Elementos básicos de Estadística para Ciencias Sociales

A. Pierdant y J. Rodríguez

Colección Docencia y Metodología

UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD XOCHIMILCO División de Ciencias Sociales y Humanidades

ELEMENTOS BÁSICOS DE ESTADÍSTICA PARA CIENCIAS SOCIALES

Elementos básicos de estadística para ciencias sociales

Alberto Isaac Pierdant Rodríguez
Jesús Rodríguez Franco

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Casa abierta al tiempo UNIDAD XOCHIMILCO División de Ciencias Sociales y Humanidades

Casa abierta al tiempo

Primera edición: diciembre de 2006

UAM-Xochimilco
Calzada del Hueso 1100
Col. Villa Quietud, Coyoacán
C.P. 04960 México, DF.

ISBN 970-31-0731-1

Impreso y hecho en México / *Printed and made in Mexico*

Índice

Anexo II

Introducción

El avance acelerado de la sociedad moderna y del conocimiento, sumado a la aparición de herramientas como la computadora, la fibra óptica, la internet, la biogenética, así como de nuevas ramas de conocimiento, han obligado a los individuos a participar en un proceso intenso de asimilación de todo cambio tecnológico que se produzca en la sociedad. Uno de éstos es el uso intensivo y común de las computadoras y la estadística en todos los ámbitos. Precisamente es el ámbito de las ciencias sociales el que mayor participación, uso y explotación le ha dado a estas herramientas, aunque, justo es decirlo, no siempre en forma adecuada. Este texto propone un análisis, combinación y utilización de estas herramientas, con el objeto de dar al lector una clara comprensión de los fenómenos sociales y, al mismo tiempo, satisfacer las crecientes exigencias de la sociedad.

Si bien es cierto que el profesional de las ciencias sociales cuenta actualmente con una gran variedad de herramientas, la estadística y las computadoras son elementos que le han permitido ampliar aún más su horizonte de trabajo. Sin embargo, el especialista social se hace continuamente la siguiente pregunta, ¿cómo puedo obtener de estas herramientas una mayor ayuda en tiempos donde la velocidad de respuesta a los problemas sociales requiere día a día de una respuesta certera y rápida? Una respuesta fácil a esta problemática planteada, implicará el uso de la estadística e informática, con un mejor nivel de conocimiento de ellas, y una instrumentación más pragmática en el ámbito de las ciencias sociales. El presente libro, por lo tanto, pretende mostrar ideas al profesional de las ciencias sociales que le permitan combinar estas herramientas proporcionándole la rapidez de respuesta que hoy la sociedad le está requiriendo.

¿Qué es la estadística?

CON EL HOMBRE nacen prácticamente las matemáticas y la estadística. La necesidad de contar dio origen a la primera, de la cual surgió, de manera intuitiva, la segunda.

La estadística es una ciencia relativamente nueva que tiene por objeto la colección e interpretación de datos. Estas actividades se remontan a la época del Viejo Testamento y los registros que ya elaboraban los babilonios y los romanos sobre la población.

Los estudios de la población, en sentido estrictamente científico, aparecen a finales del siglo XVIII y principios del XIX. Godofredo Achewald (1719-1772) introduce por primera vez el nombre de "estadística", especificando así a la asignatura universitaria encargada de la descripción de las cosas del Estado. Pero fue Adolph Quetelet (1796-1874) quien aplicó, también por primera vez, métodos modernos al estudio de un conjunto de datos.

Aunque formalmente la palabra estadística surge en el tiempo a partir de la interpretación de tres vocablos: *status*, *statera* y *staat*[1] –como la aritmética estatal que permitía asistir al gobernante en el conteo de la riqueza y número de súbditos con el objeto de recaudar impuestos o presupuestar la guerra–, hoy esta ciencia moderna puede interpretarse[2] como:

> Un método que permite organizar, sintetizar, presentar, analizar, cuantificar e interpretar grandes volúmenes de datos, de tal manera que se puedan obtener conclusiones válidas (dar información) sobre los fenómenos en estudio.

Este amplio concepto ha llevado a los especialistas en estadística a un acuerdo, que permite clasificar a esta materia en dos grandes ramas, definidas como

[1] *Status* (latín): situación, posición, estado. *Statera* (griego): balanza, ya que la estadística mide o pesa hechos. *Staat* (alemán): se refiere al Estado como expresión de unidad política superior.

[2] Véase C. Meza, A. Morales y R. Magaña, *Introducción al método estadístico*, México, UAM-Xochimilco, 1980.

estadística descriptiva y *estadística inferencial.* Ambas desempeñan funciones distintas pero complementarias en el análisis estadístico.

Estadística descriptiva

Es aquella rama de la estadística que trata del resumen y descripción de los datos. Este resumen puede ser tabular, gráfico o numérico. Su análisis y descripción se limita exclusivamente a los datos coleccionados, es decir, a los datos que forman una muestra, por lo que ésta, no puede inferir o generalizar acerca de la totalidad de los elementos que constituyen la población de estudio.

Estadística inferencial

Esta otra rama de la estadística tiene como objetivo generalizar o inferir conclusiones útiles sobre la totalidad de las observaciones (la población) a partir del análisis de los datos coleccionados (la muestra). La inferencia estadística constituye la base teórica del muestreo, es decir, permite conocer el todo con cierta aproximación a partir del estudio de una parte.

Es importante mencionar que la descripción de un todo o población con base en los principios de la estadística inferencial no da una certeza completa en sus medidas sumarias, ya que éstas están sujetas a un posible error, a causa de que las unidades seleccionadas (muestra), más o menos numerosas, no hayan sido seleccionadas de acuerdo con ciertos procedimientos y que la variabilidad de las características de estudio sea más o menos grande.

En las anteriores definiciones hemos utilizado una serie de conceptos muy comunes en estadística; es conveniente especificarlos para que ello permita dar un mayor entendimiento a las mismas.

Conceptos empleados en estadística

- *Dato*: es un número o una medida que ha sido recopilada como resultado de una observación. Los datos pueden ser producto de un conteo, una medición o una denominación; por ejemplo: número de personas en una población, número de empresas en un país, el peso, género y nombre de una persona, etcétera.

- *Variable*: para obtener estadísticas manejamos conjuntos que poseen un determinado o indeterminado número de unidades (personas, objetos, etcétera). Las unidades de estudio tienen determinadas características; por ejemplo, para un ciudadano mexicano, podríamos señalar: género, edad, estatura, peso, lugar de nacimiento, estrato social, grado de escolaridad, religión, estado civil, etcétera; todas y cada una de estas característica, que adquieren diferentes valores en cada persona, lugar o cosa y que son susceptibles de una medición, reciben el nombre de variables. De esta manera, el estudio de los habitantes de una población, requeriría, tal vez, del uso de variables como género, edad, estatura, peso, estrato social, ingreso, religión, etcétera. Como puede observarse, la variable es una construcción que el investigador genera para analizar una realidad.
- *Población*: es el conjunto formado por un número determinado o indeterminado de unidades (personas, objetos, etcétera) que comparten características comunes a un objeto de estudio. Por ejemplo, en un estudio de las preferencias de los votantes en una elección presidencial, la población estaría formada por todas las personas registradas en un padrón electoral.
- *Muestra*: es cualquier subconjunto seleccionado de una población, que sigue ciertos criterios establecidos en la teoría del muestreo. La muestra es el elemento básico sobre el cual se fundamenta la posterior inferencia acerca de la población de donde procede.
- *Parámetro*: es la medida estadística que cuantifica una característica que ha sido estudiada para una población. Este valor estadístico se considera verdadero, ya que su origen parte del estudio de cada uno de los datos que constituyen a la población.
- *Estadígrafo o estadística*: es la medida que cuantifica una característica estudiada en una muestra. Por ejemplo, si tenemos una muestra que enumera a 100 posibles clientes de un producto, y se les pregunta su opinión sobre el mismo, si 70 de éstos lo prefieren, tendremos entonces, que la proporción muestral será 0.70, y ésta será, en consecuencia, una estadística.

¿Por qué se estudia estadística?

LAS DEFINICIONES ANTERIORES nos han dado una idea general de lo que realiza la estadística como rama de las matemáticas aplicadas, sin embargo, este conjunto de técnicas que tienen aplicación en las más diversas disciplinas no siempre son aceptadas por los individuos encargados de aplicarlas. Por supuesto, esta última aseveración dependerá del conocimiento adecuado que sobre este herramental estadístico tenga el usuario. Así, los individuos que requieren elaborar estadísticas se preguntan, en muchas ocasiones, por qué se estudian estas técnicas. Como indica F. Holguín:

> La estadística proporciona los elementos básicos para fundamentar, en una investigación,
>
> - cómo planear la obtención de los datos para que de ellos se puedan extraer conclusiones confiables
> - cómo analizar estos datos
> - qué tipo de conclusiones pueden obtenerse con los datos disponibles
> - cuál es la confianza que nos merecen los datos.[3]

La estadística, como se puede observar, por medio de sus dos ramas, nos permite realizar estudios de tipo descriptivo y explicativo prácticamente en todas las áreas del conocimiento humano.

Pero, ¿cómo podemos aplicar estas técnicas a una investigación? La respuesta es simple: siguiendo un método. Son muchos los autores que sugieren un método para la investigación estadística. A continuación propondremos un resumen muy sencillo elaborado con varios de estos métodos. La propuesta que aquí hacemos tiene como objetivo facilitar la aplicación de la estadística en nuestras actividades, y para ello incluimos herramientas que permiten simplificar las tareas requeridas, siendo una de ellas la moderna computadora.

[3] Véase Fernando Holguín Quiñones, *Estadística descriptiva aplicada a las ciencias sociales*, México, UNAM, 1981, "El concepto de estadística en investigación", pp. 22 y s.

ELEMENTOS PARA REALIZAR UNA INVESTIGACIÓN USANDO ESTADÍSTICA

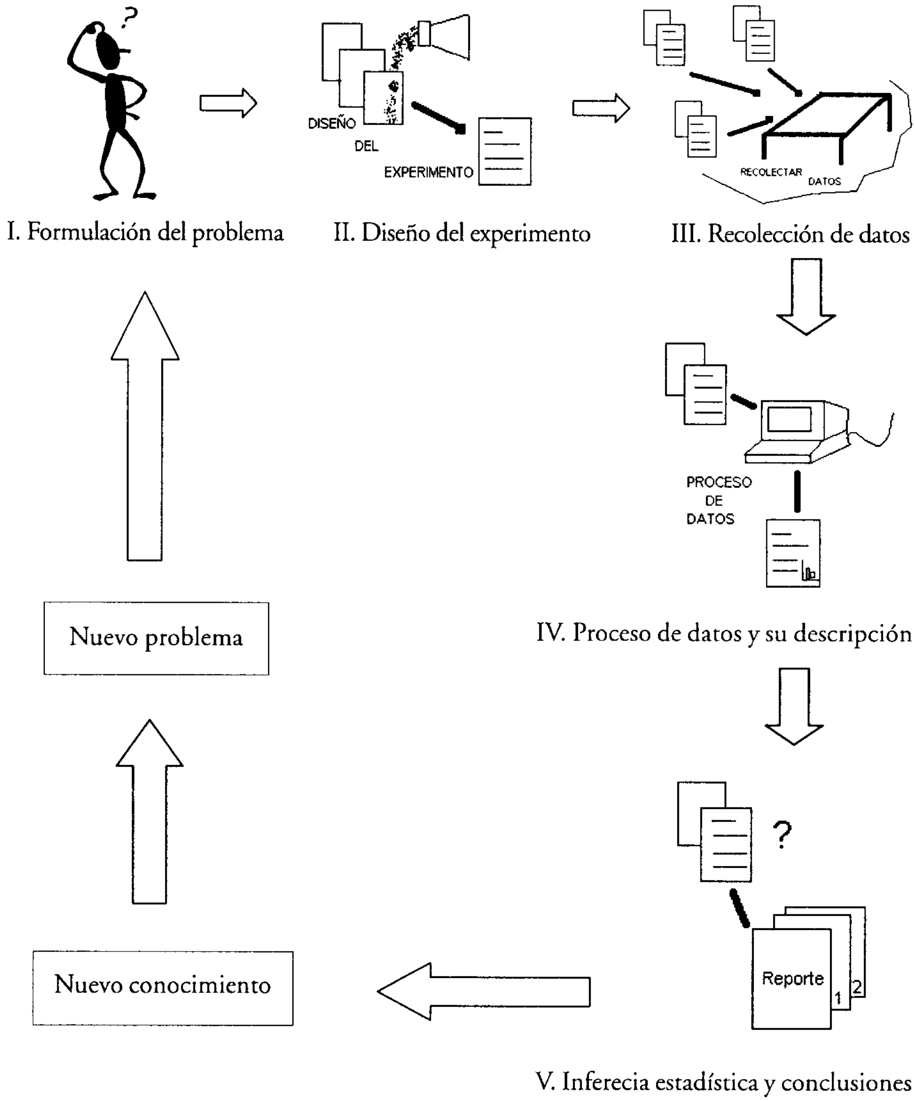
- *Formulación del problema.* Consiste en la identificación y especificación adecuada de un problema de investigación. En esta etapa es importante establecer con precisión la o las hipótesis, el o los objetivos del estudio, su alcance, y la población de datos asociada al mismo.

- *Diseño del experimento.* En esta segunda etapa, el investigador deberá seleccionar la técnica de recolección de datos (observación directa, entrevistas, encuesta, investigación documental) que le permitan obtener la información a un mínimo costo (dinero y tiempo) posible. En esta etapa, el investigador deberá definir el tamaño de la muestra, la calidad requerida y el tipo de datos que le permitan resolver el problema planteado de la manera más eficiente.

- *Recolección de datos.* Es, tal vez, la etapa de mayor importancia en la investigación, ya que la calidad de los datos obtenidos depende de una buena recolección, por lo tanto ésta deberá sujetarse a reglas estrictas que permitan obtener la información deseada.

- *Proceso de datos y su descripción.* Consiste en la elaboración de cuadros estadísticos de trabajo, cuadros estadísticos de referencia, gráficas y cálculo de medidas estadísticas apropiadas al proceso descriptivo o inferencial seleccionado. Esto es, se exponen los datos muestrales mediante representaciones tabulares, gráficas y medidas estadísticas con el objeto de hacer una descripción de los resultados.

- *Inferencia estadística y conclusiones.* Esta etapa proporciona una contribución muy importante, ya que en ella se define el nivel de confianza y significancia del proceso inferencial, lo cual sirve como orientación a quien o quienes deben tomar una decisión sobre el tema objeto de estudio. Esto último permite al investigador establecer una conclusión sobre el problema y, en algunas ocasiones, elaborar sugerencias para la solución del mismo.



Variables

Como ya ha sido indicado, la primer etapa en el desarrollo de una investigación estadística comprende la formulación del problema, en la cual debe señalarse la hipótesis o las hipótesis de investigación, y por lo tanto, la variable o variables de estudio que le posibiliten comprobarlas, es decir, el investigador debe contar con elementos que le permitan probar sus hipótesis. En estadística, estos elementos los constituyen las variables de la investigación. Reiteramos: una variable es cada una de estas característica (cualidad, rasgo, atributo o propiedad) que toma diferentes valores en cada persona, lugar o cosa y que es susceptible de una medición. De esta forma, en estadística, las variables se pueden clasificar en dos grandes grupos, definidos como variables cuantitativas o métricas y variables cualitativas o no métricas.

Las *variables cuantitativas o métricas* son aquellas cuya determinación está asociada a una unidad de medida; por ejemplo, la estatura de una persona, el número de habitantes de una población, el ingreso mensual de los individuos de un país, etcétera. Estas variables se subclasifican a su vez en variables discretas y continuas.

Las *variables discretas o discontinuas* son aquellas que cuantifican la característica por medio de valores enteros y nunca mediante fracciones de los mismos. Como ejemplo de este tipo de variables tenemos, el número de clientes de un banco, el número de hijos en una familia, el número de alumnos en un grupo de la universidad, el número de personas en una población rural, el número de automóviles en una ciudad, etcétera.

Las *variables continuas* son aquellas que pueden tomar cualquier valor numérico, es decir, un valor entero o fraccionario en un intervalo previamente especificado. Así, por ejemplo, la variable tiempo en una investigación podría medirse en intervalos de horas, o bien en horas y minutos, o bien en horas, minutos y segundos, según sea el requerimiento de la misma.

Las *variables cualitativas o no métricas* especifican y miden cualidades en los individuos, lugares o cosas a partir de su descripción con palabras. Ejemplos de ellas son: la variable género (hombre, mujer), la variable religión (católica, protestante, etcétera), la variable idioma (español, inglés, francés, etcétera), la variable estatus social (alto, medio, bajo), la variable calidad de un servicio (bueno, regular, malo), etcétera. Al igual que las variables cuantitativas, pueden subclasificarse en variables cualitativas *nominales* y *ordinales*.

Variables nominales. Son aquellas variables no métricas usadas para describir una característica que no puede ser cuantificada numéricamente; por ejemplo, el nombre de una persona, el género de los individuos en un grupo universitario, su idioma, religión, la licenciatura que cursan, etcétera. En algunos casos, a estas variables se les asignan números de acuerdo con una regla específica, pero estos números sólo se emplean para diferenciar a los distintos objetos o categorías. Así, la variable género podría clasificarse mediante dos categorías: 1 para los hombres y 2 para las mujeres. De esta manera, en una investigación todos los individuos 1 serán hombres y todos los 2 serán mujeres. Esta numeración en las variables nominales no permite ninguna operación aritmética o algebraica.

Variables ordinales. Son aquellas variables no métricas que permiten describir la característica de una persona, objeto o lugar, a partir de diferenciarla en diversas categorías establecidas en orden de supremacía de acuerdo con un criterio jerárquico. La diferencia que se establece entre las categorías ordinales no tiene significado cuantitativo, sólo expresa que una situación es mejor que otra, pero no cuantifica (la valuación es subjetiva). Un ejemplo de ello es la evaluación que un cliente puede realizar respecto del servicio en un restaurante. El servicio fue malo, regular, bueno o muy bueno. Como se puede observar, la variable mide diversas categorías que han sido establecidas de acuerdo con un criterio pero no puede establecer en ellas una cuantificación numérica, ya que la evaluación realizada por el cliente es subjetiva.

La clasificación de las variables anteriormente expuesta, que parte del punto de vista de la estadística, no es única, ya que cada disciplina científica crea alguna denominación para las variables que en ella se manejan comúnmente. Así, lo que en el momento de la operacionalización es una variable nominal, ordinal, intervalar o de razón, en el momento de su representación gráfica puede ser una variable discreta o continua, y al realizar su análisis puede ser dependiente o independiente. Por ejemplo, en las ciencias sociales es común establecer relaciones entre variables experimentales, las cuales se clasifican, desde el punto de vista metodológico, en variables *dependientes* y variables *independientes*.

La *variable dependiente* es aquella cuyos valores están condicionados por la variable o las variables independientes con las que tiene relación. Por lo tanto, la variable o las variables independientes son la causa iniciadora de la acción, es decir, condicionan de acuerdo con sus valores a la variable dependiente. Proponemos un ejemplo de tipo económico que nos permitirá establecer una relación con este tipo de variables. Consideramos el comportamiento del ahorro de un individuo en una sociedad. El modelo económico que explica su ahorro podría ser:

$$\text{Ahorro} = \text{Ingreso} - \text{Gasto}$$

AHORRO ($) =	INGRESO ($)	- GASTO ($)	OBSERVACIÓN
900	1000	100	Ahorro alto
200	1000	800	Ahorro bajo
0	1000	1000	No hay ahorro
-200	1000	1200	No hay ahorro. Déficit cubierto con crédito

En este modelo, el ahorro, la variable dependiente, presentará una situación específica de acuerdo con el comportamiento que tengan las variables independientes de la relación. En el cuadro anterior se muestran cuatro escenarios del ahorro con base en el comportamiento del ingreso y gasto.

Independientemente del área de la ciencia y de la terminología que en ella se utilice para describir a las variables, éstas son susceptibles de clasificarse según la taxonomía establecida para su manejo estadístico. Sin embargo, no sólo es importante identificarlas y clasificarlas, sino también deberán definirse adecuadamente a partir del criterio establecido por el investigador. Por lo tanto, esta es la última etapa en la definición de las variables que serán usadas en una investigación, y de acuerdo con Hernández, Fernández y Baptista,[4] su definición deberá establecerse en dos niveles, especificados como nivel conceptual y nivel operacional.

Nivel conceptual. Consiste en definir el término o variable con otros términos. Por ejemplo, el término "poder" podría ser definido como "influir más en los demás que lo que éstos influyen en uno". Este tipo de definición es necesaria pero insuficiente para definir una variable, debido a que no nos relaciona directamente con la realidad, ya que, como se puede observar, siguen siendo conceptos.

Nivel operacional. Constituye el conjunto de procedimientos que describen las actividades que un observador realiza para recibir las impresiones sensoriales que indican la existencia de un concepto teórico (conceptual) en mayor o menor grado, es decir, consiste en especificar las actividades u operaciones necesarias que se deben realizar para medir una variable.

[4] Véase S.R., Hernández, *et al*, *Metodología de la investigación*, México, McGraw-Hill, 1991, pp. 99-102.

VARIABLE = AUSENTISMO LABORAL

Nivel conceptual: "El grado en el cual un trabajador no se reportó a trabajar a la hora en la que estaba programado para hacerlo".

Nivel operacional: "Revisión de las tarjetas de asistencia al trabajo durante el último mes".

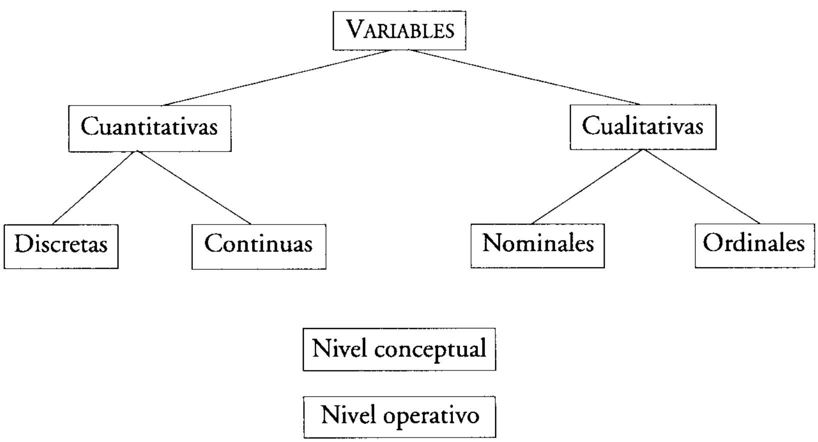
VARIABLE = GÉNERO

Nivel conceptual: "Condición orgánica que distingue al macho de la hembra".
Nivel operacional: "Asignación de la condición orgánica: hombre o mujer".

Con estas dos definiciones el estudiante o investigador estará ahora en posibilidad de acotar adecuadamente las variables para un manejo estadístico de acuerdo con el interés que se tiene en ellas, así sea la realización de un estudio o una investigación. Un par de ejemplos de ello, se mostraron con anterioridad.

A continuación se muestra de manera gráfica la clasificación de las variables desde el punto de vista de la estadística, así como la definición conceptual y operacional que el investigador debe proponer de ellas.

En esta etapa, se recomienda al investigador o estudiante especificar las variables que habrá que utilizar en su investigación, no únicamente por su tipo desde el punto de vista estadístico, sino principalmente definir conceptual y operacionalmente cada una de ellas, de tal forma que éstas queden perfectamente detalladas en la investigación que se está realizando. Con esto último ya definido, el investigador puede pasar a la etapa de recolección de datos, lo que posteriormente le permitirá llegar al proceso de datos y su descripción, tema del siguiente capítulo.



Escalas de medición

Una vez que hemos especificado por un lado las variables que se manejarán en una investigación, y por otro, su descripción a nivel conceptual, es muy conveniente también establecer, en conjunto con el nivel operacional, la llamada escala de medición, la cual nos permite definir con precisión la forma en la que el investigador medirá en la práctica sus variables.

Los niveles o escalas de medición están definidos mediante cuatro tipos generales: la escala nominal, la escala ordinal, la escala de intervalos y la escala de razón.

Escala nominal

Es el tipo más limitado de medición que puede tener una variable. Se emplea para hacer referencia a los datos que sólo pueden clasificarse en categorías, es decir, se aplica a aquellas variables que no pueden medirse mediante escalas numéricas, sino únicamente a partir de contar cada una de las características (se realiza un conteo de datos).

Esto último nos indica que en este nivel de medición no existe un orden particular para organizar a los distintos grupos que forman los valores de la variable. Estas categorías o grupos son mutuamente excluyentes, es decir, que un dato al clasificarse sólo puede pertenecer a un grupo de la clasificación. Por ejemplo, al preguntársele a un alumno de la universidad la licenciatura que está cursando, éste sólo puede responder administración (u otra), pero no puede respondernos que cursa administración y medicina. En el caso de clasificar a un automóvil por marca, no puede ser al mismo tiempo, chevrolet y renault, clasificándose sólo en una.

En este tipo de escalas, los grupos o categorías deben ser exhaustivos, lo cual significa que todos los datos de la población o muestra a clasificar pueden ser incluidos en su grupo o categoría respectiva. Así por ejemplo, los alumnos de la División de Ciencias Sociales y Humanidades en la UAM-Xochimilco, deberán clasificarse para las licenciaturas de administración, economía, psicología, comunicación, sociología, política y gestión social, como se muestra en el cuadro estadístico A.

CUADRO A
División de Ciencias Sociales y Humanidades, alumnos por licenciatura

Licenciatura	Alumnos
Administración	425
Economía	350
Comunicación	475
Sociología	370
Psicología	320
Política y Gestión Social	140
Total	2 080

Escala ordinal

Esta escala presenta diferentes niveles de medida entre sus categorías, una mayor que otra, de tal forma que todas tienen diferente valor subjetivo. Esta medida diferente tiene dos características importantes, la primera consiste en, como ya indicamos, que el valor que toma la variable es subjetivo, y la segunda, es aquella que obliga a clasificarla en las categorías establecidas, en un orden específico. Por ejemplo, suponga que un investigador desea medir la calidad del servicio que presta el transporte público Metro de la Ciudad de México, y para ello pregunta a una muestra de individuos lo siguiente:

¿Cómo considera usted la calidad del servicio del Metro de la Ciudad de México?

Mala   Regular   Buena   Muy Buena

La clasificación de sus respuestas seguirá un orden de acuerdo con el valor que cada individuo le asigne en la muestra; así por ejemplo, el investigador podría mostrar sus resultados a la pregunta como los indicados en el cuadro B o bien como los del cuadro C.

CUADRO B

Calidad del servicio	Respuestas
Muy Bueno	50
Bueno	150
Regular	35
Malo	15

CUADRO C

Calidad del servicio	Respuestas
Malo	15
Regular	35
Bueno	150
Muy Bueno	50

La escala ordinal de medición presenta al igual que la escala nominal, las mismas características, es decir, sus categorías o grupos de clasificación son mutuamente excluyentes y exhaustivos.

Escala de intervalo

Esta escala de medición presenta las mismas características básicas que tiene la escala ordinal, salvo que en ésta es posible establecer valores numéricos constantes en las diversas categorías, y por ello establecer medidas o cuantificaciones entre unas y otras. Un ejemplo de este tipo de escala es la medición de la temperatura ambiente; una temperatura normal será para un determinado lugar 23° centígrados, pero bien podría estar en un intervalo entre 25° y 28°. Observe que el cero (0°) en este tipo de escala es arbitrario, ya que 0° en el ambiente no significa que no haya temperatura, sino sólo un determinado estado de frío.

Otro ejemplo de escala de intervalo nos lo muestra el cuadro que califica la habilidad de los operarios para el armado de un componente electrónico en una empresa del valle de Toluca (cuadro D).

CUADRO D
Calificación de habilidad

Puntuaciones	Obreros
91 - 100	30
81 - 90	25
71 - 80	15
61 - 70	13
50 - 60	10
menos de 50	4

En este ejemplo, un operario puede tener una habilidad de 75, otro 85 y otro de 95, la diferencia de habilidad entre uno y otro es cuantificable (10 unidades). También podemos observar que un valor cero no tendría significado, ya que finalmente todos los operarios tienen algún grado de habilidad.

En esta escala se dice que el valor cero es arbitrario.

Escala de razón

Es el nivel de medición más alto. Esta escala tiene todas las características que presenta la escala de intervalo, es decir, las categorías se especifican con números. El tamaño de éstas es conocido y constante; son también mutuamente excluyentes y exhaustivas. Su gran diferencia con respecto a la escala de intervalo es, por un lado, que en la escala de razón el punto cero sí es significativo, y por otro, el cociente o razón entre dos números de la escala también lo es.

Como ejemplo considere la variable siguiente: gasto diario en transporte de un alumno de la universidad.

CUADRO E
Gasto diario en transporte

Gasto ($)	Alumnos
0 - 10	30
11 - 20	25
21 - 30	5
31 - 40	3

El alumno puede tener un gasto de $7.50 al día, pero también podría suceder que tenga un gasto de $0.00, lo que nos indicaría que, quizá, el alumno camina diariamente a la universidad. Es decir, el cero, tiene significado.

Considere ahora un alumno que gasta $20.00 diarios en transporte y otro que gasta sólo $10.00. El primero gasta el doble que el segundo, es decir, el cociente entre ambas cantidades tiene un significado al realizar un análisis.

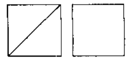
El proceso de datos y su descripción

UNA VEZ QUE EL INVESTIGADOR ha diseñado su experimento, recogido los datos del estudio mediante un instrumento de recolección (cuestionario, entrevista, observación, etcétera), la siguiente etapa del proceso estadístico consistirá en agruparlos y procesarlos.

La etapa de agrupación de datos tiene como objetivo el condensarlos en una primera fase antes de obtener las medidas estadísticas que los sumaricen o condensen aún más. Esta primera agrupación, que en estadística se define como presentación tabular, consiste en agruparlos mediante la elaboración de cuadros estadísticos o tablas estadísticas.

Para agrupar datos deberá realizarse, primeramente, la definición de cuáles y cuántos grupos o clases se tienen en cada variable de la investigación; para ello, utilizaremos la definición operacional de las variables. Después se procederá al conteo y clasificación de cada dato en los grupos o clases determinados previamente. Así, por ejemplo, en la característica género, se establecen primeramente dos clases o categorías, definidas como hombre y mujer; una vez hecho esto se deberá contar y clasificar a cada hombre o mujer entrevistada en el estudio. Para tener esta primera agrupación, el investigador o estudiante cuenta actualmente con la tabulación manual o bien el proceso electrónico de datos.

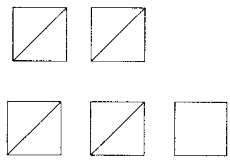
Tabulación manual. Consiste en hacer el recuento mediante el uso exclusivo de papel y lápiz. Se realiza el conteo en hojas tabulares, registrando el dato en la clase correspondiente conforme éste aparece en la fuente. Para ello pueden emplearse dos sistemas: el de diagonales o rayas y el de cuadrados.

- DIAGONALES (RAYAS)  5 UNIDADES
- CUADROS  9 UNIDADES

La experiencia demuestra que el sistema de cuadros presenta un menor número de errores en el conteo, ya que es más fácil percibir cuando se han completado 5 unidades; mientras que el sistema de diagonales puede dar lugar a errores en los que se incluye un mayor o un menor número de unidades en lugar de las 5.

Un ejemplo de esta tabulación se muestra a partir de la clasificación por género de 24 personas en la forma siguiente:

- HOMBRES: 10
- MUJERES: 14



La tabulación manual puede efectuarse además mediante el traslado de los datos a hojas o cuadros de concentración. Estas hojas o cuadros permiten tener a la vista los datos que se encuentran dispersos en los cuestionarios o tarjetas del investigador; con ellas es más simple realizar su conteo. El registro de estos datos se simplifica por medio de señales elementales: (.) (X) (✓); y los totales se muestran al pie de las columnas como se puede apreciar en el siguiente cuadro de concentración.

CUADRO 1

	Sexo		Nacionalidad		Carrera que cursa				Trabaja	
Cuest.	M	F	Mex.	Ext.	Eco.	Soc.	Com.	Adm.	Sí	No
1		X	X		X					X
2	X		X			X				X
3		X		X				X		X
4		X		X				X	X	
....										
Total	10	10	18	2	5	5	5	5	5	15

La tabulación manual es recomendable cuando el número de cuestionarios o datos recolectados no es muy grande y las categorías o clases especificadas sean poco numerosas. En estas condiciones, la tabulación manual suele ser más ventajosa que la tabulación de procesamiento electrónico.

Proceso electrónico de datos. Es recomendado en aquellas investigaciones que requieren de operaciones a gran escala, que son complicadas y con alto grado de repetición. Este sistema es preferido al manual por la gran velocidad y eficiencia con la que la computadora procesa los datos; sin embargo, a pesar de estas ventajas, se requiere de una cantidad relativamente grande de trabajo preliminar en la elaboración del programa de computadora así como en la captura de los datos para su proceso. En la actualidad, el investigador cuenta con una gran variedad de programas de computadora (paquetes) que le permiten procesar los datos eficientemente, entre algunos de ellos se tiene SPSS (Statistical Package for Social Science), SPSS for Windows, SAS (Statistical Analysis System), MINITAB, etcétera. A continuación se muestra un cuadro de concentración elaborado con el paquete SPSS.

¿TRABAJA ACTUALMENTE?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Sí	4	80.0	80.0	80.0
	No	1	20.0	20.0	100.0
	Total	5	100.0	100.0	

Cuadros estadísticos

El resultado del proceso de tabulación o condensación de datos se presenta en lo que en estadística se llaman cuadros estadísticos, también conocidos con el nombre incorrecto de tablas estadísticas, error producto de la traducción inglesa. Los cuadros estadísticos sólo tienen como objetivo ser depósitos de datos o bien contener datos ya procesados, es decir, información, que el investigador utiliza como herramienta en sus análisis. Los cuadros estadísticos tienen como principales ventajas presentar una gran cantidad de información, la posibilidad de leer valores exactos y la sencillez de elaboración.

Con base en el uso que el investigador le dé a un cuadro estadístico, éstos pueden ser clasificados en dos tipos: cuadros de trabajo y cuadros de referencia.

Cuadros de trabajo. Son aquellos que contienen datos producto de una tabulación. En otras palabras, son cuadros depositarios de datos. Éstos son utilizados por el investigador para obtener, a partir de ellos, las medidas estadísticas requeridas. Un ejemplo de este tipo de cuadros se muestra a continuación (Cuadro 2).

CUADRO 2
Producción de oro y plata en México
(toneladas)

Año	Oro	Plata
1990	8.5	2 351.60
1991	8.9	2 223.60
1992	10.4	2 317.40
1993	11.1	2 415.80
1994	14.4	2 325.40

FUENTE: Departamento de Información de Negocios de Banamex, con datos del Instituto Nacional de Estadística, Geografía e Informática (INEGI).

En el cuadro anterior se observa simplemente la presentación de los datos de la producción de oro y plata de 1990 a 1994 como resultado de un conteo de la producción nacional de esos metales sin incluir en él ningún tipo de cálculo.

Cuadros de referencia. Son aquellos que tienen como finalidad ayudar al investigador en el análisis formal de las interrelaciones que tienen las variables que están en estudio, es decir, contienen información ya procesada de los cuadros de trabajo (proporciones, porcentajes, tasas, coeficientes, etcétera). El Cuadro 3 muestra un cuadro de referencia elaborado a partir del cuadro de trabajo 2.

El cuadro de referencia es, por lo tanto un cuadro analítico en el cual se han transformado los datos absolutos. Para el Cuadro 3, la transformación de los datos absolutos de producción se ha plasmado en un cálculo de variación anual.[5]

CUADRO 3
Producción de oro y plata en México

Año	Oro (toneladas)	Variación anual (%)	Plata (toneladas)	Variación anual (%)
1990	8.5	-0.8	2 351.60	2.0
1991	8.9	4.6	2 223.60	-5.4
1992	10.4	16.5	2 317.40	4.2
1993	11.1	6.8	2 415.80	4.2
1994	14.4	29.4	2 325.40	-3.7

FUENTE: Departamento de Información de Negocios de Banamex, con datos del INEGI.

[5] El cálculo de la variación anual se explica detalladamente en el apartado "Incrementos".

Elementos para la elaboración de cuadros estadísticos[6]

La construcción de cuadros estadísticos de trabajo o de cuadros de referencia requiere prácticamente de los mismos elementos en su elaboración, ya que ambos presentan las mismas características estructurales, por lo que los elementos que a continuación se describen deberán ser utilizados en la conformación de éstos indistintamente.

1. *Número del cuadro.* Es el primer elemento de todo cuadro estadístico. Tiene como objeto permitir una fácil y rápida referencia al mismo.
2. *Título.* Es el segundo elemento del cuadro estadístico. En él se deberá indicar el contenido del cuadro, su circunscripción espacial, el periodo o espacio temporal y las unidades en las que están expresados los datos.
3. *Nota en el título* (encabezado). Elemento complementario del Título. Se emplea sólo en aquellos cuadros en los que se requiere proporcionar información adicional; incluso como un todo o relativa a la parte principal del mismo.
4. *Casillas cabecera.* Contienen la denominación de cada característica o variable que se clasifica. Por ejemplo, para los cuadros 2 y 3, las casillas de cabecera están formadas por oro, plata y variación anual %.
5. *Columnas.* Son las subdivisiones verticales de las casillas cabecera. Se incluyen tantas columnas en una casilla cabecera como categorías le correspondan. Por ejemplo, al clasificar la variable género en un cuadro estadístico, se tendría una columna para hombres y otra para mujeres.
6. *Renglones.* Son las divisiones horizontales que corresponden a cada criterio en que es clasificada una variable. Por ejemplo, en los cuadros 2 y 3, la producción de oro y plata es clasificada para cada año en el periodo 1990 a 1994.
7. *Espacio entre los renglones.* Tienen por objeto hacer más clara la presentación de los datos, facilitando así su lectura. Un ejemplo de éste se muestra en el Cuadro 3.
8. *Líneas de cabecera.* Son las líneas que se trazan para dividir las casillas de cabecera de los renglones.
9. *Cabeza del cuadro.* Está formada por el conjunto de casillas cabecera y encabezados de columnas.
10. *Casillas.* Es la intersección que forman cada columna con cada renglón en el cuadro. Las casillas contienen datos o bien los resultados de cálculos efectuados con ellos.
11. *Cuerpo del cuadro.* Está formado por todos los datos sin considerar la cabeza del cuadro y los renglones de totales.

[6] El contenido de este subtema se ha tomado, con modificaciones, del propuesto por F. Holguín Quiñones en *Estadística descriptiva aplicada a las ciencias sociales, op. cit.*, pp. 44-46.

12. *Renglón de totales*. Es un elemento opcional en los cuadros estadísticos. Se agrega el renglón de totales sólo en aquellos cuadros donde es posible sumar el contenido de las columnas. Esta característica del cuadro también es válida para las filas, es decir, es posible tener columna de totales.
13. *Línea final de cuadro*. Es la línea que se traza al final del cuerpo del cuadro y, en su caso, al final del renglón de totales. Divide los datos y cálculos de la nota de cabecera, notas al pie del cuadro y la fuente.
14. *Notas al pie del cuadro*. Se usan para calificar o explicar un elemento particular en el cuadro que presente una característica distinta de clasificación. Por ejemplo, si en el cuadro 3, la producción de oro de 1994 fue la obtenida durante el primer semestre del año, esto se deberá indicar en una nota al pie del cuadro.
15. *Fuente*. Es el último elemento en un cuadro estadístico. Tiene por objeto indicar el origen de los datos.

Para mostrar cada uno de los elementos que constituyen un cuadro estadístico se han elaborado, mediante una hoja electrónica de cálculo,[7] dos ejemplos: el cuadro de trabajo número 4 y el cuadro de referencia número 5, ambos relativos al gasto federal en educación durante tres generaciones en México.

Para complementar la elaboración de cuadros estadísticos, el lector podrá consultar la siguiente dirección electrónica: http://www.geocities.com/aipierdant/cuadros/.

Elaboración de cuadros estadísticos con Excel

Las hojas electrónicas como Excel permiten la elaboración rápida de cuadros y gráficas estadísticas. En esta sección se describirá brevemente el procedimiento de construcción de cuadros estadísticos. Si usted no tiene experiencia con la hoja electrónica de Excel le sugerimos revisar el anexo II, en el cual se da una introducción al manejo de este paquete, mostrándose los conceptos y comandos básicos usados en él. Si usted requiere de más información respecto del tema, le sugerimos recurrir a la amplia bibliografía elaborada sobre el mismo.

El proceso de construcción de cuadros estadísticos es similar para cuadros de trabajo o de referencia. Los elementos considerados deberán utilizarse como una guía en la construcción de cualquier tipo de cuadro. Como ejemplo utilizaremos el cuadro de trabajo número 4.

[7] Las hojas electrónicas de cálculo aparecen con las microcomputadoras y no son más que programas diseñados para elaborar cuadros y gráficas, así como para realizar cálculos numéricos de diversos tipos (aritméticos, algebraicos, estadísticos, financieros, etcétera).

