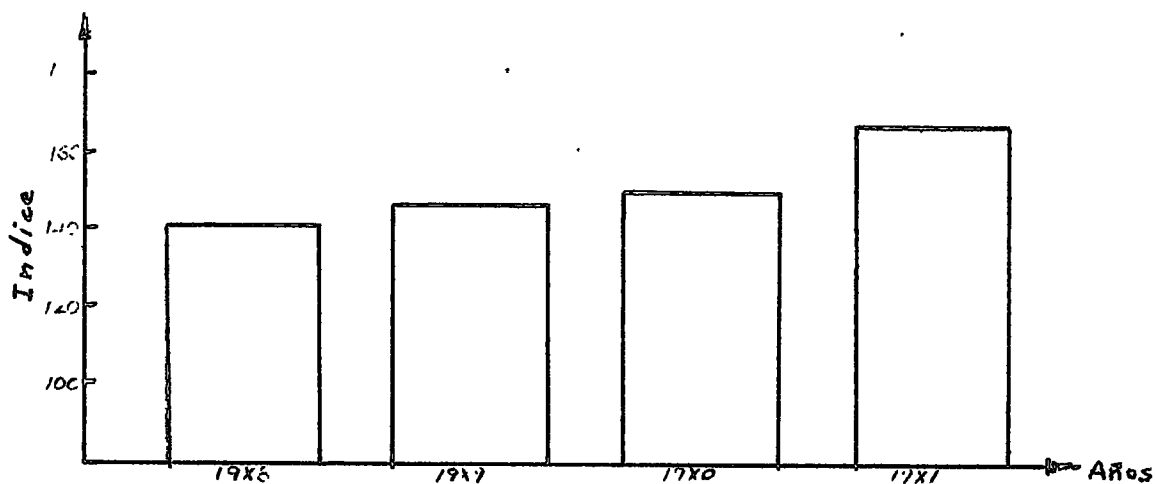


4.5 Gráficas de Barras. Son de 3 tipos:

- a. Simples
- b. Compuestas
- c. Superpuestas

a) Simple: Tomando los datos de la curva estadística. Ejemplo

19:



NOTA:

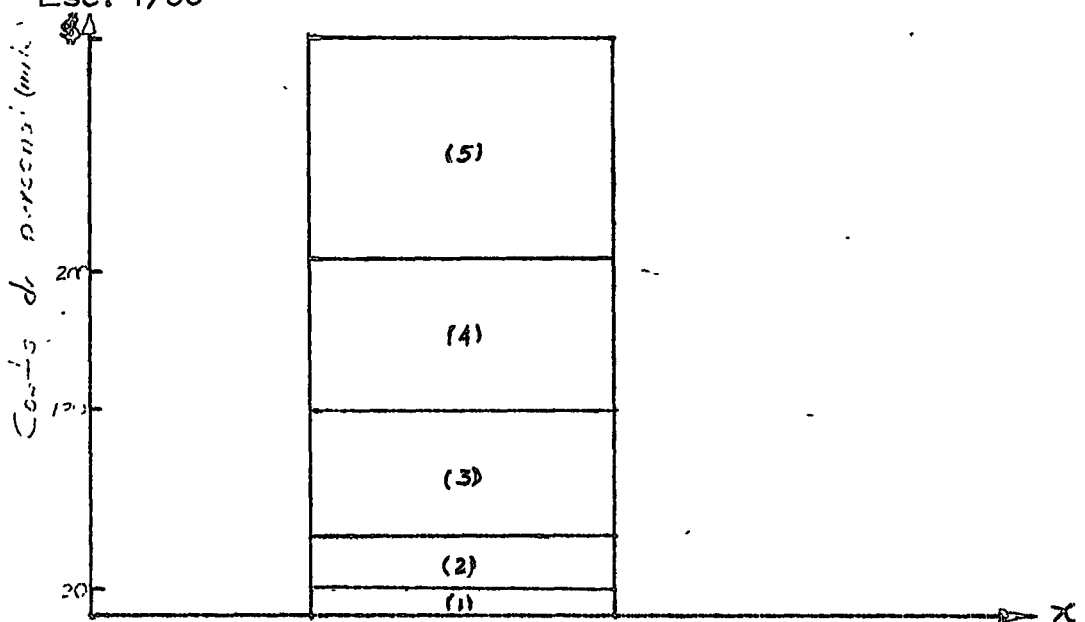
Entre barra y barra se deja la mitad del ancho de la barra, en general aunque no es norma fija.

b) Compuestas. Ejemplo 20: Costos de personal de una Empresa en (Miles \$).

Sueldos	\$150	(5)
Horas extras	93	(4)
Feriados	75	(3)
35% reserva	20	(1)
Vacaciones	35	(2)
	\$375	

Convenciones

Esc. 1/50



Se ordenan los datos en forma ascendente y se representa el (1)

Luego para representar el (2) se toma a partir del (1) $20+35=55$

Luego para representar el (3) $= 20 + 35 + 75 = 130$

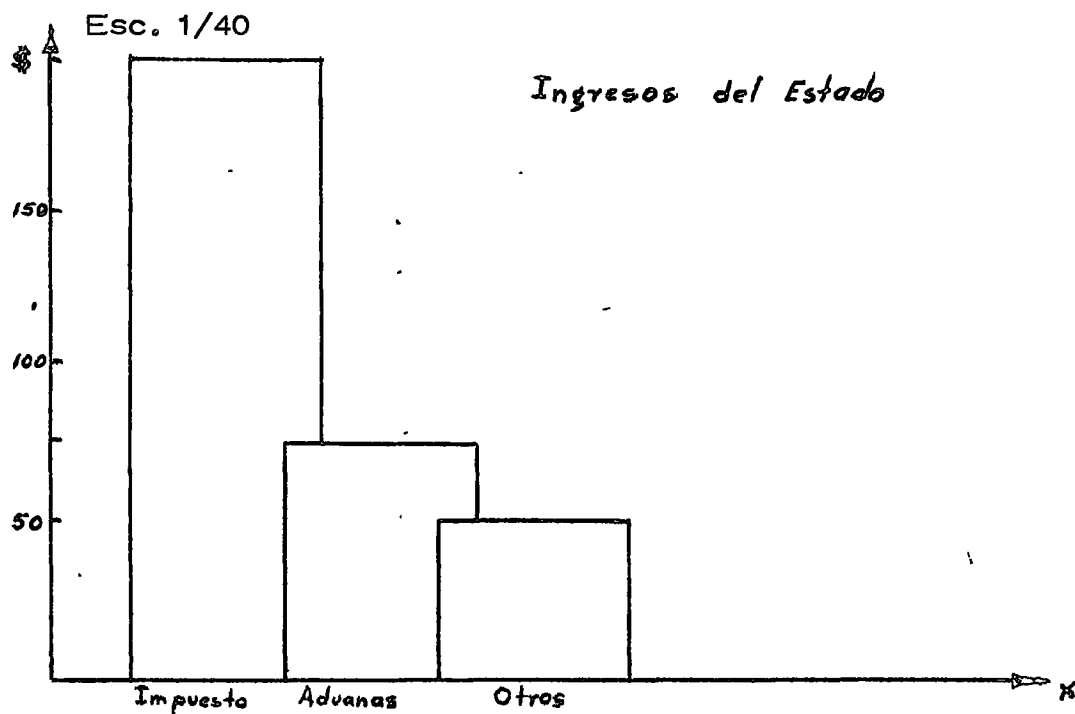
Luego para representar el (4) $= 20 + 35 + 75 + 93 = 223$

Luego para representar el (5) $= 20 + 35 + 75 + 93 + 150 = 375$

c) Superpuestas. Ejemplo 21: Ingresos del Estado.

(1)	Impuesto a la renta	\$200
(2)	Aduanas	75
(3)	Otros	50
	TOTAL	<u>\$325</u>

Convenciones



Se toma el mayor de los sumandos (200), en primer lugar y a continuación en forma descendente las restantes (75 y 50)

De estas gráficas descritas han surgido otras complementarias, que en esencia se reducen a combinación de dos o más de las anteriores. Suele por ejemplo: presentarse una serie de fenómenos; mediante una serie de círculos en los cuales la importancia de los fenómenos radica en el área de cada círculo.

En estadísticas que tienen relación con población suelen emplearse siluetas humanas; estas siluetas pueden ser completas o partidas, según lo que representen.

En estadísticas agropecuarias suelen representarse los fenómenos, mediante animales u objetos agropecuarios especiales. Estas clases de gráficas tienden más, a una importancia psicológica que real aunque naturalmente dejan en el observador un claro concepto, sobre lo que trata de describir la gráfica.

5. CUADROS ESTADISTICOS

Contienen una serie de datos estadísticos dispuestos en forma de fi las y columnas, con el fin de facilitar los cálculos y análisis.

5.1 Partes de que consta. En un cuadro o tabla estadística se debe distinguir:

- a) TITULO: Frase que designe el fenómeno colectivo, o aspectos de el que se describe en la tabla. Deberá responder a las preguntas: Qué?, dónde?, cuándo?.
- b) CABEZAS: Partes del cuadro que designan el contenido de las columnas.
- c) COLUMNAS: Partes del cuadro que contienen una serie vertical de datos.
- d) COLUMNA PRINCIPAL: Parte del cuadro que designa el contenido de las líneas.
- e) LINEA: Parte que contiene la serie de datos dispuesta en forma horizontal.
- f) CASILLAS: Parte del cuadro que contiene un número o un signo.
- g) NOTAS U OBSERVACIONES: Se utiliza para explicar el origen de la serie, para aclarar cualquier aspecto que se presente obscuro en la tabla.

Ejemplo 22:

INDICE NACIONAL DE PRECIOS AL CONSUMIDOR (1)←(Título)

(Julio 1954 a Junio 1955 = 100)

Período	Empleados					Obreros (Cabezas)				
	Alimen- tos	Vivien- da	Vestua- rio	Misce- láneo	Total	Alimen- tos	Vivien- da	Vestua- rio	Misce- láneo	Total
(Columna Principal)	Columna)									
19X7 Promedio	382.5	305.8	331.5	388.5	359.6	388.1	327.3	338.8	377.0	367.3
19X8 Promedio	409.6	327.9	353.7	418.5	386.8	416.1	353.2	361.0	407.0	394.5
19X9 Promedio	433.6	352.4	382.2	452.8	414.0	440.3	383.1	388.2	442.1	421.9
19X0 Promedio	459.3	379.9	419.5	489.9	444.4	465.0	413.6	424.5	473.5	450.2
19X1 Promedio	517.1	415.0	465.3	539.4	492.8	523.7	462.5	468.6	523.6	503.4
19X9 Marzo	417.5	346.0	369.1	446.0	402.9	420.5	376.4	376.1	437.6	408.7
Junio	433.6	351.3	381.8	451.2	413.3	442.8	381.6	387.8	440.7	422.5
Septiembre	440.4	357.0	390.2	457.3	420.0	447.3	388.5	395.6	446.1	428.1
Diciembre	452.3	363.9	399.3	463.4	428.9	459.8	394.0	404.9	450.2	437.1
19X0 Marzo	448.3	372.7	406.8	484.8	435.3	451.2	404.9	411.6	470.0	439.6
Junio	463.1	378.5	419.6	490.1	445.7	473.9	410.4	424.4	473.2	453.9
Septiembre	463.1	386.2	425.6	494.7	449.3	468.0	421.9	430.2	478.4	455.0
Diciembre	475.1	391.2	442.3	505.5	459.9	476.5	430.7	446.6	487.4	464.5
19X1 Febrero	485.9	398.8	449.3	520.5	470.1	486.4	441.6	453.8	509.1	476.3
Marzo	491.9	403.8	451.1	528.1	475.4	491.3	444.7	454.8	516.3	480.9
Abril	504.3	405.2	456.3	530.2	481.6	509.8	446.7	459.4	518.1	490.7
Mayo	509.6	407.9	462.2	535.1	486.3	518.4	452.6	464.5	519.0	497.1
Junio	513.8	409.4	465.5	536.5	489.2	521.9	454.6	468.1	520.4	499.8
Julio	522.4	417.0	466.6	541.3	495.8	528.9	465.9	469.5	523.9	506.7
Agosto	531.1	418.3	469.1	544.1	500.7	537.1	468.5	472.0	526.3	512.3
Septiembre	531.5	425.6	470.9	548.4	504.2	539.3	477.9	473.1	532.7	516.8
Octubre	539.7	427.3	476.7	556.2	510.4	550.9	479.1	479.9	538.3	524.5
Noviembre	545.7	432.7	483.6	558.6	515.5	557.4	486.9	487.0	539.2	530.3
Diciembre	546.6	436.5	488.7	561.1	517.9	558.0	492.9	493.1	541.7	532.9
19X2 Enero	555.4	447.0	489.4	565.1	525.0	564.6	499.9	494.4	542.7	538.1
Febrero	559.5	448.0	498.6	579.8	531.2	568.5	502.0	504.3	558.4	544.2
Marzo	562.3	456.8	500.2	592.1	537.3	571.1	508.2	505.7	570.3	549.1
Abril	575.9	457.7	505.7	596.0	544.5	585.7	509.1	512.3	574.0	557.7
Mayo	576.5	463.6	512.9	596.8	547.6	590.2	517.7	518.6	574.7	562.8
Junio	584.4	466.5	521.4	599.2	552.8	598.3	519.7	531.9	577.7	569.0
Julio	591.7	473.6	523.5	605.0	559.3	605.4	526.9	534.2	583.6	575.6

(1) Departamento Administrativo Nacional de Estadística. Para conocer el detalle por ciudades, véase el cuadro 9.5.1.

(Nota u observación).

5.2 Clases de Cuadros.

1) SIMPLES

- 2) DOBLE ENTRADA a) Tabla de contingencia
 b) Tabla de correlación
 c) Tablas mixtas

3) COMPLEJOS

5.2.1 Los SIMPLES expresan una sola serie, ésta puede ser cronológica, espacial, cualitativa o cuantitativa; tal como se definieron en el párrafo anterior. Ejemplo 23:

DEPOSITOS DE AHORRO

AÑOS	BANCO BOGOTA
19X7	175.267
19X8	203.608
19X9	231.820
19X0	281.166
19X1	316.150

FUENTE: Revista Banco de la República.

5.2.2 DOBLE ENTRADA: Expresan dos series. Son los utilizados en estadística bidimensional ya que facilitan el estudio de 2 variables en un plano.

a) Tabla de Contingencia: Contienen una serie cualitativa tanto por las cabezas como por la columna principal.

Ejemplo 24: Importancia de los Grupos de Productos dentro de las Importaciones Extra Comunitarias.

Productos Países	Alimen tación	Energé ticos	Materia Prima	Producto Químico
BELGICA Y LUXEMBURGO	1.418	2.395	1.875	1.474
FRANCIA	1.495	2.214	1.414	2.388
HOLANDA	1.114	970	1.243	1.676

FUENTE: Revista Banco República.

b) Tabla de Correlación: Serie cuantitativa por la cabeza y la columna principal. Ejemplo 25: Estatura y Edad de 36 Personas.

Edad \ Altura	1	2	3	4
	1	3	2	1
50	1	3	2	1
60	3	1	2	4
70	1	4	2	3
80	0	1	3	5

FUENTE: DANE.

- c) Tablas Mixtas: Contiene una serie cualitativa por la cabeza y una cuantitativa por la columna principal o viceversa. Ejemplo 26: Índice Nacional de Precios al Consumidor.

Años \ Productos	Alimentos	Vivienda	Vestuario	Misceláneos
	382.5	305.8	331.5	388.5
19X7	409.6	327.9	353.7	418.5
19X8	433.6	352.4	382.2	452.8
19X9	459.3	379.9	419.5	489.9
19X0	517.1	415.0	465.3	539.4
19X1				

FUENTE: DANE.

5.2.3 COMPLEJOS: Expresan más de dos variables, comunmente se utilizan en estadística tridimensional y pluridimensional; en forma parcial relacionando dos variables.

El estudio de ellas se verá en forma detallada en el segundo tomo.

CAPITULO III

ESTADISTICA DE ATRIBUTOS

TABLA DE EQUIVALENCIAS

S	: Conjunto universal
$\subset$	: Contenido en
$\supset$	: Contiene a
$\in$	: Pertenece o es un elemento
$=$	: Igualdad
ϕ	: Conjunto vacío o nulo
$\forall$	: Para todo
$A \wedge B$	: Dos conjuntos cualquiera
$\bar{A}$ o $\bar{B}$	: Suceso contrario o complemento de A o de B.
$A \cup B$ o $A + B$	: Unión o suma de dos conjuntos
$A \cap B$ o A , B	: Intersección
$\Rightarrow$	: Implicación
$\Leftrightarrow$	: Equivalencia.

1. NOCIONES SOBRE TEORIA DE CONJUNTOS

El término CONJUNTO es comunmente empleado en Matemáticas para designar como tal a la colección o reunión de objetos que -

tengan ciertas características especiales. Cada uno de los objetos que componen el conjunto se denomina ELEMENTO

Un conjunto puede estar compuesto por una colección de números, puntos de una recta, personas nacidas en una determinada región, etc. Cuando el número de elementos puede contarse o medirse se denomina conjunto NUMERABLE o FINITO.

Ejemplo 1: El conjunto de las edades de los miembros de un club.
El conjunto de los Triángulos Isósceles que se puedan formar en un plano.

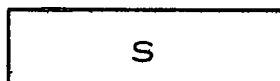
Cuando el número de elementos no pueda contarse se denomina conjunto INNUMERABLE o INFINITO.

Ejemplo 2: El conjunto de los números naturales. El conjunto de las personas vivas que nacieron en el año 1.300.

1.1 Definiciones. El conjunto que contiene todos los posibles elementos o resultados de un experimento, objeto de discusión, se denomina conjunto UNIVERSAL o ESPACIO. En general suele designarse por S , y se representa comunmente por medio de diagramas de Venn-Euler o de Venn, que representan

un conjunto con un área plana, por lo general delimitada por uno o varios círculos que expresan la relación existente entre varios eventos.

Ejemplo 3: El conjunto (S) de los establecimientos de Industria existentes en Colombia en 1970. Puede representarse así:



Todo conjunto es un subconjunto del conjunto universal, siempre que éste haya sido definido perfectamente.

Si dos conjuntos A y B son tales que todo elemento del conjunto B también es un elemento del conjunto A, se puede afirmar que "B está contenido en A" ($B \subset A$), lo que es equivalente a decir que "A contiene a B" ($A \supset B$). En forma similar, si X es un elemento del conjunto A podremos decir que "X pertenece a A" ($X \in A$).

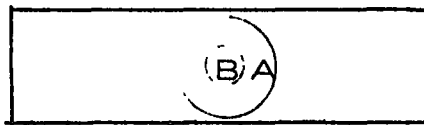
Ejemplo 4: Sean:

$$A = \{1, 3, 5, 6, 7, 8\}$$
$$B = \{8, 3, 6, 5\}$$

Se puede afirmar que $B \subseteq A$ ya que todo elemento del conjunto B también pertenece al conjunto A.

En igual forma se puede decir que "7 es un elemento del conjunto A" o también, $7 \in A$.

* La representación mediante un diagrama de EULER sería:



Si todo elemento del conjunto A pertenece también al conjunto B y viceversa, esto es, todo elemento del conjunto B pertenece también al conjunto A, se dice que los conjuntos A y B son iguales ($A = B$).

Ejemplo 5: Sean:

$$A = \{ \triangle, *, \dots \}$$

$$B = \{ *, \dots, \triangle \}$$

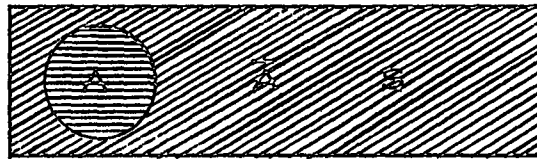
Puede decirse que los conjuntos A y B son IGUALES o IDENTICOS.

Cuando un conjunto no tiene elementos se dice que es un conjunto VACIO o NULO ($\emptyset$). El vacío es un subconjunto de todo conjunto finito. Todo conjunto A está asociado con otro conjunto $\bar{A}$ (no A) llamado complemento del primero, el cual tiene todos los elementos del conjunto universal S que no tienen la propiedad A. También suele llamarse suceso CONTRARIO u opuesto.

Ejemplo 6: Sea A el conjunto de los establecimientos de Industria con 5 personas ocupadas existentes en Colombia en 1970. Entonces $\bar{A}$ (no A) sería el conjunto de los establecimientos de Industria que tienen un número de personas ocupadas diferente de 5, con lo cual, se puede afirmar que:

$$A + \bar{A} = S.$$

Representando gráficamente equivale a:



1.2 Operaciones con Conjuntos. Dados 2 conjuntos A y B se pueden definir algunas operaciones con estos conjuntos:

La ADICION de 2 conjuntos A y B de un conjunto universal S dan origen al conjunto de los elementos que pertenecen a A, B o ambos. La notación utilizada es:

$$A \cup B \text{ o } A + B$$

Ejemplo 7: Sean:

$$A = \{1, 2, 3, 8, 10, 14\}$$

$$B = \{8, 3, 14, 4, 5, 9\}$$

$$A \cup B = \{1, 2, 3, 4, 5, 8, 9, 10, 14\}$$

Gráficamente tendríamos:



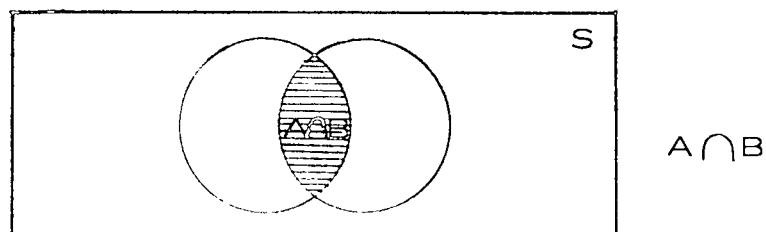
Cuando los elementos son comunes a ambos conjuntos A y B se interpreta como la INTERSECCION o parte común. La simbología empleada es:

$$A \cap B \text{ o } A \cdot B$$

Ejemplo 8: Sean A y B como en el ejemplo 7, entonces:

$$A \cap B = \{8, 3, 14\}$$

Su representación gráfica sería:



Si se generalizan las definiciones anteriores a una sección de conjuntos $A_1, A_2, A_3, \dots$ del conjunto universal, se puede definir la ADICION como:

$$\sum_{i=1}^{\infty} A_i = A_1 + A_2 + A_3 + \dots = \bigcup_{i=1}^{\infty} A_i$$

Es decir: la adición es el conjunto de los elementos pertenecientes al menos a uno de los conjuntos A_i . La adición también es conocida como SUCESO TOTAL o UNIVERSO. Se tiene definida la intersección o producto análogamente:

$$\prod_{i=1}^{\infty} A_i = A_1 \cdot A_2 \cdot A_3 \cdot \dots = \bigcap_{i=1}^{\infty} A_i$$

como el subconjunto de los elementos que pertenecen a todos los A_i . También se conoce como SUCESO COMPUESTO. Algunas veces puede presentarse el caso en el que A y B no tengan ele -

mentos comunes, con lo cual su producto es nulo; es decir:

$$A \cap B = 0 \text{ ó } A \cdot B = 0$$

También puede suceder que los conjuntos A y B sean iguales ($A = B$), en cuyo caso, la diferencia entre ambos será nula.

$$A - B = 0 \text{ ó } B - A = 0$$

Cuando B sea un subconjunto de A se presentan 2 situaciones diferentes:

$$\text{a) } A \cup B = A \quad \text{y} \quad \text{b) } A \cap B = B$$

1.3 Propiedades e Implicaciones. Los conjuntos pueden tratarse como operadores lineales y, como tales, cumplen algunas de las propiedades para la suma y el producto que es necesario enunciar:

a) Conmutatividad:

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

b) Asociatividad:

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(A \cap B) \cap C = A \cap (B \cap C)$$

c) Distributividad : $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Si la existencia de una propiedad (A) conlleva la existencia de otra (B) - se dice que la propiedad (A) IMPLICA la propiedad (B) dentro del mismo - conjunto universal S.

La simbología utilizada sería :

$$A \Rightarrow B$$

Ejemplo 9 : Sean :

A = Propiedad de ser economista

B = Propiedad de saber leer.

Por lo tanto la expresión " A implica B " o " Si A entonces B " ($A \Rightarrow B$) nos indicará que si una persona es economista entonces sabe leer, siendo S, el conjunto de los profesionales colombianos.

Este tipo de asociación es el que se conoce con el nombre de IMPLICACION.

En general podremos afirmar que las implicaciones son proposiciones verdaderas para todos los valores de las variables que intervienen.

La expresión:

$$\forall x (Px \implies Qx)$$

debe leerse:

Para todos los x si Px , entonces Qx ; o para todo x , Px impli
ca Qx .

Esta proposición es verdadera si y solamente si $Q \subset P$ donde Q
y P son los verdaderos conjuntos de Qx y Px respectivamente.

Ejemplo 10:

$$P = \{ x/x \text{ es impar} \}$$
$$Q = \{ x/x^2 \text{ es impar} \}$$

Puede asegurarse que $Px \implies Qx$.

Si del conjunto universal definimos los conjuntos A y B de tal man
nera que $A \implies B$, pero a su vez, $B \implies A$, se concluye que se ne
cesita la ocurrencia de ambos para obtener un suceso total. La
simbología utilizada es:

$$A \iff B$$

Ejemplo 11: Tenemos la siguiente implicación:

$$a) \quad \log \left(\frac{a}{b} \right) = \log a - \log b$$

$$b) \log a - \log b = \log \left(\frac{a}{b} \right)$$

$$c) \log \left(\frac{a}{b} \right) \Longleftrightarrow \log a - \log b.$$

2. CLASIFICACION POR ATRIBUTOS

La observación de un determinado fenómeno, ya sea físico, químico, económico, social, etc. es susceptible de una cuantificación y frecuentemente el observador o analizador se ve enfrentado con problemas de tipo cuantitativo si las magnitudes se expresan en datos, o también de tipo cualitativo si la persona que está interesada en el fenómeno se dedica a contar la presencia o la no presencia (ausencia) de una determinada cualidad, es decir, que se encuentra con 2 tipos de estadísticas: la primera se asocia con la Estadística de VARIABLES y la segunda que trata de la estadística de ATRIBUTOS.

Cuando se trata con estadísticas de Atributos y se hace el conteo de las personas o cosas que poseen una característica definida, simplemente lo que hacemos es cuantificar esos resultados, por lo cual es necesario contar el número de personas o cosas que poseen el atributo como una clase independiente, y enfrentarla al

número de aquellos que no la poseen para formar un universo.

Cuando el observador realiza una clasificación por atributos como fué definida antes, se encuentra con 2 clases, es decir, la clase de las personas o cosas que poseen el atributo y, además, las que carecen de él; dichas 2 clases son mutuamente excluyentes ya que ninguna unidad de observación puede poseer ambas características a la vez; cada una de estas clases puede ser dividida nuevamente en 2 subclases y éstas continuarán el proceso hasta que se llegue a una descripción precisa del fenómeno; es claro que no solo se puede fraccionar cada clase en dos subclases sino que se pueden dividir en 3 o más de ellas. En el primer caso estamos tratando con una clasificación dicotómica (división en 2) anotar presencia o ausencia; y en el segundo caso, estamos tratando con una clasificación múltiple.

Ejemplo 12: Supongamos que se va a clasificar el número de personas existentes en una nación. Este personal podremos considerarlo desde 2 puntos de vista para formar 2 clases, mutuamente excluyentes, la clase de los profesionales y la clase de los no profesionales, pero a su vez cada una de ellas, puede subdividirse en

mujeres y hombres con lo cual ya hemos formado 4 clases distintas; mediante una nueva subdivisión podremos clasificar el grupo de los profesionales hombres en 2 clases más, ya sea que estén trabajando actualmente o estén inactivos. Las mismas divisiones podremos hacer en los otros tres grupos, con lo cual completamos 8 clases; si se continúa el proceso podremos formar tantas clases como lo exijan las necesidades de la investigación. Si cada clase se divide en 3 o más subclases tendremos una clasificación múltiple la cual será tratada posteriormente.

2.1 Terminología: Cuando consideramos un fenómeno en el cual cada clase se divide en dos subclases, -continuándose la subdivisión en la misma forma-, se obtiene finalmente una clasificación DICOTOMICA. Tal es el caso del ejemplo 12 enunciado antes.

En cuanto se refiere explícitamente a la terminología, algunos autores coinciden en designar con las letras mayúsculas (A, B, C, D, E, . . .) los fenómenos que posean una característica o atributo positivo y con letras griegas (α , β , γ , δ . . .) la ausencia de atributos; aunque también se acostumbra subrayar en la parte superior las letras mayúsculas del abecedario

para indicar la negación ($\bar{A}$, $\bar{B}$, $\bar{C}$, $\bar{D}$, ...), es decir, para representar atributos negativos de tal manera que si A representa el atributo o cualidad de ser profesional, entonces, $\bar{A}$ (no A) representa la cualidad de ser no profesional. Esta nomenclatura será la utilizada en el presente capítulo. Cada letra representa una CLASE y el orden de ellas depende del número de las diferentes letras que la forman (yuxtaposición).

Ejemplo 13: En el ejemplo 12 se puede representar a N como el número de personas en la nación objeto de estudio; con A , la clase o propiedad de ser profesional, y con $\bar{A}$ la de los no profesionales. Podremos designar por B la propiedad de ser hombre y $\bar{B}$ la de mujer. Entonces la clase $A B$ representa la propiedad de ser profesional y hombre.

$A \bar{B}$ Ser profesional y mujer

$\bar{A} B$ Ser no profesional y hombre

$\bar{A} \bar{B}$ Ser no profesional y mujer

Podremos de esta forma obtener clases de un orden superior por yuxtaposición de letras.

2.2 Frecuencias y Orden de las Clases: Al número de elementos o de observaciones que están en posesión de una característica, dentro de una clase cualquiera, se llamará una FRECUENCIA DE CLASE, y se denota encerrando entre paréntesis la clase correspondiente o pospuesta de una N mayúscula, así:

- (A.) o (A) N = Número de observaciones que poseen el atributo A.
(AB.) o (AB) N = Número de observaciones que poseen A y B.
($\overline{A}\overline{B}\overline{C}$.) o ($\overline{A}\overline{B}\overline{C}$) N = Número de observaciones que poseen A pero no los B y C, etc.

Cuando expresamos los atributos de esta manera estamos haciendo una cuantificación. De acuerdo al número de atributos de cada clase formada, se obtienen grupos que varían en cuanto al orden se refiere; si $P = 1, 2, 3, \dots$, es el número de atributos de una clase, así mismo formaremos grupo 1 hasta de ORDEN P.

Si $P = 1$ tendremos clases de orden 1

A o $\overline{A}$

Si $P = 2$ tendremos clases de orden 2

AB $\overline{A}\overline{B}$ $\overline{A}B$ $A\overline{B}$

./.

Si $P = 3$ tendremos clases de orden 3

$ABC \quad A\bar{B}\bar{C} \quad \bar{A}BC \quad \bar{A}\bar{B}C \quad \bar{A}\bar{B}\bar{C} \quad \bar{A}BC \quad \bar{A}\bar{B}C \quad \bar{A}\bar{B}\bar{C},$

Si se cuantifican las frecuencias de cada grupo se formarán frecuencias de orden P .

$(P = 1, 2, 3, \dots)$

Ejemplo 14:

Frecuencias orden 1: $(A), (\bar{A})$

Frecuencias orden 2: $(AB), (A\bar{B}), \text{etc.}$

Frecuencias orden 3: $(ABC), (\bar{A}\bar{B}C), \text{etc.}$

En la misma forma como se definieron las clases pueden definirse las frecuencias de cada clase considerándolas de la siguiente forma: una frecuencia POSITIVA se refiere al número de observaciones en los cuales solo intervienen atributos positivos; será NEGATIVA en el caso contrario; será MIXTA si se combinan positivas y negativas; CONTRARIAS cuando a los atributos positivos de unas corresponden negativas de las otras; frecuencias ULTIMAS, corresponden a las frecuencias de los grupos de orden P de acuerdo al número de atributos considerados.

13 Total de las Frecuencias de Clase: De acuerdo al número (P) de atributos bajo consideración, observamos que se forman clases y sus respectivas frecuencias incluyendo al número total de observaciones (n) como de orden 0, ya que no se especifica ningún atributo en especial, es decir, que formaremos grupos de orden 0, 1, 2, 3, ... P con sus correspondientes frecuencias.

Ejemplo 15: Consideramos un número n de personas y al mismo tiempo, la presencia de P = 3 atributos A, B y C.

El orden y el número de frecuencias sería:

ORDEN 0 : n

ORDEN 1 : (A), (B), (C), ($\bar{A}$), ($\bar{B}$), ($\bar{C}$)

ORDEN 2 : (A B), (A $\bar{B}$), ($\bar{A}$ B), ($\bar{A}$ $\bar{B}$)

(A C), (A $\bar{C}$), ($\bar{A}$ C), ($\bar{A}$ $\bar{C}$)

(B C), (B $\bar{C}$), ($\bar{B}$ C), ($\bar{B}$ $\bar{C}$)

ORDEN 3 : (A B C) (A B $\bar{C}$) (A $\bar{B}$ C) (A $\bar{B}$ $\bar{C}$)

($\bar{A}$ B C) ($\bar{A}$ B $\bar{C}$) ($\bar{A}$ $\bar{B}$ C) ($\bar{A}$ $\bar{B}$ $\bar{C}$)

Un análisis detallado por grupo nos muestra el número de frecuencias que existen en cada uno estableciéndose en general lo siguiente:

Sean: r el orden de la frecuencia; n , el número de observaciones y P , el número de atributos diferentes. Entonces, el número de frecuencias del orden r vendrá dado por la expresión:

$$\frac{P(P-1)(P-2)\dots(P-r+1)}{r!} 2^r \quad r \geq 2$$

Considerando a n como frecuencia de orden 0 y a $2P$ como el número de frecuencias de orden 1.

El número total de frecuencias de clase que se obtienen al estudiar P atributos obedece al desarrollo del binomio :

$$(1+2)^P = 3^P.$$

Ejemplo 16: En el caso de 3 atributos se considera el número o total de frecuencias de clase del ejemplo anterior:

$$(1+2)^P = 1 + 2P + \frac{P(P-1)}{2!} 2^2 + \frac{P(P-1)(P-2)}{3!} 2^3$$

$$\dots \frac{P(P-1)(P-2)\dots(P-r+1)}{r!} 2^r \dots$$

$$(1+2)^3 = 1 + 2 \cdot 3 + \frac{3(2)}{2!} 2^2 + \frac{3(2)(1)}{3!} 2^3$$

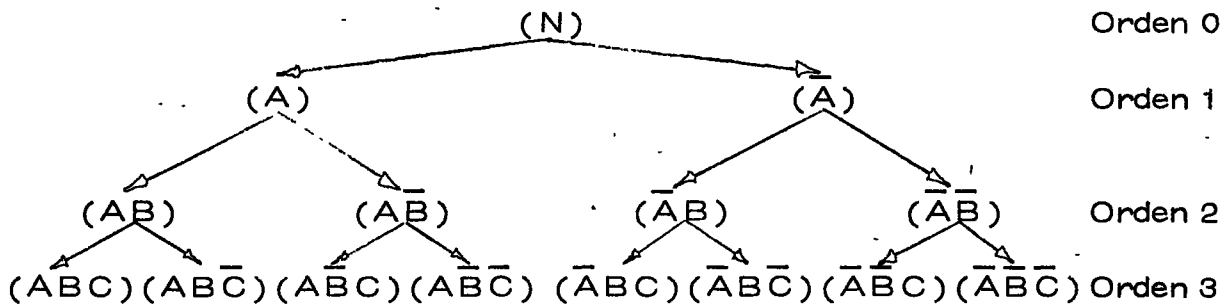
$$3^3 = 1 + 6 + 12 + 8 = 27$$

Es decir: se obtienen 27 frecuencias en total; 1 es de orden 0;

6 de orden 1; 12 de orden 2 y 8 de orden 3.

En toda tabulación de frecuencias se deben tener en cuenta 4 as
pecios:

- a) Toda frecuencia de clase debe poder expresarse en términos de las frecuencias de orden superior. Observamos una clasificación dicotómica, así:



Podremos concluir lo siguiente:

$$\begin{aligned}
 N &= (A) + (\bar{A}) & A &= (ABC) + (ABC\bar{C}) + (A\bar{B}C) + (A\bar{B}\bar{C}) \\
 (A) &= (AB) + (A\bar{B}) & \bar{A} &= (\bar{A}BC) + (\bar{A}B\bar{C}) + (\bar{A}\bar{B}C) + (\bar{A}\bar{B}\bar{C}) \\
 (\bar{A}) &= (\bar{A}B) + (\bar{A}\bar{B}) \\
 (AB) &= (ABC) + (ABC\bar{C})
 \end{aligned}$$

Cada una de estas ecuaciones puede intercambiar sus términos de acuerdo con el enunciado y la frecuencia que nos interese encontrar.

Ejemplo 17: En un sector de 1.000 habitantes existen 3 cines, A, B y C. A través de un estudio muestral, se obtuvieron los siguientes resultados en cuanto a la preferencia de asistencia a un determinado teatro:

$$(A\bar{B}\bar{C}) = 273 \quad (A\bar{B}C) = 59 \quad (\bar{A}BC) = 51$$

$$(ABC) = 190 \quad (\bar{A}\bar{B}\bar{C}) = 103$$

$$(A\bar{B}C) = 182 \quad (\bar{A}B\bar{C}) = 131 \quad (\bar{A}\bar{B}C) = 11$$

Se pide calcular todas las frecuencias de clase:

FRECUENCIAS DE ORDEN 1

$$(A) = (ABC) + (A\bar{B}\bar{C}) + (A\bar{B}C) + (A\bar{B}\bar{C})$$

$$(A) = 190 + 273 + 59 + 182$$

$$(A) = 704$$

$$(\bar{A}) = (\bar{A}BC) + (\bar{A}\bar{B}\bar{C}) + (\bar{A}\bar{B}C) + (\bar{A}\bar{B}\bar{C})$$

$$(\bar{A}) = 51 + 131 + 11 + 103$$

$$(\bar{A}) = 296$$

$$(B) = (ABC) + (\bar{A}BC) + (A\bar{B}\bar{C}) + (\bar{A}\bar{B}\bar{C})$$

$$(B) = 190 + 51 + 273 + 131$$

$$(B) = 645$$

$$(\bar{B}) = (A\bar{B}C) + (\bar{A}\bar{B}C) + (A\bar{B}\bar{C}) + (\bar{A}\bar{B}\bar{C})$$

$$(\bar{B}) = 59 + 11 + 182 + 103$$

$$(\bar{B}) = 355$$

$$(C) = (ABC) + (\bar{A}B\bar{C}) + (A\bar{B}C) + (\bar{A}\bar{B}C)$$

$$(C) = 190 + 51 + 59 + 11$$

$$(C) = 311$$

$$(\bar{C}) = (AB\bar{C}) + (A\bar{B}\bar{C}) + (\bar{A}B\bar{C}) + (\bar{A}\bar{B}\bar{C})$$

$$(\bar{C}) = 273 + 182 + 131 + 103$$

$$(\bar{C}) = 689.$$

FRECUENCIAS DE ORDEN 2

$$(AB) = (ABC) + (AB\bar{C})$$

$$(AB) = 190 + 273$$

$$(AB) = 463$$

$$(A\bar{B}) = (A\bar{B}C) + (A\bar{B}\bar{C})$$

$$(A\bar{B}) = 59 + 182$$

$$(A\bar{B}) = 241$$

$$(\bar{A}B) = (\bar{A}BC) + (\bar{A}B\bar{C})$$

$$(\bar{A}B) = 51 + 131$$

$$(\bar{A}B) = 182$$

$$(\bar{A}\bar{B}) = (\bar{A}\bar{B}C) + (\bar{A}\bar{B}\bar{C})$$

$$(\bar{A}\bar{B}) = 11 + 103$$

$$(\bar{A}\bar{B}) = 114$$

$$(AC) = (ABC) + (A\bar{B}C)$$

$$(AC) = 190 + 59$$

$$(AC) = 249$$

$$(A\bar{C}) = (AB\bar{C}) + (A\bar{B}\bar{C})$$

$$(A\bar{C}) = 273 + 182$$

$$(A\bar{C}) = 455$$

$$(\bar{A}C) = (\bar{A}BC) + (\bar{A}\bar{B}C)$$

$$(\bar{A}C) = 51 + 11$$

$$(\bar{A}C) = 62$$

$$(\bar{A}\bar{C}) = (\bar{A}B\bar{C}) + (\bar{A}\bar{B}\bar{C})$$

$$(\bar{A}\bar{C}) = 131 + 103$$

$$(\bar{A}\bar{C}) = 234$$

$$(B C) = (A B C) + (\bar{A} B C)$$

$$(B C) = 190 + 51$$

$$(B C) = 241$$

$$(\bar{B} C) = (A \bar{B} C) + (\bar{A} \bar{B} C)$$

$$(\bar{B} C) = 59 + 11$$

$$(\bar{B} C) = 70$$

$$(B \bar{C}) = (A B \bar{C}) + (\bar{A} B \bar{C})$$

$$(B \bar{C}) = 273 + 131$$

$$(B \bar{C}) = 404$$

$$(\bar{B} \bar{C}) = (A \bar{B} \bar{C}) + (\bar{A} \bar{B} \bar{C})$$

$$(\bar{B} \bar{C}) = 182 + 103$$

$$(\bar{B} \bar{C}) = 285$$

FRECUENCIAS DE ORDEN CERO

$$N = (A) + (\bar{A})$$

$$N = 704 + 296 = 1.000$$

$$N = (B) + (\bar{B})$$

$$N = 645 + 355 = 1.000$$

$$N = (C) + (\bar{C})$$

$$N = 311 + 689 = 1.000$$

b) El número de frecuencias últimas es del orden 2^P siendo P el número de atributos de que se trate.

