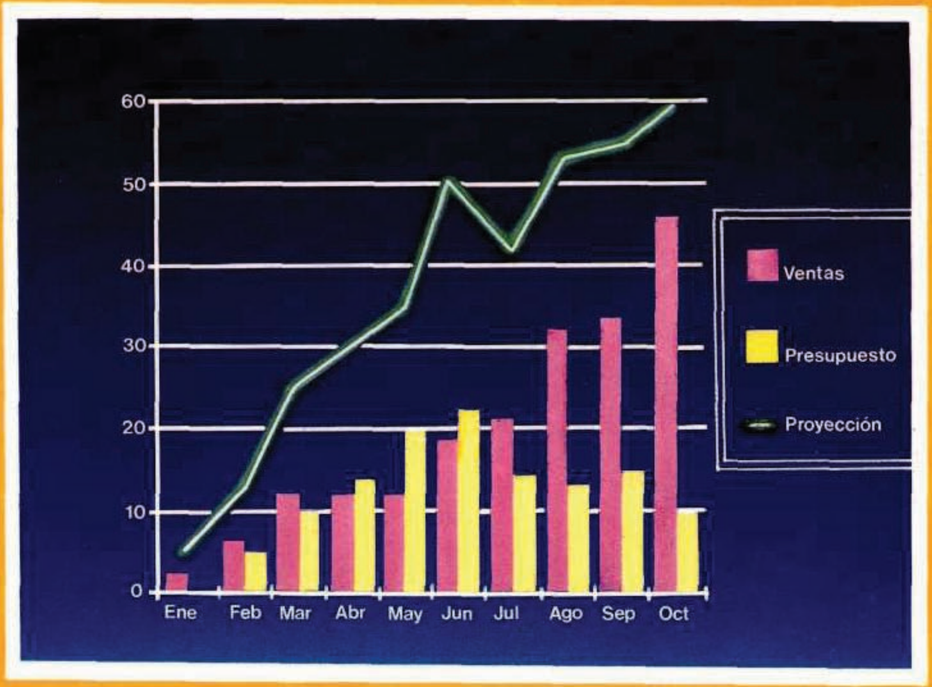
Schaum

ESTADÍSTICA

APLICADA A ADMINISTRACIÓN Y ECONOMÍA

Segunda Edición

Leonard Kazmier
Alfredo Díaz Mata



McGraw-Hill

ESTADÍSTICA APLICADA A ADMINISTRACIÓN Y ECONOMÍA

ESTADÍSTICA APLICADA A ADMINISTRACIÓN Y ECONOMÍA

Segunda Edición

LEONARD J. KAZMIER Ph. D.

Universidad Estatal de Arizona

Coautoría y Traducción:

Alfredo Díaz Mata
Lic. en Administración de Empresas FCA UNAM
Coordinador de Matemáticas
Facultad de Contaduría y Administración
Universidad Nacional Autónoma de México

Revisión Técnica:

Guillermina Eslava Gómez
Doctora en Estadística e
Investigadora del Instituto
de Investigación en Matemáticas
Aplicadas y Sistemas
IIMAS
Universidad Nacional Autónoma de México

McGRAW-HILL

MÉXICO - BOGOTÁ - BUENOSAIRES - CARACAS - GUATEMALA - LISBOA
MADRID - NUEVA YORK - PANAMÁ - SAN JUAN - SANTIAGO - SÃO PAULO
AUCKLAND - HAMBURGO - LONDRES - MILÁN - MONTREAL - NUEVA DELHI
PARÍS - SAN FRANCISCO - SINGAPUR - ST. LOUIS
SIDNEY - TOKIO - TORONTO

ESTADÍSTICA APLICADA A ADMINISTRACIÓN Y ECONOMÍA

McGRAW-HILL INTERAMERICANA DE MÉXICO, S.A. de C.V.
Atlacomulco 499-501, Fracc. Ind. San Andrés Atoto
53500 Naucalpan de Juárez, Edo. de México
Miembro de la Cámara Nacional de la Industria Editorial, Reg. Núm. 1890

ISBN 968-422-787-6
(ISBN 968-451-317-8 primera edición)

Traducido de la segunda edición en inglés de
SCHAUM'S OUTLINE OF BUSINESS STATISTICS

ISBN 0-07-033533-8

1234567890 G.f. 90 9123456780

Impreso en México Printed in Mexico

Esta obra se terminó de
imprimir en diciembre de 1990
en Gráfica Futura 2000, S.A. de C.V.
Calle 28 No. 90
Col. Federal, México, D.F.

Se tiraron 5000 ejemplares

Leonard J. Kazmier es profesor de Sistemas de Información y Decisión en la Universidad estatal de Arizona. Terminó sus estudios de licenciatura y maestría en la Universidad estatal de Wayne, Detroit y obtuvo el Doctorado en la Universidad de Ohio State. Es autor y coautor de libros sobre conceptos de administración, análisis estadísticos y aplicaciones en computadora. Es socio fundador del Instituto de Decission Sciences, y es también miembro de la Academia de Administración y de la Asociación Americana de Estadística. Ha impartido cátedra en las Universidades: estatal de Wayne, Notre Dame y estatal de Arízona.

Contenido

Prefacio

Este libro cubre los métodos básicos de la estadística descriptiva, inferencial y el análisis estadístico de decisión que regularmente son incluidos en cursos introductorios e intermedios sobre estadística para administración y economía. El propósito de este libro es el de presentar los conceptos y métodos en forma clara y concisa sobre el análisis estadístico.

A lo largo del libro se presenta sólo la teoría necesaria en favor de presentar ejemplos muy concretos. Porque este libro fue escrito especialmente para aquellos cuyo primer interés es la aplicación de las técnicas estadísticas, las demostraciones matemáticas han sido omitidas.

Aunque el libro fue diseñado como un suplemento a los libros de texto sobre estadística aplicada, su contenido es tan completo que le permite ser usado como texto. En cuanto al contenido, se han incluido tanto los métodos de análisis estadístico clásicos, como los contemporáneos, el material está agrupado y organizado de tal forma que le permite llevar una secuencia con la mayoría de los libros de texto sobre métodos estadísticos aplicados en administración y economía. Los problemas resueltos que se presentan al final de cada capítulo, incluyen todo el desarrollo para llegar a la solución, mientras que los problemas complementarios incluyen únicamente la respuesta. Todo profesor de la materia, puede obtener un Manual de Soluciones que contiene todos los resultados a los problemas complementarios, escribiendo una carta directamente al autor en papel membretado de la institución donde labora.

Deseo agradecer a la compañía Minitab su autorización para incluir el programa de computadora Minitab en este libro. Este programa es compatible con la mayoría de paquetes de software disponibles sobre análisis estadístico. Este es un sistema conocido que frecuentemente se menciona en los libros de texto sobre estad ística en los negocios.

Estoy también en deuda con el editor ejecutivo de la reciente Sir Ronald A. Fisher, F.R.S., al Dr. Frank Yates, F.R.S., y al grupo Longman, Ltd., de Londres por su autorización para adaptar y reproducir las tablas III y IV de su libro "Statistical Tabres of Biológical Agricultural and Medical Research", especial agradecimiento a mi hija Marian Lamb, por su esmerada y complaciente corrección y preparación del manuscrito.

LEONARD J. KAZMIER

Análisis de datos de negocios

1

1.1 DEFINICIÓN DE ESTADÍSTICA DE NEGOCIOS

La *estadística* se refiere a las técnicas mediante las cuales se recopilan, organizan y analizan datos cuantitativos. El punto central del análisis estadístico en los negocios es la administración de la toma de decisiones.

1.2 ESTADÍSTICA DESCRIPTIVA E INFERENCIA ESTADÍSTICA

La *estadística descriptiva* incluye las técnicas que se relacionan con el resumen y la descripción de datos numéricos. Estos métodos pueden ser gráficos o pueden incluir análisis mediante cálculos (véanse capítulos 2, 3 y 4).

EJEMPLO 1. Puede describirse y darle significado al volumen mensual de ventas de un producto durante el año pasado elaborando un diagrama de barras o una gráfica lineal (como se presenta en la sección 2.8). Pueden destacarse las ventas relativas por mes calculando un numero índice para cada mes, de manera que la desviación a partir de 100 para cualquier mes dado indique la desviación porcentual de las ventas en ese mes en comparación con las ventas mensuales promedio durante todo el año.

La inferencia estadística comprende aquellas técnicas por medio de las cuales se toman decisiones sobre una población estadística basadas en una muestra o en juicios de los administradores. Debido a que esas decisiones se toman en condiciones de incertidumbre, se requiere el uso de conceptos de probabilidad. Considerando que las características medidas en una muestra se denominan *estadísticas muéstrales,* las características medidas en una población estadística, o universo, se llaman *parámetros poblaclonales.* El proceso de medir las características dé todos los miembros de una población definida recibe el nombre de censo. Los capítulos 5,6 y 7 se ocupan de los conceptos de probabilidad y la mayoría de los capítulos siguientes se ocupan de la aplicación de estos conceptos en la inferencia estadística.

EJEMPLO 2. Con el objeto de estimar el voltaje necesario para que un dispositivo eléctrico falle, puede someterse una muestra de esos dispositivos a voltajes cada vez más altos hasta que falten. Con base en los resultados de esta muestra, puede calcularse la probabilidad de falla para los otros dispositivos de la población muestreada en los diversos niveles de voltaje.

1.3 ESTADÍSTICA CLÁSICA Y ANÁLISIS BAYESIANO DE DECISIONES

Los métodos de la *estadística clásica* se refieren al análisis de datos muestreados (objetivo) con propósitos de hacer inferencias, excluyendo todo juicio u opinión personal (véanse capítulos 8 a 15). El *análisis Bayesiano de decisiones* incorpora el uso de juicios de los administradores en el análisis estadístico y también pone énfasis en las posibles ganancias o pérdidas económicas asociadas con decisiones alternativas (véanse capítulos 18-20).

EJEMPLO 3. Por medio del enfoque clásico de la inferencia estadística, podría calcularse el nivel incierto de ventas de un producto nuevo solamente sobre la base de estudios de mercado llevados a cabo en un conjunto de lugares seleccionados de acuerdo con los requerimientos del muestreo científico. Con el enfoque Bayesiano, se obtendría y utilizaría la opinión de los administradores que han tenido experiencia con productos similares como base para estimar un volumen de ventas. Esta estimación subjetiva se podría entonces combinar con datos muestrales objetivos para obtener una estimación combinada de los volúmenes de venta.

1.4 VARIABLES DISCRETAS Y VARIABLES CONTINUAS

Una *variable discreta* sólo puede tener valores observados en puntos aislados a lo largo de una escala. En la estadística de negocios, esa información suele presentarse a través del proceso de *conteo;* de ahí que los valores se expresen generalmente como números enteros. Una *variable continua* puede suponer un valor en cualquier punto fraccionario de un intervalo especificado. Los datos continuos se generan por el proceso de *medición.*

EJEMPLO 4. Como ejemplo de datos discretos se cita el número de personas por hogar, las unidades de un artículo en un inventario y el número de componentes ensamblados que se han encontrado defectuosos. Ejemplo de datos continuos son el peso de un embarque, el tiempo transcurrido antes de que falle un dispositivo y el número promedio de personas por hogar en una comunidad grande. Obsérvese que el *número promedio* de personas puede ser un valor fraccionario y es, entonces, una variable continua, aun cuando el *número* de personas por hogar es una variable discreta.

1.5 OBTENCIÓN DE DATOS A TRAVÉS DE EXPERIMENTOS Y ENCUESTAS

Una manera de obtener datos es a través de la observación directa. Un *experimento estadístico* es una forma de observación directa en la que se controlan algunos o todos los factores que pueden influir sobre la variable que se estudia.

EJEMPLO 5. Pueden compararse dos métodos de ensamblar un componente, haciendo que un grupo de empleados utilice uno de ellos y que un segundo grupo de empleados utilice el otro. Se iguala cuidadosamente a los miembros del primer grupo con los del segundo en términos de factores como edad y experiencia.

En algunas situaciones, no es posible obtener datos en forma directa, sino que, más bien, la información debe obtenerse a partir de respuestas individuales. Una *encuesta estadística* es el proceso de recopilar datos pidiendo a personas que proporcionen información. Los datos pueden obtenerse con métodos como la entrevista personal o telefónica, o a través de un cuestionario escrito.

EJEMPLO 6. Un analista de la Secretaría del Trabajo puede necesitar determinar qué aumentos o reducciones en el nivel de empleo tienen planeados las empresas de cierto estado. Un método común para obtener esa clase de datos consiste en efectuar una encuesta entre las empresas.

1.6 MÉTODOS DE MUESTREO ALEATORIO

El *maestreo aleatorio* es aquél en el que cada uno de los elementos de la población de interés, o *población objeto,* como se le conoce, tiene una probabilidad conocida, y frecuentemente igual, de ser elegido para la muestra. A las muestras aleatorias se les denomina también *muestras probabillsticas o muestras científicas.* Son cuatro los principales métodos de muestreo aleatorio: aleatorio simple, sistemático, estratificado y por conglomerados.

Una muestra *aleatoria simple* es aquélla en la que los elementos se escogen en forma individual y al azar de la totalidad de la población. Esta selección al azar es similar a la que se realiza en la extracción aleatoria de números en una lotería. Sin embargo, en el muestreo estadístico, por lo general se utiliza un programa computarizado de *tabla de números aleatorios* o un *generador de números aleatorios* para identificar los elementos numerados de la población que se eligen para la muestra.

EJEMPLO 7. El apéndice 1 muestra una tabla abreviada de número» aleatorios. Suponga que se desea tomar una muestra aleatoria simple de 10 cuentas por cobrar de una población de 90 dé esas cuentas, las cuales se numeran del 1 al 90. Se empezarla en la tabla de números aleatorios "a ciegas", literalmente cerrando los ojos y señalando una posición inicial. Después, se leerían grupos de números, de dos en dos, en cualquier dirección, para determinar las cuentas que han de formar parte de la muestra. Suponga que se comienza leyendo pares de números, comenzando con el que está en la primera columna del renglón 6. Los 10 números de las cuentas para la muestra serian 66, 06, 59, 94, 78, 70, 08, 67,12 y 65. Sin embargo, como sólo hay 90 cuentas, no puede incluirse el número 94. Por ello, se incluye en la muestra el número siguiente, el 11. Si se repite cualquiera de los números seleccionados, sólo se le incluye una vez en la muestra.

Una *muestra sistemática* es una muestra aleatoria en la cual se eligen lo elementos de la población a intervalos uniformes, a partir de un listado ordenado, tal como elegir cada décima cuenta por cobrar para la muestra. La primera de las cuentas de la muestra se elegirla al azar (quizá utilizando una tabla de números aleatorios). Un problema especifico del muestreo sistemático es la existencia de cualquier factor periódico o cíclico en la lista dé la población que pudiera conducir a un error sistemático en los resultados muestrales.

EJEMPLO 8. Si en una comunidad en la que se estudia lo apropiado de la iluminación pública cada doceava casa se encuentra en una esquina, una muestra sistemática tendría un sesgo sistemático si se incluye en la investigación cada decimosegunda casa. En este caso, o todas o ninguna de las casas investigadas se encontrarían en esquina.

En el *muestreo estratificado,* lo primero que hace el investigador es clasificar los elementos de la población en subgrupos separados de acuerdo con una o más características importantes. Después, se obtiene por separado una muestra aleatoria simple o sistemática de cada estrato. Puede utilizarse este tipo de muestreo para asegurar una representación proporcional de diversos subgrupos en la muestra. Además, es común que el tamaño de la muestra que se requiere para lograr determinado nivel de precisión en el muestreo estratificado sea menor que con muestreo aleatorio simple, con la consiguiente reducción en los costos del muestreo.

EJEMPLO 9. En un estudio de las actitudes de estudiantes con respecto a las políticas de hospedaje en las instalaciones escolares, es razonable pensar que pueden existir diferencias importantes entre los estudiantes de licenciatura y los de posgrado, y entre hombres y mujeres. Por lo tanto, debe considerarse un esquema de muestreo estratificado, escogiendo una muestra diferente en cada uno de los cuatro estratos: hombres que estudian la licenciatura, mujeres estudiantes de licenciatura, hombres graduados en posgrado y mujeres graduadas en posgrado.

El *muestreo por conglomerados* es un tipo de muestreo aleatorio en el que los elementos de la población se dividen en forma natural en subgrupos. Así, se eligen al azar los subgrupos que forman la muestra.

EJEMPLO 10. Si un analista de la Secretaría del Trabajo necesita estudiar los salarios diarios que se pagan en un área metropolitana, seria difícil obtener una lista de todos los asalariados de la población objetivo. Sin embargo, podría obtenerse fácilmente una lista de las *empresas* de la región. Con esta lista, el analista puede tomar una muestra aleatoria de las empresas identificadas, que representan *conglomerados* de trabajadores, y obtener los salarios que estas empresas les pagan.

1.7 UTILIZACIÓN DE COMPUTADORAS PARA GENERAR NÚMEROS ALEATORIOS

Existen muchos programas de cómputo disponibles que generan números aleatorios dentro de un intervalo especificado de valores. En el problema 1.9 se ilustra el uso de esta clase de programas.

Problemas resueltos

ESTADÍSTICA DESCRIPTIVA E INFERENCIA ESTADÍSTICA

1.1 Indique cuáles de los términos u operaciones siguientes se relacionan con muestras o con muestreo (M) o con una población (P): (a) grupos de medidas llamados *parámetros, (b)* uso de *inferencia estadística,* (c) hacer un censo, (d) juzgar la calidad de un embarque de fruta inspeccionando varias de las cajas.

(a) P, (b) M. (c) P, (d) M

ESTADÍSTICA CLÁSICA Y ANÁLISIS BAYESIANO DE DECISIÓN

1.2 Indique cuáles de los siguientes tipos de información podrían usarse exclusivamente en el análisis Bayesiano de decisión (B) y cuáles podrían usarse indistintamente en el análisis clásico o en el Bayesiano (CB): (a) juicio de administradores sobre el nivel probable de ventas para un producto nuevo, *(b)* resultados de una encuesta a una muestra de clientes antiguos, (c) pronósticos de analistas financieros sobre promedios del mercado de valores.

(a) B, (*b)* CB. (c) B

VARIABLES DISCRETAS Y CONTINUAS

1.3 Para los siguientes tipos de valores, identifique si corresponden a variables disaetas (D) o variables continuas (C): (a) el peso del contenido de un paquete de cereales, *(b)* el diámetro de un rodamiento, (c) el número de artículos defectuosos que se producen, (d) el número de personas pensionadas en un área geográfica, (e) el número promedio de posibles clientes que han sido visitados por un vendedor durante el mes pasado, (f) el total de las ventas, en pesos.