Número del cuadro

Título

Nota en el título

CUADRO 4

Gasto federal en educación durante tres generaciones[1] *en México*

Casillas cabecera

Línea de cabecera

	Generación 1977-1992	1978-1993	1979-1994
Gasto federal en educación[2]	159 253	160 029	160 274
Básica	115 244	113 595	111 230
Media	23 079	23 210	23 694
Superior	20 930	23 224	25 350
Tamaño de la generación[3]	17 684	18 681	19 575
Básica	12 629	13 536	14 126
Media	3 842	3 939	4 180
Superior	1 213	1 206	1 269

Renglones

Renglón de total

Línea final de cuadro

Notas al pie

[1] Se supone un promedio de 16 años de instrucción (seis de básica, seis de media y cuatro de superior).
[2] Millones de nuevos pesos de 1994.
[3] Miles de matriculados en el sistema de educación pública.

Fuente

FUENTE: Departamento de Estudios Económicos de Banamex, con datos de informes de gobierno.

Número del cuadro

Título

Nota en el título

CUADRO 5

Gasto federal en educación durante tres generaciones[1] *en México* (%)

Casillas cabecera

Línea de cabecera

	Generación 1977-1992	1978-1993	1979-1994
Gasto federal en educación[2]	100.0	100.0	100.0
Básica	72.4	71.0	69.4
Media	14.5	14.5	14.8
Superior	13.1	14.5	15.8
Tamaño de la generación[3]	100.0	100.0	100.0
Básica	71.4	72.5	72.2
Media	21.7	21.1	21.4
Superior	6.9	6.5	6.5

Renglones

Renglón de total

Línea final de cuadro

Nota al pie

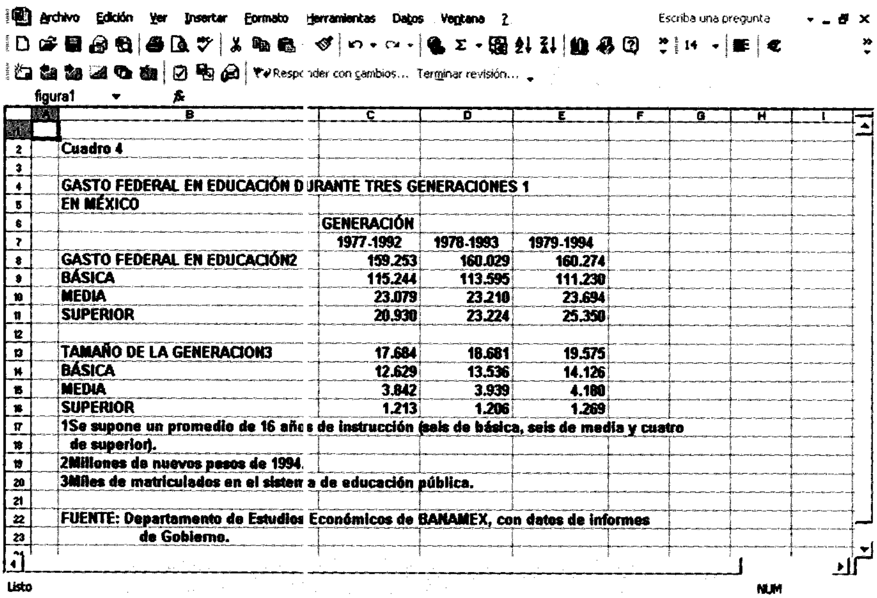
[1] Se supone un promedio de 16 años de instrucción (seis de básica, seis de media y cuatro de superior).

Fuente

FUENTE: Departamento de Estudios Económicos de Banamex, con datos de informes de gobierno.

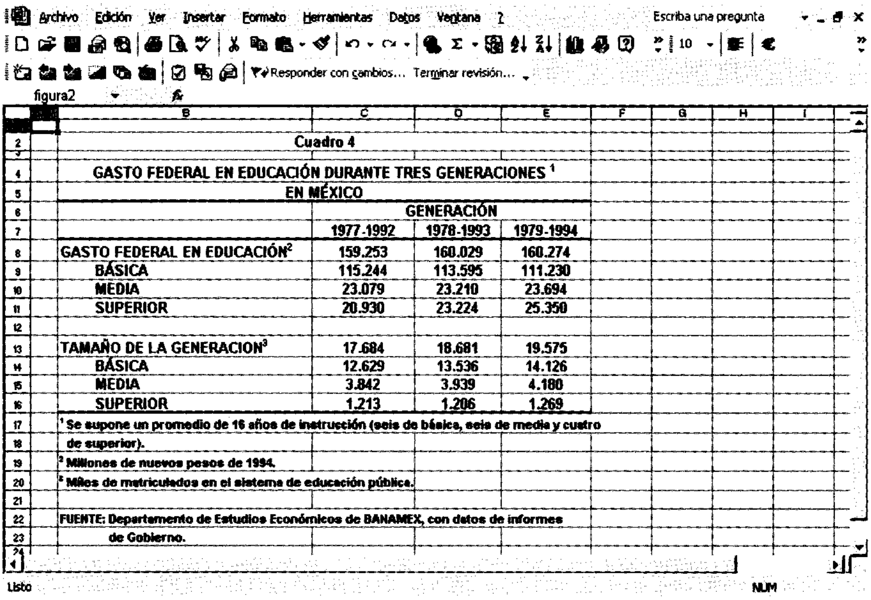
- Seleccione en la hoja el área de construcción del cuadro (columnas y renglones a utilizar).
- Determine el número de columnas y renglones requeridos. Considere que cada columna en el cuadro corresponde a una columna en la hoja, y así, respectivamente, para los renglones del mismo.
- Ubique el número de cuadro, título, subtítulo y unidades de medida en las celdas de la primer columna en su renglón correspondiente, como se indica en la Figura 1.
- Capture en las columnas correspondientes las casillas de cabecera (Figura 1).
- Agregue en el renglón respectivo las fórmulas o datos necesarios en el cuadro estadístico, tal como se indica en la Figura 1.
- Capture en su caso las notas al pie y la fuente, dejando entre ambas una línea de separación (Figura 1).
- Dé formato final al cuadro (Figura 2), e inicie con la alineación de columnas del número de cuadro, título, subtítulo y unidades, así como la selección del tipo de letra y tamaño. Posteriormente, dé formato (alineación, tipo de letra, tamaño de letra, etcétera) a las casillas de cabecera y trace las líneas respectivas. Como tercer paso en este punto, dé formato a los renglones de cuadro. Finalmente, trace la línea final de cuadro y dé formato a las notas al pie y la fuente.

FIGURA 1



	A	B	C	D	E
1					
2		Cuadro 4			
3					
4		GASTO FEDERAL EN EDUCACIÓN DURANTE TRES GENERACIONES 1			
5		EN MÉXICO			
6			GENERACIÓN		
7			1977-1992	1978-1993	1979-1994
8		GASTO FEDERAL EN EDUCACIÓN2	159.253	160.029	160.274
9		BÁSICA	115.244	113.595	111.230
10		MEDIA	23.079	23.210	23.694
11		SUPERIOR	20.930	23.224	25.350
12					
13		TAMAÑO DE LA GENERACION3	17.684	18.681	19.575
14		BÁSICA	12.629	13.536	14.126
15		MEDIA	3.842	3.939	4.180
16		SUPERIOR	1.213	1.206	1.269
17		1Se supone un promedio de 16 años de instrucción (seis de básica, seis de media y cuatro			
18		de superior).			
19		2Millones de nuevos pesos de 1994.			
20		3Miles de matriculados en el sistema de educación pública.			
21					
22		FUENTE: Departamento de Estudios Económicos de BANAMEX, con datos de informes			
23		de Gobierno.			

FIGURA 2



Cuadro 4

GASTO FEDERAL EN EDUCACIÓN DURANTE TRES GENERACIONES [1] EN MÉXICO

	GENERACIÓN		
	1977-1992	1978-1993	1979-1994
GASTO FEDERAL EN EDUCACIÓN[2]	159.253	160.029	160.274
BÁSICA	115.244	113.595	111.230
MEDIA	23.079	23.210	23.694
SUPERIOR	20.930	23.224	25.350
TAMAÑO DE LA GENERACION[3]	17.684	18.681	19.575
BÁSICA	12.629	13.536	14.126
MEDIA	3.842	3.939	4.180
SUPERIOR	1.213	1.206	1.269

[1] Se supone un promedio de 16 años de instrucción (seis de básica, seis de media y cuatro de superior).

[2] Millones de nuevos pesos de 1994.

[3] Miles de matriculados en el sistema de educación pública.

FUENTE: Departamento de Estudios Económicos de BANAMEX, con datos de informes de Gobierno.

Ejercicios

Utilice Excel o cualquier paquete de hoja de cálculo de que disponga y elabore como ejercicios los siguientes cuadros estadísticos.

CUADRO E1

Crecimiento de televisión restringida en el Distrito Federal

Años	Hogares con tv restringida	Hogares con Cablevisión	Hogares con Multivisión
1989	87 489	85 400	2 089
1990	126 483	100 400	26 083
1991	189 175	115 400	73 775
1992	311 456	146 000	165 456
1993	418 049	158 000	260 049
1994	651 033	233 000	418 033

FUENTE: Empresas de televisión restringida: Cablevisión y Multivisión.

CUADRO E2
Exportación de crudo
(miles de barriles diarios)

Tipo	1983	1984	1985	1986	1987	1988/p
Maya	859.1	904.2	829.4	716.5	818.9	768.3
Istmo	677.9	620.4	604.9	573.1	526.1	466.7
Olmeca	0	0	0	0	0	71.7
Total	1537.0	1524.6	1434.3	1289.6	1345.0	1306.7

p/ Cifra preliminar
FUENTE: Pemex, Informe de Labores 1987-1988 y Memorias de Labores.

CUADRO E3
Poder adquisitivo del salario 1940-1997

Presidente	Sexenio	Variación %	Horas de trabajo para adquirir una canasta básica
M. Ávila Camacho	40-46	-19.6	13.21
M. Alemán Valdés	46-52	-10.8	15.38
A. Ruiz Cortines	52-58	19.2	12.3
A. López Mateos	58-64	50.4	8.03
G. Díaz Ordaz	64-70	34.4	6.10
L. Echeverría A.	70-76	32.0	5.15
J. López Portillo	76-82	12.8	5.24
M. de la Madrid	82-88	-98.0	9.19
C. Salinas de G.	88-94	-56.0	16.39
E. Zedillo P.	94 al 97	-22.0	25.13

FUENTE: INEGI y Secofi.

CUADRO E4

Fuentes de Ingreso del Departamento del Distrito Federal

(millones de pesos)

FUENTE	1995	1996	1997
Contribuciones	9 337	11 488	15 914
Participaciones en Ingresos Federales	7 392	10 603	11 202
Organismos y Empresas controladas por el DDF	1 805	3 006	3 489
Financiamiento	1 091	5 100	5 950
TOTAL	19 625	30 197	36 555

FUENTE: Secretaría de Finanzas del DDF.

CUADRO E5

Promedio de escolaridad de la PEA según sexo

(grados escolares)[1]

AÑO	TOTAL	HOMBRES	MUJERES
1991	6.6	6.5	7
1993	6.7	6.6	7.1
1995	7.1	7.1	7.3
1996	7.4	7.3	7.6
1997	7.6	7.5	7.7
1998	7.5	7.5	7.7
1999	7.5	7.4	7.7
2000	7.7	7.6	8
2001	7.8	7.7	8.1
2002	7.9	7.8	8.2
2003	8	7.9	8.3
2004	8.2	8.1	8.4

[1] Con el fin de ofrecer una serie anual amplia y comparable, este tabulado presenta información sólo del segundo trimestre de cada año.

FUENTE: INEGI/STPS, Encuesta Nacional de Empleo.

CUADRO E6

Población económicamente activa según sexo,[1] 1991 a 2004

AÑO	TOTAL	HOMBRES	MUJERES
1991	31 229 048	21 630 013	9 599 035
1993	33 651 812	23 243 466	10 408 346
1995	36 195 641	24 347 607	11 848 034
1996	36 831 734	24 814 965	12 016 769
1997	38 584 394	25 394 098	13 190 296
1998	39 562 404	26 146 569	13 415 835
1999	39 648 333	26 295 840	13 352 493
2000	40 161 543	26 418 355	13 743 188
2001	40 072 856	26 415 550	13 657 306
2002	41 085 736	26 888 135	14 197 601
2003	41 515 672	27 277 029	14 238 643
2004	43 398 755	28 013 539	15 385 216

[1] Con el fin de ofrecer una serie anual amplia y comparable, este tabulado presenta información sólo del segundo trimestre de cada año.

FUENTE: INEGI/STPS, Encuesta Nacional de Empleo.

Porcentajes, proporciones, razones, coeficientes e incrementos

LA ELABORACIÓN DE CUADROS ESTADÍSTICOS, y estadísticas en general, implican el uso de ciertos conocimientos aritméticos que permitan obtener medidas de comparación de los datos que han sido condensados. Estas herramientas que analizan las características de las variables clasificadas de un problema particular son: los porcentajes, las proporciones, los coeficientes y las razones.

Porcentajes

Un porcentaje es la relación que se establece entre cada una de las partes que forman un todo, entre el todo o total multiplicado por 100; en otras palabras, es la relación que se establece entre un subconjunto de un conjunto, dividido entre todos los elementos que forman el conjunto de estudio multiplicado por 100. El porcentaje se representa con el símbolo %. Por lo tanto, ese todo o total representa el 100 por ciento, y cada una de las relaciones obtenidas al dividir la parte entre el total y multiplicarla por cien representa un tanto de cien, y es definido como tanto por ciento. Por ejemplo, si una pequeña población rural está formada por 20 hombres y 60 mujeres, el cociente que resulta de dividir el número de hombres en relación al total de personas y multiplicado por 100 nos indica el número de hombres por cada 100 personas que hay en esa población.

Hombres:	20
Mujeres:	60
Total	80

$$\% \text{ de hombres} = \frac{\text{número de hombres}}{\text{total de personas}} (100) = \frac{20}{80} (100) = 25\%$$

$$\% \text{ de mujeres} = \frac{\text{número de mujeres}}{\text{total de personas}} (100) = \frac{60}{80} (100) = 75\%$$

En el ejemplo, por cada 100 habitantes que existan en esta población, 25 serán hombres y 75 serán mujeres.

Entonces, el cálculo del porcentaje de un subconjunto "n" de "N", que es mutuamente excluyente en relación con otros subconjuntos, podrá expresarse matemáticamente como:

$$\% \, n = \frac{\text{número de elementos de n}}{\text{total de elementos en el universo N}} \, (100)$$

Por lo tanto, la suma de los porcentajes de todos los subconjuntos mutuamente exclusivos que forman un universo será siempre igual a 100 por ciento, como se muestra en el ejemplo de los habitantes de la población rural.

La principal utilidad de los porcentajes en estadística es poder obtener comparabilidad, ya que las cifras absolutas impiden en muchas ocasiones lograrla, en virtud de que oscurecen las relaciones. La comparabilidad es posible porque los números absolutos se reducen a una escala que es fácil de multiplicar y dividir, transforman al conjunto que forma el número base (total de elementos del universo) en la cifra 100 que es fácilmente divisible y multiplicable por otros números, lo cual permite la determinación de su magnitud relativa.

Los porcentajes son especialmente útiles cuando se comparan dos o más conjuntos numéricos. Por ejemplo, del Cuadro 5, es posible comparar el gasto en educación por generación, observándose que sólo la educación superior ha incrementado sus recursos económicos en las últimas tres generaciones al pasar de 13.1, en la generación 1977-1992 a 15.8 por ciento en la generación 1979-1994, a pesar de que su tamaño de matrícula ha disminuido de 6.9 a 6.5 por ciento, respectivamente.

En general, los porcentajes se presentan como enteros aunque se pueden calcular usando uno, dos, tres o más decimales. La costumbre en estadística es presentarlos en forma entera; en casos contados podrán presentarse con un decimal y muy pocas veces con dos o más decimales.

Finalmente, es importante mencionar que es un error manipular los porcentajes como si fueran números absolutos, es decir, no se pueden sumar, promediar o combinar cuando se han obtenido de bases diferentes. Retomando el ejemplo del Cuadro 5 y analizando las cifras del Cuadro 4, no podemos promediar el gasto en educación superior, ya que este promedio representaría (13.1+14.5+15.8)/3 = 14.46%, cuando en realidad el gasto promedio en las tres generaciones es de,

$$[(20{,}930+23{,}224+25{,}350)/(159{,}253+160{,}029+160{,}274)]*100 = 14.49\%$$

¿Cómo se calculan los porcentajes en las cuadros estadísticos?

La elaboración de cuadros estadísticos busca generalmente poner en relación dos o más características objeto de una investigación, así por ejemplo podemos comparar carrera universitaria y género, género y votación, trabajo y género, edad, delincuencia y zona de la ciudad, afiliación a partido político y nivel de ingresos, etcétera. En todos estos ejemplos buscamos saber, como ya indicamos, si existe algún tipo de relación o bien si no la hay.

El investigador deberá establecer, en primer término, independientemente de la relación objeto de estudio, el sentido en que deben calcularse los porcentajes en el cuadro estadístico, y para ello deberá tomar en cuenta la siguiente regla:

LOS PORCENTAJES DEBEN CALCULARSE EN EL SENTIDO DEL FACTOR QUE SE CONSIDERA COMO LA CAUSA

Por ejemplo, si consideramos que el género es la causa que da origen a la elección de un determinado tipo de carrera universitaria, entonces calcularemos para el cuadro estadístico 6, los porcentajes en el sentido del factor género.

CUADRO 6
Estructura por género y carrera del Grupo SB09-05/O de la UAM-Xochimilco
(alumnos)

	GÉNERO		
LICENCIATURA	HOMBRES	MUJERES	TOTAL
Administración	3	2	5
Economía	3	1	4
Sociología	4	3	7
Psicología	4	5	9
Comunicación	3	4	7
TOTAL	17	15	32

FUENTE: Elaboración propia con datos hipotéticos.

Los resultados del cálculo se muestran en el Cuadro 6.1

CUADRO 6.1

Estructura por género y licenciatura del Grupo SB09-05/O de la UAM-Xochimilco

(%)

	GÉNERO		
LICENCIATURA	HOMBRES	MUJERES	TOTAL
Administración	18	13	16
Economía	18	7	12
Sociología	23	20	22
Psicología	23	33	28
Comunicación	18	27	22
TOTAL	100	100	100

FUENTE: Elaboración propia con datos del Cuadro 6.

AL CALCULAR LOS PORCENTAJES EN UN SENTIDO,
LA COMPARACIÓN DEBE HACERSE EN SENTIDO CONTRARIO

En el cuadro se calcularon los porcentajes en el sentido del factor género, entonces, la comparación debe hacerse en el sentido de la variable licenciatura. En el grupo SB09-05/O, administración ha sido seleccionada por 18 por ciento de los hombres y sólo 13 por ciento por las mujeres; es decir, los hombres seleccionan en 5 por ciento más esta licenciatura que las mujeres. De los alumnos del grupo sólo 16 por ciento seleccionó administración. Comparaciones similares pueden realizarse para cada uno de los renglones del cuadro.

Si los porcentajes se calculan en sentido horizontal (Cuadro 6.2), es decir, por licenciatura, deberán interpretarse por género, entonces podemos observar que de los alumnos del grupo inscritos en administración, 60 por ciento son hombres y 40 por ciento son mujeres: 20 por ciento más hombres que mujeres.

Como puede observarse, calcular los porcentajes en un cuadro estadístico en un sentido u otro proporciona no sólo diferentes resultados sino también distintas interpretaciones en las cifras. En este punto es importante mencionar que no siempre es posible calcular los porcentajes en ambos sentidos (vertical y horizontalmente), ya que el analista deberá determinar siempre, en la realización de un análisis, cuál es el factor causal de la relación, de tal forma que éste se establezca proporcionando un sentido estrictamente lógico en el análisis.

CUADRO 6.2

Estructura por licenciatura y género del Grupo SB09-05/O de la UAM-Xochimilco

(%)

	GÉNERO		
LICENCIATURA	HOMBRES	MUJERES	TOTAL
Administración	60	40	100
Economía	75	25	100
Sociología	57	43	100
Psicología	44	56	100
Comunicación	43	57	100

FUENTE: Elaboración propia con datos del Cuadro 6.

Como ejemplo de esto último se muestra a continuación el Cuadro 7. En éste, el cálculo de los porcentajes se ha pensado en forma vertical, es decir, para el factor personal docente y población estudiantil, siendo su interpretación por licenciatura; así, el factor personal docente y población estudiantil, en el tronco interdivisional y divisional cuentan con 14.6 por ciento de los docentes que atienden al 28.8 por ciento del alumnado de la división. Si el cálculo se hiciera en sentido horizontal, estadísticamente esto no tendría ningún significado, ya que no tiene sentido sumar datos de docentes con alumnos para tratar de establecer alguna relación.

CUADRO 7

UAM-Xochimilco. Personal docente y población estudiantil por licenciatura en la DCSH

LICENCIATURA	DOCENTES	%	ALUMNOS	%
Tronco Div. e Int.	45	14.6	1 160	28.8
Administración	23	7.5	718	17.8
Economía	66	21.4	422	10.5
Sociología	56	18.2	280	7.0
Psicología	60	19.5	842	21.0
Comunicación	58	18.8	597	14.9
TOTAL	308	100	4 019	100

FUENTE: Elaboración propia con datos del Informe de estadística escolar básica 94/O de la Coordinación de Sistemas Escolares de la UAM-Xochimilco.

Proporciones

Matemáticamente, una proporción es la igualdad de dos razones cuyo objetivo es establecer la relación entre una parte con respecto al todo. En las proporciones no se multiplica el cociente resultante por 100, ya que la relación se establece respecto de la unidad. Las proporciones y los porcentajes ofrecen la misma información, aunque estos últimos se emplean más por ser más fácil su comprensión.

Matemáticamente, la proporción se define como:

$$\text{proporción de n} = \frac{\text{número de elementos de n}}{\text{total de elementos en el universo N}}$$

Retomando el ejemplo de la población rural usado en la definición de porcentaje (pág. 41), calcularemos ahora las proporciones.

Hombres:	20
Mujeres:	60
Total	80

$$\text{proporción de hombres} = \frac{\text{número de hombres}}{\text{total de personas}} = \frac{20}{80} = \frac{1}{4}\ (0.25)$$

$$\text{proporción de mujeres} = \frac{\text{número de mujeres}}{\text{total de personas}} = \frac{60}{80} = \frac{3}{4}\ (0.75)$$

En la población existe un hombre por cada cuatro habitantes y tres mujeres por cada cuatro habitantes.

La suma de todas las proporciones de un conjunto universo suman uno cuando los subconjuntos son mutuamente excluyentes como puede observarse en el ejemplo anterior.

$$\text{proporción de hombres} + \text{proporción de mujeres} = 1$$

$$0.25 + 0.75 = 1$$

Las proporciones son muy utilizadas en los cálculos estadísticos; sin embargo, en el reporte o presentación final de las cifras se acostumbra emplear porcentajes, ya que éstos se obtienen multiplicando la proporción por 100.

Razones

Razón es el cociente indicado de dos cantidades. Si se consideran dos magnitudes y se establece entre ellas una proporcionalidad "f", como se indica en la figura siguiente:

$$\begin{array}{l} X_1 \xrightarrow{f} f\,[X_1] \\ X_2 \xrightarrow{f} f\,[X_2] \end{array}$$

entonces, a esta relación

$$\frac{f\,[X_1]}{f\,[X_2]} = K$$

se le denomina razón.

En estadística, entenderemos a la razón como la relación que se da entre dos subconjuntos o dos conjuntos. Si retomamos el ejemplo de la población rural (pág. 41) formada por 80 habitantes, entonces la razón existente de hombres a mujeres es 1 a 3, es decir, un hombre por cada tres mujeres.

20 hombres / 60 mujeres =1/3

o bien: 3.3 hombres a 10 mujeres = 33 hombres a 100 mujeres.

Las razones, por lo tanto, pueden multiplicarse o dividirse por un mismo número y no se alteran, lo que permite expresarlas, en ocasiones, como números enteros.

En demografía es muy empleada la razón de hombres a mujeres, lo que nos indica el número de hombres por cada 100 mujeres. A esta relación se le conoce con los nombres de relación de masculinidad, índice de masculinidad, razón de masculinidad o sex ratio. Matemáticamente, puede expresarse

$$\text{razón de masculinidad} = \frac{Hx}{Mx} K$$

donde:

Hx representa el número de hombres de edad x,
Mx representa el número de mujeres de edad x,
K representa una constante (generalmente 100 o 1000).

Coeficientes

Los coeficientes, también conocidos con los nombres de tasas e índices, son indicadores muy similares a un porcentaje. En un coeficiente, el numerador indica el número de veces que un evento específico ocurre durante un lapso o periodo particular, y, en el denominador, el número de veces que el evento está sujeto al riesgo de que ocurra o acontezca. Por lo general, el coeficiente o tasa es multiplicado por un número que usualmente es mil, 10 mil o 100 mil.

Entre algunos de los coeficientes más conocidos están el de mortalidad general, nupcialidad, natalidad, delincuencia, fertilidad general y específica, índice de profesionales, estudiantes, afiliación a grupos políticos, etcétera.

$$\text{tasa de mortalidad general} = \frac{\text{número de defunciones en un área determinada, durante un año dado}}{\text{Población del área a mitad del año [1 de julio]}} [1000]$$

$$\text{tasa de mortalidad infantil} = \frac{\text{Defunciones en una área y tiempo determinados de niños menores de 1 año}}{\text{Número de nacidos vivos en el área y tiempo del numerador}} [1000]$$

Los coeficientes o índices generales también reciben los nombres de tasas crudas o brutas, en virtud de que aparece en el denominador la población y no el riesgo de que acontezca un evento.

Se pueden construir índices más específicos dependiendo de las necesidades del investigador. A continuación se muestra, en el cuadro estadístico 8, un ejemplo de este tipo de coeficientes.

CUADRO 8
Universidad Autónoma Metropolitana-Xochimilco
Recursos Humanos

				Coeficientes		
Años	Población estudiantil[1]	Personal docente[2]	Personal administrativo	Alumno/ Docente	Alumno/ P. Admvo.	P. Admvo./ P. Docente
1974	948	213	167	4.5	5.7	0.8
1979	7937	443	589	17.9	13.5	1.3
1984	10348	897	1048	11.5	9.9	1.2
1989	10745	955	1061	11.3	10.1	1.1
1994	11916	973	930	12.2	12.8	1.0

[1] Incluye a los alumnos de licenciatura y posgrado de los trimestres de otoño.
[2] Incluye personal docente de tiempo completo, medio tiempo y tiempos parciales.

FUENTE: Elaboración propia con datos del *Informe de Actividades 1994-1995* de Jaime Kravzov Jinich, a la sazón, rector de unidad.

En el Cuadro 8 se observa que en 1974 existían 4.5 alumnos por profesor, mientras que para 1994 esta relación es de 12.2 alumnos, la cual prácticamente ha permanecido casi constante, sin cambios significativos a partir de 1984.

Incrementos

En estadística es común analizar el comportamiento que tienen los fenómenos en el tiempo a partir del comportamiento de las variables asociadas a ellos, lo que permite determinar cambios en éstos, es decir, determinar si crecen, decrecen o permanecen estables y, además, precisar la magnitud del incremento o decremento.

Los cambios de comportamiento de un fenómeno pueden expresarse mediante porcentajes en la forma siguiente:

$$\text{incremento porcentual} = \frac{\text{Valor del último dato - Valor del dato base}}{\text{Valor del dato base}} [100]$$

$$\text{incremento porcentual} = \frac{Vu - Vb}{Vb} [100]$$

En este tipo de cálculos es importante tener cuidado en respetar los signos, especialmente en el caso de los incrementos negativos, pues éstos nos indican un decrecimiento en el fenómeno al pasar del periodo base al periodo de estudio.

En algunas ocasiones el cálculo de los incrementos porcentuales dan como resultado valores altos: 500%, -800%, 1500%, etcétera, lo que nos indica un cálculo correcto pero constituyen, técnicamente, estadísticas muy pobres. En estos casos se recomienda indicar los incrementos o decrementos en término de número de veces que creció o disminuyó un fenómeno. Por ejemplo, usando la información del Cuadro 8, observemos que la matrícula en la UAM-Xochimilco creció de 1974 a 1994, 1 156.96%, lo que nos indica que en 20 años ésta creció 11.56 veces la magnitud que tenía en 1974.

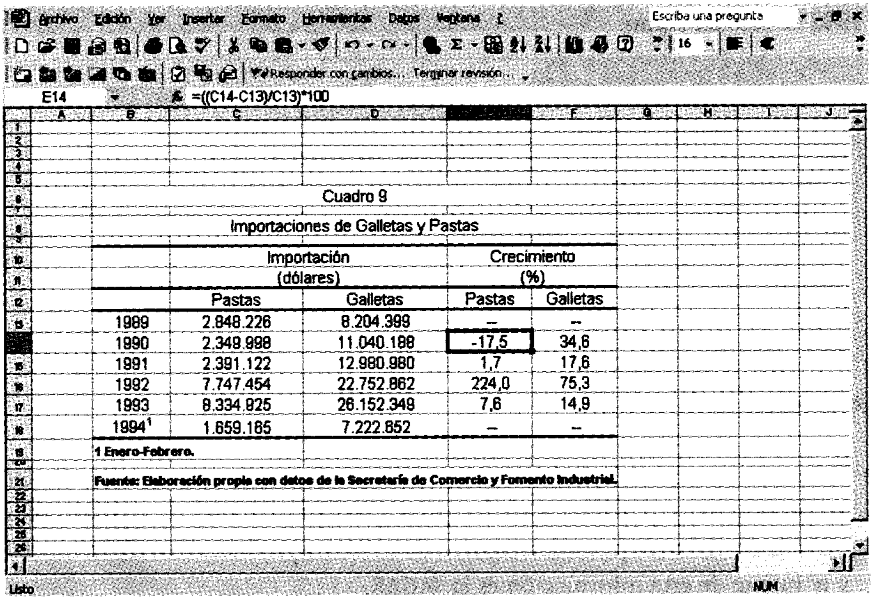
$$\text{Incremento \%} = [(11916 - 948) / 948] (100) = 1\ 156.96\%$$

Para mostrar el cálculo de los incrementos en un cuadro estadístico utilizaremos la información estadística de importaciones de galletas y pastas proporcionada por la Secretaría de Comercio y Fomento Industrial (Figura 3).

Los pasos para la construcción del Cuadro 9 son:

- Capture el título, las casillas de cabecera, los renglones, las cifras de importaciones, las notas al pie y la fuente siguiendo los lineamientos indicados en el capítulo anterior.
- Introduzca la fórmula de incremento en las celdas respectivas; por ejemplo, la celda E14 tiene la fórmula: =((C14-C13)/C13)*100, que permite conocer el incremento de las importaciones de pastas en 1990 respecto de 1989; la E15 tendrá el incremento de las importaciones de pastas de 1991 respecto de 1990: =[(C15-C14)/C14]*100, y así sucesivamente.
- Recuerde que no es posible, en este caso particular, obtener un incremento para el primer año base de comparación. Ni tampoco para el último año; puesto que la información de 1994 es sólo para enero-febrero y los incrementos calculados son para años completos.

FIGURA 3



Cuadro 9

Importaciones de Galletas y Pastas

	Importación (dólares)		Crecimiento (%)	
	Pastas	Galletas	Pastas	Galletas
1989	2.848.226	8.204.399	–	–
1990	2.349.998	11.040.188	-17,5	34,6
1991	2.391.122	12.980.980	1,7	17,6
1992	7.747.454	22.752.862	224,0	75,3
1993	8.334.925	26.152.349	7,6	14,9
1994[1]	1.659.165	7.222.652	–	–

1 Enero-Febrero.

Fuente: Elaboración propia con datos de la Secretaría de Comercio y Fomento Industrial.

En los resultados del cuadro podemos observar el decrecimiento de las importaciones de pastas de 1989 a 1990 en 17.5%, para volver a crecer 1.7% en 1991. El año de 1992 representa un crecimiento espectacular respecto de 1991 en las importaciones tanto de pastas como de galletas: 224% y 75.3%, respectivamente.

Finalmente, en estadística es muy frecuente calcular incrementos promedio de los datos que se tienen para diferentes periodos, sin embargo, deberá tenerse cuidado de no calcularlos aritméticamente, ya que esto es un error. Si se supone un crecimiento lineal, el promedio puede calcularse con la siguiente fórmula que permite obtener un valor aproximado:

$$\text{incremento medio} = \left(\frac{V_1 - V_0}{V_1 + V_0}\right)\left(\frac{2}{n}\right)K$$

donde:

V_0 = valor de la variable en el periodo base
V_1 = valor de la variable en el periodo último
n = número de periodos (años, semestres, bimestres, etcétera)
K = es una constante (generalmente 100)

Para el ejemplo, el incremento medio de las importaciones de pastas entre 1989 y 1993 se calcula como

$$\text{incremento medio} = \left(\frac{8{,}334{,}925 - 2{,}848{,}226}{8{,}334{,}925 + 2{,}848{,}226}\right)\left(\frac{2}{4}\right)100$$

$$\text{incremento medio} = 24.5\%\ \text{anual}$$

Ejercicios

- Elabore un cuadro estadístico en el que se muestre la estructura de producción, en proporción y porcentaje, de los principales productores de algodón pluma en México, del ciclo primavera/verano de 1996. La producción se da en toneladas y la fuente de esta información es la SAGAR.

Baja California (55 760)
Coahuila (19 332)
Chihuahua (40 683)
Sonora (82 530)
Campeche (7 255)

Solución

CUADRO E7
Estructura de producción de los principales productores de algodón pluma en México
(primavera/verano, 1996)

PRODUCTOR	PRODUCCIÓN (TON)	PROPORCIÓN	(%)
Baja California	55 760	0.271	27.1
Coahuila	19 332	0.094	9.4
Chihuahua	40 683	0.198	19.8
Sonora	82 530	0.402	40.2
Campeche	7 255	0.035	3.5

FUENTE: SAGAR.

- Determine la participación en el mercado de las empresas de televisión restingida, Cablevisión y Multivisión, para 1993 y 1994. Los clientes de Cablevisión, en 1993, fueron: 158 000, y en 1994: 233 000. Los clientes de Multivisión fueron, en 1993: 260 049, y en 1994: 418 033.

Solución

En 1993, la participación de Cablevisión en el mercado fue de 37.8%, y la de Multivisión fue de 62.2%.

En 1994, la participación de Cablevisión en el mercado fue de 35.8%, y la de Multivisión fue de 64.2%.

De 1993 a 1994 Cablevisión perdió 2% del mercado, el cual fue absorbido por Multivisión.

- Determine el porcentaje de crecimiento de las contribuciones (impuestos) que recibió el Departamento del Distrito Federal de 1995 a 1996 y de 1996 a 1997. Los datos proporcionados están dados en millones de pesos.

Concepto	1995	1996	1997
Contribuciones	9 337	11 488	15 914

FUENTE: Secretaría de Finanzas de DDF.

Solución

De 1995 a 1996, las contribuciones crecieron 23.04%

$$\text{Incremento}_{95\text{-}96} = [(11{,}488 - 9{,}337)/\ 9{,}337]\ (100)$$

De 1996 a 1997, las contribuciones crecieron 38.53%

$$\text{Incremento}_{96\text{-}97} = [(15{,}914 - 11{,}488)/11{,}488]\ (100)$$

- Determine el crecimiento de la población económicamente activa por género de 1999 a 2004.

Población económicamente activa según sexo,[1] *1999 a 2004*

AÑO	TOTAL	HOMBRES	MUJERES
1999	39 648 333	26 295 840	13 352 493
2000	40 161 543	26 418 355	13 743 188
2001	40 072 856	26 415 550	13 657 306
2002	41 085 736	26 888 135	14 197 601
2003	41 515 672	27 277 029	14 238 643
2004	43 398 755	28 013 539	15 385 216

[1] Con el fin de ofrecer una serie anual amplia y comparable, este tabulado presenta información sólo del segundo trimestre de cada año. Los datos de los demás trimestres, incluyendo los más recientes, se pueden consultar en los productos disponibles de esta encuesta.

FUENTE: INEGI/STPS, Encuesta Nacional de Empleo.

- Determinar el porcentaje de votación obtenido por cada partido en la delegación Coyoacán, según elección.

CUADRO E8

Votación según elección en la delegación Coyoacán

PARTIDO POLÍTICO	JEFE DE GOBIERNO 2000	JEFE DELEGACIONAL 2000	JEFE DELEGACIONAL 2003
PRI	76 822	85 798	26 759
PRD	129 843	132 215	111 042
PT	5 973	8 715	4 179
APC (PAN-PV)[1]	122 475	105 030	61 976
PV			15 538
Otros partidos		22 477	16 265
CDPPN	1 046	1 592	
PCD	2 754	4 493	
PSN	414	619	607
PARM	979	1 722	
PAS	556	811	698
DSPPN	12 631	10 734	
CC	2 569	2 506	
Convergencia			3 364
Otros partidos[2]			11 596
Votos BLANCOS	1 351	2 274	2 241
Votos NULOS	3 685	3 853	6 256
Votos CC	143 155	150 951	
TOTAL	361 098	382 839	260 521

[1] En 2003 únicamente PAN.
[2] Otros partidos: MP, PLM, FC, CC3.

FUENTE: elaboración propia con datos del *Atlas digital electoral 2005. Resultados 1999-2003*, Instituto Electoral del Distrito Federal, agosto 2005, México.

Distribución de frecuencias y gráficas en estadística

HEMOS COMENTADO que los datos analizados en estadística descriptiva pueden presentarse comúnmente mediante tres formas básicas: textual, cuadros estadísticos y gráficas. En los capítulos anteriores se han dado elementos que permiten construir cuadros estadísticos y obtener de ellos ciertas medidas de comparación (porcentajes, razones, incrementos, etcétera); sin embargo, es necesario formalizar la construcción de los cuadros estadísticos y las medidas de condensación que de ellos se pueden obtener por medio de lo que llamaremos una distribución de frecuencias.

Una distribución de frecuencias o tabla de frecuencias no es más que la presentación tabular de las frecuencias con que ocurre cada característica (subclase o categoría) en las que ha sido dividida una variable. Esta característica puede estar determinada por una cualidad o un intervalo, por lo tanto, la construcción de un cuadro de frecuencia o tabla de frecuencias puede desarrollarse tanto para una variable cuantitativa como para una variable cualitativa.

Distribución de frecuencias para variables cuantitativas

Recordemos que las variables cuantitativas o métricas pueden ser de dos tipos: continuas o discretas. En el primer caso la construcción de una tabla de distribución de frecuencias requiere de la aplicación de un proceso simple y de la definición de algunos conceptos. En el segundo este proceso es aún más sencillo.

Distribución de frecuencias para variables cuantitativas continuas

Cuando la variable es continua, la construcción de una tabla de frecuencias presenta como su punto de mayor importancia la determinación del número de intervalos o clases que la formarán. Una clase o intervalo de clase es el elemento en la tabla que permite condensar en mayor grado un conjunto de datos con el propósito de hacer un resumen de ellos. El número de casos o mediciones que quedan dentro de un intervalo recibe el nombre de frecuencia del intervalo, y generalmente se denota como f_i. La diferencia entre el extremo mayor y el extremo menor del intervalo se llama longitud o ancho del intervalo. Para construir una tabla de distribución de frecuencias de una variable continua es conveniente utilizar los pasos siguientes:

- Se determina, primeramente, el número de intervalos o clases en la tabla, en función al número de datos a condensar, para ello podemos utilizar dos criterios de selección:

  El primero consiste en que el investigador selecciona el número de intervalos o clases con base en el número de datos a clasificar utilizando la tabla siguiente:

NÚMERO DE DATOS A CLASIFICAR	NÚMERO DE INTERVALOS
De 10 a 100	De 4 a 8
De 100 a 1 000	De 8 a 11
De 1 000 a 10 000	De 11 a 20

  El segundo consiste en calcular la fórmula de Sturges, que determina un número aproximado de intervalos "k". Aunque ésta no siempre resulta muy adecuada, es una relación muy utilizada.

$$k = 1 + 3.322 \log(n)$$

  donde:

  n es el número de datos a condensar en la tabla.

- Una vez seleccionado el número de intervalos "k", se procede a determinar su longitud, ancho o tamaño del intervalo (t_i). Observe que esta longitud es la misma para todos los intervalos en la tabla de frecuencias. Esto último es con la finalidad de facilitar los cálculos mediante métodos simplificados.

$$t_i = (\text{dato mayor - dato menor}) / k$$

Si el valor de t_i no es entero, el investigador puede usar la fracción o bien seleccionar el número par más cercano a este cociente.

Los intervalos no deben ser muy grandes al grado que enmascaren la distribución, ni tan pequeños que casi no contribuyan a facilitar los cálculos.

La diferencia entre el dato mayor y el menor del conjunto que se analiza recibe el nombre de amplitud.

Cuando el investigador desee utilizar intervalos con anchos desiguales deberá considerar que los cálculos serán más laboriosos por no ser aplicables los métodos simplificados que son los utilizados en este libro.

