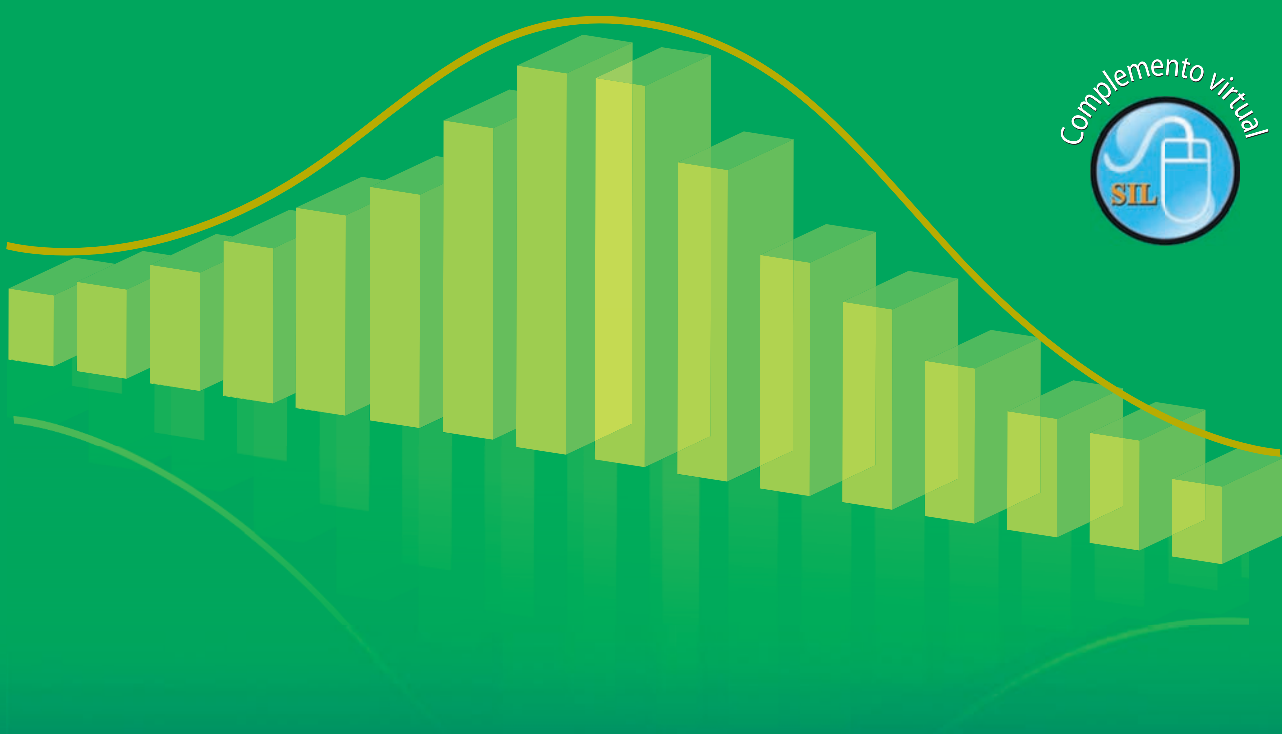
Décima tercera edición

Estadística y muestreo



Complemento virtual

SIL

Ciro Martínez Bencardino

ECOE EDICIONES

BENCARDINO

ción (Norte de
bia). Economista
Jorge Tadeo
, D.C.
iversidad de los
C.). Técnicas
ES-Santiago de
Laboral
o Piedras y
l de Puerto Rico).

Décima tercera edición

Estadística y muestreo

Ciro Martínez Bencardino

Catalogación en la publicación – Biblioteca Nacional de Colombia

Martínez Bencardino, Ciro
Estadística y muestreo / Ciro Martínez Bencardino. -- 13ª. ed. -- Bogotá : Ecoe Ediciones, 2012.
900 p. – (Ciencias exactas. Matemáticas)

ISBN 978-958-648-702-3

1. Estadística matemática 2. Muestreo (Estadística) I. Título II. Serie

CDD: 519.52 ed. 20 CO-BoBN– a746986

Colección: Textos universitarios
Área: Matemáticas
Primera edición: Bogotá, D.E., noviembre de 1978
Segunda edición: Bogotá, D.E., enero de 1982
Tercera edición: Bogotá, D.E., enero de 1984
Cuarta edición: Bogotá, D.E., julio de 1987
Quinta edición: Bogotá, D.E., agosto de 1990
Sexta edición: Santa Fe de Bogotá, D.C., julio de 1992
Séptima edición: Santa Fe de Bogotá, D.C., enero de 1995
Octava edición: Santa Fe de Bogotá, D.C., febrero de 1997
Novena edición: Santa Fe de Bogotá, D.C., febrero de 1998
Reimpresión: Santa Fe de Bogotá, D.C., abril de 1999
Décima edición: Bogotá, D.C., octubre de 2000
Onceava edición: Bogotá, D.C., enero de 2002
Reimpresión: Bogotá, D.C., septiembre de 2003
Décimo segunda edición: Bogotá, D.C., septiembre de 2005
Reimpresión: Bogotá, D.C., enero de 2007
Décima tercera edición: Bogotá, D.C., 2012

ISBN:978-958-648-702-3

Coordinación editorial: Alexander Acosta Quintero
Diagramación: Raúl Enrique Rodríguez
Portada: Edwin Nelson Penagos
Impresión: Imagen Editorial Impresores
imagenimvega@yahoo.com

Impreso y hecho en Colombia

A mis padres, hermanos,
esposa, hijos y nietos, por
ellos vivo, investigo y trabajo.

Como un homenaje póstumo a
la memoria del doctor Carlos
Alzate Giraldo, fundador de
ECOE.

ÍNDICE GENERAL

CAPÍTULO 10

ÍNDICE DE TABLAS

TABLAS

FÓRMULAS

CONTENIDO DEL SIL

1. Solucionario de los ejercicios para resolver en el libro - Capítulos 1 al 13
2. Aplicación de EXCEL en los procesos Estadística
3. Resumen de fórmulas de aplicación en la Estadística Descriptiva y en la Inferencia
4. Teoría y aplicaciones, sobre: Desarrollo de un sistema de ecuaciones; Sumatoria y Productoria; Porcentajes, Tasas y Razones

PRESENTACIÓN

Gracias a la generosidad de profesores y estudiantes que lo han preferido como texto o como fuente de consulta, se ofrece esta nueva Edición, con nueva presentación mucho más cómoda y fácil de manejar, con una revisión y actualización de los contenidos, así como de estilo, que esperamos sea del agrado de todos aquellos que de alguna manera tendrán la oportunidad de consultarla.

En esta décima tercera edición se suprimieron algunos temas que han perdido vigencia en los programas, en contraprestación se incluyeron y mejoraron otros, manteniendo las exigencias y la actualización en la enseñanza de esta materia de acuerdo a los diferentes niveles, así como en los objetivos que se exigen en las carreras intermedias y profesionales. En esta ocasión se ha suprimido el CD que siempre lo acompañaba, ahora se ofrece la consulta a la Página web w**ww.ecoeediciones.com** donde encontrará entre otros, a) Proceso continuado de la aplicación del Excel en todos los temas de esta publicación. b) Solucionario con todo el proceso que conlleva el desarrollo de los problemas propuestos, c) Además incluye un tema que fue suprimido pero que lo seguimos considerándolo importante en proceso integral del conocimiento de esta materia, como son la Sumatoria, Productoria, y el uso de los Porcentajes, Tasas y Razones

El objetivo principal y la razón de esta publicación sigue siendo el estudiante, a fin de facilitarle el aprendizaje de esta valiosa herramienta, de ahí la equivocación de algunos que la consideran "elemental", comentarios que nos alegra, pues no está dirigido a profesionales de la Estadística con grandes conocimientos matemáticos, por el contrario se pretende ayudar a aquellos que apenas se inician en esta disciplina, eliminando procesos que no solo dificultan su aprendizaje, que de alguna manera hacen perder el interés sobre aquellos temas que tienen gran importancia en su aplicación en tantos campos del conocimiento.

Creo firmemente que los contenidos de esta publicación van a ser un valioso apoyo, para quienes creen que los programas de computador la han desplazado, este último solo permite observar la base de datos en su conjunto y los resultados obtenidos que son los requeridos, pero elimina los procedimientos que fueron aplicados para su obtención.

Sea la oportunidad para manifestar nuevamente, mis más sinceros agradecimientos a profesores y estudiantes que me han hecho llegar comentarios, críticos algunos, que han sido de gran ayuda al mejoramiento de los contenidos en cada edición, como aquellos que reconocen esta labor, lo que me da aliento para seguir en esta tarea de divulgación.

El Autor.

1

CAPÍTULO

Conceptos generales

TÉRMINOS Y DEFINICIONES

El ser humano aprende en la medida en que participa en el descubrimiento y la invención. Debe tener la libertad para opinar, para equivocarse, para rectificarse, para ensayar métodos y caminos para explorar

Ernesto Sábato

CONTENIDO

- Términos, definiciones, conceptos generales.
- Finalidad de la Estadística,
- Etapas en una investigación de carácter estadístico.
- Monografías y Encuestas Diferencias.
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Entender y manejar correctamente cada uno de los términos que deberán ser utilizados.
- Conocer la importancia que tiene la Estadística no sólo como asignatura, sino en el ejercicio profesional.
- Poder realizar pequeñas investigaciones estadísticas, aplicando los conocimientos recientemente adquiridos, en cada una de las etapas del proceso.
- Distinguir con claridad la diferencia entre una Monografía y una Encuesta, de carácter estadístico.

ASPECTOS GENERALES

RESEÑA HISTÓRICA

La *estadística* que muchos la consideran como algo nuevo, dado que hoy en día todo se fundamenta en el manejo de grandes cantidades de información o datos de carácter estadísticos, pero su aplicación o uso se remonta a la antigüedad, pues hay vestigios de su inicio como en China donde el emperador Yao (2.239 A de C.) dispuso la realización de un Censo; algo similar ocurrió en la época del Rey Herodes, relacionado con un empadronamiento cuando se habló del nacimiento de Jesús.

La *estadística* como disciplina, tuvo su inicio en Alemania a mediados del Siglo XVII, en el reinado de Godofredo de Achewall (1.719-1.772) se utilizó la palabra estadística y se separó de la Sociología. Hoy en día ha alcanzado un alto desarrollo, con el apoyo de tantos investigadores dedicados a perfeccionar métodos estadísticos y a través del avance tecnológico, permitiendo la selección de aquellos procedimientos que buscan mejorar los resultados, al mismo tiempo disminuyendo los márgenes de error.

La palabra *estadística,* para algunos, proviene de la palabra status, que significa estado; para otros se deriva del vocablo griego statera que significa balanza, otra más confiable proviene de la palabra alemana Stara que significa estado, la encargada de hacer estadísticas.

APLICACIÓN DE LA ESTADÍSTICA

En la vida diaria, los diversos fenómenos de orden económico, social, político, educacional e incluso biológico, aparecen, se transforman y finalmente desaparecen.

Para tan abundante y complejo material informativo recogido, se hace necesario contar con un registro ordenado y continuo, a fin de determinar o seleccionar los datos requeridos por el estudio o investigación, en cuanto a lo sucedido, sucede o puede suceder. Para ello se requiere contar con un método o conjunto de reglas o principios, que nos permita la observación, el ordenamiento, la cuantificación y el análisis de dichos fenómenos. Este método se le denomina *Estadística.*

> La palabra *estadística* se refiere a un sistema o método usado en la recolección, organización, análisis y descripción numérica de la información. También se puede decir que la estadística estudia el comportamiento de hechos o fenómenos de grupo.

Se consideran dos fases en el campo de la estadística. En primer lugar, está la fase que sólo se limita a la descripción de un conjunto de datos sin llegar a conclusiones o a generalizar con respecto a un grupo mayor. A esta fase se le da el nombre de *Estadística Descriptiva o Deductiva*. En la segunda, ella implica el análisis mediante la cual trata de llegar a conclusiones acerca de un grupo más grande o población, basado en la información de un grupo menor o muestra, procedimiento o técnica denominada como *Estadística Inductiva* o *Inferencia*.

En un principio se consideraba que la función primordial de la *Estadística*, era la descripción de las características de un grupo, actividad que la hacía confundir con el papel que cumple la historia, consistente en observar y describir el hecho. En su origen, las estadísticas eran históricas; hoy en día, la estadística, además de ser descriptiva, es analítica, considerándose ésta cómo la función más importante que ella realiza, ya que además, permite obtener conclusiones para un grupo mayor, denominado población, partiendo de una investigación realizada en un grupo menor, conocido como muestra, cuyos elementos o unidades, en la mayoría de los casos, se seleccionan aleatoriamente o al azar.

La *estadística descriptiva o deductiva*, tiene como finalidad, colocar en evidencia aspectos característicos del grupo (promedios, variabilidad de los datos, etc.) que sirven para realizar comparaciones, sin pretender sacar conclusiones de tipo más general. Esta descripción, por lo general se realiza a través de la elaboración de cuadros, gráficos, cálculo de promedios, proporciones, varianzas de una o más variables relacionadas.

La *estadística analítica o inductiva*, busca dar explicaciones al comportamiento de un conjunto de observaciones, probar la significación o validez de los resultados, además, intenta descubrir las causas que lo originan, siendo de gran aplicación en el muestreo, logrando de esta manera, obtener conclusiones que se extienden mas allá de las muestras estadísticas mismas.

Uno de los objetivos principales de la *estadística*, consiste en realizar inferencias acerca de los valores estadísticos de la población, denominados parámetros, a través de la información obtenida mediante una muestra, que nos permite el cálculo de estimadores.

> La *estadística descriptiva-analítica*, se podría definir: como un conjunto sistemático de procedimientos utilizados para observar y describir numéricamente un fenómeno, además, descubrir las leyes que regulan la aparición, transformación y desaparición de los mismos.

También se puede considerar la *estadística*, como aquel método que permite no sólo describir el hecho o fenómeno, sino también, que permite deducir, evaluar y sacar conclusiones acerca de una población, con los resultados obtenidos a través de una muestra.

Generalmente asociamos la *estadística* con cifras sobre un campo en particular. Como por ejemplo: el número de nacimientos, defunciones, transacciones comerciales, valor de las acciones en el mercado de valores, volumen físico y monetario de las importaciones y exportaciones, beneficios y utilidades de las empresas, demanda presente o potencial de algún producto.

> *Estadísticas* (en plural). Se refiere a un ordenamiento sistemático de datos, presentados en forma de tablas o cuadros y gráficas. En otras palabras, las *estadísticas* son datos agrupados en forma metódica y presentados en publicaciones, elaboradas por diversas empresas o entidades privadas o del estado, buscando ser consultadas por los interesados.

Se clasifican en:

Estadísticas o fuentes primarias. Son datos obtenidos mediante encuestas directas, a través de la utilización de cuestionarios o como resultado de la observación directa, técnica muy utilizada en estudios de carácter científico o en investigaciones de mercados. Se podrá decir, que son datos publicados por quien recoge directamente la información primaria.

Éstas podrían clasificarse en*: personales* (entrevistas, correo, experimentos, etc.); *unipersonales* (auditoría, análisis de rastreo o de contenido, simulación); mixta (observaciones).

Estadísticas o fuentes secundarias. En éstas, los datos se obtienen de publicaciones, las cuales podrán ser reproducidas en forma total o parcial. Se les considera de gran valor para cualquier tipo de investigación. Son consideradas como *fuentes secundarias*: bibliotecas, centros de documentación, folletos, revistas. archivos, etc.

Estadísticas o series temporales. Generalmente denominadas como series de tiempo o cronológicas. La información es obtenida y ordenada en forma cronológica, a medida que se va produciendo a través del tiempo, ya sean días, meses, semestres, años. Cuando la investigaciones se realizan aisladamente, no periódicas, se les denomina: investigaciones *atemporales o aisladas*.

Las estadísticas, también se pueden clasificar como *internas o externas.* Las primeras se originan y se elaboran en la misma fuente de información, por ejemplo en una empresa, son registros internos, tales como: la producción y ventas diarias, semanales, etc. salarios, horas trabajadas, etc. Las *externas*, son registros de información originados fuera de la empresa: opinión de los consumidores, precios de la competencia, etc.

Clases de estadísticas. Se clasifican de acuerdo con el tema, sector o materia de estudios, así por ejemplo: estadísticas industriales, agropecuarias, de transporte, comunicaciones, salud, educación, etc.

Estadístico. Nos referimos a la persona que trabaja en la elaboración y análisis de información de carácter estadístico. En nuestro medio utilizamos el término estadígrafo, para referirnos a esa persona, cuando en realidad su significado es el de una medida estadística.

FINALIDAD DE LA ESTADÍSTICA

En términos generales, se considera que su finalidad es suministrar información y su utilidad dependerá en gran parte del fin que se proponga y de la forma como se obtengan los datos. La creciente especialización lograda por las empresas, ha obligado a la utilización continua de métodos estadísticos, como en la producción masiva, en el control de calidad, en éste último, evitando la inspección de cada artículo y reemplazado por selección de muestras. Por ejemplo, en el control realizado sobre dureza, resistencia y duración, que por lo general implica la destrucción del artículo, lo cual hace necesario la aplicación de las técnicas de muestreo, a fin de reducir su número, como forma de obtener conclusiones que puedan ser extensivas al total de la producción de la cual se extrajo la muestra.

Otras aplicaciones que se pueden hacer, son las investigaciones realizadas con el objeto de conocer los cambios en los gastos o gustos de los consumidores, situaciones en la demanda: exceso en las existencias; capacidad de compra en un grupo familiar. Precisando sobre la finalidad de la *estadística*, podríamos mencionar:

- *Conocer la realidad en una observación o fenómeno.* Al cuantificarlo, estamos conociendo su situación actual, es así como al tener información por horas, semanalmente, etc. sobre la producción de un artículo, se sabrá si se está cumpliendo con lo programado o, si por el contrario, estamos por debajo o por encima de lo establecido, pues ello afectará de alguna manera las existencias del producto o de la materia prima, ventas, precios, costos, etc.

- *Determinar lo típico o normal de esa observación*. Cuando se cuantifica la característica de un fenómeno, se obtiene un valor denominado *promedio*, es ésta la forma de referirnos al grupo cuantificando así su comportamiento. Si decimos que el promedio de unidades vendidas de un artículo A, en 20 almacenes, en el mes de marzo fue de 1.200 unidades, es una forma de determinar el comportamiento del grupo; pero con este valor típico o normal del conjunto no se quiere decir que los 20 almacenes vendieron la misma cantidad; posiblemente algunos más, otros menos, pero el promedio fue de 1.200.
- *Determinar los cambios que presenta el fenómeno.* Cualquier fenómeno de carácter estadístico presenta variaciones a través del tiempo, pero requiere una observación continua para poder determinar la magnitud del cambio. La demanda, la producción, las ventas, los precios, son fenómenos que constantemente registran cambios que deben ser tenidos en cuenta por la administración de una empresa.
- *Relacionar dos o más fenómenos*. Desde el punto de vista de correlación, se puede determinar si existe una relación válida entre dos o más características de una misma observación, o entre dos o más fenómenos. Tal es el caso de la relación de ingresos y gastos para un grupo de familias o la relación entre las unidades producidas, vendidas y el precio.
- *Determinar las causas que originan el fenómeno.* Los cambios que se observan en un fenómeno pueden tener una o varias causas de origen. El aumento de las exportaciones de un artículo puede ser causado por un mejoramiento del precio en el mercado externo, por un volumen de producción superior a la demanda interna, o por otras causas que producen cambios en el comportamiento del fenómeno.
- *Hacer estimativos sobre el comportamiento futuro del fenómeno*. En numerosas ocasiones se requiere proyectar, para estimar el comportamiento futuro de un fenómeno. La proyección a dos o tres años de la población de una región, sirve para determinar el mercado potencial del producto, es decir, el conocimiento, aunque sea estimado, de la demanda en los años próximos.
- *Obtener conclusiones de un grupo menor (muestra), para hacerlas extensivas a un grupo mayor (población).* Al realizar mediciones en una muestra, cuyo tamaño debe ser el óptimo deseado, los resultados obtenidos pueden ser considerados como el comportamiento que se obtendrá si se hubiera trabajado con todos los elementos que constituyen la población, de la cual se extrajo la muestra.

Además de lo anterior, la *estadística* facilita una serie de instrumentos o técnicas que, al ser utilizadas correctamente, permiten determinar el grado de validez y confiabilidad, ya sea de las predicciones o de las conclusiones obtenidas a partir de muestras.

COLECTIVOS INVESTIGADOS POR LA ESTADÍSTICA

La conformación de un grupo o colectivo alrededor de cierta característica, o características, que permiten ser investigadas en cuanto a su comportamiento, se denomina fenómeno. Los fenómenos pueden ser de carácter económico, social, politico, u otros, considerando como tales: exportaciones, importaciones, ventas, producción, defunciones, nacimientos, accidentes laborales, etc.

Los fenómenos de carácter estadístico deben reunir ciertos requisitos, como *manifestarse al exterior para poder ser observados*, lo cual se logra mediante registros. Se consideran como ejemplos: las nóminas de pago, donde quedan consignados el número de empleados, las categorias, sueldos y otras tantas características que podrian ser observadas; los manifiestos de aduana, en donde se registran, tanto las importaciones como las exportaciones; las licencias de construcción concedidas por las Secretarías de Obras Públicas; los registros de nacimientos, defunciones, escrituras de venta, hipotecas, todas ellas realizadas en las Notarías, etc. Además, las características de estos fenómenos *deben ser cuantificables* en tal forma que permitan determinar la intensidad con que se producen. Un ejemplo clarifica mejor lo anterior: si consideramos las exportaciones de café realizadas por las diferentes firmas exportadoras en un período determinado, encontramos que se trata de un fenómeno que se manifiesta al exterior y que se puede observar en los registros de exportación, y por otra parte, se cuantifica el número y valor de las remesas realizadas, además se podrá obtener la intensidad o sea la cantidad de sacos o kilos correspondientes a cada remesa o al total exportado en dicho período.

Fuera de las anteriores condiciones implícitas en un fenómeno estadístico, podemos anotar otras que se consideran dentro del campo de estudio de la estadística:

- *Colectivos.* La estadística estudia el comportamiento de los fenómenos de grupo, prescindiendo de aquellos fenómenos individuales que pueden ser considerados como resultados de casos aislados. Si bien es cierto que para estudiar el colectivo se requiere información individualizada, las conclusiones que se obtienen de la investigación no se refieren a cada elemento individualmente, sino al conjunto de los elementos considerados como grupo.

- *Frecuente repetición.* Aquellos fenómenos que se presentan una vez y no vuelven a repetirse en las mismas condiciones, no son estudiados por la estadística. Por el contrario, se requiere que el fenómeno sea de frecuente repetición. Tal es el caso de las transacciones comerciales, inventarios, accidentes de trabajo, nacimientos, etc.

- *Distinta frecuencia.* Si observamos las ventas realizadas por un almacén, encontramos que se trata de un fenómeno de distinta frecuencia, porque de lunes a sábado se repite dicha operación, pero el valor de esas ventas varían de un día a otro, de semana a semana o de mes a mes, dependiendo del período examinado. Si el fenómeno tiene la misma frecuencia, no requerirá de la estadística, pues basta su registro una sola vez y se tendrá igual información en cada período.

- *Manifestarse al exterior.* El contrabando, el cultivo y comercio de la marihuana, las transacciones en dólares en el mercado negro, son fenómenos que diariamente se suceden pero no son registrados. Por lo tanto, se dice que no se manifiestan al exterior para ser observados estadísticamente.

- *Distantes en el tiempo.* Las ventas, en cierta cantidad, de determinados artículos estacionales, tales como pólvora, juguetes, artículos para celebrar el dia de la secretaria, la primera comunión o la recolección de un determinado producto agrícola, son fenómenos que no se suceden todos los días. Por el contrario, se producen en una determinada época o fecha y hay que esperar cierto tiempo para que se presente nuevamente.

- *Distantes en el espacio.* El pago de cheques por el *Banco Comercial,* es un fenómeno que podemos observar en diferentes sitios, en las sucursales que tiene en una ciudad, en todo el departamento o en todo el pais. Las exportaciones e importaciones que se realizan por los diferentes puertos del país. Son ejemplos que muestran al fenómeno originado en diferentes sitios, sin ser indispensable que se presente en un mismo lugar.

- *Cualitativos que pueden cuantificarse.* Algunas características cualitativas investigadas, requieren ser cuantificadas para que sean consideradas dentro del campo estadístico. Sin embargo no todas son cuantificables, tales como el grado de religiosidad, la moral y tantas otras y por tal razón quedan fuera de la acción estadística. Son ejemplos de hechos cualitativos, los elementos clasificados por: sexo, estado civil, profesión, ocupación, primaria, secundaria, universitaria, por nacionalidad, mes de nacimiento, colores, etc.

- *Aquellos fenómenos que son accidentales,* tanto en el tiempo como en el espacio, no son propios de la investigación estadística, se presentan una sola vez y no vuelven a ocurrir.

ALGUNOS TÉRMINOS Y CONCEPTOS QUE SE DEBEN CONOCER Y MANEJAR

A continuación se definen en forma sencilla y fácil de entender, algunos términos que serán utilizados en el desarrollo de los diferentes capítulos, en los cuales se ha dividido este libro.

Vale la pena recalcar la importancia del conocimiento y el manejo de los términos que a continuación se expondrán, ya que nos permitirá hablar un solo lenguaje, precisar lo que se va a hacer y entender su uso en los diversos aspectos que conlleva el desarrollo de los temas del presente libro. Se recomienda leer detenidamente las primeras páginas de los capítulos 7 y 13, donde encontrará más información que le ayude a comprender mejor estos términos.

Población. Es un conjunto de medidas o el recuento de todos los elementos que presentan una característica común. El término población se usa para denotar el conjunto de elementos del cual se extrae la muestra.

Lo ideal sería, que el número de elementos o unidades de observación que constituyen la muestra, denominada también población por muestrear o *población* muestreada, fuera igual al contenido total en la *población* o *población objetivo.* Pero como no ocurre así, las conclusiones consideradas válidas para la muestra pueden ser extendidas a la población, garantizándolas estadísticamente mediante la indicación de la validez del proceso.

Los *elementos* o *unidades* que integran la población o la muestra pueden corresponder a personas objetos o cosas. Además, el *elemento* puede ser una entidad simple (una persona) o una entidad compleja (una familia) que se denomina *unidad*, la que, a su vez, puede estar compuesta de elementos o de unidades, que en algunos casos se les define como *conglomerados*.

Según sea el número de elementos, la población puede considerarse como *finita* o *infinita.* Es una población *finita* cuando el número de elementos que la compone es limitado; *infinita* cuando consta de ilimitado número de elementos. Sin embargo, algunas veces, una población finita con un número grande de elementos, el comportamiento en el cálculo del tamaño de muestra resulta igual al de una población infinita, ya que existe un tamaño muestral óptimo tal que, a partir de un determinado valor calculado, éste no se altera por más que aumente el tamaño poblacional.

Marco: se denomina *marco, marco muestral o marco de referencia*, a la lista, mapa o cualquier otro material aceptable, que contenga todas las *unidades* o *elementos* perfectamente identificados y actualizados, donde se selecciona la muestra.

Características: es importante resaltar el hecho de que a pesar de encontrarse una población constituida por un grupo de elementos o unidades, a la estadística no le interesa el elemento o la unidad en si, sino sus características.

Éstos son algunos ejemplos de lo que podrÍan denominarse elementos de una población y características de la misma:

ELEMENTOS	CARACTERÍSTICAS
Personas	Salarios
Personas	Horas trabajadas
Personas	Cargos
Departamentos	Ventas
Hogar	Consumo

Las *características* (o caracteres). Corresponden a ciertos rasgos, cualidades o propiedades que poseen los elementos que constituyen la población o la muestra.

Algunos caracteres son mensurables y se describen numéricamente; por tal motivo se denominan caracteres *cuantitativos* o *variables*, tales como: estatura, peso, ingreso, valor, producción, ventas, etc. Otros se expresan mediante *palabras* por no ser mensurables pero si cuantificables, tales como: profesión, cargo, marcas, calidad, etc. y se denominan *caracteres cualitativos o atributos*.

Las variables o aspectos cuantitativos se clasifican en: *variables discretas* y *continuas*. La primera se refiere a aquélla que sólo puede tomar unos determinados valores, siendo imposible que llegue a tomar valores intermedios entre dos consecutivos, es decir, toma únicamente valores enteros: 1, 2, 3 etc., tal es el caso del número de hijos por familia, nunca se tendrá 2 hijos y medio o dos hijos y cuarto, etc. En el segundo, se toman todos los valores infinitos posibles en un intervalo, es decir, se admiten valores fraccionarios, como el número de años de una persona: 20 años, tres meses, cinco días etc. La distinción que generalmente se hace entre variables *discretas* y *continuas* es más teórica que práctica, pues en la práctica se podrá decir que todas las variables son discretas, pero dada la facilidad de agruparlos en intervalos, como por ejemplo: edades entre 6 y 10 años, más de 10 y 20 años y así sucesivamente hacen realidad esa distinción.

Si analizamos o describimos una característica (atributo o variable) en forma independiente nos referiremos a *distribuciones unidimensionales* o *univariante*. Si relacionamos dos características entre si (variables; variable y atributo; o dos atributos) nos referimos a *distribuciones bidimensionales* o *bivariante*. En el caso de relacionar más de dos características se dirá que la *distribución* es *pluridimensional* o *multivariante*.

Investigación total. Denominada también como *censo* o *investigación exhaustiva*, es aquélla en la cual se toma la totalidad de los elementos o unidades que conforman la población objeto de estudio.

No siempre se puede realizar una investigación total, y debemos observar una parte de ellas, debido a circunstancias como las que se detallan a continuación:

- *Poblaciones* muy grandes o infinitas.
- *Tiempo* requerido demasiado grande.
- *Costos* tan elevados que sería imposible con los recursos disponibles.
- *Recursos humanos* con los que no se cuenta para el desarrollo de la investigación.
- *Destrucción* del elemento como ocurre en el control de calidad de un determinado producto, como por ejemplo, probar su resistencia o dureza.

- *Homogeneidad* de la característica, como sucede en la preparación de alimentos o en los exámenes de laboratorio.

Muestra o investigación parcial. Se define como un conjunto de medidas o el recuento de una parte de los elementos pertenecientes a la población. Los elementos se seleccionan aleatoriamente, es decir, todos los elementos que componen la población tienen la misma posibilidad de ser seleccionados.

Para que una muestra sea representativa de la población, se requiere que las unidades sean *seleccionadas al azar,* ya sea utilizando métodos tales como: el sorteo, *las tablas de números aleatorios, la selección sistemática o cualquier otro método que sea al azar.*

Una muestra *no probabilística* corresponde a la selección de las unidades o elementos, según el criterio o conveniencia del investigador. En estos casos, algunas unidades tienen mayor posibilidad que otras de ser seleccionadas; por tal razón no se puede determinar la validez, ni la confianza que merecen dichos resultados. Casi por lo general, las unidades se seleccionan en forma *caprichosa, voluntaria*, por cuotas, tal como sucede en las encuestas de opinión, que en la mayoría de los casos no tienen credibilidad, de ahi que se le conozca como *Muestra circunstancial o errática.*

Podemos concluir, en forma resumida, mediante un esquema, que la observación de una población se puede hacer de diferentes maneras, tal como se presenta a continuación. Este tema se amplia en los capítulos 7 y 13, por lo tanto se recomienda consultarlo.

- Observación
  - Exhaustiva o total o censo
  - Parcial
    - Subpoblación
    - Muestra
  - Mixta
- Aleatoria
  - Simple,
  - Estratificada
  - Por etapas, áreas o conglomerados
  - Por fases
  - Sistemática
  - Mixta
- No aleatoria
  - Voluntaria
  - Por cuotas
  - Intencional o a juicio

Parámetros. Son todas aquellas medidas que describen numéricamente la característica de una *población*. También se les denomina *valor verdadero*, ya que una característica poblacional tendrá un solo parámetro (media, varianza, etc.). Sin embargo, una población puede tener varias características y, por tanto, varios parámetros. Algunos lo denominan como *valor estadístico de la población.*

Estimador o *Estadígrafo*. La descripción numérica de una característica correspondiente a los elementos de una *muestra*, se denomina *estadígrafo*. De una población se pueden obtener (M) número de muestras posibles y en cada una de ellas se puede cuantificar la característica, obteniéndose por lo general, valores diferentes para cada muestra, a pesar de ser utilizado el mismo estadígrafo o medida.

También se le conoce como *Estimador Puntual* si se trata de un *promedio, varianza, proporción*, etc. Como por lo general, existe una diferencia entre el estimador y el parámetro, denominado *error*, es aconsejable utilizar el *estimador por intervalos,* dentro del cual deberá estar contenido el parámetro con cierto margen de seguridad, establecido por el investigador.

La diferencia entre el *estimador* y el *parámetro*, como ya se dijo, se le denomina como *error de muestreo,* y lo fija el investigador para determinar el tamaño de la muestra, con base en la experiencia y el conocimiento que tenga sobre el comportamiento y la variación que tenga la característica o características principales relacionados con el objetivo mismo de la investigación.

Errores ajenos al muestreo: no son cuantificables (como el anterior) y son resultados de errores sistemáticos que se producen a lo largo de la investigación, como: en la preparación, formulación, recopilación presentación, entre otras.

LA PROYECCIÓN Y PREPARACIÓN DE INVESTIGACIONES ESTADÍSTICAS

En nuestro diario quehacer profesional, en la mayoría de las casos, nos corresponde analizar información estadística para la toma de decisión en el desarrollo de nuestro trabajo y gran parte del libro trata de colaborar con el lector, presentando técnicas que les permitan hacerlo, en gran manera acertada; pero otras veces, no pocas, hay necesidad de realizar la investigación a fin de obtener los datos necesarios.

El método estadístico permite recolectar u observar mediante registros, que se ordenan, clasifican, cuantifican y presentan mediante cuadros, gráficas, permitiendo la depuración de datos en dos aspectos, en su presentación clara y en su simplificación, ya que de esta manera facilita manejar grandes cantidades de información presentadas en forma resumida, tarea que realizamos a través de la estadística descriptiva; luego la inferencia estadística permite el análisis y la búsqueda de relaciones que puede existir entre ellas ya que trabajamos sobre abstracciones, situación muy diferente comparada con el *método experimental* que descansa sobre realidades físicas.

Algunos problemas que nos planteamos en la vida diaria, tendrán contestación a través del análisis de los datos que se han registrado, a medida que se producen en un estricto orden cronológico: precios, ventas, accidentes, exportaciones, inventarios al finalizar el año, etc. Cuando no existen esos registros es necesario la realización de encuestas, ya sea tomando toda la población (censo) o una parte de ella (muestra) para deducir el comportamiento de las características de la totalidad de la población. En este caso podemos considerar algunas etapas, no necesariamente todas, que deben ser tenidas en cuenta para su desarrollo.

El proceso de elaboración estadística lo dividimos en seis (6) fases:

A. Planeamiento o preparación
B. Recolección
C. Crítica y codificación
D. Procesamiento
E. Análisis e interpretación (numérico y gráfico)
F. Publicación

PLANEAMIENTO Y PREPARACIÓN DE UNA INVESTIGACIÓN DE TIPO ESTADÍSTICO

Toda recolección de datos estadísticos de un hecho, requiere el establecimiento previo de un plan que detalle los aspectos que la investigación va a abarcar, que fije los procedimientos a seguir, y que resuelva de antemano las posibles dificultades que se presume hayan de presentarse.

La implantación de una nueva información estadística en una empresa privada u oficial o en un caso particular, exige el estudio y la redacción de un proyecto al respecto.

En general, un proyecto o plan de investigación estadística debe contemplar los siguientes aspectos:

1) El objeto de la investigación	4) Los procedimientos
2) La finalidad	5) El material
3) La fuente de información	6) El costo y su financiación

El objeto de la investigación

Con el término de objeto de la investigación, nos referimos al hecho o fenómeno que vamos a observar y registrar numéricamente. En una investigación de salarios: el salario; en un censo de población: los habitantes; en una estadística industrial: la producción, en una información de créditos: las operaciones crediticias.

En el proyecto, al analizar qué es o qué vamos a investigar, nos proponemos:

a) Definir el objeto o finalidad de la investigación
b) Determinar su naturaleza cuantitativa y cualitativa
c) Determinar la posibilidad de su investigación, y
d) Limitar el objeto investigable

El proyecto de una investigación estadística tiene que atender los fines que se persiguen con la investigación, porque así como sean éstos, habrán de ser los demás puntos del proyecto: la fuente de información, los procedimientos, el material y en consecuencia última, el presupuesto.

Dos ejemplos, el uno bioestadístico y el otro económico, aclaran la necesidad de establecer, a priori, el punto de mira que se persigue con la recopilación estadística.

Se organiza una estadística de natalidad. Se puede aspirar a muy diferentes fines. Por ejemplo, se trata de conocer la proporción de los sexos con un propósito biológico, o si no (y este es otro fin), con un propósito demográfico de la población que entra a la vida. En el supuesto, el fin nos obliga a preguntar por el sexo de los hijos nacidos, muertos o vivos; en segundo caso, nos bastará con indagar los nacidos vivos.

Se prospecta una estadística de salarios. El objetivo de esta investigación, entre otros, puede ser, calcular el salario medio general simplemente, o bien obtener el salario más frecuente en cada uno de los oficios de cierta rama industrial.

a) Definir el objeto o finalidad de la investigación: vamos a emprender un censo de edificios: ¿qué es un edificio para efectos de la recolección de los datos? En una investigación de salarios: ¿a qué salarios nos referimos? En una investigación industrial, ¿qué hemos de entender como "industria", como "empresa", como "fábrica"?

   ¿Indagaremos cualquier industria, toda clase de salarios, toda especie de edificios? o por el contrario, exigiremos determinados requisitos particulares? Lo anterior nos permite identificar *A quién* se investiga o sea el elemento o la unidad. Veamos algunas formas de definirlas:

   Unidad de vivienda: "Es la unidad final considerada como una estructura separada e independiente destinada para alojar una o más personas, con acceso directo a la calle, o a un pasillo o escalera, de tal forma que sus ocupantes puedan entrar o salir sin pasar por los cuartos de otras unidades de vivienda".

Hogar: "Constituye la unidad final de observación y se define como una o más personas bajo la autoridad de un sólo jefe que ocupan en común la totalidad o parte de una vivienda y alimentos, a menos que circunstancias especiales lo impidan (por ejemplo el horario de trabajo) y no se aprovisionen conjuntamente de otros bienes y servicios de primera necesidad, sin interesar el parentesco entre ellos".

La primera cuestión por tanto, al tratar el objeto de la investigación en una prospectación estadística, es la fijación precisa del concepto de lo que se aspira a indagar. Decir con claridad y exactitud lo que la Estadística va a recoger. La unidad o elemento de investigación debe ser: *clara, adecuada, mensurable y comparable.*

b) *Determinar la naturaleza cuantitativa y cualitativa del objeto de la investigación:* o sea, establecer los aspectos cuantitativos del fenómeno que se pretende investigar y consecuencialmente pueden registrarse por medio de números, y otros que por ser cualitativos se han de recoger mediante anotaciones literarias, bien de su existencia o de su ausencia (si, no), bien de su clase o categoría (1a, 2a, 3a), su origen (importación de Estados Unidos, Inglaterra, Argentina), su destino (créditos a la agricultura, a la ganadería, a la industria manufacturera), su marca (café brasilero, colombiano, mexicano), su grado de bondad (bueno, regular, malo), etc.

c) *Determinar la posibilidad de investigación*: de los aspectos establecidos en el considerando anterior, no basta conocer el objeto a investigar, y sus modalidades cuantitativas y cualitativas; se necesita examinar si una y otras pueden ser conocidas con precisión, si se exteriorizan, si pueden contarse, si admiten su existencia y su intensidad.

En un proyecto de estadística sobre la morbilidad de una población, ,cómo comprobaremos la presencia o la falta de la enfermedad?; el ahorro popular no depositado en cuentas de ahorros sino atesorado personalmente, ¿cómo podremos conocerlo para registrarlo estadísticamente?; el alcoholismo en un barrio obrero, divisas extranjeras en las transacciones en la bolsa negra, la declaración de las edades, los consumos familiares de víveres, los precios de las mercancías de importación, se exteriorizan de alguna manera, constan en algún sitio, como para poder anotar el hecho en su cantidad o en su calidad o en ambos aspectos?

d) *Limitar el objeto investigable:* conocemos lo que vamos a investigar, sus modalidades cualitativas y cuantitativas y la posibilidad de registrarlas. A continuación, el proyecto tiene que delimitar la investigación. Todo objeto a investigar en si, en el tiempo y en el espacio y a veces en el número, debe tener sus méritos.

Por imposibilidad práctica o por ser innecesaria la observación completa, la estadística reduce sus trabajos a un doble aspecto:

Primero: Limitando el objeto mismo de la investigación y

Segundo: Limitando el campo de la investigación

Limita el propio hecho o fenómeno que investiga, cuando se concreta a ciertos aspectos o modalidades, prescindiendo de todos los demás. Y ello por no serle posible abarcar todos los caracteres del fenómeno.

Estadística de salarios. Limita a las tasas de salarios en dinero, en los hombres de 18 años o más, en la industria de transformación. Se prescinde de los salarios percibidos por el obrero en trabajo extra, por trabajos reducidos, primas de días feriados, bonificaciones, etc.; se prescinde de las tasas de salarios en la agricultura y ganaderas; en la minería y en los servicios se hace caso omiso de los salarios femeninos y de los masculinos, en los trabajadores menores de 18 años.

Estadística de créditos bancarios. Se limita a los préstamos y descuentos comerciales e industriales que conceden los bancos. No se recogen los préstamos hipotecarios o de "amortización gradual", se pregunta la clase de actividad económica, a que se destina, sobre si son personales o con garantía prendaria, cuantía, interés o descuento y sus plazos de devolución. Pero se dejan de indagar muchísimos otros aspectos de la empresa a la que se le concede el préstamo, del particular prestatario, del destino geográfico, del capital prestado, etc. Al admitir como necesarias unas características del hecho que vamos a investigar y rechazar otras, el proyecto está limitando al objeto mismo de la investigación: los créditos bancarios, los salarios, en los ejemplos propuestos.

La *Estadística* puede interesarse por conocer la producción algodonera del país, pero no necesita ni le es posible llegar al detalle de los kilogramos recogidos en cada hacienda o región, le bastará con estimar aproximadamente las toneladas o incluso los miles de toneladas. La *Estadística* nos proporciona el dato sobre la cotización de las monedas extranjeras hasta las milésimas o diezmilésimas del peso; aquí la aproximación no puede ser mayor del tercero o cuarto decimal de la unidad monetaria.

Según la clase de investigación de que se trate, la exactitud podrá ser mayor o menor, y ésta deberá fijarse en el Proyecto de Investigación. Es una forma de limitar el objeto de la recolección estadística.

Estudiamos ahora la limitación del campo de la investigación. Comprende cuatro aspectos: el *espacio*, el *tiempo*, el *número de casos* y la *combinación de los tres anteriores*.

Limitación del espacio o cobertura geográfica. Toda estadística nacional presupone una limitación geográfica de orden internacional. Una limitación internacional, e incluso más de una estadística municipal, está circunscrita a determinados barrios del área urbana del municipio. Los censos nacionales, por su propia naturaleza, no están limitados en el espacio nacional. Pero, por ejemplo, la estadística sobre los índices de precios al consumidor, sólo se realiza en un reducido número de ciudades.

Como siempre, la limitación de este género se efectúa por falta de recursos para recoger los datos en todos los lugares, o por ser innecesaria la recopilación completa para efectos de la observación y conocimiento suficiente del fenómeno que investiga la estadística. El proyecto ha de concretarse expresamente los lugares que va a contemplar la investigación.

Limitación del tiempo. También el tiempo se debe establecer; no todas las investigaciones pueden ser continuas; algunas se repiten, son periódicas; otras se celebran una vez en forma accidental y no vuelven a realizarse. En el proyecto se fija la continuidad, la periodicidad o la accidentalidad de la investigación. En los dos últimos casos se trata de una limitación del tiempo.

El censo de población, como los censos industriales o agrícolas no se celebran todos los días, ni todos los meses ni siquiera es necesario que sean anuales, basta y así está convenido internacionalmente, que sean decenales.

La criminalidad, por el contrario, tiene lugar en cualquier momento con características muy variadas y su registro policial por otra parte es imprescindible. La estadística de delitos, muy bien, podrá ser continua. La estadística de mortalidad, como otras demográficas, deberá ser permanente.

Esta limitación cronológica atemporal, se refiere a lo que en técnica estadística se denomina *"el momento estadístico";* la época, la fecha de una investigación.

Esta limitación, en la mayor o menor periodicidad, depende de la variabilidad del fenómeno que se investiga. A mayores modificaciones en el fenómeno, más breve el tiempo transcurrido entre una y otra investigación; los hechos de naturaleza estática se investigan muy de tiempo (estado de las edificaciones, higiénico, sanitario de la población, características de la industria, de la ganadería, de las explotaciones agrícolas, mineras, de la enseñanza); los hechos de grandes fluctuaciones se investigan en cortos intervalos de tiempo (cotizaciones de bolsa, precios de víveres, nacimientos, producción industrial, comercio exterior, costo de vida, ingresos y egresos del Estado, etc.).

El proyecto dirá el período o el carácter continuo o accidental de la investigación.

Limitación del número de casos. La observación, a veces, puede abarcar todos los casos que se presentan; en otras ocasiones no es posible, y por lo tanto, hay que limitar su número. Los cálculos y estimaciones estadísticas parten generalmente del estudio de un número de observaciones. Así se precede comúnmente en las investigaciones científicas de tipo médico, antropológico, económico, sanitario, social, etc.

En la recolección de precios, no se hace de todos los artículos que se ofrecen y demandan en el mercado, sino de un número parcial de ellos; en la estadística de los salarios se investiga a un grupo de obreros o empleados.

El proyecto también tiene que fijar el número exacto o aproximado de casos que deben indagarse.

Limitación coordinada del espacio, del tiempo y del número de casos: por último y es lo que más frecuentemente sucede, en un proyecto de investigación estadística hay que limitar a la vez, el espacio y el tiempo, aquél y el número de casos, éstos y el tiempo, o los tres juntos.

Cuando en un proyecto se recomienda cualquiera de esas reducciones de la investigación, se plantea el problema de la elección de la época, del lugar y de las propias unidades que sometemos al examen estadístico.

Sólo existe una norma general, para efectuar esas elecciones: los casos, los lugares y las épocas designadas deberán ser normales. No deberán encerrar características, aspectos excesivamente intensos o excesivamente débiles del hecho o fenómeno de que se trate.

Los procedimientos prácticos son dos:

Elegir al azar sin buscar intencionadamente los casos extremos, los lugares o las fechas en que el fenómeno presente una situación anormal.

Si se pretende realizar un análisis sobre las condiciones del trabajo en la industria manufacturera de un país, y decidimos la investigación, por ejemplo, a doscientas veinte fábricas y talleres, diez de cada departamento, efectuaremos una elección al azar; asi del Directorio de la Industria Manufacturera entresacamos de cada departamento las diez primeras empresas fabriles según el orden alfabético de sus razones sociales; probablemente, entrarán pequeños talleres y grandes instalaciones, producciones muy diversas, condiciones de trabajo distintas, formas de asociación de los propietarios, en fin, las más variadas circunstancias y no empresas de igual naturaleza.

Elegir por casos extremos, en los cuales el hecho social económico tiene intensidades máximas y mínimas con el objeto de buscar lo normal en el punto medio de los límites extremos.

En el ejemplo anterior se pueden recoger las informaciones de la gran industria por un lado, y de la pequeña por otro. De esta forma conoceremos los casos y situaciones extremas, y podremos de ellos deducir los valores y situaciones intermedias. Es una manera indirecta de hallar lo normal o típico partiendo de lo anormal o atípico.

Las fuentes de información

Hasta ahora hemos analizado lo que se desea investigar y el por qué de esa investigación.

A continuación, es necesario preguntarse: ¿dónde es posible recoger alguna manifestación externa del fenómeno económico?, ¿en dónde podremos registrar datos sobre los saldos de moneda depositados en los bancos?, ¿en dónde cifras sobre importaciones?, ¿en dónde información sobre precios de víveres, etc.?

El lugar, el documento, la oficina, la persona que constituye la fuente informativa podrá ser:

- **Directa e Indirecta**

Investigaciones directas e indirectas. Directamente recogemos datos de un acontecimiento de cualquier índole, cuando acudimos a él, lo observamos y anotamos su presencia o su ausencia y su intensidad mediante un número. Nos proponemos conocer el canje de cheques entre los bancos, acumulamos los cheques, los contamos en su número, sumamos su valor y lo anotamos. Hemos observado y anotado directamente el mismo hecho cuyo conocimiento deseábamos.

Indirectamente investigamos en estadística, cuando se recurre a un hecho distinto del que nos interesa, para después deducir de éste el valor del que en definitiva deseábamos conocer. En estadística agrícola frecuentemente se emplea este procedimiento. Del rendimiento de determinada hectárea se deduce la producción total. De la cosecha probable se deduce la siembra.

De la transmisión de herencias durante la vida media de un hombre en un país, el capital y la renta nacional; de los consumos, la población; de los precios del café hemos visto deducir estimativamente la importación y el consumo de determinados artículos suntuarios. Son inducciones lógicas, cálculos aproximados, estimaciones que constantemente se realizan en los negocios.

Se da el caso de que las cifras obtenidas indirectamente son más necesarias y frecuentes en los países de pequeño desarrollo estadístico por la misma falta de muchos renglones desconocidos. Y su proyección y aproximación suficiente depende precisamente de la abundancia, extensión y calidad de las estadísticas directas. Puesto que la garantía de esta clase de cálculos se funda en varias pruebas por varias estadísticas directas.

Se refieren sobre todo, a la investigación de elevado costo que directamente no se puede realizar, las que por su naturaleza reservada, consumos de bebidas alcohólicas, vicios sociales, la pregunta directa se hace difícil.

Muchas veces, se estima y otras se calcula. Se estima subjetivamente cuando por ejemplo, un agrónomo tasa la cosecha, valora las tierras. Se calcula objetivamente cuando, por ejemplo, el demógrafo nos da la cifra de habitantes en una fecha intermedia a dos censos directamente obtenidos; cuando los consumos de cierta población se deducen de los consumos directamente investigados en un grupo parcial de consumidores; cuando de las necesidades de materias primas de las grandes empresas se calcula la cantidad necesaria para todo un pais; por generalización de una parte al todo.

También, en caso de pasar a la intensidad de un hecho partiendo de la intensidad de otro; el consumo de carne y leche a la población; de la importación de autos, a las necesidades de llantas; de la importación y exportación de libros, a la frecuencia de la lectura; del impuesto a la cerveza al consumo; de la pluviosidad a los precios, etc. Para que este cálculo dé resultados suficientemente buenos, se requiere imprescindiblemente que entre el hecho conocido y el desconocido exista una simple relación, cierta proporcionalidad conocida o una dependencia de causa a efecto, de ésta a aquélla.

Llamamos fuente de información estadística *directa* allí donde el hecho económico se produce. Son ejemplos, el hogar de la familia obrera, para investigar sus consumos y el costo de su vivir; los peajes, para obtener los datos de la circulación por carreteras; en las notarías para determinar el número de nacimientos, defunciones, matrimonios y otras informaciones estadísticas; los balances comerciales, para conocer los resultados de los ejercicios semestrales; el taller, para registrar el tiempo empleado en cada operación productora.

Son fuentes de información estadística *indirecta,* aquéllas donde el hecho se manifiesta indirectamente, en donde se refleja. Las rentas departamentales para determinar el consumo de bebidas alcohólicas; la policía para indagar datos sobre criminalidad; las listas o nóminas de salarios para obtener información de las cesantías de éstos; el degüello del ganado para estimar la población consumidora; los manifiestos de importación y exportación para averiguar el consumo interno.

En el proyecto hay que fijar concretamente la *fuente* o *fuentes* de información que conviene o se necesita adoptar.

Las mejores son las *fuentes directas*. Pero no siempre son posibles. Una estadística de los delitos contra la propiedad no podría organizarse de forma que el registro estadístico se efectúe en el lugar del robo. Necesariamente la fuente informativa ha de ser la comisaría o inspección de policía, donde el hecho se registra.

De aquí nace la desconfianza sobre la estadística, por la diferencia en los registros: una persona que ha sido atracada, violada, atropellada, etc., cree que denunciando el hecho no le representa ningún beneficio, sólo perderá tiempo, de ahí que los resultados publicados no muestran la magnitud del problema, sino parte del mismo, lo cual origina la idea de que las estadísticas son mentirosas, en realidad se debe decir, que son deficientes, debido al no registro de lo que sucede.

Cuando sea posible, debe emplearse una fuente directa. Cuando no, se empleará una indirecta. Frecuentemente éstas se usan como complementarias de las primeras, y a efecto de establecer un control sobre las directas.

Algunos autores clasifican la fuente de información como *Primaria*, cuando se obtiene directamente de la investigación, realizada a través de una encuesta, y *Secundaria*, cuando se trata de información complementaria, publicada por la misma institución o cualquier otra.

También la clasifican en *Interna* cuando la información se produce dentro de la institución, *Externa* en el caso de ser obtenida por fuera.

La anterior clasificación puede dar lugar a confusiones en cuanto a la denominación de directa o indirecta.

Los procedimientos de investigación

¿Qué características debe reunir la investigación? ¿bajo qué formas haremos la recolección de los datos?

En este aspecto del proyecto, comúnmente los autores incluyen varios considerandos que se ajusta a estos interrogantes: ¿cómo investigaremos?, ¿quién debe hacer la investigación y por qué sistema?

¿Cómo debe realizarse una investigación? Es decir, ¿cuáles deben ser las normas generales y particulares de la estadística que proyectamos?

Suelen resumirse en estos puntos:

- *Claridad y publicidad*. Toda la investigación, en todos sus pormenores, debe ser clara y conocida por todos los que vayan a intervenir: observadores y observados.

  La claridad se hermana con la comprensión. La estadística, primero en el período de investigación, luego en el de elaboración y más tarde en presentación y el análisis, debe ser entendida por todo el mundo: investigador, investigado, empleado que elabora los datos y el público que va a leerlos y quiera utilizarlos.

- *Sencillez*. Sencillez en todo. En los formularios, en las instrucciones, en el proyecto, en la finalidad, en los cuadros de exposición, en los comentarios literarios de análisis, en las operaciones de cálculo, en los gráficos. La sencillez es conducción esencial de la claridad; difícilmente se dará aquélla sin ésta.

- *Utilidad*. Toda estadística que se inicie debe tener alguna aplicación práctica de interés. De otro modo nos hallaremos ante una acumulación molesta de datos que perjudicarán, por muchos motivos, a la información realmente aprovechable, a los negocios y a la política económica de un país.

¿Cuándo deberá efectuarse la recopilación estadística? En el denominado “Momento Estadístico” Una estadística podrá ser *continua*, *periódica* o *aislada*. En el primer caso, nada tenemos que decir respecto del momento estadístico apropiado, porque son todos. Es el caso de la estadística de criminalidad, constantemente funcionan las comisarías de policía recogiendo los datos sobre los delitos que vayan denunciándose.

Pero la mayoría de las estadísticas económicas y demográficas, son más o menos periódicas, se producen en intervalos grandes o pequeños, según el hecho de que se trate; es la cuestión del tiempo limitado. Se le denomina como *series cronológicas o series de tiempo*.

Ahora se nos presenta el problema de elegir ese tiempo; el período, la duración de la investigación aislada y la fecha de ésta o fechas de aquéllos.

Las investigaciones pueden ser, según la clasificación de Conrado Gini y de otros autores: *ocasional, periódica* y *continua*.

- ❑ Es *ocasional* una recolección de datos en circunstancias extraordinarias, cuando eventualmente se presenta un problema, o se agita su solución. Una investigación del *costo de la vida* o de los salarios cuando se plantea una huelga; cuando se denuncia el incumplimiento de la legislación laboral; al discutirse reforma de la tributación; del proteccionismo aduanero, etc. Se ejecuta la indagación numérica en esa ocasión y luego no se vuelve a efectuar. Son investigaciones *intencionales* o *aisladas*.

- ❑ Es *periódica* aquella investigación que se repite de tiempo en tiempo, en lapsos regulares. Los censos en períodos decenales; la estadística industrial, con periodicidad anual; la del comercio exterior, mensual; los precios de los víveres, semanal; semestral, la estadística de los dividendos de las sociedades anónimas; diaria, la información de los dividendos de las sociedades anónimas, la información de los precios del café, del ganado sacrificado en el matadero de Bogotá, D.C. y de las cotizaciones de las divisas extranjeras; dos veces al día, las estadísticas de cotizaciones y operaciones en la Bolsa.
- ❑ Es *continua* la investigación automática, es decir, la que se produce sin interrupción. Las informaciones demográficas (la natalidad, la mortalidad, la nupcialidad); las de criminalidad, las del tráfico por carreteras y ríos, las de los transportes férreos, de ciertos espectáculos (cines), de artículos de consumo.
- ❑ Es registro permanente, a medida que el hecho tiene lugar.

Son normas generales, puesto que particulares no se pueden señalar, sino en la simple discusión concreta de cada estadística, son:

- ❑ Periodicidad y duración suficiente para observar el hecho o fenómeno.
- ❑ Fechas normales, cuando el hecho o fenómeno presente características habituales, corrientes o cuando no sea ni muy intenso ni muy débil.

Por ejemplo, una investigación sobre los ingresos y gastos familiares en Europa tiene que durar todo el año porque las cuatro estaciones climáticas afectan el presupuesto del hogar, cuatro características consumidoras y cuatro gastos distintos; en el trópico se considera suficiente un mes de duración, porque la variabilidad de las compras es pequeña en los cuatro trimestres del año.

Los fenómenos que cambian mucho y en cortos lapsos, exigen una periodicidad de investigación grande, las cotizaciones en la bolsa se registran mañana y tarde. Las temperaturas y demás condiciones meteorológicas, tres y cuatro veces al día, cuando no continuamente. Pero los precios de las mercancías al por mayor suelen investigarse semanalmente. La estadística bancaria se publica de un mes a otro, como la de la producción minera y del comercio exterior. Las modificaciones entre dos semanas o entre dos días no son tan grandes como para anotarlas.

Los datos sobre el estado de desarrollo de la educación se conocen anualmente. En períodos semestrales los balances de las sociedades anónimas. Decenales, los censos de población.

Las estadísticas *discontinuas*, pueden tener una mayor o menor periodicidad en función de su mayor o menor variabilidad de fenómenos que se alteran mucho, frecuentes períodos; en los que se alcanzan a notar cambios muy de tarde en tarde, períodos bastante separados. Pero además, el "momento estadístico" encierra la elección concreta de la fecha o fechas. Y decimos que ésta debe ser un período normal. Investigar la ocupación obrera, por ejemplo, cuando la actividad del trabajo se halle en su grado habitual y no cuando por un motivo extraordinario y pasajero haya un número excesivo de desocupados o una actividad también excesiva.

El costo de vida no podría investigarse únicamente durante el mes de diciembre o el de enero, porque el último mes del año por exceso de gastos y el primero del año por defecto de gastos, serían notablemente anormales.

Ni el censo de población podría levantarse el 24 o 25 de diciembre, porque un núcleo no despreciable de gentes, se halla en esas fechas en un lugar fuera de su habitual o legal de residencia o domicilio.

Por último, no hay que olvidar en el proyecto de investigación el período preliminar de preparación, que proceda a la recopilación de los datos. Su iniciación y duración deben determinarse de antemano. Por tal motivo se recomienda la elaboración de un calendario de trabajo que incluya a cada una de las diferentes etapas de investigación con sus fechas respectivas de iniciación y terminación.

Todos estos elementos intervienen en los puntos siguientes: el personal, el material, la propaganda y el costo.

¿Quién debe hacer la investigación? La pregunta de quién debe hacer la investigación comprende un doble aspecto: primero la entidad u organismo, segundo el personal. Ambas hay que fijarlas en el proyecto.

Entidad: pueden ser instituciones u organismos públicos oficiales o privados, particulares, de los gobiernos o de las personas o asociaciones particulares.

Ambos tienen ventajas y ambos, unos frente a otros, tienen inconvenientes, veamos:

Principales desventajas de las entidades públicas:

- ❑ Infunden recelo y suspicacias sobre la finalidad que los gobiernos persiguen con la estadística: temor fiscal, temor policial, etc.
- ❑ Frecuentemente el costo de las investigaciones es más elevado que en las entidades privadas.

Principales ventajas de las entidades privadas:

- ❑ Son más indicadas para las encuestas ocasionales de tipo científico, problemas sociales, médicos, higiénicos, etc.
- ❑ Ofrecen al público más confianza que los gobiernos: los datos se conceden con mayor sinceridad.

Los inconvenientes de las entidades privadas corresponden a las ventajas de las instituciones públicas: menos recursos económicos, sobre todo en los países de escaso desarrollo de las investigaciones científicas; más reducida organización administrativo-territorial y falta de poder coactivo.

En las naciones de gran proceso estadístico, como Estados Unidos, Francia, Italia, Inglaterra y Alemania, son numerosas las entidades privadas que realizan investigaciones de esta clase. En América del Centro y del Sur, todavía es deficiente el desarrollo de la investigación no oficial.

Desde el punto de vista de la Administración Pública no es discutible la conveniencia y la necesidad de que los gobiernos dispongan de un servicio de Estadística Nacional y Seccional; las estadísticas generales o parciales de los países sólo pueden llevarse por los gobiernos, máxime teniendo en cuenta que para el público las informaciones deben ser gratuitas y generalmente conocidas.

En la actualidad se vienen realizando encuestas a través de los medios escritos (periódicos, revistas), radiales y de TV., algunos mediante entrevistas, otros en forma voluntaria con llamadas al medio de información o escritas, siendo de mayor uso del Internet, telefónicas o del celular, sin embargo los resultados son poco confiables, dado que no hay selección al azar del informante, procedimiento válido para una muestra aleatoria.

Personal. También en el proyecto tiene que tratarse el personal que ha de ser reclutado para la investigación. Se fijará el número, las atribuciones y las categorías de los distintos puestos. Alrededor del tema del personal, plantea el problema delicadísimo de su elección, de su instrucción y de su remuneración.

Nos parece conveniente hacer estas advertencias al respecto:

- La instrucción, requiere cursos de preparación, sobre todo en donde la cultura y formación estadística no sean grandes.
- La elección debe atender a la instrucción, puede ser mediante examen previo, de aptitud y vocación y a la imparcialidad de los aspirantes al puesto de estadística, es decir, que el empleado sea ajeno a los intereses particulares que puedan entrar en juego en las informaciones estadísticas.

Aparte de eso, todo el personal que trabaje en estadística debe ser constantemente controlado en sus tareas, sobre todo cuando éstas se refieren a cálculos, a clasificaciones, recuentos y construcción de tablas. El control también debe existir para el personal de recolectores, con el fin de comprobar si los datos que ellos recogen son ciertos o no.

Desgraciadamente el personal que se vincula, lo hace en forma transitoria, sólo quiere resolver su situación laboral por un tiempo menor, mientras consiga mejores condiciones salariales y de estabilidad.

Sistemas de investigación: se distinguen varios procedimientos de investigación:

- Las recopilaciones *automáticas* de datos por declaración espontánea del sujeto de la investigación. Se refiere generalmente a las investigaciones por medio de registros o inscripciones obligatorias; como en el caso de las estadísticas demográficas (natalidad, nupcialidad, mortalidad, migración). Su carácter automático estriba en el hecho de producirse en cumplimiento de una disposición legal o de una costumbre, de una tradición. La estadística del movimiento de divisas extranjeras, por ejemplo, es automática, porque se forma en virtud de la obligación de solicitar la compra de monedas extranjeras y acudir a su cambio o venta a la Oficina de Control de Cambios del Banco del Estado. La estadística del comercio exterior, la de las edificaciones, la estadística de los recaudos gubernamentales por concepto de los impuestos, son otros ejemplos de estadística automática, que automáticamente y sin necesidad de buscarlo, se producen mediante la anotación de los hechos en oficinas administrativas: Recaudación de impuestos, Aduanas, Banco del Estado, Secretarías de Obras Públicas Departamentales, etc.