Ejemplo 18: Dadas las siguientes frecuencias, calcular las frecuencias últimas, referentes a $P = 3$ atributos.

$$N = 23.713$$

$$(A) = 1.618$$

$$(B) = 2.015$$

$$(C) = 770$$

$$(A\bar{B}) = 587$$

$$(\bar{A}C) = 428$$

$$(B\bar{C}) = 335$$

$$(ABC) = 156.$$

El problema se limita a encontrar las frecuencias últimas, que como hemos dicho deben ser:

$$2^3 = 8$$

$$(A) = (AB) + (A\bar{B})$$

$$(AB) = (A) - (A\bar{B})$$

$$(AB) = 1618 - 587$$

$$(AB) = 1031$$

$$(B) = (AB) + (\bar{A}B)$$

$$(\bar{A}B) = (B) - (AB)$$

$$(\bar{A}B) = 2015 - 1031$$

$$(\bar{A}B) = 984$$

$$(B) = (BC) + (B\bar{C})$$

$$(BC) = (B) - (B\bar{C})$$

$$(BC) = 2015 - 335$$

$$(BC) = 1680$$

$$(C) = (AC) + (\bar{A}C)$$

$$(AC) = (C) - (\bar{A}C)$$

$$(AC) = 770 - 428$$

$$(AC) = 342$$

$$(AB) = (ABC) + (AB\bar{C})$$

$$(AB\bar{C}) = (AB) - (ABC)$$

$$(AB\bar{C}) = 1031 - 156$$

$$(AB\bar{C}) = 875$$

$$(AC) = (ABC) + (A\bar{B}C)$$

$$(A\bar{B}C) = (AC) - (ABC)$$

$$(A\bar{B}C) = 342 - 156$$

$$(A\bar{B}C) = 186$$

$$(A\bar{B}) = (A\bar{B}C) + (A\bar{B}\bar{C})$$

$$(A\bar{B}\bar{C}) = (A\bar{B}) - (A\bar{B}C)$$

$$(A\bar{B}\bar{C}) = 587 - 186$$

$$(A\bar{B}\bar{C}) = 401$$

$$(BC) = (ABC) + (\bar{A}BC)$$

$$(\bar{A}BC) = (BC) - (ABC)$$

$$(\bar{A}BC) = 1680 - 156$$

$$(\bar{A}BC) = 1524$$

$$(\bar{A}B) = (\bar{A}BC) + (\bar{A}B\bar{C})$$

$$(\bar{A}B\bar{C}) = (\bar{A}B) - (\bar{A}BC)$$

$$(\bar{A}B\bar{C}) = 984 - 1524$$

$$(\bar{A}B\bar{C}) = -540$$

$$(\bar{A}C) = (\bar{A}BC) + (\bar{A}\bar{B}C)$$

$$(\bar{A}\bar{B}C) = (\bar{A}C) - (\bar{A}BC)$$

$$(\bar{A}\bar{B}C) = 428 - 1524$$

$$(\bar{A}\bar{B}C) = -1.096$$

$$(\bar{A}\bar{B}) = (\bar{A}\bar{B}C) + (\bar{A}\bar{B}\bar{C})$$

$$(\bar{A}\bar{B}\bar{C}) = (\bar{A}\bar{B}) - (\bar{A}\bar{B}C)$$

Sabemos que:

$$(\bar{A}\bar{B}) = (\bar{A}) - (\bar{A}B)$$

$$(\bar{A}) = N - (A)$$

$$(\bar{A}) = 23.713 - 1618$$

$$(\bar{A}) = 22.095$$

$$(\bar{A}\bar{B}) = 22.095 - 984$$

$$(\bar{A}\bar{B}) = 21.111$$

Luego:

$$(\bar{A}\bar{B}\bar{C}) = 21.111 - (-1.096)$$

$$(\bar{A}\bar{B}\bar{C}) = 21.111 + 1.096$$

$$(\bar{A}\bar{B}\bar{C}) = 22.207$$

Dado que las frecuencias $(\bar{A}B\bar{C})$ y $(\bar{A}\bar{B}C)$ resultaron negativas y no es posible que lo sean (lo que se verá más adelante), se concluye que los datos son inconsistentes.

c) El número de frecuencias positivas también es de 2^P en todos los órdenes, así:

Orden 0 = 1	N
Orden 1 = 3	(A) , (B) , (C)
Orden 2 = 3	(A B) , (A C) , (B C)
Orden 3 = 1	(A B C)

En el ejemplo 18 se ve que las frecuencias positivas son:

N	= 23.713
(A)	= 1.618
(B)	= 2.015
(C)	= 770
(A B)	= 1.031
(A C)	= 342
(B C)	= 1.680
(A B C)	= 156

d) Toda frecuencia negativa o mixta puede expresarse en función de las frecuencias positivas:

$$\begin{aligned}
 (\bar{A}) &= N - (A) \\
 (A\bar{B}) &= (A) - (AB) \\
 (\bar{A}B) &= (B) - (AB), \text{ etc.}
 \end{aligned}$$

Ejemplo 19: Dados los siguientes datos, calcular:

$$(A\bar{B}\bar{C}), (\bar{A}\bar{B}\bar{C}), (\bar{A}B), (B\bar{C}) \text{ y } (\bar{A})$$

$$N = 400$$

$$(A) = 125$$

$$(B) = 69$$

$$(C) = 102$$

$$(AB) = 12$$

$$(AC) = 30$$

$$(BC) = 24$$

$$(ABC) = 10$$

$$(\bar{A}) = N - (A)$$

$$(\bar{A}) = 400 - 125$$

$$(\bar{A}) = 275$$

$$(\bar{A}B) = (B) - (AB)$$

$$(\bar{A}B) = 69 - 12$$

$$(\bar{A}B) = 57$$

$$(B\bar{C}) = (B) - (BC)$$

$$(B\bar{C}) = 69 - 24$$

$$(B\bar{C}) = 45$$

$$(A\bar{B}) = (A\bar{B}C) + (A\bar{B}\bar{C})$$

$$(AC) = (A\bar{B}C) + (ABC)$$

$$(A\bar{B}C) = (AC) - (ABC)$$

$$(A\bar{B}C) = 30 - 10$$

$$(A\bar{B}C) = 20$$

$$(A\bar{B}) = (A) - (AB)$$

$$(A\bar{B}) = 125 - 12$$

$$(A\bar{B}) = 113$$

$$(A\bar{B}\bar{C}) = (A\bar{B}) - (A\bar{B}C)$$

$$(A\bar{B}\bar{C}) = 113 - 20$$

$$(A\bar{B}\bar{C}) = 93$$

$$(\bar{A}\bar{B}\bar{C}) = (\bar{B}\bar{C}) - (A\bar{B}\bar{C})$$

Tenemos:

$$(\bar{B}) = (\bar{B}C) + (\bar{B}\bar{C})$$

$$(\bar{B}\bar{C}) = (\bar{B}) - (\bar{B}C)$$

$$(\bar{B}) = N - (B) \vee (C) = (\bar{B}C) + (BC)$$

Reemplazando:

$$(\overline{BC}) = [N - (B)] - [(C) - (BC)]$$

$$(\overline{BC}) = N - (B) - (C) + (BC)$$

$$(\overline{BC}) = 400 - 69 - 102 + 24$$

$$(\overline{BC}) = 253$$

Luego:

$$(\overline{ABC}) = 253 - 93$$

$$(\overline{ABC}) = 160$$

3. RELACION DE ATRIBUTOS

3.1 Consistencia de Atributos: Cuando se hace un estudio y se analizan los resultados obtenidos, éstos deben corresponder a la realidad, ya que la información se extrae de una misma población, con un mismo período de referencia; los resultados así obtenidos pueden ser INCONSISTENTES debido a los posibles errores cometidos ya sea porque se extrajo información de diferentes universos o debido a que no se tuvo en cuenta el período de referencia; en resumen, se presenta inconsistencia cuando no hay un universo bien definido o cuando no existe una identidad de tiempo.

Ejemplo 20: Supóngase que se obtuvieron los siguientes resultados en una investigación acerca de los gustos de los consumidores para tres artículos:

$$N = 200 \quad (AB) = 60$$

$$(A) = 125 \quad (AC) = 90$$

$$(B) = 100 \quad (BC) = 80$$

$$(C) = 165 \quad (ABC) = 48$$

Calcular la frecuencia ($\overline{ABC}$).

Se sabe que:

$$(\overline{ABC}) = N - (A) - (B) - (C) + (AB) + (AC) + (BC) - (ABC)$$

$$(\overline{ABC}) = 200 - 125 - 100 - 165 + 60 + 90 + 80 - 48$$

$$(\overline{ABC}) = -8.$$

Estos resultados son inconsistentes porque se obtiene un valor negativo para ($\overline{ABC}$), y no es posible que esto suceda; puede haber un error en cuanto al universo o al período de referencia.

3.1.1 Condiciones de Consistencia. Se enuncian a continuación algunas normas que se deben tener en cuenta para la consistencia de atributos: Ninguna frecuencia de clase puede ser negativa; ninguna frecuencia de clase debe ser mayor

que el total de datos; la suma de las frecuencias de las clases debe ser igual al número total de elementos considerados.

Ejemplo 21 : Como puede observarse en el desarrollo de los ejercicios anteriores se tiene que :

$(\bar{A}\bar{B}\bar{C})$ debe ser $\Rightarrow 8$ Ver ejemplo 17

$(A \supset B) \supset N \Rightarrow 463 \supset 1.000$ Ver ejemplo 17

$(A) + (\bar{A}) = N \Rightarrow 704 + 296 = 1.000$ Ver ejemplo 17

Ejemplo 22: En una ciudad de Colombia, el porcentaje de personas (frecuencia) que lee las revistas A, B y C, así como sus combinaciones, es el siguiente:

A: 9,8%	A y B: 5,1%
B: 22,9%	A y C: 3,7% A y B y C: 2,4%
C: 12,1%	B y C: 6,0%

Calcular el porcentaje (frecuencias) de personas que :

- Lee al menos una de las tres revistas
- lee A, o B, o ambas
- lee solo A y no B ni C

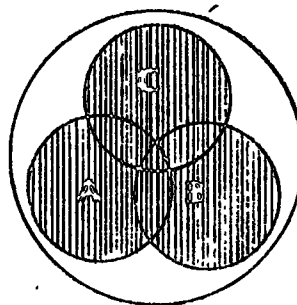
* $\supset$ Se lee "no mayor que"

- d) lee solo B y no A ni C
- e) lee solo C y no A ni B
- f) lee solo A y B y no C
- g) lee solo B y C y no A
- h) lee solo A y C y no B
- i) no lee revista alguna
- j) Calcular, entre las personas que leen al menos una revista, cuál es el porcentaje de las que leen dos revistas y solo dos.

Solución:

$$\begin{aligned}
 (A) &= 9,8\% & (AB) &= 5,1\% & (ABC) &= 2,4\% \\
 (B) &= 22,9\% & (AC) &= 3,7\% \\
 (C) &= 12,1\% & (BC) &= 6,0\%
 \end{aligned}$$

a)



$$\begin{aligned}
 &(ABC) + (\overline{A}BC) + (A\overline{B}C) + (AB\overline{C}) + (\overline{A}\overline{B}C) + (\overline{A}B\overline{C}) + \\
 &(\overline{A}BC) = ?
 \end{aligned}$$

Sabemos $(ABC) = 2.4\%$

Hay que averiguar $(\bar{A}BC)$, $(A\bar{B}C)$, $(\bar{A}\bar{B}C)$, $(A\bar{B}\bar{C})$,
 $(\bar{A}B\bar{C})$ y $(\bar{A}\bar{B}\bar{C})$

Procedimiento:

$$(ABC) + (\bar{A}BC) = (AC)$$

$$(A\bar{B}C) = (AC) - (ABC)$$

$$(A\bar{B}C) = 3.7\% - 2.4\%$$

$$(A\bar{B}C) = 1.3\%$$

$$(A\bar{B}C) + (ABC) = (AB)$$

$$(A\bar{B}\bar{C}) = (AB) - (ABC)$$

$$(A\bar{B}\bar{C}) = 5.1\% - 2.4\%$$

$$(A\bar{B}\bar{C}) = 2.7\%$$

$$(\bar{A}BC) + (ABC) = (BC)$$

$$(\bar{A}B\bar{C}) = (BC) - (ABC)$$

$$(\bar{A}B\bar{C}) = 6.0\% - 2.4\%$$

$$(\bar{A}B\bar{C}) = 3.6\%$$

$$(\bar{A}\bar{B}C) + (\bar{A}\bar{B}\bar{C}) = (\bar{A}\bar{B})$$

$$\text{Pero: } (\bar{A}\bar{B}) + (AB) = (A)$$

$$(A\bar{B}) = (A) - (AB)$$

$$(A\bar{B}) = 9.8\% - 5.1\%$$

$$(A\bar{B}) = 4.7\%$$

$$(A\bar{B}\bar{C}) = (A\bar{B}) - (A\bar{B}C)$$

$$(A\bar{B}\bar{C}) = 4.7\% - 1.3\%$$

$$(A\bar{B}\bar{C}) = 3.4\%$$

$$(\bar{A}\bar{B}\bar{C}) + (\bar{A}BC) = (\bar{A}\bar{B})$$

$$\text{Pero: } (\bar{A}\bar{B}) + (AB) = (B)$$

$$(\bar{A}\bar{B}) = (B) - (AB)$$

$$(\bar{A}\bar{B}) = 22.9\% - 5.1\%$$

$$(\bar{A}\bar{B}) = 17.8\%$$

$$(\bar{A}\bar{B}\bar{C}) = (\bar{A}\bar{B}) - (\bar{A}BC)$$

$$(\bar{A}\bar{B}\bar{C}) = 17.8\% - 3.6\%$$

$$(\bar{A}\bar{B}\bar{C}) = 14.2\%$$

$$(\bar{A}\bar{B}C) + (\bar{A}BC) = (\bar{A}C)$$

$$\text{Pero: } (\bar{A}C) + (AC) = (C)$$

$$(\bar{A}C) = (C) - (AC)$$

$$(\bar{A}C) = 12.1\% - 3.7\%$$

$$(\bar{A}C) = 8.4\%$$

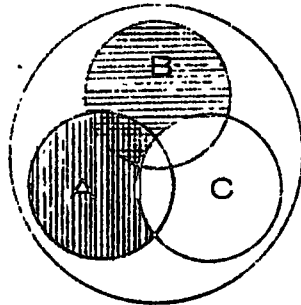
$$(\overline{A}\overline{B}C) = (\overline{A}C) - (\overline{A}BC)$$

$$(\overline{A}\overline{B}C) = 8.4\% - 3.6\%$$

$$(\overline{A}\overline{B}C) = 4.8\%$$

Reemplazando: $(ABC) + (\overline{A}BC) + (\overline{A}\overline{B}C) + (\overline{A}B\overline{C}) +$
 $(AB\overline{C}) + (\overline{A}B\overline{C}) + (\overline{A}B\overline{C}) = 32.4\%$

b)



$$(AB) + (\overline{A}B) + (A\overline{B}) = ?$$

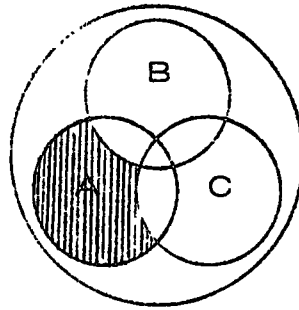
Sabemos que $(AB) = 5.1\%$

$$(\overline{A}B) = 17.8\% \text{ (parte (a))}$$

$$(A\overline{B}) = 4.7\% \quad \text{"}$$

Reemplazando: $(AB) + (\overline{A}B) + (A\overline{B}) = 27.6\%$

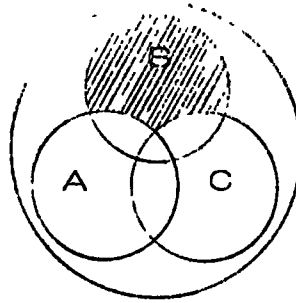
c)



$$(A \bar{B} \bar{C}) = ?$$

$$(A \bar{B} \bar{C}) = 3.4\% \text{ (por la parte (a))}$$

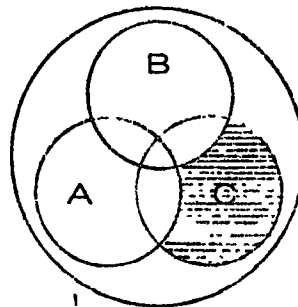
d)



$$(\bar{A} B \bar{C}) = ?$$

$$(\bar{A} B \bar{C}) = 14.2\%$$

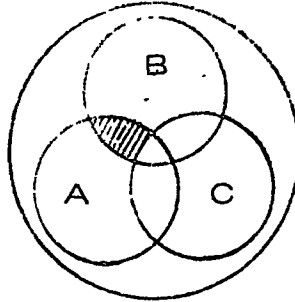
e)



$$(\bar{A}\bar{B}C) = ?$$

$$(\bar{A}\bar{B}C) = 4.8\%$$

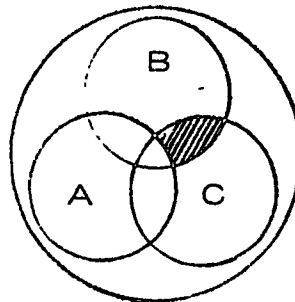
f)



$$(AB\bar{C}) = ?$$

$$(AB\bar{C}) = 2.7\%$$

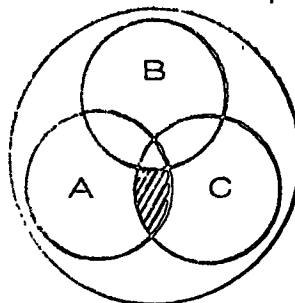
g)



$$(\bar{A}BC) = ?$$

$$(\bar{A}BC) = 3.6\%$$

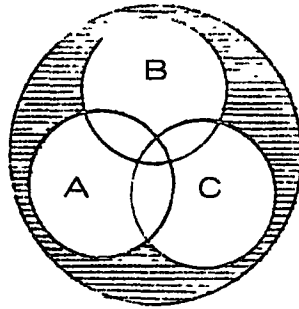
h)



$$(A \bar{B} C) = ?$$

$$(A \bar{B} C) = 1.3\%$$

i)



$$(\bar{A} \bar{B} \bar{C}) = ?$$

$$(\bar{A} \bar{B} \bar{C}) + (\bar{A} \bar{B} C) = (\bar{A} \bar{B})$$

$$(\bar{A} \bar{B}) + (A \bar{B}) = (\bar{B})$$

$$(\bar{B}) + (B) = N$$

$$(\bar{B}) = N - (B)$$

$$(\bar{B}) = 100\% - 22.9\%$$

$$(\bar{B}) = 77.1\%$$

$$(\bar{A} \bar{B}) = (\bar{B}) - (A \bar{B})$$

$$(\bar{A} \bar{B}) = 77.1\% - 4.7\%$$

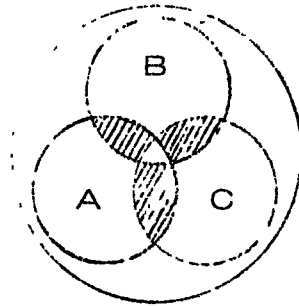
$$(\bar{A} \bar{B}) = 72.4\%$$

$$(\bar{A} \bar{B} \bar{C}) = (\bar{A} \bar{B}) - (\bar{A} \bar{B} C)$$

$$(\bar{A} \bar{B} \bar{C}) = 72.4\% - 4.8\%$$

$$(\bar{A} \bar{B} \bar{C}) = 67.6\%$$

J)



$$\frac{(AB) + (AC) + (BC) - 3(ABC)}{(ABC) + (\overline{A}\overline{B}C) + (\overline{A}B\overline{C}) + (A\overline{B}\overline{C}) + (A\overline{B}C) + (A\overline{B}C) + (\overline{A}BC)} =$$

$$\frac{.60\%}{32.4\%} = 0.23\%$$

Ejemplo 23: En una ciudad hondureña de 50.000 habitantes existen 3 cines. A través de un estudio muestral, los estadísticos han concluido que los porcentajes de personas que asisten a los teatros designados por A, B y C son los siguientes:

A : 22,6% A y B : 6,7% A y B y C : 2,4%

B : 40,3% A y C : 7,6%

C : 28.7% B y C : 10,9%

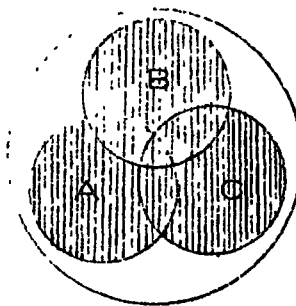
Un comerciante interesado en proyectar avisos sobre cierto producto ha hecho las siguientes preguntas:

- ¿ Cuántas personas asisten al menos a un teatro ?
- ¿ Cuántas asisten al teatro A o B o a ambos ?
- ¿ Cuántas asisten solo al A; solo al B; solo al C ?
- ¿ Cuántas no asisten a los teatros ?
- De entre las personas que asisten al teatro A, ¿ cuántas asisten también al teatro B ?
- De entre las personas que asisten al teatro C, ¿ cuántas asisten además solo al teatro B ?

Solución:

$$\begin{array}{lll}
 (A) = 22,6\% & (AB) = 6,7\% & (ABC) = 2,4\% \\
 (B) = 40,3\% & (AC) = 7,6\% & \\
 (C) = 28,7\% & (BC) = 10,9\% &
 \end{array}$$

a)



$$(ABC) + (A\bar{B}\bar{C}) + (\bar{A}BC) + (\bar{A}\bar{B}C) + (AB\bar{C}) + (A\bar{B}C) + (\bar{A}BC) = ?$$

Sabemos que:

$$(ABC) = 2.4\%$$

Hay que averiguar:

$$(\bar{A}BC), (A\bar{B}C), (\bar{A}BC), (\bar{A}\bar{B}C), (\bar{A}\bar{B}C) \text{ y } (\bar{A}\bar{B}C)$$

Procedimiento:

$$(ABC) + (\bar{A}BC) = (AC)$$

$$(\bar{A}BC) = (AC) - (ABC)$$

$$(\bar{A}BC) = 7.6\% - 2.4\%$$

$$(\bar{A}BC) = 5.2\%$$

$$(ABC) + (A\bar{B}C) = (AB)$$

$$(A\bar{B}C) = (AB) - (ABC)$$

$$(A\bar{B}C) = 6.7\% - 2.4\%$$

$$(A\bar{B}C) = 4.3\%$$

$$(\bar{A}BC) + (ABC) = (BC)$$

$$(\bar{A}BC) = (BC) - (ABC)$$

$$(\bar{A}BC) = 10.9\% - 2.4\%$$

$$(\bar{A}BC) = 8.5\%$$

$$(\bar{A}\bar{B}C) + (A\bar{B}C) = (\bar{A}\bar{B})$$

$$\text{Pero: } (\bar{A}\bar{B}) + (AB) = (A)$$

$$(\bar{A}\bar{B}) = (A) - (AB)$$

$$(\bar{A}\bar{B}) = 22.6\% - 6.7\%$$

$$(\bar{A}\bar{B}) = 15.9\%$$

$$(\overline{A}\overline{B}\overline{C}) = (\overline{A}\overline{B}) - (\overline{A}\overline{B}C)$$

$$(\overline{A}\overline{B}\overline{C}) = 15.9\% - 5.2\%$$

$$(\overline{A}\overline{B}\overline{C}) = 10.7\%$$

$$(\overline{A}\overline{B}\overline{C}) + (\overline{A}\overline{B}C) = (\overline{A}\overline{B})$$

Pero: $(\overline{A}\overline{B}) + (A\overline{B}) = (B)$

$$(\overline{A}\overline{B}) = (B) - (A\overline{B})$$

$$(\overline{A}\overline{B}) = 40.3\% - 6.7\%$$

$$(\overline{A}\overline{B}) = 33.6\%$$

$$(\overline{A}\overline{B}\overline{C}) = (\overline{A}\overline{B}) - (\overline{A}\overline{B}C)$$

$$(\overline{A}\overline{B}\overline{C}) = 33.6\% - 8.5\%$$

$$(\overline{A}\overline{B}\overline{C}) = 25.1\%$$

$$(\overline{A}\overline{B}\overline{C}) + (\overline{A}\overline{B}C) = (\overline{A}\overline{C})$$

Pero: $(\overline{A}\overline{C}) + (A\overline{C}) = (C)$

$$(\overline{A}\overline{C}) = (C) - (A\overline{C})$$

$$(\overline{A}\overline{C}) = 28.7\% - 7.6\%$$

$$(\overline{A}\overline{C}) = 21.1\%$$

$$(\overline{A}\overline{B}\overline{C}) = (\overline{A}\overline{C}) - (\overline{A}\overline{B}C)$$

$$(\overline{A}\overline{B}\overline{C}) = 21.1\% - 8.5\%$$

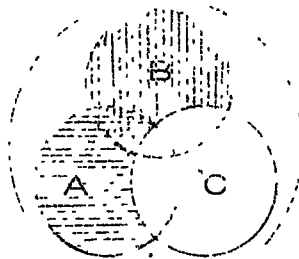
$$(\overline{A}\overline{B}\overline{C}) = 12.6\%$$

Reemplazando:

$$(ABC) + (\overline{A}BC) + (A\overline{B}C) + (\overline{A}\overline{B}C) + (ABC) + (\overline{A}BC) + (\overline{A}BC) = 68.8\%$$

$$68.8\% \text{ de } 50.000 = 34.400 \text{ personas.}$$

b)



$$(AB) + (\overline{A}\overline{B}) + (\overline{A}B) = ?$$

$$(AB) = 6.7\%$$

$$(\overline{A}\overline{B}) = 15.9\%$$

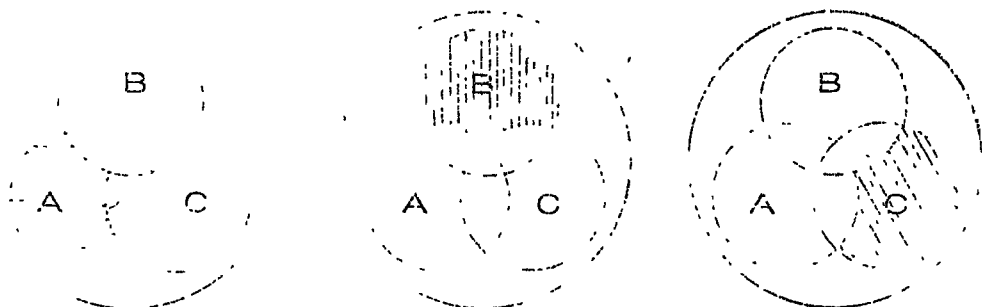
$$(\overline{A}B) = 33.6\%$$

Reemplazando:

$$(AB) + (\overline{A}\overline{B}) + (\overline{A}B) = 56.2\%$$

$$56.2\% \text{ de } 50.000 = 28.100 \text{ personas.}$$

c)



$$(\overline{A}\overline{B}\overline{C}) = ? \quad (\overline{A}\overline{B}C) = ? \quad (\overline{A}B\overline{C}) = ?$$

$$(\overline{A}\overline{B}\overline{C}) = 10.7\%$$

$$10.7\% \text{ de } 50.000 = 5.350 \text{ personas.}$$

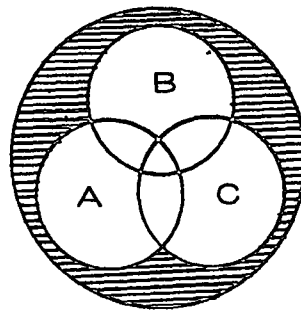
$$(\overline{A}\overline{B}C) = 25.1\%$$

$$25.1\% \text{ de } 50.000 = 12.550 \text{ personas.}$$

$$(\overline{A}B\overline{C}) = 12.6\%$$

$$12.6\% \text{ de } 50.000 = 6.300 \text{ personas.}$$

d)



$$(\overline{A}\overline{B}\overline{C}) = ?$$

$$(\overline{A}\overline{B}\overline{C}) + (\overline{A}\overline{B}C) = (\overline{A}\overline{B})$$

$$(\overline{A}\overline{B}) + (\overline{A}B) = (\overline{A})$$

$$(\overline{A}) + (A) = N$$

$$(\overline{A}) = N - (A)$$

$$(\overline{A}) = 100\% - 40.3\%$$

$$(\overline{A}) = 59.7\%$$

$$(\overline{AB}) = (\overline{B}) - (A\overline{B})$$

$$(\overline{AB}) = 59.7\% - 15.9\%$$

$$(\overline{AB}) = 43.8\%$$

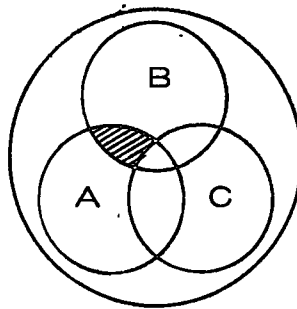
$$(\overline{ABC}) = (\overline{AB}) - (\overline{ABC})$$

$$(\overline{ABC}) = 43.8\% - 12.6\%$$

$$(\overline{ABC}) = 31.2\%$$

$$31.2\% \text{ de } 50.000 = 15.600 \text{ personas.}$$

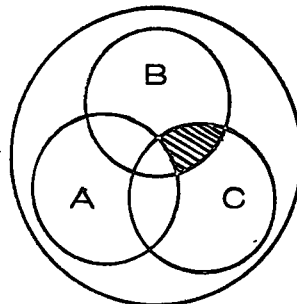
e)



$$\frac{(AB)}{(A)} = \frac{6.7}{22.6} = 29.64\%$$

$$29.64\% \text{ de } 5.350 = 1.586 \text{ personas}$$

f)



$$\frac{(ABC)}{(C)} = \frac{8.5\%}{28.7\%} = 29.62\%$$

29.62% de 6 300 = 1.866 personas.

3.2 Criterio de Independencia: Será necesario conocer cuando dos atributos tienen alguna clase de dependencia o si por el contrario no existe dependencia, en cuyo caso se dice que los atributos son INDEPENDIENTES. Las formas para el criterio de independencia son diversas, según el tipo de atributos que se estén relacionando, pero por lo general puede decirse que:

$$\frac{(AB)}{(B)} = \frac{(A\bar{B})}{(\bar{B})} \quad (1)$$

Esta relación y las demás que pueden obtenerse al considerar 2 atributos; se puede comprender más fácilmente si agrupamos las frecuencias en un cuadro de filas y columnas de la siguiente manera:

ATRIBUTO	ATRIBUTO		TOTALES
	B	$\bar{B}$	
A	(AB)	(A $\bar{B}$)	(A)
$\bar{A}$	($\bar{A}B$)	($\bar{A}\bar{B}$)	($\bar{A}$)
TOTALES	(B)	($\bar{B}$)	N

$$\frac{(\bar{A}B)}{(B)} = \frac{(\bar{A}\bar{B})}{(\bar{B})} \quad (2)$$

$$\frac{(A\bar{B})}{(A)} = \frac{(\bar{A}\bar{B})}{(\bar{A})} \quad (3)$$

$$\frac{(AB)}{(A)} = \frac{(\bar{A}B)}{(\bar{A})} \quad (4)$$

Establiendo relaciones entre las filas y las columnas se afirma:

$$\frac{(AB) + (A\bar{B})}{(B) + (\bar{B})} = \frac{(AB)}{(B)} = \frac{(A)}{N} \quad (\text{Filas}) \quad (5)$$

$$\frac{(AB)}{(A)} = \frac{(AB) + (\bar{A}B)}{(A) + (\bar{A})} = \frac{(B)}{N} \quad (\text{Columnas}) \quad (6)$$

De la ecuación (5) se concluye:

$$\frac{(AB)}{(B)} = \frac{(A)}{N} \Rightarrow (AB) = \frac{(A)(B)}{N} \quad (7)$$

Lo cual significa, que si hay independencia, la proporción de los (AB) es igual al producto de los (A) por los (B) dentro del universo.

Si a los miembros de la ecuación (7) se les resta una misma cantidad la ecuación no se altera.

$$\begin{aligned} (A) - (AB) &= (A) - \frac{(A)(B)}{N} \\ (A\bar{B}) &= \frac{N(A) - (A)(B)}{N} \Rightarrow (A\bar{B}) = \frac{(A) [N - (B)]}{N} \\ \Rightarrow (A\bar{B}) &= \frac{(A)(\bar{B})}{N} \quad (8) \end{aligned}$$

Análogamente se obtienen criterios de independencia para las demás frecuencias.

Transformando los términos de la ecuación (8) se puede escribir:

$$(\bar{B}) - (A\bar{B}) = (\bar{B}) - \frac{(A)(\bar{B})}{N}$$

$$(\bar{A}\bar{B}) = \frac{N(\bar{B}) - (A)(\bar{B})}{N} \Rightarrow (\bar{A}\bar{B}) = \frac{(\bar{B}) [N - (A)]}{N}$$

$$(\bar{A}\bar{B}) = \frac{(\bar{A})(\bar{B})}{N} \quad (9)$$

Se concluye sin lugar a equivocaciones que:

$$(\bar{A}B) = \frac{(\bar{A})(B)}{N} \quad (10)$$

multiplicando miembro a miembro (7) y (9); (8) y (10)

$$(AB)(\bar{A}\bar{B}) = \frac{(A)(B)(\bar{A})(\bar{B})}{N^2} \quad (11)$$

$$(A\bar{B})(\bar{A}B) = \frac{(A)(\bar{B})(\bar{A})(B)}{N^2} \quad (12)$$

Luego:

$$(AB)(\bar{A}\bar{B}) = (A\bar{B})(\bar{A}B)$$

Puede deducirse lo siguiente:

$$\frac{(AB)}{(A\bar{B})} = \frac{(\bar{A}\bar{B})}{(\bar{A}B)} \quad \text{ó} \quad \frac{(AB)}{(\bar{A}B)} = \frac{(A\bar{B})}{(\bar{A}\bar{B})}$$

Conclusión:

La razón de los B y los $\bar{B}$ es la misma dentro del conjunto de los A y $\bar{A}$

La razón de los A y $\bar{A}$ dentro del conjunto B es la misma dentro de los $\bar{B}$.

Ejemplo 24: Se tienen las propiedades A y B se sabe que de A se tienen 25 unidades, de la B se tienen 30 unidades; de la $\bar{A}$ igual a 8; de la $\bar{B}$ igual a 15; de AB igual a 8; de $\bar{A}\bar{B}$ igual a 5, de $\bar{A}B$ igual a 4; de $A\bar{B}$ igual a 6; se pregunta:

1. Existen relaciones de independencia ?
2. No existen relaciones de independencia ?
3. Los datos son consistentes ?
4. Los datos son inconsistentes ?

(A)	= 25		25	8	?
(B)	= 30	$\bar{B}$	6	5	15
($\bar{A}$)	= 8	B	8	4	30
($\bar{B}$)	= 15		A	$\bar{A}$	

No hay relación de independencia.

(AB) = 8

($\bar{A}\bar{B}$) = 5

($\bar{A}B$) = 4

($A\bar{B}$) = 6

S. existiera relación de independencia, el cuadro quedaría:

	14	9	23
$\bar{B}$	6	5	11
B	8	4	12
	A	$\bar{A}$	

Pero de aquí se puede sacar otra conclusión:

$$(A\bar{B}) = \frac{(14)(11)}{23} = 6,6 \neq 6.$$

$$(AB) = \frac{(14)(12)}{23} = 7,3 \neq 8$$

$$(\bar{A}\bar{B}) = \frac{(9)(11)}{23} = 4,3 \neq 5 ; (\bar{A}B) = \frac{(9)(12)}{23} = 4,6 \neq 4.$$

No existe criterio de independencia entre estos atributos. Los datos son consistentes según lo estudiado anteriormente.

Ejemplo 25:

$$A = 40\%$$

$$B = 60\%$$

$$\bar{A}\bar{B} = 15\%$$

Se pregunta:

1. Obtener los subconjuntos y decir si son consistentes.
2. Decir si existe el criterio de independencia.

1D
274

	40%	60%	100%
$\bar{B}$	25%	15%	40%
B	15%	45%	60%
	A	$\bar{A}$	

1. Los datos son consistentes.

2. No hay relaciones de independencia, ya que:

$$(AB) \neq \frac{(A)(B)}{N}, (AB) \neq \frac{(40)(60)}{100}$$

$$15\% \neq 24\%$$

En caso que existiera independencia, el cuadro quedaría:

$$(AB) = \frac{(A)(B)}{N}, 24\% = 24\%$$

	0,40	0,60	1,00
$\bar{B}$	0,16	0,24	0,40
B	0,24	0,36	0,60
	A	$\bar{A}$	

Ejemplo 26: Dadas las frecuencias de primer orden: $(A)=0,20$;

$(\bar{A}) = 0,80$; $(B) = 0,20$; $(\bar{B}) = 0,80$. Encontrar las frecuencias

de segundo orden: (AB) , $(\bar{A}B)$, $(A\bar{B})$, $(\bar{A}\bar{B})$.

$$(AB) = (A)(B) = 0,20 \times 0,20 = 0,04$$

$$(\bar{A}B) = (\bar{A})(B) = 0,80 \times 0,20 = 0,16$$

$$(A\bar{B}) = (A)(\bar{B}) = 0,20 \times 0,80 = 0,16$$

$$(\bar{A}\bar{B}) = (\bar{A})(\bar{B}) = 0,80 \times 0,80 = 0,64$$

	0,20	0,80	1,00
$\bar{B}$	0,16	0,64	0,80
B	0,04	0,16	0,20
	A	$\bar{A}$	

Luego existe independencia entre (A) y (B)

Ejemplo 27: Se tiene la siguiente tabla:

	40	160	200
$\bar{B}$	($A\bar{B}$)	($\bar{A}\bar{B}$)	160
B	(AB)	($\bar{A}B$)	40
	A	$\bar{A}$	

Se pide calcular:

$$(A\bar{B}) = ?$$

$$(AB) = ?$$

$$(\bar{A} \bar{B}) = ?$$

$$(\bar{A} B) = ?$$

$$(A \bar{B}) = \frac{(A)(\bar{B})}{N} = \frac{40 \cdot 160}{200} = 32$$

$$(\bar{A} \bar{B}) = \frac{(\bar{A})(\bar{B})}{N} = \frac{160 \cdot 160}{200} = 128$$

$$(AB) = \frac{(A)(B)}{N} = \frac{40 \cdot 40}{200} = 8$$

$$(\bar{A} B) = \frac{(\bar{A})(B)}{N} = \frac{160 \cdot 40}{200} = 32$$

La tabla de asociación de atributos quedaría:

	40	160	200
$\bar{B}$	32	128	160
B	8	32	40
	A	$\bar{A}$	

Ejemplo 28: Si A es independiente de B, demostrar que $\bar{A}$ también es independiente de B.

Solución:

$$(AB) = \frac{(A)(B)}{N} \implies (\bar{A}B) = \frac{(\bar{A})(B)}{N}$$

./.

Demostración:

$$(B) - (AB) = (B) - \frac{(A)(B)}{N}$$

$$(\bar{A}B) = \frac{(B)N - (A)(B)}{N}$$

$$(A\bar{B}) = \frac{(B)[N - (A)]}{N}$$

$$(\bar{A}\bar{B}) = \frac{(B)(\bar{A})}{N}$$

Ejemplo 29: Si A y B son excluyentes ¿ deben ser independientes?
 res? Si A y B son independientes ¿ deben ser excluyentes?

Solución:

a) No

$$(A)(B) = (AB)N$$

$$\neq 0 \neq 0$$

b) No

$$(AB) \neq 0$$

A un grupo de 100 personas se les pregunta acerca de sus preferencias por las flores. A 30 personas les gustan las azucenas; a 20 las begonias y a 40 los crisantemos; a 6 les gustan las dos primeras de estas flores; a 12 la primera y la última; y a 8 las dos últimas. Finalmente, a una le gustan las tres flores mencionadas.

A base de lo anterior, se pide indicar si los gustos por estas flores

- a) Son independientes dos a dos
- b) Son mutuamente independientes.

Solución:

$$N = 100 \quad (AB) = 6$$

$$(A) = 30 \quad (AC) = 12$$

$$(B) = 20 \quad (BC) = 8$$

$$(C) = 40 \quad (ABC) = 1$$

$$(AB) = \frac{(A)(B)}{N}$$

$$6 = \frac{30 \times 20}{100} = 6$$

$$(AC) = \frac{(A)(C)}{N}$$

$$12 = \frac{40 \times 30}{100} = 12$$

$$(BC) = \frac{(B)(C)}{N}$$

$$8 = \frac{20 \times 40}{100} = 8$$

Son independientes 2 a 2

$$(ABC) = \frac{(A)(B)(C)}{N}$$

$$1 \neq \frac{30 \times 20 \times 40}{100} = 240$$

No son mutuamente independientes

3.3 Asociación de Atributos.

Definición: 2 atributos (A) y (B) están asociados en un colectivo N si se cumple la siguiente desigualdad:

$$(AB) > \frac{(A)(B)}{N}$$

También se dice que están asociados positivamente. Cuando la desigualdad se da en el otro sentido se dice que están asociados o asociados negativamente, es decir, en otros términos:

$$(AB) < \frac{(A)(B)}{N}$$

A partir de estas definiciones podremos encontrar los criterios que nos van a decir de qué forma están asociados 2 atributos ya que la expresión numérica de (AB) dirá si hay asociación o disociación, y en qué medida.

