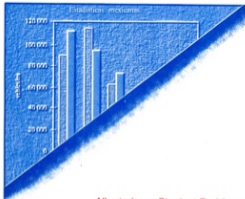




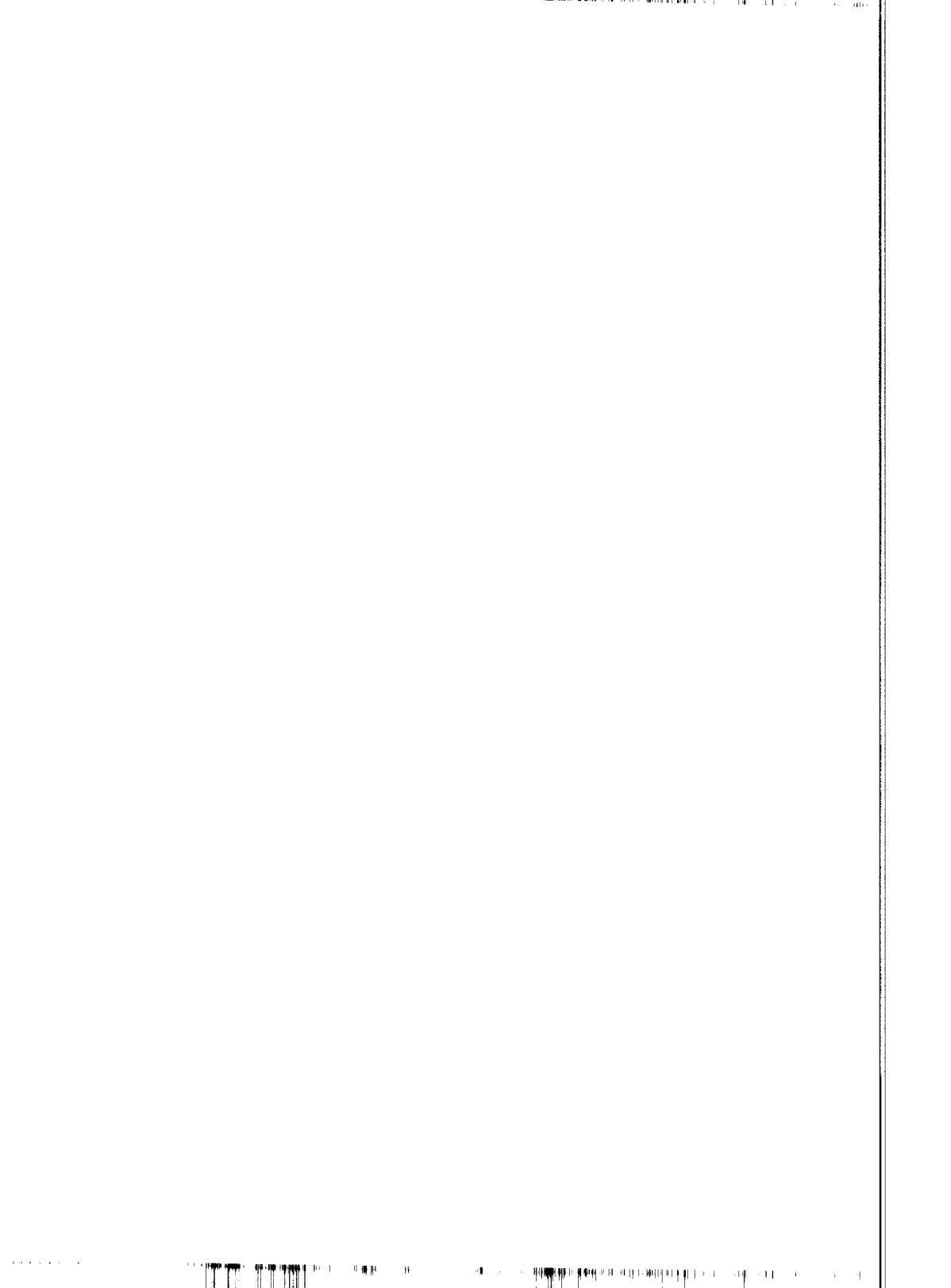
ESTADÍSTICA DESCRIPTIVA USANDO EXCEL

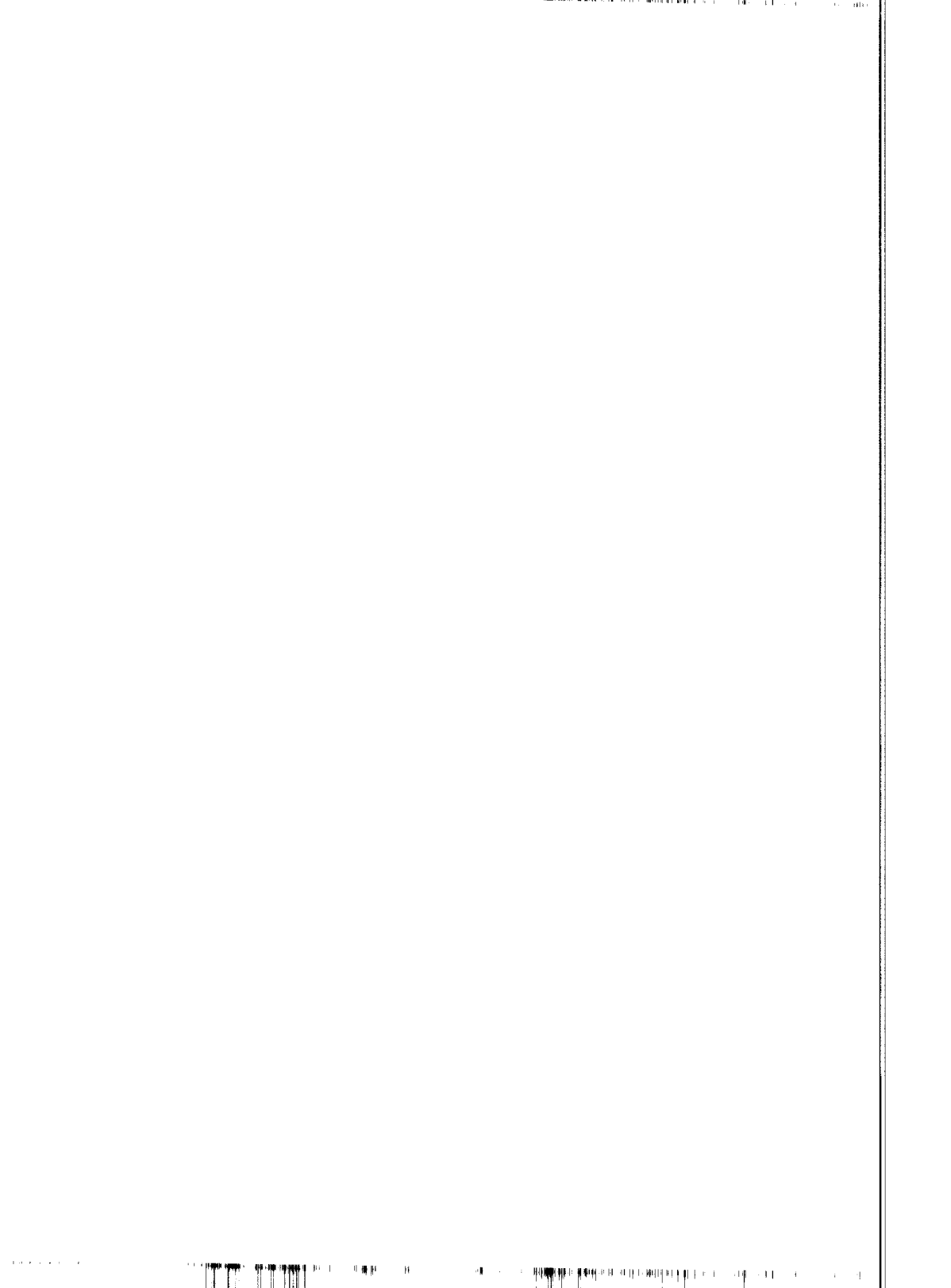


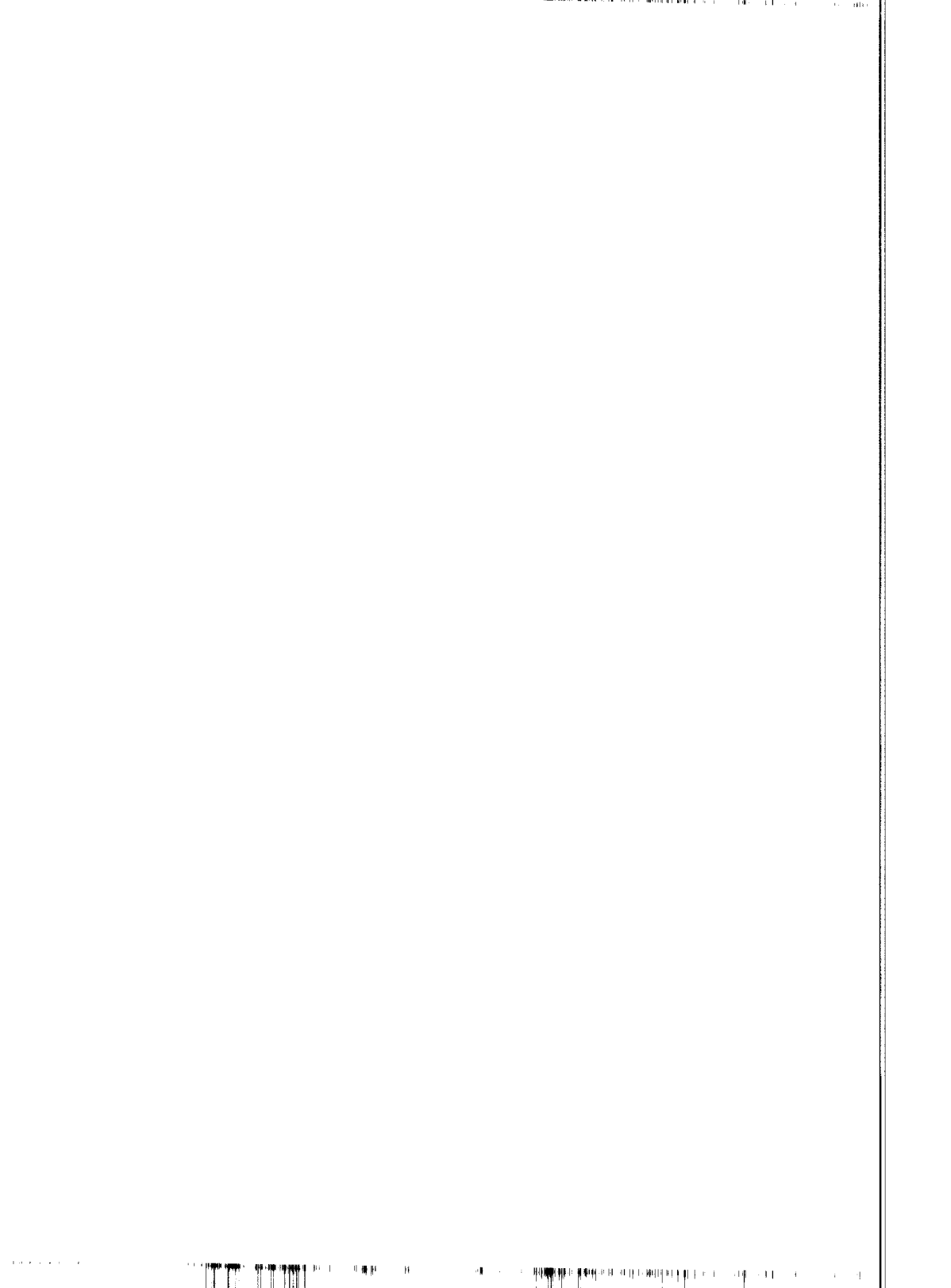
Alberto Isaac Pierdant Rodríguez



UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD XICOMILCO División de Ciencias Sociales y Humanidades



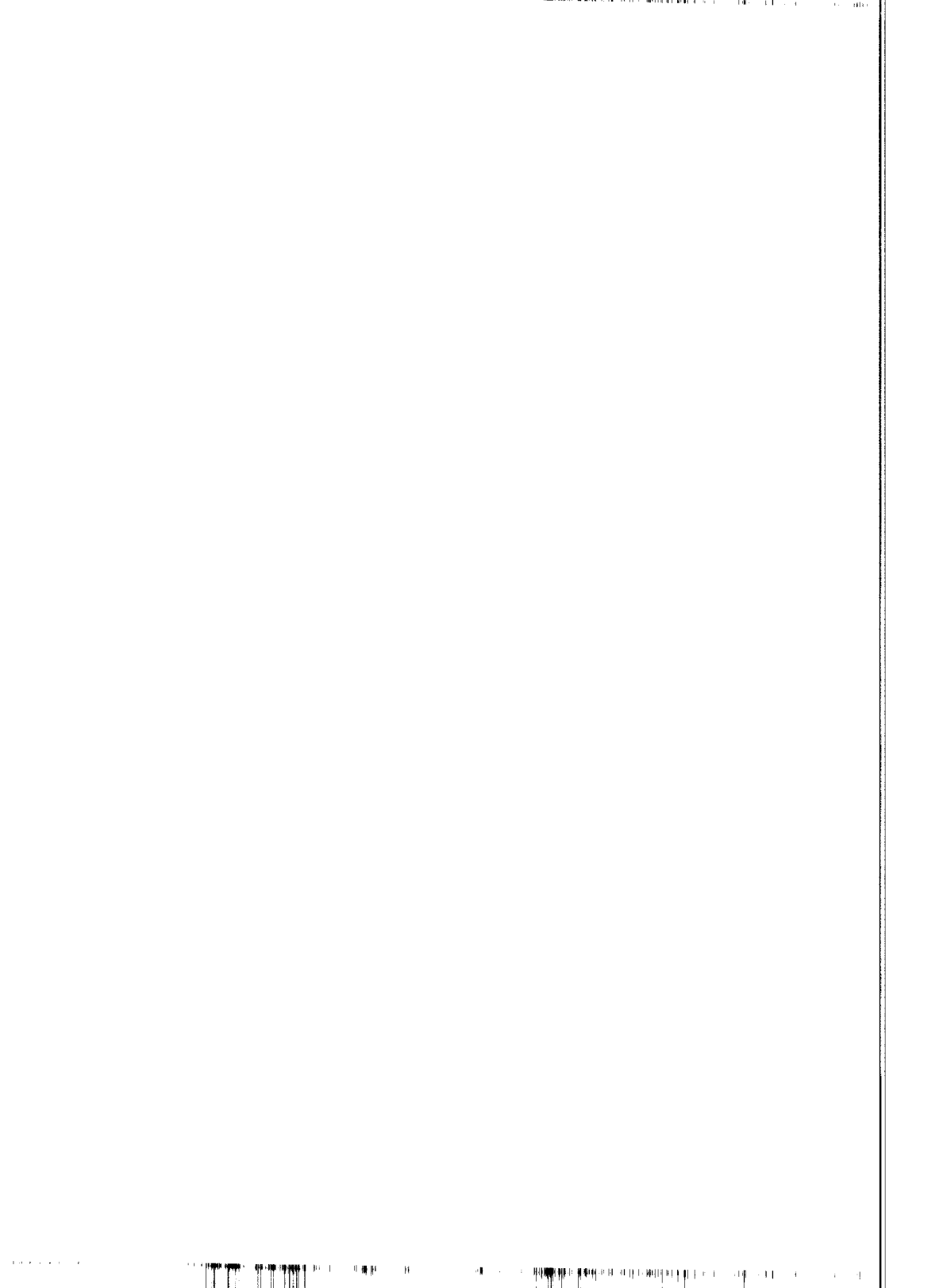






Casa abierta a tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
UNIDAD XOCHIMILCO División de Ciencias Sociales y Humanidades



ESTADÍSTICA DESCRIPTIVA
USANDO EXCEL

ALBERTO ISAAC PIERDANT RODRÍGUEZ

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Rector General, doctor Julio Rubio Oca

Secretaria General, maestra Magdalena Fresán Orozco

UNIVERSIDAD AUTÓNOMA METROPOLITANA-XOCHIMILCO

Rector, químico Jaime Kravzov Jinich

Secretaria de la Unidad, maestra Marina Altagracia Martínez

DIVISIÓN DE CIENCIAS SOCIALES Y HUMANIDADES

Director, doctor Guillermo Villaseñor García

Secretario Académico, licenciado Gerardo Zamora Fernández de Lara

Responsable de Publicaciones, licenciada Araceli Soní Soto

Edición: Salvador González Vilchis

Asistencia editorial: Águeda Arroyo / Norma Castañeda Bustamante

Comité editorial

Ignacio Gatica

Dolly Espíndola / José Flores Salgado / Ma. Isabel García Rodríguez /

María Eugenia Salazar

Primera edición, 1997

D.R. © 1997. Universidad Autónoma Metropolitana

Unidad Xochimilco

Calzada del Hueso 1100

Col. Villa Quietud, Coyoacán

04960. México DF.

ISBN 970-654-051-2

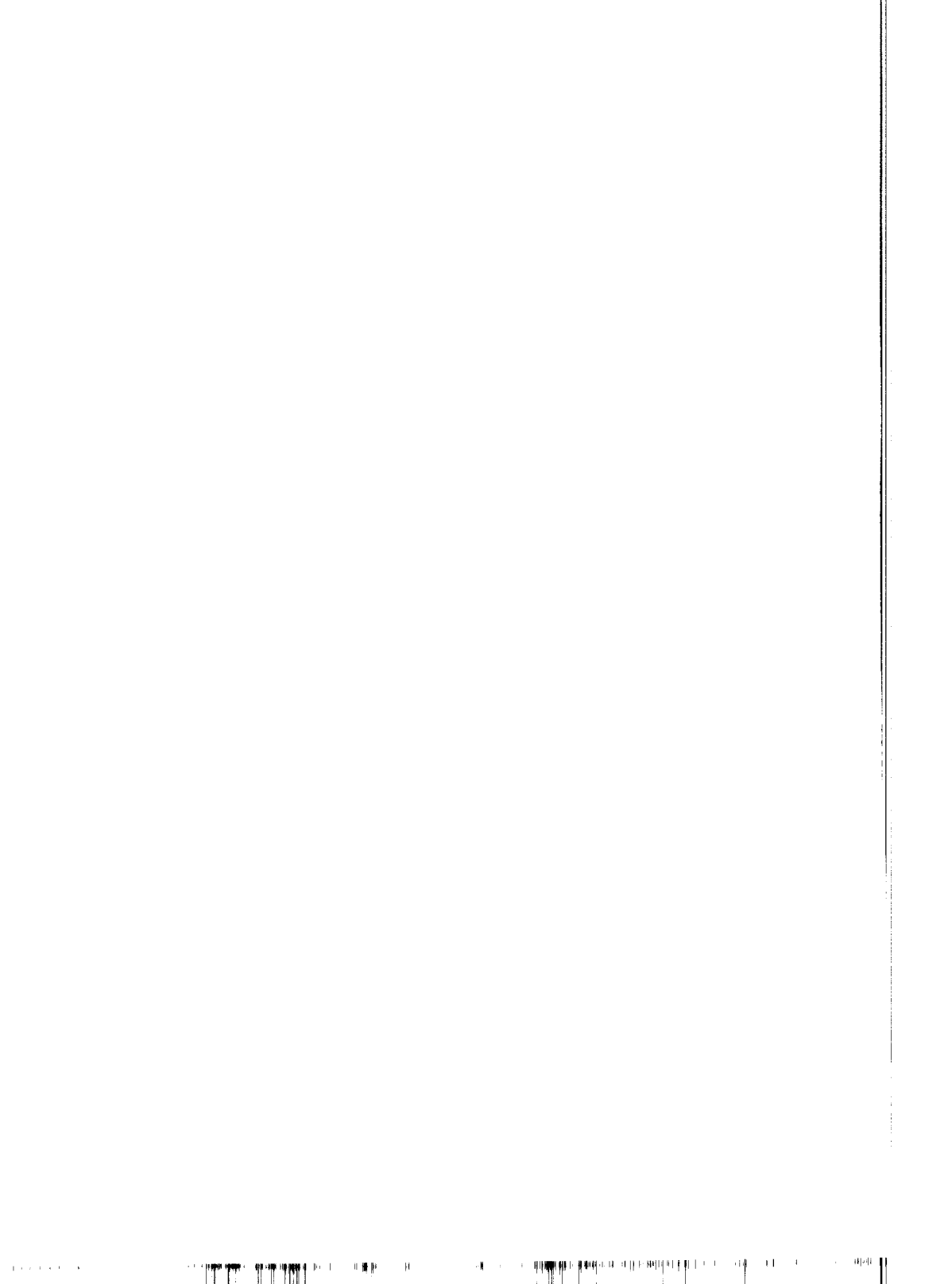
Impreso y hecho en México / *Printed and made in Mexico*