- Una vez determinado el número de intervalos y su tamaño, el paso siguiente consiste en indicar el límite inferior de la primera clase, el cual puede ser un valor igual o ligeramente menor al dato de valor mínimo del conjunto de datos. Una vez hecho esto, le sumamos el valor del ancho del intervalo para fijar el límite superior de esta clase considerando en ello los valores de los límites.
  Indicamos el límite inferior de la segunda clase agregando una unidad al límite superior de la primera clase. El límite superior de esta segunda clase será la suma del ancho del intervalo al límite superior de la clase anterior.
  A partir de esta dinámica se construyen todos los intervalos de clase; debemos tomar en consideración que, con esta técnica de construcción, siempre, el primer intervalo contiene al menor de los datos y el último al mayor.
  Los intervalos hasta aquí construidos reciben el nombre de intervalos de clase o intervalos ficticios.
- Posteriormente se construyen los intervalos reales de clase. Para ello, se resta media unidad a los límites inferiores de los intervalos ficticios (falsos) y se agrega media unidad a los límites superiores de los mismos.
- Con el establecimiento de los límites reales de clase en la tabla, se efectúa la clasificación de los datos en cada intervalo para determinar así la frecuencia de cada clase (f_i).
- Finalmente se construye la tabla de frecuencias definitiva, la cual contiene, en la primera columna, la clase; en la segunda, los intervalos reales de clase y, en la tercera, las frecuencias de clase, también llamadas frecuencias absolutas.

Para ejemplificar el proceso de construcción de una tabla de distribución de frecuencias de una variable continua utilizaremos los datos de las calificaciones obtenidas por 25 estudiantes en un curso de estadística.

calificaciones:

9, 7.5, 8, 7, 8, 7, 6.3, 9, 6, 5, 8, 8, 6.5, 6,
8, 8, 7, 7.4, 7.6, 5.5, 9, 7.2, 9, 7.4, 8.4

1. Determinamos el número de intervalos de clase. En este caso usaremos la fórmula de Sturges: $k=1+3.322 \log(n)$, n= número de datos.

$$k = 1 + 3.322 \log (25)$$

$$k = 1 + 3.322\ (1.39794)$$

$$k = 1 + 4.6439 = 5.64, \text{ es decir,}$$
tomamos $k = 6$ intervalos

2. Determinamos el ancho de los intervalos.

Dato mayor = 9 Dato menor = 5 k = 6

$t_i = (9 - 5)/6 = 0.66$, es decir, tomamos $t_i = 0.7$

3. Creamos el primer intervalo y los intervalos de clase sucesivos (intervalos falsos o ficticios).

5 a 5.6 (5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6) Lim. Inf. = 5, Lim. Sup. = 5.6
5.7 a 6.3 Lim. Inf = 5.6 + 0.1 = 5.7, Lim. Sup = 5.6 + 0.7 = 6.3
6.4 a 7.0 Lim. Inf = 6.3 + 0.1 = 6.4, Lim. Sup = 6.3 + 0.7 = 7.0
7.1 a 7.7 .....
7.8 a 8.4
8.5 a 9.1

4. Creamos los intervalos reales de clase, en este caso, restando y agregando 0.05 unidades, respectivamente, a cada límite de los intervalos ficticios.

4.95 - 5.65
5.65 - 6.35
6.35 - 7.05
7.05 - 7.75
7.75 - 8.45
8.45 - 9.15

5. Clasificamos los datos en los intervalos reales y creamos la tabla de distribución de frecuencias.

CLASE	INTERVALO REAL DE CLASE	FRECUENCIA (F_I)
1	4.95 - 5.65	2
2	5.65 - 6.35	3
3	6.35 - 7.05	4
4	7.05 - 7.75	5
5	7.75 - 8.45	7
6	8.45 - 9.15	4

La elaboración de una tabla de distribución de frecuencias generalmente se complementa con el cálculo de los siguientes elementos:

Marca de clase (m_i): constituida por el punto medio del intervalo de clase. Para calcularla es necesario sumar los dos límites del intervalo real, y dividir esta suma entre dos.

Frecuencia acumulada de la clase i (F_i): se llama frecuencia acumulada de la clase *i* al número resultante de sumar la frecuencia de la clase *i* con la frecuencia de las clases que la anteceden. Se denota generalmente como F_i. La última clase o intervalo en la tabla de frecuencias contiene como frecuencia acumulada el total de los datos.

Este cálculo tiene como objetivo informar del número de datos que se hallan distribuidos en los intervalos que anteceden al intervalo *i*, incluido éste.

Frecuencia relativa de la clase i (f_i/n): es el cociente entre la frecuencia absoluta de la clase *i* (f_i) y el número total de datos (n). Se expresa matemáticamente como

$$f_i/n = \text{frecuencia en la clase } i \text{ / total de datos}$$

Esta frecuencia muestra la proporción del número de casos que se han presentado en el intervalo *i* respecto del total de casos en la investigación. Si a este cociente se le multiplica por 100 entonces obtenemos una frecuencia relativa para cada clase expresada como porcentaje; esta última frecuencia, en %, permite hacer un análisis del comportamiento de los datos.

Frecuencia acumulada relativa de la clase i (F_i/n): es el cociente entre la frecuencia acumulada de la clase *i* (F_i) y el número total de datos(n). Se expresa matemáticamente como

$$F_i/n = \text{frecuencia acumulada en la clase } i \text{ / total de datos}$$

Esta frecuencia muestra la proporción del número de casos que se han acumulado hasta el intervalo *i* respecto del total de casos en la investigación. Si a este cociente se le multiplica por 100, entonces obtenemos un porcentaje denominado frecuencia acumulada relativa porcentual (acumulado porcentual). La acumulación de esta proporción en el último intervalo mostrará un porcentaje de 100%, es decir, la acumulación de todos los datos.

Con base en la definición de estos elementos la tabla de distribución de frecuencia de nuestro ejemplo quedaría reestructurada de la forma siguiente:

CUADRO 10

Clase	Intervalo de clase	f_i	m_i	F_i	f_i/n	F_i/n	fi %	Fi %
1	4.95 - 5.65	2	5.3	2	2/25	2/25	8	8
2	5.65 - 6.35	3	6.0	5	3/25	5/25	12	20
3	6.35 - 7.05	4	6.7	9	4/25	9/25	16	36
4	7.05 - 7.75	5	7.4	14	5/25	14/25	20	56
5	7.75 - 8.45	7	8.1	21	7/25	21/25	28	84
6	8.45 - 9.15	4	8.8	25	4/25	25/25	16	100

Si utilizamos porcentajes, podemos observar que sólo 8% de los estudiantes reprobó el examen de estadística, y que 12% obtuvo una calificación prácticamente en el límite si se considera una escala aprobatoria de 6 a 10. Sin embargo, 80% de los estudiantes aprobaron el examen citado.

De los estudiantes aprobados, 16% obtuvo una calificación suficiente (6.35 a 7.05); 20% una calificación regular, entre 7.05 y 7.75; 28% una buena calificación (7.75 a 8.45) y sólo 16% una calificación superior a 8.45. 64% de los alumnos del grupo obtuvieron una calificación superior a 7.

Distribución de frecuencias para variables cuantitativas discretas

En el caso de *variables discretas*, la construcción de una tabla de distribución de frecuencias sigue los lineamientos establecidos para una variable continua con la salvedad de que en este tipo de tablas no existen intervalos ni marcas de clase, lo cual simplifica su construcción. La tabla de frecuencias para variables discretas clasificará en la primera columna las subclases de la variable, en la siguiente indicará los casos o frecuencias en ellas, en la tercera calculará la frecuencia relativa, en la cuarta la frecuencia acumulada y en la quinta la frecuencia acumulada relativa, como se muestra en el Cuadro 11 del ejemplo siguiente.

Suponga que se cuenta con los datos del número de hijos por familia (350 familias) en el municipio de Chalco.

Datos (hijos por familia): 1, 4, 4, 3, 2, 1, 4, 3, 3, 2...

CUADRO 11

Número de hijos por familia en Chalco, Estado de México, 1995

Número de hijos	frecuencia (fi)	fi / n	Fi	Fi / n
1	20	20/350	20	20/350
2	120	120/350	140	140/350
3	200	200/350	340	340/350
4	10	10/350	350	350/350

FUENTE: datos hipotéticos.

Es importante recordar que las frecuencias relativas se pueden expresar como proporciones (como en el ejemplo) o como porcentajes.

Para la variable del Cuadro 11 puede indicarse que, en una muestra de 350 familias, sólo 5.7% de ellas tiene un hijo; 34% dos hijos; 57% tres y, sólo 3% cuatro hijos. Esto permite indicar que 97% de las familias muestreadas tienen tres hijos o menos.

Como puede observarse de los ejemplos de la variable continua y de la variable discreta, el uso de las tablas de distribución de frecuencias no sólo permite hacer una condensación de los datos sino también desarrollar con ello una primera interpretación o análisis de los mismos.

Distribución de frecuencias para variables cualitativas

La construcción de tablas de frecuencia para variables cualitativas o no métricas requiere sólo del conteo del número de elementos o individuos que caen dentro de cierta cualidad o bien dentro de determinada característica. En estos casos la tabla se construye de la manera siguiente:

- En la primera columna se registran las cualidades o características.
- En la segunda columna se anotan las frecuencias absolutas.
- En la tercera columna se registran las frecuencias relativas.

NOTA: Para datos cualitativos no existen intervalos de clase ni frecuencias acumuladas ya que ello carecería de sentido.

Como ejemplo de distribución de frecuencias para datos cualitativos mostraremos la estructura por licenciatura del grupo SB09-05/O de la UAM-Xochimilco (Cuadro 12).

CUADRO 12
Alumnos por licenciatura del grupo SB09-05/O, UAM-Xochimilco

LICENCIATURA	ALUMNOS (f_i)	f_i / n (%)
Administración	4	4/22 (18)
Economía	2	2/22 (9)
Psicología	2	2/22 (9)
Sociología	4	4/22 (18)
Comunicación	10	10/22 (46)

FUENTE: datos hipotéticos.

En el ejemplo, 46% de los alumnos del grupo cursarán la licenciatura en comunicación; 18% administración; 18% sociología; y sólo 9% economía y psicología.

Gráficas

Como complemento a este primer análisis que realiza el investigador por medio de las tablas de distribución de frecuencias, existe la posibilidad de construir gráficas de diversos tipos que le permiten explicar más fácilmente el comportamiento de los datos estudiados. Una gráfica permite mostrar, explicar, interpretar y analizar de manera clara y efectiva los datos estadísticos mediante formas geométricas tales como líneas, áreas, volúmenes, superficies, etcétera. Las gráficas permiten además la comparación de magnitudes, tendencias y relaciones entre los valores que adquiere una variable.

Las gráficas tienen gran utilidad como medios de divulgación del análisis estadístico, ya que las relaciones visuales se captan con facilidad y resulta sencillo recordarlas.

Histogramas y polígonos de frecuencias

Un histograma de frecuencias es un gráfico de rectángulos continuos que tienen su base en el eje de las abscisas (eje horizontal o eje de las equis); con anchura igual cuando se trata de representar el comportamiento de una variable continua. En este caso el punto central de la base de los rectángulos equivale al punto medio de cada clase (marca de clase).

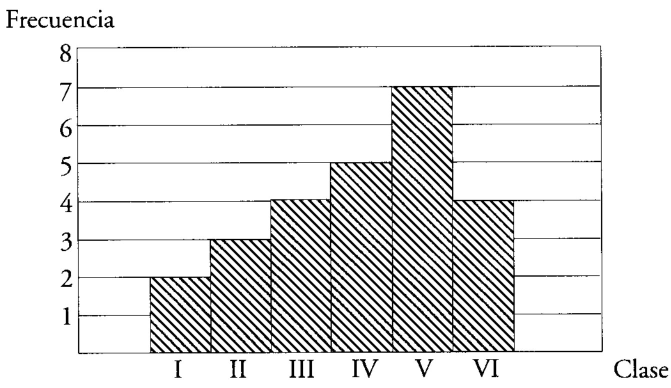
Las alturas de los rectángulos ubicadas en el eje de la ordenadas (de las Y o eje vertical) corresponden a las frecuencias de las clases.

El área de los rectángulos así formados es proporcional a las frecuencias de las clases.

Los histogramas de frecuencias pueden construirse no sólo con las frecuencias absolutas, sino también con las frecuencias acumuladas y las frecuencias relativas. En este último caso el histograma recibe el nombre de histograma de frecuencias relativas, histograma de porcentajes o histograma de proporciones, según el caso.

GRÁFICA 1

Calificaciones de 25 estudiantes en un curso de estadística



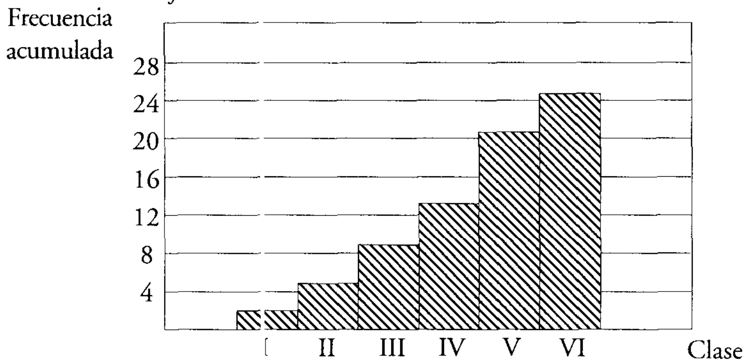
I.	4.95 - 5.65	IV.	7.05 - 7.75
II.	5.65 - 6.35	V.	7.75 - 8.45
III.	6.35 - 7.05	VI.	8.45 - 9.15

Histograma de frecuencias

FUENTE: datos del Cuadro 10.

GRÁFICA 2

Calificaciones de 25 estudiantes en un curso de estadística



I.	4.95 - 5.65	IV.	7.05 - 7.75
II.	5.65 - 6.35	V.	7.75 - 8.45
III.	6.35 - 7.05	VI.	8.45 - 9.15

Histograma de frecuencias acumuladas

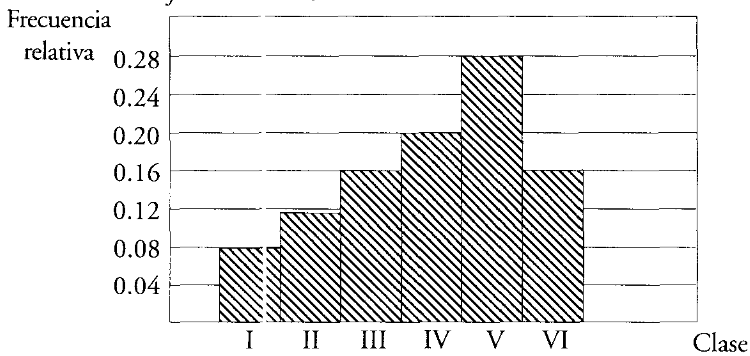
FUENTE: datos del Cuadro 10.

Los dos primeros histogramas, el de frecuencia absoluta y el de frecuencia acumulada (gráficas 1 y 2), se forman con los intervalos de clase en el eje de las abscisas y las frecuencias absoluta o acumulada en el eje de las ordenadas, respectivamente.

Para el histograma siguiente (Gráfica 3), la frecuencia que se usa en el eje de las ordenadas es la relativa, lo que implica tener un histograma de frecuencias relativas.

GRÁFICA 3

Calificaciones de 25 estudiantes en un curso de estadística



I.	4.95 - 5.65	IV.	7.05 - 7.75
II.	5.65 - 6.35	V.	7.75 - 8.45
III.	6.35 - 7.05	VI.	8.45 - 9.15

Histograma de frecuencias acumuladas

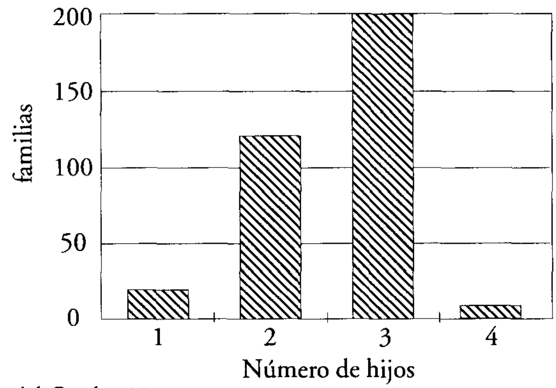
FUENTE: datos del Cuadro 10.

También con los datos del Cuadro 10 es posible construir un histograma de frecuencias acumuladas relativas.

A continuación se muestra un gráfico (Gráfica 4) con los datos de la variable discreta del Cuadro 11, es decir, un gráfico de columnas simples, ya que en estos casos, la variable no presenta continuidad y por lo tanto su gráfico no se puede llamar histograma. Las variables discretas se pueden representar gráficamente mediante gráficos de columnas o de barras simples.

GRÁFICA 4

Hijos por familia en Chalco, Estado de México, 1995



FUENTE: datos del Cuadro 11.

Los histogramas y las gráficas de columnas tienen como función mostrar no únicamente una representación visual de los datos sino, fundamentalmente, tres características de su comportamiento. Éstas son:

1. La forma o distribución que presenta el histograma o gráfica de barras (según sea la variable); es decir, si hay simetría o asimetría en la distribución de los datos.
2. Si existe algún intervalo o barra donde se acumulan los datos o aparece una tendencia posicional de los mismos.
3. Principalmente en el caso de variables cuantitativas continuas, el grado de dispersión o variabilidad de los datos.

En el ejemplo de las calificaciones (de 25 estudiantes en un curso de estadística, Gráfica 3) puede observarse que 64% de los datos tienden a acumularse en los intervalos IV, V y VI, lo que nos indica que no hay una simetría en la distribución, sino un pequeño sesgo hacia la derecha de la misma (calificaciones buenas). Por otro lado la dispersión de los datos no es muy grande.

En el ejemplo de la variable discreta, número de hijos por familia en Chalco, es posible observar que los datos se concentran en los casos de las familias con dos y tres hijos (Gráfica 4). La variabilidad en estos datos respecto de los casos de dos y tres hijos tampoco es muy grande.

En Excel el proceso de construcción de un histograma puede simplificarse con los siguientes pasos:

1. Se utilizan dos columnas de la hoja electrónica, la primera deberá contener los intervalos reales de clase y la segunda la frecuencia absoluta (f_i), la frecuencia acumulada (F_i) o la frecuencia relativa, según el histograma que se desea construir.
2. Se selecciona el conjunto de celdas de estas dos columnas (rango con la información de la gráfica), y se oprime el botón del asistente que permite crear gráficos (Figura 4).

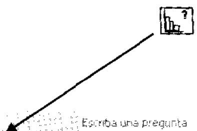
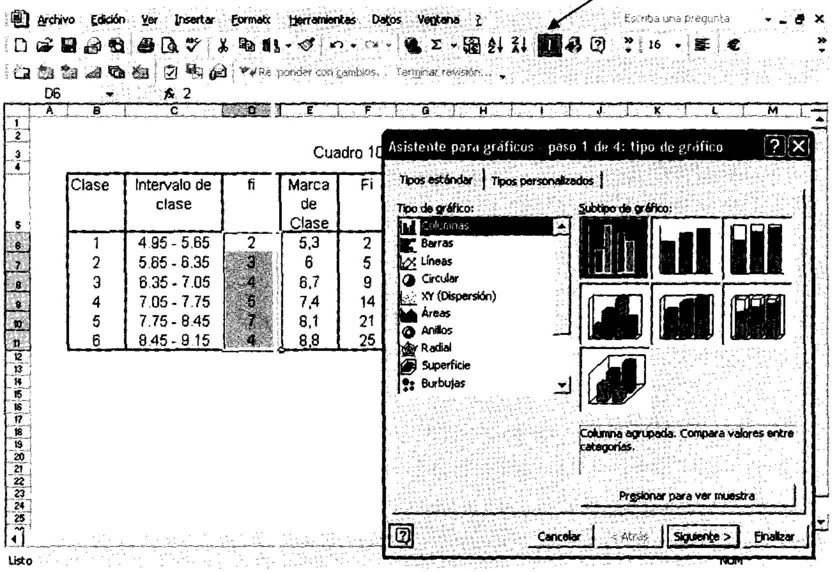
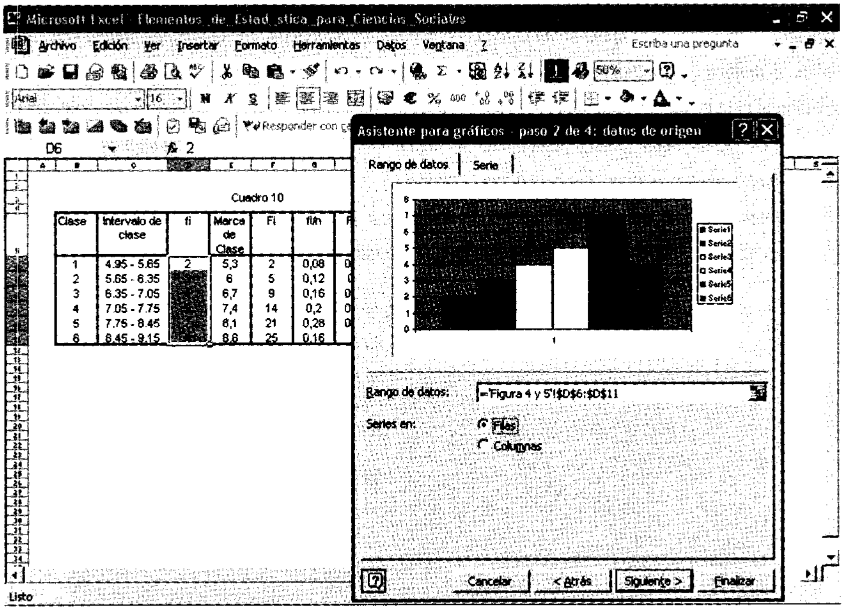


FIGURA 4



3. Aparece el paso 1 del asistente de gráficas. Seleccionamos la opción estándar, es decir, gráfica de columnas, subtipo de gráfica 1. Oprimimos el botón "Siguiente" para pasar al segundo paso del asistente.
4. En este segundo paso deberá seleccionarse la opción de "filas". Posteriormente en la categoría de "Serie", en la opción definida como: Rótulos del eje de categorías (X), dar un espacio con la barra espaciadora y oprimir "Siguiente" (véase Figura 5).

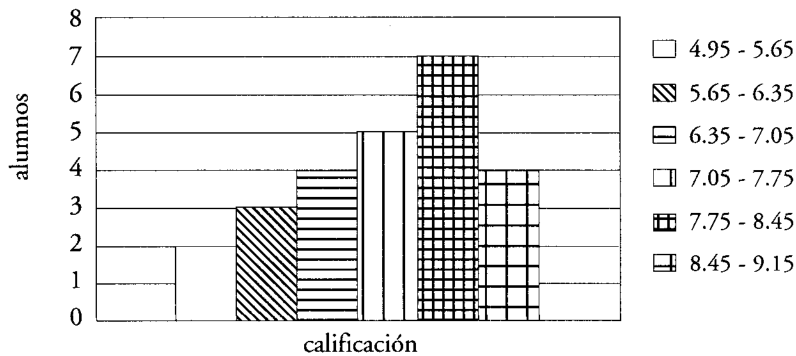
FIGURA 5



5. En el paso 3 del asistente, deberemos especificar el título de la gráfica, la etiqueta del eje X y la del eje de las Y. Posteriormente oprimir “Siguiente”.
6. En el paso 4 (último) del asistente, deberemos indicar si queremos que el gráfico se agregue como un objeto en la misma hoja electrónica o bien se agregue en otra hoja nueva. En este momento oprimimos “Finalizar” y con ello obtenemos un histograma como el que se muestra en la Gráfica 4a.

GRÁFICA 4A

Calificaciones del examen de estadística

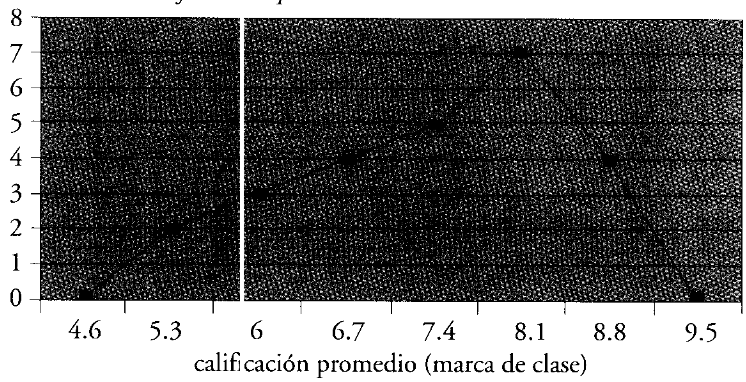


Polígono de frecuencias

Es un gráfico de línea que se construye sobre el sistema de coordenadas cartesianas al colocar sobre cada marca de clase un punto a la altura igual a la frecuencia absoluta asociada a esa clase; posteriormente, estos puntos se unen por segmentos de recta. Para que el polígono quede cerrado se debe considerar un intervalo ficticio más al inicio y otro al final con frecuencias cero.

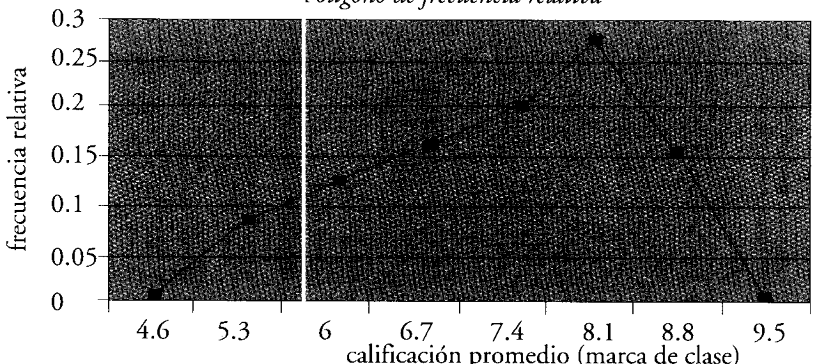
A continuación se muestra, en la Gráfica 5, el polígono de frecuencias de las calificaciones de los 25 estudiantes de un curso de estadística (Cuadro 10).

GRÁFICA 5
Calificaciones promedio del examen de matemáticas



Los polígonos de frecuencia también se pueden construir utilizando las frecuencias relativas de una distribución de frecuencias; estos gráficos se denominan polígonos de frecuencias relativas (Gráfica 6) o polígonos porcentuales, si están expresados en porcentajes.

GRÁFICA 6
Polígono de frecuencia relativa

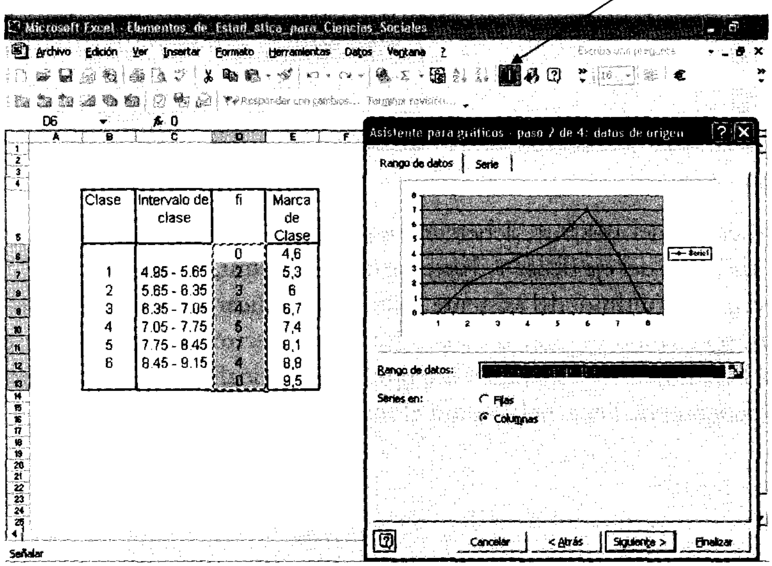


FUENTE: datos del Cuadro 10.

En Excel, la construcción de los polígonos de frecuencias es relativamente simple:

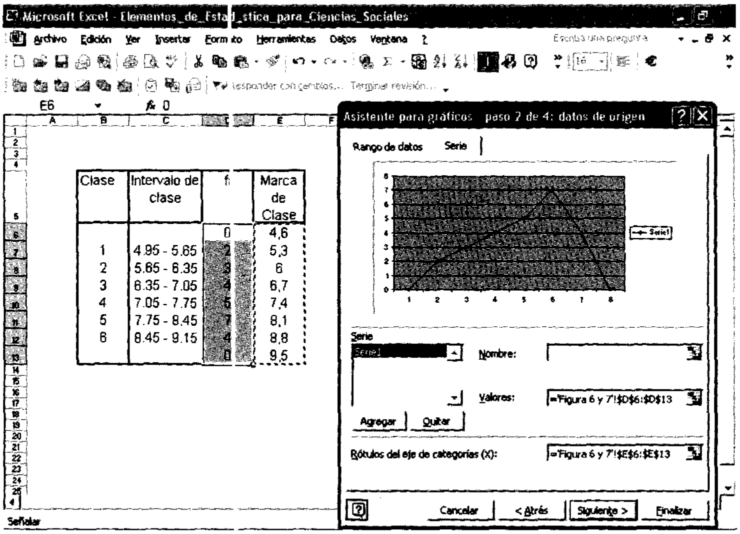
1. Se utiliza una columna de la hoja electrónica, la cual deberá contener la frecuencia absoluta o la frecuencia relativa, según el polígono que se desea construir. A estas dos series deberá agregarse una marca de clase antes de la primer clase con frecuencia cero, y otra más, al final, también con frecuencia cero. Esto último se hace con la finalidad de tener un gráfico cerrado.
2. Se selecciona el conjunto de celdas que forman la columna (rango con la información de la gráfica), y se oprime el botón del asistente que permite crear gráficos (véase Figura 6).

FIGURA 6



3. Aparece el paso 1 del asistente de gráficas. Seleccionamos la opción "líneas", subtipo de gráfica 1, 2 o bien la 4. Oprimimos el botón "siguiente" para pasar al segundo paso del asistente.
4. En este segundo paso deberá estar seleccionada la opción de "Serie". Posteriormente en la categoría de "Rótulos del eje de categorías (X)", deberán seleccionarse las celdas que corresponden a las "Marcas de clase", como se muestra en la Figura 7. Finalmente, como último paso en esta etapa debemos oprimir "Siguiente".

FIGURA 7



5. En el paso 3 del asistente, deberemos especificar el título de la gráfica, la etiqueta del eje X y la del eje de las Y. Posteriormente oprimir, "Siguiente".
6. En el paso 4 (último) del asistente, deberemos indicar si queremos que el gráfico se agregue como un objeto en la misma hoja electrónica o bien se agregue en otra hoja nueva. En este momento oprimimos "Finalizar" y con ello obtenemos un polígono de frecuencias como el mostrado en la Gráfica 5.

Ojivas

La gráfica que se construye para una frecuencia acumulada o una frecuencia acumulada relativa se llama ojiva.

La ojiva es un polígono abierto en el extremo superior, que se obtiene al unir por segmentos de recta, los puntos situados a una altura igual a la frecuencia acumulada o la frecuencia acumulada relativa de cada clase (eje Y) con los límites reales superiores de éstas (eje X). Para llevar a cabo su construcción se requiere crear un primer intervalo ficticio (o falso) con frecuencia acumulada cero. Este gráfico nos permite analizar cuántas observaciones están por debajo de un determinado valor.

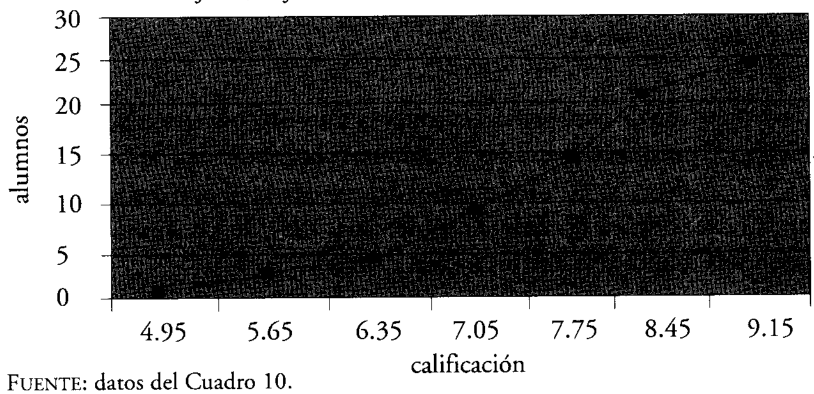
La ojiva es un polígono abierto que, como ya indicamos, se puede construir no sólo con la frecuencia acumulada, sino también con la frecuencia acumulada relativa, o bien, con la frecuencia acumulada relativa expresada en porcentaje (ojiva porcentual).

Con los datos del Cuadro 10 se construyen dos ojivas, la ojiva (con la frecuencia acumulada) y la ojiva porcentual, gráficas 7 y 8 respectivamente.

Para construir la Gráfica 7, debemos crear un intervalo ficticio inicial (4.25 - 4.95) cuya frecuencia acumulada es cero, por lo que debajo del límite real superior (4.95) del intervalo ficticio no hay un acumulado de alumnos. Debajo del segundo límite real superior (5.65) hay acumulados dos alumnos; debajo del tercer límite real superior (6.35) hay acumulados cinco alumnos; debajo del cuarto límite real superior (7.05) hay acumulados nueve alumnos; y así sucesivamente, hasta el último intervalo debajo del cual se acumulan las 25 calificaciones.

GRÁFICA 7

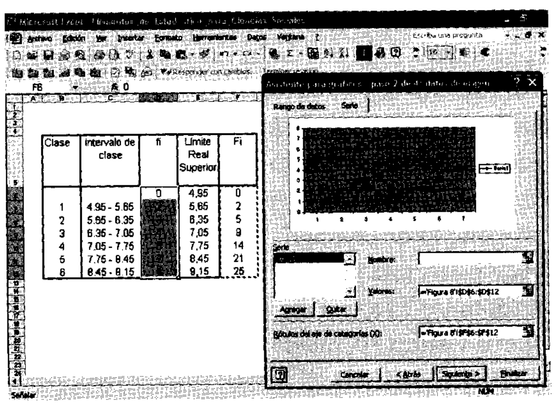
Ojiva (calificaciones del examen de matemáticas)



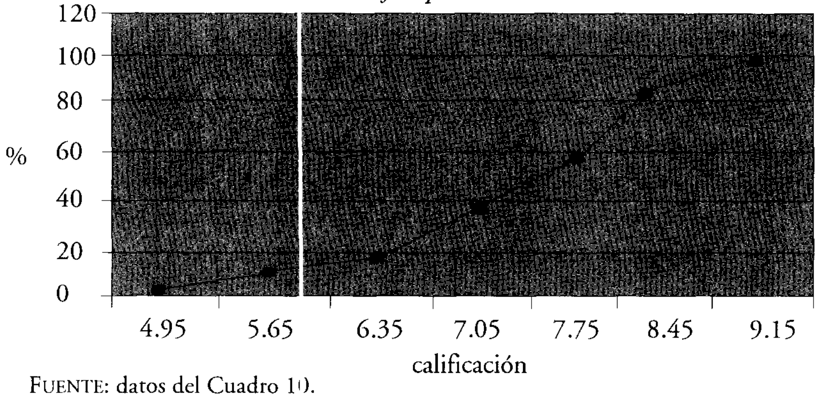
FUENTE: datos del Cuadro 10.

La construcción del gráfico anterior en Excel, sigue la mecánica utilizada para la construcción de un polígono de frecuencias, con la salvedad de que en este caso, se utilizan las frecuencias acumuladas o las frecuencias acumuladas relativas (Figura 8) y los límites reales superiores de cada clase.

FIGURA 8



GRÁFICA 8
Ojiva porcentual



FUENTE: datos del Cuadro 10.

La interpretación de estos gráficos es simple y muy útil; por ejemplo, de la primera ojiva (Gráfica 7) puede observarse que 14 alumnos obtuvieron una calificación inferior o igual a 7.75, es decir, 56% de los alumnos del curso (este último dato se obtuvo con la segunda ojiva, Gráfica 8) Sólo cinco de ellos, 20% de los alumnos del grupo, obtuvieron una calificación inferior o igual a (6.35).

Gráficas de columnas (barras) simples para datos cualitativos

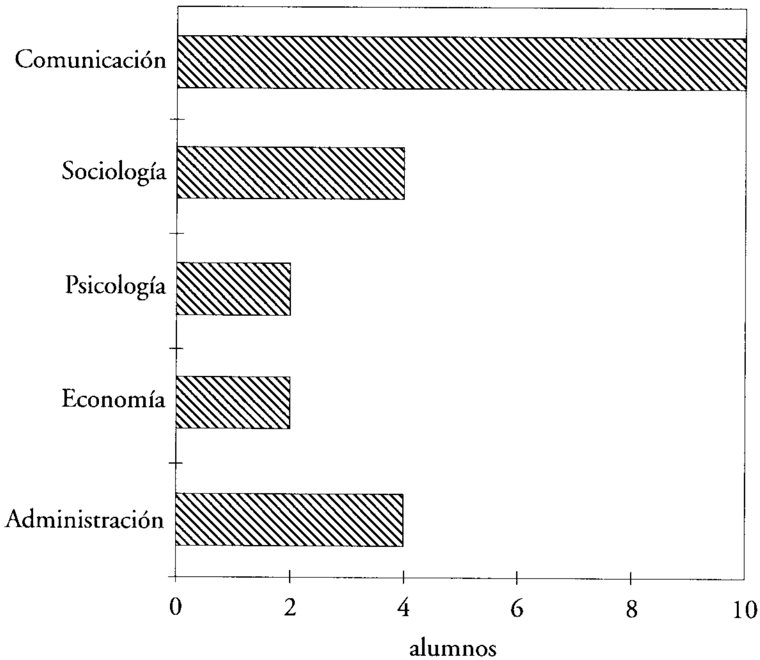
Como ya se indicó, las variables cualitativas no tienen intervalos de clase por carecer de sentido. Tampoco en ellas se calcula la frecuencia acumulada. Por lo tanto, para las variables cualitativas sólo existe la construcción de las gráficas de columnas (barras) simples, tanto para las frecuencias absolutas de cada categoría como para las relativas de éstas. *Para las variables cualitativas no existen el polígono de frecuencias ni las ojivas.*

Una gráfica de columnas simples para datos cualitativos está formada por rectángulos que representan a cada una de las categorías (características) que forman la variable y cuya altura estará determinada por la frecuencia absoluta o la frecuencia relativa que presente la categoría. Los rectángulos se dibujan separados para enfatizar que entre ellos existe una diferencia cualitativa y no cuantitativa. Los rectángulos en este gráfico pueden trazarse horizontal o verticalmente.

Para mostrar un ejemplo de este tipo de gráficos se utilizará la información del cuadro estadístico 12, "Estructura por licenciatura de los estudiantes del grupo SB09-05/O en la UAM-Xochimilco". La Gráfica 9 muestra un ejemplo de gráfica de barras simples, mientras que la Gráfica 10, muestra una gráfica de columnas simples para la misma variable.

GRÁFICA 9

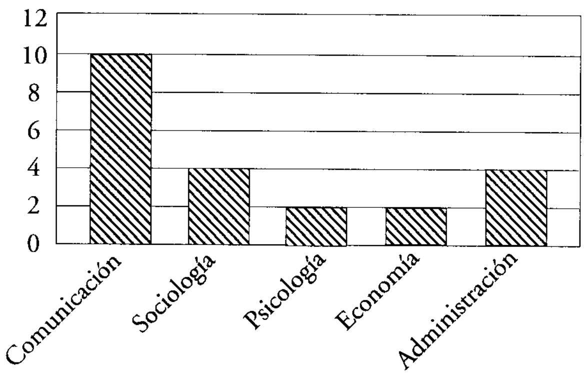
Estructura por licenciatura del grupo SB09-05/O, UAM-Xochimilco



FUENTE: datos del Cuadro 12.

GRÁFICA 10

Estructura por licenciatura del grupo SB09-05/O, UAM-Xochimilco



FUENTE: datos del Cuadro 12.

Otras representaciones gráficas

En estadística es muy común presentar los resultados de un estudio mediante el uso de gráficas, por eso no sólo encontraremos histogramas, polígonos de frecuencia, gráfica de columnas (barras) simples y ojivas, sino también otras formas gráficas de presentar datos y cálculos estadísticos. A continuación se enumeran algunos tipos de estas representaciones gráficas.

- Gráfica de barra o columna simple
- Gráfica de sectores
- Gráfica de barras agrupadas
- Gráfica de barras de desviaciones
- Mapas estadísticos
- Gráficas pictóricas

Gráfica de barra o columna simple

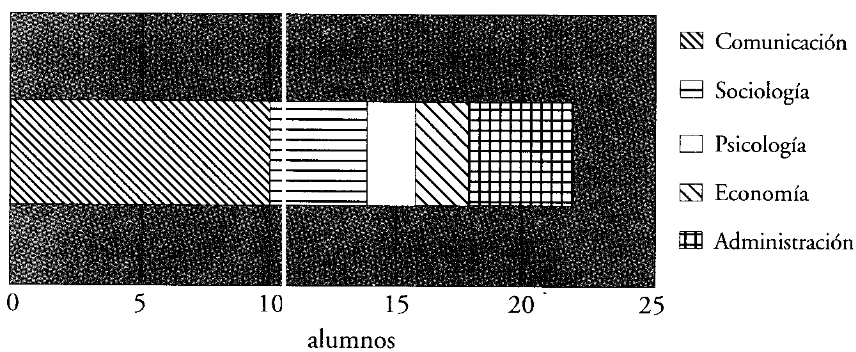
Es la más sencilla de las gráficas, y consiste en representar datos mediante una barra o columna simple, la cual puede ser colocada horizontal o verticalmente.

Este gráfico permite comparar las proporciones que guardan cada una de las partes respecto del todo, por lo que pueden construirse usando valores absolutos, proporciones o bien porcentajes. Se suelen utilizar cuando se comparan gráficamente las distribuciones de conceptos iguales en dos o más periodos.