(a) C, (b) C, (c) D, (d) D, (e) C, (f) D *(Nota:* aunque los valores monetarios son datos discretos, cuando se trata de cantidades grandes, se consideran como continuos.)

OBTENCIÓN DE DATOS MEDIANTE EXPERIMENTOS Y ENCUESTAS

1.4 Señale cuáles de los siguientes procedimientos de recopilación de datos pueden considerarse como experimentos (E) y cuáles como encuestas (EN): (a) Una investigación política sobre la forma en que las personas piensan votar en ciertas elecciones, (b) se entrevista a los clientes de un centro comercial sobre las razones que tienen para hacer sus compras en ese lugar, (c) se comparan dos métodos para comercializar una póliza de seguros, utilizando cada uno de los métodos en áreas geográficas comparables.

(a) EN, (b) EN, (c) E

1.5 En las mediciones estadísticas, tal como los cuestionarios, la *confiabilidad* se refiere a la consistencia del instrumento de medición, y *validez* se refiere a la precisión del instrumento. Así, si un cuestionario arroja resultados similares cuando se le aplica a dos grupos equivalentes de encuestados, entonces se dice que el cuestionario es confiable. ¿El hecho de que un cuestionario sea confiable garantiza su validez?

La confiabilidad de un instrumento de medición no garantiza que sea válido para un propósito específico. Un instrumento confiable es consistente en repetidas mediciones, pero todas éstas pueden incluir un componente común de error, o sesgo. (Véase el siguiente problema resuelto.)

1.6 Con referencia al problema 1.5, ¿un Instrumento que no es confiable puede tener validez para un propósito especifico?

Un instrumento que no es confiable no puede tener validez para *ningún* propósito especifico. Sin confiabilidad, los resultados que se obtienen no son consistentes. Se puede ilustrar este concepto mediante una analogía con un objetivo de tiro con rifle. Si los hoyos de las balas se agrupan cercanos entre si en el objetivo son señal de confiabiliad (consistencia) en los disparos. En este caso, puede mejorarse la validez (precisión) ajusfando la mirilla para que los hoyos tiendan a agruparse en el centro del objetivo. Por otro lado, si los hoyos se encuentran muy dispersos, serian señal de falta de confiabilidad y, bajo estas circunstancias, un ajuste en la mirilla no puede conducir a disparos más certeros.

MÉTODOS DE MUESTREO ALEATORIO

1.7 Para realizar inferencias estadísticas, se desea tener una muestra *representativa.* Sin embargo, los métodos de la inferencia estadística sólo requieren que se obtenga una muestra *aleatoria.* ¿Por qué?

No existe ningún método de muestreo que pueda garantizar que se obtendrá una muestra representativa. Lo mejor que puede lograrse es evitar sesgos consistentes o sistemáticos utilizando el muestreo aleatorio (probabilístico). Aunque son raras las ocasiones en que una muestra aleatoria es representativa de la población objeto de la cual se obtiene, el uso de este procedimiento garantiza que las diferencias entre la muestra y la población se deben sólo al azar.

1.8 Una compañía petrolera desea determinar los factores que afectan la elección de estaciones de servicio por parte de los clientes en un área de prueba y, para ello, ha obtenido los nombres, direcciones y la información personal disponible sobre todos los propietarios de automóviles registrados en esa área Describa la forma en que podría obtenerse una muestra aleatoria a partir de esta lista, utilizando los cuatro métodos de muestreo aleatorio que se describieron en este capitulo.

Para una *muestra aleatoria simple,* podrían numerarse los nombres de la lista de manera secuencial y, después, se podría elegir a los elementos de la muestra utilizando una tabla de números aleatorios. Para una *muestra sistemática,* se podría entrevistar a cada enésima persona de la lista (cada quinta persona, por ejemplo), comenzando al azar con alguno de los primeros cinco nombres. Para una/nuestra *estratificada,* pueden clasificarse los nombres según el tipo de automóvil que poseen, según su valor, o según su sexo o edad y, después, sacar una muestra aleatoria simple de cada uno de los estratos. Para una *muestra por conglomerados,* podría entrevistarse a los propietarios de automóviles registrados que residan en cuadras seleccionadas al azar del área de prueba. Como tendría esta base geográfica, a este tipo de muestra por conglomerados se le podría denominar, alternativamente, *muestra por área.*

UTILIZACIÓN DE COMPUTADORAS PARA GENERAR NÚMEROS ALEATORIOS

1.9 Un analista económico desea obtener una muestra aleatoria simple de 30 empresas de las 435 que están listadas en un área geográfica. Las empresas están identificadas secuencialmente con números del 001 al 435. Utilizando algún programa de computación, obtenga 30 números de identificación, que determinen las empresas que deben incluirse en la muestra.

```
MTB > RANDOM 30 DIGITS, PUT IN C1;
SUBO INTEGERS FROM 1 TO 435.
MTB > NAME FOR C1 IS 'SAMPLE'
MTB > PRINT 'SAMPLE'
SAMPLE
    8    56   184    94   275    78   135   172   303   296    94
   90   405   169   239   114   100    26   212   331   215    18
  141   326   290   297   201   140   248    94
```

Fig. 1-1 Listado obtenido con Minitab. (Derechos reservados por la Universidad Estatal de Pennsylvania.)

En la figura 1-1 se muestra el listado de los números de identificación de las 30 empresas elegidas al azar. Observe que, por casualidad, la empresa número 94 fue elegida tres veces. Por ello, se requerirla obtener los números de dos empresas más para alcanzar el tamaño deseado de 30 empresas.

Problemas complementarios

ESTADÍSTICA DESCRIPTIVA E INFERENCIA ESTADÍSTICA

1.10 Indique cuáles de los términos u operaciones siguientes se relacionan con muestras o muestreo (M), y aquéllos que se relacionan con una población (P): (a) Universo, (b) grupos de medidas denominadas *estadísticas,* (c) aplicación de conceptos de probabilidad, (d) inspección de todos los artículos que se fabrican.

Resp. (a) P, (b) M, (c) M, (d) P

ESTADÍSTICA CLÁSICA Y ANÁLISIS BAYESIANO DE DECISIÓN

1.11 Indique cuáles de los siguientes tipos de información pueden ser sólo usados en el análisis Bayesiano de decisión, (B) y los que pueden emplearse tanto en el análisis clásico como en el Bayesiano (CB): (a) Datos objetivos, *(b)* datos de una muestra, (c) datos subjetivos, (d) juicios de administradores.

Resp. (a) CB, (b) CB, (c) B, (d) B

VARIABLES DISCRETAS Y CONTINUAS

1.12 Para los siguientes valores, identifique cuáles corresponden a variables discretas (D) y cuáles a variables continuas (C): (a) Número de unidades de un artículo en existencia, (b) relación de activos circulantes con pasivos circulantes, (c) tonelaje total embarcado, (d) cantidad embarcada, en unidades, (e) volumen de tráfico en un camino de cuota, (f) asistencia a la reunión anual de una compañía.

Resp. (a) D, (b) C, (c) C, (d) D, (e) D, (f) D

OBTENCIÓN DE DATOS MEDIANTE EXPERIMENTOS Y ENCUESTAS

1.13 Señale cuáles de los siguientes procedimientos de recopilación de datos se considerarían experimentos (E) y cuáles encuestas (EN): (a) comparar los resultados de un nuevo método para entrenar agentes de boletos de avión con el método tradicional, (b) evaluar dos conjuntos diferentes de instrucciones de ensamble para un juguete haciendo que dos grupos comparables de niños ensamblen el juguete utilizando diferentes instrucciones, (c) hacer que una revista envíe a sus suscriptores un cuestionario pidiéndoles que evalúen un producto recientemente adquirido.

Resp. (a) E, (b) E, (c) EN

MÉTODOS DE MUESTREO ALEATORIO

1.14 Identifique si se utiliza el método de muestreo aleatorio simple (A) o el sistemático (S) en los siguientes casos: (a) Utilización de una tabla de números aleatorios, (b) entrevistar a cada centesimo adulto que entra a un parque de diversiones, comenzando aleatoriamente con la quincuagésima persona que entra al parque.

Resp. (a) A, *(b)* S.

1.15 Para las siguientes situaciones de muestreo grupal, determine si se utilizaría el muestreo aleatorio estratificado (ES) o el muestreo por conglomerados (C): (a) Estimación de las preferencias electorales de la gente que vive en diversas comunidades, (b) estudio de las actitudes de los consumidores, suponiendo que existen diferencias importantes debidas a edad y sexo.

Resp. (a) C, *(b)* ES

UTILIZACIÓN DE COMPUTADORAS PARA GENERAR NÚMEROS ALEATORIOS

1.16 Un auditor desea obtener una muestra aleatoria simple de 50 de las 5250 cuentas por cobrar de una empresa grande. Las cuentas están numeradas secuendalmente del 0001 al 5250. Utilice algún programa de computadora para obtener un listado de tos 50 números aleatorios que se requieren.

Presentaciones estadísticas

2

2.1 DISTRIBUCIONES DE FRECUENCIAS

Una *distribución de frecuencias* es una tabla en la cual se agrupan en clases los valores posibles para una variable y se registra el número de valores observados que corresponde a cada clase. Los datos organizados en una distribución de frecuencias se denominan *datos agrupados.* Por el contrario, para los *datos no agrupados,* se enumeran todos los valores observados de la variable aleatoria.

EJEMPLO 1. En la Tabla 2.1 se muestra una distribución de frecuencias de salarios diarios. Nótese que las cantidades están dadas hasta el peso más cercano. Cuando se trata de redondear una fracción que es "exactamente 0.5" (exactamente $0.50 en este caso), usualmente se redondea al número entero par más cercano. Así, un salario diario de $2599.50 se redondearía a $2600.00, como parte del proceso de reporte de datos.

Tabla 2.1 Distribución de frecuencias de los salarios de 100 obreros no calificados

Salario diario	Número de obreros (f)
$2400-2599	7
2600-2799	20
2800-2999	33
3000-3199	25
3200-3399	11
3400-3599	4
	Total 100

2.2 INTERVALOS DE CLASE

Para cada una de las clases de una distribución de frecuencias, los *límites nominales de clase* superior e inferior indican los valores incluidos dentro de la clase.(Véase la primera columna de la Tabla 2.1.) En contraste, las *fronteras de clase* o *límites exactos,* son los puntos específicos de la escala de medición que sirven para separar clases adyacentes cuando se trata de variables continuas. Los límites exactos de clase pueden determinarse identificando los puntos que están a la mitad entre los límites superior e inferior, respectivamente, de las clases adyacentes. El *intervalo de clase* indica el rango de los valores incluidos dentro de una clase y puede ser determinado restando ei límite exacto inferior de clase de su límite exacto superior. Cuando no se identifican límites exactos, puede determinarse el intervalo de clase restando el límite nominal inferior de una clase del límite nominal inferior de la clase inmediata siguiente. Finalmente, para ciertos propósitos, los valores de una clase se representan a menudo por el *punto medio de clase,* que puede ser determinado sumando la mitad del intervalo de clase a su limite inferior.

EJEMPLO 2. En la Tabla 2.2 se presentan los límites exactos de clase y los correspondientes puntos medios de clase de la distribución de frecuencias de la Tabla 2.1.

EJEMPLO 3. Calculado mediante los dos procedimientos, el intervalo de clase de la primera clase de la Tabla 2.2 es \$2,599.50 - \$2,399.50 - \$200 (restando el limite de clase exacto inferior del límite de clase exacto superior) o \$2,600 - \$2,400 = \$200 (restando el límite nominal inferior de clase del limite nominal inferior de la clase inmediata siguiente).

Tabla 2.2 Salarios diarios de 100 obreros no calificados

Salario diario (límites de clase)	Límites exactos de clase*	Punto medio de clase	Número de obreros
$2400-2599	2399.50-2599.50	2499.50	7
2600-2799	2599.50-2799.50	2699.50	20
2800-2999	2799.50-2999.50	2899.50	33
3000-3199	2999.50-3199.50	3099.50	25
3200-3399	3199.50-3399.50	3299.50	11
3400-3599	3399.50-3599.50	3499.50	4
			Total 100

* En general, sólo se expresa un dígito decimal significativo en los limites exactos de clase. Sin embargo, debido a que con unidades monetarias la medida unitaria más precisa después de "el peso más cercano" se define comúnmente como "el centavo más cercano", en esta tabla se incluyen dos dígitos decimales.

Por razones de cálculo, generalmente es deseable que todos los intervalos de clase en una distribución de frecuencias dada sean iguales. Una fórmula que puede utilizarse para determinar el intervalo de clase aproximado es:

$$\text{Intervalo aproximado} = \frac{\begin{array}{c}\text{valor mayor en los}\\ \text{datos no agrupados}\end{array} - \begin{array}{c}\text{valor menor en los}\\ \text{datos no agrupados}\end{array}}{\text{número de clases que se desean}} \tag{2.7}$$

EJEMPLO 4. Para los datos originalmente no agrupados que se agruparon en la Tabla 2.1, supóngase que el salario más alto observado era \$3580 y el más bajo era de \$2420. Dado el objetivo de tener seis clases con intervalos de clase iguales:

$$\text{intervalo aproximado} = \frac{3{,}580 - 2{,}420}{6} = \$193.33$$

El tamaño de clase conveniente más próximo es, entonces, \$200.

Para datos distribuidos de manera muy irregular, como los datos anuales de salario para diversas ocupaciones, pueden ser convenientes los *intervalos desiguales de clase.* En este caso, se utilizan intervalos de clase mayores para los rangos de valores en que hay relativamente pocas observaciones.

2.3 HISTOGRAMAS Y POLÍGONOS DE FRECUENCIAS

Un *histograma* es la gráfica de barras de una distribución de frecuencias. Como se indica en la figura. 2-1, por lo general se colocan sobre el eje horizontal de la gráfica los limites exactos de clase, en tanto que sobre el eje vertical se coloca el número de observaciones.

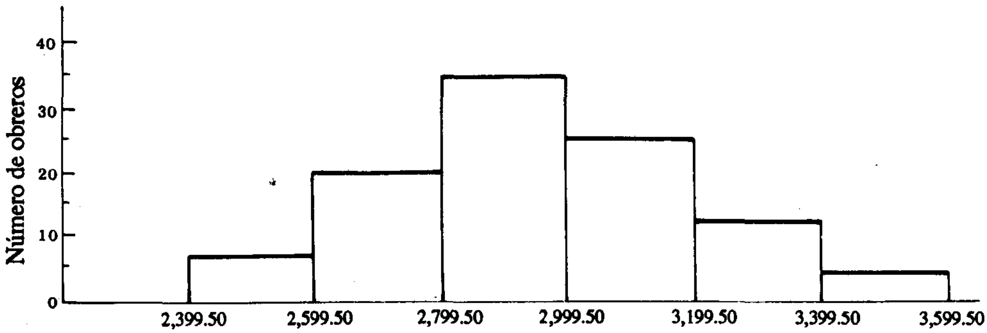


Fig. 2-1

EJEMPLO 5. En la figura 2-1 se ilustra un histograma para la distribución de frecuencias de los salarios diarios de la Tabla 2.2. Un *polígono de frecuencias* es la gráfica lineal de una distribución de frecuencias. Como se muestra en la figura 2-2, los dos ejes de esta gráfica son similares a los del histograma excepto que, comúnmente, se coloca el punto medio de clase sobre el eje horizontal. El número de observaciones en cada clase se representa por un punto en el punto medio de la clase y estos puntos están unidos por una serie de segmentos de linea para formar una "figura de varios lados", o polígono.

EJEMPLO 6. La figura 2-2 muestra un polígono de frecuencias para la distribución de los salarios diarios de la Tabla 2-2.

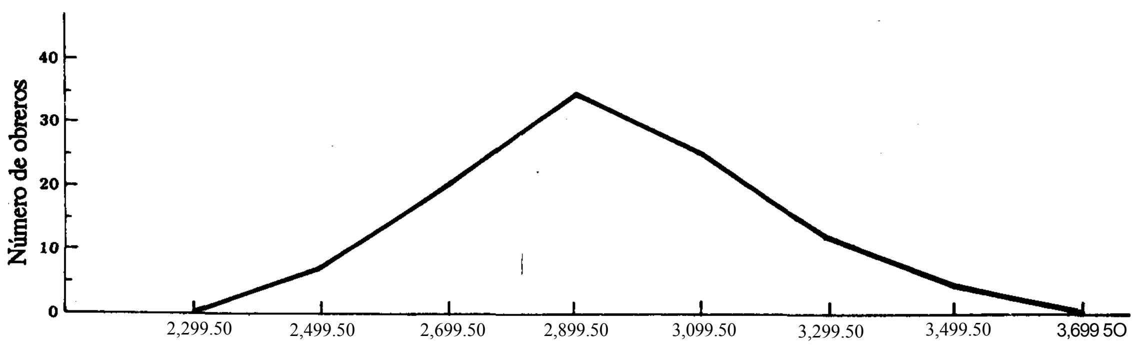


Fig. 2-2

2.4 CURVAS DE FRECUENCIAS

Una curva de frecuencias es un polígono de frecuencias suavizado.

EJEMPLO 7. La figura 2-3 representa una curva de frecuencias para la distribución de los salarios diarios de la Tabla 2.2.

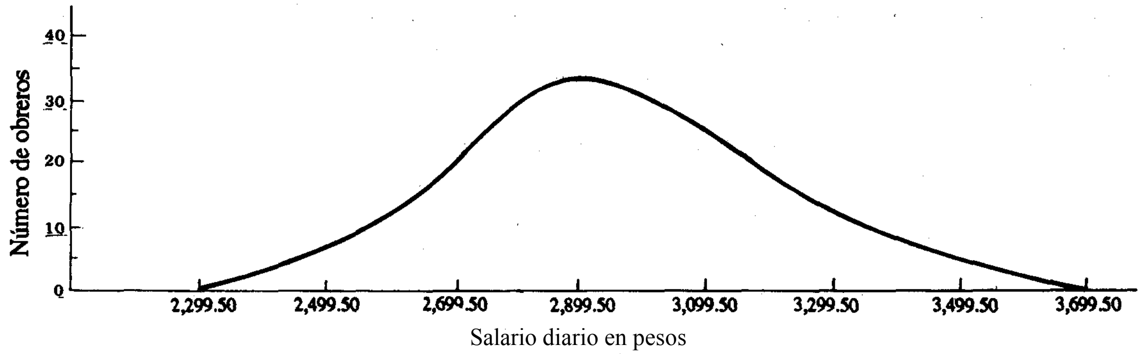


Fig. 2-3

En términos de asimetría, una curva de frecuencias puede sen (1) *asimétrica negativa:* asimétrica con la "cola" hacia la izquierda, (2) *positivamente sesgada:* asimétrica con la "cola" hacia la derecha, o (3) *simétrica.*

EJEMPLO 8. En la figura 2-4 se ilustra gráficamente el concepto de asimetría de una curva de frecuencias.