Las propiedades que se han visto, merecen un especial tratamiento, en estudios relacionados con el campo científico o determinadas técnicas, y es aquí donde se puede apreciar la utilidad de los criterios de asociación. En Economía se puede proponer el estudio de uno o más cultivos en lo relativo a la siembra o evolución de un producto, usando fertilizantes o no. Otro ejemplo el de una droga determinada, puede servir para explicar la asociación entre dos atributos, la efectividad de una vacuna, una vitamina, etc.

La intensidad de asociación entre los atributos (A) y (B) depende de los dos límites en los cuales puede encontrarse (AB) siendo el superior igual a la frecuencia (A) o (B) pero en todo caso siempre será el menor de los dos; para el límite inferior puede tomar el valor cero o el valor (A), (B) - N; en todo caso, el mayor de ellos. En la medida en que (AB) se aproxime o aleje del valor (A) (B) / N habrá asociación o desociación más acentuada.

Ejemplo 30: En una fiesta hay más mujeres que hombres, hay más hombres rubios que mujeres morenas; hay menos mujeres rubias solteras que hombres morenos casados. Demostrar que hay más mujeres rubias casadas que hombres morenos solteros.

Datos:

$$\left. \begin{array}{l} (M) > (\bar{M}) \\ (\bar{M}R) > (\bar{M}\bar{R}) \\ (\bar{M}\bar{R}S) > (MRS) \end{array} \right\} \text{donde: } \begin{array}{l} M = \text{mujer} \\ R = \text{rubio} \\ S = \text{soltero} \end{array}$$

Solución:

Se desea demostrar que:

$$(M\bar{R}S) > (\bar{M}RS)$$

Sabemos que:

$$(M) > (\bar{M})$$

$$(MR) + (M\bar{R}) > (\bar{M}R) + (\bar{M}\bar{R})$$

pero

$$(\bar{M}R) > (M\bar{R}) \text{ reemplazamos:}$$

$$(MR) + (\bar{M}R) > (\bar{M}R) + (\bar{M}\bar{R})$$

$$(MR) > (\bar{M}\bar{R})$$

$$(MR\bar{S}) + (MRS) > (\bar{M}\bar{R}\bar{S}) + (\bar{M}\bar{R}S)$$

pero

$$(\bar{M}\bar{R}\bar{S}) > (MRS) \text{ reemplazando:}$$

$$(MR\bar{S}) + (\bar{M}\bar{R}\bar{S}) > (\bar{M}\bar{R}\bar{S}) + (\bar{M}\bar{R}S)$$

$$(MR\bar{S}) > (\bar{M}\bar{R}S)$$

Ejemplo 31: Si A se presenta entre los B con mayor frecuencia que entre los $\bar{B}$, demuéstrese que B se presenta con mayor frecuencia entre los A que entre los $\bar{A}$.

$$\text{Si: } \frac{(AB)}{(B)} > \frac{(A\bar{B})}{(\bar{B})}; \text{ demostrar que: } \frac{(AB)}{(A)} > \frac{(\bar{A}B)}{(\bar{A})}$$

$$\frac{(AB)}{(B)} > \frac{(A\bar{B})}{(\bar{B})}$$

∴

$$\frac{(AB)}{(A\bar{B})} > \frac{(B)}{(\bar{B})}$$

$$\frac{(AB)}{(AB) + (A\bar{B})} > \frac{(B)}{(B) + (\bar{B})}$$

$$\frac{(AB)}{(A)} > \frac{(B)}{(N)}$$

$$\frac{(AB)}{(B)} > \frac{(A)}{(N)}$$

$$\frac{(AB)}{(B) - (AB)} > \frac{(A)}{(N) - (A)}$$

$$\frac{(AB)}{(\bar{A}\bar{B})} > \frac{(A)}{(\bar{A})}$$

$$\frac{(AB)}{(A)} > \frac{(\bar{A}\bar{B})}{(\bar{A})}$$

Otra forma:

$$\frac{(AB)}{(B)} > \frac{(\bar{A}\bar{B})}{(\bar{B})}$$

$$\frac{(AB)}{(B) - (AB)} > \frac{(\bar{A}\bar{B})}{(\bar{B}) - (\bar{A}\bar{B})}$$

$$\frac{(AB)}{(\bar{A}B)} > \frac{(\bar{A}\bar{B})}{(\bar{A}B)}$$

$$\frac{(AB)}{(\bar{A}\bar{B})} > \frac{(\bar{A}\bar{B})}{(\bar{A}\bar{B})}$$

$$\frac{(AB)}{(AB) + (\bar{A}\bar{B})} > \frac{(\bar{A}\bar{B})}{(\bar{A}\bar{B}) + (\bar{A}\bar{B})}$$

$$\frac{(AB)}{(A)} > \frac{(\bar{A}\bar{B})}{(\bar{A})}$$

Ejemplo 32:

Si: $(A) = (B) = (\bar{A}) = (\bar{B}) = \frac{n}{2}$

a) Demostrar que $(AB) = (\bar{A}\bar{B})$ y $(\bar{A}B) = (A\bar{B})$

b) ¿Qué valor tiene (AB) si $(\bar{A}\bar{B}) = n/4$?

Solución:

a) Si: $(B) = (\bar{A})$; será

$$(AB) + (\bar{A}B) = (\bar{A}\bar{B}) + (\bar{A}B)$$

$$(AB) = (\bar{A}\bar{B})$$

Si: $(A) = (B)$; será:

$$(AB) + (A\bar{B}) = (A\bar{B}) + (\bar{A}B)$$

$$(A\bar{B}) = (\bar{A}B)$$

b) $(A) = (B) = \frac{n}{2}$

$$(AB) + (\bar{A}B) = (B)$$

$$(AB) = (B) - (\bar{A}B)$$

$$(AB) = \frac{n}{2} - \frac{n}{4} = \frac{n}{4}$$

3.3.1 Coeficiente de Asociación: El coeficiente de asociación (expresado mediante el símbolo C) especifica de una manera precisa si existe o no asociación entre atributos y, si existe, determinar el signo.

Ya se ha estudiado anteriormente el criterio de independencia como:

$$(AB)_i = \frac{(A)_i (B)_i}{N}$$

Cuando existe asociación se tiene que:

$$(AB) > \frac{(A)_i (B)_i}{N} \Rightarrow (AB) > (AB)_i$$

la desigualdad anterior puede convertirse en igualdad si se suma una cantidad r

$$(AB) = (AB)_i + r$$

si se restan de un mismo miembro los términos anteriores:

$$(A) - (AB) = (A) - \left[\frac{(A)(B)}{N} + r \right]$$

$$(A\bar{B}) = \frac{(A) [N - (B)]}{N} - r = \frac{(A)(\bar{B})}{N} - r$$

$$(A\bar{B}) = (A\bar{B})_i - r$$

Análogamente se tienen para las frecuencias $(\bar{A}B)$ y $(\bar{A}\bar{B})$ las siguientes igualdades:

$$(\bar{A}B) = (\bar{A}B)_i - r$$

$$(\bar{A}\bar{B}) = (\bar{A}\bar{B})_i + r$$

Entonces el valor de r puede expresarse así:

$$r = (AB) - (AB)_i = (\bar{A}\bar{B})_i - (\bar{A}\bar{B}) = (A\bar{B})_i - (A\bar{B}) = (\bar{A}B) - (\bar{A}B)_i$$

La constante r puede ser hallada de otra manera:

$$r = (AB) - (AB)_i \implies r = (AB) - \frac{(A)(B)}{N}$$

$$r = \frac{N(AB) - (A)(B)}{N}$$

$$r = \frac{1}{N} \left\{ \left[(AB) + (A\bar{B}) + (\bar{A}B) + (\bar{A}\bar{B}) \right] (AB) - \left[(AB) + (A\bar{B}) \right] \left[(AB) + (\bar{A}B) \right] \right\}$$

Realizando las operaciones indicadas se llega a:

$$r = \frac{1}{N} \left[(AB)(\bar{A}\bar{B}) - (A\bar{B})(\bar{A}B) \right]$$

Reemplazando N por su equivalente:

$$C = \frac{(AB)(\bar{A}\bar{B}) - (A\bar{B})(\bar{A}B)}{(AB)(\bar{A}\bar{B}) + (A\bar{B})(\bar{A}B)}$$

C se llama el coeficiente de asociación, tomando valores en el intervalo $-1, 1$ así:

Si $r = 0$ hay independencia

Si $-1 \leq r \leq 0$ Asociación negativa (disociación)

Si $0 < r \leq 1$ Asociación positiva (asociación)

Existe otro coeficiente que permite calcular el grado de asociación entre dos atributos. Este coeficiente recibe el nombre de COLIGACION.

$$Y = \frac{1 - \sqrt{\frac{(A\bar{B})(\bar{A}B)}{(AB)(\bar{A}\bar{B})}}}{1 + \sqrt{\frac{(A\bar{B})(\bar{A}B)}{(AB)(\bar{A}\bar{B})}}}$$

Ejemplo 33: En una región X se entrevistaron 40 familias con el fin de conocer el tipo de vivienda y el número de hijos. Se obtuvieron los resultados que aparecen a continuación. Calcular el coeficiente de asociación.

Familia \ Vivienda	B	$\bar{B}$	TOTALES
A Con Hijos	20	0	20
$\bar{A}$ Sin Hijos	17	3	20
TOTALES	37	3	40

$$C = \frac{(20)(3) - (17)(0)}{(20)(3) + (17)(0)} = +1$$

El coeficiente igual a + 1 indica que la asociación es perfecta.

En otra región los resultados fueron:

Famílias \ Vivienda	B	$\bar{B}$	Total
	Urbana	Rural	
A Con Hijos	5	15	20
$\bar{A}$ Sin Hijos	5	15	20
TOTAL	10	30	40

$$C = \frac{(5)(15) - (5)(15)}{(5)(15) + (5)(15)} = 0$$

Un coeficiente nulo significa que los resultados son independientes.

En otra región se obtuvo:

Vivienda Familia	B	$\bar{B}$	Totales
	Urbana	Rural	
A Con Hijos	10	10	20
$\bar{A}$ Sin Hijos	0	20	20
TOTAL	10	30	40

$$C = \frac{(10)(20) - (10)(0)}{(10)(20) + (10)(0)} = +1$$

El resultado obtenido indica que existe asociación completa en las frecuencias negativas.

Otra región estudiada arrojó los resultados siguientes:

Vivienda Familia	B	$\bar{B}$	Total
	Urbana	Rural	
A Con Hijos	20	0	20
$\bar{A}$ Sin Hijos	0	20	20
TOTAL	20	20	40

$$C = \frac{(20)(20) - (0)(0)}{(20)(20) + (0)(0)} = +1$$

Este es un caso extremo de asociación perfecta en las frecuencias positivas y negativas.

En otra región los resultados quedaron así:

Familia \ Vivienda	B	$\bar{B}$	Total
	Urbana	Rural	
A Con Hijos	11	9	20
$\bar{A}$ Sin Hijos	20	0	20
TOTAL	31	9	40

$$C = \frac{(11)(0) - (9)(20)}{(11)(0) + (9)(20)} = -1$$

Este resultado significa que existe disociación entre las frecuencias negativas.

Para otra región el resultado fue:

Familia \ Vivienda	B Urbana	$\bar{B}$ Rural	Total
A Con Hijos	0	0	0
$\bar{A}$ Sin Hijos	20	20	40
TOTAL	20	20	40

$$C = \frac{(0)(0) - (20)(20)}{(0)(0) + (20)(20)} = -1$$

Este resultado indica que existe disociación completa en las frecuencias positivas y negativas.

3.4 Asociaciones Parciales: Pueden los atributos A y B estar asociados con un nuevo atributo C, o con más de ellos; la asociación entre A y B puede ser marcada o definida, pero si se considera el atributo C, la relación puede cambiar; por lo tanto, a este tipo de relación existente entre los A y B con el atributo C se le denomina una ASOCIACION PARCIAL, que puede ser de la forma siguiente:

$$(ABC) > \frac{(AC)(BC)}{(C)} \Rightarrow \text{Asociación}$$

$$(ABC) < \frac{(AC)(BC)}{(C)} \Rightarrow \text{Disociación}$$

$$(ABC) = \frac{(AC)(BC)}{(C)} \Rightarrow \text{Independencia}$$

Si el número de atributos se ha aumentado ahora a 4, la asociación se mide así:

$$(ABCD) \neq \frac{(ACD)(BCD)}{(CD)}$$

proceso que se continúa cuando el número de atributos en consideración sea mayor.

Ejemplo 34: Asociación entre vacunación e inmunidad contra el cólera en seis epidemias distintas.

	No Atacaos	Atacados	Totales
Vacunados	276	3	279
No vacunados	473	66	539
Totales	749	69	818
Vacunados	192	4	196
No vacunados	113	34	147
Totales	305	38	343
Vacunados	5.751	27	5.778
No vacunados	6.351	198	6.549
Totales	12.102	225	12.327
Vacunados	4.087	5	4.092
No vacunados	113.856	1.144	115.000
Totales	117.943	1.149	119.092
Vacunados	8.332	8	8.340
No vacunados	84.444	556	85.000
Totales	92.776	564	93.340
Vacunados	4.870	5	4.875
No vacunados	153.096	904	154.000
Totales	157.966	909	158.875

La cuestión importante que aquí se plantea es saber hasta qué punto la vacunación protege contra el cólera. Así que la comparación más natural es:

Porcentaje de vacunados que no fueron atacados

Porcentaje de no vacunados que fueron atacados

$$\frac{(AB)}{(A)}, \frac{(\bar{A}B)}{(\bar{A})} \text{ para todas las epidemias } \frac{(AB)}{(A)} > \frac{(\bar{A}B)}{(\bar{A})}$$

PROPORCION DE NO ATACADOS

Epidemias	Vacunados	No Vacunados	Diferencia
1	0.9892	0.8776	0.1116
2	0.9796	0.7687	0.2109
3	0.9953	0.9698	0.0255
4	0.9988	0.9901	0.0087
5	0.9990	0.9935	0.0055
6	0.9990	0.9941	0.0049

En todos los casos existe asociación positiva de vacunación e inmunidad.

3.5 Asociaciones Ilusorias: Se da este nombre a la asociación existente entre los atributos A y C; B y C, aunque haya una independencia entre los atributos A y B, es decir, que los atributos A y B son independientes, pero -

cúando se asocian con un nuevo atributo C la combinación de ellos con éste no es independiente.

Puede presentarse el caso en el cual el atributo A está asociado positivamente con el atributo C, y éste esté asociado negativamente con B; entonces, existe una asociación negativa ilusoria entre los atributos A y B. Algunas veces puede presentarse una distorsión en los resultados, debido a que el observador comete errores de confusión en la clasificación y tiende a anotar la presencía de A como si fuera B y viceversa; siendo C el ob-servador, se tienen resultados poco satisfactorios.

4. CLASIFICACION MULTIPLE

- 4.1 Consideraciones Generales. En lugar de dividir el colectivo en dos partes, como hasta aquí se ha hecho, a través de una clasificación dicotómica, se puede dividir en mayor número de partes; mediante un procedimiento análogo, se amplía la clasificación dicotómica de un colectivo de hombres de ojos azules y hombres de ojos no azules, estableciendo una clasificación triple:

1) hombres de ojos azules; 2) hombres de ojos pardos
y 3) hombres de ojos, ni azules ni pardos.

En general se puede dividir el colectivo en relación con
S rúbricas o S clases:

$$A_1, A_2, A_3, A_4, \dots, A_t$$

a su vez éstas se dividen en otras clases:

$$B_1, B_2, B_3, \dots, B_t$$

a su vez éstas se agrupan en otras:

$$C_1, C_2, C_3, C_4, \dots, C_u$$

Lo mismo que se hizo en relación con la clasificación
dicotómica, puede aplicarse a la clasificación múltiple.

Estos principios aplicables a la clasificación múltiple se
pueden resolver de la siguiente manera:

a) Existen $S \times T \times U \dots$ número de frecuencias de cla
se. últimas.

b) El número total de clases, incluyendo N y las clases últimas es:

$$(S + 1)(T + 1)(U + 1).$$

c) Los datos serán consistentes, si ninguna frecuencia última es negativa.

d) Los datos quedan completamente clasificados, mediante $S \times T \times U$. . . frecuencias de clase, algebraicamente independientes.

4.2 Tablas de Contingencia.

Supongamos que la clasificación múltiple de los A es de orden S y la B de orden T, habrá pues $S \times T$ clases de $A_m B_n$.

Generalizado en parte la notación de los capítulos precedentes representamos por (A_m) la frecuencia de los individuos A_m y por $(A_m B_n)$ la frecuencia de clase de los individuos $A_m B_n$.

De esta manera podremos construir una tabla de T filas y S columnas.

ATRIBUTOS	A_1	A_2	-	-	A_{s-1}	A_s	TOTAL
B_1	$(A_1 B_1)$	$(A_2 B_1)$	-	-	$(A_{s-1} B_1)$	$(A_s B_1)$	(B_1)
B_2	$(A_1 B_2)$	$(A_2 B_2)$	-	-	$(A_{s-1} B_2)$	$(A_s B_2)$	(B_2)
$\vdots$							
B_t	$(A_1 B_t)$	$(A_2 B_t)$	-	-	$(A_{s-1} B_t)$	$(A_s B_t)$	(B_t)
TOTAL	(A_1)	(A_2)	-	-	(A_{s-1})	(A_s)	N

4.3 Asociación en las Tablas de Contingencia. La asociación en estas tablas se determina a través de los coeficientes que reciben su mismo nombre. Si los A y los B son independientes se cumplirá la siguiente condición de identidad:

$$(A_m B_n) = \frac{(A_m)(B_n)}{N} = (A_m B_n)_i$$

de no cumplirse la identidad anterior es decir:

$$(A_m B_n) \neq (A_m B_n)_i$$

porque hay asociación entre los A y los B y se representa por la diferencia de los dos términos, siendo igual esta diferencia a δ

$$\delta = (A_m B_n) - (A_m B_n)_i$$

sin embargo, se han ideado unos coeficientes especiales que indican el grado de asociación. Un coeficiente es una expresión matemática con la cual se mide el grado de asociación para una clasificación dicotómica o múltiple ordenada ; si ésta se encuentra compuesta por una tabla de m filas y n columnas se dice que dicho coeficiente se encuentra localizado entre 2 límites.

$$[-1, 1]$$

Pearson propuso un coeficiente de contingencia, cuyo valor se aproxima a cada uno de los límites, en la medida que esté variando tanto el número de filas, como el número de columnas. Si el número de clases que integran la tabla tiende a ser grande, entonces el coeficiente se aproximará a su límite superior (1). Este coeficiente está dado por la expresión:

$$C = \sqrt{\frac{X^2}{N + X^2}}$$

haciendo la deducción matemática de este coeficiente, se presenta de la siguiente manera, para efectos de cálculo:

$$C = \sqrt{\frac{S - N}{S}} \quad \text{en donde } S = \sum \left\{ \frac{(A_m B_n)^2}{(A_m B_n)_i} \right\}$$

4.4 Interpretación del Coeficiente de Concingencia de Pearson.

Cuando los A y los B son independientes, su resultado es igual a cero. Este coeficiente fluctúa entre - 1 y + 1 :

$$(0 \text{ y } 1) \text{ ó } (0 \text{ y } -1)$$

Cuando es igual a 1, la asociación es perfecta, también puede adquirir valores entre 0 y - 1, sería el caso de la disociación; cuando su valor es igual a - 1 se dice que la disociación es perfecta.

Ejemplo 35: Se ha hecho la siguiente clasificación de un colectivo de hombres, clasificados de acuerdo al color del pelo y de los ojos. Al construir la respectiva tabla de contingencia, se ha llegado a los siguientes resultados:

Color Cabello (A _m) Color Ojos (B _n)	Rubio (A ₁)	Castaño (A ₂)	Negro (A ₃)	Rojo (A ₄)	Total
Azules (B ₁)	1.768	807	189	47	2.811
Gris-Verde (B ₂)	946	1.387	746	53	3.132
Pardos (B ₃)	115	438	288	16	857
TOTAL	2.829	2.632	1.223	116	6.800

Calcular el coeficiente de contingencia de Pearson, de la tabla anterior:

$$C = \sqrt{\frac{S - N}{S}}; S = \sum \left\{ \frac{(A_m B_n)^2}{(A_m B_n)_i} \right\}; (A_m B_n)_i = \frac{(A_m)_i (B_n)_i}{N}$$

$$(A_1 B_1)_i = \frac{2.829 \times 2.811}{6.800} = 1.169; (A_2 B_2)_i = \frac{2.632 \times 3.132}{6.800} = 1.212$$

$$(A_1 B_2)_i = \frac{2.829 \times 3.132}{6.800} = 1.303; (A_2 B_3)_i = \frac{2.632 \times 857}{6.800} = 332$$

$$(A_1 B_3)_i = \frac{2.829 \times 857}{6.800} = 357; (A_3 B_1)_i = \frac{1.223 \times 2.811}{6.800} = 506$$

$$(A_2 B_1)_i = \frac{2.632 \times 2.811}{6.800} = 1.088; (A_3 B_2)_i = \frac{1.223 \times 3.132}{6.800} = 563$$

$$(A_3 B_3)_i = \frac{1.223 \times 857}{6.800} = 154$$

$$(A_4 B_1)_i = \frac{116 \times 2.811}{6.800} = 48$$

$$(A_4 B_2)_i = \frac{116 \times 3.132}{6.800} = 53$$

$$(A_4 B_3)_i = \frac{116 \times 857}{6.800} = 15$$

$(A_m B_n)_i$			
1.169	1.088	506	48
1.303	1.212	563	53
357	332	154	15

$$S = \sum \left\{ \frac{(A_m B_n)^2}{(A_m B_n)_i} \right\} ; \quad \frac{(1.768)^2}{1.169} = 2.674$$

$$\frac{(946)^2}{1.303} = 687 ; \quad \frac{(189)^2}{506} = 71$$

$$\frac{(115)^2}{357} = 37 \quad ; \quad \frac{(288)^2}{154} = 539$$

$$\frac{(807)^2}{1.088} = 599 \quad ; \quad \frac{(16)^2}{15} = 17$$

$$\frac{(1387)^2}{1.212} = 1.587 \quad ; \quad \frac{(47)^2}{48} = 46$$

$$\frac{(23)^2}{332} = 578 \quad ; \quad S = 7.876$$

$$\frac{(746)^2}{563} = 988 \quad ;$$

$$C = \sqrt{\frac{7.876 - 6.800}{7.876}}$$

$$\frac{(53)^2}{53} = 53 \quad ;$$

$$C = \sqrt{\frac{1.076}{7.876}}$$

$$C = \sqrt{0.137} \quad ; \quad C = 0.370$$

4.3 Coefficiente de Tschuprow Otro coeficiente que permite establecer el grado de contingencia es el de Tschuprow, quien propuso un coeficiente de contingencia que se aproximara a 1 cuando el número de columnas es igual al número de filas, de tal manera que para aplicarlo tendría un mejor significado cuando la tabla fuera $m \times m$ o $n \times n$.

El cual se define de la siguiente forma:

$$T^2 = \frac{\varphi^2}{\sqrt{(S-1)(T-1)}}$$

y para efectos de cálculo:

$$T^2 = \frac{C^2}{(1 - C^2) \sqrt{(S-1)(T-1)}}$$

Ejemplo 36: Del ejercicio anterior calcular el coeficiente de Tschuprow:

$$T^2 = \frac{0.1370}{(1 - 0.1370) \sqrt{(3)(2)}} = \frac{0.1370}{(0.8630) \sqrt{6}}$$

$$T^2 = \frac{0.1370}{(0.8630)(2.44)} = \frac{0.1370}{2.1057} = 0.0651$$

$$T = \sqrt{0.0651} = 0.255$$

Ejemplo 37: Con los datos de la siguiente tabla calcular el coeficiente de contingencia de Pearson y de Tschuprow:

$\begin{matrix} A_m \\ B_n \end{matrix}$	A_1	A_2	A_3	A_4	TOTAL
B_1	2	1	12	9	24
B_2	4	5	10	8	27
B_3	3	6	11	7	27
TOTAL	9	12	33	24	78

$$C = \frac{S - N}{S} \qquad S = \frac{(A_m B_n)^2}{(A_m B_n)_i}$$

$$(A_m B_n)_i = \frac{(A_m)(B_n)}{N}$$

$$(A_m B_n)_i$$

2,76	3,69	10,15	7,38
3,11	4,15	11,42	8,31
3,11	4,15	11,42	8,31

$$S = \sum \left\{ \frac{(Am Bn)^2}{(Am Bn)_i} \right\}; \quad \frac{4}{2,76} = 1,45 \quad \frac{25}{4,15} = 6,02 \quad \frac{121}{11,42} = 10,60$$

$$\frac{16}{3,11} = 5,14 \quad \frac{36}{4,15} = 8,67 \quad \frac{81}{7,38} = 10,98$$

$$\frac{9}{3,11} = 2,89 \quad \frac{144}{10,15} = 14,19 \quad \frac{64}{8,31} = 7,70$$

$$\frac{1}{3,69} = 0,27 \quad \frac{100}{11,42} = 8,75 \quad \frac{49}{8,31} = 5,90$$

$$S = 82.56$$

$$C = \sqrt{\frac{82.56 - 78}{82.56}} \quad C = \sqrt{\frac{4.56}{82.56}}$$

$$C = \sqrt{0.055} = 0.234$$

El coeficiente de contingencia de Pearson es 0.234, significa que si hay asociación entre A y B.

$$T^2 = \frac{C^2}{(1 - C^2) \sqrt{(S - 1)(T - 1)}}$$

$$T^2 = \frac{0.055}{(1 - 0.055) \sqrt{(3)(2)}}$$

$$T^2 = \frac{0.055}{(0.945) \sqrt{6}} = \frac{0.055}{(0.945)(2.44)}$$

$$T^2 = \frac{0.055}{2.306} = 0.024$$

$$T = \sqrt{0.024} = 0.155$$

El coeficiente de contingencia de Tschuprow, 0.155 significa que hay asociación entre A y B.

Ejemplo 38: El gobierno de un país solicitó o respecto del grado de asociación que habría entre el estado de las viviendas (adelanto de la edificación y sanidad de ellas) y el carácter urbano de las comunas (o municipios) Los estadísticos encargados del trabajo lograron construir la siguiente tabla:

(Miles)

Viviendas (Am) Comunas (Bn)	Habita bles	Inhabita bles	En cons- trucción	Total
Total	493	272	35	800
Comunas Rurales	78	163	11	252
Otras Comunas Rurales	201	71	16	288
Comunas de la Capital	214	38	8	260

Se pide:

- Calcular los coeficientes de contingencia de Pearson y de Tschuprow.
- Determinar los límites de variación de cada uno de los coeficientes de contingencia que se consideran en la pregunta anterior.

Solución:

a)

$$C = \sqrt{\frac{S - N}{S}}$$

$$S = \sum \left\{ \frac{(A_m B_n)_i^2}{(A_m B_n)_i} \right\}$$

$(A_m B_n)_i = \frac{(A_m) \cdot (B_n)_i}{N}$	$(A_m B_n)_i^2$	$\frac{(A_m B_n)_i^2}{(A_m B_n)_i}$
$(A_1 B_1)_1 = \frac{493 \times 252}{800} = \frac{124.236}{800} = 155$	$(A_1 B_1)_1^2 = (78)^2 = 6.084$	$\frac{6.084}{153} = 39$
$(A_1 B_2)_1 = \frac{493 \times 288}{800} = \frac{141.984}{800} = 177$	$(A_1 B_2)_1^2 = (20)^2 = 40.401$	$\frac{40.401}{183} = 22$
$(A_1 B_3)_1 = \frac{493 \times 260}{800} = \frac{126.180}{800} = 157$	$(A_1 B_3)_1^2 = (214)^2 = 45.796$	$\frac{45.796}{160} = 286$

$(A_2 B_1)_i = \frac{272 \times 252}{800} = \frac{68.544}{800} = 86$	$(A_2 B_1)^2 = (163)^2 = 26.569$	$\frac{26.569}{86} = 309$
$(A_2 B_2)_i = \frac{272 \times 288}{800} = \frac{78.336}{800} = 98$	$(A_2 B_2)^2 = (71)^2 = 5.041$	$\frac{5.041}{98} = 51$
$(A_2 B_3)_i = \frac{272 \times 260}{800} = \frac{70.720}{800} = 88$	$(A_2 B_3)^2 = (38)^2 = 1.444$	$\frac{1.444}{88} = 16$
$(A_3 B_1)_i = \frac{35 \times 252}{800} = \frac{8.820}{800} = 11$	$(A_3 B_1)^2 = (11)^2 = 121$	$\frac{121}{11} = 11$
$(A_3 B_2)_i = \frac{35 \times 288}{800} = \frac{10.080}{800} = 13$	$(A_3 B_2)^2 = (16)^2 = 256$	$\frac{256}{13} = 20$
$(A_3 B_3)_i = \frac{35 \times 260}{800} = \frac{9.100}{800} = 11$	$(A_3 B_3)^2 = (8)^2 = 64$	$\frac{64}{11} = 6$
		$= 966 = S$

$$C = \sqrt{\frac{966 - 800}{966}}$$

$$T^2 = \frac{C^2}{(1 - C^2) \sqrt{(s-1)(t-1)}}$$

$$C = \sqrt{\frac{166}{966}}$$

$$T^2 = \frac{0,172}{(0,828) \sqrt{(3-1)(3-1)}}$$

$$C = \sqrt{0,172}$$

$$T^2 = \frac{0,172}{(0,828) \sqrt{4}}$$

$$C = 0,415$$

$$T^2 = \frac{0,172}{0,828 (2)}$$

$$C^2 = 0,172$$

$$T^2 = \frac{0,172}{1,656} = 0,104$$

$$T^2 = 0,104$$

$$b) \quad 0 \leq C^2 \leq 1 - \frac{1}{q}$$

$$0 \leq C^2 \leq 1 - \frac{1}{3}$$

$$0 \leq C^2 \leq \frac{2}{3}$$

$$0 \leq T^2 \leq \frac{\min(s_1 - 1, t_2 - 1)}{\sqrt{(s_1 - 1)(t_2 - 1)}}$$

$$0 \leq T^2 \leq \frac{2}{2}$$

$$0 \leq T^2 \leq 1$$

Ejemplo 39: El Departamento de Orientación Pedagógica de un colegio secundario examinó a un grupo de alumnos y a sus respectivos padres, clasificándolos en cada caso según la materia (Historia, Idiomas o Matemáticas) para la que ellos tenían la mayor facilidad. El resultado fue:

Alumnos Padres	Historia	Idiomas	Matemáticas	Total
Total	22	32	20	74
Matemáticas	6	8	6	20
Idiomas	7	15	9	31
Historia	9	9	5	23

- a) Determinar el grado de relación entre la facilidad para las materias que se indican, de los alumnos respecto de sus respectivos padres.
- b) Qué relación existe entre los alumnos con la mayor facilidad para las matemáticas y los padres con máxima facilidad para la historia ?

Solución:

$$a) C = \sqrt{\frac{S - N}{S}}$$

$$S = \sum \left\{ \frac{(Am Bn)^2}{(Am Bn)_i} \right\}$$

$(Am Bn)_i = \frac{(Am)(Bn)}{n}$	$(Am Bn)^2$	$S = \sum \frac{(Am Bn)^2}{(Am Bn)_i}$
$(A_1 B_1)_i = \frac{20 \times 22}{74} = \frac{440}{74} = 5.94$	$(A_1 B_1)^2 = (6)^2 = 36$	$\frac{36}{5.94} = 6.06$
$(A_1 B_2)_i = \frac{20 \times 32}{74} = \frac{640}{74} = 8.64$	$(A_1 B_2)^2 = (8)^2 = 64$	$\frac{64}{8.64} = 7.40$
$(A_1 B_3)_i = \frac{20 \times 20}{74} = \frac{400}{74} = 5.40$	$(A_1 B_3)^2 = (6)^2 = 36$	$\frac{36}{5.40} = 6.66$

$(A_2 B_1)_i = \frac{22 \times 31}{74} = \frac{682}{74} = 9.21$	$(A_2 B_1)^2 = (7)^2 = 49$	$\frac{49}{9.21} = 5.32$
$(A_2 B_2)_i = \frac{32 \times 31}{74} = \frac{992}{74} = 13.40$	$(A_2 B_2)^2 = (15)^2 = 225$	$\frac{225}{13.40} = 16.79$
$(A_2 B_3)_i = \frac{20 \times 31}{74} = \frac{620}{74} = 8.37$	$(A_2 B_3)^2 = (9)^2 = 81$	$\frac{81}{8.37} = 9.67$
$(A_3 B_1)_i = \frac{22 \times 23}{74} = \frac{506}{74} = 6.83$	$(A_3 B_1)^2 = (9)^2 = 81$	$\frac{81}{6.83} = 11.85$
$(A_3 B_2)_i = \frac{32 \times 23}{74} = \frac{736}{74} = 9.94$	$(A_3 B_2)^2 = (9)^2 = 81$	$\frac{81}{9.94} = 8.14$
$(A_3 B_3)_i = \frac{20 \times 23}{74} = \frac{460}{74} = 6.21$	$(A_3 B_3)^2 = (5)^2 = 25$	$\frac{25}{6.21} = 4.02$
$S = 75.91$		

$$\frac{(Am Bn)^2}{(Am Bn)_i} = 6.66 + 5.32 + 16.79 + 9.67 + 11.85 + 8.14 + 4.02 + 6.06 + 7.40$$

$$S = 75.91$$

$$C = \sqrt{\frac{75.91 - 74}{75.91}} = \sqrt{\frac{1.91}{75.91}} = \sqrt{0.025} = 0.16$$

$$T^2 = \frac{C^2}{(1 - C^2) (\sqrt{(s-1)(t-1)})} = \frac{0.025}{(0.975) \cdot (2) \cdot 1.95} = 0.012$$

$$b) (A_3 B_3) = 5$$

$$(A_3 B_3) = \frac{(A_3)(B_3)}{N} + r$$

$$5 = \frac{(20)(23)}{74} + r$$

./.

$$5 = \frac{460}{74} + r$$

$$5 - \frac{460}{74} = r$$

$$5 - 6.21 = r$$

$$- 1.21 = r$$

$$r = - 1.21$$

Existe asociación negativa o disociación.

Estos coeficientes de contingencia son muy útiles en diversos trabajos, sin embargo toda tabla deberá ser siempre analizada cuidadosamente , para ver si presentan particularidades claramente significativas en una distribución de frecuencias.

- 4.7 Clasificaciones Homogéneas y Heterogéneas. Las clasificaciones estadísticas para efectos de establecer los grados de asociación o de contingencia, solamente pueden ser de carácter homogéneo. Se llaman clasificaciones homogéneas aquellas que se establecen de acuerdo a un principio de división preciso para todas las subclases. Sin embargo hay muchas clasificaciones completamente heterogéneas, como las clasificaciones biológicas en órdenes, géneros y especies; las clasificaciones por profesiones, por clases de cultivos, por edades, por trabajos, etc.
- 4.8 Clasificación Múltiple. Como Sucesión de Clasificaciones Dicotómicas. Desde el punto de vista teórico, la clasificación múltiple

puede considerarse como una sucesión de clasificaciones dicotómicas.

En efecto se divide primeramente el colectivo en individuos de ojos azules e individuos de ojos no azules; se obtienen dos clases. - Volviendo a dividir en individuos de ojos pardos e individuos con ojos no pardos, se tienen cuatro clases. Pero al aplicar esta operación de clasificación de individuos, a las clases de individuos de ojos azules, una clase resulta nula, porque no puede haber individuos que tengan ojos azules y pardos a la vez. Y otra clase es precisamente la clase de individuos con ojos azules.

Luego se tienen tres clases:

- a) La clase de individuos con ojos azules
- b) La clase de individuos con ojos pardos
- c) La clase que forman el resto de individuos.

Si ahora se divide dicotómicamente cada una de estas clases, incluyendo en una a los individuos de ojos grises y los de ojos no grises, se obtiene de nuevo otras cuatro clases.

CAPITULO IV

DISTRIBUCIONES DE FRECUENCIAS

1. INTRODUCCION

Como el objetivo fundamental de las estadísticas, es el de tratar de resumir en una pequeña información una serie de valores determinados; se intentará entonces, de buscar aquellos valores que representen a todas las observaciones, mediante la "reducción, simplificación, concentración o resumen" de los datos o de la información base.

Hay que tener especial cuidado en cuanto a este criterio de selección, ya que la demasiada concentración lleva a errores de apreciación y a conclusiones falsas, lo mismo que la poca concentración. Esta consolidación de datos estará en razón directa con la índole de investigación que se considera y por lo tanto, si para una determinada muestra o conglomerado se requieren seis (6) grupos diferentes de concentración, se distribuye el total de las observaciones en esos seis grupos. Si la misma observación se enmarca por ejemplo en unos nueve grupos, se

tendrá una mejor información porque la concentración alrededor de cada grupo será mayor; por lo tanto, a mayor número de grupos, se tendrá una mejor información y a menor número de grupos una deficiente información.

Los problemas concretos se deben tener en cuenta para la conformación de los grupos. Una buena información más o menos resumida, depende de características propias del suceso o fenómeno investigado. Como complemento a la selección en grupos se construirá un modelo, en el cual aparezcan simultáneamente los diferentes grupos y la importancia de cada uno de ellos. Esa importancia se medirá en función del número de datos que convergen hacia cada grupo; por lo tanto el grupo más importante será aquel que tiene un mayor número de datos, si se observa que algunos grupos tienen un bajo número de observaciones, se deben desplazar a otros valores más importantes o simplemente eliminarlos; para lo cual es necesario tener un amplio conocimiento sobre el fenómeno y tener la certeza de que su eliminación no afectará los resultados finales.

La conformación de grupos en la forma descrita, es lo que se llama DISTRIBUCIONES DE FRECUENCIA. Se utilizan variables

(cuyas características se trataron en el capítulo II) como: peso, estatura, color de los ojos, ingresos, consumo e inversión, etc. que pueden ser cuantitativas o cualitativas según la naturaleza del estudio.

Si la variable es cualitativa los grupos tendrán función de cualidad o atributo; si es cuantitativa, es posible medirla; puede ser de tipo continuo si los valores de ésta se presentan en cualquiera de los puntos a lo largo del recorrido de la variable, (es una función derivable); o de tipo discreto si los valores de las variables son observados en ciertos y determinados puntos del recorrido de la variable y no es posible encontrarlo en forma continua, (función no derivable).

En resumen, la "variable" continua puede ocurrir en cualquier punto, en tanto que la "discreta" se presenta a saltos, escalones o intervalos.

2. DETERMINACION DE INTERVALOS Y DE GRUPOS DE CLASE.

La primera etapa para cualquier investigación estadística está dada por su naturaleza; una vez determinada, se procede a elegir el número de elementos que servirán de base, para las posterior-

res conclusiones. Estos elementos se ubican en tantos grupos diferentes, como convengan a las circunstancias, denominados " INTERVALOS O - CLASES " y el conjunto de éstos, determina lo que se denomina Valores diferentes de la variable ; es decir que cada intervalo tiene un valor característico de la variable o función que se está estudiando : como se ve visto, los distintos grupos deben tener valores diferentes según el número elegido previamente ; si se determina constante la diferencia entre ellos, pueden calcularse los intervalos en una forma racional.

Para determinar estas clases hay que elegir el número de grupos, clases o intervalos ; para distribuir las observaciones se identifica la variable o la función, y dentro de ésta los diferentes valores que se van a tomar ; se toma el valor inferior y el superior y por su diferencia se obtiene el RAN - GO o RECORRIDO DE LA VARIABLE (R). Siendo $x_{\text{máx}}$, el límite superior x_{min} , el límite inferior, la diferencia entre $x_{\text{máx}}$ y x_{min} da el recorrido de la variable. El número de clases o de intervalos se elige previamente y se designa con la letra m ; la amplitud de cada intervalo se representa con la letra c y su cálculo es el siguiente :

$$C = \frac{x_{\text{máx}} - x_{\text{mín}}}{m} = \text{AMPLITUD DEL INTERVALO}$$

Si m es grande, la amplitud entre un intervalo y el siguiente es relativamente pequeño, es decir, la información para cada intervalo será muy detallada. Si por el contrario m es un valor pequeño, se presentará una gran distancia entre un intervalo y el siguiente, y por lo tanto la información para cada uno de ellos, será concentrada y por consiguiente muy general. Entre estos dos criterios de acuerdo a las necesidades, se elige el m más apropiado para cada caso en particular. Por las anteriores deducciones se infiere en principio que la amplitud de los intervalos debe ser constante en general, aunque no necesariamente.

Si la característica no está homogéneamente distribuída a lo largo de todo el recorrido de la variable, sino que por ejemplo, se observa gran concentración de los datos entre el límite inferior y un límite intermedio al superior, se eligen dentro de estos dos extremos magnitudes de intervalos constantes; y si a partir de este límite y el superior las observaciones son escasas, se toman intervalos de amplitudes no constantés. En general, las amplitudes de los intervalos estarán dadas por las características fundamentales de la presentación de los datos originales.

Hay dos extremos posibles: Cuando la distribución de los datos es homogénea, es conveniente que las amplitudes sean constantes; si la distribución de los datos es heterogénea, los distintos intervalos tienen amplitudes diferentes.

En la práctica, cuando la diferencia entre máx , ... mín no es exactamente divisible por m , para facilitar los cálculos, se hace una pequeña adición de valores en los extremos de:

$$(> \text{máx} + a) \text{ ó } (< \text{mín} - a)$$

en forma tal, que dicha diferencia sea exactamente divisible por m o su resultado función de 0,5; 0,05, etc.