Índice

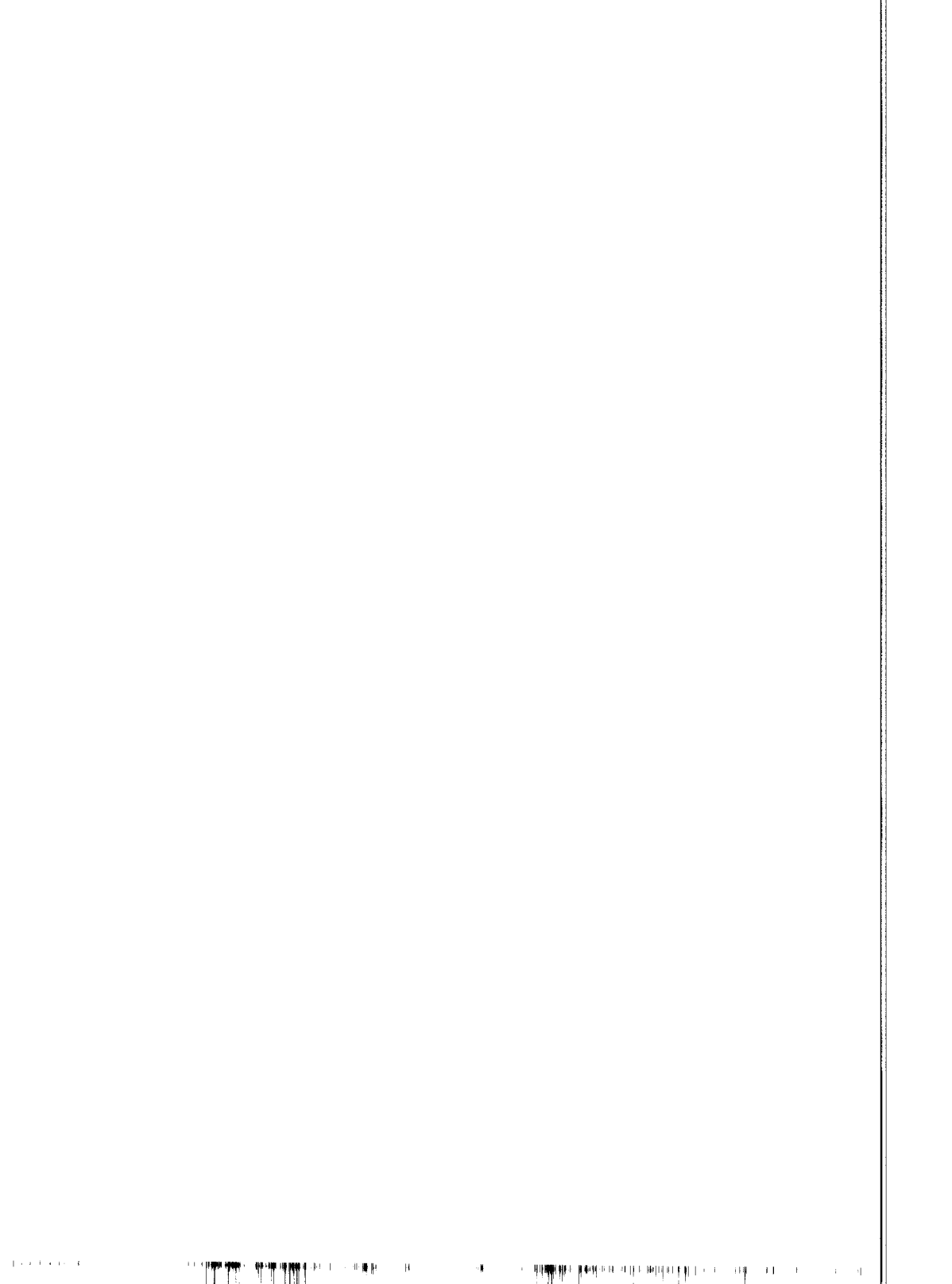
Introducción	13
¿ Qué es la estadística?	15
Estadística descriptiva e inferencial	16
¿Con qué objeto se estudia la estadística?	19
Elementos para realizar una investigación usando estadística	20
Variables	21
El proceso de datos y su descripción	
Tabulación manual	27
Proceso electrónico de datos	29
Cuadros estadísticos	29
Elementos para la elaboración de cuadros estadísticos	31
Elaboración de cuadros estadísticos usando Excel	33
Porcentajes, proporciones, razones, coeficientes e incrementos	39
Porcentajes	39
¿Cómo se calculan los porcentajes en las tablas estadísticas?	41
Proporciones	44
Razones	46

Coeficientes	47
Incrementos	49
Distribución de frecuencias y gráficas en estadística	53
Distribución de frecuencias para variables cuantitativas	53
Distribución de frecuencias para variables cualitativas	60
Gráficas	62
Histogramas y polígonos de frecuencias	62
Ojivas	69
Gráfica de barra o columna simple	73
Gráfica de sectores	74
Gráfica de barras agrupadas	75
Gráfica de barras de desviaciones	76
Mapas estadísticos y gráficas pictóricas	78
Medidas descriptivas de la distribución de frecuencias	81
Medidas de posición o de tendencia central	82
Media aritmética	82
Media ponderada	89
Media geométrica	92
Mediana	93
Moda	98
Cuartiles, deciles y percentiles	101
Medidas de dispersión o de variabilidad	107
Amplitud o rango	108
Desviación absoluta promedio	109
Varianza	111
Desviación estándar	115
Dispersión relativa: el coeficiente de variación	119
Medida de sesgo (asimetría)	120
Medida de curtosis	124
Un ejemplo del uso de las funciones estadísticas en Excel	128

Resumen de fórmulas estadísticas y funciones en Excel	131
Anexo	
Elementos básicos de la hoja electrónica Excel	137
Introducción	137
Movimientos rápidos en la hoja	139
Principales componentes de la ventana de Excel	140
Submenús de Excel	142
Gráficas en Excel	149
Barras de herramientas	151
Bibliografía	155



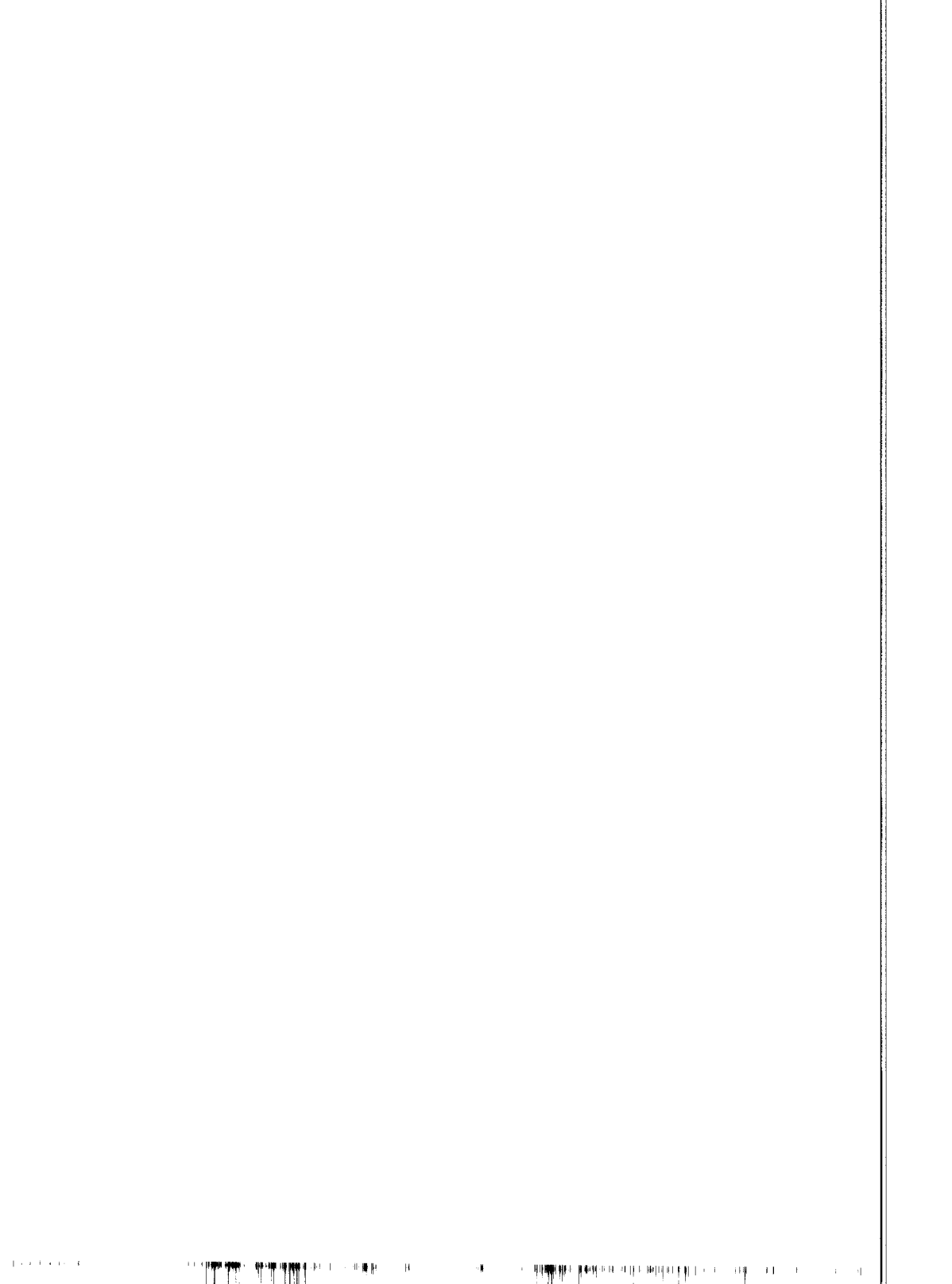
Para
Alberto Isaac Pierdant Castellanos
una semilla que dará un gran fruto



Introducción

El avance tecnológico de las sociedades posindustriales aunado a la aparición de herramientas como la computadora, han obligado a los individuos a asimilar más rápidamente cualquier cambio que se produzca en la sociedad. Uno de estos cambios es el uso intensivo y común de las computadoras y la estadística en todos los ámbitos de la sociedad prácticamente impactándola en mayor o menor grado. Es precisamente el ámbito de las ciencias sociales, el que mayor participación, uso y explotación le ha dado a estas herramientas, en muchas ocasiones en forma no muy adecuada, por lo que este texto propone un análisis, combinación y utilización de las mismas con el objetivo de dar al individuo una mejor comprensión de los fenómenos sociales, satisfaciendo al mismo tiempo las crecientes exigencias de la sociedad.

Si bien es cierto que el profesional de las ciencias sociales cuenta hoy con un gran variedad de herramientas, la estadística y las computadoras son elementos que le han permitido ampliar aún más su horizonte de trabajo. Sin embargo, ¿cómo poder obtener de estas herramientas una mayor ayuda en tiempos donde la velocidad de respuesta a los problemas sociales requiere día a día de una mayor rapidez? Una respuesta fácil a esta problemática implica el uso de la estadística con ayuda de la informática. El presente texto por lo tanto, pretende mostrar ideas al profesional de las ciencias sociales que le permitan combinar estas herramientas proporcionándole la rapidez de respuesta requerida.



¿Qué es la estadística?

Las matemáticas nacen prácticamente con el hombre debido a la necesidad natural de contar y, con ella, en forma intuitiva, nace la estadística.

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 a los registros que ya elaboraban los babilonios y los romanos, sobre la población.

Los estudios de la población en sentido estrictamente científico aparecen por primera vez a finales del siglo XVIII y principios del siglo 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 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 interpretarse² como:

Un método que permite organizar, sintetizar, presentar, analizar, cuantificar e interpretar grandes volúmenes de datos, de tal manera que se

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-X, 1980.

En las anteriores definiciones hemos utilizado una serie de conceptos muy comunes en estadística, por lo que es conveniente especificarlos de tal forma que permitan dar un mayor entendimiento a las mismas.

- *Dato*: es un número o una medida que ha sido recopilada como resultado de una observación. Los datos puede 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 de una persona, el sexo de una persona, el nombre de una persona, etcétera.
- *Variable*: para obtener estadísticas manejamos conjuntos que poseen un determinado o indeterminado número de unidades (personas, objetos, etc.). Las unidades de estudio tienen determinadas características; por ejemplo, para un ciudadano mexicano, podríamos señalar: sexo, edad, estatura, peso, lugar de nacimiento, estrato social, grado de escolaridad, religión, estado civil, etcétera; todas y cada una de estas características, que adquieren diferentes valores en cada persona, lugar o cosa y que son susceptibles de una medición, reciben el nombre de variables. De esta forma, el estudio de los habitantes de una población, requeriría tal vez, del uso de variables como: sexo, edad, estatura, peso, estrato social, 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, etc.) 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, siguiendo 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 se ha tomado.

puedan obtener conclusiones válidas (dar información) sobre los fenómenos en estudio.

Este amplio concepto ha llevado a los especialistas en estadística a llegar a un acuerdo que permite clasificar a esta materia en dos grandes ramas, *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, por lo que ésta no puede inferir o generalizar acerca de la totalidad de donde provienen dichas observaciones (la población).

Estadística inferencial

Es la otra rama de la estadística que tiene como objetivo generalizar o inferir conclusiones útiles sobre la totalidad de las observaciones (población) a partir del análisis de los datos coleccionados (muestra). En otras palabras, 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, esto a causa de que las unidades seleccionadas (muestra) sean más o menos numerosas y hayan sido seleccionadas de acuerdo a ciertos procedimientos y que la variabilidad de las características de estudio sea más o menos grande.

- *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 se considera una muestra de 100 posibles clientes de un producto y se les entrevista para conocer las preferencias sobre el mismo, obteniéndose como resultado que 70 prefieren el producto, entonces la proporción muestral será 0.70, constituyendo ésta una estadística.

¿Con qué objeto se estudia la 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 que esta última aseveración dependerá del conocimiento adecuado que sobre este instrumental estadístico tenga el usuario. Así, los individuos que requieren elaborar estadísticas se preguntan, en muchas ocasiones, con qué objeto 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 puede observarse, 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

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

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.

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 muy 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 de tal forma 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 inferencial seleccionado. Esto es, se exponen los datos muestrales mediante de 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 tanto, la variable o variables de estudio que permitan 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 características (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 pueden clasificarse 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, dentro de 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 sexo (masculino, femenino), la variable religión (católica, protestante, etc.), la variable idioma (español, inglés, francés, etc.), la variable estatus social (alto, medio, bajo), la variable calidad de un servicio (bueno, regular, malo), etcétera. Al igual que las variables cuantitativas, éstas 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 sexo de los individuos en un grupo universitario, su idioma, su religión, la carrera que cursan, etcétera. En algunos casos, a estas variables se les asignan números, de acuerdo a una regla específica, pero estos números sólo se emplean para diferenciar a los distintos objetos o categorías. Por ejemplo, la variable sexo podría clasificarse mediante dos categorías: 1 para los hombres y 2 para las mujeres. De esta forma, en una investigación todos los individuos 1 serían hombres y todos los 2 serían 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 a un criterio jerárquico. La diferencia que se establece entre las categorías ordinales no tiene significado cuantitativo, sólo expresan, por ejemplo, que una situación es mejor que otra, pero no cuantifican. Un

ejemplo de ello, es la evaluación que un cliente puede realizar sobre el servicio en un restaurante. El servicio fue malo, regular, bueno y muy bueno. Como puede observarse, la variable mide diversas categorías que han sido establecidas de acuerdo a 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 acostumbra hacer 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 en el momento de su análisis puede ser dependiente o independiente. Por ejemplo, en el área de las ciencias sociales es común establecer relaciones entre variables experimentales, y es entonces, que en estas ciencias, las variables se clasifican desde el punto de vista metodológico en: *variables dependientes* y *variables independientes*.

La *variable dependiente* es aquella cuyos valores están condicionados por los valores que toman 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 a sus valores a la variable dependiente. Proponemos un ejemplo de tipo económico que nos permite establecer una relación. Consideremos el comportamiento del ahorro de un individuo en una sociedad. El modelo económico que explica su ahorro podría ser:

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

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

Ahorro = (\$)	Ingreso (\$)	- Gasto (\$)	Observación
900	1 000	100	Ahorro alto
200	1 000	800	Ahorro bajo
0	1 000	1 000	No hay ahorro
-200	1 000	1 200	No hay ahorro. Déficit cubierto con crédito

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. Ésta, por lo tanto, 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 puede observarse, 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 deben realizarse para medir una variable.

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

Con estas dos definiciones, el estudiante o investigador estará ahora en posibilidad de acotar adecuadamente las variables para un manejo estadístico, de acuerdo al interés que se tiene en ellas, para la realización de un estudio o investigación. Un par de ejemplos de ello, los mostraremos a continuación.

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”.

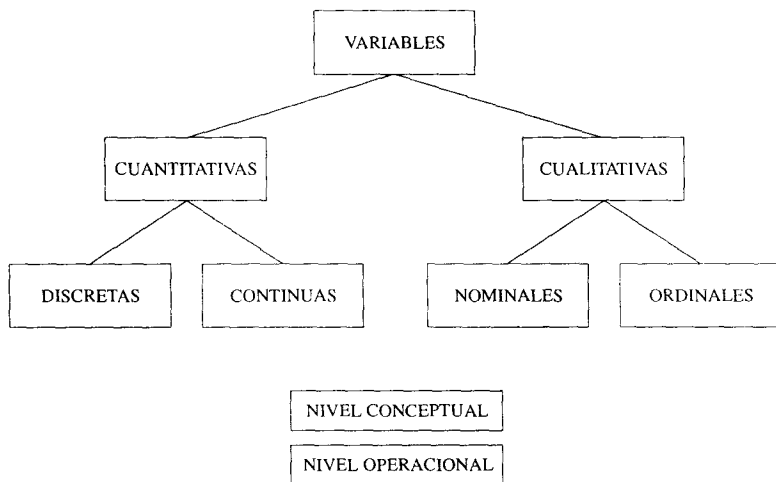
Nivel operacional: “Revisión de las tarjetas de asistencia al trabajo durante el último bimestre”.

Variable = “Sexo”.

Nivel conceptual: “Condición orgánica que distingue al macho de la hembra”.

Nivel operacional: “Asignación de la condición orgánica: masculino o femenino”.

En la siguiente página 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.



El proceso de datos y su descripción

Una vez que el investigador ha diseñado su experimento, y a través de un instrumento de recolección, colectado los datos, la siguiente etapa del proceso estadístico consiste en agrupar 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 de datos se obtiene mediante la formación de cuadros o tablas estadísticas, lo cual se define como *presentación tabular*.

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, utilizamos la definición operacional de las variables. Posteriormente se procede al conteo y clasificación de cada dato en los grupos o clases determinados previamente. Así, por ejemplo, en la característica: sexo, se establecen primeramente dos clases o categorías, definidas como: masculino y femenino; una vez hecho esto se deberá contar y clasificar a cada hombre o mujer entrevistada en el estudio. Para lograr tener esta primera agrupación, el investigador o estudiante cuenta con la tabulación manual o 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 cuándo 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 sexo de 24 personas en la siguiente forma:

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 tienen por objeto tener a la vista los datos que se encuentran dispersos en los cuestionarios o tarjetas del investigador, facilitando al mismo tiempo su conteo. El registro de estos datos, se acostumbra hacer por medio de señales simples (:; ✓; x), y los totales se muestran al pie de las columnas como se puede apreciar en el siguiente cuadro de concentración.

Cuadro 1

Cuest	Sexo		Nacionalidad			Carrera que cursa			Trabaja	
	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	
...										
<i>Total</i>	<i>10</i>	<i>10</i>	<i>18</i>	<i>2</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>5</i>	<i>15</i>

La tabulación manual es conveniente emplearla cuando el número de cuestionarios o datos recolectados no es muy grande y el número de categorías o clases especificados para ser tabulados sea poco numeroso. Bajo 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 que evaluará el cuestionario aplicado, así como en la captura que se tiene que hacer 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.

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 tienen como objetivo ser sólo 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: el permitir 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, estos pueden ser clasificados en dos tipos: *cuadros de trabajo* y *cuadros de referencia*.

Cuadros de trabajo. Son aquellos cuadros estadísticos que contienen datos producto de una tabulación. En otras palabras, son cuadros depositarios

de datos. Estos cuadros 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)

<i>año</i>	<i>oro</i>	<i>plata</i>
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.

En el cuadro de trabajo 2, se observa simplemente la presentación de los datos de la producción de oro y plata del año 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, etc.). El cuadro 3 muestra un cuadro de referencia elaborado a partir del cuadro de trabajo número 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 han plasmado en un cálculo de variación anual.⁵

⁵ El cálculo de la variación anual se explica detalladamente en el tema: Incrementos (*infra*: 49).

Cuadro 3
Producción de oro y plata en México
(toneladas)

<i>año</i>	<i>oro</i>	<i>variación anual %</i>	<i>plata</i>	<i>variación anual %</i>
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: *Idem.*

Elementos para la elaboración de cuadros estadísticos⁶

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 relativa al cuadro como un todo o a la parte principal del mismo.

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

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: sexo 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.
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 los renglones, 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 fuera sólo la obtenida durante el primer semestre del año, esto debería indicarse 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 cálculo⁷ dos ejemplos: el cuadro de trabajo número 4 y el cuadro de referencia número 5, sobre el gasto federal en educación durante tres generaciones en México.

Elaboración de cuadros estadísticos usando 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 sugiero revisar el anexo (*infra*: 137), 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 sobre el tema, le sugiero 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.

- Seleccione en la hoja el área de construcción del cuadro (columnas y renglones).

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

Cuadro 4 ← Número del cuadro

Gasto federal en educación durante tres generaciones¹ en México ← Nota en el título

Casillas 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

← Línea de cabecera

← Renglones

← Renglón de total

← Línea final de cuadro

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.