Retomando los datos del cuadro estadístico 12, el gráfico de barra o columna simple con valores absolutos se muestra en la Gráfica 11.

GRÁFICA 11

Estructura por licenciatura del grupo SB09-05/O, UAM-Xochimilco



FUENTE: datos del Cuadro 12

Gráfica de sectores

Se emplean para mostrar la relación existente entre los componentes de un todo, es decir, la proporción. Cada uno de los sectores del círculo representa una parte de un agregado o de un total. Este tipo de gráfica recibe también el nombre de gráfica circular o gráfica de pastel.

Para construir estos gráficos, el analista deberá contar con la proporción o el porcentaje de cada una de las partes del todo y multiplicarla por 360° o 3.6°, respectivamente. Así, una proporción de 0.65 equivaldría a 234° en la gráfica, y un porcentaje de 25% equivaldría a 90°.

Para mostrar un ejemplo de estos gráficos se utilizarán los datos del cuadro estadístico 13, el cual trata sobre la exportación de vehículos por planta armadora en México en los años 1992 y 1993. Los datos de la Gráfica 12 corresponden a 1992. Se puede elaborar un gráfico similar con los datos de 1993.

CUADRO 13

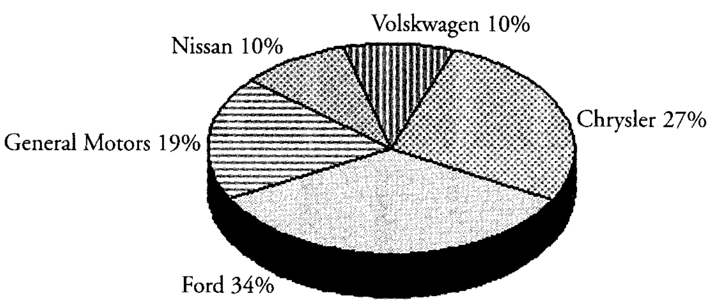
Exportaciones mexicanas de vehículos enero-octubre (unidades)

COMPAÑÍA	1992	1993
Chrysler	90 178	113 629
Ford	115 333	94 066
General Motors	62 227	73 348
Nissan	32 590	37 851
Volskwagen	32 573	60 487
Dina	0	340
TOTAL	332 901	379 721

FUENTE: Departamento de Información de Negocios de Banamex, diciembre de 1993.

GRÁFICA 12

Exportaciones mexicanas de vehículos, enero-octubre 1992



FUENTE: Departamento de Información de Negocios de Banamex, diciembre de 1993.

En el gráfico de sectores anterior se muestra la participación de cada planta armadora en las exportaciones de vehículos para el periodo enero-octubre de 1992.

Observe aquí que los elementos básicos usados para elaborar un cuadro estadístico (número de gráfica, título, fuente, etcétera) son aplicables también en la elaboración de gráficas.

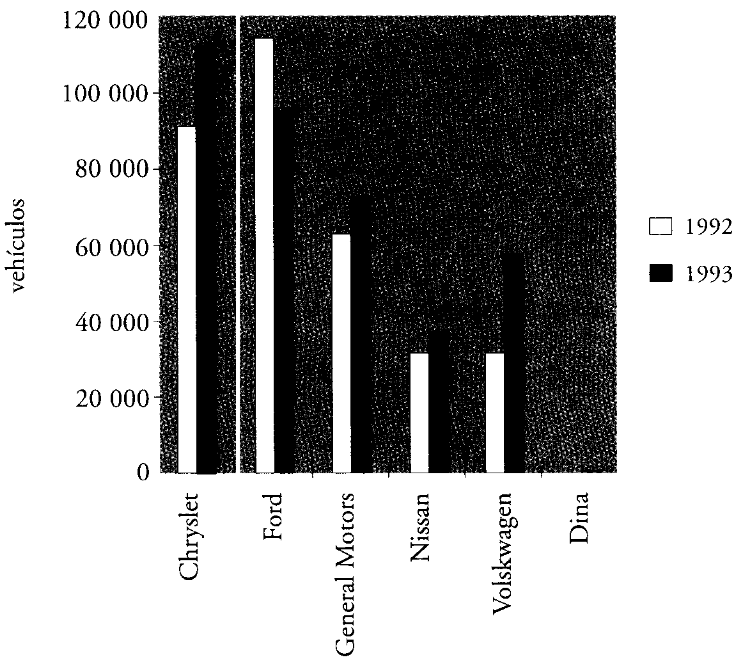
Gráfica de barras agrupadas

Se utilizan cuando se requiere comparar una variable que presenta diferentes categorías o bien cuando se desea comparar dos o más variables que, a su vez, se dividen en dos o más categorías. Como su nombre lo indica, están formadas por barras que se agrupan por categorías; las barras se pueden presentar en forma horizontal o vertical.

Utilizando la información del Cuadro 13 a continuación se muestra un gráfico de barras agrupadas que permite mostrar el comportamiento de las exportaciones de vehículos por planta armadora para los periodos enero-octubre de 1992 y 1993 en México.

GRÁFICA 13

Exportaciones mexicanas de automóviles



Gráficas de barras de desviaciones

Es un gráfico de barras que se puede elaborar vertical u horizontalmente. Siempre tiene como referencia el valor cero, a partir del cual se construyen las barras: hacia la derecha o izquierda si el eje es horizontal, o bien, hacia arriba o hacia abajo, si el eje considerado es el vertical. Este tipo de gráficas permite la representación de una variable que contiene variaciones positivas y negativas; por ejemplo, en una empresa las pérdidas y ganancias en el tiempo, los incrementos positivos y decrementos negativos de la producción, etcétera.

Como un ejemplo de este tipo de gráfico, se mostrará la información de crecimiento promedio anual del primer semestre para la industria minero-metalúrgica mexicana de los años 1985 a 1995 proporcionada por el Instituto Nacional de Estadística, Geografía e Informática (INEGI) (Gráfica 14).

CUADRO 13A

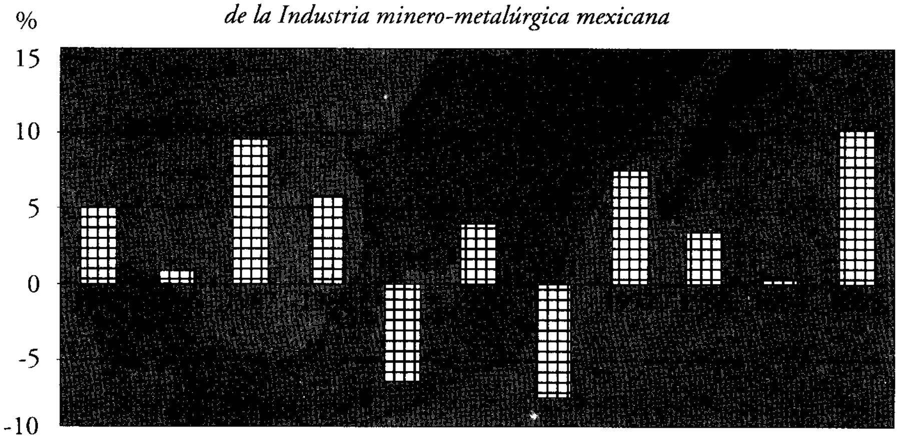
Crecimiento promedio anual del primer semestre de la Industria minero-metalúrgica mexicana (%)

1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
5	1	9	7	-7	4	-8	7	3	0.5	10.2

FUENTE: INEGI.

GRÁFICA 14

Crecimiento promedio anual del primer semestre de la Industria minero-metalúrgica mexicana



FUENTE: elaboración propia con datos del INEGI.

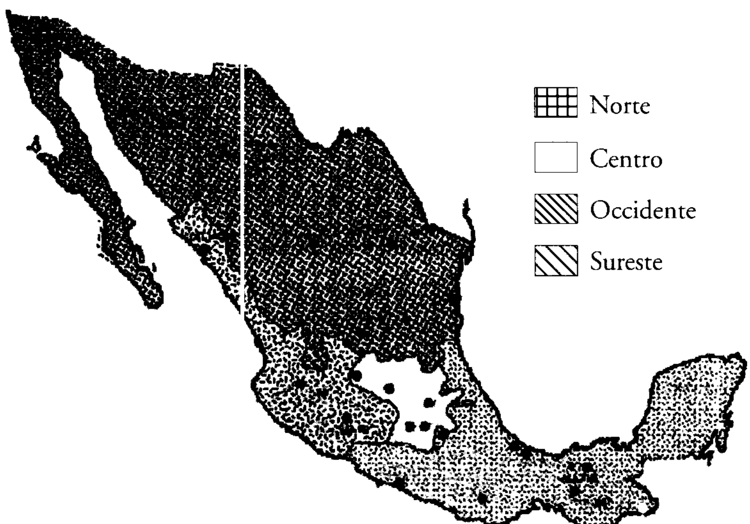
Mapas estadísticos

Tienen como objetivo representar relaciones espaciales en mapas geográficos. Entre algunos de los tipos más usados tenemos los mapas sombreados, punteados, de isolíneas, de gráficas sobrepuestas (barras, líneas, flujos, etcétera), y con combinaciones de dos o más de los anteriores tipos.

En los mapas sombreados la intensidad de la sombra indica la magnitud del fenómeno en el área; los punteados enfatizan las magnitudes absolutas. Los mapas de isolíneas se emplean frecuentemente para mostrar la distribución de fenómenos meteorológicos, económicos y demográficos. En el mapa de la República se muestra la forma en la que nuestro país fue dividido en cuatro regiones por los investigadores de Banamex en 1987 para elaborar la Segunda Encuesta Nacional sobre los Valores de los Mexicanos.

En los mapas de gráficas sobrepuestas es fácil representar las relaciones espaciales de variables, lo que permite una mejor comprensión de las mismas. Por ejemplo, si deseamos representar las magnitudes de emigración e inmigración en los estados del país, se pueden sobreponer dos gráficas de columnas en cada estado que indiquen dichas magnitudes.

Regiones y ciudades, muestreo de cuotas
II Encuesta Nacional sobre los Valores de los Mexicanos, 1987



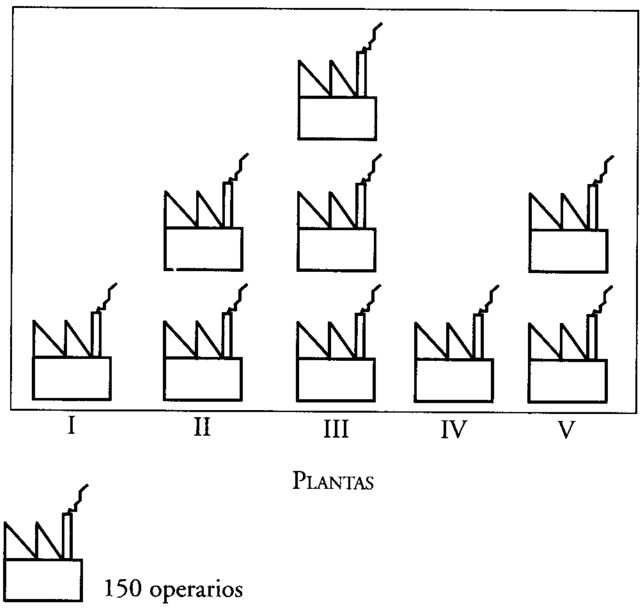
Gráficas pictóricas

Son muy empleadas por ser atractivas y de fácil comprensión para el público que no tiene una preparación estadística, por lo que en estudios formales no son utilizadas. Estas gráficas pueden ser de dos tipos.

Gráficas con símbolos pictóricos de tamaño proporcional. En ellas las figuras deben tener una magnitud proporcional a los valores que representan.

Gráficas de unidades pictóricas. En estas gráficas cada símbolo representa un valor definido y uniforme. Por ejemplo, si una unidad o símbolo representa 100 estudiantes, dos símbolos representan 200 estudiantes, etcétera. La Gráfica 15 muestra un ejemplo de este tipo de gráficas.

GRÁFICA 15
Número de operarios por planta



Ejercicios

- Se obtuvo una muestra de 20 niños en la delegación Tlahuac, y se les preguntó cuánto tiempo diario (en horas) veían la televisión. Los resultados obtenidos fueron:

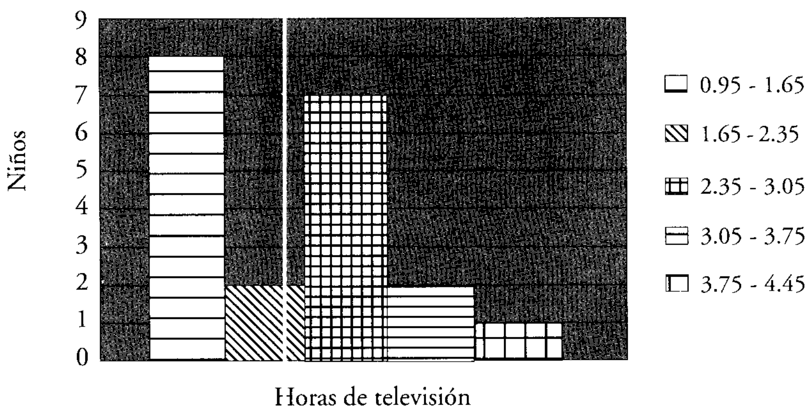
2.5, 3.0, 1.5, 1.5, 1.5, 3.5, 1.0, 2.5, 1.5, 1.0, 4.0, 1.5, 3.0, 3.5, 3.0, 2.0, 2.5, 2.0, 1.5, 3.0

Con estos datos construya una tabla de distribución de frecuencias, un histograma y un polígono porcentual.

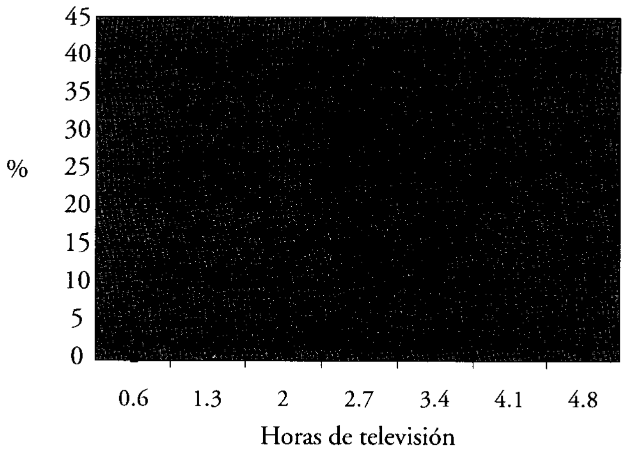
Solución:

HORAS DE TELEVISIÓN	NIÑOS	ACUMULADO	ACUMULADO (%)
0.95 - 1.65	8	8	40
1.65 - 2.35	2	10	50
2.35 - 3.05	7	17	85
3.05 - 3.75	2	19	95
3.75 - 4.45	1	20	100

HISTOGRAMA



POLÍGONO PORCENTUAL



- En una prueba de habilidad psicomotora aplicado a niños de primaria en la ciudad de Puebla, resultaron las puntuaciones siguientes:

34, 28, 29, 22, 33, 30, 31, 32, 30, 24, 22, 24, 29, 24, 28, 24, 34, 28, 29, 30, 32, 35, 33, 22, 19, 24, 18, 23, 20, 20, 22, 21, 20, 30, 31, 32

Con estos datos construya una tabla de distribución de frecuencias, un histograma, un histograma porcentual y un polígono de frecuencias.

- Las calificaciones en un examen de estadística en la UAM-Xochimilco se muestran en la siguiente tabla de frecuencias:

CLASE	INTERVALO REAL DE CLASE	FRECUENCIA (fi)
1	4.95 - 5.65	4
2	5.65 - 6.35	6
3	6.35 - 7.05	8
4	7.05 - 7.75	10
5	7.75 - 8.45	14
6	8.45 - 9.15	8

Construir un histograma, un histograma porcentual, un polígono de frecuencias y una ojiva.

- Los ingresos anuales de las familias en una comunidad del estado de Oaxaca se muestran en la siguiente tabla.

CLASE	INGRESO ANUAL (PESOS)	FRECUENCIA (fi)
1	4500 - 5000	4
2	5000 - 5500	9
3	5500 - 6000	10
4	6000 - 6500	7
5	6500 - 7000	3

Construir un histograma, un histograma porcentual, un polígono de frecuencias y una ojiva.

- El número de hermanos de los alumnos del grupo SC-05 en el trimestre 05/I se muestra en la siguiente tabla.

NÚMERO DE HERMANOS	FRECUENCIA (fi)
0	2
1	4
2	6
3	7
4	3
5 o más	3

Construir con estos datos una gráfica de barras simples y una gráfica de sectores.

- Se entrevistaron alumnos de ciencias sociales, con el fin de conocer su opinión acerca del nivel académico de sus cursos de matemáticas. Las respuestas a esta variable de investigación se clasificaron como: Bueno (1), Regular (2) y Bajo (3). Los resultados obtenidos en la entrevista son:

1, 3, 2, 2, 1, 2, 2, 3, 4, 2, 2, 1, 3, 2, 2, 3, 1, 1, 3, 1, 2, 1, 1, 3, 3, 3, 1, 2, 2, 3, 2,1, 3, 2, 2, 2, 3, 2, 1, 3, 2, 1, 1, 1, 3, 2, 2, 2, 3, 3, 3, 1, 1, 1, 2, 2

Elaborar un cuadro de distribución de frecuencias para esta variable. Trazar un gráfico de sectores y un gráfico de barra simple.

Medidas descriptivas de la distribución de frecuencias

HASTA EL MOMENTO, el análisis de un grupo de datos, ya sea en bruto o bien agrupados en un cuadro de distribución de frecuencias, se ha realizado por medio de algunos cálculos aritméticos simples (proporciones, porcentajes, índices, etcétera) o bien a partir de observar su comportamiento en una gráfica, donde sólo podemos discernir sus tendencias o patrones de comportamiento. Se podría hacer la pregunta siguiente: ¿existe alguna medida más exacta respecto del comportamiento de un conjunto de datos? La respuesta es sí, ya que podemos, en este punto, utilizar números individuales que permiten describir ciertas características de dicho conjunto, y que reciben, en estadística, el nombre de *estadísticos descriptivos.*

Los estadísticos descriptivos constituyen un resumen de la característica que se estudia, lo que permite tomar una decisión más rápida y satisfactoria sin la necesidad de consultar nuevamente todas las observaciones.

Los estadísticos se pueden clasificar, de acuerdo con sus características en:

1. Medidas de tendencia central y de posición
2. Medidas de dispersión
3. Medida de sesgo (asimetría)
4. Medida de curtosis

Las primeras tienen como objetivo buscar el punto medio o típico del conjunto de datos, o bien una ubicación específica dentro de una distribución de datos, de ahí que éstas reciban el nombre de medidas de tendencia central, en el caso de que su objetivo sea localizar un punto medio, y de posición o localización, en el caso de buscar una ubicación específica en una distribución de datos.

Las segundas se refieren al esparcimiento o grado de dispersión que tienen los datos respecto de una medida de posición o bien respecto de sus datos extremos. A estas medidas se les conoce como de dispersión o de variabilidad.

La medida de sesgo o asimetría tiene por objeto mostrar si la distribución de frecuencias de un conjunto de datos es simétrica o asimétrica respecto de una medida de tendencia central.

Y, finalmente, la medida de curtosis, que es aquella que nos permite cuantificar el tamaño del pico o afilamiento que presenta una distribución de frecuencias, y cuya interpretación es más simple realizarla de forma gráfica.

Todas y cada una de estas medidas serán analizadas en los siguientes puntos de este capítulo.

Medidas de tendencia central y de posición

Tienen por objeto encontrar el punto central, o bien, un punto específico en la distribución de un conjunto de datos. Estas medidas pueden clasificarse en:

- Media aritmética
- Media ponderada
- Media geométrica
- Mediana
- Moda
- Cuartiles, deciles y percentiles

Media aritmética

La media aritmética, media o promedio es, tal vez, la medida de posición más utilizada, y se define como la suma de los valores observados de una variable cuantitativa (discreta o continua), dividida por el número total de las observaciones. De una manera formal, decimos que, si $X_1, X_2, X_3, X_4 \ldots X_n$ son *n* observaciones numéricas del fenómeno que se está estudiando, entonces, la media aritmética o promedio de estas *n* observaciones se denota y define de la siguiente manera:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n} \qquad \text{I}$$

Esta media o promedio corresponde a lo que en estadística se llama media muestral (I), ya que su cálculo proviene precisamente de una muestra de la variable cuantitativa. Se simboliza con una X que tiene una raya en la parte superior, y se lee equis barra.

Si las observaciones del fenómeno corresponden a los datos de toda una población (N), entonces la media aritmética corresponde a lo que se denomina media poblacional (II), la cual constituye un parámetro que denotamos como:

$$\mu = \frac{X_1 + X_2 + X_3 + \ldots + X_N}{N} \qquad \text{II}$$

y cuyo símbolo es la letra griega mu: μ

Si utilizamos la notación de sumatoria $\sum$, las ecuaciones I y II quedarían especificadas como:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad \text{Ia}$$

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} \qquad \text{IIa}$$

Donde X_i son todas las observaciones de la variable desde la primera X_1 hasta la X_n o X_N, según corresponda.

Para mostrar un ejemplo del cálculo de este estadístico, se utilizará una muestra de 20 estaturas de estudiantes universitarios del grupo SB09-05/O de la UAM-Xochimilco.

Estaturas en metros y centímetros

1.60	1.71
1.59	1.70
1.65	1.68
1.64	1.66
1.55	1.70
1.54	1.69
1.58	1.68
1.57	1.67
1.70	1.57
1.69	1.56

Utilizando la ecuación I para el cálculo del estadístico tenemos:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_{20}}{20}$$

$$\overline{X} = \frac{1.60 + 1.59 + 1.65 + \ldots + 1.56}{20}$$

$$\overline{X} = 1.64$$

El estadístico nos indica que la estatura media del grupo de 20 estudiantes es de 1.64m, lo que no significa que todos los estudiantes de la muestra tengan esta estatura, sino que ésta representa a ese grupo de estudiantes.

Mediante el uso de la hoja electrónica de cálculo Excel, tenemos que la media aritmética de ese conjunto de datos se calcula con la función:

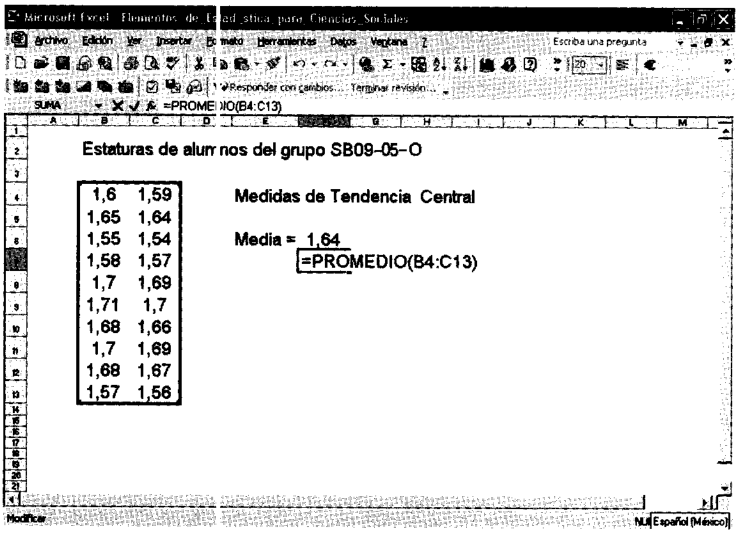
= PROMEDIO (conjunto de celdas que contienen los datos)

cuya versión en inglés es:

= AVERAGE (conjunto de celdas que contienen los datos)

Como se muestra en la Figura 9.

FIGURA 9



Observe en la Figura 9 que cada dato u observación es colocado en una celda, para posteriormente calcular la media muestral con la función ubicada en la celda F6 definida como: =PROMEDIO (B4:C13) o bien =AVERAGE (B4:C13) en la versión inglesa.

Esta función permite el cálculo de una media muestral y el de una media poblacional para datos que no han sido agrupados en una tabla de distribución de frecuencias. Una condición importante para aplicar estas funciones en Excel es que los datos deben estar ordenados en celdas contiguas, evitando tener celdas intermedias en blanco. Con base en esta consideración, le sugerimos al usuario de la hoja electrónica usar un solo renglón o columna si el número de datos a analizar es impar, y dos o más columnas si el número de datos es par.

En aquellos casos en que la variable en estudio se clasifique utilizando una tabla de distribución de frecuencias, las relaciones para el cálculo de la media aritmética son:

Media muestral o media poblacional para datos agrupados (Ib)

$$\overline{X}, \mu = \frac{\sum_{i=1}^{k} f_k \; X_k}{n} \qquad \text{Ib}$$

En esta ecuación para datos agrupados (Ib), los términos f_k y X_k representan:

f_k = son las frecuencias de cada clase o intervalo,
k = 1, 2, 3... n clase o intervalo

X_k = son las marcas de clase de cada intervalo,
k = 1, 2, 3... n marca de clase de cada intervalo

Para mostrar un ejemplo del cálculo de la media para datos agrupados usaremos la tabla de frecuencias del Cuadro 10 de este libro a través de la hoja electrónica Excel, la cual muestra la clasificación de las calificaciones de 25 estudiantes en un examen de estadística (Figura 10).

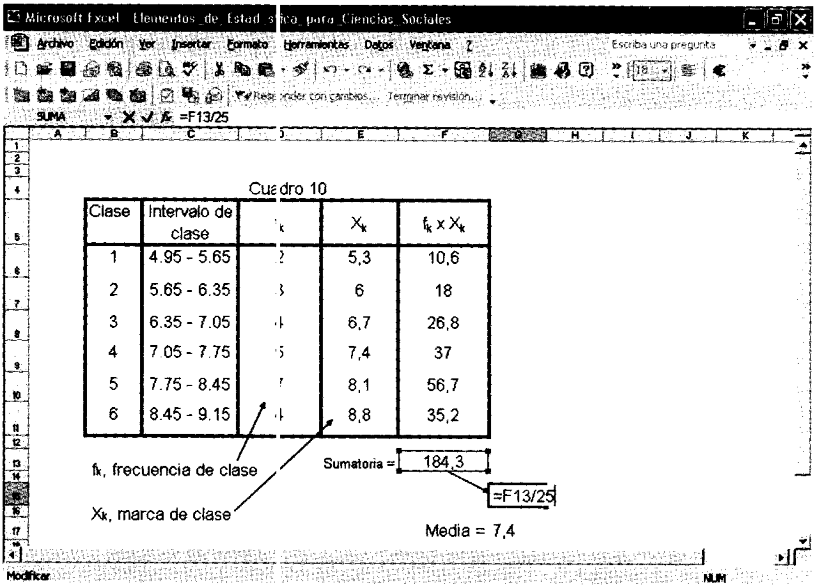
De la relación Ib y de la Figura 10 se desprende que las operaciones realizadas fueron:

- primero, el producto de cada marca de clase con su frecuencia respectiva
- segundo, la sumatoria de estos productos
- tercero, la división de la sumatoria entre el total de datos, es decir,

$$\text{Media} = [(2)(5.3) + (3)(6) + (4)(6.7) + (5)(7.4) + (7)(8.1) + (2)(8.8)] / 25$$

$$\text{Media} = 7.4$$

FIGURA 10



Clase	Intervalo de clase	f_k	X_k	$f_k \times X_k$
1	4.95 - 5.65	2	5,3	10,6
2	5.65 - 6.35	3	6	18
3	6.35 - 7.05	4	6,7	26,8
4	7.05 - 7.75	5	7,4	37
5	7.75 - 8.45	7	8,1	56,7
6	8.45 - 9.15	4	8,8	35,2

Si se calculara la media directamente a partir de los datos que dan origen a la Tabla 10, se encontraría que dicho valor es: 7.5. Esta pequeña diferencia no es rara, y es a causa de que los datos no se distribuyen simétricamente en cada intervalo. Por lo tanto, el cálculo de la media para datos no agrupados y para datos agrupados *diferirá* siempre en una pequeña cifra.

La media aritmética, media o promedio, es un número individual único que puede obtenerse de un conjunto de observaciones de una variable cuantitativa; sin embargo, este número, que trata de representar a ese conjunto de datos, puede proporcionar al analista información estadística falsa cuando en dichos datos encontramos los llamados valores extremos. Un dato extremo es aquel que sale del ámbito normal del conjunto de datos que se analiza. Por ejemplo, considere los datos sobre el gasto en transporte diario de un grupo de 6 estudiantes: \$3.50, \$3.20, \$3.80, \$2.90, \$3.00, \$50.00. El último dato en el conjunto anterior sale del ámbito del resto de los datos, por lo tanto se trata de un dato extremo.

Veamos cómo afectan los datos extremos en el cálculo de una media considerando los datos de gasto diario en transporte de los estudiantes.

Media que incluye el dato extremo:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_6}{6}$$

$$\overline{X} = \frac{3.5 + 3.2 + 3.8 + 2.9 + 3.0 + 50.0}{6}$$

$$\overline{X} = 11.07$$

Media sin incluir el dato extremo:

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_5}{5}$$

$$\overline{X} = \frac{3.5 + 3.2 + 3.8 + 2.9 + 3.0}{5}$$

$$\overline{X} = 3.28$$

En el primer caso, el valor extremo genera una media de gasto de \$11.07 que no es representativa del conjunto de datos, es decir, sesga el resultado hacia el dato extremo. Mientras que en el segundo caso, la media obtenida \$3.28 sí es representativa del conjunto (el dato extremo quedó excluido), es decir, los estudiantes gastan en transporte diariamente un promedio de \$3.28.

Como se puede observar, los valores extremos sesgan el valor de la media, esto es, la media es sensible a los datos extremos en un conjunto de observaciones. Por este motivo es importante que el analista que utiliza este estadístico esté consciente de ello y tome en consideración esta característica que presenta la media aritmética en estos casos.

Media ponderada

La media ponderada es un estadístico que permite obtener el promedio de un conjunto de datos, en el que se toma en cuenta la importancia que tiene cada dato dentro del cálculo global. Así, por ejemplo, al calcular una media aritmética, se asigna el mismo peso a cada dato en ella, como se puede observar en el siguiente ejemplo:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_5}{5}$$

$$\bar{X} = \frac{3.5 + 3.2 + 3.8 + 2.9 + 3.0}{5}$$

$$\bar{X} = 3.28$$

En esta media aritmética cada dato tiene un mismo peso en el cálculo global, es decir, 1/5 de 3.5, 1/5 de 3.2, así hasta 1/5 de 3.0. La media entonces se puede calcular así:

$$\text{media} = 1/5(3.5) + 1/5(3.2) + 1/5(3.8) + 1/5(2.9) + 1/5(3.0) = 3.28$$

y el ponderador de cada dato es por lo tanto igual a 1/5. La media ponderada, a diferencia de la media aritmética, permite asignar un ponderador o peso específico distinto a cada dato según su importancia en el cálculo global, así, esta medida se puede expresar matemáticamente como

$$\bar{X}w = \frac{\sum_{i=1}^{e} w_i X_i}{\sum_{i=1}^{e} w_i}$$

donde:

$\bar{X}w$ = es la media ponderada

w_i = el peso asignado a cada elemento "i" en la media ponderada $i = 1, 2, 3 \ldots e$

X_i = es cada uno de los elementos a ponderar $i = 1, 2, 3 \ldots e$

Un buen ejemplo de uso de una media ponderada es el cálculo de la calificación final que los profesores de cada módulo hacen trimestralmente en la UAM-Xochimilco.

Suponga que un estudiante de administración del quinto módulo obtuvo las siguientes calificaciones en el trimestre: 10 en investigación, 6 en matemáticas y 6 en el taller de contabilidad. Una media aritmética daría como resultado 7.3, equivalente a una S, ya que cada parte del módulo tendría un peso de 33.3%. Sin embargo, los módulos se ponderan con base en el siguiente criterio: 75% la investigación, 20% los contenidos de matemáticas y 5% el taller de contabilidad. Con esto último el alumno obtiene una calificación de 9 puntos, la cual equivale a MB.

ponderadores   elementos a ponderar

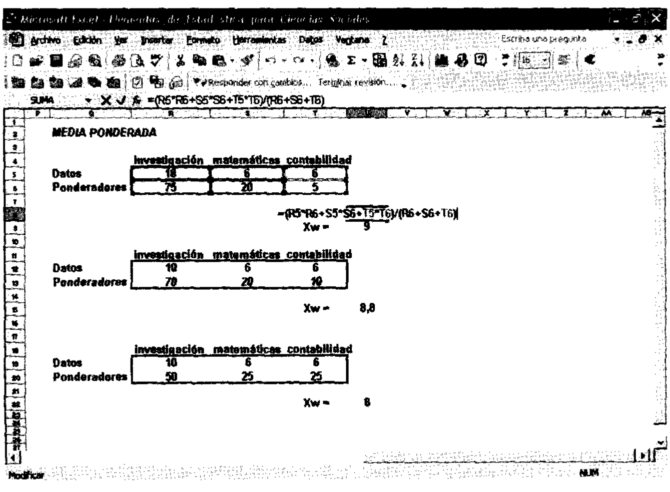
$$\overline{X}w = \frac{[75]\,[10] + [20]\,[6] + [5]\,[6]}{75 + 20 + 5} = 9$$

Como se puede observar, la media ponderada asigna un peso específico a cada parte que constituye un todo, a diferencia, como ya indicamos, de la media aritmética, la cual asigna el mismo peso a cada elemento del todo.

Tal vez la dificultad más grande que existe para aplicar este estadístico es determinar, precisamente, el valor que deben tener los ponderadores para cada elemento.

La hoja electrónica de cálculo facilita considerablemente el cálculo de medias ponderadas, ya que permite probar diversos ponderadores en un mismo problema, como se muestra en la Figura 3 (p. 51).

FIGURA 11



En la Figura 11 la ecuación de la media ponderada está ubicada en la celda U8, definida como: =(R5*R6+S5*S6+T5*T6)/(R6+S6+T6); mientras que las ecuaciones de las otras medias ponderadas se ubican en las celdas U15 y U22 respectivamente. La construcción de estas últimas ecuaciones se realiza en forma similar a la de la celda U8.

Observe, finalmente, que un cambio en los ponderadores genera medias ponderadas distintas.

Media geométrica

Este estadístico se emplea principalmente cuando una variable presenta un comportamiento distinto en el tiempo, es decir, se pretende evaluar un promedio de comportamiento para diferentes periodos, conjugandose para su cálculo dos factores, el valor de la variable y el tiempo. En estos casos, una media normal o una media ponderada no pueden medir con precisión dicho comportamiento.

La media geométrica se define como la raíz *n* del producto de los valores que toma la variable en el tiempo, donde *n* es el número de factores empleados en el producto. Matemáticamente, la media geométrica queda indicada como:

$$MG = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \ldots X_n}$$

n = número de factores

$X_1, X_2 \ldots Xn$ = son los *n* datos en el tiempo usados en el cálculo de la media geométrica.

Un ejemplo típico de uso de una media geométrica en el área de la economía o de la administración es el cálculo de una tasa promedio de interés; por ejemplo, suponga que un analista desea saber cuál fue el interés promedio generado con los certificados de la tesorería (Cetes) en 1993. Con base en los datos de los indicadores económicos proporcionados por el Banco de México para 1993, las medias aritmética y geométrica serían:

Datos enero-diciembre, 1993
(tasa de rendimiento neto)

16.72	17.73	17.47	16.17	15.04	15.5	13.85	13.68	13.71	13.13	14.38	11.78

Media

$$\bar{X} = \frac{16.72 + 17.73 + 17.47 + \ldots + 11.78}{12} = 14.93$$

Media geométrica

$$MG = \sqrt[12]{[16.72]\,[17.73]\,[17.47]\ldots\,[11.78]}$$

$$MG = 14.83$$

Como se puede observar, la tasa de rendimiento medio en 1993, para los Cetes, fue de 14.83%, distinta a la tasa errónea que se obtiene con una simple media aritmética.

En Excel, el cálculo de este tipo de estadístico se puede realizar usando la función: =MEDIA.GEOM ($dato_1$, $dato_2$, $dato_3$ ... $dato_n$), y en la versión en inglés, =GEOMEAN ($dato_1$, $dato_2$, $dato_3$ ... $dato_n$). Para el ejemplo, la función estaría definida en una celda como:

=MEDIA.GEOM (16.72, 17.73, 17.47... 11.78)

Mediana

La mediana es un estadístico cuyo valor proporciona el elemento central de un conjunto de datos ordenados respecto de la magnitud de los valores, ya sea en forma ascendente o descendente. Este elemento central divide al conjunto de datos en dos partes iguales, 50% de los datos se encuentran por debajo de este valor y el otro 50% por arriba de él.

Si el conjunto de datos ordenados contiene un número impar de elementos, entonces el de la mitad será la mediana. La relación que nos permite ubicar el dato en el conjunto es $U_{me} = (n+1)/2$, donde n es el número de datos en el conjunto.

Por ejemplo, considere el siguiente conjunto de datos:

3, 11, 4, 4, 8, 5, 6, 8, 9

Se ordena, primeramente, al conjunto en forma ascendente (o descendente):

3, 4, 4, 5, 6, 8, 8, 9, 11

Se calcula la ubicación de la mediana:

$$U_{me}=(9_{\text{datos}}+1)/2 = 5\text{to. dato}$$

Se ubica el quinto dato (5to.) en el conjunto ordenado contando en forma ascendente (o descendente). Para el ejemplo, es el dato cuyo valor es 6, por lo tanto la mediana (Me) será:

$$\text{Me} = 6$$

Este valor 6 divide al conjunto exactamente a la mitad, es decir, 50% de los datos tienen valores menores a 6 y el otro 50% mayores a éste.

En aquellos casos en que el conjunto de datos contenga un número par de elementos, entonces la relación a utilizar para calcular la mediana en el conjunto es: Me= ([Dato(n/2)] + [Dato(n/2)+1])/2, donde *n* es el número de datos en el conjunto.

Por ejemplo, considere el conjunto de datos siguiente:

5, 15, 18, 5, 9, 7, 12, 11

Se ordena el conjunto en forma ascendente:

5, 5, 7, 9, 11, 12, 15, 18

Se determinan los valores de los datos n/2 y del dato (n/2)+1

[Dato (n/2)]= [Dato(8/2)]= [Dato(4)] = 9

[Dato(n/2)+1]= [Dato(8/2)+1]= [Dato(4)+1]= [Dato(5)] = 11

Finalmente, se calcula la mediana:

Me= ([Dato(n/2)] + [Dato(n/2)+1])/2

Me= (9 + 11)/2 = 20/2 = 10

Me= 10

Este valor 10 divide al conjunto exactamente a la mitad, lo que indica que 50% de los datos tienen un valor menor de 10 y otro 50% un valor mayor a éste.

Como se puede observar, el estadístico encuentra el punto central del conjunto de datos sin considerar los valores que éstos tengan. Esto puede ser contradictorio al comparar su comportamiento frente a la media aritmética. Si tomamos al conjunto de datos 1, 2, 3, 6, 8 y calculamos la media y la mediana, los resultados obtenidos serían media=4 y mediana=3. Si cambiamos el quinto dato del conjunto de 8 a 88, entonces la mediana seguiría siendo igual a 3, pero la media sería 20. Con esto concluimos que la mediana es un estadístico no sensible a los datos numéricos extremos, ya que no los toma en cuenta en su cálculo, mientras que la media es muy sensible a ellos, como ya se había indicado anteriormente.

La mediana, por otro lado, es un estadístico que tiene la ventaja de poder analizar datos no cuantitativos a diferencia de la media aritmética y encontrar de ellos su punto central. Por ejemplo, considere las opiniones vertidas sobre un producto por cinco individuos: malo, regular, malo, bueno, bueno.

Ordenando las opiniones en forma ascendente tenemos:

malo, malo, regular, bueno, bueno

La mediana será por lo tanto: regular, lo que indica que 50% de los individuos clasifican al producto por debajo de esta categoría y el otro 50% por arriba de ella.

Mediana para datos agrupados

Cuando el conjunto de datos está agrupado en una tabla de distribución de frecuencias, el procedimiento de cálculo se realiza entonces por medio de una interpolación utilizando los pasos siguientes:

1. Se ubica el intervalo de clase que contiene a la mediana mediante el cálculo del cociente del número de datos en el cuadro (n) entre dos; es decir: n/2.
2. Una vez ubicado el intervalo que contiene la mediana, se procede a determinar las variables siguientes:

Lim = límite real inferior de la clase que contiene la mediana.
n = número total de datos en la tabla de frecuencias.
Fac = frecuencia acumulada hasta la clase que antecede a la que contiene la mediana.
f = frecuencia absoluta de la clase que contiene a la mediana.
T = tamaño del intervalo de la clase que contiene la mediana.

3. Se calcula la mediana utilizando la relación de interpolación siguiente:

$$Me = Lim + \left(\frac{\frac{n}{2} - Fac}{f} \right) T$$

Como un ejemplo del cálculo de este estadístico, utilicemos los datos de las estaturas (en pulgadas) de los trabajadores de una empresa en Vancouver, Canadá.

CUADRO 14

Estatura de trabajadores en Vancouver, Canadá (pulgadas)

CLASE	INTERVALO	FREC. ABSOLUTA (f_K)	FREC. ACUMULADA (F_K)
1	53 - 55	2	2
2	56 - 58	5	7
3	59 - 61	8	15
4	62 - 64	16	31
5	65 - 67	12	43
6	68 - 70	5	48
7	71 - 73	2	50

FUENTE: datos hipotéticos.