Ejemplo 1: En una distribución el límite inferior ($< \text{mín}$) es 15, el límite superior ($> \text{máx}$) es 43 y el número de clases (m) es 5 calcular la amplitud del intervalo (C):

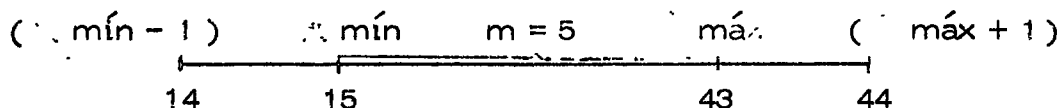
$$C = \frac{> \text{máx} - < \text{mín}}{m}$$

$$C = \frac{43 - 15}{5} = \frac{28}{5} \quad C = 5,6$$

se calcula nuevamente C , haciendo adición de valores en los límites extremos:

$$C = \frac{(\text{máx} + 1) - (\text{mín} - 1)}{m} = \frac{44 - 14}{5} = \frac{30}{5} = 6$$

Gráficamente sería:



Este artificio matemático se utilizará siempre agregando unidades por fuera de los límites reales, con el objeto de que ninguna observación pueda ser eliminada. Con los anteriores conceptos se podrá construir una primera columna correspondiente a lo que se denomina DISTINTOS VALORES DE LA VARIABLE.

Una vez determinados los valores de C , m y R se tabulan los datos, para lo cual es necesario calcular los límites (superior e inferior) de las diferentes clases. En general se denomina Y'_{j-1} el (límite) LIMITE INFERIOR de cada clase y Y'_j el LIMITE SUPERIOR ($j = 1, 2, 3, \dots, m$) de tal manera que C_j queda determinado por la diferencia entre Y'_j y Y'_{j-1} para cada una de las clases.

Los valores inferiores de los intervalos de clase estarán representados por:

$$Y'_0, Y'_1, Y'_2, Y'_3, \dots, Y'_{m-1};$$

superiores por:

$$Y_1, Y_2, Y_3, \dots, Y_m;$$

los promedios de los límites de los intervalos de clase, toman el nombre de "Marcas de clase".

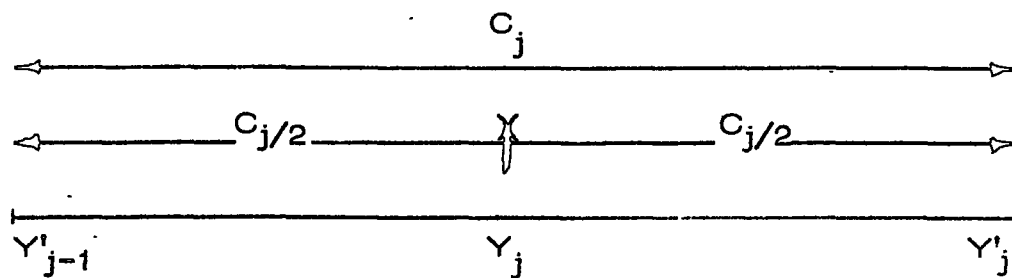
La simbología para cada una de ellas se representa en general por:

$$Y_j \quad (j = 1, 2, \dots, m)$$

La distribución de frecuencias, hasta este momento quedaría - constituida así:

	C L A S E S	M A R C A S D E C L A S E
j	$Y'_{j-1} - Y'_j$	Y_j
1	$Y'_0 - Y'_1$	Y_1
2	$Y'_1 - Y'_2$	Y_2
.	.	.
.	.	.
.	.	.
m	$Y'_{m-1} - Y'_m$	Y_m

Los límites superior e inferior pueden determinarse fácilmente conociendo la marca de clase (Y_j) y la amplitud de clase (C_j). En forma general, para la clase j -ésima tendremos la relación entre sus límites, marca de clase y amplitud, según la gráfica siguiente:



De la gráfica anterior se establecen las siguientes ecuaciones:

$$(1) \quad Y_j = Y'_{j-1} + C_j/2$$

$$(2) \quad Y_j = Y'_j - C_j/2$$

Ejemplo 2: Se tiene una distribución de frecuencias que se refiere a una observación de 19 áreas de fincas.

Hay tres formas posibles con relación a sus límites, así:

INTERVALOS DE
CLASE BASICOS

$$\begin{aligned} Y'_0 &= 30 & Y'_1 &= 34 \\ Y'_1 &= 34 & Y'_2 &= 38 \\ Y'_2 &= 38 & Y'_3 &= 42 \\ Y'_3 &= 42 & Y'_4 &= 46 \\ Y'_4 &= 46 & Y'_5 &= 50 \\ Y'_5 &= 50 & Y'_6 &= 54 \end{aligned}$$

INTERVALOS DE CLASE, DELI-
MITADOS EN EL EXTREMO IN-
FERIOR

$$\begin{aligned} Y'_0 &= 30,1 & Y'_1 &= 34 \\ Y'_1 &= 34,1 & Y'_2 &= 38 \\ Y'_2 &= 38,1 & Y'_3 &= 42 \\ Y'_3 &= 42,1 & Y'_4 &= 46 \\ Y'_4 &= 46,1 & Y'_5 &= 50 \\ Y'_5 &= 50,1 & Y'_6 &= 54 \end{aligned}$$

INTERVALOS DE CLASE DELIMITADOS EN EL EXTREMO

SUPERIOR

$$\begin{aligned} Y'_0 &= 30 & Y'_1 &= 33,9 \\ Y'_1 &= 34 & Y'_2 &= 37,9 \\ Y'_2 &= 38 & Y'_3 &= 41,9 \\ Y'_3 &= 42 & Y'_4 &= 45,9 \\ Y'_4 &= 46 & Y'_5 &= 49,9 \\ Y'_5 &= 50 & Y'_6 &= 53,9 \end{aligned}$$

Se observa que en la primera forma cada Y'_j después de Y'_{j-1} participa como límite superior de un intervalo y como límite inferior del siguiente.

En la segunda forma cada Y'_j es excluido, como límite superior del siguiente intervalo.

En la tercera, cada Y'_j es excluido como límite superior de cada intervalo y se incluye como límite inferior del siguiente.

En general, se puede adoptar el sistema en el cual el guión de cada intervalo se interpreta como signo menos, a excepción del último valor del último intervalo.

Esta variable es continua porque hay densidad en los datos a lo largo del recorrido y por su naturaleza.

3. CONCEPTOS Y GRAFICOS DE LAS FRECUENCIAS

3.1 Frecuencias Absolutas. Se denomina frecuencia al número de veces que se presenta el fenómeno dentro de los límites determinados por cada grupo, y suelen denominarse frecuencias absolutas por estar medidas en las unidades de la variable.

Las frecuencias se determinan en la columna correspondiente a los valores de la variable, observando en el experimento el número de veces que un determinado suceso ocurre con

determinado valor de la variable, es decir, haciendo un recuento o tabulación en los distintos intervalos.

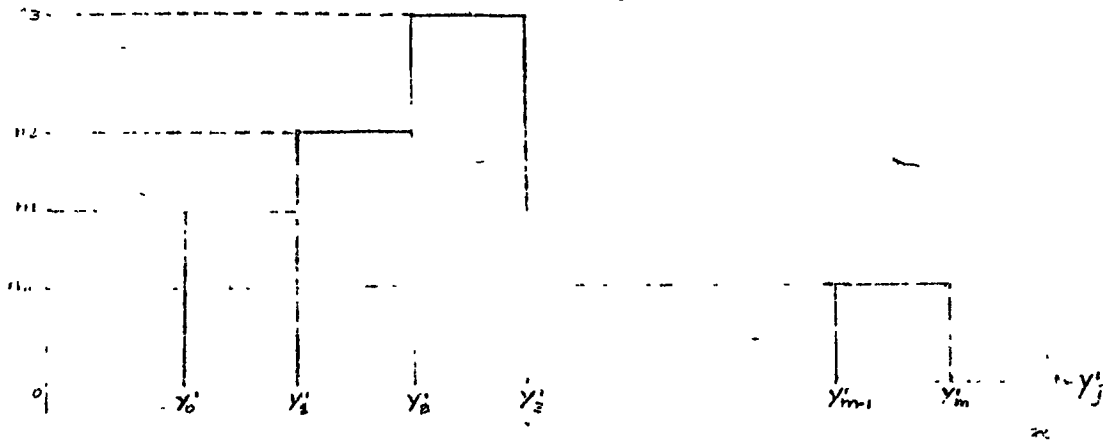
La simbología utilizada para esta clase de frecuencias será:

$$n_j (j = 1, 2, 3, \dots, m)$$

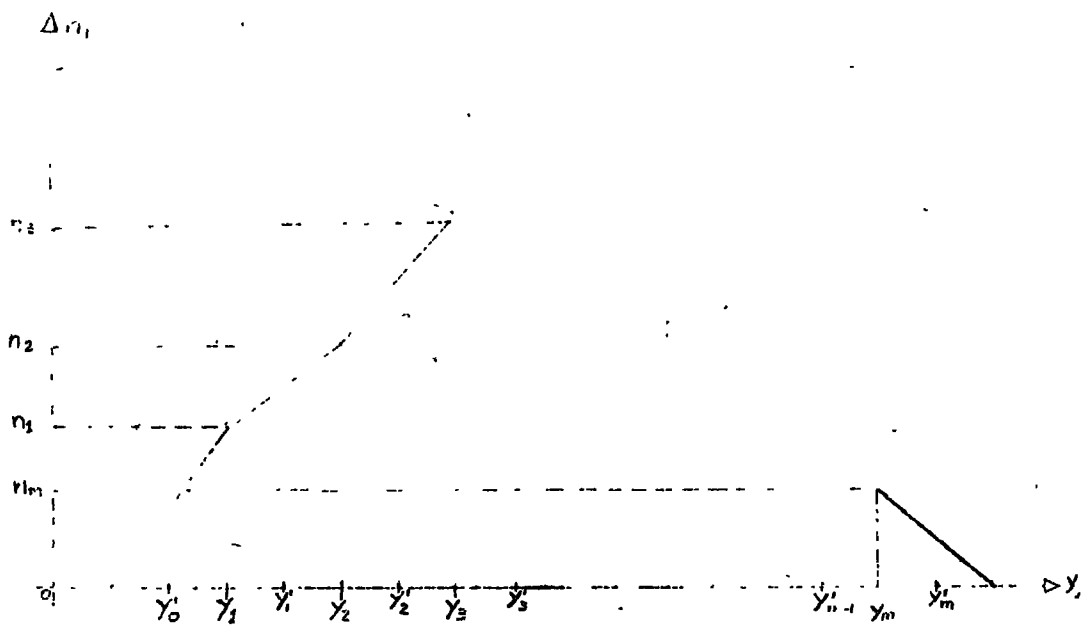
con lo cual la distribución en términos generales quedaría:

j	$Y'_{j-1} - Y'_j$ $C_j = K$	Y_j	n_j
1	$Y'_0 - Y'_1$	Y_1	n_1
2	$Y'_1 - Y'_2$	Y_2	n_2
3	$Y'_2 - Y'_3$	Y_3	n_3
4	$Y'_3 - Y'_4$	Y_4	n_4
.	.	.	.
.	.	.	:
.	.	.	.
m	$Y'_{m-1} - Y'_m$	Y_m	n_m

La representación en gráfica de barras de las frecuencias absolutas se llama HISTOGRAMA DE FRECUENCIAS:



Cuando se tienen en cuenta los puntos medios o marcas de clase se se obtiene un POLIGONO DE FRECUENCIAS:



Ejemplo 3: Precios de las Acciones de 50 Empresas en la Bolsa de Bogotá en Enero 19 X

$X_1 = 5,33$	$X_{11} = 7,24$	$X_{21} = 26,09$	$X_{31} = 25,93$	$X_{41} = 22,00$
$X_2 = 2,91$	$X_{12} = 16,77$	$X_{22} = 14,10$	$X_{32} = 7,00$	$X_{42} = 5,00$
$X_3 = 5,91$	$X_{13} = 13,08$	$X_{23} = 15,68$	$X_{33} = 6,15$	$X_{43} = 19,58$
$X_4 = 2,90$	$X_{14} = 12,81$	$X_{24} = 8,95$	$X_{34} = 3,00$	$X_{44} = 2,80$
$X_5 = 11,43$	$X_{15} = 12,40$	$X_{25} = 12,40$	$X_{35} = 15,86$	$X_{45} = 2,40$
$X_6 = 13,49$	$X_{16} = 16,98$	$X_{26} = 12,15$	$X_{36} = 13,55$	$X_{46} = 11,72$
$X_7 = 23,09$	$X_{17} = 9,47$	$X_{27} = 28,00$	$X_{37} = 3,64$	$X_{47} = 4,53$
$X_8 = 6,19$	$X_{18} = 8,98$	$X_{28} = 16,06$	$X_{38} = 6,87$	$X_{48} = 13,20$
$X_9 = 29,84$	$X_{19} = 11,00$	$X_{29} = 31,90$	$X_{39} = 12,74$	$X_{49} = 13,00$
$X_{10} = 26,89$	$X_{20} = 10,60$	$X_{30} = 12,00$	$X_{40} = 10,30$	$X_{50} = 22,00$

RANGO O RECORRIDO (R)

Fórmula General:

$$R = X \text{ máx} - X \text{ mín}$$

$$R = 31,90 - 2,40$$

$$R = \$29,50$$

Número de Clases (m): Tal como se explicó antes el número (m) de clases puede ser asignado por el investigador, teniendo

do en cuenta que, no exista demasiada concentración, pero tampoco, que exista dispersión dentro de los límites que conforman cada clase. Para el ejemplo se asigna a m un número total de 7 clases diferentes:

INTERVALO DE CLASE (C)

Fórmula General:

$$C = \frac{(\text{X máx} + 1) - (\text{X mín} - 1)}{m}$$

$$C = \frac{(31.90 + 1) - (2.40 - 1)}{7}$$

$$C = \$ 4.50$$

CONSTRUCCION DE CLASES

Fórmula General:

$$Y'_j = Y'_{j-1} + C_j$$

$$Y'_1 = Y'_0 + C \quad 1.40 + 4.50 = \$ 5.90$$

$$Y'_2 = Y'_1 + C \quad 5.90 + 4.50 = \$ 10.40$$

$$Y'_3 = Y'_2 + C \quad 10.40 + 4.50 = \$ 14.90$$

$$Y'_4 = Y'_3 + C \quad 14.90 + 4.50 = \$ 19.40$$

$$Y'_5 = Y'_4 + C \quad 19.40 + 4.50 = \$ 23.90$$

$$Y'_6 = Y'_5 + C \quad 23.90 + 4.50 = \$ 28.40$$

$$Y'_7 = Y'_6 + C \quad 28.40 + 4.50 = \$ 32.90$$

MARCA DE CLASE (Y_j)

Fórmula General:

$$Y_j = \frac{Y'_{j-1} + Y'_j}{2}$$

$$Y_1 = \frac{Y'_0 + Y'_1}{2} = \frac{1.40 + 5.90}{2} = \$ 3.65$$

$$Y_2 = \frac{Y'_1 + Y'_2}{2} = \frac{5.90 + 10.40}{2} = \$ 8.15$$

$$Y_3 = \frac{Y'_2 + Y'_3}{2} = \frac{10.40 + 14.90}{2} = \$ 12.65$$

$$Y_4 = \frac{Y'_3 + Y'_4}{2} = \frac{14.90 + 19.40}{2} = \$ 17.15$$

$$Y_5 = \frac{Y'_4 + Y'_5}{2} = \frac{19.40 + 23.90}{2} = \$ 21.65$$

$$Y_6 = \frac{Y'_5 + Y'_6}{2} = \frac{23.90 + 28.40}{2} = \$ 26.15$$

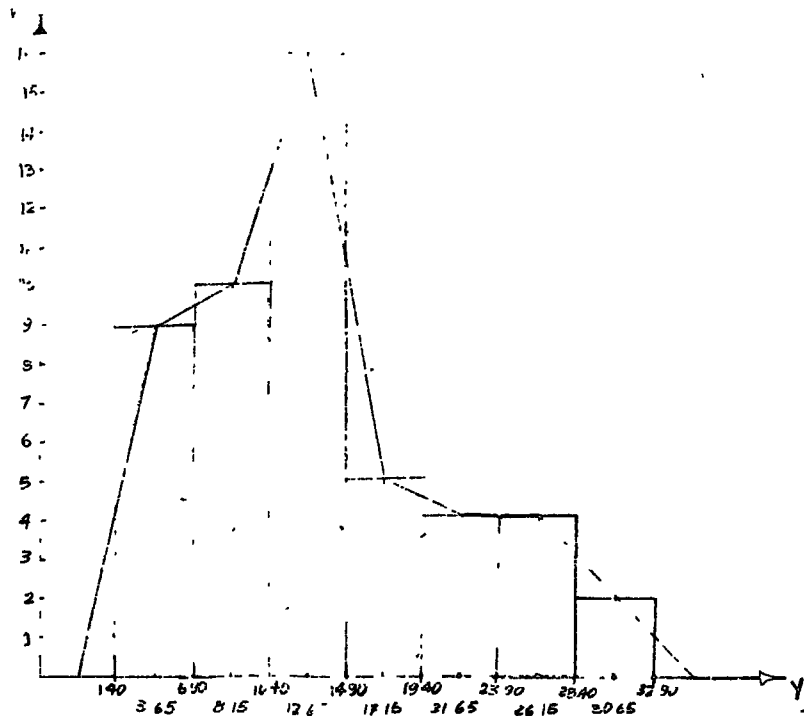
$$Y_7 = \frac{Y'_6 + Y'_7}{2} = \frac{28.40 + 32.90}{2} = \$ 30.65$$

hasta aquí se han considerado los pasos que se deben seguir para la construcción de la tabla de frecuencias. Para evitar tener que ordenar la serie de mayor a menor o viceversa, los datos originales se someten a una tabulación dentro de los límites de cada clase para obtener de esta manera, las frecuencias absolutas de la siguiente forma:

j	$Y'_{j-1} - Y'_j$	TABULACION	Y_j	n_j
1	1,40 - 5,90	≡ IIII	3,65	9
2	5,90 - 10,40	≡ ≡	8,15	10
3	10,40 - 14,90	≡ ≡ ≡ I	12,65	16
4	14,90 - 19,40	≡	17,15	5
5	19,40 - 23,90	IIII	21,65	4
6	23,90 - 28,40	IIII	26,15	4
7	28,40 - 32,90	II	30,65	2
				<u>50</u>

GRAFICO No. 1

HISTOGRAMA Y POLIGONO DE FRECUENCIAS



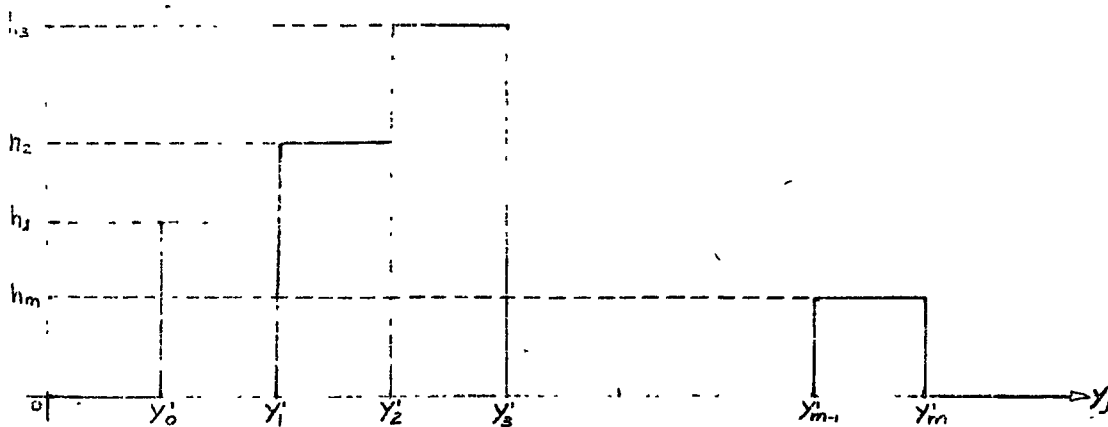
3.2 Frecuencias Relativas (h_j). Se entiende por frecuencias relativas, a los cocientes obtenidos entre el valor de una determinada frecuencia absoluta y el total de las observaciones $1]$. Todas las frecuencias relativas serán fracciones menores o iguales a 1.

1] La probabilidad clásica de un evento, se define como el número de casos favorables del evento sobre el número de casos posibles; de esta manera la frecuencia relativa h_j ; se puede entender también como la probabilidad empírica de que el evento esté en la clase correspondiente.

les a la unidad. Estos valores relativos de la frecuencia sirven de base para análisis de proporción de datos alrededor de ciertos y determinados valores; su propiedad fundamental es:

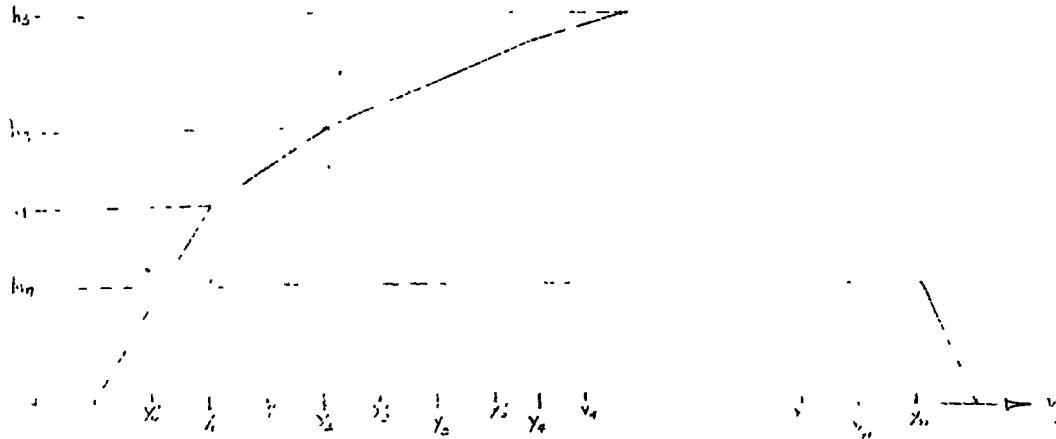
La suma de todas las frecuencias relativas, es igual a la unidad.

La representación gráfica de las frecuencias relativas se hace mediante el HISTOGRAMA DE FRECUENCIAS RELATIVAS:



o mediante el POLIGONO DE FRECUENCIAS RELATIVAS, cuando se tienen en cuenta los puntos medios o marcas de clase :

Δ'



Para una misma distribución los gráficos de frecuencias ab
solutas y relativas son semejantes; difieren en la escala ver
tical.

3 3 Frecuencias Absolutas Acumuladas (N_j). Si se tiene una
distribución o una tabla de frecuencias y se quiere saber el
total de datos u observaciones desde el origen hasta un valor
de la variable dado, hay que acumular o agregar las frecuenci
as correspondientes a los intervalos anteriores obteniendo
así las frecuencias absolutas acumuladas. Estas frecuencias
acumuladas se designan por N_j y tienen como característic
as fundamentales las siguientes: Una frecuencia acumulada de

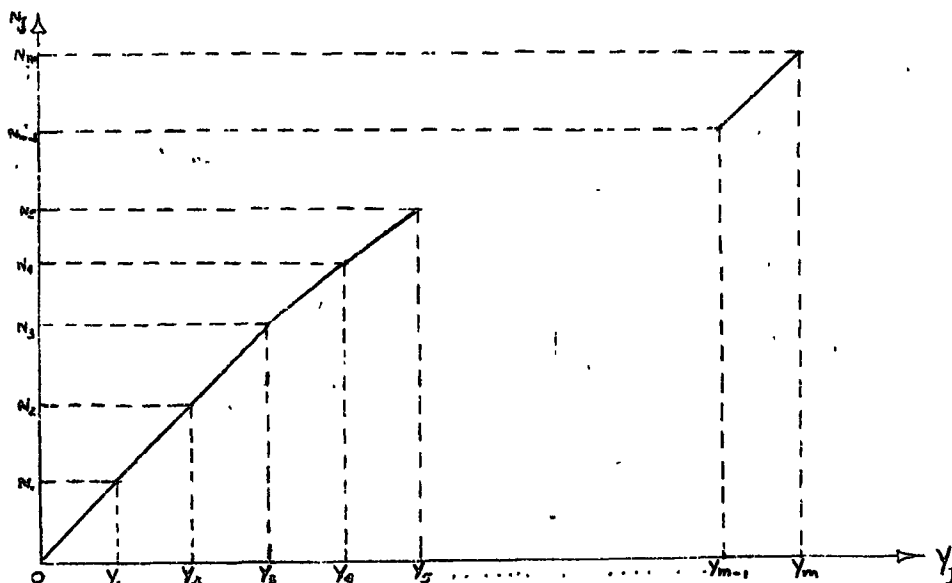
un punto determinado tendrá que ser mayor o igual que la frecuencia correspondiente al intervalo anterior, pero nunca inferior; la frecuencia absoluta acumulada última tendrá que ser igual al total de las observaciones; la suma de la columna de las frecuencias absolutas acumuladas, no tienen ninguna razón lógica, ya que éstas representan una acumulación de ellas. Cualquier frecuencia acumulada puede expresarse así:

$$N_j = \sum_{i=1}^j n_i$$

Resumiendo, las propiedades de N_j son las siguientes:

$$n_1 = N_1 \leq N_2 \leq N_3 \leq \dots \leq N_m = n$$

su representación gráfica se denomina POLIGONO DE FRECUENCIAS ACUMULADAS o FUNCION DE DENSIDAD DE FRECUENCIAS:



Para el polígono de frecuencias absolutas acumuladas, se lleva al eje horizontal los intervalos de clase, y al eje vertical los valores de las frecuencias absolutas acumuladas.

Se levanta el límite superior de cada intervalo, una barra igual al valor de la frecuencia absoluta acumulada, finalmente se unen los extremos superiores y la línea que los una constituirá el polígono de frecuencias.

3 4 Frecuencias Relativas Acumuladas (H_j). Se obtienen en forma análoga a las frecuencias absolutas acumuladas, partiendo de las frecuencias relativas; sirven para determinar la proporción (el porcentaje) acumulado de un valor X cualquiera, dentro de la distribución, que sea mayor o menor que él, tendremos que adicionar o sustraer sucesivamente las frecuencias relativas hasta ese valor. En términos generales vienen expresados por la relación entre las frecuencias acumuladas absolutas y el total de elementos de la muestra.

$$H_j = \frac{N_j}{n} \quad j = 1, 2, 3, \dots, m$$

$$H_j = \sum_{i=1}^j h_i$$

Cumple las siguientes propiedades:

a) $H_j > H_{j-1}$

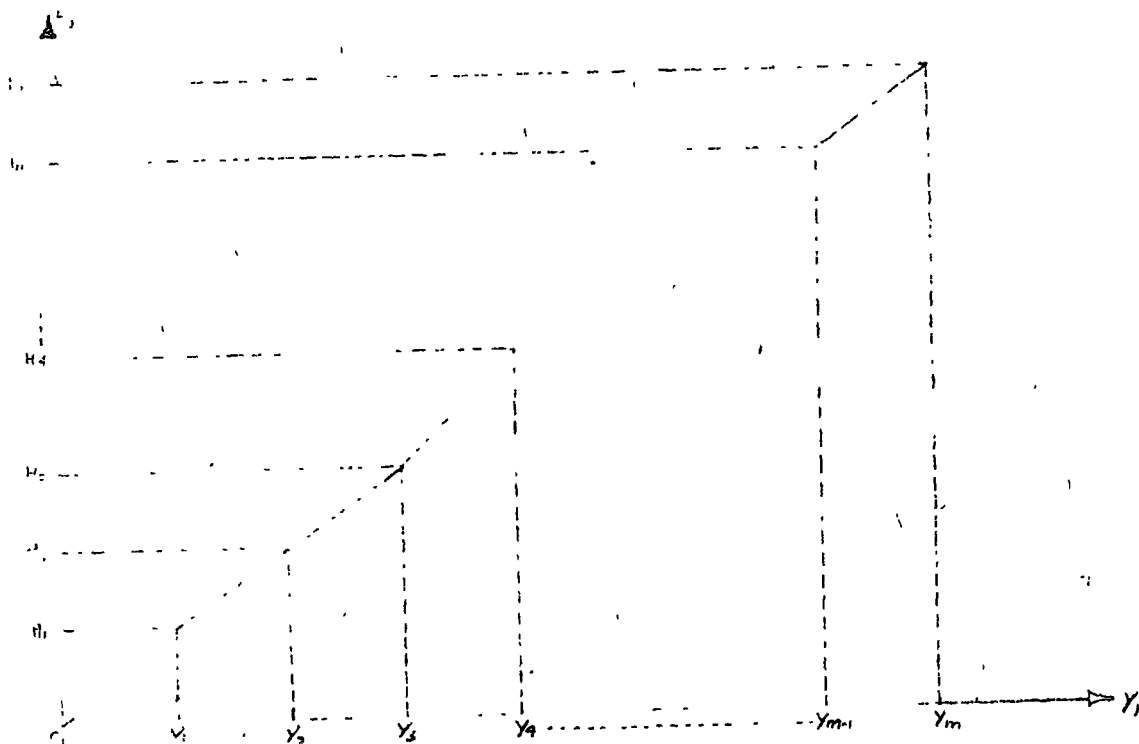
b) $H_m = 1$

c) $H_1 = h_1$

como conclusión:

d) $h_1 = H_1 < H_2 < H_3 < \dots < H_m = 1$

La representación gráfica se llama POLIGONO DE FRECUENCIAS RELATIVAS ACUMULADAS:



Esta representación es similar a la gráfica de frecuencias absolutas acumuladas, pero difiere en la escala vertical que debe variar entre los límites 0 y 1 (proporción) ó 0 y 100 (porcentaje).

Ejemplo 4: En base al ejemplo 3; se pueden aplicar los conocimientos anteriores, así:

FRECUENCIAS RELATIVAS (n_j)

Fórmula General:

$$h_j = n_j/n$$

$$h_1 = 9/50 = 0.18 = 18\%$$

$$h_2 = 10/50 = 0.20 = 20\%$$

$$h_3 = 16/50 = 0.32 = 32\%$$

$$h_4 = 5/50 = 0.10 = 10\%$$

$$h_5 = 4/50 = 0.08 = 8\%$$

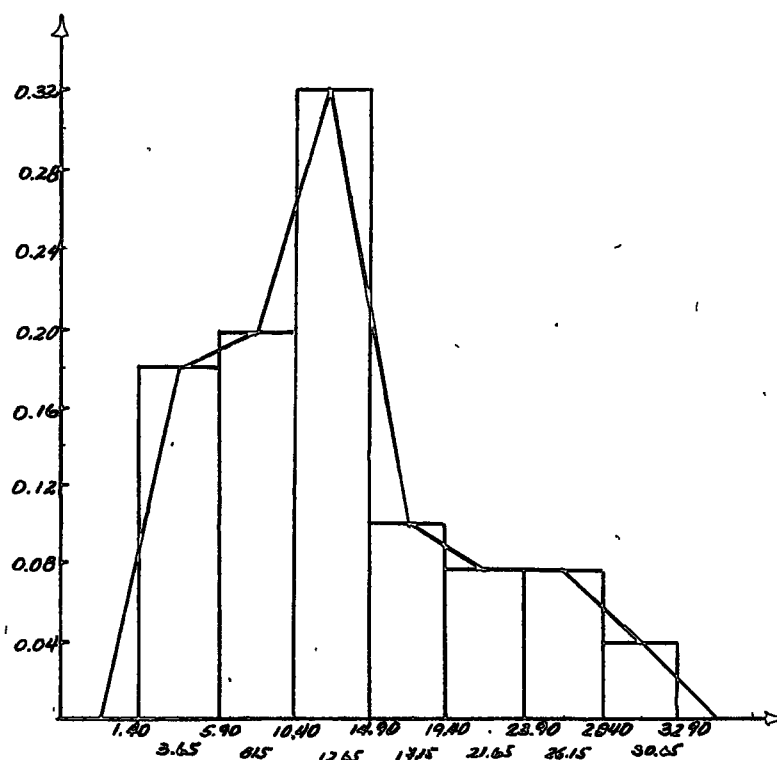
$$h_6 = 4/50 = 0.08 = 8\%$$

$$h_7 = 2/50 = 0.04 = 4\%$$

$$\sum_{j=1}^m H_j = 1.00 \quad 100\%$$

GRAFICO N° 2

HISTOGRAMA Y POLIGONO DE FRECUENCIAS RELATIVAS



FRECUENCIAS ACUMULADAS (N_j)

Fórmula General :

$$N = \sum_{j=1}^j n_j$$

$$N_1 = n_1 = 9 = 9$$

$$N_2 = N_1 + n_2 = 9 + 10 = 19$$

$$N_3 = N_2 + n_3 = 19 + 16 = 35$$

$$N_4 = N_3 + n_4 = 35 + 5 = 40$$

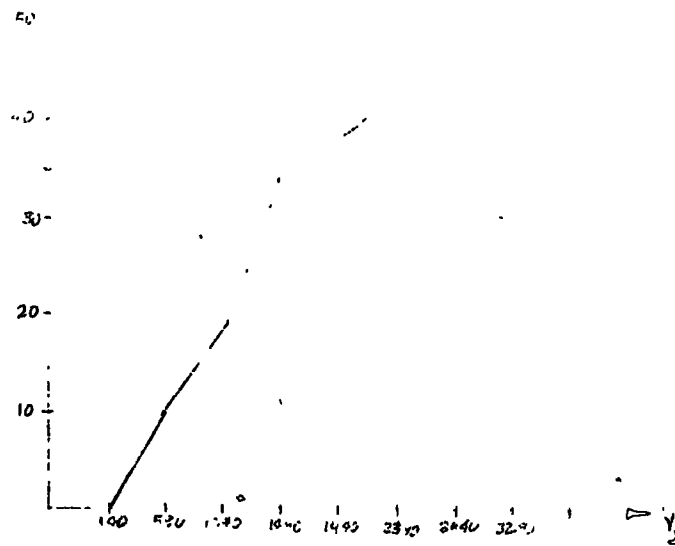
$$N_5 = N_4 + n_5 = 40 + 4 = 44$$

$$N_6 = N_5 + n_6 = 44 + 4 = 48$$

$$N_7 = N_6 + n_7 = 48 + 2 = 50$$

GRAFICO No. 3

N_j FRECUENCIAS ABSOLUTAS ACUMULADAS



FRECUENCIAS RELATIVAS ACUMULADAS

Fórmula General:

$$H_j = \sum_{i=1}^j h_i$$

$$H_1 = h_1 = 0.18 = 0.18\%$$

$$H_2 = H_1 + h_2 = 0.18 + 0.20 = 0.38\%$$

$$H_3 = H_2 + h_3 = 0.38 + 0.32 = 0.70\%$$

$$H_4 = H_3 + h_4 = 0.70 + 0.10 = 0.80\%$$

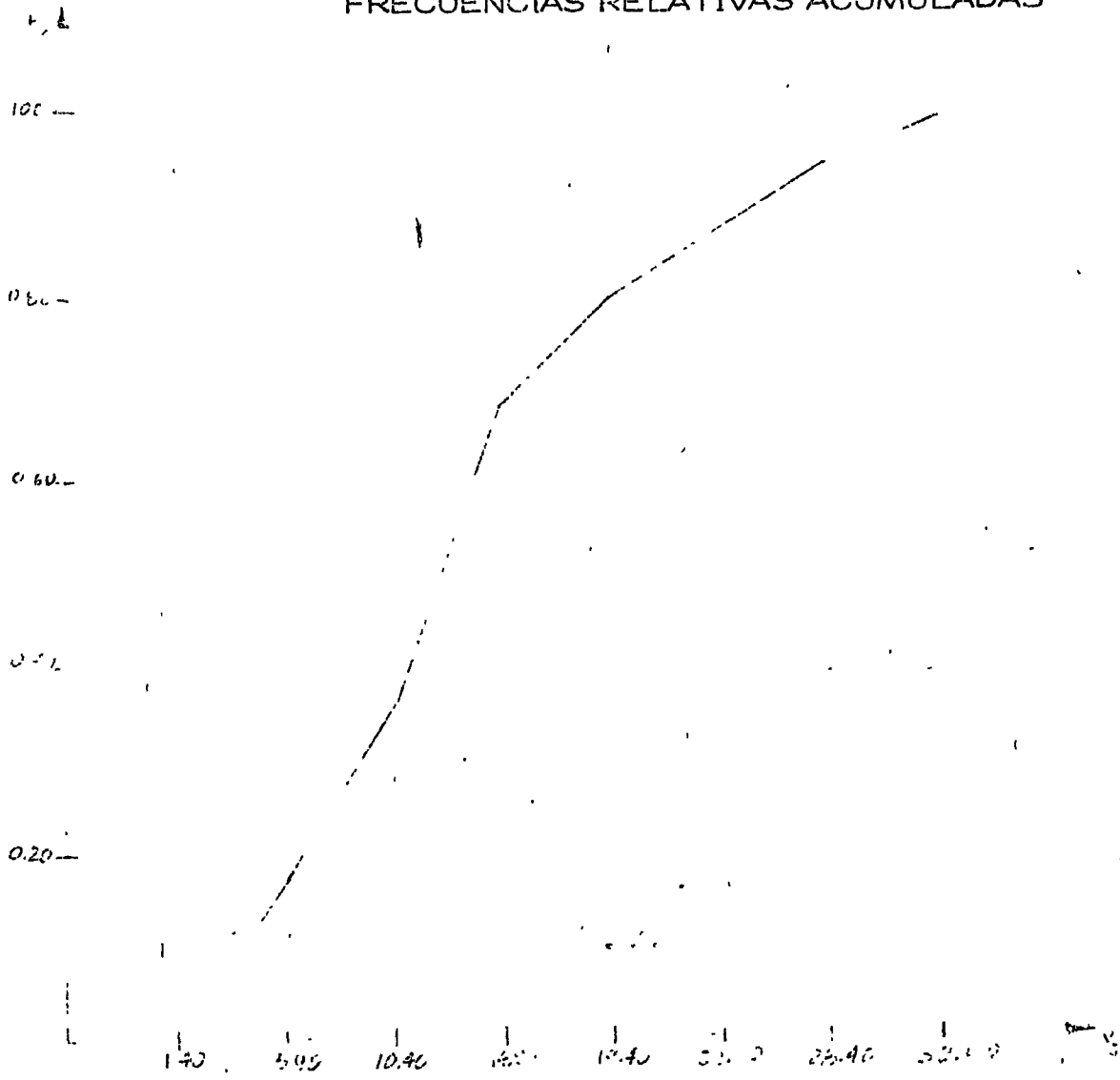
$$H_5 = H_4 + h_5 = 0.80 + 0.08 = 0.88\%$$

$$H_6 = H_5 + h_6 = 0.88 + 0.08 = 0.96\%$$

$$H_7 = H_6 + h_7 = 0.96 + 0.04 = 1$$

GRAFICO No. 4

FRECUENCIAS RELATIVAS ACUMULADAS



4. DISTRIBUCIONES DE FRECUENCIA PARA VARIABLES DISCRETAS.

Para la conformación de la distribución de frecuencias, cuando el tipo de variable objeto de estudio es discreta, se debe tener en cuenta:

- a) Si el rango (R) de la variable, a juicio del investigador, es pequeño y se ajusta a sus objetivos se toma como valores diferentes de la variable (clases o grupos) cada uno de ellos. En la práctica estos valores diferentes de la variable se deben tomar como marcas de clase (Y_j).
- b) Al contrario, si (R) es grande, se deben determinar las clases en forma similar al tratamiento hecho para la variable continua, teniendo en cuenta, que los valores superior e inferior de las clases deben ser valores enteros.

Una vez terminadas las marcas de clase para este tipo de variables, el tratamiento que se aplica a las series tanto discretas como continuas es el mismo, pero difieren en la interpretación de los parámetros; algunos de los cuales, deben tomar valores estrictamente enteros.

El procedimiento, así como la representación gráfica de este tipo de variables se puede apreciar de una manera más amplia en el Anexo No. 1.

5. MOMENTOS ESTADÍSTICOS UNIDIMENSIONALES

Momentos:

En general se define como momento estadístico a la media (aritmética) de las desviaciones con respecto a un determinado valor, elevadas a una potencia cualquiera. La variable puede ser una serie de datos originales o agrupados y la potencia a que esté elevada indica el orden del momento. En general se consideran momentos para DATOS ORIGINALES y momentos para DATOS AGRUPADOS. Algunos de ellos tienen un significado por si solos, otros no lo tienen.

5.1 Momentos para DATOS ORIGINALES: Se estudian aquí 2 tipos de momentos, así:

Momentos con respecto a un origen cero (a_r).

Momentos con respecto a la media (aritmética) (m_r).