En términos de kurtosis, una curva de frecuencia puede ser: (1) *platikúrtka:* plana, con las observaciones distribuidas de manera relativamente uniforme en todas las clases, (2) *leptokúrtica:* puntiaguda, con las observaciones concentradas en un estrecho rango de valores, o (3) *mesokúrtica* ni plana ni puntiaguda, en términos de la distribución de los valores observados.

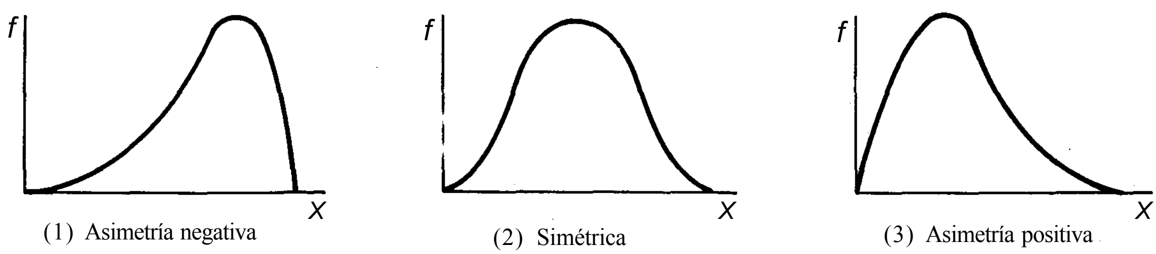


Fig. 2-4

EJEMPLO 9. En la figura 2-5 se muestran tipos de curvas de frecuencias en términos de kurtosis.

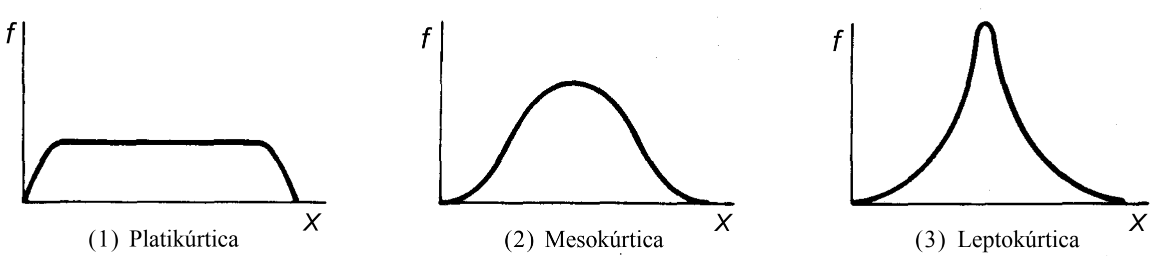


Fig. 2-5

2.5 DISTRIBUCIONES DE FRECUENCIAS ACUMULADAS

Una distribución de frecuencias acumuladas identifica el número de observaciones acumuladas incluidas bajo el limite exacto superior de cada clase de la distribución. Puede determinarse la frecuencia acumulada para una clase agregando la frecuencia observada para dicha clase a la frecuencia acumulada de la clase precedente.

EJEMPLO 10. En la Tabla 2.3 se ilustra el cálculo de las frecuencias acumuladas.

Tabla 2.3 Cálculo de las frecuencias acumuladas de los datos de salarios diarios de la Tabla 2.2

Salario diario	Límite exacto superior de clase	Número de trabajadores	Frecuencia acumulada (fa)
$2400-2599	2599.50	7	7
2600-2799	2799.50	20	20 + 7= 27
2800-2999	2999.50	33	33 + 27= 60
3000-3199	3199.50	25	25 + 60= 85
3200-3399	3399.50	11	11 + 85= 96
3400-3599	3599.50	4	4 + 96 = 100
		Total 100	

La gráfica de una distribución de frecuencias acumuladas se denomina *ojiva. Para* las distribuciones acumuladas del tipo "y menor que" esta gráfica indica la frecuencia acumulada debajo de cada límite exacto de clase de la distribución de frecuencias. Cuando dicha gráfica de línea está suavizada, se le denomina *curva ojiva.*

EJEMPLO 11. En la figura 2-6 se presenta una curva ojiva para la distribución de frecuencias acumuladas de la Tabla 2.3.

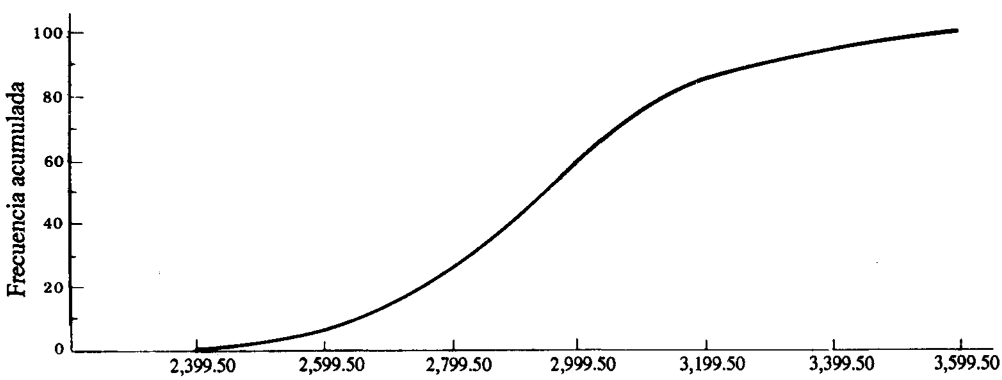


Fig. 2-6

2.6 DISTRIBUCIONES DE FRECUENCIAS RELATIVAS

Una *distribución de frecuencias relativas* es aquélla en la que el número de observaciones de cada clase se convierte en una frecuencia relativa dividiéndolo entre el número total de observaciones en la distribución. Por esto, cada frecuencia relativa es una proporción y se le puede convertir en porcentaje multiplicándola por 100.

Una de las ventajas que presenta la construcción de la distribución de frecuencias relativas reside en que la distribución acumulada y la ojiva correspondiente indican la proporción acumulada (o porcentaje) de observaciones presentes hasta los diversos valores posibles de la variable. Un valor *percentil* es el porcentaje de observaciones acumulado hasta un valor designado de una variable. (Véanse problemas 2.14 y 2.16 a 2.20.)

2.7 LA DISTRIBUCIÓN DE FRECUENCIAS DEL TIPO "Y MENOR QUE"

Considérense los siguientes limites nominales de clase para dos clases adyacentes:

5 y menor que 8
8 y menor que 11

Para estos limites nominales de clase planteados como "y menor que", los limites exactos de clase son idénticos en valor a los respectivos limites nominales. Asi, los limites exactos superior e inferior para la primera clase mencionada son 5.0 y 8.0, respectivamente.(Vóanse los problemas 2.21 y 2.22.) Por lo general, es más fácil manejar en computadora este enfoque de "y menor que", pues en ocasiones ilustra una forma más "natural" de recopilar datos cuando se miden cantidades de tiempo. Por ejemplo, es común que las edades de las personas se reporten como la edad cumplida en el último cumpleaños, y no la edad que se cumple en el cumpleaños próximo.

2.8 GRÁFICAS DE BARRAS Y GRÁFICAS DE LÍNEA

Una *gráfica* (o diagrama) *de barras* ilustra, mediante barras, cantidades de frecuencias para diferentes categorías de datos. La diferencia entre una gráfica de barras y un histograma es que éste se refiere siempre a datos de una distribución de frecuencias, en tanto que una gráfica de barras ilustra cantidades para cualquier tipo de categorías.

EJEMPLO 12. En la gráfica de barras de la figura 2-7 se ilustran los montos (en billones de pesos) de los créditos otorgados por la banca múltiple a empresas y particulares, durante los meses de marzo a agosto de 1989.

Una gráfica de barras y componentes incluye subdivisiones de las barras. Por ejemplo, se podría subdividir cada una de las barras de la figura 2-7, en partes (quizá diferenciadas mediante colores o algún tipo de rayado para distinguir los montos de los créditos otorgados cada mes a empresas de los que se otorgan a particulares. (Véase el problema 2.24.)

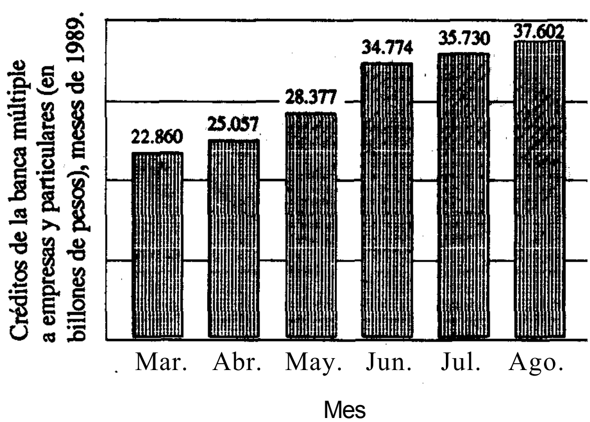


Fig. 2.7 (Fuente: Indicadores Económicos, Banco de México)

Cuando las categorías que se utilizan representan un segmento de tiempo, como es el caso de los datos de la figura 2-7, también se les puede representar mediante *una gráfica de línea,* la cual ilustra mediante segmentos de linea los cambios en cantidades con respecto al tiempo.

EJEMPLO 13. En la figura 2-8 se presentan los datos de la figura 2-7, en forma de gráfica de linea.

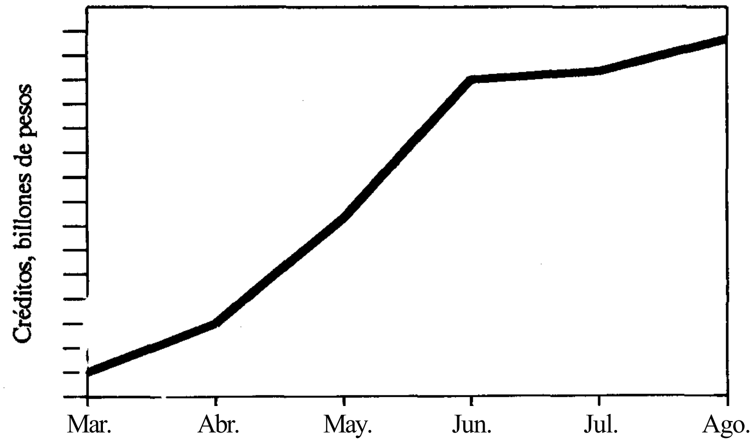


Fig. 2-8

2.9 GRÁFICAS DE PASTEL

Las *gráficas de pastel* son especlalmente apropiadas para ilustrar las divisiones de una cantidad total, tal como la distribución de los egresos o los ingresos de un; compañía

Una *gráfica de pastel en porcentajes* (o porcentual) es aquélla en la que ios valores se convierten a porcentajes para que resulte más fácil compararlos. (Véase el problema 2.26.)

EJEMPLO 14 La figura 2-9 es una gráfica de pastel porcentual que ilustra la participación de los diversos valores operados en el mercado de dinero de la Bolsa Mexicana de Valores, al 31 de enero de 1990.

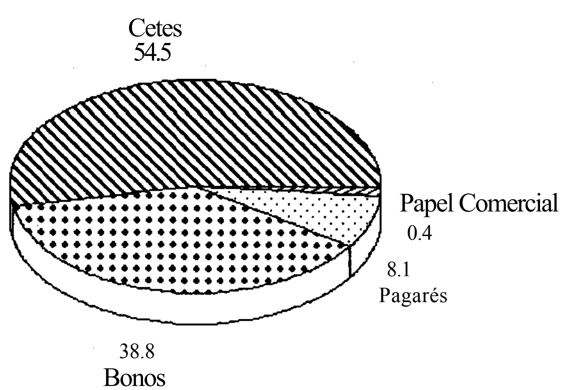


Fig. 2-9

2.10 RESULTADOS POR COMPUTADORA

Existen programas de computadora que permiten formar distribuciones de frecuencias a partir de datos no agrupados y que producen el histograma asociado. Normalmente estos programas determinan el intervalo de clase conveniente, o permiten al usuario especificar el tamaño de intervalo que se desea utilizar. Debido a la conveniencia para interpretar puntos medios de clase no fraccionarios y límites exactos de clase, por lo general se utilizan distribuciones de frecuencias del tipo de "y menor que". (Véanse los problemas 2.27 y 2.28.)

Problemas resueltos

DISTRIBUCIONES DE FRECUENCIAS, INTERVALOS DE CLASE Y MÉTODOS GRÁFICOS ASOCIADOS

2.1 Con referencia a la Tabla 2.4,

(a) ¿Cuáles son los límites nominales inferior y superior de la primera clase?
(b) ¿Cuáles son los límites exactos inferior y superior de la primera clase?
(c) El intervalo de clase es igual en todas las clases de la distribución. ¿Cuál es el tamaño del intervalo?

Tabla 2.4 Distribución de frecuencias de rentas mensuales de 200 departamentos (en miles de pesos)

Renta mensual (miles de pesos)	Número de departamentos
$350-379	3
380-409	8
410-439	10
440-469	13
470-499	33
500-529	40
530-559	35
560-589	30
590-619	16
620-649	12
	Total 200

(d) ¿Cuál es el punto medio de la primera clase?
(e) ¿Cuáles son los límites exactos inferior y superior de la clase en la que se tabuló la mayor cantidad de rentas de departamentos?
(f) Suponga que se reporta una renta de $439 500. Identifique los limites nominales inferior y superior de la clase en la que se incluiría esta renta.

(a) $350 000 y $379 000
(b) $349 500 y $379 500 *Nota*.Como en el ejemplo 2, dos dígitos adicionales están expresados en lugar de un dígito adicional, como usualmente se maneja en su límite de clase comparado con los límites nominales.
(c) Concentrándose en el intervalo de valores de la primera clase,
$379 500 - $349 500 = $30 000 (restando el limite exacto inferior de la clase del límite exacto superior).
$380 000 - $350 000 = $30 000 (restando el límite nominal inferior de clase del límite nominal inferior de la clase adyacente superior)

(d) $\$349\,500+\frac{30000}{2}-\$349\,500+\$15\,000=\$364\,500$

(e) \$499 500 y \$529 500

(f) \$440 000 y \$469 000 *(Nota:* los \$439 500 se redondean a \$440 000, como la cifra más próxima, de acuerdo con la "regla del número par" descrita en la sección 2.1.)

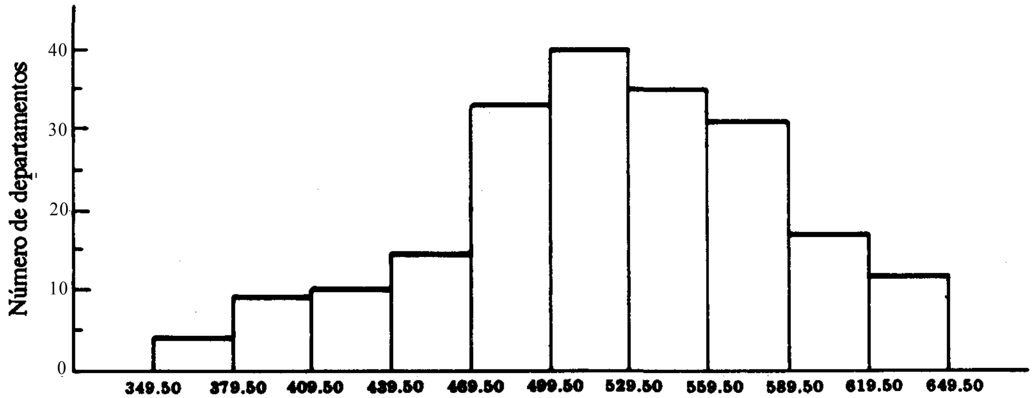


Fig. 2-10

2.2 Construya un histograma para los datos de la Tabla 2.4

En la figura 2-I0, aparece un histograma para los datos de la Tabla 2.4.

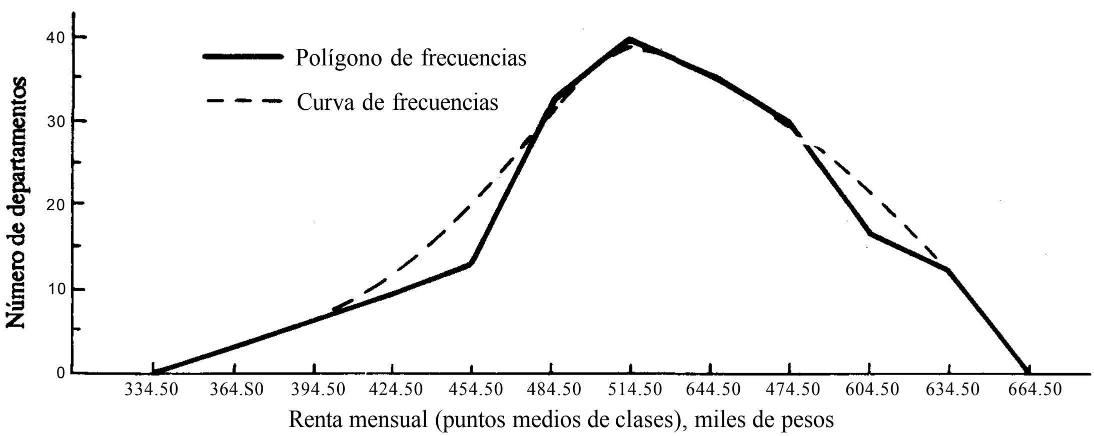


Fig. 2-11

2.3 Construya un polígono de frecuencias y una curva de frecuencias para los datos de la Tabla 2.4.

La figura 2-11 es el polígono de frecuencias de la curva de frecuencias para los datos de la Tabla 2.4.

2.4 Describa la curva de frecuencias de la figura 2-11, desde el punto de vista de la asimetría.

Parece que la curva de frecuencias es asimétrica negativa.

2.5 Construya una distribución de frecuencias acumuladas para los datos de la Tabla 2.4. Véase la Tabla 2.5.

2.6 Presente la distribución de frecuencias acumuladas de la Tabla 2.5 en forma gráfica mediante una ojiva.

La ojiva para los datos de la Tabla 2.5 se muestran en la figura 2-12.