Siguiendo los pasos establecidos anteriormente obtenemos:

1. Se ubica el intervalo de clase que contiene a la mediana:

$$n/2 = 50_{datos} / 2 = 25$$

la mediana se ubica aproximadamente en el dato 25, el cual se localiza en la clase 4 intervalo (62-64), ya que al acumular los datos, el tercer intervalo sólo contiene hasta el dato 15, y el cuarto comprende hasta el dato 31.

2. Se determinan las variables de la fórmula.

Puesto que la cuarta clase debe contener a la mediana, las variables de la fórmula toman los valores siguientes:

Lim = 61.5
n = 50
Fac = 15
f = 16
T = 3

3. Calculamos la mediana con la fórmula establecida.

$$Me = 61.5 + \left(\frac{\frac{50}{2} - 15}{16} \right) 3$$

$$Me = 63.37$$

El valor encontrado, nos indica que 50% de los trabajadores tiene una estatura menor a 63.37 pulgadas y otro 50% mayor a ésta.

En Excel existe una función estadística que permite calcular la mediana de datos numéricos no agrupados. La función para la versión en español e inglés se define como:

= MEDIANA (rango del conjunto de datos) —— español

= MEDIAN (rango del conjunto de datos) —— inglés

Por ejemplo, si el conjunto de datos es: 1, 3, 5, 4, 7, 6, 8. Si éstos se ubican por columna en las celdas de la A1 a la A7, la mediana podría calcularse en la celda B1 como:

= MEDIANA (A1:A7), observándose un resultado de 5

Moda

Es una medida de tendencia central que difiere de la media, aunque se parece un poco a ella. Es un estadístico que no se calcula por medio de los procesos ordinarios de la aritmética. La moda (Mo) se define como el valor que más se repite dentro de un conjunto de datos.

Es un estadístico muy útil cuando se desea cuantificar las características de variables cualitativas. Por ejemplo, los empleados de una compañía se pueden clasificar por género, estado civil, ocupación, etcétera. En este caso no tiene sentido la media o la mediana del género de los empleados, la media o la mediana de su estado civil, etcétera, por lo que se determina la característica predominante en el conjunto (la que más se repite), es decir, la moda.

La moda, a diferencia de los estadísticos anteriores, puede no ser única y puede, además, no existir en un conjunto de datos. Para mostrar esto último observe los ejemplos siguientes.

Dado el conjunto: 1, 2, 3, 3, 4, 5, 6
Mo = 3
datos unimodales

Dado el conjunto: 1, 1, 2, 3, 4, 4, 5, 6
Mo = 1 y Mo = 4
datos bimodales

Dado el conjunto: 1, 2, 3, 4, 5, 6
datos sin moda

Moda para datos agrupados

Para determinar la moda de un conjunto de datos agrupados en una tabla de distribución de frecuencias, debemos realizar el siguiente procedimiento:

1. Localice la clase modal, es decir, aquella que presenta mayor frecuencia absoluta. Pueden existir dos o más clases modales o bien no existir clase modal en el caso de que todas las clases tengan la misma frecuencia.
2. Una vez ubicada la clase modal, calculamos la moda por interpolación mediante el uso de la relación:

$$Mo = Lim + \left(\frac{d_1}{d_1 + d_2} \right) T$$

donde:

Lim = límite real inferior de la clase modal (la clase de mayor frecuencia absoluta)
d_1 = diferencia entre la frecuencia de la clase modal y la frecuencia de la clase que la antecede
d_2 = diferencia entre la frecuencia de la clase modal y la frecuencia de la clase que le sigue
T = tamaño del intervalo de la clase modal

Para ejemplificar el uso de este procedimiento consideremos el Cuadro 14 de las estaturas de trabajadores en Vancouver, Canadá, propuesto en el tema anterior.

1. La cuarta clase es la clase modal ya que ésta presenta la mayor frecuencia absoluta (f_4=16).
2. Se calculan las variables de la ecuación de interpolación y la moda.

$$Lim = 61.5,\ d_1 = 8,\ d_2 = 4 \text{ y } T = 3$$

$$Mo = 61.5 + \left(\frac{8}{8+4} \right) 3$$

$$Mo = 63.49$$

El dato con mayor frecuencia en la tabla es 63.49, es decir, la moda, lo que indica que de los trabajadores canadienses, la estatura que más se repite en ellos es 63.49 pulgadas.

En Excel existe, al igual que en el caso de la media y la mediana, una función que permite calcular la moda para datos no agrupados. Esta función se define como:

= MODA (rango usado por el conjunto de datos) —— español

= MODE (rango usado por el conjunto de datos) —— inglés

Por ejemplo, si el conjunto de datos siguiente: 1, 1, 3, 2, 5, 4, 6, 7, 1, 8 está ubicado en una columna, en las celdas que van de B1 a B10, entonces su moda puede calcularse en C3 como: = MODA (B1:B10), cuyo resultado es 1.

Cuando el conjunto tiene más de una moda (bimodal), Excel sólo encuentra una, la primera que esté en orden ascendente. Y en aquellos casos donde el conjunto no tenga moda se observará el mensaje: #N/A, en la celda de cálculo.

MEDIDAS DE POSICIÓN

Cuartiles, deciles y percentiles

Además de las medidas de tendencia central antes estudiadas (media, mediana, moda, media geométrica y media ponderada), existen otras que pueden ser más prácticas en ciertas situaciones donde el investigador no busca una ubicación central, sino una posición específica en la distribución de sus datos. Estas medidas de posición son los cuartiles, deciles y percentiles; éstos, en cierta forma, son una extensión de la mediana.

Cuartiles

Los cuartiles de una sucesión de datos ordenados son aquellos números que dividen la sucesión en cuatro partes porcentualmente iguales. Hay tres cuartiles, denotados como Q1, Q2 y Q3. El segundo cuartil (Q2) es, precisamente, la mediana. El primer cuartil es el valor por debajo del cual queda un cuarto (25%) de los valores de la sucesión ordenada, mientras que para el tercer cuartil es 75% de los datos.

Puesto que estas medidas adquieren mayor importancia cuando los datos están agrupados en una tabla de distribución de frecuencias, nos limitaremos entonces a mostrar su fórmula de cálculo para estos casos.

Los cuartiles se calculan con la fórmula:

$$Qk = Lk + \left(\frac{k\,[n/4] - Fk}{fk} \right) T$$

donde:

k = 1, 2, 3 el número de cuartil a calcular
Lk = límite real inferior de la clase donde se ubica el cuartil k. La clase del cuartil k se determina de manera similar que en el caso de la mediana [k (n/4)].
n = número de datos
Fk = Frecuencia acumulada de la clase que antecede a la clase del cuartil k
fk = frecuencia absoluta de la clase del cuartil k
T = tamaño del intervalo de la clase del cuartil k

Por ejemplo, utilizando el Cuadro 14, el de la información de estaturas de trabajadores canadienses, el cálculo del primer cuartil sería:

1. Se ubica la clase que contiene al cuartil 1.

$$k\,(n/4) = 1\,(50/4) = \text{dato } 12.5$$

Se busca el dato 12.5 con ayuda de la frecuencia acumulada.

2. El dato buscado, para este ejemplo, se ubica en la clase 3.
3. Se determinan las variables de la fórmula de cálculo del cuartil.

$$L_1 = 58.5,\ n = 50,\ F_1 = 7,\ f_1 = 8 \text{ y } T = 3$$

4. Se sustituyen los valores en la fórmula y se obtiene el valor del cuartil 1.

$$Q_1 = 58.5 + \left(\frac{1\,[50/4] - 7}{8} \right) 3$$

$$Q_1 = 60.56$$

Esto último indica que 25% de los trabajadores en esta muestra miden menos de 60.56 pulgadas.

En el caso de que los datos no estén agrupados podemos usar para el cálculo de un cuartil la función que proporciona la hoja electrónica Excel. Esta función está definida como:

= CUARTIL (rango de datos, número de cuartil deseado) —— español

= QUARTILE (rango de datos, número de cuartil deseado) —— inglés

Por ejemplo, si tenemos el siguiente conjunto de datos: 1, 3, 4, 2, 3, 5, 5, 6, 6, 7, capturados por columna en la celdas de C2 a C11, entonces para calcular el cuartil 1 la función usada en la celda D2 sería: = CUARTIL (C2:C11, 1) dando como resultado 3, es decir, $Q_1 = 3$, lo que significa que 25% de los datos son menores a 3.

Deciles

Los deciles son números que dividen a una sucesión ordenada de datos en diez partes porcentualmente iguales. Los deciles se calculan del decil 1 (D_1) al decil 9 (D_9). Su fórmula de cálculo para datos agrupados en una tabla de distribución de frecuencias es:

$$Dk = Lk + \left(\frac{k\,[n/10] - Fk}{fk} \right) T$$

donde:

k = 1, 2, 3, 4, 5, 6, 7, 8, 9 el número de decil a calcular
Lk = límite real inferior de la clase que contiene al decil k. La clase del decil k se determina de manera similar que en el caso de la mediana (k (n/10)).
n = número de datos
Fk = Frecuencia acumulada de la clase que antecede a la clase del decil k
fk = frecuencia absoluta de la clase del decil k
T = tamaño del intervalo de la clase del decil k

Por ejemplo, del Cuadro 14, el de estaturas de trabajadores canadienses, determine el decil 7.

1. Se ubica la clase que contiene al decil 7.

$$k\,(n/10) = 7\,(50/10) = \text{dato } 35$$

Se busca el dato 35 con ayuda de la frecuencia acumulada.

2. El dato buscado, para este ejemplo, se ubica en la clase 5.
3. Se determinan las variables de la fórmula de cálculo del decil.

$$L_7 = 64.5,\ n = 50,\ F_7 = 31,\ f_7 = 12 \text{ y } T = 3$$

4. Se sustituyen los valores de las variables en la fórmula y se obtiene el valor del decil 7.

$$D_7 = 64.5 + \left(\frac{7\,[50/10] - 31}{12}\right)3$$

$$D_7 = 65.5$$

Este estadístico nos indica que 70% de los trabajadores de la muestra tienen una estatura inferior a las 65.5 pulgadas.

Percentiles

Son una medida estadística muy utilizada cuando se desea clasificar o ubicar características en las personas (peso, estatura, etcétera). Estas medidas son números que dividen una sucesión de datos ordenados en cien partes porcentualmente iguales. Cuando los datos se agrupan en una tabla de frecuencias, los percentiles se calculan mediante la fórmula:

$$Pk = Lk + \left(\frac{k\,[n/100] - Fk}{fk} \right) T$$

donde:

k = 1, 2, 3, 4, 5, 6, 7... 99, el número de percentil a calcular
Lk = límite real inferior de la clase que contiene al percentil k. La clase del percentil k se determina de manera similar que en el caso de la mediana [k (n/100)]
n = número de datos
Fk = Frecuencia acumulada de la clase que antecede a la clase del percentil k
fk = frecuencia absoluta de la clase del percentil k
T = tamaño del intervalo de la clase del percentil k

Utilizando el Cuadro 14, de estaturas de trabajadores canadienses, determine el percentil 80.

1. Se ubica la clase que contiene al percentil 80.

$$k\,(n/100) = 80\,(50/100) = \text{dato } 40$$

Se busca el dato 40 con ayuda de la frecuencia acumulada

2. El dato buscado, para este ejemplo, se ubica en la clase 5.
3. Se determinan las variables de la fórmula de cálculo del percentil.

$$L_{80} = 64.5,\ n = 50,\ F_{80} = 31,\ f_{80} = 12 \text{ y } T = 3$$

4. Se sustituyen los valores en la fórmula y se obtiene el valor del percentil 80.

$$P_{80} = 64.5 + \frac{80\,[50/100] - 31}{12}\,3$$

$$P_{80} = 66.75$$

El percentil calculado indica que 80% de los trabajadores de la muestra tienen una estatura inferior a las 66.75 pulgadas.

En el caso de que los datos no estén agrupados podemos usar para el cálculo de un percentil la función que proporciona la hoja electrónica Excel. Esta función está definida como:

= PERCENTIL (rango de datos, número de percentil deseado en decimales)

= PERCENTILE (rango de datos, número de percentil deseado en decimales)

Por ejemplo, si tenemos el siguiente conjunto de datos: 1, 3, 4, 2, 3, 5, 5, 6, 6, 7, capturados por columna en la celdas de C2 a C11, entonces para calcular el percentil 20 la función usada en la celda D2 sería: = PERCENTIL (C2:C11, 0.20), dando como resultado 2.8, es decir, P_{20}=2.8, lo que significa que 20% de los datos son menores a 2.8.

Medidas de dispersión o de variabilidad

En el punto anterior de este capítulo indicamos que las medidas de tendencia central estudiadas sólo nos permiten ubicar una tendencia hacia un centro en un conjunto de datos pero no permiten conocer si éstos se encuentran alrededor de estas medidas o bien están muy alejados de ellas; tampoco nos indican si el conjunto tiene datos extremos que sesguen los valores de los estadísticos. Por eso, todo análisis estadístico se inicia con el cálculo de las medidas de tendencia central, y se complementa con un análisis de la variabilidad o dispersión de los datos.

Una medida de variabilidad, por lo tanto, será un número que indique el grado de dispersión (esparcimiento) en un conjunto de datos con respecto a un estadístico de tendencia central (generalmente la media aritmética). Si este valor es pequeño (respecto de la unidad de medida) entonces hay una gran uniformidad de los datos; por el contrario, un gran valor indica poca uniformidad, y finalmente un valor cero nos indica que todos los datos son iguales.

La variabilidad, al igual que el centro de la distribución de un conjunto de datos, se puede estudiar o determinar de distintas formas según el fin perseguido por el investigador o usuario de la estadística, sin embargo podemos indicar que la medidas de variabilidad más comunes son:

- El rango (amplitud o recorrido)
- La desviación absoluta promedio
- La varianza
- La desviación estándar
- La dispersión relativa: el coeficiente de variación

Rango, amplitud o recorrido

Es la más elemental de las medidas de variabilidad y se le clasifica como una medida de distancia. El rango o recorrido es la diferencia entre el valor máximo y el valor mínimo de un conjunto de datos; o bien, en una distribución de frecuencias, el límite real superior de la última clase menos el límite real inferior de la primera clase, cuando la clasificación se ha hecho en forma ascendente en valor.

Matemáticamente, el rango se define como:

$$R = Dmy - Dme$$

donde:

Dmy = dato de mayor valor en el conjunto de datos.

Dme = dato de menor valor en el conjunto de datos.

Es una medida de variabilidad muy útil cuando sólo se desea conocer la diferencia entre los datos extremos de un conjunto, sin embargo presenta dos grandes problemas. El primero es que, al analizar únicamente dos datos, el mayor y el menor, no informa del comportamiento del resto de los datos que forman el conjunto. El segundo es que, al no tomar en cuenta ninguna medida de tendencia central, no informa nada acerca de cómo se comportan los datos respecto del centro.

Finalmente, el rango o recorrido es un estadístico muy sensible a los valores extremos en una sucesión de datos, por lo que su utilidad puede ser escasa en muchos casos prácticos.

Considere los siguientes datos de peso (kg) de siete adultos: 55, 45, 57, 46, 48, 50, 70. La amplitud en este problema es 70-45 = 25 kg. Esta medida nos indica que entre el adulto de mayor peso y el de menor peso hay 25 kg, pero no nos indica cuál es el comportamiento de variabilidad del resto de los datos, ni cuál es su comportamiento respecto de un centro.

En Excel este estadístico de dispersión puede ser calculado restando a la función que encuentra el valor máximo [MAX()], la función que encuentra el valor mínimo [MIN()] en ese mismo conjunto. Por ejemplo, si los datos de peso de los siete adultos del ejemplo anterior se ubican en las celdas de A1 a A7 en una hoja electrónica, la amplitud se puede calcular en la celda C2 con la fórmula = MAX (A1:A7) - MIN (A1:A7), obteniéndose como resultado 25.

Desviación absoluta promedio

Este estadístico es la media aritmética de los valores absolutos de las desviaciones respecto de la media aritmética o de la mediana. Las desviaciones se definen como la diferencia entre el estadístico de tendencia central usado (media aritmética o mediana) y cada uno de los datos en el conjunto de estudio. De esta forma la desviación absoluta promedio, cuando el estadístico de posición es la media aritmética, se define matemáticamente como:

$$DAP = \frac{\sum_{i=1}^{n} |X_i - \overline{X}|}{n} = \frac{\sum_{i=1}^{n} |d_i|}{n}$$

donde:

$|d_i| = |X_i - \overline{X}|$ son los valores absolutos de las desviaciones de cada dato Xi respecto de la media aritmética

n n, es el número de datos en el conjunto

Para el cálculo de este estadístico, la mediana es preferida en algunas ocasiones sobre la media, ya que se puede demostrar que la suma de los valores absolutos de las desviaciones respecto de ésta, es menor que la suma de las desviaciones respecto de cualquier otro valor. En la práctica las desviaciones se toman respecto de la media aritmética. Si la distribución del conjunto de datos que se analiza es simétrica, entonces la media es igual a la mediana y de esta manera se obtiene la misma desviación absoluta promedio.

Para mostrar el procedimiento de cálculo de este estadístico usaremos como ejemplo el conjunto de datos siguiente: 3, 3, 5, 5, 5, 7, 7, 8, 8, 9.

1. Determine la media del conjunto de datos.

$$\bar{X} = 6$$

2. Calcule el valor absoluto de las desviaciones.

$$|3 - 6| = 3$$
$$|3 - 6| = 3$$
$$|5 - 6| = 1$$
$$\ldots$$
$$|8 - 6| = 2$$
$$|9 - 6| = 3$$

3. Calcule la media de las desviaciones cuyo resultado es la desviación absoluta promedio.

$$DAP = \frac{|3 - 6| + |3 - 6| + |5 - 6| + \ldots + |9 - 6|}{10} = \frac{18}{10} = 1.8$$

Si la distribución de estos datos es normal y simétrica, entonces 68% de las observaciones quedan comprendidas entre la [$\bar{X}$ - DAP] y la [$\bar{X}$ + DAP], es decir, entre 4.2 y 7.8. Prácticamente, en este conjunto, 7 de 10 datos (70%) quedan comprendidos en este intervalo, por lo que no existe una dispersión muy grande de los datos alrededor de la media aritmética.

Observe finalmente que el cálculo de este estadístico involucra a todos y cada uno de los datos del conjunto de estudio ponderando por igual a cada elemento e indicando a que distancia de la media (o mediana) se halla en promedio cada observación.

En los casos donde los datos están dados en una tabla de distribución de frecuencias, la desviación absoluta promedio se calcula mediante la relación siguiente:

$$DAP = \frac{\sum_{i=1}^{k} |X_i - \overline{X}| f_i}{n} = \frac{\sum_{i=1}^{k} |d_i| f_i}{n}$$

donde:

X_i = marca de clase del intervalo i
$\overline{X}$ = media
f_i = frecuencia absoluta del intervalo
n = número total de datos en la tabla
k = número total de clases o intervalos

En Excel el cálculo de la desviación absoluta promedio para un conjunto de datos no agrupados se puede obtener a partir del uso de la función:

= DESVPROM (rango de datos) —— español

= AVEDEV (rango de datos) —— inglés

Si el conjunto de datos del ejemplo anterior se captura en una hoja electrónica en las celdas A1 a A10, la desviación absoluta promedio puede calcularse en la celda C2 como = DESVPROM (A1:A10), cuyo resultado es: 1.8.

Varianza

Es un estadístico que se puede definir como la media aritmética de las desviaciones respecto de la media elevada al cuadrado. En esencia es similar a la desviación absoluta promedio, salvo que en este caso eliminamos el uso del valor absoluto cambiándolo por la elevación al cuadrado de cada una de las desviaciones. Este procedimiento provoca, por un lado, que todas las desviaciones sean positivas, lo que evita el uso del valor absoluto y, por otro, que el promedio obtenido de las desviaciones elevadas al cuadrado resulte siempre en unidades cuadradas; así, si el conjunto de datos está medido en kilogramos, la varianza de esos datos se medirá en kilogramos al cuadrado (kg^2).

La varianza se simboliza matemáticamente con: S^2 para una muestra. En este texto usaremos σ^2 (sigma al cuadrado) para la población.

Matemáticamente, la varianza se calcula –para una muestra– mediante la siguiente relación:

$$S^2 = \frac{\sum_{i=1}^{n} [X_i - \overline{X}]^2}{n-1} = \frac{\sum_{i=1}^{n} [d_i]^2}{n-1} \qquad 2.I$$

donde:

$[d_i]^2 = [X_i - \overline{X}]^2$ son las desviaciones al cuadrado de cada dato X_i respecto de la media aritmética

n es el número total de datos en la muestra

Si la varianza que se desea calcular considera a todos los datos de una población, entonces la relación a utilizar es:

$$\sigma^2 = \frac{\sum_{i=1}^{N} [X_i - \mu]^2}{N} = \frac{\sum_{i=1}^{N} [d_i]^2}{N} \qquad 2.II$$

donde:

$[d_i]^2 = [X_i - \mu]^2$ son las desviaciones al cuadrado de cada dato X_i respecto de la media de la población (μ)

N es el número total de datos en la población

La ecuación 2.I permite calcular la varianza de un conjunto de datos que forman parte de una muestra, mientras que la ecuación 2.II calcula la varianza de los datos de toda una población.

Por otro lado, si la muestra o la población que se analiza está agrupada en una tabla de distribución de frecuencias, entonces la varianza muestral se calcula con la ecuación 2.III, y la varianza poblacional con la ecuación 2.IV.

$$S^2 = \frac{\sum_{i=1}^{k} [X_i - \overline{X}]^2 f_i}{n-1} = \frac{\sum_{i=1}^{k} [d_i]^2 f_i}{n-1} \qquad 2.III$$

$$\sigma^2 = \frac{\sum_{i=1}^{k} [X_i - \mu]^2 f_i}{N} = \frac{\sum_{i=1}^{k} [d_i]^2 f_i}{N} \qquad 2.IV$$

donde:

Xi = marca de clase del intervalo i
fi = frecuencia absoluta del intervalo i
$\overline{X}$ = media aritmética de la muestra
μ = media aritmética de la población
n = número total de datos en la muestra
N = número total de datos en la población
k = número total de clases o intervalos

En la hoja electrónica Excel, el cálculo de una varianza muestral o poblacional, para datos no agrupados, se obtiene mediante las funciones siguientes:

VARIANZA MUESTRAL

= VAR (rango de datos de la muestra)

VARIANZA POBLACIONAL

= VARP (rango de datos de la población)

Para mostrar el procedimiento de cálculo de este estadístico usaremos como ejemplo el conjunto de datos definidos en el problema de la desviación absoluta promedio: 3, 3, 5, 5, 5, 7, 7, 8, 8, 9.

Puesto que se trata de una muestra usaremos la ecuación 2.I:

1. Calculamos la media de los datos.

$$\overline{X} = 6$$

2. Calculamos las desviaciones elevadas al cuadrado.

$$|3-6|^2 = 9$$
$$|3-6|^2 = 9$$
$$|5-6|^2 = 1$$
$$\dots$$
$$|8-6|^2 = 4$$
$$|9-6|^2 = 9$$

3. Calculamos la varianza (la media de las desviaciones elevadas al cuadrado).

$$S^2 = \frac{9+9+1+1+1+1+1+4+4+9}{10-1}$$

$$S^2 = 4.44 \text{ unidades}^2$$

Esta medida nos indica que la dispersión de los datos con respecto a la media muestral son cuatro punto cuarenta y cuatro unidades cuadradas. Físicamente, el estadístico no nos dice mucho acerca de la variabilidad de los datos pues mientras la media aritmética se mide en unidades, la varianza se mide en unidades cuadradas.

Si la muestra de los datos del ejemplo anterior se captura en una hoja electrónica en las celdas A1 a A10, la varianza muestral puede calcularse en la celda C2 como: = VAR (A1:A10) cuyo resultado es 4.444 (varianza corregida o cuasivarianza).

Desviación estándar

Dada la dificultad de poder medir con la varianza el grado de dispersión en un conjunto de datos, se decide entonces crear un nuevo estadístico a partir de ésta mediante el cálculo de su raíz cuadrada; es decir, la desviación estándar; por lo tanto, las unidades en las que se mide este nuevo estadístico de dispersión serán las mismas que tienen las observaciones y su media aritmética.

Más aún, se dice que la desviación estándar es la raíz cuadrada del promedio de los cuadrados de las desviaciones que existen de las observaciones respecto de la media aritmética. En matemáticas, el cálculo de la desviación estándar [S o σ (sigma)] para una muestra y una población se expresan en las ecuaciones 2.V y 2.VI, respectivamente:

$$S = \sqrt{\frac{\sum_{i=1}^{n} [X_i - \bar{X}]^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} [d_i]^2}{n-1}} \qquad 2.V$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} [X_i - \mu]^2}{N}} = \sqrt{\frac{\sum_{i=1}^{N} [d_i]^2}{N}} \qquad 2.VI$$

donde:

Xi = observación i en la muestra o en la población
$\bar{X}$ = media aritmética de la muestra
μ = media aritmética de la población
n = número total de datos en la muestra
N = número total de datos en la población
d_i = desviaciones de cada observación con respecto a la media muestral o con respecto a la media poblacional, respectivamente.

En aquellos casos en que los datos se agrupan en una tabla de distribución de frecuencias, el cálculo de la desviación estándar para una muestra o para una población se efectúan con las relaciones 2.VII y 2.VIII, según sea el caso.

$$S = \sqrt{\frac{\sum_{i=1}^{k} [X_i - \bar{X}]^2 f_i}{n-1}} = \sqrt{\frac{\sum_{i=1}^{k} [d_i]^2 f_i}{n-1}} \qquad 2.VII$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k} [X_i - \mu]^2 f_i}{N}} = \sqrt{\frac{\sum_{i=1}^{k} [d_i]^2 f_i}{N}} \qquad 2.VIII$$

donde:

Xi = marca de clase del intervalo i
fi = frecuencia absoluta del intervalo i
$\overline{X}$ = media aritmética de la muestra
μ = media aritmética de la población
n = número total de datos en la muestra
N = número total de datos en la población
k = número total de clases o intervalos

Para mostrar el procedimiento de cálculo de la desviación estándar, continuaremos utilizando el mismo conjunto de datos de los ejemplos de los dos estadísticos anteriores: 3, 3, 5, 5, 5, 7, 7, 8, 8, 9.

Puesto que se trata de una muestra de datos no agrupados se utiliza la ecuación 2.V:

1. Calculamos la media de los datos.

$$\overline{X} = 6$$

2. Calculamos las desviaciones elevadas al cuadrado.

$$(3 - 6)^2 = 9$$

$$(3 - 6)^2 = 9$$

$$(5 - 6)^2 = 1$$

$$\dots$$

$$(8 - 6)^2 = 4$$

$$(9 - 6)^2 = 9$$

3. Calculamos la varianza (la media de las desviaciones elevadas al cuadrado).

$$S^2 = \frac{9 + 9 + 1 + 1 + 1 + 1 + 1 + 4 + 4 + 9}{10 - 1}$$

$$S^2 = 4.44 \text{ unidades}^2$$

4. Calculamos la desviación estándar, es decir, la raíz cuadrada de la varianza.

$$\sqrt{S^2} = \sqrt{4.44 \text{ unidades}^2}$$

$$S = 2.11 \text{ unidades}$$

Con este valor se pueden hacer dos observaciones. La primera es que las unidades en que está dada la desviación estándar son las mismas que la de los datos y la media aritmética. La segunda, si la distribución de estos datos es normal y simétrica, entonces 68% de las observaciones deben quedar comprendidas entre la $[\overline{X}\text{-}S]$ y la $[\overline{X}+S]$, es decir, entre 3.89 y 8.11. Para nuestro ejemplo, 70% de los datos quedan comprendidos en este intervalo, lo que parece indicar que no existe una dispersión muy grande de los datos alrededor de esta media. Desafortunadamente, la desviación estándar es una medida de dispersión absoluta, es decir, no nos indica qué tan grande o qué tan pequeña es la dispersión de los datos respecto de la media aritmética, por lo que el analista deberá recurrir a una medida de dispersión relativa que le permita llegar a una conclusión. Una de estas medidas de dispersión relativa es el coeficiente de variación, del cual hablaremos más adelante.

En la hoja electrónica Excel, el cálculo de una desviación estándar muestral o poblacional se obtiene mediante las funciones:

DESVIACIÓN ESTÁNDAR MUESTRAL

= DESVEST (rango de datos de la muestra) —— español

= STDEV (rango de datos de la muestra) —— inglés

DESVIACIÓN ESTÁNDAR POBLACIONAL

= DESVESTP (rango de datos de la población) —— español

= STDEVP (rango de datos de la población) —— inglés

La fórmula para calcular la desviación estándar muestral también se modifica en el denominador de *n* a *n-1*, es decir, esta desviación se calcula con la varianza corregida o cuasivarianza.

Para la muestra de datos del ejemplo anterior, el cálculo de la desviación estándar de la muestra en una hoja electrónica que tiene los datos capturados en las celdas A1 a A10 es: = DESVEST (A1:A10) cuyo resultado es 2.11.

Dispersión relativa: el coeficiente de variación

Como ya indicamos, la desviación estándar es una medida de variación absoluta que no permite concluir qué tan grande o pequeña es la dispersión de los datos, sin embargo, combinada con la media aritmética da origen a una medida de dispersión relativa llamada coeficiente de variación.

El coeficiente de variación (CV) es la medida relativa que nos da una idea general de la magnitud de la desviación estándar en relación con la magnitud de la media aritmética. Esta relación expresa la desviación estándar como porcentaje de la media aritmética, de ahí que sus unidades se midan en "por ciento". Matemáticamente puede expresarse como:

$$CV = \frac{S}{\overline{X}} 100\%$$

El coeficiente de variación es una medida de variabilidad que se emplea fundamentalmente para:

1. Comparar la variabilidad entre dos grupos de datos, ya sea que tengan la misma unidad de medida o distinta. Por ejemplo, un conjunto medido en kilogramos y otro en metros.
2. Comparar el comportamiento de dos grupos de datos obtenidos por dos o más personas distintas.
3. Comparar dos grupos de datos que tienen distinta media aritmética.
4. Determinar si cierta media aritmética es consistente con cierta varianza.

Como ejemplo ilustrativo consideremos los resultados obtenidos en módulo por dos grupos distintos en el tronco divisional del área de ciencias sociales y humanidades en la UAM-Xochimilco.

Estadístico	Grupo SB05	Grupo SB13
Media (promedio)	8.2	9.5
Desviación estándar	0.5	0.5

Si sólo analizamos la desviación estándar de los dos grupos debemos aceptar que la variabilidad de éstos, en cuanto a las calificaciones obtenidas, es la misma.

Sin embargo, las medias son diferentes, lo que permite distinguir a cada grupo, es decir, nuestro primer análisis debe ser revisado. Una forma de hacerlo es calculando el coeficiente de variación en ambos grupos.

GRUPO SB05

CV = (0.5/8 2) 100 = 6.1%

GRUPO SB13

CV = (0.5/9 5) 100 = 5.2%

Como sucede en cualquier medida de variabilidad: a mayor valor, más variabilidad. En este caso, el Grupo SB05 presenta una variabilidad relativa mayor que la del grupo SB13, es decir, este último fue más homogéneo en cuanto a rendimiento que el grupo SB05.

En Excel, este coeficiente se puede calcular combinando la función de la media con la desviación estándar. Para el caso de una muestra el coeficiente se calcularía de la manera siguiente:

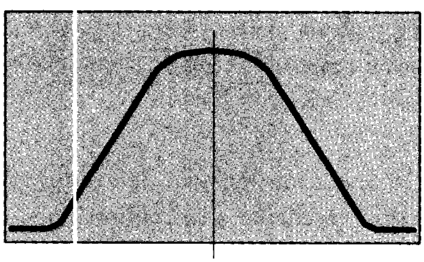
= (DESVEST (rango de datos)/PROMEDIO (rango de datos))*100

El rango de datos para ambas funciones debe ser el mismo.

Medida de sesgo (asimetría)

Las curvas que se pueden utilizar para representar las observaciones de un conjunto de datos pueden ser simétricas o asimétricas (sesgadas). Las curvas simétricas, como la que se muestra en la Figura 12, son aquellas que, si trazamos una línea vertical desde el punto más alto de ella (la cima) al eje horizontal, el área total de esta curva será dividida en dos partes iguales. Cada parte es el espejo de la otra.

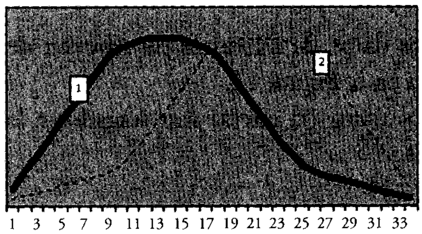
FIGURA 12



Por otro lado, las curvas en la Figura 13 son sesgadas. Ello por que los valores de sus distribuciones de frecuencias se concentran en el extremo inferior o en el extremo superior de la escala de medición situada sobre el eje horizontal. Los valores no tienen una distribución igual. La curva 1 está sesgada a la derecha (tiene asimetría positiva), pues disminuye gradualmente hacia el extremo superior de la escala. Lo contrario sucede con la curva 2: está sesgada a la izquierda (tiene asimetría negativa) puesto que disminuye gradualmente hacia el extremo inferior de la escala.

La medida estadística que cuantifica el sesgo de un conjunto de datos se llama coeficiente de sesgo (CS).

FIGURA 13



El CS se define y denota para datos no agrupados mediante la relación siguiente:

$$CS = \frac{n}{[n - 1]\,[n - 2]} \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}}{S} \right)^3$$

donde:

Xi = observación (dato) i
n = número total de datos en la muestra
$\overline{X}$ = media aritmética de la muestra
S = desviación estándar de la muestra

Y en el caso de que los datos se agrupen en una tabla de distribución de frecuencias la relación de cálculo de la asimetría será

$$CS = \frac{\sum_{i=1}^{k} f_i\,[X_i - \overline{X}]^3/n}{S^3}$$

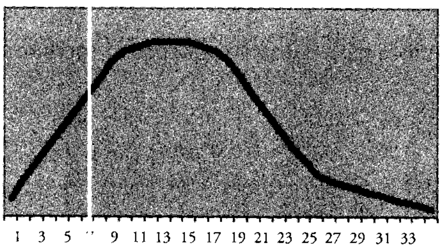
donde:

Xi = marca de clase del intervalo i
fi = frecuencia absoluta del intervalo i
$\overline{X}$ = media aritmética de la muestra
S = desviación estándar de la muestra
n = número total de datos en la muestra
k = número total de clases o intervalos

Para ambas relaciones de cálculo del coeficiente de sesgo, su interpretación deberá realizarse del modo siguiente:

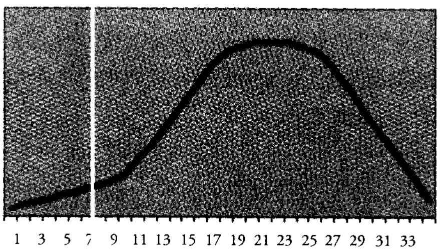
- Si CS = 0, entonces los datos (la curva) se distribuyen de manera simétrica, como la que se muestra en la Figura 12.
- Si CS > 0, entonces los datos (la curva) son sesgados a la derecha, como se muestra en la Figura 14.

FIGURA 14



- Si CS < 0, entonces los datos (la curva) son sesgados a la izquierda, como se muestra en la Figura 15.

FIGURA 15

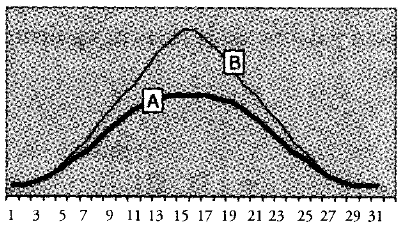


En Excel, el coeficiente de sesgo se calcula para datos no agrupados con la función: = COEFICIENTE.ASIMETRIA (rango de datos). Por ejemplo, si el conjunto de datos es: 3, 4, 5, 2, 3, 4, 5, 6, 4, 7 su coeficiente de sesgo se calcula con la función = COEFICIENTE.ASIMETRIA (3, 4, 5, 2, 3, 4, 5, 6, 4, 7) o bien, si los datos ocupan las celdas A1 a A10, = COEFICIENTE.ASIMETRIA (A1:A10). El resultado en ambos casos es: CS = 0.359, mostrando con ello que los datos tienen un pequeño sesgo a la derecha.

Medida de curtosis (afilamiento)

Es la medida estadística de un conjunto de observaciones que permite determinar su grado de pico en la curva de distribución, es decir, las curvas de distribuciones de frecuencias que se construyen difieren, en muchos casos, tan sólo por el hecho de que una tiene un pico mayor que otra como se puede observar en la Figura 16. En esta figura las dos curvas (A y B) tienen la misma localización y dispersión, y ambas son simétricas, sin embargo presentan diferente grado de pico: tienen diferentes grados de curtosis.

FIGURA 16



Existen diferentes grados de curtosis, pero en estadística normalmente se utilizan tres clases generales. Una curva que se distribuye simétricamente de forma normal, como la de la Figura 12, recibe el nombre de curva mesocúrtica. Una curva que tenga más pico, como la mostrada en la Figura 17, recibe el nombre de curva leptocúrtica, y una curva que tenga menos pico, como la mostrada en la Figura 18, recibe el nombre de curva platocúrtica.

FIGURA 17

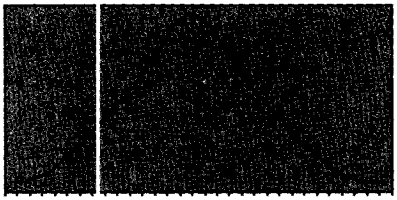
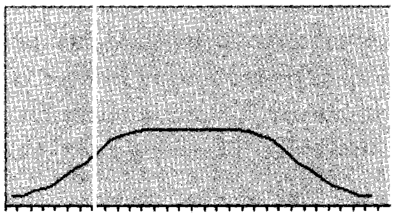


FIGURA 18



La curtosis de una distribución se determina a partir del llamado coeficiente de curtosis (CC), el cual se denota y define de la manera siguiente:

Para datos no agrupados

$$CC = \left(\frac{n\,[n+1]}{[n-1]\,[n-2]\,[n-3]} \sum_{i=1}^{n} \left(\frac{[X_i - \bar{X}]}{S} \right)^4 \right) - \frac{3\,[n-1]^2}{[n-2]\,[n-3]}$$

donde:

Xi = observación (dato) i
n = número total de datos en la muestra
$\bar{X}$ = es la media aritmética de la muestra
S = desviación estándar de la muestra

Para datos agrupados en una tabla de distribución de frecuencias

$$CC = \frac{\sum_{i=1}^{k} f_i \, [X_i - \overline{X}]^4 / n}{S^4}$$

donde:

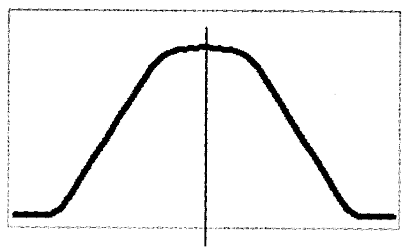
Xi = marca de clase del intervalo i
fi = frecuencia absoluta del intervalo i
$\overline{X}$ = media aritmética de la muestra
S = desviación estándar de la muestra
n = número total de datos en la muestra
k = número total de clases o intervalos

Para ambas relaciones este coeficiente se interpreta de la manera siguiente:

• Si CC = 3, entonces los datos (la curva) se distribuyen de manera simétrica en forma de una normal estandarizada, como la que se muestra en la Figura 19.

El número 3 se establece de manera teórica en estudios de estadística matemática.

FIGURA 19



• Si CC > 3, entonces los datos (la curva) presentan un pico mayor a los de la curva normal estandarizada, como se muestra en la Figura 17 (curva leptocúrtica).
• Si CC < 3, entonces los datos (la curva) se presentan más aplanados que los de la curva normal, como se muestra en la Figura 18 (curva platocúrtica).

En Excel el coeficiente de curtosis se calcula para datos no agrupados mediante la función: = CURTOSIS (rango de datos).

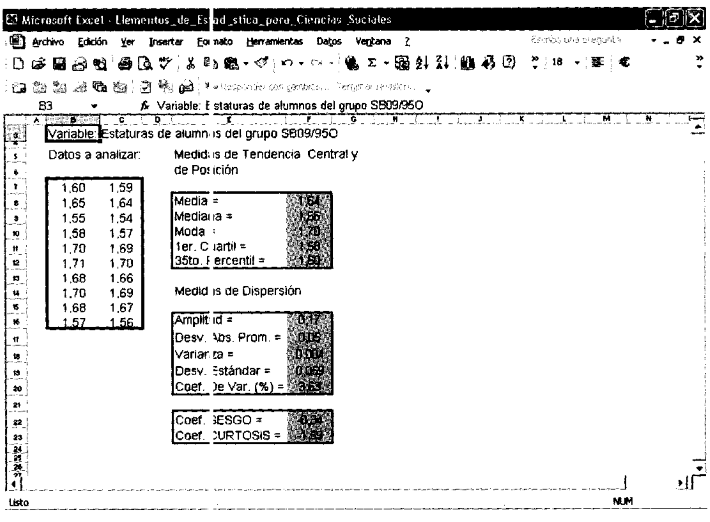
Por ejemplo, si el conjunto de datos es: 3, 4, 5, 2, 3, 4, 5, 6, 4, 7, su coeficiente de curtosis se calcula con la función: = CURTOSIS (3, 4, 5, 2, 3, 4, 5, 6, 4, 7) o bien, si los datos ocupan las celdas A1 a A10, = CURTOSIS (A1:A10). El resultado en ambos casos es CC = -0.1518, mostrando con ello que los datos forman una curva platocúrtica.