5.1.1 Los momentos a_r para una serie de datos origina

les pueden estudiarse teniendo en cuenta que X_i sean los diferentes datos; n el tamaño de la muestra, entonces:

$$a_r = M \left[(X - 0)^r \right] = \frac{\sum_{i=1}^n (X_i - 0)^r}{n}$$

Obteniéndose de esta forma los diferentes momentos de acuerdo al orden r

$$(r = 1, 2, 3, \dots)$$

Si $r = 1$

$$a_1 = M(X) = \frac{\sum_{i=1}^n X_i}{n} = \text{Media aritmética de la variable.}$$

$r = 2$

$$a_2 = M(X^2) = \frac{\sum_{i=1}^n X_i^2}{n}$$

$r = 3$

$$a_3 = M(X^3) = \frac{\sum_{i=1}^n X_i^3}{n}$$

$$r = n$$

$$a_n = M(X^n) = \frac{\sum_{i=1}^n X_i^n}{n}$$

5.1.2 Dados n (tamaño de la muestra) X_i los diferentes valores y $\bar{X}$ la media de tales datos podremos definir los momentos con respecto a la media m_r ($r=1,2,3, \dots$), así:

$$m_r = M[(X - \bar{X})^r] = \frac{\sum_{i=1}^n (X_i - \bar{X})^r}{n}$$

Sea $Z_i = (X_i - \bar{X})$ desviaciones con respecto a la media aritmética, entonces:

$$m_r = M(Z^r) = \frac{\sum_{i=1}^n Z_i^r}{n}$$

Se obtienen momentos de diferente orden:

$$r=1 \quad m_1 = M(Z) = \frac{\sum_{i=1}^n Z_i}{n} = 0$$

(primera propiedad de la media).

$$r=2 \quad m_2 = M(Z^2) = \frac{\sum_{i=1}^n Z_i^2}{n} = S^2$$

(varianza de los datos).

$$r = 3 \quad m_3 = M(Z^3) = \frac{\sum_{i=1}^n Z_i^3}{n}$$

$$r = n \quad m_n = M(Z^n) = \frac{\sum_{i=1}^n Z_i^n}{n}$$

5.2 Momentos para DATOS AGRUPADOS: Existen 4 clases de momentos para la tabla de frecuencias:

5.2.1 Con respecto a un origen cero (a_r).

5.2.2 Con respecto a la media aritmética (m_r).

5.2.3 Con respecto al origen de trabajo (m'_r).

5.2.4 Con respecto a un origen de trabajo medidos en unidades de intervalo (m''_r).

5.2.1 Momentos con respecto a un origen cero (a_r). Para datos agrupados se tienen las variables Y_j (marca de clase) y, n_j (frecuencia absoluta), entonces los momentos con respecto a un origen (cero) a_r se definen como:

$$a_r = M \left[(Y - 0)^r \right] = \frac{\sum_{j=1}^m (Y_j - 0)^r n_j}{n}$$

$$\text{Si } r = 1 \quad a_1 = M(Y) = \frac{\sum_{j=1}^m Y_j n_j}{n}$$

(media aritmética).

$$r = 2 \quad a_2 = M(Y^2) = \frac{\sum_{j=1}^m Y_j^2 n_j}{n}$$

$$r = 3 \quad a_3 = M(Y^3) = \frac{\sum_{j=1}^m Y_j^3 n_j}{n}$$

$$r = n \quad a_n = M(Y^n) = \frac{\sum_{j=1}^m Y_j^n n_j}{n}$$

Ejemplo 5:

CALCULO DE LOS MOMENTOS PARA LOS DATOS AGRUPADOS
DEL EJERCICIO

CLASES	Y_j	n_j	$Y_j n_j$	$Y_j^2 n_j$	$Y_j^3 n_j$	$Y_j^4 n_j$
1,40 - 5,90	3,65	9	32,85	119,9025	437,6441	1597,401
5,90 - 10,40	8,15	10	81,50	664,225	5413,434	44119,485
10,40 - 14,90	12,65	16	202,40	2560,360	32388,554	409712,640
14,90 - 19,40	17,15	5	85,75	1470,6125	25221,004	432540,218
19,40 - 23,90	21,65	4	86,60	1874,890	40591,3685	878803,117
23,90 - 28,40	26,15	4	104,60	2735,290	71527,833	1 870452,832
28,40 - 32,90	30,65	2	61,30	1878,845	57586,5992	1 765029,259
		50	655,00	11304,1250	233166,4368	5 402254,952

Según los resultados anteriores:

$$a_1 = \frac{\sum_{j=1}^m Y_j n_j}{n} = \frac{655,00}{50} = \$ 13,1$$

$$a_2 = \frac{\sum_{j=1}^m Y_j^2 n_j}{n} = \frac{11304,125}{50} = 226,082$$

$$a_3 = \frac{\sum_{j=1}^m Y_j^3 n_j}{n} = \frac{233166,4368}{50} = 4663,328$$

$$a_4 = \frac{\sum_{j=1}^m Y_j^4 n_j}{n} = \frac{5402254,952}{50} = 108045,099$$

5.2.2 Momentos con respecto a la Media Aritmética (m_r).

Se tienen Y_j y n_j como en la definición anterior; además $Z_j = Y_j - a$ (desviaciones con respecto a la media aritmética), entonces los momentos con respecto a la media se definen como:

$$m_r = M[(Y - a)^r] = M(Z^r) = \frac{\sum_{j=1}^m Z_j^r n_j}{n}$$

$$\text{Si } r=1 \quad m_1 = M(Z) = \frac{\sum_{j=1}^m Z_j n_j}{n} = 0$$

$$r=2 \quad m_2 = M(Z^2) = \frac{\sum_{j=1}^m Z_j^2 n_j}{n} = S^2$$

(varianza de la distribución).

$$r=3 \quad m_3 = M(Z^3) = \frac{\sum_{j=1}^m Z_j^3 n_j}{n}$$

$$r=n \quad m_n = M(Z^n) = \frac{\sum_{j=1}^m Z_j^n n_j}{n}$$

Ejemplo 6:

CALCULO DE LOS MOMENTOS m_r

CLASES	Y_j	n_j	Z_j	$Z_j n_j$	$Z_j^2 n_j$	$Z_j^3 n_j$	$Z_j^4 n_j$
1,40 - 5,90	3,65	9	-9,45	-85,05	803,7225	-7595,1776	71774,4283
5,90 - 10,40	8,15	10	-4,95	-49,50	245,0250	-1212,8737	6003,7248
10,40 - 14,90	12,65	16	-0,45	-7,20	3,2400	-1,4580	0,6561
14,90 - 19,40	17,15	5	4,05	20,25	82,0125	332,1506	1345,2099
19,40 - 23,90	21,65	4	8,55	34,20	292,4100	2500,1056	21375,9020
23,90 - 28,40	26,15	4	13,05	52,20	681,2100	8889,7905	116011,7660
28,40 - 32,90	30,65	2	17,55	35,10	616,0050	10810,8877	189731,0844
		50		- 0 -	2723,625	13723,425	406242,7715

Sea $Z_j = Y_j - a_1$

$$m_1 = m(Z) = \frac{\sum_{j=1}^m Z_j n_j}{n} = 0$$

$$m_2 = m(Z^2) = \frac{\sum_{j=1}^m Z_j^2 n_j}{n} = \frac{2723,625}{50} = 54,4725$$

$$m_3 = m(Z^3) = \frac{\sum_{j=1}^m Z_j^3 n_j}{n} = \frac{13723,425}{50} = 274,4685$$

$$m_4 = m(Z^4) = \frac{\sum_{j=1}^m Z_j^4 n_j}{n} = \frac{406242,7715}{50} = 8124,8554$$

5.2.3 Momento con respecto a un origen de trabajo (m'_r).

Sean Y_j y n_j definidos como antes; ot, un origen de trabajo cualquiera y $Z'_j = Y_j - ot$ las desviaciones con respecto a ot. Los momentos se definen así:

$$m'_r = M[(Y_j - ot)^r] = M(Z')^r = \frac{\sum_{j=1}^m (Z'_j)^r n_j}{n}$$

$$\text{Si } r=1 \quad m'_1 = M[Z'] = \frac{\sum_{j=1}^m Z'_j n_j}{n}$$

$$r=2 \quad m'_2 = M \left[(Z')^2 \right] = \frac{\sum_{j=1}^m (Z'_j)^2 n_j}{n}$$

$$r=n \quad m'_n = M \left[(Z')^n \right] = \frac{\sum_{j=1}^m (Z'_j)^n n_j}{n}$$

Ejemplo 7:

CALCULO DE LOS MOMENTOS m'_r

CLASES	Y_j	n_j	Z'_j	$(Z'_j)n_j$	$(Z'_j)^2 n_j$	$(Z'_j)^3 n_j$	$(Z'_j)^4 n_j$
1,40 - 5,90	3,65	9	- 4,9	- 81	729,0	-6561,00	59049,00
5,90 - 10,40	8,15	10	- 4,5	- 45	202,5	- 911,25	4100,625
10,40 - 14,90	<u>12,65</u>	16	-	-	-	-	-
14,90 - 19,40	17,15	5	4,5	22,5	101,25	455,625	2050,3125
19,40 - 23,90	21,65	4	9,0	36,0	324,00	2916,00	26244,00
23,90 - 28,40	26,15	4	13,5	54,0	729,00	9841,00	132860,25
28,40 - 32,90	30,65	2	18,0	36,0	648,00	11664,00	209952,00
		50		22,5	2733,75	17404,875	434256,1875

Sea $ot = 12,65$, , $Z'_j = Y_j - ot$ entonces:

$$m'_1 = M(Z') = \frac{\sum_{j=1}^m Z'_j n_j}{n} = \frac{22,5}{50} = 0,45$$

$$m'_2 = M \left[(Z')^2 \right] = \frac{\sum_{j=1}^m (Z'_j)^2 n_j}{n} = \frac{2733,75}{50} = 54,675$$

$$m'_3 = M \left[(Z')^3 \right] = \frac{\sum_{j=1}^m (Z'_j)^3 n_j}{n} = \frac{17404,875}{50} = 348,0975$$

$$m'_4 = M \left[(Z')^4 \right] = \frac{\sum_{j=1}^m (Z'_j)^4 n_j}{n} = \frac{434256,1875}{50} = 8685,1237$$

5.2.4 Momentos con respecto a un Origen de Trabajo, medidos en Unidades de Intervalo (m''_r).

Considerando Y_j y n_j como antes, sea:

$$Z''_j = \frac{Y_j - ot}{C_j}$$

desviaciones con respecto a ot medido en unidades de intervalos, luego los momentos m''_r se definen:

$$m''_r = M \left[\left(\frac{Y - ot}{C} \right)^r \right] = M \left[(Z'')^r \right] = \frac{\sum_{j=1}^m Z''_j{}^r n_j}{n}$$

$$r = 1 \quad m''_1 = M \left[Z'' \right] = \frac{\sum_{j=1}^m Z''_j n_j}{n}$$

$$r=2 \quad m''_2 = M \left[(Z'')^2 \right] = \frac{\sum_{j=1}^m (Z''_j)^2 n_j}{n}$$

$$r=3 \quad m''_3 = M \left[(Z'')^3 \right] = \frac{\sum_{j=1}^m (Z''_j)^3 n_j}{n}$$

$$r=n \quad m''_n = M \left[(Z'')^n \right] = \frac{\sum_{j=1}^m (Z''_j)^n n_j}{n}$$

Ejemplo 8:

CALCULO DE MOMENTOS m''_r

CLASES	Y_j	n_j	Z'_j	$Z'_j n_j$	$(Z'_j)^2 n_j$	$(Z'_j)^3 n_j$	$(Z'_j)^4 n_j$
1,40 - 5,90	3,65	9	-3	-27	81	-243	729
5,90 - 10,40	8,15	10	-2	-20	40	-80	160
10,40 - 14,90	12,65	16	-1	-16	16	-16	16
14,90 - 19,40	17,15	5	-	-	-	-	-
19,40 - 23,90	21,65	4	1	4	4	4	4
23,90 - 28,40	26,15	4	2	8	16	32	64
28,40 - 32,90	30,65	2	3	6	18	54	162
				-45	175	-249	1135

Sea:

$$Z''_j = \frac{Y_j - ot}{C_j}, \quad \wedge, \quad ot = 17,15$$

$$m''_1 = M(Z'') = \frac{\sum_{j=1}^m Z''_j n_j}{n} = \frac{-45}{50} = -0.9$$

$$m''_2 = M[(Z'')^2] = \frac{\sum_{j=1}^m (Z''_j)^2 n_j}{n} = \frac{175}{50} = 3.5$$

$$m''_3 = M[(Z'')^3] = \frac{\sum_{j=1}^m (Z''_j)^3 n_j}{n} = \frac{249}{50} = 4.98$$

$$m''_4 = M[(Z'')^4] = \frac{\sum_{j=1}^m (Z''_j)^4 n_j}{n} = \frac{1135}{50} = 22.7$$

5.3 Relaciones entre Momentos: La relación entre los distintos tipos de momentos es una forma de abreviar el cálculo de - aquellos momentos que tienen un significado propio (m_r) y además para estimar parámetros. En forma general estas relaciones obedecen al desarrollo de un binomio de Newton, cuya expresión es la siguiente:

$$m_r = \sum_{k=0}^r \binom{r}{k} (-1)^k a_{(r-k)} a_1^k$$

para m_r en función de a_r ; la expresión es la misma para - los m_r reemplazando a_r por m'_r .

La expresión para m_r en función de m''_r es la siguiente:

$$m_r = \sum_{k=0}^r \binom{r}{k} (-1)^k (C_j)^k m''_{r-k} (m''_1)^k$$

Se expresan y comprueban tales relaciones hasta $r = 4$ que son las más utilizadas:

a) m_r en función de a_r

$$m_1 = a_1 - a_1 = 0$$

$$m_2 = a_2 - a_1^2 = S^2$$

$$m_3 = a_3 - 3 a_2 a_1 + 2 a_1^3$$

$$m_4 = a_4 - 4 a_3 a_1 + 6 a_2 a_1^2 - 3 a_1^4$$

Comprobacion:

$$m_2 = S^2 = a_2 - (a_1)^2$$

$$m_2 = 226,082 - (13,1)^2$$

$$m_2 = 226,082 - 171,61$$

$$m_2 = 54,472$$

$$m_3 = a_3 - 3 a_2 a_1 + 2 a_1^3$$

$$m_3 = 4663,328 - 3 (226,082) (13,1) + 2 (13,1)^3$$

$$m_3 = 4663,328 - 8885,0226 + 4496,182$$

$$m_3 = 274,4874$$

$$m_4 = a_4 - 4 a_3 a_1 + 6 a_2 a_1^2 - 3 a_1^4$$

$$m_4 = 108045,099 - 4 (4663,328) (13,1) + 6 (226,082) (171,61) - 3 (13,1)^4$$

$$m_4 = 108045,099 - 244358,3872 + 232787,592 - 88349,9763$$

$$m_4 = 8124,3275$$

Estos valores pueden observarse en la correspondiente tabla. Los errores, aunque mínimos, son errores de REDONDEO.

b) m_r en función de m'_r

$$m_2 = m'_2 - (m'_1)^2 = S^2$$

$$m_3 = m'_3 - 3 m'_2 m'_1 + 2 m'_1^3$$

$$m_4 = m'_4 - 4 m'_3 m'_1 + 6 m'_2 m'_1^2 - 3 m'_1^4$$

Comprobación:

$$m_2 = 54,675 - (0,45)^2$$

$$m_2 = 54,675 - 0,202$$

$$m_2 = 54,473$$

$$m_3 = 348,0975 - 3 (54,675) (0,45) + 2 (0,45)^3$$

$$m_3 = 274,4685$$

$$m_4 = 8685,1237 - 4 (348,0975) (0,45) + 6 (54,675) (0,202) - 3 (0,45)^4$$

$$= 8685,1237 - 626,5752 + 66,2658 - 0,1227$$

$$m_4 = 8124,6916$$

Las diferencias se deben a errores de redondeo.

c) m_r en función de m''_r

$$m_2 = C^2 [m''_2 - (m''_1)^2]$$

$$m_3 = C^3 [m''_3 - 3 m''_2 m''_1 + 2 (m''_1)^3]$$

$$m_4 = C^4 [m''_4 - 4 m''_3 m''_1 + 6 m''_2 m''_1^2 - 3 (m''_1)^4]$$

Comprobación:

$$m_2 = 20,25 [3,5 - (- 0,9)^2]$$

$$m_2 = 20,25 [3,5 - 0,81]$$

$$m_2 = 54,4725$$

$$m_3 = 91,125 [- 4,98 - 3 (3,5) (- 0,9) + 2 (- 0,9)^3]$$

$$m_3 = 91,125 [- 4,98 + 9,45 - 1,458]$$

$$m_3 = 91,125 (3,0120)$$

$$m_3 = 274,4685$$

$$m_4 = 410,0625 \left[22,7 - 4(-4,98)(-0,9) + 6(3,5)(-0,9)^2 - 3(-0,9)^4 \right]$$

$$m_4 = 410,0625 \left[22,7 - 17,928 + 17,010 - 1,9683 \right]$$

$$m_4 = 8124,8454$$

En resumen las relaciones serían:

$$m_r \begin{cases} (a_{(r)} - a_1)^r \\ (m'_{(r)} - m'_1)^r \\ C^r (m''_{(r)} - m''_1)^r \end{cases}$$

6. MEDIDAS DE TENDENCIA CENTRAL - PROMEDIOS.

Un promedio es, en general, un valor típico que se presume representativo de un conjunto de datos. Este valor típico deberá estar localizado alrededor de la mitad del grupo, y no en un lugar extremo de la distribución.

Los promedios por su naturaleza tienen muchos usos. Todas las ciencias modernas, desde la astronomía y la física, hasta la Psicología y la Economía, se basan en datos medios o promedios

No pueden compararse dos o más grupos de hechos análogos sino comparando sus promedios.

A pesar de su utilidad se deben tener muchas precauciones; una sería, analizar la homogeneidad de los datos en base a su dispersión. Tal sucede en el concepto que se pueda formar sobre los ingresos de un país, en base al ingreso per-cápita (ingreso por habitante); por haber ingresos demasiado bajos y demasiado altos, no es representativo y por lo tanto no debe usarse como tal.

En estadística aplicada se utilizan varias clases de promedios clasificados en dos grupos :

6.1 Matemáticos: aquellos que pueden someterse a un tratamiento algebraico; pertenecen a esta clasificación la media aritmética, la media geométrica y la media armónica.

6.2 De posición: se calculan teniendo en cuenta los datos individuales previamente ordenados de acuerdo a su tamaño; entre estos están: la mediana, la moda y las fractilas.

6.1.1 Media Aritmética. Se caracteriza por estar localizada en un punto dentro de los límites del recorrido de

la variable. La media aritmética es aquel valor obtenido mediante la suma total de todos los elementos, dividida por el número de observaciones. Si los elementos individuales, son llamados X_i , la media aritmética de todos estos elementos estará dada por la suma total de todos los elementos X_i , así:

$$\frac{\sum_{i=1}^n X_i}{n} = a = M(x) = \bar{X}$$

La media aritmética es representada por uno de estos símbolos a , $M(x)$, $\bar{X}$.

Ejemplo 9: Se tiene el número de personas ocupadas por 10 establecimientos en una zona industrial de Bogotá en el año 19X. Calcular la media aritmética.

$$X_1 = 52$$

$$X_6 = 43$$

$$X_2 = 45$$

$$X_7 = 44$$

$$X_3 = 47$$

$$X_8 = 49$$

$$X_4 = 31$$

$$X_9 = 53$$

$$X_5 = 35$$

$$X_{10} = 43$$

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} = \frac{442}{10} = 44,2$$

El promedio aritmético es 44 personas ocupadas. No hay que confundir, el valor obtenido de la media aritmética, con el hecho de que dicho valor sea representativo en la distribución analizada. Para determinar si es representativo o no, se estudiarán posteriormente ciertas funciones que van a dar el criterio para calificarla. En general suele decirse por ejemplo, que el promedio de sueldos de una empresa es de determinada cantidad en dinero, e inmediatamente se piensa que es el común en los salarios de la empresa; esto puede ser cierto o no, según la forma de distribución de dichos salarios.

El caso puede generalizarse para cualquier concepto, si se logra construir una tabla de frecuencias, agrupando en cada intervalo un cierto número de ellas; se podrá calcular la media aritmética en forma indirecta multiplicando cada uno de los valores iguales por sus frecuencias; luego, sumar los diferentes valores entre sí y dividir por n.

En la práctica se presentará una diferencia despreciable entre el valor obtenido a base de los datos originales X_i y los datos agrupados Y_j , esta diferencia se debe a que los datos dispersos están tomados con sus verdaderos valores, en tanto que los agrupados tienen como característica para cada intervalo, un determinado Y_j que no es precisamente el número exacto, en cuanto a las observaciones reales; sin embargo, para una serie suficientemente grande, la diferencia obtenida entre un sistema y otro es mínima y por lo tanto puede emplearse el sistema de agrupamientos con el cual se hará un notable ahorro de tiempo sin sacrificar la precisión deseada.

La fórmula para el cálculo de la media con datos agrupados sería:

$$M(y) = \frac{\sum_{j=1}^m Y_j n_j}{n} = \frac{\sum_{j=1}^m Y_j n_j}{\sum_{j=1}^m n_j}$$

Para datos reducidos en la tabla de distribución de frecuencias, cuando la variable es continua, está representada por

la marca de clase o punto medio (Y_j).

$$M(Y) = \sum_{j=1}^m Y_j h_j; \text{ ya que } h_j = \frac{n_j}{n}$$

Expresados ya en términos de frecuencia relativa.

La media aritmética es un estadígrafo que representa aquel valor de la variable, en torno al cual, la frecuencia tiende a ser máxima y por lo tanto dicho valor parece ser el más común en el experimento.

Ejemplo 10 : De acuerdo al ejemplo 3 se tiene:

Fórmula General:

$$a = M(Y) = \bar{Y} = \frac{1}{n} \sum_{j=1}^m Y_j n_j$$

Y'_{j-1}	—	Y'_j	Y_j	n_j	$Y_j n_j$
1.40	—	5.90	3.65	9	32.85
5.90	—	10.40	8.15	10	81.50
10.40	—	14.90	12.65	16	202.40
14.90	—	19.40	17.15	5	85.75
19.40	—	23.90	21.65	4	86.60
23.90	—	28.40	26.15	4	104.60
28.40	—	32.90	30.65	2	61.30
				50	655.00

$$M(Y) = \frac{1}{n} \sum_{j=1}^m Y_j n_j = \frac{655}{50} = \$ 13,1$$

Este resultado indica que el precio promedio de las acciones en la bolsa de Bogotá es de 13,1 pesos.

Ejemplo 11: Siguiendo el mismo ejemplo 3 se calculará la media aritmética en términos de frecuencias relativas:

$$a = M(Y) = \bar{Y} = \sum_{j=1}^m Y_j h_j$$

Y'_{j-1}	—	Y'_j	Y_j	h_j	$Y_j h_j$
1,40	—	5,90	3,65	0,18	0,657
5,90	—	10,40	8,15	0,20	1,630
10,40	—	14,90	12,65	0,32	4,048
14,90	—	19,40	17,15	0,10	1,715
19,40	—	23,90	21,65	0,08	1,732
23,90	—	28,40	26,15	0,08	2,092
28,40	—	32,90	30,65	0,04	1,226
				1,00	13,100

$$M(Y) = a = \bar{Y} = \$ 13,1$$

6.1.1.1 Media Aritmética Ponderada: En el cálculo de los promedios hay que tener en cuenta el peso o ponderación de los datos para que los resultados sean correctos.

Cuando se calculan los promedios de una distribución de frecuencias, la ponderación vendrá dada por cada frecuencia de la Serie,

La media aritmética ponderada se define, como la suma de los productos de los datos, multiplicados por su factor de ponderación y dividida por el total de la suma de los factores de ponderación.

Para el caso de datos originales la fórmula sería:

$$\bar{X} = \frac{X_1 W_1 + X_2 W_2 + \dots + X_m W_m}{W_1 + W_2 + \dots + W_m}$$

o sea:

$$\bar{X} = \frac{\sum_{i=1}^m X_i W_i}{\sum_{i=1}^m W_i}$$

Siendo W_i el factor de ponderación.

Ejemplo 12: Se poseen los datos de los salarios diarios de los trabajadores de una empresa dedicada al ramo de la construcción. Calcular el salario promedio diario.

Clase de trabajo	Y_j	n_j	$Y_j n_j$
Pintores	\$ 60	2	120
Carpinteros	\$ 40	5	200
Albañiles	\$ 30	10	300
Ayudantes	\$ 18	20	360
Plomero	\$ 12	40	480
		<u>77</u>	<u>1460</u>

$$\bar{Y} = \frac{\sum_{j=1}^5 Y_j n_j}{\sum_{j=1}^5 n_j} = \frac{1.460}{77} = 18,96$$

El promedio del salario diario ponderado es de \$ 18,96.

6.1.1.2 Propiedades de la Media Aritmética.

Primera propiedad:

La suma algebraica de las desviaciones (1) con respecto a la Media Aritmética es igual a cero cualquiera que sea la distribución.

$$Z_j = Y_j - a$$

$$Z_j n_j = Y_j n_j - a n_j$$

$$\sum_{j=1}^m Z_j n_j = \sum_{j=1}^m Y_j n_j - a \sum_{j=1}^m n_j$$

como:
$$a = \frac{\sum_{j=1}^m Y_j n_j}{n} \Rightarrow a n = \sum_{j=1}^m Y_j n_j$$

luego:
$$\sum_{j=1}^m Z_j n_j = a n - a n$$

entonces:
$$\sum_{j=1}^m Z_j n_j = 0$$

Ejemplo 13: Se comprueba esta propiedad en el ejemplo 3 que se ha venido desarrollando:

(1) Desviación es la diferencia entre los valores de la variable (X_i o Y_j) y un valor constante cualquiera. Es la misma definición de desviación dada para el cálculo de los momentos.

Clases					Z_j	
Y'_{j-1}	—	Y'_j	Y_j	n_j	$(Y_j - \bar{Y})$	$Z_j n_j$
1,40	—	5,90	3,65	9	- 9,45	- 85,05
5,90	—	10,40	8,15	10	- 4,95	- 49,50
10,40	—	14,90	12,65	16	- 0,45	- 7,20
14,90	—	19,40	17,15	5	4,05	20,25
19,40	—	23,90	21,65	4	8,55	34,20
23,90	—	28,40	26,15	4	13,05	52,20
28,40	—	32,90	30,65	2	17,55	35,10
				50		0

$$\sum_{j=1}^7 (Y_j - \bar{Y}) n_j = \sum_{j=1}^7 Z_j n_j = 0$$

Como consecuencia, la media aritmética de las desviaciones respecto a la media aritmética (Momento M_1) es cero.

$$M [Z] = m_1 = \frac{\sum_{j=1}^m Z_j n_j}{n} = \frac{0}{n} = 0$$

Segunda propiedad:

La suma de los cuadrados de las desviaciones con respecto a la media aritmética es mínima comparada con cualquier otra desviación:

$$Z'_j = Y_j - ot.$$

$$\sum_{j=1}^m Z_j'^2 n_j = \sum_{j=1}^m (Y_j - ot)^2 n_j$$

$$\sum_{j=1}^m Z_j'^2 n_j = \sum_{j=1}^m [(Y_j - a) + (a - ot)]^2 n_j$$

$$\sum_{j=1}^m Z_j'^2 n_j = \sum_{j=1}^m [(Y_j - a)^2 + 2(a - ot)(Y_j - a) + (a - ot)^2] n_j$$

$$\sum_{j=1}^m Z_j'^2 n_j = \sum_{j=1}^m (Y_j - a)^2 n_j + 2(a - ot) \sum_{j=1}^m (Y_j - a) n_j + (a - ot)^2 \sum_{j=1}^m n_j$$

$$\sum_{j=1}^m Z_j'^2 n_j = \sum_{j=1}^m Z_j^2 n_j + 0 + n(a - ot)^2$$

Si $a \neq ot$ entonces $n(a - ot)^2 > 0$

$$\Rightarrow \sum_{j=1}^m Z_j'^2 n_j \geq \sum_{j=1}^m Z_j^2 n_j$$

Ejemplo 14: (en base al ejemplo 3).

Clases $Y'_{j-1} - Y'_j$	Y_j	n_j	Z_j	Z_j^2	$Z_j^2 n_j$	Z'_j	$Z_j'^2$	$Z_j'^2 n_j$
1,40-5,90	3,65	9	-9,45	89,30	803,70	-9,0	81,00	729,00
5,90-10,40	8,15	10	-4,95	24,50	245,00	-4,5	20,25	202,50
10,40-14,90	12,65	16	-0,45	0,20	3,20	0,0	0,00	0,00
14,90-19,40	17,15	5	4,05	16,40	82,00	4,5	20,25	101,25
19,40-23,90	21,65	4	8,55	73,10	292,40	9,0	81,00	324,00
23,90-28,40	26,15	4	13,05	170,30	681,20	13,5	182,25	729,00
28,40-32,90	30,65	2	17,55	308,00	616,00	18,0	324,00	648,00
		50			2.723,50			2.733,75

luego,

$$\sum_{j=1}^m Z_j^2 n_j < \sum_{j=1}^m Z_j'^2 n_j$$

$$2723,50 < 2733,75$$

y se puede comprobar que a medida que el origen de trabajo difiera más de la media, esta desigualdad es mayor.

Tercera propiedad:

La media aritmética de una muestra dividida en sub-muestras es igual, a la media aritmética ponderada de las sub-muestras, tomando como ponderación los tamaños de las mismas.

Dada una muestra tamaño n y una Media $\bar{a}$ se consideran 2 sub-muestras tamaños n_1 y n_2 mutuamente excluyentes, con medias a_1 y a_2 entonces:

$$n = n_1 + n_2$$

$$a_1 = \frac{\sum_{j=1}^{m_1} Y_{1j} n_{1j}}{n_1} \quad \wedge \quad a_2 = \frac{\sum_{j=1}^{m_2} Y_{2j} n_{2j}}{n_2}$$

$$a = \frac{a_1 n_1 + a_2 n_2}{n_1 + n_2} \quad \Rightarrow$$

$$a = \frac{\sum_{j=1}^{m_1} \left[\frac{Y_{1j} n_{1j}}{n_1} \right] n_1 + \sum_{j=1}^{m_2} \left[\frac{Y_{2j} n_{2j}}{n_2} \right] n_2}{n_1 + n_2}$$

$$a = \frac{\sum_{j=1}^{m_1} Y_{1j} n_{1j} + \sum_{j=1}^{m_2} Y_{2j} n_{2j}}{n}$$

Ejemplo 15: (en base al ejemplo 3).

Sean $X1$, $X2$, $X3$, - - - $Xn1$ los precios de las acciones menores de \$ 14,90 y $X1$, $X2$, - - - $Xn2$ los precios de las acciones mayores o

iguales de \$ 14,90. Definidas las 2 sub-muestras de esta forma se conforman sus correspondientes distribuciones de frecuencias utilizando la misma metodología expuesta anteriormente:

Primera Sub-Muestra:

C L A S E S	Y_{1j}	n_{1j}	$Y_{1j} n_{1j}$
1,40 — 5,90	3,65	9	32,85
5,90 — 10,40	8,15	10	81,50
10,40 — 14,90	12,61	16	202,40
		35	316,75

luego:

$$a_1 = \frac{\sum_{j=1}^3 Y_{1j} n_{1j}}{n_1} = \frac{316,75}{35}$$

$$a_1 = \$ 9,05$$

Segunda Sub-Muestra:

C L A S E S	Y_{2j}	n_{2j}	$Y_{2j} n_{2j}$
14,90 — 19,40	17,15	5	85,75
19,40 — 23,90	21,65	4	86,60
23,90 — 28,40	26,15	4	104,60
28,40 — 32,90	30,65	2	61,30
		15	338,25

luego:

$$a_2 = \frac{\sum_{j=1}^{m_2} Y_{2j} n_{2j}}{n_2}$$

$$a_2 = \frac{338,25}{15} = \$ 22,55$$

En conclusión:

$$a = \frac{a_1 n_1 + a_2 n_2}{n}$$

$$a = \frac{316,75 + 338,25}{50} = \$ 13,10$$

Como se puede observar se obtiene el mismo valor que el calcu
lado inicialmente para la media aritmética. Debe aclararse que
las dos sub-muestras, en general, pueden escogerse de diferente
tes formas y tamaños.

Para este caso se han elegido las dos sub-muestras de tal manera
ra que el trabajo se facilitara y los dos cálculos fueran más corre
tos pero, en la práctica se pueden considerar más de 2 sub-muestr
tras; todos los cálculos conducen siempre a obtener el resultado

original para la media aritmética.

Cuarta propiedad:

La media aritmética de una constante es igual a dicha constante

Ejemplo: Supóngase que el precio de las acciones de 8 empresas en un determinado día fue de \$ 11,78, discriminadas así:

$$X_1 = 11,78$$

$$X_5 = 11,78$$

$$X_2 = 11,78$$

$$X_6 = 11,78$$

$$X_3 = 11,78$$

$$X_7 = 11,78$$

$$X_4 = 11,78$$

$$X_8 = 11,78$$

La media aritmética para esta serie de datos según la definición sería:

$$M(X) = \frac{\sum_{i=1}^n X_i}{n} = \frac{94,24}{8} = 11,78$$

En general:

$$M(K) = \frac{\sum_{i=1}^n K}{n} = \frac{nk}{n} = K$$

Quinta propiedad:

La media aritmética del producto de una constante por una variable

es igual a la constante por la Media Aritmética de la variable

$$M(KY) = \frac{\sum_{j=1}^m KY_j n_j}{n} = \frac{K \sum_{j=1}^m Y_j n_j}{n}$$

$$M(KY) = KM(Y)$$

Ejemplo 16: Debido a una devaluación el gobierno toma medidas económicas y decreta que el precio de las acciones sea cotizado al doble. La nueva distribución de frecuencias no altera la estructura que se tenía antes pero, tanto los límites de las clases, así como las marcas de clases, quedan multiplicados por la constante 2; la nueva distribución de frecuencia quedaría así:

J	$2(Y'_{j-1} - Y'_j)$	$2Y_j$	n_j	$2Y_j n_j$
1	2,80 - 11,80	7,30	9	65,70
2	11,80 - 20,80	16,30	10	163,00
3	20,80 - 29,80	25,30	16	404,80
4	29,80 - 38,80	34,30	5	171,50
5	38,80 - 47,80	43,30	4	173,20
6	47,80 - 56,80	52,30	4	209,20
7	56,80 - 65,80	61,30	2	122,60
			50	1.310,00

$$M(2Y) = \frac{\sum_{j=1}^n 2Y_j n_j}{n} = \frac{1.310,00}{50} = \$ 26,20$$

$$M(2Y) = 2M(Y)$$

$$M(2Y) = 2(13,1) \Rightarrow M(KY) = KM(Y)$$

Sexta propiedad:

La media aritmética de la suma de dos variables homogéneas es igual a la suma de las medias de dichas variables.

$$M \left[\sum_{k=1}^r Y_k \right] = \sum_{k=1}^r M[Y_k]$$

Ejemplo 17: En base a la información del ejemplo 3, se supone un aumento de 2% en los valores de la totalidad de acciones ($Y_j^2 =$ marcas de clase; Y_j^2 : 2% de aumento; Y_j^* : Variable total).

$Y_{j-1}^1 - Y_j^1$	Y_{1j}	Y_{2j}	Y_j^*	n_j	$Y_{1j} n_j$	$Y_{2j} n_j$	$Y_j^* n_j$
1,40 - 5,90	3,65	0,073	3,723	9	32,85	0,657	33,507
5,90 - 10,40	8,15	0,163	8,313	10	81,50	1,630	83,130
10,40 - 14,90	12,65	0,253	12,903	16	202,40	4,048	206,448
14,90 - 19,40	17,15	0,343	17,493	5	85,75	1,715	87,465
19,40 - 23,90	21,65	0,433	22,083	4	86,60	1,732	88,332
23,90 - 28,40	26,15	0,523	26,673	4	104,60	2,092	106,692
28,40 - 32,90	30,65	0,613	31,263	2	61,30	1,226	62,526
				50	655,00	13,100	668,100

$$M[Y^*] = \frac{\sum_{j=1}^7 Y_j^* n_j}{50} = \frac{668,100}{50} = 13,362$$

$$M[Y_1] = \frac{\sum_{j=1}^7 Y_{1j} n_j}{50} = \frac{655,00}{50} = 13,100$$

$$M[Y_2] = \frac{\sum_{j=1}^7 Y_{2j} n_j}{50} = \frac{13,100}{50} = 0,262$$

$$M[Y^*] = M[Y_1 + Y_2] = M[Y_1] + M[Y_2]$$

$$13,362 = 13,100 + 0,262$$

$$13,362 = 13,362$$

6.1.1.3 Procedimientos Abreviados de Cálculo: Estos procedimientos abreviados también son llamados "Métodos Indirectos" y son utilizados para el cálculo de la media aritmética con da tos agrupados, cuando los datos originales son muy exten-
sos, para simplificar el trabajo.

1. Primer Método Abreviado:

$$a = ot + \frac{\sum_{j=1}^m Z'_j n_j}{n}$$

Sabiendo que:

$$\frac{\sum_{j=1}^m (Y_j - ot) n_j}{n} = \frac{\sum_{j=1}^m Z'_j n_j}{n}$$

Demostración:

ot = Origen de trabajo

$$K + Z_j = Z'_j \quad (1)$$

$$K = a - ot \quad (2)$$

$$Z_j = Y_j - a \quad (3)$$

$$Z_j + K = Y_j - a + a - ot \quad (4)$$

$$Y_j - ot = Z'_j$$

$$Z_j + K = (Y_j - a) + (a - ot) = Z'_j \quad (5)$$

$$Z'_j = Z_j + K \quad (6)$$

$$Z'_j n_j = Z_j n_j + K n_j \quad (7)$$

$$\sum_{j=1}^m Z'_j n_j = \sum_{j=1}^m Z_j n_j + K \sum_{j=1}^m n_j \quad (8)$$

$$\sum_{j=1}^m Z'_j n_j = 0 + Kn \quad (9)$$

$$\sum_{j=1}^m Z'_j n_j = Kn \quad (10)$$

$$\frac{\sum_{j=1}^m Z'_j n_j}{n} = K$$

Pero:

$$K = a - ot$$

$$ot + K = a$$

Luego:

$$a = ot + \frac{\sum_{j=1}^m Z'_j n_j}{n}$$

2. Segundo Método Abreviado:

$$a = ot + C \frac{\sum_{j=1}^m Z''_j n_j}{n}$$

La media aritmética es igual al origen de trabajo, más el producto del intervalo de clase por la media de las desviaciones con respecto al origen de trabajo y medidas las desviaciones tomando como unidad el intervalo de clase.

Demostración:

$$C Z''_j = Z_j + CK_1$$

$$Z_j = Y_j - a$$

$$CK_1 = a - ot \quad K_1 = \frac{a - ot}{c}$$

$$Z_j + CK_1 = CZ''_j \quad (1)$$

$$CZ''_j = Z_j + CK_1 \quad (2)$$

$$CZ''_j n_j = Z_j n_j + CK_1 n_j \quad (3)$$

$$\sum_{j=1}^m CZ''_j n_j = \sum_{j=1}^m Z_j n_j + CK_1 \sum_{j=1}^m n_j \quad (4)$$

$$C \sum_{j=1}^m Z''_j n_j = CK_1 n \quad (5)$$

$$\frac{C \sum_{j=1}^m Z''_j n_j}{Cn} = K_1 \quad (6)$$

$$\frac{\sum_{j=1}^m Z''_j n_j}{n} = K_1 \quad (7)$$

pero:

$$C K_1 = a - ot \quad (8)$$

$$ot + C K_1 = a \quad (9)$$

$$a = ot + C K_1 \quad (10)$$

$$a = ot + C \frac{\sum_{j=1}^m Z''_j n_j}{n} \quad (11)$$

Ejemplo 18:

Primer procedimiento.

Fórmula general:

$$a = ot + \frac{\sum_{j=1}^m Z'_j n_j}{n}$$

Reemplazando: (Ver ejemplo 7)

$$a = ot + \frac{\sum_{j=1}^m Z'_j n_j}{n} = 12.65 + \frac{22.5}{50} = \$ 13.10$$

Ejemplo 19:

Segundo procedimiento.

Fórmula general:

$$a = ot + \frac{C \sum_{j=1}^m Z''_j n_j}{n}$$

$Y'_{j-1} - Y'_j$	Y_j	n_j	Z''_j	$Z''_j n_j$
1.40 - 5.90	3.65	9	- 2	- 18
5.90 - 10.40	8.15	10	- 1	- 10
10.40 - 14.90	12.65	16	0	0
14.90 - 19.40	17.15	5	1	5
19.40 - 23.90	21.65	4	2	8
23.90 - 28.40	26.15	4	3	12
28.40 - 32.90	30.65	2	4	8

$$\sum_{j=1}^m Z''_j n_j = 5$$

$$\frac{\sum_{j=1}^m Z''_j n_j}{n} = \frac{5}{50} = 0.10$$

$$a = ot + C \frac{\sum_{j=1}^m Z''_j n_j}{n}$$

$$a = 12,65 + 4.5 (0,10)$$

$$a = 12,65 + 0,45$$

$$a = 13,10$$

En caso de tener en una distribución intervalos desiguales no es práctico utilizar este método.

6.1.1.4 Ventajas y Desventajas del Promedio Aritmético.

Las principales ventajas son:

- 1 - Es la medida más usada
- 2 - Es de fácil comprensión
- 3 - Su cálculo es fácil y perfectamente determinado; solo son necesarios para su cálculo los valores totales y el número de elementos.
- 4 - En su determinación influyen todos los valores de la serie.
- 5 - Es susceptible de toda clase de tratamiento algebraico.

La única desventaja en algunos casos sería que su valor puede estar demasiado afectado por los valores extremos de la serie y por ello no es representativo.

6.1.2 Media Geométrica: Se define como la raíz enésima de la productoria de los n valores de la variable ; se presenta por Mo .