2.7 En la Tabla 2.6 se enlistan los tiempos que requieren 30 empleados para terminar una labor típica de ensamble y que han solicitado una transferencia promocional a otro puesto que requiere ensamble de precisión. Supóngase que se desea organizar estos datos en cinco clases de tamaños iguales. Determine el tamaño conveniente del intervalo.

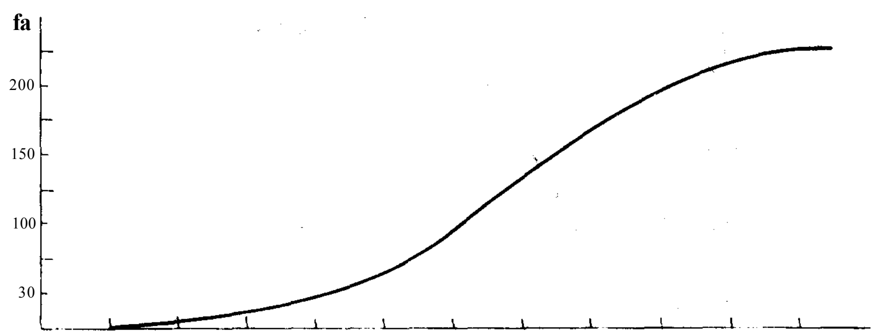
$$\text{Intervalo aproximado} = \frac{\text{valor mayor en los datos no agrupados} \quad \text{valor menor en los datos no agrupados}}{\text{número de clases que se desean}}$$

$$= \frac{18-9}{5} = 1.80$$

En este caso, resulta conveniente redondear el intervalo a 2.0.

Tabla 2.5 Distribución de frecuencias acumuladas de las rentas mensuales de departamentos

Renta mensual (miles de Pesos)	Límites exactos de clase	Número de departamentos	Frecuencia acumulada (fa)
$350-379	$349.50-379.50	3	3
380-09	379.50-409.50	8	11
410-439	409.50-439.50	10	21
440-469	439.50-469.50	13	34
470-499	469.50-499.50	33	67
500-529	499.50-529.50	40	107
530-559	529.50-559.50	35	142
560-589	559.50-589.50	30	172
590-619	589.50-619.50	16	188
620-649	619.50-649.50	12	200
		Total 200	

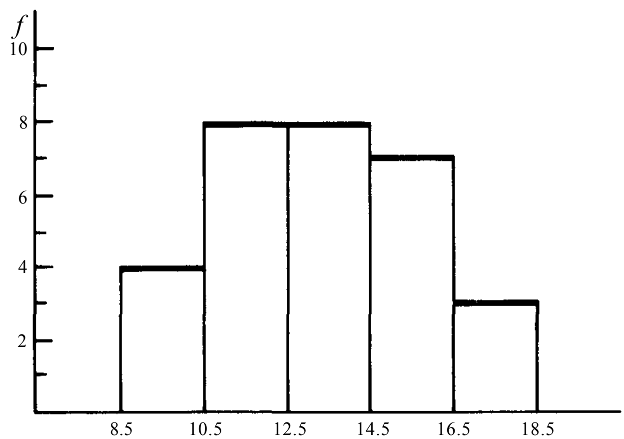


Rentas mensuales (límites exactos de clases), miles de pesos.

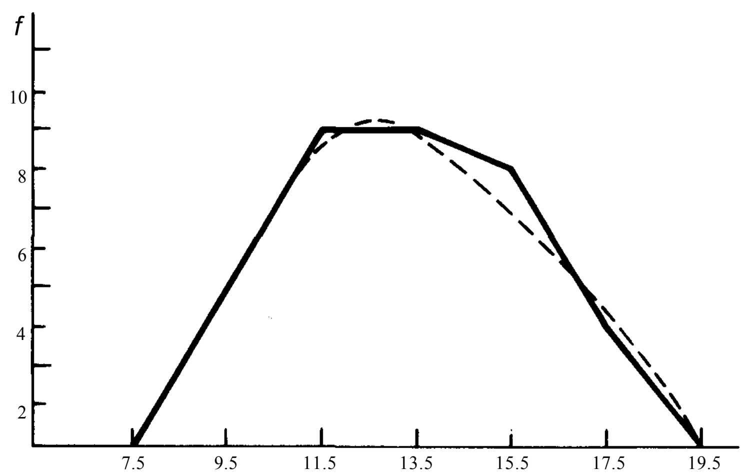
Fig. 2-12

Tabla 2.6 Tiempo de ensamble para 30 empleados (minutos)

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



Tiempo, minutos
Fig. 2-13



Tiempo, minutos
Fig. 2-14

2.8 Construya la distribución de frecuencias para los datos de la Tabla 2.6 utilizando un intervalo de clase de 2.0 para todas las clases, y fijando el límite nominal inferior de la primera clase en 9 minutos.

En la Tabla 2.7 aparece la distribución solicitada.

2.9 En la Tabla 2.7, con referencia a la clase que tiene el menor número de empleados, identifique (a) sus límites exactos, (b) su intervalo, (c) su punto medio.

(a) 16.5-18.5, (b) 18.5-16.5 - 2.0, (c) 16.5 + 2.0/2 =17.5.

2.10 Construya un histograma para la distribución de frecuencias de la Tabla 2.7

El histograma se presenta en la figura 2-13.

Tabla 2.7 Distribución de frecuencias para los tiempos de ensamble

Tiempo, minutos	Número de empleados
9-10	4
11-12	8
13-14	8
15-16	7
17-18	3
	Total 30

2.11 Construya un polígono de frecuencias y una curva de frecuencias para los datos de la Tabla 2.7.

El polígono de frecuencias y la curva de frecuencias aparecen en la figura 2-14.

2.12 Describa la curva de frecuencias de la figura 2-14 en términos de asimetría.

La curva de frecuencias es casi simétrica, pero tiene una ligera asimetría positiva.

2.13 Construya una distribución de frecuencias acumuladas para la distribución de frecuencias de los tiempos de ensamble de la Tabla 2.7, utilizando límites exactos para identificar cada clase e incluyendo porcentajes acumulados, así como también frecuencias acumuladas en la tabla.

Véase la Tabla 2.8, en la cual se presenta la distribución de frecuencias acumuladas.

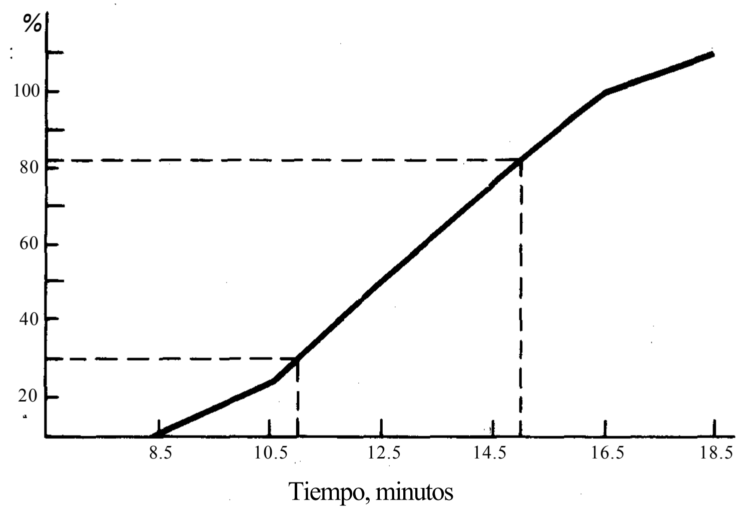


Fig. 2-15

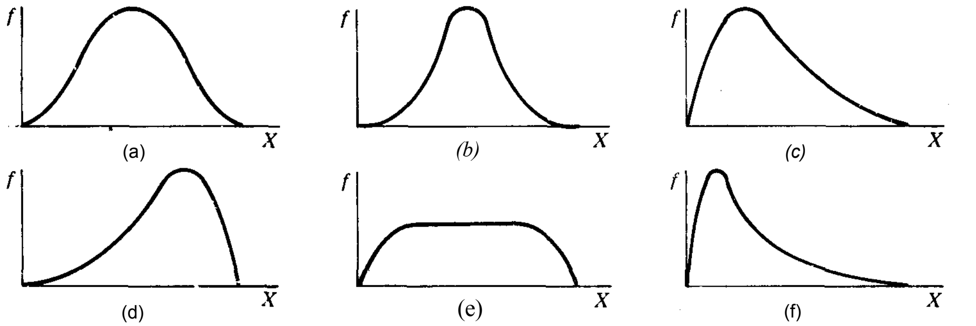


Fig. 2-16

Tabla 2.8 Distribución de frecuencias acumuladas para los tiempos de ensamble.

Tiempo, minutos	/	*fa*	% acum.
8.5-10.5	4	4	13.3
10.5-12.5	8	12	40.0
12.5-14.5	8	20	66.7
14.5-16.5	7	27	90.0
16.5-18.5	3	30	100.0

2.14 Con referencia a la distribución de frecuencias acumuladas de la Tabla 2.8.
(a) Construya una ojiva porcentual para esos datos
(b) ¿En qué punto percentil se encuentra un tiempo de ensamble de 15 minutos?
(c) ¿Cuál es el tiempo de ensamble del vigésimo percentil de la distribución?

(a) La ojiva se presenta en la figura 2-15.
(b) Tai como se identifica mediante las líneas punteadas de la porción superior de la figura, el percentil aproximado para 15 minutos de tiempo de ensamble es 72.
(c) Tal como se muestra con las líneas punteadas de la parte inferior de la figura, el tiempo aproximado en el percentil vigésimo es de 11 minutos.

FORMAS DE CURVAS DE FRECUENCIAS

2.15 Suponiendo que la curva de frecuencias (a) de la figura 2-16 es, al mismo tiempo, simétrica y mesokúrtica, describa las curvas (b), (c), (d), (e) y (f) en términos de asimetría y kurtosis.

La curva (b) es simétrica y ieptokúrtica; la curva (c) tiene asimetría positiva y es mesokúrtica. La curva *(d)* tiene asimetría negativa y es mesokúrtica; la curva (e) es simétrica y platikúrtica y la curva *(f)* es asimétrica positiva y es ieptokúrtica.

DISTRIBUCIONES DE FRECUENCIAS RELATIVAS

2.16 Utilizando las instrucciones revisadas en la sección 2.6, determine (a) las frecuencias relativas y (b) las proporciones acumuladas para los datos de la Tabla 2.9.

En la Tabla 2.10 se presentan las frecuencias relativas y las proporciones acumuladas para los datos de la Tabla 2.9.

(a) Véase la figura 2-17.

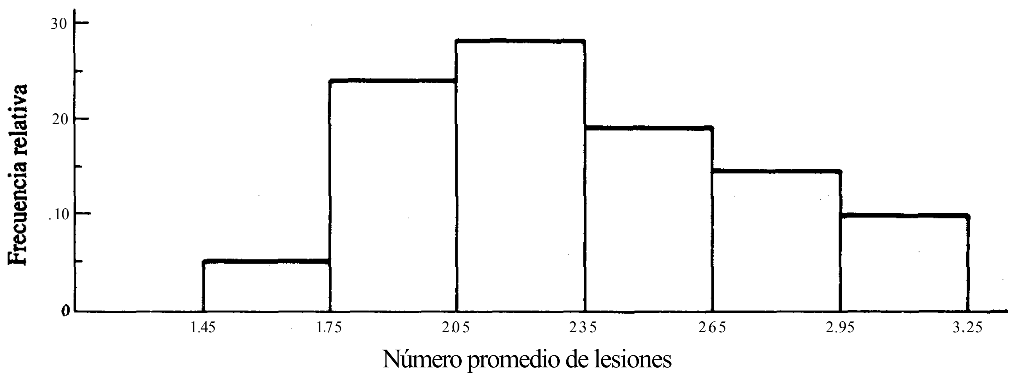


Fig. 2-17

(b) Véase la figura 2-18.

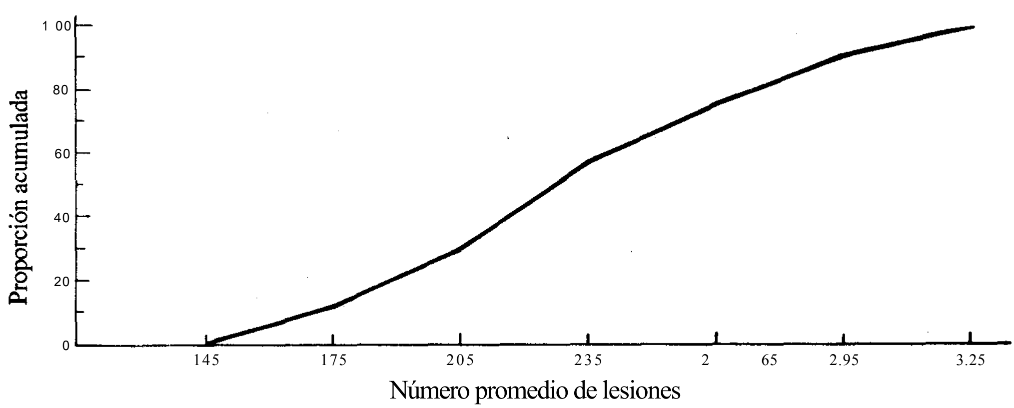


Fig. 2-18

Tabla 2.9 Número promedio de lesiones por millar de horas-hombre en una industria específica.

Número promedio de lesiones por millar de horas-hombre	Número de empresas
1.5-1.7	3
1.8-2.0	12
2.1-2.3	14
2.4-2.6	9
2.7-2.9	7
3.0-3.2	5
	Total 50

Tabla 2.10 Frecuencias relativas y proporciones acumuladas para el número promedio de lesiones

Número promedio de lesiones por millar de horas-hombre	Número de empresas	(a) Frecuencia relativa	(b) Proporción acumulada
1.5-1.7	3	0.06	0.06
1.8-2.0	12	0.24	0.30
2.1-2.3	14	0.28	0.58
2.4-2.6	9	0.18	0.76
2.7-2.9	7	0.14	0.90
3.0-3.2	5	0.10	1.00
	Total 50	Total 1.00	

2.17 Con referencia a la Tabla 2.10, construya (a) un histograma para la distribución de frecuencias relativas y (b) una ojiva para las proporciones acumuladas

2.18 (a) Con referencia a la Tabla 2.10, ¿qué proporción de empresas se encuentra en la categoría de haber tenido un promedio de al menos 3.0 lesiones por millar de horas-hombre? (b) ¿Qué porcentaje de empresas se encontraba en o por debajo de un promedio de 2.0 lesiones por millar de horas-hombre?

(a) 0.10, (b) 6% + 24% - 30%

2.19 (a) Con referencia a la Tabla 2.10, ¿cuál es el valor percentil asociado con un promedio de 2.95 (aproximadamente 3.0) lesiones por millar de horas-hombre? (b) ¿Cuál es el número promedio de accidentes en el percentil 58?

(a) Percentil 90, (b) 2.35

2.20 Mediante interpolación gráfica en una ojiva pueden determinarse los percentiles acumulados para diversos valores de la variable y viceversa. Con respecto a la figura 2-18, (a) ¿cuál es el percentil aproximado asociado con un promedio de 2.5 accidentes? (6) ¿Cuál es el número promedio aproximado de accidentes en el punto percentil cincuenta?

(a) Percentil 65 (Esta es la altura aproximada de la ojiva que corresponde a 2.50 sobre el eje horizontal.)
(b) 2.25 (Este es el punto aproximado sobre el eje horizontal que corresponde a una altura de 0.50 en la ojiva.)

LA DISTRIBUCIÓN DE FRECUENCIAS DEL TIPO "Y MENOR QUE"

2.21 Identifique los límites exactos de clase para los datos de la Tabla 2.11.

Tabla 2.11 Tiempo que se requiere para procesar órdenes de alimentos en un restaurante

Tiempo, minutos	Número de órdenes
5 y menor que 8	10
8 y menor que 11	17
11 y menor que 14	12
14 y menor que 17	6
17 y menor que 20	2
	Total 47

De la Tabla 2.11,

Tabla 2.12 Tiempo que se requiere para procesar y preparar las órdenes de alimentos (con límites exactos de clase)

Tiempo, minutos	Número de órdenes	Limites exactos de clase
5 y menor que 8	10	5.0-8.0
8 y menor que 11	17	8.0-11.0
11 y menor que 14	12	11.0-14.0
14 y menor que 17	6	14.0-17.0
17 y menor que 20	2	17.0-20.0
	Total 47	

2.22 Construya un polígono de frecuencias para la distribución de frecuencias de la Tabla 2.12. El polígono de frecuencias aparece en la figura 2-19.

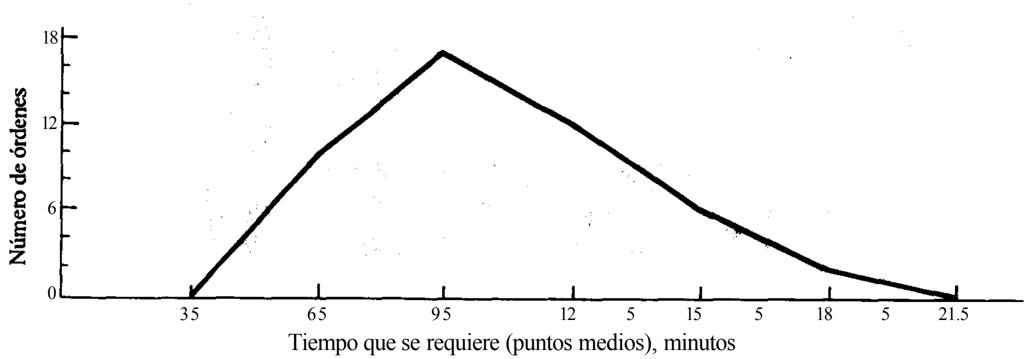


Fig. 2-19

GRÁFICAS DE BARRAS

2.23 En la Tabla 2.13 se reporta el personal ocupado en la industria maquiladora de exportación. *(Fuente:* revista Expansión, 25 de octubre de 1989, p. 24.) Construya una gráfica de barras verticales que ilustre el total del personal que laboraba en la industria maquiladora de 1984 a 1988.

Tabla 2.13 Personal ocupado en la industria maquiladora de exportación en México, (Fuente: revista Expansión, 25 de octubre de 1989)

Año	Total	Obreros	Técnicos	Empleados
1984	199,684	165,505	22,381	11,798
1985	211,968	173,874	25,042	13,052
1986	249,833	203,894	30,367	15,572
1987	305,253	248,638	36,740	19,875
1988	369,489	301,379	44,312	23,798

La gráfica de barras aparece en la figura 2-20.

2.24 Construya una gráfica de barras componentes para los datos de la Tabla 2.13, de manera que pueda apreciarse la división entre obreros (O), técnicos (T) y empleados (E), para cada año.