CUADRO DE TRABAJO

Título

Cuadro 5

Número del cuadro

Nota en el título

**Gasto federal en educación durante tres generaciones¹
en México (%)**

	GENERACIÓN		
	1977-1992	1978-1993	1979-1994
GASTO FEDERAL EN EDUCACIÓN	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	100.00	100.0	100.00
BÁSICA	71.4	72.5	72.2
MEDIA	21.7	21.1	21.4
SUPERIOR	6.9	6.5	6.5

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

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

CUADRO DE REFERENCIA

- 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).
- Capture los datos en los renglones del cuadro (figura 1).
- Capture en su caso las notas al pie y la fuente, dejando entre ambas una línea de separación (figura 1).
- Agregue en el renglón respectivo las fórmulas o cálculos necesarios en el cuadro estadístico, como se indica en la figura 2.
- De formato final al cuadro (figura 2), iniciando con la alineación a 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, etc.) 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

File Edit View Insert Format Tools Data Window Help							
H	I	J	K	L	M	N	
4	Cuadro 4						
5							
6	GASTOS FEDERAL EN EDUCACIÓN DURANTE TRES GENERACIONES ¹						
7	EN MÉXICO						
8		GENERACIÓN					
9		1977-1992	1978-1993	1979-1994			
10	GASTO FEDERAL EN EDUCACIÓN ²						
11	BÁSICA	115 244	113 595	111 230			
12	MEDIA	23 079	23 210	23 694			
13	SUPERIOR	20 930	23 224	25 350			
14							
15	TAMAÑO DE LA GENERACIÓN ³						
16	BÁSICA	12 629	13 536	14 126			
17	MEDIA	3 842	3 939	4 180			
18	SUPERIOR	1 213	1 206	1 269			
19	¹ Se supone un promedio de 16 años de instrucción (seis de básica, seis de media y cuatro de superior).						
20							
21	² Millones de nuevos pesos de 1994.						
22	³ Miles de matriculados en el sistema de educación pública.						
23							
24	FUENTE: Departamento de Estudios Económicos de Banamex, con datos de informes de gobierno.						
25							

Figura 2

	H	I	J	K	L
3		Cuadro 4			
4					
5	GASTOS FEDERAL EN EDUCACIÓN DURANTE TRES GENERACIONES ¹				
6		EN MÉXICO			
7		GENERACIÓN			
8		1977-1992	1978-1993	1979-1994	
9	GASTO FEDERAL EN EDUCACIÓN ²	159 253	160 029	160 274	
10	BÁSICA	115 244	113 595	111 230	
11	MEDIA	23 079	23 210	23 694	
12	SUPERIOR	20 930	23 224	25 350	
13					
14	TAMAÑO DE LA GENERACIÓN ³	17 684	18 681	19 575	
15	BÁSICA	12 629	13 536	14 126	
16	MEDIA	3 842	3 939	4 180	
17	SUPERIOR	1 213	1 206	1 269	
18	¹ Se supone un promedio de 16 años de instrucción (seis de básica, seis de media y cuatro				
19	de superior).				
20	² Millones de nuevos pesos de 1994.				
21	³ Miles de matriculados en el sistema de educación pública.				
22					
23	FUENTE: Departamento de Estudios Económicos de Banamex, con datos de informes				
24	de gobierno.				
25					

Porcentajes, proporciones, razones, coeficientes e incrementos

La elaboración de cuadros estadísticos y estadísticas en general implican el uso de cierto herramental aritmético que permita obtener medidas de comparación de los datos que han sido condensados. Estas herramientas que permiten analizar las características 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 o total 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, 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 cien, nos indica el número de hombres por cada 100 personas que hay en esa población.

Hombres	: 20
Mujeres	: <u>60</u>
Total	80

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

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

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:

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

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 el de 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 decimales.

Finalmente, es importante mencionar que es un error manipular los porcentajes como si fueran números absolutos; es decir: cuando éstos se han obtenido de bases diferentes no se pueden sumar, promediar o combinar. 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 los cuadros estadísticos?

La elaboración de cuadros estadísticos busca generalmente poner en relación dos o tres características, objeto de una investigación, así por ejemplo, podemos comparar: carrera universitaria y sexo, sexo y votación, trabajo y sexo, edad, delincuencia y zona de la ciudad, afiliación en un 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 sexo 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 sexo.

Cuadro 6
Estructura por sexo y carrera del grupo SB09/950 de la UAM-X
(alumnos)

carrera	Sexo		Total
	hombres	mujeres	
administración	3	2	5
economía	3	1	4
sociología	4	3	7
psicología	4	5	9
comunicación	3	4	7
<i>Total</i>	<i>17</i>	<i>15</i>	<i>32</i>

Fuente: Elaboración propia con datos hipotéticos.

Los resultados del cálculo se muestran en el cuadro 6.1

Cuadro 6.1
Estructura por sexo y carrera del grupo SB09/950 de la UAM-X
(%)

carrera	Sexo		Total
	hombres	mujeres	
administración	18	13	16
economía	18	7	12
sociología	23	20	22
psicología	23	33	28
comunicación	18	27	22
<i>Total</i>	<i>100</i>	<i>100</i>	<i>100</i>

Fuente: *Idem.*

*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 sexo, entonces, la comparación debe hacerse en el sentido de la variable carrera. En el grupo SB09/950, la carrera de administración ha sido seleccionada por el 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 carrera que las mujeres. De los alumnos del grupo sólo el 16 por ciento seleccionó la carrera de administración. Comparaciones similares pueden realizarse para cada uno de los renglones del cuadro.

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

Cuadro 6.2
Estructura por sexo y carrera del grupo SB09/950 de la UAM-X
 (%)

carrera	Sexo		Total
	hombres	mujeres	
administración	60	40	100
economía	75	25	100
sociología	57	43	100
psicología	44	56	100
comunicación	43	57	100

Fuente: *Idem.*

Como puede observarse, el calcular los porcentajes en un cuadro estadístico en un sentido u otro proporciona no sólo diferentes resultados sino también diferentes interpretaciones de las cifras. En este punto es muy 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.

Como un ejemplo de esto último se muestra a continuación el cuadro 7. En éste, el cálculo de los porcentajes se han pensado en forma vertical, es decir, siendo su interpretación por licenciatura; así, el factor personal docente y población estudiantil, el tronco interdivisional y divisional cuentan con el 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 no tendría ningún significado el análisis, ya que no tiene sentido sumar datos de docentes con alumnos, para tratar de establecer alguna relación.

Cuadro 7
UAM-X. Personal docente y población estudiantil por licenciatura
en la DCSH
1994

licenciatura	docentes	%	alumnos	%
tronco int. y div.	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
<i>Total</i>	<i>308</i>	<i>100</i>	<i>4 019</i>	<i>100</i>

Fuente: Elaboración propia con datos del informe de estadística escolar básica 94o de la coordinación de sistemas escolares de la UAM-X.

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 a la unidad. Las proporciones y los porcentajes ofrecen la misma información, sólo que estos últimos se emplean más ampliamente por ser más fácil su comprensión.

Matemáticamente la proporción se define como:

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

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

$$\begin{array}{l} \text{hombres : } 20 \\ \text{mujeres : } \underline{60} \\ \text{total} \quad \quad 80 \end{array}$$

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

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

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.

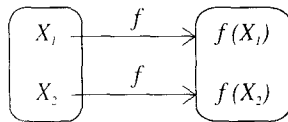
$$\text{proporción de hombres} + \text{proporción de mujeres} = 1$$

$$0.25 + 0.75 = 1$$

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:



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 formada por 80 habitantes, entonces la razón existente de hombres a mujeres es 1 a 3, es decir, un hombre por cada tres mujeres.

$$20 \text{ hombres} / 60 \text{ mujeres} = 1/3$$

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 como:

$$\text{razón de masculinidad} = \frac{H_x}{M_x} K$$

donde:

H_x representa el número de hombres de edad x ,

M_x representa el número de mujeres de edad x ,

K representa una constante (generalmente 100 o 1000).

Coefficientes

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 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 profesionistas, estudiantes, afiliación a grupos políticos, etcétera.

$$\text{tasa de mortalidad general} = \frac{\text{número de defunciones en una área determinada, durante un año dado}}{\text{población del área a mitad del año (1o. de julio)}} (1000)$$

$$\text{tasa de mortalidad infantil} = \frac{\text{defunciones en una \u00e1rea y tiempo determinados, de ni\u00f1os menores de 1 a\u00f1o}}{\text{n\u00famero de nacidos vivos en el \u00e1rea y tiempo del numerador}} (1000)$$

Los coeficientes o \u00edndices generales tambi\u00e9n reciben los nombres de tasas crudas o brutas, en virtud de que aparece en el denominador la poblaci\u00f3n y no el riesgo de que acontezca un evento.

Se pueden construir \u00edndices m\u00e1s espec\u00edficos dependiendo de las necesidades del investigador. A continuaci\u00f3n se muestra, en el cuadro estad\u00edstico 8, un ejemplo de este tipo de coeficientes.

Cuadro 8
Universidad Aut\u00f3noma Metropolitana-Xochimilco
Recursos Humanos

a\u00f1os	poblaci\u00f3n estudiantil ¹	personal docente ²	personal administrativo	coeficientes		
				alumno / docente	alumno / P.admvo.	P.admvo. / P.docente
1974	948	213	167	4.5	5.7	0.8
1979	7 937	443	589	17.9	13.5	1.3
1984	10 348	897	1 048	11.5	9.9	1.2
1989	10 745	955	1 061	11.3	10.1	1.1
1994	11 916	973	930	12.2	12.8	1.0

¹ Incluye a los alumnos de licenciatura y posgrado de los trimestres de oto\u00f1o.

² Incluye personal docente de tiempo completo, medio tiempo y tiempos parciales.

Fuente: Elaboraci\u00f3n propia con datos del Informe de actividades 1994-1995 del rector Jaime Kravzov Jinich.

Del cuadro 8, se observa que en 1974 exist\u00edan 4.5 alumnos por profesor, mientras que para 1994 esta relaci\u00f3n es de 12.2 alumnos, la cual pr\u00e1cticamente 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, lo que permite determinar cambios en ellos: 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 siguiente forma:

$$\text{incremento porcentual} = \frac{\text{valor del último dato} - \text{valor del dato base}}{\text{valor del dato base}} (100)$$

$$\text{incremento porcentual} = \frac{V_u - V_b}{V_b} (100)$$

En este tipo de cálculos es importante tener cuidado de 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%, 1 500%, 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, observamos que la matrícula en la UAM-X creció de 1974 a 1994 en 1 156.96%, lo que nos indica que en 20 años ésta creció 11.56 veces la magnitud que tenía en 1974.

$$\text{incremento}\% = ((11\ 916 - 948) / 948) (100) = 1\ 156.96\%$$

Para mostrar la construcción de un cuadro estadístico en donde son calculados incrementos, usaremos la información estadística de importaciones de galletas y pastas proporcionada por la Secretaría de Comercio y Fomento Industrial (cuadro 9).

File Edit View Insert Format Tools Data Window Help							
E88			=((C88-C87)/(C87)*100				
A	B	C	D	E	F	G	
78							
79							
80	Cuadro 9						
81							
82	Importaciones de Galletas y Pastas						
83							
84	Importación			Crecimiento			
85	(dólares)			(%)			
86		Pastas	Galletas	Pastas	Galletas		
87	1989	2 848 226	8 204 399	--	--		
88	1990	2 349 998	11 040 188	-17.5	34.6		
89	1991	2 391 122	12 980 980	1.7	17.6		
90	1992	7 747 454	22 752 862	224.0	75.3		
91	1993	8 334 925	26 152 349	7.6	14.9		
92	1994 ¹	1 659 165	7 222 652	--	--		
93	1 Enero-Febrero.						
94							
95	Fuente: Elaboración propia con datos de la Secretaría de Comercio y Fomento Industrial.						
96							
97							
98							
99							
100							
101							

- 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 incrementos en la celdas respectivas; por ejemplo, la celda E88 tiene la fórmula: $=((C88-C87)/C87)*100$, que permite conocer el incremento de las importaciones de pastas en 1990 respecto a 1989; la E89, tendrá el incremento de las importaciones de pastas de 1991 con respecto a 1990: $=((C89-C88)/C88)*100$, y así sucesivamente.
- Recuerde que no es posible obtener un incremento para el primer año base de comparación. En este caso particular, tampoco para

el último, dado que la información de 1994 es sólo para enero-febrero y los incrementos calculados son para años completos.

En los resultados del cuadro podemos observar 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 a 1991 en las importaciones tanto de pastas como de galletas, 224% y 75.3% respectivamente.

Indicaremos, finalmente, que es muy frecuente calcular incrementos promedio de los datos estadísticos, 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:

$$\text{incremento porcentual} = \frac{V_1 - V_0}{V_1 + V_0} \frac{2}{n} K$$

donde:

V_0 es el valor de la variable en el periodo base

V_1 es el valor de la variable en el periodo último

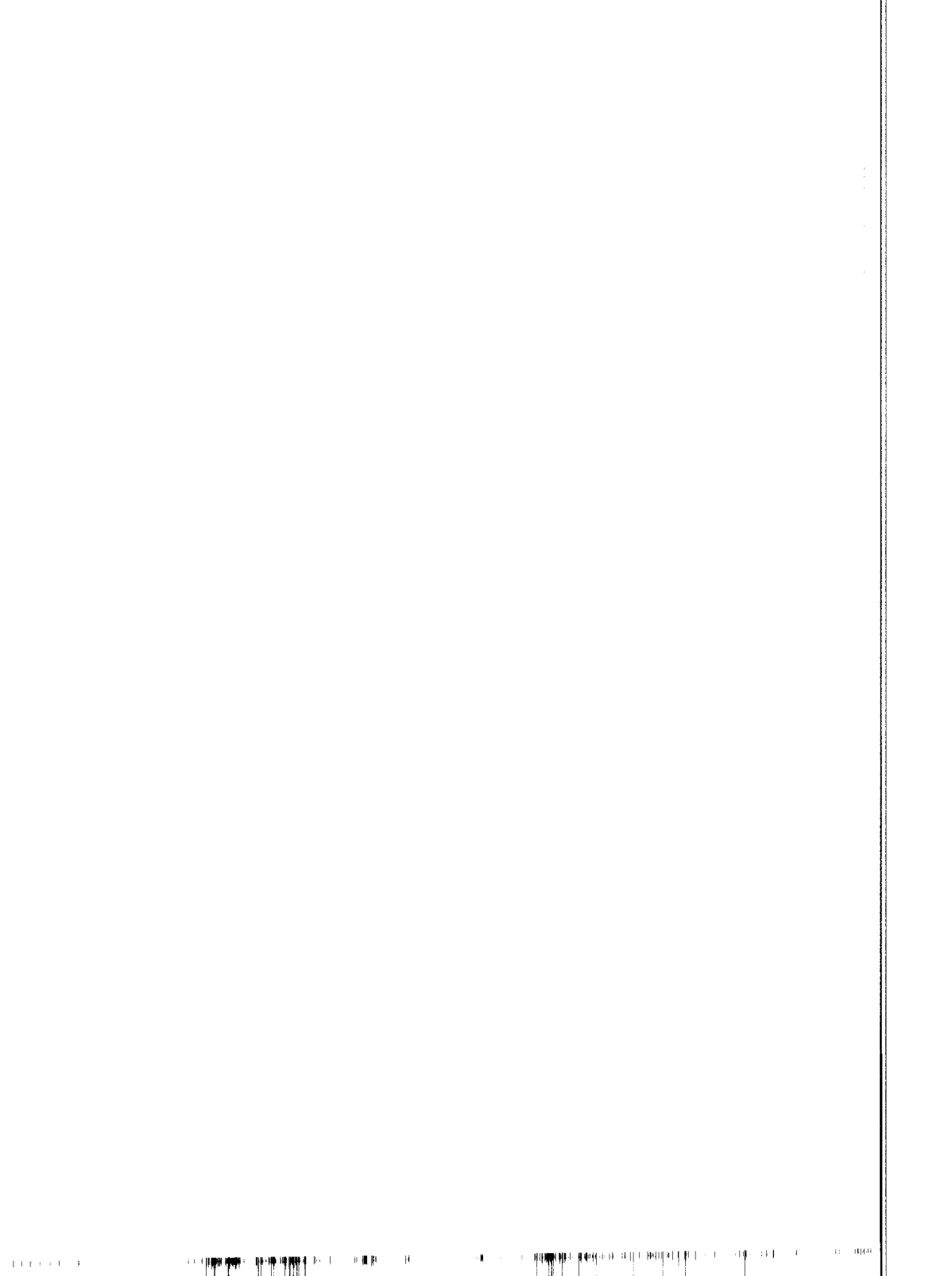
n es el número de periodos (años, semestres, bimestres, etc.)

K es una constante (generalmente 100)

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

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

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



Distribución de frecuencias y gráficas en estadística

En estadística descriptiva hemos comentado que los datos analizados pueden presentarse comúnmente mediante tres formas básicas: en forma textual, mediante cuadros estadísticos y mediante 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, etc.), 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) 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 caso este proceso es aun más sencillo.

Distribución de frecuencias para variables cuantitativas continuas

Cuando la variable es continua, la construcción de una tabla de frecuencia 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 reciben el nombre de *frecuencia del intervalo*, y se denota generalmente 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 para una variable continua es conveniente utilizar los siguientes pasos:

- 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 utilizando la siguiente tabla:

número de datos	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 o ancho. Observe que esta longitud o ancho es la misma para todos los intervalos en la tabla de frecuencia. Esto último se hace con la finalidad de facilitar los cálculos mediante métodos simplificados.

Para calcular el ancho del intervalo se aplica la siguiente relación:

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

Si el valor de t_i no es entero, el investigador puede manejarlo usando 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 de datos que se analiza recibe el nombre de *amplitud*.

Cuando el investigador desee utilizar intervalos con anchos desiguales deberá tomar en consideración que los cálculos serán más laboriosos por no ser aplicables los métodos simplificados.

- 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 inferior de la misma. Bajo esta dinámica se construyen todos los intervalos de clase, tomando en consideración que el primer intervalo debe contener al menor de los datos y el último el mayor.

Los intervalos hasta aquí contruidos reciben el nombre de intervalos de clase o intervalos ficticios.

- Se construyen los intervalos reales de clase, restando media unidad a los límites inferiores de los intervalos ficticios y agregando 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 columna, los intervalos reales de clase y, en la tercera columna, las frecuencias de clase, también llamadas frecuencias absolutas.

Para ejemplificar el proceso de construcción de una tabla de distribución de frecuencias para una variable continua utilizaremos 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, 8.4, 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).

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

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

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

2. Determinamos el ancho de los intervalos

$$\text{Dato mayor} = 9 \qquad \text{Dato menor} = 5$$

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

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

5 a 5.6 (5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6)
5.7 a 6.3
6.4 a 7.0
7.1 a 7.7
7.8 a 8.4
8.5 a 9.1

4. Creamos los intervalos reales de clase

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 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 se complementa generalmente 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 y dividirlos entre dos.

Frecuencia acumulada de la clase i (absoluta): 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 contiene como frecuencia acumulada el total de los datos.

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

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

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

Esta frecuencia muestra la proporción del número de casos que se han presentado en el intervalo i respecto al total de casos en la investigación. Si a este cociente se le multiplica por 100 entonces obtenemos un porcentaje, el cual permite hacer un análisis del comportamiento de los datos.

Frecuencia acumulada relativa de la clase i: es el cociente entre la frecuencia acumulada de la clase i y el número total de datos. Se expresa matemáticamente como:

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

Esta frecuencia muestra la proporción del número de casos que se han acumulado hasta el intervalo i respecto al total de casos en la investigación. Si a este cociente se le multiplica por 100 entonces obtenemos un porcentaje. 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 siguiente forma:

Cuadro 10

Clase	Intervalo de clase	f_i	m_i	F_i	f_i/n	F_i/n	$f_i \%$	$F_i \%$
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 el 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.

Para los estudiantes aprobados, el 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 alta. El 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 la construcción de la tabla. 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 la siguiente tabla de ejemplo.

Cuadro 11
Número de hijos por familia en Chalco, Estado de México
1995

Número de hijos	frecuencia (f_i)	f_i/n	F_i	F_i/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.

Para esta variable puede indicarse que en una muestra de 350 familias sólo 5.7% de ellas tiene un solo hijo; 34% dos hijos; 57% tres y, sólo 3% cuatro hijos. Esto permite indicar que el 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. La tabla se construye en estos casos 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 un ejemplo de distribución de frecuencias para datos cualitativos mostraremos la estructura por carrera del grupo SB09/950 de la UAM-Xochimilco.

Cuadro 12

carrera	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)

En el ejemplo, 46% de los alumnos del grupo cursarán la carrera de comunicación; 18% la de 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 sencilla, 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

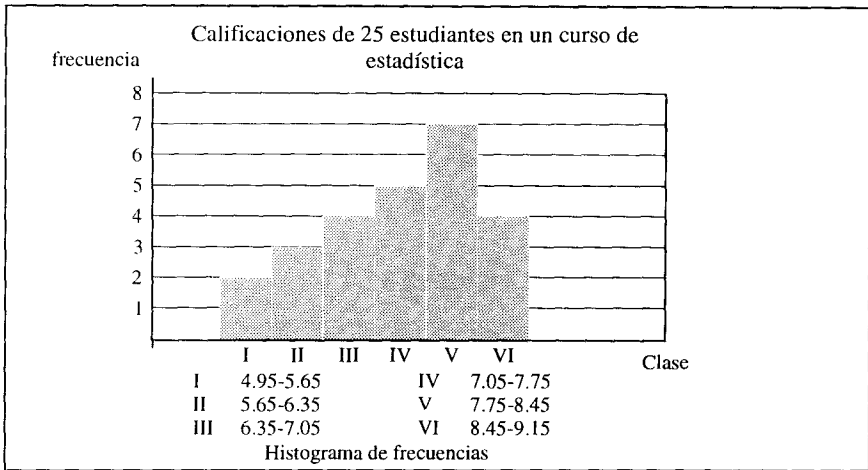
Un histograma de frecuencias es un gráfico de rectángulos que tiene su base en el eje de las abscisas (eje horizontal o eje de las equis); con anchura igual cuando se trata de representar el comportamiento de una variable discreta y anchura proporcional a la longitud del intervalo cuando se desea representar una variable continua. En este último caso el punto central de la base de los rectángulos equivale al punto medio de cada 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.