  Es un buen sistema para recoger informaciones de esta clase, puesto que la obligatoriedad legal o la fuerza de la costumbre influyen en la exactitud sinceridad y generalización de las informaciones.

- Las recopilaciones *intencionales* de datos, obtenidas mediante el empleo de un agente que ex profeso vaya a la fuente de información para registrar los datos: el censo de la población, las encuestas de hogares sobre ingresos y gastos, las encuestas sobre sistemas de trabajo, sobre condiciones de una determinada industria, etc. En ellas se utilizan los cuestionarios, boletines, formularios, cédulas o patrones de inscripción de los datos.

La organización y establecimiento de este sistema de compilar datos se efectúa con una finalidad propiamente estadística, a diferencia de las recopilaciones automáticas en las cuales el motivo de la declaración no es propiamente estadístico. El registro civil, demográfico, la inscripción catastral, de la propiedad, de las sociedades comerciales, etc. La estadística, en estos casos, lo que hace es adjudicar a los registros de otra naturaleza la obtención de estadísticas, aprovechando los servicios del Estado.

Las investigaciones sistemáticas, por ser exclusivamente estadísticas, ofrecen las ventajas de poder organizarse conforme a una técnica completa y abarcar todos los aspectos que requieren las necesidades del servicio de estadística.

- Las recopilaciones *combinadas automáticas e intencionales*. En ellas se añaden a las informaciones del servicio ajeno al de la estadística, las preguntas que exige una buena investigación estadística. En el registro eclesiástico se anotan automáticamente los matrimonios y los nacimientos, se aprovecha esta circunstancia para poder indagar algunas cuestiones que al registro de la Iglesia no le interesarían.
- Algunos estadísticos, clasifican las investigaciones, en *extensivas* e *intensivas* y éstas en *monográficas y encuestas*. La diferencia se basa en el número de datos que abarcan y en los detalles a que llegan.

Cuando la recopilación se limita a las preguntas esenciales, a los datos muy generales, tal como ocurre en los censos, en la estadística bancaria, en la del comercio exterior y en la fiscal; y su extensión es completa por proceder de todo el campo de la investigación: se denomina *extensiva*. La mayoría de las recolecciones de datos periódicos y continuos son de esta naturaleza.

Cuando, por el contrario, la investigación se detiene en aspectos muy particulares del hecho, se hacen muchas preguntas y muy minuciosas y cuando se reduce a un sector del campo de la investigación, se le llama *intensiva*. La mayoría de las investigaciones ocasionales son intensivas.

Ejemplos reales de las primeras son los censos, la estadística de precios y salarios, las monetarias, entre muchas otras. Ejemplos de la segunda, intensivas, son las encuestas campesinas, de costo de vida, de alimentación, etc.

- *Investigaciones completas, incompletas y más que completas,* continuamos la división didáctica del profesor Conrado Gini, división necesaria a efectos de la enseñanza del método estadístico, y más que nada para poder efectuar una crítica y corrección técnica.

Son *completas*, las investigaciones que recogen todos los casos. Indagan todo el campo de la observación; cuantas unidades de las operaciones bancarias, (es otro ejemplo de información completa), entran en todos los balances de la banca. La producción de sal, la de cemento, los transportes aéreos, la estadística fiscal, son otros tantos ejemplos de estadísticas completas; se tomarían todas las fábricas de cemento, todos los aeropuertos y todas las oficinas que tenga el Ministerio de Hacienda.

Son *incompletas*, las investigaciones que sólo atienden a una parte de las unidades estadísticas, bien por no ser posible recoger la totalidad de los datos, o bien por no ser necesario para el fin que se persigue. Si la estadística incompleta no es representativa del conjunto, no es típica como para generalizar los resultados parciales al conjunto de los casos; su valor es nulo, o bien pequeño. En caso contrario, cuando el circulo estudiado numéricamente puede sustituir al total, la estadística incompleta es de extraordinaria utilidad.

Son *más que completas*, las investigaciones que repiten varias veces una misma anotación, a veces intencionalmente como control, otras inadvertidamente, dando origen a un error.

Pero no faltan ocasiones en las cuales el estadístico voluntariamente hace una misma pregunta en un formulario de dos formas distintas para verificar la una con la otra: salario medio de una empresa por un lado, y número de jornadas y valor total de los salarios, por otro; el primer dato se deduce de los segundos. Edad y fecha de nacimiento, etc.

- ❑ *Investigaciones representativas y no representativas*. En la crítica de este problema hemos de ocuparnos de la responsabilidad de un grupo, de una muestra estadística. Bástenos advertir ahora que la condición esencial para que un grupo limitado represente el total de casos, es que el grupo parcial contenga elementos de la misma característica del total y con iguales frecuencias.

La encuesta de costo de vida obrera hecha sólo entre los obreros de una fábrica de cervezas, no serviría como muestra de toda la clase obrera de una ciudad, porque los salarios de aquella empresa no reúnen las mismas características, en sus ingresos, gastos, prestaciones sociales, clase de trabajos, oficios, residencia, que se observan en toda la masa de trabajadores de la ciudad. Incluso, tal vez, la composición demográfica de las familias de aquel sector de obreros sea distinta de la composición familiar de todo el obrerismo del país. Nos hallaríamos ante una estadística poco representativa.

- ❑ *Investigaciones obligatorias voluntarias*. Son obligatorias en cuanto a la rendición de datos; por lo general, las que organizan los gobiernos, porque valiéndose de la atribución legal para exigir las aclaraciones, así se dispone. El registro civil, base de la estadística demográfica; la declaración de las importaciones y exportaciones, son obligatorias, como la mayoría de las estadísticas que llevan las entidades que el Estado ha creado para ello.

Sin perjuicio de esto, los organismos del Estado también pueden y suelen organizar investigaciones voluntarias, ante las cuales el público puede contestar o dejar de hacerlo.

Las recopilaciones *voluntarias* de datos, frecuentemente se llevan a cabo por las instituciones privadas y se refieren comúnmente a las monografías y encuestas científicas. La radio, prensa y las revistas suelen invitar a sus lectores a opinar sobre algunos problemas candentes o a declarar un dato de su vida o negocio particular.

Pues bien, de estos sistemas, el proyecto, para el caso particular, tendrá que decir cuál interesa más y cuál debe emplearse.

Algunos autores consideran tres sistemas importantes de recolección: *correo*, *entrega personal del cuestionario* y *la entrevista*; otros sistemas de menor importancia corresponden a: *internet, teléfono*, *fax* y *panel*.

Todos ellos presentan ventajas y desventajas. Se considera que la *entrevista* presenta más ventajas que los otros sistemas, siendo la más importante: mayor número de cuestionarios recolectados, mayor número de respuestas, permite aclarar el objetivo de la investigación y las dudas del informante. Entre las desventajas se menciona: mayor costo, más tiempo para la recolección y una cantidad mayor de personas, además, la influencia negativa que puede presentar el entrevistador.

La organización. Es éste otro de los problemas que debe estudiarse en un proyecto; qué forma se va a dar a la organización de la estadística de que se trate. Existen dos tendencias: *centralista* y *descentralista*.

La centralización. Es incuestionablemente necesaria en el orden técnico. La técnica, las normas generales, la dirección de toda investigación debe ser central y general para todo el territorio, para el campo de cada investigación y para todos los organismos inferiores y el personal que intervenga.

Es el único derecho y medio de conseguir unidad en los sistemas y de poder totalizar los datos de todo un municipio, de todo un departamento, o del conjunto del país. Si se deja a cada sección que realice las estadísticas que quiera y por los medios que se le ocurran, las diferencias regionales impedirán reunir en totales nacionales los datos e incluso harán imposible las comparaciones interregionales.

La descentralización. Debe emplearse en el orden administrativo, en el orden ejecutivo, para seleccionar personal, para cooperar en el presupuesto de la estadística, para aportar el material, para recoger los datos, para controlar la bondad de las investigaciones, para, en fin, servir de intermediaria entre las diversas categorías de dependencias.

Claro está que a la centralización, en ningún caso, puede corresponder uniformidad absoluta en lo esencial y en lo accidental. La dirección central debe tener en cuenta las particularidades regionales del hecho o fenómeno que se investiga estadísticamente, con el fin de ajustar el proyecto a las distintas características geográficas.

El material estadístico

Está constituido por todos los útiles, documentos o instrumentos necesarios para llevar a buen fin la investigación. El proyecto tiene que enumerarlo en sus clases y en su número. Podemos dividirlo en: material *impreso* e *instrumental*.

Material general para la investigación y material para la investigación de que se trate.

Material impreso. Se refiere a los formularios o cuestionarios, boletines, hojas de inscripción, registros, circulares, pliegos de instrucciones.

El formulario, boletín, boleta o cuestionario, que de todas las formas suele denominársele, es el instrumento que sirve para anotar las respuestas o el producto de la observación personal del hecho que se investiga.

Los formularios pueden ser:

- *Colectivos.* Destinados a la inscripción de los datos de varias unidades estadísticas, principalmente cuando esas unidades son personales. El formulario del censo de población es colectivo, puesto que en uno se reúne la información de todos los habitantes de una vivienda u hogar.

Tiene la ventaja de facilitar la recepción de los mismos en la oficina que investiga, puesto que los colectivos son menos en cuanto al número y supone por esto mismo, una economia en la impresión. Exigen, generalmente, un investigador que haga las anotaciones y que lleve personalmente el formulario.

- *Individuales.* Destinados a la inscripción de los datos de un solo elemento o unidad estadística: una persona, una fábrica, un almacén, una sociedad industrial, una nueva construcción, un hospital, una escuela, etc.

Ofrecen la gran ventaja de facilitar el agrupamiento, el recuento o escrutinio, con sólo amontonar los formularios de una misma naturaleza: por departamentos, por clases, por industrias, por países de origen, por países de destino, por sexo, por importancia de un carácter cuantitativo como capital, producción, consumo, importe del salario, etc.

Contenido de un formulario. En general todo formulario debe contener:

- El *encabezamiento* que comprende: el nombre de la institución u organismo que realiza la investigación: Universidad Nacional, Investigaciones Económico-Sociales, Federación Nacional de Cafeteros, Departamento de Investigación de Mercadeo, etc. Un número que lo personifique de los demás. Esta enumeración puede ser doble: 1. Numeración de orden igual para todos los ejemplares de un mismo formulario. 2. Numeración de orden, dentro de un mismo formulario, para cada unidad estadística.

El título del formulario, que nos diga de qué investigación se trata: “Investigación sobre la propiedad”. “Investigación sobre precios”. Además cómo, cuándo y dónde se realizó la investigación.

- El *cuerpo* abarca la indicación de la clase, dirección y nombre de la unidad o grupo estadístico que se investiga. Además las preguntas generales relativas a toda la unidad. No siempre es necesario dedicar la primera parte del cuerpo a generalidades, que tiene por objeto recoger información sobre el nombre del informante, dirección, sexo, ocupación, cuando se trata de personas; o nombre de la entidad, razón social, dirección, actividad económica, etc. De acuerdo al objetivo, se puede prescindir de este tipo de preguntas.

También comprende las *preguntas particulares* relativas a la unidad o unidades examinadas. El cuerpo de un formulario puede disponerse en serie o en cuadro. En forma de serie o cuestionario, una pregunta después de la otra, se emplea frecuentemente cuando el formulario es individual; en forma de cuadro, cuando es colectivo. No es extraña tampoco la combinación de ambos sistemas en un mismo formulario. Incluye preguntas de *control*, *filtro*, *abiertas*, *cerradas* que pueden ser de selección *múltiple* o *dicotómicas*.

Las preguntas pueden ser *cerradas*, cuando se responde si o no; *abiertas*, en el caso de que se pida una opinión; *selección múltiple*, si se da una serie de alternativas para que escoja una o varias; *control*, para determinar la veracidad del informante; y preguntas *filtro*, que nos determinan si se debe dar por terminada la entrevista o si debe pasar a otra parte del formulario.

Las preguntas deben ser: claras y sencillas; utilizando el mismo vocablo del informante; se inicia con las más fáciles y se termina con las más complejas; no se deben hacer preguntas que no vayan a tener respuesta.

Las *observaciones*: en todo formulario conviene destinar un espacio libre al final, en la última columna del cuadro, para que el investigador o el investigado puedan anotar las aclaraciones y observaciones que crean convenientes, como complemento de las preguntas impresas, de por si rígidas.

- Las *instrucciones*: casi siempre, la brevedad y concisión de las preguntas de un formulario exigen explicaciones de su sentido o sobre la manera de dar respuesta. Para facilitar la labor del encargado de llenar el documento, se incluyen algunas instrucciones, bien en el mismo formulario (al pie, en la cabeza o al dorso), con sus correspondientes llamadas entre paréntesis (1), (2), etc., bien en pliego aparte cuando son extensas. En este último caso, ninguna pregunta puede ser omitida, aún cuando no haya necesidad, por la claridad y precisión de la redacción, deberá acompañarse una enumeración metódica en forma de código y con ejemplos: 10, 20, 30, etc.

No hay criterio unificado, sobre el lugar del formulario donde deben colocarse las instrucciones. Algunos consideran que debe ser después del encabezamiento, para obligar al informante a leerlas antes de iniciar la contestación, pero otros, aconsejan ubicarlas ya sea al final o en cartilla separada del formulario aduciendo una mejor presentación.

- La fecha de envío del formulario a la entidad que recoge los datos.
- La firma responsable del investigador o agente cuando lo hay, o del investigado cuando éste informe directamente: propietario, administrador, gerente de una empresa.

También suele incluirse al pie, un renglón para la firma o visto bueno del jefe o coordinador de la investigación, el nombre o firma del entrevistador.

Trazado y redacción de los formularios. Al proyectar una nueva investigación, se debe diseñar el material que vaya a repartirse y someterlo a discusión y aprobación como el resto del proyecto.
Las normas del diseño y redacción son las siguientes:

- Debe ser *sucinto*, es decir, limitado a las preguntas esenciales, justamente las que se necesitan para los fines de la investigación y que en realidad puedan obtenerse de la fuente informativa.
- Debe prescindirse de toda pregunta *indiscreta* que levante suspicacias y temores o que moleste al investigado.
- Debe *ser claro*, fácilmente comprensible por cuantos, de una u otra manera, tengan que manejarlo. No ofrecer dudas sobre la forma de contestar cada pregunta; que no admita sino una sola interpretación.
- *Categoría*, generalmente las respuestas deben limitarse a un número o a una afirmación (sí) o una negación (no), por lo tanto las preguntas deben ser cerradas.
- Debe procurarse evitar los *juicios personales* del investigador o del investigado, como cuando se deja a criterio del calificador juzgar la *importancia* o la *bondad* de un hecho (grande, mediano, o pequeño); (bueno, regular, malo).

Otros aspectos: la cuestión de los formularios trae consigo otros considerandos que deben resolverse en el proyecto. Estos son: *clase de papel, forma de impresión* (mimeografiados, en imprenta), tamaño general y espacios intensos; anulación de casillas en los cuadros por no tener que contestar, inclusión de preguntas de control, diversidad de colores en las tintas, rayados fuertes y débiles, etc. Todos los problemas relativos a las labores de imprenta.

Pero lo esencial es, en la etapa de prospección, calcular el número de formularios que se han de necesitar para la investigación previa y para la definitiva.

Todo formulario antes de imprimirlo, debe ser ensayado con una recolección real de datos. Por muy perfecto que se suponga el modelo en una oficina, la práctica enseña a perfeccionarlo, a corregir la redacción, a suprimir lo imposible e innecesario, a añadir lo que se dejó en olvido. Por otra parte permite establecer el tiempo necesario para su diligenciamiento, la reacción del informante y la claridad de las instrucciones. Lo anterior implica la realización de una *encuesta preliminar o piloto*. El material impreso lo podemos dividir en dos: *material de recolección* y *material* de *elaboración.*

Equipos. En la recolección de datos, a veces, y en la elaboración posterior casi siempre se requiere de muchos instrumentos, aparatos, máquinas y útiles, que quien proyecta debe tener en cuenta, en su número y clase.

En cualquier investigación, generalmente se necesitan máquinas calculadoras, sumadoras, de escribir, tablas de cálculo, computadores con programas estadísticos (Microstat, SPSS, SAS, Estagraf, TSP, etc.) material de dibujo y los útiles propios de toda oficina.

En las grandes investigaciones (censos, comercio exterior, estadística fiscal, de criminalidad, etc.), se necesitan equipos que permitan el procesamiento de grandes cantidades de información.

En ciertas investigaciones especiales, se requiere también de un instrumento especial, sin el cual no se podrían recoger los datos. Si es una investigación antropométrica, requiere escalas cromáticas de la piel, del pelo y de los ojos, cinta métrica, balanza, etc. Si es de una investigación meteorológica, termómetros, barómetros, pluviómetros, etc. Si se trata de llevar en una empresa un registro estadístico del horario de entradas y salidas del personal, se recurrirá a un reloj de marcar. Lo mismo en un aparcadero sobre entrada y salida. En un almacén, la estadística de ventas e ingresos se lleva en la caja registradora y por último, los datos del movimiento de viajeros en los buses se conocen por el control mecánico de los usuarios.

El presupuesto de la investigación

El presupuesto comprende el cálculo del costo y su financiación; las partidas del costo, en general, suelen ser las siguientes:

a. Organización

Estudios preliminares	Asesorías
Trabajos geográficos	Formulación del plan
Plan de propaganda	Impresión del formulario
Adiestramiento del personal	Contratación de servicios auxiliares

Materiales y equipos.

b. Trabajos de campo (recolección)

c . Sistematización

d. Publicación.

Ambos aspectos, la forma de allegar los recursos y la estimación previa de los gastos, constituyen el último punto de un proyecto de Investigación Estadística.

La *preparación*. Una vez redactado el plan de la Investigación Estadística y antes de llevarlo a la práctica, requiere ser sometido a un ensayo denominado preparación.

Todo proyecto de esta clase, conviene ser discutido y aprobado, en último término, por un grupo de entendidos y técnicos en estadística, puesto que, por regla general, los proyectos enteramente individuales no suelen ser perfectos, adolecen de algún defecto que el encargado de realizarlo no alcanzó a ver, la critica de un plan estadístico no sólo debe estar hecha por los peritos en procedimientos estadísticos, sino también por los técnicos en las materias que van a investigarse.

En materia de estadística agrícola, se necesita el parecer de los agrónomos y de los agricultores. En una investigación industrial el de los administradores y empresarios. En una información comercial el del gremio de comerciantes y para una estadística bancaria, o sindical, no puede prescindirse de la opinión de los banqueros o de los sindicatos, respectivamente.

Aprobado el plan con las modificaciones del consejo de técnicos, se inicia la preparación del mismo. Ésta comprende la práctica y el ensayo o prueba en los siguientes aspectos:

- ❑ *Adiestramiento y elección* del personal por medio de cursos más o menos breves y entrevistas.
- ❑ *Ensayo de investigación sobre el terreno*. Esto supone el ensayo correspondiente a la época y fecha del material, del propio personal, utilizado de la fuente de información que genera, de la organización y de los programas a los sistemas que se utilizan.
- ❑ *Elección y preparación de la fuente informativa*. No basta con adiestrar personal de recolección e investigadores. Se requiere, al mismo tiempo, instruir a los sujetos o a las entidades que van a ser investigados, sobre los registros contables o previamente estadísticos que deben llevar, sobre la manera de contestar los cuestionarios, sobre la fecha de visita de los recolectores, sobre la obligación de rendir los datos, etc.
- ❑ *Preparación del material de la encuesta*. La impresión, conexión, el reparto de material, su adquisición cuando se trata de instrumentos, son gestiones que como deben estar terminadas antes de dar comienzo a la recolección propiamente dicha de los datos, entran dentro del período preparatorio.
- ❑ *Propaganda*. La labor de anunciar que se va a proceder a una investigación estadística, las disposiciones del ánimo del público afectado por la investigación, el aislarla de todo

intervencionismo perjudicial del Estado, de todo recelo contra la solicitud de datos o de todo temor por las consecuencias que puedan presumirse; el dar a conocer el interés general de los datos estadísticos, incluso y principalmente para aquél que los rinde, llena esta función de propaganda de las estadísticas, cuya realización encaja necesariamente en la fase preparatoria.

- La preparación y actualización del *listado de informantes* o *marco de referencia* que son objeto de investigación; esta labor debe desarrollarse en su totalidad.

- En algunas investigaciones se requiere del uso de la *cartografía*, por lo tanto es necesario haberla preparado o actualizado.

- *Financiación*. La consecución del dinero necesario para cubrir todos los gastos, incluso la elaboración posterior, no debe dejarse para más tarde de esta etapa de preparación. Desgraciadamente, son frecuentes las investigaciones que no se terminan, aun después de haberse hecho el esfuerzo y el gasto parcial, necesarios en la recopilación de los datos.

- Hay aspectos, además de los anteriores que merecen especial atención por ejemplo: el *calendario de trabajo* que consiste en una descripción de las diferentes etapas de la investigación, indicando la fecha de inicio y de término de cada etapa; en algunos casos se utilizan las barras Gantt que se colorean a medida que se realiza el trabajo. La *formulación de hipótesis* en aquellas investigaciones por muestreo en las que se desea hacer comprobaciones; el *examen de la documentación* requerida antes de comenzar la investigación, siendo posible que la investigación haya sido realizada y no se requiera de un nuevo estudio; en otros casos el proceso será únicamente de actualización, y en situaciones extremas se podrían utilizar las experiencias, corrigiendo los errores cometidos; finalmente la *unidad de investigación*, es decir, a quién va dirigido: una persona, familia, hogar, manzana, barrio, empresa, etc.

La duración del período preparatorio depende de la clase de investigación. Hay estadísticas que requieren un corto plazo para realizarlas (precios, salarios, producción, transportes) y otras exigen meses (censos de población, industriales, ganaderos y agrícolas, encuestas de hogares).

RECOLECCIÓN

Preparada la investigación, comienza la recolección definitiva de los datos.

De esta tarea depende todo el resultado posterior de la estadística. Si está mal hecha, la elaboración resultará incorrecta e incluso imposible de efectuar, y si se realiza, dará origen a un análisis erróneo y a interpretaciones falsas.

La investigación externa de los hechos, recopilación del material básico, puede efectuarse por muy diversos sistemas y con muy distintos fines.

Puede referirse a los datos preliminares o a los datos definitivos. La investigación preliminar coincide con las labores de preparación que hemos visto antes. Su necesidad se deja sentir más cuanto más extensa y complicada sea la investigación. Presta un triple servicio: enseña, controla los resultados y sirve para la organización.

Su valor como medio de control, porque los datos previstos, su distribución y sus peculiaridades indican la calidad del material (formularios), el personal indispensable, la propaganda conveniente, la elección de las fuentes informativas, la manera de repartir y recoger el material, la fecha apropiada, la duración, y los procedimientos a regir en la sistematización de los datos.

La etapa de recolección comprende varios aspectos:

- Distribución del material o instrumento de recolección
- La recolección propiamente dicha
- Control del número de formularios recolectados
- Control sobre la calidad de las informaciones recogidas

CRÍTICA Y CODIFICACIÓN

De la investigación se obtienen los datos individuales, colectivos o aislados. Cada cifra presenta la intensidad de un hecho particular; el salario de un obrero, la producción de una fábrica o los préstamos otorgados por una entidad de crédito.

Es un conjunto de operaciones de revisión y corrección de la información recolectada, que nos permita agruparla y procesarla, de tal manera que nos facilite la elaboración de cuadros, gráficas y análisis, necesarios en la publicación.

Todo ese conjunto de material recogido en los formularios antes de ser totalizado y utilizado, requiere un examen crítico, severo, con el objeto de comprobar si cumple con estas condiciones indispensables; ¿son exactos o erróneos los datos?, ¿son completos, precisos o imprecisos, representativos o no representativos?

> El objeto de la crítica, es clasificar el material primario que precede de la misma investigación, en tres grupos: material bueno, material incorrecto pero corregible y material incorregible o desechable.

La clase y la importancia del error cometido, determinan la admisión o no del dato primitivo resultante de la observación o declaración. Además, se critican los datos numéricos directos para establecer las causas de los errores y modificar o perfeccionar las fuentes y los métodos de observación.

Difícilmente se consigue, por primera vez, informaciones perfectas, en cuanto a la exactitud, a la extensión y consecuencialmente en cuanto a la responsabilidad de las investigaciones parciales.

La necesidad de procesar las informaciones, recogidas en los cuestionarios, ha obligado a traducir las respuestas en códigos. Por ejemplo el código para la pregunta estado civil:

Soltero	1	Divorciado	4
Casado	2	Separado	5
Viudo	3	Otro	6

Cuando el número de respuestas pasa de 9, es preciso utilizar cifras de dos dígitos.

EJEMPLO A

CLASIFICACIÓN POR PROFESIÓN	
Abogados	01
Arquitectos	02
...	
...	
Zootecnistas	20

EJEMPLO B

CLASIFICACIÓN POR CATEGORÍAS SOCIO-PROFESIONALES	
0	Agricultores
00	Agricultores (sin ninguna indicación)
01	Propietarios explotadores
02	Colonos
03	Aparceros

TABULACIONES O PROCESAMIENTO

Puede ser manual, mecánica o computarizada y su elección dependerá:

- ❑ De la cantidad de formularios que se van a utilizar.
- ❑ Del número de preguntas que tenga el formulario.
- ❑ Del tiempo y de los recursos, ya sean financieros o de equipo disponible.

Cuando la tabulación se acuerda desde el principio, como parte integrante de la planeación general de la investigación, es de suponer que todo el proceso será totalmente satisfactorio, lo cual ha sido demostrado por la experiencia.

Del proceso que origina una base de datos, de la cual se obtienen *listados*, son revisados a fin de detectar las inconsistencias que se presenten en los procesos anteriores. Una vez hechas las correcciones, se procede a elaborar los *cuadros*, con el fin de facilitar el cruce y *análisis* de la información, la elaboración de *gráficas*, las conclusiones y las recomendaciones, si las hay.

Hoy en día, el procesamiento de datos, la obtención de resultados, la aplicación de ciertas medidas estadísticas, se realizan con mayor rapidez y confianza si se utilizan paquetes estadísticos elaborados para tal fin. Entre otros, como ya se dijo: el SAS, el SPSS, el Microstat, Minitab, el Stagraf, TSP, etc. Por otra parte, el uso de la calculadora ayuda a agilizar los cálculos en la aplicación de las diferentes medidas, cuando se están desarrollando ejercicios en una clase de estadística o no se dispone del computador.

ANÁLISIS E INTERPRETACIÓN

Esta etapa, se puede considerar como la más importante del informe, ya que el análisis de los datos tendrá que ver con la formulación del objetivo mismo de la investigación y de las hipótesis establecidas; sin embargo, este proceso de análisis tendrá menos dificultad, si el investigador tiene pleno conocimiento de los problemas que son inherentes al planeamiento de una investigación.

En este proceso, se debe considerar la elaboración de distribuciones o tablas de frecuencias obtenidas a través de una sistematización de la información para poder ser presentada en forma de cuadros y gráficos. Con los anteriores resultados, se procede luego a hacer un resumen y a aplicar las diferentes medidas que hemos denominado *estadígrafos o estimadores, puntuales*, cuando son aplicados a las características de las unidades en la muestra o como *parámetros* aplicados en las características de la población, entre los cuales figuran las medidas de *dispersión y los promedios, incluyendo en éstos los porcentajes y proporciones*.

Con las cifras resultantes, se pueden hacer comparaciones con otros estudios, para poder llegar a mejores conclusiones.

De esta última fase de la metodología se puede decir que encierra dos aspectos:

a) Análisis y evaluación estadística de los resultados.
b) Análisis y evaluación técnica de acuerdo con la naturaleza de la investigación.

Estos dos aspectos permitirán determinar el grado de consistencia y confiabilidad de los resultados obtenidos de la investigación.

El profesor Jhon W. Best en su libro *Cómo investigar en educación* nos da una posible guía del análisis, sugiriendo los siguientes puntos:

1. Título
   a) ¿Es claro y conciso?
   b) ¿No promete más de lo que el estudio puede proporcionar?
2. El problema
   a) ¿Se halla establecido con claridad?
   b) ¿Está bien delimitado?
   c) ¿Se reconoce su significado?
   d) ¿Las preguntas son específicas o se han establecido las hipótesis con claridad?
   e) ¿Se establecen supuestos y limitaciones?
   f) ¿Se definen los términos importantes?
3. Revisión de la bibliografía relacionada:
   a) ¿Es de amplitud adecuada?
   b) ¿Se destacan los hallazgos importantes?
   c) ¿Está bien organizada?
   d) ¿Se procura un resumen efectivo?
4. Procedimientos utilizados:
   a) ¿Se describe detalladamente el diseño experimental?
   b) ¿Es adecuado este diseño?
   c) ¿Se describen las muestras?

d) ¿Se reconocen las variables relevantes?

e) ¿Se procuran controles adecuados?

f) ¿Son idóneos los instrumentos de recolección de datos?

g) ¿Se establecen la validez y la fiabilidad?

h) ¿Es adecuado el tratamiento estadístico?

5. Análisis de los datos:

a) ¿Es adecuado el uso de tablas y figuras?

b) ¿Es concisa y clara la exposición del texto?

c) ¿Es lógico y perceptible el análisis de las relaciones de datos?

d) ¿Se interpreta con precisión el análisis estadístico?

6. Resumen y conclusiones:

a) ¿Se replantea el problema?

b) ¿Se describen en detalle los procedimientos?

c) ¿Se presentan concisamente los hallazgos?

d) ¿Es objetivo el análisis?

e) ¿Los datos presentados y analizados justifican los hallazgos y conclusiones?

PUBLICACIÓN

Corresponde a la fase final de la investigación, y con ella se propone hacer llegar a las personas interesadas, el resultado total del estudio, teniendo en cuenta todos los aspectos considerados en el proceso, de tal forma que los datos sean comprensibles, con la correspondiente validez que merezcan las conclusiones.

En términos generales se puede decir que un informe deberá contener:

- Planteamiento del problema
- Objetivo de la investigación.
- Hipótesis que se quieren probar.
- Breve exposición de la metodología adoptada, diseño y tamaño de la muestra. Proceso de selección de las unidades de información y de recolección.
- Se podrá incluir en el informe, copia del formulario utilizado en la recolección, aún relacionando y justificando, en forma muy sucinta, las preguntas que se consideran más importantes dentro de la investigación.

- Descripción de los resultados en forma de cuadros y gráficas, acompañados del análisis y comparaciones obtenidas a través de los datos.
- Conclusiones y recomendaciones. Estas últimas cuando así lo exija la investigación.
- En algunos casos, el informe tiene una parte final, denominada apéndice, en donde se incluyen cuadros más generales, que permiten aclarar o comprobar rápidamente cualquier información más detallada. También se puede incluir documentación complementaria al informe.

MONOGRAFÍAS Y ENCUESTAS

Al dividir las investigaciones en *extensivas* e *intensivas,* indicamos que estas últimas se refieren, ante todo a estos dos métodos de recolección de datos tan en boga hoy y que denominamos *Monografías* y *Encuestas*.

Las monografías se iniciaron en los sistemas de observación social y económica por el Ingeniero y Sociólogo francés Federico Le Play (1806-1882) a mediados del siglo *XIX*. Se aplicó la monografía por el propio Le Play, por sus continuadores en las investigaciones de la situación física, económica, sanitaria, moral, social, etc., de las familias obreras. Luego se extendieron estos mismos estudios a otras clases sociales y más tarde a monografías de los municipios, regiones y talleres industriales.

Como podemos deducir etimológicamente de la palabra monografía, se trata de la descripción de un caso, para nosotros, de una unidad estadística. Desde este punto de vista, estrictamente, la monografía no es investigación estadística, puesto que esta última, por su misma naturaleza, tiene que referirse a un grupo grande, el mayor número posible de casos, salvo que se apliquen métodos de muestreo aleatorio, con lo cual se logra validez de los resultados obtenidos.

Al analizar un solo caso, la monografía se detiene en multitud de aspectos y cuestiones que directa e indirectamente se relacionan o afectan al sujeto u objeto investigado. Intensifica el examen, recoge muchos detalles, se hacen muchas preguntas, que sólo en un estudio profundo interesan. El fundamento científico de este método arranca de la razón de “muestra”, “sampling” en inglés, de la posibilidad de representar un conjunto o masa por un solo caso; es la tipicidad individual. Sus defensores nos dicen que así como el químico toma una cantidad (un litro) de agua para analizar la composición química de un embalse y el ingeniero de minas se conforma con granos de mineral para calificar la riqueza de un filón, el sociólogo, el economista y el hombre de negocios pueden, igualmente, limitarse a una muestra social o económica, a un ejemplo de lo sucedido en la vida del comercio y de las finanzas para luego generalizar al conjunto de lo observado en su caso.

El *estudio de un sujeto elegido como tipo*. Un individuo, una familia, una empresa, una fábrica, un almacén, una hacienda, etc. En realidad el procedimiento monográfico se diferencia, además, de la estadística porque indaga aspectos cualitativos que no se pueden reducir a cifras, sin prescindir de preguntas cuantitativas de números e intensidades, pero también incluye inscripciones cualitativas y descripciones literarias, subjetivas, impresiones de observador o del observado.

Es un sistema de observación directa y de generalización indirecta: de un caso a todos.

Dificultades. Aún cuando se discute el hecho de que un caso, una familia, un edificio, un taller, etc., pueda representar bien todas las familias, oficios, talleres, etc., es evidente que existe un caso más *representativo* que todos los demás; la muestra mejor, el que al sustituir al conjunto comete menor error, compensa exactamente o tiende a comprender los errores, por ser unos por defecto y otros por exceso. El análisis monográfico de ese caso precisamente y no de otro, no deja de tener interés.

Pero la dificultad estriba en determinar de antemano en su conjunto, cuál es ese caso más representativo de la unidad a investigar. El tropiezo del método monográfico, es la elección a priori de la unidad a investigar: ¿Cómo podemos saber que la familia, el taller, el municipio, o la empresa elegida al azar o con orientaciones parciales, representa bien a la colectividad de familias, talleres, municipios o empresas?

Esta crítica, con razón ha hecho que el sistema monográfico puro, apenas o incorrectamente se emplee en la actualidad.

A pesar de lo expuesto anteriormente, es un gran auxiliar de la estadística propiamente dicha. La investigación estadística afecta a un gran número de casos, en sus manifestaciones cuantitativas o cualitativas traducibles a número, y siempre indagando los aspectos esenciales del problema, fenómeno o hecho del mundo económico en nuestro caso. La monografía, si es representativa, confirma las conclusiones de la estadística y las reviste de detalles minuciosos, cualitativos, de descripciones sagaces y de interpretaciones, que ella –la Estadística– por su naturaleza mecánica y estrictamente objetiva, no estaba en condiciones de captar.

DIFERENCIAS DE LA ESTADÍSTICA CON LA MONOGRAFÍA

- La estadística observa y registra fenómenos colectivos, en su conjunto o en su mayor parte posible.
- La monografía, estudia intencionalmente sólo un hecho. La estadística, es exclusivamente numérica. La monografía es descriptiva o literal sin perjuicio de poder emplear el número.
- El análisis matemático y gráfico que emplea la estadística, no caben en la monografía.

La Encuesta. Para subsanar el defecto de la monografía, se comenzó a aplicar un procedimiento llamado por algunos autores franceses “Polomonográfico”. Si un caso es insuficiente, estudiemos varios, si una monografía no basta, recojamos varias; de éstas saquemos la más representativa, a quien le atribuiremos, a su vez, la representación de todo el grupo. De esta forma, además de obtener una sola muestra bien elegida, de un grupo, el grupo parcial nos permite estudiar las variaciones que entre las unidades se presentan. El grupo parcial nos permite conocer lo normal, lo típico y lo anormal o atípico. Luego generalizaremos ambos conocimientos: el normal y el anormal. A este sistema se le llama hoy *encuesta*.

La Encuesta, es pues, una investigación estadística parcial o incompleta cuando recoge datos numéricos y deja de ser una investigación estadística propiamente dicha, aún cuando luego pueda someterse a un recuento o elaboración estadística, cuando se limite a anotaciones literarias. Ejemplo de estas últimas son las que suelen organizar los periódicos y revistas, y muchas instituciones profesionales y técnicas; preguntan el parecer, la opinión, por ejemplo, sobre la situación económica, sobre temas de la política, de los deportes, aficiones y gustos artísticos, etc.

A las encuestas se les suele dividir, en cuanto al organismo que las realiza, en: *privadas y públicas*, y a estas últimas en *parlamentarias y estadísticas*.

Las encuestas *privadas* son muy corrientes en los países anglosajones en donde existe una inclinación grande entre el público a dar respuesta a las preguntas de una encuesta. Son menos frecuentes en los países latinos. La prensa, con el objeto de pulsar la opinión y dar la publicidad y las entidades científicas privadas, con el propósito de alcanzar avances en las ciencias, organizan esta clase de investigaciones.

La distinción de las encuestas *públicas* y *oficiales*, en *parlamentarias* y *estadísticas*, se funda, principalmente, en el hábito de los sistemas democráticos de estudiar la realidad de los problemas sobre los cuales tiene que legislar. Así, el Congreso, Asambleas, los Concejos, encargan a sus miembros realizar encuestas alrededor de los tópicos de discusión. Por lo general los senadores, representantes, diputados, concejales, buscan interrogar a las personas interesadas, a observar visualmente la realidad y luego informan.

La Estadística, por el contrario, recoge datos numéricos, realiza anotaciones. Son ejemplos constantemente repetidos, las encuestas sobre ingresos y gastos que llevan a cabo en casi todos los países, así como otras sobre problemas sociales.

La sociología emplea de ordinario el sistema de observación por encuesta.

En cuanto al procedimiento de recoger los datos y hacer las preguntas de una encuesta, se distinguen dos: *orales* y *escritos*.

Las encuestas parlamentarias, por lo regular son orales. Las estadísticas tienen que ser escritas.

Las escritas, unas veces se organizan repartiendo cuestionarios entre el público en general, para que quien desee las responda.

Otras veces se remiten los formularios ex profeso a determinadas fuentes de información, determinadas personas, entidades, empresas, que se consideran de interés o representativas de su sector. También se utiliza el método de agentes investigadores, (reportajes) que acuden personalmente al campo de la observación con el cuestionario a recoger los datos y opiniones.

Las encuestas suelen ser muy frecuentes en los períodos difíciles, en las crisis económicas o cuando se agudiza un problema social o económico, huelgas, epidemias, plagas en la agricultura, alzas de precios, de empleo, pliegos de peticiones obreras, alzas de impuestos, etc. Por eso resultan investigaciones ocasionales.

Las diferencias con la investigación estadística propiamente dicha, son las mismas que hemos señalado para las monografías, con la particularidad de que en la encuesta se observa un mayor número de casos y suelen emplearse para conocer opiniones de personas o entidades en una materia.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. ¿Cuáles son los nuevos términos aprendidos en este capítulo?

2. ¿Qué significado tiene cada uno de los términos recordados?

3. Ilustrar con ejemplos lo que se entiende por población, población finita, población infinita, muestra, característica, variable.

4. Mediante ejemplos, explicar la diferencia entre la estadística descriptiva y la estadística inductiva.

5. ¿Por qué es útil la estadística en el campo para el cual se está preparando?

6. Dar como mínimo tres ejemplos de fenómenos económicos, considerados dentro del campo de la investigación estadística.

7. Dar tres ejemplos de fenómenos no considerados dentro del campo de la estadística.

8. "La estadística estudia el comportamiento de fenómenos colectivos y nunca de una observación individual". Comentar este principio.

9. Señalar el literal que más concuerde con cada una de las siguientes aseveraciones.

   9.1 Una buena muestra industrial debe ser:

   a) Pequeña.
   b) Grande.
   c) Representativa de la población objetivo.
   d) Incluir sólo los establecimientos grandes.
   e) Incluir sólo los establecimientos pequeños.

   9.2 La clasificación de los establecimientos dedicados al comercio al detal se hace teniendo en cuenta:

   a) Su función.
   b) El valor de las ventas.
   c) El valor de la producción.
   d) El valor del trabajo realizado.

   9.3 Una población infinita es aquélla en donde:

   a) El número de elementos está completamente delimitado.
   b) El número de elementos pasa de 1.000.
   c) El número de elementos está escondido.
   d) El número de elementos no está delimitado.

10. Contestar verdadero o falso, según el caso:

   a) La palabra población, significa en la metodología estadística lo mismo que en cualquier otra disciplina.

   b) La teoría estadística es de naturaleza general y puede aplicarse a cualquier campo del conocimiento.

   c) A las medidas que se obtienen de una muestra, se les da el nombre de parámetros, mientras que a las obtenidas de una población se les denomina estadígrafos.

   d) Se conoce como fuente primaria, aquella información que se obtuvo inicialmente, es decir directamente de la persona, empresas o entidad investigada.

   e) Una muestra aleatoria, es aquélla en la cual ciertos elementos tienen mayor posibilidad que otros de ser seleccionados.

11. Señalar el literal más adecuado para las siguientes observaciones:

11.1 La investigación preliminar permite:

a) Establecer la hipótesis. b) Determinar la muestra.
c) Coordinar el personal de campo. d) Ninguna de los anteriores.

11.2 Antes que nada, la investigación estadística requiere

a) Que exista un objetivo. b) Que se hayan trazado planes.
c) Que se tenga un problema. d) Ninguna de los anteriores.

11.3 El costo de una encuesta, por correo, es generalmente:
a) Igual al de una encuesta por medio de entrevistas personales.
b) Mayor al de una encuesta por medio de entrevistas personales.
c) Menor al de una encuesta por medio de entrevistas personales.
d) Imposible de medir en relación con el costo de una encuesta por medio de entrevistas personales.

11.4 En el diseño del cuestionario, las preguntas más difíciles deben colocarse:
a) Al principio, para salir inmediatamente de la parte más difícil.
b) En el centro para que sean precedidas y seguidas por preguntas fáciles.
c) Al final, luego que se haya establecido un clima de confianza, al comenzar por las más fáciles hasta llegar a las difíciles.
d) Ninguna de los anteriores.

12. Contestar verdadero o falso, según el caso:

a) Código es la representación cualitativa de un hecho cuantitativo.

b) Después de elaborar el formulario se define el objetivo de la investigación.

c) Al recolectar información por medio de entrevistadores se tiene la ventaja de que estos pueden observar el sitio de la operación que se está llevando a cabo.

d) Se conoce como fuente primaria, aquélla que obtuvo inicialmente la información directamente de la persona o entidad.

e) Al diseñar un cuestionario no es de gran importancia la forma de hacer la pregunta, siempre que ésta sea clara.

f) No hay posibilidad alguna de que en una encuesta, por correo, se interpreten mal las preguntas de un cuestionario, siempre y cuando la persona que la reciba sepa leer.

g) El examen de la documentación y metodología se efectúa después de tabulada la información.

13. Se ha dicho que en una investigación se consideran tres etapas; las que a su vez se subdividen en otras fases. ¿Cuáles son? ¿Podría usted reagrupar los titulares de este capítulo en un índice de temas de acuerdo con estas etapas?

14. Mencionar algunos aspectos técnicos y materiales que deben tenerse en cuenta en el diseño de un formulario.

15. Un ejemplo de característica cualitativa pueden ser datos sobre:

a) Salarios b) Pulsaciones por minuto c) Gasto mensual en alimentación.
d) Ocupación e) Temperatura.

16. Una muestra es aleatoria cuando las unidades se seleccionan:
    a) En forma caprichosa. b) Por conveniencia.
    c) A través de un censo. d) En forma repetitiva.
    e) De tal manera que todas tengan la misma posibilidad.

17. Por población o universo se entiende:
    a) Un recuento de unidades.
    b) Un conjunto de seres humanos.
    c) Un conjunto de datos.
    d) Un conjunto de medidas o el recuento de todas las unidades que tienen una característica común.
    e) Ninguna de las anteriores.

18. Cualquier medida aplicada a la característica de las unidades en la población se denomina:
    a) Parámetro. b) Estimador.
    c) Estadístico. d) Variable.
    e) Población.

19. Dentro de los hechos o fenómenos que no caen dentro del campo de la estadística están:
    a) Los de frecuente repetición. b) Los de distinta frecuencia.
    c) Los colectivos. d) Los individuales.
    e) Los cualitativos que pueden cuantificarse.

20. La estadística descriptiva tiene como objetivo:
    a) Probar la significación de los resultados.
    b) Ser herramienta indispensable en el muestreo.
    c) Descubrir las causas que originan el hecho.
    d) Lograr conclusiones más allá de las muestras.
    e) Efectuar comparaciones sin sacar conclusiones de tipo más general.

21. Se debe responder *verdadero* si el enunciado es siempre válido. En caso contrario, se deberá sustituir la palabra *subrayada* por otra, con la cual el enunciado tenga validez.
    a) Parámetro, es el resultado al aplicar una medida a las características de las unidades de una *población.*
    b) El recuento de los empleados de una empresa, de acuerdo al cargo, es un ejemplo de características *cuantitativas.*
    c) La estadística *descriptiva,* es el "estudio" de una muestra para hacer estimaciones acerca de los valores estadísticos de la población, de la cual se tomó la muestra.
    d) Una muestra aleatoria significa que, cada elemento tiene una probabilidad *diferente* al ser seleccionado.
    e) La *inferencia,* es un ordenamiento sistemático de la información en cuadros y gráficas que observamos en las diferentes publicaciones e informes.

22. Explique brevemente los siguientes puntos.
    a) Factores o impedimentos que obligan a realizar una investigación parcial, en vez de una investigación total.

b) Marco muestral.
c) Hechos que abarca y no abarca la estadística.

23. En los tres (3) ejemplos siguientes, determinar en cada uno de ellos ¿cuál es la población?; ¿cuál es la muestra?; ¿cuál es la unidad?; ¿cuál es la característica?; ¿cuál es cualitativa o cuantitativa?; ¿cuál de las variables es discreta o continua?
    a) Se realiza un estudio en 350 hogares de la clase media de la ciudad Bellavista para conocer el tipo de aceite o grasa usada en la cocina. Los resultados fueron: 130 hogares utilizan el ajonjolí; maíz en 90 hogares; girasol en 75 hogares; etc.
    b) El laboratorio de control de calidad de una empresa realiza un test de rapidez de acción de un pesticida de jardín, en 50 plantas infestadas. Los resultados fueron observados cada hora, habiéndose obtenido algunos datos del número de plantas totalmente libres de plaga, después de los períodos de tiempo que se indican: 6 horas, 6 plantas; 7 horas, 9 plantas; 8 horas. 5 plantas; etc.
    c) En un plantel de 800 niños de ambos sexos, de 5 a 12 años, se realizó un test de aceptación a 20 niños, utilizando una escala de 10 puntos, para medir el grado de aceptación de un nuevo producto que fabrica la compañía Chocolatera la Avispa S.A.

24. Diga brevemente que diferencia hay entre:
    a) Parámetro y estimador
    b) Población y muestra
    c) Atributo y variable
    d) Muestreo aleatorio y no aleatorio

25. Qué entiende usted por:
    a) Estadística
    b) Estadísticas
    c) Fuentes o estadísticas primarias
    d) Fuentes o estadísticas secundarias
    e) Estadísticas internas y externas
    f) Unidad o elemento
    g) Objetivo de la investigación
    h) Proceso de recolección
    i) Muestreo aleatorio y no aleatorio
    j) Encuestas y monografías.

Síntesis de capítulo

Existe una diferencia entre los términos *estadística* y *estadísticas*; la primera se refiere a la técnica de recolección, procesamiento y análisis del dato, la segunda, corresponde al ordenamiento sistemático de la información en forma de cuadros y gráficas que aparecen en publicaciones.

La estadística cumple dos funciones: a) la primera, de análisis *descriptivo* en forma cuantitativa de las características observadas en el fenómeno, y b) la segunda, *inferencia* estadística o inducción, lográndose generalizaciones para un grupo mayor denominado población, partiendo de un grupo menor o muestra.

Se dan una serie de términos estadísticos que hay que conocer para un buen desarrollo del curso de estadística, entre otros: población, muestra, variables y atributos.

Se considera que la finalidad de la estadística, por ejemplo, en una empresa es suministrar información que permita: a) tener una visión general de ella en su conjunto, para que la administración pueda formular directrices con pleno conocimiento de causa; b) descubrir las relaciones de causa a efecto en las diversas manifestaciones económicas de la empresa; c) reconocer y separar, en vista del control, lo normal y anormal, observando las fluctuaciones internas ligadas estrechamente con las condiciones externas, para obtener una mayor orientación a la actividad de la empresa.

También, se establecen las condiciones que deben presentar los fenómenos para que sean considerados dentro del campo de la estadística

El campo de acción de la Estadística es muy amplio; sin embargo, no todos los fenómenos son abarcados, únicamente aquellos que reúnen ciertas condiciones, a saber: fenómenos colectivos o de grupos; de frecuente repetición; de distinta frecuencia; distantes en el espacio; distantes de tiempo; cualitativos que pueden cuantificarse.

En cambio, quedan fuera del campo de acción de la estadística, los enumerados a continuación: Fenómenos individuales; que no se exteriorizan; accidentales en el tiempo y en el espacio; cualitativos que no pueden cuantificarse.

Son numerosas las finalidades de la estadística; entre las más importantes podemos anotar:

Conocer la realidad acerca de un fenómeno; determinar lo que sea normal o típico; observar los cambios que presentan; relacionar dos o más hechos; determinar las causas que originan; hacer estimaciones sobre el comportamiento futuro; partiendo de un grupo menor, obtener conclusiones para un grupo mayor (inferencia estadística).

En esta unidad se quiere resaltar, que no todas las investigaciones son indispensables de ser llevadas a cabo, debido al hecho de que en, algunos casos, se dispone de información obtenida de investigaciones realizadas con anterioridad, por la misma entidad o por otras entidades. En otras ocasiones, dicha información puede ser considerada como complementaria a la nueva investigación.

Una investigación requiere en la primera etapa, ante todo, de un planeamiento que comprenda:

- Fijar los objetivos de la investigación, determinando si el fenómeno puede ser observado mediante la aplicación de métodos estadísticos.
- Establecer la unidad o unidades de observación
- Determinar, si se trata de un censo o de una muestra.
- Si la encuesta se va a hacer mediante entrevista, entrega personal del cuestionario, por correo, teléfono, observación, panel, internet, fax, etc. Elaborar el presupuesto.
- Diseñar el calendario de trabajo. Diseñar el formulario e instrucciones.
- Seleccionar y capacitar a las personas que van a trabajar en la investigación. Hacer una preencuesta.
- En la segunda etapa, le corresponde como actividad la recolección.

- Distribución y recolección de formularios.
- Verificación del número de formularios enviados y recogidos. Efectuar un control sobre la calidad de los datos.

La tercera y última etapa comprende: la sistematización, análisis y presentación de los resultados

- Elaboración de códigos.
- Crítica y codificación
- Elaboración de cuadros de tabulación.
- Selección del proceso de tabulación, si es manual, mecánico o computarizado.
- Elaboración de cuadros y gráficas.
- Análisis y comparación de los datos.
- Publicación.

Las estadísticas y la estadística, que para muchos significan lo mismo, son términos empleados frecuentemente, siempre vinculados a cifras que miden algo. Así oímos hablar de la inflación mediante el crecimiento del índice de precios al consumidor, lo mismo de la tasa de desempleo o de criminalidad, llegando al extremo de desafiar y criticar duramente esas cifras, sin tener medianamente conocimiento de cómo fueron obtenidas ¿Cuál fue el procedimiento utilizado? ¿Qué significado tienen? y algo más, no pensamos si la interpretación dada es la correcta. Todo esto nos muestra lo poco que sabemos de esta herramienta para la interpretación y análisis de una información.

Por la anterior razón, se comenzó por definir aquellos términos que son indispensables para entender la teoría estadística, tales como población, unidad, muestra, característica, etc., que a primera vista parecen sencillos, en algunos casos resultan complejos. Veamos, en el salón de clase se dice que se va a realizar una encuesta a un grupo de familias de un barrio determinado de la ciudad, preguntémosle al estudiante: ¿Qué se entiende por familia, como unidad de investigación? resultando, un término definido de diferentes maneras. El investigador seguramente ha tomado alguna de ellas, desgraciadamente algunos de los que proporcionan información estadística no definen los conceptos que supuestamente miden, prestándose a muchas interpretaciones con los resultados presentados.

La parte metodológica, en la enseñanza de la estadística, parece que ha perdido su importancia si observamos el desarrollo y contenido de los de programas, lo cual podemos constatar en los últimos programas académicos. Se ha despertado un gran interés por la aplicación de modelos matemáticos que sólo son parte del proceso, algunos se centran en deducciones y comprobaciones, en vez de ayudar a comprender la estadística, sólo se ha logrado aumentar el grado de dificultad en su aprendizaje, dejando a un lado el conocimiento de la importancia que tiene la etapa de planeación y de recolección para producir una información consistente, que permita la aplicación de los métodos estadísticos tanto descriptivos como los de inferencia, a fin de realizar un buen análisis de los resultados y cumplir con los objetivos propuestos.

2

CAPÍTULO

Distribuciones de frecuencias

El fin primordial de la educación es formar los hombres capaces de hacer cosas nuevas y no de repetir simplemente lo que otras generaciones han hecho.
Jean Piaget

CONTENIDO

- Elaboración de tablas o cuadros de frecuencias.
- Elaboración de gráficos aplicados a la teoría estadística.
- Elaboración de gráficas generalmente utilizadas en informes.
- Recomendaciones para la elaboración de tablas y gráficos.
- Síntesis del Capítulo.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Seleccionar y elaborar gráficas que visualicen mejor la información.
- Aplicar las técnicas y recomendaciones en la elaboración de tablas o cuadros.
- Poder analizar y sacar conclusiones sobre las informaciones que se tienen, tanto en tablas o en gráficas
- Distinguir la importancia y diferencias entre el texto, cuadros y gráficas.

GENERALIDADES

La *distribución de frecuencias*, es un método utilizado para organizar y resumir información. Bajo este método, los datos recolectados se ordenan y clasifican, indicándonos la frecuencia o sea el número de veces que se repiten.

Se podrá decir también, que nos permite manejar grandes cantidades de información en espacios reducidos, en formas de cuadros o tablas, complementadas con gráficas.

Por *población o universo,* se entiende como un conjunto de medidas para ser aplicadas a una característica cuantitativa, o como el recuento de todas las unidades que presentan una característica común, siendo ésta cualitativa. También se puede definir a la población como un conjunto de elementos o unidades. Lo que se estudia en una *unidad o elemento* son sus *características*.

Cuando se toman todas las unidades o elementos de la población, se habla de una *investigación exhaustiva* o censo. Si sólo se investiga una parte, se le considera como *investigación parcial o muestra.*

La *muestra,* para que sea representativa de la población, requiere que las unidades o elementos sean seleccionadas al *azar*, en tal forma que cada una de ellas tenga la misma posibilidad de ser seleccionada.

Se usan letras mayúsculas o letras del alfabeto griego como símbolos en poblaciones, en cambio, en la muestras se emplean letras minúsculas.

Los caracteres de los *elementos* de una población pueden ser *cualitativos o cuantitativos*, Los datos cualitativos, denominados también *atributos*, son todos aquellos elementos que pueden ser descritos cualitativamente, es decir mediante palabras; son ejemplos de atributos: la clasificación de los alumnos de una universidad por lugar de origen; clasificación de un grupo de personas por ocupación, por cargo, por sexo, etc.

Los caracteres *cuantitativos* denominados *variables*, son todas aquellas características susceptibles de ser expresadas cuantitativamente, es decir, mediante números. Ejemplo: peso, estatura, edad, número de hijos, salarios, etc.

Las variables se dividen en *discretas y continuas.* Es de tener en cuenta que esta clasificación tiene más valor teórico que práctico.

Las *variables discretas* son aquéllas que admiten solamente valores enteros, es decir, no tienen valores intermedios. Ejemplo: el número de hijos por familia será discreta, ya que no se podrá decir que una familia tiene dos hijos y medio; el número de empleados por departamento en una empresa, etc.

Las *variables continuas* son aquéllas que admiten valores fraccionarios, pudiéndose establecer intervalos. Ejemplo: la estatura de una persona que mide un metro con setenta centímetros; que pesa sesenta kilos, una libra y cuatro onzas, etc.

Nota: Se hará un pequeño resumen sobre aquellos términos que más serán utilizados en los capítulos posteriores y que fueron presentados en forma amplia en el anterior capítulo.

ELABORACIÓN DE TABLA DE FRECUENCIAS

VARIABLE DISCRETA

Ante todo es conveniente familiarizarnos con ciertos símbolos que utilizaremos tanto para la variable discreta como para la variable continua.

n	=	n	=	Tamaño de la muestra.
N	=	N	=	Tamaño de la población o universo.
X_i	=	x_i	=	Identificación para cada valor observado. (minúscula en la muestra)
f_i	=	n_i	=	Frecuencias absolutas.
f_i/n	=	h_i	=	Frecuencias relativas.
F_i	=	N_i	=	Frecuencias absolutas acumuladas.
H_i	=	H_i	=	Frecuencias relativas acumuladas.
X_i	=	y_i	=	Identifica la variable discreta o las marcas de clase en la continua.
$X'_{i-1}-X'_i$	=	$y'_{i-1}-y'_i$	=	Identifica a la variable continua con sus intervalos.
i	=	c	=	Amplitud del intervalo
m		m	=	Número de valores de la variable o de intervalos.

Se ha considerado conveniente presentar las dos simbologías más utilizadas en los textos de Estadística, de tal manera que el estudiante pueda consultar o aplicar indistintamente cualquiera de ellas.

Ejemplo 1. Supongamos que se tiene una población constituida por 2.000 cajas y deseamos examinarlas con el fin de determinar el número de piezas o elementos defectuosos que contiene cada caja. Por diferentes razones, se desea que la investigación no sea exhaustiva, es decir, no examinar la totalidad de las 2.000 cajas o universo sino, por el contrario, seleccionar una muestra de tamaño 30, correspondiente a una investigación parcial. $N = 2.000$ $n = 30$

Los resultados obtenidos en esta encuesta, se anotan a continuación. Siendo x_1, la primera caja examinada y 3, el número de piezas defectuosas encontradas en ella y así sucesivamente:

$x_1 = 3$ $x_2 = 2$ $x_3 = 0$ $x_4 = 2$ $x_5 = 3$ $x_6 = 3$ $x_7 = 1$ $x_8 = 1$ $x_9 = 0$

$x_{10} = 1$ $x_{11} = 3$ $x_{12} = 3$ $x_{13} = 4$ $x_{14} = 4$ $x_{15} = 3$ $x_{16} = 2$ $x_{17} = 4$ $x_{18} = 2$

$x_{19} = 4$ $x_{20} = 2$ $x_{21} = 4$ $x_{22} = 0$ $x_{23} = 0$ $x_{24} = 1$ $x_{25} = 3$ $x_{26} = 4$ $x_{27} = 2$

$x_{28} = 3$ $x_{29} = 1$ $x_{30} = 2$

En la anterior información o conjunto de datos, observamos que no están ordenados y los denominamos d*atos originales* o *no ordenados*. Quizás, dado que la muestra es muy pequeña, podemos darnos cuenta, en forma muy ligera del comportamiento de la característica. Pero si en vez de tener 30 observaciones, fueran 200 o 2.000 es muy poco lo que podemos decir o conocer de la variable; por lo tanto se hace necesario el ordenamiento o la clasificación de esos datos, mediante un proceso que llamaremos *tabulación* o *procesamiento de datos.*

En primer lugar, se determinan los valores que toma la variable, en cuanto al número de piezas defectuosas, siendo en este caso de 0 1 2 3 4. También se hubiera podido establecer los valores extremos (0 y 4) y luego los valores intermedios (1, 2, 3); procedemos en seguida al conteo, a fin de determinar cuántas veces se repite cada valor de la variable, resultados que denominaremos *frecuencias o frecuencias absolutas.*

Tabla 2.1 Tabulación

NÚMERO DE PIEZAS	NÚMERO DE CAJAS	TABULACIÓN A	TABULACIÓN B
0	4	IIII	□
1	5	卌	⧄
2	7	卌 II	⧄ ∟
3	8	卌 III	⧄ ⊔
4	6	卌 I	⧄ I
Σ	30	30	30

Se recomienda formar grupos de cinco, 卌 卌 卌 III que nos representará 18 observaciones, siendo más fácil contar de cinco en cinco, que hacerlo de uno en uno, IIIIIIIIIIIIIIIIII. También se pueden formar cuadros de cinco, lo cual nos quedaría así: ⧄ ⧄ ⧄ ⊔ para un total de 18.

En el ejemplo 1 de la Tabla 2.1 se puede interpretar, diciendo que en la muestra de 30 cajas encontramos cuatro de ellas sin piezas defectuosas; 5 cajas con una; 7 cajas con dos; 8 cajas con tres; y 6 con cuatro piezas defectuosas. Sin embargo, a veces, es mejor expresarlo en términos porcentuales, diciendo que el 13% de las cajas examinadas no tenían piezas defectuosas, que el 17% tenían una; que el 23%, dos; el 27%, tres; y el 20% con cuatro piezas o unidades defectuosas, como se puede observar en el cuadro siguiente:

En el caso en que la variable no emplee valores en forma continua, solamente se debe considerar en la distribución aquellos que aparecen en la base de datos, como se puede observar en la Tabla 2.2, donde se omiten los valores 2 4 6 7 y 9.

Tabla 2.2 Distribución de frecuencias

NÚMERO DE PIEZAS DEFECTUOSAS
0
1
3
5
8
10

Tabla 2.3

Distribución de frecuencias - variable discreta

y_i	n_i	h_i	N_i	H_i
0	4	0,13	4	0,13
1	5	0,17	9	0,30
2	7	0,23	16	0,53
3	8	0,27	24	0,80
4	6	0,20	30	1,00
Σ	30	1,00	-	-
X_i	f_i	f_i/n	F_i	H_i

y_i	n_i	h_i	N_i	H_i
y_1	n_1	h_1	N_1	H_1
y_2	n_2	h_2	N_2	H_2
y_3	n_3	h_3	N_3	H_3
y_4	n_4	h_4	N_4	H_4
y_5	n_5	h_5	N_5	H_5
Σ	n	1,00	-	-
X_i	f_i	f_i/n	F_i	H_i

Las frecuencias relativas se obtienen de la siguiente manera:

$$h_1 = \frac{n_1}{n} = \frac{f_1}{n} \qquad h_2 = \frac{n_2}{n} = \frac{f_2}{n} \qquad h_3 = \frac{n_3}{n} = \frac{f_3}{n} \qquad h_4 = \frac{n_4}{n} = \frac{f_4}{n} \qquad h_5 = \frac{n_5}{n} = \frac{f_5}{n}$$

$$f_1/n = h_1 = \frac{4}{30} = 0{,}13 \qquad f_2/n = h_2 = \frac{5}{30} = 0{,}17 \quad \ldots \text{ etc.}$$

Las *frecuencias absolutas acumuladas* las utilizaremos más adelante para el cálculo de algunas medidas. Se obtienen así:

$$\begin{aligned}
N_1 &= n_1 & &\rightarrow & &= N_1 = n_1 = 4\\
N_2 &= n_1 + n_2 & &\rightarrow & &= N_1 + n_2 = 4 + 5 = 9\\
N_3 &= n_1 + n_2 + n_3 & &\rightarrow & &= N_2 + n_3 = 9 + 7 = 16\\
N_4 &= n_1 + n_2 + n_3 + n_4 & &\rightarrow & &= N_3 + n_4 = 16 + 8 = 24\\
N_5 &= n_1 + n_2 + n_3 + n_4 + n_5 & &\rightarrow & &= N_4 + n_5 = 24 + 6 = 30
\end{aligned}$$

$$\begin{aligned}
F_1 &= f_1 & &\rightarrow & &= f_1\\
F_2 &= f_1 + f_2 & &\rightarrow & &= F_1 + f_2\\
F_3 &= f_1 + f_2 + f_3 & &\rightarrow & &= F_2 + f_3\\
F_4 &= f_1 + f_2 + f_3 + f_4 & &\rightarrow & &= F_3 + f_4\\
F_5 &= f_1 + f_2 + f_3 + f_4 + f_5 & &\rightarrow & &= F_4 + f_5
\end{aligned}$$

Las frecuencias relativas acumuladas resultan asi.

$$\begin{aligned}
H_1 &= h_1 & &\rightarrow & &= H_1 = h_1 = 0{,}13 & &= F_1/n\\
H_2 &= h_1 + h_2 & &\rightarrow & &= H_1 + h_2 = 0{,}13 + 0{,}17 = 0{,}30 & &= F_2/n\\
H_3 &= h_1 + h_2 + h_3 & &\rightarrow & &= H_2 + h_3 = 0{,}30 + 0{,}23 = 0{,}53 & &= F_3/n\\
H_4 &= h_1 + h_2 + h_3 + h_4 & &\rightarrow & &= H_3 + h_4 = 0{,}53 + 0{,}27 = 0{,}80 & &= F_4/n\\
H_5 &= h_1 + h_2 + h_3 + h_4 + h_5 & &\rightarrow & &= H_4 + h_5 = 0{,}80 + 0{,}20 = 1{,}00 & &= F_5/n
\end{aligned}$$

Ejemplo 2. Con la información de la tabla No. 2.3 determinar los valores que toma:

Solución

$$y_3 = X_3 = 2 \qquad N_2 = F_2 = 9 \qquad n_5 = f_5 = 6$$

$$H_3 = \frac{F_3}{n} = 0{,}53 \qquad h_4 = \frac{f_4}{n} = 0{,}27 \qquad x_{16} = 2$$

VARIABLE CONTINUA

Ejemplo 3. Consideremos que se seleccionó una muestra de 30 alumnos, a fin de conocer su peso en kilos; para facilitar el trabajo redondeamos las cifras.