En la figura 2-21 se presenta la gráfica de barras componentes para los datos de la Tabla 2.13.

GRÁFICAS DE LÍNEA

2.25 Construya una gráfica de linea para el total de personal empleado que se reporta en la Tabla 2.13.

La gráfica de líneas se muestra en la figura 2-22.

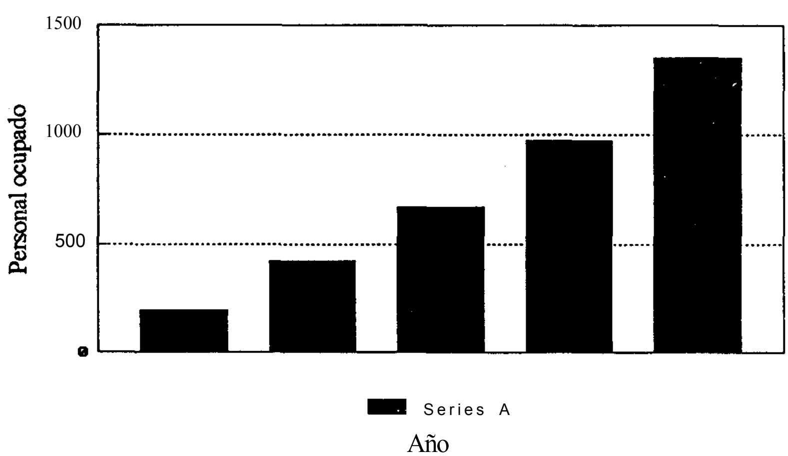


Fig. 2-20

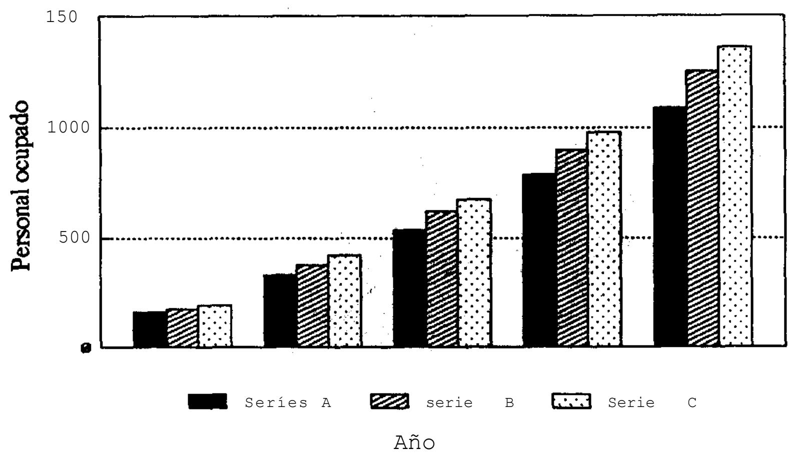


Fig. 2-21

GRÁFICAS DE PASTEL

2.26 La Tabla 2.14 se basa en datos publicados en los Indicadores Económicos del Banco de México, y son datos preliminares correspondientes a diciembre de 1988. Construya una gráfica de pastel porcentual de las exportaciones de México, de acuerdo con su tipo.

Tabla 2.14 Exportaciones mexicanas en diciembre de 1988, millones de dólares

Categoría	Cantidad
Petroleras	$560.1
Agropecuarias	143.0
Extractivas	49.9
Manufactureras	951.1
	Total 1,704.1

La gráfica de pastel se presenta en la figura 2-23

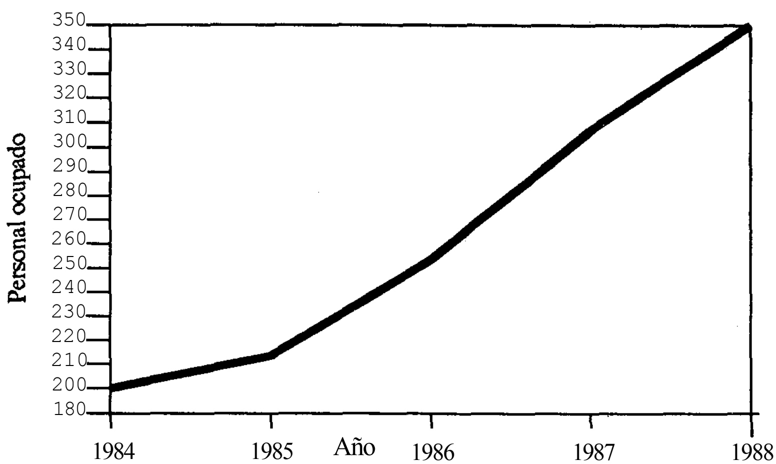


Fig. 2-22

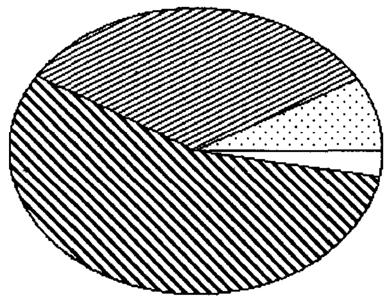


Fig. 2-23

RESULTADOS POR COMPUTADORA

2.27 Utilice un paquete de cómputo para formar una distribución de frecuencias para producir un histograma de los datos no agrupados de la Tabla 2.6, que contiene una muestra de los tiempos que 30 empleados requirieron para llevar a cabo una tarea de ensamble.

(a) Identifique el punto medio de la primera clase.
(b) Determine el tamaño del intervalo de clase.
(c) Determine los límites de clase exactos, superior e inferior, de la primera clase.

En la figura 2-24 se presentan los datos introducidos a la computadora y los resultados.

```
MTB ❯ SET ASSEMBLY TIMES INTO Cl
DOTA❯   10   14   15   13   17   16   12   14   11   13   15   18    9
DOTO❯   14   14    9   15   11   13   11   12   10   17   16   12   11
DOTO❯   16   12   14   15
DOTO❯ END
MTB ❯ NAME FOR Cl IS 'TIME'
MTB ❯ HISTOGRAM FOR 'TIME'

Histogram of TIME    N = 30

Midpoint   Count
       9       2  **
      10       2  **
      11       4  ****
      12       4  ****
      13       3  ***
      14       5  *****
      15       4  ****
      16       3  ***
      17       2  **
      18       1  *
```

Fig. 2-24 Resultado de Minitab para el problema 2.27.

(a) 9.0
(b) Con referencia *a* los puntos medios de las clases adyacentes: 10.0 - 9.0 - 1.0.
(c) 8.5-9.5

2.28 Con referencia al problema anterior, repita el análisis especificando que el punto medio de la primera clase debo quedar en 10.0, con un intervalo de clase de 2.0.

(a) Determine los límites exactos de clase, superior e inferior, de la primera clase.

(b) Identifique el tipo de distribución de frecuencia que se forma.

(c) Determine los límites nominales de clase, inferior y superior, de la primera clase.

En la figura 2-25 se presentan los datos introducidos a la computadora, y los resultados obtenidos utilizando Minitab.

(a) 9.0-11.0.

(b) Aunque no resulta evidente, se trata de una distribución de frecuencias del tipo "y menor que". Por ejemplo, revisando manualmente los resultados, puede observarse que los valores observados de "9", que se encuentran en el límite exacto inferior de la primera clase, se encuentran efectivamente incluidos en esa clase. Sin embargo, los valores observados de "11", que se encuentran en el límite exacto superior de la primera clase, no se incluyen en esa clase.

(c) 9 y menos que 11.

Problemas complementarios

DISTRIBUCIONES DE FRECUENCIAS, INTERVALOS DE CLASE Y MÉTODOS GRÁFICOS ASOCIADOS

2-29 En la Tabla 2.15 se presenta una distrit ución de frecuencias del rendimiento de gasolina en 25 viajes de autos de una compañía muestreados al azar, (a) ¿Cuáles son los límites nominales superior e inferior de la última clase? (b) ¿Cuáles son los límites exactos superior e interior de la última clase? (c) ¿Cuál es el intervalo de clase que se utiliza? *(d)* ¿Cuál es el punto medio de la última clase? (e) Suponga que se determina que el rendimiento por litro en un viaje fue de 2.7. Señale los límites inferior y superior de la clase en la que se incluye este resultado.

Resp. (a) 20.0 y 21.9, (b) 19.95 y 21.95, (c) 2.0, (d) 20.95, (e) 12.0 y 13.9.

```
MTB > HISTOGRAM FOR 'TIME';
SUBO>START AT 10.00;
SUBO> INCREMENT = 2.0.

Histogram of TIME   N = 30

Midpoint   Count
   10.00       4  ****
   12.00       8  ********
   14.00       8  ********
   16.00       7  *******
   18.00       3  ***
```

Fig. 2-25 Resultado de Minitab para el problema 2.28.

Tabla 2.15 Rendimiento de 25 viajes de automóviles de una compañía automotriz

Kilómetros por litro	Número de viajes
10.0-11.9	3
12.0-13.9	5
14.0-15.9	10
16.0-17.9	4
18.0-19.9	2
20.0-21.9	1
	Total 25

2.30 Construya un histograma para los datos de la Tabla 2.15.

2.31 Construya un polígono de frecuencias y una curva de frecuencias para los datos de la Tabla 2.15.

2.32 Describa la curva de frecuencias que construyó en el problema 2.31, desde el punto de vista de la asimetría.

Resp. Parece que la curva de frecuencias muestra cierta asimetría positiva.

2.33 Construya una distribución de frecuencias acumuladas para los datos de la Tabla 2.15, y construya una ojiva para presentar esta distribución en forma gráfica.

2.34 En la Tabla 2.16 se presentan las cantidades otorgadas en 40 préstamos personales. Suponga que se desea ordenar esas cantidades en una distribución de frecuencias, que tenga un total de 7 clases. Suponiendo intervalos de clase iguales, ¿cuál sería un intervalo de clase conveniente para esta distribución de frecuencias?

Resp. $400 000.

2.35 Construya la distribución de frecuencias para los datos de la Tabla 2.16, comenzando la primera clase con un límite inferior de $300 000 y utilizando un intervalo de clase de $ 400 000

2.36 Elabore un histograma para la distribución de frecuencias construida en el problema 2.35.

2.37 Construya un polígono de frecuencias y una curva de frecuencias para la distribución construida en el problema 2.35.

2.38 Describa la curva de frecuencias que se construyó en el problema 2.37, en términos de asimetría.

Resp. Se nota claramente que la curva de frecuencias tiene asimetría positiva.

Tabla 2.16 Montos de 40 préstamos personales (en miles de pesos)

$932	$1,000	$356	$2,227
515	554	1,190	954
452	973	300	2,112
1,900	660	1,610	445
1,200	720	1,525	784
1,278	1,388	1,000	870
2,540	851	1,890	630
586	329	935	3,000
1,650	1,423	592	334
1,219	727	655	1,590

2.39 Construya una distribución de frecuencias acumuladas para la distribución de frecuencias que se elaboró en el problema 2.35, y trace una ojiva para esos datos.

FORMA DE LAS CURVAS DE FRECUENCIAS

2.40 Describa cada una de las siguientes curvas en términos de asimetría o kurtosis, según sea apropiado: (a) Una curva de frecuencias que tiene una "cola" hacia la derecha, (b) una curva de frecuencias relativamente puntiaguda, (c) una curva de frecuencias que es relativamente plana, *(d)* una curva de frecuencias que tiene una "cola" hacia la izquierda.

Resp. (a) Asimetría positiva, (b) leptokúrtica, (c) platikúrtica, (d) asimetría negativa.

DISTRIBUCIONES DE FRECUENCIAS RELATIVAS

2.41 Construya una tabla de frecuencias relativas para la distribución de frecuencias que se presenta en la Tabla 2.17.

Tabla 2.17 Vida media de herramientas de corte en un proceso industrial

Horas antes del reemplazo	Número de herramientas
0.0-24.9	2
25.0-49.9	4
50.0-74.9	12
75.0-99.9	30
100.0-124.9	18
125.0-149.9	4
	Total 70

2.42 Construya un histograma para la distribución de frecuencias relativas que se preparó en el problema 2.41.

2.43 Con referencia a la Tabla 2.17, (a) ¿qué porcentaje de herramientas de corte duraron cuando menos 125 horas? (b) ¿Qué porcentaje de herramientas de corte tuvieron una vida útil de cuando menos 100 horas?

Resp. (a) 6%, (b) 31 %

2.44 Prepare una tabla de proporciones acumuladas para la distribución de frecuencias de la Tabla 2.17.

2.45 Con referencia a la tabla que se construyó en el problema 2.44, (a) ¿cuál es el tiempo útil de las herramientas asociado con el percentil 26 de la distribución? (b) ¿Cuál es el percentil asociado con un tiempo útil de las herramientas de aproximadamente 100 horas?

Resp. (a) 74.95 horas ≅ 75 horas, (6) percentil 69

2.46 Construya una ojiva para las proporciones acumuladas que se calcularon en el problema 2.44.

2.47 Con referencia a la ojiva que se preparó en el problema 2.46, determine en forma aproximada los siguientes valores, mediante interpolación gráfica: (a) el tiempo útil de las herramientas en el percentil 50 de la distribución, (b) el percentil asociado con un tiempo útil de 60 horas.

Resp. (a) aprox. 89 horas, (b) aprox. el percentil 16

LA DISTRIBUCIÓN DE FRECUENCIAS DEL TIPO "Y MENOR QUE"

2.48 Con referencia a la distribución de frecuencias de la Tabla 2.18, determine (a) el límite nominal inferior de la primera dase, (b) el límite nominal superior de la primera clase, (c) el límite exacto inferior de la primera clase, (d) el límite exacto superior de la primera clase, (e) el punto medio de la primera clase.

Resp. (a) 18, (b) 20, (c) 18.0, (d) 20.0, (e) 19.0

Tabla 2.18 Edades de una muestra de solicitantes para un programa de capacitación

Edad	Número de solicitantes
18 y menor que 20	5
20 y menor que 22	18
22 y menor que 24	10
24 y menor que 26	6
26 y menor que 28	5
28 y menor que 30	4
30 y menor que 32	2
	Total 50

2.49 Construya un polígono de frecuencias para la distribución de frecuencias de la Tabla 2.18.

GRÁFICAS DE BARRAS

2.50 Los datos de la Tabla 2.19 fueron tomados de la revista Expansión, correspondiente al 22 de noviembre de 1989. Construya una gráfica de barras verticales para estos datos.

Tabla 2.19 Miles de toneladas métricas de yeso exportadas por la Compañía Occidental Mexicana, 1984-1988.

Año	Miles de toneladas métricas de yeso exportadas
1984	1830
1985	1940
1986	2161
1987	2004
1988	2211

GRÁFICAS DE LÍNEA

2.51 Construya una gráfica de línea para las toneladas métricas de yeso exportadas por la Compañía Occidental Mexicana que se presentan en la Tabla 2.19.

GRÁFICAS DE PASTEL

2.52 Los datos de la Tabla 2.20 fueron tomados de los Indicadores Económicos del Banco de México, publicados en noviembre de 1989, y muestran las fuentes de los ingresos del gobierno mexicano en noviembre de 1988 (datos preliminares). Construya una gráfica de pastel porcentual con esas fuentes.

Tabla 2.20 Ingresos totales del gobierno mexicano en noviembre de 1988, miles de millones de pesos

Ingresos provenientes de Pemex		\$12,986.4
Ingresos por tributaciones		46,347.0
Otros ingresos no tributarios		10,101.3
	Total	\$69,434.7

RESULTADOS POR COMPUTADORA

2.53 Utilice algún paquete de cómputo para formar una distribución de frecuencias y producir un histograma de los datos no agrupados de la Tabla 2.16, para las cantidades otorgadas en 40 préstamos personales.
(a) Identifique el punto medio de la primera clase.
(b) Determine el intervalo de clase que se utiliza.
(c) Determine los límites de clase exactos, superior e inferior, para la primera clase.

Resp (a) Utilizando Minitab, 400 000, (b) 400 000, (c) 400 000-600 000

2.54 Con referencia al problema 2.53, vuelva a realizar el análisis especificando que el punto medio de la primera clase es 500 000, con un intervalo de clase de 400 000.
(a) Determine los limites de ciase exactos, superior e inferior, para la primera clase.
(b) Identifique el tipo de distribución de frecuencias que se forma.
(c) Determine los limites nominales de clase, superior e inferior, para la primera dase.

Resp. (a) 300 000-700 000, (b) Distribución del tipo "y menor que", (c) 300 000 y menor que 700 000

Descripción de datos de negocios: Medidas de posición

3

3.1 MEDIDAS DE POSICIÓN EN CONJUNTOS DE DATOS

Una medida de posición es un valor que se calcula para un grupo de datos y que se utiliza para describirlos de alguna manera. Normalmente se desea que el valor sea representativo de todos los valores incluidos en el grupo y, por ello, se desea alguna clase *de promedio.* En sentido estadístico, un "promedio" es una *medida de tendencia central* para un conjunto de valores. En este capitulo se cubren los diversos procedimientos estadísticos que se refieran a medidas de posición.

3.2 LA MEDIA ARITMÉTICA

La *media aritmética,* o *promedio aritmético,* se define como la división de la suma de todos los valores entre el número de valores.

En estadística es normal representar una medida descriptiva de una población, o *parámetro poblacional,* mediante letras griegas, en tanto que se utilizan letras romanas para las medidas descriptivas de muestras, o *estadísticas muestrales.* Así, la media aritmética para una población de valores se presenta mediante el símbolo μ (que se pronuncia "mu"), en tanto que la media aritmética de una muestra de valores se representa mediante el símbolo X (que se lee "x barra") Las fórmulas para la media de una población y de una muestra son:

$$\mu = \frac{\Sigma X}{N} \tag{3.1}$$

$$\bar{X} = \frac{\Sigma X}{n} \tag{3.2}$$

En cuanto a operaciones se refiere, las dos fórmulas son idénticas; en ambos casos se suman todos los valores (ΣX) y después se divide este total entre el número de valores que son. Sin embargo, la diferencia en los denominadores se debe a que en análisis estadístico, la *N* normalmente indica el número de elementos de una población, en tanto que la *n* señala el número de elementos de una muestra.

EJEMPLO 1. En determinado mes, 8 vendedores de artículos electrónicos vendieron los siguientes números de aparatos: 8, 11, 5, 14, 8, 11, 16, 11. Considerando a este mes como a la población estadística que interesa, el número promedio de unidades vendidas es:

$$\mu = \frac{\Sigma X}{N} = \frac{84}{8} = 10.5 \text{ unidad}$$

Nota: Cuando se reportan estas medidas, por lo general se les menciona con un dígito adicional al nivel original de medición.