Para su cálculo ocurren dos casos :

- a) Uno para observaciones originales
- b) Y uno para datos agrupados en variable discreta o en variable continua,

- a) Para datos originales es igual al valor de un radical que tiene por índice el tamaño de la muestra y por cantidad subradical el producto de todos los datos originales

$$1 - Mo(X) = \sqrt[n]{\prod_{i=1}^n X_i}$$

$$2 - Mo(X) = \sqrt[n]{X_1 X_2 X_3 \dots X_n}$$

El cálculo de la media geométrica se facilita reduciendo la fórmula a su forma logarítmica, como

$$\text{Log } Mo = \frac{\log X_1 + \log X_2 + \dots + \log X_n}{n}$$

$$\text{Log } Mo = \frac{1}{n} \sum_{i=1}^n \text{Lg } X_i \quad \text{de donde :}$$

$$Mo = \text{Antilog } \frac{\sum_{i=1}^n \text{Log } X_i}{n}$$

- b) Para datos agrupados será definida como el valor de un radical que tiene por índice el tamaño de la muestra y por cantidad subradical el producto de cada valor de la variable, elevado a una potencia n_j ; es decir el valor de la frecuencia absoluta de cada clase.

$$Mo(Y) = \sqrt[n]{\prod_{j=1}^m Y_j^{n_j}}$$

$$Mo(Y) = \sqrt[n]{Y_1^{n_1} Y_2^{n_2} Y_3^{n_3} \dots Y_n^{n_m}}$$

$$\text{Log } Mo(Y) = \frac{\sum_{j=1}^m (\text{Log } Y_j) n_j}{n}$$

$$Mo = \text{Antilog } \frac{\sum_{j=1}^m (\log Y_j) n_j}{n}$$

Para variable continúa será:

Fórmula General:

$$\text{Log Mo} = \frac{1}{n} \sum_{j=1}^m (\log Y_j) n_j$$

$$\text{Mo} = \text{Antilog } M (\text{Log } Y)$$

Ejemplo 20:

CLASES $Y'_{j-1} - Y'_j$	Y_j	n_j	Logaritmo (Y_j)	$(\text{Log } Y_j) n_j$
1,40 - 5,90	3,65	9	0,56229	5,06061
5,90-10,40	8,15	10	0,91116	9,1116
10,40-14,90	12,65	16	1,10209	17,63344
14,90-19,40	17,15	5	1,23426	6,17130
19,40-23,90	21,65	4	1,33546	5,34184
23,90-28,40	26,15	4	1,41747	5,66988
28,40-32,90	30,65	2	1,48643	2,97286

$$\begin{array}{ccc} 50 & m & \\ & \sum_{j=1} & (\text{Log } Y_j) n_j = 51,96153 \end{array}$$

Aplicando la fórmula:

$$\text{Log Mo} = \frac{51,96153}{50} = 1,03923$$

$$\text{Mo} = \text{Antilog } 1,03923$$

$$M_o = \$ 10,94$$

Lo cual indica que el promedio de las acciones de las 50 empresas es \$ 10,94.

6.1.2.1 Propiedades y Usos.

- 1 - La media geométrica será cero ($M_o = 0$) cuando uno de los valores de la variable sea igual a cero.
- 2 - Si uno o varios de los valores de la variable son negativos, la media geométrica no tiene sentido, porque da origen a raíces imaginarias que en ningún momento explican el problema.

Las siguientes son las principales aplicaciones o usos de la media geométrica:

- 1 - Se utiliza para calcular promedios de números índices, razón de capital a pasivo, etc. y en general valor relativo.
- 2 - Para conocer el crecimiento promedio de una serie cronológica.

3 - Cuando se utilizan los datos en función de sus logaritmos y se desea saber el promedio de aquellos.

6.1.3 Media Armónica : Se define como el recíproco de la media aritmética de los recíprocos de las observaciones y se representa por M_{-1}

$$M_{-1} = \frac{1}{\frac{1/X_1 + 1/X_2 + \dots + 1/X_n}{n}}$$

$$M_{-1} = \frac{1}{\frac{\sum (1/X_j)}{n}} = \frac{n}{\sum (1/X_j)}$$

Así : La media armónica de los números 3, 4, 6 es :

$$M_{-1} = \frac{3}{1/3 + 1/4 + 1/6} = \frac{3}{\frac{4 + 3 + 2}{12}}$$

$$M_{-1} = \frac{3}{9/12} = 4$$

Dois (2) obreros A y B gastan 30 y 20 minutos respectivamente en hacer una corbata. Cuál es el tiempo medio requerido para terminar una corbata?

Se tiene que : $n = 2$

$$M - 1 = \frac{2}{1/30 + 1/20} = \frac{2}{\frac{2 + 3}{60}}$$

$$M - 1 = \frac{2}{\frac{5}{60}} = 24$$

Entonces el tiempo medio requerido para terminar una corbata es de 24 minutos.

Para datos agrupados, la fórmula es :

$$M - 1 = \frac{1}{\frac{n_1/Y_1 + n_2/Y_2 + \dots + n_j/Y_j}{n}}$$

$$M - 1 = \frac{1}{\frac{\sum(n_j/Y_j)}{n}} = \frac{n}{\sum(n_j/Y_j)}$$

$Y'_{j-1} - Y'_j$	Y_j	n_j	n_j / Y_j
1,40 - 5,90	3,65	9	9/ 3,65 = 2 465
5,90 - 10,40	8,15	10	10/ 8,15 = 1,227
10,40 - 14,90	12,65	16	16/ 12,65 = 1,265
14,90 - 19,40	17,15	5	5/ 17,15 = 0 291
19,40 - 23,90	21,65	4	4/ 21,65 = 0,185
23,90 - 28,40	26,15	4	4/ 26,15 = 0 153
28,40 - 32,90	30,65	2	2/ 30,65 = 0 065
		50	5,651

$$\sum_{j=1}^7 n_j / Y_j = 5,651$$

Aplicando la fórmula se tiene:

$$M_{\text{ar}} = \frac{50}{5,651} = \$ 8.848$$

Se acostumbra utilizar la media armónica:

- 1 - Cuando se desea obtener un valor representativo de un conjunto de datos expresados en forma de tasas (tantas unidades de un tipo por cada unidad de otra especie).
- 2 - Para promediar datos inversamente proporcionales

- 3 - Cuando se trata de estudiar rendimientos respecto a un patrón fijo.

En general los promedios se relacionan mediante la siguiente desigualdad:

$$M < M_o < M$$

6.2 Promedios de Posición. Dentro de una distribución de frecuencias, algunas veces, se desean obtener valores con una posición determinada, que permita dividir la distribución en partes iguales, en cuanto a las observaciones. Cada parte recibe un nombre y una numeración diferente en forma creciente, siendo las más importantes la división en dos, cuatro, diez y cien partes, llamadas FRACTILAS

- También se puede calcular un valor que corresponda al valor máximo o punto en el cual se sitúa el mayor número de datos, que se llamará un valor MODAL o MODA de la distribución.

6.2 Mediana. La mediana se define como aquel valor que supera a no más de la mitad de las observaciones, y a

su vez es superado por no más de la mitad de las observaciones. Se representa por M_e .

Se puede determinar:

- a) Cuando se tienen datos originales.
- b) Agrupados.

6.2.1.1 Cálculo en Datos Originales. En el caso de datos no agrupados hay que distinguir si el número de observaciones es impar o si es par.

- a) Si en el número de observaciones es impar, la mediana estará representada por el valor central de las observaciones dispuestas en orden creciente o decreciente en valor.

Ejemplo 22: Tenemos los siguientes salarios:

50		30	90
70	Se ordenan en orden	45	70
90	creciente o decreciente	50 ← M_e →	50
30		70	45
45		90	30

La mediana corresponde al valor central 50.

- b) En el caso de que el número de observaciones sea par, podemos establecer el valor de la mediana como el promedio o media aritmética de los dos datos centrales (siempre y cuando estén ordenados en forma creciente o decreciente), aunque cualquier valor comprendido entre ese intervalo es una mediana, incluyendo los extremos del intervalo.

Ejemplo 23:

$$\begin{array}{r} 1 \\ 3 \\ 4 \\ 7 \\ 10 \\ 12 \end{array} \quad M_e = 5,5 \quad \frac{4 + 7}{2} = \frac{11}{2} = 5,5$$

$n = 6 \quad n/2 = 6/2 = 3 \text{ observaciones}$

Cualquier valor comprendido en el intervalo 4 - 7, es una mediana, incluyendo los extremos.

- 4 Cumple con la definición, no supera a 3 observaciones que es el valor medio del número de observaciones.
- 7 Tampoco es superado por más de 3 observaciones y por consiguiente cumple con la definición.

Esto supone que hay un infinito número de medianas; para obviar esta dificultad se tomará siempre el valor medio - del intervalo. En este ejemplo la mediana es 5,5. Estas formas de determinar la MEDIANA, solo son aplicables cuando el tamaño de la muestra es reducido; tratándose de muestras grandes hay que recurrir a la agrupación de datos y a las frecuencias absolutas acumuladas.

6.2.1.2 Cálculo en Datos Agrupados. En el cálculo de la mediana con datos agrupados se presentan dos casos cuando la variable es discreta y otros dos cuando la variable es continua.

1) Cuando la variable es DISCRETA se presentan 2 alternativas que se observan en el polígono de frecuencias acumuladas:

a)

$$N_{j-1} < n/2 < N_j$$

b)

$$N_{j-1} = n/2$$

- a) En esta situación el valor que corresponde a $N_j - 1$ (en las frecuencias acumuladas) es Y_{j-1} (en la variable), lo mismo ocurre con N_j cuyo valor es Y_j , luego la mediana se determina por el valor de la variable correspondiente a Y_j .
- b) El segundo caso se presenta cuando $N_j - 1 = n/2$. En este caso el valor de la mediana queda determinado por el Y'_{j-1} y Y'_j correspondientes:

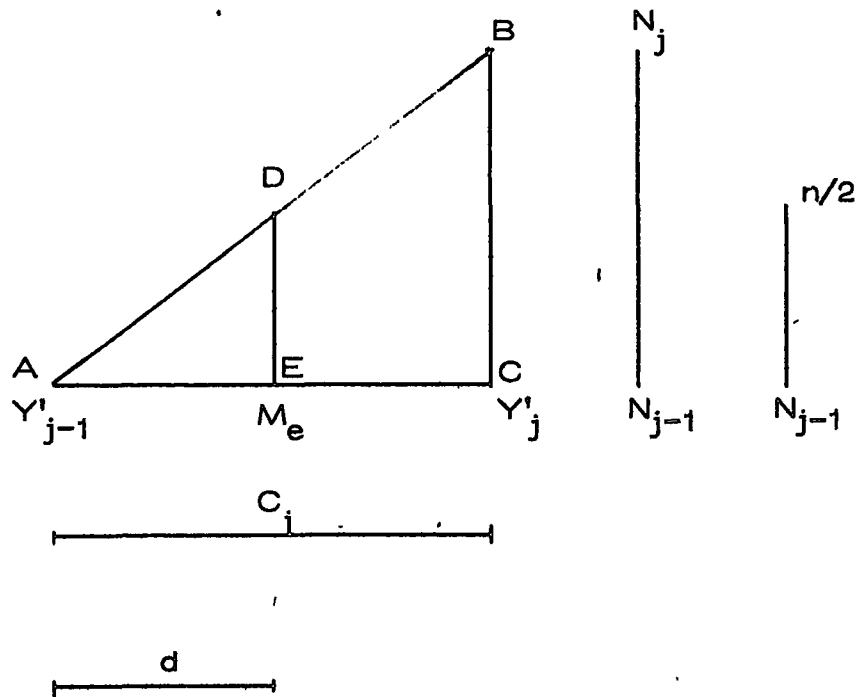
$$M_e = \frac{Y'_{j-1} + Y'_j}{2}$$

El cálculo de la mediana para variable discreta puede comprenderse al examinar el ejercicio del Anexo No. 1.

- 2) Cuando la variable en estudio sea del tipo CONTINUA, como en el caso anterior, se presentan dos posibilidades para las cuales, se debe tener en cuenta el gráfico de las frecuencias absolutas acumuladas.

- a) Supóngase que el polígono de frecuencias acumuladas es de la forma:

exactamente el valor de M_e ; extrayendo de la gráfica los triángulos formados, se tiene:



Los triángulos ADE y ABC son semejantes, por tal razón:

$$\frac{AE}{AC} = \frac{DE}{BC} \quad AE = \frac{AC \times DE}{BC}$$

Reemplazando por los valores de cada uno:

$$d = \frac{C_j \times (n/2 - N_{j-1})}{N_j - N_{j-1}}$$

$$d = C_j \left[\frac{n/2 - N_{j-1}}{N_j - N_{j-1}} \right]$$

De donde se deduce la fórmula de la mediana para variable CONTINUA en datos agrupados, cuando $N_{j-1} < n/2$

$$M_e = Y'_{j-1} + d$$

$$M_e = Y'_{j-1} + C_j \left[\frac{n/2 - N_{j-1}}{N_j - N_{j-1}} \right]$$

b) Si en la gráfica de frecuencias absolutas acumuladas:

$$N_{j-1} = n/2$$

se concluye que $d = 0$; en consecuencia:

$$M_e = Y'_{j-1} + d$$

$$M_e = Y'_{j-1}$$

Ejemplo 24: En la práctica de acuerdo al ejercicio desarrollado, su cálculo sería:

CLASES	$\dot{Y}_j$	n_j	N_j
1,40 - 5,90	3,65	9	9
5,90 - 10,40	8,15	10	19
10,40 - 14,90	12,65	16	35
14,90 - 19,40	17,15	5	40
19,40 - 23,90	21,65	4	44
23,90 - 28,40	26,15	4	48
28,40 - 32,90	30,65	2	50
50			

Se tiene:

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

luego:

$$N_{j-1} < 25 < N_j$$

$$N_2 < 25 < N_3$$

El valor Y'_{j-1} está dado por el límite inferior de la clase a la que corresponde la frecuencia absoluta acumulada (N_j) inmediatamente superior a $n/2$, lo mismo ocurre con el valor de C_j en el caso de que no sea constante.

$$M_e = 10,40 + 4,5 \left[\frac{25 - 19}{35 - 19} \right]$$

$$M_e = 10,40 + 4,5 \left[\frac{6}{16} \right]$$

Se puede observar que el denominador del segundo término está dado por la frecuencia absoluta de la clase en la cual se encuentra la mediana ($N_j - N_{j-1} = n_j$).

$$\begin{aligned} M_e &= 10,40 + 4,5 (0,375) \\ &= 10,40 + 1,6875 \\ &= 12,0875 \approx \$ 12,09 \\ M_e &= \$ 12,09 \end{aligned}$$

lo cual indica que el 50% de los precios de las acciones, son menores de \$ 12,09 y el otro 50% es superior a este valor. Para este ejercicio los precios de las acciones de 25 empresas serán menores de \$ 12,09 y las restantes tendrán precios mayores de \$ 12,09.

6.2.1.3 Características de la Mediana.

- a) Es una medida de posición que se sitúa en la mitad de la distribución donde su valor es superado por la mitad de las observaciones y además supera a la mitad.

- b) El valor de la mediana no es afectado por los valores extremos; esto se explica debido a su definición.
- c) La suma de las desviaciones es mínima cuando estas desviaciones lo son con respecto a la mediana en comparación con un origen de trabajo "ot" diferente de la mediana.

6.2.2 Fractilas. Como se dijo antes, corresponden a la división en partes iguales de la distribución. Supóngase que se quiere dividir la distribución en S partes iguales; sea $r = 1, 2, \dots, n$, S el orden de cada una de éstas y P el símbolo particular para cada división; de tal manera que se tiene a las siguientes relaciones generalizadas de la mediana con los mismos casos:

$$P_{r/s} = Y'_{j-1} + C_j \left[\frac{\frac{rn}{S} - N_{j-1}}{N_j - N_{j-1}} \right]$$

Los valores de Y'_{j-1} , C_j , N_j y N_{j-1} quedan determinados al localizar en la distribución el valor de $\frac{rn}{S}$ en las frecuencias acumuladas absolutas.

Supóngase ahora que:

$$S = 4 \rightarrow r = 1, 2, 3, 4,$$

$P_{r/s}$ se llamarán cuartiles o cuartías y dividen la distribución en 4 partes iguales de tal manera que cualquier valor supere unas observaciones y sea superado por los demás. Considerando el cuartil uno para variable continua, en el cual:

$$N_{j-1} < \frac{n}{4}$$

$$P_{1/4} = Y'_{j-1} + C_j \cdot \frac{\frac{n}{4} - N_{j-1}}{N_j - N_{j-1}}$$

Su valor supera exactamente a una cuarta parte de las observaciones y es superado por las tres cuartas partes de las mismas.

Si:

$$S = 10 \rightarrow r = 1, 2, 3, \dots, 10$$

cada $P_{r/s}$ se llamará un DECIL con las mismas propiedades enumeradas anteriormente; escogiendo alguno de ellos, ($r = 6$), estará definido el DECIL 6:

$$P_{6/10} = Y'_{j-1} + C_j \left[\frac{\frac{6n}{10} - N_{j-1}}{N_j - N_{j-1}} \right]$$

Si,

$$S = 100 \Rightarrow r = 1, 2, 3, 4, \dots, 100$$

cada uno se llamará un PERCENTIL, definido en los mismos términos:

$$P_{77/100} = Y'_{j-1} + C_j \left[\frac{\frac{77n}{100} - N_{j-1}}{N_j - N_{j-1}} \right]$$

Especial mención merecen algunas de estas divisiones tales como:

$$P_{2/4}, P_{5/10}, \text{ y } P_{50/100}$$

ya que en estos puntos se halla un valor central equivalente a la mediana de la distribución y se expresa como $P_{1/2}$.

Ejemplo 25:

CLASES	Y_j	n_j	N_j
1,40 - 5,90	3,65	9	9
5,90 - 10,40	8,15	10	19
10,40 - 14,90	12,65	16	35
14,90 - 19,40	17,15	5	40
19,40 - 23,90	21,65	4	44
23,90 - 28,40	26,15	4	48
28,40 - 32,90	30,65	2	50

$$Q_1 = P_{1/4} \Rightarrow \frac{rn}{4} = \frac{50}{4} = 12,50$$

$$P_{1/4} = 5,90 + 4,50 \left[\frac{12,50 - 9}{19 - 9} \right]$$

$$P_{1/4} = 5,90 + 4,50 \left[\frac{3,50}{10} \right]$$

$$P_{1/4} = 5,90 + 4,50 (0,35)$$

$$P_{1/4} = 5,90 + 1,575$$

$$P_{1/4} = 7,475 \approx \$ 7,48$$

Cálculo de $P_{3/4}$:

$$\frac{rn}{S} = \frac{3(50)}{4} = 37,50$$

$$P_{3/4} = 14,90 + 4,5 \left[\frac{37,50 - 35,00}{40,00 - 35,00} \right]$$

$$P_{3/4} = 14,90 + 4,5 \left[\frac{2,50}{5,00} \right]$$

$$P_{3/4} = 14,90 + 2,25$$

$$P_{3/4} = \$ 17,15$$

Los resultados obtenidos para $P_{1/4}$ y $P_{3/4}$ indican que el 50% de los datos deben estar comprendidos entre \$ 7,48 y \$ 17,15 ya que 7,48 es superado por 3/4 partes de los datos y 17,15 por 1/4, entonces entre estos dos valores estará 1/2 de ellos.

Cálculo de $P_{1/10}$:

$$\frac{rn}{S} = \frac{1 (50)}{10} = 5$$

$$P_{1/10} = 1,40 + 4,5 \left[\frac{5 - 0}{9 - 0} \right]$$

$$P_{1/10} = 1,40 + 4,5 (.555)$$

$$P_{1/10} = 1,40 + 2,499$$

$$P_{1/10} = 3,8999 \approx 3,90$$

$$P_{1/10} = \$ 3,90,$$

Cálculo de $P_{9/10}$:

$$\frac{rn}{S} = \frac{9 (50)}{10} = 45$$

$$P_{9/10} = 23,90 + 4,5 \left[\frac{45 - 44}{48 - 44} \right]$$

$$P_{9/10} = 23,90 + 4,5 \quad (0,25)$$

$$P_{9/10} = 23,90 + 1,125$$

$$P_{9/10} = \$ 25,025 \approx \$ 25,02$$

Según los resultados obtenidos el 80% de los datos estarán comprendidos entre los límites \$ 3,90 y \$ 25,02; a su vez se sabe que 3,90 supera a la décima parte de las observaciones y \$25,02 supera a 9/10 partes de las mismas y es superado por una décima parte.

Para el cálculo de $P_{10/100}$ y $P_{90/100}$ se observa que sus valores y la interpretación, es la misma que para $P_{1/10}$ y $P_{9/10}$.

En general la metodología de cálculo de las fractilas sigue los mismos principios estudiados en la mediana de acuerdo al tipo de variable que se esté estudiando, esto es, si es discreta o continua y tomando siempre como referencia para la localización de N_j y N_{j-1} , y la expresión:

$$\frac{rn}{S}$$

6.2.3 Modo o Moda. Se define la moda, mode, promedio típico o valor modal (Md) como el valor más frecuente dentro de la serie y corresponde al máximo de la distribución.

Para datos originales su cálculo se basa en la observación del valor que más se repite cuando los datos se hayan ordenado en forma creciente; al igual que la mediana no es afectada por los valores extremos.

En una distribución de frecuencias es necesario que las frecuencias absolutas inmediatamente anterior y posterior a la clase modal, sean menores que la frecuencia absoluta de esa clase:

$$a) \quad n_j > n_{j-1}$$

$$b) \quad n_j > n_{j+1}$$

Luego:

$$n_{j-1} < n_j > n_{j+1}$$

En algunos casos la desigualdad anterior se presenta en 2 o más clases por lo cual la distribución puede ser bimodal, trimodal, plurimodal de acuerdo al número de máximos que

se presenten pero, de todas maneras, debe tomarse como valor modal el máximo absoluto.

Para una distribución de variable DISCRETA se define la moda como el valor de la marca de clase (Y_j) en la cual se halla la mayor frecuencia absoluta, o sea que:

$$Md = Y_j$$

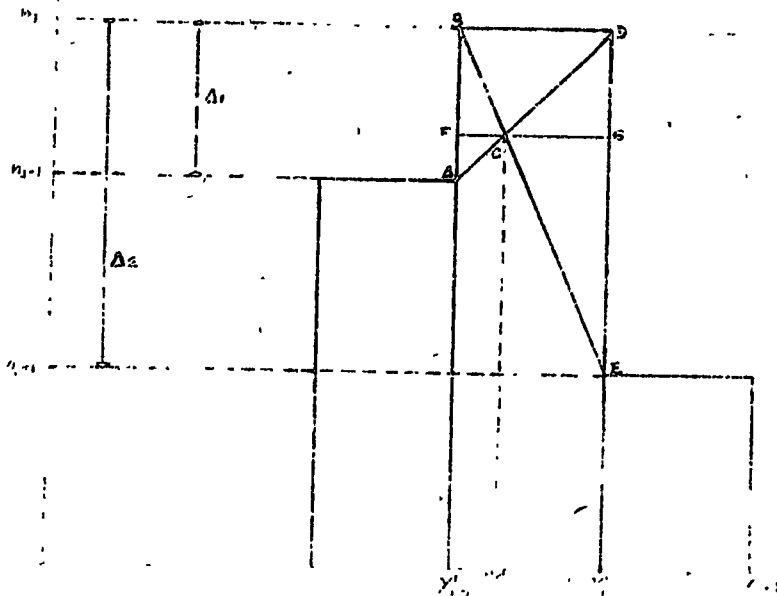
Para su cálculo ver Anexo No. 1.

Si la variable es CONTINUA se conocen 2 caminos para obtener el valor de Md:

- a) Una fórmula empírica que no tiene justificación de tipo matemático y que solo la experiencia ha enseñado; tal fórmula expresa el valor de la moda en los términos siguientes:

$$Md = Y'_{j-1} + C_j \left[\frac{n_{j+1}}{n_{j-1} + n_{j+1}} \right]$$

- b) Una deducción de tipo matemático basados en el histograma de frecuencias absolutas tanto de la clase modal como de las adyacentes de la siguiente forma :



Se forman los triángulos BCA y DCE que por ser semejantes, se puede establecer la siguiente relación:

$$\frac{FC}{BA} = \frac{CG}{DE}$$

escrita en otros términos determina la Md :

$$\frac{Md - Y'_{j-1}}{\Delta_1} = \frac{Y'_j - Md}{\Delta_2}$$

./.

$$(M_d - Y'_{j-1}) \Delta_2 = (Y'_j - M_d) \Delta_1$$

$$M_d \Delta_2 - Y'_{j-1} \Delta_2 = Y'_j \Delta_1 - M_d \Delta_1$$

$$M_d \Delta_2 + M_d \Delta_1 = Y'_j \Delta_1 + Y'_{j-1} \Delta_2$$

$$M_d = \frac{Y'_j \Delta_1 + Y'_{j-1} \Delta_2}{\Delta_1 + \Delta_2}$$

Pero:

$$Y'_j = Y'_{j-1} + C_j \Rightarrow Y'_j \Delta_1 = Y'_{j-1} \Delta_1 + C_j \Delta_1$$

Reemplazando:

$$M_d = \frac{Y'_{j-1} \Delta_1 + C_j \Delta_1 + Y'_{j-1} \Delta_2}{\Delta_1 + \Delta_2}$$

$$M_d = \frac{Y'_{j-1} (\Delta_1 + \Delta_2) + C_j \Delta_1}{(\Delta_1 + \Delta_2)}$$

Separando cada término se tiene finalmente:

$$M_d = Y'_{j-1} + C_j \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right]$$

Reemplazando Δ_1 y Δ_2 (ver figura) por sus equivalentes, en
tonces:

$$Md = Y'_{j-1} + C_j \left[\frac{n_j - n_{j-1}}{(n_j - n_{j-1}) + (n_j - n_{j+1})} \right]$$

que es la expresión para la obtención de la moda. Se debe tener en cuenta que aunque el cálculo de la moda por uno u otro procedimiento produce resultados algo diferentes, cualquiera de los dos sirve para representarla.

Cuando la amplitud de las clases de la distribución de frecuencias sea desigual, es necesario tener cuidado al establecer la moda, porque la mayor frecuencia puede presentarse debido a que la clase supuesta modal, tenga una amplitud mayor, comparada con otra en donde se encuentre verdaderamente la moda. Para evitar esto es necesario tener en cuenta la relación existente entre la frecuencia absoluta y la amplitud de clase, ya que la moda se encuentra en aquella clase en la cual esta relación sea mayor.

Se puede igualmente estudiar el concepto contrario que es el denominado antimoda o aquel valor de la variable que corresponde a

la mínima frecuencia; luego el planteamiento sería el inverso de la moda.

Ejemplo 26:

Cálculo de la Moda para Variable Continua.

CLASES	Y_j	n_j
1,40 - 5,90	3,65	9
5,90 - 10,40	8,15	10
10,40 - 14,90	12,65	16
14,90 - 19,40	17,15	5
19,40 - 23,90	21,65	4
23,90 - 28,40	26,15	4
28,40 - 32,90	30,65	2

50

Aplicando la fórmula deducida en a el valor de la moda sería:

$$Md = 10,40 + 4,50 \left[\frac{5}{10 + 5} \right]$$

$$Md = 10,40 + 4,50 \left[\frac{5}{15} \right]$$

$$M d = 10,40 + \frac{22,50}{15}$$

$$M d = 10,40 + 1,50$$

$$M d = \$ 11,90$$

Aplicando la fórmula deducida en b se tiene:

$$M d = 10,40 + 4,50 \left[\frac{16 - 10}{(16 - 10) + (16 - 5)} \right]$$

$$= 10,40 + 4,50 \left[\frac{6}{6 + 11} \right]$$

$$= 10,40 + 4,50 \left[\frac{6}{17} \right]$$

$$= 10,40 + 4,50 (0, 3529)$$

$$= 10,40 + 1,5881$$

$$M d = 11,9881 \approx \$ 11,99$$

El mayor valor que alcanzan las acciones de las 50 empresas es de \$ 11,99.

7. MEDIDAS DE DISPERSION

En los capítulos anteriores se estudiaron las funciones o estadígrafos de posición dado que en general los datos tienden a agruparse; se puede también calcular en que magnitud se alejan del valor o valores centrales, o se concentran respecto a éste, es decir, medir el grado de dispersión que tiene la serie objeto de estudio respecto del valor central.

El estudio de la dispersión en estadística descriptiva se enfoca desde dos puntos de vista:

7.1 Dispersión Absoluta: La que se obtiene relacionando valores absolutos de una misma serie. La comparación entre parámetros de dispersión absoluta se hace con series que se refieren a un fenómeno análogo. Las principales medidas de dispersión absoluta son:

La varianza

La desviación standard

La desviación media

El rango o recorrido de la variable

Desviación mediana

Recorrido intercuartílico

7.2 Dispersión Relativa: Resulta de una relación determinada entre parámetros absolutos de la distribución. Las siguientes medidas de dispersión relativa serán objeto de este estudio:

El coeficiente de variación

Coeficiente de deformación (asimetría)

Coeficiente de apuntamiento (curtosis).

7.1.1 Varianza (S^2): Se define como la media aritmética de los cuadrados de las diferencias entre los valores de la serie y la media aritmética ($\bar{X}$); la fórmula para datos originales es:

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

Ejemplo 27: Varianza para datos originales. (véase ejemplo 3)

No. Orden	$X_i - 12,92$	$(X_i - 12,92)^2$	No. Orden	$X_i - 12,92$	$(X_i - 12,92)^2$
1	- 7.59	57.6081	26	- 0.77	0.5929
2	- 10.01	100.2001	27	15.08	227.4064
3	- 7.01	49.1401	28	3.14	9.8596
4	- 10.02	100.4004	29	19.00	361.0000
5	- 1.49	2.2201	30	- 0.92	0.8464
6	0.57	0.3249	31	31.01	169.2601
7	10.17	103.4289	32	- 5.92	35.0464
8	- 6.73	45.2929	33	- 6.77	45.8329
9	16.92	286.2864	34	- 9.92	98.4064
10	13.97	195.1609	35	2.94	8.6436
11	- 5.68	32.2624	36	0.63	0.3969
12	3.85	14.8225	37	- 9.28	86.1184
13	0.16	0.0156	38	- 6.05	36.6025
14	- 0.11	0.0121	39	- 0.18	0.0324
15	- 0.52	0.2704	40	- 2.62	6.8644
16	4.06	16.4836	41	9.08	82.4464
17	- 3.45	11.9025	42	- 7.92	62.7264
18	- 3.94	15.5236	43	6.66	44.3556
19	- 1.92	3.6864	44	- 10.12	102.4144
20	- 2.32	5.3824	45	- 10.52	110.6704
21	13.17	173.4489	46	- 1.20	1.4400
22	1.18	1.3924	47	- 8.39	70.3921
23	2.76	7.6176	48	0.28	0.0784
24	- 3.97	15.7609	49	0.08	0.0064
25	- 0.52	0.2704	50	9.08	82.4464

TOTAL = 2,882,8003

$$S^2 = \frac{2,882,8003}{50}$$

$$S^2 = 57,6562$$

Para datos agrupados se toman las diferencias entre las marcas de clase (valores de Y_j) y la media aritmética ($\bar{Y}$), elevando al cuadrado y multiplicando este resultado por n_j .
Luego:

$$S^2 = \frac{\sum_{j=1}^m (Y_j - \bar{Y})^2 n_j}{n} = \frac{\sum_{j=1}^m Z_j^2 n_j}{n}$$

La varianza da una primera aproximación sobre la idea de dispersión, al comparar dos o más series referidas a fenómenos análogos; a mayor valor de la varianza mayor es la dispersión de las observaciones y lo contrario. La varianza es por su definición una función cuadrada.

Ejemplo 28:

CLASES	Y_j	n_j	Z_j	$Z_j n_j$	$Z_j^2 n_j$
1,40 - 5,90	3,65	9	- 9,45	- 85,05	803,72
5,90 - 10,40	8,15	10	- 4,95	- 49,50	245,03
10,40 - 14,90	12,65	16	- 0,45	- 7,20	3,24
14,90 - 19,40	17,15	5	4,05	20,25	82,01
19,40 - 23,90	21,65	4	8,55	34,20	292,41
23,90 - 28,40	26,15	4	13,05	52,20	687,21
28,40 - 32,90	30,65	2	17,55	35,10	616,01
		50			2.723,63

$$S^2 = \frac{2.723,63}{50}$$

$$S^2 = 54,47$$

Como se dijo anteriormente este valor de la varianza

$$S^2 = 54,47$$

es una primera cuantificación de la dispersión de los datos; para este caso se puede considerar como un valor más bien grande (este jui cio es dado principalmente en base al tipo de datos que se están tratando y al valor obtenido anteriormente para la media aritmética), luego es posible que la distribución tienda a ser heterogénea.

El valor $S^2 = 54,47$ no se acostumbra a expresarse acompañado de ningún tipo de unidad, porque no tiene significado real el hacerlo; como ejemplo de esto tendríamos el valor S^2 expresado en unidades de pesos cuadrados.

La comparación de mayor o menor homogeneidad (o heterogeneidad) se hace más fácil cuando se tienen dos o más varianzas de distribuciones análogas, que de acuerdo a los valores obtenidos muestran cuales de las distribuciones en mención son más homogéneas y cuales más heterogéneas.

Es bueno anotar, que la varianza calculada para datos originales y agrupados es generalmente diferente, en razón a que al agrupar los datos y tomar un valor (marca de clase) como representativo de un grupo de ellos, se está cambiando en una forma leve la es tructura de homogeneidad o heterogeneidad de la distribución.

Otra forma de cálculo para la varianza se puede obtener partiendo de la definición:

$$Z_j = Y_j - \bar{Y}$$

$$Z_j^2 = (Y_j - \bar{Y})^2 = Y_j^2 - 2\bar{Y}Y_j + \bar{Y}^2$$

$$Z_j^2 n_j = Y_j^2 n_j - 2\bar{Y}Y_j n_j + \bar{Y}^2 n_j$$

$$S^2 = \frac{\sum_{j=1}^m Z_j^2 n_j}{n} = \frac{\sum_{j=1}^m Y_j^2 n_j}{n} - 2\bar{Y} \frac{\sum_{j=1}^m Y_j n_j}{n} + \bar{Y}^2 \frac{\sum_{j=1}^m n_j}{n}$$

Donde:

$$\bar{Y} = \frac{\sum_{j=1}^m Y_j n_j}{n} \text{ y } \frac{\sum_{j=1}^m n_j}{n} = 1$$

Luego:

$$S^2 = \frac{\sum_{j=1}^m Y_j^2 n_j}{n} - 2\bar{Y}^2 + \bar{Y}^2$$

$$S^2 = \frac{\sum_{j=1}^m Y_j^2 n_j}{n} - \bar{Y}^2$$

$$S^2 = M[Y^2] - (M[Y])^2.$$

Para datos sencillos sin agrupar, es:

$$S^2 = M[X^2] - (M[X])^2$$

Este método es mucho más práctico, especialmente cuando resulta muy engorroso el cálculo de las diferencias y luego el elevarlas al cuadrado.

Ejemplo 29:

CLASES	Y_j	n_j	$Y_j n_j$	$Y_j^2 n_j$
1,40 - 5,90	3,65	9	32,85	119,90
5,90 - 10,40	8,15	10	81,50	664,23
10,40 - 14,90	12,65	16	202,40	2560,36
14,90 - 19,40	17,15	5	85,75	1470,61
19,40 - 23,90	21,65	4	86,60	1874,89
23,90 - 28,40	26,15	4	104,60	2735,29
28,40 - 32,90	30,65	2	61,30	1878,85
		50	655,00	11304,13

$$S^2 = \frac{304,3}{50} - \left[\frac{655,00}{50} \right]^2$$

$$S^2 = 226,08 - (3,10)^2$$

$$S^2 = 226,08 - 7,6$$

$$S^2 = 54,47$$

17.3.1.1 Procedimientos Abreviados de Cálculo: Al igual que la media aritmética, la varianza se puede obtener por métodos abreviados para facilitar su cálculo

Primer Método Abreviado:

Se basa en las desviaciones con respecto a un origen de trabajo; para la deducción de la fórmula se parte de las ecuaciones siguientes:

$$Z_j = Y_j - a \quad (1)$$

$$K = a - ot \quad (2)$$

$$Z_j^0 = Y_j - ot \quad (3)$$

sumando (1) y (2) se tiene que:

$$Z_j + K = Y_j - \bar{X} + \bar{X} - ot = Z'_j$$

$$\Rightarrow Z_j = Z'_j - K$$

$$Z_j^2 = (Z'_j - K)^2 = Z_j'^2 - 2KZ'_j + K^2$$

$$Z_j^2 n_j = Z_j'^2 n_j - 2KZ'_j n_j + K^2 n_j$$

$$S^2 = \frac{\sum_{j=1}^m Z_j^2 n_j}{n} = \frac{\sum_{j=1}^m Z_j'^2 n_j}{n} - 2K \frac{\sum_{j=1}^m Z'_j n_j}{n} + K^2 \frac{\sum_{j=1}^m n_j}{n}$$

Recordando los métodos abreviados de la media aritmética:

$$K = \frac{\sum_{j=1}^m Z'_j n_j}{n}$$

De donde:

$$S^2 = \frac{\sum_{j=1}^m Z_j'^2 n_j}{n} - 2K^2 + K^2$$

$$S^2 = \frac{\sum_{j=1}^m Z_j'^2 n_j}{n} - K^2$$

$$S^2 = \frac{\sum_{j=1}^m Z_j'^2 n_j}{n} - \left[\frac{\sum_{j=1}^m Z_j' n_j}{n} \right]^2$$

$$S^2 = M [Z'^2] - [M [Z']]^2$$

Ejemplo 30:

CLASES	Y_j	n_j	Z'_j	$Z'_j n_j$	$Z_j'^2 n_j$
1,40 - 5,90	3,65	9	- 9,0	- 81,0	729,00
5,90 - 10,40	8,15	10	- 4,5	- 45,0	202,50
10,40 - 14,90	12,65	16	0	0,0	0,00
14,90 - 19,40	17,15	5	4,5	22,5	101,25
19,40 - 23,90	21,65	4	9,0	36,0	324,00
23,90 - 28,40	26,15	4	13,5	54,0	729,00
28,40 - 32,90	30,65	2	18,0	36,0	648,00
				22,50	2.733,75

$$S^2 = \frac{2.733,75}{50} - \left[\frac{22,50}{50} \right]^2$$

$$S^2 = 54,6750 - (0,45)^2$$

$$S^2 = 54,6750 - 0,2025$$

$$S^2 = 54,4725$$

Segundo Método Abreviado:

Está basado en las desviaciones respecto a un origen de trabajo, medidas en unidades de intervalo; se parte de las ecuaciones siguientes:

$$Z''_j = \frac{Y_j - ot}{c} \quad (1)$$

$$K_1 = \frac{a - ot}{c} \quad (2)$$

$$Z_j = Y_j - a \quad (3)$$

donde (1) y (2) se pueden escribir:

$$C Z''_j = Y_j - ot \quad (4)$$

$$C K_1 = a - ot \quad (5)$$

Sumando (3) y (5):

$$Z_j + CK_1 = Y_j - a + a - ot = C Z''_j$$

$$\Rightarrow Z_j = C Z''_j - CK_1$$

$$Z_j^2 = C^2 (Z''_j - K_1)^2 = C^2 (Z''_j^2 - 2K_1 Z''_j + K_1^2)$$

$$S^2 = \frac{\sum_{j=1}^m Z_j^2 n_j}{n} = C^2 \left[\frac{\sum_{j=1}^m Z''_j^2 n_j}{n} - 2K_1 \frac{\sum_{j=1}^m Z''_j n_j}{n} + K_1^2 \frac{\sum_{j=1}^m n_j}{n} \right]$$

En los métodos abreviados de cálculo de la media aritmética también se definió que:

$$K_1 = \frac{\sum_{j=1}^m Z''_j n_j}{n}$$

De donde:

$$S^2 = C^2 \left[\frac{\sum_{j=1}^m Z''_j^2 n_j}{n} - 2K_1^2 + K_1^2 \right]$$

$$S^2 = C^2 \left[\frac{\sum_{j=1}^m Z''_j^2 n_j}{n} - K_1^2 \right] = C^2 \left[\frac{\sum_{j=1}^m Z''_j^2 n_j}{n} - \left[\frac{\sum_{j=1}^m Z''_j n_j}{n} \right]^2 \right]$$

$$S^2 = C^2 \left[M [Z''^2] - [M [Z'']]^2 \right]$$

Ejemplo 31:

CLASES	Y_j	n_j	Z''_j	$Z''_j n_j$	$Z''_j^2 n_j$
1,40 - 5,90	3,65	9	- 2	- 18	36
5,90 - 10,40	8,15	10	- 1	- 10	10
10,40 - 14,90	12,65	16	0	0	0
14,90 - 19,40	17,15	5	1	5	5
19,40 - 23,90	21,65	4	2	8	16
23,90 - 28,40	26,15	4	3	12	36
28,40 - 32,90	30,65	2	4	8	32
		50		5	135

$$S^2 = (4,50)^2 \left[\frac{135}{50} - \left[\frac{5}{50} \right]^2 \right]$$

$$S^2 = (4,50)^2 \left[2,70 - (0,10)^2 \right]$$

$$S^2 = 20,25 \left[2,70 - 0,01 \right]$$

$$S^2 = 20,25 \times 2,69$$

$$S^2 = 54,4725$$

7.1.1.2 Propiedades de la Varianza:

A) La varianza es siempre una cantidad no negativa:

$$S^2 \geq 0$$

B) La varianza de una constante es cero:

$$S^2 = V[K] = M[(K - M(K))^2]$$

$$= M[(K - K)^2]$$

$$= M[0]$$

$$V[K] = 0$$

Es decir que para una serie de valores iguales, no existe desviación alguna.