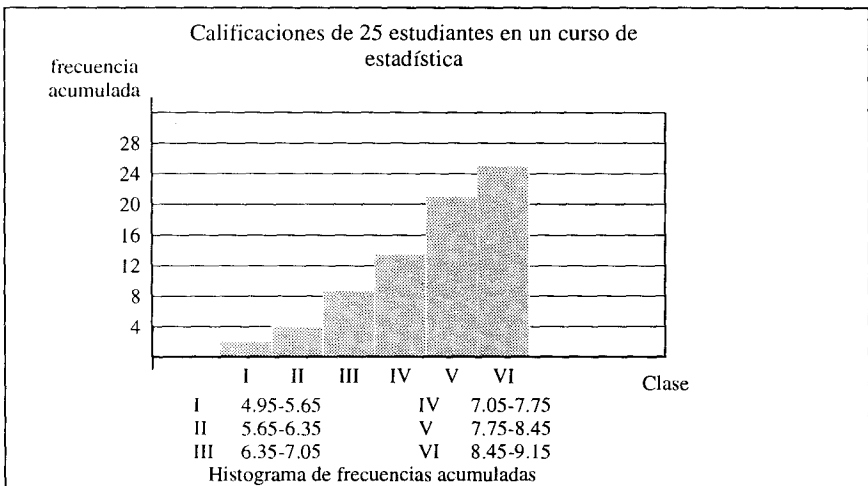
Los histogramas de frecuencias pueden construirse no sólo con las frecuencias absolutas, sino también con las frecuencias acumuladas y las frecuencias relativas. En este último caso el histograma recibe el nombre de histograma de frecuencias relativas, histograma de porcentajes o histograma de proporciones,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

Gráfica 1



Fuente: Datos del cuadro 10.

Gráfica 2

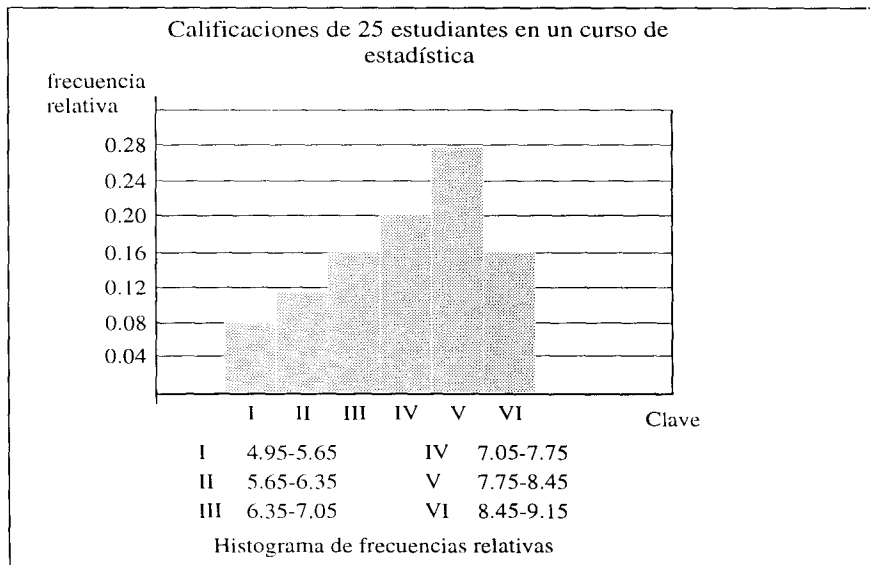


Fuente: *Idem.*

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

Gráfica 3

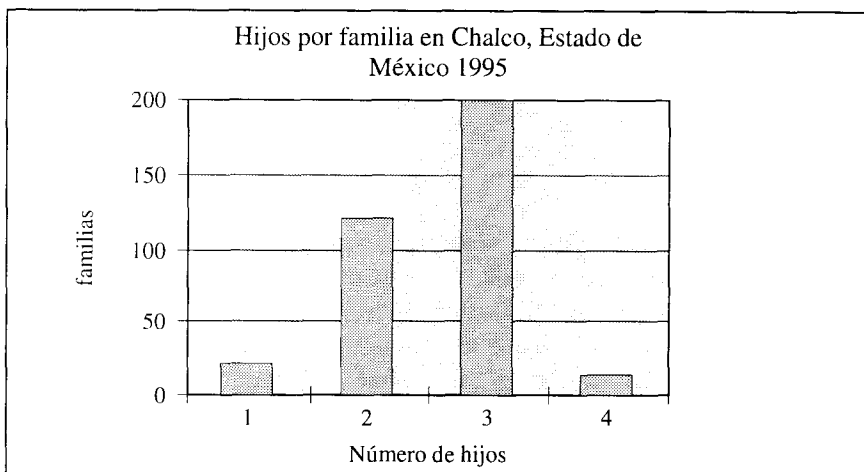


Fuente: *Idem*.

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

A continuación se muestra un histograma de frecuencias absolutas con los datos del cuadro 11, es decir, un histograma para una variable discreta.

Gráfica 4



Fuente: Datos del cuadro 11.

Los histogramas tienen como función, en el caso de variables cuantitativas (discretas o continuas) mostrar no únicamente una representación visual de los datos sino fundamentalmente tres características de su comportamiento. Éstas son:

1. la forma que presenta el histograma,
2. si existe algún intervalo donde se acumulan los datos o aparece una tendencia posicional de los mismos, y, por último,
3. el grado de dispersión o variabilidad de los datos.

En el ejemplo de las calificaciones (de 25 estudiantes en un curso de estadística) puede observarse que el 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.

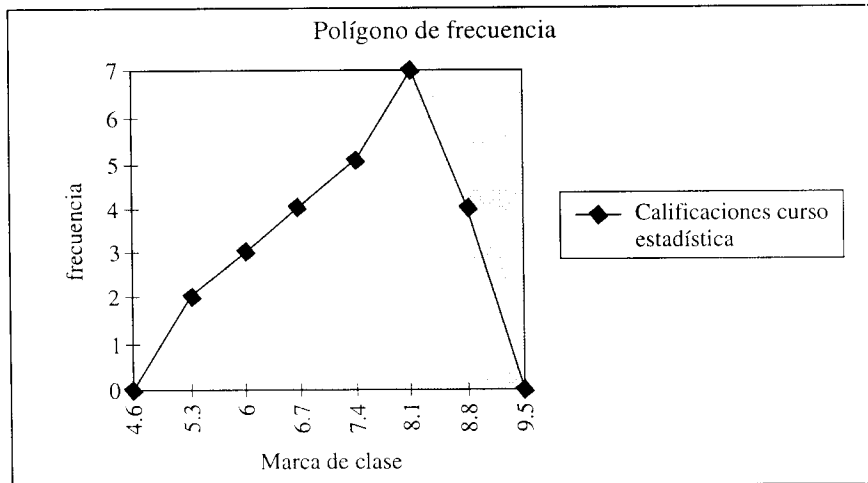
En el ejemplo de la variable discreta: número de hijos por familia en Chalco, es posible observar que los datos se concentran en los casos de las familias con dos y tres hijos. La variabilidad en estos datos respecto a los casos de dos y tres hijos tampoco es muy grande.

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 asociada a esa clase; posteriormente, estos puntos se unen por segmentos de recta. Para que el polígono quede cerrado se debe considerar un intervalo más al inicio y otro al final con frecuencias cero.

A continuación se muestra el polígono de frecuencias de las calificaciones de los 25 estudiantes de un curso de estadística (cuadro 10).

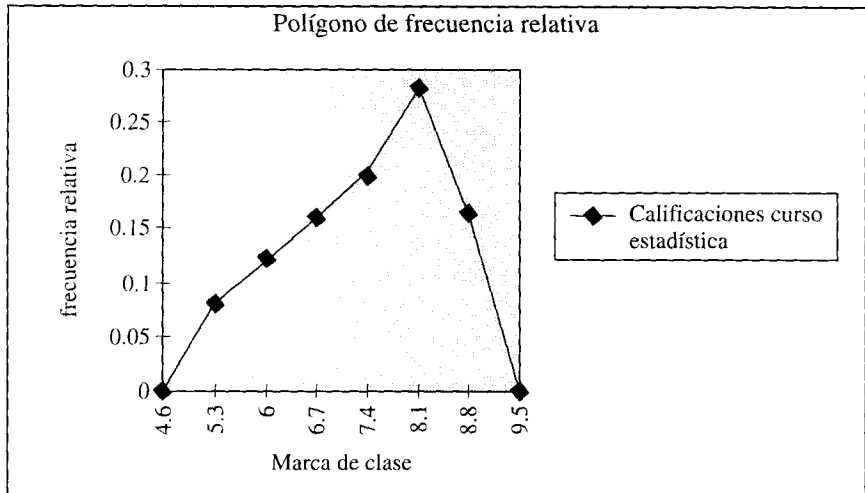
Gráfica 5



Fuente: Datos del cuadro 10

Los polígonos de frecuencia también se pueden construir utilizando las frecuencias relativas de la distribución de frecuencias; estos gráficos se denominan polígonos de frecuencias relativas (gráfica 6).


Gráfica 6



Fuente: *Idem.*

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

1. Se utilizan dos columnas de la hoja electrónica; la primera deberá contener la marca de clase, y, la segunda, 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 de estas dos columnas (rango con la información de la gráfica), y se oprime el botón del asistente que permite crear gráficos .
3. La forma del cursor cambia a una pequeña cruz con una gráfica. Seleccione un área en la hoja donde desee ubicar al polígono. Oprima el botón izquierdo del mouse y, sin soltar, genere un recuadro donde se ubicará la gráfica.
4. Siguiendo los pasos del asistente construya el gráfico.

En forma abreviada estos pasos son:

En la primer ventana del asistente se debe confirmar si el conjunto de celdas seleccionado es el deseado. Se confirma y se pasa al siguiente punto.

La segunda ventana pasa automáticamente al tipo de gráfico, el cual, como estándar, presenta un gráfico de columnas, entonces, deberá seleccionarse el gráfico de línea. De este último puede seleccionarse el tipo de gráfica marcada con el número 1.

La siguiente ventana muestra el polígono y dos opciones. La primera permite indicar si las series a graficar están dadas en columnas o renglones, y la segunda indica si la primer columna o renglón es el deseado para el eje de las abscisas (X 's).

La última ventana del asistente permite agregarle el título, la etiqueta para el eje de las ordenadas (Y 's) y la etiqueta para el eje de las abscisas (X 's).

Finalmente el nombre de la serie puede modificarse mediante los comandos: Formato/Selección de series/Nombres y Valores. En la ventana de este último, en el renglón de nombre, se da el nuevo nombre a la serie.

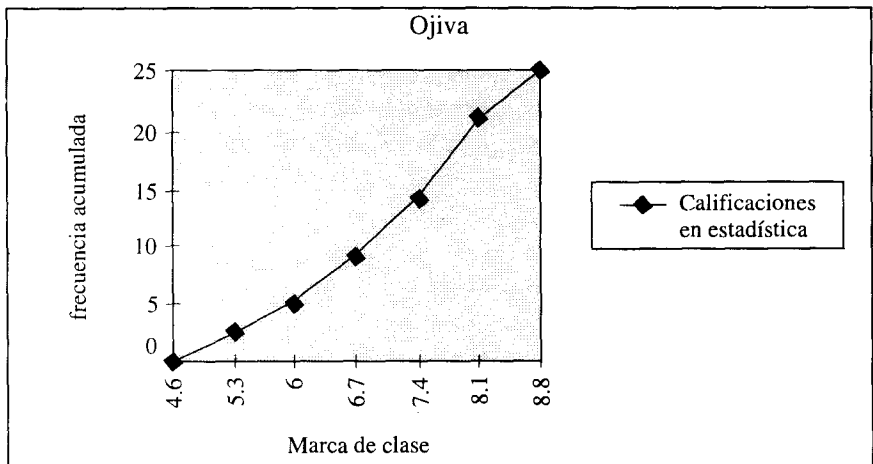
Ojivas

La ojiva es el polígono que se obtiene al unir, por segmentos de recta, los puntos situados a una altura igual a la frecuencia acumulada a partir de la marca de clase, en la misma forma en que se realizó para construir el polígono de frecuencias. La ojiva también es un polígono que se puede construir con la frecuencia acumulada relativa.

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

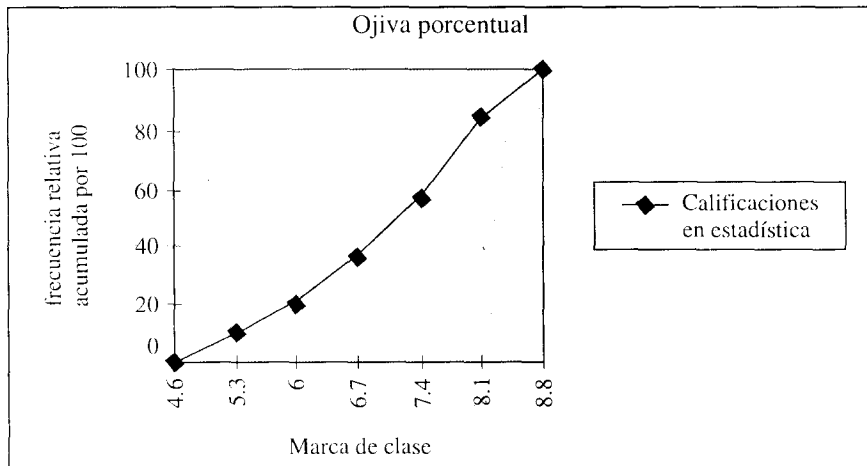
A partir de Excel la mecánica para la construcción de estos gráficos es similar a la utilizada para el polígono, salvo que, en este caso, se elimina la última marca de clase que tiene frecuencia cero, ya que la ojiva no es un gráfico cerrado.

Gráfica 7



Fuente: *Idem.*

Gráfica 8



Fuente: *Idem.*

La interpretación de estos gráficos es simple y muy útil; por ejemplo, de la primera ojiva puede observarse que 15 alumnos obtuvieron una calificación inferior o igual a 7.4: 60% de los alumnos del curso (este último dato se obtuvo con la segunda ojiva), y sólo cinco de ellos, 20%, una calificación inferior o igual a 6.

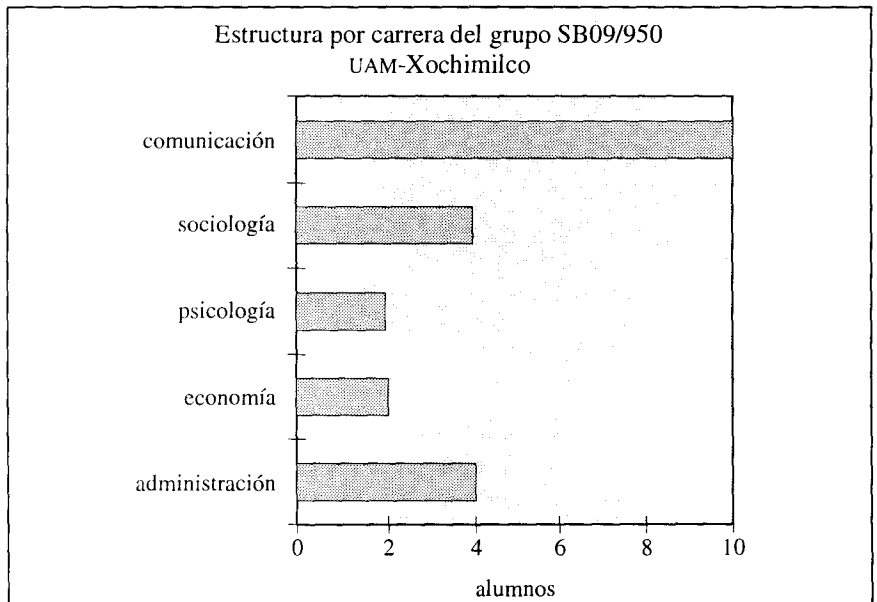
Histogramas para datos cualitativos

Como ya se indicó previamente, las variables cualitativas no tienen intervalos de clase por carecer éstos de sentido. Tampoco en ellas se calcula la frecuencia acumulada. Por lo tanto, para las variables cualitativas sólo existe la construcción de los histogramas de frecuencia absoluta y los histogramas porcentuales o de frecuencia relativa. *Para variables cualitativas no existe polígono de frecuencias.*

Un histograma de frecuencias para datos cualitativos también está formado por rectángulos que 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.

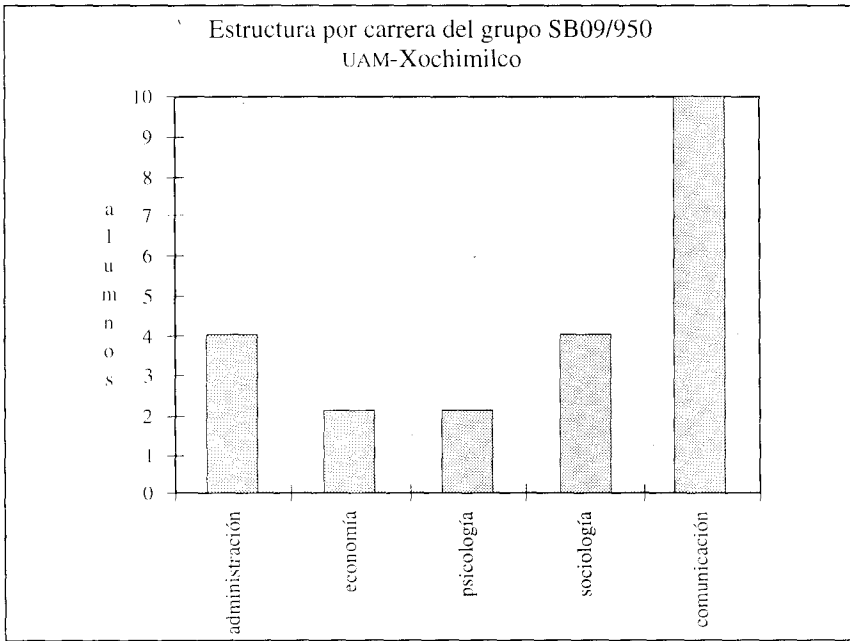
Para mostrar un ejemplo de este tipo de gráficos se utilizará la información del cuadro estadístico 12, “Estructura por carrera de los estudiantes del grupo SB09/950 en la UAM-Xochimilco”. Los gráficos 9 y 10 muestran histogramas para una variable cualitativa utilizando tanto el eje horizontal como el eje vertical para su trazo.

Gráfica 9



Fuente: Datos del cuadro 12.

Gráfica 10



Fuente: *Idem.*

Otras representaciones gráficas

En estadística es muy común presentar los resultados de un estudio mediante el uso de gráficas, es por ello, que no sólo encontraremos histogramas o polígonos de frecuencia, sino también otras formas gráficas de representar datos y cálculos estadísticos. A continuación se listan 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

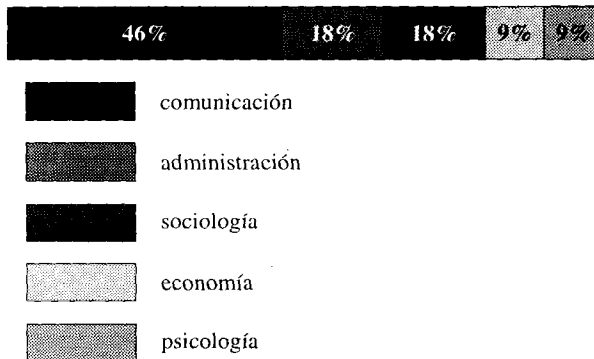
Es la más sencilla de las gráficas y consiste en representar datos mediante una barra o columna simple, la cual puede ser colocada horizontal o verticalmente.

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

Retomando los datos del cuadro estadístico 12, el gráfico de barra o columna simple con porcentajes se muestra como:

Gráfica 11

Estructura por carrera del grupo
SB09/950



Fuente: *Idem.*

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 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°.

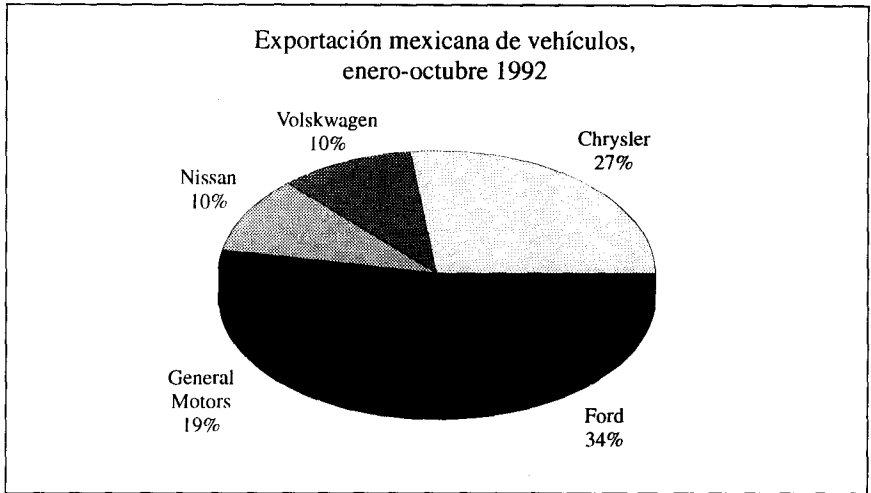
Para mostrar un ejemplo de estos gráficos se utilizarán los datos del cuadro estadístico 13, el cual trata sobre la exportación de vehículos por planta armadora en México en los años 1992 y 1993. Los datos de la gráfica 12 corresponden a 1992. Se puede elaborar un gráfico similar con los datos de 1993.

Cuadro 13
Exportaciones mexicanas de vehículos, enero-octubre
(unidades)

<i>compañía</i>	<i>1992</i>	<i>1993</i>
Chrysler	90 178	113 629
Ford	115 333	94 066
General Motors	62 227	73 348
Nissan	32 590	37 851
Volkswagen	32 573	60 487
Dina	0	340
<i>Total</i>	<i>332 901</i>	<i>379 721</i>

Fuente: Departamento de información de negocios de Banamex, dic. 1993.