3.3 LA MEDIA PONDERADA

La *media ponderada* o *promedio ponderado* es una media aritmética, en la cual se considera a cada uno de tos valores de acuerdo con su importancia en el grupo. Las fórmulas para la media ponderada muestral y poblacional son idénticas:

$$\mu_p \text{ o } X_p = \frac{\Sigma(pX)}{\Sigma p} \tag{3.3}$$

En términos de operaciones, cada uno de ios valores del grupo (X) se multiplica por el tactor de ponderación apropiado (p) y después se suman estos productos y la suma se divide entre los pesos (o ponderaciones).

EJEMPLO 2. En una compañía que maneja 4 productos, los márgenes de utilidad correspondientes a cada uno de ellos durante el año fiscal anterior fueron; producto A, 4.2%; producto B, 5.5%; producto C, 7.4%; y producto D, 10.1%: El margen de utilidad promedio, *no ponderado,* es

$$\mu = \frac{\Sigma X}{N} = \frac{27.2}{4} = 6.8\%$$

Sin embargo, este promedio no ponderado es incorrecto porque se vendieron cantidades distintas de los 4 productos. Suponiendo los totales de ventas que aparecen en la Tabla 3.1, se encuentra que el promedio ponderado describe en forma correcta el promedio global.

Tabla 3.1 Margen de utilidad y volumen de ventee para 4 productos

Producto	Margen de utilidades	Ventas	pX
A	4.2%	\$30,000,000	\$1,260,000
B	5.5	20,000,000	1,100,000
C	7.4	5,000,000	370,000
D	10.1	3.000,000	303,000
		ΣpX = \$58,000,000	$\Sigma(pX)$ = \$3,033,000

$$\mu_p = \frac{\Sigma(pX)}{\Sigma p} \quad \frac{\$3,033,000}{\$58,000,000} = 5.2\%$$

3.4 LA MEDIANA

La mediana de un grupo de datos es el valor del dato que ocupa un lugar de cuando se les agrupa a todos en orden ascendente o descendente. Para un grupo con un número par de elementos, se supone que la mediana se encuentra a la mitad entre los dos valores adyacentes al centro. Cuando el conjunto de datos contiene un número grande de valores, resulta útil la siguiente fórmula para determinar la posición de la mediana en el conjunto ordenado:

$$\text{Med} = X_{[(n/2)+(1/2)]} \tag{3-4}$$

EJEMPLO 3. Los 8 vendedores que se describieron en el ejemplo 1 vendieron el siguiente número de aparatos, en orden ascendente: 5, 8, 8,11,11,11,14,16. El valor de la mediana es:

$$\text{Med} = X_{[(n/2)+(1/2)]} = X_{[(8/2)+(1/2)]} = X_{4.5} = 11.0$$

El valor de la mediana se encuentra entre los valores cuarto y quinto de este conjunto ordenado. Como los dos son iguales a "11" en este caso, la mediana es 11.0.

3.5 LA MODA

La *moda* es el valor que se presenta con mayor frecuencia en un conjunto de datos. A una distribución que tiene una sola moda se le denomina *unimodal.* Para un conjunto de datos poco numerosos, en los que no se repite ningún valor, no existe moda. Cuando dos valores no adyacentes tienen frecuencias máximas similares, se dice que la distribución es *bimodal.* A las distribuciones de mediciones que tienen varias modas se les denomina *multimodales.*

EJEMPLO 4 Los 8 vendedores que se describieron en el ejemplo 1 vendieron el siguiente número de aparatos: 8,11, 5,14,8,11,16 y 11. La moda para este grupo de valores es el valor que tiene la mayor frecuencia, o moda = 11.

3.6 RELACIÓN ENTRE LA MEDIA, LA MEDIANA Y LA MODA

Las diferencias entre los valores de la media, la mediana y la moda permiten saber la forma de la curva de frecuencias en términos de asimetría. Para una distribución unimodal simétrica, el valor de la media, la mediana y la moda es igual [véase la Fig. 3-1 *(a)].* Para una distribución asimétrica positiva, la media es el mayor valor de los tres y la mediana es mayor que la moda pero menor que la media [véase Fig. 3-1 (6)]. Para una distribución asimétrica negativa, la media es el menor valor de los tres y la mediana es inferior a la moda pero mayor que la media [véase la Fig. 3-1 (c)]. El coeficiente de asimetría de Pearson, que se describe en la sección 4.9, es una medida conocida de asimetría que utiliza la diferencia observada entre la media y la mediana de un grupo de valores.

EJEMPLO 5. Para los datos de ventas considerados en los ejemplos 1,3 y 4, puede observarse que la media es 10.5, la mediana es 11.0 y la moda es de 11.0 unidades, lo cual indica que la distribución tiene una ligera asimetría negativa (está sesgada hacia la izquierda).

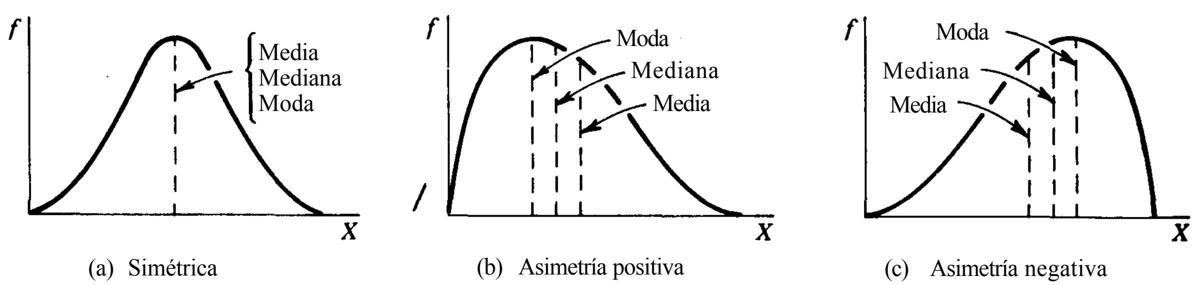


Fig. 3-1

3.7 CUARTILES, DECILES Y PERCENTILES

Los cuartiles, deciles y percentiles se parecen mucho a la media porque también gubdividen una distribución de mediciones de acuerdo con la proporción de frecuencias observadas. Mientras que la mediana divide a la distribución en dos mitades, los cuartiles la dividen en cuatro cuartos, los deciles la dividen en diez décimos y los puntos percentiles la dividen en cien partes. La fórmula *3.4* de la mediana, modificada de acuerdo con el punto fraccionario de interés es, por ejemplo,

$$Q_1 \text{ (primer cuartil)} = X_{[(n/4)+(1/2)]} \tag{3.5}$$

$$D_3 \text{ (tercer decil)} = X_{[(3n/10)+(1/2)]} \tag{3.6}$$

$$P_{70} \text{ (percentil 70)} = X_{[(70n/100)+(1/2)]} \tag{3.7}$$

EJEMPLO 6. Los ocho vendedores que se describieron en el ejemplo 1 vendieron las siguientes cantidades de aparatos electrónicos, en orden ascendente: 5,8,8,11,11,11,14 y 16. Encuentre la posición del tercer cuartil para está distribución.

$$Q_3 = X_{[(3n/4)+(1/2)]} = X_{[(24/4)+(1/2)]} = X_{6.5} = 12.5$$

En este caso el valor se encuentra entre los valores sexto y séptimo del grupo ordenado.

3.8 LA MEDIA ARITMÉTICA PARA DATOS AGRUPADOS

Cuando se agrupan datos en una distribución de frecuencias, se utiliza el punto medio de cada clase como aproximación de todos los valores contenidos en ella. El punto medio se representa con el símbolo X_0 ,en donde el subíndice c se debe a "clase*, y se utiliza la letra *f* para representar la frecuencia observada de valores en la clase respectiva. Por ello, las fórmulas para la media de la población y de la muestra para datos agrupados son:

$$\mu = \frac{\Sigma(fX_c)}{\Sigma f} \quad \text{o, en forma más simple} \quad \mu = \frac{\Sigma(fX)}{N} \tag{3.8}$$

$$\bar{X} = \frac{\Sigma(fX_c)}{\Sigma f} \quad \text{o, en forma más simple} \quad \bar{X} = \frac{\Sigma(fX)}{n} \tag{3.9}$$

En términos de cálculo, ambas fórmulas señalan que cada punto medio de clase se multiplica (X_c) por la frecuencia de clase correspondiente *(f),* luego se suman estos productos (Σ) para después dividir esta suma entre el número total de observaciones (Σf) representadas en la distribución de frecuencias.

EJEMPLO 7. Los datos agrupados de la Tabla 3.2 provienen de la distribución de frecuencias que se presentó en la sección 2.1. La media se representa con el símbolo $\bar{X}$ debido a que se supone que el grupo de trabajadores es una muestra de una población mayor.

Tabla 3.2 Salarios diarios de 100 trabajadores no calificados (redondeados al peso más cercano)

Salario diario	Punto medio de clase *(X)*	Número de trabajadores *(f)*	*fX*
\$2400-2599	\$2499.50	7	\$17496.50
2600-2799	2699.50	20	53990.00
2800-2999	2899.50	33	95683.50
3000-3199	3099.50	25	77487.50
3200-3399	3299.50	11	36294.50
3400-3599	3499.50	4	13998.00
		Total 100	$\Sigma(fX)$ - \$294,950.00

$$\bar{X} = \frac{\Sigma(fX)}{n} = \frac{294950}{100} = \$2949.50$$

3.9 LA MEDIANA PARA DATOS AGRUPADOS

Para datos agrupados, en primer lugar es necesario determinar la *clase* que contiene el valor de la mediana, para después determinar la posición de la mediana dentro de la clase mediante interpolación. La clase que contiene la mediana es la primera cuya frecuencia acumulada iguala o excede la mitad del total de observaciones. Una vez que se identiflca esta clase, se determina el valor específico mediante la siguiente fórmula:

$$\text{Med} = L_l + \left(\frac{\frac{N}{2} - fa_A}{f_c}\right) i \tag{3.70}$$

en donde L_l límite exacto inferior de la clase que contiene la mediana
N número total de observaciones en la distribución de frecuencias (*n* para una muestra)
fa_A la frecuencia acumulada de la dase que precede ("antes") a la clase que contiene la mediana
f_c número de observaciones en la clase que contiene la mediana
i tamaño del intervalo de clase

EJEMPLO 8. En la Tabla 3.2 se repiten los datos agrupados de la Tabla 3.3. En este caso, la clase que contiene la mediana es la que incluye el valor 100/2 = 50. La primera cuya frecuencia acumulada es igual o superior a 50 es la clase que tiene los límites nominales \$2800-2999; así, la interpolación para determinar el valor específico de la mediana se realiza en esta clase.

Tabla 3.3 Salarios semanales de 100 trabajadores no calificados

Salario diario	Número de trabajadores *(f)*	Frecuencia acumulada *(fa)*
\$2400-2,599	7	7
2600-2,799	20	27
2800-2,999	33	60
3000-3,199	25	85
3200-3,399	11	96
3400-3,599	4	100
	Total 100	

$$\text{Med} = L_l + \left(\frac{\frac{n}{2} - fa_A}{f_c} \right) i = 2977.5 + \left(\frac{50 - 27}{33} \right) 200 = \$2938.89$$

3.10 LA MODA PARA DATOS AGRUPADOS

Para datos agrupados en una distribución de frecuencias con intervalos de clase iguales, primero se identifica la *clase* que contiene la moda determinando cuál de ellas tiene el mayor número de observaciones. Algunos profesionales de la estadística consideran que la moda es el punto medio de la clase modal. Sin embargo, la mayoría de ellos interpolan dentro de la clase modal, de acuerdo con la siguiente fórmula:

$$\text{Moda} = L_l + \left(\frac{d_1}{d_1 + d_2} \right) i \tag{3.11}$$

en donde L_l = limite exacto inferior de la clase que contiene la moda
d_1 = diferencia entre la frecuencia de la clase modal y la frecuencia de la clase precedente
d_2 = diferencia entre la frecuencia de la clase modal y la frecuencia de la dase siguiente
i = tamaño del intervalo de clase

EJEMPLO 9. Con referencia a los datos agrupados de la Tabla 3.3, la clase modal es la que tiene limites nominales \$2800-2999. Por ello.

$$\text{Moda} = L_l + \left(\frac{d_1}{d_1 + d_2} \right) i = 2799.50 + \left(\frac{13}{13 + 8} \right) 200 = \$2923.31$$

3.11 CUARTILES, DECILES Y PERCENTILES PARA DATOS AGRUPADOS

Para datos agrupados, la fórmula de la mediana (3.70) se modifica de acuerdo con el punto fraccionario de interés. Al utilizar esta fórmula modificada, en *primer lugar se determina la clase que contiene él punto de interés,* de acuerdo con las frecuencias acumuladas, y después se lleva a cabo la interpolación como se hizo antes. Algunos ejemplos de fórmulas para este caso son:

$$Q_l \text{ (primer cuartil} = L_l + \left(\frac{\frac{n}{4} - fa_A}{f_c} \right) i \tag{3.72}$$

$$D_3 \text{ (tercer decil)} = L_l + \left(\frac{\frac{3n}{10} - fa_A}{f_c} \right) i \tag{3.13}$$

$$P_{70} \text{ (percentil 70)} = L_l + \left(\frac{\frac{70n}{100} - fa_A}{f_c} \right) i1 \tag{3.14}$$

EJEMPLO 10. Pueden utilizarse las frecuencias acumuladas de la Tabla 3.3 para determinar el punto percentil 90 de esos salarios. En este caso, debe observarse que la clase que contiene este valor es la primera, cuya frecuencia acumulada excede **90*n***/100, o 90. Esta es la clase que tiene los límites nominales \$3200-3399, por lo que el limite inferior que se utiliza en la fórmula es \$3195.

$$P_{90} = L_I + \left(\frac{\frac{90n}{2} - fa_A}{f_c}\right) i = 3199.50 + \left(\frac{90 - 85}{11}\right) 200 = \$3290.41$$

3.12 RESULTADOS POR COMPUTADORA

Las fórmulas especiales para datos agrupados que se describieron en las últimas secciones de este capítulo son importantes por dos razones. La primera es que los datos que se obtienen de fuentes secundarias, tales como reportes gubernamentales, por lo general se encuentran agrupados en distribuciones de frecuencias y, por ello, los cálculos con esos datos requerirían el uso de las fórmulas para datos agrupados. La segunda utilidad de estas fórmulas es que facilitan los cálculos con conjuntos grandes de datos agrupándolos primero en una distribución de frecuencias. Aun cuando los valores calculados de esta manera, tal como la media y la mediana, son *aproximaciones* de los respectivos valores de los datos no agrupados (debido a errores ocasionados por la agrupación, o el redondeo), por lo general se considera compensatorio el ahorro de tiempo que se obtiene al realizar los cálculos. Con la amplia disponibilidad de computadoras, se ha eliminado ya esta última razón para utilizar fórmulas de datos agrupados.

Existen paquetes de computación que permiten determinar diversas medidas de posición, o tendencia central. En el problema 3.24 se ¡lustra el cálculo de la media y la mediana utilizando una computadora.

Problemas resueltos

LA MEDIA, LA MEDIANA Y LA MODA

3.1 Las siguientes cifras son los importes del consumo de 15 personas en un restaurante, en orden ascendente: \$1000, 1000, 2500, 2500. 2500, 3500, 4000, 5300, 9000,12,500,13,500, 24,500,27,500, 30,900, y 41,000. Determine: (a) la media (b) la mediana y (c) la moda, para esos importes.

(a) $\bar{X} = \frac{\Sigma X}{n} = \frac{180000}{15} = \$1.21$

(b) Med $= X_{[(n/2)+(1/2)]} = X_{[(15/2)+(1/2)]} = X_8 =$ \$5300

(c) Moda = el valor mas frecuente =\$2500

3.2 ¿Cómo se describiría la distribución del problema 3.1 con respecto a la asimetría?

Siendo que la media es considerablemente mayor que la mediana y la moda, resulta evidente que la distribución de valores tiene asimetría positiva. Es decir, está sesgada a la derecha.

3.3 Si se le pidiera una descripción de los datos del problema 3.1 reportando la cantidad "típica" de importe por cliente del restaurante en la muestra, ¿qué medida de tendencia central, o promedio, reportaría? ¿Por qué?

Para una distribución muy asimétrica como ésta, la media aritmética no resulta descriptiva del valor típico del grupo. La elección entre la media y la moda como una mejor medida depende del grado en el que las observaciones se concentran en el punto modal. En este caso, el grado de concentración no es tan grande, por lo que "$25000" no es en realidad una descripción adecuada del importe típico. Por estas razones, parece que la mejor elección para estos datos es la mediana. Por ello, puede considerarse que "$53000" es típico en el sentido de que aproximadamente la mitad de los importes fueron inferiores a esta cantidad y la otra mitad superiores.

3.4 Una muestra de 20 trabajadores de una compañía pequeña obtuvieron los siguientes salarios para un mes determinado: $240.000,240,000,240.000,240,000,240,000,240,000,240,000,240,000,255,000,255,000,265,000 265,000, 280,000, 280,000, 290,000, 300,000, 305,000, 325,000. 330.000 y 340,000. Calcule (a) la media, (b) la mediana y (c) la moda, para este conjunto de salarios.

(a) $\bar{X} = \frac{\Sigma X}{n} = \frac{5\,410\,000}{20} = \$270\ 500$

(b) Median = $X_{[(n/2)+(1/2)]} = X_{[(20/2)+(1/2)]} = X_{10.5} = \$260000$

(c) Moda = el valor mas frecuente - $240 000

3.5 Para los salarios del problema 3.4, describa su distribución en términos de asimetría.

Como la media es mayor que la mediana y la moda, la distribución tiene asimetría positiva, o sesgada a la derecha.

3.6 Si estuviera usted en cada una de las siguientes situaciones, señale qué medida de "promedio" reportaría para los datos del problema 3.4, y en qué sentido puede considerarse "típico" cada valor, (a) Como vicepresidente responsable de las negociaciones colectivas con los trabajadores, (b) como presidente de los representantes de los trabajadores.

(a) Se indinaría por elegir la media aritmética para apoyar la posición de la compañía y podría usted mencionar que esta media toma en consideración todos los salarios individuales. También se podría (incorrectamente) plantear que el promedio aritmético es el único promedio "real".

(b) Podría elegirse ya sea la moda o la mediana, en ese orden, como descripción de lo que es "típico" en esta empresa. La moda es "típica" porque resulta evidente que muchos más trabajadores se encuentran en este nivel de salarios que en cualquier otro. La mediana es "típica" en el sentido de que la mitad de los trabajadores ganan menos de esa cantidad y la otra mitad más.