Ejemplo 32: El precio de ocho acciones de la Bolsa en un determinado día fue \$ 11,78, con lo cual se tendría:

$$X_1 = 11,78$$

$$X_5 = 11,78$$

$$X_2 = 11,78$$

$$X_6 = 11,78$$

$$X_3 = 11,78$$

$$X_7 = 11,78$$

$$X_4 = 11,78$$

$$X_8 = 11,78$$

Según el ejemplo desarrollado para las propiedades de la media se tiene que $\bar{X} = 11,78$

Luego desarrollado:

$$S^2 = M [(X_i - \bar{X})^2]$$

$$X_1 - \bar{X} = 11,78 - 11,78 = 0$$

$$X_2 - \bar{X} = 11,78 - 11,78 = 0$$

$$X_3 - \bar{X} = 11,78 - 11,78 = 0$$

$$X_4 - \bar{X} = 11,78 - 11,78 = 0$$

$$X_5 - \bar{X} = 11,78 - 11,78 = 0$$

$$X_6 - \bar{X} = 11,78 - 11,78 = 0$$

$$X_7 - \bar{X} = 11,78 - 11,78 = 0$$

$$X_8 - \bar{X} = 11,78 - 11,78 = 0$$

$$S^2 = \frac{\sum_{j=1}^7 (X_i - \bar{X})^2}{50} = \frac{0}{50} = 0$$

- C) La varianza del producto de una constante por una variable, es igual al producto del cuadrado de la constante por la varianza de la variable.

$$\begin{aligned}
 V[KY] &= M \left[(KY - M[KY])^2 \right] \\
 &= M \left[(KY - KM[Y])^2 \right] \\
 &= M \left[K^2 (Y - M[Y])^2 \right] \\
 &= K^2 \left[M (Y - M[Y])^2 \right] \\
 &= K^2 V[Y]
 \end{aligned}$$

Ejemplo 33: Se tiene en cuenta el ejemplo de propiedades de la media aritmética:

$2(Y'_{j-} - Y'_j)$	$2Y_j$	n_j	$2Y_j - 26,20$	$(2Y_j - 26,20)n_j$	$(2Y_j - 26,20)^2 n_j$
2,80 - 11,80	7,30	9	- 18,90	- 170,10	3214,89
11,80 - 20,80	16,30	10	- 9,90	- 99,00	980,10
20,80 - 29,80	25,30	16	- 0,90	- 14,40	12,96
29,80 - 38,80	34,30	5	8,10	40,50	328,05
38,80 - 47,80	43,30	4	17,10	68,40	1169,64
47,80 - 56,80	52,30	4	26,10	104,40	2724,84
56,80 - 64,80	61,30	2	35,10	70,20	2464,02
50			10.894,50		

$$V[2Y] = \frac{10.894,50}{50} = 217,89$$

$$V [2Y] = (2)^2 V [Y]$$

$$217,89 = 4 \times 54,47$$

$$217,89 \approx 217,88$$

La diferencia se debe a la aproximación anterior, que se hizo de la varianza.

- D) La varianza de una constante más una variable, es igual a la varianza de la variable, dado que las diferencias se mantienen pero a diferente nivel.

$$\begin{aligned} V [K+Y] &= M \left\{ [(K+Y) - M [K+Y]]^2 \right\} \\ &= M \left\{ [K+Y - M [K] - M [Y]]^2 \right\} \\ &= M \left\{ [K+Y - K - M [Y]]^2 \right\} \\ &= M \left\{ [Y - M [Y]]^2 \right\} \\ &= V [Y] \end{aligned}$$

Ejemplo 34: Supóngase que a causa de una medida del gobierno, todas las acciones subieran en 5 centavos; luego:

CLASES	$Y_j + 0,05 = Y_j^*$	n_j	$Y_j^* n_j$	$Y_j^{*2} n_j$
1,40 - 5,90	3,65 + 0,05 = 3,70	9	33,30	123,21
5,90 - 10,40	8,15 + 0,05 = 8,20	10	82,00	672,40
10,40 - 14,90	12,65 + 0,05 = 12,70	16	203,20	2580,64
14,90 - 19,40	17,15 + 0,05 = 17,20	5	86,00	1479,20
19,40 - 23,90	21,65 + 0,05 = 21,70	4	86,80	1883,56
23,90 - 28,40	26,15 + 0,05 = 26,20	4	104,80	2745,76
28,40 - 32,90	30,65 + 0,05 = 30,70	2	61,40	1884,98
		50	657,50	11369,75

$$S^2 = \frac{\sum_{j=1}^m Y_j^{*2} n_j}{n} - \left[\frac{\sum_{j=1}^m Y_j^* n_j}{n} \right]^2$$

$$S^2 = \frac{11.369,75}{50} - \left[\frac{657,50}{50} \right]^2$$

$$S^2 = 227,395 - (13,15)^2$$

$$S^2 = 227,395 - 172,9225$$

$$S^2 = 54,4725$$

$$S^2 = 54,47$$

Que es el valor calculado anteriormente para la varianza.

E) Para el tratamiento de submuestras, difiere el cálculo de la varianza, del que se aplicó para obtener la media total, como ponderación de las medias de las submuestras; se puede probar que: para dos sub-muestras de tamaño n_1 y n_2 , medias $\bar{Y}_1$ y $\bar{Y}_2$ y varianzas S_1^2 y S_2^2 respectivamente la varianza de la muestra total está dada por :

$$S^2 = \frac{S_1^2 n_1 + S_2^2 n_2}{n_1 + n_2} + \frac{(\bar{Y}_1 - \bar{Y})^2 n_1 + (\bar{Y}_2 - \bar{Y})^2 n_2}{n_1 + n_2}$$

El primer término de la fórmula, es la media aritmética ponderada de las varianzas de las dos submuestras, que se denomina Intravarianza (S_w^2); el segundo término es la varianza de las medias de las submuestras que se denomina Inter-varianza (S_b^2).

En general se puede escribir para (t) submuestras:

$$S^2 = \frac{\sum_{r=1}^t S_r^2 n_r}{\sum_{r=1}^t n_r} + \frac{\sum_{r=1}^t (\bar{Y}_r - \bar{Y})^2 n_r}{\sum_{r=1}^t n_r}$$

$$S^2 = S_w^2 + S_b^2$$

Ejemplo 35: Se considera la misma subdivisión en submuestras que se hizo para la 3a. propiedad de la media aritmética.

Primera Sub-muestra.

$Y'_{1j-1} - Y'_{1j}$	Y_{1j}	n_{1j}	$Y_{1j} - 9,05$	$(Y_{1j} - 9,05)n_j$	$(Y_{1j} - 9,05)^2 n_j$
1,40 - 5,90	3,65	9	- 5,40	- 48,60	262,44
5,90 - 10,40	8,15	10	- 0,90	- 9,00	8,10
10,40 - 14,90	12,65	16	3,60	57,60	207,36
			35		477,90

$$S_1^2 = \frac{477,90}{35} = 13,6543$$

Segunda Sub-muestra.

$Y'_{2j-1} - Y'_{2j}$	Y_{2j}	n_{2j}	$Y_{2j} - 22,55$	$(Y_{2j} - 22,55)n_j$	$(Y_{2j} - 22,55)^2 n_j$
14,90 - 19,40	17,15	5	- 5,40	- 27,00	145,80
19,40 - 23,90	21,65	4	- 0,90	- 3,60	3,24
23,90 - 28,40	26,15	4	3,60	14,40	51,84
28,40 - 32,90	30,65	2	8,10	16,20	131,22
			15		332,10

$$S_2^2 = \frac{332,10}{15} = 22,14$$

./.

La varianza total vendría dada por:

$$S^2 = \frac{S_1^2 n_1 + S_2^2 n_2}{n_1 + n_2} + \frac{(\bar{Y}_1 - \bar{Y})^2 n_1 + (\bar{Y}_2 - \bar{Y})^2 n_2}{n_1 + n_2}$$

$$S^2 = S_w^2 + S_b^2$$

$$S_w^2 = \frac{13,6543 \times 35 + 22,14 \times 15}{35 + 15}$$

$$S_w^2 = \frac{477,9005 + 332,10}{50} = \frac{810,0005}{50}$$

$$S_w^2 = 16,200010$$

$$S_b^2 = \frac{(9,05 - 13,10)^2 \times 35 + (22,55 - 13,10)^2 \times 15}{35 + 15}$$

$$S_b^2 = \frac{(-4,05)^2 \times 35 + (9,45)^2 \times 15}{50} = \frac{16,4025 \times 35 + 89,3025 \times 15}{50}$$

$$S_b^2 = \frac{574,0875 + 1339,5375}{50} = \frac{1.913,6250}{50}$$

$$S_b^2 = 38,2725$$

$$S^2 = 16,2000 + 38,2725$$

$$S^2 = 54,4725$$

7.1.2 Desviación Típica (S, σ)

La desviación típica o la desviación estandar, se define como la raíz cuadrada positiva de la varianza:

$$S = \sqrt{V[Y]} = \sqrt{\frac{\sum_{j=1}^m (Y_j - \bar{Y})^2 n_j}{n}}$$

Conviene observar que la varianza expresa la dispersión de la variable en unidades distintas a las que corresponden a la variable originalmente; si la variable está expresada en cms., la varianza quedará expresada en cms², si la variable está en kilogramos la varianza quedará en kilogramos cuadrados; si la variable está dada en pesos (\$) la varianza da en pesos cuadrados; por esta razón la varianza no tiene un significado real en el tratamiento mismo del fenómeno, esta dificultad es superada al tomar su raíz cuadrada (desviación estandar) de esta manera la dispersión queda expresada en las mismas uni

dades en que ha sido dada inicialmente la variable.

Ejemplo 36:

$$S = \sqrt{S^2} = \sqrt{54,4725} = 7,3806$$

$S \approx 7,38$ pesos.

La interpretación de la desviación estandar es muy similar a la que se hace de la varianza, con la ventaja de que esta desviación está medida en las unidades de variable; en la práctica se acostumbra a groso modo y como prueba inicial compararla con la media aritmética mediante la relación:

$$\bar{Y} \approx 3 S$$

Si $3 S > \bar{Y}$ se considera que la distribución en este caso es cada vez más heterogénea y si $3 S < \bar{Y}$ la distribución es cada vez más homogénea. Lo anterior se cumple siempre y cuando el rango de la variable empiece en cero o en un valor próximo a él.

Es de gran utilidad para comparar el grado de dispersión en tre distribuciones análogas.

7.1.3 Desviación Media. Es la media aritmética de las desviaciones respecto a la media aritmética ($\bar{Y}$) tomadas en valor absoluto.

$$D M = M [| Z |] = \frac{\sum_{j=1}^m | Y_j - \bar{Y} | n_j}{n} = \frac{\sum_{j=1}^m | Z_j | n_j}{n}$$

No es susceptible de tratamiento algebraico, sin embargo, su fácil cálculo la hace muy utilizada como una aproximación a la cuantificación de la dispersión de la variable sobre todo en aquellos casos en que el tiempo apremia; su valor se aproxima al de la desviación estandar para simplificar el trabajo y generalmente se toma con el signo $\pm$ para indicar su oscilación alrededor de la media.

Ejemplo 37:

C L A S E S	Y_j	n_j	$ Z_j $	$ Z_j n_j$
1,40 - 5,90	3,65	9	9,45	85,05
5,90 - 10,40	8,15	10	4,95	49,50
10,40 - 14,90	12,65	16	0,45	7,20
14,90 - 19,40	17,15	5	4,05	20,25
19,40 - 23,90	21,65	4	8,55	34,20
23,90 - 28,40	26,15	4	13,05	52,20
28,40 - 32,90	30,65	2	17,55	35,10

50

283,50

$$D M = \frac{\sum_{j=1}^m |Z_j| n_j}{n}$$

$$D M = \frac{283,50}{50}$$

$$D M = 5,67 \text{ pesos}$$

Como se puede ver el valor de la desviación media es siempre menor que el de la desviación estandar. Aunque ésta es una forma de cuantificar la dispersión más rápida y fácil, también es menos exacta.

7.1.4 Se define la desviación mediana como la media aritmética de los valores de las desviaciones tomados respecto a la mediana, expresadas en valor absoluto, igualmente su resultado se toma $\pm$ como oscilación alrededor de la mediana:

$$DMe = M \left[|Y_j - Me| \right] = \frac{\sum_{j=1}^m |Y_j - Me| N_j}{n}$$

Ejemplo 38:

CLASES	Y_j	n_j	$ Y_j - Me $	$ Y_j - Me n_j$
1,40 - 5,90	3,65	9	8,44	75,96
5,90 - 10,40	8,15	10	3,94	39,40
10,40 - 14,90	12,65	16	0,56	8,96
14,90 - 19,40	17,15	5	5,06	25,30
19,40 - 23,90	21,65	4	9,56	38,24
23,90 - 28,40	26,15	4	14,06	56,24
28,40 - 32,90	30,65	2	18,56	37,12

50

281,20

$$DMe = \frac{\sum_{j=1}^m |Y_j - Me| n_j}{n} ; Me = 12,09$$

$$DMe = \frac{95,21}{50} = 5,6042 \approx 5,62$$

$$DMe = 5,62 \text{ pesos.}$$

7.1.5 Recorrido: Inicialmente definido cuando se estudió la construcción de la distribución de frecuencias como la diferencia entre la mayor y la menor de las observaciones, al tener ya la distribución de frecuencias el recorrido está dado por la diferencia entre el límite superior del último intervalo de clase y el límite inferior del primero.

$$R = Y'_m - Y'_o$$

El recorrido da una idea muy relativa y global sobre la dispersión de la serie, pues no dice nada sobre la conformación misma (interna) de ésta; suele usarse frecuentemente en control de calidad en la medición de características cuyos resultados son más bien homogéneos dentro de las muestras que se toman para controlar especificaciones determinadas.

Ejemplo 39: Observando el ejemplo 38, se tiene que:

$$R = 32,90 - 1,40 = 31,50 \text{ pesos.}$$

Se tiene así que, dentro de un intervalo de 31,50 pesos es tán todos los valores de las acciones de las empresas o - también, que la diferencia entre el máximo valor que pueden tomar las acciones de una empresa y el menor valor de éstas es 31,50 pesos.

7.1.6 Recorrido Intercuartílico: Se usa para eliminar la influen cia de los valores extremos, está dado por la diferencia en tre el tercer y el primer cuartil.

$$RQ = Q_3 - Q_1$$

El 50% de las observaciones tienen sus valores dentro de un intervalo de tamaño RQ.

Ejemplo 40: En relación al ejemplo 25, se tiene:

$$RQ = 17,15 - 7,48$$

$$RQ = \$9,67$$

Podemos decir entonces que el 50% de los valores de las ac ciones de las empresas están incluidos en el intervalo 7,48 a 17,15.

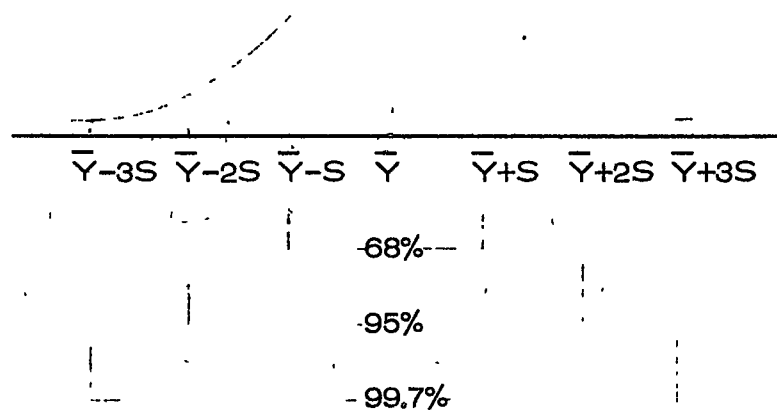
7.2 Medidas de Dispersión Relativa:

7.2.1 Coeficiente de Variación: Se define como el cociente entre la desviación estandar y la media aritmética, dado en porcentaje:

$$C.V. = \frac{S}{\bar{Y}} (100)$$

Es de gran utilidad para la comparación de distribuciones que están medidas en unidades diferentes, por ejemplo una desviación estandar para una distribución de \$ 500 por hogar y una desviación estandar de 0,62 hijos por hogar, son valores que por si mismos no permiten hacer ningún tipo de comparación, ni saber cual es más o menos homogéneo con relación al otro; en cambio con el coeficiente de variación que es un número abstracto es posible hacerlo. Observando la forma como se relaciona la desviación estándar con la media aritmética tenemos que para una distribución más normal (1) se dan los siguientes intervalos:

(1) Siempre y cuando que el dominio de la variable inicie en cero o en un valor cercano a él.



La distribución normal es una distribución monótonamente cre-
ciente hasta un punto máximo y luego monótonamente decre-
ciente en la misma magnitud en que aumenta inicialmente. Es
simétrica y unimodal por lo cual: $M_d = M_e = M$

Dentro del dominio de la variable la media aritmética en gene-
ral debe contener aproximadamente tres veces a la desviación
estandar, luego se puede deducir que la desviación estandar en
una distribución normal debe aproximarse a la tercera parte -
de la media aritmética, de donde se obtiene que el punto de com-
paración para los coeficientes de variación está dado por el
33% que es el C V, de una curva normal.

$$C.V. = \frac{1}{3} = 33\%$$

Se dice que una distribución que tenga un C.V. menor que 33% tiende a ser cada vez más homogénea y si mayor del 33% a ser cada vez más heterogénea, o que a medida que el coeficiente de variación se aproxima a cero la distribución es más homogénea y en caso contrario heterogéneo.

Ejemplo 4: En relación al ejemplo 34. $S = 7,3806$

$$C.V. = \frac{S}{\bar{Y}} \times 100$$

$$C.V. = \frac{7,3806}{13,10} \times 100$$

$$C.V. = 56,34\%$$

Lo cual indica que los datos de la distribución están bastante dispersos y los parámetros como la media, varianza, etc. que tienen en cuenta a todos los datos se ven notablemente afectados por los valores extremos de la distribución. Por consiguiente hay que ser cautelosos en la utilización del promedio y demás medidas que sean muy afectadas por los valores extremos.

Es conveniente si las condiciones lo permiten, hacer estudios por sub-grupos, de tal manera que a cada uno le corresponda una media y un coeficiente de variación, que serán más representativos a dicho nivel.

8. COEFICIENTE DE DEFORMACION (ASIMETRIA)

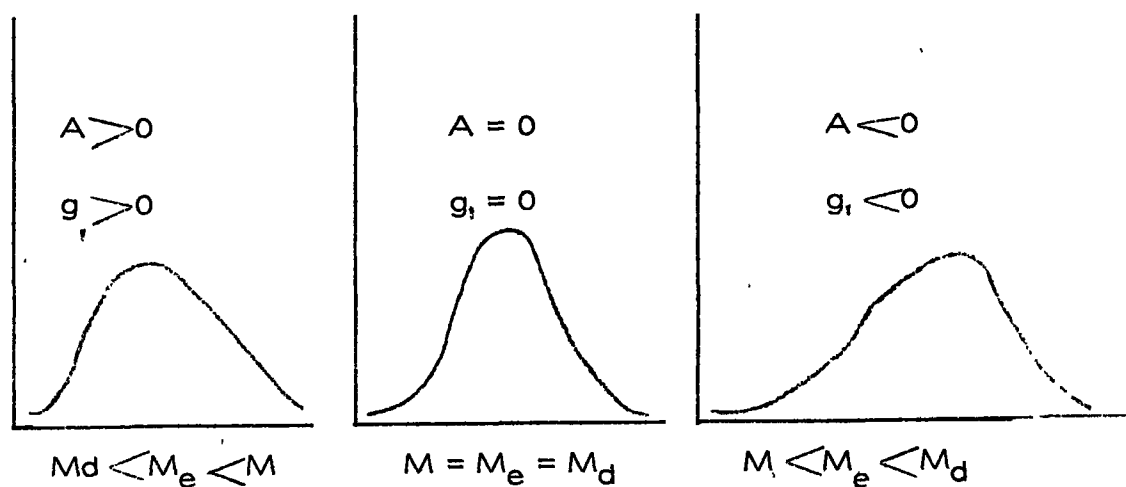
Se trata de explicar ahora el grado de deformación de la distribución para lo cual se acostumbra establecer una comparación con una distribución teórica normal, se trata ante todo de cuantificar el grado de deformación de la distribución para lo cual Karl Pearson y A. Bawley empíricamente han deducido y llegado a proponer la forma de cuantificación. El carácter empírico de la deducción de estos parámetros hace que no sean siempre justificables, las dos fórmulas usualmente utilizadas son:

$$A = \frac{\bar{X} - Md}{S}$$

$$A_1 = \frac{3(\bar{X} - Me)}{S}$$

La comparación de estas estadígrafas se hace en base a que en una distribución simétrica:

$$M = M_e = M_d$$



La asimetría se representa por la mayor longitud de la falda o cola que presente el diagrama o histograma de la distribución y se hablará de deformación hacia la izquierda o hacia la derecha según el caso, por lo tanto se tiene que para valores de $A < 0$ la distribución estaría recargada hacia la derecha y para valores de $A > 0$ la distribución estaría recargada hacia la izquierda, sin embargo, se debe tener cuidado al

realizar el análisis valiéndonos de estos estadígrafos, pues el mismo carácter de su procedencia ocasiona que en algunos casos en que el valor de la asimetría es bajo se observe en la gráfica que existe realmente simetría o que el signo es el contrario al que da en la cuantificación, pero para una asimetría considerable estos estadígrafos no tienen ningún problema en su aplicación.

Pearson plantea su cálculo en base a las desviaciones con respecto a la media aritmética (m_r) mediante la relación:

$$b_1 = \frac{m_3^2}{m_2^3}$$

el coeficiente b_1 nos indicará en caso de ser diferente de cero la existencia de la deformación pero no el sentido de ella:

$$g_1 = \sqrt{b_1} = \frac{m_3}{s^3}$$

el coeficiente g_1 para una distribución simétrica es igual a cero.

Ejemplo 42:

$$A' = \frac{M - Md}{S}$$

$$A = \frac{13,10 - 11,90}{7,3806} = \frac{1,2}{7,3806}$$

$$A = 0,1626$$

$$A_1 = \frac{3 (M - Me)}{S}$$

$$A_1 = \frac{3 (13,10 - 12,09)}{7,3806} = \frac{3 (1,01)}{7,3806}$$

$$A_1 = \frac{3,03}{7,3806}$$

$$A_1 = 0,4105$$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(274,4685)^2}{(54,4725)^3}$$

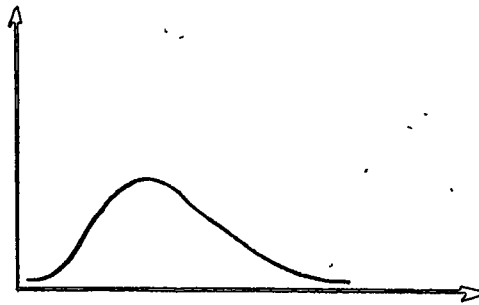
$$b_1 = \frac{75\,332,9575}{161,633,7030}$$

$$b_1 = 0,4661$$

$$g_1 = \sqrt{b_1}$$

$$g_1 = \sqrt{0,4661} = 0,6827$$

Los anteriores son números abstractos que no se deben expresar en las unidades de la variable:



$$A_1 = 0,4105 > 0$$

$$g_1 = 0,6827 > 0$$

Los resultados nos indican que la distribución está inclinada hacia la izquierda, es decir que los valores están concentrados sobre los valores menores de las acciones.

Observando los resultados de las fórmulas dadas para el cálculo de la deformación, se debe anotar que en la práctica se utiliza A en el caso de que la distribución esté recargada a la iz

quierda y A_1 cuando está recargada a la derecha, por lo cual para nosotros es más válido el valor A . El valor b_1 confirmará el hecho de que la distribución está deformada y el g_1 de que su deformación es hacia la derecha, o hacia la izquierda.

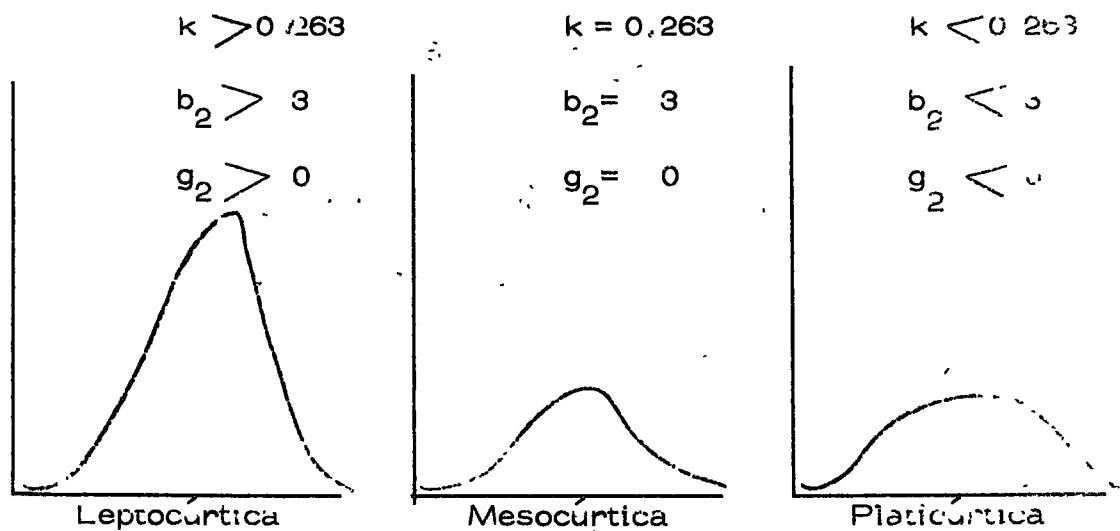
9. CURTOSIS O ESTADIGRAFOS DE APUNTAMIENTO

Mediante los cuales se trata de cuantificar el grado homogeneidad o heterogeneidad de los datos respecto de sus valores centrales, en otras palabras se trata de analizar la concentración comparando la distribución en la misma forma que en el caso anterior con una distribución teórica normal.

Los estadígrafos más utilizados en este caso son:

$$C = \frac{1/2 (Q_3 - Q_1)}{P_{90/100} - P_{10/100}}$$

Para la distribución teórica normal (simétrica) se considera que $C = 0.263$ valor contra el cual se van a hacer las comparaciones:



Si $C = 0.263$ la distribución no es apuntada ni aplanada, es normal (simétrica) y se le denomina mesocúrtica, leptocúrtica en caso de ser apuntada la distribución (concentrada) $k > 0.263$, platicúrtica en caso de ser aplanada la distribución (existe considerable dispersión) $k < 0.263$.

También se puede determinar mediante los coeficientes b_2 de Pearson y g_2 de Fisher:

$$b_2 = \frac{m_4}{m_2^2}$$

para $b_2 = 3$ la distribución es mesocúrtica (normal); menor que 3 es platicúrtica y mayor de 3 la curva es leptocúrtica:

$$g_2 = b_2 - 3$$

para $g_2 = 0$ la curva es mesocúrtica (normal), g_2 menor que cero, es platocúrtica y g_2 mayor que cero es leptocúrtica.

Ejemplo 43:

$$C = \frac{1/2 (17,15 - 7,48)}{25,02 - 3,90}$$

$$C = \frac{1/2 (9,67)}{21,12}$$

$$C = \frac{9,67}{42,24}$$

$$C = 0,2289$$

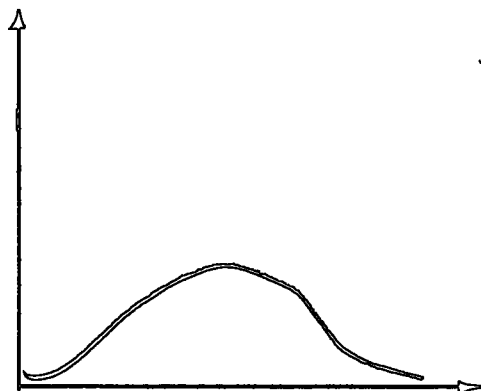
$$b_2 = \frac{m_4}{m_2^2} = \frac{8124,8554}{(54,4725)^2}$$

$$b_2 = \frac{8224,8554}{2967,2533} = 2,7382$$

$$g_2 = b_2 - 3$$

$$g_2 = 2,7382 - 3$$

$$g_2 = -0,2618$$



$$C = 0,2289 < 0,263$$

$$b_2 = 2,7382 < 3$$

$$g_2 = -0,2618 < 0$$

Para la distribución objeto de nuestro estudio se tiene que de acuerdo a los diferentes parámetros utilizados para cuantificar el apuntamiento (curtosis) hemos obtenido que nuestra distribución es más bien aplanada, aunque no de una manera notable; de acuerdo a lo anterior se puede también observar que la distribución debe tender a ser heterogénea, lo cual ya ha sido comprobado con las otras medidas de dispersión.

CAPITULO V

PROBABILIDADES

Es necesario como introducción a este capítulo tener en cuenta una serie de definiciones :

Experimento aleatorio: es aquel que realizado bajo las mismas condiciones no se pueden predecir los mismos resultados.

Espacio Muestral: es el conjunto de todos los puntos muestrales de un experimento aleatorio.

Evento: subconjunto del espacio muestral (suceso).

Punto Muestral: un elemento del espacio muestral.

La definición más elemental de probabilidad y que se conoce como definición clásica de probabilidad es:

"De un espacio muestral, compuesto por una serie de sucesos igualmente posibles; la probabilidad de que se presente un suceso A determinado, el cual a su vez puede presentarse en n' de los n sucesos posibles", está dada por la relación :

$$P(A) = \frac{n'}{n} = \frac{\text{No. de sucesos favorables}}{\text{No. de sucesos posibles}}$$

Sin embargo se considera que esta definición es de carácter regular por basarse la probabilidad en la igual posibilidad de todos los sucesos.

Se tiene también una definición de probabilidad empírica como la frecuencia relativa de aparición de determinado suceso A'' , la probabilidad se define entonces como el límite de la frecuencia relativa cuando el número de observaciones tiende a infinito.

Sin embargo es conveniente anotar que las anteriores definiciones tienen problemas en su utilización y se prefiere más bien dar una definición de tipo axiomático.

$$P(A) = \frac{n(A)}{n(S)}$$

$n(A)$: número de elementos del conjunto A .

$n(S)$: número de elementos del espacio muestral

En algunas ocasiones es relativamente fácil determinar cuál es el número de casos favorables y cuál es el número de casos posibles,

sin embargo en la mayor parte de las ocasiones esto no es así, tendremos que recurrir al análisis combinatorio que es la rama de las matemáticas que nos ayuda a determinar el número de resultados posibles de un experimento sin necesidad de enumerar cada caso particular.

Principio Fundamental: si un evento A puede ocurrir de n_1 maneras y otro evento B puede ocurrir de n_2 maneras, el número de maneras en que los eventos pueden realizarse es $n_1 \times n_2$.

Ej: la boleta para una rifa se compone de una letra y un número de un dígito, con cuantas boletas se hace la rifa?

$$26 \times 10 = 260$$

Siendo n_1 = número de letras del abecedario

n_2 = número de dígitos

Este principio se generaliza para cualquier número de eventos.

Variaciones: se llaman variaciones de clase k de los n objetos disponibles, a los distintos grupos ordenados que se puedan formar tomando k objetos de entre los n disponibles; cada uno de los grupos de las variaciones se diferencian entre sí por la naturaleza de alguno de sus elementos, o si ellos son iguales se distinguen por el orden en que se encuentran.

n^V_k = Variaciones de n cosas tomadas de k en k

$$n^V_k = n(n-1)(n-2)\dots(n-k+1)$$

lo que también se puede escribir en forma factorial como:

$$n^V_k = \frac{n!}{(n-k)!}$$

Ej : la preselección para un equipo de ajedrez consta de 10 jugadores, se va a conformar un equipo de 4 para un torneo, cuántos equipos diferentes pueden formarse si los jugadores pueden ocupar cualquiera de los 4 tableros ?

$$10^V_4 = \frac{10!}{(10-4)!} = \frac{10!}{6!} = 5,040$$

Se tendría así 5,040 formas diferentes de conformar el equipo.

Permutaciones: son las variaciones de clase n de los n objetos disponibles.

P_n permutaciones de n objetos

$$P_n = n!$$

Ej : Cuántas señales pueden transmitirse con 5 banderas de diferente color, si se toman simultáneamente y se colocan en línea recta.

$$P_5 = 5! = 120$$

En caso de existir objetos repetidos

$n \text{ } ^P_{n_1, n_2, \dots, n_r}$: permutaciones de n objetos de los cuales hay n_1 objetos iguales; n_2 objetos iguales; ...; n_r objetos iguales.

Existen n_r clases de objetos y además

$$n_1 + n_2 + \dots + n_r = n$$

$$n \text{ } ^P_{n_1, n_2, \dots, n_r} = \frac{n!}{n_1! n_2! \dots n_r!}$$

Ej: Cuántas señales se podrían hacer si se tuvieran 3 banderas rojas, 2 azules y 4 blancas; si se tomar todas simultáneamente.

$$9 \text{ } ^P_{3, 2, 4} = \frac{9!}{3! 2! 4!} = 1.260 \text{ señales}$$

Combinaciones: se llaman combinaciones de clase k de los n objetos diferentes, a los distintos grupos que se puedan formar tomando

k objetos de entre los n disponibles; las combinaciones se diferencian entre sí por la naturaleza de alguno de sus elementos (en este caso ya no interesa el orden);

$${}^nC_k = {}^nC_k = \text{combinaciones de } n \text{ cosas tomadas de } k \text{ en } k$$

$${}^nC_k = {}^nC_k = \frac{n!}{(n-k)! k!}$$

Ej: un estudiante tiene que resolver 10 preguntas de 13 en un examen. Cuántas maneras de escoger tiene?

$$\binom{13}{0} = \frac{13!}{(13-0)! 0!} = \frac{13 \times 12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \times 11 \times 12 \times 1} = 286$$

- Cuántos si las dos primeras son obligatorias?

$$\binom{2}{2} \binom{11}{8} = 1 \times \frac{11!}{(11-8)! 8!} = \frac{11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8} = 165$$

- Cuántas si una de las dos primeras es obligatoria?

$$\binom{2}{1} \binom{11}{9} = \frac{2!}{(2-1)! 1!} \times \frac{11!}{(11-9)! 9!} = 2 \times \frac{11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9} = 220$$

$$= 2 \frac{11 \times 10}{1 \times 2} = 110$$

- Cuántas sí tiene que contestar exactamente 3 de las 5 primeras ?

$$\binom{5}{3} \binom{8}{7} = \frac{5!}{(5-3)! 3!} \times \frac{8!}{(8-7)!}$$

$$= \frac{5 \times 4}{1 \times 2} \times \frac{8}{1} = 80$$

- Cuántas sí tiene que contestar por lo menos 3 de las 5 primeras ?

$$\binom{5}{3} \binom{8}{7} + \binom{5}{4} \binom{8}{6} + \binom{5}{5} \binom{8}{5} = \frac{5!}{(5-3)! 3!} \times \frac{8!}{(8-7)! 7!} +$$

$$\frac{5!}{(5-4)! 4!} \times \frac{8!}{(8-4)! 6!} + \frac{5!}{(5-5)! 5!} \times \frac{8!}{(8-5)! 5!}$$

$$= \frac{5 \times 4}{1 \times 2} \times \frac{8}{1} + \frac{5}{1} \times \frac{8 \times 7}{1 \times 2} + \frac{1}{1} \times \frac{8 \times 7 \times 6}{1 \times 2 \times 3}$$

$$= 80 + 140 + 56 = 276$$

NOTA: $\binom{n}{k} = n^C k$

Teorema aditivo de Probabilidad: Se tienen dos eventos que se interceptan. El número de eventos simples en la unión de dos eventos A y B estaría dado por:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

De acuerdo a la definición de probabilidad

$$P(A \cup B) = \frac{n(A \cup B)}{n}$$

$$P(A \cup B) = \frac{n(A)}{n} + \frac{n(B)}{n} - \frac{n(A \cap B)}{n}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Ej: si se arroja un dado cuál es la probabilidad de que aparezca un número par o un número divisible entre 3

$$A = \{2, 4, 6\} \quad n(A) = 3$$

$$B = \{3, 6\} \quad n(B) = 2$$

$$A \cap B = \{6\} \quad n(A \cap B) = 1$$

$$P(A \cup B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{2}{3}$$

En caso de que A y B sean mutuamente excluyentes

$$A \cap B = \emptyset$$

$$P(A \cap B) = P(\emptyset) = 0$$

de donde se obtendría en este caso que

$$P(A \cup B) = P(A) + P(B)$$

Ej: se extrae una carta aleatoriamente de una baraja de 52 cartas.

Cuál es la probabilidad de obtener un as o un rey?

$$P(A \cup R) = \frac{4}{52} + \frac{4}{52} = \frac{2}{13}$$

- Un corazón o un trébol

$$P(C \cup T) = \frac{13}{52} + \frac{13}{52} = \frac{1}{2}$$

- De obtener un as o un corazón (en este caso si tenemos intersección).

$$P(A \cup C) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{4}{13}$$

Para generalizar podemos escribir para eventos disyuntos que

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

$$= \sum_{i=1}^n P(A_i)$$

Probabilidad Condicional: Sea S un espacio muestral y B un evento de S , es posible definir otro espacio muestral como los eventos de B no teniendo en cuenta a los demás, con lo que se obtiene otro espacio de eventos simples B con $n(B)$ número de eventos posibles. Si A es un evento cualquiera del espacio S , en el nuevo espacio B solo aparecerán $n(A \cap B)$ eventos simples favorables de la intersección de A y B ; de esta manera la probabilidad quedará definida:

$P(A|B)$: probabilidad de que se presente A dado que B ya sucedió.

$$P(A|B) = \frac{N(A \cap B)}{N(B)} = \frac{P(A \cap B)}{P(B)}$$

Ej : se escogen al azar dos dígitos diferentes entre los dígitos 1 a 9.

Si la suma es impar, ¿cuál es la probabilidad de que 2 sea uno de los números escogidos?

1 : la suma es impar $n(1) = 5 \times 4 = 20$

2 : dos es el número escogido $n(2) = 1 \times 5 = 5$

$$P(2|1) = \frac{5}{20} = \frac{1}{4}$$

También podemos expresar la anterior relación como:

$$P(A \cap B) = P(B) \cdot P(A|B)$$

y esta fórmula final es la que se conoce con el nombre de teorema de la multiplicación.

Además tenemos que:

$$A \cap B = B \cap A$$

luego

$$P(A \cap B) = P(B \cap A) = P(A) P(B|A)$$

generalizando podemos escribir que para n eventos

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 A_2) \dots P(A_n|A_1 \dots A_{n-1})$$

Independencia: Dos eventos se dicen independientes si uno no influye en la aparición del otro o viceversa, por lo tanto A y B son dos eventos independientes si la probabilidad de que A y B se presenten simultáneamente o en secuencia es igual al producto de las probabilidades de cada uno de ellos.

$$P(A \cap B) = P(A) P(B)$$

y relacionándolo con lo expuesto anteriormente tendremos que:

$$P(A \cap B) = P(B) P(A | B)$$

reemplazando

$$P(A) \cancel{P(B)} = P(A \cap B) \cancel{P(B)}$$

$$P(A) = P(A | B)$$

para eventos independientes, del mismo modo

$$P(B) = P(B | A)$$

Ej: un dado se lanza dos veces sea

A : impar el primer lanzamiento

B : par el segundo lanzamiento

comprobar que los dos eventos son independientes

$$n(A) = 3 \times 6 = 18$$

$$n(B) = 6 \times 3 = 18$$

$$n(A \cap B) = 3 \times 3 = 9$$

$$P(A) = \frac{18}{36} = \frac{1}{2}$$

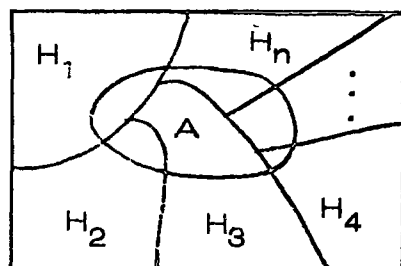
$$P(B) = \frac{18}{36} = \frac{1}{2}$$

$$P(A|B) = \frac{9}{18} = \frac{1}{2}$$

$$P(A \cap B) = \frac{9}{36} = \frac{1}{4}$$

$$P(A) P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

Teorema de Bayes:



Sea A un evento que ocurre cuando alguno de los eventos H_i ($i = 1, 2, \dots, n$) ha ocurrido.

Además se tiene que $H_1, H_2, \dots, H_n$ forman una partición del espacio muestral

es decir:

$$H_i \cap H_j = \emptyset \text{ para } i \neq j$$

Se conocen $P(H_i)$ y $P(A|H_i)$ para todo valor de i , lo cual se denomina probabilidad a priori; se desea calcular la probabilidad $P(H_i|A)$ de que ocurriendo A haya sido causado específicamente por alguna de las H_i , que viene a ser la probabilidad a posteriori.

De lo anterior tendríamos que

$$P(H_i|A) = \frac{P(A \cap H_i)}{P(A)}$$

pero además tenemos que:

$$P(A \cap H_i) = P(H_i) P(A|H_i)$$

se tiene también

$$A \subset (H_1 \cup H_2 \cup H_3 \cup \dots \cup H_n)$$

de donde podemos escribir

$$A = A \cap (H_1 \cup H_2 \cup H_3 \cup \dots \cup H_n)$$

y según la propiedad distributiva tenemos

$$A = (A \cap H_1) \cup (A \cap H_2) \cup \dots \cup (A \cap H_n).$$

como los H_i son mutuamente excluyentes

$$P(A) = P(A \cap H_1) + P(A \cap H_2) + \dots + P(A \cap H_n)$$

$$= \sum_{i=1}^n P(A \cap H_i)$$

$$P(A) = \sum_{i=1}^n P(H_i) P(A|H_i)$$

de donde tendremos

$$P(H_i|A) = \frac{P(H_i) P(A|H_i)}{\sum_{i=1}^n P(H_i) P(A|H_i)}$$

Ej: Una urna contiene 9 bolas blancas y 1 negra; y una segunda urna contiene 5 bolas blancas y 20 negras. Una bola extraída de una de las urnas resulta blanca; ¿cuál es la probabilidad de que la bola venga de la primera urna?