Gráfica 12



Fuente: *Idem.*

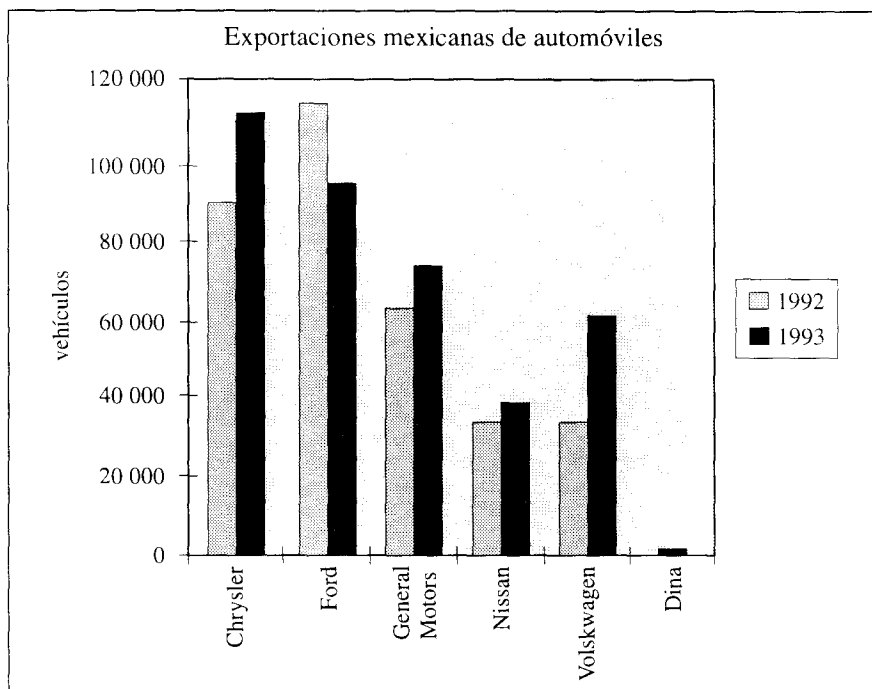
En el gráfico de sectores anterior se muestra la participación de cada planta armadora en las exportaciones de vehículos para el periodo enero-octubre de 1992.

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 pueden presentarse en forma horizontal o vertical.

Utilizando la información del cuadro 13 se muestra, a continuación, 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



Fuente: *Idem*.

Gráficas de barras de desviaciones

Es un gráfico de barras que puede elaborarse vertical u horizontalmente. Tiene, siempre, 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, como es el caso, por ejemplo, en una empresa, de las pérdidas y ganancias en el tiempo, los incrementos positivos y decrementos negativos de la producción, etcétera.

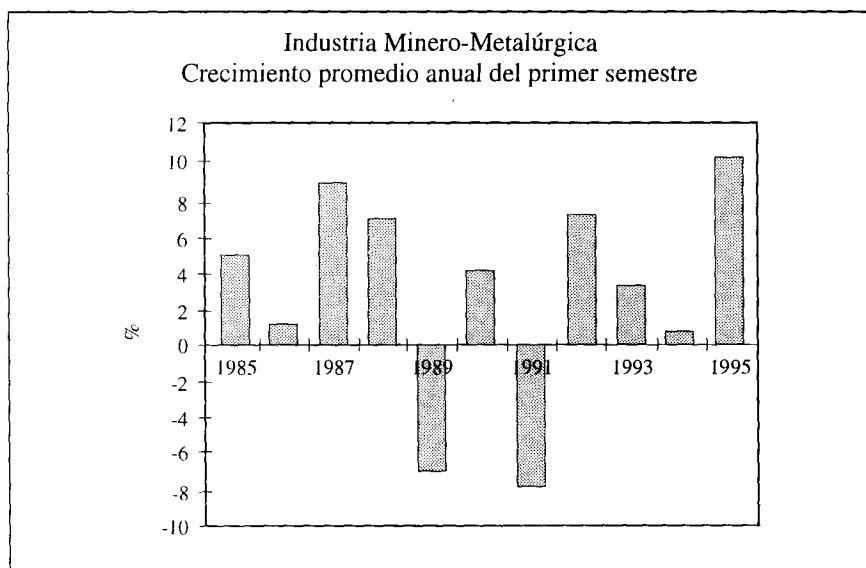
Como un ejemplo de este tipo de gráfico, se mostrará la información de crecimiento promedio anual del primer semestre para la industria minero metalúrgica mexicana de los años 1985 a 1995 proporcionada por el Instituto Nacional de Estadística, Geografía e Informática.

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



Fuente: *Idem.*

Mapas estadísticos

Tienen como objetivo representar relaciones espaciales en mapas geográficos. Entre algunos de los tipos más usados de mapas estadísticos tenemos: mapas sombreados, mapas punteados, mapas de isolíneas, mapas con gráficas sobrepuestas (barras, líneas, flujos, etc.) y mapas 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 mapas 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 siguiente página se muestra la forma en la que nuestro país fue dividido en cuatro regiones por los investigadores de Banamex en 1987 para elaborar la II 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, pueden sobreponerse dos gráficas de columnas en cada estado, que indiquen dichas magnitudes.

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.

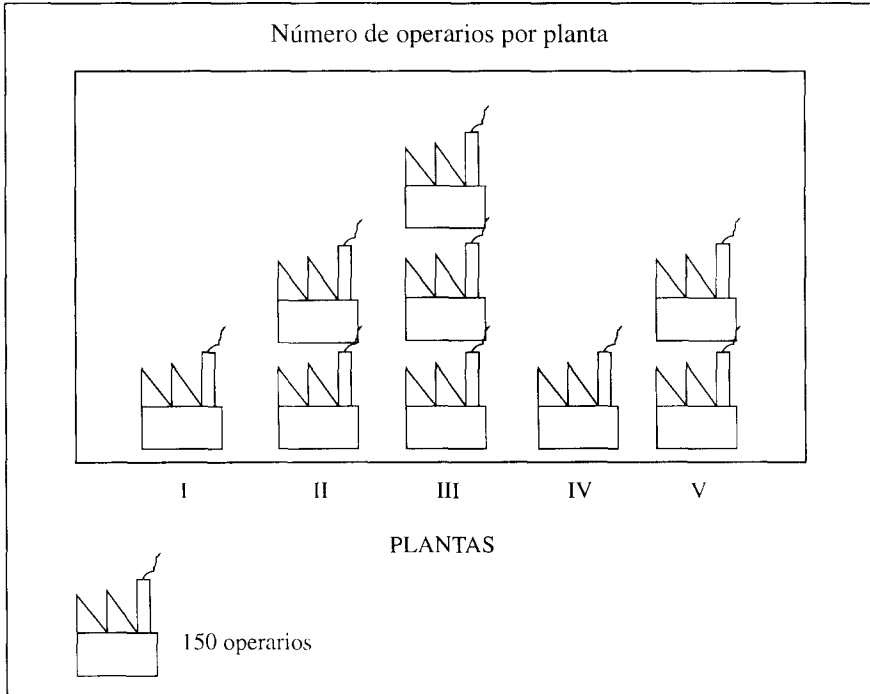
Gráficas con símbolos pictóricos de tamaño proporcional. En ellos las figuras deben tener una magnitud proporcional a los valores que representan.

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.

REGIONES Y CIUDADES, MUESTREOS DE CUOTAS
 II ENCUESTA NACIONAL SOBRE LOS VALORES DE LOS MEXICANOS 1987



Gráfica 15



Medidas descriptivas de la distribución de frecuencias

Hasta el momento, el análisis de un grupo de datos, ya sea en bruto o bien agrupado en un cuadro de distribución de frecuencias, se ha realizado por medio de algunos cálculos aritméticos simples (proporciones, porcentajes, índices, etc.) 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 siguiente pregunta: ¿existe alguna medida más exacta sobre el 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 pueden clasificarse de acuerdo con sus características en:

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

Las primeras tienen como objetivo buscar el punto medio o típico del conjunto de datos, de ahí que éstas reciban el nombre de medidas de posición, localización o de tendencia central.

Las segundas, se refieren al esparcimiento o grado de dispersión que tienen los datos con respecto a una medida de posición o bien respecto a 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 a una medida de posición.

Y finalmente, la medida de curtosis que se aplica a las distribuciones de frecuencia nos permiten cuantificar el tamaño del pico que presentan en el gráfico.

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

Medidas de posición o de tendencia central

Como ya se indicó, tienen por objeto encontrar el punto central 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 $x_1, x_2, x_3, x_4, \dots, x_n$ 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:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad \text{..... I}$$

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 que 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 es un parámetro que denotamos como:

$$\mu = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} \quad \text{..... II}$$

y cuyo símbolo es la letra griega *mu* : μ

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

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n} \quad \text{..... Ia}$$

$$\mu = \frac{\sum_{i=1}^N x_i}{N} \quad \text{..... IIa}$$

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 de este estadístico, se utilizará una muestra formada por 20 estaturas de estudiantes universitarios del grupo SB09/950 de la UAM-X.

Estaturas en metros y centímetros

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

Utilizando la ecuación I para el cálculo del estadístico tenemos:

$$\bar{X} = \frac{X_1 + x_2 + x_3 + \dots + x_{20}}{20}$$

$$\bar{X} = \frac{1.60 + 1.65 + 1.55 + \dots + 1.56}{20}$$

$$\bar{X} = 1.64$$

El estadístico nos indica que la estatura media del grupo de 20 estudiantes es de 1.64 mt.

Utilizando la hoja electrónica de cálculo Excel tenemos que la media de ese conjunto de datos se calcula con la función:

=AVERAGE (conjunto de celdas que contienen los datos)

cuya versión en español es:

=PROMEDIO (conjunto de celdas que contienen los datos)

Figura 3

	A	B	C	D	E	F	G	H
1								
2								
3		VARIABLE:	Estaturas de los alumnos del grupo SB09/950					
4								
5		DATOS A ANALIZAR:						
6					MEDIDAS DE POSICIÓN			
7		1.60	1.59					
8		1.65	1.64		MEDIA =	1.64		
9		1.55	1.54					
10		1.58	1.57					
11		1.70	1.69					
12		1.71	1.70					
13		1.68	1.66					
14		1.70	1.69					
15		1.68	1.67					
16		1.57	1.56					
17								
18								
19								
20								
21								
22								
23								
24								
25								

Observe en la figura 3 que cada dato u observación es colocado en una celda, para posteriormente calcular la media muestral con la función ubicada en la celda F8 definida como: =AVERAGE (B7:C16).

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 son:

media muestral o media poblacional para datos agrupados (Ib)

$$\bar{X}, \mu = \frac{\sum_{k=1}^k f_k X_k}{n} \quad \text{..... Ib}$$

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,
 $k = 1, 2, 3, \dots n$ intervalo

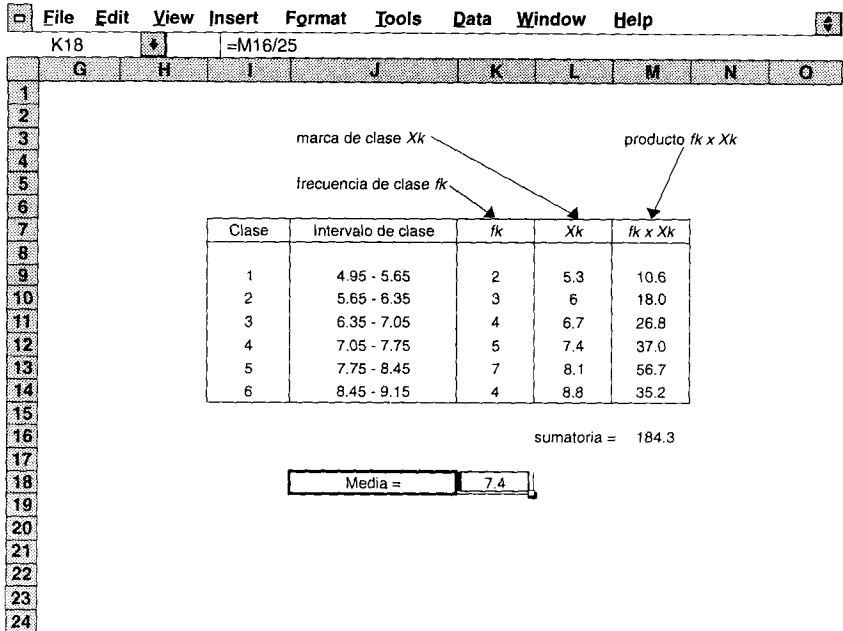
X_k = son las marcas de clase de cada intervalo,
 $k = 1, 2, 3, \dots n$ marca de clase

Para mostrar un ejemplo del cálculo de la media para datos agrupados usaremos la tabla de frecuencias del cuadro 10 de este texto a través de la hoja electrónica Excel, la cual muestra la clasificación de las calificaciones de 25 estudiantes en un examen de estadística (figura 4).

De la relación Ib y del gráfico anterior 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,

Figura 4



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

$$\text{Media} = 7.4$$

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 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

cuando en dichos datos encontramos los llamados valores extremos. Un dato extremo es aquel que sale fuera 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 seis estudiantes: \$3.5, \$3.2, \$3.8, \$2.9, \$3.0, \$50.0. El último dato del conjunto anterior sale fuera 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 incluyendo el dato extremo:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_6}{6}$$

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

$$\bar{X} = 11.07$$

Media sin incluir el dato extremo:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_5}{5}$$

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

$$\bar{X} = 3.28$$

En el primer caso el valor extremo genera una media (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 (excluyendo por supuesto el dato extre-

mo), es decir, los estudiantes gastan en transporte diariamente un promedio de \$3.28.

Como puede observarse los valores extremos sesgan el valor de la media, o sea, la media es sensible a los datos extremos en un conjunto de observaciones. Por este motivo es muy 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 cual 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 puede observarse en el siguiente ejemplo:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_5}{5}$$
$$\bar{X} = \frac{3.5 + 3.2 + 3.8 + 2.9 + 3.0}{5}$$
$$\bar{X} = 3.28$$

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 puede calcularse también como:

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

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 puede expresarse 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 = al peso asignado a cada elemento i
en la media ponderada $i=1,2,3,\dots,e$

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

El mejor ejemplo de uso de una media ponderada es el cálculo de la calificación final que los profesores de cada módulo hacen cada trimestre en la UAM-X.

Suponga que un estudiante de administración del 5to. 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 al 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

$$\bar{X}_w = \frac{(75)(10) + (20)(6) + (5)(6)}{75 + 20 + 5} = 9$$

Como puede observarse, la media pondera, 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 el determinar precisamente el valor que deben de tener los ponderadores para cada elemento a ponderar.

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 5.

Figura 5

	O	P	Q	R	S	T	U	V	W	
1										
2		MEDIA PONDERADA								
3										
4				investigación	matemáticas	contabilidad				
5		Datos		10	6	6				
6		Ponderadores		75	20	5				
7										
8							Xw = 9			
9										
10										
11				investigación	matemáticas	contabilidad				
12		Datos		10	6	6				
13		Ponderadores		70	20	10				
14										
15							Xw = 8.8			
16										
17										
18				investigación	matemáticas	contabilidad				
19		Datos		10	6	6				
20		Ponderadores		50	25	25				
21										
22							Xw = 8			
23										

En esta gráfica la ecuación de la media ponderada se ubica en la celda U8, la cual se define como: $= (R5 * R6 + S5 * S6 + T5 * T6) / (R6 + S6 + T6)$, siendo las ecuaciones de las medias ponderadas en U15 y U22 similares a ésta.

Media geométrica

Este estadístico se emplea principalmente cuando una variable presenta un comportamiento distinto en el tiempo, es decir, cuando se conjugan para el cálculo de un promedio 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, donde n es el número de factores empleados en el producto. Matemáticamente, la media geométrica queda indicada como:

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

n = número de factores

X_1, X_2, \dots, X_n = son los n datos 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 el año de 1993, la media 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

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

Media geométrica

$$MG = \sqrt[12]{(16.72)(17.73)(17.47) \dots (11.78)}$$

$$MG = 14.83$$

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

En Excel, el cálculo de este tipo de estadístico puede realizarse usando la función: =GEOMEAN(dato1,dato2,dato3,...,dato n), y en la versión en español, =MEDIA.GEOM(dato1,dato2,dato3,...,dato n). Para el ejemplo, la función estaría definida en una celda como:

$$=GEOMEAN (16.72,17.73,17.47,...11.78)$$

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: $Ume = (n+1)/2$, 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 5to. dato 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:

$$Me = 6$$

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 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 siguiente conjunto de datos:

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 $(n/2)+1$

$$[dato(n/2)] = [dato(8/2)] = [dato(4)] = 9$$

$$[\text{dato}(n/2)+1]= [\text{dato}(8/2)+1]= [\text{dato}(4)+1]= [\text{dato}(5)] = 11$$

Finalmente se calcula la mediana

$$Me= ([\text{dato}(n/2)] + [\text{dato}(n/2)+1])/2$$

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

$$Me= 10$$

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

Como puede observarse, el estadístico encuentra el punto central del conjunto de datos sin considerar los valores que estos 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 sobre un producto vertidas 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 el 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, entonces su procedimiento de cálculo se realiza por medio de una interpolación mediante los siguientes pasos.

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

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 clase 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 siguiente relación:

$$Me = LIM + \frac{\frac{n}{2} - Fac}{f} T$$

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)	frec. acumulada (Fac)
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

Siguiendo los pasos establecidos anteriormente tenemos:

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

$$n/2 = 50 \text{ datos}/2 = 25$$

la mediana se ubica aproximadamente en el dato 25, el cual se localiza en el intervalo 4 (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.

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

$$Lim = 61.5$$

$$n = 50$$

$$Fac = 15$$

$$f = 16$$

$$T = 3$$

3. Calculamos la mediana con la fórmula establecida.

$$Me = 61.5 + \frac{\frac{50}{2} - 15}{16} \cdot 3$$

$$Me = 63.37$$

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

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

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

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

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

=MEDIAN (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 pueden clasificarse por sexo, estado civil, ocupación, etcétera. En este caso no tiene sentido la media o la mediana del sexo 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 siguientes ejemplos.

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 que están agrupados en una tabla de distribución de frecuencias, debemos seguir el siguiente procedimiento:

1. Localice la clase modal. La clase modal es aquella que presenta mayor frecuencia absoluta.
2. Una vez ubicada la clase modal, calculamos la moda por interpolación mediante el uso de la relación:

$$Mo = Lim + \frac{d_1}{d_1 + d_2} t$$

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 = longitud 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 clase modal es la cuarta, ya que ésta presenta la mayor frecuencia absoluta ($f_4 = 16$).
2. Se calcula la moda.

$$Lim = 61.5$$

$$d_1 = 8$$

$$d_2 = 4$$

$$t = 3$$

$$Mo = 61.5 + \frac{8}{8+4} 3$$

$$Mo = 63.49$$

El dato con mayor frecuencia en la tabla es 63.49, es decir, la moda, lo que indica que la mayoría de los trabajadores canadienses de la muestra analizada tienen una estatura aproximada de 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:

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

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

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

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

Cuartiles, deciles y percentiles

Además de las medidas de posición antes estudiadas (media, mediana, moda) existen otras que pueden ser más prácticas en ciertas situaciones. Estas medidas son los cuartiles, deciles y percentiles.

Los cuartiles, los deciles y los percentiles son en cierta forma una extensión de la mediana.

Cuartiles

Los cuartiles de una sucesión de datos ordenados son aquellos números que dividen la sucesión en cuatro partes porcentuales iguales. Hay tres cuartiles, denotados como: Q_1 , Q_2 y Q_3 . El segundo cuartil (Q_2), 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.

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

Los cuartiles se calculan con la fórmula:

$$Q_k = L_k + \frac{k(n/4) - F_k}{f_k} T$$

donde:

k = 1, 2, 3 el número de cuartil a calcular

L_k = límite real inferior de la clase del 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

F_k = Frecuencia acumulada de la clase que antecede a la clase del cuartil k

f_k = frecuencia absoluta de la clase del cuartil k

T = tamaño del intervalo de la clase del cuartil k

Por ejemplo, utilizando el cuadro 14 con 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 con ayuda de la frecuencia acumulada el dato 12.5

2. El dato buscado se ubica en la clase: 3.

3. Se determinan las variables de la fórmula de cálculo del cuartil.

$$\begin{aligned}
L_1 &= 58.5 \\
n &= 50 \\
F_1 &= 7 \\
f_1 &= 8 \\
T &= 3
\end{aligned}$$

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

$$Q_1 = 58.5 + \frac{1(50/4) - 7}{8} 3$$

$$Q_1 = 60.56$$

Esto último indica que el 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:

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

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

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: =QUARTILE(C2:C11,1) dando como resultado: 3, es decir, $Q_1=3$, lo que significa que el 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:

$$Dk = Lk + \frac{k(n/10) - Fk}{fk} T$$

donde:

k = 1,2,3,4,5,6,7,8,9 el número de decil a calcular

Lk = límite real inferior de la clase del 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, de la tabla 14 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 con ayuda de la frecuencia acumulada el dato 35

2. El dato buscado se ubica en la clase: 5.

3. Se determinan las variables de la fórmula de cálculo del decil.

$$L_7 = 64,5$$

$$n = 50$$

$$F_7 = 31$$

$$f_7 = 12$$

$$T = 3$$

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

$$D_7 = 64.5 + \frac{7(50/10) - 31}{12}$$

$$D_7 = 65.5$$

Este estadístico nos indica que el 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, etc.). Estas medidas son números que dividen una sucesión de datos ordenados en 100 partes porcentualmente iguales. Cuando los datos se agrupan en una tabla de frecuencias, los percentiles se calculan mediante la fórmula:

$$Pk = Lk + \frac{k(n/100) - Fk}{fk} T$$

donde:

k = 1,2,3,4,5,6,7, ... 99 el número de percentil a calcular

Lk = límite real inferior de la clase del 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 la tabla 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 con ayuda de la frecuencia acumulada el dato 40

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

$$L_{80} = 64.5$$

$$n = 50$$

$$F_{80} = 31$$

$$f_{80} = 12$$

$$T = 3$$

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

$$P_{80} = 65.5 + \frac{80 (50/100) - 31}{12} 3$$

$$P_{80} = 66.75$$

El percentil calculado indica que el 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:

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

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

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: =PERCENTILE (C2:C11,0.20) dando como resultado: 2.8 , es decir, $P_{20}=2.8$, lo que significa que el 20% de los datos son menores a 2.8.