$x_1 = 74$	$x_2 = 67$	$\boxed{x_3 = 94}$	$x_4 = 70$	$x_5 = 69$	$x_6 = 61$	$x_7 = 71$	$x_8 = 79$	$\boxed{x_9 = 47}$
$x_{10} = 85$	$x_{11} = 82$	$x_{12} = 55$	$x_{13} = 65$	$x_{14} = 88$	$x_{15} = 52$	$x_{16} = 58$	$x_{17} = 76$	$x_{18} = 57$
$x_{19} = 72$	$x_{20} = 66$	$x_{21} = 48$	$x_{22} = 56$	$x_{23} = 63$	$x_{24} = 71$	$x_{25} = 60$	$x_{26} = 64$	$x_{27} = 68$
$x_{28} = 83$	$x_{29} = 74$	$x_{30} = 92$						

El primer paso a seguir consiste en determinar el valor máximo y el mínimo que toma la variable. En este ejemplo tenemos: $x_{max} = 94$ y $x_{min} = 47$. La diferencia entre el valor máximo y el mínimo se denomina recorrido. En el presente caso el recorrido es de 47.

$$\boxed{x_{máx} - x_{mín} = \text{Recorrido o rango}}$$

$$94 - 47 = 47$$

Introducimos dos nuevos símbolos que son: m = *número de intervalos y*

C = amplitud del intervalo = $\lceil i \rfloor$

El valor de m se puede determinar de varias maneras:

- Un número arbitrario que sea mayor o igual a cinco, y menor o igual a 16
  $$5 \leq m \leq 16 \quad \text{ó} \quad 5 \leq m \leq 18$$
- Métodos muy utilizados en clase, consiste en la aplicación de la fórmula Sturges
  $$m = 1+3{,}3 \log n$$
- De acuerdo a la clasificación que requiera la variable o la forma en que se ha venido presentando la información.
- Algunos utilizan un procedimiento, no muy recomendado, calculando m mediante la fórmula $m = \sqrt{\log n}$

En la práctica, se determina m atendiendo a varios factores, tales como: finalidad del estudio, grado de variabilidad de los datos, necesidad de efectuar comparaciones con otros estudios que obligue a mantener los intervalos y su número. Sin embargo, se recomienda que el número de intervalos, hasta donde sea posible, no sea menor de 5 ni mayor de 16. Algunos lo admiten hasta 18.

En este ejercicio se tendrá que $m = 1 + 3{,}3 \log 30 = 5{,}29 \simeq 6$, lo aproximamos a 6, pero se hubiera podido considerar que $m = 5$. La aproximación no tiene importancia alguna en estos casos. Esta fórmula recibe el nombre de Sturges, su descubridor.

En cuanto la amplitud $\lceil C \rfloor$ o $\lceil i \rfloor$ que debe tomar cada intervalo de la distribución, dependerá del criterio establecido para presentar la información. Puede variar el valor de la amplitud para cada intervalo, sin embargo en clase y con fines de explicación es mejor trabajar con una amplitud constante, para ello se tendrá que:

$$\boxed{i = C = \frac{Rango}{m}}$$

Reemplazando se tiene que:
$C = 47 \div 6 = 7{,}83$. Siempre aproximamos al número inmediatamente superior, por pequeña que sea la fracción. Si C hubiese resultado 7,06 o 7,10, se aproxima a 8. Si usted no aproxima al número superior por pequeña que sea la fracción al calcular amplitud, es muy posible que le queden valores observados por fuera del rango.

Para nuestro ejercicio fijamos m igual a 6, con lo cual podemos determinar el valor de $\boxed{C} = \lceil i \rceil$

$$\boxed{C = \frac{X_{\text{máx}} - X_{\text{mín}}}{m}} \qquad C = \frac{97-47}{6} = \frac{47}{6} = 7{,}83 \qquad C = \lceil i \rceil = 7{,}83 \cong 8$$

Siempre que el resultado de *C* sea un valor entero con alguna fracción, por pequeña que ésta sea, debe ser aproximada siempre al valor inmediatamente superior; según esto, el valor de *C* será de 8, con el cual el recorrido se aumenta en una unidad, puesto que 6 x 8 = 48. La unidad de incremento se le puede restar al límite inferior o aumentarse al límite superior. En este caso se lo restamos a 47, quedando el límite inferior en 46, a partir de este valor comenzamos a agregar el valor de *C*, para formar los distintos intervalos.

En la variable continua. Se designa con $\boxed{y'_{i-1} - y'_i}$ ó $\boxed{X'_{i-1} - X'_i}$, la columna correspondiente a los distintos valores que toma la variable. Al primer valor $\boxed{y'_{i-1}}$ ó $\boxed{X'_{i-1}}$ se denomina límite inferior del intervalo y $\boxed{y'_i}$ ó $\boxed{X'_i}$ al límite superior; la diferencia entre estos límites corresponde al valor de la amplitud del intervalo $\boxed{C}$ ó $\boxed{i}$. Es preferible con fines de trabajo, que la amplitud sea constante, sin embargo, en la práctica no es muy frecuente que esto se presente.

La tabla de frecuencias para este ejercicio nos quedará de la siguiente forma:

Tabla 2.4

Distribución de frecuencias - variable continua

$y'_{i-1} - y'_i$	n_i	h_i	N_i	H_i	y_i
46,1 – 54	3	0,10	3	0,10	50
54,1 – 62	6	0,20	9	0,30	58
62,1 – 70	8	0,27	17	0,57	66
70,1 – 78	6	0,20	23	0,77	74
78,1 – 86	4	0,13	27	0,90	82
86,1 – 94	3	0,10	30	1,00	90
Σ	30	1,00			-
$X'_{i-1} - X'_i$	f_i	f_i/n_i	F_i	H_i	X_1

$y'_{i-1} - y'_i$	n_i
$y'_0 - y'_1$	y_1
$y'_1 - y'_2$	y_2
$y'_2 - y'_3$	y_3
$y'_3 - y'_4$	y_4
$y'_4 - y'_5$	y_5
$y'_5 - y'_6$	y_6
Σ	-
$X'_{i-1} - X'_i$	f_i

Es de anotar que a cada límite inferior del intervalo se le ha agregado 0,1, lo que permite una fácil tabulación y así se sabe a cuál intervalo corresponde el valor de x_i, obtenido. **El 0,1 se omitirá para el cálculo de las marcas de clase y de los diferentes estadígrafos o medidas que se requiera calcular.**

El cuadro de tabulación con el cual se elaboró la tabla de frecuencias, se estableció de la siguiente forma:

Tabla 2.5

Tabulaciones de datos - variable continua

$y'_{i-1} - y'_i$	n_i	TABULACIÓN A	TABULACIÓN B
46,1 – 54	3	⊔	III
54,1 – 62	6	⊠ I	卌 I
62,1 – 70	8	⊠ ⊔	卌 III
70,1 – 78	6	⊠ I	卌 I
78,1 – 86	4	☐	IIII
86,1 – 94	3	⊔	III
Σ	30	30	30

Las *marcas de clase* se pueden obtener de 3 maneras:

a) Se promedia cada intervalo, de la siguiente forma:

$$X_1 = y_1 = \frac{y'_0 + y'_1}{2} = \frac{46+54}{2} = 50 = \frac{X'_0 + X'_1}{2}$$

$$X_2 = y_2 = \frac{y'_1 + y'_2}{2} = \frac{54+62}{2} = 58 = \frac{X'_1 + X'_2}{2}$$

$$X_3 = y_3 = \frac{y'_2 + y'_3}{2} = \frac{62+70}{2} = 66 = \frac{X'_2 + X'_3}{2}$$

$$X_4 = y_4 = \frac{y'_3 + y'_4}{2} = \frac{70+78}{2} = 74 = \frac{X'_3 + X'_4}{2} \text{..., etc.}$$

b) Se obtiene la primera marca de clase por el método anterior y si la amplitud C es constante, se le suma a la primera marca de clase obtenida y así sucesivamente.

$$y_1 = \frac{y'_0 + y'_1}{2} = \frac{46+54}{2} = 50 = \frac{X'_0 + X'_1}{2} \qquad ; X_2 = X_2 = 50 + 8 = 58 \qquad ; X_3 = X_3 = 58 + 8 = 66$$

c) Se divide la amplitud de cada intervalo por dos y se le suma al límite inferior del intervalo o se le resta al límite superior del intervalo.

$$\frac{i}{2} = \frac{C}{2} = \frac{8}{2} = 4$$

$$\boxed{y'_0 + \frac{C}{2} = y_1} \qquad X_1 = y_1 = 46 + 4 = 50 \qquad X_2 = y_2 = 54 + 4 = 58, \textit{etc.}$$

$$\boxed{y'_1 - \frac{C}{2} = y_1} \qquad X_1 = y_1 = 54 - 4 = 50 \qquad X_2 = y_2 = 62 - 4 = 58, \textit{etc.}$$

Nota: una distribución es simétrica cuando las frecuencias absolutas o relativas, equidistantes a un punto central son iguales, además, la gráfica que la representa tiene forma acampanada o convexa como se observa con los datos que se presentan a continuación:

Tabla 2.6

A continuación se presenta dos ejemplos de Distribuciones simétricas:

a)

$y'_{i-1} - y'_i$	n_i	h_i
46,1 – 54	2	0,05
54,1 – 62	6	0,15
62,1 – 70	12	0,30
70,1 – 78	12	0,30
78,1 – 86	6	0,15
86,1 – 94	2	0,05
Σ	40	1,00
$X'_{i-1} - X'_i$	f_i	f_i/n

b)

y_i	n_i	h_i
2	2	0,10
4	5	0,25
6	6	0,30
8	5	0,25
10	2	0,10
Σ	20	1,00
X_i	f_i	f_i/n

Propiedades de las frecuencias

- Las frecuencias absolutas son siempre valores enteros.
- La suma de las frecuencias absolutas es igual a n.

$$\sum_{i=1}^{n} n = n \qquad \sum_{i=1}^{5} n = n = n_1 + n_2 + n_3 + n_4 + n_5 = n$$

$$\sum_{i=1}^{5} n_i = 30 = 4 + 5 + 7 + 8 + 6 = 30$$

$$\sum_{i=1}^{n} f_i = n \qquad \sum_{i=1}^{5} f_i = f_1 + f_2 + f_3 + f_4 + f_5 = n \qquad \sum_{i=1}^{5} f_i = 30$$

- Las frecuencias relativas son siempre valores fraccionarios.

$$0 < h_i < 1 \qquad 0 < f_i/n < 1$$

- La suma de las frecuencias relativas es igual a 1

$$\Sigma h_i = 1 \qquad \sum_{i=1}^{5} h_i = h_1 + h_2 + h_3 + h_4 + h_5 = 1$$

$$\sum^{5} h_i = 0,13 + 0,17 + 0,23 + 0,27 + 0,20 = 1,00$$

$$\Sigma f_i/n = 1 \qquad \sum_{i=1}^{5} f_i/n = \frac{f_1}{n} + \frac{f_2}{n} + \frac{f_3}{n} + \frac{f_4}{n} + \frac{f_5}{n} = 1$$

- El último valor de las frecuencias absolutas acumuladas es igual a n.

$N_m = n \qquad N_5 = n \quad 30 = 30 \qquad F_m = n \qquad F_5 = n = 30$

- El último valor correspondiente a las frecuencias relativas acumuladas debe ser igual a 1.

$H_m = 1 \qquad H_5 = 1 \qquad F_m/n = 1 \qquad F_6/n = 1$

Ejemplo 3.

a. Indique dos ejemplos de distribuciones unidimensionales, dos de bidimensionales y dos de multidimensionales, relacionadas con la economía del país. b.Señale cuatro distribuciones de variable discreta, tres de variable continua y tres de atributos.

Solución:

a.	Unidimensional (univariante)	1. Producción
		2. No. de explotaciones agrícolas
		3. Ingresos
	Bidimensional	1. Producción e inversión (bivariante)
		2. No. de explotaciones agrícolas y hectáreas cultivadas.
		3. Ingresos y gastos.
	Pluridimensional (polivariante)	1. Producción; inversión; horas laborales
		2. No. de explotaciones agrícolas, hectáreas cultivadas número de obreros
		3. Ingreso, gasto, No. de familias
b.	Variable discreta	1. No. de hijos por familia
		2. No. de hermanos por alumno.
		3. No. de billetes por numeración
		4. No. de piezas defectuosas.
	Variable continua	1. Ingresos
		2. Peso en Kg.
		3. Longitud.
c.	Atributos	1. Exportaciones por puertos colombianos.
		2. Número de alumnos por Universidad en Caracas.
		3. Consumo de cigarrillos por marcas.

Ejemplo 4.

En los ejemplos siguientes señalar con una **A** las series constituidas por datos cualitativos (atributos) y con una **V** los datos cuantitativos (variables).

Solución:

a. Distribución de alumnos por mes de nacimiento **(A)**
b. Distribución de alumnos por nacionalidad **(A)**
c. Distribución de profesionales por estatura y peso. **(V)**
d. Distribución de oficiales por sueldo. **(V)**
e. Distribución de accidentes por causa. **(A)**

Ejemplo 5.

Señale con **C** las series de variable continua y con **D** las de variable discreta.

Solución:

a. Distribución de obreros por salarios (**C**)

b. Distribución de fallecimientos por edades (**C**)

c. Distribución de alumnos por número de hermanos (**D**)

d. Distribución de alumnos por estatura. (**C**)

Ejemplo 6.

Leer el siguiente texto: *"Una vez recolectados los datos en forma ordenada, es necesario presentarlos en forma tal que se facilite su comprensión y su posterior análisis y utilización. Para ello se ordenan en cuadros numéricos y luego se representan en gráficos, para variable discreta mediante diagramas de frecuencias, tanto para absolutas o relativas"*.

a. Formar una tabla de frecuencias absolutas (n_i), según el número de letras que componen cada palabra (*rr* y *ll*, debe considerarse una sola letra).

b. Considerando el número de letras por palabra. ¿Qué valor tienen x_4 x_{27} x_{12} x_{34} x_2?

c. En la tabla de frecuencias del punto a. ¿Qué valores poseen y_2 y_4 y_7?

d. Análogamente ¿Qué valores poseen n_2 n_4 n_7?

e. ¿Qué valores poseen h_2 h_4 h_7?

f. Calcular las frecuencias relativas acumuladas (H_i) y las frecuencias absolutas acumuladas (N_i).

g. Puede ocurrir que $h_4 > h_7$ (en general). ¿Qué $H_4 < H_7$?

Solución:

x_1 = una = 3 letras x_2 = vez = 3 letras, etc.

a)

$x_1 = 3$	$x_{11} = 12$	$x_{21} = 2$	$x_{31} = 7$	$x_{41} = 8$
$x_2 = 3$	$x_{12} = 2$	$x_{22} = 9$	$x_{32} = 9$	$x_{42} = 8$
$x_3 = 12$	$x_{13} = 5$	$x_{23} = 8$	$x_{33} = 1$	$x_{43} = 9$
$x_4 = 3$	$x_{14} = 3$	$x_{24} = 1$	$x_{34} = 5$	$x_{44} = 2$
$x_5 = 5$	$x_{15} = 3$	$x_{25} = 11$	$x_{35} = 2$	$x_{45} = 11$
$x_6 = 2$	$x_{16} = 2$	$x_{26} = 4$	$x_{36} = 11$	$x_{46} = 5$
$x_7 = 5$	$x_{17} = 8$	$x_{27} = 3$	$x_{37} = 2$	$x_{47} = 4$
$x_8 = 8$	$x_{18} = 2$	$x_{28} = 2$	$x_{38} = 8$	$x_{48} = 9$
$x_9 = 2$	$x_{19} = 11$	$x_{29} = 7$	$x_{39} = 4$	$x_{49} = 1$
$x_{10} = 9$	$x_{20} = 1$	$x_{30} = 2$	$x_{40} = 8$	$x_{50} = 9$

y_i	n_i	h_i	N_i	H_i
1	4	0,08	4	0,08
2	11	0,22	15	0,30
3	6	0,12	21	0,42
4	3	0,06	24	0,48
5	5	0,10	29	0,58
7	2	0,04	31	0,62
8	7	0,14	38	0,76
9	6	0,12	44	0,88
11	4	0,08	48	0,96
12	2	0,04	50	1,00
Σ	50	1,00	-	-
X_i	f_i	f_i/n	F_i	N_i/n

b) $x_4 = 3$ $x_{27} = 3$ $x_{12} = 2$ $x_{34} = 5$ $x_2 = 3$

c) $y_2 = 2 = \boxed{X_2}$ $y_4 = 4 = \boxed{X_4}$ $y_7 = 8 = \boxed{X_7}$

d) $n_2 = 11 = \boxed{f_2}$ $n_4 = 3 = \boxed{f_4}$ $n_7 = 7 = \boxed{f_7}$

e) $h_2 = 0{,}22$ $h_4 = 0{,}06$ $h_7 = 0{,}14$

$\boxed{f_2/n = 0{,}22}$ $\boxed{f_4/n = 0{,}06}$ $\boxed{f_7/n = 0{,}14}$

f) En general puede ocurrir que $h_4 > h_7$ ó $h_7 > h_4$
Pero siempre se tendrá $H_4 < H_7$ y nunca se dará que $H_4 > H_7$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

1. En un grupo de familias considerando el número de hijos, se han obtenido los siguientes valores:

2	0	2	4	4	6	6	4	6	7	4	4	7	4	2	0	4	6	7	7

Construir una tabla con base en los anteriores datos. Calcular las columnas correspondientes a: n_i h_i N_i y H_i.

2. Dé contestación a los siguientes puntos, señalando con una *X* si es cierto o falso.

	Cierto	Falso
a. $H_5 = 0{,}36$ $N_4 = 30$ $n_5 = 6$ $n = 50$	( )	( )
b. Para calcular las marcas de clase, se suma el límite inferior al superior del intervalo y se divide entre dos.	( )	( )
c. El número de accidentes según sus causas, es una variable discreta.	( )	( )
d. Si $H_6 = 0{,}7$ y $H_4 = 0{,}3$ un 40% de los valores de la variable es menor que Y_6 y mayor que Y_4.	( )	( )
e. $H_5 = 1{,}10$ $H_4 = 0{,}80$ $h_5 = 0{,}30$	( )	( )
f. La suma de las frecuencias absolutas es igual a 1.	( )	( )

3. Reconstruya la siguiente distribución simétrica.

y_i	n_i	H_i
10	6	0,12
20	-	0,32
30	-	-
40	-	-
50	-	-
X_i	f_i	F_i/n

4. Se realiza un estudio en la ciudad de Guayaquil a 150 familias de clase media, para conocer el tipo de aceite o manteca usados en la cocina. Los resultados son los siguientes: Maíz, 14 hogares; Soya, 65 hogares; Ajonjolí, 21 hogares; compran aceite al detal sin especificar tipo, 17 hogares; manteca de cerdo, 21 hogares; grasas de origen vegetal, 6 hogares; Oliva, 13 hogares.
a) ¿Cuál es la población?
b) ¿Cuál es el tamaño de la muestra?
c) ¿Qué carácter tiene la población?
d) ¿Cuál es la característica?
e) ¿Cuántas clases de características tiene la distribución?
f) Construya una distribución de frecuencias en el mismo orden indicado.
g) ¿Cuál es la 5ª clase?
h) ¿Cómo se explica que la suma de frecuencias sea superior al número de hogares?

5. Una fábrica de gaseosas proyecta lanzar al mercado un nuevo sabor. Se realiza un test de aceptación de dicho sabor en una muestra de 30 niños, utilizando una escala de 10 puntos, para medir el grado de aceptación. Los puntos obtenidos en los 30 niños fueron los siguientes:

2	6	8	7	4	5	10	6	6	7	6	7	3	8	7	6	8	6	5	4	7	8	5	7	6	7	2	7	2	7

La muestra estuvo compuesta por igual número de niños de ambos sexos, de 5 a 12 años de edad, residentes en el Barrio San Eduardo de la ciudad de Maracaibo.

a) ¿Cuál es la población?
b) ¿Cuál es la muestra?
c) ¿Cuál es la característica?
d) ¿La característica es cualitativa o cuantitativa?
e) ¿Cuántas clases tiene el atributo?
f) ¿Qué clase de escala se ha utilizado en la medición de la variable?
g) Construya una distribución de frecuencias
h) ¿Por qué la suma de la frecuencias es diferente a 150?

6.Contestar si son ciertas o falsas las siguientes relaciones y por qué.

a.	$m = 6$	$h_1 = 0,2$	$h_4 = 0,2$	$H_2 = 0,6$	$H_3 + H_4 = 1,9$
	$m = 6$	$f_1 = 0,2$	$f_4 = 0,2$	$F_2/n = 0,6$	$F_3/n + F_4/n = 1,9$
b.	$h_2 = 0,20$	$n = 80$	$n_2 = 16$		
	$f_2/n = 0,20$	$n = 80$	$f_2 = 16$		
c.	$m = 6$	$n_2 + n_4 = 15$	$n_1 + n_3 = 25$	$n = 50$	$n_5 + n_6 = 20$
	$m = 6$	$f_2 + f_4 = 15$	$f_i + f_3 = 25$	$N = 50$	$f_5 + f_6 = 20$

7. Determine si las siguientes características son cualitativas o cuantitativas. En el caso de ser cuantitativas, determine si son discretas o continuas.
a) Almacenes de la localidad de acuerdo a la especialidad (ropa, zapatos, etc.).
b) La cantidad de artículos para el hogar comprado en el mes, por 30 hogares.
c) El número de unidades por marca, comprado por las amas de casa en un barrio de la ciudad.
d) Ocupación de los familiares y amigos.
e) Pagos mensuales en servicios por los propietarios en un conjunto residencial.
f) La marca de microcomputadora personal que usan los compañeros de la universidad.
g) La cantidad de tiempo que utilizan por semana los compañeros de la universidad en la microcomputadora personal.

8. Qué entiende usted por:
a) Distribución de frecuencias
b) Variable discreta
c) Frecuencia absoluta
d) Frecuencia relativa

9. Una encuesta realizada a 30 fumadores para determinar el número de cigarrillos que encienden (fuman) en un día corriente arrojó los siguientes resultados:

3	7	5	10	8	4	5	8	10	8	8	4	5	3	10	5	7	10	8	5	5	12	8	4	4	3	5	8	12	10

Con los datos anteriores elabore una tabla de frecuencias.

10. Elabore una distribución de frecuencias para atributos; tome para ello cualquier característica que vea en el salón o en el instituto.

11. Diga si son ciertas o no las siguientes afirmaciones. ¿Por qué?
    a. El número de accidentes de trabajo por días de la semana es un ejemplo de variable discreta.
    b. La suma de las frecuencias absolutas acumuladas, debe ser el doble al valor del tamaño de la muestra.
    c.

$H_5 = 0{,}8$	$h_4 = 0{,}2$	$H_3 = 0{,}3$	$h_5 = 0{,}3$
$F_1/n = 0{,}8$	$f_4 = 0{,}2$	$F_3/n = 0{,}3$	$F_5 = 0{,}3$

12. Con los datos siguientes, elaborar una tabla de frecuencias sabiendo que la distribución es simétrica.

$m = 7$	$c = 10$	$n_1 = 8$	$n_2 + n_5 = 62$	$y_3 n_3 = 1.260$	$h_3 = 0{,}21$	$H_6 = 0{,}96$
$m = 7$	$i = 10$	$f_i = 8$	$f_2 + f_5 = 62$	$X_3 f_3 = 1.260$	$f_3/n = 0{,}21$	$F_6/n = 0{,}96$

13. Las estaturas (en centímetros) de los socios de un club juvenil de Panamá, son las siguientes:

153	123	129	132	147	138	137	134	131	147
138	128	134	148	125	139	146	145	148	135
152	128	146	143	138	138	122	146	137	151
145	124	132	138	144	141	137	146	138	146
152	136	160	159	157	150	160	142	148	130

Se pide agrupar los datos en una tabla de frecuencias con 6 intervalos.

14. Los siguientes datos corresponden a la distribución de frecuencias de los gastos en publicidad (millones de pesos) de 50 empresas comerciales, durante el último trimestre de 2011. Dichos gastos se agruparon en cuatro clases de amplitud constante, de la cual se sabe:

$y_1 = 3{,}5$	$y'_4 = 8{,}75$	$n_1 = 4$	$N_2 = 20$	$n_3 = 25$
$X_1 = 3{,}5$	$X'_4 = 8{,}75$	$f_1 = 4$	$F_2 = 20$	$f_3 = 25$

Se pide elaborar una tabla de frecuencias.

15. Una empresa dedicada a la producción de ambientadores en forma de atomizadores, realiza una investigación entre amas de casa del barrio El Recuerdo de esta ciudad, para lo cual selecciona 50 de ellas, en cuanto al tiempo de permanencia del aroma en su casa, con los siguientes resultados: 3 horas, 3 amas; 4 horas, 7 amas; 5 horas, 10 amas; 6 horas, 16 amas; 7 horas, 9 amas; 8 horas, 5 amas.
    a) ¿Cuál es la población?
    b) ¿Cuál es la muestra?
    c) ¿Cuál es la característica?
    d) Diga si la característica es cualitativa o cuantitativa.
    e) Diga si la variable es discreta o continua.

16. En una empresa, se hace un estudio sobre el número de horas que el personal deja de laborar durante la semana. Los resultados son: 10 horas, 3 obreros; 3 horas, 7 obreros; 9 horas, 1 obrero; 4 horas, 2 obreros; 8 horas, 5 obreros; 5 horas, 8 obreros; 6 horas, 6 obreros.
    a) ¿Cuál es la población?
    b) ¿Cuál es la variable?

c) ¿Que clase de variable es?
d) ¿Cuántos valores toma la variable?
e) ¿Cuales son los valores de:

y_3	n_2	H_5	h_4
X_3	f_2	F_5/n	f_4/n

17. Con los siguientes datos, correspondientes a los saldos (miles de $) de cuentas pendientes:

77	70	65	62	53	78	41	48	74	63	34	38	69	
79	76	55	59	69	78	75	34	68	56	61	41	83	
54	49	68	48	64	84	74	68	73	69	31	69	78	64

Elabore una tabla de frecuencias, calculando m, mediante la fórmula: $m = 1 + 3{,}3 \ \log n$
Determine los valores de:

y_3	y_5	H_4	N_5	C
X_3	X_5	F_4/n	F_5	i

18. Con los siguientes datos:

$y_2 = 22$	$y'_5 = 50$	c = constante	$m = 6$
$n = 150$	$n_1 = n_6$	$n_2 = n_5 = n_1 + 5$	$n_3 = n_4 = 30$
$X_2 = 22$	$X'_5 = 50$	i = constante	$m = 6$
$n = 150$	$f_1 = f_6$	$f_2 = f_5 = f_1 + 5$	$f_3 = f_4 = 30$

Elaborar una tabla de frecuencias.

19. Según la regla de Sturges $m = 1 + 3{,}3 \ \log n$ ¿Cuántas marcas de clase calcularía usted cuando:
a) $n = 50$
b) $n = 200$
c) $n = 1.000$
d) La amplitud del intervalo es la diferencia que hay entre el límite superior e inferior de cada intervalo.

20. ¿Qué opinión le merece a usted, si elaboramos una tabla de frecuencias y utilizamos:
a) Dos marcas o número de clases
b) 28 marcas de clase
c) 12 marcas de clase

21. Contestar si son ciertos o falsos los siguientes puntos:
a) Las frecuencias absolutas son valores que admiten fracciones
b) La suma de las frecuencias relativas puede ser mayor a 1
c) Se dice que $0 > h_1 > 1$
d) El último término de la frecuencia absoluta acumulada es igual a n.

22. Decir si son ciertas o no las siguientes relaciones:
a) $m = 6$ $h_1 = 0{,}2$ $h_4 = 0{,}2$ $H_2 = 0{,}6$ $H_3 + H_4 = 1{,}9$
b) $H_4 = 0{,}20$ $H_5 = 0{,}12$ $h_5 = -0{,}08$
c) $m = 4$ $H_5 + h_6 = H_6$

23. Con el ejercicio 18 ¿Qué porcentaje de los valores:
    a) Son superiores a 42?
    b) Inferiores a 34?
    c) Entre 18 y 50?

RECOMENDACIONES EN LA ELABORACIÓN DE CUADROS O TABLAS PARA INFORMES

Las anteriores tablas de frecuencias nos permiten desarrollar con más facilidad la aplicación de *medidas* de *posición, dispersión, asimetría* y *apuntamiento*, nociones básicas en la estadística descriptiva, sin embargo en los informes puede haber cambios en la forma de presentación de los datos.

Aunque no se pueden fijar normas rígidas en la presentación de los cuadros de informes, si es posible hacer algunas recomendaciones.

- Todo cuadro debe estar numerado, en especial, cuando el informe presenta dos o más cuadros.
- El título debe indicar: ¿qué?, ¿cómo?, ¿dónde? y ¿cuándo?.
- El cuadro consta de un encabezamiento, en el cual van los títulos que deben escribirse en mayúsculas y los subtítulos con minúsculas.
- En el cuerpo del cuadro va la información.
- En un cuadro, ninguna casilla debe quedar en blanco, siendo necesario emplear símbolos.
- El pie, se ubica abajo de la línea inferior del cuadro y se coloca en él, la fuente, nota o llamada que se crean necesarias.
- Por presentación, el cuadro no debe estar cerrado a los lados, sin embargo no se hace.
- El cuadro deberá estar cerrado en la parte superior e inferior con líneas más gruesas.
- Se repetirá el título del cuadro en todas las páginas por él ocupadas, utilizando la palabra continúa, continuación o conclusión, según el caso.
- El cuadro podrá tener subtotales y totales

Cuadro N° Título

Encabezamiento	MAYÚSCULAS	MAYÚSCULAS			MAYÚSCULAS
		minúsculas	minúsculas	minúsculas	
Cuerpo					

Pie

Veamos ahora, las clases y la estructura que se debe tener en cuenta en la elaboración de cuadros o tablas de informe.

Los cuadros estadísticos pueden ser clasificados como:

1. De simple entrada
   - Simple
     - Homogéneos
     - Heterogéneos
   - Complejos
2. De doble entrada
   - Corriente
   - Correlación

La presentación de las informaciones obtenidas en encuestas, se puede realizar de varias formas, ya sean aisladas o combinadas, manteniendo el siguiente orden:

a) Textual,
b) Cuadros o tablas,
c) Gráficas.

a) La presentación en *texto o textual*, tiene aplicación bastante limitada; se le utiliza principalmente para informes de empresas, artículos, reportajes y otras publicaciones de igual naturaleza. En este caso, los datos aparecen intercalados con los comentarios o la simple interpretación de ellos.
b) Los *cuadros* o *tablas* corresponden a arreglos sistemáticos de los datos en filas o columnas. Los cuadros son un buen complemento del texto en los informes.

ELABORACIÓN DE GRÁFICAS

Hemos visto que la estadística descriptiva cumple con algunas funciones, ya sea suministrando técnicas para recolectar, ordenar, clasificar mediante cuadros o tablas, de las informaciones obtenidas como las que acabamos de ver, pero por si solas no es suficiente, de ahí la necesidad de complementarlas por medio de *gráficas*, que nos permitan ver con mayor ligereza la información que se ha querido presentar mediante cuadros.

A continuación en términos generales, se dan algunas recomendaciones para la elaboración de gráficas.

- La mejor gráfica es la más simple.
- Las gráficas deben ser tan sencillas y claras, de tal manera que sean comprensibles sin la ayuda de las descripciones del texto.
- Las gráficas deben seguir y nunca preceder a la exposición del texto.
- Las gráficas no sustituyen al cuadro, al contrario deben complementarse.
- Las gráficas más comunes se elaboran teniendo como base los ejes de coordenadas cartesianas.
- La finalidad de las gráficas es visualizar mejor la información.
- La gráfica es considerada como el medio de expresión de la estadística, más llamativa y sugestiva, a la vez que presenta la ventaja de dejar en la memoria una expresión más duradera que los cuadros o el texto, en un menor tiempo de lectura.
- Si el informe tiene dos o más gráficas deberán numerarse.

- Toda gráfica debe tener título que indique con claridad el contenido de la misma. Las líneas y símbolos usados, deben ser los estrictamente indispensables.
- Las líneas que representan los datos o las variables, deben ser más gruesas que las escalas.
- La línea vertical (ordenada) representa las frecuencias, y se debe comenzar de cero.
- Las características cualitativas y cuantitativas, por lo general, van en la línea horizontal (abscisa).
- La lectura de la escala del eje horizontal se hace de izquierda a derecha. La del eje vertical debe hacerse de abajo hacia arriba.
- Cuando la gráfica presenta más de una característica o variable, deberá diferenciarse por medio de leyendas, notas o signos convencionales.
- En toda gráfica se debe explicar la fuente de donde fueron obtenidos los datos; además aclarar las escalas, leyendas, notas y convenciones que ayuden a identificar las características presentadas.

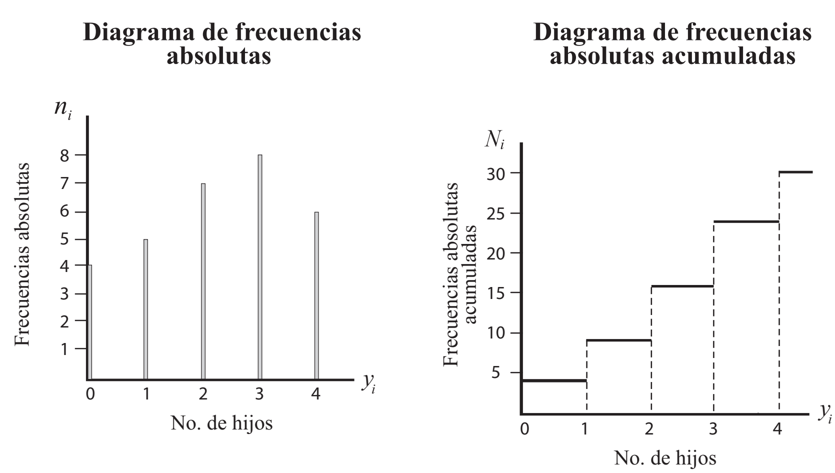
Hemos dividido las gráficas en dos grandes grupos: aquéllas que generalmente se utilizan para el desarrollo de la teoría estadística, tales como los *diagramas de frecuencias y de dispersión, los histogramas, polígonos y ojivas*. El segundo grupo lo conforman las gráficas que frecuentemente vemos en las publicaciones, llámense informes, periódicos, revistas, publicidad, etc., cómo son las *gráficas: circulares,* de *barras,* de *líneas, pictogramas, cartogramas* y *muchas más.*

GRÁFICAS APLICADAS AL DESARROLLO DE LA *TEORÍA ESTADÍSTICA*

Diagramas de frecuencias. Son muy utilizados para representar las frecuencias absolutas y relativas, incluyendo las acumuladas que ocurren con respecto a una variable aleatoria discreta. Se representan por líneas delgadas ya sean verticales u horizontales, colocando las frecuencias.

Ejemplo 1. Con los datos de la tabla 2.1 elaboraremos los respectivos *diagramas* de *frecuencias*.

Figura 2.1



Histograma. Son diagramas de frecuencias unidimensionales, en los cuales se levantan rectángulos de áreas, proporcionales a las frecuencias de clases sobre los intervalos del eje horizontal, por lo tanto es utilizado para representar a la variable continua, cuando la amplitud es constante.

La representación gráfica, para las frecuencias relativas y absolutas en variables continuas, se hace mediante *histogramas de frecuencias*. En el eje de las abscisas se señalan los intervalos que constituyen la base del rectángulo, y en el eje vertical u ordenada corresponde a las frecuencias absolutas o número proporcional de ellas, formando la altura del rectángulo (Figura No. 2.2). Sin embargo, en la confección de este gráfico se pueden presentar grupos irregulares o sea, cuando la amplitud es diferente para cada intervalo, presentando unos espacios pequeños y otros demasiados grandes. Este gráfico aparentemente correcto (Figura No. 2.3), es en realidad falso y engañoso. Para obtener un gráfico correcto es preciso dividir la frecuencia absoluta, por la amplitud del intervalo, es decir:

$$\text{Altura} = \frac{n_i}{C_i}$$ (Ver figura 2.4 y tabla 2.10)

El histograma de frecuencias se representa por medio de rectángulo o áreas:

Tabla 2.8

Tabla de frecuencias

$y'_{i-1} - y'_i$	n_i	h_i	N_i	H_i
46,1- 54	3	0,10	3	0,10
54,1- 62	6	0,20	9	0,30
62,1- 70	8	0,27	17	0,57
70,1- 78	6	0,20	23	0,77
78,1- 86	4	0,13	27	0,90
86,1-94	3	0,10	30	1,00
Σ	30	-	1,00	-
$X'_{i-1} - X'_i$	f_i	f_i/n	F_i	F_i/n

Figura 2.2

Histograma de frecuencias

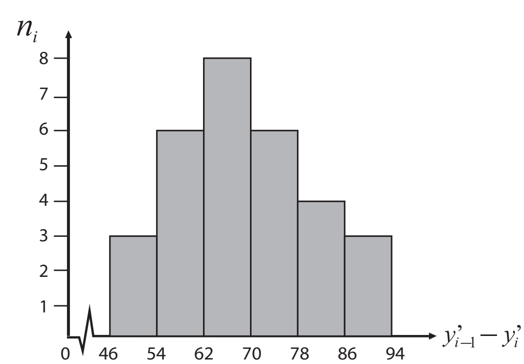


Tabla 2.9

Obreros clasificados por edades en una gráfica

EDAD (AÑOS)	No. OBREROS
16,1 - 25	55
25,1 - 28	47
28,1 - 31	32
31,1 - 40	26
Σ	**160**

Figura 2.3

Histograma elaborado incorrecto

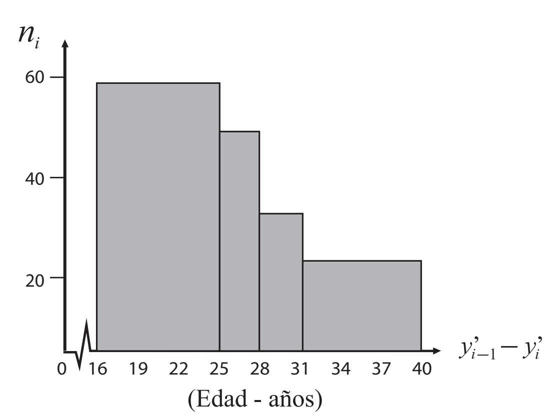


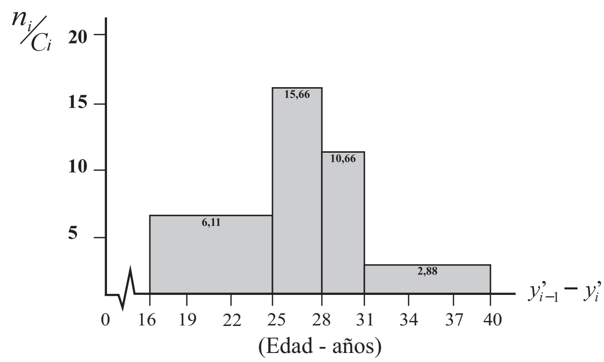
Tabla 2.10

Obreros clasificados por edades en una gráfica

AMPLITUD	FRECUENCIAS	ALTURA
C_i	n_i	n_i / C_i
9	55	6,11
3	47	15,66
3	32	10,66
9	26	2,88
i	f_i	f_i/i

Figura 2.4

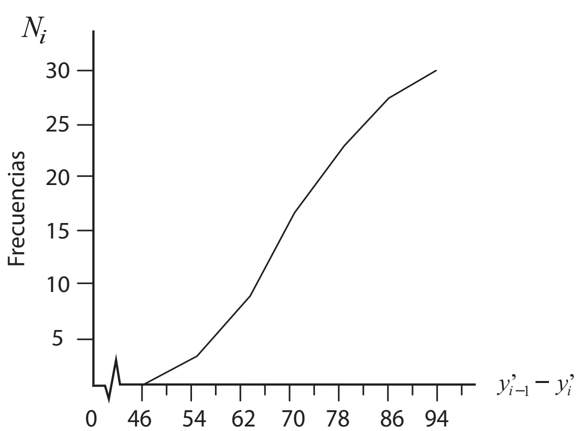
Histograma correcto



Ojiva. La representación gráfica para las frecuencias absolutas y relativas acumuladas en una variable continua se hace a través de una *ojiva ascendente*. Para ello se determinan los puntos de intersección entre cada valor de la variable y su respectiva frecuencia; luego se unen con trazos rectilíneos. (Figura No 2.5). Prácticamente es un *polígono de frecuencias*, con la diferencia de ser aplicado a una tabla de frecuencias absolutas acumuladas.

Figura 2.5

Ojiva



Polígono de frecuencias. En la variable continua es bastante utilizado este diagrama, fijando puntos, utilizando las marcas de clase y las frecuencias, luego se unen dando una línea quebrada. Si en el histograma de frecuencias unimos los puntos medios en la parte superior de cada rectángulo, obtenemos el *polígono de frecuencias* (figura No. 2.6 y 2.7).

Figura 2.6

Polígono de frecuencias

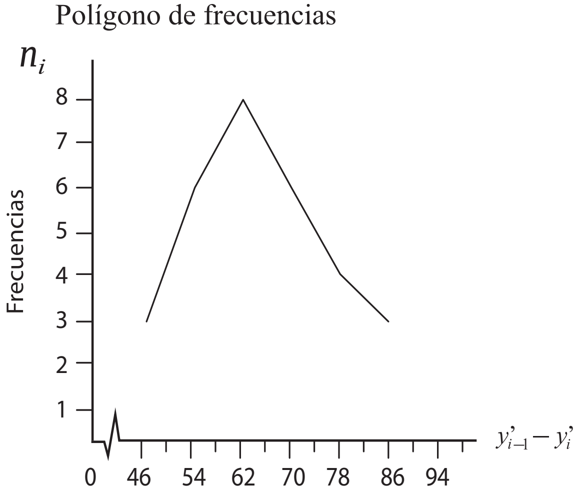
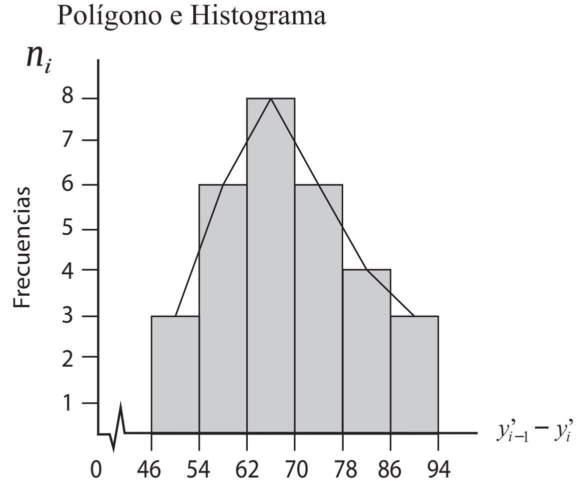


Figura 2.7

Polígono e Histograma



EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

24. Responda a los siguientes puntos, diciendo si son ciertas o falsas las siguientes afirmaciones:
   a) Al hacer una gráfica, la variable a estudiar va en el eje de las abscisas y las frecuencias en la ordenada.
   b) El polígono de frecuencias se emplea para representar a la variable continua.
   c) Entre más características tenga la gráfica, se le considera como la mejor.
   d) Si los intervalos no son iguales, nos facilita la elaboración del histograma por lo tanto será mucho más representativo.

25. Con los datos del ejercicio No. 9 construya los diagramas de frecuencias tanto con las frecuencias absolutas, como con las acumuladas.

26. Con los datos del ejercicio No. 13, se pide elaborar el histograma de frecuencias.

27. Con los datos del ejercicio No. 15, se pide graficar el conjunto de observaciones.

28. Con los datos del ejercicio No. 18, se pide elaborar el histograma de frecuencias.

29. Elaborar un histograma y polígono utilizando la siguiente distribución:

$y'_{i-1} - y'$	40,1 – 52	4,1 – 20	20,1 – 24	24,1 – 32	32,1 – 40		$X'_{i-1} - X'_i$
n_i	30	16	20	10	24	100	f_i

30. Explique brevemente qué entiende usted por:
   a) Marcas de clase
   b) Variable continua
   c) Intervalo de clase
   d) Histograma de frecuencias
   e) Polígonos de frecuencias
   f) Amplitud del intervalo.

31. En una investigación realizada a 800 empresas sobre valor de las ventas, en una quincena, se obtuvo la siguiente distribución.

VENTAS	PORCENTAJE DE EMPRESAS
8 – 18	0,30
18 – 48	0,25
48 – 98	0,18
98 – 148	0,14
148 – 200	0,13

Se pide:
a) Completar la distribución de frecuencias
b) Cuántas empresas venden menos de 98 mill $
c) Elaborar el histograma y polígono de frecuencias.

32. Los ingresos semanales (miles de $) de 50 trabajadores de la construcción fueron los siguientes:

170	172	183	164	166	164	160	168	173	176
184	190	163	151	158	166	166	172	170	186
170	166	169	160	158	155	160	157	163	165
170	172	178	183	170	166	165	172	168	158
158	168	166	159	154	166	172	168	163	172

a) Se pide elaborar una tabla de frecuencias.
b) Construir un histograma, el polígono y la ojiva.

33. Con los siguientes datos sin agrupar, correspondientes a saldos (en miles de pesos) de cuentas pendientes:

77	70	65	62	53	78	41	48	74	63
34	38	69	64	79	76	55	59	69	78
75	34	68	56	61	41	83	54	49	68
48	64	84	74	68	73	69	31	69	78

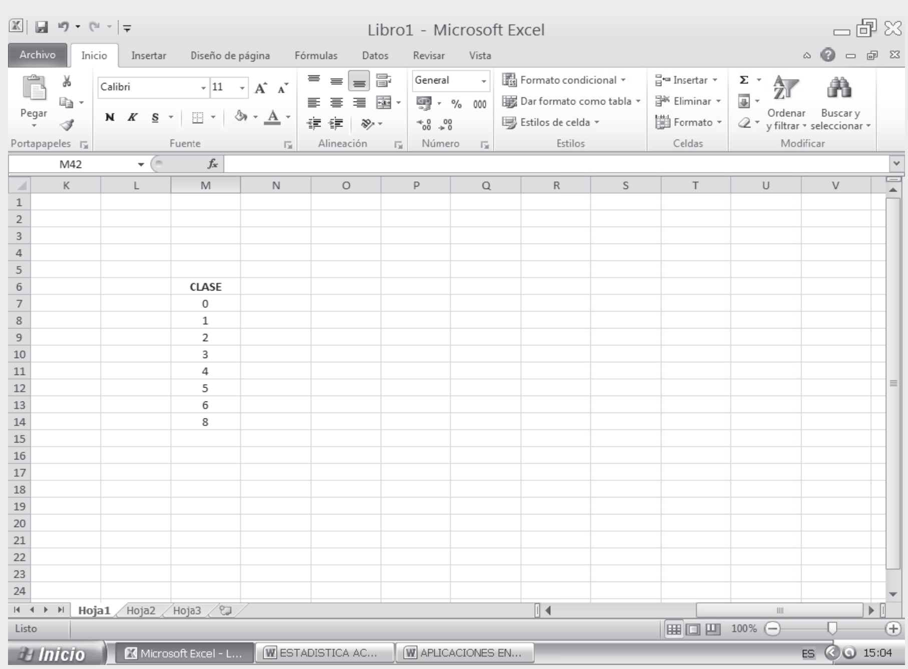
Elaborar una tabla de frecuencias y dibujar la ojiva.

APLICACIONES DE LA FUNCIÓN EXCEL EN LA ELABORACIÓN DE TABLAS O CUADROS

Complemento virtual SIL

Acudimos a la información que aparece en la tabla No. 1 (ver SIL: Aplicaciones estadísticas en Excel). Para elaborar nuestra tabla, seleccionamos otras celdas donde aparezcan los datos ya organizados, por ejemplo, ubicamos el cursor desde M6 hasta M15, tecleamos el título CLASE y a continuación (debajo) en orden ascen-dente se teclean los valores que toma nuestra variable, siendo 0 1 2 3 4 5 6 8 como lo podemos observar en la aplicación siguiente:

Figura No. 1. Ubicación de Datos



- Guiándonos por la Tabla No. 1 del SIL, realizamos la selección de la información que encontramos en esta tabla y la llevamos a una hoja de Excel.
- Hacemos clic en ANÁLISIS DE DATOS en el grupo ANÁLISIS de la ficha DATOS. Si el comando ANÁLISIS DE DATOS no está disponible, debemos hacer CLIC en el botón y en OPCIONES EN EXCEL seleccionamos COMPLEMENTOS y damos CLIC en IR. A continuación, aparece un cuadro de diálogo en el que activamos la opción HERRAMIENTAS PARA ANÁLISIS y finalmente hacemos CLIC en ACEPTAR. El proceso anterior nos permite ver el ícono ANÁLISIS DE DATOS en la tabla de cálculo, tal como se presenta a continuación:

Figura No. 2. Selección Comando Análisis de Datos

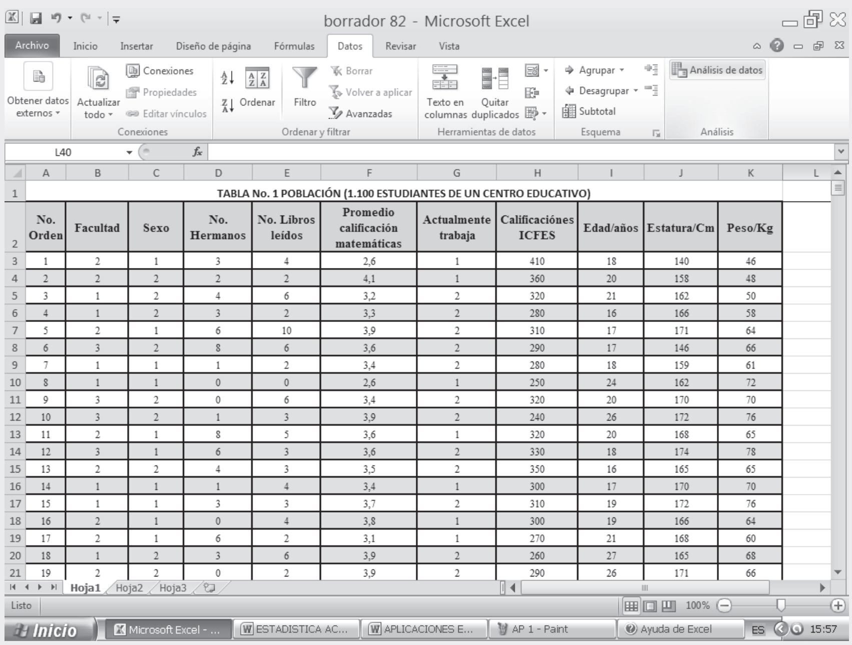
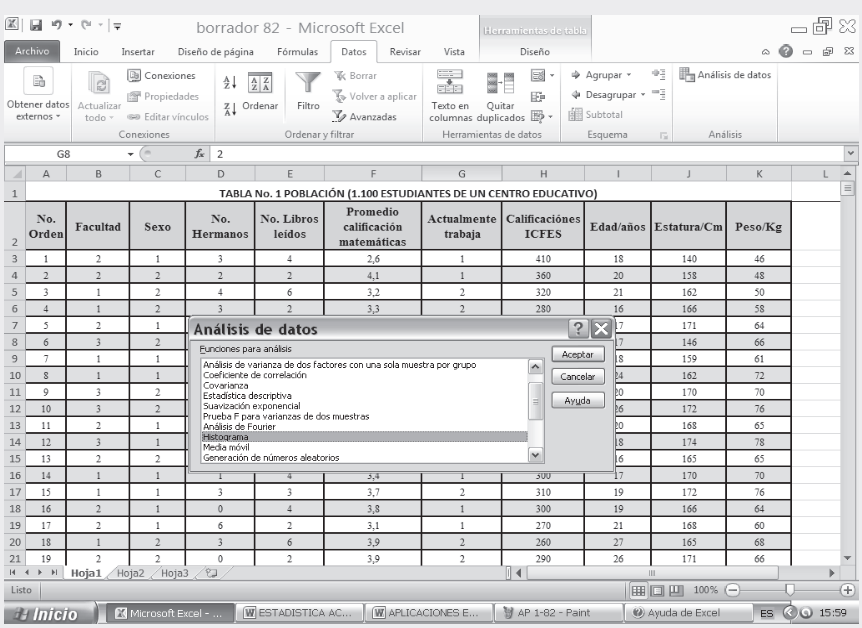


TABLA No. 1 POBLACIÓN (1.100 ESTUDIANTES DE UN CENTRO EDUCATIVO)

No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciónes ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1	3	4	2,6	1	410	18	140	46
2	2	2	2	2	4,1	1	360	20	158	48
3	1	2	4	6	3,2	2	320	21	162	50
4	1	2	3	2	3,3	2	280	16	166	58
5	2	1	6	10	3,9	2	310	17	171	64
6	3	2	8	6	3,6	2	290	17	146	66
7	1	1	1	2	3,4	2	280	18	159	61
8	1	1	0	0	2,6	1	250	24	162	72
9	3	2	0	6	3,4	2	320	20	170	70
10	3	2	1	3	3,9	2	240	26	172	76
11	2	1	8	5	3,6	1	320	20	168	65
12	3	1	6	3	3,6	2	330	18	174	78
13	2	2	4	3	3,5	2	350	16	165	65
14	1	1	1	4	3,4	1	300	17	170	70
15	1	1	3	3	3,7	2	310	19	172	76
16	2	1	0	4	3,8	1	300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
18	1	2	3	6	3,9	2	260	27	165	68
19	2	2	0	2	3,9	2	290	26	171	66

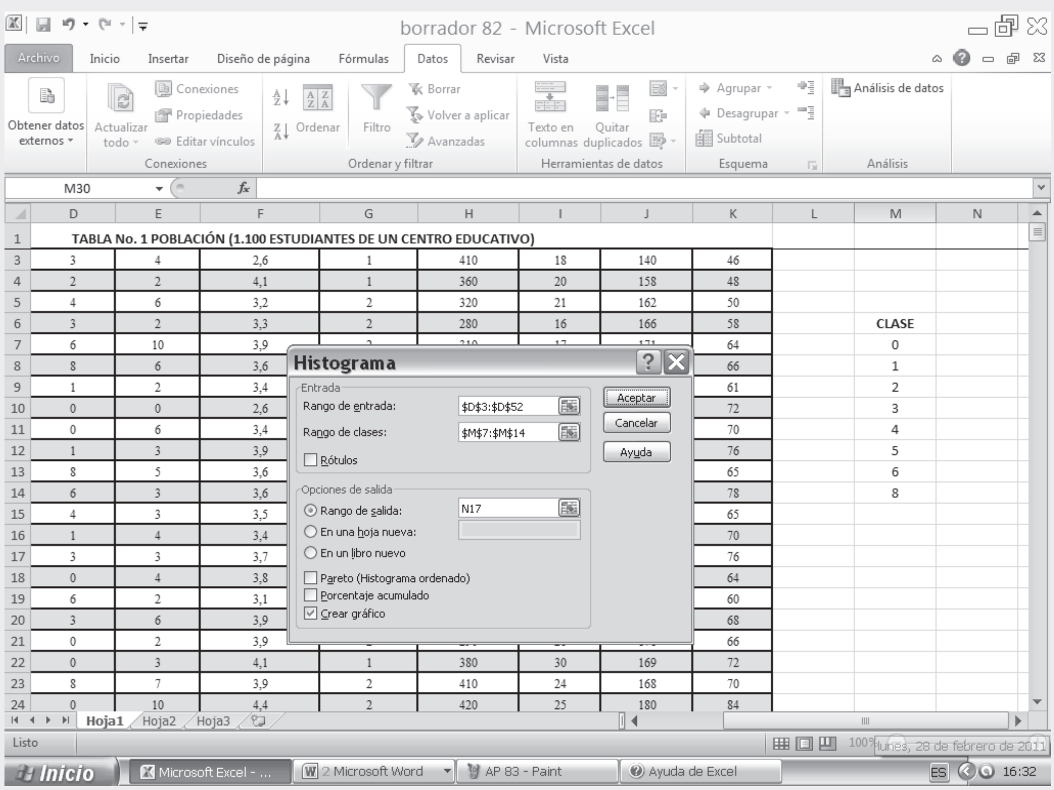
Figura No. 3. Opción Histograma



Con base en la cuarta columna denominada “Número de hermanos”, ubicada en la celda D, que aparece en la Tabla No.1 del ejemplo que seguimos, realizamos el HISTOGRAMA. Nos aparece un cuadro de diálogo denominado HISTOGRAMA, el cual diligenciamos tal como se especifica a continuación:

- ❑ En el RANGO DE ENTRADA seleccionamos el rango correspondiente a D3:D52
- ❑ Con el cursor nos ubicamos en el RANGO DE CLASES, que lo hemos establecido como M7:M14 (ver figura No. 1)
- ❑ Nos situamos en la casilla RANGO DE SALIDA y la activamos haciendo CLIC en el círculo que se antepone. Tecleamos en esta casilla la celda en la cual queremos que se presenten los resultados, en nuestro caso seleccionamos la celda N17, columna que presentará las distribuciones de frecuencias correspondientes.
- ❑ Si además del cuadro nos interesa crear el gráfico, EXCEL lo realiza como HISTOGRAMA, cuando en realidad es un gráfico de BARRAS. Para obtener la representación gráfica, se debe seleccionar la opción CREAR GRÁFICO.
- ❑ Finalmente hacemos CLIC en ACEPTAR.

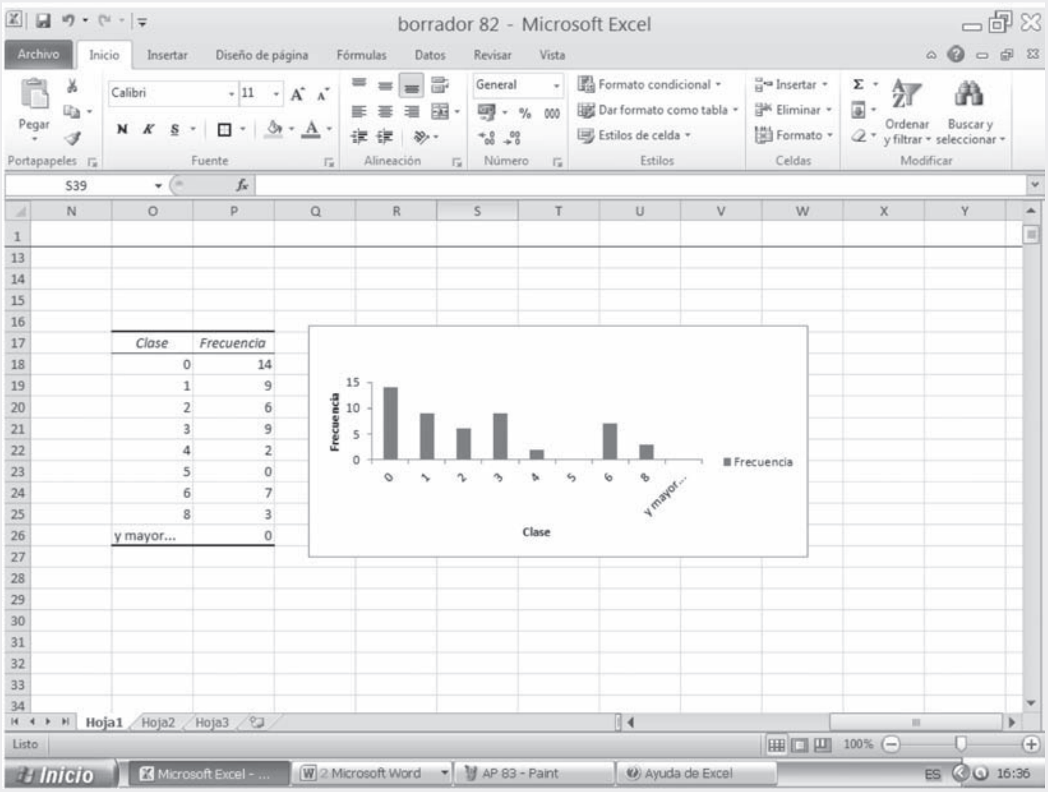
Figura No. 4. Datos de Histograma.



Una vez hecho el CLIC en ACEPTAR; obtenemos en pantalla los resultados esperados, es decir, se tendrá el cuadro de frecuencias con sus respectivos valores de clases, desde N17 hasta N27, a su lado debe aparecer la gráfica por haberse seleccionado la opción CREAR GRÁFICO.

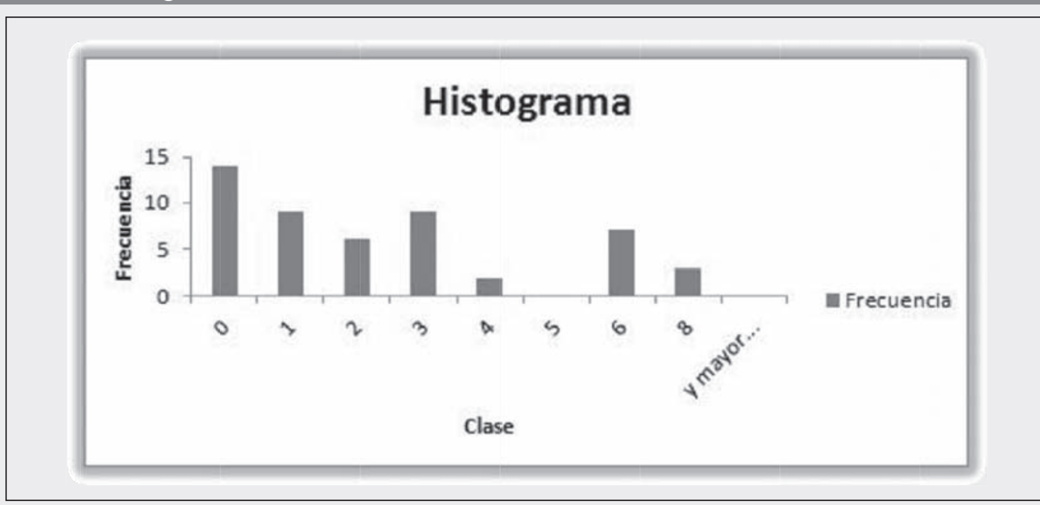
A continuación se presentan los resultados, en el cual aparece la TABLA DE FRECUENCIAS acompañada de un HISTOGRAMA similar a la GRÁFICA DE BARRAS.

Figura No. 5. Resultados Histograma.



Luego de finalizado el procedimiento anterior, se procede a trasladar la figura a un procesador de texto para efectos de presentación.

Gráfica 1. Histograma



Nota:

(1) La línea que puede aparecer en este tipo de gráfico, corresponderá a un DIAGRAMA DE FRECUENCIAS PORCENTUALES, cuando señalamos o hacemos CLIC en PORCENTAJE ACUMULADO.

(2) Observe que Excel incluye una última clase como mayor por si existen valores superiores.