A: Sale blanca

H_1 : se elige la 1a. urna

H_2 : se elige la 2a. urna

$$P(H_1) = P(H_2) = \frac{1}{2}$$

$$P(A|H_1) = 9/10$$

$$P(A|H_2) = 5/25$$

$$P(H_1|A) = \frac{P(H_1) P(A|H_1)}{P(H_1) P(A|H_1) + P(H_2) P(A|H_2)}$$

$$P(H_1|A) = \frac{1/2 \times 9/10}{1/2 \times 9/10 + 1/2 \times 5/25}$$

$$= \frac{9/20}{9/20 + 1/10} = \frac{9/20}{11/20}$$

$$P(H_1|A) = \frac{9}{11}$$

A N E X O

Para efectos de ilustrar el desarrollo de un ejercicio sobre variable discreta, se tomaron los datos referentes al número de aves de corral, en cientos, existentes en las 443 veredas del departamento del Meta, según el Censo Nacional Agropecuario de 1971 .

La tabulación y la obtención de los diferentes valores de la tabla que aparece a continuación, siguen la misma metodología que para el ejercicio sobre variable continua.

TABLA DE TRABAJO

	1	2	3	4	5	6	7	8
N Orden	Aves de Corral (en cientos) Clases $Y'_{j-1} - Y'_j$	Y_j	Nº de Veredas n_j	h_j	N_j	H_j	$\log Y_j$	$(\log Y_j) n_j$
1	0 — 3	1,50	125	0,2822	125	0,2822	0,176091	22,011375
2	3 — 6	4,50	130	0,2934	255	0,5756	0,653130	84,906900
3	6 — 9	7,50	71	0,1603	326	0,7359	0,875061	62,129330
4	9 — 12	10,50	43	0,0970	369	0,8329	1,021181	43,910783
5	12 — 15	13,50	17	0,0384	386	0,8713	1,130334	19,215678
6	15 — 18	16,50	21	0,0474	407	0,9187	1,217484	25,567164
7	18 — 21	19,50	10	0,0226	417	0,9413	1,290035	12,900350
8	21 — 24	22,50	6	0,0135	423	0,9548	1,352183	3,112098
9	24 — 27	25,50	3	0,0068	426	0,9616	1,406540	4,219620
10	27 — 30	28,50	3	0,0068	429	0,9634	1,454845	4,364535
11	30 — 33	31,50	3	0,0068	432	0,9752	1,493311	4,494933
12	33 — 36	34,50	1	0,0022	433	0,9774	1,537819	1,537819
13	36 — 39	37,50	10	0,0226	443	1,0000	1,574031	15,740310
		257,50	443	1,0000				309,111395

	9	10	11	12	13	14	15	16
Nº Orden	n_j/Y_j	$Y_j n_j$	$Y_j^2 n_j$	$Y_j^3 n_j$	$Y_j^4 n_j$	Z_j $(Y_j - \bar{Y})$	$Z_j n_j$	$Z_j^2 n_j$
1	83,33	187,50	281,2500	421,8750	632,8125	-6,0135	-751,6875	4.520,2728
2	28,88	585,00	2632,5000	11846,2500	53308,1250	-3,0135	-391,7550	1.180,5537
3	9,47	532,50	3993,7500	29953,1250	224648,4375	-0,0135	-0,9585	0,0129
4	4,09	451,50	4740,7500	49777,8750	522667,6875	2,9865	128,4195	383,5248
5	1,26	229,50	3098,2500	41826,3750	564656,0625	5,9865	101,7705	609,2491
6	1,27	346,50	5717,2500	94334,6250	1556521,3125	8,9865	188,7165	1.695,9008
7	0,51	195,00	3802,5000	74148,7500	1445900,6250	11,9865	119,8650	1.436,7618
8	0,27	135,00	3037,5000	68343,7500	1537734,3750	14,9865	39,9190	1.347,5711
9	0,12	76,50	1950,7500	49744,1250	1268475,1875	17,9865	53,9595	970,5425
10	0,10	85,50	2436,7500	69447,3750	1979250,1875	20,9865	62,9595	1.321,2995
11	0,09	94,50	2976,7500	93767,6250	2953680,1875	23,9865	71,9595	1.726,0565
12	0,03	34,50	1190,2500	41063,6250	1416695,0625	26,9865	26,9865	728,2712
13	<u>0,27</u>	<u>375,00</u>	<u>14062,5000</u>	<u>527343,7500</u>	<u>19775390,6256</u>	29,9865	<u>299,8650</u>	<u>8.991,9018</u>
	129,69	3328,50	49920,7500	1.152019,1250	33299560,6875		0,0195	24,911,9185

	17	18	19	20	21	22
Nº Orden	$z_J^3 n_j$	$z_J^4 n_j$	z_j' (Y_j -ot.)	$z_j' n_j$	$z_j'^2 n_j$	$z_j'^3 n_j$
1	- 27182,6605	163462,9289	- 18	- 2250	40500	- 729000
2	- 3557,5986	10720,8234	- 15	- 1950	29250	- 438750
3	- 0,0002	0,000003	- 12	- 952	10224	- 122688
4	1145,3969	3420,7279	- 9	- 387	3483	- 31347
5	3547,2697	21834,3800	- 6	- 102	612	- 3672
6	15240,2128	136956,1722	- 3	- 63	189	- 567
7	17221,7456	206428,4535	0	0	0	0
8	20195,3742	302657,9754	3	18	54	162
9	17456,6635	313984,2784	6	18	108	648
10	27729,4529	581944,1642	9	27	243	2187
11	41402,0554	993090,4010	12	36	432	5184
12	19653,4903	530378,9147	15	15	225	3375
13	<u>269635,6640</u> 402587,0660	<u>8085429,8390</u> 11350309,0586	18	<u>180</u> - 5310	<u>3240</u> 88560	<u>58320</u> -1256148

	23	24	25	26	27	28
N ² Orden	$Z_j^{(4)} n_j$	$\frac{Y_j - \text{ot}}{C}$	$Z_j^{(1)} n_j$	$Z_j^{(2)} n_j$	$Z_j^{(3)} n_j$	$Z_j^{(4)} n_j$
1	1 122 000	- 6	- 750	4,500	- 27,000	162,000
2	6.581 250	- 5	- 650	3,250	- 16,250	31,250
3	1 472,256	- 4	- 284	1,136	- 4,544	18,176
4	282.123	- 3	- 120	387	- 1,161	3,483
5	22,032	- 2	- 34	68	- 136	272
6	1.701	- 1	- 21	21	- 21	21
7	0	0	0	0	0	0
8	486	1	6	6	6	6
9	3.888	2	6	12	24	48
10	19.683	3	9	27	81	243
11	62.208	4	12	48	192	768
12	50.625	5	5	25	125	625
13	<u>1.642 300</u> 22,6 8,012	6	<u>60</u> -1,770	<u>360</u> 9,840	<u>2.160</u> - 46,524	<u>12.960</u> 279,852

	29	30
Nº Orden	$Y_j - M_e$	$(Y_j - M_e) n_j$
1	3,72	465,00
2	0,72	93,60
3	2,28	161,88
4	5,28	227,04
5	8,28	140,76
6	11,28	236,88
7	14,28	142,80
8	17,28	103,68
9	20,28	60,84
10	23,28	69,84
11	26,28	78,84
12	29,28	29,28
13	32,28	<u>322,80</u> 2.133,24

MEDIA ARITMETICA

Fórmula General :

$$\sum_{j=1}^{13} Y_j n_j \quad \sum_{j=1}^{13} \quad (\text{Col. No. 10})$$

$$M(Y) = \frac{\quad}{n} = \frac{\quad}{443}$$

$$M(Y) = \frac{3.328.50}{443} = 7,51$$

$$M(Y) = 7,51 \text{ cientos de aves de corral}$$

La media aritmética es igual a 7,51 y nos indica que las 443 veredas del Departamento del Meta tienen un promedio de 751 aves de corral.

METODOS ABREVIADOS

Primer Método :

$$a = ot + \frac{\sum_{j=1}^{13} Z'_j n_j}{n}$$

Teniendo que :

$$M(Z'_j) = \frac{\sum_{j=1}^{13} (Y_j - ot) n_j}{n} = \frac{\sum_{j=1}^{13} (Col. N^{\circ} 20)}{443}$$

Reemplazando en la fórmula anterior se tiene :

$$a = 19,50 + \frac{(- 5310)}{443}$$

$$a = 19,50 - \frac{5310}{443}$$

$$a = 19,50 - 11,99$$

$$a = 7,51 \text{ cientos de aves de corral.}$$

Por consiguiente se obtiene un promedio de 751 aves de corral.

Segundo Método.:

$$a = ot. + C \frac{\sum_{j=1}^{13} Z''_j n_j}{n}$$

Sabiendo :

$$M[Z''_j] = C \frac{\sum_{j=1}^{13} \frac{(Y_j - ot.)}{C} n_j}{n} = 3 \frac{\sum_{j=1}^{13} (Col N^{\circ} 25)}{443}$$

$$a = 19,50 + 3 \frac{(- 1.770)}{443}$$

$$a = 19,50 + 3 (- 3,995)$$

$$a = 19,50 - 11,99$$

$$a = 7,51 \text{ ciento de aves de corral.}$$

Por consiguiente se obtiene un promedio de 751 aves de corral

MEDIA GEOMETRICA

$$M_o = \text{antilog.} \frac{\sum_{j=1}^{13} (\log Y_j) n_j}{n} = \text{antilog.} \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 8)}{443}$$

$$M_o = \text{antilog.} \frac{309,11189}{443}$$

$$M_o = \text{antilog.} \quad 0.697770$$

$$M_o = 4,9862$$

La media geométrica es de 498

MEDIA ARMONICA

$$M_{-1} (Y) = \frac{n}{\sum_{j=1}^{13} n_j / Y_j} = \frac{443}{\sum_{j=1}^{13} (\text{Col. N}^\circ 9)}$$

$$M_{-1}(Y) = \frac{443}{129,69}$$

$$M_{-1}(Y) = 3,416$$

MEDIDAS DE POSICION

Obsérvese la posición de :

m/s

en la columna 5 del cuadro resumen .

1. LA MEDIANA

Sea :

$$\frac{n}{2} = \frac{443}{2} = 221,5$$

$$M_e = Y'_{j-1} + C \left[\frac{n/2 - N_{j-1}}{N_j - N_{j-1}} \right] = 3 + 3 \left[\frac{221,5 - 125}{130} \right]$$

$$M_e = 3 + 3 \left[\frac{96,5}{130} \right] = 3 + 2,22$$

$$M_e = 5,22$$

Se dice entonces que el 50% de las veredas tienen 522 aves de corral.

2. CUARTILES O CUARTILAS

$$P_{1/4} = Y'_{j-1} + C_j \left[\frac{\frac{rn}{4} - N_{j-1}}{N_j - N_{j-1}} \right]$$

Para :

$$\frac{1 \cdot n}{4} = \frac{443}{4} = 110,75 \quad Y'_{j-1} = 0$$

$$P_{1/4} = 0 + 3 \left[\frac{110,75 - 0}{125} \right] = \frac{332,25}{125}$$

$$P_{1/4} = 2,6580$$

El 25% de las veredas del departamento tienen aproximadamente 266 aves de corral.

$$P_{3/4} = Y'_{j-1} + C \left[\frac{\frac{3n}{4} - N_{j-1}}{N_j - N_{j-1}} \right]$$

$$\frac{3n}{4} = \frac{3(443)}{4} = 332,25$$

$$P_{3/4} = 9 + 3 \left[\frac{332,25 - 326}{43} \right] = 9 + 3 \left[\frac{6,250}{43} \right] =$$

$$P_{3/4} = 9 + \left[\frac{18,750}{43} \right] = 9,43604$$

De las 443 veredas el 25% tiene más de 944 aves de corral y el 75% tiene menos de 944.

3.

DECILES O DECILAS

$$P_{1/10} = Y'_{j-1} + C \left[\frac{n/10 - N_{j-1}}{N_j - N_{j-1}} \right]$$

$$\frac{rn}{S} = \frac{1 \cdot n}{10} = 44,3 \quad Y'_{j-1} = 0, \quad N_{j-1} = 0$$

$$P_{1/10} = 0 + 3 \left[\frac{44,3 - 0}{125} \right] = \frac{132,9}{125}$$

$$P_{1/10} = 1,0632$$

La décima parte o el 10% de las veredas del departamento alcanzan a 106 aves de corral.

$$P_{9/10} = Y'_{j-1} + C \left[\frac{9 \cdot n/10 - N_{j-1}}{N_j - N_{j-1}} \right]$$

$$\frac{m}{S} = \frac{9 (443)}{10} = 398,7$$

$$P_{9/10} = 15 + 3 \left[\frac{398,7 - 386}{21} \right] = 15 + 3 \left[\frac{12,70}{21} \right]$$

$$15 + \left[\frac{38,10}{21} \right] = 15 + 1,8143$$

$$P_{9/10} = 16,8143$$

El 10% de las veredas tienen más de 1681 aves de corral mientras que el 90% alcanza una cifra inferior.

Como se explicó antes :

$$P_{9/10} = P_{90/100} \text{ y } P_{1/10} = P_{10/100}$$

con lo cual los valores obtenidos tienen el mismo significado .

4. EL MODO O MODA

Véase columna 3 del cuadro resumen.

$$Md = Y'_{j-1} + C \left[\frac{(n_j - n_{j-1})}{(n_j - n_{j-1}) + (n_j - n_{j+1})} \right]$$

$$= 3 + 3 \left[\frac{5}{5 + 59} \right] = 3 + 3 \left[\frac{5}{64} \right]$$

$$Md = 3 + 15 / 64 = 3,2344$$

El valor que más veces se repite es 324 aves de corral que corresponde al máximo absoluto de la distribución.

MOMENTOS ESTADÍSTICOS UNIDIMENSIONALES

1. RESPECTO A UN ORIGEN (CERO) a_r

$$a_r = M(Y^r) = \frac{\sum_{j=1}^{13} Y_j^r \cdot n_j}{n}$$

$$a_1 = M(Y) = \frac{\sum_{j=1}^{13} Y_j \cdot n_j}{N} = \frac{\sum_{j=1}^{13} (Col. N^{\circ} 10)}{443} = \frac{3328,50}{443} = 7,5135$$

$$a_1 = 7,5135$$

Como se explicó anteriormente, el promedio de aves de corral fue de 751 aproximadamente para todas las veredas

$$a_2 = M(Y^2) = \frac{\sum_{j=1}^{13} Y_j^2 \cdot n_j}{n} = \frac{\sum_{j=1}^{13} (Col. N^{\circ} 11)}{443} = \frac{49,920,75}{443} = 112,68792$$

$$a_2 = 112,68792$$

$$a_3 = M(Y^3) = \frac{\sum_{j=1}^{13} Y_j^3 \cdot n_j}{n} = \frac{\sum_{j=1}^{13} (Col. N^{\circ} 12)}{443} = \frac{1152019,125}{443} = 2600,4946$$

$$a_3 = 2600,4946$$

$$a_4 = M(Y^4) = \frac{\sum_{j=1}^{13} Y_j^4 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 13)}{443} = \frac{33299560,6875}{443}$$

$$a_4 = 75.168,3085.$$

2. RESPECTO A LA MEDIDA ARITMETICA m_r

$$m_r = M(Y - a_1)^r = M(Z^r) = \frac{\sum_{j=1}^{13} Z_j^r n_j}{n}$$

$$m_1 = M(Z) = \frac{\sum_{j=1}^{13} Z_j n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 15)}{443} = \frac{0,0195}{443} = 0,00004$$

Teóricamente este momento debería ser cero, nó lo es, debido a las aproximaciones ya que la media aritmética no da un valor entero , pero puede observarse que la cifra tiende a cero.

$$m_2 = S^2 = M(Z^2) = \frac{\sum_{j=1}^{13} Z_j^2 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 16)}{443} = \frac{24911,9185}{443}$$

$$m_2 = 56,2345.$$

En este punto se encuentra localizada la varianza de la distribución.

$$m_3 = M(Z^3) = \frac{\sum_{j=1}^{13} Z_j^3 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 17)}{443} = \frac{402587,0660}{443}$$

$$m_3 = 908,7744.$$

$$m_4 = M(Z^4) = \frac{\sum_{j=1}^{13} Z_j^4 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 18)}{443} = \frac{11350309,0586}{443}$$

$$m_4 = 25621,4651$$

RESPECTO A UN ORIGEN DE TRABAJO

$$m'_r = M \left[(Y_j - \text{ot.})^r \right] = M(Z'^r) = \frac{\sum_{j=1}^{13} Z_j'^r n_j}{n}$$

$$m'_1 = M(Z') = \frac{\sum_{j=1}^{13} Z_j' n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 20)}{443} = \frac{-5310}{443} = -11,9865$$

$$m'_2 = M(Z'^2) = \frac{\sum_{j=1}^{13} Z_j'^2 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 21)}{443} = \frac{88560}{443}$$

$$m'_2 = 199,9097.$$

$$m'_3 = M(Z'^3) = \frac{\sum_{j=1}^{13} Z_j'^3 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 22)}{443} = \frac{-1256,148}{443}$$

$$m'_3 = -2835,5485.$$

$$m'_4 = M(Z'^4) = \frac{\sum_{j=1}^{13} Z_j'^4 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 23)}{443} = \frac{22.608,012}{443}$$

$$m'_4 = 51,0339,3270.$$

RESPECTO A UN ORIGEN DE TRABAJO (ot.) MEDIDO EN UNIDADES DE INTERVALO

$$m''_r = M \left[\left(\frac{Y_j - \text{ot.}}{C_j} \right)^r \right] = M \left[Z''^r \right] = \frac{\sum_{j=1}^{13} Z''^r_j n_j}{n}$$

$$m''_1 = M(Z'') = \frac{\sum_{j=1}^{13} Z''_j n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 25)}{443} = \frac{-1770}{443}$$

$$m''_1 = -3,9955.$$

$$m''_2 = M(Z''^2) = \frac{\sum_{j=1}^{13} Z''^2_j n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 26)}{443} = \frac{9840}{443}$$

$$m''_2 = 22,2122.$$

$$m''_3 = M(Z''^3) = \frac{\sum_{j=1}^{13} Z''^3_j n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 27)}{443} = \frac{-46524}{443}$$

$$m''_3 = -105,0203.$$

$$m''_4 = M(Z''^4) = \frac{\sum_{j=1}^{13} Z''^4_j n_j}{n} = \frac{\sum_{j=1}^{13} \text{Col. N}^\circ 28}{443} = \frac{279852}{443}$$

$$m''_4 = 631,7201$$

RELACIONES ENTRE MOMENTOS - COMPROBACION

a) m_r en función de a_r

$$m_r = \sum_{k=0}^r (-1)^k \binom{r}{k} a_{r-k} a_1^k$$

$$m_1 = a_1 - a_1 = 0$$

$$m_2 = a_2 - (a_1)^2$$

$$m_2 = 112,688 - (7,5135)^2$$

$$m_2 = 112,688 - 56,4527$$

$$m_2 = 56,2353$$

$$m_3 = a_3 - 3 a_2 a_1 + 2 a_1^3$$

$$m_3 = 2600,4946 - 3 (112,688) (7,5135) + 2 (7,5135)^3$$

$$m_3 = 2600,4946 - 2540,0439 + 848,3147$$

$$m_3 = 908,7654.$$

$$m_4 = a_4 - 4 a_3 a_1 + 6 a_2 a_1^2 - 3 a_1^4$$

$$m_4 = 75.168,3085 - 73.155,2648 + 38.169,2514 - 9.560,7198$$

$$m_4 = 25.621,5753$$

b) m_r en función de m'_r

$$m_r = \sum_{k=0}^r \binom{r}{k} (-1)^k m'^{r-k}_1 m'^k_1$$

$$m_1 = m'_1 - m'_1 = 0$$

$$m_2 = m'_2 - (m'_1)^2$$

$$m_2 = 199,91 - (-11,9865)^2$$

$$m_2 = 199,91 - 143,6761$$

$$m_2 = 56,2339$$

$$m_3 = m'_3 - 3 m'_2 m'_1 + 2 (m'_1)^3$$

$$m_3 = (-2.835,5485) - 3 (199,9097) (-11,9865) + 2 (-11,9865)^3$$

$$m_3 = 2.835,5485 + 7.188,6528 - 3444,3496$$

$$m_3 = 908,7547$$

$$m_4 = m'_4 - 4 m'_3 m'_1 + 6 m'_2 m'^2_1 - 3 m'^4_1 = 51214,4740 - 4 (-2.835,5485)(-11,9865) + 6 (199,9097)(-11,9865)^2 - 3 (-11,9865)^4$$

$$m_4 = 51214,4740 - 135.953,2084 + 172.333,5960 - 61.928,5446$$

$$m_4 = 223.548,0700 - 197.881,7530$$

$$m_4 = 25.666,3170$$

c) m_r en función de m''_r

$$m_r = \sum_{k=0}^r \binom{r}{k} (-1)^k C^k m''_{r-k} m''_1{}^k$$

$$m_1 = C (m''_1 - m''_1) = 0$$

$$m_2 = C^2 [m''_2 - (m''_1)^2]$$

$$m_2 = 9 [22,2122 - (-3,9955)^2]$$

$$m_2 = 9 (22,2122 - 15,9640)$$

$$m_2 = 9 (6,2482) , \quad m_2' = 56,2338$$

$$m_3 = C^3 \left[m''_3 - 3 m''_2 m''_1 + 2 (m''_1)^3 \right]$$

$$m_3 = 27 \left[- 105,0203 - 3 (22,2122)(-3,9955) + 2 (-3,9955)^3 \right]$$

$$m_3 = 27 \left[- 105,0203 + 266,2464 - 127,5684 \right]$$

$$m_3 = 27 \left[- 232,5887 + 266,2464 - 127,5684 \right]$$

$$m_3 = 27 \left[- 232,5887 + 266,2464 \right] = \left[27 (33,6577) \right]$$

$$m_3 = 908,7579$$

$$m_4 = C^4 \left[m''_4 - 4 m''_3 m''_1 + 6 m''_2 (m''_1)^2 - 3 (m''_1)^4 \right]$$

$$m_4 = 81 \left[631,7201 - 4 (- 105,0203) (- 3,9955) + \right. \\ \left. 6 (22,2122) (- 3,9955)^2 - 3 (- 3,9955)^4 \right]$$

$$m_4 = 81 \left[631,7201 - 4 (419,6086) + 6 (354,5955) - 764,5479 \right]$$

$$m_4 = 81 \left[631,7201 - 1.678,4344 + 2.127,5730 - 764,5479 \right]$$

$$m_4 = 81 \left[2.759,2931 - 2.442,9823 \right]$$

$$m_4 = 81 \left[316,3108 = 25.621,1748 \right]$$

MEDIDAS DE DISPERSION

Varianza :

$$S^2 = \frac{\sum_{j=1}^{13} (Y_j - \bar{Y})^2 n_j}{n} = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 16)}{443} = \frac{24.911,9181}{443}$$

$$S^2 = 56,2346$$

Primera cuantificación de la dispersión; de acuerdo a lo expuesto en el Capítulo IV se puede considerar como un valor alto y por lo tanto, se dice que existe heterogeneidad en la distribución.

Otra fórmula de cálculo para la varianza es :

$$S^2 = \frac{\sum_{j=1}^{13} Y_j^2 n_j}{n} - (\bar{Y})^2 = \frac{\sum_{j=1}^{13} (\text{Col. N}^\circ 11)}{443} - (\bar{Y})^2$$

$$= \frac{49920,7500}{443} - (7,5135)^2$$

$$S^2 = 112,6879 - 56,4527 = 56,2352.$$

Primer Método Abreviado :

$$S^2 = \frac{\sum_{j=1}^{13} Z_j^2 n_j}{n} - \left[\frac{\sum_{j=1}^{13} Z_j n_j}{n} \right]^2 =$$

$$\begin{aligned} & \frac{\sum_{j=1}^{13} (\text{Col. N}^{\circ} 21)}{443} - \left[\frac{\sum_{j=1}^{13} (\text{Col. N}^{\circ} 20)}{443} \right]^2 \\ &= \frac{88.560}{443} - \left[\frac{- 5310}{443} \right]^2 \end{aligned}$$

$$S^2 = 199,9097 - (11,9865)^2 = 199,9097 - 143,6762$$

$$S^2 = 56,2335$$

Segundo Método Abreviado :

$$\begin{aligned} S^2 &= C^2 \left\{ \frac{\sum_{j=1}^{13} \tilde{z}_j^2 n_j}{n} - \left[\frac{\sum_{j=1}^{13} z_j'' n_j}{n} \right]^2 \right\} \\ S^2 &= (3)^2 \left\{ \frac{\sum_{j=1}^{13} (\text{Col. N}^{\circ} 26)}{443} - \left[\frac{\sum_{j=1}^{13} (\text{Col. N}^{\circ} 25)}{443} \right]^2 \right\} \\ S^2 &= (3)^2 \left\{ \frac{9843}{443} - \left[\frac{- 1770}{433} \right]^2 \right\} \end{aligned}$$

$$S^2 = 9 (22,2122 - 15,9640) = 9 (6,2482)$$

$$S^2 = 56,2338$$

Desviación Típica :

$$S = \sqrt{S^2} = \sqrt{\frac{\sum_{j=1}^{13} (Y_j - \bar{Y})^2 n_j}{n}}$$

$$S = \sqrt{56,2346}$$

$$S = 7,4990 \text{ cientos de aves de corral.}$$

Se observa que el valor de la desviación estándar (S) es muy aproximado al de la media aritmética ($\bar{Y}$), lo cual indica que existe una gran dispersión respecto a la media; esto se entiende perfectamente por la forma como en una distribución normal se plantean los intervalos en base a la desviación estándar.

Desviación Media :

$$DM = \frac{\sum_{j=1}^{13} |Y_j - \bar{Y}| n_j}{n} = \frac{\sum_{j=1}^{13} |Z_j| n_j}{n} = \frac{5898}{443}$$

$$DM = 13,3138.$$

Desviación M_e :

$$DM_e = \frac{\sum_{j=1}^{13} |Y_j - M_e| n_j}{n} = \frac{2.133,24}{443} = 4.815$$

Recorrido :

$$R = Y'_{13} - Y'_0$$

Para la distribución de frecuencias queda indefinido este intervalo, dado que el límite superior del último intervalo es indefinido.

Para datos originales el intervalo está definido así :

$$R = X_{\max} - X_{\min} = 120,00 - 0,16 = 119,84 \text{ cientos aves de corral.}$$

Es decir, que en un intervalo de 119,84 cientos de aves de corral están contenidos todos los valores que pueden tomar las veredas del departamento del Meta en número de aves de corral por vereda.

Recorrido Intercuartílico :

$$\begin{aligned} RQ &= P_{3/4} - P_{1/4} = 9,43604 - 2,6580 \\ &= 6,77804 \end{aligned}$$

El 50% de las veredas tienen entre 2,6580 cientos de aves de corral

y 9,43604 cientos de aves de corral.

Coefficiente de Variación :

$$C.V. = \frac{S}{\bar{Y}} \times 100$$

$$C.V. = \frac{7,4990}{7,5135} \times 100$$

$$\approx 99,81 \%$$

Este coeficiente de variación tan alto, indica una gran dispersión (heterogeneidad) de los datos de la distribución.

Coefficiente de Deformación (Asimetría) :

$$A = \frac{\bar{X} - Md}{S} = \frac{7,5135 - 3,2344}{7,4990}$$

$$= \frac{4,2791}{7,4990} = 0,5706$$

$$A_1 = \frac{3 (\bar{X} - M_e)}{S} = \frac{3 (7,5135 - 5,22)}{7,4990} = \frac{3 (2,2935)}{7,4990} =$$

$$A_1 = \frac{6,8805}{7,4990} = 0,91752$$

$$b_1 = \frac{m_3^2}{m_2^3} = \frac{(908,7744)^2}{(56,2345)^3} = \frac{825870,9101}{177831,4272} = 4,6441$$

$$g_1 = \sqrt{b_1} = \frac{m_3}{S^3} = \frac{908,7744}{(7,4990)^3}$$

$$g_1 = \frac{908,7744}{421,7063} = 2,1550$$



$$g_1 = 2,1550 > 0$$

Estos parámetros indican que la curva se encuentra notoriamente recargada hacia la derecha, es decir, concentrados los datos sobre los

menores valores de la distribución.

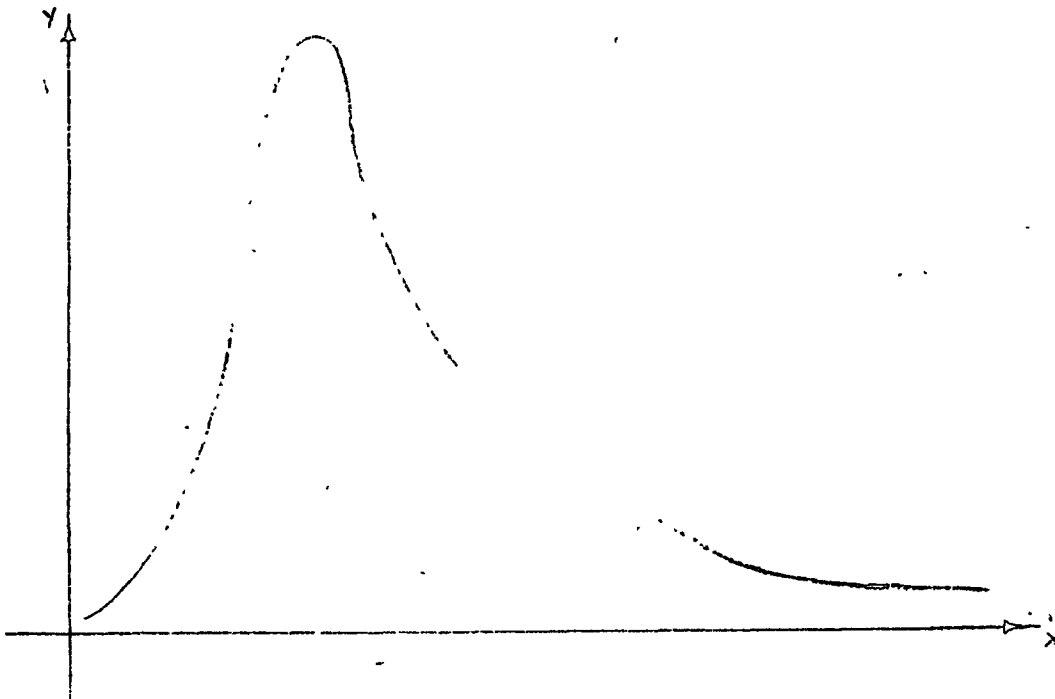
Coeficiente de Apuntamiento (Curtosis) :

$$C = \frac{1/2 (P_{3/4} - P_{1/4})}{P_{90/100} - P_{10/100}} = \frac{1/2 (9,43604 - 2,6580)}{16,8143 - 1,0632}$$
$$= \frac{6,77804}{2 \times 15,7511} = \frac{6,77804}{31,5022} = 0,21516$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{25621,4651}{(56,2345)^2} = \frac{25621,4651}{3162,3190}$$

$$b_2 = 8,1021$$

$$g_2 = b_2 - 3 = 8,1021 - 3 = 5,1021$$



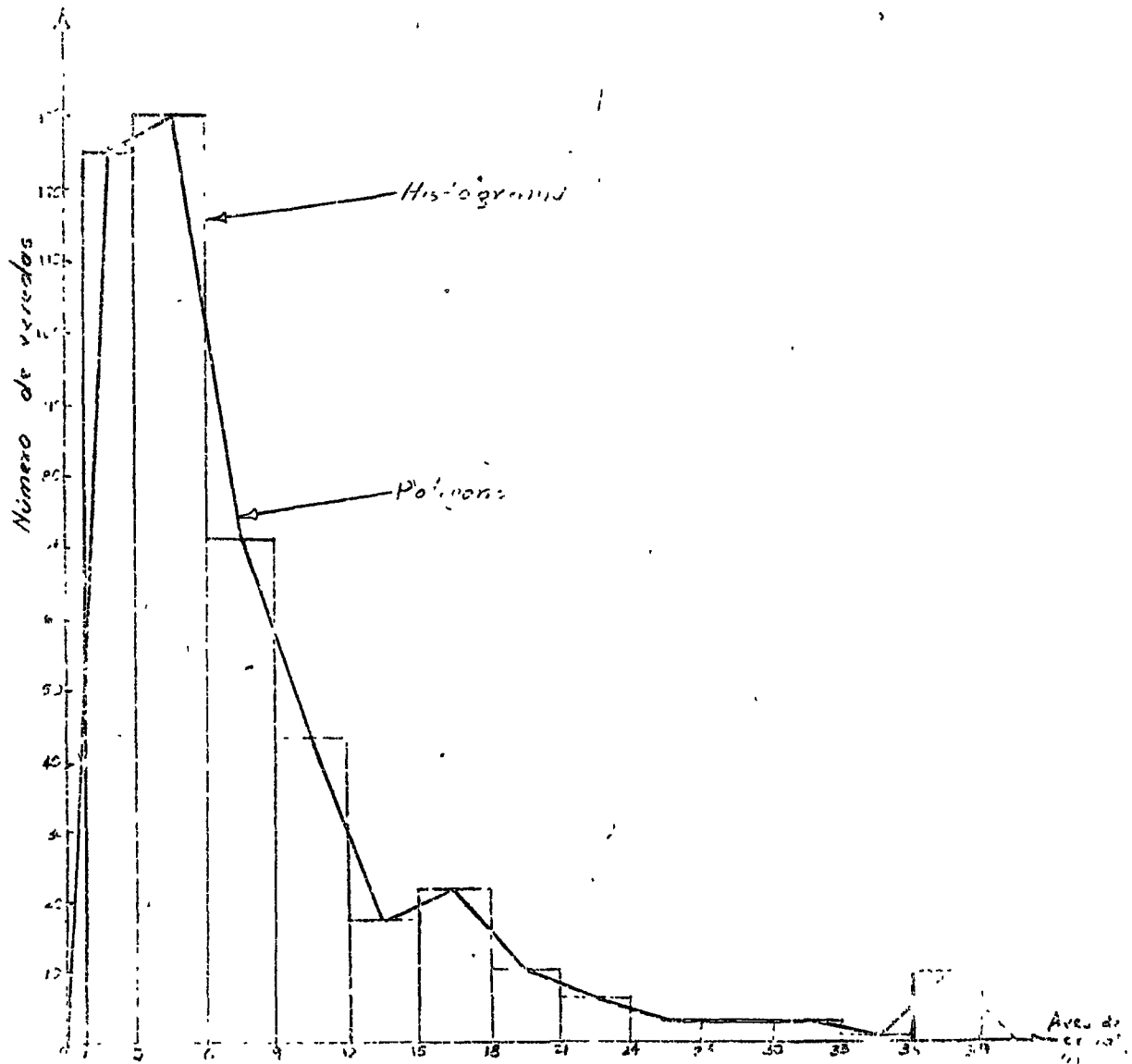
Según el valor de :

$$C = 0,21516$$

indica que la distribución es ligeramente aplanada, sin embargo, los cálculos de los coeficientes b_2 y g_2 muestran que la distribución está notoriamente concentrada (apuntada), son mucho más confiables los valores de los segundos parámetros, porque el valor de C es ante todo empírico, por lo cual se presentan muchos problemas en su uso

En resumen, la distribución se halla recargada hacia la derecha y concentrada notoriamente sobre los valores extremos.

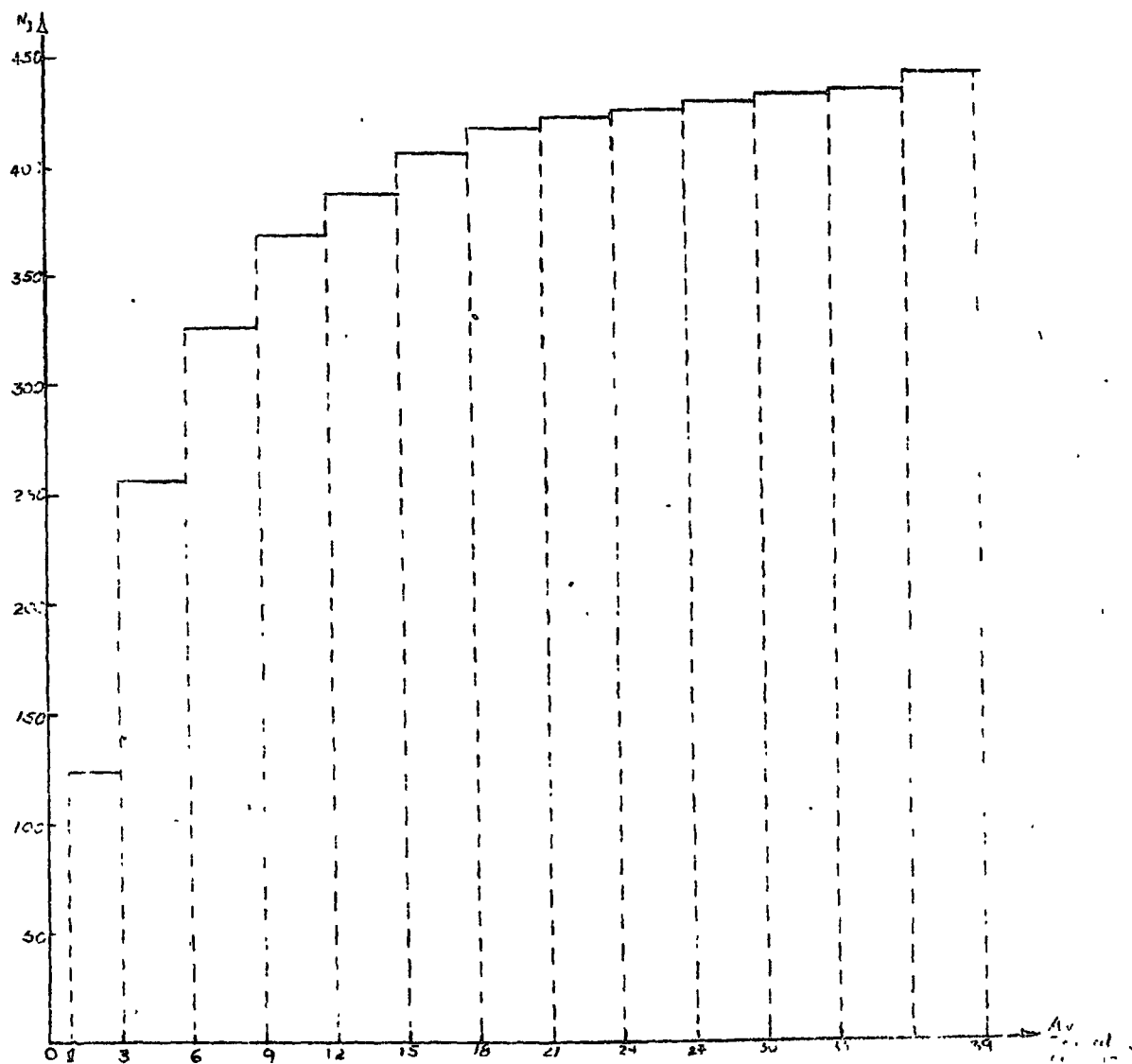
HISTOGRAMA Y POLIGONO DE FRECUENCIAS ABSOLUTAS



(*) Ver cuadro resumen.

(*) Para el histograma y polígono de frecuencias relativas, se obtiene la misma gráfica pero cambia la escala vertical la cual debe estar comprendida en el intervalo (0,0022 y 0,2934).

POLIGONO DE FRECUENCIAS ACUMULADAS ABSOLUTAS



(*) Ver cuadro resumen

(*) Para el polígono de frecuencias relativas acumuladas se obtiene la misma gráfica cambiando la escala vertical la cual debe estar comprendida en el intervalo (0,2922 y 1,0000).

BIBLIOGRAFIA

- D'OTTONE R, Horacio D; Cooperativa de Cultura y Publicaciones, Ltda.
Santiago de Chile, 1971.
- PIATIER, André; Estadística y Observación Económica; Tomo I, Ediciones Ariel S.A. Barcelona, 1967.
- OSTLE, Bernard; Estadística Aplicada, Editorial Lemusa Wiley, S.A.
Mexico 1968.
- YULE, G. UDNY y KENDALL, M.G. Introducción a la Estadística Matemática, Aguilar, S.A. de Ediciones, Madrid (España), 1960.
- CROXTON, Frederik E. y COWDEN, Dudley J.; La Estadística Aplicada a los Negocios, Editorial Hispano Europea, Barcelona, 1963.
- CANSADO, Enrique : Estadística General, Centro Interamericano de Enseñanza Estadística (CIENES), Santiago de Chile, 1959.
- HOLLMAN, Leonardo ; Naciones de Estadística de Empresas, Ediciones Deusto, 1970.
- SANÇHEZ Arango, Javier; Estadística Básica I, Editorial EAFIT, Medellín 1969.

CRAMER, Harold ; Métodos Matemáticos de Estadística, Ediciones Aguilar, S.A., Madrid (España) 1968.

CANSADO, Enrique; Problemas de Estadística tomados del Curso de Estadística General del CIEF, Santiago de Chile, Agosto de 1960.
Conferencias y apuntes de clase como experiencia en la Cátedra.