Un ejemplo del uso de las funciones estadísticas en Excel

El cálculo de medidas de estadística descriptiva con las hojas electrónicas es muy simple, particularmente con la hoja electrónica Excel, ya que basta con capturar los datos de la variable (o variables) a analizar, y utilizar las funciones estadísticas adecuadas que permitan calcular tanto medidas de tendencia central y de posición (media, mediana, moda, cuartil, percentil) como medidas de dispersión (amplitud, desviación absoluta promedio, varianza, desviación estándar y coeficiente de variación); el coeficiente de sesgo y el coeficiente de curtosis.

En la Figura 20 se muestran los datos de estatura de 20 alumnos del grupo SB09 del trimestre 95/O. Estos datos son capturados en el área o rango de celdas indicadas; para el ejemplo, el rango está formado de la celda B7 a la C16. Puesto que las funciones estadísticas han sido ya definidas en la hoja a partir de la celda F8, Excel calcula inmediatamente los estadísticos solicitados para dichos datos.

FIGURA 20



La tarea final para el estudiante será, entonces, interpretar las medidas calculadas, realizando de esta forma un análisis descriptivo de la variable en estudio.

En el cuadro siguiente se resume el contenido por celda del modelo de la Figura 20 y las funciones asociada para la hoja electrónica de Excel.

CELDA	FUNCIÓN
F8	= PROMEDIO (B7:C16)
F9	= MEDIANA (B7:C16)
F10	= MODA (B7:C16)
F11	= CUARTIL (B7:C16, 1)
F12	= PERCENTIL (B7:C16, 0.35)
F16	= MAX (B7:C16)-MIN (B7:C16)
F17	= DESVPROM (B7:C16)
F18	= VAR (B7:C16)
F19	=DESVEST (B7:C16)
F20	=(F19/F8)*100
F22	= COEFICIENTE.ASIMETRIA (B7:C16)
F23	= CURTOSIS (B7:C16)

El subprograma "Estadística Descriptiva" de Excel

En Excel el subprograma que permite calcular estadísticos descriptivos sin necesidad de utilizar las funciones estadísticas recibe el nombre de "Estadística Descriptiva". Con él, es posible calcular los estadísticos siguientes:

ESTADÍSTICO	DESCRIPCIÓN
Media aritmética	La suma de datos dividida entre el número de datos que forman el conjunto (promedio).
Error estándar	El error estándar de la media del conjunto de datos.
Mediana	El dato de en medio cuando el conjunto de datos está ordenado en forma ascendente (o descendente).
Moda	El dato que se repite más en el conjunto (dato de mayor frecuencia).

continúa

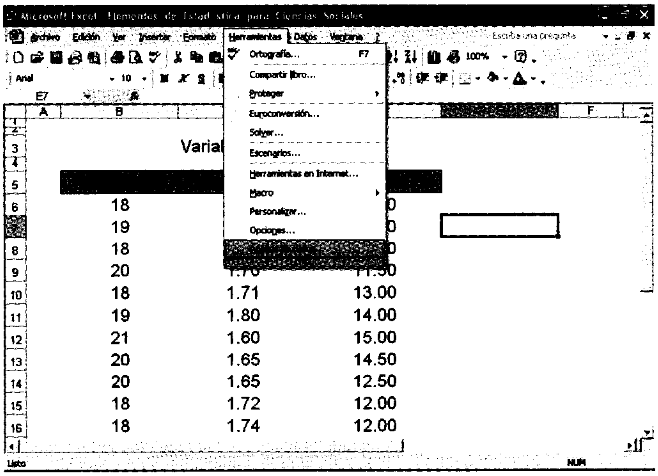
ESTADÍSTICO	DESCRIPCIÓN
Desviación estándar	La raíz cuadrada positiva de la varianza.
Varianza	La suma de las desviaciones (cada dato menos la media aritmética) elevadas al cuadrado, divididas entre el número de datos menos uno.
Curtosis	El nivel de pico o afilamiento de la distribución de los datos comparado con la distribución normal.
Skewness	El grado de asimetría de la distribución de los datos alrededor de su media aritmética.
Rango o amplitud	El dato más grande en el conjunto de datos, menos el dato más pequeño en el mismo conjunto.
Mínimo	El dato de menor valor en el conjunto de datos.
Máximo	El dato de mayor valor en el conjunto de datos.
Suma	La suma de los valores de los datos del conjunto.
Cuenta	El número de datos que forma el conjunto.

Para calcular los estadísticos descriptivos de un conjunto de datos con este subprograma es necesario realizar el procedimiento que a continuación se describe:

• Capture en una hoja electrónica en fila o columna, todos los datos a ser analizados mediante estadística descriptiva. Un ejemplo del uso de este subprograma utiliza las variables de edad, estatura y gasto diario ($) de 20 alumnos de la universidad como se muestra en la Figura 21.

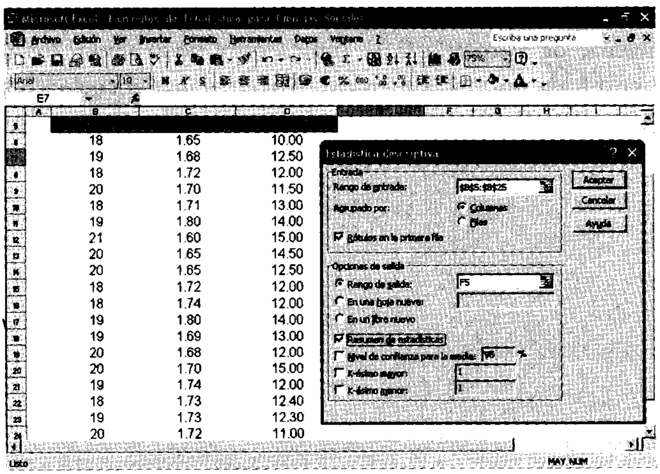
• Seleccione del menú de HERRAMIENTAS la opción ANÁLISIS DE DATOS como se muestra en la Figura 21.

FIGURA 21



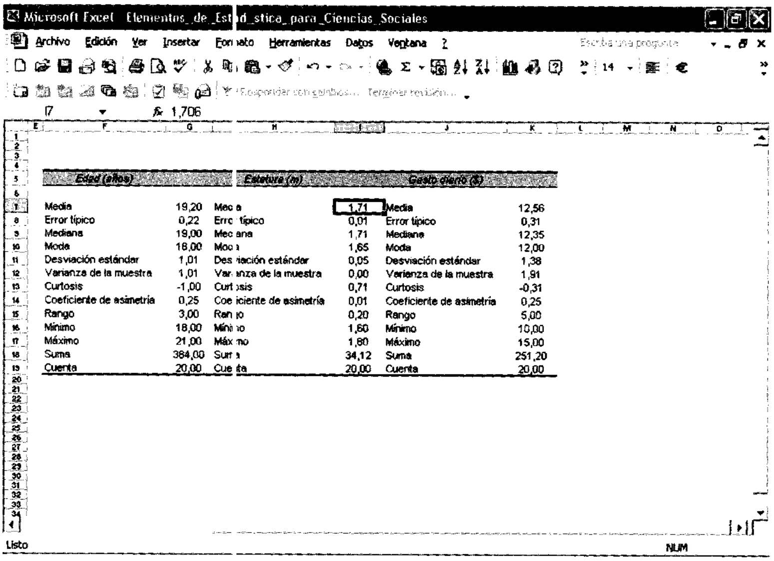
- De las opciones de *Análisis de datos* se debe seleccionar: "Estadística Descriptiva".
- Con el cursor ubicado en *Rango de entrada*, seleccionar el conjunto de datos a analizar. En el ejemplo el rango comprende de la celda B5 a D25, como se muestra en la Figura 14. Observe que en el primer renglón se incluyen las etiquetas de las variables, por lo que, deberá seleccionarse la opción: "Rótulos en la primera fila".
- En la opción *Rango de salida*, se debe ubicar el cursor en la celda a partir de la cual deseamos obtener los resultados. Para nuestro ejemplo, esta celda es la F5 (Figura 14).
- Seleccionar la casilla denominada *Resumen de estadísticas* (Figura 14).

FIGURA 22



- Oprimir *Aceptar* con lo que obtenemos los cálculos que se muestran en la Figura 23.

FIGURA 23



Edad (años)		Estatura (m)		Gasto diario ($)	
Media	19,20	Mec a	1,71	Media	12,56
Error típico	0,22	Errc típico	0,01	Error típico	0,31
Mediana	19,00	Mec ana	1,71	Mediana	12,35
Moda	18,00	Moo ı	1,65	Moda	12,00
Desviación estándar	1,01	Des iación estándar	0,05	Desviación estándar	1,38
Varianza de la muestra	1,01	Var. ınza de la muestra	0,00	Varianza de la muestra	1,91
Curtosis	-1,00	Curt ɔsis	0,71	Curtosis	-0,31
Coeficiente de asimetría	0,25	Coe iciente de asimetría	0,01	Coeficiente de asimetría	0,25
Rango	3,00	Ran ɪo	0,20	Rango	5,00
Mínimo	18,00	Míni ıo	1,60	Mínimo	10,00
Máximo	21,00	Máx mo	1,80	Máximo	15,00
Suma	384,00	Sur ı	34,12	Suma	251,20
Cuenta	20,00	Cue ıta	20,00	Cuenta	20,00

En la Figura 23 se muestran los estadísticos calculados para nuestro ejemplo. En ella podemos observar que la estatura promedio (media aritmética) de los estudiantes del ejemplo es de 1.71m, la desviación estándar es de 0.05m, es decir, 99% de las estaturas se encuentran ent:e: 1.61 y 1.81m. También podemos observar que el estudiante de mayor estatura en el grupo medía 1.80m y el de menor estatura 1.60m, lo que establece una amplituc o rango entre estos datos de 0.20m, es decir, la diferencia de estatura entre el alumno ınás alto y el más bajo es de 20cm.

El coeficiente de asimetría 0.01 indica muy poco sesgo, los datos casi se distribuyen en forma normal (simétrica). Mientras que el coeficiente de curtósis 0.71, indica que la curva de distribución de estos datos está más aplanada que una curva normal, es decir, es una curva platocúrtica.

Ejercicios

1. El gasto diario ($) de una muestra de 20 alumnos de la universidad se indica a continuación:

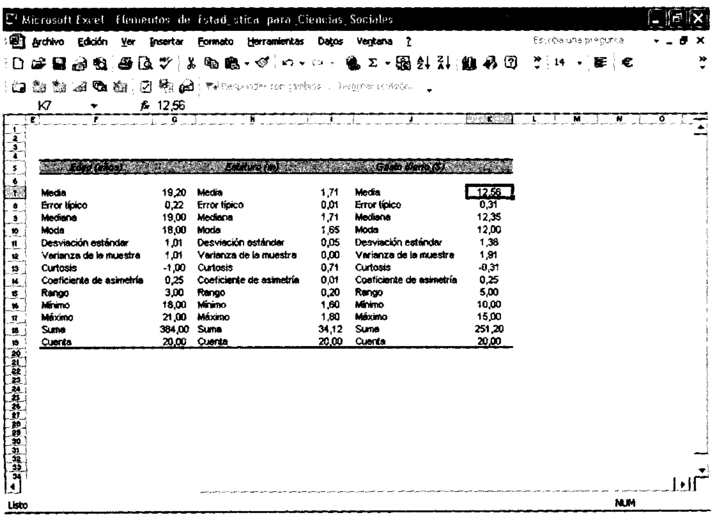
10.00	13.00
14.00	12.30
12.00	11.50
12.00	12.50
12.50	12.00
15.00	11.00
14.00	13.00
12.40	12.00
12.00	15.00
14.50	10.50

a) Calcule el tercer cuartil, el 35 percentil y el decil 3.
b) Calcule la media muestral y la desviación estándar muestral.

Solución:

La solución a este problema se muestra en la hoja electrónica siguiente, en la cual se han utilizado funciones para obtener los cálculos solicitados (Figura 24).

FIGURA 24



Media	19,20	Media	1,71	Media	12,56
Error típico	0,22	Error típico	0,01	Error típico	0,31
Mediana	19,00	Mediana	1,71	Mediana	12,35
Moda	18,00	Moda	1,65	Moda	12,00
Desviación estándar	1,01	Desviación estándar	0,05	Desviación estándar	1,38
Varianza de la muestra	1,01	Varianza de la muestra	0,00	Varianza de la muestra	1,91
Curtosis	-1,00	Curtosis	0,71	Curtosis	-0,31
Coeficiente de asimetría	0,25	Coeficiente de asimetría	0,01	Coeficiente de asimetría	0,25
Rango	3,00	Rango	0,20	Rango	5,00
Mínimo	18,00	Mínimo	1,60	Mínimo	10,00
Máximo	21,00	Máximo	1,80	Máximo	15,00
Suma	384,00	Suma	34,12	Suma	251,20
Cuenta	20,00	Cuenta	20,00	Cuenta	20,00

2. En una prueba de habilidad aplicada a obreros en la ciudad de Toluca resultaron las puntuaciones siguientes:

34 28 29 22 33 30 31 32 30 24 22 24 29 24 28 24

34 28 29 30 32 35 33 22 19 24 18 23 20 20 22 21

a) Calcule el segundo cuartil, el 45 percentil y el decil 6.
b) Calcule la media muestral y la desviación estándar muestral.
c) Concluya sobre la habilidad de estos obreros.

3. Los ingresos anuales en una comunidad del estado de Oaxaca se muestran en el siguiente cuadro. Determine al ingreso promedio, su dispersión mediante la desviación estándar y la dispersión relativa, para finalmente emitir una conclusión.

CLASE	INGRESO ANUAL (pesos)	JEFES DE FAMILIA
1	4,950 - 5,650	7
2	5,650 - 6,350	9
3	6,350 - 7,050	10
4	7,050 - 7,750	7
5	7,750 - 8,450	3

4. Determine la media, desviación estándar y dispersión relativa de un grupo de 12 alumnos en su examen de módulo y concluya en cuanto a su nivel de aprovechamiento.

7.5, 8, 7, 8.5, 8, 7.5, 8, 7, 7, 9, 8, 8.5

5. El peso de sacos de exportación de un producto industrial se indica en la siguiente muestra de 30 semanas. Se desea saber si la máquina llenadora de estos sacos está operando adecuadamente. El llenado adecuado no debe exceder una dispersión de 1.5%. Analice estadísticamente la muestra mediante las medidas de posición y dispersión convenientes y construya un histograma y polígono de frecuencias. Concluya e indique si la máquina de llenado está operando convenientemente.

20.2	20.3	20.2	21.0	19.9	19.8	20.5	21.0	20.1	20.4
20.3	20.0	20.9	20.4	19.7	19.8	20.9	20.1	20.2	20.1
20.2	20.2	20.1	20.9	19.8	19.7	20.4	20.0	20.1	20.0

Sugerencia: calcule media y desviación estándar.

6. El número de horas y el horario con mayor número de telespectadores son dos factores que influyen en la publicidad televisiva. Una muestra de 50 familias con el número de horas que ven televisión, produjo los datos siguientes:

   3.0, 6.0, 7.5, 15.0, 12.0, 6.5, 8.0, 4.0, 5.5, 6.0, 5.0, 12.0, 1.0, 3.5, 3.0, 7.5, 5.0, 10.0, 8.0, 3.5, 9.0, 2.0, 6.5, 1.0, 5.0, 4.5, 1.0, 6.0, 1.5, 8.5, 3.0, 7.5, 9.5, 4.5, 7.0, 3.0, 2.5, 3.0, 1.0, 11.5, 4.5, 5.5, 5.0, 3.5, 7.5, 6.0, 11.5, 14.5, 7.0, 5.5

a) Construya una tabla de distribución de frecuencias.
b) Elabore los histogramas y polígonos conocidos.
c) Calcule el segundo cuartil, el 25 percentil y el decil 7.
d) Calcule la media muestral y la desviación estándar muestral.
e) Concluya sobre el número de horas que ven la televisión las familias de la muestra.

7. Los ingreso mensuales que reciben 15 ejecutivos medios en la ciudad de Guadalajara se muestran a continuación:

10 000	10 500	9 900	11 000	11 500
10 500	10 300	11 000	10 500	11 400
11 500	10 000	11 000	10 300	11 100

Determine:

- El ingreso promedio mensual de estos ejecutivos.
- La desviación estándar de sus ingresos.
  La dispersión relativa.
- Concluya con base en la información estadística anterior.

Cálculo de la media aritmética y la desviación estándar por el método abreviado

EL MÉTODO ABREVIADO se emplea para el cálculo de la media aritmética y la desviación estándar porque tiene la ventaja de ahorrar tiempo y trabajo cuando se manejan datos numéricos relativamente pequeños, cuyos cálculos no son tan laboriosos como sucede con el método tradicional revisado en el capítulo anterior.

Cálculo de la media aritmética

Dado un conjunto de datos (una muestra), este método se inicia proponiendo o estimando una media aritmética, a esta estimación se le aplica posteriormente una corrección para encontrar el valor real o verdadero de la media aritmética.

Para explicar los pasos requeridos en el procedimiento de cálculo de este estadístico nos auxiliaremos con el cuadro estadístico 15.

CUADRO 15

Cálculo de la media aritmética y desviación estándar (método abreviado)

1	2	3	4	5	6	7
No.	Intervalo	X_i	f_i	x'	f_i x'	f_i $(x')^2$
1	239.5 - 244.5	242	1	-6	-6	36
2	244.5 - 249.5	247	3	-5	-15	75
3	249.5 - 254.5	252	2	-4	-8	32
4	254.5 - 259.5	257	4	-3	-12	36
5	259.5 - 264.5	262	4	-2	-8	16
6	264.5 - 269.5	267	6	-1	-6	6
7	269.5 - 274.5	272	10	0	0	0
8	274.5 - 279.5	277	8	1	8	8
9	279.5 - 284.5	282	5	2	10	20
10	284.5 - 289.5	287	4	3	12	36
11	289.5 - 294.5	292	2	4	8	32
12	294.5 - 299.5	297	1	5	5	25
Σ =			50		-12	322

Fuente: datos hipotéticos.

X_i = marca de clase del intervalo i
x' = desviación de un puntaje X_i de la media estimada
fi = frecuencia absoluta del intervalo i

PROCEDIMIENTO:

1. Encontrar la media estimada ($\overline{X}_E$) tomando el valor de la marca de clase del intervalo más cercano al centro de la distribución o la del intervalo que tiene la mayor frecuencia. En el Cuadro 15 el intervalo 269.5 - 274.5 es el que tiene la mayor frecuencia y es casi el centro de la distribución, por lo que consideramos el valor de la marca de clase de este intervalo como la media estimada ($\overline{X}_E$ = 272).
2. Llenamos la columna (5) x', anotando la desviación que existe entre cada clase respecto de la clase donde está ubicada la media estimada, en unidades de intervalos de clase. La media estimada con valor de 272 está ubicada en el intervalo (7), por lo que no existe una desviación entre la marca de clase de este intervalo y la media estimada, entonces anotamos el valor de cero. Todas las clases menores a la clase (7) con media estimada en 272, tomarán valores negativos en x' y las superiores a 272 serán valores positivos. En el intervalo (6) la desviación es de -1, porque está desviado un intervalo de clase; el intervalo (5) está desviado dos intervalos de clase, es decir, -2 unidades en la columna x', y así sucesivamente hasta el intervalo (1), el cual está desviado seis intervalos de clase, colocando el -6 en la columna x'. Los valores de marca de clase superiores a 272 se encuentran a partir del intervalo (8), colocando en la columna x' el valor de 1, y así sucesivamente encontramos las desviaciones 2, 3, 4 y 5 de intervalos de clase desde la media estimada.
3. Calculamos la columna (6) f_ix', multiplicando cada x' de la columna (5) por la frecuencia correspondiente de la columna (4). Para el renglón (1) la desviación es de -5 por su frecuencia 1, se obtiene como resultado -5 el cual como ya se indicó se coloca en la columna (6). En la misma forma calculamos estas multiplicaciones para cada uno de los renglones subsecuentes.
4. El siguiente paso consiste en encontrar la corrección en unidades de intervalo de clase (c), realizando la suma algebraica de las f_ix' positivas y negativas, y dividiendo esta suma por n, como se indica en la siguiente expresión:

$$c = \frac{\sum_{i=1}^{k} (fx'^{-} + fx'^{+})}{n}$$

De la columna (6) se suman los valores positivos (43), y los valores negativos (-55). En la columna (4) se encuentra en el renglón de $\sum$ el total de datos (n) que forman la muestra (50).

5. Se obtiene la corrección de puntaje (ct_i), multiplicando el tamaño del intervalo de clase (T) por la corrección en unidades de intervalo de clase (c),

$$ct_i = (5)(-0.24) = -1.2$$

6. Finalmente encontramos el valor de la media aritmética ($\overline{X}$) sumando algebraicamente la media estimada ($\overline{X}_E$) y la corrección de puntaje (ct_i).

$$\overline{X} = \overline{X}_E + ct_i$$

$$\overline{X} = 272 - 1.2 = 270.8$$

La media aritmética para este ejemplo por el método abreviado es de 270.8, valor similar al que podría obtenerse si se calcula utilizando el procedimiento establecido en los capítulos anteriores.

Cálculo de la desviación estándar

Para explicar el proceso de cálculo de la desviación estándar también usaremos el cuadro estadístico 15.

En este caso el método abreviado utiliza la fórmula:

$$s = T\sqrt{\frac{\sum_{i=1}^{k} f_i x'^2}{n-1} - C^2}$$

T = tamaño del intervalo
c^2 = corrección en unidades de intervalo de clase elevada al cuadrado
n = número de datos

PROCEDIMIENTO:

1. Se multiplican los valores de la columna (5) x' por los de la columna (6) f_i x', obteniendo como resultado la columna (7) f_i $(x')^2$. En el intervalo (1) el valor de x' es de -6 y el de f_i x' es –6, entonces el valor en f_i $(x')^2$ es de 36. Este proceso se realiza para los demás intervalos del cuadro 15, posteriormente se suman todos los resultados de la columna f_i $(x')^2$ obteniéndose 322.
2. La corrección en unidades de intervalo de clase (c) elevada al cuadrado:

$$c = \frac{\sum_{i=1}^{k} (fx'^{-} + fx'^{+})}{n}$$

$$c^2 = (-0.24)^2 = 0.0576$$

3. Sustituyendo los valores calculados en la fórmula de la desviación estándar por el método abreviado se obtiene:

$$s = (5)\sqrt{\frac{322}{50 - 1} - 0.0576} = (5)\sqrt{6.57143 - 0.0576} = 12.76$$

4. Si se desea conocer el valor de la varianza (s^2), sólo se requiere elevar al cuadrado el valor de la desviación estándar calculado en el paso anterior.

$$s^2 = (12.76)^2 = 162.846$$

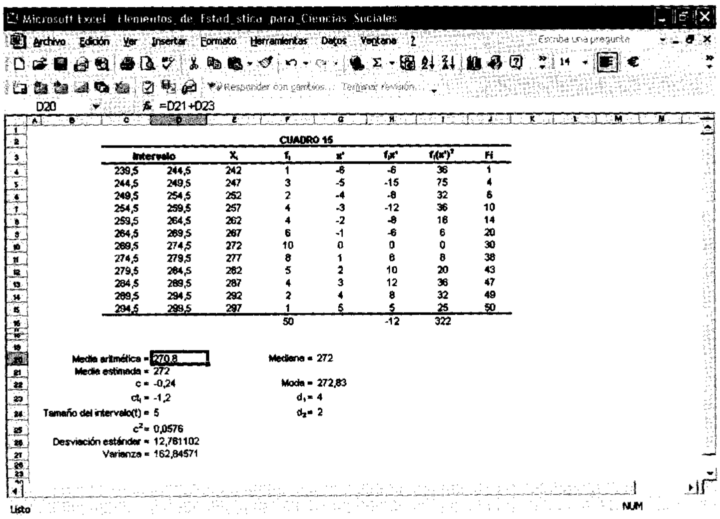
Cálculo de la media aritmética y la desviación estándar con el método abreviado mediante la hoja de cálculo de Excel

Con base en los datos del ejercicio del Cuadro 15 se describe a continuación el procedimiento a seguir en la hoja electrónica para calcular la media aritmética y la desviación estándar por este método.

Media aritmética

1. Abrir una hoja de cálculo.
2. A partir de la celda C3 anote los títulos en cada una de las columnas de la hoja de trabajo como se muestra en la Figura 25.
3. Escribir el límite inferior del intervalo de la primera clase en la celda C4.
4. En la celda D4 colocar la fórmula: =C4+D24 donde D24 es el tamaño del intervalo.
5. Ubicarse en la celda C5 y asignarle el valor de la celda D4 como: = D4, después copiar la fórmula en D4 a D5. Copiar C5 y D5 hasta la celda C15 y D15.
6. Ubicarse en la celda E4 y construir manualmente la fórmula: = (C4+D4)/2; la cual calcula el valor de la marca de clase del primer intervalo.
7. Se copia la fórmula de la celda E4, hasta la celda E15.
8. Se escriben los datos de la frecuencia en la columna f_i, de la celda F4 a la celda F15 (1, 3, 2, 4, 4, 6, 10, 8, 5, 4, 2, 1).
9. Posteriormente ubicarse en la celda F16.
10. Pulsar el icono de funciones de la barra de herramientas estándar, enseguida aparece una ventana que permite seleccionar diferentes categorías (véase Figura 26).
11. Seleccionar la categoría de funciones matemáticas y trigonométricas.
12. Seleccionar la función de suma.
13. Aparece una ventana de diálogo, que solicita el rango de suma (véase Figura 27). El rango puede indicarse con el seleccionador de rango (con ayuda del ratón) o escribiéndolo directamente.

FIGURA 25



D20 =D21+D23

CUADRO 15

Intervalo		X_i	f_i	x'	f_ix'	$f_i(x')^2$	Fi
239,5	244,5	242	1	-6	-6	36	1
244,5	249,5	247	3	-5	-15	75	4
249,5	254,5	252	2	-4	-8	32	6
254,5	259,5	257	4	-3	-12	36	10
259,5	264,5	262	4	-2	-8	16	14
264,5	269,5	267	6	-1	-6	6	20
269,5	274,5	272	10	0	0	0	30
274,5	279,5	277	8	1	8	8	38
279,5	284,5	282	5	2	10	20	43
284,5	289,5	287	4	3	12	36	47
289,5	294,5	292	2	4	8	32	49
294,5	299,5	297	1	5	5	25	50
			50		-12	322	

Media aritmética = 270,8 Mediana = 272
Media estimada = 272
c = -0,24 Moda = 272,83
ct_i = -1,2 d_1 = 4
Tamaño del intervalo(t) = 5 d_2 = 2
c^2 = 0,0576
Desviación estándar = 12,761102
Varianza = 162,84571

FIGURA 26

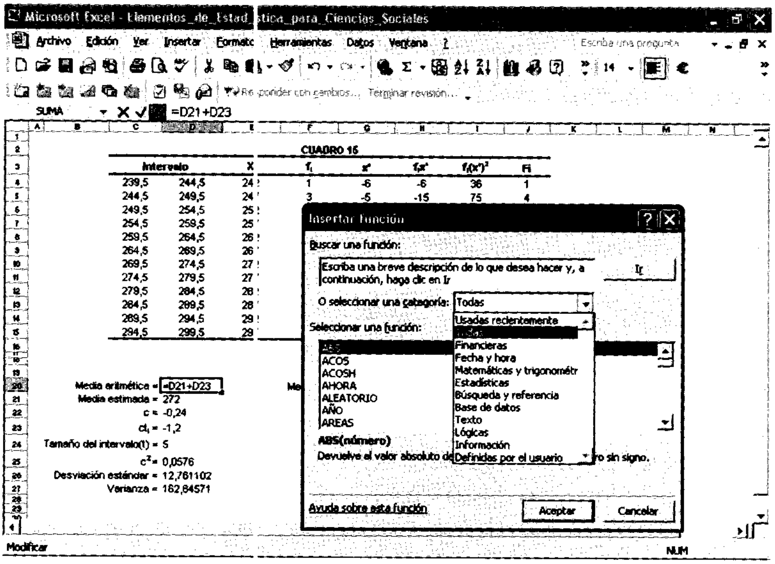
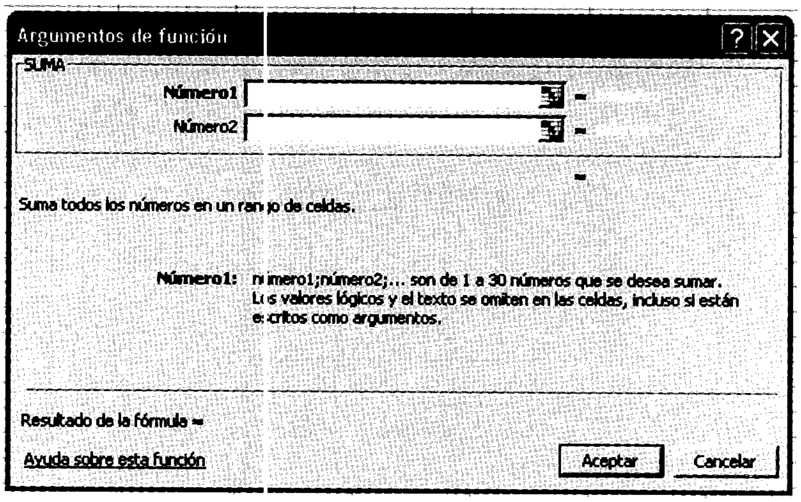


FIGURA 27



14. Anotar títulos en la celda C20 media aritmética =, en la celda C21, media estimada =, en la celda C22 escribir c =, en la celda C23 escribir ct_i =, en la celda C24 escribir tamaño del intervalo de clase (T)=, y en la celda C25 escribir c^2 =.
15. En la celda D21 anotar el valor de media estimada 272, y en la celda D24 anotar 5, el tamaño del intervalo de clase.
16. Llenar la columna de x', primero se coloca un cero en la celda G10, en forma consecutiva escribir los valores de desviación negativa correspondiente, desde la celda G9 hasta G4 (-1, -2, -3, -4, -5, -6), posteriormente anotar los valores positivos en forma consecutiva de la celda G11 hasta la G15 (1, 2, 3, 4, 5).
17. En la celda H4 en forma manual anotar el producto de la celda F4 y la celda G4: =F4*G4, después copiar la fórmula hasta la celda H15.
18. En la celda H16, realizar la sumatoria de la celda H4 hasta la celda H15.
19. En D22, escribir la operación: = H16/F16.
20. En la celda D23 multiplicar la celda D22 por la celda D24: = D22*D24.
21. Seleccionar la celda D20 para calcular el valor de la media aritmética por el método abreviado sumando las celdas D21 y D23: = D21+D23.

Desviación estándar

1. Ubicarse en la celda D25 y elevar al cuadrado el valor de la celda D22: = (D22)^2.
2. Seleccionar la celda I4 y multiplicar en ella la celda G4 por la celda H4: =G4*H4, posteriormente copiar esta fórmula hasta la celda I15.
3. En la celda I16 ubicar la sumatoria desde la celda I4 hasta I15.
4. Anotar los títulos de desviación estándar en la celda C26 y de la varianza en la celda C27.
5. En la celda D26 en forma manual anotar la fórmula de la desviación estándar del método abreviado

=(((I16/(F16-1))-D25)^0.5)*D24

desviación estándar = 12.761102

6. En la celda D27 elevar al cuadrado el valor de la celda D26: = D26^2, encontrando con ello el valor de la varianza.

varianza = 162.84571

Mediana

1. Escribir en la celda F20 Mediana =.
2. Construir la columna de frecuencia acumulada (F_i), primero copiando el valor de la celda F4 en la celda J4: = F4.
3. En la celda J5 escribir la fórmula: = J4+F5, posteriormente copiar esta fórmula hasta la celda J15.
4. En la celda G20 escribir la fórmula de la mediana[1] en forma manual como:

=(((F16/2)-J9)/F10)*D24)+C10

mediana = 272

Moda

1. Ubicarse en la celda F22 y escribir el título Moda =, en la celda F23 escribir d_1 =, y en la celda F24 escribir d_2 =.
2. En G23 escribir la fórmula: = F10-F9.
3. En G24 escribir la fórmula: = F10-F11.
4. En la celda G22 en forma manual escribir la fórmula de la moda:[2]

=(G23/(G23+G24)*D24)+C10

moda = 272.83

Método abreviado de cálculo de la media aritmética y la desviación estándar con datos agrupados en frecuencia

Para explicar los pasos requeridos en el procedimiento de cálculo de estos estadísticos nos auxiliaremos con el Cuadro 16, donde los datos a analizar han sido agrupados de acuerdo con la frecuencia que presentan en el problema que a continuación se describe.

PROBLEMA
Se cuenta con los datos de edad de 27 jóvenes y adolescentes y se desea conocer su promedio de edad y la dispersión que ésta presenta.

[1] La fórmula empleada en este cálculo está en la página 98.
[2] La fórmula empleada en este cálculo está en la página 100.

Datos (años cumplidos)
22, 20, 21, 18,19, 19, 18, 18, 18, 16, 16, 17, 16,
17, 17, 17, 15, 15, 11, 12, 12, 17, 14, 14, 13, 15, 13

CUADRO 16

(1)	(2)	(3)	(4)	(5)	(6)
No	X_i	f_i	x'	f_i x'	f_i x')^2
1	11	1	-6	-6	36
2	12	2	-5	-10	50
3	13	2	-4	-8	32
4	14	2	-3	-6	18
5	15	3	-2	-6	12
6	16	3	-1	-3	3
7	17	5	0	0	0
8	18	4	1	4	4
9	19	2	2	4	8
10	20	1	3	3	9
11	21	1	4	4	16
12	22	1	5	5	25
Σ		27		-19	213

FUENTE: datos hipotéticos.

El proceso de cálculo de la media aritmética y la desviación estándar cuando los datos están agrupados en frecuencia es:

1. Primero encontramos la media estimada ($\overline{X}_E$) tomando como referencia el dato de mayor frecuencia. En el Cuadro 16, renglón 7, se encuentra el dato 17, el cual tiene la mayor frecuencia (5), y es casi el centro de la distribución. La media estimada para este caso es diecisiete ($\overline{X}_E = 17$).
2. Llenamos la columna (4) x', anotando la desviación que existe entre cada dato con respecto al dato donde está ubicada la media estimada. El dato 17 del renglón (7) no está desviado respecto de la media estimada puesto que corresponde con el valor de ésta, entonces anotamos el valor de cero en la columna x'. Todos los valores menores a 17, al llenar la columna x' tomarán valores negativos y los superiores a 17 serán valores positivos. En el renglón (6) el dato es 16, colocamos un -1 en la columna x', en el reglón (5) el dato es 15, colocamos un -2 en la columna x', así sucesivamente hasta el primer renglón en el cual el dato está desviado seis posiciones y se anota un -6 en la columna x'. Los datos superiores

a 17 se encuentran a partir del renglón (8); para este caso colocamos en la columna x' el valor de 1, y así sucesivamente encontramos las desviaciones 2, 3, 4 y 5 de cada dato respecto de la media estimada.

3. Calculamos la columna (5) f_i x', multiplicando cada x' de la columna (4) por la frecuencia correspondiente de la columna (3). Para el renglón (1) la desviación es de -5 por su frecuencia 1, se obtiene como resultado -5, el cual se coloca en la columna (6); así continuamos realizando estas multiplicaciones para cada uno de los renglones subsecuentes y los resultados los anotamos como ya indicamos en la columna (5).
4. Encontrar la corrección de puntaje (c) realizando la suma algebraica de las f_i x' positivas y negativas, y posteriormente las dividimos por n, como se indica en la fórmula siguiente:

$$c = \frac{\sum_{i=1}^{k} (fx'^{+} + fx'^{-})}{n}$$

En la columna (5) la suma de los valores positivos es 20, y la de los valores negativos -39. En la columna (3) se encuentra el renglón: $\sum$, se suma el total de datos (n) que forman la muestra, para el ejemplo: 27.

$$c = -\frac{20 - 39}{27} = \frac{-19}{27} = -0.7037$$

5. Finalmente encontrar el valor de la media aritmética $\overline{X}$ sumando algebraicamente la media estimada $\overline{X}_E$ y la corrección de puntaje *c*.

$$\overline{X} = \overline{X}_E + c$$

$$\overline{X} = 17 - 0.7037 = 16.2963$$

$$\overline{X} = 16.3$$

Para calcular la desviación estándar debemos utilizar la fórmula del método abreviado siguiente:

$$s = \sqrt{\frac{\sum_{i=1}^{k} f_i x'^2}{n-1} - c^2}$$

6. De esta fórmula se observa que, primero, se multiplican los valores de la columna (4) x' por los de la columna (5) f_i x' obteniendo como resultado la columna (6) f_i $(x')^2$. En la columna (6) renglón (1) el valor de x' es de -6 y el de f_i x' es -6, entonces el valor en f_i $(x')^2$ es de 36; este proceso se realiza para los demás renglones de la columna en el Cuadro 16. Posteriormente se suman los resultados de la columna obteniéndose 213.
7. La corrección en unidades (c) se eleva al cuadrado:

$$c^2 = (-0.7037)^2 = 0.49519$$

8. Finalmente se sustituyen los valores en la fórmula de la desviación estándar por el método abreviado, establecida anteriormente.

$$s = \sqrt{\frac{213}{27-1} - 0.49519} = \sqrt{7.6971} = 2.774$$

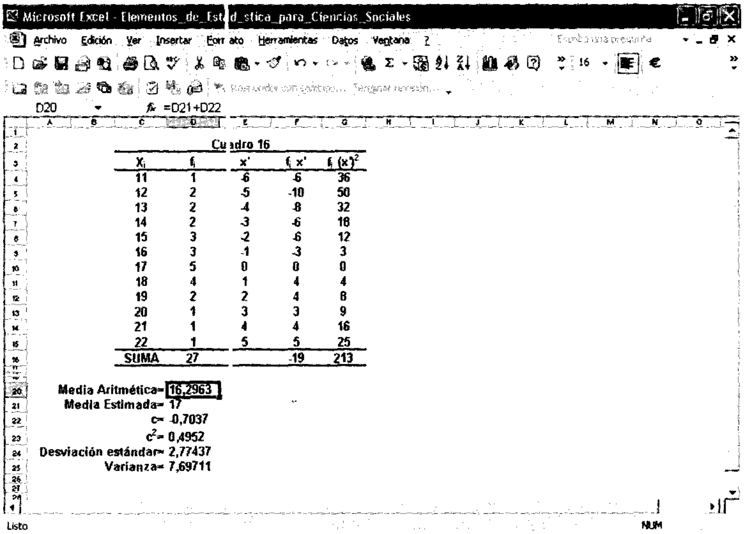
$$s = 2.774$$

Para calcular el valor de la varianza s^2, sólo se eleva al cuadrado el valor de la desviación estándar:

$$s^2 = (2.774)^2 = 7.697$$

En la Figura 28 se muestra en Excel el ejercicio elaborado con el Cuadro 16. El proceso de construcción y los cálculos realizados en él siguen los procedimientos establecidos previamente para el Cuadro 15.

FIGURA 28


Microsoft Excel - Elementos_de_Estadística_para_Ciencias_Sociales

D20 =D21+D22

Cuadro 16

X_i	f_i	x'	f_i x'	f_i $(x')^2$
11	1	-6	-6	36
12	2	-5	-10	50
13	2	-4	-8	32
14	2	-3	-6	18
15	3	-2	-6	12
16	3	-1	-3	3
17	5	0	0	0
18	4	1	4	4
19	2	2	4	8
20	1	3	3	9
21	1	4	4	16
22	1	5	5	25
SUMA	27		-19	213

Media Aritmética= 16,2963
Media Estimada= 17
c= -0,7037
c^2= 0,4952
Desviación estándar= 2,77437
Varianza= 7,69711

Listo

Cálculo de la desviación estándar con datos no agrupados

El método abreviado de cálculo tiene una aplicación inmediata cuando los datos a analizar no están agrupados, es decir, el investigador o analista cuenta con los datos originales. En estos casos es muy recomendable usar los métodos abreviados ya que nos permiten ahorrar tiempo, trabajo de cálculo y no existe la necesidad de ordenar los datos por magnitud. El Cuadro 17 contiene un ejemplo con 10 datos y nos sirve de apoyo para explicar la aplicación de estos métodos en estos casos.

CUADRO 17

Cálculo de la desviación estándar de datos no agrupados

(1)	(2)	(3)	(4)
No.	Datos (X_i)	x'	$(x')^2$
1	210	210	44100
2	250	250	62500
3	190	190	36100
4	310	310	96100
5	180	180	32400
6	270	270	72900
7	220	220	48400
8	280	280	78400
9	200	200	40000
10	250	250	62500
Σ	2360	2360	573400

FUENTE: datos hipotéticos.

x' = denota la desviación de un puntaje X_i de la media estimada

PROCEDIMIENTO de cálculo de la desviación estándar por el método abreviado para datos no agrupados.