(3) Como se puede observar, en vez de HISTOGRAMA, aparece una gráfica de barras. Se recomienda hacer el histograma en forma correcta utilizando la función gráficas de barras, al principio aparecen barras separadas y utilizando área SERIES ofrece dos alternativas (FILAS) (COLUMNAS) con una de ellas toma la forma adecuada.

GRÁFICAS UTILIZADAS EN LA PRESENTACIÓN DE INFORMES

Los gráficos, son utilizados con el fin de visualizar mejor la información, no sustituyen al CUADRO, antes por el contrario lo complementan. Algunos lo consideran como algo decorativo del informe, pero esto no es cierto, ya que ayuda a mostrar la información mucho más rápida y agradable, que la obtenida a través del TEXTO o del mismo CUADRO.

El EXCEL es una buena herramienta para la elaboración de GRÁFICOS, ofreciéndonos variadas formas de presentación, tales como:

COLUMNAS; BARRAS; LINEAL; CIRCULAR; DISPERSIÓN; ÁREAS; ANILLOS RADIAL; SUPERFICIE; BURBUJAS; COTIZACIONES

De las anteriores opciones, las primeras cuatro (4) son las más utilizadas, las que a su vez son muy sencillas, para visualizar mejor la información.

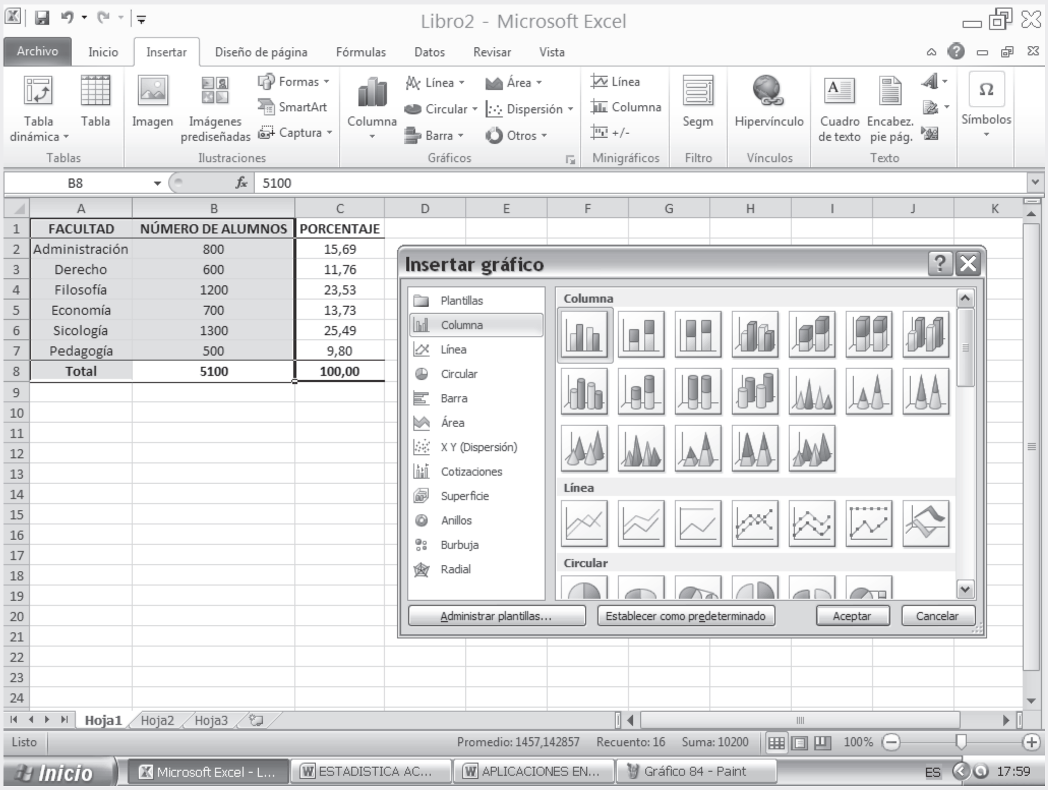
Ejemplo 1. Para indicar el proceso que se sigue en la elaboración de una GRÁFICA consideramos como información (cualitativa) la matrícula de 5.100 alumnos, distribuidos por facultades, en una institución universitaria de la capital.

Cuadro 2.1. Alumnos matriculados por facultad en el instituto X (primer semestre de 2011)

FACULTAD	NÚMERO DE ALUMNOS	%
Administración	800	15,69
Contaduría	600	11,76
Derecho	1.200	23,53
Economía	700	13,73
Ingeniería	1.300	25,49
Psicología	500	9,80
Total	**5.100**	**100,00**

- Digitamos la anterior información en una hoja de cálculo EXCEL, recomendando aumentar el ancho de las columnas (la primera) a fin de lograr una mejor presentación.
- Activamos las CELDAS que contienen la información o datos que deseamos graficar, incluidos los TÍTULOS y RÓTULOS. En este caso consideramos las columnas: **Facultades y Número de alumnos.**
- Procedemos a hacer CLIC sobre la pestaña del grupo GRÁFICOS en la ficha INSERTAR. Una forma más rápida consiste en seleccionar directamente sobre el grupo GRÁFICOS el que más se adapte según el tipo de información que tengamos

Figura No. 6. Insertar Gráfico.



- ❑ Los datos pueden estar organizados en FILAS o COLUMNAS. Debemos tener en cuenta el rango de datos que vayamos a seleccionar; en este caso, seleccionamos el rango de A2:B8.
- ❑ Seleccionamos el TIPO DE GRÁFICO que deseemos crear, haciendo CLIC en él. En nuestro caso consideramos el gráfico de COLUMNAS, optando por uno de los DISEÑOS de las diecinueve (19) posibilidades que se presentan.
- ❑ Después de crear un gráfico, es posible modificar cualquiera de sus elementos en la ficha HERRAMIENTAS DE GRÁFICO, de acuerdo a las pestañas de DISEÑO, PRESENTACIÓN y FORMATO. De este modo, podemos cambiar la forma en que se presentan los ejes, el título del gráfico, mover u ocultar la leyenda así como incluir elementos de gráfico adicionales. Esto nos permite AGREGAR o QUITAR aquello que deseamos que aparezca o desaparezca.
- ❑ El gráfico se puede ver de dos formas, como objeto y como una hoja nueva.
- ❑ Como una HOJA NUEVA: Mostrando el gráfico en la misma hoja, denominada HOJA DE GRÁFICO, hacemos CLIC sobre el gráfico ya insertado y en la pestaña de DISEÑO, encontramos el fichero UBICACIÓN; Hacemos CLIC en MOVER GRÁFICO y elegimos la opción HOJA NUEVA.
- ❑ Como OBJETO, seguimos el procedimiento anterior sólo que esta vez elegimos la opción OBJETO EN, especificando la hoja en la que deseamos insertar el gráfico. En este caso el gráfico no genera una hoja nueva sino un cambio de ubicación, por ejemplo de hoja 1 a hoja 2. Lo anterior se presenta en la figura siguiente:

Figura No. 7. Gráfico de Columnas

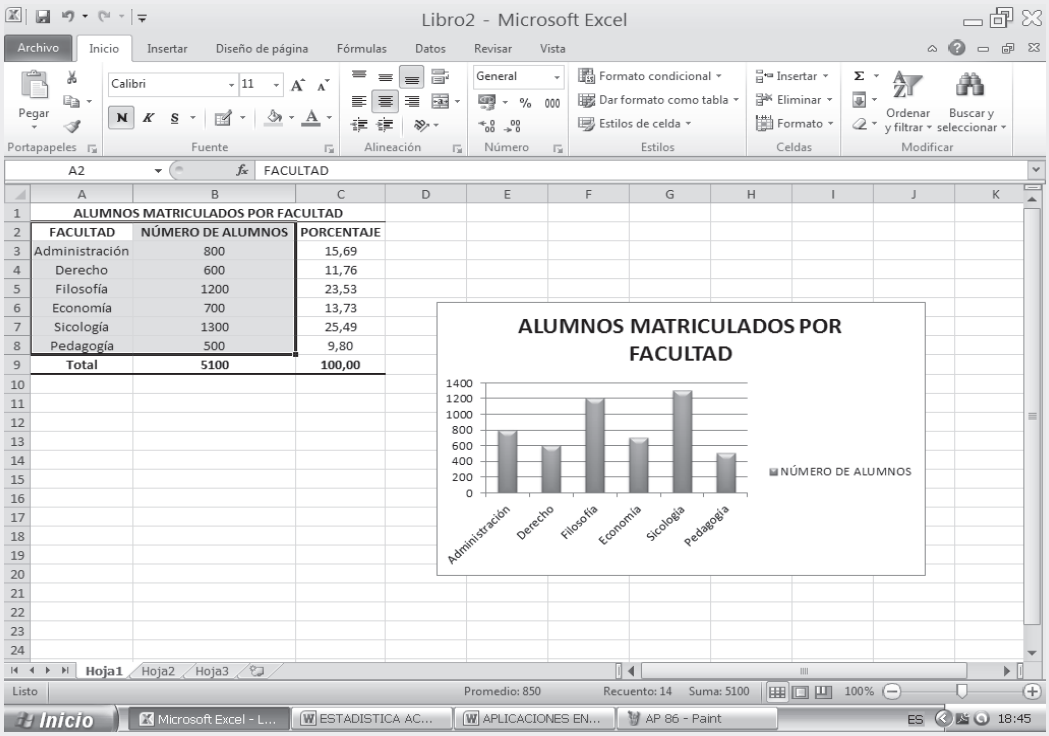
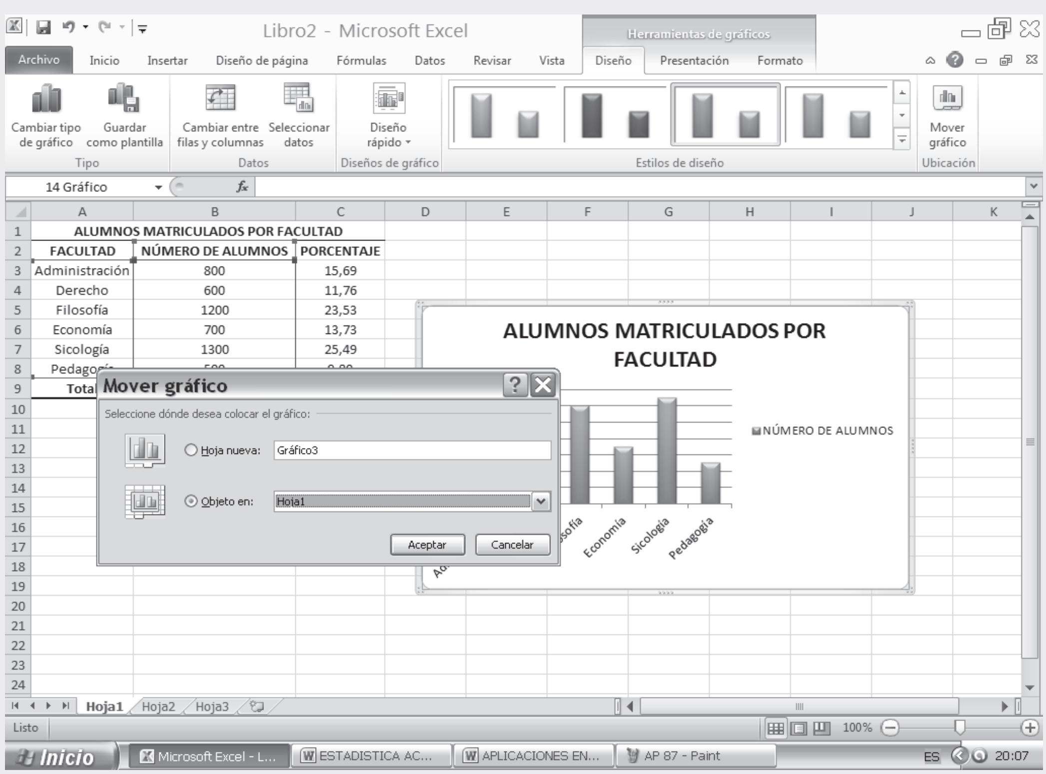


Figura No. 8. Opciones de pegado de Gráfico



- Finalmente, mejoramos la presentación del gráfico, haciendo CLIC en el área del gráfico, le damos opción COPIAR, y luego lo pegamos donde quiera que estemos realizando el informe o trabajo tal como se presenta a continuación:

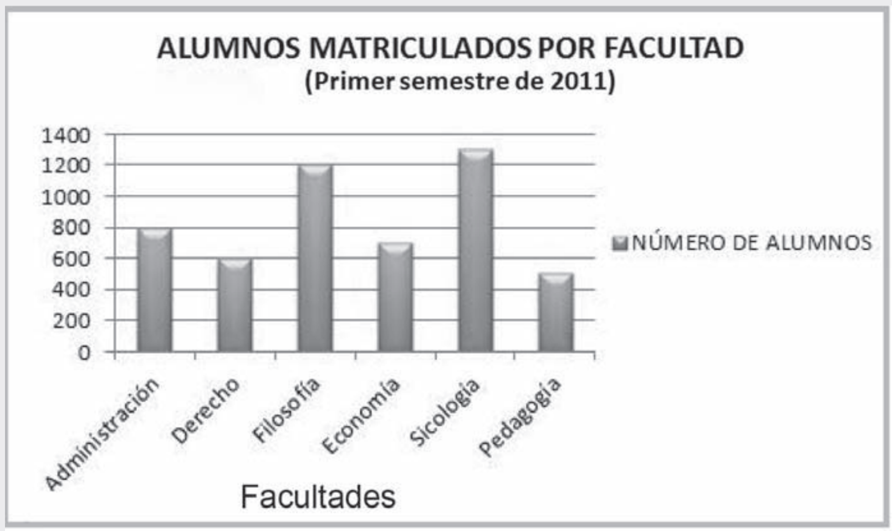


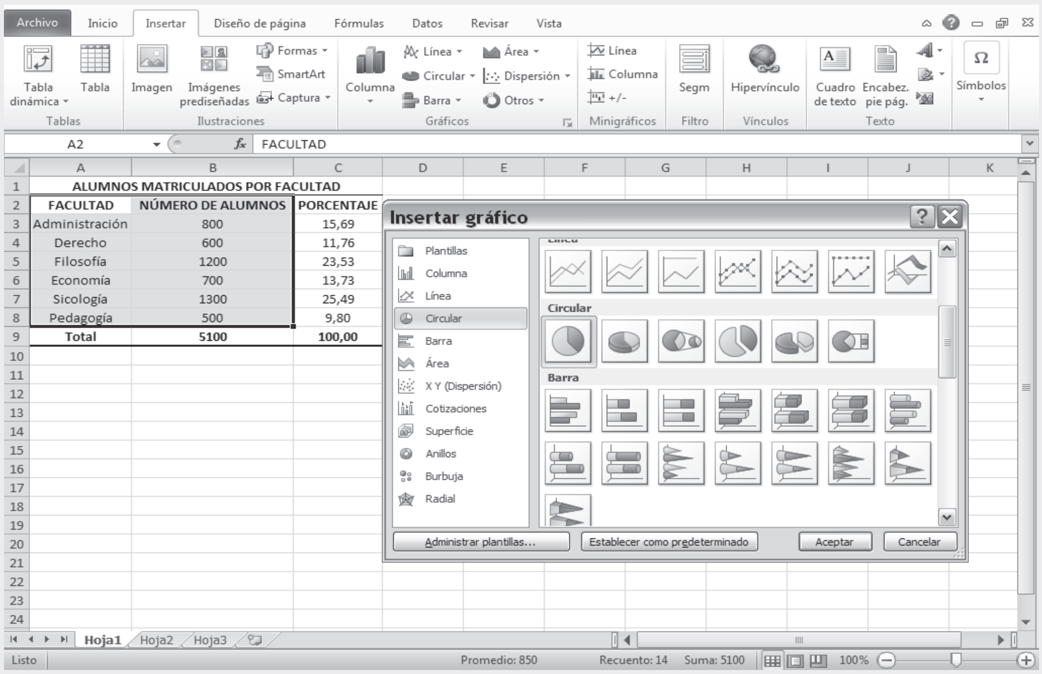
GRÁFICO CIRCULAR O PASTEL

El uso de este gráfico CIRCULAR, se hace muy conveniente cuando la(s) característica(s) que se quieren visualizar son ATRIBUTOS o CUALITATIVOS, tal como en el ejercicio que se realizó anteriormente al aplicar la gráfica de BARRAS, para lo cual consideramos la matrícula de 5100 alumnos en una institución universitaria de la capital.

Vale la pena tener en cuenta que este tipo de gráfico no es muy indicado en aquellos casos donde la característica presenta numerosos aspectos, igual como en el ejercicio que nos ocupa, si la institución tuviera más de 10 facultades, esto haría un gráfico muy recargado y no se lograría una buena visualización de la información. El proceso que se sigue en la graficación es muy similar al de las barras. Veámoslo:

- Tomamos la misma información para la gráfica de barra.
- Activamos las CELDAS respectivas, incluyendo los TÍTULOS, en este caso FACULTADES y NÚMERO DE ALUMNOS.
- Nos ubicamos en la ficha INSERTAR haciendo CLIC en la pestaña GRÁFICOS para cambiar el tipo de gráfico.
- Seleccionamos CIRCULAR y uno de los SUBTIPO DE GRÁFICO. En nuestro ejemplo, optamos por el primero de las seis (6) opciones que nos ofrece y hacemos CLIC en ACEPTAR
- De ahí en adelante seguimos los pasos similares a los mostrados en el GRÁFICO DE BARRAS, de acuerdo con la siguiente presentación, tal como se puede observar en las aplicaciones que aparecen a continuación:

Figura No. 9. Insertar Gráfico.



❑ Seleccionando la opción HOJA NUEVA, nos resulta la siguiente GRÁFICA en una hoja completa.

Figura No. 10. Resultados de Salida

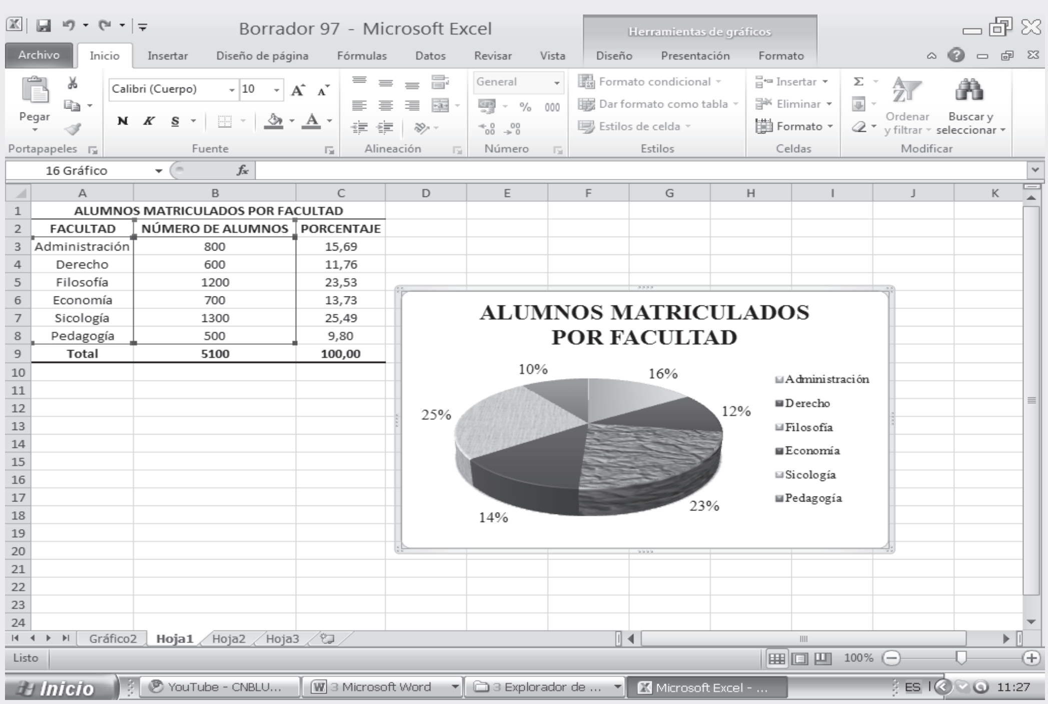
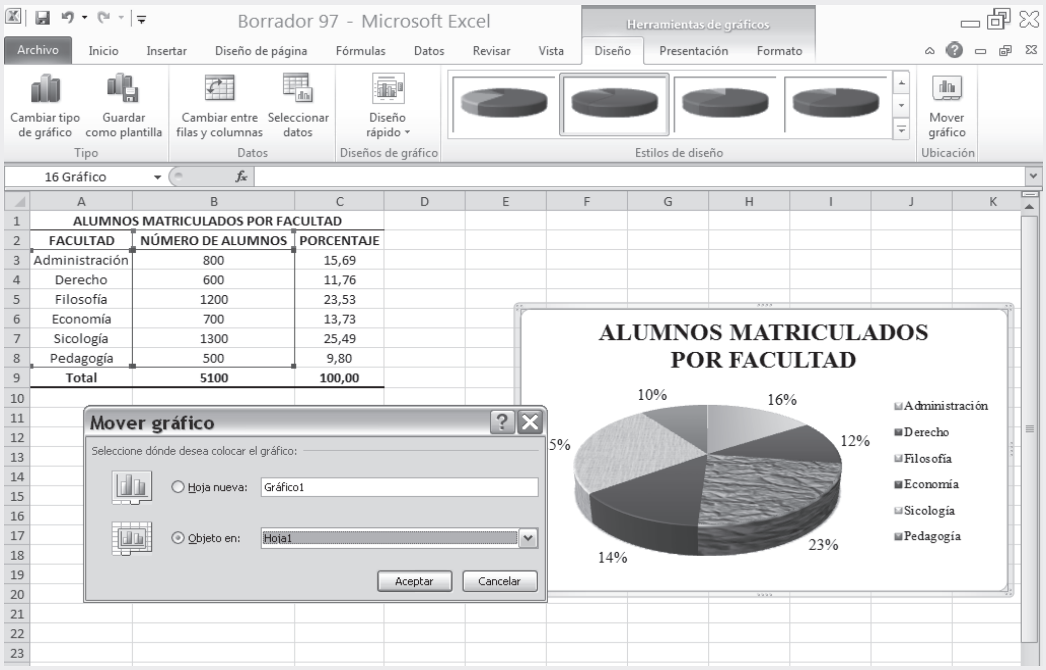


Figura No. 11. Opciones de Pegado



- Si tecleamos la opción de ETIQUETA DE DATOS, es posible mostrar la distribución que presentan las diferentes categorías. Esto es, nos indica la cantidad de ALUMNOS que hay por cada facultad.
- Finalmente, se puntea la gráfica y se pega en el informe o tarea que se esté realizando.

Figura No. 12. Gráfico Circular en una Hoja Nueva.

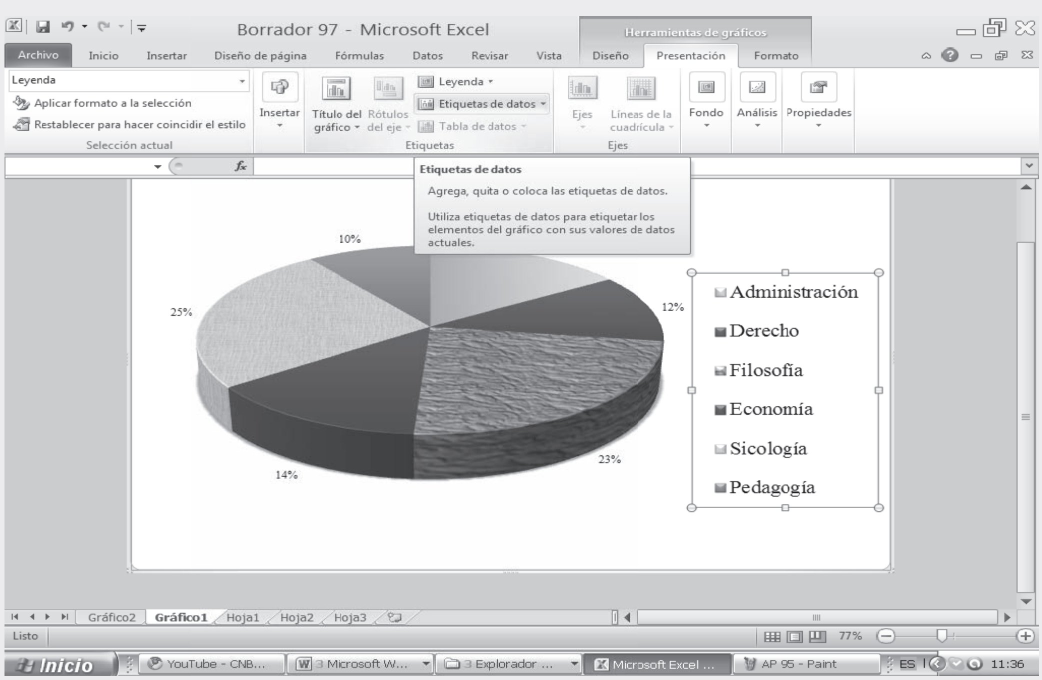


Gráfico No. 3. Gráfico Circular.

GRÁFICO DE LÍNEAS

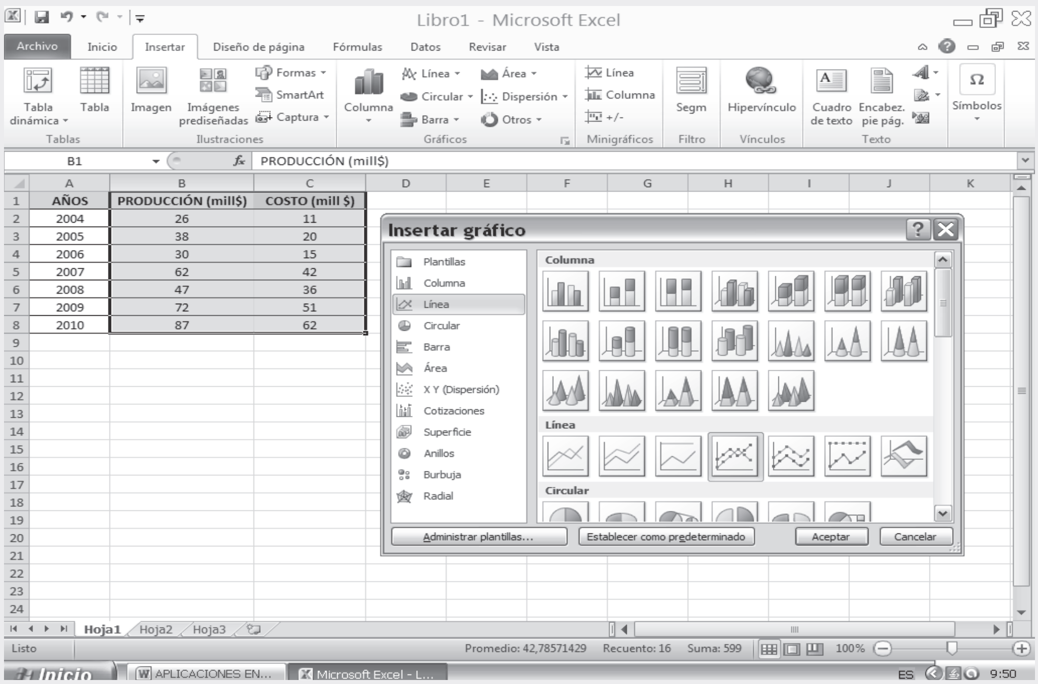
Son de gran aplicación en las denominadas SERIES DE TIEMPO O SERIES CRONOLÓGICAS, donde una de las variables corresponde al TIEMPO (*X*) (años, meses, días, etc.) y la segunda la variable investigada (Y), en nuestro caso PRODUCCIÓN (ciento mill $) y/o COSTO (ciento mill $). Recordemos que en el eje *horizontal o abscisa* van los valores correspondientes a la *variable tiempo* y en el eje vertical u ordenada los valores de la variable (Y) *producción y/o costo.*

Cuadro 2.2 Producción y Costo (ciento mill $) en la empresa X

AÑOS	PRODUCCIÓN (ciento mill $)	COSTO (ciento mill $)
2004	260	110
2005	380	200
2006	300	150
2007	620	420
2008	470	360
2009	720	510
2010	870	620

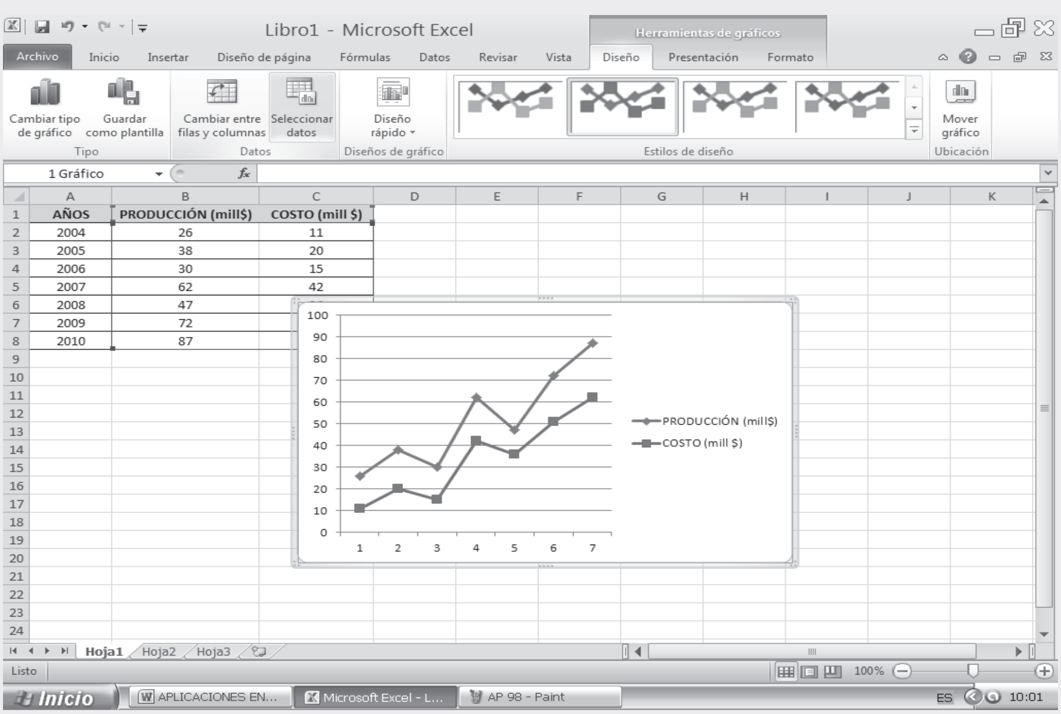
- ❑ Una vez construida la tabla en una hoja electrónica EXCEL, se deben activar las celdas correspondientes a PRODUCCIÓN y COSTOS, al momento de hacer la selección de los datos de gráfico.
- ❑ En la ficha INSERTAR hacemos CLIC en el grupo Gráficos y seleccionamos el tipo de gráfica LÍNEAS.
- ❑ Activamos el subtipo de gráfico cuatro (4) de las siete (7) opciones que nos muestran.

Figura No. 13. Insertar Gráfico.



Nota: Importante para agregar los AÑOS en el eje vertical, luego de incrustada la gráfica, vamos al menú HERRAMIENTAS DE GRÁFICO y en la pestaña DISEÑO seleccionamos el ÍCONO SELECCIONAR DATOS

Figura No. 14. Resultados de Salida



En la opción ETIQUETAS DEL EJE HORIZONTAL que se encuentra dentro del cuadro SELECCIONAR ORIGEN DE DATOS, elegimos la opción EDITAR y seleccionamos el rango de datos correspondiente a los AÑOS.

Figura No. 15. Selección Origen de Datos.

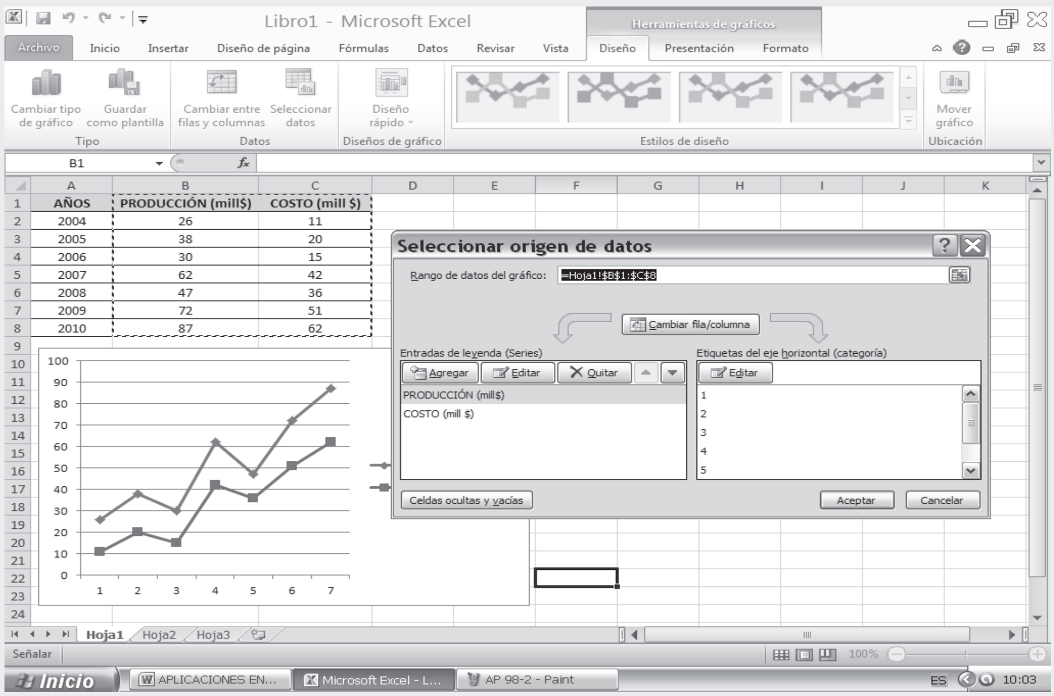


Figura No. 16. Selección Años de la producción y costo de la Empresa X. (Pantalla Adic.)

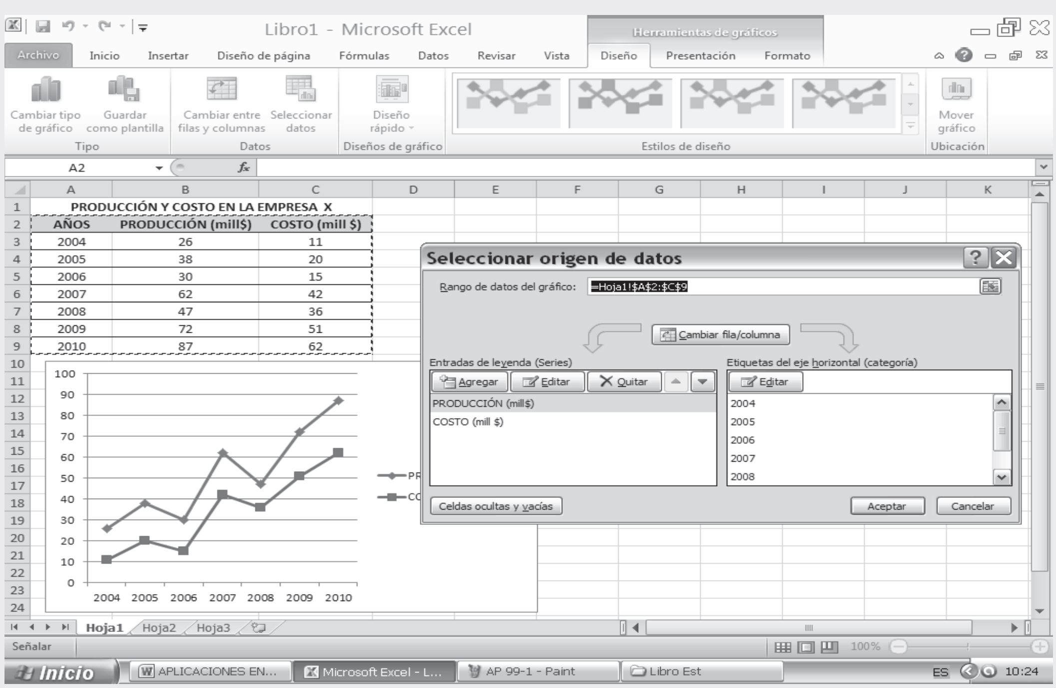
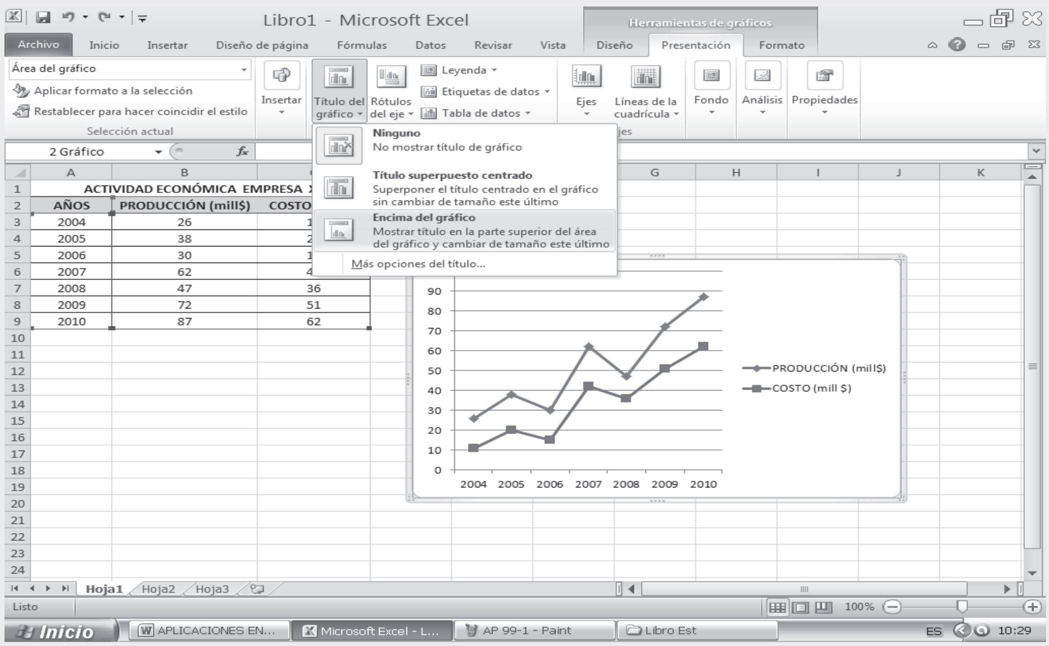
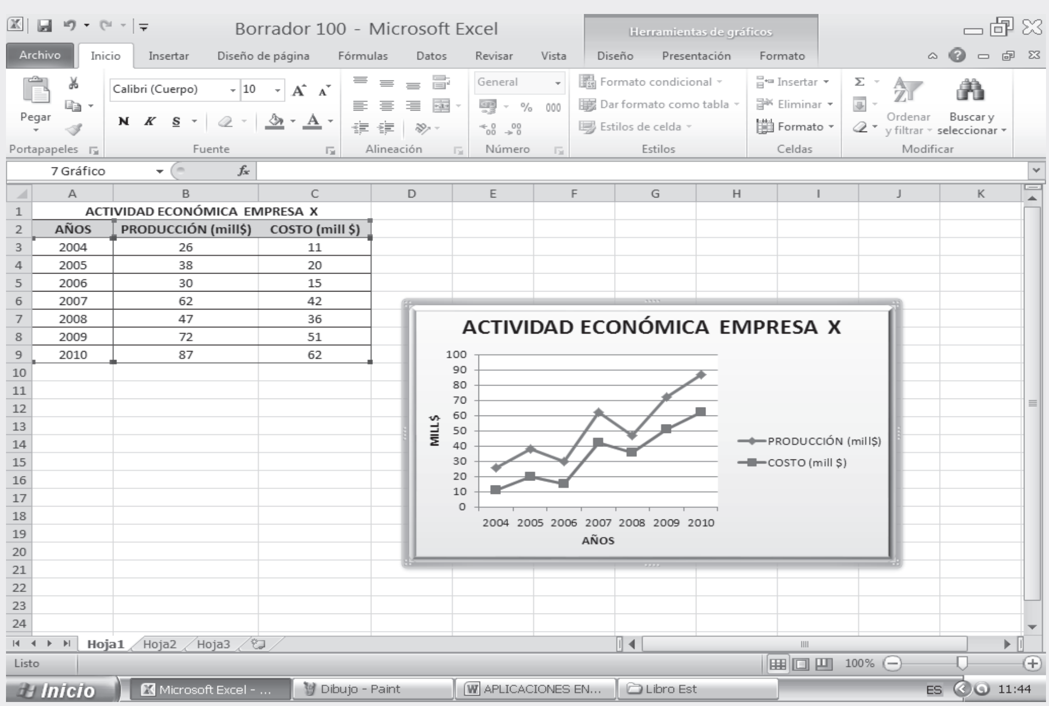


Figura No. 17. Título del Gráfico. (Pantallazo Adicional).



- Finalmente, agregamos el titulo y hacemos los ajustes de diseño que consideremos pertinentes.

Figura No.18. Resultados Gráfica de Líneas.



- Luego de este paso se selecciona el gráfico, y se copia en el informe o tarea que se esté realizando.

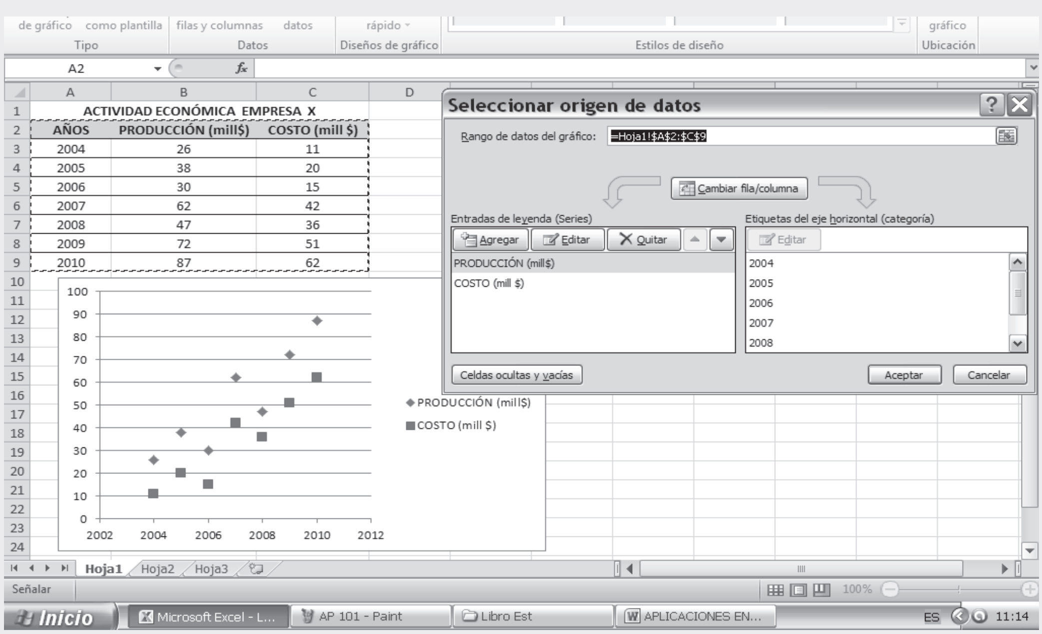
Gráfica 3. Gráfica de Líneas.

NUBE DE PUNTOS O DIAGRAMAS DE DISPERSIÓN

Esta gráfica es utilizada en el ANÁLISIS DE REGRESIÓN (ver capítulo 10), es decir cuando se trabaja simultáneamente con dos variables, a fin de visualizar mejor la información y decidir cuál es la función (lineal, parabólica, exponencial, logarítmica) que más se ajusta a ese conjunto de puntos

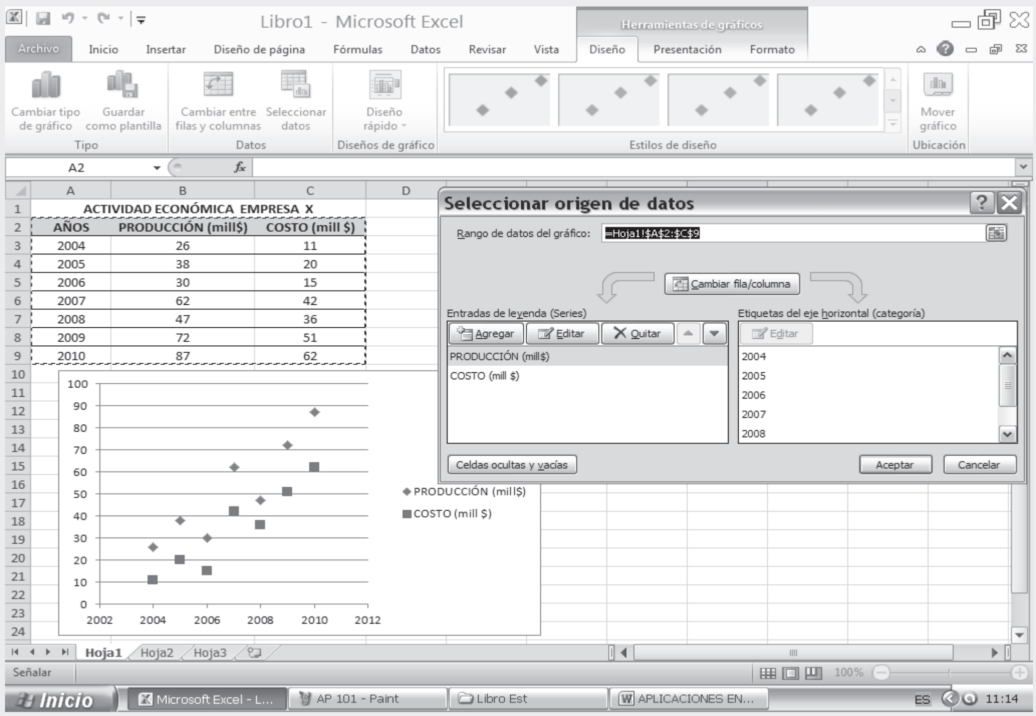
- ❑ Con los datos utilizados en el cuadro No. 2.2 (ver página 97), seleccionamos las variables PRODUCCIÓN y COSTO para elaborar la representación gráfica.
- ❑ En el cuadro de diálogo INSERTAR GRÁFICO que nos aparece, seleccionamos la opción XY DISPERSIÓN y hacemos CLIC en el primer (1) SUBTIPO DE GRÁFICO de las 5 opciones posibles.

Figura No. 19. Insertar Gráfico.



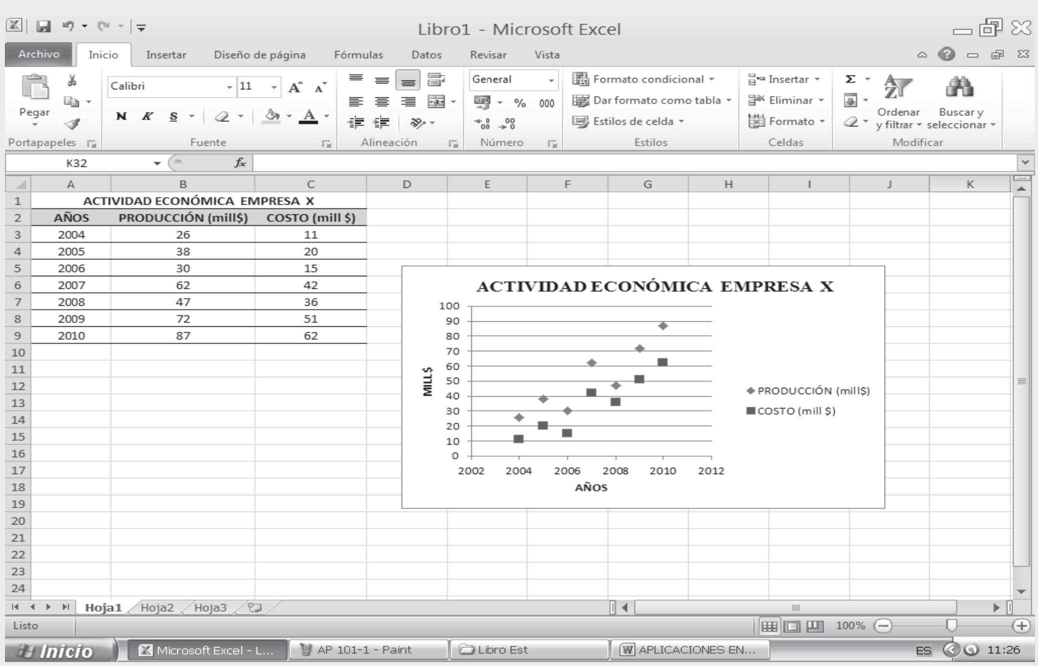
- De los datos de origen se obtienen los siguientes resultados:

Figura No.20. Salida de Resultados.



- Ahora procedemos a colocar el título del gráfico y el de los ejes, como se puede observar en la aplicación siguiente:

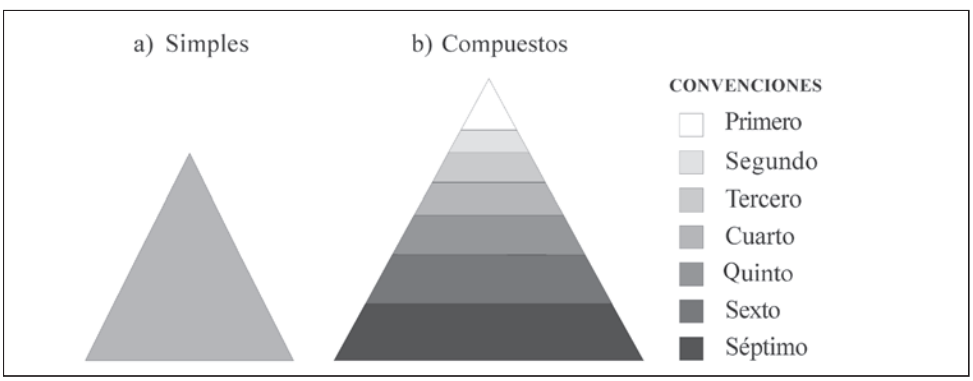
Figura No. 21. Inserción Título del Gráfico y de los Ejes.



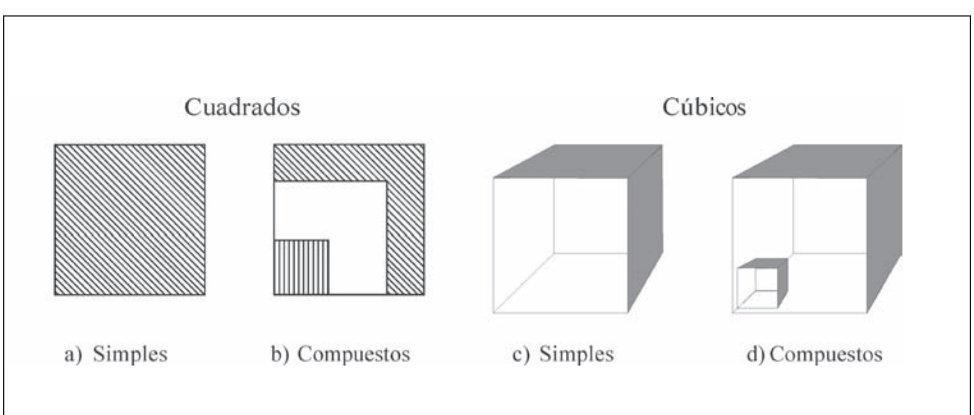
Gráfica No. 4. Gráfico de Dispersión XY

OTRAS GRÁFICAS (NO SE ENCUENTRAN EN EL PROGRAMA EXCEL)

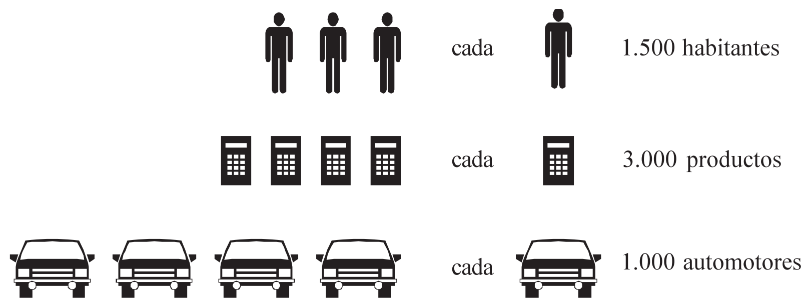
Gráficas triangulares



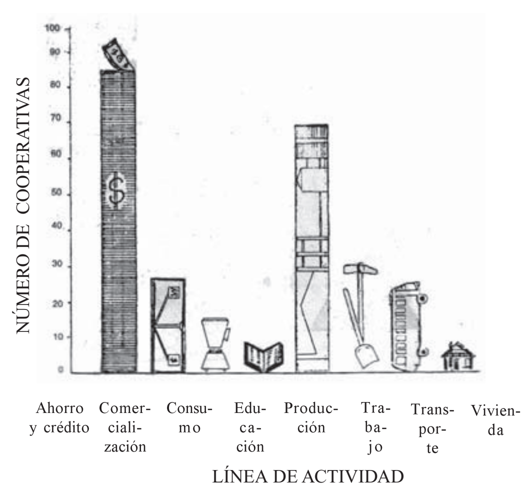
Gráficos cuadrados



Pictogramas



Número de entidades Cooperativas creadas según línea de actividad
de julio 1º de 2006 a 30 junio de 2011



Cartogramas

Gráficas semilogarítmicas

Estas gráficas se emplean en casos concretos, cuando se requiere destacar las variaciones relativas más que las absolutas; tradicionalmente utilizamos la *escala aritmética* espaciando los valores de la escala vertical (ordenada) en proporción a las diferencias absolutas; en cambio en la *escala logarítmica* los espacios corresponden proporcionalmente a las diferencias logarítmicas, siendo menores estos espacios. Consideremos la representación de un determinado índice para un período, siendo inicialmente de 400 y al finalizar el período, de 460; el incremento en valor absoluto es de 60; pero si tenemos otro índice que representa otra variable con la cual estamos comparando para el mismo período, iniciando con 800 y finalizando en 920, podemos observar que el incremento es de 120, es decir el doble del anterior, pero porcentualmente el incremento en ambos casos es igual, es decir, del 1 %. Estos dos indicadores representados gráficamente en una *escala aritmética* (valor absoluto) muestran la variación, en cambio en la *logarítmica* representan la misma magnitud siendo esta última la que concuerda con la realidad. Es necesario anotar en la gráfica, que solamente la escala vertical es de tipo logarítmico y la horizontal es aritmética.

Para la confección de un gráfico *semilogarítmico* se requiere desarrollar los siguientes pasos:

a) Se debe establecer el valor máximo y mínimo que tomará la escala vertical, en este caso hacemos referencia al índice (no hay problema, si los valores tomados en la escala son ligeramente superiores e inferiores a los extremos obtenidos en la información a representar).

b) Se calcula la diferencia de los logaritmos de los valores máximo y mínimo obtenidos en el paso anterior.

c) Se establece cuántos centímetros tendrá la escala del gráfico correspondiente.

d) Se divide el valor dado en centímetros, establecido en el punto c, por el resultado obtenido en el punto b, es decir, (c/ b).

e) Multiplicamos el logaritmo correspondiente al valor máximo dado en la escala, en el punto a, por el resultado obtenido en el punto d y se le resta el valor del punto c.

f) Luego, para cada valor (en nuestro caso el índice) de la tabla se obtiene su respectivo logaritmo y se le multiplica por el valor del punto d y se le resta el valor obtenido en el punto.

g) Los valores resultantes son dados en centímetros y corresponden a la ordenada por cada valor que tome la abscisa (eje horizontal).

Gráficas logarítmicas

El gráfico logarítmico, se emplea cuando las dos variables muestran un crecimiento geométrico, por lo tanto en ambos ejes (ordenada y abscisa) se utilizará la escala logarítmica y sirve para comparar los cambios relativos de una variable, conforme a los cambios relativos de otra variable. Esta representación es muy usual en economía cuando se trabaja con oferta y demanda, precios relativos y cantidades, bien sea de oferta o de demanda; en estos dos ejemplos, estamos comparando los valores relativos de una con respecto a valores relativos de otra.

Se dice, que si las dos variables relacionadas presentan un crecimiento geométrico se deberá utilizar el papel logarítmico, pero si una de ellas presenta un crecimiento aritmético mientras que la otra variable lo hace en forma geométrica se empleará el papel semilogarítmo, con lo cual se logra una representación gráfica correcta.

Gráficas Gantt

Tienen como fin presentar las tareas o etapas de un proceso, relacionándolas con el tiempo de su ejecución, de gran interés en el planeamiento o actividad empresarial, siendo su aplicación tan amplia y de tanto éxito que puede ser desarrollada en una empresa que tenga 2 ó 3 obreros, como en aquellos establecimientos comerciales o industriales de gran magnitud.

- ❑ Los elementos que conlleva a la elaboración del gráfico Gantt es tan sencillo, que es fácil de comprenderlo, representando a la vez el espacio, tiempo y la tarea a ejecutar. Podríamos clasificar estos gráficos en tres formas:
- ❑ Los que tratan de resaltar la relación existente entre lo realizado y lo que podría haberse hecho, tal es el caso de un obrero o máquina al realizar una determinada labor, podríamos estar determinando el tiempo inactivo.
- ❑ Cuando se desea establecer un reparto equitativo de la tarea con el fin de evitar la inactividad (obrero o máquina) y de esta manera lograr su ejecución según el orden establecido de antemano.
- ❑ Para comparar de una manera continua la ejecución de una tarea de acuerdo al plan trazado con anterioridad y dar a conocer las causas que han impedido su cumplimiento.

Gráficas de volumen

Constantemente observamos en los diferentes medios de comunicación (periódicos, revistas, T.V.) la utilización de gráficas de volúmenes y pictogramas, en estos últimos, como ya lo vimos se emplean figuras tales como personas, animales, fábricas, aviones, casas, etc., debiendo en ambos casos ser proporcionales a las cantidades para que sean representativos.

Veamos un pequeño ejemplo en el cual se quieren representar los gastos de una pequeña empresa, tomando tres aspectos: salarios, publicidad y alquiler, utilizando cubos. En este caso si se conoce o se establece la longitud de un lado (arista), se podrá calcular el volumen siendo a^3, supongamos que la arista para el cubo que representa los salarios es de 4 cms., se tendrá que el volumen es $4^3 = 64$ cm^3, que representa a un gasto en salarios de 40 millones de pesos; ahora bien, si el gasto en publicidad es de 8 millones de pesos se tendrá que $40/8 = 5$ veces más pequeño proporcionalmente y su volumen será $64/5 = 12,4$ cm^3; finalmente consideramos alquileres por un valor de $5 millones siendo $40/5 = 8$ veces más pequeño y su volumen será $64/8 = 8$ cm^3. De acuerdo a los resultados anteriores la arista para cada caso será igual:

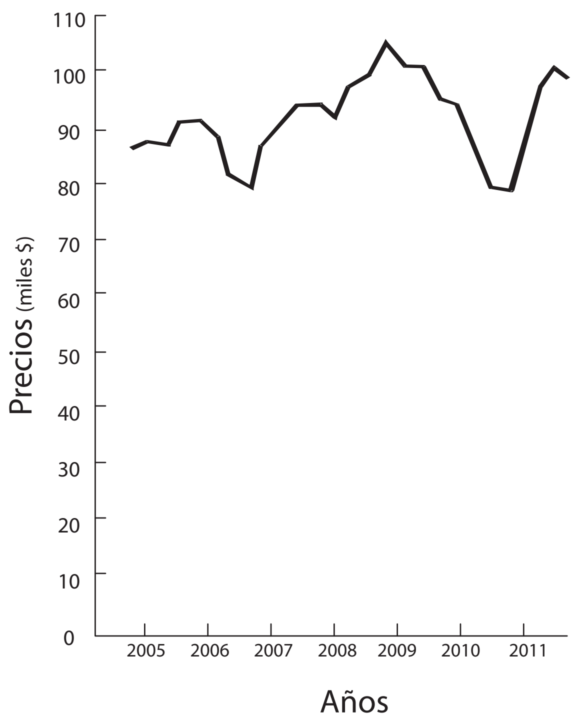
$$\text{Salarios} = \sqrt[3]{64} = 4\text{cms.} \qquad \text{Publicidad} : \sqrt[3]{12,4} = 2,31\text{cms.}$$

$$\text{Alquileres} = \sqrt[3]{8} = 2,20\,\text{cms.}$$

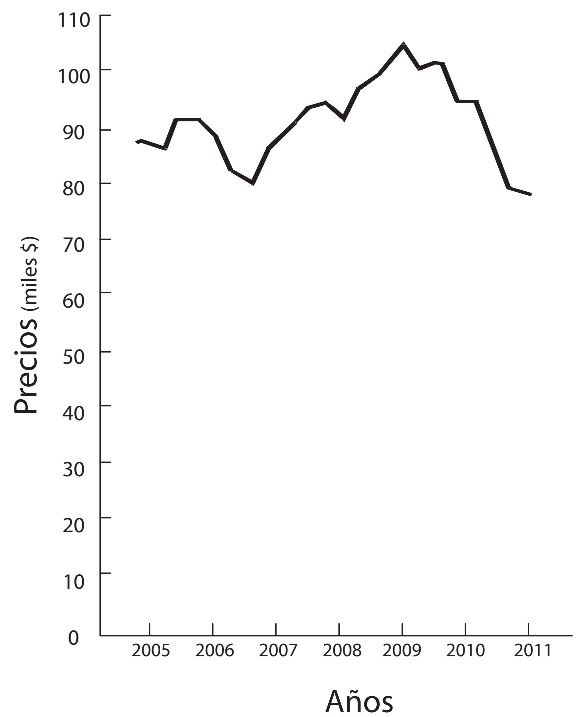
Gráficas engañosas

Con frecuencia se presentan gráficas engañosas al falsear una de las escalas, veamos estas dos representaciones gráficas lineales presentadas en forma desfavorable, casi con fluctuaciones suaves tratando de disimular las bajas súbitas.

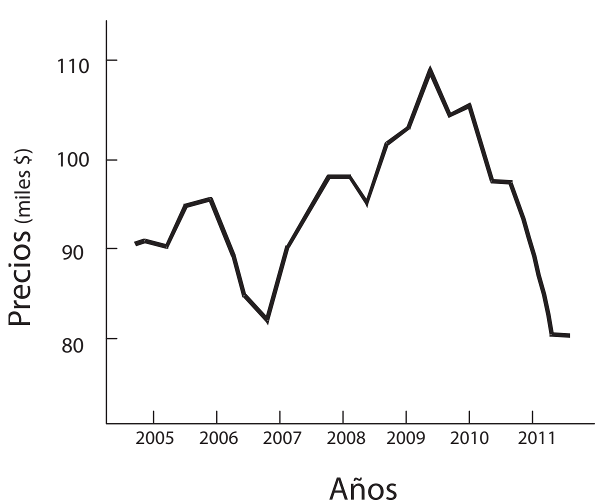
Gráfica A



Gráfica B



Gráfica C



Se puede observar en la GRÁFICA B, como al omitirse el último año hay cambio total con respecto a la anterior, desaparece la recuperación que ofrece la variable para el año 2011, dejando la sensación de que la tendencia de esa serie es a la baja. En ambos casos las escalas están mal trazadas y las gráficas mal presentadas.

Veamos ahora otras situaciones que se podrían presentar con los mismos datos, distorsionando totalmente la información que se desea presentar en forma clara y objetiva. La gráfica C es la forma correcta de presentación de la serie.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

34. Contestar verdadero si el enunciado es verídico. Si no lo es, la palabra en negrilla debe sustituirse por una expresión con la cual el enunciado sea válido.
    a) Un cuadro o tabla es la mejor forma de visualizar la información.
    b) El recuento de los empleados en una empresa, de acuerdo al cargo, es un ejemplo de variable discreta.
    c) El pictograma es la representación estadística por medio de mapas, dentro de los cuales se ubican símbolos.
    d) El histograma es un tipo de representación en gráfica de rectángulos de una distribución continua.
    e) El grosor de la lámina de metal que la compañía A utiliza en sus procesos de manufactura es un ejemplo de atributos.
    f) La población o universo es un subconjunto de unidades extraídas al azar en tal forma que nos permita tener información sobre un grupo mayor.