LA MEDIA PONDERADA

3.9 Suponga que tiene una lista de los precios de un articulo de consumo popular en 22 áreas metropolitanas que varían considerablemente en tamaño y ventas de ese articulo. Se calculan (a) la mediana y (b) la media aritmética para esos datos. Describa el significado que tendría cada uno de esos valores.

(a) La mediana señalarla el precio promedio, o típico, en el sentido de que la mitad de las áreas metropolitanas mostrarían precios del articulo por debajo de ese valor, y en la otra mitad los precios estarían por encima.

(b) El significado de la media aritmética como valor no ponderado es cuestionable. Resulta claro que *no* señalaría los precios promedio que pagan todos los compradores del articulo y que viven en las 22 áreas metropolitanas, porque las áreas metropolitanas pequeñas estarían representadas de igual manera que las áreas grandes en ese promedio. Con el objeto de determinar un promedio apropiado para todos los compradores de todas las zonas, se tendría que calcular un promedio ponderado (utilizando las ventas del artículo como pesos).

3.10 Con respecto a la Tabla 3.4, determine el porcentaje global de artículos defectuosos ensamblados durante la semana muestreada.

Tabla 3.4 Porcentaje de artículos defectuosos en un departamento de ensamble en una (muestra de una semana)

Turno	Porcentaje defectuoso	Número de artículos, en millares (p)	P^X
1	1.1	210	231.0
2	1.5	120	180.0
3	2.3	50	115.0
		$\Sigma p = 380$	$\Sigma(pX) = 526.0$

Utilizando la fórmula (3.3)

$$\bar{X}_w = \frac{\Sigma(pX)}{\Sigma p} \quad \frac{526.0}{380} = 1.4\% \text{ defectuosodefectuosos}$$

CUARTILES, DECILES Y PERCENTILES

3.11 Determine los valores de (a) el segundo cuartil, (b) el segundo decil, y (c) el punto percentil 40, para los importes presentados en el problema 3.1.

(a) $Q_2 = X_{[(2n/4)+(1/2)]} = X_{(7.5+0.5)} = X_8 = \$5300$

(Nota: por definición, el segundo cuartil es siempre el mismo punto que la mediana)

(b) $D_2 = X_{[(2n/10)+(1/2)]} = X_{(3.0+0.5)} = X_{3.5} = 2500$

(Nota: este es el valor que se encuentra a la mitad entre los montos de las ventas tercera y cuarta, en orden ascendente.)

(c) $P_{40} = X_{[(40n/100)+(1/2)]} = X_{(6.0+0.5)} = X_{6.5} = 3750 \approx \$0.38$

3.12 Para las medidas del problema 3.4, determine los valores de (a) el tercer cuartil, (b) el noveno decil, (c) el punto percentil 50 y (d) el punto percentil 84, para este grupo de salarios.

(a) $Q_3 = X_{[(3n/4)+(1/2)]} = X_{15.5} = \$295{,}000$
(Nota: este es el valor que se encuentra a la mitad entre los salarios que ocupan los lugares 15 y 16, en orden ascendente)
(b) $D_9 = X_{[(9n/10)+(1/2)]} = X_{18.5} = \$327{,}500$
(c) $P_{50} = X_{[(50n/100)+(1/2)]} = X_{10.5} = \$260{,}000$
(Nota: por definición, el punto percentil 50 es siempre igual a la mediana)
(d) $P_{84} = X_{[(84n/100)+(1/2)]} = X_{(16.8+0.5)} = X_{17.3} = \$311{,}000$
(Nota: este es el valor que se encuentra a *tres décimos* de distancia entre los salarios decimosexto y decimo -octavo, en orden ascendente)

LA MEDIA, LA MEDIANA Y LA MODA PARA DATOS AGRUPADOS

3.13 En la Tabla 3.5 se reproduce la distribución de frecuencias dada originalmente en el problema 2.1. Determine la renta mensual promedio en términos de *(a)* la media, *(b)* la mediana y (c) la moda. Suponga que éstos son *todos* los departamentos de una área geográfica dada.

Tabla 3.5 Distribución de frecuencias para las rentas mensuales de departamentos

(1) Renta mensual (en millares de pesos)	(2) Punto medio de clase *(X)*	(3) Número de departamentos *(f)*	(4) *fX*	(5) Frecuencia acumulada *(fa)*
\$350-379	\$364.50	3	\$1,093.50	3
380-409	394.50	8	3,156.00	11
410-439	424.50	10	4,245.00	21
440-469	454.50	13	5,908.50	34
470-499	484.50	33	15,988.50	67
500-529	514.50	40	20,580.00	107
530-559	544.50	35	19,057.50	142
560-589	574.50	30	17,235.00	172
590-619	604.50	16	9,672.00	188
620-649	634.50	12	7.614.00	200
		Total 200	$\Sigma(fX)$=\$104,550.00	

(a) $\mu = \dfrac{\Sigma(fX)}{N} = \dfrac{104{,}550}{200} = \$522.75$

(b) Med $= B_L + \left(\dfrac{\frac{N}{2} - cf_B}{f_c}\right) i = 499.50 + \left(\dfrac{100-67}{40}\right) 30 = \$524.25$

(Nota: \$499.50 es el límite exacto inferior de la clase que contiene a la medición N/2 es decir, la centésima medición.)

(c) Moda $= L_I + \left(\dfrac{d_1}{d_1 + d_2}\right) i = 499.50 + \left(\dfrac{7}{7+5}\right) 30 = \$517.00$

(Nota: \$499.50 es el limite exacto inferior de la clase que contiene la frecuencia más alta.)

3.14 Comente la forma de la distribución de las rentas de los departamentos del problema 3.13.

En la figura 2-11 se presenta el polígono de frecuencias para estos datos, y parece señalar cierto grado de asimetría negativa. La comparación de los valores de la media y la mediana apoya también esta conclusión, puesto que la media es menor que la mediana. Sin embargo, como la media es mayor que la moda, se concluye que la cantidad de asimetría es pequeña.

3.15 Al realizar auditorías anuales, en un despacho contable se lleva un registro del tiempo que se requiere para auditar 50 cuentas, tal como se señala en la Tabla 3.6. Calcule (a) la media, (b) le mediana y (c) la moda, para los tiempos de auditoría que se requieren en esta muestra de registros.

Tabla 3.6 Tiempo requerido para auditar saldos de cuentas.

(1) Tiempo de auditoría, al minuto más cercano	(2) Punto medio de clase (X)	(3) Número de registros (f)	(4) fX	(5) Frecuencia acumulada *(fa)*
10-19	14.5	3	43.5	3
20-29	24.5	5	122.5	8
30-39	34.5	10	345.0	18
40-49	44.5	12	534.0	30
50-59	54.5	20	1,090.0	50
		Total 50	$\Sigma(fX) = 2{,}135.0$	

(a) $\bar{X} = \frac{\Sigma(fX)}{n} = \frac{2{,}135.0}{50} = 42.7 \text{ min}$

(b) $+\left(\frac{\frac{n}{2} - cf_B}{f_c}\right)i = 39.5 + \left(\frac{25-18}{12}\right)10.00 = 45.3 \text{ min}$

(Nota: 39.5 es el límite exacto inferior de la clase que contiene a la medición n/2, o vigesimoquinta)

(c) Moda $= L_l + \left(\frac{d_1}{d_1 + d_2}\right)i = 49.5 + \left(\frac{8}{8+20}\right)10.0 = 52.36 \text{ min}$

(Nota: 49.5 es el límite exacto inferior de la clase que contiene la mayor frecuencia. Observe también que, en este caso, $d_2 = 20 - 0 = 20$)

3.16 Comente la forma de la distribución de los tiempos de auditoría reportados en el problema 3.15.

La media es menor que la mediana o la moda. Por ello, resulta evidente que la distribución tiene asimetría negativa; es decir, está sesgada a la izquierda.

3.17 En la Tabla 3.7 se reproducen los datos que se reportaron originalmente en el problema 2.16. Determine (a) la media, (6) la mediana y (c) la moda, para esta muestra de 50 empresas.

Tabla 3.7 Número promedio de lesiones por millar de horas-hombre en una industria específica

Número promedio de lesiones	Punto medio de clase (X)	Número de empresas (f)	fx	Frecuencia acumulada (fa)
1.5-1.7	1.6	3	4.8	3
1.8-2.0	1.9	12	22.8	15
2.1-2.3	2.2	14	30.8	29
2.4-2.6	2.5	9	22.5	38
2.7-2.9	2.8	7	19.6	45
3.0-3.2	3.1	5	15.5	50
		Total 50	$\Sigma(fX) = 116.0$	

(a) $\bar{X} = \dfrac{\Sigma(fX)}{n} = \dfrac{116.0}{50} = 2.32$ lesiones

(b) Med $= L_I + \left(\dfrac{\frac{n}{2} - fa_A}{f_c}\right) i = 2.05 + \left(\dfrac{25-15}{14}\right) 0.3 = 2.26$ lesiones

(Nota: 2.05 es el límite exacto inferior de la clase que contiene la medición n/2 o vigesimoquinta)

(c) Moda $= L_I + \left(\dfrac{d_1}{d_1 + d_2}\right) i = 2.05 + \left(\dfrac{2}{2+5}\right) 0.3 = 2.14$ lesiones

(Nota: 2.05 es el limite exacto inferior de la clase que contiene a la medición n/2, o vigesimoquinta.)

3.18 Comente la forma de la distribución para los datos del problema 3.17.

La media muestral es mayor que la mediana y la moda y, por lo tanto, la distribución tiene asimetría positiva: es decir, está sesgada a la derecha.

3.19 En la Tabla 3.8 se reproduce la distribución de frecuencias del problema 2.21. Determine (a) la media, (b) la mediana y (c) la moda, para estos datos. No a la base para determinar los límites de clase o limites exactos, y los puntos medios de clase para una distribución de frecuencias del tipo "y menor que", como en este caso.

Tabla 3.8 Tiempo requerido para procesar y preparar órdenes en un restaurante

Tiempo, minutos	Límites exactos de clase	Punto medio de clase (X)	Número de órdenes	fX	Frecuencia acumulada (fa)
5 y menor que 8	5.0-8.0	6.5	10	65.0	10
8 y menor que 11	8.0-11.0	9.5	17	161.5	27
11 y menor que 14	11.0-14 0	12.5	12	150.0	39
14 y menor que 17	14.0-17.0	15.5	6	93.0	45
17 y menor que 20	17.0-20.0	18.5	2	37.0	47
			Total 47	$\Sigma(fX) = 506.5$	

(a) $\bar{X} = \frac{\Sigma(fX)}{n} = \frac{506.5}{47} = 10.8$ min

(b) Med $= L_I + \left(\frac{\frac{n}{2} - fa_A}{f_c}\right)i = 8.0 + \left(\frac{23.5 - 10}{17}\right)3.0 = 10.4$ min

(Nota: 8.0 es el límite exacto inferior de la clase que contiene a la medición *n/2,* o 23.5ª.)

(c) Moda $= L_I + \left(\frac{d_1}{d_1 + d_2}\right)i = 8.0 + \left(\frac{7}{7+5}\right)3.0 = 9.75 \cong 9.8$ **min**

(Nota: 8.0 es el limite exacto inferior de la clase que contiene la mayor frecuencia.)

3.20 Comente la forma de la distribución para los tiempos de procesamiento analizados en el problema 3.19.

Como la media es la mayor de las medidas de tendencia central, la distribución tiene asimetría positiva.

CUARTILES, DECILES Y PERCENTILES PARA DATOS AGRUPADOS

3.21 Con referencia a la Tabla 3.5 (página 38). determine los valores de (a) el primer cuartil, (b) el tercer decil, (c) el punto percentil 80, y *(d)* el percentil 17, para el promedio de las rentas de los apartamentos.

(a) $Q_1 = L_I + \left(\frac{\frac{N}{4} - fa_A}{f_c}\right)i = 469.50 + \left(\frac{50 - 34}{33}\right)30 = \$484.05$

(Nota: \$469.50 es límite exacto inferior de la clase que contiene la medición n/4, o 50ª.)

(b) $= L_I + \left(\frac{\frac{3N}{10} - fa_A}{f_c}\right)i = 469.50 + \left(\frac{60 - 34}{33}\right)30 = \$493.14$

(Nota: \$469.50 es el límite exacto inferior de la clase que contiene la medición 3N/10, o 60ª.

(c) $P_{80} = L_I + \left(\frac{\frac{80N}{100} - fa_A}{f_c}\right)i = 559.50 + \left(\frac{160 - 142}{30}\right)30 = \$577.50$

(Nofa: \$559.50 es el limite exacto inferior de la clase que contiene la medición 80N/100 o 160ª.)

(d) $P_{17} = L_I + \left(\frac{\frac{17N}{100} - fa_A}{f_c}\right)i = 439.50 + \left(\frac{34 - 21}{13}\right)30 = \$469.50$

(Nota 1: \$439.50 es el límite exacto inferior de la clase que contiene la medición 17N/100, o 34ª.)
(Nota 2: el percentil 17 se encuentra en el límite exacto superior de la clase. Esto es lógico, porque se requiere atravesar toda la clase para llegar al 34° elemento de la distribución.)

3.22 Con referencia a la Tabla 3.6 (página 38), determine los valores de (a) el tercer cuartil, (b) el primer decil y (c) el punto percentil 90, para los tiempos requeridos para auditar los saldos de cuentas.

(a) $Q_3 = L_I + \left(\frac{\frac{3n}{4} - fa_A}{f_c}\right)i = 49.5 + \left(\frac{37.5 - 30}{20}\right)10 = 53.25 \cong 53.2$ **min**

(Wofa: 49.5 es el límite exacto inferior de la clase que contiene la medición 3n/4, o 37.5ª.)

(b) $D_1 = L_I + \left(\frac{\frac{n}{10} - fa_A}{f_c}\right)i = 19.5 + \left(\frac{5 - 3}{5}\right)10 = 23.5$ **min**

(Nota: 19.5 es el límite exacto inferior de la clase que contiene la medición n/10, o 5ª.)

(c) $P_{90} = L_I + \left(\frac{\frac{90n}{100} - fa_A}{f_c}\right)i = 49.5 + \left(\frac{45 - 30}{20}\right)10 = 57.0$ **min**

(Wofa: 49.5 es el límite exacto inferior de la clase que contiene la medición 90n/100, o 45ª.)

3.23 Para la distribución de frecuencias del tipo "y menor que" reportada en la Tabla 3.8, determine los valores de (a) el segundo cuartil, *(b)* el noveno decil y (c) el punto percentil 75, para los tiempos que se requieren para procesar y preparar las órdenes en el restaurante.

(a) $Q_2 = L_l + \left(\frac{\frac{2n}{4} - fa_A}{f_c}\right) i = 8.0 + \left(\frac{23.5 - 10}{17}\right) 3.0 = 10.4 \text{ min}$

(b) $D_9 = L_l + \left(\frac{\frac{9n}{10} - fa_A}{f_c}\right) i = 14.0 + \left(\frac{42.3 - 39}{6}\right) 3.0 \cong 15.6 \text{ min}$

(c) $P_{75} = L_l + \left(\frac{\frac{75n}{100} - fa_A}{f_c}\right) i = 11.0 + \left(\frac{35.25 - 27}{12}\right) 3.0 = 13.1 \text{ min}$

RESULTADOS POR COMPUTADORA

3.24 Utilice una computadora para calcular la media y la mediana de los datos de la Tabla 2.6 (página 16), que contiene los tiempos que una muestra de 30 empleados requieren para llevar a cabo un trabajo de ensamble. En los problemas 2.27 y 2.28 se agruparon esos datos en una distribución de frecuencias utilizando una computadora, y se produjo el histograma correspondiente.

En la figura 3-2 se presentan los datos introducidos a la computadora y los resultados. Como puede observarse, el tiempo promedio de ensamble es 13.3 minutos y el tiempo mediano de ensamble es 13.5 minutos.

```
MTB >  SET ASSEMBLY TIMES INTO Cl
DATA>    10    14    15    13    17    16    12    14    11    13    15    18     9
DATA>    14    14     9    15    11    13    11    12    10    17    16    12    11
DATA>    16    12    14    15
DATA>  END
MTB > NAME FOR Cl IS 'TIME'
MTB > MEAN OF 'TIME'
   MEAN    =      13.300
MTB > MEDIAN OF 'TIME'
   MEDIAN -      13.500
```

Fig. 3-2

PROBLEMAS COMPLEMENTARIOS

LA MEDIA, LA MEDIANA Y LA MODA

3.25 El número de automóviles que vendió cada uno de los 10 vendedores de una distribuidora en un mes especifico, en orden ascendente, es: 2,4, 7,10,10,10,12,12,14 y 15. Determine, para la población (a) la media, (b) la mediana y (c) la moda, para el número de automóviles vendidos.

Resp. (a) 9.6, (b) 10.0, (c) 10.0

3.26 ¿Cuál de los valores del problema 3.25 describe de mejor manera el volumen de ventas "típico" por vendedor?

Resp. 10.0

3.27 Se determinó que los pesos de una muestra de cartas procesadas en una oficina postal, pesadas hasta el gramo más próximo, son: 21,18,30,12,14,17,28,10,16 y 25. Determine (a) la media, (b) la mediana y (c) la moda para estos pesos.

Resp. (a) 19.1, (b) 17.5, (c) no existe moda

3.28 ¿Cómo puede obtenerse la moda para los pesos de las cartas que se describieron en el problema 3.27?

Resp. Puede obtenerse la moda construyendo una distribución de frecuencias para los datos. En ese caso, se requerirla un tamaño de muestra mayor.

3.29 Se presentan los siguientes resultados de un examen practicado a 20 estudiantes inscritos en un curso de análisis de decisiones, en orden ascendente: 39, 46, 57, 65, 70, 72, 72, 75, 77, 79, 81,81, 84, 84, 84, 87, 93, 94, 97 y 97. Determine (a) la media, *(b)* la mediana y (c) la moda para esas calificaciones.

Resp. (a) 76.7, *(b)* 80.0, (c) 84.0

3.30 Describa la distribución de las calificaciones del problema 3.29 en términos de asimetría.

Resp. Tiene asimetría negativa.

3.31 El número de accidentes ocurridos durante determinado mes en 13 departamentos de manufactura de una planta industrial fueron: 2,0,0,3,3,12,1,0, 8,1,0, 5,1. Calcule (a) la media, (b) la mediana y (c) la moda, para el número de accidentes por departamento.