1. Se establece la media estimada $\overline{X}_E$. Esta media toma el valor de cero $\overline{X}_E = 0$.
2. Cada dato X_i permanece sin ningún cambio. Columna (3).
3. La corrección en unidades (c) se calcula realizando la diferencia de la media verdadera y la media estimada.

$$c = \frac{\sum_{i=1}^{n} X_i}{n} - \overline{X}_E = \frac{2360}{10} - 0 = 236$$

4. Se eleva al cuadrado cada dato de la columna (3) x', y su resultado se ubica en la columna (4) $(x')^2$. Finalmente se suman todos los resultados de la columna (4).
5. Se aplica la fórmula para el cálculo de la desviación estándar por el método abreviado:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x')^2}{n-1} - C^2}$$

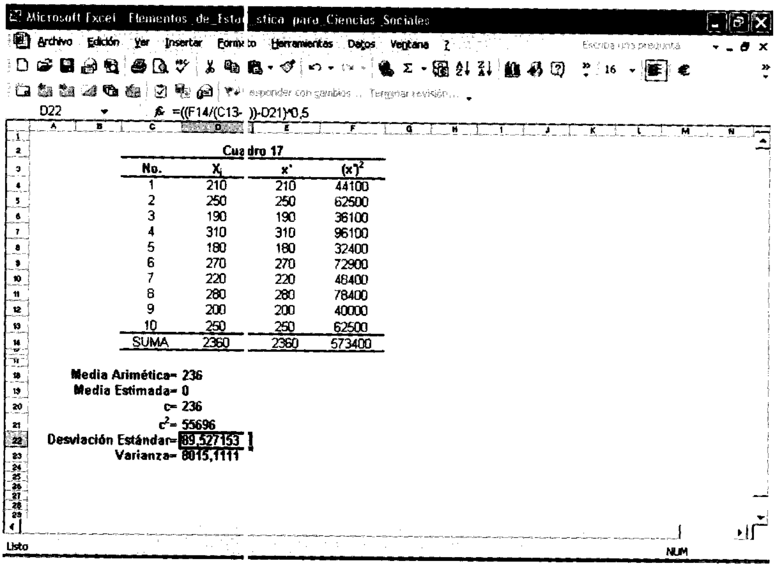
sustituyendo valores se tiene:

$$s = \sqrt{\frac{573400}{10-1} - 55696} = \sqrt{8015.11} = 89.527$$

$$s = 89.527$$

La Figura 29 muestra en Excel el ejercicio elaborado con el Cuadro 17.

FIGURA 29



No.	X_i	x'	(x')²
1	210	210	44100
2	250	250	62500
3	190	190	36100
4	310	310	96100
5	180	180	32400
6	270	270	72900
7	220	220	48400
8	280	280	78400
9	200	200	40000
10	250	250	62500
SUMA	2360	2360	573400

Cálculo de la desviación estándar por el método abreviado con datos ordenados por magnitud

El Cuadro 18 contiene un ejemplo con 8 datos y nos sirve de apoyo para explicar el cálculo de la desviación estándar por el método abreviado cuando los datos han sido ordenados por magnitud.

CUADRO 18

(1)	(2)	(3)	(4)
No.	X_i	x	x^2
1	3	-3	9
2	5	-2	4
3	6	-1	1
4	7	0	0
5	8	1	1
6	8	2	4
7	9	3	9
8	10	4	16
$\sum$	56		44

FUENTE: datos hipotéticos.

x = desviación de un puntaje X_i desde la media verdadera $\overline{X}$ de la distribución.

El PROCEDIMIENTO de cálculo es el siguiente:

1. Ordenar por magnitud los datos del problema y numerarlos como se indica en la columna (1) del cuadro.
2. Calcular el valor de la media aritmética.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{56}{8} = 7$$

media aritmética = 7

3. Llenar la columna (3) x. Esta columna es el resultado de anotar la diferencia que existe entre cada renglón de datos y el renglón donde se ubica la media aritmética (renglón 4). En el renglón 4 se coloca el número cero porque no hay desviación del puntaje 7 a la media aritmética. Para los datos menores al valor de la media aritmética las desviaciones son negativas y para los mayores positivas. Por ejemplo para el primer dato, la diferencia es: 1-4 = -3, para el segundo: 2-4 = -2, y así sucesivamente.
4. Se eleva al cuadrado la columna (3) x, y el resultado se anota en la columna (4) x^2, posteriormente se realiza la sumatoria de todos los resultados de esta columna.
5. La fórmula para el cálculo de la desviación estándar por el método abreviado para datos ordenados por magnitud es:

$$s = \sqrt{\frac{\sum_{i=1}^{n} x^2}{n-1}}$$

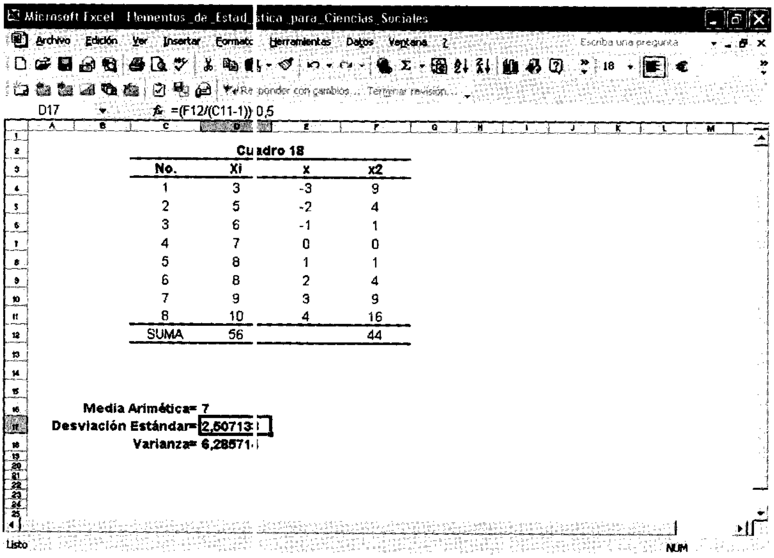
sustituyendo valores:

$$s = \sqrt{\frac{44}{8-1}} = \sqrt{6.28} = 2.5$$

$$s = 2.5$$

La Figura 30 muestra en Excel el ejercicio elaborado con el Cuadro 18.

FIGURA 30



Cuadro 18

No.	Xi	x	x2
1	3	-3	9
2	5	-2	4
3	6	-1	1
4	7	0	0
5	8	1	1
6	8	2	4
7	9	3	9
8	10	4	16
SUMA	56		44

Ejercicios

- Veinte vendedores realizaron visitas a diferentes empresas durante una semana, obteniendo el siguiente número de ventas: 2, 2, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5, 5, 6, 6, 6, 7, 7, 8, 9. Calcular: media aritmética, desviación estándar y varianza utilizando el método abreviado. También encontrar el valor de la mediana y moda.
- En dos manzanas de la colonia José López Portillo, en la delegación Iztapalapa, se tomó una muestra de 10 departamentos con todos los servicios y de tres recamaras. Se realizaron diferentes preguntas a las personas que los habitan y una de ellas es la siguiente: ¿cuánto pagan de renta mensual en miles de pesos? Se obtuvieron las siguientes respuestas: 1.4, 2.4, 1.7, 1.5, 1.8, 1.8, 2.3, 2.1, 2.0, 1.6. Encontrar la renta promedio y la desviación estándar con el método abreviado.
- La producción de lápices con características novedosas por hora de 300 obreros de una compañía manufacturera de Tecate, Baja California, tiene variaciones en sus unidades producidas de 56 a 65. La distribución de frecuencia de la producción en la última semana se muestra en el cuadro siguiente:

No.	X_i	f
1	56	9
2	57	15
3	58	35
4	59	36
5	60	54
6	61	60
7	62	35
8	63	25
9	64	23
10	65	8
Σ		300

Encontrar el valor de la media aritmética y la desviación estándar por el método abreviado.

- Un cobrador de la empresa Hilos y Tejidos Elegantes, tiene en sus registros el número de días que tarda en cobrar (X_i) cada una de las 30 líneas de crédito que tiene a su cargo. En el cuadro siguiente se muestra esta distribución de frecuencia (X_i).

X_i	4	5	10	13	14	19	20	26	28	30	35	38	42	60	90	
f_i	1	1	2	1	2	2	2	3	5	3	2	2	1	2	1	$\Sigma = 30$

Calcular el valor de la media aritmética y la desviación estándar por el método abreviado.

- Con el siguiente cuadro de distribución de frecuencias, calcular el valor de la media aritmética, la desviación estándar y la varianza por el método abreviado.

No.	Intervalo de clase	f
1	2.25 – 2.75	4
2	2.75 – 3.25	6
3	3.25 – 3.75	5
4	3.75 – 4.25	5
5	4.25 – 4.75	3
6	4.75 – 5.25	1
7	5.25 – 5.75	1
Σ		25

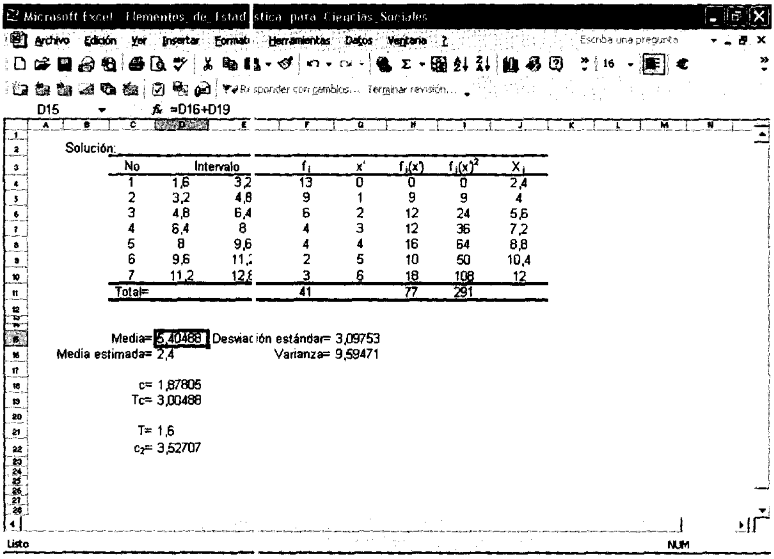
- De las cien empresas más importantes en México en 1999 se tomó una muestra en forma aleatoria de 41 empresas para conocer el número de empleados que tenía contratados cada una de ellas. Los datos obtenidos se muestran a continuación (el número de empleados está dado en cientos):

1.5, 1.6, 1.8, 2.0, 2.3, 2.4, 2.4, 2.6, 2.6, 2.6, 2.6, 2.7, 2.8, 3.2,
3.3, 3.7, 3.7, 3.8, 3.8, 3.9, 3.9, 4.7, 4.8, 5.0, 5.4, 5.4, 5.5, 5.5,
6.6, 6.6, 7.4, 7.7, 8.1, 8.7, 9.2, 9.4, 9.9, 10.0, 10.2, 11.6, 11.7

- Construir un cuadro de distribución de frecuencias considerando el tamaño del intervalo de clase de 1.6 unidades y el límite inferior del primer intervalo en 1.6.
- Calcular el valor de la media aritmética, la desviación estándar y la varianza por el método abreviado.

La solución de este problema se muestra a continuación (Figura 31).

FIGURA 31


Solución:

No	Intervalo		f_i	x'	$f_i(x')$	$f_i(x')^2$	X_i
1	1,6	3,2	13	0	0	0	2,4
2	3,2	4,8	9	1	9	9	4
3	4,8	6,4	6	2	12	24	5,6
4	6,4	8	4	3	12	36	7,2
5	8	9,6	4	4	16	64	8,8
6	9,6	11,[illegible]	2	5	10	50	10,4
7	11,2	12,[illegible]	3	6	18	108	12
Total=			41		77	291	

Media= 6,40488 Desviación estándar= 3,09753
Media estimada= 2,4 Varianza= 9,59471

c= 1,87805
Tc= 3,00488

T= 1,6
c_2= 3,52707

Resumen de fórmulas estadísticas y funciones en Excel

Medidas de posición o tendencia central

ESTADÍSTICO	FÓRMULA	FUNCIÓN EN EXCEL (EJEMPLO)
Media aritmética muestral para datos no agrupados	$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$	= PROMEDIO (rango de datos) = PROMEDIO (A1:A10)
Media aritmética poblacional	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$	= PROMEDIO (rango de datos) = PROMEDIO (A1:A10000)
Media aritmética muestral o poblacional para datos agrupados	$\bar{X}, \mu = \frac{\sum_{i=1}^{k} f_k X_k}{n}$	No existe pero se puede construir a partir de la tabla de distribución.
Media ponderada	$\bar{X}w = \frac{\sum_{i=1}^{e} w_i X_i}{\sum_{i=1}^{e} w_i}$	No existe pero se puede construir a partir de una tabla con datos y ponderadores.
Media geométrica	$MG = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \ldots X_n}$	= MEDIA. GEOM (dato 1, dato 2, dato 3... dato_n) = MEDIA. GEOM (A1:A10)
Mediana para número de datos impar no agrupados	Ume = (n+1)/2 Ubica a la mediana con datos ordenados	= MEDIANA (rango del conjunto de datos) =MEDIANA(A1:A10)
Mediana para número de datos par no agrupados	Me= [Dato (n/2) + Dato (n/2) + 1]/2	= MEDIANA (rango del conjunto de datos) = MEDIANA (A1:A10)
Mediana para datos agrupados	$Me = Lim + \left(\frac{\frac{n}{2} - Fac}{f}\right)T$	No existe. Se puede construir con una macroinstrucción.

Medidas de posición o tendencia central (continúa)

Estadístico	Fórmula	Función en Excel (ejemplo)
Moda para datos no agrupados	Dato con mayor frecuencia en el conjunto	= MODA (rango de datos) = MODA (A1:A10)
Moda para datos agrupados	$Mo = Lim + \left(\frac{d_1}{d_1 + d_2}\right)T$	No existe. Se puede construir con una macroinstrucción
Cuartiles	$Qk = Lk + \left(\frac{k\,[n/4] - Fk}{fk}\right)T$ Datos agrupados	Datos no agrupados = CUARTIL (rango de datos, número de cuartil deseado) = CUARTIL (A1:A10,1)
Deciles	$Dk = Lk + \left(\frac{k\,[n/10] - Fk}{fk}\right)T$ Datos agrupados	No existe. Se puede usar la función PERCENTIL en datos no agrupados
Percentiles	$Pk = Lk + \left(\frac{k\,[n/100] - Fk}{fk}\right)T$ Datos agrupados	Datos no agrupados = PERCENTIL (rango de datos, número de percentil deseado en decimales) = PERCENTIL (A1:A10, 0.33)

Medidas de dispersión o variabilidad

Estadístico	Fórmula	Función en Excel (ejemplo)
Amplitud o Rango	Amp = Dmy - Dme Datos no agrupados: diferencia entre el valor máximo y el valor mínimo Datos agrupados: límite real superior de la última clase menos el límite real inferior de la primera clase (agrupación ascendente)	Datos no agrupados = MAX (rango de datos) - MIN (rango de datos) = MAX (A1:A7) - MIN (A1:A7)

Medidas de dispersión o variabilidad (continúa)

Estadístico	Fórmula	Función en Excel (ejemplo)
Desviación absoluta promedio para datos no agrupados	$DAP = \frac{\sum_{i=1}^{n} \lvert X_i - \bar{X} \rvert}{n}$	= DESVPROM (rango de datos) = DESVPROM (A1:A10)
Desviación absoluta promedio para datos agrupados	$DAP = \frac{\sum_{i=1}^{k} \lvert X_i - \bar{X} \rvert f_i}{n}$	No existe. Se puede construir con una macroinstrucción.
Varianza para una muestra con datos no agrupados	$S^2 = \frac{\sum_{i=1}^{n} [X_i - \bar{X}]^2}{n-1}$	= VAR (rango de datos) = VAR (A1:A10)
Varianza para una población con datos no agrupados	$\sigma^2 = \frac{\sum_{i=1}^{N} [X_i - \mu]^2}{N}$	= VARP (rango de datos) = VARP (A1:A16000)
Varianza para una muestra con datos agrupados	$S^2 = \frac{\sum_{i=1}^{k} [X_i - \bar{X}]^2 f_i}{n-1}$	No existe. Se puede construir con una macroinstrucción.
Varianza para una población con datos agrupados	$\sigma^2 = \frac{\sum_{i=1}^{k} [X_i - \mu]^2 f_i}{N}$	No existe. Se puede construir con una macroinstrucción.
Desviación estándar para una muestra con datos no agrupados	$S = \sqrt{\frac{\sum_{i=1}^{n} [X_i - \bar{X}]^2}{n-1}}$	= DESVEST (rango de datos) = DESVEST (A1:A10)
Desviación estándar para una población con datos no agrupados	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} [X_i - \mu]^2}{N}}$	= DESVESTP (rango de datos) = DESVESTP (A1:A16000)

Medidas de dispersión o variabilidad (continúa)

ESTADÍSTICO	FÓRMULA	FUNCIÓN EN EXCEL (EJEMPLO)
Desviación estándar para una muestra con datos agrupados	$S = \sqrt{\dfrac{\sum_{i=1}^{k} [X_i - \overline{X}]^2 f_i}{n-1}}$	No existe. Se puede construir con una macroinstrucción.
Desviación Estándar para una población con datos agrupados	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{k} [X_i - \mu]^2 f_i}{N}}$	No existe. Se puede construir con una macroinstrucción.
Dispersión relativa: Coeficiente de variación	$CV = \dfrac{S}{\overline{X}} 100\%$	= [DESVEST (rango de datos)/ PROMEDIO (rango de datos)]*100 = [DESVEST (A1:A10)/ PROMEDIO (A1:A10)]*100

Medida de sesgo (asimetría)

ESTADÍSTICO	FÓRMULA	FUNCIÓN EN EXCEL (EJEMPLO
Coeficiente de Sesgo (de Asimetría) para datos no agrupados	$CS = \dfrac{n}{[n-1][n-2]} \sum_{i=1}^{n} \left(\dfrac{X_i - \overline{X}}{S} \right)^3$	= COEFICIENTE.ASIMETRIA (rango de datos) = COEFICIENTE.ASIMETRIA (A1:A10)
Coeficiente de Sesgo (de Asimetría) para datos agrupados	$CS = \dfrac{\sum_{i=1}^{k} f_i [X_i - \overline{X}]^3 / n}{S^3}$	No existe. Se puede construir con una macroinstrucción.

Medida de curtosis

ESTADÍSTICO	FÓRMULA
Coeficiente de Curtosis para datos no agrupados	$CC = \left(\frac{n[n+1]}{[n-1][n-2][n-3]} \sum_{i=1}^{n} \left(\frac{[X_i - \bar{X}]}{S} \right)^4 \right) - \frac{3[n-1]^2}{[n-2][n-3]}$
	FUNCIÓN EN EXCEL (EJEMPLO)
	= CURTOSIS (rango de datos) = CURTOSIS (A1:A10)
Coeficiente de Curtosis para datos agrupados	FÓRMULA
	$CC = \frac{\sum_{i=1}^{k} f_i [X_i - \bar{X}]^4 / n}{S^4}$
	FUNCIÓN EN EXCEL (EJEMPLO)
	No existe. Se puede construir con una macroinstrucción.

Algunas consideraciones para el diseño de encuestas, cuestionarios y muestras en ciencias sociales*

*Alberto Isaac Pierdant Rodríguez***
*Jesús Rodríguez Franco****

Introducción

En el área de las ciencias sociales, los datos que se obtienen sobre un problema de investigación, y su consecuente análisis, representan para el investigador una tarea importante que debe definirse con la mayor precisión posible. La obtención de datos e información relevantes para el desarrollo de estudios en ciencias sociales (estudios de mercado, de opinión, económicos, etcétera) y cuyo origen proviene directamente de la fuente, se ven involucrados en el uso y aplicación de una encuesta.

La encuesta es un proceso que permite obtener datos e información de un conjunto de individuos, objetos o hechos con la finalidad de probar una o varias hipótesis de investigación. Es decir,

> [...] la encuesta se utiliza para captar información acerca de un cierto grupo o población de objetos. Estas encuestas son mediciones en un momento determinado, por lo que no puede establecerse que sus resultados sean indicadores precisos de lo que ocurrirá meses después. Su calidad está condicionada por un gran número de factores que pueden afectar su grado de predicción [Pimienta, 2000:263-264].

Es por ello que el especialista en ciencias sociales debe establecer un adecuado criterio en su diseño, desarrollo y validación, al aplicarlos a cualquier problema del

* Artículo publicado en *Quehacer científico. Un panorama actual en la UAM-Xochimilco*, tomo I, Universidad Autónoma Metropolitana, Unidad Xochimilco, México, 2004.

** Maestro en Ingeniería. Profesor e investigador del Departamento de Política y Cultura de la Universidad Autónoma Metropolitana-Xochimilco.

*** Maestro en Ciencias. Profesor e investigador del Departamento de Política y Cultura de la Universidad Autónoma Metropolitana-Xochimilco.

ámbito social. Este documento tiene como principal objetivo mostrar algunas consideraciones que deben tomarse en cuenta al desarrollar este proceso.

Nuestro análisis y propuesta de consideraciones se centra en tres aspectos. Primero, la encuesta en sí misma, desde su etapa de diseño hasta la publicación de los resultados de una investigación mediante un reporte. Segundo, la elaboración de una herramienta fundamental en ella: el cuestionario, y finalmente, los elementos matemáticos de muestreo básicos requeridos en este tipo de estudios.

La encuesta

En las diversas áreas de trabajo de las ciencias sociales, la sociología, la economía, la administración, etcétera, es común realizar encuestas sin conocimiento de ello o con una metodología de diseño elemental, lo que genera resultados pobres en una investigación. Por lo que en esta primera etapa nos proponemos revisar un método simple de diseño de encuestas desarrollado por A.N. Oppenheim (1997), el cual hemos adaptado con cierto éxito a nuestras investigaciones.

Este método requiere de ciertos conocimientos técnicos y de un trabajo continuo de análisis que perfeccione paso a paso el estudio, por lo que no se debe pensar que una sola propuesta o línea de investigación es la única posible. Aun ésta deberá perfeccionarse a medida que se avanza con las etapas de este proceso.

El método,[1] con algunas adaptaciones y observaciones nuestras, establece las siguientes etapas:

1. Decidir las metas y objetivos del estudio. Deberán establecerse primeramente las metas generales del estudio, de éstas surgirán los objetivos específicos, y de estos últimos se podrán establecer de manera práctica las metas que son posibles de alcanzar con el estudio.
   En las áreas de las ciencias sociales este es un proceso que puede tomar hasta 80 por ciento del tiempo dedicado a una investigación específica; por ello, en esta etapa debemos determinar con precisión cuáles serán nuestras variables a medir, cómo las vamos a medir, es decir qué escala o indicador usaremos y, por supuesto, el conjunto de preguntas para cada una de ellas que nos permitan explicarlas.
2. Revisar la literatura relevante sobre el tema, discutiéndola con especialistas y organizaciones interesadas. Verificar si existen estudios similares al nuestro, cuáles

[1] Para un mayor detalle, el método original puede consultarse en el capítulo: "Introduction to survey design" del libro de Oppenheim (1997).

fueron las experiencias obtenidas, de tal forma que ello nos permita desarrollar un mejor proyecto.

3. Hacer una conceptualización preliminar del estudio, seguida de una serie de entrevistas exploratorias o a profundidad con los especialistas o individuos que nos puedan aportar ideas. En este momento es importante que sean revisados los objetivos y las metas que nos hemos propuesto originalmente, lo que nos permitirá determinar su vigencia en el estudio.
4. Seleccionar el diseño del estudio (encuesta descriptiva o encuesta analítica), verificar la factibilidad del tiempo, costos y personal requerido para el mismo. Una encuesta descriptiva es aquella que enumera las características de un problema, es decir, es una encuesta del tipo de un censo. Por ejemplo, una encuesta descriptiva industrial, nos indicaría, cuántas industrias hay, de qué tamaño, cuál o cuáles son sus productos, número de empleados, capital invertido, tipo de maquinaria, etcétera. Estas encuestas tienen como principal objetivo contar; cuando esto no es posible, es decir, estudiar a todos los elementos de una población que se encuentra bajo análisis, entonces se establece una muestra o subconjunto de ella, que la represente (muestra representativa) y que permita inferir el comportamiento de toda la población.
   Los estudios de mercado que realizan muchas organizaciones privadas o públicas, caen en muchos casos en esta categoría.
   La encuesta analítica busca responder a preguntas del tipo ¿por qué?, más que a preguntas de tipo ¿cuántos? En este sentido, este tipo de encuestas tienen como objetivo examinar las diferencias entre grupos y las relaciones que se pueden dar entre las variables de estudio. En ellas es común investigar el por qué esto está sucediendo y si esto está relacionado con algún otro factor. En ellas se realiza un análisis de causalidad.
   En relación con el tiempo requerido del estudio, sería bueno determinar el número de semanas o meses que tomará, su costo para nuestra organización y el personal necesario que permitirá llevarlo a cabo.
5. Decidir cuáles hipótesis serán investigadas. En este punto deberemos hacer nuestras suposiciones o hipótesis acordes a la situación que se presenta en el estudio. Es decir, hacer nuestras hipótesis operacionales, mencionando en ellas las variables que tendrán que ser medidas.
   Por ejemplo, si queremos saber si el diseño de un nuevo producto impacta o no a un consumidor, debemos establecer variables como: color, tamaño, peso, costo, marca, etcétera, que midan con cierta precisión dicho impacto.
6. Diseñar o adaptar los instrumentos y técnicas de investigación necesarios como, por ejemplo, un cuestionario, cuestionarios por correo, programa de entrevistas, escalas de actitud, métodos proyectivos, etcétera.

Recordemos en este punto que el principal instrumento de recolección de datos en México es el cuestionario; éste no es simplemente una lista de preguntas que debe ser llenada; es, como indican los investigadores en la materia, un instrumento de recolección de datos específicos. Las preguntas[2] y su diseño se establecen a partir de los objetivos y metas que han sido definidas para nuestro estudio.

7. Realizar las pruebas piloto de los instrumentos de recolección de datos que permitan hacer las revisiones y correcciones necesarias de los mismos. Es decir, probar nuestra primera versión del cuestionario o entrevista propuesta.
   También en esta etapa debemos probar la forma de aplicación del instrumento, el tiempo requerido para su aplicación, el diseño adecuado para poder ser procesado más tarde en computadora, así como otros aspectos que juzguemos necesarios y que permitan obtener un buen resultado del proyecto.
8. Diseñar la muestra. Es decir, obtener una muestra de la población de estudio que sea representativa de ella. Si dicha muestra es así, el analista podrá responder a la pregunta, ¿de quién? Por otro lado, se debe verificar si existen muestras previas que puedan ser utilizadas, o bien determinar si se requiere de un grupo de control, etcétera. Recordemos que en la práctica es difícil, y muchas veces muy costoso, tratar de obtener los datos de todos los individuos o cosas que pretendemos estudiar, por lo que generalmente los estudios en ciencias sociales se hacen mediante el uso de muestras, es decir, una parte o subconjunto seleccionado de una población que sigue ciertos criterios establecidos en la teoría del muestreo.

> En una encuesta, la población es el agregado o colección de elementos que poseen las características que se desean investigar; ésta puede delimitarse espacial y temporalmente. A cada elemento de la población sujeta a investigación se le llama unidad de muestreo y al elemento de la población del cual se obtienen los datos se le denomina unidad de información; por ejemplo, en una encuesta de ingreso-gasto la unidad de investigación es la familia y la de información puede ser el padre, la madre o el jefe de familia, dependiendo esto de la persona que administre el ingreso familiar. Si el sondeo se hace en todos y cada uno de los elementos que conforman la población se habla de un censo, si sólo se hizo en una parte de ésta se habla de un muestreo [Pimienta, 2000:264].

La teoría del muestreo[3] nos permite establecer con cierta certidumbre la muestra representativa del estudio.

[2] Para un adecuado diseño de preguntas en cuestionarios puede consultarse a Foddy (1999).

[3] Sugerimos consultar los textos de Lohr (2000) y Cochran (1998). Y para un conocimiento más detallado del tema, el libro de Des Raj (1980).

9. Definir la muestra. Es decir, seleccionar a las personas que serán entrevistadas. Con base en lo establecido en el diseño de la muestra, el tamaño de la muestra y el tipo de muestreo a utilizar en nuestro estudio, deberemos establecer la mecánica de selección de los individuos que formarán nuestra muestra, así como el procedimiento que tendremos que utilizar para aplicar el instrumento (cuestionario, entrevista, etcétera) que hemos seleccionado previamente.
10. Hacer el trabajo de campo. Es decir, recolectar los datos, mediante las entrevistas o cuestionarios previamente diseñados y probados (véase punto 7) que han sido aceptados para su aplicación por el investigador.
11. Procesar los datos. En esta etapa deberemos codificar las respuestas, capturar los datos en una computadora (si esto es posible) y determinar los análisis y las pruebas estadísticas necesarias que permitan probar las hipótesis que hemos establecido para el estudio.
12. Hacer el análisis estadístico (tal vez una de las etapas más complejas del estudio). Consiste básicamente en realizar todos los cálculos y pruebas estadísticas que han sido definidas para cada una de las variables y sus posibles relaciones.
13. Estructurar los resultados y probar las hipótesis. Analizar los resultados estadísticos obtenidos por variable del estudio y verificar si las hipótesis establecidas son verdaderas. Esta etapa suele complementarse con la construcción de cuadros estadísticos y gráficas,[4] herramientas que permiten mostrar de manera simple el análisis que hemos desarrollado de los datos obtenidos.
14. Escribir el reporte del estudio. Reportar en el documento todas y cada una de las etapas que comprendió el estudio. Principalmente mostrar los resultados y los análisis realizados con ayuda de cuadros estadísticos y gráficas, lo que permitirá establecer las conclusiones e interpretaciones del estudio.

Como puede observarse, este método propuesto es complejo en su operación. Su uso por primera vez puede generar varios errores del investigador novato; por ejemplo, se puede establecer una pobre definición de metas y objetivos, una mala selección en cuanto al tipo de estudio a realizar, un desconocimiento del método de muestreo seleccionado, etcétera. Por ello, si se pretende aplicar este método, sugerimos que como primer paso se realice una adecuada planeación del proyecto de encuesta, lo que le permitirá detectar problemas previos a su ejecución, ya que en esta etapa dicho diseño se encuentra todavía en el papel, y ello hace posible una modificación o bien una mejora del mismo. Un diseño pobre de encuesta, en cualquiera de las etapas,

[4] Una revisión sobre la construcción de cuadros estadísticos y gráficas puede realizarse en Pierdant (2000), o bien puede consultarse en línea un resumen del procedimiento de construcción de cuadros estadísticos en [www.geocities.com/aipiertant/cuadros].

fallará para darnos resultados confiables sobre el problema que estemos estudiando, ya que se generarán muchos vacíos que nos impiden llegar a conclusiones adecuadas. Asimismo, se generará información irrelevante, desperdicio de materiales y recursos que generalmente son escasos en nuestras organizaciones.

El cuestionario

Una segunda consideración es la referente al diseño del instrumento de recolección de datos. Éste puede ser un cuestionario, o bien, podemos obtener los datos mediante una entrevista. En este punto, e independientemente del instrumento seleccionado, deberán cuidarse con bastante detalle su proceso de diseño, su construcción y su aplicación, ya que de ello depende la obtención adecuada de datos útiles que permitirán generar información relevante.

Cada cuestionario elaborado como herramienta de investigación debe ser "considerado como único, por definición" (Posner, 2002:10); en este sentido, el diseño de un cuestionario sólo podrá iniciarse si se parte de la base de que ya se cuenta con la especificación completa de las variables que se necesitan medir, así como de las escalas de medición que serán utilizadas para ello. Por supuesto que dichas variables han sido establecidas con base en los objetivos y metas ya propuestos para la encuesta.

No existe un procedimiento o metodología única que nos permita construir un cuestionario, por lo cual proponemos algunas ideas que consideramos útiles en este proceso.

Primeramente deberán agruparse las variables de investigación formando módulos que proporcionen información sobre un tema específico; por ejemplo, podrán plantearse un conjunto de preguntas que nos proporcionen un perfil mínimo de nuestros encuestados, es decir, que nos permitan conocer quién nos respondió. En cada módulo deberán plantearse las preguntas que respondan en la mejor forma a cada variable. De esta manera, el investigador social deberá definir los diversos módulos que formarán el cuestionario con base en los objetivos y metas establecidas en la encuesta.

Tampoco existe un método que permita definir con precisión cada una de las preguntas en un cuestionario; en este sentido, una guía útil la hemos obtenido de los trabajos de William Foddy (1999) y A.N. Oppenheim (1997).

Al diseñar preguntas para los módulos que forman un cuestionario, es conveniente vigilar los siguientes aspectos:

- Evitar hacer preguntas basadas en hechos.
- Utilizar un lenguaje adecuado para el entrevistado.
- Establecer cada pregunta lo más claro posible.
- Cuidar el formato de la pregunta.
- Cuidar el contexto cultural del entrevistado.
- Tratar de tener una secuencia de preguntas que genere interés en el entrevistado.
- Establecer un balance entre preguntas abiertas y preguntas cerradas.
- Diseñar adecuadamente el área de respuestas.

Todo cuestionario puede incluir tres tipos de preguntas: preguntas abiertas sin clasificación –aquellas donde el entrevistador trata de registrar la respuesta completa del entrevistado–, preguntas abiertas con clasificación –aquellas donde el entrevistador usa una clasificación de respuestas para la pregunta, pero esta clasificación no es leída al entrevistado–, y preguntas cerradas o de formato estructurado –donde existe una clasificación definida para la respuesta de la pregunta.

Las ventajas y desventajas que presentan cada uno de estos tipos de preguntas son tratadas adecuadamente por Oppenheim (1997:112-115). Esto último permite al investigador social seleccionar el tipo de pregunta indicada para cada módulo.

Para los especialistas de las ciencias sociales, las preguntas en cada módulo pueden diseñarse mediante la técnica de convergencia (también llamada del embudo). Esta técnica está constituida por cuatro etapas, definidas en la siguiente forma.

1. Inicie con una pregunta amplia sobre el tema. Por ejemplo: ¿cuál es su opinión sobre el café? o ¿qué piensa usted de la gente que toma café?
2. Elabore preguntas más restringidas sobre el tema. ¿Toma usted café en ocasiones? o ¿toma usted café desde joven?
3. Elabore preguntas todavía más restringidas que las anteriores. ¿Usted cree que el café puede dañar en algún sentido? o ¿qué desventajas hay al tomar café?
4. Elaborar las preguntas del problema de estudio pero procurando no plantearlas bajo una dirección específica (hasta donde sea posible). Algunas personas consideran que el café es malo para su organismo, pero otras piensan que esto no las afecta. ¿Qué piensa usted de esto?

En el proceso es posible utilizar preguntas filtro, las cuales pueden o no excluir al informante de un módulo particular dentro de nuestro cuestionario.

Como resultado de este proceso se obtiene un primer cuestionario, el cual deberá someterse a una prueba piloto (en algunas ocasiones a más de una) que permita comprobar su validez y eficacia en la obtención de datos para el estudio para el que ha sido diseñado. Con la verificación de cada pregunta en aspectos como su orden, vocabulario, contexto cultural utilizado y su formato, el investigador procederá a realizar las modificaciones pertinentes, obteniendo con ello el cuestionario definitivo del estudio. La encuesta está lista para la siguiente etapa: el muestreo.

El muestreo

La tercera consideración importante es el diseño de la muestra.

> Las encuestas por muestreo se clasifican en dos grandes grupos: encuestas a partir de muestras probabilísticas y encuestas a partir de muestras no probabilísticas. En la práctica se habla de muestreo probabilístico y de muestreo no probabilístico. En un muestreo de tipo probabilístico, a partir de la muestra se pueden hacer inferencias sobre el total de la población; en uno no probabilístico, solamente la población investigada, es decir, únicamente sobre los elementos estudiados [Pimienta, 2000:264].

Una diferencia fundamental entre estos tipos de encuestas es la selección que se hace de la muestra y con ello la representatividad que ésta tenga de la población de estudio.

> [En la encuesta probabilística] se habla de una selección aleatoria de la muestra en la que cada elemento de la población tiene una probabilidad conocida no nula de ser seleccionado, con lo cual cada elemento de la muestra representa a un sector de la población y su totalidad a toda la población. La selección se puede hacer mediante un proceso mecánico similar al de una lotería, aunque es difícil manejar una lotería imparcial su equivalente práctico es la selección mediante las denominadas tablas de números aleatorios. En el muestreo no probabilístico las muestras no son aleatorias –de ahí que con frecuencia se diga que no son representativas– sino de tipo casual o fortuito [Pimienta, 2000:264].

En el ámbito de las ciencias sociales, el especialista debe contar con los conocimientos básicos de la teoría del muestreo, ya que ello le permitirá determinar el tipo de muestreo a aplicar (aleatorio simple, estratificado, por conglomerados, muestreo no probabilístico, etcétera), el tamaño que debe tener su muestra con base en el tamaño que tiene la población de estudio,[5] el proceso de selección de los elementos

[5] En este sentido, el tamaño de la muestra en muchas ocasiones implica que en las encuestas por muestreo realizadas en ciencias sociales, el investigador especifique el tamaño de la muestra en función de un porcentaje determinado del tamaño de la población para obtener un resultado confiable. Esto último es un procedimiento incorrecto, ya que la exactitud de las encuestas por muestreo no depende

que forman la muestra y el proceso de aplicación del instrumento seleccionado (cuestionario o entrevista) para la recolección de datos.

Con estos elementos, el especialista procede a desarrollar su plan de muestreo para recolectar los datos requeridos por el estudio. Finalmente, con la disponibilidad de estos datos se cuenta entonces con los elementos necesarios para realizar su proceso.

El proceso de datos

Una cuarta consideración consiste en establecer el procedimiento a emplear para procesar los datos. Actualmente, el uso de las microcomputadoras nos permite procesar de manera simple y rápida una gran variedad de datos. En el caso de utilizar un cuestionario como instrumento de recolección, nosotros sugerimos al especialista de las ciencias sociales, diseñarlo de tal forma que éste pueda ser procesado electrónicamente. En el mercado existe gran variedad de programas de computadora que pueden ser utilizados en esta etapa. Nuestra propuesta es utilizar la versión más reciente del paquete de análisis estadístico SPSS (Statistical Package for the Social Science) que exista en el mercado.

Análisis estadístico

Finalmente el análisis de los datos. En esta etapa deberá cuidarse con mucho detalle el análisis estadístico que se realice para cada variable y con ello la aprobación o rechazo de las hipótesis de investigación que se han propuesto. Esto último permitirá llegar a una serie de conclusiones sobre nuestro estudio que nos serán de gran utilidad.

Conclusión

Diseñar y desarrollar una encuesta no es tarea fácil, requiere de conocimiento adecuado en la materia, mente analítica y deseo de realizar bien la tarea. Es por ello que sugerimos desarrollar un adecuado trabajo de planificación de nuestra investigación, tomando como base la metodología que hemos propuesto. En ella, cada etapa deberá ser evaluada y rediseñada en su caso. Esto permitirá al especialista de las ciencias sociales contar

---

de un porcentaje de la totalidad de elementos que haya sido consultada, sino del número absoluto de éstos. Recordemos que "una muestra demasiado grande implica un despilfarro de recursos y una muy pequeña disminuye la utilidad de los resultados" (Cochran, 1998:104).

con una guía ordenada, misma que –al ser revisada detalladamente– permitirá determinar su aceptación o bien su corrección, de tal forma que cada una de estas fases puedan ser desarrolladas adecuadamente, lo que finalmente permitiría lograr los objetivos y metas establecidas para el estudio. En la aplicación de esta guía podremos recurrir a trabajos anteriores, a los especialistas en la materia o a experiencia previas que nos permitan diseñar un proceso de investigación nuevo y enriquecido. Todos estos elementos con los que contará el especialista en ciencias sociales deberán complementarse con sólidos conocimientos de muestreo, proceso de datos y análisis estadístico.

Bibliografía

Cochran G., William (1998), *Técnicas de muestreo*, CECSA, México.

Ferran A., Magdalena (2001), *SPSS para Windows. Análisis estadístico*, Osborne McGraw-Hill, España.

Foddy, William (1999), *Constructing Questions for Interviews and Questionaires*, Cambridge University Press, UK.

Fournier G., María de L. (2000), "Elaboración y análisis de encuestas de opinión", *Reflexiones finiseculares*, UAM-Xochimilco, México, pp. 107-138.

Oppenheim, A.N. (1997), *Questionnaire Design, Interviewing and Attitude Measurement*, Pinter Publishers, UK.

Pérez, César (2001), *Técnicas estadísticas con SPSS*, Prentice Hall Editores, España.

Pierdant R., Alberto (2000), *Estadística descriptiva con Excel 97*, Colección la Llave núm. 16, UAM-Xochimilco, México.

—— y Rodríguez F., Jesús, "Cuadros estadísticos. Ideas y consejos", en www.geocities.com/aipierdant/cuadros.

Pimienta L., Rodrigo (2000), "Encuestas probabilíticas *vs.* no probabilíticas", *Política y Cultura*, verano 2000, núm. 13, UAM-Xochimilco, México, pp. 263-276.

Posner M., Charles (2002), "Quantitative methods: The construction of questionnaires", Topic V, *Notes for the seminar of the programme of PhD in Education*, University of London/UAM-Xochimilco, México, pp 10.

Raj, Des (1980), *Teoría del muestreo*, FCE, México.

Robson, C. (1993), "Interviews and Questionnaires", *Real World Research*, Oxford, Blackwell, UK, cap. 9, pp. 227-267.

Sharon L., Lohr (2000), *Muestreo: diseño y análisis*, Thomson, México.