35. El costo estimado, en miles de pesos ($) por kilómetro, para la operación de un automóvil (Renault, Mazda, Chevrolet) modelo 2002, se muestra en la siguiente tabla:

Depreciación	$236
Mantenimiento	212
Gasolina	85

Seguros	$1060
Parqueadero	38
Impuestos	37

Construya una gráfica circular y otra de barras

36. ¿Qué ventajas tienen las gráficas en un resumen de datos?

37. Represente los siguientes datos en millones de $ en ventas:

AÑOS	2006	2007	2008	2009	2010	2011
y_i	2.000	1.800	4.600	1.200	4.600	8.230

38. Representar los siguientes datos en millones de $, en costos y ventas:

AÑOS	2007	2008	2009	2010	2011
y_i	120	180	75	240	200
x_i	150	170	100	200	250

39. Opinión de los electores sobre si van a votar para Congreso y Presidente en las próximas elecciones. Se pide elaborar el gráfico correspondiente:

OPINIÓN	CONGRESO %	PRESIDENTE %
Afirmativo	22,24	48,32
Negativo	77,76	51,68
Total	100,00	100,00

40. Elabore la gráfica correspondiente.

AÑOS	2007	2008	2009	2010	2011
Producción (miles ton.)	380	450	700	740	720
Ventas (miles ton.)	260	350	600	520	508

41. Con los siguientes datos presentados en una tabla de frecuencias:

$y'_{i-1} - y'_i$	n_i
4,1 – 24	36
24,1 – 32	20
32,1 – 36	18
36,1 – 48	22
48,1 – 52	14
Σ	110

Se pide elaborar:
a) Un histograma
b) Un polígono de frecuencias

42. Suponga que los precios de 5 artículos para dos períodos han sido:

ARTÍCULOS	2010	2011	VARIACIÓN % 2010 – 2011
Vehículos (usados)	$12.500.000	$11.300.000	–9,60
Computadores	1.700.000	1.100.000	–35,29
Betamax	250.000	350.000	40,00
Celular	1.200.000	1.500.000	25,00
Antenas parabólicas	300.000	410.000	36,67

Elaborar una gráfica de barras en dos direcciones (positivo y negativo) o en dos sentidos:

43. El consejo de administración de una gran cooperativa de vivienda desea investigar la posibilidad de contratar un supervisor para el parque de juegos infantiles. Se hizo una encuesta en la totalidad de las 616 familias de la cooperativa y cada familia tuvo un solo voto, cualquiera que fuera el tamaño del apartamento. Los resultados fueron:

¿Debe la cooperativa contratar un supervisor?

Si	146
No	91
No está seguro	158
No Respondió	221

Convierta los datos en porcentajes y construya:
a) Una gráfica de barras
b) Una gráfica circular
c) Una gráfica de componentes de porcentajes

44. Los siguientes datos corresponden al tiempo (minutos) que han necesitado 30 clientes de un Banco para llevar a cabo una transacción bancaria.

20	42	62	32	28	20	38	26	30	18	42	54	16	32	42
36	18	56	41	16	14	42	34	14	24	51	49	24	18	56

Se pide elaborar:
a) Una tabla de frecuencias
b) Histograma y el polígono en una sola gráfica.

45. Una empresa petrolera "Compañía X" en su informe anual mencionó las siguientes ventas netas y el costo de ventas desde 2006 (en miles de millones de pesos).

VARIABLE	2006	2007	2008	2009	2010	2011
Ventas netas	19.116	15.586	13.534	21.344	27.342	30.620
Costo de ventas	15.776	12.895	18.287	18.476	20.698	25.382

Se pide:
a) Elaborar un cuadro que represente esta información para ser publicada en una revista de la facultad.
b) Haga como mínimo dos gráficas que permitan representar la información anterior.

46. Con los siguientes datos correspondientes a una variable discreta, dibujar el diagrama de frecuencias, tanto para frecuencias absolutas como acumuladas.

y_i :	2	4	6	8	10	12	
n_i:	7	15	8	10	16	4	=60

47. Un almacén al cierre de sus ventas encuentra que el porcentaje de ellas por articulo se distribuye así: 42% camisas; 8% corbatas; 34% pantalones; 5% calcetines; 11% otros artículos. Se pide e1aborar:
a) Un gráfico de barras
b) Una gráfica circular.

48. Los siguientes datos corresponden a1 número de retardos (en la mañana) por parte de los empleados de una empresa, durante el mes:

2	5	0	0	6	3	5	0	0	2	7	4	4	2	3	8	4	3	3	1
3	2	3	2	4	2	2	3	1	3	2	0	0	1	4	2	2	0	3	5

Se pide elaborar una tabla de frecuencias y los diagramas correspondientes.

49. Las calificaciones de 30 alumnos que presentaron examen de admisión a la facultad, utilizando una escala de 0 a 100, fueron:

83	64	51	46	82	91	73	65	82	61	74	64	75	81	94
65	42	81	56	61	72	65	39	54	70	93	42	46	54	72

Elaborar un cuadro de frecuencias y dibujar los gráficos correspondientes (histograma, polígono y ojiva).

50. Una oficina de publicidad fue contratada con el fin de presentar en una revista de ventas, la distribución porcentual de ventas (valor) por jornadas (mañana, tarde, noche), realizada por 4 grandes almacenes de la capital, cuyos resultados fueron:

ALMACENES	MAÑANA (9 – 1)	TARDE (1 – 5)	NOCHE (5 – 10)
Carulla	22,5	42,6	34,9
Carrefour	38,6	45,4	16,0
Éxito	30,3	36,8	32,9
Vivero	46,4	32,8	20,8

Con la información del cuadro adjunto, se pide elaborar dos (2) gráficas, con el fin de seleccionar una, para que aparezca en la revista ¿Cuáles elaboraría usted?

51. Con los siguientes datos (miles mill $), el gerente de una compañía, le propone hacer dos gráficos, con el fin de escoger uno de ellos. ¿Cuales haría usted?

AÑOS	2007	2008	2009	2010	2011
Ventas	260	350	600	520	510
Producción	380	450	700	740	630

52. Una investigación realizada a 120 estudiantes en un instituto universitario requiere elaborar una gráfica.

DEPORTES	Ajedrez	Baloncesto	Balompié	Natación	Ciclismo	Tenis
ALUMNOS	13	47	62	34	20	8

53. Dibuja un histograma y al polígono de frecuencias, partiendo de la siguiente información obtenida de una tabla de frecuencias.

$m = 6$	$h_1 = 0{,}10$	$H_2 = 0{,}25$	$y_3 = 60$	$H_3 + H_4 = 0{,}95$
	$h_6 = 2(h_1)$	$h_4 = 0{,}20$	$n_1 = 4$	$c = 10$

$m = 6$	$f_1/n = 0{,}10$	$F_2/n = 0{,}25$	$X_3 = 60$	$F_3/n + F_4/n = 0{,}95$
	$f_6/n = 2(f_1/n)$	$f_4/n = 0{,}20$	$f_1 = 4$	$i = 10$

Síntesis de capítulo

Dá la sensación en numerosos informes, que el objetivo final es la presentación de *cuadros* o *tablas,* para mostrar lo dispendioso que ha sido al proceso de recolección y la abundancia de resultados obtenidos, sin preocuparse por lo que ellos digan. E1 procesamiento de la información es una de las etapas importantes en la elaboración de estadísticas, pues mediante ellas organizamos, relacionamos y en parte describimos el comportamiento de un conjunto de datos.

Un *cuadro* por si solo, no es explicativo, su elaboración se constituye en un instrumento que se emplea como complemento del texto explicativo; sin embargo esto no basta, pues a su vez debe estar acompañado de una buena *gráfica*. Esta última tiene la gran ventaja de visualizar mejor la información pero nunca sustituye al *cuadro*, ni al texto.

La *gráfica* por fallas en confección, generalmente por desconocimiento de las técnicas necesarias para una buena representación, se convierte en una herramienta que a menudo se utiliza para engañar incautos, produciendo gráficas aparentemente llamativas pero desorientadoras.

En este capítulo vimos unas gráficas generalmente utilizadas para exp1icar la teoría estadística, ellas son: los *diagramas* en la variable discreta; los *histogramas*, *polígonos* y *ojivas* en la variable continua. Otras gráficas, tales como las circulares, *pictogramas, barras, cartogramas*, etc., que se emplean en los informes y que son las que frecuentemente observamos en los diarios o revistas que presentan estadísticas.

3

CAPÍTULO

Medidas de posición y de tendencia central

Los sabios son los que buscan sabiduría: los necios piensan ya haberla encontrado.
Napoleón Bonaparte

CONTENIDO

- Términos, definiciones, conceptos generales.
- Finalidad, usos, aplicaciones, ventajas y desventajas de cada uno
- Promedios: Media, Mediana, Moda, Media geométrica, Armónica, Cuadrática y Cúbica.
- Cuartiles, Deciles y Percentiles. Centro recorrido
- Síntesis de la Unidad.
- Ejercicios para Resolver, resueltos en el **Sistema de Información en Línea SIL.**

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- Conocer, manejar y aplicar en cada caso particular el respectivo promedio.
- Entender el uso, las ventajas y desventajas que ofrece ese promedio.
- Saber aprovechar las propiedades que ellas tienen.

ASPECTOS GENERALES

En el capítulo anterior, se dijo que la *estadística* cumplía una función *descriptiva* haciéndolo mediante cuadros y gráficas, ahora lo complementamos con la aplicación de una serie de *medidas de posición* y de *dispersión*, que nos permiten la descripción de un hecho o de un conjunto de observaciones.

Las medidas de *posición* o de *tendencia central,* denominadas también como *promedios*, nos permiten determinar la posición de un valor respecto a un conjunto de datos, el cual lo consideramos como representativo o típico, para el total de las observaciones.

Estas medidas aplicadas a las características de las unidades en una muestra, se les denomina estimadores o *estadígrafos*. En cambio, aplicadas a las características de los elementos de una población, se les conoce como *parámetros* o *valores estadísticos* de la población.

Si con el resultado obtenido, en una encuesta aplicada en una zona o barrio de la ciudad, afirmamos que el consumo promedio de leche por familia es de dos litros por semana, estamos representando una gama o variedad de consumos, que van desde familias que no consumen, hasta un consumo superior a dos litros. Con esta información hacemos referencia al comportamiento del consumo de leche en una zona de la ciudad; también, el resultado puede ser comparado con los consumos promedios de otros barrios, o el consumo promedio por persona, o establecer la relación que puede haber entre el consumo y los niveles de ingresos.

Dependiendo de la naturaleza de los datos, la misma necesidad que se tiene y/o la característica que ellos representan, requieren de la aplicación de un promedio especial dentro de los diferentes tipos que se expondrán a continuación, pero para ello deberán tenerse en cuenta algunas reglas que serán básicas en la selección de estas herramientas de descripción.

CARACTERÍSTICAS, USO, VENTAJAS Y DESVENTAJAS DE LAS MEDIDAS DE TENDENCIA CENTRAL Y DE POSICIÓN.

Un *promedio* es un valor en la escala de las X correspondiente a una distribución de frecuencias, cuyo valor se calcula para representarla en calidad de medida típica del grupo.

Algunas condiciones esenciales para que este valor pueda cumplir su cometido:

- Debe estar representada por una sola cifra.
- Debe ser objetiva y definida por una fórmula algebraica, de tal manera que cualquiera que sea la persona que trabaje con la distribución, se obtenga siempre el mismo valor.
- Debe ser descriptiva de los datos, de manera que su significado sea fácil de entender.
- No debe ser una abstracción matemática, que únicamente pueda ser entendida por personas muy versadas en esta ciencia, ya que una de las condiciones de la estadística, es simplificar los datos y no hacerlos más complejos.
- Debe ser fácil de calcular, pero esta condición no debe preferirse en perjuicio de otras ventajas.
- Debe depender de cada uno de los elementos del grupo, de tal manera que si se altera alguno de éstos, consecuentemente se altera el valor del promedio ya que

por definición, ésta es representante típica de todos los miembros del grupo y no solamente del alguno de ellos.

- A pesar de que cada uno de los valores de los elementos del grupo debe tener influencia en el valor de la medida, ésta debe ser de tal calidad que no se deje afectar demasiado por alguno o algunos pocos valores.
- Debe tener lo que los estadísticos llaman estabilidad en el muestreo.
- Debe ser de uso fácil para cálculos matemáticos posteriores.

El *promedio* es un término muy utilizado por todos, aunque indeterminado. Se dice generalmente que un valor promedio, intenta representar o resumir las características de un conjunto de valores. La expresión corriente de promedio, suele en la mayoría de los casos referirse a la *media aritmética.*

MEDIA ARITMÉTICA: $\overline{X} = \mu$

Es la *medida de posición* o *promedio* más conocida, la más utilizada y entendida por todos, por su gran estabilidad es la preferida en el muestreo, sus fórmulas admiten tratamiento algebraico. Su desventaja principal, es ser muy sensible a cambios en sus valores u observaciones, también, cuando alguno de sus valores extremos es demasiado grande o pequeño

Ventajas

- El *promedio aritmético* es, en sí, la medida más entendida y la más utilizada. Es un valor tal, que si se calculara para un grupo en el cual todos los elementos fueran iguales, cada uno de ellos sería igual a su promedio aritmético, por lo cual podemos deducir que el promedio de *n* elementos, es un nuevo elemento formado, correspondiernte a una parte igual a 1/*n* de cada uno de los elementos originales.
- Esta medida se define mediante una ecuación matemática muy fácil de entender y algunas veces se puede obtener cuando no es posible calcular otros tipos de medidas de *tendencia central* o de *posición* y, aún en el casos de no conocerse los valores individuales de la serie, por ejemplo: si 10 familias consumen 20 litros de leche, el promedio aritmético será de 2 litros por familia.
- El promedio aritmético es extraordinariamente estable en el muestreo.
- Es altamente sensible a cualquier cambio en los datos de la distribución.
- Como ventaja final, se podrá anotar que es excepcionalmente adaptable cuando se trata de hacer cálculos matemáticos posteriores con él.

Desventajas

- Es muy sensible a valores muy pequeños o grandes, especialmente cuando estos últimos se encuentren incluidos en la distribución que se está estudiando, puede resultar en un promedio que no represente lo típico para el total del grupo.
- Cuando la distribución es marcadamente asimétrica, de tal forma que el *promedio aritmético*, la *mediana* y el *modo* difieran en forma apreciable, debe considerarse siempre la posibilidad de que el promedio aritmético pueda no ser el valor único representativo de la serie.

- Otro inconveniente o desventaja, es cuando la distribución tiene la forma de U, es decir, parabólica; éste corresponde a los valores menos comunes en la serie y por tanto, puede darnos una idea irreal de la distribución.

Esta medida, por lo general, se representa mediante una equis con una raya o trazo colocado en su parte superior, sin embargo se utilizan otros símbolos como se pueden observar a continuación.

$$a \quad a_y \quad a_x \quad M_{[x]} \quad M_{[y]} \quad \bar{x} \quad \bar{y} \quad M_{[x]} \quad \mu$$

Se define como la "*suma de todos los valores observados, divididos por el número total de observaciones*" De esta forma definida, sólo se aplica en datos sin agrupar, tambien denominados como datos originales.

Media aritmética simple

Se trabaja con datos sin agrupar u originales, siendo:

$$\bar{x} = \frac{\Sigma x_i}{n} \quad \text{Donde} \quad \bar{x} = \frac{x_1 + x_2 + x_3 \ldots\ldots\ldots x_n}{n} \quad \text{(Minúsculas en las muestras)}$$

$$\mu = \bar{X} = \frac{\Sigma X_i}{N} \; ; \; \text{Siendo} \quad \mu = \frac{X_1 + X_2 + X_3 \ldots\ldots X_N}{N} \quad \text{(Mayúsculas y letras griegas en la población)}$$

Ejemplo 1. Aplicando la fórmula dada para calcular la media aritmética, en datos no agrupados, de acuerdo a los datos del Ejemplo 1, de la Unidad 2, correspondiente a una muestra realizada en 30 cajas, para determinar el número de unidades con piezas defectuosas, se tendrá que:

$$\bar{x} = \frac{\Sigma x_i}{n} \qquad \bar{x} = \frac{3 + 2 + 0 + \ldots\ldots + 2}{30} = \frac{67}{30} = 2{,}23 \quad \text{Aproximadamente 2 unidades defectuosas por caja}$$

Ejemplo 2. Apliquemos la anterior fórmula para datos no agrupados con los datos dados, en el caso de una variable continua (Ejemplo 2 de la Unidad 2), correspondiente a una muestra de 30 alumnos, con la cual se quiere investigar su peso promedio. El resultado es.

$$\bar{x} = \frac{\Sigma x_i}{n} \qquad \bar{x} = \frac{74 + 67 + 94 + \ldots.. + 92}{30} = \frac{2.067}{30} = 68{,}9$$

Media aritmética ponderada.

Se aplica en datos agrupados, es decir, aquellos que se encuentran organizados en una *tabla de frecuencias,* siendo las frecuencias absolutas sus ponderaciones..

Ejemplo 3. Supongamos que dispone de información para 10 0bservaciones: 8, 2, 8, 6, 2, 2, 6, 8, 2, 4. La media aritmética será:

$$\bar{x} = \frac{8 + 2 + 8 + \ldots\ldots + 4}{10} = \frac{48}{10} = 4{,}8$$

Ahora si calculamos la media aritmética de las mismas 10 observaciones, pero ordenadas, el resultado obtenido será el mismo.

$$\bar{x} = \frac{2+2+2+2+4+6+6+8+8+8}{10} = \frac{48}{10} = 4{,}8$$

Resultado que sería exactamente igual a:

$$\bar{x} = \frac{2(4)+4(1)+6(2)+8(3)}{10} = \frac{8+4+12+24}{10} = 4{,}8$$

¿Qué se ha hecho? Simplemente se han organizado los datos en una tabla de frecuencias

y_i	n_i	$y_i\, n_i$
2	4	8
4	1	4
6	2	12
8	3	24
Σ	10	48
X_i	f_i	$X_i\, f_i$

$$\boxed{\bar{y} = \frac{\Sigma y_i n_i}{n}} \quad \text{o} \quad \boxed{\bar{X} = \frac{\Sigma X_i f_i}{N}}$$

$$\bar{y} = \frac{y_1 n_1 + y_2 n_2 + \ldots\ldots}{n_1 + n_2 + \ldots\ldots\ldots} = \frac{2(4)+4(1)\ \ldots\ldots}{10} = \frac{48}{10} = 4{,}8$$

Ejemplo 4. En el ejemplo de las 30 cajas, seleccionadas con el fin de establecer el promedio de unidades defectuosas por caja, se podrá obtener también aplicando la fórmula de cálculo para datos agrupados, como se observa a continuación:

y_i	n_i	$y_i\, n_i$
0	4	0
1	5	5
2	7	14
3	8	24
4	6	24
Σ	30	67
X_i	f_i	$X_i\, f_i$

Con la información de la tabla 2.1- Unidad 2
La *Media aritmética* trabajando con datos agrupados será:

$$\boxed{\bar{y} = \frac{\Sigma y_i n_i}{n}} \quad \text{o} \quad \boxed{\bar{X} = \frac{\Sigma X_i f_i}{N}}$$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{67}{30} = 2{,}23 \ \text{Unidades defectuosas}$$

Ejemplo 5. Algo similar lo podemos hacer, para el cálculo de la media aritmética en la variable continua. Se obtiene en primer lugar las marcas de clase y luego se trabaja exactamente, como se hizo en la variable discreta.

$y'_{i-1} - y'_i$	n_i	y_i	$y_i n_i$
46,1 − 54	3	50	150
54,1 − 62	6	58	348
62,1 − 70	8	66	528
70,1 − 78	6	74	444
78,1 − 86	4	82	328
86,1 − 94	3	90	270
Σ	30	--	2.068
$X'_{i-1} - x'_i$	f_i	X_i	$X_i\, f_i$

$$\boxed{\bar{y} = \frac{\Sigma y_i n_i}{n}} \quad \text{o} \quad \boxed{\bar{X} = \frac{\Sigma X_i f_i}{N}}$$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{2.068}{30} \equiv 68{,}93 \ \text{Kilos}$$

Cálculo de la Media, utilizando Frecuencias Relativas

Esta medida también se puede calcular, utilizando como ponderaciones a las frecuencias relativas, con las cuales se obtendrá el mismo resultado:

$\boxed{\bar{y} = \frac{\Sigma y_i n_i}{n}}$ $\bar{y} = \frac{y_1 n_1 + y_2 n_2 + ... + y_m n_m}{n_1 + n_2 + ...}$ Se sabe que $h_i = \frac{n_i}{n}$ por lo tanto se tiene que

$\bar{y} = y_i h_1 + y_2 h_2 + ...y_m h_m \Rightarrow$ $\boxed{\bar{y} = \Sigma y_i h_i}$ ó $\boxed{\bar{x} = \Sigma x_i (f_i / n)}$

Ejemplo 6. Considerando las tablas de frecuencias (2.1 y 2.4) calculamos las medias aritméticas, en cada una de ellas, trabajando con las frecuencias relativas, cuyos resultados deberían ser iguales, pero en algunos casos se observan algunas diferencias, debido a la aproximación a dos dígitos, que se hizo para las frecuencias relativas.

(2.1)

y_i	h_i	$y_i h_i$
0	0,13	0
1	0,17	0,17
2	0,23	0,46
3	0,27	0,81
4	0,20	0,80
Σ	1,00	2,24
X_i	f_i	$X_i\left(f_i/n\right)$

$\bar{y} = \Sigma y_i h_i = 2{,}24$

$\boxed{\bar{x} = \Sigma x_i (f_i / n)}$

(2.2)

$\bar{y} = \Sigma y_i h_i = 68{,}88$

y_i	h_i	$y_i h_i$
50	0,10	5,00
58	0,20	11,60
66	0,27	17,82
74	0,20	14,80
82	0,13	10,66
90	0,10	9,00
Σ	1,00	68,88
X_i	f_i	$X_i\left(f_i/n\right)$

DESVIACIONES

Se considera necesario mencionar las *Desviaciones con respecto a la media aritmética*, las cuales se definen como las diferencias que hay entre los distintos valores que toman la variable y la media aritmética, tanto en *datos sin agrupar* como en *datos agrupados*, por lo general simbolizada por Z_i o d_i

$\boxed{Z_i = x_i - \bar{x}}$ (En datos sin agrupar) $\boxed{Z_i = y_i - \bar{y}}$ (Para datos agrupados

Ejemplo 7. Procedamos a considerar una nueva fuente de información con 8 datos, sin agrupar: 6, 10, 4, 10, 8, 2, 6, 2. En primer lugar calculamos la media y luego sus desviaciones.

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{48}{8} = 6: \quad \begin{aligned} z_1 &= x_1 - \bar{x} = 6 - 6 = 0 \Leftrightarrow \boxed{d_1} \\ z_2 &= x_2 - \bar{x} = 10 - 6 = 4 \Leftrightarrow \boxed{d_2} \\ z_3 &= x_3 - \bar{x} = 4 - 6 = -2 \Leftrightarrow \boxed{d_3} \end{aligned}$$

Ejemplo 8. Si calculamos las desviaciones con respecto a la media, en una tabla con datos correspondiente a una variable discreta (2.1) en la cual la media fue de 2,23, se tendrá:

y_i	n_i	$y_i - \bar{y}$
0	4	–2,23
1	5	–1,23
2	7	–0,23
3	8	0,77
4	6	1,77
Σ	30	--
X_i	f_i	$X_i - \bar{X}$

$$\begin{aligned} z_1 &= y_1 - \bar{y} = 0 - 2{,}23 = -2{,}23 \\ z_2 &= y_2 - \bar{y} = 1 - 2{,}23 = -1{,}23 \\ z_3 &= y_3 - \bar{y} = 2 - 2{,}23 = -0{,}23 \\ z_4 &= y_4 - \bar{y} = 3 - 2{,}23 = 0{,}77 \\ z_5 &= y_5 - \bar{y} = 4 - 2{,}23 = 1{,}77 \end{aligned}$$

$\boxed{d_i = x_i - \bar{x}}$

Ejemplo 9. Para el cálculo de estas desviaciones en la variable continua, se debe trabajar con las marcas de clase. La media aritmética de los datos de la tabla 2.2 fue de 68,93.

$y'_{i-1} - y'_i$	y	$y_i - \bar{y}$
46,1 – 54	50	–18,93
54,1 – 62	58	–10,93
62,1 – 70	66	– 2,93
70,1 – 78	74	5,07
78,1 – 86	82	13,07
86,1 – 94	90	21,07
Σ	--	--
$X'_{i-1} - X'_i$	X_i	$X_i - \bar{X}$

$$z_1 = y_1 - \bar{y} = 50 - 68{,}93 = -18{,}93$$
$$z_2 = y_2 - \bar{y} = 58 - 68{,}93 = -10{,}93$$
$$z_3 = y_3 - \bar{y} = 66 - 68{,}93 = -\ 2{,}93$$
$$z_4 = y_4 - \bar{y} = 74 - 68{,}93 = +\ 5{,}07$$
$$z_5 = y_5 - \bar{y} = 82 - 68{,}93 = +13{,}07$$
$$z_6 = y_6 - \bar{y} = 90 - 68{,}93 = +21{,}07$$

PROPIEDADES DE LA MEDIA ARITMÉTICA

Se hace necesario conocer y manejar correctamente las propiedades que presenta la Media aritmética, pues facilita la realización de ciertas operaciones, algunas de ellas necesarias para el desarrollo de la teoría estadística y otras, como simplificación de cálculos.

a) La suma de las desviaciones respecto a la media, siempre serán iguales a cero.

$\boxed{\Sigma Z_i = \Sigma(X_i - \bar{X}) = 0}$ ó $\boxed{\Sigma d_i = \Sigma(X_i - \bar{X}) = 0}$ En datos sin agrupar

$\boxed{\Sigma Z_i n_i = \Sigma(y_i - \bar{y}) n_i}$ ó $\boxed{\Sigma d_i f_i = \Sigma(X_i - \bar{X}) f_i = 0}$ En datos agrupados

Aplicación

y_i	n_i	$y_i - \bar{y}$	$(y_i - \bar{y}) n_i$
0	4	-2,23	-8,92
1	5	-1,23	-6,15
2	7	-0,23	-1,61
3	8	0,77	6,16
4	6	1,77	10,62
Σ	30	--	0
X_i	f_i	$X_i - \bar{X}$	$(X_i - \bar{X}) f_i$

$\boxed{\Sigma Z_i n_i = \Sigma(y_i - \bar{y}) n_i = 0}$ $\quad \bar{y} = \dfrac{\Sigma y_i n_i}{n} = \dfrac{67}{30} \equiv 2{,}23$

Nota: Si sumamos los valores de la última columna de la tabla, debería haber dado cero, pero resulta que nos da 0,10, en vez de 0, ello se debe que el valor de la media es 2,23333 que usted puede comprobarlo trabajando con todos estos decimales.

b) La media aritmética de una variable por una constante, es igual a la constante por la media aritmética de la variable.

Aplicación

y_i	n_i	$y_i n_i$
0	4	0
1	5	5
2	7	14
3	8	24
4	6	24
Σ	30	67
X_i	f_i	$X_i f_i$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{67}{30} \equiv 2{,}23$$

K = 4

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{268}{30} \equiv 8{,}93$$

$$\boxed{\bar{y} = k\bar{y} = 4(2{,}23) = 8{,}92}$$

Por las aproximaciones en la media inicial.

y_i	n_i	$y_i n_i$
0	4	0
4	5	20
8	7	56
12	8	96
16	6	96
Σ	30	268
X_i	f_i	$X_i f_i$

La variable de la primera tabla, está dada en Kilos, y en la segunda tabla convertida en libras y se desea calcular la media aritmética en libras.

$$\overline{X} = \frac{\Sigma X_i f_i}{N}$$

c) *La media aritmética de una constante es igual a la constante* $\boxed{\Longrightarrow M_{(K)} = K}$. Supongamos que un estudiante durante el semestre le hicieron 6 evaluaciones y en cada una de ellas la calificación obtenida fue 4, su promedio será exactamente cuatro (4)

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{4+4+4+4+4+4}{6} = \frac{24}{6} = 4 \qquad \bar{x} = \frac{\Sigma k}{n} = \frac{nk}{n} = k$$

(Propiedad vista en la sumatoria)

d) *La media aritmética de una variable más una constante, es igual a la media aritmética de la variable.* Esta propiedad de la suma, es válida para la diferencia

Aplicación

y_i	n_i	$y_i n_i$
0	4	0
1	5	5
2	7	14
3	8	24
4	6	24
Σ	30	67
X_i	f_i	$X_i f_i$

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{67}{30} \equiv 2{,}23$$

K = 4

$$\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{187}{30} \equiv 6{,}23$$

$$\bar{y} = \bar{y} + k = 2{,}23 + 4 = 6{,}23$$

Por las aproximaciones en la media inicial.

$$M_{[x+k]} = x + k \qquad M_{[x-k]} = x - k$$

y_i	n_i	$y_i n_i$
4	4	16
5	5	25
6	7	42
7	8	56
8	6	48
Σ	30	187
X_i	f_i	$X_i f_i$

e) *La media aritmética de la suma de dos variables, es igual a la suma de las dos medias correspondiente a las dos variables.* Algo similar sucede con la diferencia o cuando se toma más de dos de dos variables.

$$M_{[x+y]} = M_{[x]} + M_{[y]} = \bar{x} + \bar{y} \qquad M_{[x-y]} = M_{[x]} - M_{[y]} = \bar{x} - \bar{y}$$

f) *La media aritmética de dos muestras, es igual, a la media ponderada de las submuestras, siendo sus ponderaciones los tamaños de esas submuestras.*

Aplicación

Muestra

y_i	n_i	$y_i n_i$
50	3	150
58	6	348
66	8	528
74	6	444
82	4	328
90	3	270
Σ	30	2.068
X_i	f_i	$X_i f_i$

Submuestra 1

y_i	n_i	$y_i n_i$
50	3	150
58	6	348
66	8	528
Σ	17	1.026
X_i	f_i	$X_i f_i$

Submuestra 2

y_i	n_i	$y_i n_i$
74	6	444
82	4	328
90	3	270
Σ	13	1.942
X_i	f_i	$X_i f_i$

Muestra	**Submuestra 1**	**Submuestra 2**
$n = 30;\quad \bar{y} = \dfrac{2.068}{30} = 68{,}93;$	$n_1 = 17;\quad \bar{x}_1 = \dfrac{1.026}{17} = 60{,}35;$	$n_2 = 13;\quad \bar{x}_1 = \dfrac{1.042}{13} = 80{,}15$

$$\boxed{\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \ldots\ldots}{n_1 + n_2 + \ldots\ldots}} \qquad \bar{x} = \frac{60{,}35(17) + 80{,}15(13)}{17+13} = \frac{1.026 + 1.042}{30} = \frac{2.068}{30} = 68{,}93$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en **Sistema de Información en Línea SIL.**

1. Con los siguientes datos referentes a una distribución simétrica, se pide calcular la media aritmética aplicando uno de los métodos que usted conoce.

X_i	y_i:	10	20	30	40	5	
f_i	n_i:	6	10	18	10	6	= 50

2. Con los siguientes datos

153	123	129	132	147	138	137	134	131	147	158	128	134	148	125	139	146
145	148	135	152	128	146	143	138	138	122	146	137	151	145	124	132	138
144	141	137	146	138	146	152	136	160	159	157	150	160	142	148	130	

Se pide:

a) Calcular la media aritmética para datos no agrupados u originales.

b) Elaborar una tabla de frecuencias.

c) Calcular la media aritmética para datos agrupados, primero utilizando las frecuencias absolutas y luego las relativas.

d) Comprobar que la suma de las desviaciones con respecto a la media es cero.

3. Con la tabla anterior, se pide:

a) Dividir la tabla en dos, cada una con tres intervalos y calcular la media para cada una. Luego con esas medias aplicar una de sus propiedades, para calcular una sola media.

b) Utilizando la tabla completa, multiplique por dos a la variable, calculando la nueva Media; luego aplique la propiedad para abreviar el procedimiento. ¿Cuál?

4. Con la siguiente distribución de frecuencias, calcular la media ponderada.

$y'_{i-1} - y'_i$	6,1-12	12,1-18	18,1-24	24,1-30	30,1-36	36,1-42	$X'_{i-1} - X'_i$
N_i	8	12	40	70	90	100	F_i

5. Si la media de 75 artículos es 52,6 galones y, la de 25 artículos es de 48,4 galones, encuentre la media de los 100 artículos.

6. De 500 estudiantes de secundaria, cuya estatura media es de 1,57 metros, 150 de ellos son mujeres. Si la estatura media de las mujeres es de 1,52 metros ¿Cuál es la estatura media de los varones?

7. Un grupo de 200 estudiantes, cuya estatura media es de 160,96 centímetros, se dividen en dos grupos, uno con una estura media de 163,4 centímetros y otro con 157,3 centímetros. ¿Cuántos estudiantes hay en cada grupo?

8. Se realiza una prueba a dos secciones de un curso de matemáticas, con un promedio general de 60,98. La sección I tiene una media de 57,30 y la sección II de 65,30. Si hay 27 estudiantes en la sección I.¿Cuántos hay en la sección II?

9. En una competencia atlética con 45 participantes, 20 compiten el primer día y los restantes al día siguiente. Si en el primer día, los 20 obtienen un promedio de 48,4 puntos. ¿Cuál es el menor número de puntos que deben obtener en promedio los restantes atletas, de modo que el puntaje medio de todo el equipo sea por lo menos igual a 55,0?

10. Un grupo de 100 atletas viaja en dos aviones. El primero lleva 40 atletas y el segundo los restantes. Se sabe que el peso medio de todos los atletas es de 186,3 libras y que el de los del segundo avión es tres libras menos que el de los atletas del primer avión ¿Cuál es el peso medio de los atletas en cada avión?

11. En un examen tomado a tres secciones en un curso de estadística de 91 alumnos, el puntaje medio general fue de 69,3. Los puntajes medios de las secciones 1 y 2 fueron 70,4 y 64,2 respectivamente. Se perdieron los archivos con las notas de la sección 3, pero los monitores recuerdan que las secciones 1 y 2 tenían exactamente el mismo número de alumnos, mientras que el monitor de la sección 3, afirma que su sección tenía 5 estudiantes menos que la 1.¿Cuál fue el promedio de la sección 3?

12. La media aritmética de los salarios pagados en un mes a los empleados de una empresa, ascendió a $920.000. La media aritmética de los salarios pagados a los hombres y a las mujeres, fueron respectivamente $970.000 y $840.000. Determinar el porcentaje de hombres y mujeres empleados en dicha empresa?.

13. El salario medio mensual por obrero en la empresa A fue de $938.000 durante el 2010. Para el presente año la empresa aumentó a cada uno de sus empleados la suma $78.000. Admitiendo que los salarios no se han modificado desde entonces. ¿Cuál es el promedio mensual?

14. En cierto examen el puntaje medio de los estudiantes del curso A es 68,4 y los del B 71,2. Si el puntaje medio del grupo es 70,0 ¿Cuál es la relación, entre el número de los alumnos, en los cursos A y B?

15. En un curso hay 35 hombres con una media de 17,5 años y 15 mujeres, las que en promedio son 22% más jóvenes. ¿Cuál es la edad promedio del curso?.

16. El precio promedio de un centenar de artículos es $18.750, los artículos se dividen en dos grupos, con medias $17.580 y $19.780 ¿Cuántos artículos hay en cada grupo?

17. Se sabe que ninguna de las sucursales de una empresa comercial, tiene más de 19 empleados o menos de 17. La mayoría tiene 18 empleados, pero el 25% tiene 19 empleados y cada una de las 10 sucursales tiene 17 empleados. ¿Cuál es el promedio de empleados por sucursal?

18. De 10 familias investigadas con teléfono, auto y antena parabólica, se obtuvo la siguiente información, en miles de pesos, sobre los gastos mensuales en:
a) Costo del teléfono por familia $56 64 38 60 42 28 55 70 42 y 63 (Miles pesos).
b) Costo promedio mensual en mantenimiento $83
c) La distribución de costos de servicio mensual de la parabólica, es

COSTOS (Miles de pesos)	80,1-120	120,1-160	160,1-200	200,1-240	240,1-280
Nº DE FAMILIAS	1	3	2	3	1

Se pide calcular el costo total promedio mensual por familia, de los 3 servicios.

19. Los siguientes datos corresponden a los salarios mensuales, en miles de pesos, pagados por una empresa a su personal.

SALARIOS (Miles pesos)	500	600	700	800	900
Nº EMPLEADOS	10	16	35	26	13

En la empresa se presenta un conflicto laboral. El gerente propone un aumento del 7% mensual, para cada uno de los empleados. La Junta directiva propone un aumento de $49.000, para cada empleado. Se solicita su asistencia técnica, para que se discuta que es más ventajoso para la empresa y qué para cada grupo de empleados, según su clasificación por nivel de salarios. ¿Cómo actuaría usted?

20.

a)	$H_5 = 0{,}36$	$N_4 = 30$	$n_5 = 6$	$n = 5$
b)	$y_1 = 17$	$m = 5$	$c = 6$	$\bar{y} = 125$
c)	$y_i n_i = 240$	$\bar{y} = 8$	$n = 40$	
d)	$H_5 = 1{,}10$	$H_4 = 0{,}80$	$h_3 = 0{,}30$	

21. Señale y justifique la respuesta correcta a las siguientes afirmaciones: a) El promedio de las notas de un curso fue de 5,8. Las 25 mujeres obtuvieron un promedio de 3,8 y los hombres de 6,4. Luego el curso tiene más de 90 alumnos. b) Si $H_6 = 0{,}7$ y $H_3 = 0{,}3$ un 40% de las observaciones de la variable es menor que y_6 y mayor que y_4 c) El ingreso mensual medio de los obreros agrícolas es $635.000, el de los mineros es $754.000 y el de los industriales es $864.000, es decir, el ingreso medio del conjunto de obreros es de $754.000. d) Para calcular la media geométrica, es preciso que todos los intervalos sean iguales. e) El 75% de los empleados públicos son hombres, mientras que en el sector privado es el 81%. Esto es, hay un 22% de mujeres empleadas en ambos sectores. f) Si se multiplica por 2 todas las frecuencias relativas, el valor de la media aritmética también se duplica.

22. Contestar si es *cierto o falso* a los siguientes puntos: a) La suma de las desviaciones es igual a cero, cuando se toma con respecto a la media. b) Un grupo de valores puede tener más de una media aritmética. c) Para calcular las marcas de clase, se suma el límite inferior al superior del intervalo y se divide por dos.

23. Un laboratorio de control de calidad de una empresa, quiere lanzar al mercado un nuevo spray ambientador, para determinar el tiempo de permanencia del aroma; los resultados fueron observados cada hora, habiéndose obtenido:10 horas, 5 apartamentos; 3 horas, 3 apartamentos; 5 horas, 10 apartamentos; 4 horas, 7 apartamentos; 6 horas, 16 apartamentos; y al final, 8 horas, 9 apartamentos. a) Cuál es la población? b) Cuál es la muestra? c) Cuál es la variable d) Qué tipo de variable es? e) Cuál es el promedio de duración del aroma? f) Elabore una gráfica que la represente.

24 En una empresa constructora de vivienda, los jornales mensuales de los obreros tienen una media de $655.000. Como solución a un conflicto laboral se proponen dos alternativas: a) un aumento del 8%; b) un aumento del 5%, más una bonificación mensual de $25.000 a cada obrero. ¿Cuál de las dos alternativas le conviene aceptar al sindicato, si se quiere que la distribución de los salarios sea lo mejor posible?

25. El promedio de salario mensual para un grupo de obreros es de $662.000. El gerente de la firma resuelve reconocerles un aumento diario por cada obrero de $3.000; si consideramos los meses de 30 días. ¿Cuál será el nuevo promedio de salario mensual?

26. Un almacén de artículos electrónicos, se dedica a la venta de cuatro líneas de producto. Se sabe que el año anterior el total de ventas (Millones de $) y margen de utilidad por línea fueron los siguientes:

LÍNEA DE PRODUCTOS	VENTAS	MARGEN DE UTILIDAD
1 - Televisores…………	214.000	12,6 %
2 - Neveras…………….	90.000	5,8%
3 - Equipos de sonido….	183.000	9,3%
4 - Lavadores…………..	75.000	4,6%
TOTAL……...…….	562.000	-------

Se pide calcular: a) La media aritmética simple del margen de utilidad. b) La media ponderada del margen de utilidad. c) ¿Cuál de las dos medias es más representativa? ¿Por qué?

27. Una firma de producción múltiple, tiene cuatro líneas de productos. Durante el mes de octubre del presente año, los resultados obtenidos al realizar las operaciones de rutina de calidad, fueron las siguientes:

LÍNEAS DE PRODUCCIÓN	UNIDADES PRODUCIDAS	PORCENTAJES DEFECTUOSOS
A	8.300	0,8
B	12.600	1,1
C	24.300	2,6
D	15.800	1,4
TOTAL	61.000	-------

Determinar el promedio porcentual de artículos defectuosos producidos por dicha firma en el mes de octubre del presente año.

28. Dos fábricas tienen 80 y 120 empleados, respectivamente. El salario promedio mensual para el total de empleados es de $620.000. Sabiendo, además, que los empleados de la primera de ellas, ganan en promedio $58.500 más que los empleados de la segunda.¿Cuál es el salario promedio mensual de los empleados en cada una de las fábricas?

29. Un grupo de 400 empleados se divide en operarios y técnicos, con un salario promedio de $980.600. Los salarios promedios para los dos grupos son $725.730 y $1.076.500. ¿Cuántos operarios y cuántos técnicos hay?

30. Tres almacenes tienen un total de 80 vendedores. Los dos primeros pagan mensualmente $925.000 y $870.000 respectivamente y tienen además, el mismo número de vendedores. Se quiere saber, cuál es el salario promedio de los vendedores del tercer almacén, si se sabe, que el salario promedio de los vendedores de los tres almacenes es $890.000 y además, tiene 10 vendedores menos que el primer almacén.

31. Un grupo de empleados tienen los siguientes salarios mensuales (Miles de $).

$ 730	510	640	680	480	640	710	600	780

Los empleados piden un reajuste del 15%, pero el empresario ofrece un aumento del 10%, más una bonificación mensual de $20.000. ¿Cuál cree usted que debe ser la decisión más ventajosa para el empleado?

32. Conteste cada una de las siguientes preguntas, investigando.
a) ¿Qué significa el término “promedio”?

b) ¿Qué ventajas y desventajas tiene la aplicación de la media aritmética?
c) ¿Se pondera la media aritmética? ¿Por qué o por qué no?

33. Un inversionista compró acciones de varias empresas en la Bolsa de Bogotá, así: 500 acciones a \$2.500 cada una; 1.200 a \$5.800 cada una; 600 a \$10.000 c/u y 2.500 a \$800 c/u. a) ¿Cuál es la inversión total? b) ¿Cuál es el precio promedio por acción?

34. A un profesor de estadística le asignaron 5 cursos con igual número de alumnos, al finalizar el semestre, los promedios de calificación de cada curso fueron: 3,6; 4,1; 3,2; 3,8; y 4,6. a) ¿Cuál es la calificación promedio con ese profesor? b) Ahora suponga, que el número de alumnos por curso fue; 32; 40; 26; 34; y 15 ¿Cuál es el promedio de calificación por curso?

MEDIANA M_E

Es considerada también, al igual que la *Media*, como una *medida de tendencia central*. Su importancia es menor, sus fórmulas son rígidas ya que no admiten tratamiento algebraico, es por eso que tendremos que aplicar seis procedimientos de cálculo diferente, que lo hace engorroso, como lo veremos a continuación:

Se define como "*aquel valor de la variable que supera a no más de la mitad de las observaciones, al mismo tiempo, es superado por no más de la mitad de las observaciones"* en otras palabras, se puede definir como el "valor central". Se simboliza por M_e

Ventajas

- Esta medida tiene una definición bastante rígida y el concepto de ella es tan claro, que cualquier persona lo entiende, aún sin ser familiar su término.
- Los datos deben estar ordenados, pero los valores extremos no tienen ninguna incidencia en su cálculo.
- Tiene menor estabilidad en el *muestreo* que el promedio aritmético, pero mayor estabilidad que otras medidas.
- Hay situaciones en que la única medida de *tendencia central* que puede utilizarse es la *mediana*, especialmente cuando los valores extremos de una distribución de variable continua no están definidos

Desventajas

- No es tan conocida como la *media aritmética*.
- Se hace necesario ordenar los datos para poderla calcular.
- Esta medida no se adapta a cálculos posteriores aritméticos, por cuanto que si obtenemos las medianas de diferentes grupos, no se podrá calcular una mediana para el conjunto.
- Tampoco es sensible a cambios en los valores de los elementos, manteniéndose el valor central y los valores ordenados.

Cálculo en datos sin agrupar u originales

a) *Número impar de observaciones.* Cuando ésta medida la aplicamos en los datos originales o sin agrupar, lo primero que se debe hacer, es ordenarlos de menor a mayor o de mayor a menor.

Ejemplo 1. Supongamos se tienen los siguientes datos: y $x_1=2$ $x_2=18$ $x_3=4$ $x_4=12$ Observemos que la serie es *impar*, ya que $n=5$ por lo tanto ordenamos los datos, por ejemplo, de menor a mayor:

2	4	6	12	18

La *mediana* será igual al valor central $M_e = 6$

b) *Número par de observaciones.* Ahora calcularemos la mediana, cuando se tenga un número par de observaciones. En estos casos, encontramos dos valores en el centro de la serie, por tal razón la mediana deberá ser el promedio de ellos.

Ejemplo 2. Supongamos que en vez de 5 datos se tengan 8 datos así:

$x_1 = 8 \quad x_2 = 16 \quad x_3 = 4 \quad x_4 = 2 \quad x_5 = 20 \quad x_6 = 3 \quad x_7 = 12 \quad x_8 = 20$

Ordenamos los datos de menor a mayor o viceversa;

2	3	4	8	12	16	20	20

$$\Rightarrow M_e = \frac{x_{j-1} + x_j}{2} = \frac{8+12}{2} = 10$$

En los dos casos anteriores, correspondiente al cálculo de la mediana para datos sin agrupar, la localización de ella, luego de ser ordenados, se facilita aplicando el siguiente procedimiento:

$\frac{n+1}{2} = \frac{5+1}{2} = 3^{a}$ *posición*, la *mediana* en el primer caso o ejemplo 1, será $M_e = x_j = 6$ y en el ejemplo 2, será:

$\frac{n+1}{2} = \frac{8+1}{2} = 4{,}5^{a}$ *posición*, es decir, se encuentra entre la 4ª y la 5ª observación y la *mediana* será igual a $M_e = x_j = 10$, el valor promedio de los dos valores centrales.

Cálculo en datos agrupados o en tablas de frecuencias

Veamos las fórmulas que se aplican cuando la variable es discreta y cuando es continua. En cada caso, tiene dos procedimientos

Variable discreta

Cuando $N_{j-1} = n/2$ La *mediana* será igual a: $M_e = \frac{y_{j-1} + y_j}{2}$

Cuando $N_{j-1} < n/2$ La *mediana* será igual a: $M_e = y_j$

Variable continua

Cuando $N_{j-1} = n/2$ La *mediana* será igual a: $M_e = y'_{j-1}$

Cuando $N_{j-1} < n/2$ La *mediana* será igual a: $M_e = y'_{j-1} + c\left[\frac{n/2 - N_{j-1}}{n_j}\right]$

Pasos a seguir en el cálculo de la Mediana Variable discreta

Ejemplo 3

<u>Caso A</u> $\frac{n}{2} = \frac{30}{2} = 15$ Se localiza en la columna N_J

Cuando $N_{j-1} = n/2$ o sea 15 = 15

Se dice que $\boxed{M_e = \frac{y_{j-1} + y_j}{2}}$

$$M_e = \frac{2+3}{2} = \frac{5}{2} = 2{,}5$$

	y_i	n_i	N_J	
	0	4	4	
	1	5	9	
y_{j-1}	$\rightarrow$ 2	6	15 $\rightarrow$	N_{j-1}
y_j	$\rightarrow$ 3	7	22 $\rightarrow$	N_j
	4	8	30	
	Σ	30	---	
	X_i	f_i	N_i	

Es decir que la *mediana* se encuentra localizada entre los valores 2 y 3, siendo igual a 2,5. También se puede comprobar el anterior resultado, transformando la distribución de frecuencias en una variable que nos muestre los datos originales o que éstos se encuentren sin agrupar.

0 0 0 0 1 1 1 1 1 2 2 2 2 2 $\boxed{2\ 3}$ 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4

La *mediana* se localiza entre los valores ubicados entre la posición 15^a y 16^a

<u>**Caso B**</u> $\frac{n}{2} = \frac{30}{2} = 15$ No se localiza en la columna

Cuando $N_{j-1} < n/2$ o sea que $9 < 15$

Se dice que $\boxed{M_e = y_j}$ $M_e = 2$

	y_i	n_i	N_j	
	0	4	4	
	1	5	9 $\rightarrow$	N_{j-1}
y_j	$\rightarrow$ 2	7	16 $\rightarrow$	N_j
	3	8	24	
	4	6	30	
	Σ	30	---	
	X_i	f_i	N_j	

La mitad de las observaciones es 15, por lo tanto la 15^a y la 16^a corresponden al valor de 2, el cual se toma como el valor de la *Mediana*, como se puede ver a continuación:

Tal como se hizo en el ejercicio anterior, en datos sin agrupar, el valor central lo ocupa el 2

0 0 0 0 1 1 1 1 1 2 2 2 2 2 $\boxed{2\ 2}$ 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4

Mediana Datos agrupados: Variable discreta

a) $M_e = \frac{y_{j-1} + y_j}{2}$ Cuando $N_{j-1} = n/2$

b) $M_e = y_j$ Cuando $N_{j-1} = n/2$

Pasos a seguir en el cálculo de la mediana en la Variable continua

Datos agrupados: Variable continua

a) $M_e = y'_{j-1}$ Cuando $N_{j-1} = n/2$

b) $M_e = y'_{j-1} + c\left[\frac{n/2 - N_{j-1}}{n_j}\right]$ Sí $N_{j-1} < n/2$

Ejemplo 4 .

Caso A $\frac{n}{2} = \frac{40}{2} = 20$

Cuando $\boxed{N_{j-1} = n/2}$ en este caso $N_{j-1} = 20$

y $N_j = 30$ así que el valor de la *mediana*

$\boxed{M_e = y'_{j-1} = 70}$

Es decir corresponde al valor central que toma la variable

$y'_{i-1} - y_i$	y_i	n_j	N_J
46,1 – 54	50	3	3
54,1 – 62	58	8	11
62,1 – 70	66	9	20 → N_{j-1}
y'_{j-1} → 70,1 – 78	74	10	30 → N_j
78,1 – 86	82	8	38
86,1 – 94	90	2	40
Σ	--	40	---
$X'_{i-1} - X'_i$	x_i	f_i	N_J

Caso B $\frac{n}{2} = \frac{40}{2} = 20$ No se localiza en la columna

En este caso $\boxed{N_{j-1} < n/2}$ por lo tanto se tendrá que $N_{j-1} = 11$ y $N_j = 22$

La *mediana* será igual a:

$$\boxed{M_e = y'_{j-1} + c\left[\frac{n/2 - N_{j-1}}{n_j}\right]}$$

$y'_{i\,1} - y_i$	y_i	n_j	N_J
46,1 – 54	50	3	3
54,1 – 62	58	8	11 → N_{j-1}
y'_{j-1} → 62,1 – 70	66	11	22 → N_j
70,1 – 78	74	7	29
78,1 – 86	82	9	38
86,1 – 94	90	2	40
Σ	--	40	
$X'_{i-1} - X'_i$	x_i	f_i	N_j

$$M_e = 62 + 8\left[\frac{20-11}{11}\right] = 62 + \left[\frac{72}{11}\right] = 62 + 6{,}55 = 68{,}55$$

Algunos autores, utilizan otras simbologías para realizar el anterior cálculo, pero el procedimiento y resultado son siempre iguales

$$M_e = L_i + i\left[\frac{n/2 - F_{j-1}}{f_i}\right]$$

Pasos a seguir en el cálculo de la *Mediana*

a) Obtener en primer lugar las frecuencias absolutas.

b) Localizar la mitad de las observaciones, por medio de la expresión $n/2$.

c) El resultado anterior lo buscamos en la columna de las frecuencias absolutas acumuladas. Si no aparece, se toma el valor inmediatamente anterior y lo simbolizamos por N_{j-1} y al inmediatamente superior o posterior por N_j. En caso contrario, cuando dicho valor aparece registrado en esa columna, será simbolizado N_{j-1} y el inmediatamente superior posterior por N_j.

d) Cuando aparece $n/2$, se dice que es igual a N_{j-1} ($N_{j-1} = n/2$); en caso contrario, que no aparezca, se dirá que $N_{j-1} < n/2$. En ambos casos, dependiendo de si la variable es discreta o continua, se aplica la fórmula respectiva, tal como aparece en el resumen anterior.

MODO – MODA O VALOR MODAL M_D

Se define como *“el valor de la variable que más se repite”* o *“aquel valor que presenta la máxima frecuencia”*. Puede suceder que una distribución tenga *dos Modas*, en este caso se dice que la distribución *es Bimodal*, en el caso que haya más de dos modas, se dice que es *plurimodal* o *multimodal*. Es la única medida de posición que puede ser utilizada en atributos, es decir, cuando la característica es cualitativa. *Modo, Moda o valor Modal*, medida originalmente concebida, como aquel valor de la variable, que presenta el mayor número de observaciones, es decir, el valor de la variable que más veces se repite. Se empleó formalmente en 1.984 por Karl Pearson, ya que no era utilizado anteriormente.

Desventajas

- El hecho que el *modo* nos indique el punto de mayor concentración, lo hace tal vez, la mejor medida de *tendencia central*. Cuando una distribución es muy *asimétrica*, claramente se ve que el *Modo* es el más representativo del grupo, y en algunos casos, si el *modo* y el promedio *aritmético* son significativamente diferentes al valor, es preferible usar el *modo*.
- En series *plurimodales*, el modo permite dividir la distribución con fines de estratificación.
- Se considera de uso riguroso en variables continuas, aunque puede extenderse a casos discontinuos.

Ventajas

- Se le considera con cierto grado de dificultad de cálculo en datos agrupados, siendo su resultado de poca confianza, especialmente cuando la amplitud no es constante, en este caso, si se hace necesario se debe seguir un determinado proceso.
- Es muy inestable en el muestreo.
- No puede ser utilizado en procesos algebraicos posteriores.
- No es sensible, especialmente cuando hay cambios en los valores de la variable, salvo que no afecten a su propio valor.

Datos sin agrupar u originales

Ejemplo 1. Consideremos los siguientes datos: 5, 10, 8, 5, 10, 18, 5, 12, 5, 12 calculemos la media, mediana y la moda. $n = 10$

$$Media \rightarrow \bar{x} = \frac{\sum x_i}{n} = \frac{90}{10} = 9$$

Mediana → Ordenamos los datos de mayor a menor o de menor a mayor

$$5 \quad 5 \quad 5 \quad 5 \quad \underbrace{\boxed{8 \;|\; 10}}_{M_e} \quad 10 \quad 12 \quad 12 \quad 18$$

Posición: $\frac{n+1}{2} = \frac{10+1}{2} = 5{,}5^a$ La mediana se encuentra entre la 5^a y la 6^a observación

Siendo: $M_e = \frac{8+10}{2} = 9$ $\quad M_e = \frac{x_{j-1} + x_j}{2}$

Moda → Corresponderá a 5, siendo el valor de la variable que más se repite $\rightarrow M_d = x_j = 5$

Ejemplo 2. Con dos casos nuevos y con diferentes valores, calculemos otra vez las tres (3) medidas que se aplicaron en el ejemplo anterior.

Caso 1 Con los siguientes datos 5 10 8 12 18 5 12 5 12

$$Media \rightarrow \bar{x} = \frac{8}{9} \cong 9,6$$

Mediana → Ordenamos los datos de mayor a menor o de menor a mayor

Posición: $\frac{n+1}{2} = \frac{9+1}{2} = 5^{a}$ 5 5 5 8 $\boxed{10}$ 12 12 12 18

$M_e = x_j = 10$

Moda → ncontramos en esta serie de 9 datos, que el 5 y el 12 se repiten veces n mero mayor de repetición por lo tanto ay dos modas y en este caso se dice que es *bimodal.* Si ay mas de dos modas se abla de *Plurimodal* o *Multimodal.*

$M_d = 5$ y $M_d = 12$

Caso 2 Con los datos 13 4 8 12 10 12 16 6 11

$$\text{Media} \rightarrow \bar{x} = \frac{80}{10} = 8$$

Mediana → Ordenamos los datos de mayor a menor o de menor a mayor

$$M_e = \frac{10+11}{2} = 10,5$$

Moda → No hay, dado que ninguno de los valores de la variable se repite más de una vez.

Datos agrupados

En datos agrupados, tanto en la variable discreta como en la continua, la *Moda* corresponderá a aquel valor que representa la mayor frecuencia. Hay que observar en la variable continua que la amplitud del intervalo sea constante.

a) *Variable discreta*

	y_i	n_i	
	0	4	
	1	5	
	2	7	
$y_j \rightarrow$	3	8	$\rightarrow n_j$
	4	6	
	Σ	30	
	X_i	f_i	

$M_d = y_j = 3$

b) *Variable continua*

$y'_{i-1} - y'_i$	y_i	n_j	
46,1 – 54	50	3	
54,1 – 62	58	8	$\rightarrow n_{j-1}$
62,1 – 70	66	11	$\rightarrow n_j$
70,1 – 78	74	7	$\rightarrow n_{j+1}$
78,1 – 86	82	9	
86,1 – 94	90	2	
Σ	--	40	
$X'_{i-1} - X'_i$	X_i	f_i	

Amplitud constante

$M_d = y_j = 66$

Como se observa en el ejercicio anterior, primero localizamos el valor de la frecuencia más alta y luego el valor al frente de la frecuencia, en la columna de la variable, corresponderá al valor del Modo. La frecuencia la simbolizaremos por n_j, la frecuencia inmediatamente anterior a ese por n_{j-1} y la frecuencia inmediatamente superior por n_{j+1}.

El requisito primordial al calcular el *Modo*, es que la frecuencia absoluta debe ser un valor mayor que n_{j-1} y además, menor a: n_{j+1}. $\boxed{n_{j-1} < n_j < n_{j+1}}$

Existen otras fórmulas, menos importantes, para calcular el *Modo*, diseñadas especialmente para ser aplicadas cuando la variable es continua y cuando la amplitud del intervalo es constante.

a) $$M_d = y'_{j-1} + c\left[\frac{n_{j+1}}{n_{j-1}+n_{j+1}}\right]$$

b) $$M_d = y'_{j-1} + c\left[\frac{n_j - n_{j-1}}{(n_j - n_{j-1}) + (n_j - n_{j+1})}\right]$$

c) $$M_d = y'_{j-1} + c\left[\frac{n_{j+1} - n_{j-1}}{2(2n_j - n_{j-1} - n_{j+1})}\right]$$

d) $$M_d = y'_{j-1} + c\left[\frac{n_{j-1} - n_{j+1}}{n_j + n_{j+1} + n_{j-1}}\right]$$

Con los datos correspondientes al último ejercicio, de una variable continua, que fue utilizado para calcular la M oda, con datos agrupados, será utilizado en la aplicación de las fórmulas anteriores.

a) $$M_d = 62 + 8\left[\frac{7}{8+7}\right] = 62 + \frac{56}{15} = 62 + 3{,}73 = 65{,}73$$

b) $$M_d = 62 + 8\left[\frac{11-8}{(11-8)+(11-7)}\right] = 62 + 8\left[\frac{3}{7}\right] = 62 + 3{,}43 = 65{,}43$$

c) $$M_d = 62 + 8\left[\frac{7-8}{2(2(11)-8-7)}\right] = 62 + 8\left[\frac{-1}{14}\right] = 62 - 0{,}57 = 61{,}43$$

d) $$M_d = 62 + 8\left[\frac{8-7}{11+7+8}\right] = 62 + 8\left[\frac{1}{26}\right] = 62 + 0{,}31 = 62{,}31$$

Hemos calculado la Moda en datos agrupados, correspondiente a una variable continua, con una amplitud constante. Ahora veamos el procedimiento a seguir cuando no lo es. Por ejemplo:

Observemos en la tabla adjunta, que el mayor valor registrado en la columna de n_i/c_i es 4,50, por lo tanto el valor correspondiente al Modo, estará al frente, en la columna de las marcas de clase (y_i). En este caso, se tendrá que:

$$M_d = y_J = 83$$

$y'_{i-1} - y'_i$	y_i	n_j	c_i	n_j/c_i
46,1 – 50	48	3	4	0,75
50,1 – 62	56	8	12	0,67
62,1 – 77	69,5	11	15	0,73
77,1 – 82	79,5	7	5	1,40
82,1 – 84	83	9	2	4,50
84,1 – 90	87	12	6	2,00
Σ	--	50	--	--
$X'_{i-1} - X'_i$	X_i	f_i	i	f_i/c_i

Relación entre la *Media, Mediana* y el *Modo*.

Son tres situaciones que se pueden presentar entre estas tres medidas de posición.

$\boxed{M_x = M_e = M_d}$ $\boxed{M_x < M_e < M_d}$ $\boxed{M_x > M_e > M_d}$

En un polígono de frecuencias, la Moda se ubica en la cima, es decir en la parte superior de la curva o línea poligonal. Como término parroquial, corresponde al llamado horas pico, siendo el tiempo de mayor congestión vehicular dentro de la ciudad o de mayor consumo de energía, durante el día.

Algunos calculan la moda mediante la fórmula: $\boxed{M_d = M_x - 3(M_x - M_d)}$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

35. Con el siguiente ejercicio, referente a los datos obtenidos sobre el número de hijos por familia, se pide calcular la media, mediana y modo. (n = 20)

y_i	0	2	4	6	7	X_i
n_i	2	3	7	4	4	f_i

36. Con el siguiente ejercicio, referente a una distribución asimétrica, se pide calcular: la media, mediana y modo. (n = 50)

$y'_{i-1} - y'_i$	2,75 - 4,25	4,25 - 5,75	5,75 - 7,25	7,25 - 8,75	$X'_{i-1} - X'_i$
n_i	4	16	25	5	f_i

37. Con la siguiente serie de números, calcule: 2.000 2.500 3.000 3.250 4.000 5.000 6.500 7.500 50.000 a) La media b) Mediana c) Moda
d) ¿Qué factor de la variable afecta la media en esta serie?
e) ¿Cuál de las medidas calculadas estima usted más representativa?

38. Los sueldos mensuales de 70 empleados de una oficina, son los siguientes:

OCUPACIÓN	Nº DESEMPLEADOS	SUELDOS
Recepcionista…………	2	642.000
Mecanógrafa…..………	12	751.000
Secretaria………………	8	758.000
Aux. Contabilidad……	10	794.000
Técnicos Eléctricos….	24	801.000
Téc. Mecánicos………	14	911.000

Se pide calcular la media, mediana y modo.

39. Utilizando la siguiente distribución sobre resistencia a la tensión de láminas de acero (en kgrs/) obtenga la: media, mediana y modo.

RESISTENCIA	10 a 20	20 a 30	30 a 40	40 a 50
% DE LÁMINAS	0,20	?	?	0,12

Nota: Se sabe que la resistencia promedio es de 29,4 kgrs/

40. El presidente de un sindicato, clasificó a los compañeros de trabajo según el salario mensual en la siguiente forma:

SALARIO MENSUAL (Miles $)	Nº DE TRABAJADORES
Menor o igual a 600	30
600,1 - 800	60
800,1 - 1.000	45
1.000,1 - 1.200	15
1.200,1 y más	50

41. Calcular el Modo utilizando las diferentes fórmulas presentadas aquí.

SALARIOS ($)	Nº EMPLEADOS
648.000,1 – 668.000	120
668.000,1 – 688.000	38
688.000,1 – 708.000	22
708.000,1 – 728.000	10
728.000,1 – 748.000	6
748.000,1 – 768.000	4

42. Con base en los siguientes datos y medidas de tendencia central, determinar si hay simetría o asimetría y cuál es su dirección.
a) $\bar{x} = 68$ $M_e = 74$ $M_d = 80$ c) $\bar{x} = 74$ $M_e = 74$ $M_d = 74$
b) $\bar{x} = 80$ $M_e = 74$ $M_d = 68$ d) $\bar{x} = 74$ $M_e = 74$ $M_d = 60$

43. Consideremos que un vendedor de cigarrillos, vende de lunes a sábado el siguiente número de cajetillas. ¿Cuál es la media, la mediana y la moda?
a) 55 60 65 60 65 65 b) 55 60 64 70 65 50 c) 40 36 48 35 100 35

44. Con los siguientes datos, correspondientes a dos submuestras, se pide calcular:

x_i	4	6	7	6	10	6	4	15	6
y_i	10	18	12	20	8	10			

a) La mediana, moda y media aritmética para cada una de las variables.
b) La media conjunta de las dos submuestras.