Medidas de dispersión o de variabilidad

En el punto 1 de este capítulo indicamos que las medidas de posición estudiadas sólo nos permiten ubicar una tendencia central 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. Es por ello que todo análisis estadístico se inicia con el cálculo de las medidas de tendencia central o de posición 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 en un conjunto de datos. 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, puede estudiarse o determinarse 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:

- La amplitud o rango
- La desviación absoluta promedio
- La varianza

- La desviación estándar
- Dispersión relativa: el coeficiente de variación

Amplitud o rango

Es la más elemental de las medidas de variabilidad, y se le clasifica como una medida de distancia. La amplitud o rango 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, la amplitud se define como:

$$Amp = D_{my} - D_{me}$$

donde:

D_{my} = dato mayor en el conjunto de datos

D_{me} = dato menor 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, que al analizar únicamente dos datos (el mayor y el menor) no informa del comportamiento del resto. El segundo, que al no tomar en cuenta ninguna medida de posición, no informa nada acerca de cómo se comportan los datos respecto al centro.

Finalmente, el rango 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 a un centro.

En Excel este estadístico de dispersión puede ser calculado restando a la función que encuentra el valor máximo ($MAX()$) en un conjunto de datos, 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 o de la mediana. Las desviaciones se definen como la diferencia entre el estadístico de tendencia central usado (media 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, se define matemáticamente como:

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

donde:

$|di| = |X_i - \bar{X}|$ son los valores absolutos de las desviaciones de cada dato X_i respecto a la media.

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 a ésta es menor que la suma de las desviaciones respecto de cualquier otro valor. En la práctica las

desviaciones se toman respecto a la media. 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 siguiente conjunto de datos: 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.

$$|3 - 6| = 3$$

$$|3 - 6| = 3$$

$$|5 - 6| = 1$$

....

$$|8 - 6| = 2$$

$$|9 - 6| = 3$$

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

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

Si la distribución de estos datos es normal y simétrica, entonces el 68% de las observaciones quedan comprendidas entre la $\bar{X} - DAP$ y la $\bar{X} + DAP$, es decir, entre 4.2 y 7.8. En este conjunto prácticamente 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.

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 qué distancia de la media (o mediana) se halla en promedio cada observación.

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

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

donde:

- X_i = marca de clase del intervalo i
- \bar{X} = media
- f_i = frecuencia absoluta del intervalo i
- n = número total de datos en la tabla

En Excel, el cálculo de la desviación absoluta promedio para un conjunto de datos se obtiene a partir del uso de la función:

= AVEDEV (rango de datos) --- inglés

= DESVPROM (rango de datos) --- español

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: =AVEDEV(A1:A10) cuyo resultado es: 1.8.

Varianza

Es un estadístico que puede definirse como la media aritmética de las desviaciones respecto de la media elevadas 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 matemáticamente se calcula, para una muestra, a través de la siguiente relación:

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} = \frac{\sum_{i=1}^n (d_i)^2}{n} \quad \dots 2.I$$

donde:

$(d_i)^2 = (X_i - \bar{X})^2$ son las desviaciones al cuadrado de cada dato X_i respecto a la media de la muestra

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 es:

$$S^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N} = \frac{\sum_{i=1}^N (d_i)^2}{N} \quad \dots 2.II$$

donde:

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

N es el número total de datos de la población

La ecuación 2.I por lo tanto 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.

$$S^2 = \frac{\sum_{i=1}^k (X_i - \bar{X})^2 f_i}{n} = \frac{\sum_{i=1}^k (d_i)^2 f_i}{n} \quad \dots 2.III$$

$$S^2 = \frac{\sum_{i=1}^k (X_i - \mu)^2 f_i}{N} = \frac{\sum_{i=1}^k (d_i)^2 f_i}{N} \quad \dots 2.IV$$

donde:

X_i es la marca de clase del intervalo i

f_i es la frecuencia absoluta del intervalo i

\bar{X} es la media de la muestra

μ es la media de la población

n es el número total de datos en la muestra

N es el número total de datos en la población

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

Varianza muestral

$$= \text{VAR (rango de datos de la muestra)}$$

Varianza poblacional

$$= \text{VARP (rango de datos de la población)}$$

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 2.1:

1. Calculamos la media de los datos.

$$\bar{X} = 6$$

2. Calculamos las desviaciones elevadas al cuadrado.

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

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

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

....

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

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

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

$$S^2 = \frac{9 + 9 + 1 + 1 + 1 + 1 + 1 + 1 + 4 + 4 + 9}{10} = \frac{40}{10}$$

$$S^2 = 4 \text{ unidades}^2$$

Esta medida nos indica que la dispersión de los datos respecto a la media muestral son cuatro unidades cuadradas. Físicamente el estadístico no nos dice mucho acerca de la variabilidad de los datos pues mientras éstos se miden 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. Esta diferencia con respecto al cálculo manual es a causa de que en la fórmula en Excel el denominador se cambia de n a $n-1$, es decir, $40/9=4.444$ calculándose así lo que se conoce con el nombre de *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 éste mediante el cálculo de su raíz cuadrada; es decir, la desviación estándar, el nuevo estadístico, no es mas que la raíz cuadrada de la varianza, y por lo tanto, las unidades en las que se mide serán las mismas que tienen las observaciones.

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 a la media. Matemáticamente la desviación estándar para una muestra y una población se expresan en las ecuaciones 2.V y 2.VI respectivamente:

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (di)^2}{n}} \quad \dots 2.V$$

$$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}} = \sqrt{\frac{\sum_{i=1}^N (d_i)^2}{N}} \quad \dots 2.VI$$

donde:

X_i es la observación i en la muestra o en la población

\bar{X} es la media de la muestra

μ es la media de la población

n es el número total de datos en la muestra

N es el número total de datos en la población

d_i son las desviaciones de cada observación en relación con 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.

$$S = \sqrt{\frac{\sum_{i=1}^k (X_i - \bar{X})^2 f_i}{n}} = \sqrt{\frac{\sum_{i=1}^k (d_i)^2 f_i}{n}} \quad \dots 2.VII$$

$$S = \sqrt{\frac{\sum_{i=1}^k (X_i - \mu)^2 f_i}{N}} = \sqrt{\frac{\sum_{i=1}^k (d_i)^2 f_i}{N}} \quad \dots 2.VIII$$

donde:

X_i es la marca de clase del intervalo i
 f_i es la frecuencia absoluta del intervalo i
 \bar{X} es la media de la muestra
 μ es la media de la población
 n es el número total de datos en la muestra
 N es el número total de datos en la población

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 no agrupada en una tabla de distribución se utiliza la ecuación 2.V:

1. Calculamos la media de los datos.

$$\bar{X} = 6$$

2. Calculamos las desviaciones elevadas al cuadrado.

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

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

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

....

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

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

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

$$S^2 = \frac{9 + 9 + 1 + 1 + 1 + 1 + 1 + 1 + 4 + 4 + 9}{10} = \frac{40}{10}$$

$$S^2 = 4 \text{ unidades}^2$$

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

$$\sqrt{s^2} = \sqrt{4 \text{ unidades}^2}$$

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

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. La segunda, si la distribución de estos datos es normal y simétrica, entonces el 68% de las observaciones deben quedar comprendidas entre la $\bar{X} - S$ y la $\bar{X} + S$, es decir, entre 4 y 8. Para nuestro ejemplo, el 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 la 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 a la media, 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

$$=STDEV(\text{rango de datos de la muestra})$$

Desviación estándar poblacional

$$=STDEVP(\text{rango de datos de la población})$$

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: =STDEV(A1:A10) cuyo resultado es: 2.1.

Dispersión relativa: el coeficiente de variación

Como ya indicamos previamente la desviación estándar es una medida de variación absoluta que no permite concluir qué tan grande o qué tan pequeña es la dispersión de los datos, sin embargo combinada con la media 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. Esta relación expresa la desviación estándar como porcentaje de la media, de ahí que sus unidades se midan en “por ciento”. Matemáticamente puede expresarse como:

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

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 bien, distinta unidad de medida. 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.
4. Determinar si cierta media es consistente con cierta varianza.

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

	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 en el módulo que el grupo SB05.

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

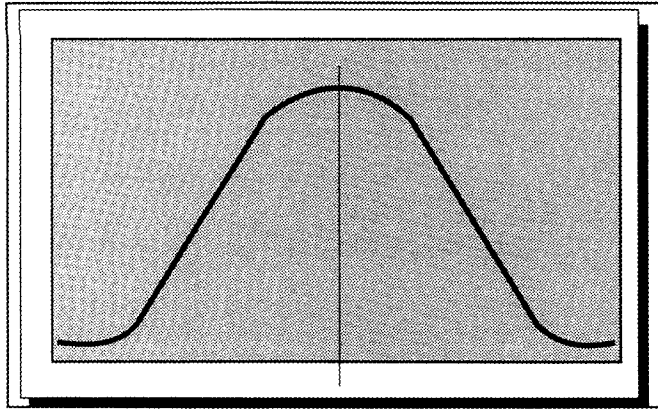
$$=(STDEV (\text{rango de datos}) /AVERAGE (\text{rango de datos})) * 100$$

El rango de datos para ambas funciones es el mismo.

Medida de sesgo (asimetría)

Las curvas que pueden utilizarse para representar las observaciones de un conjunto de datos pueden ser simétricas o asimétricas (sesgadas). Las curvas

Figura 6



simétricas como la que se muestra en la figura 6, son tales 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 la imagen a espejo de la otra.

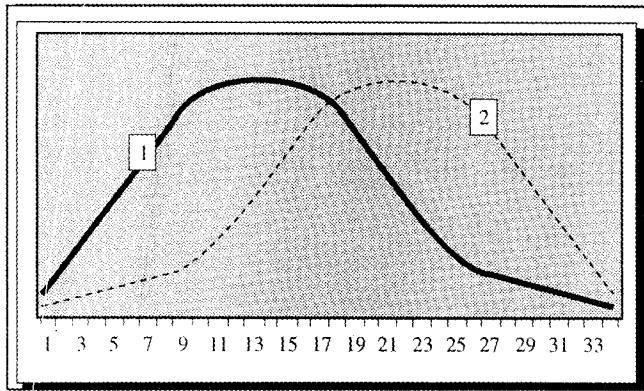
Por otro lado, las curvas en la figura 7 son curvas sesgadas. Son sesgadas porque los valores de sus distribuciones de frecuencias se concentran en el extremo inferior o en el extremo superior de la escala de medición situada sobre el eje horizontal. Los valores no tienen una distribución igual. La curva 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) dado 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)*.

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

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

Figura 7



donde:

X_i es la observación (dato) i

n es el número total de datos en la muestra

\bar{X} es la media de la muestra

S es la 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:

$$CS = \frac{\sum_{i=1}^k f_i (X_i - \bar{X})^3 / n}{S^3}$$

donde:

X_i es la marca de clase del intervalo i

f_i es la frecuencia absoluta del intervalo i

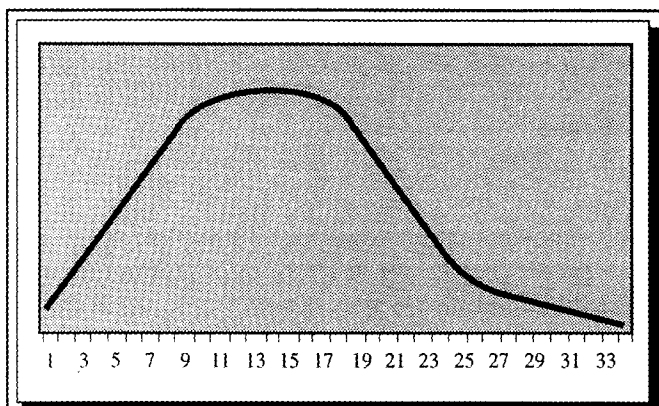
\bar{X} es la media de la muestra

S es la desviación estándar de la muestra
 n es el número total de datos en la muestra

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

- Si $CS=0$, entonces, los datos (la curva) se distribuye de manera simétrica, como la que se muestra en la figura 6.
- Si $CS>0$, entonces, los datos (la curva) son sesgados a la derecha, como se muestra en la figura 8.

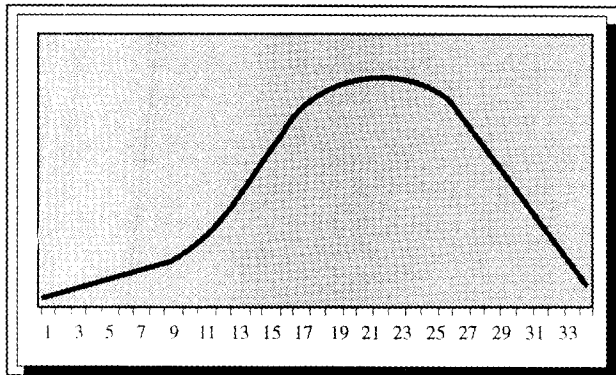
Figura 8



- Si $CS<0$, entonces, los datos (la curva) son sesgados a la izquierda, como se muestra en la figura 9.

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

Figura 9



Medida de curtosis

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 puede observarse en la figura 10. 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, es decir, tienen diferentes grados de curtosis.

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 6, recibe el nombre de curva *mesocúrtica*. Una curva que tenga más pico, como la mostrada en la figura 11, recibe el nombre de curva *leptocúrtica*, y una curva que tenga menos pico, como la mostrada en la figura 12, recibe el nombre de curva *platocúrtica*.

Figura 10

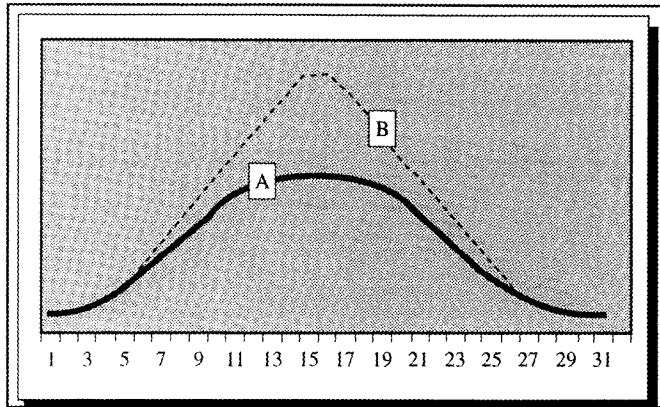
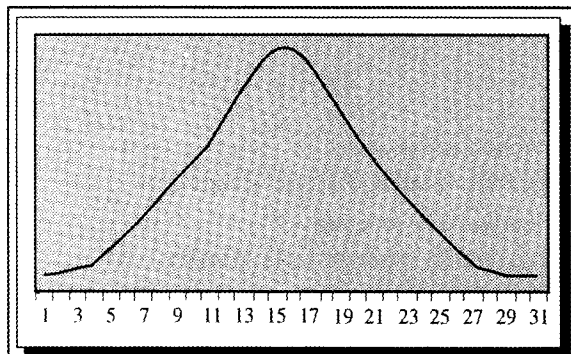


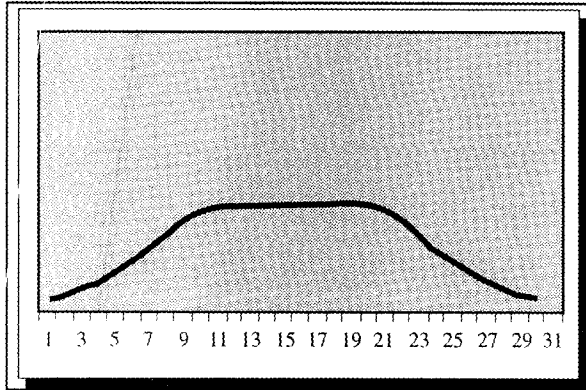
Figura 11



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

Para datos no agrupados.

Figura 12



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

donde:

X_i es la observación (dato) i

n es el número total de datos en la muestra

\bar{X} es la media de la muestra

S es la desviación estándar de la muestra

Para datos agrupados en una tabla de distribución.

$$CC = \frac{\sum_{i=1}^k f_i (X_i - \bar{X})^4 / n}{S^4}$$

donde:

X_i es la marca de clase del intervalo i

f_i es la frecuencia absoluta del intervalo i

\bar{X} es la media de la muestra

S es la desviación estándar de la muestra

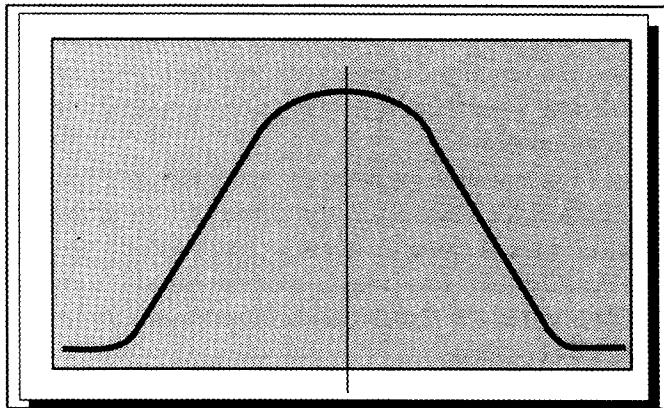
n es el número total de datos en la muestra

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 13.

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

Figura 13



- 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 11.
- Si $CC < 3$, entonces los datos (la curva) se presentan más aplanados que los de la curva normal, como se muestra en la figura 12.

En Excel, el coeficiente de curtosis se calcula para datos no agrupados mediante la función: =KURT (*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: =KURT (3,4,5,2,3,4,5,6,4,7) o bien, si los datos ocupan las celdas A1 a A10: =KURT (A1:A10). El resultado en ambos casos es: $CC = -0.1518$, mostrando con ello que los datos forman una curva platocúrtica.

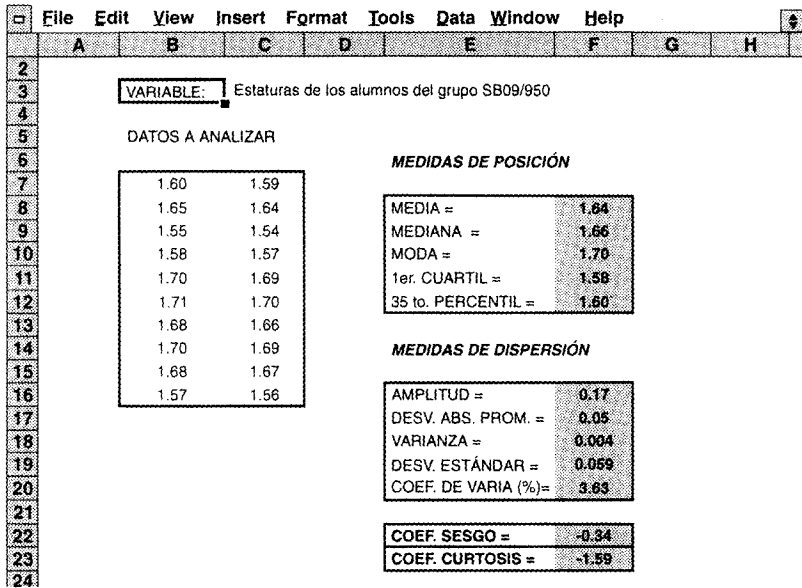
Un ejemplo del uso de las funciones estadísticas en Excel

El cálculo de medidas de estadística descriptiva con las hojas electrónicas es muy simple, y 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 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 14 se muestran los datos de estatura de 20 alumnos del grupo SB09 del trimestre 950. 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 celda 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.

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

Figura 14



En la siguiente tabla se muestra el contenido por celda del modelo de la figura 14 y las funciones asociada para la versión 5.0 de la hoja electrónica Excel.