Wright, G. y Fowler, C. (1986), *Investigative Design and Statistics*, Penguin, Londres.

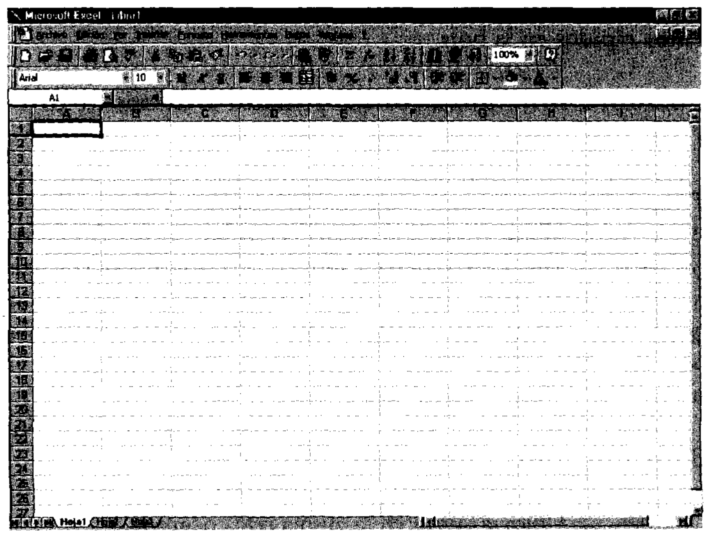
ANEXO II

Elementos básicos de la hoja de cálculo Excel

Introducción

Una hoja de cálculo es un programa de computadora que en la pantalla tiene la apariencia de una hoja cuadriculada; a cada hilera vertical de cuadros le llamaremos columna y a cada hilera horizontal renglón. Del lado izquierdo de la pantalla aparecerán los números de los renglones y en la parte superior los nombres de las columnas. Para ubicarnos en esta hoja de trabajo, a cada cuadro le denominaremos celda y su ubicación llevará el nombre de la columna y el número del renglón, es decir, si estamos en la columna A y el renglón 1, entonces el cruce de éstos nos ubica dentro de la hoja en la celda A1, como vemos en la Figura 1.

FIGURA 1



En la pantalla aparece siempre marcada una celda con un borde en negro y un pequeño cuadro en su parte inferior derecha al que denominaremos el "cursor" de EXCEL. Al ubicar el cursor en una celda particular, la barra de edición indica la columna y el renglón de ubicación, así como su contenido.

Cada celda puede contener un dato en forma de un número, un mensaje, una fórmula, o bien una función. El dato podrá ser agregado a la hoja tecleando en la celda la información deseada y oprimiendo al final la tecla Intro (ENTER ↵) o la tecla de navegación deseada (flechas a la derecha →, izquierda ←, arriba ↑ o abajo ↓).

Esta hoja electrónica tiene diversos usos: podemos hacer desde cálculos simples –como una suma– hasta cálculos complejos –como estadísticas, cálculo financiero, cálculos de ingeniería, etcétera. A las celdas que tienen la información que necesitamos y que se van a utilizar dentro de las funciones o fórmulas les llamaremos referencias, y a la posición del cursor dentro de la hoja, celda activa; por ejemplo, si tenemos un valor numérico en la celda A1, otro en la celda B1, y queremos el resultado de la suma de los valores de estas celdas en la celda C1, entonces, para realizar la suma, la celda activa será C1 y en ésta se teclea la información siguiente: = A1+B1 y al final se presiona la tecla ENTER, con lo que obtenemos el resultado deseado de la operación.

Como se puede observar en la Figura 1, toda celda ocupa una ubicación única en la hoja electrónica, por ejemplo: F5, A33, etcétera; es decir, no existen en la hoja celdas duplicadas. Dicha celda, como ya indicamos, puede contener un mensaje o un letrero, una cantidad numérica o bien una fórmula. A los mensajes o letreros les llamaremos etiquetas; a las cantidades numéricas, valores numéricos; algunas funciones especiales pueden, incluso, tener valores lógicos.

Movimientos rápidos en la hoja

Excel cuenta con diversos comandos que permiten al usuario ubicarse en diversas partes de la hoja en forma rápida, entre ellos tenemos:

TECLA	FUNCIÓN
HOME	Ubica el cursor al inicio del renglón (en la columna A).
END y (→)	Ubica el cursor en la última celda ocupada en el renglón.
END y (↓)	Ubica el cursor en la última celda ocupada en la columna.
END y (↑)	Ubica el cursor en la primera celda ocupada en la columna.
PgUp	Ubica el cursor una página anterior.
PgDn	Ubica el cursor una página posterior.
F5	Ubica el cursor en una celda especificada (debemos indicarla).
Clic del ratón	Ubica el cursor en la celda indicada.
CTRL y HOME	Ubica el cursor al inicio de la hoja (celda A1).
CTRL y END	Ubica el cursor en la última celda ocupada.

Dentro del área de trabajo, en la Figura 1, observamos del lado derecho una barra sombreada que tiene dos flechas, una en la parte inferior y otra en la superior; al final de la figura también observamos otra barra similar en forma horizontal, a estas barras les llamaremos barras de desplazamiento. Éstas nos permiten ubicarnos en las diversas áreas de una hoja electrónica. Si tenemos ratón o *mouse* (🖱) esto lo lograremos con los siguientes pasos: colocamos el cursor del *mouse* sobre el cuadro que se encuentra en las barras y oprimir el botón derecho del *mouse* sin soltarlo y se desliza el cursor sobre la barra, luego soltamos el botón y automáticamente estamos ubicados en otra sección de la hoja. Momentáneamente aparecerá en la pantalla un reloj (⌛) que indica que Excel está trabajando. En múltiples ocasiones el reloj se aparece en la pantalla lo que indica que esperemos unos segundos para que podamos seguir tecleando.

Principales componentes de la ventana de Excel

En la hoja electrónica Excel se muestran seis secciones o áreas de la ventana. Éstas son:

- Barra de la aplicación (MicroSoft Excel)
- Barra de menú de comandos
- Barras de herramientas
- Barra de fórmula
- Hoja electrónica o área de trabajo
- Barra de estado

En la Figura 2 se muestran cada uno de estos componentes.

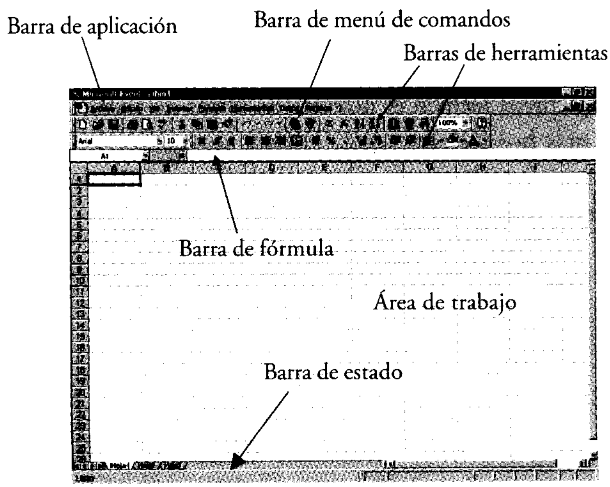


FIGURA 2

1. *Barra de la aplicación.* Es la barra donde se muestra el nombre de la aplicación, el nombre del libro que se está utilizando y los botones de control de la ventana: minimizar, maximizar y cerrar.

2. *Barra del menú de comandos.* Muestra los comandos del menú principal de EXCEL: ARCHIVO, EDICIÓN, VER, INSERTAR, FORMATO, HERRAMIENTAS, DATOS, VENTANA y ? (ayuda o asistente).

3. *Barras de herramientas.* Son conjuntos organizados de instrucciones que han sido asignados a pequeños dibujos o iconos. Para ejecutar una instrucción nos ubicamos con el ratón o *mouse* en alguno de estos pequeños dibujos y oprimimos el botón izquierdo, con lo cual se ejecutará la instrucción de inmediato, es decir, son comandos rápidos representados por medio de botones, y los dibujos indican de qué tipo es la instrucción asociada a ellos.

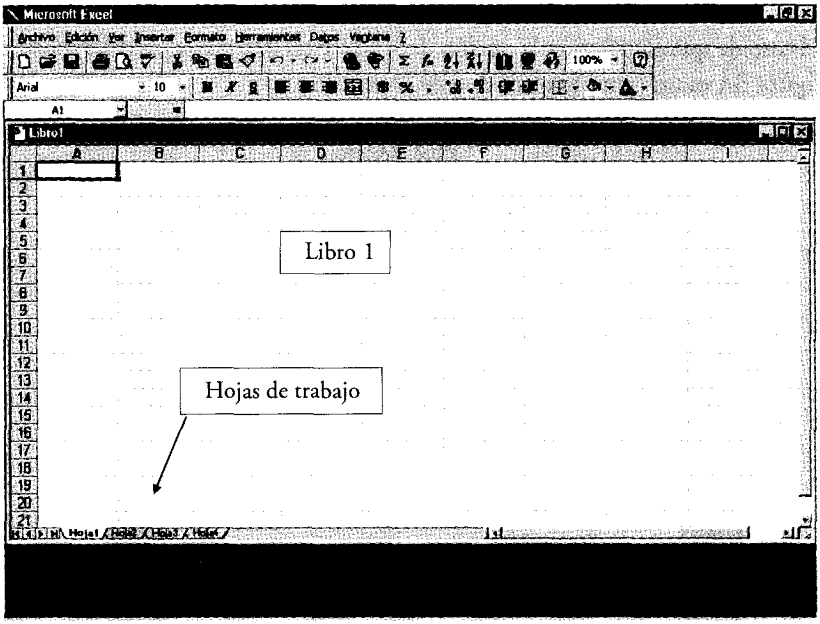
4. *Barra de fórmula.* Muestra el contenido de la celda y su ubicación. Nos indica si en la celda existe un texto, un número, una fórmula, una función o bien una combinación de estas características.

5. *Área de trabajo.* Está formada por un cuadriculado en el que las columnas se denotan por letras y los renglones por números. El cruce de columnas y renglones forman las celdas de trabajo de la hoja electrónica.

6. *Barra de estado.* Nos indica el tipo de información, si está lista para trabajar, si se está editando una celda, etcétera. Indica lo que se hace en la hoja activa y las teclas que están activadas (CAPS, NUM, INS, DESP, etcétera), así como la situación de la celda en la hoja.

En Excel los archivos reciben el nombre de libros de trabajo. Cada uno puede contener una o más hojas electrónicas. Éstas a su vez pueden tener un nombre particular, según sea asignado por el usuario. En la Figura 3 se muestra un libro 1 con cuatro hojas de trabajo (hoja 1 a hoja 4).

FIGURA 3



Submenús de Excel

A continuación se describen brevemente los principales submenús de Excel, así como sus principales comandos asociados. Éstos pueden ser aplicados indepen-dientemente en cada hoja de trabajo que tenga un libro.

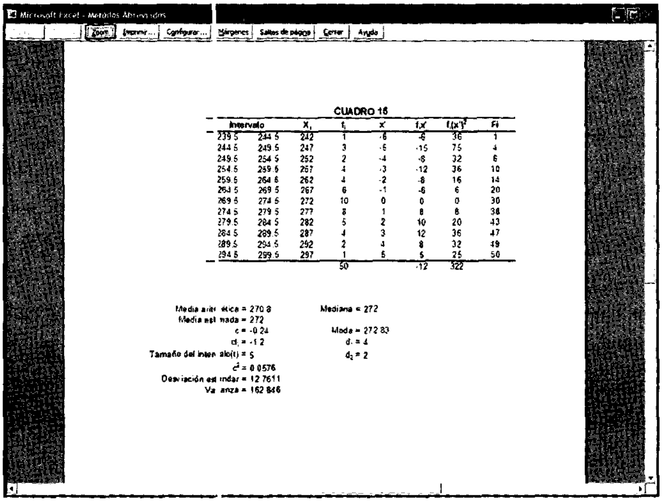
1. ARCHIVO. Permite el manejo de los archivos en la hoja electrónica. Sus principales subcomandos son:

NUEVO	Crea un nuevo libro de trabajo (archivo).
ABRIR	Abre un libro de trabajo ya existente.
CERRAR	Cierra el libro de trabajo activo.
GUARDAR	Guarda el libro con el mismo nombre, es decir, lo sobreescribe.
GUARDAR COMO	Guarda el libro en uso por primera vez, o guardar las modificaciones con otro nombre.
GUARDAR ÁREA DE TRABAJO	Guarda el grupo de libros definidos como grupo de trabajo.
VISTA PRELIMINAR	Reproduce en la pantalla la hoja electrónica como si estuviera impresa.
IMPRIMIR	Permite indicar las características de impresión de la hoja, márgenes, tamaño, etcétera.
SALIR	Para salir de la aplicación

Ejemplo:

En el caso de que eligiésemos la opción de VISTA PRELIMINAR la pantalla mostrada en la Figura 3A es un ejemplo claro de este comando.

FIGURA 3A



2. EDICIÓN. Permite realizar operaciones de trabajo con los datos que se encuentran en las celdas de la hoja electrónica.

DESHACER	Anula la última acción ejecutada y la celda(s) así como la hoja queda sin cambio.
CORTAR	"Extrae" el rango de celdas seleccionado de su posición actual, lo coloca dentro del portapapeles, para después trasladarlo a otra sección de la hoja activa u otra hoja mediante el comando PEGAR.
COPIAR	"Registra" en el portapapeles el rango de celdas seleccionado para posteriormente copiarlo a la sección deseada de la hoja mediante el comando PEGAR o al oprimir la tecla ENTER.
PEGAR	Coloca la información que se encuentra en el portapapeles a partir de la celda activa.
BORRAR	Elimina columnas, renglones o celdas seleccionadas.
PEGADO ESPECIAL	Al copiar las celdas entre hojas de cálculo, establece una liga entre las distintas hojas. La hoja que recibe las celdas es dependiente y la que proporciona los datos es la independiente.

3. INSERTAR. Este menú maneja la creación de las fórmulas o secuencias de valores, referencias, nombres, funciones u operadores, y crea un nuevo valor a partir de los datos existentes. Permite dar nombres a rangos, localización de celdas distintas, localización de textos y reemplazo de textos.

   Definiremos primero los tipos de operadores que maneja Excel:

- *Aritméticos.* Realizan las operaciones matemáticas más comunes.

+	SUMA
-	RESTA
*	MULTIPLICACIÓN
/	DIVISIÓN
%	PORCENTAJE
^	EXPONENCIACIÓN

- *De texto.* El único operador es el & (*ampersand*) que nos permite unir varios textos o unir valores alfanuméricos con textos.

- *Lógicos.* Compara dos valores y al resultado le asigna un valor lógico, es decir, verdadero o falso.

=	IGUAL
<	MENOR QUE
>	MAYOR QUE
>=	MAYOR O IGUAL QUE
<=	MENOR O IGUAL QUE
<>	DISTINTO A

- *De rango.* Une las referencias, ya que toma en cuenta el rango desde la primera referencia hasta la segunda (A1:B3 significa que toma en cuenta la información que tiene la hoja desde la celda A1 hasta la celda B3).
  Une referencias aisladas (A1, A7, B9 significa que toma en cuenta la información de la celda A1, la de la celda A7 y la de la celda B9).

Insertar

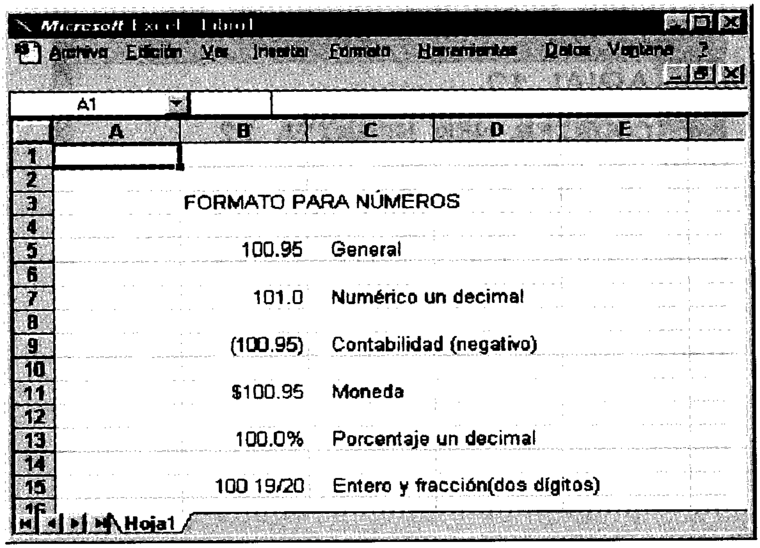
CELDAS	Agrega celdas a la hoja electrónica, desplazando a otras celdas los datos o información.
FILAS	Agrega uno o varios renglones a la hoja electrónica, desplazando a otras filas los datos o información.
COLUMNAS	Agrega una o varias columnas a la hoja electrónica, desplazando a otras columnas los datos o información.
FUNCIÓN	Introduce una función propia de Excel a la celda activa. Presenta la sintaxis requerida para su correcta operación.
NOMBRE	Crea, elimina o cambia nombres a un rango de celdas seleccionado.
NOTAS	Agrega, borra o edita notas en las celdas de la hoja.
HOJA DE CÁLCULO	Agrega una nueva hoja de cálculo al libro de trabajo.

4. FORMATO. Estos comandos permiten asignar a los datos o información en una celda o en un conjunto de celdas un formato determinado, es decir, determina tamaño de la celda (ancho y altura); el tipo, tamaño y el color de la letra; el color del fondo de la celda, etcétera.

CELDA	Asigna características al dato que está en la celda o celdas seleccionadas. Por ejemplo, tipo de letra, tamaño, alineación de la información en la celda, color, formato numérico, etcétera. Activa y desactiva la protección de las celdas.
FILA	Permite modificar tanto la altura, como el ancho de la fila, también restablece la altura de la fila al alto estándar.
COLUMNA	Ajusta el ancho de las columnas seleccionadas.
HOJA	Cambia nombre y oculta hojas seleccionadas.
AUTOFORMATO	Da formato de tabla a un rango de datos.
MODELO	Aplica o define un modelo de celda.
OBJETO	Pasa los objetos seleccionados al primer plano.

Tenemos diversos formatos para la presentación de los datos numéricos al seleccionar la opción NÚMERO del menú FORMATO/CELDA. A continuación se presentan unos ejemplos del uso de este comando en la Figura 4.

FIGURA 4



Otro ejemplo de formato lo podemos observar en la Figura 5, la columna "A" es más ancha que el resto de las columnas, esto se logra utilizando el menú ANCHO DE COLUMNA, para seleccionar el ancho adecuado a nuestra necesidad.

En forma análoga tenemos varias opciones de alineamiento para los datos ubicados en las celdas de la hoja electrónica [centrado (A4), izquierda (A3), derecha (A5)].

FIGURA 5

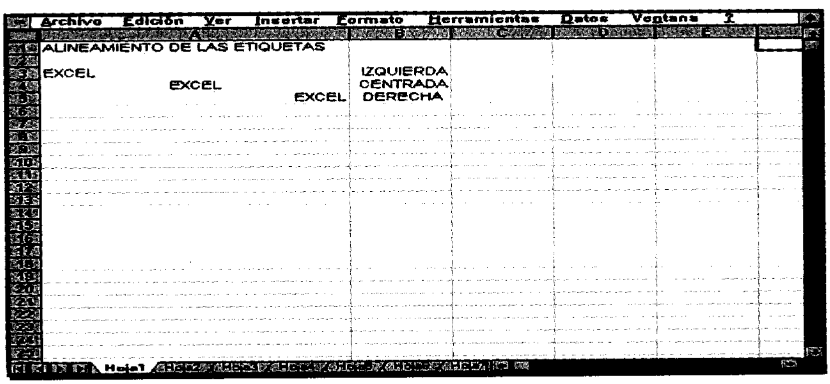
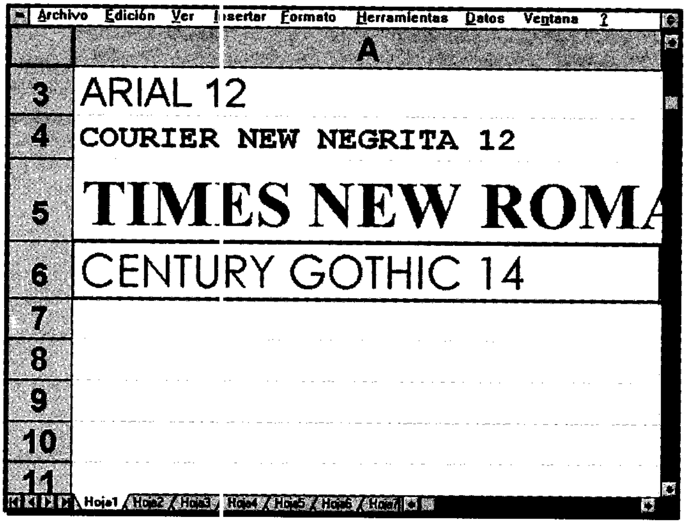


FIGURA 6



En el ejemplo de la Figura 6 vemos que los renglones tienen diversas alturas, esto se debe a que los tipos de letra presentados los modifican de acuerdo con su tamaño.

En el caso del comando Bordes nos crea un marco para los datos. El procedimiento es simple, primero se selecciona el conjunto de celdas a formatear (rango) y después selecciona el tipo de borde para éste.

Para "acomodar" alguna etiqueta en una columna, primero introducimos la etiqueta en una celda, segundo, seleccionamos el número de celdas hacia abajo en donde queremos que nos acomode la etiqueta, tercero, elegimos el comando justificar. En esta forma la etiqueta quedará ordenada en una sola columna. Cabe mencionar que esto dependerá del tamaño que se le asigne a la columna.

5. ? (AYUDA). Es un menú que permite consultar en forma rápida mediante ejemplos el uso de comandos y procedimientos de operación del paquete.

AYUDA DE MS EXCEL	Despliega el contenido de Excel, para consultar temas.
CONTENIDO E ÍNDICE	Muestra una introducción completa al Excel. Despliega lecciones rápidas para que el usuario se familiarice con Excel.
AYUDA LOTUS 123	Muestra las equivalencias en los comandos con Excel.
ACERCA DE MS EXCEL	Despliega información sobre la versión de Excel, memoria total y memoria disponible.

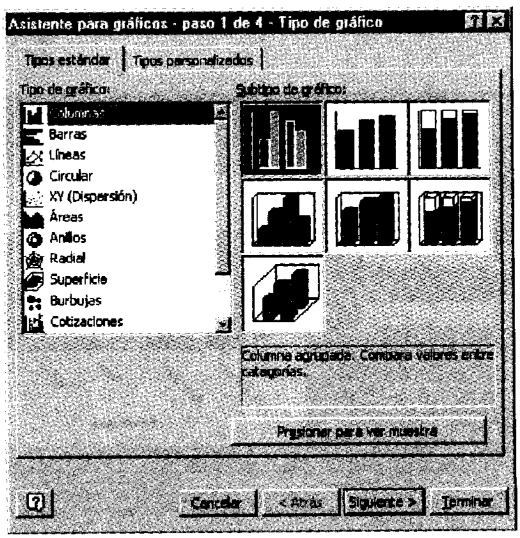
Gráficas en Excel

Excel cuenta con una gran variedad de gráficos, entre éstos se tienen:

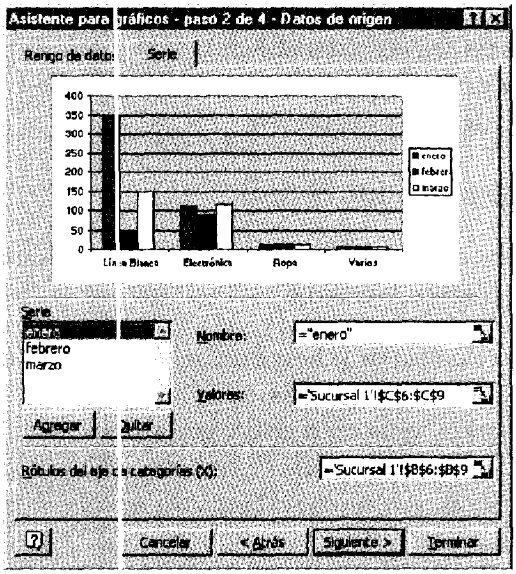
- Área (en dos y tres planos)
- Barras-Columnas (en dos y tres planos)
- Líneas (en dos y tres planos)
- De sectores-pie (en dos y tres planos)
- De tipo XY
- Combinadas (barras y líneas, etcétera)

En Excel, la construcción de una gráfica puede simplificarse en los siguientes pasos, si usamos el asistente de gráficos:

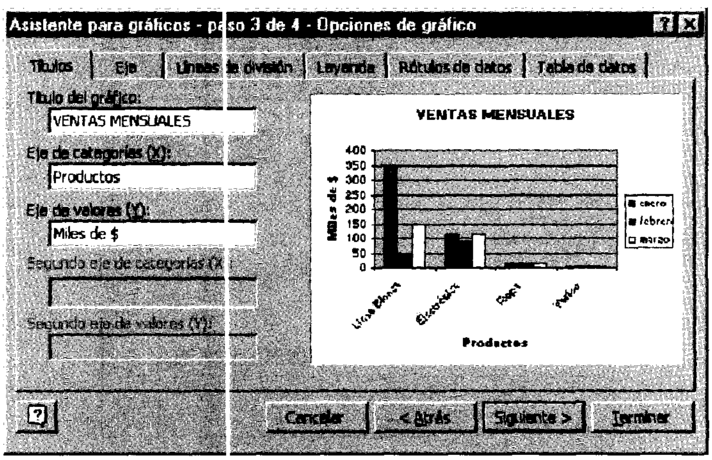
1. Se utilizan dos columnas de la hoja electrónica, la primera deberá contener los datos del eje "X" y la segunda los correspondientes al eje "Y", o bien, se pueden utilizar más columnas según sea el gráfico que se desea construir.
2. Se selecciona el conjunto de celdas de estas dos columnas (rango con la información de la gráfica), y se oprime el botón del *asistente de gráficas* que permite crear gráficos.
3. Aparece el paso 1 del asistente de gráficas. En él aparece seleccionada la opción estándar, es decir, gráfica de columnas, subtipo de gráfica 1. En este paso podemos seleccionar otro tipo de gráfica, y a su vez un subtipo específico. Oprimimos el botón "siguiente" para pasar al segundo paso del asistente.



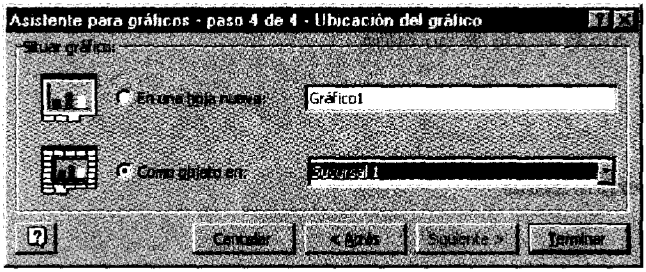
4. En este segundo paso deberá seleccionarse la opción de "filas" o "columnas" según se desee mostrar los datos. Posteriormente en la categoría de series, se definen las series a incluir en la gráfica y sus etiquetas. Oprimir "siguiente" para pasar al paso 3.



5. En el paso 3 del asistente, deberemos especificar el título de la gráfica, la etiqueta del eje "X" y la del eje de las "Y". También se pueden definir aquí otras características de la gráfica como: poner o quitar líneas de división, leyendas a cada serie en la gráfica, rótulos (valores) a los datos y una tabla de datos. Posteriormente oprimir, "siguiente".



6. En el paso 4 (último) del asistente, deberemos indicar si queremos que el gráfico se agregue como un objeto en la misma hoja electrónica o bien se agregue en otra hoja nueva. En este momento oprimimos "terminar" y con ello obtenemos una gráfica.



Al finalizar una gráfica, Excel permite cambiar su tamaño dentro de la hoja electrónica, para ello deberá seleccionarse la gráfica (dando un clic sobre ella), y con ayuda del *mouse* (oprimiendo el botón derecho sin soltar) arrastrar sus extremos para cambiar de tamaño. También es posible cambiarla de ubicación, para esto seleccionamos la gráfica, ubicamos el cursor del ratón sobre ella, oprimimos el botón derecho sin dejar de oprimir, con lo que aparece un cursor con cuatro flechas, en ese momento arrastramos la gráfica a su nueva ubicación.

Excel permite trabajar 17 tipos de gráficas diferentes de las cuales 10 se trabajan en 2 ejes y 7 son en tres dimensiones (3-D).

Modificar características de una gráfica. Para agregar y modificar diversas características de su gráfica, primero seleccione la gráfica, posteriormente ubique el cursor en la sección de la gráfica a modificar y oprima dos veces el botón para que aparezca la ventana de características asociada que permite hacer las modificaciones deseadas.

Borrar una gráfica. Para borrar una gráfica deberemos primero seleccionarla y posteriormente oprimir la tecla suprimir o delete: "Supr" o "Del".

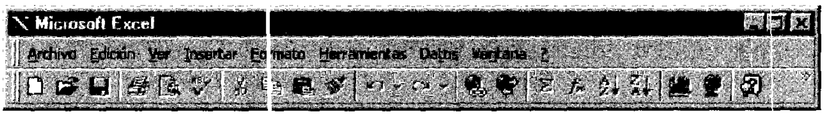
Elaborar gráfica sin ayuda del ratón. Para elaborar una gráfica a partir de comandos sin utilizar el *mouse* se deberá seguir el proceso que se describe a continuación:

- Lo primero es seleccionar el conjunto de datos (rango de datos) a graficar.
- En el menú INSERTAR se selecciona GRÁFICO. Aparece la ventana del paso 1 del asistente de gráficas.
- Con ayuda de las teclas de navegación se selecciona el tipo de gráfica deseada. Se usa la tecla del "Tabulador" para pasar a las diferentes opciones. A partir de aquí pueden aplicarse los pasos ya indicados para la construcción de un gráfico cuando se cuenta con la ayuda de un ratón.

Barras de herramientas

Dentro del Menú VER, <BARRAS DE HERRAMIENTAS> puede indicar a Excel cuáles barras de herramientas desea visualizar el usuario en la pantalla. Como ejemplo de éstas, se muestran a continuación las barras siguientes: estándar, para gráficas y de formato. Para cada una de ellas se indicará alguna función asociada a cada icono.

BARRA ESTÁNDAR



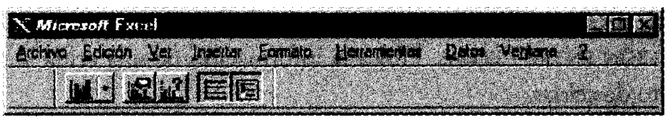
1 2 3 4 5 6 7 8 9 10 11 a b c d e f g

1. NUEVO. Crea un nuevo libro de trabajo.
2. ABRIR. Despliega una caja de diálogo para abrir libros existentes.
3. GUARDAR. Guarda cambios hechos al libro activo.
4. IMPRIMIR. Imprime la hoja o el área de la hoja electrónica seleccionada.
5. VISTA PRELIMINAR. Muestra en pantalla como queda la impresión de la hoja electrónica.
6. VERIFICACIÓN ORTOGRÁFICA. Verifica ortográficamente los textos en las celdas de la hoja electrónica.

7. CORTAR. Corta el contenido de las celdas seleccionadas, gráfica u objeto al portapapeles.
8. COPIAR. Copia el contenido de las celdas seleccionadas, o bien una gráfica u objeto al portapapeles.
9. PEGAR. Pega en la celda indicada el contenido del portapapeles.
10. COPIAR FORMATO. Copia el formato de la celda o celdas seleccionadas.
11. DESHACER. Elimina la última acción o comando ejecutado.

a) AUTOSUMA. Suma en la celda seleccionada el contenido de las celdas indicadas.
b) PEGAR FUNCIÓN. Agrega a la celda seleccionada una función de Excel.
c) ORDENAR ASCENDENTE. Ordena el contenido de las celdas seleccionadas en forma ascendente.
d) ORDENAR DESCENDENTE. Ordena el contenido de las celdas seleccionadas en forma descendente.
e) ASISTENTE PARA GRÁFICOS. Llama al asistente para poder elaborar una gráfica.
f) MAPA. Llama al asistente de mapas. Permite copiar mapas a hojas electrónicas.
g) AYUDA OFFICE. Llama al asistente de ayuda de office.

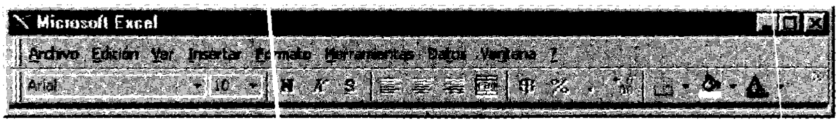
HERRAMIENTAS PARA GRÁFICOS



1 2 3 4 5

1. TIPO DE GRÁFICA. Despliega un apuntador especial que puede usar para seleccionar una gráfica en particular.
2. TIPO DE GRÁFICA PREFERIDA. Selecciona la gráfica preferida del usuario.
3. ASISTENTE DE GRÁFICAS. Guía al usuario para la elaboración de una gráfica.
4. LÍNEAS HORIZONTALES. Agrega o elimina líneas horizontales a la gráfica.
5. ETIQUETAS. Coloca las etiquetas de las series usadas en la gráfica.

HERRAMIENTAS DE FORMATO



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

1. COMANDO FUENTES. Enumera las fuentes (*fonts*) o tipos de letra disponibles. Escoja el tipo de letra que desee aplicar a la selección.
2. COMANDO TAMAÑO DE FUENTE. Lista los tamaños disponibles para la fuente mostrada en la caja de nombre de *Font.* Escoja el tamaño que quiera.
3. NEGRITAS. Pone negritas en la celda o celdas seleccionadas, caja de texto o texto en gráficas.
4. ITÁLICA. Pone formato itálico en la celda o celdas seleccionadas, en la caja de texto seleccionada, botón, o texto en gráfica seleccionado.
5. SUBRAYADO. Subraya el texto en las celdas seleccionadas, o el texto seleccionado en una gráfica, caja de texto, o botón.
6. ALINEACIÓN IZQUIERDA. Ubica la información de la celda o celdas al lado izquierdo de la misma(s).
7. CENTRADO. Ubica la información de la celda o celdas al centro de la misma(s).
8. ALINEACIÓN DERECHA. Ubica la información de la celda o celdas al lado derecho de la misma(s).
9. ALINEACIÓN ENTRE CELDAS. Ubica la información de una celda al centro de un conjunto de celdas.
10. FORMATO MONEDA. Agrega o elimina el formato de $ a las celdas seleccionadas.
11. FORMATO PORCENTAJE. Agrega o elimina el formato de % a las celdas seleccionadas.
12. FORMATO MILLARES. Agrega o elimina el formato de miles (000) a las celdas seleccionadas.
13. AUMENTAR DECIMALES. Agrega decimales a las celdas seleccionadas.
14. PALETA DE BORDES. Agrega o elimina diferentes tipos de bordes a las celdas seleccionadas.
15. PALETA DE COLOR DE RELLENO. Agrega o elimina un color al fondo de las celdas seleccionadas.
16. PALETA COLOR DE FUENTE. Agrega o elimina un color a la fuente de las celdas seleccionadas.

Bibliografía

Anderson, D.; Sweeney, D. y Williams, T. *Estadística para administración y economía*", Thomson, México, 2004.

Chao L., Lincon. *Estadística para las ciencias administrativas*, Mc Graw Hill, México, 1993.

Berenson, M., Levine, D. y Krehbiel, T. *Estadística para administración*", Pearson Educación, México, 2001.

Budnick S., Frank. *Matemáticas aplicadas para administración, economía y ciencias sociales*, Mc Graw Hill, México, 1990.

Downie, N.M. y Heath, R.W. *Métodos estadísticos aplicados*, HARLA, México, 1973.

Garrett E. Henry. *Estadística en psicología y educación*", Paidós, Buenos Aires, 1970.

González Martínez, María D. y Oliver D., Jackson, *Introducción a la teoría de gráficas en educación*, ANUIES, México, 1979.

Grolier Electronic Publishing, INC. *The 1995 Grolier Multimedia Encyclopedia*, CD-Rom, Grolier Electronic Publishing, INC., USA, 1995 y 1997.

Haussler F., Ernest y Richard S., Paul, *Matemáticas para administración y economía*, Iberoamérica, México, 1992.

Hernández, S. R.; Fernández, C.C. y Baptista L. P. *Metodología de la investigación*, Mc Graw Hill, México, 1991.

Holguín Quiñones, Fernando. *Estadística Descriptiva Aplicada a las Ciencias Sociales*, UNAM, México, 1981.

Levin, I. R.; Rubin, S. D.; Balderas, M.; Del Valle, J. y Gómez, R. *Estadística para administración y economía*, Pearson Educación, México, 2004.

Lopes A., Paulo. *Probabilidad y estadística*, Pearson Educación, Colombia, 2000.

Mendenhall, William. *Estadística para administradores*, Iberoamérica, México, 1990.

Meza L., Carlos; Morales A., Andrés y Magaña C., Rogelio. *Introducción al método estadístico*, UAM-Xochimilco, México, 1980.

Microsoft Corporation. *Microsoft Excel User's Guide, Version 5.0*, Microsoft Corporation, USA, 1994.

——. *Microsoft Excel Function Reference, Version 4.0*, Microsoft Corporation, USA,1992.

——. *Obtenga resultados con Microsoft Office para Windows 95*, Microsoft Corporation, USA, 1995.

——. *Microsoft Excel 7, User's Manual*, Microsoft Corporation, USA, 1996.

——. *Microsoft Office 2003*, Microsoft Corporation, USA, 2003.

Nuñez del Prado B., Arturo. *Estadística básica para planificación*, Siglo XXI, México, 1987.

Pierdant R., Alberto I. *Estadística descriptiva con Excel 97*, UAM, México, 2000.

—— y Rodríguez F., Jesús. *Reflexiones Finiseculares. Las matemáticas en las ciencias sociales*, capítulo "Aplicaciones matemáticas con Excel", UAM, México, 2001.
Sanzo, Richard. *Ratio Analisis for Small Business*, Small Business Administration, USA, 1970.
Stevenson J., William. *Estadíatica para administración y economía*, HARLA, México, 1981.
Stokes J., Charles. *Economics for Managers*, Mc Graw Hill, USA, 1979.
VOX, *Diccionario enciclopédico*, núm. 22, "Matemáticas", Bibliograf, Barcelona, 1981.
Webster L., Allen. *Estadística aplicada a los negocios y la economía*, McGraw-Hill, Colombia, 2000.

Hemerografía

Banamex-Accival, revista *Examen de la situación económica de México*, vol. LXIX, núm. 817, diciembre, 1993, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXX, núm. 825, agosto, 1994, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXXI, núm. 836, julio, 1995, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXIX, núm. 817, diciembre, 1993, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXXI, núm. 838, septiembre, 1995, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXXI, núm. 839, octubre, 1995, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXXIV, núm. 874, septiembre, 1998, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXXV, núm. 881, mayo, 1999, Banamex, México.
——. Revista *Examen de la situación económica de México*, vol. LXXV, núm. 882, junio, 1999, Banamex, México.
Banco Mundial. *La página del Banco Mundial*, sección "Reporte Anual 1995", en www.wordbank.org/html/extdr/annrept.html.
——. *La página del Banco Mundial*, sección "Reporte Anual 1996", en www.wordbank.org/html/extdr/annrept.html.
——. *La página del Banco Mundial*, sección "Reporte Anual 1997", en www.wordbank.org/html/extdr/annrept.html.
Bancomext. *Mexico Today*, Banco de Comercio Exterior, México, 1992.
Kravzov Jinich, Jaime. *Informe de Actividades 1994-1995*, informe del rector, UAM-Xochimilco, México, 1995.
Pierdant R., Alberto I. "Modelos económicos y de planificación utilizando hojas electrónicas de cálculo", revista *Política y Cultura*, DCSH, núm. 4, UAM-Xochimilco, México, 1995.

——. "Las hojas electrónicas de cálculo, una herramienta para la enseñanza de las matemáticas", ponencia, IX Congreso Departamental de Investigación, Departamento de Producción Económica, DCSH, UAM-Xochimilco, Oaxtepec, México, 1995.

——. "Una base de datos en Excel", *Boletín Informativo*, núm. 2, Departamento de Producción Económica, DCSH, UAM-Xochimilco, México, 1996.

—— "Solución de modelos de optimización lineal con Excel", *Emprendedores*, núm. 73, Facultad de Contaduría y Administración, UNAM, México, enero-febrero, 2002.

Elementos básicos de estadística para ciencias sociales se terminó de imprimir en diciembre de dos mil seis. La edición consta de mil ejemplares; en su formación se utilizaron tipos de la familia AGaramond, 11/13.

Váksu. Entrepalabras, editores
Tel. 5594 9341 Cel. 04455 1569 8399
vaksu_entrepalabras@yahoo.com.mx

Este libro proporciona al especialista e investigador en el área de las ciencias sociales conceptos básicos de estadística que le permitirán realizar un primer análisis sobre sus trabajos o investigaciones. Integra temas de aritmética aplicada, construcción de cuadros estadísticos, medidas de tendencia central, dispersión, sesgo y curtosis, métodos abreviados de cálculo de los estadísticos descriptivos, así como ejemplos de aplicaciones mediante el uso de hojas electrónicas de cálculo.

$$S = \sqrt{\frac{\sum_{i=1}^{n}[x_i - \bar{x}]^2}{n-1}}$$

ISBN 970310731-1

9709783107319

Publicaciones