45. Con los siguientes datos calcular la media, mediana y el modo.

$y'_{i-1} - y'_i$	3,1-8	5,1-9	10,1-16	16,1-24	24,1-30	30,1-32	$X'_{i-1} - X'_i$
n_i	8	8	14	16	18	10	f_i

46 Una muestra realizada a 12 profesores de tiempo completo, en Universidades privadas de la capital, se encontró que su sueldo quincenal en miles de $, es de: 1.800, 720, 750, 1.680, 900, 720, 840. 810, 720, 810, 720, 840. Calcular la media, mediana y la moda de estos salarios ¿Cuál cree usted que sea el mejor promedio? ¿Por qué?

47. Si al realizar los cálculos en una distribución, se obtienen los siguientes resultados: $\bar{x} = 36$ $M_e = 32$; $M_d = 23$ y se desea corregirlos ¿Cuáles serían los nuevos promedios en los casos siguientes: a) Cada uno de los valores se incrementó en 5 unidades? b) La variable está dada en años y se desea que esté dada en meses?

48. Una corporación financiera, en los últimos 9 meses ha hecho captaciones en dinero con interés anual de: 7,9% 8,3% 6,1% 6,8% 7,4% 8,6% 8,2% 7,6% y 8% ¿Cuál es la media, mediana y moda de estos valores?

49. Se tiene el control de llegada (en minutos) de un empleado a su oficina, cuando llega temprano se tendrá tiempo a su favor (+) y sus retardos podrán ser compensados (-) Se le realizó un control durante 8 días hábiles, así: 12 -8 15 -12 -6 10 -8 y -18. Calcular la media, mediana y la moda.

50. Con los siguientes datos, calcular la media, mediana y moda.

$y'_{i-1} - y'_i$	2,1-5	5,1-9	9,1-16	16,1-20	20,1-28	28,1-36	$X'_{i-1} - X'_i$
n_i	120	15	8	6	6	4	f_i

MEDIA GEOMÉTRICA M_G M_O G

Se considera como una medida de posición o promedio, pudiendo utilizarse cualquiera de estos símbolos y el resultado obtenido deberá ser menor al de la *Media aritmética.* Se aplica en todos aquellos casos en que la variable muestra un crecimiento geométrico, como sucede con la población o que se encuentre relacionada con ella, además, como en el caso de un capital colocado a una tasa de interés compuesto, durante un determinado período de tiempo.

La *Media geométrica* se define como "*la raíz enésima del producto de los valores que toma la variable*"

$$M_0 = \sqrt[n]{\pi x_i} = \sqrt[n]{x_1 . x_2 . x_3 \ldots . x_n}$$

Observemos que el signo π se lee *productoria*, siendo este símbolo la forma de expresar que los valores correspondientes a la variable deben estar multiplicados entre sí.

Ventajas

- Se define rígidamente mediante una fórmula matemática.
- Se utiliza cuando se quiere dar importancia a pequeños valores que toma la variable.
- Es sensible a cualquier cambio que se haga en los valores de la distribución.
- Su valor no es muy influenciado por valores extremos grandes, tal como sucede con los otros promedios.
- Se hace indispensable, cuando se desea calcular el promedio en una serie de valores que están dados en progresión geométrica o aproximadamente.
- Su resultado puede ser utilizado en procesos estadísticos posteriores, puesto que los promedios geométricos en diferentes muestras, pueden ser combinados para el total de la muestra.

Desventajas

- Es bastante complicado el proceso de cálculo.
- Si cualquier valor de la serie es cero, el promedio será cero, dependiendo de la fórmula empleada.
- Si un valor de la serie es negativo, su valor será negativo o imaginario.

Cálculo en datos sin agrupar.

Ejemplo 1. Supongamos que se tienen 5 observaciones (n = 5) cuyos valores son: x_i: 6 4 8 2 5. Se pide calcular la media geométrica

Solución:

$$M_0 = \sqrt[5]{6.4.8.2.5} = \sqrt[5]{1.920} = 4{,}54$$

> Esta fórmula presenta algunos problemas en su aplicación: a) Si un valor de la variable es cero el producto será cero y su raíz cero. b) Si uno de los valores es negativo, el producto será negativo, por lo tanto se tendrá una raíz imaginaria. c) Sí el valor de n es grande, será tedioso su cálculo.

Frente a la anterior situación es preferible utilizar los logaritmos. Veamos el procedimiento:

$$M_0 = \sqrt[n]{\pi x_i} = \pi x_i^{1/n} \qquad \Leftrightarrow \qquad \boxed{\log M_0 = \frac{\Sigma \log x_i}{n}}$$

Reemplacemos y se tendrá que $\log M_0 = \dfrac{\Sigma \log x_i}{n} = \dfrac{\log 6 + \log 4 + \log 8 + \log 2 + \log 5}{5}$

$$\log M_0 = \frac{0{,}77815 + 0{,}60206 + 0{,}90309 + 0{,}30103 + 0{,}69897}{5} = \frac{3{,}28330}{5} = 0{,}65666$$

M_0 = *anti* log de 0,65666 = 4,54 El mismo resultado obtenido anteriormente.

Cálculo en datos agrupados.

Ejemplo 2.

y_i	n_i	$\log y_i$	$n_i \log y_i$
1	4	0	0
2	5	0,30103	1,50515
3	7	0,47712	3,33985
4	8	0,60206	4,81648
6	6	0,77815	4,66891
Σ	30	--	14,33039
X_i	f_i	$\log X_i$	$f_i \log X_i$

Solución:

$$\log M_0 = \frac{\Sigma n_i \log y_i}{n} = \frac{14{,}33039}{30} \equiv 0{,}477680$$

$$M_0 = anti \log 0{,}477680 = 3$$

$$\boxed{\log M_0 = \frac{\Sigma f_i \log x_i}{n}}$$

Ejemplo 3.

y_i	n_i	$y_i^{n_i}$
1	4	1
2	5	32
3	7	2.187
4	8	65.536
6	6	46.656
Σ	30	--
X_i	f_i	$X_i^{f_i}$

Solución:

$$M_0 = \sqrt[n]{\pi y_i^{n_i}}$$

$$M_0 = \sqrt[30]{1x32x2.187x65.536x46.656} = 3{,}00$$

Como lo podemos observar, en los dos ejercicios que se acaban de realizar los resultados son exactamente iguales. Este último cálculo hizo ver que el anterior procedimiento fuera se considerara más engorroso.

Ejemplo 4. La media aritmética de 3 números es 8, la mediana es 8 y la media geométrica es $M_0 = \sqrt[3]{480}$. Se pide calcular los valores de esos tres números.

Solución:

$$M_1 = 8 \quad M_e = 8$$
$$24 = x_1 + x_2 + x_3$$

$$M_0 = \sqrt[3]{480}$$
$$24 = x_1 + 8 + x_3$$
$$24 - 8 = x_1 + x_3$$
$$16 = x_1 + x_3$$

$$\sqrt[3]{480} = \sqrt[3]{x_1 x_2 x_3}$$
$$480 = x_1.(8).x_3$$
$$60 = x_1.x_3$$
$$x_1 = \frac{60}{x_3}$$

Reemplazando $16 = x_1 + x_3 \quad \Leftrightarrow \quad 16 = \frac{60}{x_3} + x_3 \quad \Leftrightarrow 16x_3 = 60 + x_3^2$

Despejando se tiene que $x_3^2 - 16x_3 + 60 = 0$, de donde: $\quad x_1 = 6 \quad x_2 = 8 \quad x_3 = 10$

(Teniendo en cuenta dos números que sumados den 16 y multiplicados 60)

Ejemplo 5. Supongamos que un país tuvo en 1997 una población de 8 millones y subió a 12 millones para el 2011. ¿Cuál será el promedio (media) de poblacion para ese período?

$M_0 = \sqrt{x_1.x_2} = \sqrt{8(12)} = \sqrt{96} = 9{,}8$ Millones de personas

Ejemplo 6. Supongamos que un capital de $50.000 sea colocado a una tasa anual del 24% el primero de enero del 2.008. Si se capitaliza los días primero de enero, calcule el promedio del dinero invertido entre el 31 de diciembre del 2.008 y el 31 de diciembre del 2011.

Solución:

AÑOS	x_i	$\log x_i$
2008	50.000	4,698970
2009	62.000	4,792392
2010	76.880	4,885813
2011	93.331,2	4,979235
Σ		19,356410

$$\log M_0 = \frac{\Sigma \log x_i}{n}$$
$$\log M_0 = \frac{19{,}356410}{4} = 4{,}8391025$$
$$M_0 = anti\log\ 4{,}8391025 = 69.040{,}27$$

(Promedio de dinero invertido durante el período)

MEDIA ARMÓNICA: $M_H \quad M_{-1}$

Este promedio, menos importante que los tres anteriores mencionados, se aplica en especial cuando la variable está dada en forma de tasas o cuando se trata calcular la velocidad media. Vale la pena decir, que esta medida le da gran importancia a aquellos valores pequeños de la variable, contrario a la media aritmética.

Se define, diciendo que: *dada una serie de datos, el inverso de la media armónica, es igual a la media aritmética del inverso de los valores de la variable.*

$\frac{1}{M_H} = M\left[\frac{1}{x_i}\right]$ Por lo tanto se tendrá que $M_H = \frac{n}{\Sigma\left[\frac{1}{x_i}\right]}$ ó $\boxed{M_H = \frac{N}{\Sigma\left[\frac{1}{x_i}\right]}}$

Ventajas

- Se usa con preferencia en el cálculo de la velocidad media.
- De gran utilidad cuando la variable está dada en forma de tasas.

Desventajas

- Que un valor de la variable sea cero.
- El promedio armónico está rígidamente definido por una fórmula matemática, su valor depende de cada uno de los datos de la distribución y su resultado no puede ser utilizado en cálculos posteriores.

Datos no agrupados

Ejemplo 1. Supongamos que se dispone de 10 observaciones, cuyos valores son:

x_i = 8 2 8 6 2 2 6 8 2 4 Con ellos se desea calcular la media armónica, aplicando la fórmula anterior.

$$M_{-1} = \frac{10}{\frac{1}{8}+\frac{1}{2}+\frac{1}{8}+\frac{1}{6}+\frac{1}{2}+\frac{1}{2}+\frac{1}{6}+\frac{1}{8}+\frac{1}{2}+\frac{1}{4}} = 3{,}38$$

Si comparamos este resultado, con el de la media aritmética aplicada a los mismos datos, siendo igual a 4,8, notaremos que es inferior, hecho que se dará en todos los ejercicios que realicemos, siendo: $M_{-1} < M_x$

Datos agrupados

Ejemplo 2. Con los datos correspondientes a una tabla de frecuencias, tal como aparece a continuación, tanto para la *variable discreta* como la continua, la Media armónica será:

1- VARIABLE DISCRETA

y_i	n_i	n_i/y_i
1	4	4,00
2	5	2,50
3	7	2,33
4	8	2,00
6	6	1,00
Σ	30	11,83
X_i	f_i	f_i/X_i

(1) $M_H = 30/11{,}83 =$

$M_H = 2{,}54$

(2) $M_H = 50/0{,}70 =$

$M_H = 71{,}43$

2 - VARIABLE CONTINUA

$y'_{i-1} - y'_i$	y_i	n_i	n_i/y_i
46,1 – 50	50	3	0,06
50,1 – 62	58	8	0,14
62,1 – 77	66	11	0,17
77,1 – 82	74	7	0,09
82,1 – 84	82	9	0,11
84,1 – 90	90	12	0,13
Σ	--	50	0,70
$X'_{i-1} - X'_i$	X_i	f_i	f_i/X_i

Lo mismo que en datos sin agrupar, los anteriores resultados en datos agrupados, deben ser inferiores a los obtenidos aplicando la media aritmética.

Fórmula aplicada en datos agrupados

$$M_H = \frac{n}{\sum\left[n_i/y_i\right]}$$

$$\boxed{M_H = \frac{N}{\sum\left[f_i/x_i\right]}}$$

Ya se mencionó al definir la *media armónica*, que una de sus aplicaciones se realizaba al calcular la *velocidad media*, es decir, cuando hay necesidad de relacionar tiempo con velocidad. Supongamos disponer de la siguiente información:

$\boxed{S_i = V_i t_i}$ S_i= *espacio* V_i= *velocidad* t_i= *tiempo* V_m= *velocidad media*

$$V_m = \frac{S}{\sum\left[S_i / V_i \right]} \qquad \text{Siendo} \quad S = S_1 + S_2 + \ldots\ldots\ldots\ldots \quad \text{Total de espacio}$$

Ejemplo 3. Consideremos, un auto recorre 60 km a una velocidad promedio de 100 km. por hora y luego, los siguientes 40 km. a una velocidad media de 80 km. por hora. ¿Cuál será la velocidad media del recorrido total?

Solución:

$$V_1 = 100 \qquad V_2 = 80 \qquad S_1 = 60 \qquad S_2 = 40$$

$$V_m = \frac{100}{60/100 + 40/80} = \frac{100}{0{,}6 + 0{,}5} = \frac{100}{1{,}1} = 90{,}9 \text{ Velocidad media}$$

A continuación veremos una forma errónea en la aplicación de la media aritmética. En este caso en particular, debe aplicarse la media armónica

Ejemplo 4. Un grupo de trabajadores producen 120 papeleras para escritorio, en un promedio diario de 12 papeleras. Una vez terminado ese contrato, se dedican a producir otras 120 papeleras, a razón de 8 por día. Se desea determinar la productividad diaria en la elaboración de las 240 papeleras.

Solución:
Si calculamos la media aritmética obtendríamos un resultado erróneo

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{12+18}{2} = 15 \text{ Papeleras diarias}$$

Se observa que las primeras papeleras se realizan en 10 días (120/12 = 10) y las siguientes papeleras en 6,67 días (120/8 = 6,67); las 240 papeleras se harían en 16,67 días; si son 16,67 y el rendimiento por día es de 15 se tendrían un total de 250 papeleras, cantidad diferente a las 240 papeleras del problema

Mediante la media armónica se tiene: $M_{-1} = \dfrac{2}{1/12 + 1/18} = \dfrac{2}{0{,}13888} = 14{,}4$

La aplicación de la media no es correcta, debido a que existe una relación inversamente proporcional. Con la media armónica, el número de papeleras producidas diariamente es de 14,4, que a la vez multiplicadas por 16,67 días nos daría un total de producción de 240 papeleras durante ese período, diferente a las 250 como resultado al aplicar la media.

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **Sistema de Información en Línea SIL.**

51. Calcule la media geométrica y la media armónica en la siguiente distribución de seis intervalos de amplitud constante, de la cual se sabe:

$n = 150$	$y_5 = 60$	$n_1 = n_6$	$n_3 = n_4 = 30$	$n_2 = n_5 = n_1 + 5$	$\overline{y} = 45$
$N = 150$	$X_5 = 60$	$f_1 = f_6$	$f_3 = f_4 = 30$	$f_2 = f_5 = f_1 + 5$	$\overline{X} = 45$

52 Halle la media aritmética, media geométrica, modo y mediana con los siguientes datos sin agrupar: 3 6 13 22 46 89 y conteste ¿Cuál de estas medidas es mejor y por qué?

53. Con los siguientes datos: 2 5 8 12 y 20
    a) Calcular la media aritmética, media geométrica y media armónica.
    Observar que $\boxed{M_1 > M_0 > M_{-1}}$
    b) Calcular la media, la mediana y la moda. Observar en cuál de las dos situaciones se presenta:
    $\boxed{M_1 > M_e > M_d}$ o $\boxed{M_1 < M_e < M_d}$

54. Los gastos (miles de millones de $) en publicidad de 50 empresas comerciales, durante el mes de diciembre de 2011 agruparon en 4 clases de amplitud constante, de la cual se sabe:

$y_1 = 3{,}5$	$y'_4 = 8{,}75$	$n_1 = 4$	$N_2 = 20$	$n_3 = 25$
$X_1 = 3{,}5$	$X'_4 = 8{,}75$	$f_1 = 4$	$F_2 = 20$	$f_3 = 25$

Se pide calcular la media aritmética y la media geométrica.

55. Una persona maneja su automóvil durante 400 kilómetros. Los primeros 120 km viaja a razón de 60 km por hora; los siguientes 120 a 100 km por hora; el 25% del total lo hace a razón de 80 km por hora. ¿A qué velocidad debe viajar el resto, para tener en total una velocidad promedio de 70 km por hora?

56. Un avión vuela una distancia de 900 km. Si cubre el primero y el último tercio del viaje a 250 km por hora y el tercio medio a 300 km por hora, se pide calcular su velocidad media.

57. Una persona viaja durante 4 días. Diariamente recorre 200 km, pero maneja el primero y el último día a 50 km por hora, el segundo a 55 y el tercero a 60 km por hora. ¿Cuál es la velocidad media durante el viaje?

58. Una persona viaja en coche de la ciudad "A" a la "B" con una velocidad media de 40 km por hora y vuelve de la ciudad "B" a la "A" con una velocidad media de 60 km por hora, hallar la velocidad media del viaje completo.

59. Las ciudades A, B y C, son equidistantes entre sí. Un motorista viaja de A a B a 30 km por hora, de B a C a 40 km por hora y de C a A a 50 km por hora. Determinar el promedio de velocidad para el viaje completo.

60. Durante un mes se construyeron 134 kilómetros de carretera en la siguiente forma: la primera semana 3,6% del total, la segunda semana 7,6% del total, la tercera semana 15,3%, la cuarta semana,

24,5% y la última 49%. Halle la medida de tendencia central que represente mejor el promedio de la distribución, en kilómetros por semana.

61. Se sabe que dos obreros gastan en la ejecución de un trabajo 50 y 40 minutos respectivamente, ¿cuál es el tiempo medio requerido para hacerlo en conjunto?

62. Un grupo de trabajadores produce 140 papeleras para piso con una productividad de 14 papeleras diarias; una vez terminado ese contrato se dedica a producir otras 140 papeleras razón de 10 por día. Se desea determinar la productividad diaria en la elaboración de las 280 papeleras.

63. Una persona va a tres tiendas del barrio a comprar azucar; los precios son como siguen:

TIENDA	PRECIO
A	\$2.250
B	\$2.830
C	\$2.570

Si la persona compra azúcar en dos formas diferentes. La primera forma consiste en que compra 3 paquetes en cada tienda. La segunda forma corresponde a comprar en cada tienda el equivalente de \$10.000 en azúcar.

a) ¿Cuál es el precio promedio por paquete que la persona paga cuando compra el azúcar de la primera forma?
b) ¿Cuál es el precio promedio por paquete cuando lo compra de la segunda forma?

64. En una tabla de frecuencias, se clasifican 50 datos utilizando 4 intervalos de igual magnitud. Se pide calcular la mediana y la media armónica sabiendo que:

$y_2 = 50$ $\quad n_1 = 4$ $\quad N_2 = 20$ $\quad n_3 = 25$ $\quad \bar{y} = 62,4$

65. Se sabe que dos obreros gastan en la ejecución de un trabajo 150 y 140 minutos respectivamente ¿Cuál es el tiempo medio requerido para hacerlo en conjunto?
66. Un fabricante dispone de \$100.000, para la compra de cierta materia prima. Durante tres años invierte la misma cantidad de dinero. Si el precio promedio por kilo en los tres años sucesivos sube de \$1.200 a \$1.800 y a \$3.600. ¿Cuál es el precio promedio que ha pagado el fabricante en dicho período?

67. Con la siguiente distribución:

$y'_{i-1} - y'_i$:	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35
n_i :	3	7	15	25	10

$X'_{i-1} - X'_i$
f_i

Se pide calcular la mediana y la media armónica.

MEDIA CUADRÁTICA: M_2

Se usa con muy poca frecuencia esta medida de tendencia central. Es un promedio que involucra a los cuadrados de las cantidades promediadas. La desviación estándar o típica de la cual hablaremos en la próxima unidad, es una media cuadrática. Es utilizada a menudo con el fin de reducir la media al mismo grado de los valores promediados, siendo la razón por la cual se calcula la raíz cuadrada de la misma, después de haber realizado el procedo de su promedación.

Se define como *la raíz cuadrada de la media aritmética, de los cuadrados de los valores que toma la variable”.*

Fórmulas: Datos sin agrupar $\boxed{\Leftrightarrow M_2 = \sqrt{\frac{\Sigma x_i^2}{n}}}$

Datos agrupados: $\boxed{M_2 = \sqrt{\frac{\Sigma y_i^2 n_i}{n}}}$ $\boxed{M_2 = \sqrt{\frac{\Sigma X_i^2 f_i}{N}}}$

Ventajas

- Es de importancia en algunos problemas de probabilidad, siendo teóricamente recomendable trabajar con el cuadrado de los valores.

Desventajas

- Su fórmula está definida rígidamente por una fórmula matemática.
- Se deja influenciar mucho por valores extremos, especialmente cuando éstos son demasiado grandes.

Datos sin agrupar

Ejemplo 1. Supongamos que se tienen 6 observaciones, cuyos valores son: $x_i =$ 2 6 5 4 6 8 y se desea calcular la media cuadrática de esos seis números.

$$M_2 = \sqrt{\frac{2^2 + 6^2 + 5^2 + 4^2 + 6^2 + 8^2}{6}} = 5{,}49 \qquad \text{La media hubiese sido } \bar{x} = \frac{31}{6} = 5{,}17$$

Ejemplo 2. Consideremos los datos de la tabla siguiente para calcular la media aritmética y la media cuadrática.

y_i	n_i	$y_i n_i$	$y_i^2 n_i$
1	4	4	4
2	5	10	20
3	7	21	63
4	8	32	128
6	6	36	216
Σ	30	103	431
X_i	f_i	$X_i f_i$	$X_i^2 f_i$

Media: $\bar{y} = \frac{\Sigma y_i n_i}{n} = \frac{103}{30} = 3{,}43$ $\boxed{M_2 = \sqrt{\frac{\Sigma X_i^2 f_i}{N}}}$

Media cuadrática: $M_2 = \sqrt{\frac{\Sigma y_i^2 n_i}{n}} = \sqrt{\frac{431}{30}} = 3{,}79$

En los dos casos, donde hemos calculado las dos medias, permite observar que $M_1 < M_2$ (La Media aritmética es menor que la media cuadrática)

$3{,}43 < 3{,}79$

MEDIA CÚBICA: M_3

Este promedio, al igual que el anterior es poco conocido y por lo tanto de uso limitado. Se define como "*la raíz cúbica de la media aritmética de los cubos de los valores de la variable*"

$$M_3 = \sqrt[3]{\frac{\Sigma x_i^3}{n}} \quad \boxed{M_3 = \sqrt[3]{\frac{\Sigma X_i^3}{N}}} \qquad\qquad M_3 = \sqrt[3]{\frac{\Sigma y_i^3 n_i}{n}} \quad \boxed{M_3 = \sqrt[3]{\frac{\Sigma X_i^3 f_i}{N}}}$$

Datos sin agrupar Datos agrupados

Datos sin agrupar

Ejemplo 1. Consideremos nuevamente los datos no agrupados, que fueron utilizados para calcular la Media cuadrática $x_i = 2 \quad 6 \quad 5 \quad 4 \quad 6 \quad 8$.

$$M_3 = \sqrt[3]{\frac{\Sigma x_i^3}{n}} = \sqrt[3]{\frac{2^3 + 6^3 + 5^3 + 4^3 + 6^3 + 8^3}{6}} = 5{,}75 \qquad \boxed{M_1 < M_2 < M_3}$$

Veamos: 3,43 < 3,79 < 5,75 por lo tanto se cumple la anterior relación. Si relacionamos estas tres medidas, con la media geométrica y la media armónica, observamos que: $M_{-1} < M_0 < M_1 < M_2 < M_3$ que el lector podrá comprobar desarrollando un ejercicio que involucre el cálculo de estas medidas.

Datos agrupados

Ejemplo 2. Con los siguientes datos, correspondiente a una tabla de frecuencias, se podrá calcular la Media cúbica. Veamos como es el proceso:

y_i	n_i	$y_i n_i$	$y_i^2 n_i$	$y_i^3 n_i$
1	4	4	4	4
2	5	10	20	40
3	7	21	63	189
4	8	32	128	512
6	6	36	216	864
Σ	30	103	431	1.609
X_i	f_i	$X_i f_i$	$X_i^2 f_i$	$X_i^3 f_i$

$$M_3 = \sqrt[3]{\frac{1^3(4) + 2^3(5) + ...6^3(6)}{30}} = \sqrt[3]{\frac{1.609}{30}} = 3{,}77$$

$$M_3 = \sqrt[3]{\frac{\Sigma y_i^3 n_i}{n}} \qquad \boxed{M_3 = \sqrt[3]{\frac{\Sigma X_i^3 f_i}{N}}}$$

Ejemplo 3. Consideremos que la media aritmética de tres números es 7, su mediana es 6 y su media geométrica es $\sqrt{216}$. Se pide calcular la media cuadrática y la cúbica.

Solución:

$$\bar{x} = 7 \qquad M_e = 6 \qquad M_0 = \sqrt{216}$$

$$\bar{x} = \frac{x_1 + 6 + x_3}{3} = 7 \Rightarrow 21 = x_1 + 6 + x_3 \Rightarrow 15 = x_1 + x_3$$

$$M_0 = \sqrt[3]{216} = \sqrt[3]{x_1.x_2.x_3} = \sqrt[3]{x_1(6)x_3} \Leftrightarrow 216 = x_1(6)x_3$$

$$\frac{216}{6} = x_1.x_3 \Rightarrow 36 = x_1.x_3 \Rightarrow x_1 = \frac{36}{x_3} \qquad \text{Reemplazando } 15 = \frac{36}{x_3} + x_3 \text{ se tiene que}$$

$$15x_3 = 36 + x_3^2 \Leftrightarrow x_3^2 - 15x_3 + 36 = 0 \qquad \text{Siendo: } x_1 = 3 \quad x_2 = 6 \quad \text{y} \quad x_3 = 12$$

La media cuadrática será: $M_2 = \sqrt{\frac{3^2 + 6^2 + 12^2}{3}} = \sqrt{\frac{9 + 36 + 144}{3}} \cong 7{,}94$

La media cúbica será: $M_3 = \sqrt[3]{\frac{3^3 + 6^3 + 12^3}{3}} = \sqrt[3]{\frac{27 + 216 + 1.728}{3}} = 8{,}69$

RELACIÓN ENTRE LAS MEDIDAS DE TENDENCIA CENTRAL O PROMEDIOS.

a) Cuando tratamos la media aritmética, la mediana y la moda, mencionamos algunas relaciones existentes entre ellas

$\bar{x} = M_e = M_d$ En este caso la Distribución se considera *Simétrica*

$\bar{x} > M_e > M_d$ Se dice que la Distribución es *Asimétrica positiva*

$\bar{x} < M_e < M_d$ Es un caso de Distribución *Asimétrica negativa*

Los tres casos anteriores serán nuevamente mencionados en el Capítulo siguiente.

b) Si la distribución es convexa y moderada asimétrica, la *Mediana* se ubica entre la *Media* y el *Modo*, quedando aproximadamente, dos veces más lejos de este último que del primero. $\bar{x} - M_d = 3(\bar{x} - M_e)$ y $M_d = 3\,M_d - 2\,\bar{x}$

c) En cualquier distribución, donde los elementos originales difieren en tamaño, las siguientes medidas diferirán en valor, de acuerdo al siguiente orden.

$$M_{-1} < M_0 < M_1 < M_2 < M_3$$

CENTRO RECORRIDO C_R

Esta medida poca utilizada, se define como la media aritmética de los valores extremos que toma la variable.

$$C_r = \frac{x_{máx} + x_{mín}}{2} \qquad C_r = \frac{y_m + y_1}{2} \qquad C_r = \frac{y'_m + y'_0}{2}$$

(Datos sin agrupar) (Variable discreta) (Variable continua)

Ejemplo 1. Supongamos que se tiene la siguiente información, correspondiente a datos sin agrupar:
x_i = 2 6 5 4 6 8 12 7 6 10 Su centro recorrido será:

$$C_r = \frac{x_{máx} + x_{mín}}{2} = \frac{12+2}{2} = 7$$

Ejemplo 2. Su aplicación tanto en la variable discreta como en la continua, se puede observar a continuación.

a)

y_i	n_i
1	4
2	5
3	7
4	8
6	6
Σ	30
X_i	f_i

$$C_r = \frac{y_m + y_1}{2} \qquad C_r = \frac{6+1}{2} = 3{,}5$$

b)

$$C_r = \frac{y'_m + y'_0}{2} \qquad C_r = \frac{94+46}{2} = 70$$

En ninguno de los dos casos se tiene en cuenta, el valor que tomen las frecuencias absolutas.

$y'_{i-1} - y'_i$	y_i	n_i
46,1 – 54	50	3
54,1 – 62	58	8
62,1 – 70	66	11
70,1 – 78	74	7
78,1 – 86	82	9
86,1 – 94	90	2
Σ	--	40
$X'_{i-1} - X'_i$	X_i	f_i

CUARTILES, DECILES Y PERCENTILES

Cuando la distribución está constituida por un número grande de intervalos o de marcas de clase, haciéndose necesario calcular un promedio sobre una parte de ella, en estos casos, la distribución puede ser distribuida en cuatro, en diez o en cien partes. En el primer caso nos referiremos a *Cuartiles*, en el segundo a *Deciles* y en el tercero a *Percentiles* o *Centiles.*

En el primero o sea los cuartiles, se podrá obtener: Q_1 como *primer cuartil*, Q_2 *segundo cuartil* y Q_3 correspondiente al *tercer cuartil*.

El *primer cuartil,* es aquel valor de la variable que supera al 25% de las observaciones y a la vez es superado por el 75%.

El *segundo cuartil,* es aquel valor de la variable que supera al 50% de las observaciones y a la vez es superado por el 50%. Es decir, corresponde a la Mediana

En los *deciles, el séptimo decil,* es aquel valor de la variable que supera al 70% de las observaciones y a la vez es superado por el 30%.

Y en los percentiles, el *percentil cuarenta*, es aquel valor de la variable que supera al 40% de las observaciones y a la vez es superado por el 60%.

Tal como lo mencionamos, el segundo cuartil, el quinto decil y el percentil cincuenta, en una misma distribución presenta el mismo resultado, siendo en los tres casos igual al de la mediana.

Datos no agrupados

Ejemplo 1. Con los siguientes datos: 16 10 4 8 12 10 8 20 4 13 12 22 16 26 20, se pide calcular a) Primer y tercer cuartil. b) Cuarto y sexto decil y c) el 30 y 90 percentil.

a) Lo primero que se hace, es ordenar los datos de menor a mayor o viceversa.

4 4 8 8 10 10 12 12 13 16 16 20 20 22 26 $n = 15$

a) Para el *primer cuartil*, aplicamos el siguiente procedimiento, muy parecido al calcular la mediana, es decir: $\dfrac{1(n+1)}{4} = \dfrac{1(15+1)}{4} = 4^{a}$ *posición* $Q_1 = x_j = 8$

Para el *tercer cuartil*: $\dfrac{3(n+1)}{4} = \dfrac{3(15+1)}{4} = 12^{a}$ *posición* $Q_3 = x_j = 20$

b) *Cuarto decil* : $\dfrac{4(n+1)}{10} = \dfrac{4(15+1)}{10} = 6{,}4^{a}$ *posición.* Observemos que el proceso correspondiente al resultado obtenido es diferente a los dos anteriores, pues aparece con la posición de 6,4, es decir, que debe estar entre la sexta y la séptima posición, siendo el valor de 10 la sexta posición y la diferencia entre la 7ª y 6ª es 12-10=2, este resultado se multiplica por la fracción 0,4, así 2(0,4)=0,8 y se lo agregamos al valor anterior de 10, siendo: $D_4 = x_j = 10{,}8$

El *sexto decil*, será: $\dfrac{6(n+1)}{10} = \dfrac{6(15+1)}{10} = 9{,}6^{a}$ *posición*. El valor correspondiente a la novena posición es 13 y se tiene además, 0,60(16-13) = 0,6(3) = 1,8. Por lo tanto $D_6 = 14{,}8$

c) *Percentil 30* será: $\dfrac{30(n+1)}{100} = \dfrac{30(15+1)}{100} = 4{,}8^{a}$ *posición*. El valor de la cuarta posición es 8, además se tiene que, 0,8(10 – 8) = 1,6. Por lo tanto $P_{30} = 9{,}6$

El *Percentil 90* será: $\frac{90(n+1)}{100} = \frac{90(15+1)}{100} = 14{,}4^{a}$ *posición*. El valor en la posición 14 es de 22, además, 0,4(26 – 22) = 1,6, se tendrá que P_{90}= 23,6

Datos agrupados

El cálculo de los *cuartiles, deciles* y *percentiles* es bastante parecido al procedimiento que se había utilizado para calcular la *Mediana*, salvo una pequeña variación, al establecer la posición donde se debe ubicar la respectiva medida.

Los pasos a seguir en los cálculos son:

- Se determinan las *frecuencias absolutas acumuladas.*
- Determinamos la *posición* donde se localiza.
- Observamos si el valor obtenido en el paso anterior, aparece o no en dicha columna, correspondiente a las frecuencias absolutas acumuladas.
- Se aplica la fórmula respectiva, dependiendo de la variable trabajada.

$y'_{i-1} - y'_i$	n_i	N_i	y_i
2,1 – 6	3	3	4
6,1 – 10	12	15	8
10,1 – 14	10	25	12
14,1 – 18	18	43	16
18,1 – 22	10	53	20
22,1 – 26	6	59	24
26,1 – 30	20	79	28
30,1 – 34	8	87	32
34,1 – 38	3	90	36
Σ	90	--	-----
$X'_{i-1} - X'_i$	f_i		X_i

Se pide calcular:

a) *Cuartil tres*. La posición se da así: $\frac{3(n)}{4} = \frac{3(90)}{4} = 67{,}5^{a}$ *posición*. Observemos que este valor no aparece en la columna de las frecuencias absolutas acumuladas, por tal razón, seleccionamos el valor inmediatamente anterior (59) como N_{j-1} y al valor inmediatamente superior teniendo a 79 como N_j. Ahora veamos las fórmulas a aplicar dependiendo de la variable.

Variable Discreta $\rightarrow Q_3 = y_j = 28$. Como se observa, se toma el valor de y_j al frente de N_j si consideramos a las marcas de clase como *variable discreta*.

Variable continua. En este caso nos ubicamos al frente de los intervalos y aplicamos la fórmula parecida a la utilizada al calcular la *Mediana,* con una ligera modificación:

$$Q_3 = y'_{j-1} + c\left[\frac{3n/4 - N_{j-1}}{n_j}\right] = 26 + 4\left[\frac{67{,}5 - 59}{20}\right] = 26 + 1{,}7 = 27{,}7$$

Observemos que el valor de 26 fue seleccionado en el intervalo correspondiente a N_j

b) *Decil seis* Determinamos la posición así: $\frac{6(n)}{10} = \frac{6(90)}{10} = 54^{a}$ Cómo este valor no aparece en la columna de las frecuencias absolutas acumuladas, se toma a N_{j-1}=53 y a N_j =59.

Ahora observemos el procedimiento a seguir en ambas variables.

Variable Discreta $\rightarrow D_6 = y_j = 24$ (*Marcas de clase*).

Variable Continua $\rightarrow D_6 = 22 + 4\left[\frac{54 - 53}{6}\right] = 22 + 0{,}66 = 22{,}66$

c) *Percentilil* 80. La posición es: $\frac{80(n)}{100} = \frac{80(90)}{100} = 72^{a}$*posición.* Al igual que en los dos casos anteriores, se deberán aplicar los mismos procedimientos para esta tabla. N_{j-1}= 59 y para N_j= 79

$$\textit{Variable Discreta} \rightarrow P_{80} = y_j = 28 \ (\textit{Marcas de clase}).$$

$$\textit{Variable Continua} \rightarrow P_{80} = 26 + 4\left[\frac{72-59}{20}\right] = 26 + 2{,}6 = 28{,}6$$

EJERCICIOS MISCELÁNEOS PARA RESOLVER

Complemento virtual SIL

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el **SIL.**

68. Una compañía de aviación que ofrece un vuelo diario a una determinada región del país no cumplió con el horario de llegada en los últimos 10 días de abril, con los siguientes minutos de retardo o de anticipación (número negativo):

-3	6	4	10	-4	124	2	-1	4	1

a) Si la compañía contratara un especialista en estadística para mostrar el cumplimiento, ¿cuáles serían las medidas que utilizaría?
b) Si el objetivo fuese mostrar que ofrece un buen servicio (cumplimiento), ¿qué medida utilizaría?
Si el objetivo es mostrar un mal servicio de una compañía con la que él desea competir ¿qué medida utilizaría para lograrlo?

69. Explique brevemente:
a) Tres ventajas de la media
b) Dos desventajas de la mediana
c) Dos ventajas de la moda
d) ¿En qué condiciones serán iguales la media, mediana y moda?

70. Con los siguientes datos, se pide calcular:

$y'_{i-1} - y'_i$	n_i
6,1 — 12	2
12,1 — 15	14
15,1 — 20	5
20,1 — 28	3
28,1 — 36	7
36,1 — 40	16
40,1 — 50	3
Σ	50

a) Media aritmética
b) Mediana
c) Moda
d) Tercer decil
e) Segundo cuartil
f) Percentil sesenta

71. Un grupo de 400 empleados, que tiene una compañía, se dividen en operarios y técnicos con un salario promedio de \$1.260.960. Los salarios promedio para cada uno de los grupos son de \$857.300 y \$1.320.856 respectivamente.
a) ¿Cuántos operarios y cuántos técnicos tiene la compañía?
b) Si el gerente establece una bonificación de \$30.000 para los operarios y del 8% para los técnicos, ¿cuál será el salario promedio para los 400 empleados de la compañía?

72. ¿Qué es un promedio?. ¿cuáles podría mencionar?; ¿qué condiciones debe tener el promedio para cumplir su cometido, mencione tres?

73. Con los datos del ejercicio N° 45, se pide calcular: la media cuadrática, la cúbica y el séptimo decil.

74. Con los siguientes datos:

$x_1 = 4$	$x_2 = 10$	$x_3 = 6$	$x_4 = 4$	$x_5 = 4$	$x_6 = 12$

a) Comprobar que $M_{-1} < M_O < M_1 < M_2 < M_3$

b) Suponga que al calcular la mediana, media aritmética y moda, estos resultados provienen de una distribución de frecuencias. ¿Cuál sería su relación, es decir, simétrica o asimétrica?

75. Suponga que los salarios pagados a los empleados de una compañía, son como se representa a continuación:

CARGOS	NÚMERO	SALARIO MENSUAL
Directores	2	1.930.000
Supervisores	4	1.510.000
Economistas	6	1.370.000
Contadores	4	1.350.000
Auxiliares	26	646.000
Obreros	110	590.000

Se pide:

a) Calcular la media, mediana y moda.

b) En su opinión. ¿Cuál de estos promedios considera más representativo?

76. En una fábrica de tres secciones se sabe que en la sección A, con 120 obreros la asistencia promedio es de 240 días al año; en la sección B, que tiene 180 operarios, la asistencia media es de 216 días al año. Si la asistencia media en toda la fábrica es de 226,70 días. ¿cuántos obreros hay en la sección C, donde la asistencia promedio es de 230 días al año?

77. Durante diez días se observó en un almacén de autos, el número de estos vendidos por su empleado estrella.

DÍA:	1	2	3	4	5	6	7	8	9	10
AUTOS:	3	4	2	1	3	2	4	6	5	4

Cada auto vale $18.500.000 y al empleado se le paga un sueldo de $270.000 básico mensual más el 0,5% del valor de cada auto vendido. ¿Cuál será el sueldo promedio para el vendedor estrella en este lapso de 10 días?

78. Se sabe que el salario mensual promedio de los empleados de una empresa es de $1.052.000. El total de los empleados se aumenta en 20% con relación a los que había hasta la fecha. Los nuevos empleados tienen salario mensual promedio igual al 36,8% de salario promedio de los antiguos. Un año después se hace un aumento de $57.600 a todos los empleados de la empresa. ¿Cuál será el salario promedio de la empresa actualmente?

79. La siguiente es la población de 3 zonas de un Departamento en que se encuentra dividido el país según los censos de 2001 y de 2011

ZONA	2001	2011
1	322.867	240.776
2	204.903	267.358
3	345.604	575.864

Calcule la población total de las tres zonas del Departamento para una fecha equidistante entre los dos censos, suponiendo que:

a) El porcentaje (tasa) de aumento poblacional es constante en todo el Departamento.

b) El porcentaje (tasa) de aumento es constante en cada zona del Departamento.

80.

$y'_{i-1} - y'_i$	n_i
800 – 1.000	5
1.000 – 1.200	13
1.200 – 1.400	17
1.400 – 1.600	8
1.600 – 1.800	7
$X'_{i-1} - X'_i$	f_i

Con la siguiente distribución, calcular:

a) La media, mediana y modo.

b) La media cúbica, cuadrática, armónica y geométrica.

81. Contestar verdadero si el enunciado es verdadero. En caso contrario, la palabra subrayada debe sustituirse por una expresión con la cual el enunciado sí sea válido.
   a. La media de una muestra divide los datos en dos mitades iguales: la mitad mayores y la mitad menores que su propio valor.
   b. La media armónica es la raíz enésima del producto de los valores que toma la variable.
   c. El decil ocho representa aquel valor de la variable que supera el 80% de las observaciones y es superado por el resto 20%.
   d. Una distribución o un conjunto de datos, permite el cálculo de varias medidas de tendencia central.
   e. Un promedio aplicado o calculado en un conjunto de datos provenientes de una muestra, se denomina parámetro.

82. Explique brevemente los siguientes puntos:
   a. Cuatro condiciones para que los promedios puedan cumplir su cometido.
   b. Cuatro reglas que deben ser tenidas en cuenta para el uso de los promedios.

83. Se sabe que la media aritmética de dos números es igual a 5 y la media geométrica de los mismos es igual a 4. ¿cuál es la media cúbica?

84. En una empresa constructora de vivienda, los jornales semanales tienen una media de $169.000. Como una solución al conflicto laboral surgido, se proponen dos soluciones al conflicto: (a) un aumento del 6% en el salario semanal; (b) un aumento del 4%, más una bonificación semanal de $5.800 a cada obrero. ¿cuál de las dos alternativas mejora la situación del obrero?

85. Dos empresas tienen ocupados 600 obreros, distribuidos así, el 30% en A y el resto en la empresa B. Se sabe que el promedio de salario en esta última es de $860.000 y en A es del 20% menos que de B. ¿Cuál es el promedio de salarios para el total de obreros?

86. Se tienen dos cursos A y B, donde el primero tiene un promedio de calificación de 3,4, mientras que en el B es de 4,2. Diga qué porcentaje de alumnos tiene A y B, si el promedio de calificación para la suma de los dos cursos es de 3,7.

87. Se tiene que

$y_2 = 22$	$y'_5 = 50$	$m = 6$	C = constante	$n_2 + n_4 = 15$	
$n_1 + n_3 = 25$	$n = 50$	$n_5 + n_6 = 10$	$n_1 = 8$	$n_5 = 2$	$n_2 = 7$

Se pide calcular la media armónica y cúbica.

88. Decir si son ciertos o falsos y por qué:

a)	$\bar{x} = 68$	$M_e = 62$	$M_d = 68$		
b)	$\bar{x} = 70$	$M_o = 65$	$M_2 = 72{,}50$	$M_{-1} = 63$	$M_3 = 76{,}8$

89. Dados los números 8, 12, 16, 20, encontrar la media armónica, la media geométrica, la media cúbica y la media cuadrática.

90. Contestar ***verdadero*** si el enunciado siempre es verídico. Si no lo es, la palabra en negrilla deberá sustituirse por una expresión con la cual el enunciado si sea válido.

a) A un grupo de valores se le puede calcular **más de un** promedio.

b) Un grupo de valores puede tener **más de una media** aritmética.

c) La **media cúbica** es la raíz enésima del producto de los valores de la variable.

d) La suma de las desviaciones con respecto al **origen de trabajo** es igual a cero.

e) En una serie sencilla (datos sin agrupar) cuando el valor de n **es impar**, la mediana es igual al valor central.

f) La **media** divide a los datos de una distribución en dos mitades: la mitad mayores y la mitad menores que su propio valor.

g) La **moda** es el valor de la variable que más se repite.

91. Un industrial tiene dos fábricas, la primera con 200 trabajadores y la segunda con 300, además se sabe que en promedio el salario mensual es de $1.026.328 y que los trabajadores de la segunda fábrica ganan $86.400 menos que los de la primera fábrica. ¿Cuál es el salario promedio en cada fábrica?

92. Con los siguientes datos, $n = 200$, con frecuencias relativas y acumuladas diferentes, ya que se trata de dos problemas (A y B), se pide:

	(A)	(B)
$y'_{i-1} - y'_i$	h_i	N_i
15,1 – 25	0,04	6
25,1 – 35	0,10	23
35,1 – 45	0,21	57
45,1 – 55	0,30	110
55,1 – 65	0,21	152
65,1 – 75	0,10	190
75,1 – 85	0 04	200
Σ	1,00	–
$X'_{i-1} - X'_i$	f_i/n_i	F_i

Primero trabajar con la columna de las frecuencias relativas (A) y luego con las frecuencias absolutas acumuladas. (B)

Nota: Son dos problemas diferentes, donde las frecuencias relativas no corresponden a las frecuencias absolutas *acumuladas*.

a) Calcular la media aritmética utilizando las fórmulas ya conocidas.
b) Calcular la mediana (variable continua). Obtener las marcas de clase y calcular la mediana.
c) Calcular el modo (aplicando la fórmula general y las establecidas para variables continuas).

93. Con los datos del punto anterior se pide calcular:
a) La media geométrica
b) El modo mediante la aplicación de la siguiente fórmula: $M_d = 3M_e - 2M_1$
c) Las medias: armónica, cuadrática y cúbica.

94. De tres números se sabe que su media cúbica es $\sqrt[3]{657}$ su media aritmética es 7 y su mediana es 6. Se pide calcular los valores de esos tres números.

95. Se sabe que la media aritmética de dos números es igual a 5 y la media geométrica de los mismos es igual a 4. ¿Cuál es su media armónica?

96. Con los siguientes datos sin agrupar, calcular la media cuadrática.

x_i :	8	2	6	10	2	6	8	10	2	6

97. Hallar la media cuadrática de la siguiente distribución

$y'_{i-1} - y'_i$	2,1 – 6	6,1 – 10	10,1 – 14	14,1 – 18	18,1 – 22	Σ	$X'_{i-1} - X'_i$
n_i	3	7	10	6	4	30	f_i

98. Con los siguientes datos, calcular el primer cuartil, tercer cuartil, sexto decil y el percentil 80

a) *Variable discreta*

y_i	n_i
2	3
4	6
6	15
8	8
10	2
12	2
14	10
Σ	50
X_i	f_i

b) *Variable continua*

$y'_{i-1} - y'$	n_i
3,1 — 8	14
8,1 — 13	15
13,1 — 18	8
18,1 — 23	6
23,1 — 28	7
28,1 — 33	10
Σ	60
$X'_{i-1} - X'_i$	f_i

Síntesis de capítulo

La posición, en una distribución de frecuencias, lo da el valor de la variable denominada promedio, alrededor de la cual se agrupan las frecuencias.

A los promedios se les conoce como medidas de posición, valor promedio o valor central.

Para cada caso en particular se deben aplicar algunas de las medidas de posición o de tendencia central, de tal manera que el valor represente al conjunto de datos, los cuales tendrán algunas ventajas y desventajas en su aplicación.

Hay que tener en cuenta que estas medidas o promedios, proporcionan información acerca de un conjunto, siendo una forma de describir el comportamiento de esos datos, pero no proporcionan ninguna idea del grado de variabilidad, es decir, su verdadera o aproximada representación de ese conjunto observado.

Se ha dicho en repetidas ocasiones, que los promedios aplicados a datos correspondientes a una muestra se les denomina estimadores y aplicados a datos de una población, se les conoce como parámetros.

Al tener varias clases de promedios: media, mediana, modo, etc., la cifra resultante al aplicar una de esas medidas puede traer confusión, ya que algunas de ellas se ajustan más a determinadas situaciones que otras. Hoy por hoy, la gente utiliza más la media aritmética, algunos pocos la mediana y la moda, sin tener el cuidado de seleccionar el promedio que mejor represente el resumen de los datos mediante una cifra que no distorsione.

Estos promedios utilizados sólo pueden ocultar información, cuando no representan correctamente el conjunto de datos. En estos casos se hace necesaria la aplicación de otras medidas como las de dispersión, que ayudan a una mejor interpretación de ese valor central y que se verán en el capítulo siguiente.

Reglas para el uso de los promedios

- ❑ Cuando la serie tenga forma de progresión geométrica, debe usarse el promedio geométrico.
- ❑ Para calcular la velocidad media debe usarse la media armónica.
- ❑ Cuando la distribución sea muy asimétrica, debe considerarse la posibilidad de usar le mediana o el modo.
- ❑ Cuando la distribución tenga forma de U o sea que pueda representarse por una cóncava de extremos iguales, debe usarse el modo.
- ❑ Cuando quiera dársele importancia a valores pequeños de la variable, es aconsejable la media geométrica.
- ❑ En una distribución cuyos valores extremos no están definidos, es aconsejable la mediana o el modo.
- ❑ Cuando haya alguna razón para pensar que el promedio aritmético no representa muy bien a la distribución, debido a que valores extremos lo afectan, o por otras razones, debe considerarse la posibilidad de usar la mediana o el modo.
- ❑ Cuando la amplitud de la distribución no es constante, no debe usarse el modo.
- ❑ Cuando se requiere promediar relaciones (tasas), se debe usar la media armónica.
- ❑ En los demás casos debe usarse la media aritmética.

ANEXOS A LA UNIDAD 3

Se explicará brevemente el uso de la calculadora, así como la aplicación del programa EXCEL, con lo cual el lector podrá agilizar los cálculos correspondientes, tal como se verá a continuación.

APLICACIONES CON LA CALCULADORA

- **Calculadora CASIO** $f_x - 5000F$
  - ❑ ***Primero.*** Se oprimen las teclas MODE X, con lo cual debe aparecer en la parte superior de la pantalla, el símbolo SD
  - ❑ ***Segundo.*** Se hace necesario, antes de iniciar los cálculos, borrar la información que hubiese con anterioridad, mediante la ejecución SHIFT DEL EXE.
  - ❑ ***Tercero.*** Se comienza con la entrada o captura de la información (datos).

Ejemplo 1. Consideremos los siguientes datos sin agrupar: x_i: 8 10 20 3 7 Iniciamos digitando el 8 y luego la tecla √ DT identificada por DT, correspondiente a DATA, a continuación se repite el proceso, pero ahora con 10, luego con 20 y así sucesivamente.

Ejemplo 2. Procedamos ahora a la entrada de los datos correspondientes a una tabla de frecuencias o datos agrupados, siendo $n = 40$

x_i: 2 4 6 8 10

n_i: 3 6 10 12 9

La entrada de estos datos se hará de la siguiente forma: 2 SHIFT ; 3 √ DT Luego se hará 4 SHIFT ; 6 DATA y así sucesivamente.

Ya digitados los datos, la obtención de los resultados al aplicar las fórmulas respectivas, tanto para el primero como en el segundo ejemplo, se hará de la siguiente forma:

$$\text{SHIFT} \quad \text{PROG} \begin{cases} 1 \rightarrow \Sigma x_i^2 \text{ ó } \Sigma y_i^2 n_i \\ 2 \rightarrow \Sigma x_i \text{ ó } \Sigma y_i n_i \\ 3 \rightarrow n \end{cases}$$

$$\text{Ahora con SHIFT} \begin{cases} 1 \rightarrow \bar{x} \text{ ó } \bar{y} \\ 2 \rightarrow {}_x\sigma_n \\ 3 \rightarrow {}_x\sigma_{n-1} \end{cases}$$

- **Calculadora CASIO** $f_x - 30MS$ **G**

❑ ***Primero.*** Al oprimir la tecla MODE aparecen en pantalla tres operadores COMP SD REG
1 2 3
oprimimos en el 2 para obtener SD con el cual iniciaremos los cálculos. Vale la pena indicar que se oprime el 1 para salir, una vez realizadas todas las operaciones.

❑ ***Segundo***. Borramos toda la información que se tenga con SHIFT MODE (CLR) con el cual aparecen otros operadores, seleccionando el 1 correspondiente SCL.

❑ ***Tercero***. La entrada o captura de todos los datos, se hace igual que con la anterior calculadora.

❑ ***Cuarto***. La salida u obtención de los resultados se hace de la siguiente forma:

Con $\boxed{\text{SHIFT}}$ $\boxed{1}$ se tienen tres opciones $\begin{cases} \boxed{1} \rightarrow \Sigma x_i^2 \quad ó \quad \Sigma y_i^2 n_i \\ \boxed{2} \rightarrow \Sigma x_i \quad ó \quad \Sigma y_i n_i \\ \boxed{3} \rightarrow \boxed{n} \end{cases}$

Con $\boxed{\text{SHIFT}}$ $\boxed{2}$ hay tres opciones $\begin{cases} \boxed{1} \rightarrow \bar{x} \quad ó \quad \bar{y} \\ \boxed{2} \rightarrow {}_x\sigma_n \\ \boxed{3} \rightarrow {}_x\sigma_{n-1} \end{cases}$

- **Calculadora CASIO** $f_x - 570MS$

- ❑ ***Primero.*** Al oprimir la tecla $\boxed{\text{MODE}}$ aparecen en pantalla dos opciones $\boxed{1}$ $\boxed{\text{COMP}}$ $\boxed{2}$ $\boxed{\text{CMPLY}}$

  Oprimimos nuevamente $\boxed{\text{MODE}}$ y nuevamente se nos dan otras tres opciones $\boxed{\text{SD}}$ $\boxed{\text{REG}}$ $\boxed{\text{BASE}}$

  $\boxed{1}$ $\boxed{2}$ $\boxed{3}$

  Procedemos a seleccionar el $\boxed{1}$ correspondiente a $\boxed{\text{SD}}$

- ❑ ***Segundo***. Procedemos a capturar información, tal como se realizó en la primera calculadora. Datos sin agrupar: $\boxed{8}$ $\boxed{\text{DT}}$ luego $\boxed{10}$ $\boxed{\text{DT}}$ y así sucesivamente.

  En datos agrupados, se procede así: $\boxed{2}$ $\boxed{;}$ $\boxed{3}$ luego $\boxed{4}$ $\boxed{;}$ $\boxed{6}$ y así sucesivamente.

- ❑ ***Tercero***. Los resultados aparecen, mediante el siguiente proceso:

  Con $\boxed{\text{SHIFT}}$ $\boxed{1}$ se seleccionan así: $\boxed{1}$ Σx_i^2 $\boxed{2}$ Σx_i $\boxed{3}$ n

  Con $\boxed{\text{SHIFT}}$ $\boxed{2}$ se seleccionan así: $\boxed{1}$ $\bar{x}$ $\boxed{2}$ ${}_x\sigma_n$ $\boxed{3}$ ${}_x\sigma_{n-1}$

- **Calculadora CASIO** $f_x - 350TL$

  Con MODE nos aparecen $\boxed{1}$ $\boxed{\text{COMP}}$ $\boxed{2}$ $\boxed{\text{SD}}$ $\boxed{3}$ $\boxed{\text{REG}}$, seleccionamos la opción $\boxed{2}$ y procedemos a capturar o digitar la información, exactamente igual como se hizo en los procesos anteriores. La salida de los resultados se hará así:

  Con $\boxed{\text{SHIFT}}$ $\boxed{1}$ se obtiene $\bar{x}$ → Media aritmética

  Con $\boxed{\text{SHIFT}}$ $\boxed{2}$ se obtiene ${}_x\sigma_n$ → Desviación típica

  Con $\boxed{\text{SHIFT}}$ $\boxed{3}$ se obtiene ${}_x\sigma_{n-1}$ → Desviación típica corregida

  Ahora con la tecla u opción $\boxed{\text{RCL}}$ y las letras $\boxed{A}$ $\boxed{B}$ $\boxed{C}$ se obtienen los resultados para

  $\boxed{\text{RCL}}$ $\boxed{A}$ Σx_i^2 $\boxed{\text{RCL}}$ $\boxed{B}$ Σx_i y $\boxed{\text{RCL}}$ $\boxed{C}$ n respectivamente.

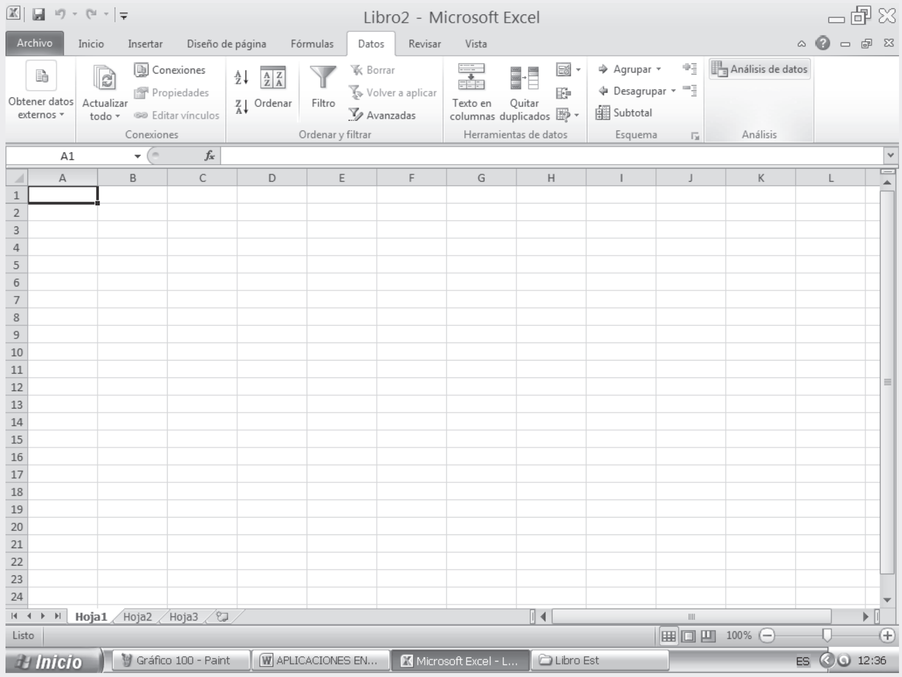
TRABAJO EN EXCEL PARA EL CÁLCULO DE LA MEDIA

Complemento virtual SIL

Para el cálculo de la media se realiza el siguiente procedimiento:

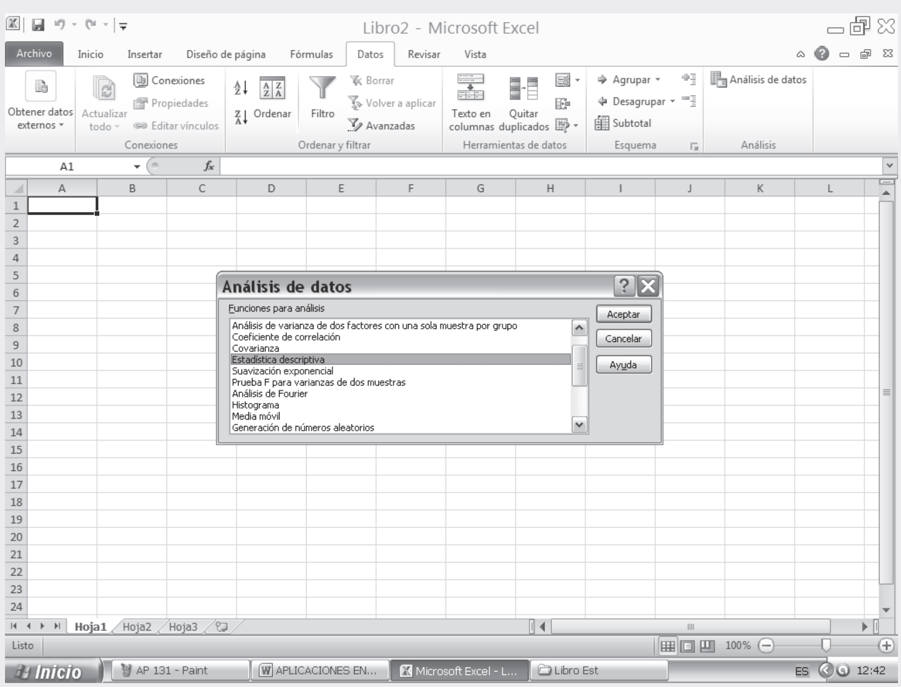
- ❑ Hacemos CLIC en ANÁLISIS DE DATOS. Para tener acceso a estas herramientas, hacemos clic en ANÁLISIS DE DATOS en el grupo ANÁLISIS de la ficha DATOS (Ver aplicativos en Excel del Capítulo 2)

Figura No. 1. Microsoft Excel



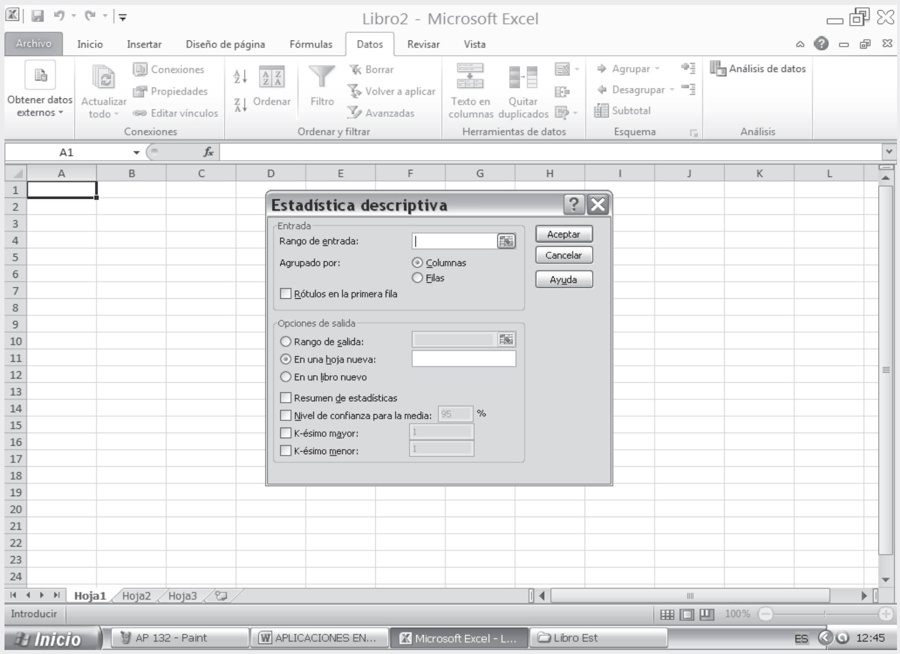
❑ Hacemos CLIC en el submenú ANÁLISIS DE DATOS. Nos aparece un cuadro de diálogo, en el cual seleccionamos una de las funciones, en nuestro caso la opción identificada como ESTADÍSTICA DESCRIPTIVA.

Figura No. 2. Funciones para análisis



❑ Luego al hacer CLIC en ACEPTAR, deberá aparecer la figura siguiente:

Figura No. 3. Estadística Descriptiva



- ❑ Tabla No. 1- Los datos que se presentan a continuación se encuentran en la segunda parte del SIL en internet. La tabla siguiente sólo muestra una parte de los datos que contiene el SIL, por tanto es conveniente para el desarrollo del siguiente ejercicio tomar la totalidad de los datos, es decir, los 1100 datos que aparecen allí.