<i>Celda</i>	<i>Función</i>
F8	= AVERAGE(B7:C16)
F9	= MEDIAN(B7:C16)
F10	= MODE(B7:C16)
F11	= QUARTILE(B7:C16,1)
F12	= PERCENTILE(B7:C16,0,35)
F16	= MAX(B7:C16)-MIN(B7:C16)
F17	=AVEDEV(B7:C16)
F18	= VAR(B7:C16)
F19	= STDE(B7:C16)
F20	= (F19/F8)*100
F22	= SKEW(B7:C16)
F23	= KURT(B7:C16)

Resumen de fórmulas estadísticas y funciones en Excel

1. Medidas de posición o tendencia central

Estadístico	Fórmula	Función en Excel (ejemplo)
Media aritmética muestral para datos no agrupados	$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$	=AVERAGE(rango de datos) =AVERAGE(A1:A10)
Media aritmética poblacional	$\mu = \frac{\sum_{i=1}^N X_i}{N}$	=AVERAGE(rango de datos) =AVERAGE(A1:A10 000)
Media aritmética muestral o poblacional para datos agrupados	$\bar{X}, m = \frac{\sum_{i=1}^k f_k X_k}{n}$	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	$MG = \sqrt[n]{X_1 \cdot X_2 \cdot X_3 \cdot \dots \cdot X_n}$	=GEOMEAN(datos 1, datos 2, dato 3,...,dato _n) =GEOMEAN(A1:A10)

Estadístico	Fórmula	Función en Excel (ejemplo)
Mediana para número de datos impar no agrupados	$Ume = (n+1)/2$ Ubica a la mediana con datos ordenados	= <i>MEDIAN</i> (rango del conjunto de datos) = <i>MEDIAN</i> (A1:A10)
Mediana para número de datos par no agrupados	$Me = (Dato(n/2) + Dato(n/2)+1)/2$	= <i>MEDIAN</i> (rango del conjunto de datos) = <i>MEDIAN</i> (A1:A10)
Mediana para datos agrupados	$Me = Lim + \frac{\frac{n}{2} - Fac}{f} T$	No existe. Se puede construir con una macroinstrucción
Moda para datos no agrupados	Dato con mayor frecuencia en el conjunto	= <i>MODE</i> (rango de datos) = <i>MODE</i> (A1:A10)
Moda para datos agrupados	$Mo = Lim + \frac{d_1}{d_1 + d_2} t$	No existe. Se puede construir con una macroinstrucción
Cuartiles	$Qk = Lk + \frac{k(n/4) - Fk}{fk} T$ Datos agrupados	Datos no agrupados = <i>QUARTILE</i> (rango de datos, número de cuartil deseado) = <i>QUARTILE</i> (A1:A10,1)
Deciles	$Dk = Lk + \frac{k(n/10) - Fk}{fk} T$ Datos agrupados	No existe. Se puede usar la función <i>PERCENTIL</i> en datos no agrupados
Percentiles	$Pk = Lk + \frac{k(n/100) - Fk}{fk} T$ Datos agrupados	Datos no agrupados = <i>PERCENTILE</i> (rango de datos, número de percentil deseado en decimales) = <i>PERCENTILE</i> (A1:A10, 0.33)

2. Medidas de dispersión o variabilidad

Estadístico	Fórmula	Función en Excel (ejemplo)
Amplitud o rango	$Amp = Dmy - Dme$ <p>Datos no agrupados: diferencia entre el valor máximo y el valor mínimo</p> <p>Datos agrupados: límite real superior de la última clase menos el límite real inferior de la primera clase (agrupación ascendente)</p>	<p>Datos no agrupados</p> $=MAX(\text{rango de datos}) - MIN(\text{rango de datos})$ $=MAX(A1:A7) - MIN(A1:A7)$
Desviación absoluta Promedio para datos no agrupados	$DAP = \frac{\sum_{i=1}^n Xi - \bar{X} }{n}$	$=AVEDEV(\text{rango de datos})$ $=AVEDEV(A1:A10)$
Desviación absoluta Promedio para datos agrupados	$DAP = \frac{\sum_{i=1}^k Xi - \bar{X} fi}{n}$	No existe. Se puede construir con una macroinstrucción.
Varianza para una muestra con datos no agrupados	$S^2 = \frac{\sum_{i=1}^n (Xi - \bar{X})^2}{n}$	$=VAR(\text{rango de datos})$ $=VAR(A1:A10)$
Varianza para una población con datos no agrupados	$S^2 = \frac{\sum_{i=1}^N (Xi - \mu)^2}{N}$	$=VARP(\text{rango de datos})$ $=VARP(A1:A16000)$
Varianza para una muestra con datos agrupados	$S^2 = \frac{\sum_{i=1}^k (Xi - \bar{X})^2 fi}{n}$	No existe. Se puede construir con una macroinstrucción

Estadístico	Fórmula	Función en Excel (ejemplo)
Varianza para una población con datos agrupados	$S^2 = \frac{\sum_{i=1}^k (X_i - \mu)^2 f_i}{N}$	No existe. Se puede construir con una macroinstrucción
Desviación estándar para una muestra con datos no agrupados	$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}}$	=STDEV(rango de datos) =STDEV(A1:A10)
Desviación estándar para una población con datos no agrupados	$S = \sqrt{\frac{\sum_{i=1}^N (X_i - \mu)^2}{N}}$	=STDEVP(rango de datos) =STDEVP(A1:A16 000)
Desviación estándar para una muestra con datos agrupados	$S = \sqrt{\frac{\sum_{i=1}^k (X_i - \bar{X})^2 f_i}{n}}$	No existe. Se puede construir con una macroinstrucción
Desviación estándar para una población con datos agrupados	$S = \sqrt{\frac{\sum_{i=1}^k (X_i - \mu)^2 f_i}{N}}$	No existe. Se puede construir con una macroinstrucción

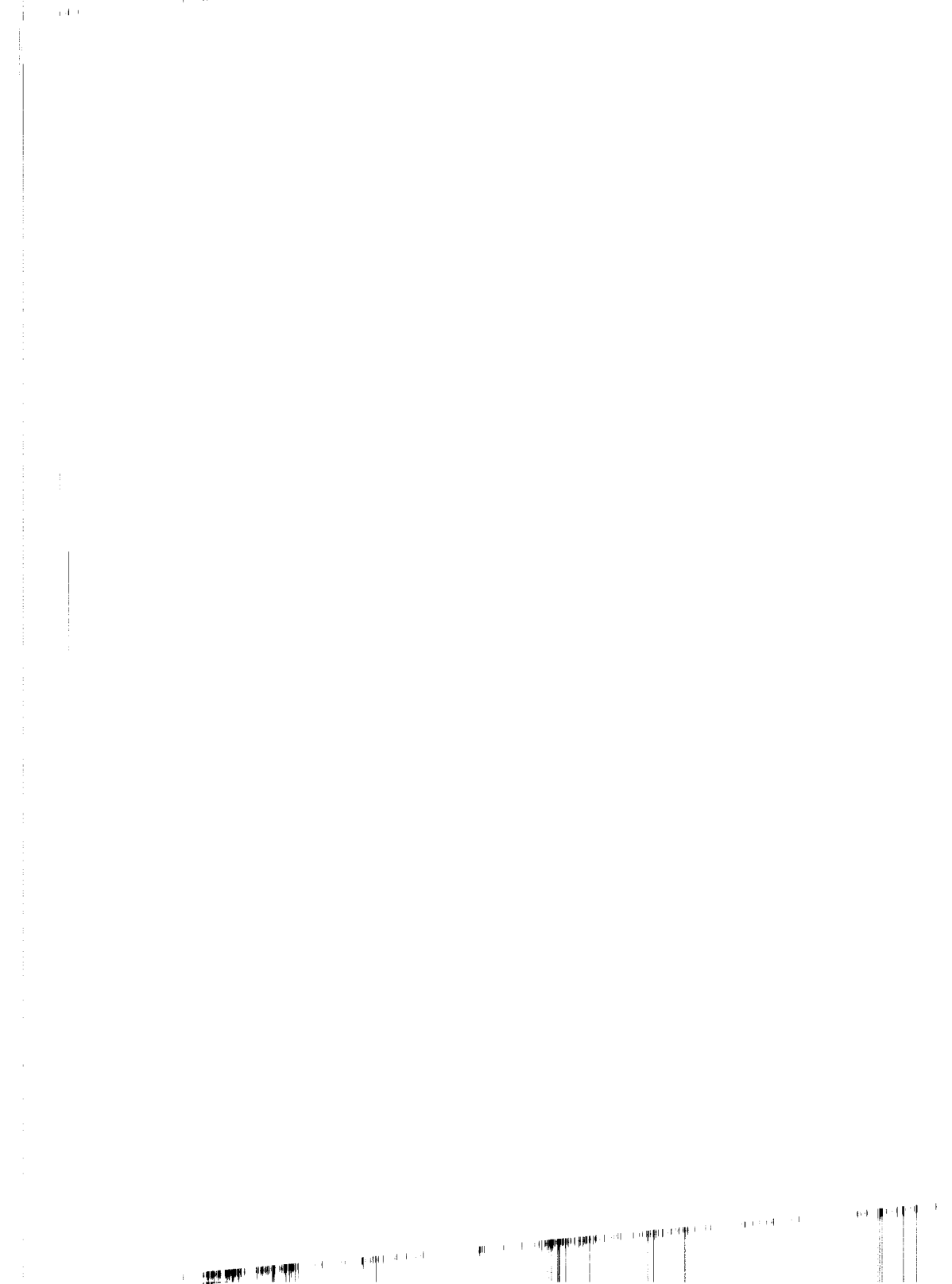
Estadístico	Fórmula	Función en Excel (ejemplo)
Dispersión relativa: coeficiente de variación	$CV = \frac{S}{\bar{X}} 100\%$	=(STDEV(rango de datos) /AVERAGE(rango de datos))* 100 =(STDEV(A1:A10)/AVE- RAGE(A1:A10))*100

3. Medida de sesgo (asimetría)

Estadístico	Fórmula	Función en Excel (ejemplo)
Coefficiente de sesgo(de asimetría) para datos no agrupados	$CS = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left[\frac{Xi - \bar{X}}{S} \right]^3$	=SKEW(rango de datos) =SKEW(A1:A10)
Coefficiente de sesgo(de asimetría) para datos agrupados	$CS = \frac{\sum_{i=1}^k fi (Xi - \bar{X})^3 / n}{S^3}$	No existe. Se puede construir con una macroinstrucción

4. Medida de curtosis

Estadístico	Fórmula	Función en Excel (ejemplo)
Coefficiente de curtosis para datos no agrupados	$CC = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left[\frac{(Xi - \bar{X})^4}{S^4} \right] \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$	=KURT(rango de datos) =KURT(A1:A10)
Coefficiente de curtosis para datos agrupados	$CC = \frac{\sum_{i=1}^k fi (Xi - \bar{X})^4 / n}{S^4}$	No existe. Se puede construir con una macroinstrucción



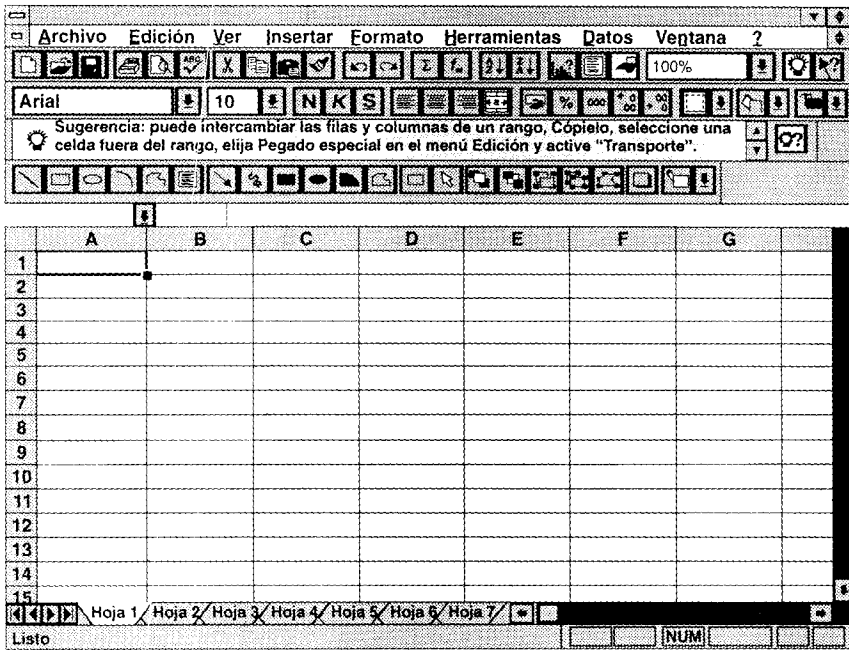
ANEXO

Elementos básicos de la hoja electrónica Excel

Introducción

Una hoja de cálculo es un programa de computadora que en la pantalla tiene la apariencia de una hoja cuadrículada; a cada hilera vertical de cuadros le llamaremos *columna* y a cada hilera horizontal le llamaremos *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 siempre aparece 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, así como su contenido.

Cada celda puede contener un dato, en forma de: un número, un mensaje, una fórmula, una función, o bien, una macroinstrucción. 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, ya que podemos hacer desde cálculos simples, como una suma, hasta cálculos complejos: estadísticas, funciones financieras, 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*, ejemplificando esto: 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 puede observarse 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*, y algunas funciones especiales, pueden, inclusive, tener valores lógicos.

I. 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:

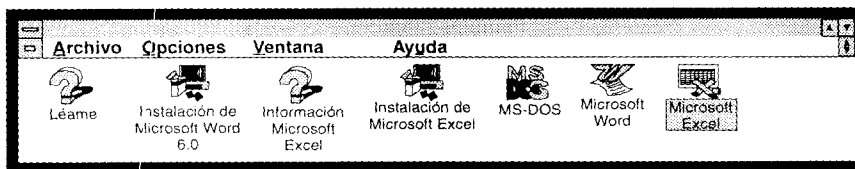
Tecla	Función
HOME	ubica el cursor al inicio del renglón (en la columna A)
END y (→)	ubica el cursor en la última celda ocupada en el renglón
END y (↓)	ubica el cursor en la última celda ocupada en la columna
END y (↑)	ubica el cursor en la primera celda ocupada en la columna
PgUp	ubica el cursor una página anterior
PgDn	ubica el cursor una página posterior
F5	ubica el cursor en una celda especificada (debemos indicar la celda deseada)
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 al final de la hoja (celda IV16384)

En la figura 1, dentro del área de trabajo, 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 (☞) lo lograremos con los siguientes pasos: colocamos el cursor del mouse sobre el cuadro que se encuentra en las barras, se oprime el botón derecho del mouse sin soltarlo y deslizar el cursor sobre la barra, luego soltamos el botón y automáticamente estamos ubicados en otra sección de la hoja. Aparecerá en la pantalla un reloj (⌚) que indica que Excel está trabajando. En múltiples ocasiones el reloj aparece en la pantalla, indica que esperemos unos segundos para que podamos seguir tecleando.

II. Principales componentes de la ventana de Excel

En la hoja electrónica Excel se muestran cinco barras de *Menú* o de *Estado*.

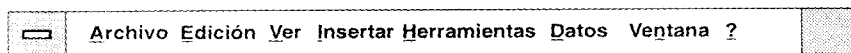
1. Barra del menú del sistema



En ellas aparece el nombre de la aplicación y un cuadro con una línea en forma horizontal; se puede accionar con la combinación de la tecla *ALT+espacio*, y posteriormente nos muestra un submenú con el cual podemos minimizar la pantalla, maximizarla, moverla, etcétera.

2. Barra del menú principal

Muestra los comandos propios de *Excel*: *archivo*, *edición*, *ver*, etcétera.



3. Barra de herramienta



Son instrucciones que le podemos dar al paquete por medio de iconos. Para ejecutar una instrucción nos ubicamos con el ratón o mouse en alguno de estos pequeños dibujos y oprimimos el clic derecho, con lo cual se ejecutará la instrucción de inmediato, es decir, son comandos rápidos representados por medio de botones; los dibujos indican de qué tipo es la instrucción asociada a ellos.

4. Barra de edición



Muestra el contenido de la celda y su ubicación, así como otras características.

5. Barra de estado



Nos indica el tipo de información; si está lista para trabajar, si se está editando una celda, etcétera. Indica también lo que se hace en la hoja activa y las teclas que están activadas (*Caps*, *num*, *ins*, etcétera), así como la situación de la celda en la hoja.

III. Submenús de Excel

A continuación se describen brevemente los principales submenús de Excel, así como sus principales comandos asociados.

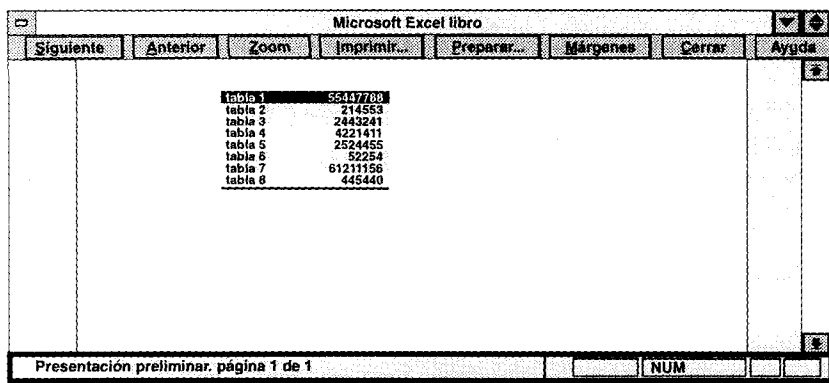
I. Archivo

Permite el manejo de los archivos en la hoja electrónica. Sus subcomandos son:

<i>Nuevo</i>	Crear una nueva hoja (archivo)
<i>Abrir</i>	Abrir una hoja ya existente
<i>Cerrar</i>	Cierra la hoja de trabajo activa
<i>Guardar</i>	Guarda el archivo con el mismo nombre; es decir, lo sobrescribe
<i>Guardar como</i>	Guarda el archivo en uso por primera vez, o guarda las modificaciones con otro nombre
<i>Guardar área de trabajo</i>	Guarda el grupo de archivos definidos como grupo de trabajo
<i>Presentación preliminar</i>	Reproduce en la pantalla el documento como si estuviera impreso
<i>Imprimir</i>	Para indicar las características de impresión de la hoja, márgenes, tamaño, etcétera
<i>Salir</i>	Para salir de la aplicación

Ejemplo:

En el caso que eligiésemos la opción de *presentación preliminar*, en la cual contamos para consulta con un pequeño menú, la pantalla mostrada es la siguiente:



2. Edición

Permite realizar operaciones de trabajo con los datos que se encuentran en las celdas de la hoja electrónica.

<i>Deshacer</i>	Anula la última acción ejecutada y el archivo queda sin cambio
<i>Cortar</i>	“Extrae” el rango de celdas seleccionado de su posición actual, lo coloca dentro de portapapeles para después trasladarlo a otra sección de la hoja activa mediante el comando <i>Pegar</i>
<i>Copiar</i>	“Registra” en el portapapeles el rango de celdas seleccionado para posteriormente copiarlo en la sección deseada de la hoja mediante el comando <i>Pegar</i> o al oprimir la tecla <i>Enter</i>
<i>Pegar</i>	Coloca la información que se encuentra en el portapapeles a partir de la celda activa
<i>Borrar</i>	Elimina columnas, renglón o celdas seleccionadas
<i>Pegado especial</i>	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:

1. Aritméticos

Realizan las operaciones matemáticas más comunes.

- + *suma*
- *resta*
- * *multiplicación*
- / *división*
- % *porcentaje*
- ^ *exponenciación*

2. De texto

El único operador es el & (ampersand) que nos permite unir varios textos o valores alfanuméricos.

3. 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*

4. 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

<i>Celdas</i>	Agrega celdas a la hoja electrónica, desplazando a otras celdas los datos o información
<i>Filas</i>	Agrega uno o varios renglones a la hoja electrónica, desplazando a otras filas los datos o información
<i>Columnas</i>	Agrega una o varias columnas a la hoja electrónica, desplazando a otras columnas los datos o información
<i>Función</i>	Introduce una función propia de Excel a la celda activa Presenta la sintaxis requerida para su correcta operación
<i>Nombre</i>	Crea, elimina o cambia nombres a un rango de celdas seleccionado
<i>Notas</i>	Agrega, borra o edita notas en las celdas de la hoja.
<i>Hoja de cálculo</i>	Agrega una nueva hoja de cálculo al archivo de trabajo.

4. Formato

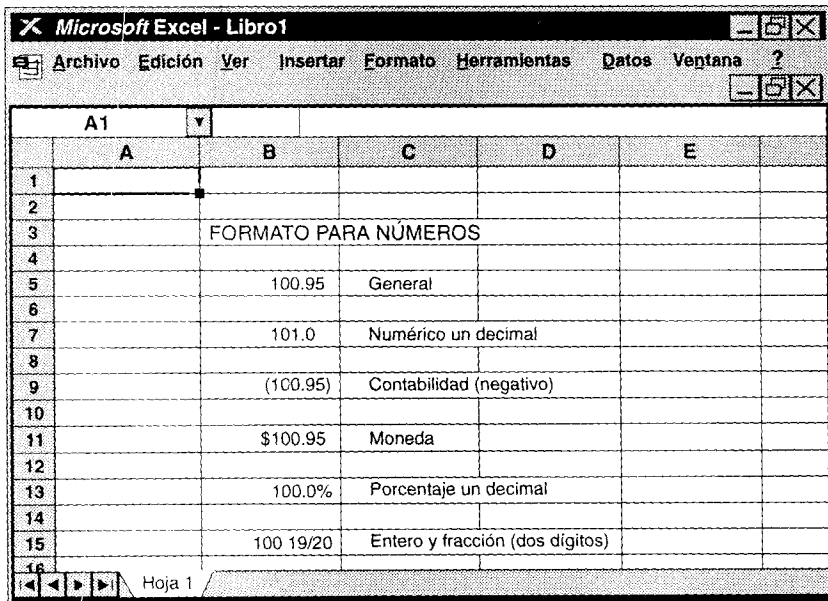
Estos comandos permiten asignar a los datos o información en una celda o en un conjunto de celdas un formato determinado, es decir, determina tamaño de la celda, su altura, el tipo de letra, el tamaño de la letra, el color de la letra, el color del fondo de la celda, etcétera.

<i>Celda</i>	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
--------------	--

<i>Fila</i>	Permite modificar tanto la altura, como el ancho de la fila, también restablece la altura de la fila al alto estándar
<i>Columna</i>	Ajusta el ancho de las columnas seleccionadas
<i>Hoja</i>	Cambia nombre y oculta hojas seleccionadas
<i>Autoformato</i>	Da formato de tabla a un rango de datos
<i>Modelo</i>	Aplica o define un modelo de celda
<i>Objeto</i>	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.

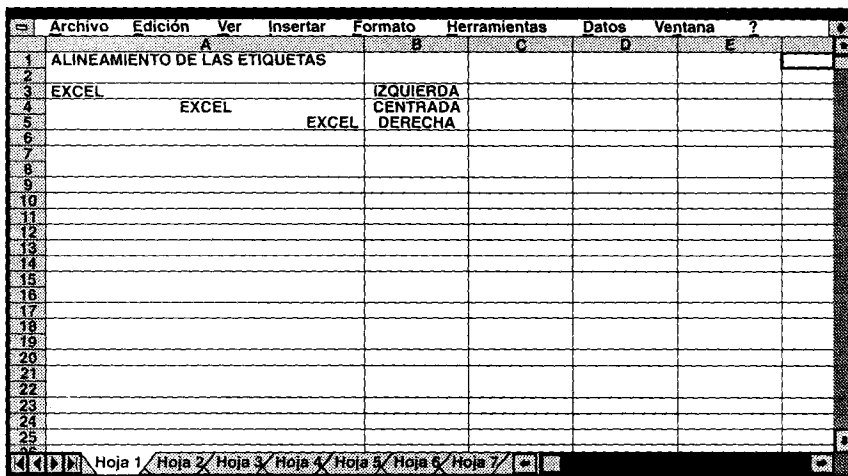
Figura 2



Como podemos observar en la figura 3, 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) y justificado.