Tabla No. 1. Población Total de una Institución Educativa

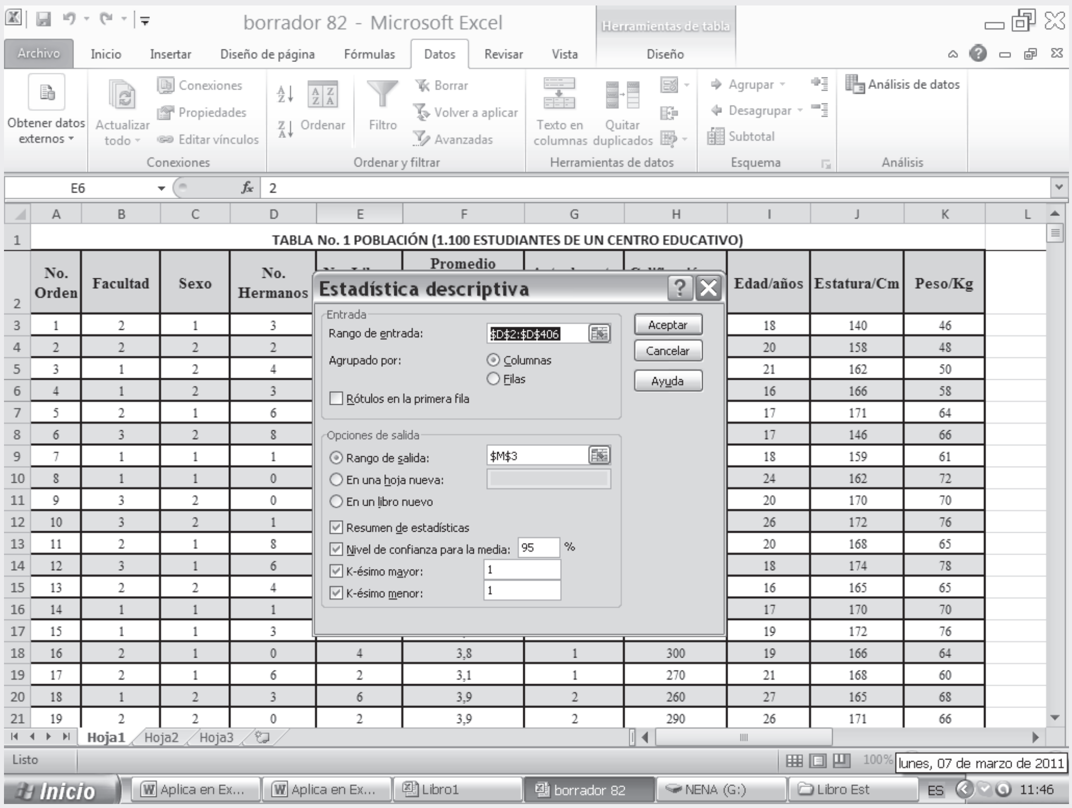
No. Orden	Facultad	Sexo	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciones ICFES	Edad/años	Estatura/Cm	Peso/Kg
1	2	1	3	4	2,6	1	410	18	140	46
2	2	2	2	2	4,1	1	360	20	158	48
11	2	1	8	5	3,6	1	320	20	168	65
16	2	1	0	4	3,8	1	300	19	166	64
17	2	1	6	2	3,1	1	270	21	168	60
75	3	2	3	3	3,1	1	280	19	154	50
147	3	1	1	8	5	1	310	17	174	83
148	2	2	4	2	4,2	2	310	16	182	70
149	1	1	6	3	2,6	1	320	20	178	88
150	3	1	2	4	2,8	1	400	24	165	70
151	1	2	0	6	3,4	2	380	20	165	58
152	1	2	0	3	2,8	1	310	20	171	59
153	1	2	3	5	4,2	2	350	22	169	64
154	2	2	4	8	4,1	2	400	23	172	68
155	2	1	2	2	4	2	400	24	163	72
156	3	2	6	2	4	2	380	20	165	58
157	1	1	5	2	3,6	2	420	26	178	88
158	1	1	6	2	3,2	1	300	20	182	84
159	3	2	2	6	3,8	2	315	19	180	70
160	1	1	4	4	4,2	1	310	18	174	81
276	1	1	0	2	4	2	350	19	162	69
277	1	1	0	6	3,8	2	320	20	168	68

276	1	1	0	2	4	2	350	19	162	69
277	1	1	0	6	3,8	2	320	20	168	68
376	2	1	0	4	3,8	1	300	19	166	64
377	2	1	6	2	3,1	1	270	21	168	60
378	1	2	3	6	3,9	2	260	27	165	68
379	2	2	0	2	3,9	2	290	26	171	66
380	3	1	0	3	4,1	1	380	30	169	72
381	2	2	8	7	3,9	2	410	24	168	70
382	3	1	0	10	4,4	2	420	25	180	84
383	1	1	1	3	4,2	1	390	22	174	80
384	2	2	6	6	4	2	370	23	166	68
385	3	1	0	11	3,9	2	290	24	162	60
386	3	2	0	8	3,9	2	300	19	151	66
387	2	2	2	2	4	1	330	18	158	60
388	1	2	3	2	3,3	2	310	17	159	58
399	1	1	2	2	4,5	1	300	28	150	56
400	3	2	0	6	3,6	2	280	17	148	46
408	1	1	0	3	3,5	2	270	17	163	65
440	1	1	4	3	2,7	2	320	15	168	76
441	3	1	0	3	3	2	290	17	171	82
442	3	1	0	6	3	2	410	18	174	86
443	1	1	4	2	3,6	1	400	19	176	84
444	2	2	7	4	3,8	1	380	20	180	70
445	3	2	2	2	4,5	2	360	17	178	69
814	2	2	3	10	2	1	280	18	165	60
815	2	2	2	8	4,3	2	270	18	160	61
816	1	1	0	2	4	2	350	19	162	69
817	1	1	0	6	3,8	2	320	20	168	68
818	1	2	2	2	3,6	1	320	17	174	68
819	3	1	3	2	4,6	2	410	16	176	69
820	2	2	1	6	1,8	1	240	17	172	69
821	3	1	3	5	3,1	2	280	17	169	76
822	3	1	8	5	1,9	2	250	16	164	70
828	2	2	0	6	4	2	370	17	174	69
869	1	1	6	3	2,6	1	320	20	178	88
870	3	1	2	4	2,8	1	400	24	165	70
871	1	2	0	6	3,4	2	380	20	165	58
997	1	1	0	6	3,8	2	320	20	168	68
998	1	2	2	2	3,6	1	320	17	174	68
999	3	1	3	2	4,6	2	410	16	176	69
1000	2	2	1	6	1,8	1	240	17	172	69
1068	3	2	2	2	3,4	2	300	26	172	70
1069	2	1	6	6	3,6	1	310	21	164	68
1075	1	2	3	2	2,6	1	270	15	165	60
1082	3	2	3	8	3,6	1	320	32	165	64
1083	1	1	6	10	2,9	2	370	30	162	70
1084	1	2	0	8	3,4	2	400	28	163	58
1085	1	2	3	6	2,6	1	270	15	168	56
1096	1	2	2	2	3,6	1	320	17	174	68
1097	3	1	3	2	4,6	2	410	16	176	69

El cuadro de diálogo que se presenta en la figura No. 3. ESTADÍSTICA DESCRIPTIVA será tecleado con la siguiente información:

- ❑ La casilla RANGO DE ENTRADA, corresponde a la variable NÚMERO DE HERMANOS, es decir, seleccionamos una muestra de rango D2:D406. Debemos tener en cuenta que esta selección incluye el nombre de la fila.
- ❑ Nuestra información esta agrupada por Columnas, en este caso activamos el campo correspondiente.

Figura No. 4. Estadística Descriptiva (Pantallazo Adicional)



- ❑ Hacemos CLIC en el campo RÓTULOS EN LA PRIMERA FILA, con el fin de que los resultados salgan clasificados.
- ❑ En la categoría OPCIONES DE SALIDA, podemos establecer dónde queremos que nos muestre el cuadro de resultados. En nuestro ejemplo, activamos la opción RANGO DE SALIDA, y seleccionamos la celda M3.
- ❑ Activamos las opciones RESUMEN DE ESTADÍSTICAS; NIVEL DE CONFIANZA PARA LA MEDIA, que para nuestro ejemplo, lo consideramos en un 95%, así que tecleamos este valor.
- ❑ También seleccionamos las opciones de K-ÉSIMO MAYOR Y K-ÉSIMO MENOR, haciendo CLIC en estos campos. En nuestro caso, tecleamos el número 1 en cada campo, esto es, en caso de que se presente un número mayor de resultados para el análisis.
- ❑ Finalmente hacemos CLIC en ACEPTAR. Al hacer CLIC en ACEPTAR, se obtiene la información, tal como puede observarse en la figura No. 4, que será profundizada mediante la aplicación de las fórmulas correspondientes en los capítulos 3 y 4 de este libro.
- ❑ Los resultados de la figura No. 5, nos muestran un cuadro resumen con los valores de la Media, Error Típico, Mediana, Asimetría, Mínimo, Máximo, Suma, así como el Conteo para la variable NÚMERO DE HERMANOS.

Figura No. 5 Cuadro de Resultados

No. Hermanos	
Media	2,68069307
Error típico	0,12714053
Mediana	2
Moda	0
Desviación estándar	2,55549299
Varianza de la muestra	6,53054443
Curtosis	0,58527109
Coeficiente de asimetría	1,04257772
Rango	12
Mínimo	0
Máximo	12
Suma	1083
Cuenta	404
Mayor (1)	12
Menor(1)	0
Nivel de confianza(95,0%)	0,24994148

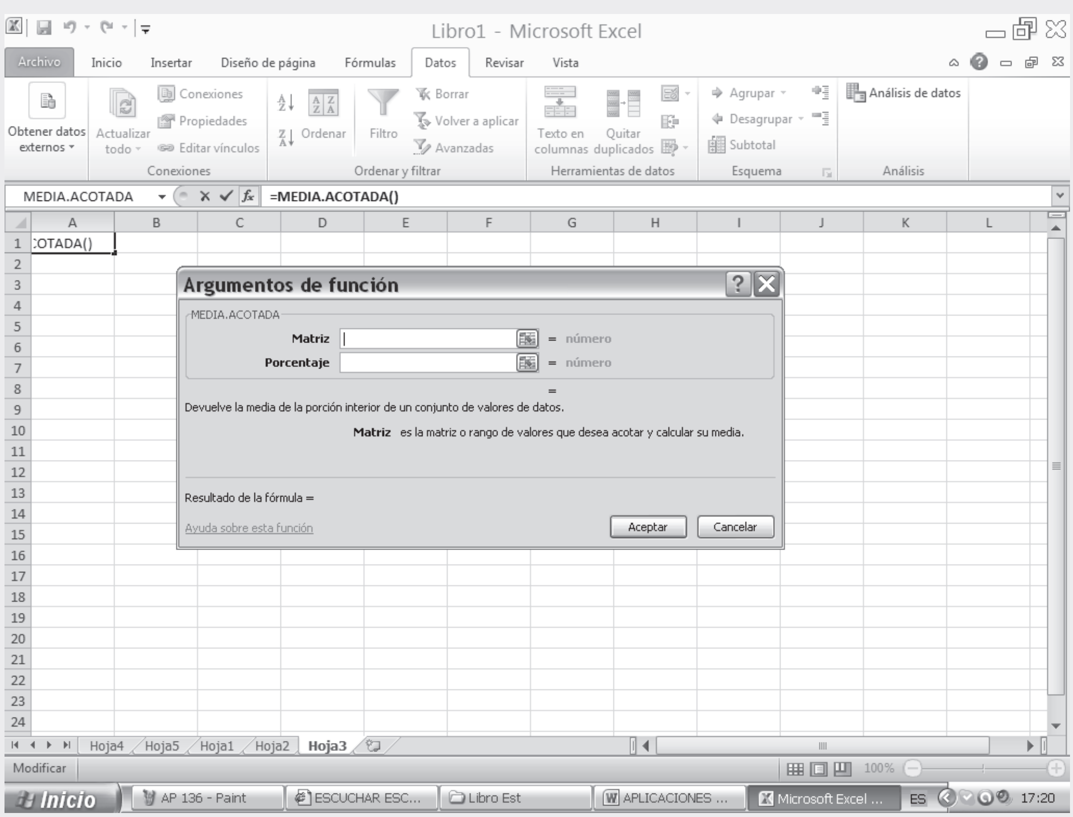
- Nota: Para lograr los resultados en todas y cada una de las opciones (Resumen de estadísticas; nivel de confianza para la media, K-ésimo mayor y K-ésimo menor), deben activarse.

Se podrá calcular en forma independiente, cada una de las medidas denominadas de TENDENCIA CENTRAL y de DISPERSIÓN, para ello sólo se va a tomar una de las variables, en este caso será el NÚMERO DE HERMANOS, de acuerdo con la información de la Tabla No. 1. El cálculo de estas medidas se presenta a continuación:

- Seleccionamos la CELDA, es decir, la activamos en el lugar donde deseamos que aparezca el resultado.

- Con el MOUSE nos movemos al ÍCONO fx, y hacemos CLIC dos veces. De este modo aparecerá la figura correspondiente a INSERTAR FUNCIÓN.

Figura No. 5. Insertar Función

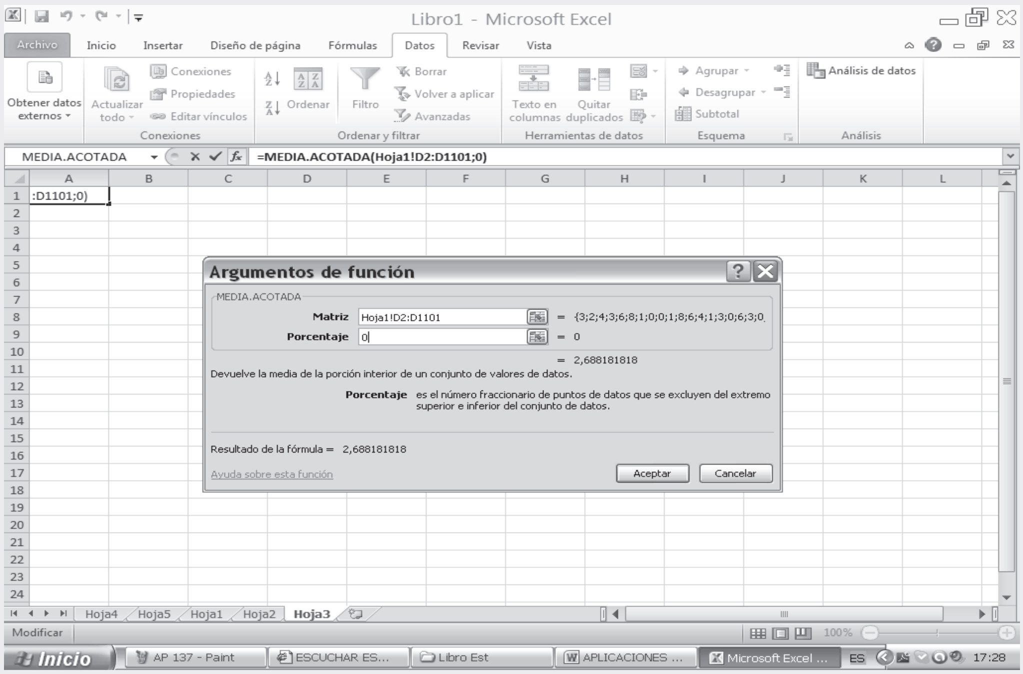


Cuadro No. 1. Resultados para cada una de las variables cuando los datos están sin agrupar

Medidas	No. Hermanos	No. Libros leídos	Promedio calificación matemáticas	Actualmente trabaja	Calificaciones ICFES	Edad/años	Estatura/Cm	Peso/Kg
Media	2,687898089	4,66515014	3,489626934	1,651501365	316,5696087	20,3963636	166,9363636	67,1172727
Error típico	0,077286891	0,08489441	0,019766201	0,014379939	1,334706027	0,12466909	0,222495241	0,27102907
Mediana	2	4	3,6	2	310	20	168	66
Moda	0	2	3,6	2	310	20	168	60
Desviación estándar	2,562150779	2,81434901	0,655272679	0,476711778	44,24706493	4,13480593	7,379332318	8,98901725
Varianza de la muestra	6,564616616	7,92056034	0,429382283	0,227254119	1957,802755	17,0966201	54,45454545	80,8024311
Curtosis	0,67110538	-0,30739702	0,063374502	-1,597435984	-0,223624021	0,48917427	1,349355724	-0,02761855
Coeficiente de asimetría	1,066995481	0,75017149	-0,27308826	-0,636768724	0,653420663	1,04230999	-0,484056991	0,3668056
Rango	12	12	3,4	1	180	18	50	42
Mínimo	0	0	1,6	1	240	15	140	46
Máximo	12	12	5	2	420	33	190	88
Suma	2954	5127	3835,1	1815	347910	22436	183630	73829
Cuenta	1099	1099	1099	1099	1099	1100	1100	1100
Mayor (1)	12	12	5	2	420	33	190	88
Menor(1)	0	0	1,6	1	240	15	140	46
Nivel de confianza(95,0%)	0,151646685	0,16657361	0,038783795	0,028215264	2,618862553	0,24461632	0,436563451	0,53179288

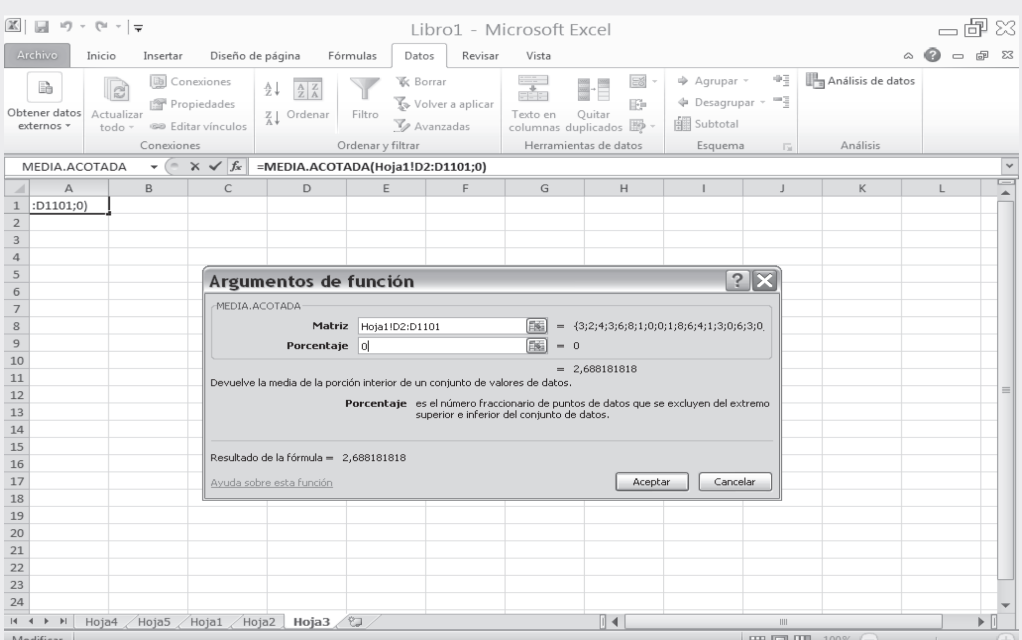
Nota. Observe cuidadosamente los errores que se pueden cometer si consideramos aquellas características que son cualitativas, tales como FACULTAD, SEXO, TRABAJA. En estos casos el resultado no nos dice absolutamente nada, pues las medidas a aplicar, entre otras, sería la proporción, para luego presentarlo como porcentajes.

Figura No.6. Media



- En la figura No. 6, se nos pide en la primera celda la MATRIZ. Se debe ir a la Tabla No. 1 y sombrear todos los datos sin soltar el botón izquierdo.
- En la casilla que dice PORCENTAJE se le da el valor de cero (0) y se observará que el resultado de la MEDIA es de 2.68, siendo el mismo resultado que se obtuvo por el procedimiento anterior, tal como lo muestra la figura No. 7.

Figura No. 7 Cálculo de la media (número de hermanos)



Nota: el procedimiento anterior es aplicado igualmente para las diferentes medidas de POSICIÓN y de DISPERSIÓN, sólo cambia cuando nos ubicamos en el nombre de la función, es decir, la medida que se requiere calcular.

En las siguientes páginas se expondrá de forma general la TEORÍA de ciertas medidas de posición y de Tendencia Central, tales como LA MEDIANA, MODA Y MEDIA GEOMÉTRICA, especificando en cada caso a través de un ejemplo el procedimiento a seguir para obtener los resultados correspondientes en la hoja de cálculo. De forma complementaria, al final del capítulo, se encuentra la unidad de ejercicios para resolver a fin de profundizar sobre estos temas.

MEDIANA- M_E

En datos sin agrupar hay dos maneras de calcularla:

a) Número de Observaciones IMPAR. En este caso se ordenan los datos de menor a mayor o viceversa, luego seleccionamos como MEDIANA al valor central.

Ejemplo: 3 5 $\boxed{12}$ 16 30

$$\underset{\text{Me}}{\downarrow}$$

b) Número PAR de observaciones. También se ordenan de mayor a menor o de menor a mayor; sólo que en este caso se tienen dos valores en el centro, así que debemos promediar.

Ejemplo: 3 5 $\boxed{12 \quad 16}$ 30 46

Me = (12 + 16) / 2 = 14

Figura No. 8. Mediana (Pantallazo Adicional)

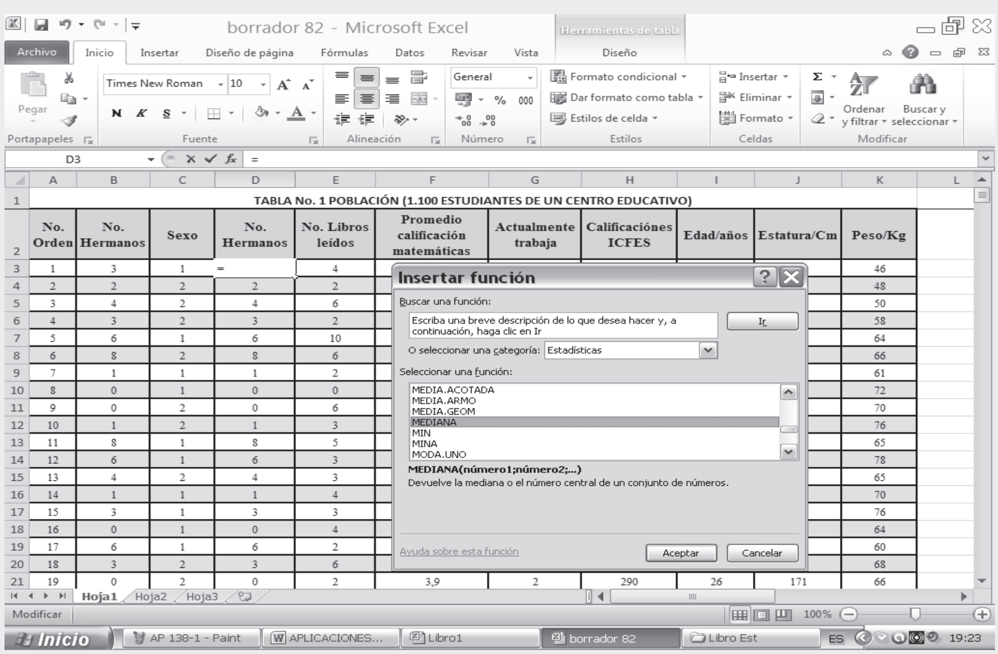
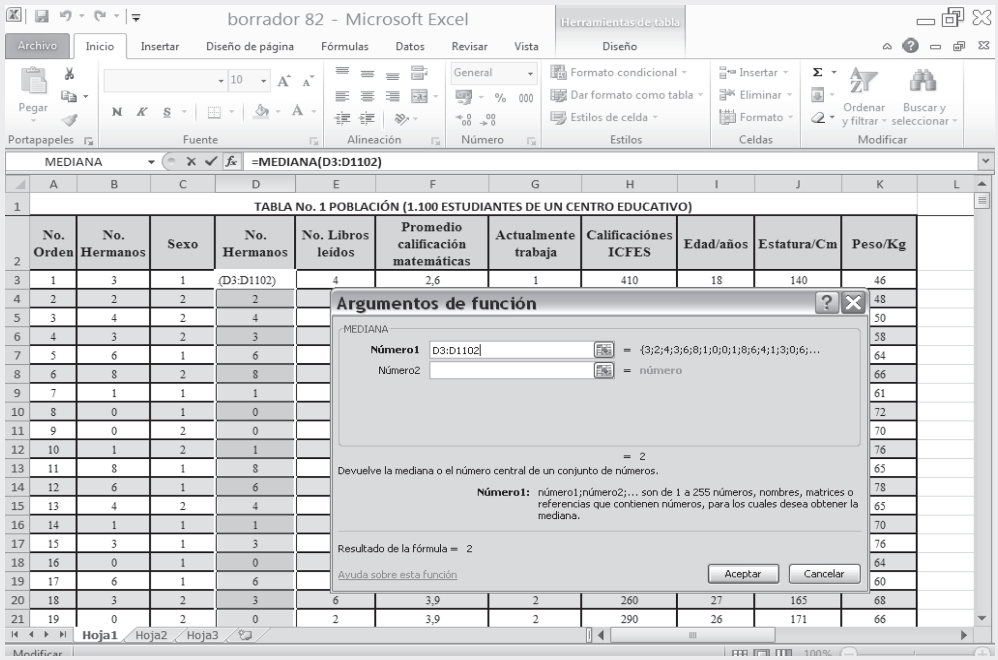


Figura No. 8. Mediana (Pantallazo Adicional)



- Observe, que para el cálculo de la MODA y demás MEDIDAS, aparece en las primeras casillas la palabra NÚMERO 1, es en ésta donde se debe copiar el rango de los valores que aparecen en la Tabla No. 1 correspondiente a la columna NÚMERO DE HERMANOS. Luego de establecer este rango es posible observar, que al compararlo con el obtenido en el anterior método, la MEDIANA sigue siendo igual a 2.0.

Nota: Este resultado se localiza en la celda que se escoja y así sucesivamente. No obstante lo anterior y para efectos de profundizar sobre estos temas, le invitamos a realizar el mismo procedimiento para las demás columnas de la Tabla No. 1.

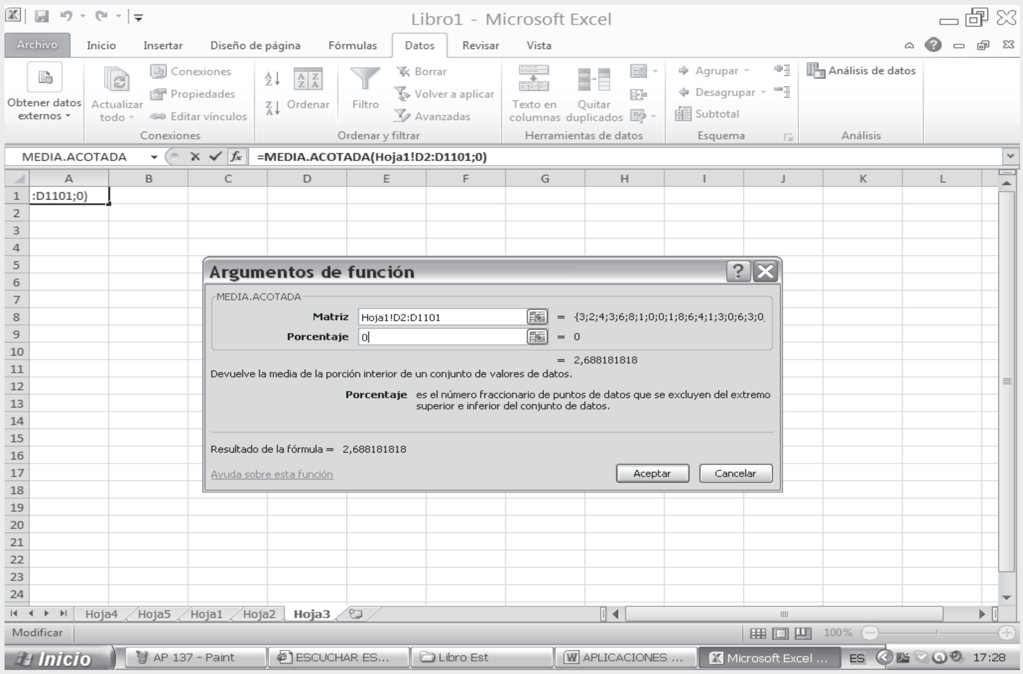
MODO = MODA = VALOR MODAL: M_D

Se define como aquel valor de la variable que más se repite. Para algunos no es aconsejable calcularla mediante la utilización de Excel en caso de que haya más de una MODA, es decir, que sea BIMODAL o PLURIMODAL. Sólo se debe aplicar si es UNIMODAL, es decir, cuando hay una sola moda. En caso de ser BIMODAL O PLURIMODAL, sólo se reconoce la primera MODA de la lista.

Los pasos son los mismos utilizados para el cálculo de la MEDIA y la MEDIANA, esto es: (Figura No. 9).

Nota: El resultado de una MODA = 0 nos indica que cero es el valor que más se repite en la muestra. Obsérvese que el resultado anterior, es el mismo al obtenido en el cuadro No. 5.

Figura No.9. Moda

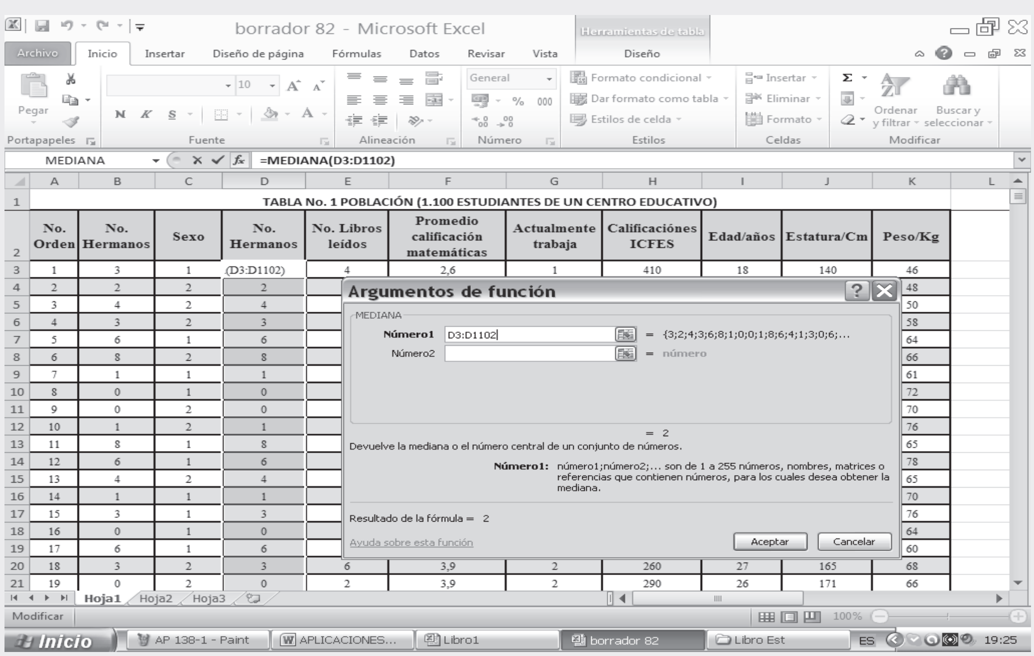


MEDIA GEOMÉTRICA : $M_G = M_O$

Se define como la "raíz enésima del producto de los valores de la variable.

Fórmula $M_o = \sqrt[n]{X_1.X_2................ X_n}$ $\qquad M_o = \sqrt[n]{\overset{n}{\underset{i=1}{\pi}} x_i}$

Figura No. 10. Estimación Moda (número de hermanos)



- ❑ Nuevamente se siguen los mismos pasos aplicados para calcular la MEDIA, MEDIANA Y MODA. Hacemos CLIC en *fx* y seleccionamos CATEGORÍA ESTADÍSTICA, luego, a la derecha elegimos la función MEDIA GEOMÉTRICA, finalmente hacemos CLIC en ACEPTAR.

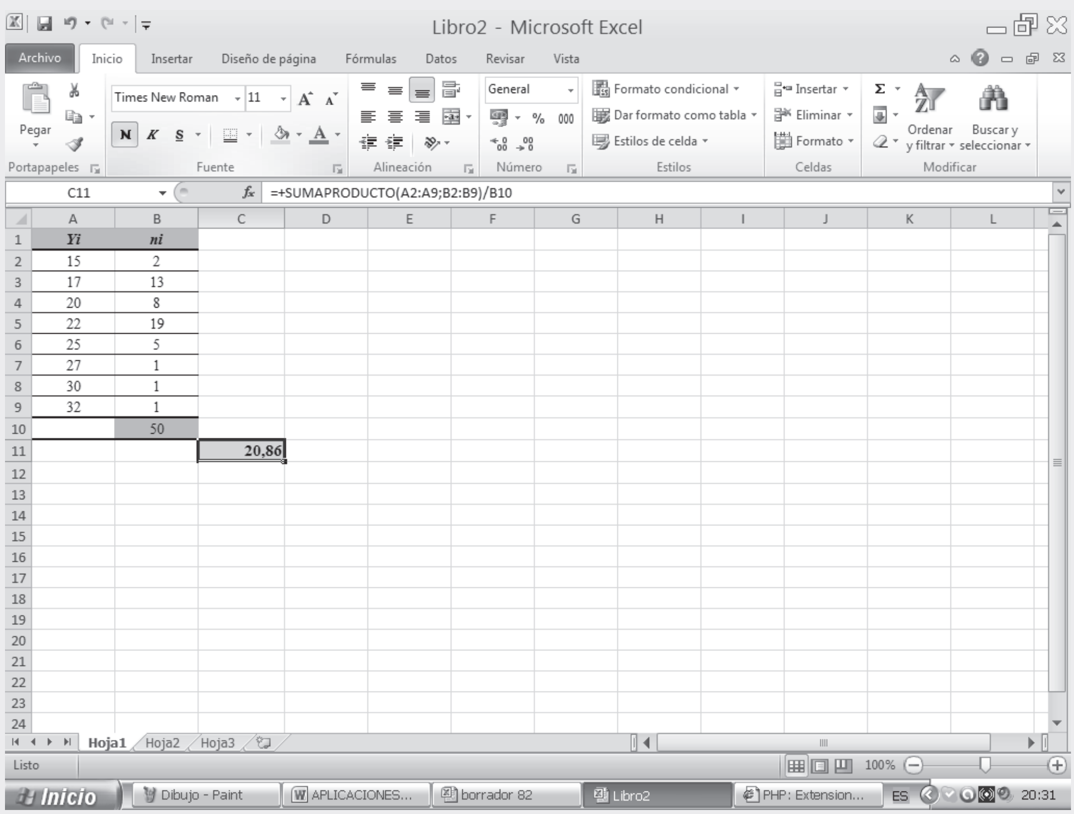
MEDIA ARITMÉTICA PONDERADA

Para utilizar Excel en el cálculo de la MEDIA PONDERADA, se realiza el siguiente procedimiento:

- ❑ Tomemos ahora la tabla que aparece a continuación, correspondiente a la variable (y_i): Edad (años); y su frecuencia absoluta, siendo (n_i) el número de veces que se repite cada valor de la variable. Para ello vamos a tener presente una tabla de datos agrupados, aplicando la fórmula que deberá ser digitada:

=SUMA PRODUCTO (A2:A9;B2:B9)/SUMA(B2:B9)

Figura No. 11. Media Aritmética Ponderada



$$\bar{y} = \frac{\sum y_i n_i}{n} = 20,86$$

4
CAPÍTULO

Medidas de dispersión, de deformación y apuntamiento

Estudia no para saber algo más sino para saberlo mejor. Ya que el saber más que los otros es fácil. Lo difícil es saber algo mejor que los otros

a

CONTENIDO

Aplicaciones prácticas de las medidas de Dispersión o de deformación, Asimetría y de Apuntamiento.

- Definiciones, fórmulas, propiedades y aplicaciones.
- Medidas de dispersión: Varianza, Desviación típica, Coeficiente de variación, Puntaje típico, Rango o recorrido.
- Medidas de asimetría.
- Medidas de apuntamiento.
- Síntesis de la unidad.
- Ejercicios para Resolver, resueltos en el Sistema de Información en Línea SIL.

COMPETENCIAS

El estudiante deberá estar en capacidad de:

- ❑ Seleccionar, comprender y aplicar cada una de las medidas, en casos que lo ameriten.
- ❑ Manejar los conceptos, fórmulas, usos e importancia que cada una de estas medidas lo exige, no sólo en los talleres, sino en casos reales.
- ❑ Comprender sus aplicaciones y agilizar sus cálculos.
- ❑ Entender los conceptos, las ventajas y desventajas en su aplicación.

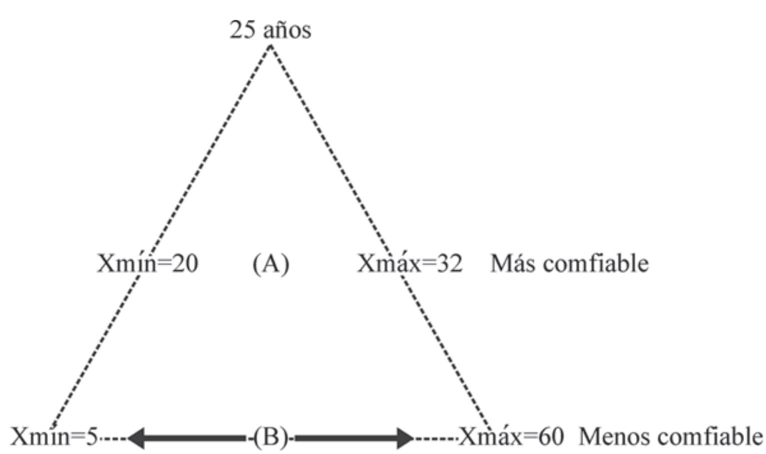
CONCEPTOS GENERALES

Estos primeros cuatro capítulos han sido dedicados a la *estadística descriptiva*. Comenzamos con el manejo de términos y conceptos, luego, el uso de la simbología para hablar un mismo lenguaje, lo cual nos permitió la elaboración de cuadros y gráficas; luego la aplicación de *promedios* y ahora, el cálculo de las *medidas de dispersión*, que será el tema de este capítulo.

Al calcular un promedio, por ejemplo la media aritmética, no sabemos hasta donde se dá la representatividad para ese conjunto de datos, sin embargo es posible determinar si hay concentración de datos alrededor del promedio, el cual nos indicaría una buena aplicación, por el contrario una gran dispersión estaría indicando poca representatividad, por lo tanto no sería confiable o adecuada para el conjunto de datos. Aún más, las medidas de dispersión son indicadas cuando queremos evaluar dos o más promedios. Para este capítulo el estudiante tiene la necesidad de repasar los contenidos del capítulo 3.

MEDIDAS DE DISPERSIÓN

Supongamos que dos distribuciones de edades tienen igual media, pero diferentes grados de dispersión, por lo tanto hace que una de ellas sea más representativa o más confiable.



Observe que 32 - 20 = 12, y 60 - 5 = 55, por lo tanto 12 < 55 permite decir, que hay menos variabilidad en el primer caso; lo cual hace que el promedio de 25 años de edad sea mucho más representativo.

VARIANZA : S^2 ó σ^2

Las medidas de dispersión más conocidas y utilizadas son la *varianza y la desviación típica o estándar*. Esta última, es la raíz cuadrada de aquélla.

La *varianza* se define como: *la media aritmética de los cuadrados de las diferencias (desviaciones) entre los valores que toma la variable y su media aritmética*. Su símbolo es *S*2 en la muestra, σ2 (sigma al cuadrado) en la población.

ato a a

1) $S^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n}$	(muestra);	$\sigma^2 = \dfrac{\Sigma(X_i - \ )^2}{N}$
2) $S^2 = \dfrac{\Sigma x_i^2 - n\bar{x}^2}{n}$		$\sigma^2 = \dfrac{\Sigma X_i^2 - N\ ^2}{N}$
3) $S^2 = \dfrac{\Sigma x_i^2}{n} - \bar{x}^2$		$\sigma^2 = \dfrac{\Sigma X_i^2}{N} - \ ^2$

Veamos con un ejercicio cómo se aplican las fórmulas anteriormente presentadas, obteniéndose un resultado igual a la aplicación en cada uno de ellos.

Ejemplo 1. Con los siguientes datos (sin agrupar) calcule la varianza por los diferentes métodos

$x_1 = 5 \qquad x_2 = 3 \qquad x_3 = 1 \qquad x_4 = 6 \qquad x_5 = 10$

1) $\boxed{S^2 = \dfrac{\Sigma(x_i - \bar{x})^2}{n}}$ $\qquad S^2 = \dfrac{\Sigma d_i^2}{n}$

$$\bar{x} = \frac{5+3+1+6+10}{5} = \frac{25}{5} = 5$$

$$S^2 = \frac{(5-5)^2 + (3-5)^2 + (1-5)^2 + (6-5)^2 + (10-5)^2}{5}$$

$$S^2 = \frac{0+4+16+1+25}{5} = \frac{46}{5} = 9{,}2$$

También se podrá calcular asi:

2) $\boxed{S^2 = \dfrac{\Sigma x_i^2 - n\bar{x}^2}{n}}$ $\qquad$ 3) $\boxed{S^2 = \dfrac{\Sigma x_i^2}{n} - \bar{x}^2}$

$$\Sigma x_i^2 = 5^2 + 3^2 + 1^2 + 6^2 + 10^2 = 25 + 9 + 1 + 36 + 100 = 171$$

$$S^2 = \frac{171 - 5(5)^2}{5} = \frac{171 - 125}{5} = \frac{46}{5} = 9{,}2$$

$$S^2 = \frac{171}{5} - 5^2 = 34{,}2 - 25 = 9{,}2$$

Procedimiento de cálculo

- ❑ Se calcula la media aritmética, por cualquiera de los procedimientos que han visto.
- ❑ En la primera fórmula, se establecen las diferencias entre los valores de la variable con respecto a su media y luego se elevan al cuadrado: $(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2$ y así sucesivamente.
- ❑ Para la segunda y tercera fórmula se calcula la Σx_i^2, se debe tener presente que corresponde a la suma de $x_1^2 + x_2^2 + x_3^2$ y así sucesivamente.

ato a a o

Las fórmulas son casi iguales a las anteriores, salvo que los valores están ponderados, es decir, multiplicados por las frecuencias absolutas o las relativas.

1) $s^2 = \dfrac{\Sigma(y_i - \bar{y})^2\, n_i}{n}$ $\qquad$ $S^2 = \dfrac{\Sigma(X_i - \bar{X})f_i}{n}$

2) $s^2 = \dfrac{\Sigma\, y_i^2\, n_i - n\bar{y}^2}{n}$ $\qquad$ $S^2 = \dfrac{\Sigma X_i^2 f_i - n\bar{X}^2}{n}$

3) $s^2 = \dfrac{\Sigma\, y_i^2\, n_i}{n} - \bar{y}^2$ $\qquad$ $S^2 = \dfrac{\Sigma X_i^2 f_i}{n} - \bar{X}^2$

Ejemplo 2. Calculemos la varianza para la siguiente distribución, aplicando las tres fórmulas anteriores. Siendo el procedimiento igual tanto al trabajar con variable discreta o continua.

y_i	n_i	$y_i\, n_i$	$y_i - \bar{y}$	$(y_i - \bar{y})n_i$	$(y_i - \bar{y})^2 n_i$	y_i^2	$y_i^2\, n_i$
5	2	10	−9,75	−19,50	190,1250	25	50
10	4	40	−4,75	−19,00	90,2500	100	400
15	8	120	0,25	2,00	0,5000	225	1.800
20	5	100	5,25	26,25	137,8125	400	2.000
25	1	25	10,25	10,25	105,0625	625	625
Σ	20	295	—	0	523,7500	—	4.875
X_i	f_i	$X_i\, f_i$	d_i	$d_i\, f_i$	$d_i^2\, f_i$	X_i^2	$X_i^2\, f_i$

$\bar{y} = \dfrac{\Sigma\, y_i\, n_i}{n}$ $\qquad$ $\bar{y} = \dfrac{295}{20} = 14,75$ $\qquad$ $\bar{X} = \dfrac{\Sigma X_i f_i}{n}$

1) $s^2 = \dfrac{\Sigma(y_i - \bar{y})^2\, n_i}{n} = \dfrac{523,75}{2} = 26,1875$ $\qquad$ $S^2 = \dfrac{\Sigma d_i^2\, f_i}{n}$

2) $s^2 = \dfrac{\Sigma\, y_i^2\, n_i - n\,\bar{y}^2}{n} = \dfrac{4.875 - 20\ (14,75)^2}{20} = 26,1875$ $\qquad$ $S^2 = \dfrac{\Sigma X_i^2\, f_i - n\bar{X}^2}{n}$

3) $s^2 = \dfrac{\Sigma\, y_i^2\, n_i}{n} - \bar{y}^2 = \dfrac{4.875}{20} - 14,75^2 = 26,1875$ $\qquad$ $S^2 = \dfrac{\Sigma X_i^2\, f_i}{n} - \bar{X}^2$

Hay dos fórmulas más, conocidas como *métodos abreviados*, utilizando las desviaciones con respecto a un valor arbitrario (cualquier valor que usted escoja) que no serán presentados y por lo tanto desarrollados.

Se pudo observar en el ejercicio anterior que en el cálculo de la *varianza*, utilizando las diferentes fórmulas, producen el mismo resultado cuando son aplicadas en un mismo conjunto de observaciones.

Propiedades de la varianza

- ***a:*** La varianza debe ser siempre un valor positivo. $S^2 > 0$.
- ***a***: La varianza de una constante es igual a 0.

$$V_{[\ ]} = M_{[\ -M(K)]}{}^2 = M_{[\ -\ ]}{}^2 = 0$$

- ***a:*** La varianza de una constante más una variable, es igual a la varianza de la variable.

$$V_{[y+\ ]} = V_{[y]} + V_{[\ ]} = V_{[y]} \qquad V_{[x+\ ]} = V_{[x]} + V_{[\ ]} = V_{[x]}$$

Válido también para la diferencia: $V_{[x-\ ]} = V_{[x]} - V_{[\ ]} = V_{[x]}$

- ***a ta:*** La varianza de una constante por una variable, es igual al producto de la constante al cuadrado por la varianza de la variable.

$$V_{[\ y]} = K^2\, V_{[y]} = K^2\, S_y^2$$

- ***ta:*** Para el cálculo de la varianza en una muestra, cuando se está trabajando con submuestras, se aplica la siguiente fórmula:

$$S^2 = \frac{s_1^2 n_1 + s_2^2 n_2 + \ldots}{n_1 + n_2 + \ldots} + \frac{(\bar{x}_1 - \bar{x})^2\, n_1 + (\bar{x}_2 - \bar{x})^2\, n_2 + \ldots}{n_1 + n_2 + \ldots}$$

sabiendo que: $\bar{x} = \dfrac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \ldots}{n_1 + n_2 + \ldots}$

Ejercicios de aplicación de las propiedades

Ejemplo 4. *Primera propiedad*: supongamos que se tienen dos distribuciones A y B y se desea aplicar la fórmula de la varianza en ambos casos.

a)

x_i :	8	2	10	6	4	$\bar{x} = 6$

b)

y_i :	0	1	2	3	4	$\bar{y} = 1{,}65$
n_i :	4	6	5	3	2	= 20

Las varianzas serán:

a) $S^2 = \dfrac{\Sigma x_i^2 - n\,\bar{x}^2}{n} = \dfrac{220 - 5(6)^2}{5} = 8$ $\qquad S^2 = \dfrac{\Sigma x^2 - n\,\bar{x}^2}{n}$

b) $S^2 = \dfrac{\Sigma\, y_i^2 n_i - n\,\bar{y}^2}{n} = \dfrac{85 - 20(1{,}65)^2}{20} = 8 = 1{,}527 \cong 1{,}53$

Observemos que en ambos casos se cumple la *primera propiedad, la varianza de una variable siempre será mayor a cero* ($S^2 > 0$). Por lo tanto siempre será positiva.

Ejemplo 5. *La segunda propiedad*, no requiere que se haga un ejercicio de aplicación, pues si consideramos que en un curso el profesor califica un trabajo o taller con 4 para todos los alumnos, el promedio será de cuatro y la varianza será igual a cero, pues ninguna nota será superior o inferior a cuatro es decir, no hay variación.

Ejemplo 6. *Tercera propiedad:* consideremos que las dos variables del ejemplo 4 están dadas en la misma unidad de medida, por ejemplo en kilogramos, y se requiere que el resultado se dé en libras (aproximadamente 2 libras por kg); para ello se tendrá que:

x_i :	16	4	20	12	8

$\bar{x} = 12$ libras y $S^2 = 32$ libras^2

$$S^2 = \frac{880 - 5(12)^2}{5} = 32 \text{ libras}^2 \qquad V_{[\ \ +y]} = \ {}^2 S^2 = 2^2(8) = 32 \text{ libras}^2$$

Algo similar ocurre al trabajar con datos agrupados. Observe que el resultado obtenido, siempre será expresado en las mismas unidades, *pero al cuadrado*.

Con los datos del ejemplo 4–b se harán los calculos

y_i	n_i	$y_i n_i$	$y_i^2 n_i$
0	4	0	0
2	6	12	24
4	5	20	80
6	3	18	108
8	2	16	128
Σ	20	66	340
X_i	f_i	$X_i f_i$	$X_i^2 f_i$

$$S^2 = \frac{\Sigma y_i^2 n_i - n\bar{y}^2}{n}$$

$$S^2 = \frac{340 - 20(3{,}3)^2}{20} = 6{,}11$$

$$\bar{y} = \frac{66}{20} = 3{,}3 \qquad \bar{X} = \frac{\Sigma x_i f_i}{n}$$

$$V_{[KX]} = K^2 S^2 = 2^2(1{,}527) = 6{,}11$$

Ejemplo 7. *Cuarta propiedad:* si procedemos a sumarle una determinada cantidad a cada valor observado, por ejemplo 5, se tendrá que x_i: 13 7 15 11 9 (ver ejemplo 4–a)

La media será: $\bar{x} = 11$ y la varianza = 8

$$V_{[X+K]} = S^2 + V_{[K]} = 8 + 0 = 8$$

Se podrá comprobar aplicando la fórmula: $S^2 = \dfrac{645 - 5(11)^2}{5} = 8 \qquad V_{[K]} = 0$

En datos agrupados al aplicar la *cuarta propiedad*, se tendrá que:

y_i	n_i
0	4
1	6
2	5
3	3
4	2
Σ	20

y_i	n_i	$y_i n_i$	$y_i^2 n_i$
5	4	20	100
6	6	36	216
7	5	35	245
8	3	24	192
9	2	18	162
Σ	20	133	915
X_i	f_i	$X_i f_i$	$X_i^2 f_i$

$$\bar{x} = \bar{y} = \frac{133}{20} = 6{,}65$$

$$M_{[y+K]} = 1{,}65 + 5 = 6{,}65$$

$$S^2 = \frac{915 - 20(6{,}65)^2}{20} = 1{,}527 \cong 1{,}53$$

Ejemplo 8. *Quinta propiedad*: finalmente consideramos que las dos variables o las dos distribuciones anteriores (a y b) corresponden a dos submuestras, siendo los resultados para cada submuestras los siguientes:

$$n_1 = 5 \qquad n_2 = 20 \qquad \bar{x}_1 = 6 \qquad \bar{x}_2 = 1{,}65 \qquad S_1^2 = 8 \qquad S_2^2 = 1{,}53$$

Se quiere calcular la media y la varianza para el conjunto de las 25 observaciones, aplicando la *quinta propiedad*.

$$\bar{x} = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2}{n} \qquad \bar{x} = \frac{6(5) + 1{,}65(20)}{25} = 2{,}52$$

$$\boxed{S^2 = \frac{S_1^2 n_1 + S_2^2 n_2}{n} + \frac{(\bar{x}_1 - \bar{x})^2 n_1 + (\bar{x}_2 - \bar{x})^2 n_2}{n}}$$

$$S^2 = \frac{(5)8 + 1{,}53(20)}{25} + \frac{(6 - 2{,}52)^2\,5 \ + (1{,}65 - 2{,}52)^2\,20}{25}$$

$$S^2 = 2{,}824 + \frac{60{,}552 + 15{,}138}{25} = 2{,}824 + 3{,}027 = 5{,}851$$

DESVIACIÓN TÍPICA : S y σ

La *desviación típica* es la raíz cuadrada de la varianza, considerada siempre con signo positivo. Es la medida de dispersión más extensamente aplicada.

$$\boxed{S = +\sqrt{S^2}} \qquad S = +\sqrt{26{,}19} \ = 5{,}12$$

La desviación típica o estándar será la medida que más se va a utilizar de aquí en adelante, no sólo en este capítulo, sino en todos los siguientes de este libro.

> La varianza se expresa siempre en unidades diferentes a las originales, es decir, si la variable se refiere al peso en kg, al calcular la varianza el resultado obtenido estará dado en kg al cuadrado, sucediendo exactamente igual con cualquiera de las unidades de medidas que utilicemos. Esta es la razón por la cual siempre estamos utilizando *la desviación típica o desviación estándar*, pues el resultado se expresa en las mismas unidades de la variable.

Supongamos que se tiene una población normalmente distribuida, con media:

$$\mu = 30 \qquad \sigma^2 = 15 \qquad \text{y} \qquad \sigma = 3{,}87$$

De acuerdo con la desigualdad de Tchebycheff, puede deducirse para cualquier distribución normal o simétrica que la media aritmética es más o menos....

Si se toma una sola vez la desviación típica al lado y lado de la media, el intervalo obtenido, incluye el 68,3% de las observaciones.

$$\mu \pm 1\sigma \quad \Rightarrow \quad 30 \pm 1(3{,}87) \begin{matrix} \nearrow \\ \searrow \end{matrix} \left.\begin{matrix} 33{,}87 \\ 26{,}13 \end{matrix}\right\} \quad 68{,}3$$

$$\mu \pm 1{,}5\sigma \quad \Rightarrow \quad 30 \pm 1{,}5(3{,}87) \begin{matrix} \nearrow \\ \searrow \end{matrix} \left.\begin{matrix} 35{,}805 \\ 24{,}1\ \ 5 \end{matrix}\right\} \quad 86{,}6$$

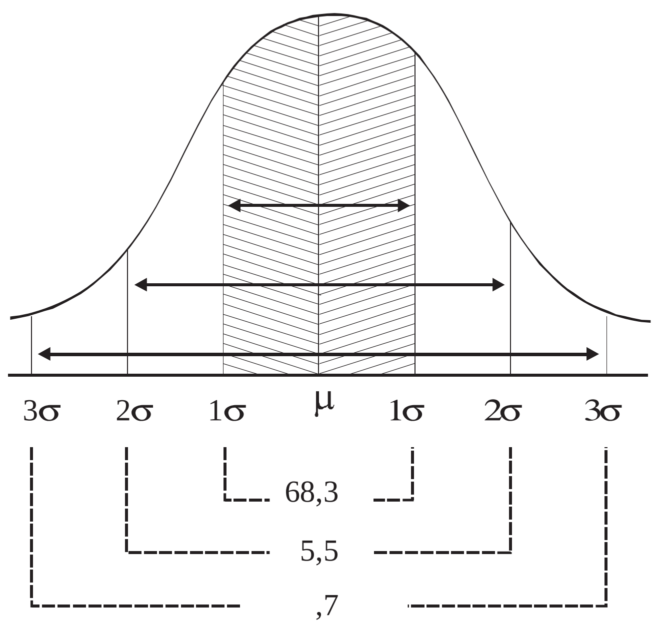
$$\mu \pm 2\sigma \quad \Rightarrow \quad 30 \pm 2(3{,}87) \begin{matrix} \nearrow \\ \searrow \end{matrix} \left.\begin{matrix} 37{,}74 \\ 22{,}26 \end{matrix}\right\} \quad 5{,}5$$

$$\mu \pm 3\sigma \quad \Rightarrow \quad 30 \pm 3(3{,}87) \begin{matrix} \nearrow \\ \searrow \end{matrix} \left.\begin{matrix} 41{,}46 \\ 18{,}3 \end{matrix}\right\} \quad ,7$$

En los próximos capítulos 7 y 8 se ampliarán estos conceptos de *limites de confianza* y, cual es su aplicación y significado.

Gráficamente para una distribución normal, teniendo en cuenta que la *media aritmética*, la *mediana* y el *modo* son iguales, corresponderán al valor central de la distribución

$$= M_e = M_d$$



APLICACIONES DEL EXCEL

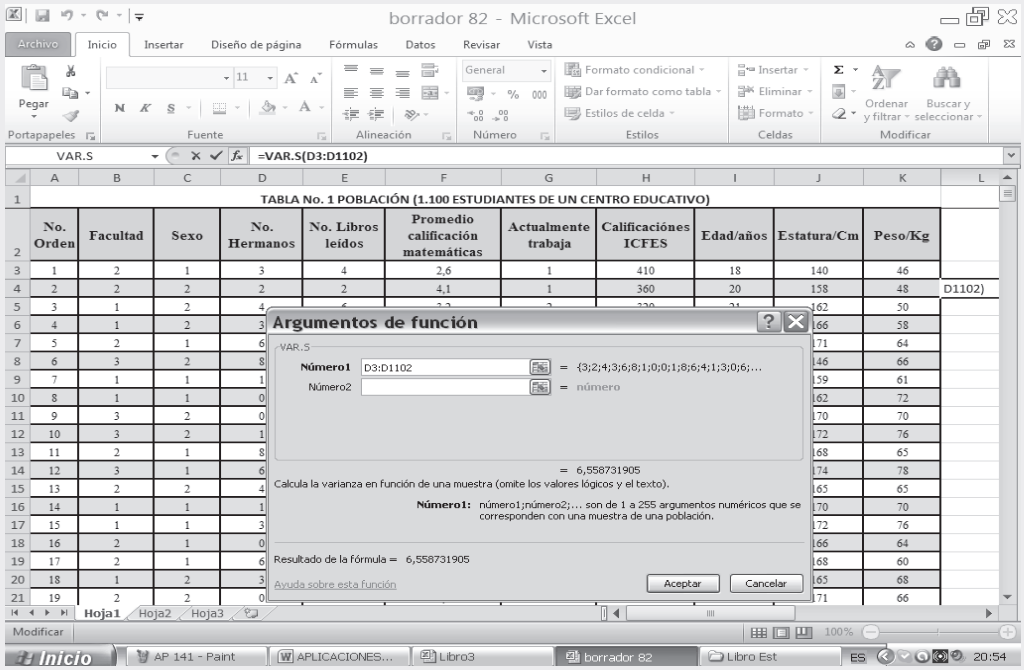
Varianza (aplicada a la muestra) $S^2 = \sigma^2$

El procedimiento es exactamente igual a los aplicados para calcular: Media, *Mediana, Moda y Media Geométrica*. Seleccionamos el renglón correspondiente a la función VAR, de acuerdo a la figura No. 1. Luego establecemos el rango separado por dos puntos pertenecientes a la variable NÚMERO DE HERMANOS y hacemos CLIC en ACEPTAR

- ❑ Observemos que el valor de la varianza por este método es el mismo obtenido por el anterior método de cálculo.

  Nota: En Excel 2010 es posible encontrar nuevas funciones de varianza. Si los datos representan la población total se utiliza la función VARP y si por el contrario sólo se toma una muestra de la población total, se utiliza la función VARM. En este caso aplicamos VARM.

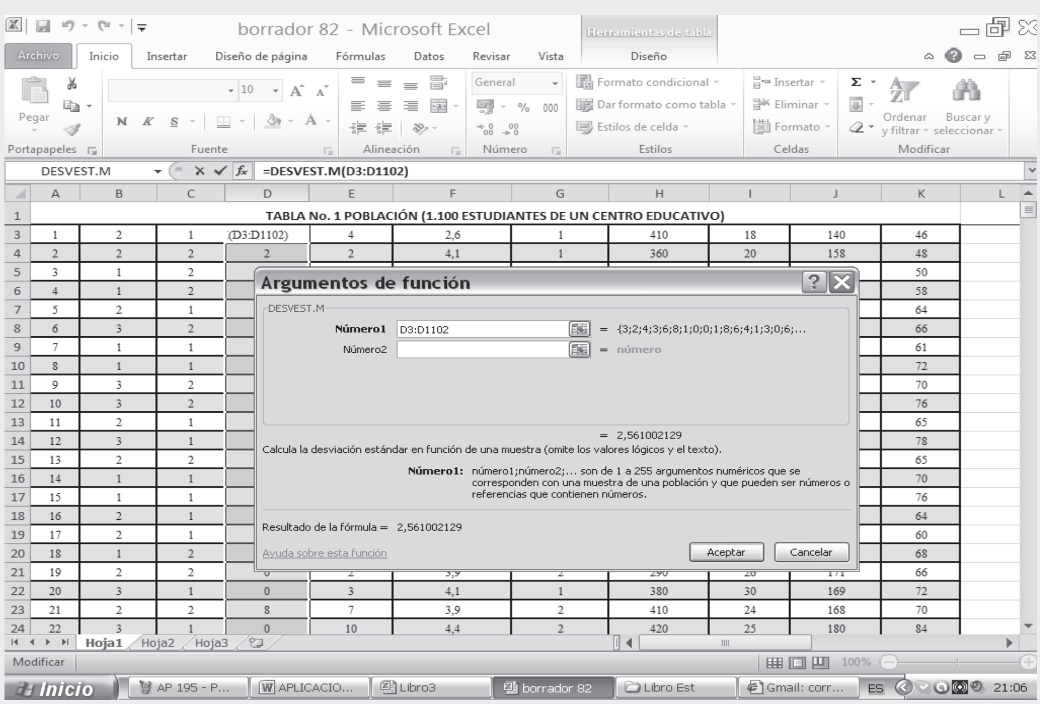
Figura No. 1. Argumentos para una Varianza



Desviación típica estándar $= \sigma$

Se define como la Raíz cuadrada de la Varianza: $S = \sqrt{S^2}$, siguiendo el mismo procedimiento que hemos venido utilizando, pero esta vez, buscamos la función DESVEST y de esta manera se obtienen los resultados siguientes:

Figura No. 2. Cálculo Desviación Típica



Al observar este resultado se ve claramente que es la raíz cuadrada de la varianza, es decir que

$$\mathbf{S = \sqrt{6{,}558} = 2{,}56}$$

Nota: Todas las medidas que aparecen en la columna NOMBRE DE LA FUNCIÓN podrán calcularse de la mism que las anteriores.

En Excel 2010 es posible encontrar nuevas funciones de desviación estándar. Si los datos representan la población total se utiliza la función DESVESTP y si por el contrario sólo se toma una muestra de la población total se utiliza la función DESVESTM. En este caso aplicamos DESVESTM.

USO DE LA CALCULADORA

El cálculo de la desviación típica por medio del uso de la calculadora, se sigue el mismo procedimiento que se indicó para el cálculo de ΣX_i ΣX_i^2 n y $\overline{X}$ (ver página 130). Se debe calcular la desviación típica para las poblaciones o en muestras ($n > 30$) ver la tecla 2, donde generalmente aparece indicado $_x\sigma_n$ que corresponde a desviación típica sin corregir.

En algunas, el valor de σ se obtiene oprimiendo SHIFT 2; en otras calculadoras será: SHIFT 2 OPCIONES y seleccione la opción deseada y ese resultado obtenido al ser elevado al cuadrado corresponde al valor de la varianza $_x\sigma_n^2$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

1. En una distribución unidimensional se sabe, que la media aritmética es igual a 3,04 y que la media cuadrática es igual a 4,26. Se pide calcular la varianza y la desviación típica de dicha distribución.

2. En la sección financiera de un diario se publicó recientemente la (distribución unidimensional de variable discreta) tabla adjunta y decía el texto del artículo, que la media aritmética era igual a 120 y la varianza igual a 25.

   Desgraciadamente la publicación apareció con dos manchas de tinta caídas en las columnas de las frecuencias absolutas, lo cual impidió comprobar directamente lo que en el artículo se decía. Sin embargo, se le pregunta ahora a usted, ¿pueden aceptarse dichos valores de la media y la varianza, teniendo en cuenta lo que puede verse en la tabla? Conteste si puede ser posible o no y porqué.

y_i :	105	110	115	120	125	130	135	140
n_i:		37	90	95	85	60		

3. Con la siguiente ecuación:
   $V_{[1/8X - 8]}$ sabiendo que $S_x = 8$. Calcular la varianza de y_i.

4. Con los siguientes datos:
   $y = 1/4\,(8x - 2)$ x = 4 $S_x = 2$ Se pide calcular: $V_{[Y]}$ y $M_{[X]}$

5. Una muestra de 70 datos da para una cierta variable una media de 120 y una desviación típica de 6, otra semejante, pero de 30 observaciones, da para la misma variable una media de 125 y una desviación típica de 5. Si se reúnen las dos muestras formando una sola de 100 datos, ¿cuál será su media y su desviación típica?