Figura 3

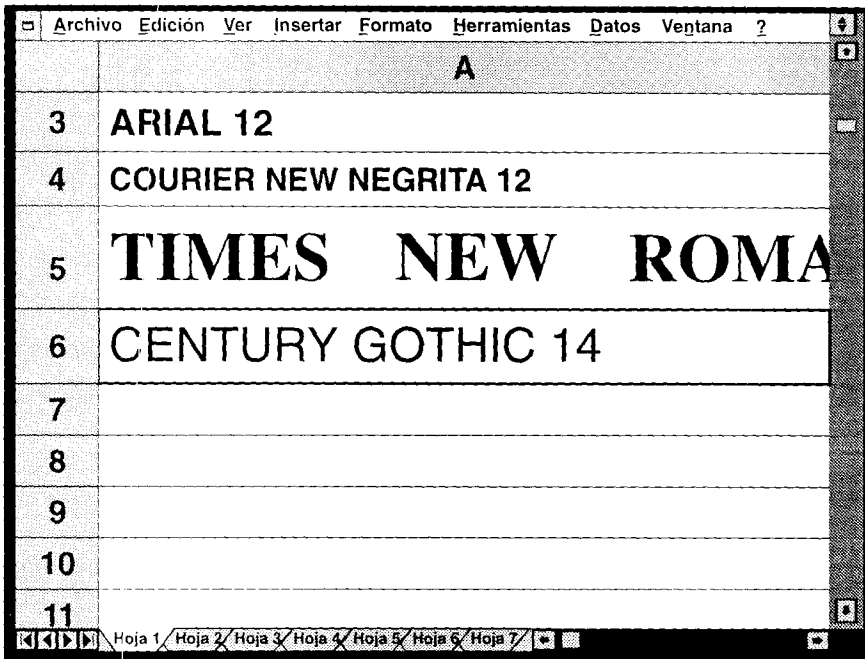


En el ejemplo de la figura 4 vemos que los renglones tienen diversas alturas, esto es a causa de que los tipos de letra presentados los modifican de acuerdo a su tamaño.

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 se 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

Figura 4



columna. Cabe mencionar que esto dependerá del tamaño que se le asigne a la columna.

5. Ayuda

Es un menú que permite consultar el uso de comandos y procedimientos de operación del paquete en forma rápida por medio de ejemplos.

Temas de ayuda de MS Excel

Despliega el contenido de Excel para consultar un tema

Asistente para ayuda

Muestra una introducción completa al Excel 5

Despliega lecciones rápidas para que el usuario se familiarice con Excel.

Ayuda Lotus 123
Acerca de MS Excel

Muestra las equivalencias en los comandos con Excel
Despliega información sobre la versión de Excel, memoria total y memoria disponible

IV. Gráficas en Excel

Excel cuenta con una gran variedad de gráficos, entre ellos se tiene:

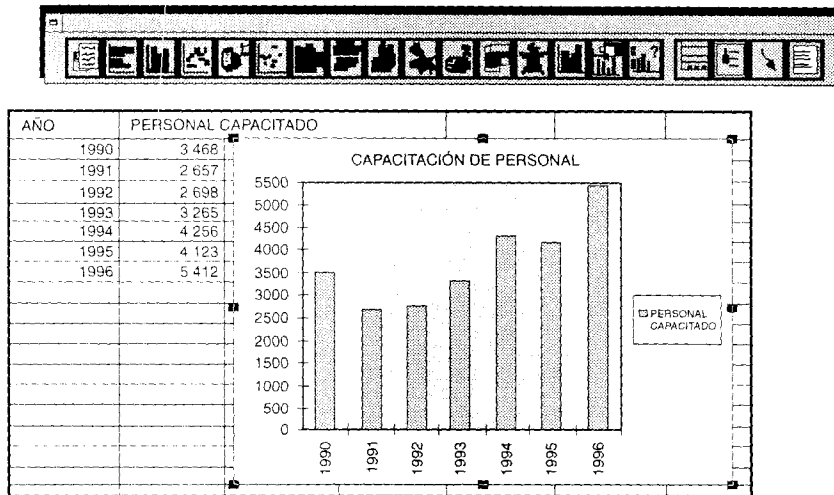
- Área (en dos y tres planos)
- Barras-columnas (en dos y tres planos)
- Líneas (en dos y tres planos)
- De sectores-pie (en dos y tres planos)
- De tipo *XY*
- Combinadas (barras y líneas, etc.)

Para construir una gráfica en Excel, en el caso de contar con ratón o mouse, será necesario seguir los siguientes pasos. En caso contrario, la construcción se realiza con otro procedimiento que se detalla más adelante.

Para elaborar un gráfico contamos con diversos procesos, pero el más simple consiste en seleccionar las series que forman el gráfico (sombrearlas), y posteriormente dar un clic sobre el icono de gráficas del menú de herramientas, con ello el cursor cambia a un formato de cruz. Con el cursor se construye un rectángulo en la hoja, en la sección donde se desea ubicar la gráfica, esto se logra al oprimir en forma sostenida el clic del mouse. Así, Excel pregunta si la primera serie en columna o renglón es el eje *X* o no, ello deberá indicarse seleccionando la opción deseada y posteriormente se obtendrá la gráfica. Si se desea modificar la gráfica se oprime doble clic para obtener el menú de gráficas que se muestra a continuación:

Este menú aparece cuando ya se ha construido una gráfica. La barra nos facilita la construcción o modificación de la gráfica.

Excel permite trabajar 17 tipos de gráficas de las cuales 10 se trabajan en 2 ejes y 7 en tres dimensiones (*3-D*).



Series de datos

Cada serie de datos estará representada en la gráfica por una línea, barra, pedazo de pie, etcétera, dependiendo del tipo de gráfica elegido.

Establecer preferido

Asimismo podemos indicar al paquete cuál es nuestro tipo preferido de gráfica (Preferred), para que éste construya siempre este tipo de gráfico con los datos indicados; o bien llamar un formato de gráfico preferido (Preferred).

Para agregar y modificar diversas características de las gráficas se utilizan los siguientes comandos que se encuentran dentro del *Menú de gráficas*:

Agregar texto: permite agregar y modificar el título de las gráficas, etiquetas en los ejes X y Y, así como en los datos en una serie específica.

Editar línea: agrega o elimina una flecha indicativa a la gráfica.

Agregar leyenda: agrega o modifica las etiquetas en las series de las gráficas.

Ejes: permite modificar la presentación de los ejes de la gráfica.

Editar series: permite agregar, modificar o eliminar series en una gráfica, así como sus etiquetas en el gráfico (opción name).

En el caso de que no tengamos mouse entonces se elige *Galería* y podremos escoger el tipo de gráfica deseada.

En la opción *3-D View* tenemos algunos puntos interesantes para visualizar la gráfica, ya que podemos seleccionar la elevación, perspectiva y rotación de nuestra gráfica.

Para elaborar una gráfica por medio de comandos sin utilizar el mouse se deberá seguir el proceso que se describe a continuación:

+ Lo primero que tenemos que hacer es seleccionar el conjunto de datos (rango de datos) para graficar.

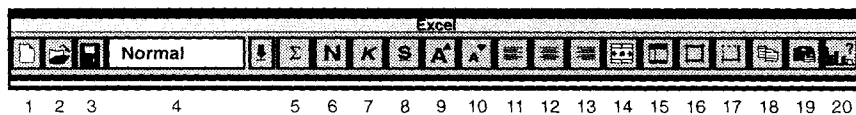
+ En el menú *Archivo* se selecciona *Nuevo*, aparece un cuadro diálogo y se elige *Gráfica*. Aparece otro cuadro y se elige *Primera serie de datos*.

+ Posteriormente aparece un nuevo menú, el *Menú de gráficas* y una barra de herramientas para construirla, la cual sólo puede ser accionada con el mouse.

V. Barras de herramientas

Dentro del Menú *Opciones*, *Barra de herramientas*, se puede indicar a Excel cuáles barras de herramientas desea visualizar el usuario dentro de 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 la función asociada a cada icono.

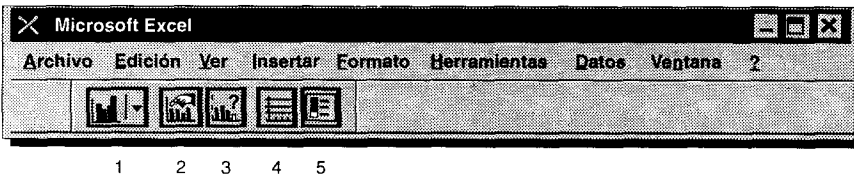
Barra estándar



- 1 *Nuevo*. Crea una nueva hoja de trabajo.
- 2 *Abrir*. Despliega una caja de diálogo para abrir archivos existentes.
- 3 *Guardar*. Guarda cambios hechos al documento activo.
- 4 *Formato*. Selecciona el tipo de formato para usar en la hoja.
- 5 *Suma*. Inserta la función *Suma* y un rango de suma propuesto, basado en los datos anteriores o a la izquierda de la celda activa.
- 6 *Negritas*. Pone negritas en la celda o celdas seleccionadas, caja de texto o texto en gráficas.
- 7 *Cursiva*. Pone formato itálico en la celda o celdas seleccionadas, en la caja de texto seleccionada, así como en el botón o texto en gráfica seleccionado.
- 8 *Subrayado*. Subraya la información seleccionada.
- 9 *Aumenta el tamaño de fuente*. Incrementa el tamaño del *font* del texto seleccionado al próximo tamaño mayor.
- 10 *Disminuye el tamaño de fuente*. Reduce el tamaño del *font* del texto seleccionado al próximo tamaño menor.
- 11 *Alineación izquierda*. Alinea hacia la izquierda el contenido de celdas, caja de textos y textos en gráficas seleccionados.
- 12 *Centrado*. Centra el contenido de una celda, cajas de textos y textos en gráficas seleccionados.
- 13 *Alineación derecha*. Alinea hacia la derecha el contenido en celdas, caja de texto y texto en gráficas seleccionados.
- 14 *Centrar en varias columnas*. Centra el texto de una celda dentro del rango de columnas seleccionadas en forma horizontal.
- 15 *Autoformato*. Aplica el último formato de tabla aplicado con el comando *Auto Format* en el menú de formato. Para aplicar el próximo formato en la lista, presione *Shift* y el botón del ratón en *Tools*.
- 16 *Línea de borde*. Añade un borde o margen en los extremos de las celdas seleccionadas.
- 17 *Botón de borde*. Añade o borra un borde o margen en la parte inferior de cada celda seleccionada.
- 18 *Copiar*. Copia la selección de celdas al portapapeles.
- 19 *Pegar formato*. Inserta sólo los formatos de las celdas copiadas en las celdas seleccionadas.

- 20 *Asistente para gráficos.* Activa la gráfica mágica para editar la gráfica seleccionada o crear una gráfica en particular a través de cinco etapas que se presentan en cuadros diálogo.

Herramientas para gráficos



- 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 *Comando fuentes.* Enlista los fonts o tipos de letra disponibles. Escoja el font que desee aplicar a la selección.
- 2 *Comando tamaño de fuente.* Enlista los tamaños disponibles para el font mostrado 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 (mismas).
- 7 *Centrado*. Ubica la información de la celda o celdas al centro de la misma (mismas).
- 8 *Alineación derecha*. Ubica la información de la celda o celdas al lado derecho de la misma (mismas).
- 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 miles*. Agrega o elimina el formato de miles (000) a las celdas seleccionadas.
- 13 *Agregar decimales*. Agrega decimales a las celdas seleccionadas.
- 14 *Eliminar decimales*. Elimina decimales a las celdas seleccionadas.
- 15 *Paleta de bordes*. Agrega o elimina diferentes tipos de bordes a las celdas seleccionadas.
- 16 *Paleta de color de fondo*. Agrega o elimina un color al fondo de las celdas seleccionadas.

Bibliografía

- Chao, L. Lincon, *Estadística para las ciencias administrativas*, 3a. México, McGraw-Hill México, 1993.
- Downie, N.M. y Heath R.W., *Métodos estadísticos aplicados*, México, Harla, 1973.
- Hernández, S. R., C.C. Fernández y L. P. Baptista, *Metodología de la investigación*, México, McGraw-Hill, 1991.
- Holguín Quiñones, Fernando, *Estadística descriptiva aplicada a las ciencias sociales*, México, UNAM, 1981.
- Levin, I. Richard, *Estadística para administradores*, 2a., México, Prentice Hall, 1988.
- Mendenhall, William, *Estadística para administradores*, México, Grupo Editorial Iberoamérica, 1990.
- Meza, L. Carlos, Andrés Morales A. y Rogelio Magaña C., *Introducción al método estadístico*, México, UAM-X, 1980.
- Microsoft Excel User's Guide, Version 5.0*, EE UU, Microsoft Corporation, 1994.
- Microsoft Excel Function Reference, Version 4.0*, EE UU, Microsoft Corporation, 1992.
- Stokes, J. Charles, *Economics for managers*, EE UU, McGraw-Hill, 1979.
- The 1995 Grolier Multimedia Encyclopedia*, EE UU, Grolier Electronic Publishing, 1995.
- Vox, Diccionario Enciclopédico*, núm. 22, "Matemáticas", Barcelona, España, Bibliograf, 1981.

Revistas

- Examen de la Situación Económica de México*, vol. LXIX, núm. 817, México, Banamex, 1993.
- , vol. LXX, núm. 825, México, Banamex, 1994.
- , vol. LXXI, núm. 836, México, Banamex, 1995.
- , núm. 838, México, Banamex, 1995.
- , núm. 839, México, Banamex, 1995.
- Banco Mundial*, “La página del Banco Mundial”, en sección *Reporte Anual 1995*, Internet, www.worldbank.org/html/extdr/annrept.html
- Bancomext*, “Mexico Today”, México, Banco de Comercio Exterior, 1992.
- Kravzov Jinich, Jaime, “Informe de actividades 1994-1995”, en *Informe del rector*, México, UAM-X, 1995.
- Pierdant, R. Alberto I., “Modelos económicos y de planificación utilizando hojas electrónicas de cálculo”, en *Política y Cultura*, núm. 4, México, DPC/UAM-X, 1995.
- , “¿Qué es una base de datos?”, en *Boletín Informativo*, núm. 1, México, DPE/UAM-X, 1995.
- , “Las hojas electrónicas de cálculo, una herramienta para la enseñanza de las matemáticas”, ponencia, IX Congreso departamental de investigación, Oaxtepec, México, DPE/UAM-X, 1995.
- , “Una base de datos en Excel”, en *Boletín Informativo*, núm. 2, México, DPE/UAM-X, 1996.

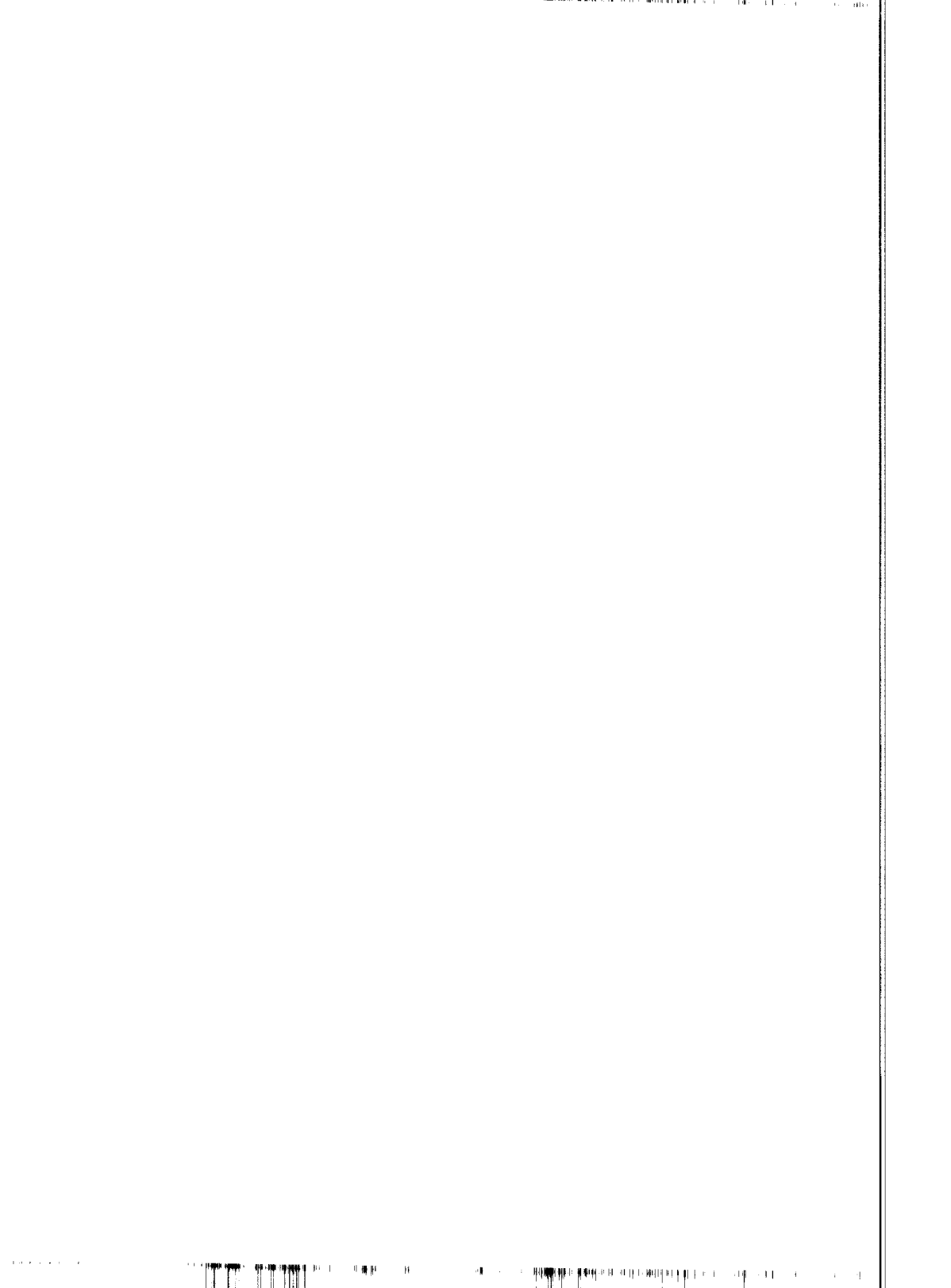


- 1 La sociedad mexicana en los ochenta
Federico Novelo Urdanivia
- 2 Material inflamable
Lauro Zavala
- 3 Lecciones sobre derecho administrativo
Mario Ayluardo Saúl
- 4 Orígenes y desarrollo de la clase obrera en México
Guillermina Bringas
- 5 Crónica sobre utopías
Héctor Cuadra (coord.)
- 6 Introducción al SPSS para microcomputadoras
Rodrigo Pimienta Lastra
- 7 Ahorro e inflación
Raúl Yáñez Alcántara
- 8 Alma y psique
Isabel Jaidar, et al.
- 9 La revolución de los ciudadanos
Jaime Osorio
- 10 Política y Estado en el pensamiento moderno
Gerardo Ávalos Tenorio / María Dolores París (coord. y comp.)
- 11 Modelos administrativos usando Excel
Alberto Isaac Pierdant Rodríguez
- 12 Estadística descriptiva usando Excel
Alberto Isaac Pierdant Rodríguez



Estadística descriptiva usando Excel de Alberto Isaac Pierdant Rodríguez, se terminó de imprimir en febrero de 1997 en Amacalli Editores, SA de CV. Av. México Coyoacán 421, Col. Xoco General Anaya, México DF. ☎ 604 72 63. La diagramación y formación estuvieron a cargo de Amacalli Editores, SA de CV / Lourdes Morales Narvaéz. La edición consta de 500 ejemplares.





La estadística descriptiva es hoy en día una herramienta básica para cualquier especialista o científico social. Su conocimiento y aplicación le permitirán abordar y entender mejor la problemática de la sociedad y proponer, con ello, mejores soluciones. Esta parte de las matemáticas es tratada, en este texto, bajo un nuevo enfoque, ya que no solamente incluye el conocimiento de la estadística sino además su aplicación directa a través del uso de la computadora mediante el programa de hoja electrónica Excel.

Este texto proporciona elementos al especialista o investigador sobre temas poco tratados en otros libros del mismo género, siendo un ejemplo el diseño y construcción de cuadros estadísticos, la elaboración de diferentes tipos de gráficas y la aplicación e interpretación de estadísticas básicas. Asimismo, simplifica la tarea del científico social, proporcionándole un mayor tiempo para la realización del análisis de los datos y estadísticos generados, liberándolo así del difícil quehacer de efectuar cálculos.

Estadística descriptiva usando Excel, es un libro cuya sencillez permite utilizarlo tanto en el bachillerato como en el nivel universitario, y aun para el especialista y el científico de las ciencias sociales.

ALBERTO ISAAC PERDANT RODRIGUEZ (Ciudad de México, 1956). Recibió del Instituto Politécnico Nacional el título de ingeniero industrial y de la Universidad Nacional Autónoma de México el grado de maestría en ingeniería con especialidad en planeación. Ha realizado estudios de especialización, destacando los cursos: Evaluación económica de proyectos de exploración regional de hidrocarburos I y II, en la Universidad de los Andes, Bogotá, Colombia, y Petroleum energy, en The Institute of Energy Economics, Japón.

Es profesor e investigador del Departamento de Política y Cultura de la Universidad Autónoma Metropolitana-Xochimilco.