6. Durante un período de 3 años, los precios de un producto fueron en promedio de \$8.000, con una desviación típica de \$120. En el período anterior de 3 años, el promedio fue de \$5.000 con una varianza de \$3.600. ¿En qué período hubo mayor estabilidad?

7. La varianza de dos números es 1 y su media aritmética 8. Calcular esos dos números.

8. Si un conjunto de n valores de x_i se sabe que: $\sum x_i = 10$ $\quad \Sigma x_i^2 = 260$ $\quad S^2 = 25$. Se pide encontrar el valor de n.

9. La media de 10 observaciones es 3 y la suma de sus cuadrados es 100. Encontrar la desviación estándar del conjunto.

10. Sumando 4 a cada uno de los números de la serie 2 6 5 9 1, se obtiene la serie 6 10 9 13 5. Comprobar que ambas series tienen la misma varianza y distintas medias.

11. Hallar la desviación típica de dos números cuya media aritmética es 9,0 y su media geométrica es 7,2.

12. Tres profesores de estadística registraron una calificación media en sus exámenes de 71 78 89 con desviaciones típicas de 9 8 7; sus clases estaban formadas por 30 25 y 15 estudiantes respectivamente. Determinar la calificación media y la desviación típica para el conjunto de los 70 estudiantes. ¿Cuáles serían estas medidas, si todos los cursos tienen el mismo número de estudiantes?

13. La distribución por edades de los inmigrantes extranjeros que en el mes de Abril de 2011 arribaron a un país X por vía marítima fueron:

EDAD (AÑOS):	0 - 10	10,1 - 20	20,1 - 30	30,1 - 40	40,1 - 50	50,1 - 60	60,1 - 70	70,1 - 80	80,1 - 90
PASAJEROS:	640	684	863	876	753	663	414	154	13

a) Calcule el coeficiente de variación de la distribución.

b) Aumente en 15 años todas las edades y calcule el coeficiente de variación ¿Qué opina al respecto?

c) Fijar los límites de confianza para la $\overline{y}$ obtenida en el punto (a), siendo las posibilidades del 86,64% y 98,76%.

14. En un conjunto de n valores de X, se sabe $\sum x_i^2 = 1.360$ $\quad \sum x_i = 40$ y $ns^2 = 1.280$. Se pide encontrar n.

15. Se toman las medidas de 80 personas las que tienen una estatura media de 170 cmts y una varianza de 7,4 cmts. Posteriormente se verificó que el metro usado en la medición tenía 9 cmts menos. Se pide rectificar los estadígrafos mencionados.

16. Dada una serie de valores de x: 3, 2, 8, 1, comprobar que siendo y = 2x + 5, la media de la serie ( ) es $\overline{y} = 2\overline{x} + 5$ y su varianza: $s_y^2 = 4\, s_x^2$

17. Responda a los siguientes puntos señalando si es cierto o falso.
    a) En una distribución simétrica la desviación media es igual a 0,7979 S.
    b) Si se multiplica la distribución por una constante, el coeficiente de variación se altera.
    c) La desviación mediana es menor o igual a la desviación media, a su vez, es menor que la desviación típica

18. Un profesor de educación física informa, que el peso promedio de un grupo de alumnos que conforman el equipo de fútbol es de 56 kilos y que su desviación típica es de 7 cmts. ¿Es correcto? ¿cual debería ser el coeficiente de variación?

19. Si se tienen dos distribuciones A y B en el primero $S^2 = 16$ y en la segunda $S^2 = 25$, ambos se dan en la misma unidad de medida. ¿En cuál de ellas hay una mayor variabilidad absoluta?

20. Si las dos varianzas del ejercicio anterior se dieran en unidades de medidas diferentes, ¿cuál de ellas tienen mayor variabilidad absoluta? Si no se puede responder ¿qué medida de dispersión utilizaría usted?

21. Si en una distribución se tiene que la varianza calculada es de 4 horas ¿cuál sería la varianza de esa misma distribución, pero en minutos?

22. Con los siguientes datos obtenidos en dos distribuciones, con datos en las mismas unidades de medida, ¿cuál será la desviación típica para el conjunto A + B?
$s_A^2 = 28{,}2 \quad s_B^2 = 14{,}6 \quad \bar{y}_A = 12 \quad \bar{y}_B = 9 \quad n_A = 20 \quad n_B = 30$

23. ¿Es posible que una varianza sea negativa? ¿Por qué?

VARIACIÓN ABSOLUTA

Dos o más varianzas y dos o más desviaciones típicas se pueden comparar entre sí, si las variables están dadas en las mismas unidades de medidas, en caso contrario, se deberá utilizar el coeficiente de variación.

COEFICIENTE DE VARIACIÓN CV

Más conocido como *variación relativa*. En ocasiones nos interesa comparar la variabilidad de dos series de datos, sin embargo podemos encontrar, al hacerlo, que ambas series están expresadas en diferentes unidades de medidas, por lo tanto no se podrán comparar sus varianzas o sus desviaciones típicas. Puede darse el caso de que estén expresadas en la misma unidad, pero nos interesa determinar la variación respecto a una base. Para resolver los anteriores problemas se usa el *coeficiente de variación*. Si es multiplicado por 100, el resultado se dará en términos porcentuales.

$$\boxed{CV = \frac{S}{\bar{x}}} \text{ (relativo)} \qquad \boxed{CV = d = \frac{S}{\bar{x}}\,100} \text{ (porcentual)}$$

De acuerdo a la fórmula anterior, es *el cociente entre la desviación típica estándar y su media aritmética*, como en ocasiones se expresa en porcentajes, por esa razón aparece multiplicado por 100. Fue propuesto por Karl Pearson (1895) con el propósito de comparar la variabilidad entre varias distribuciones de frecuencias.

$$S = 5{,}12 \qquad \bar{x} = 14{,}75 \qquad CV = d = \frac{5{,}12}{14{,}75}\,100 = 34{,}71\%$$

La aplicación de esta fórmula presenta el inconveniente de que varía para distribuciones que tienen diferentes medias aritméticas e igual varianza, es decir, que con igual dispersión tienen diferentes coeficientes de variación.

Ejemplo 1. Si consideramos dos distribuciones cuyas medias aritméticas son:

$\bar{x}_1 = 24{,}5$ y $\bar{x}_2 = 30$, además sus desviaciones típicas o estándar son idénticas, $S_1 = 2$ y $S_2 = 2$, esto nos indica que tienen la misma variación absoluta, sin embargo al calcular los coeficientes de variación, encontramos que presentan diferente variación relativa a pesar de tener la misma desviación.

o :

$$CV_1 = d_1 = \frac{S_1}{\bar{x}_1} = \frac{2}{24{,}5} = 0{,}0816 = 8{,}16\% \qquad CV_2 = d_2 = \frac{S_2}{\bar{x}_2} = \frac{2}{30} = 0{,}0666 = 6{,}66\%$$

Ejemplo 2. Para la media y la varianza de un conjunto se han hallado, respectivamente, los valores 4 y 25. ¿Qué opinión merece la media aritmética?

o :

$$\boxed{CV = d = \frac{S}{\bar{x}} \ 100} \qquad CV = d = \frac{5}{4} = 1{,}25 = 125\%$$

Esta media aritmética no es lo suficientemente representativa de la distribución, debido a la variabilidad tan alta que presenta.

Ejemplo 3. Multiplicando por 4 cada uno de los valores de la variable x_i : 3 2 0 5 se obtiene la serie x_i: 12 8 0 20. Comprobar que ambas series tienen el mismo coeficiente de variación.

o :

x_i	$4(x_i)$	x_i^2	$16(x_i^2)$
3	12	9	144
2	8	4	64
0	0	0	0
5	20	25	400
10	40	38	608

$$\bar{x}_1 = \frac{10}{4} = 2{,}5$$

$$\bar{x}_2 = \frac{40}{4} = 10$$

$$M_{KX} = 4(2{,}5) = 10 = \bar{x}_2$$

$$S_1^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2 = \frac{38}{4} - 2{,}5^2 \qquad S_1^2 = 9{,}5 - 6{,}25 = 3{,}25 \qquad S_1 = \sqrt{3{,}25} = 1{,}80$$

$$V_{[KX]} = K^2 V_{[x]} \qquad V_{[4X]} = 16\ (3{,}25) = 52$$

$$V_{[4X]} = \frac{608}{4} - 10^2 = 152 - 100 = 52 \qquad V_{[4X]} = 52 \qquad S_2 = 7{,}21$$

$$CV_1 = \frac{1{,}8}{2{,}5} = 0{,}72 = 72\% \qquad CV_2 = \frac{7{,}21}{10} = 0{,}72 = 72\% \qquad CV_1 = CV_2$$

PUNTAJE TÍPICO O ESTANDARIZADO: Z ó "t

Este estadígrafo de dispersión, muy utilizado en la distribución normal y en el análisis del coeficiente de correlación, mide la *desviación de una observación con respecto a la media aritmética en unidades de desviación típica*, determinando la posición de una observación dada, dentro de un conjunto de observaciones.

$$Z = \frac{x - \bar{x}}{S} \qquad \text{ó} \qquad t = \frac{x - \bar{x}}{S}$$

El *puntaje típico* sirve para comparar dos o más datos individuales, aunque pertenezcan a distribuciones diferentes, aún en casos en que la media y/o varianza no coincidan.

Tanto el puntaje "*t*" o "*z*" es muy utilizado en las pruebas educacionales y psicológicas, aunque hoy en día es muy frecuente en otros tipos de experimentos.

Ejemplo 1 Supongamos que una persona presenta exámenes de admisión para la Facultad de Economía, en dos universidades de San José. En la primera obtuvo una calificación de 85 y en la segunda de 60, además se sabe que el promedio de calificación para el ingreso de la universidad A es de 90, con una desviación estándar de 10, en cambio en la universidad B es de 55, con una desviación estándar de 12. Se desea saber en cuál de las universidades esta persona se desempeñó relativamente mejor?

$$Z_A = \frac{85 - 90}{10} = \frac{-5}{10} = -0{,}50 \qquad\qquad Z_B = \frac{60 - 55}{12} = \frac{5}{12} = 0{,}42$$

Lo anterior indica, que el estudiante se desempeñó mejor en la universidad B, donde logró 0,42 de desviación estándar por encima del promedio, en cambio, en B fue de – 0,50 por debajo del promedio.

En una *distribución normal la media de la variable* Z *vale* 0 *y su varianza es igual a* 1. El puntaje típico estandarizado va a ser utilizado en gran medida en el Capitulo 6, correspondiente a la Distribución Normal.

Ejemplo 2. En un examen final de estadística, la puntuación media de un grupo de 150 estudiantes fue de 78 y la varianza 64. En álgebra, sin embargo, la media final del grupo fue 73 y la desviación típica 7,6. En qué asignatura hubo mayor.

a) Dispersión absoluta b) Dispersión relativa

c) Si el estudiante consiguió una nota de 75 en estadística y 70 en álgebra ¿En qué asignatura fue su puntuación relativa superior?

o :

a) Dispersión absoluta:

$S_1^2 = 64 \qquad S_1 = 8 \qquad S_2^2 = 57{,}76 \qquad S_2 = 7{,}6$

En estadística hubo una mayor dispersión absoluta: $S_1 > S_2 \qquad 8 > 7{,}6$

b) Dispersión relativa

$$CV_1 = \frac{8}{78} = 0{,}1025 \qquad CV_1 = 10{,}25\%$$

$$CV_2 = \frac{7{,}6}{73} = 0{,}1041 \qquad CV_2 = 10{,}41\%$$

En álgebra hubo una mayor dispersión relativa: $\boxed{CV_2 > CV_1} \rightarrow 10{,}41\% > 10{,}25\%$

c) Puntuación relativa: $Z = \dfrac{x - \bar{x}}{S}$

$$Z_1 = \frac{75 - 78}{8} = \frac{-3}{8} = -0{,}37$$

$$Z_2 = \frac{71 - 73}{7{,}6} = \frac{-2}{7{,}6} = -0{,}26 \qquad Z_2 > Z_1$$

Su puntuación relativa fue superior en álgebra, dado que – 0,26 > – 0,37

DESVIACIÓN MEDIA : $_a$

Es otro estadígrafo de dispersión de menor importancia, en relación a la varianza y la desviación típica. Se define como la *media de las desviaciones respecto a la media aritmética*, *tomadas en valor absoluto.*

$$D_a = \frac{\Sigma|x_i - \bar{x}|}{n}$$ (Datos no agrupados) $$D_a = \frac{\Sigma|X_i - \bar{X}|}{N} = \frac{\Sigma|d_i|}{N}$$

$$D_a = \frac{\Sigma|y_i - \bar{y}|n_i}{n}$$ (Datos agrupados) $$D_a = \frac{\Sigma|X_i - \bar{X}|f_i}{N} = \frac{\Sigma|d_i|f_i}{N}$$

Ejemplo 1. Con los siguientes datos no agrupados, calcular la desviación media:

$x_1 = 5$ $x_2 = 3$ $x_3 = 1$ $x_4 = 6$ $x_5 = 10$

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$
5	0	0
3	−2	2
1	−4	4
6	1	1
10	5	5
25	0	12

$$\bar{x} = \frac{25}{5} = 5$$

$$D_a = \frac{\Sigma|x_i - \bar{x}|}{n} \qquad D_a = \frac{\Sigma|d_i|}{N}$$

$$D_a = \frac{12}{5} = 2{,}4$$

Ejemplo 2. Con los datos agrupados, que se presentan a continuación, calcular la Desviación Media.

El procedimiento de cálculo para datos agrupados es el siguiente:

y_i	n_i	$y_i - \bar{y}$	$\lvert y_i - \bar{y}\rvert$	$\lvert y_i - \bar{y}\rvert n_i$
5	2	−9,75	9,75	19,50
10	4	−4,75	4,75	19,00
15	8	0,25	0,25	2,00
20	5	5,25	5,25	26,25
25	0	10,25	10,25	10,25
Σ	20	—	—	77,00
X_i	f_i	d_i	$\lvert d_i\rvert$	$\lvert d_i\rvert f_i$

$$\bar{y} = 14{,}75 = \bar{X}$$

$$D_a = \frac{\Sigma|y_i - \bar{y}|n_i}{n}$$

$$D_a = \frac{77}{20} = 3{,}85$$

$$D_a = \frac{\Sigma|d_i|f_i}{n}$$

En una distribución Simétrica o Normal, la Desviación Media es igual al producto de 0,7979 por la desviación típica.

$$D_a = 0{,}7979\, S$$

En el ejercicio anterior tenemos una desviación típica de 5,12, por lo tanto la Desviación media será igual a: $D_a = 0{,}7979\ (5{,}12) = 4{,}09$

Dado que la distribución anterior no es *simétrica,* presenta el resultado obtenido alguna diferencia; sin embargo, vemos que la relación se cumple aproximadamente, por ser la distribución ligeramente *asimétrica*.

La *desviación media* se usa muy poco; en el caso de ser calculada en una *distribución de frecuencias* será necesario suponer que la *media aritmética* representa adecuadamente a los valores de la variable.

El valor obtenido al calcular la *desviación media* siempre será menor al de la *desviación típica*

$$\boxed{D_a < S}$$

COEFICIENTE DE DESVIACIÓN MEDIA : a

Esta medida de dispersión es muy parecida en su presentación y cálculo al *Coeficiente de variación*, con la diferencia de considerar a la *Desviación media* (*Da*) en vez de la *Desviación típica* (*S*). También se puede expresar en términos porcentuales.

Ejemplo 1. De acuerdo a los datos de dos ejemplos anteriores (desviación media) se pide aplicar esta medida de dispersión:

o :

Ejemplo1: $CD_a = \frac{D_a}{\bar{x}} 100 = \frac{2,4}{5} = 0,48$ o 48% (en datos no agrupados)

Ejemplo 2: $CD_a = \frac{D_a}{\bar{y}} 100 = \frac{3,85}{20} = 0,1925 = 19,25\%$ (en datos agrupados)

DESVIACIÓN MEDIANA : e

Corresponde a los denominados estadígrafos o medidas de dispersión, siendo su uso más bien limitado. Se define como la *Media de los valores absolutos de las diferencias entre los valores que toma la variable y su mediana.*

Sus fórmulas son:

$$D_e = \frac{\Sigma \left| x_i - M_e \right|}{n}$$ (Datos no agrupados)

$$D_e = \frac{\Sigma \left| y_i - M_e \right| n_i}{n}$$ (Datos agrupados) $$D_e = \frac{\Sigma \left| X_i - M_e \right| f_i}{N}$$

La desviación mediana, será menor o igual a la *Desviación media* y ésta a la vez, será menor que la *Desviación típica*: $\boxed{D \leq D_a < S}$

Ejemplos 1. Con los siguientes datos calcular la desviación mediana.

o :

a) $x_1 = 5$ $x_2 = 3$ $x_3 = 1$ $x_4 = 6$ $x_5 = 10$

Se calcula primero la mediana para ello ordenamos los valores de menor a mayor.
1 – 3 – $\boxed{5}$ – 6 – 10 Luego establecemos las diferencias entre la variable y el promedio M_e

x_i	$x_i - M_e$	$\lvert x_i - M_e \rvert$
1	– 4	4
3	– 2	2
5	0	0
6	1	1
10	5	5
Σ	0	12

$$D_e = \frac{\Sigma \lvert x_i - M_e \rvert}{n}$$

$$D_e = \frac{12}{5} = 2{,}4$$

Ejemplo 2. Con datos agrupados, se pide calcular la desviación media

	y_i	n_i	N_i	
	5	2	2	
	10	4	6	$\leftarrow N_{j-1}$
$y_J \rightarrow$	15	8	14	$\leftarrow N_j$
	20	5	19	
	25	1	20	
	Σ	20	–	
	X_i	f_i	F_i	

$$N_{J-1} < \frac{n}{2}$$

$$6 < 10$$

$$M_e = y_J = 15$$

$$\boxed{M_e = X_J = 15}$$

Continuación:

$y_i - M_e$	$\lvert y_i - M_e \rvert$	$\lvert y_i - M_e \rvert n_i$
–10	10	20
–5	5	20
0	0	0
5	5	25
10	10	10
0	–	75
$X_i - M_e$	$\lvert X_i - M_e \rvert$	$\lvert X_i - M_e \rvert f_i$

$$D_e = \frac{\Sigma \lvert y_i - M_e \rvert n_i}{n}$$

$$D_e = \frac{75}{20} = 3{,}75$$

$$\boxed{D_e = \frac{\Sigma \lvert X_i - M_e \rvert f_i}{N}}$$

EJERCICIOS PARA RESOLVER

Complemento virtual SIL

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

24. Un conjunto de 20 valores tiene una media igual a 50; otro conjunto de 20 valores tiene una media igual a 30; la desviación estándar de los 40 valores considerados conjuntamente es igual a 10. Calcular el coeficiente de variación del conjunto de los 40 números.

25. Reconstruya y calcule el coeficiente de variación de la siguiente distribución simétrica. (Para el cálculo de la media y la varianza, hacerlo mediante el método que usted conoce).

X_i:	$\Rightarrow$	y_i	10	20	30	40	50
f_i:	$\Rightarrow$	n_i	6				
F_i/n	$\Rightarrow$	H_i	0,12	0,32			

26. Hallar el coeficiente de variación de dos números, cuya media aritmética es 5 y la media geométrica es 4.

27. Se tienen los siguientes datos correspondientes a dos submuestras:
$\bar{x}_1 = 120 \quad \bar{x}_2 = 125 \quad n_1 = 70 \quad n_2 = 30 \quad S_1^2 = 36 \quad S_2 = 3$
Se pide calcular el coeficiente de variación para el conjunto.

28. Se supone que $x_1 \; x_2 \; x_3 ... x_n$ son elementos que poseen una media aritmética 10 y una desviación típica de 3. Si se considera una nueva media y_i, de cada uno de los elementos relacionados con la media x_i mediante la ecuación.

$$y_i = 4x_i + 2$$

¿Cuál es la media aritmética, la varianza y el coeficiente de variación de los nuevos valores y_i obtenidos?

29. Las notas de 50 alumnos se clasifican en una tabla de frecuencias con cuatro intervalos de igual magnitud. Se pide calcular el coeficiente de variación, sabiendo que:

$y_2 = 50$	$n_1 = 4$	$N_2 = 20$	$n_3 = 25$	$\bar{y} = 62,4$
$X_2 = 50$	$f_1 = 4$	$F_2 = 20$	$f_3 = 25$	$\bar{X} = 62,4$

30. El cuadro de pagos en dos fábricas para el año 2011 fue así:

MEDIDAS	FÁBRICA A	FÁBRICA B
Salario medio semanal	\$930.000	\$1.020.000
Desviación típica del salario semanal	\$96.000	\$97.000

a) ¿En cuál de las fábricas los jornales fueron absolutamente más variables?
b) ¿Relativamente más variables?

31. Se analizaron en el primer semestre del año 1 los gastos de una empresa de la construcción y se obtuvo un promedio de \$1.740 millones y una desviación típica de \$90 millones. Se determinó luego que los contadores de esta empresa habían prescindido de \$13 millones en los gastos por un error de apreciación. Corrigiendo las medidas enunciadas, obtenga el coeficiente de variación de los gastos de esta empresa.

32. Para cierta variable se sabe que $M_{[4+X]} = 11,2$ y $M_{[X+2]^2} = 30,25$
Calcule el coeficiente de variación.

33. En cierta región la distribución de predios por extensión tiene una media de 35,4 hectáreas y una desviación típica de 19,33 hectáreas, mientras que la distribución por canon de arrendamiento tiene una media de \$945.750 y una desviación de \$74.708. ¿Cuál de las dos distribuciones tiene mayor variabilidad?

34. Una empresa fabrica bombillas eléctricas de dos tipos (A y B), con base en muestras de la producción, se sabe que las distribuciones de la duración en horas de esas bombillas son tales que tienen las siguientes medias y varianzas.

TIPO	MEDIA	VARIANZA
A	800 horas	7.800
B	650 horas	5.400

Se pide:

a) Comparar ambas distribuciones en cuanto a su variabilidad absoluta y relativa.

b) Si se extrajo una bombilla de cada tipo y su duración fue de 700 y 630 horas respectivamente, se quiere saber cuál tipo de bombilla tiene mejor posición relativa.

35. Un almacén vende un promedio mensual de $3.000.000 (cientos $) en telas con una desviación típica de $30.000 (cientos de $).

a) El almacén paga un impuesto igual al 10% sobre las ventas. ¿Cuál será la varianza de las ventas una vez pagado el impuesto?

b) Las utilidades del almacén se calculan teniendo en cuenta que mensualmente se pagan $2.450.000 (cientos de $) por salarios, gastos, etc., además del impuesto sobre las ventas. Calcule el coeficiente de variación de las utilidades.

36. Los siguientes datos muestran el número de cuartos habitables en 60 viviendas localizadas en la ciudad de Cúcuta.

NO. DE CUARTOS:	2	4	6	8	10
NO. DE VIVIENDAS:	6	18	16	12	8

Obtenga e interprete para estos datos la desviación típica, la desviación mediana y el coeficiente de variación.

37. Los trabajadores de una empresa solicitan, en una convención colectiva, que cada salario semanal de sus afiliados sea aumentado según la ecuación:

$$\boxed{y_i = 1{,}13x + 10.750}$$

La empresa tiene 1.000 trabajadores y antes de solicitar el reajuste salarial, devengaban un promedio de $120.000 semanal.

a) ¿Cuál será el nuevo promedio de ingreso mensual de los trabajadores si la empresa acepta la petición?

b) El coeficiente de variación antes de solicitar el reajuste salarial es del 38%. ¿Cuál será el nuevo coeficiente de variación?

38. Con los datos siguientes:

$$\Sigma y_i n_i = 1.500 \quad \rightarrow \quad \Sigma x_i f_i$$
$$\Sigma y_i^2 n_i = 40.000 \quad \rightarrow \quad \Sigma x_i^2 f_i$$
$$n = 100 \quad \rightarrow \quad N$$

Se pide calcular:

a) La desviación estándar.

b) El coeficiente de variación.

c) El puntaje típico semanal, si $x = 24$

39. Al trabajar con cierta información se obtienen los siguientes datos: $\overline{x} = 36$ $S = 8$ $CV = 0{,}22$.
Si a cada uno de los valores de la variable se les aumenta en 5 ¿Cuáles serían las nuevas medidas: media, desviación estándar y coeficiente de variación?

40. Con los datos correspondientes a dos submuestras.

a)

x_i :	4	6	7	6	10	6	4	15	6	X_1
y_i :	10	18	12	20	8	10				X_2

Se pide:

b) Determinar en cada una de las submuestras la mejor posición de los siguientes valores $x = 8$ y $y = 15$.

c) Transformando las dos submuestras en una sola muestra, calcular el coeficiente de variación.

41. Con la siguiente distribución de frecuencias:

a)

$y'_{i-1} - y'_i$	n_i
8,1 – 16	3
16,1 – 24	6
24,1 – 32	10
32,1 – 40	15
40,1 – 48	4
48,1 – 56	2
Σ	40
$X'_{i-1} - X'_i$	f_i

b) $n = 8$

x_i : 8 10 7 30 25 38 47 70

Se pide determinar:

1) En cuál de las dos distribuciones (a y b) hay una mayor variabilidad absoluta.
2) En cuál hay una menor variabilidad relativa
3) Si se tiene un valor $x = 33$, en cuál de las dos distribuciones, dicho valor tiene una mejor posición relativa.

42. En un grupo de 70 personas, se sabe que 30 de ellas tienen un salario medio diario de \$27.000 y los restantes de \$32.000.

a) Se pide calcular el salario medio diario de las 70 personas.

b) Si se sabe que el coeficiente de variación es del 35%, ¿cuál es la varianza de los salarios diarios de las 70 personas?

43. Si la media de 10 observaciones es 5,7 y la suma de los cuadrados es 490. Encontrar el coeficiente de variación para el conjunto.

44. Los siguientes datos corresponden a la distribución de frecuencias de los gastos de publicidad (miles de millones de pesos) de 50 empresas comerciales, durante el último trimestre de 2011. Dichos gastos se agruparon en cuatro clases de amplitud constante, de la cual se sabe:

$y_1 = 3,5$ $y'_4 = 8,75$ $n_1 = 4$ $N_2 = 20$ $n_3 = 25$

Se pide calcular el coeficiente de variación, la desviación media y la desviación mediana.

45. Con la siguiente distribución calcule:

$y'_{i-1} - y'_i$	N_i
10,1 – 16	4
16,1 – 22	12
22,1 – 28	25
28,1 – 34	35
34,1 – 40	44
40,1 – 46	50
Σ	–
$X_{i-1} - X_i$	F_i

a) La desviación típica o estándar.
b) La desviación media.
c) La desviación mediana.
d) Coeficiente de variación.

46. En la siguiente distribución de resistencia a la tensión (en kgrs./mm^2) de láminas de acero. Obtenga la desviación mediana:

RESISTENCIA:	10 a 20	20 a 30	30 a 40	40 a 50
% DE LÁMINAS:	0,20	?	?	0,12

Nota: se sabe que la resistencia promedio es de 29,4 kgrs./mm^2.

47. La siguiente distribución de frecuencias se refiere al número de compradores de un determinado producto, en 30 barrios de clase media en Bogotá.

No. DE COMPRADORES	No. DE BARRIOS
130	3
148	6
160	5
220	3
280	2
320	4
400	7

a) Calcule la desviación mediana.
b) Calcule el coeficiente de variación mediana.
c) Si la empresa considera que el número de compradores por barrio es el doble. ¿Cuál será la nueva varianza y el coeficiente de variación?

48. Con los siguientes datos se pide:

1)

$y_{i-1} - y_i$	n_i
8,1 – 16	3
16,1 – 24	6
24,1 – 32	10
32,1 – 40	15
40,1 – 48	4
48,1 – 56	2
Σ	40
$X_{i-1} - X_i$	f_i

2) x_i: 8 10 7 30 38 25 47 70 .
b) Si se considera (1) y (2) como submuestras, ¿cuál será el coeficiente de variación para el conjunto o sea cuando .$n = 48 = 40 + 8$.
c) Con la distribución (1) determine la desviación media.
c) Con la distribución (2) determine la desviación mediana.
d) Compruebe con la distribución (a) que

$$De \leq Da < S$$

49. Con los siguientes datos: 2 5 8 12 y 20. Calcular la desviación típica, desviación media y mediana. Observar que $S > Da \geq De$

50. Calcular el coeficiente de la desviación media a los ejercicios: 47 48 y 49

51. Calcular el coeficiente de la desviación mediana a los ejercicios: 47 48 y 49

RECORRIDO INTERCUARTÍLICO, DESVIACIÓN CUARTIL Y COEFICIENTE DE DESVIACIÓN CUARTIL

Se considera una serie de medidas de dispersión cuando se ha trabajado con *cuartiles, deciles* o *percentiles.* Entre los más importantes se tienen:

El *recorrido intercuartílico*, se define como la diferencia entre el tercer cuartil y el primero.

$$Q_R = Q_3 - Q_1$$

El *recorrido interdecil*, corresponde a la diferencia entre el noveno y el primer decil.

$$D_R = D_9 - D_1$$

La *desviación cuartil o semi-recorrido intercuartílico*, se obtiene, mediante el cálculo del recorrido intercuartílico divido entre 2.

$$Q_D = \frac{Q_3 - Q_1}{2}$$

Este estadígrafo, en una distribución normal, equivale a 0,6745 veces la desviación estándar.

El *coeficiente de desviación cuartil*, es aplicado, especialmente, en aquellas distribuciones cuyos valores extremos no se encuentran definidos o cuando se desea obtener una rápida estimación de la dispersión, no necesaria para posteriores cálculos.

$$C_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

El *recorrido interdecil* se calcula estableciendo la diferencia entre el noveno y el primer decil, lo mismo, para obtener el recorrido interpercentil será la diferencia entre el percentil 99 y el percentil uno.

$$D_D = D_3 - D_1 \qquad P_D = P_{99} - P_1$$

Ejemplo 1. Con la siguiente distribución calcular el recorrido intercuartílico, la desviación cuartil y el coeficiente de desviación cuartil.

o **:**

$y'_{i-1} - y'_i$	n_i	N_i
30,1 – 38	6	6
38,1 – 46	14	20 $\leftarrow N_{j-1}$
46,1 – 54	$n_j \rightarrow 36$	56 $\leftarrow N_j$
54,1 – 62	50	106
62,1 – 70	43	149 $\leftarrow N_{j-1}$
70,1 – 78	$n_j \rightarrow 32$	181 $\leftarrow N_j$
78,1 – 86	18	199
86,1 – 94	9	208
Σ	208	-
$X'_{i-1} - X'_i$	f_i	F_i

Primer cuartil

$\frac{n}{4} = \frac{208}{4} = 52$ no está, en la columna de N_i, por lo tanto se toma a $N_{i-1} = 20$ y $N_i = 56$.

Siendo que $N_{J-1} < \frac{n}{4}$ se tendrá:

$$Q_1 = y'_{j-1} + C\left[\frac{\frac{n}{4} - N_{J-1}}{n_j}\right] \qquad Q_1 = 46 + 8\left[\frac{52-20}{36}\right] = 53{,}11$$

Tercer cuartil

$\frac{3n}{4} = \frac{3(208)}{4} = 156$ como no se encuentra en la columna de las frecuencias absolutas acumuladas, se toma a $N_{j-1} = 149 < 156$ y $N_j = 181 > 156$. Ahora, cuando se presenta $N_{j-1} < \frac{3n}{4}$ se tendrá:

$$Q_3 = y'_{j-1} + C\left[\frac{\frac{3n}{4} - N_{J-1}}{n_J}\right]$$ Reemplazando $$Q_3 = 70 + 8\left[\frac{156 - 149}{32}\right] = 71{,}75$$

Recorrido intercuartílico: $Q_D = Q_3 - Q_1$ $\quad Q_D = 71{,}75 - 53{,}11 = 18{,}64$

Desviación cuartil será igual a: $Q_{D_2} = \dfrac{Q_3 - Q_1}{2}$ $\quad Q_{D_2} = \dfrac{71{,}75 - 53{,}11}{2} = 9{,}32$

Coeficiente de desviación cuartil será igual a:

$$C_Q = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad C_Q = \frac{71{,}75 - 53{,}11}{71{,}75 + 53{,}11} = \frac{18{,}64}{124{,}86} = 0{,}1439 \text{ ó } 14{,}39\%$$

Ejemplo 2. Con la misma distribución anterior calculemos el recorrido interdecil, la desviación interdecil y el coeficiente de desviación decil.

Primer decil:

$\dfrac{n}{10} = \dfrac{208}{10} = 20{,}8$ Este valor no se encuentra en la columna de las frecuencias absolutas acumuladas, por lo tanto $N_{j-1} = 20 < 20{,}8$ y $N_j = 56 > 20{,}8$, la fórmula de cálculo que debe ser aplicada será:

$$D_1 = y'_{j-1} + C\left[\frac{\frac{n}{10} - N_{J-1}}{n_J}\right] \qquad D_1 = L_i + i\left[\frac{\frac{n}{10} - F_{J-1}}{f_J}\right]$$

Reemplazando en la fórmula anterior se tendrá: $D_1 = 46 + 8\left[\dfrac{20{,}8 - 20}{36}\right] = 46{,}18$

Noveno decil:

$\dfrac{9n}{10} = \dfrac{9(208)}{10} = 187{,}2$ Siendo que $N_{J-1} = 181 < 187{,}2$ y $N_J = 199 > 187{,}2$ se tendrá que:

$$D_9 = y'_{j-1} + C\left[\frac{\frac{9n}{10} - N_{J-1}}{n_J}\right] \qquad D_9 = 78 + 8\left[\frac{187{,}2 - 181}{18}\right] = 80{,}76$$

$$D_9 = L_i + i\left[\frac{\frac{9n}{10} - F_{J-1}}{f_J}\right]$$

El *recorrido interdecil*: $D_R = D_9 - D_1$ $\quad D_R = 80{,}76 - 46{,}18 = 34{,}58$

La *desviación interdecil*: $D_{D_2} = \dfrac{D_9 - D_1}{2}$ $\quad D_{D_2} = \dfrac{80{,}76 - 46{,}18}{2} = 17{,}29$

El *coeficiente de desviación decil*:

$$C_D = \frac{D_9 - D_1}{D_9 + D_1} \qquad C_D = \frac{80{,}76\ -\ 46{,}18}{80{,}76\ +\ 46{,}18} = \frac{17{,}29}{126{,}94} = 13{,}62\% \qquad 13{,}62\%$$

RECORRIDO U OSCILACIÓN

El recorrido, denominado también como oscilación o fluctuación, es un estadígrafo fácil de calcular y su significado muy sencillo. Se calcula tomando la diferencia entre los valores extremos, sin tener en cuenta la repetición o frecuencia de los valores que toma la variable, de ahí su gran inestabilidad, especialmente cuando se añaden o suprimen cifras extremas.

Recorrido = x_{max} $\quad x_{min}$ (para datos no agrupados)

En datos agrupados, para variable discreta el recorrido se obtendrá así: recorrido = $y_m - y_l$.

Para la variable continua se calcula aplicando recorrido = $y'_m - y'_o$

Ejemplo 1. Supongamos que se tienen 5 observaciones: (datos no agrupados)

$x_1 = 8 \qquad x_2 = 4 \qquad x_3 = 6 \qquad x_4 = 6 \qquad x_5 = 12$

Recorrido = $x_{max} - x_{min}$ $\qquad$ Rango = $12 - 4 = 8$

Ejemplo 2. Tomemos una distribución de frecuencias para una *variable discreta*

y_i	n_i
2	3
4	5
6	6
8	4
10	2
Σ	20
X_i	f_i

Recorrido = $y_m - y_1$

Recorrido = 10 – 2 = 8

Recorrido = $X_m - X_1$

Ejemplo 3. En el caso de una *variable continua* el procedimiento es casi igual al anterior.

$y'_{i-1} - y'_i$	n_i
3,1 – 7	6
7,1 – 11	4
11,1 – 15	12
15,1 – 20	8
20,1 – 23	5
Σ	35
$X'_{i-1} - X'_i$	f_i

Recorrido = $y'_m - y'_0$

Recorrido = 23 – 3 = 20

Recorrido = $X_m - X_0$

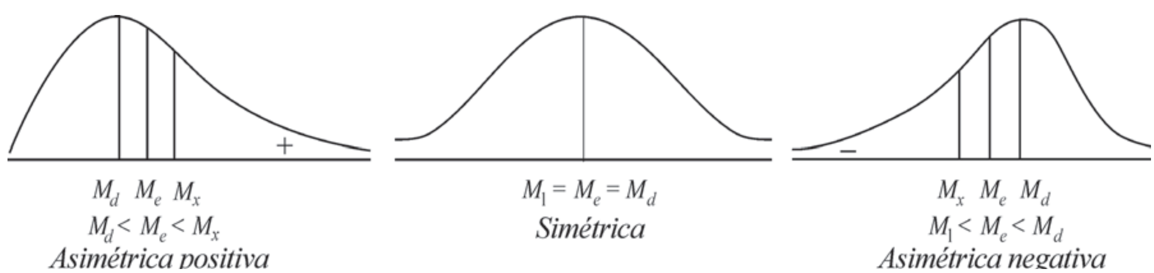
MEDIDAS DE ASIMETRÍA O DE DEFORMACIÓN

Fueron propuestas, algunas de ellas, por Karl Pearson. Una distribución simétrica no tiene riesgo; recordemos que en este caso $M_x = M_e = M_d$, estas medidas son iguales, por lo tanto consideramos que la distribución tiene la forma de una campana, denominada de Gauss o normal, ya que el promedio se ubica en todo el centro de ella.

Si las frecuencias son considerablemente altas, la distribución deja de ser *simétrica* y hablamos de distribuciones *asimétricas positivas* o *negativas* y las tres medidas (media, moda y mediana) no tienen igual valor.

En las gráficas que se presentan a continuación se observa que en la parte más alta de la distribución se ubica el Modo.

Si $M_l > M_e > M_d$ se dice que la distribución es *asimétrica negativa*, ya que la curva presenta un alargamiento hacia la derecha. Si por el contrario $M_x = M_e = M_d$, el alargamiento es hacia la izquierda y se dice que es asimétrica positiva.



Las fórmulas utilizadas para calcular el grado de asimetría son:

1) $A_S = \dfrac{M_1 - M_d}{S}$ 2) $A_S = \dfrac{3\,(M_1 - M_e)}{S}$ 3) $A_S = \dfrac{m_3}{S^3}$

En la última se utiliza el *momento de orden 3 con respecto a la media*

$$m_3 = \frac{\Sigma d_i^3\, f_i}{n} \qquad m_3 = \frac{\Sigma (y_i - \overline{y})^3\, n_i}{n}$$

El grado de asimetría, también se puede calcular aplicando la fórmula establecida por Bowley, siendo: $A_s = \dfrac{Q_3 + Q_1 - 2Q_2}{Q_3 + Q_1}$, para ello es necesario, trabajar con *cuartiles*.

Si $A_s = 0$ Simétrico Si $A_s > 0$ Asimétrico positivo Si $A_s < 0$ Asimétrico negativo, ya que el signo nos indicará hacia que lado se presenta la deformación o alargamiento de la distribución y el valor será el grado de asimetría, entre más grande sea este valor, más grande será la asimetría.

Ejemplo 1. Con los siguientes datos de una distribución de frecuencias, se puede calcular la asimetría y su grado, por las fórmulas anteriormente establecidas.

y_i	n_i	$y_i n_i$	N_i	$y_i^2 n_i$	$(y_i - \bar{y})$	$(y_i - \bar{y})^3$	$(y_i - \bar{y})^3 n_i$
2	4	8	4	16	–3,3	–35,94	–143,76
4	6	24	10 $\rightarrow N_{j-1}$	96	–1,3	–2,20	–13,20
6	5	30	15 $\rightarrow N_j$	180	0,7	0,34	1,70
8	3	24	18	192	2,7	19,68	59,04
10	2	20	20	200	4,7	103,82	207,64
Σ	20	106	–	684	–	–	111,42
X_i	f_i	$X_i f_i$	F_i	$X_i^2 f_i$	d_i	d_i^3	$d_i^3 f_i$

$\bar{y} = M_1 = \dfrac{106}{20} = 5,3$ $\quad \boxed{\bar{X} = 5,3} \quad$ $M_d = y_j = 4$ (modo)

$M_e = \dfrac{y_{j-1} + y_j}{2} \quad \boxed{M_e = \dfrac{X_{j-1} + X_j}{2}} \quad M_e = \dfrac{4+6}{2} = 5$

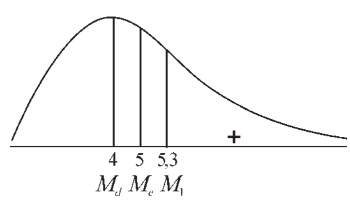
a) $A_S = \dfrac{M_1 - M_d}{S} \quad A_S = \dfrac{5,3 - 4}{2,47} = \dfrac{1,3}{2,47} = 0,53 \quad \boxed{S^2 = \dfrac{\Sigma X_i^2 f_i - n\bar{X}^2}{n}}$

$S^2 = \dfrac{\Sigma y_i^2 n_i}{n} - \bar{y}^2 = \dfrac{684}{20} - 5,3^2 = 34,2 - 28,09 = 6,11\,; \quad S = \sqrt{6,11} = 2,47$

b) $A_S = \dfrac{3(M_1 - M_e)}{S} \quad A_S = \dfrac{3(5,3 - 5)}{2,47} = \dfrac{3(0,3)}{2,47} = \dfrac{0,9}{2,47} = 0,36$

c) $A_S = \dfrac{m_3}{S^3} = \dfrac{5,57}{15,07} = 0,37 \quad m_3 = \dfrac{\Sigma(y_i - \bar{y})^3 n_i}{n} = \dfrac{111,42}{20} = 5,57 \quad \boxed{m_3 = \dfrac{\Sigma d_i^3 f_i}{N}}$

Siendo: $M_1 > M_e > M_d$
$\quad 5,3 > 5 > 4$



La distribución es asimétrica positiva

La *asimetria*, también es calculada mediante la aplicación de la fórmula de Bowley, donde

$$A_S = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 + Q_1}$$

Veamos cómo se calcula con los datos de la distribución anterior:

$Q_1 = y_5 = 4 \quad \rightarrow \quad$ siendo: $\dfrac{20}{4} = 5 \quad N_{j-1} = 4$ y $N_j = 10$

$Q_3 = \dfrac{6+8}{2} = 7 \quad \rightarrow \quad$ siendo: $\dfrac{3(20)}{4} = 15 \quad N_{j-1} = 15$ y $N_j = 18$

$Q_2 =$ mediana $= 5 \quad \rightarrow \quad A_S = \dfrac{(7+4) - 2(5)}{(7-4)} = \dfrac{1}{3} = 0,33$

Nota: la aplicación de las anteriores medidas es preferible hacerla con la variable continua, usando los intervalos de clase.

Nota: Los resultados obtenidos en la aplicación de las fórmulas anteriores, arrojan resultados diferentes sin que esto tenga importancia alguna.

Ejemplo 2. Determinar si es grande o pequeño el grado de asimetría, en una distribución cuyos estadígrafos de posición son:

$\overline{x} = 189{,}87 \qquad M_e = 189{,}16 \qquad M_d = 187{,}60$

o :

Aplicando la fórmula empírica: $\overline{x} - M_d = 3\,(\overline{x} - M_e)$

$189{,}87 - 187{,}60 = 2{,}27 \neq 2{,}13$

$3\,(189{,}87) - 189{,}16 = 2{,}03$ La diferencia nos indica que existe una pequeña asimetría.

Ejemplo 3. Tomando una distribución ligeramente asimétrica. Calcular el modo sabiendo que su media es igual a 3 y que la diferencia entre la media y mediana es igual a –2

o :

$\overline{x} - M_d = 3\,(\overline{x} - M_e) \qquad 3 - M_d = -6$

$3 - M_d = 3\,(-2) \qquad M_d = \overline{x} - 3\,(\overline{x} - M_e) \qquad M_d = 9$

Ejemplo 4. Calcular el coeficiente de asimetría en la siguiente distribución, utilizando la fórmula de Pearson.

y_i :	1	2	3	4	5	6
n_i :	2	8	3	5	7	5

o :

y_i	n_i	$y_i\,n_i$	$x_i^2\,n_i$
1	2	2	2
2	8	16	32
3	3	9	27
4	5	20	80
5	7	35	175
6	5	30	180
Σ	30	112	496
X_i	f_i	$X_i\,f_i$	$X_i^2\,f_i$

$$\overline{X} = \frac{\Sigma X_i\, f_i}{N} \qquad S^2 = \frac{\Sigma X_i^2\, f_i}{N} - \overline{X}^2$$

$$\overline{y} = \frac{\Sigma y_i\, n_i}{n} \qquad \overline{y} = \frac{112}{30} = 3{,}73 \qquad M_d = 2$$

$$S^2 = \frac{\Sigma y_i^2\, n_i}{n} - \overline{y}^2 \qquad S^2 = \frac{496}{30} - 3{,}73^2 = 2{,}62 \;\Rightarrow\; S = 1{,}62$$

$$A_S = \frac{\overline{y} - M_d}{S} \qquad A_S = \frac{3{,}73 - 2}{1{,}62} = 1{,}07 \qquad A_S = \frac{\overline{x} - M_d}{S}$$

EJERCICIOS PARA RESOLVER

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

52. Conteste si es cierto o falso y por qué:

a) En toda serie de datos $\dfrac{\Sigma x_i^2}{n}$ es mayor o igual a $\left(\dfrac{\Sigma x_i}{n}\right)^2$

b) Si para dos distribuciones $\overline{x}_1 = \overline{x}_2$ y $S_1^2 > S_2^2$ entonces $CV_1 < CV_2$

c) Si el coeficiente de asimetría es negativo, la curva es achatada.
d) Si las frecuencias de una distribución se multiplican por 8, la varianza queda multiplicada por 64.
e) La varianza está expresada en las mismas unidades de la variable.

53. Conteste los siguientes puntos diciendo si es cierto o falso y por qué:
a) El coeficiente de variación no puede ir expresado en unidades de la variable.
b) El coeficiente de variación, es la desviación estándar expresada en términos de porcentajes de $\bar{x}$.
c) Si una distribución es simétrica o moderadamente asimétrica, entre $\bar{x} \pm 2S$ está el 95,5% de las observaciones.
d) La desviación típica no se mide en las mismas unidades de la variable.

54. Con con la información sobre el número de retardos anuales, de los 500 empleados de una empresa calcular el coeficiente de asimetría de Pearson.
$\bar{x} = 9.725$ $M_e = 9.672$ $S = 1.217,50$

55. Con los siguientes datos determine la simetría o asimetría de la distribución aplicando las diferentes fórmulas vistas.

a)

y_i	n_i
5	3
7	39
9	10
11	8
13	7
15	3
Σ	70
X_i	f_i

b)

y_i	n_i
5	3
7	7
9	8
11	9
13	30
15	3
Σ	60
X_i	f_i

c)

y_i	n_i
5	5
7	10
9	20
11	20
13	10
15	5
Σ	70
X_i	f_i

Haga los respectivos histogramas y dibuje los polígonos de frecuencias.

MEDIDAS DE APUNTAMIENTO : A_p

urtosis o estadígrafos de apuntamiento. Una característica importante de la variación en algunas distribuciones, es su grado de agudeza en la cima de la curva que las representa. Esta agudeza que se observa en la región del *Modo*, comparada con las condiciones halladas para el mismo sitio en la curva normal, es lo que se llama *urtosis o curtosis*.

Si la curva es más *plana* que la *normal*, la distribución se le considera *achatada* o *planticúrtica* y si es más aguda se le denomina *apuntada* o *leptocúrtica*. Si la curva es *normal* se dice que es *mesocúrtica.*

La *curtosis* es una medida de altura de la curva y por tanto está representada por el cuarto momento de la media. En la misma forma que para la asimetría, su cálculo se efectúa en función de la desviación típica y de los momentos unidimensionales de orden cuatro con respecto a la media aritmética:

$$A_p = \frac{m_4}{S^4}$$ Siendo 3 el valor de $\frac{m_4}{S^4}$ en la curva normal,

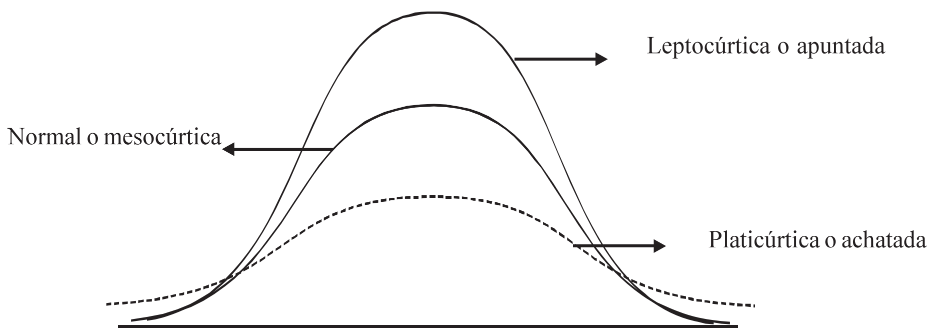
$$m_4 = \frac{\Sigma(y_i - \bar{y})^4 n_i}{n}$$

$$m_4 = \frac{\Sigma(X_i - \bar{X})^4 f_i}{N} = \frac{\Sigma d_i^4 f_i}{N}$$

Si $A_p = 3 \rightarrow$ la distribución es normal o mesocúrtica

$A_p > 3 \rightarrow$ la curva es leptocúrtica o apuntada

$A_p < 3 \rightarrow$ la curva es platicúrtica o achatada



Ejemplo 1. Calcular el grado de apuntamiento en la siguiente distribución:

y_i	n_i	$(y_i-\bar{y}_i)^3$	$(y_i-\bar{y}_i)^4 n_i$
2	4	–35,94	474,41
4	6	– 2,20	17,16
6	5	0,34	1,19
8	3	19,68	159,41
10	2	103,82	975,91
Σ	20	—	1.628,08
X_i	f_i	d_i^3	$d_i^4 f_i$

$$m_4 = \frac{\Sigma\,(y_i-\bar{y}_i)^4 n_i}{n} = \frac{1.628,08}{20} = 81,404$$

$$S^2 = \frac{\Sigma(y_i-\bar{y}_i)^2 n_i}{n} = 6,11; \quad S^4 = 37,33$$

$$\boxed{m_4 = \frac{\Sigma\, d_i^4 f_i}{N}} \qquad \boxed{S^2 = \frac{\Sigma\, d_i^2 f_i}{N}}$$

$$A_p = \frac{m_4}{S^4} = \frac{81,404}{37,33} = 2,18$$

$2,18 < 3$ la curva es achatada o platicúrtica.

Ejemplo 2. Con los siguientes datos:

$y'_{i-1} - y'_i$	n_i
10,1 – 16	4
16,1 – 22	8
22,1 – 28	13
28,1 – 34	10
34,1 – 40	9
40,1 – 46	6
Σ	50
$X'_{i-1} - X'_i$	f_i

Se pide calcular:

a) Los coeficientes de asimetría.
b) Los coeficientes de apuntamiento.

o :

a) $A_s = \dfrac{3(\bar{y} - M_e)}{S}$ $\qquad A = \dfrac{3(28{,}6 - 28{,}1)}{8{,}736} = 0{,}1717$ (asimetría positiva $A_s > 0$)

$\boxed{A_s = \dfrac{3(\bar{X} - M_e)}{S}}$

b) $A_s = \dfrac{m_4}{S^4}$ $\qquad m_4 = \dfrac{\Sigma(y_i - \bar{y})^4 n_i}{n}$ $\qquad \boxed{m_4 = \dfrac{\Sigma d_i^4 f_i}{n}}$

y_i	n_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(y_i - \bar{y})^3$	$(y_i - \bar{y})^4$	$(y_i - \bar{y})^4 n_i$
13	4	–15,6	243,36	–3.796,41	59.223,99	236.895,96
19	8	–9,6	92,16	–884,74	8.493,50	67.948,00
25	13	–3,6	12,96	– 46,66	167,98	2.183,74
31	10	2,4	5,76	13,82	33,17	331,70
37	9	8,4	70,56	592,70	4.978,68	44.808,12
43	6	14,4	207,36	2.985,98	42.998,11	257.988,66
Σ	50	—	—	—	—	610.156,18
X_i	f_i	d_i	d_i^2	d_i^3	d_i^4	$d_i^4 f_i$

$Z_i = y_i - \bar{y}$ $\qquad \boxed{d_i = X_i - X_i}$ $\qquad \bar{y} = \dfrac{1.430}{50} = 28{,}6 = \boxed{\bar{X}}$

$m_4 = \dfrac{\Sigma(y_i - \bar{y})^4 n_i}{n}$ $\qquad m_4 = \dfrac{610.156{,}18}{50} = 12.203{,}12$ $\qquad S^2 = 76{,}32$ $\qquad S^4 = 5.824{,}74$

$A_p = \dfrac{m_4}{S^4}$ $\qquad A_p = \dfrac{12.203{,}12}{5.824{,}74} = 2{,}09$ $\qquad A_p < 3 \longrightarrow 2{,}09 < 3$ achatada ó platicúrtica

EJERCICIOS MISCELÁNEOS

La gran mayoría de los ejercicios de este libro, se encuentran resueltos en el Sistema de Información en Línea SIL.

56 Con los siguientes datos:

28	42	64	32	68	25	34	40	51	62
36	56	76	40	70	28	38	46	58	64

Se pide:

a) Trabajando con los datos originales calcular la desviación típica, desviación mediana y el coeficiente de variación.

b) Agrupar los datos en una tabla de frecuencias, teniendo en cuenta que la amplitud es constante (la variable es continua) y $m = 6$, calcular el coeficiente de variación.

57. Suponga que se tiene otra distribución que presenta las siguientes medidas: $\bar{y} = 38{,}5$ y $S^2 = 36$
Comparando con los resultados del ejercicio anterior (No. 56), determinar:

a) ¿Cuál de las dos distribuciones presenta una menor variabilidad absoluta?
b) ¿Cuál de ellas presenta una mayor variabilidad relativa?
c) Si consideramos que la primera distribución presenta un valor de x = 48 y en la segunda (o sea la del punto c) el valor de x = 50, ¿cuál de estos valores presenta una mejor posición relativa?

58. a) Calcular el grado y la dirección de la asimetría con los datos de la tabla de frecuencias del punto 56 (b)
b) Calcular el grado de apuntamiento en esa distribución, determinando si es normal, achatada o apuntada.

59. Se tiene una distribución de 6 intervalos de amplitud constante, de la cual se sabe que:

$n = 150$ $n_3 = n_4 = 30$ $n_2 = n_5 = n_1 + 5$ $n_1 = n_6$ $y_5 = 50$ $\bar{y} = 41$

$N = 150$ $f_3 = f_4 = 30$ $f_2 = f_5 = f_1 + 5$ $f_1 = f_6$ $X_5 = 50$ $\bar{X} = 41$

a) Calcular el coeficiente de variación
b) Calcular el coeficiente de asimetría aplicando las diferentes fórmulas expuestas c) Calcular el coeficiente de apuntamiento.

60. Con los siguientes datos, se pide calcular:

$y'_{i-1} - y'_i$	n_i
6,1 – 12	3
12,1 – 15	10
15,1 – 20	6
20,1 – 28	4
28,1 – 36	8
36,1 – 40	16
40,1 – 50	3
Σ	50

a) Varianza
b) Desviación típica
c) Coeficiente de variación
d) Desviación media
e) Desviación mediana

61. Con los datos del punto 51 calcular: a) el grado de asimetría; b) grado de apuntamiento.

62. Se tiene dos distribuciones:
l) Corresponde a datos agrupados

$y'_{i-1} - y'_i$	n_i
2,1 – 6	22
6,1 – 10	14
10,1 – 14	10
14,1 – 18	8
18,1 – 22	4
22,1 – 26	2
$X'_{i-1} - X'_i$	F_i

2) Corresponde a datos sin agrupar

$x_1 = 12$ $x_3 = 6$ $x_5 = 23$ $x_7 = 4$ $x_9 = 6$
$x_2 = 18$ $x_4 = 20$ $x_6 = 17$ $x_8 = 10$ $x_{10} = 4$

Se pide determinar:
a) En cuál distribución se observa una mayor variabilidad absoluta?
b) En cuál distribución se observa una menor variabilidad relativa?
c) Si consideramos un valor de 18, en cuál de las distribuciones se tiene una mayor posición relativa?

63. Se pide calcular el coeficiente de variación para el total de las 70 observaciones (A+B) del ejercicio 62.
    a) ¿Qué pasaría con este coeficiente de variación, si multiplicamos a cada una de las 70 observaciones por 5?
    b) ¿Qué pasaría con este coeficiente de variación, si le sumamos 10 a cada una de las 70 observaciones?
64. Considerando únicamente la distribución (A) para datos agrupados: (Ejercicio 61)
    a) ¿La distribución es simétrica o asimétrica? ¿cuál es su grado de asimetría (si fuera la respuesta afirmativa) a qué lado se presenta?
    b) Calcule el grado de apuntamiento.

65. Los salarios mensuales que paga una fábrica a los operarios que trabajan en dos turnos, tiene las siguientes estadísticas:

MEDIDAS	TURNO I	TURNO II
Número de trabajadores	38	62
Salario medio mensual	978.000	1.203.500
Varianza	851.600	962.600

    a) Comparar los salarios en los dos turnos en cuanto a variabilidad absoluta y relativa.
    b) Obtener el coeficiente de variación para el salario de los 100 operarios.
    c) Cual será el coeficiente de variación para el año entrante si el gerente ofrece un aumento de $22.000 para los del primer turno y un 7% para los del segundo turno. El coeficiente de variación a calcular, se refiere al total de los 100 operarios.

66. Con los siguientes datos, correspondientes a una distribución de frecuencias, se pide calcular el grado de simetría o asimetría, indicando hacia qué lado, lo mismo que el grado de apuntamiento. Además calcular:

$y'_{i-1} - y'_i$	n_i
2,1 – 6	3
6,1 – 10	12
10,1 – 14	25
14,1 – 18	11
18,1 – 22	7
22,1 – 26	2
Σ	60
$X'_{i-1} - X'_i$	F_i

    a) El coeficiente de variación
    b) El puntaje típico, si se tiene queun valor de $y = 12$
    c) La desviación media y mediana

67. A los trabajadores de una empresa, el próximo año, el salario mensual les será aumentado en un 4,2% más $8.000. La empresa tiene 600 trabajadores y actualmente devengan un salario medio mensual de $810.000, con un coeficiente de variación de 0,36.
    a) ¿Cuál será la varianza de los salarios el próximo año?
    b) El próximo año, ¿cuál será el valor total de la nómina mensual?

68. Con los siguientes datos no agrupados: x_i: 6 4 8 2 10, calcular:
    a) Desviación media b) Desviación mediana

c) Desviación típica   d) Comprobar que $D_e \leq D_a < S$

e) Calcular el coeficiente de variación

69. Con la siguiente información: $x = \frac{6}{8}[16\text{ y } +16]$ $CV_y = 0{,}40$ $\overline{y} = 5$

Hallar el coeficiente de variación de x

70. Sabiendo que $y = 6 - 10x$ $V_{[x]} = 8$.   Hallar la varianza de $4x - y$

71. ¿Por qué es cero el coeficiente de asimetría para una distribución simétrica?

72. Dos obreros del mismo y trabajo muestran los siguientes resultados en un período determinado

MEDIDAS	OBREROS	
	A	B
Tiempo promedio para el desarrollo de su trabajo	42	35 minutos
Desviación típica	8	6 minutos

a) ¿Cuál grupo es el más regular en el desarrollo de su trabajo?

b) ¿Cuál grupo es el más rápido en terminar el trabajo?

73. Un estudiante obtuvo las siguientes notas en 4 de sus materias así:

Economía = 4   Matemática = 3 ,3   Inglés = 3 ,6   y   Derecho = 4 ,2

a) ¿Qué clase de opinión podríamos dar con respecto a esas notas?

Recurrimos a los profesores que en un análisis han podido reunir más información sobre el rendimiento del curso donde se encuentra el estudiante.

Promedio del curso: Economía = 4,3   Matemática = 2,8   Inglés = 3,2 y Derecho = 4,6. Además, conocen las desviaciones típicas, siendo: 0,6   0,75   0,8   0,6 respectivamente.

b) ¿Su opinión se mantiene igual respecto al rendimiento de este estudiante?

74. Supongamos que un acucioso empleado de la empresa de Acueducto de la ciudad, realiza una muestra de 60 usuarios del servicio, sobre los reclamos en los 2 últimos años por esas personas, con los siguientes resultados:

NÚMERO DE RECLAMACIONES	NÚMERO DE USUARIOS
0	26
1	10
2	8
3	6
4	4
5	3
6	2
7	1

Se pide hallar:

a) El promedio de reclamos

b) La varianza y su desviación típica

c) El coeficiente de variación

Síntesis de capítulo

En el resumen del capítulo anterior, se decía que no basta con calcular el *promedio* para describir el comportamiento de un conjunto de datos, sino que además, en numerosos casos, requiere de la aplicación de alguna de las medidas de dispersión, las que nos determinan la extensión que ocupan los valores de la variable respecto al valor promedio elegido.

Supongamos que dos personas conocedoras de estadística se dedican a lanzar dardos y en el primer lanzamiento ambos lo colocan el tercer círculo uno arriba y el otro abajo, es de imaginar que ambos se sientan satisfechos del lanzamiento, pues lograron la misma distancia con respecto al centro. Se puede sacar como conclusión que el punto central fue para ellos menos importante y la atención se centró más en la distancia o dispersión que a dar en el blanco o punto central.

El cálculo de un promedio para dos grupos puede ser igual, como por ejemplo el promedio de calificación de dos cursos de estadística con diferentes profesores, en ambos se obtuvo 3,7, pero se observa que en el primero de ellos la calificación mínima fue de 1,1 y la máxima de 4,3, en cambio con el segundo profesor la mínima fue de 3,4 y la máxima de 4,2 ¿qué profesor le recomendaría a su hermano si hubiese necesidad de elegir uno de ellos? En este ejemplo se le da importancia a la dispersión; como se puede observar, no siempre el promedio por sí mismo nos proporciona una suficiente información.

Son varias las medidas o varios los procedimientos que nos indican el mayor o menor grado de dispersión, siendo la más conocida la *desviación estándar* o *desviación típica*. La varianza es el procedimiento que nos permite llegar a la desviación típica, con el inconveniente que no se expresa en la misma unidad de medida de la variable, ya que siempre corresponde al cuadrado de ella, así se tendrá: metros, kilos, horas al cuadrado, en cambio la desviación estándar y el recorrido, se expresan en las mismas unidades. Cuando el coeficiente de variación se expresa en porcentajes, permite realizar mejor la comparación de la dispersión entre dos o más distribuciones, cuando las variables están expresadas en unidades de medidas diferentes.

Decimos que una *distribución es simétrica*, cuando la *media*, *mediana* y la *moda* calculada en una variable son iguales. El grado de *simetría* o *asimetría* en una distribución se obtiene aplicando la medida denominada coeficiente de asimetría o simplemente observando el histograma de frecuencias. El problema se presenta cuando la distribución tiene forma de U, siendo considerada no normal, ya que no tiene la forma de campana de Gauss que siempre utilizamos para indicar una distribución simétrica.

Finalmente recordemos que los valores observados, si bien es cierto se repiten, en conjunto son diferentes si las comparamos entre sí, por lo tanto esta dispersión no puede pasar desapercibida y su información radica cuando tiende a agruparse alrededor del promedio, de ahí que no solo calculamos estos últimos, pero además, debemos acompañarlo con técnicas que nos permitan estudiar la variación, cuando encontramos diferentes distribuciones con diversos grados de dispersión. Vale la pena resaltar cómo en la fabricación de artículos de precisión, una pequeña variación puede influir en su mal funcionamiento.