Resp. (a)2.8,(5)1.0, (c) 0

3.32 Describa la distribución de los accidentes reportados en el problema 3.31, en términos de asimetría.

Resp. Tiene asimetría positiva.

LA MEDIA PONDERADA

3.33 Suponga que los precios al menudeo de determinados artículos han sufrido los cambios que se muestran en la Tabla 3.9. Determine el cambio porcentual promedio de los precios al menudeo *sin* referencia al promedio de gastos que se incluye en la tabla.

Resp. 4.0%

Tabla 3.9 Cambios en los precios al menudeo de algunos artículos durante un año específico

Artículo	Aumento porcentual	Gasto mensual promedio (antes del aumento)
Leche	10%	$2000.00
Carne molida	- 6	3000.00
Ropa	- 8	3000.00
Gasolina	20	5000.00

3.34 Con referencia a la Tabla 3.9, determine el cambio porcentual promedio, ponderando el aumento porcentual de cada articulo con la cantidad promedio mensual que se gastaba en ese articulo antes del aumento.

Resp. 6.0%

3.35 ¿Cuál de los cambios porcentuales promedio de los precios que se calcularon en los problemas 3.33 y 3.34 sería más apropiado como medida del impacto de los cambios de los precios sobre este consumidor específico? ¿Por qué?

Resp. La media ponderada del problema 3.34 es más adecuada (en el ejemplo 2 se tiene la explicación).

CUARTILES, DECILES Y PERCENTILES

3.36 Determine los valores de (a) el primer cuartil, (b) el segundo decil y (c) el punto percentil 30 para los importes de las ventas del problema 3.25.

Resp. (a) 7.0, (b) 5.5, (c) 8.5

3.37 Con base en el problema 3.27, determine los pesos correspondientes a: (a) el tercer cuartil, (b) el tercer decil y (c) el punto percentil 70.

Resp. (a) 25.0 gr. (b) 15.0 gr. (c) 23.0 gr.

3.38 Determine (a) el segundo cuartil, (b) el noveno decil y (c) el punto percentil 50, para las calificaciones del examen del problema 3.29

Resp. (a) 80.0, (b) 95.5, (c) 80.0

3.39 En general, ¿qué cuartil, decil y punto percentil, respectivamente, son equivalentes a la mediana?

Resp. El segundo cuartil, el quinto decil y el punto percentil 50.

LA MEDIA, LA MEDIANA Y LA MODA PARA DATOS AGRUPADOS

3.40 Con referencia a la Tabla 2.15 (página 25), determine el rendimiento típico de gasolina que se obtiene en términos de (a) la media, (b) la mediana y (c) la moda.

Resp. (a) 28.95, (b) 28.85, (c) 28.86

3.41 Describa la forma de la distribución de los rendimientos de gasolina del problema 3.40, en términos de asimetría.

Resp. La distribución es relativamente simétrica.

3.42 La distribución de frecuencias que aparece en la Tabla 3.10 se basa en los datos considerados originalmente en el problema 2.35. Calcule (a) la media, (b) la mediana y (c) la moda, para los montos de los préstamos, con base únicamente en la distribución de frecuencias de la Tabla 3.10.

Resp. (a) $1,109.50, (b) $954.05, *(c)* $646.17

Tabla 3.10 Montos de 40 préstamos personales (en miles de pesos)

Monto de los préstamos (en miles de pesos)	Número de préstamos
$ 300-699	13
700-1,099	11
1100-1,499	6
1500-1,899	5
1900-2,299	3
2300-2,699	1
2700-3,099	1
Total	40

3.43 Describa la forma de la distribución de frecuencias para los montos de préstamos personales del problema 3.42.

Resp. Sesgada positivamente.

3.44 A partir de la Tabla 2.17 (página 27), determine el tiempo útil promedio de las herramientas de corte calculando (a) la media, (b) la mediana y (c) la moda.

Resp. (a) 87.45, (b) 89.12, (c) 89.95

3.45 Describa la distribución de frecuencias del tiempo útil de las herramientas del problema 3.44, en términos de asimetría.

Resp. Tiene una ligera asimetría negativa.

3.46 Con referencia a la distribución de frecuencias de "y menor que" de la Tabla 2.18 (página 27), calcule (a) la media, (b) la mediana y (c) la moda para las edades de los solicitantes.

Resp. (a) 23.3, (b) 22.4, (c) 21.2

3.47 Describa la forma de la distribución de la Tabla 2.18, con base en las respuestas dadas al problema 3.46.

Resp. La distribución tiene asimetría positiva.

CUARTILES, DECILES Y PERCENTILES PARA DATOS AGRUPADOS

3.48 Determine los valores de (a) el primer cuartil, (6) el primer decil y (c) el décimo punto percentil, para los datos de rendimiento de gasolina de los automóviles de la Tabla 2.15 (página 25).

Resp. (a) 13.25, (b) 11.44, (c) 218.45

3.49 Para los montos de los préstamos personales reportados en el problema 3.42, determine los valores de (a) el segundo cuartil, (b) el segundo decil y (c) el punto percentil 90.

Resp. (a) $954.05, (b) $545.64, (c) $2,032.83

3.50 Determine los valores de (a) el tercer cuartil, (b) el séptimo decil y (c) el percentil 75, para las horas de tiempo útil de las herramientas de corte de la Tabla 2.17 (página 27).

Resp. (a) 106.20, *(b)* 101.34. (c) 106.20

3.51 Determine los valores de (a) el primer cuartil, (b) el séptimo decil y *(c)* el punto percentil 80, para los datos de las edades de la Tabla 2.18 (página 18).

Resp. (a) 20.8, (b) 24.7, (c) 26.4

RESULTADOS POR COMPUTADORA

3.52 Utilice una computadora para calcular la media y la mediana de los datos de la Tabla 2.16 (página 26), que contiene los montos de 40 préstamos personales. En los problemas 2.53 y 2.54 se agruparon esos datos en una distribución de frecuencias utilizando una computadora, y se produjo el correspondiente histograma. También, en el problema 3.42 se calcularon aproximaciones de estas media y mediana con base en datos agrupados.

Resp. El monto promedio de los préstamos es $1 097 400, en tanto que la mediana es $944,500. Esto se compara con las aproximaciones del problema 3.42, basado en los datos agrupados de $1,109,500 y $954,050, respectivamente.

Descripción de datos de negocios: Medidas de variabilidad

4

4.1 MEDIDAS DE VARIABILIDAD EN CONJUNTOS DE DATOS

Las medidas de tendencia central que se describieron en el capítulo 3 son útiles para identificar el valor "típico" en un conjunto de datos. En contraste, *las medidas de variabilidad* se ocupan de describir la variabilidad entre los valores. Están disponibles diversas técnicas para medir la magnitud de la variabilidad en conjuntos de datos. Las que se describen en este capítulo son: *rango, rangos modificados, desviación media, varianza, desviación estándar* y *coeficiente de variación.*

EJEMPLO 1. Suponga que dos máquinas empacadoras distintas dan como resultado paquetes con un peso promedio de 10 gramos de café, pero que en un caso todos los paquetes se encuentran dentro de un rango de 0.10 gramos de este peso, en tanto que en el otro caso los pesos pueden variar hasta en un gramo en cualquier dirección. Medir la variabilidad, o dispersión, de las cantidades* que se empacan sería tan importante en este caso como medir el promedio.

El concepto de asimetría se describió ya en las secciones 2.4 y 3.6. El *coeficiente de asimetría* de Pearson se describe en la sección 4.9.

4.2 EL RANGO

El *rango,* o *R,* es la diferencia entre los valores mayor y menor del conjunto de datos. Así, cuando *My* representa el mayor valor del grupo y *Mn* representa el menor, el rango de los datos no agrupados es:

$$R = My - Mn \tag{4.1}$$

EJEMPLO 2. Durante un mes determinado del verano, los ocho vendedores de aparatos electrónicos de una empresa vendieron el siguiente número de ventiladores: 8, 11, 5, 14, 8, 11, 16, 11. El rango del número de unidades vendidas es

$$R = My - Mn = 16 - 5 = 11.0 \text{ unidades}$$

Nota: Con el propósito de llevar a cabo comparaciones, por lo general se reportan las medidas de variabilidad con un digito adicional al nivel original de medición.

4.3 RANGOS MODIFICADOS

Un rango *modificado* es un rango para el cual se elimina cierto porcentaje de los valores en cada uno de los extremos de la distribución. Algunos rangos modificados típicos son: el 50% *central,* el *80% central* y el *90% central.*

El procedimiento mediante el cual se determina un rango modificado consiste, primero, en ubicar los dos puntos percentiles adecuados (véanse las secciones 3.7 y 3.11) para, después, calcular la diferencia entre los valores que se encuentran en esos puntos. Por ejemplo, para el rango del 80% central, los puntos percentiles apropiados son el décimo percentil y el nonagésimo percentil, porque el 80% central de esos valores se ubica entre esos dos puntos.

EJEMPLO 3. Los datos de las ventas de aparatos eléctricos que se presentaron en el ejemplo 2, son, en orden ascendente: 5, 8, 8.11.11,11,14,16. Para calcular el rango del 50% central, en primer lugar se determinan los valores de los puntos percentiles adecuados y después se resta el valor menor del mayor.

$$P_{75} = X_{[(75\,n/100)+(1/2)]} = X_{[6+(1/2)]} = X_{6.5} = 12.5$$

$$P_{25} = X_{[(25n/100)+(1/2)]} = X_{[2+(1/2)]} = X_{2.5} = 8.0$$

$$R\ 50\%\ \text{central} = P_{75} - P_{25} = 12.5 - 8.0 = 4.5\ \text{unidades}$$

4.4 LA DESVIACIÓN MEDIA

La *desviación media* o *DM* se basa en la diferencia entre el valor absoluto de cada uno de los elementos conjuntos de datos y la media del grupo. Después, se calcula la media de ésas desviaciones (algunos especialistas en estadística utilizan la diferencia entre cada valor y la mediana). Si se calculara la media por la suma de las diferencias positivas y negativas entre cada valor y la media aritmética, la respuesta sería, siempre igual a 0. Por esta razón, son los *valores absolutos* de las diferencias lo que se suma.

$$\text{DM poblacional} = \frac{\Sigma|X - \mu|}{N} \tag{4.2}$$

$$\text{DM muestral} = \frac{\Sigma|X - \bar{X}|}{n} \tag{4.3}$$

EJEMPLO 4. Para los datos de ventas de aparatos eléctricos que se dieron en el ejemplo 2, la media aritmética es 10.5 (véase la sección 3.2). Utilizando los cálculos de la Tabla 4.4, se determina la desviación media de la siguiente manera:

$$DM = \frac{\Sigma|X - \mu|}{N} = \frac{21.0}{8} = 2.625 \cong 2.6\ \text{unidades}$$

Tabla 4.1 Hoja de trabajo para calcular la desviación media de los datos de ventas.

X	$X - \mu$	$\lvert X - \mu \rvert$
5	-5.5	5.5
8	-2.5	2.5
8	-2.5	2.5
11	0.5	0.5
11	0.5	0.5
11	0.5	0.5
14	3.5	3.5
16	5.5	5.5
		Total 21.0

Asi, puede decirse que, en promedio, las ventas de aparatos eléctricos por vendedor difieren en 2.6 unidades de la media del grupo, en cualquier dirección.

4.5 LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR.

La *varianza* es similar a la desviación media porque se basa en la diferencia entre cada uno de los valores del conjunto de datos y la media del grupo. La diferencia consiste en que, antes de sumarlas, *se eleva al cuadrado* cada una de las diferencias. Para una población, se representa la varianza mediante *v(X)* o, en forma más típica, mediante la letra σ^2 (que se lee "sigma cuadrada"); la fórmula es:

$$\upsilon(X) = \sigma^2 = \frac{\Sigma(X) - \mu)^2}{N} \qquad (4.4)$$

A diferencia de otras estadísticas muestrales que se han analizado, la varianza de una muestra no es, en términos de cálculo, completamente equivalente a la varianza de la población. Para este caso, el denominador de la fórmula de la varianza muestral es ligeramente diferente. En esencia, lo que se incluye en la fórmula es un factor de corrección, para que la varianza muestral sea un estimador insesgado para la varianza de la población (véase la sección 8.1). La varianza muestral se representa mediante s^2; su fórmula es:

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1} \qquad (4.5)$$

Por lo general, resulta difícil Interpretar el significado del valor de una varianza porque las unidades en las que se expresa son valores al cuadrado. En parte por esta razón, se utiliza con mayor frecuencia la raíz cuadrada de la varianza, representada mediante la letra griega o (o s para una muestra) y se le denomina *desviación estándar.* Las fórmulas son:

Desviación estándar poblacional:

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}} \qquad (4.6)$$

Desviación estándar muestral:

$$s = \sqrt{\frac{\Sigma(X - \bar{X})^2}{n-1}} \qquad (4.7)$$

La desviación estándar es especialmente útil cuando se le utiliza junto con la denominada distribución normal (véase la sección 4.7).

EJEMPLO 5. Para los datos de ventas de aparatos eléctricos dados en el ejemplo 2, la media aritmética es 10.5 unidades (véase la sección 3.2). Considerando que estos datos mensuales de ventas son la población estadística de interés, se determina la desviación estándar con base en los cálculos de la Tabla 4.2, de la manera siguiente:

$$\sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}} = \sqrt{\frac{86}{8}} = \sqrt{10.75} = 3.3$$

Tabla 4.2 Hoja de trabajo para calcular la desviación estándar poblacional para los datos de ventas

X	$X - \mu$	$(X - \mu)^2$
5	-5.5	30.25
8	-2.5	6.25
8	-2.5	6.25
11	0.5	0.25
11	0.5	0.25
11	0.5	0.25
14	3.5	12.25
16	5.5	30.25
		Total 86.00

4.6 CÁLCULOS ABREVIADOS DE LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR

A las fórmulas de la sección 4.5 se les denomina *fórmulas de desviaciones* porque, en cada caso, se deben determinar las desviaciones de los valores individuales con respecto a la media grupal. Existen formas alternativas que son matemáticamente equivalentes, pero que no requieren del cálculo de cada una de las desviaciones. Debido a que, por lo general, es más fácil utilizar estas fórmulas para realizar cálculos, se les denomina *fórmulas abreviadas.*

Las fórmulas abreviadas son:

Varianza poblacional:
$$\sigma^2 = \frac{\Sigma X^2 - N\mu^2}{N} \tag{4.8}$$

Desviación estándar poblacional:
$$\sigma = \sqrt{\frac{\Sigma X^2 - N\mu^2}{N}} \tag{4.9}$$

Varianza muestral:
$$s^2 = \frac{\Sigma X^2 - n\bar{X}^2}{n-1} \tag{4.10}$$

Desviación estándar muestral:
$$s = \sqrt{\frac{\Sigma X^2 - n\bar{X}^2}{n-1}} \tag{4.11}$$

$$\sigma = \sqrt{\frac{\Sigma X^2 - N\mu^2}{N}} = \sqrt{\frac{968 - 8(10.5)^2}{8}} = \sqrt{10.75} = 3.3 \text{ unidades}$$

Tabla 4.3 Hoja de trabajo para calcular la desviación estándar poblacional para los datos de ventas

X	X^2
5	25
8	64
8	64
11	121
11	121
11	121
14	196
16	256
	Total 968

4.7 USO DE LA DESVIACIÓN ESTÁNDAR

La desviación estándar es la medida de dispersión más importante, ya que se le utiliza junto con varios de los métodos de inferencia estadística que se analizan en capítulos posteriores. Una descripción de estos usos escapa al alcance del siguiente capitulo.Sin embargo, como ejemplo de su uso, considérese una distribución de frecuencias que es al mismo tiempo simétrica y mesokúrtica. En análisis estadístico a una curva de frecuencias con estas características se le denomina *curva normal.* Para una distribución que tiene *distribución normal,* se sabe que aproximadamente el 68% de las mediciones se encuentran a no más de una desviación estándar de la media y que aproximadamente 95% de las mediciones se encuentran a no más de dos desviaciones estándar de la media. Esas observaciones se presentan en los diagramas de las figuras 4-1 (a) y (b), respectivamente.

EJEMPLO 7. Se observa que las cuentas por energía eléctrica en un área residencial, para un mes determinado, tienen una distribución normal. Si se determina que la media de los consumos es de $ 8 400, con una desviación estándar de $2 400, entonces puede concluirse que el 68% de las cuentas por consumo de energía eléctrica se encuentran a no más de $2 400 de la media, o, lo que es lo mismo, entre $6 000 y $10 800. También puede concluirse que aproximadamente 95% de las cuentas se encuentran a no más de $4 800 de la media o entre $3 600 y $13 200.

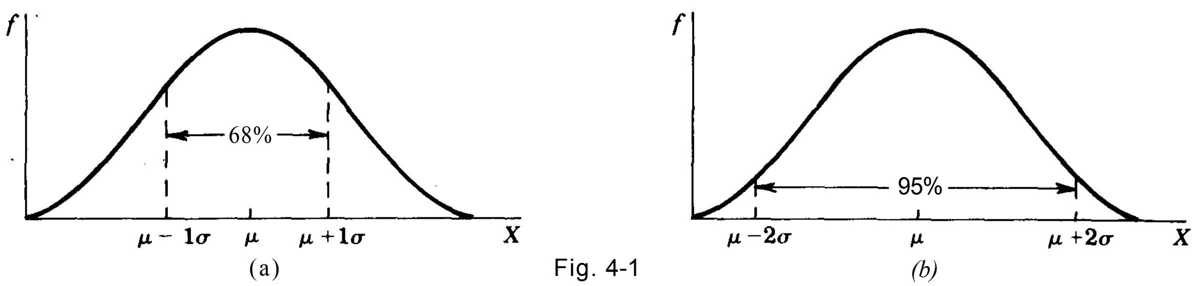


Fig. 4-1

4.8 EL COEFICIENTE DE VARIACIÓN

El *coeficiente de variación, CV,* indica la magnitud relativa de la desviación estándar con respecto a la media de la distribución, así, las fórmulas son:

Así,

Población:
$$CV = \frac{\sigma}{\mu} \tag{4.12}$$

Muestra:
$$CV = \frac{s}{\bar{X}} \tag{4.13}$$

El coeficiente de variación es útil cuando se desea comparar la variabilidad de 2 conjuntos de datos con respecto al nivel general de los valores de cada conjunto (y, por ello, respecto a la media).

EJEMPLO 8. Para 2 acciones comunes de empresas de la industria electrónica, el precio promedio de cierre en el mercado de valores durante un mes fue, para la acción A, de \$1 500, con desviación estándar de \$500. Para la acción *B*, el precio promedio fue de \$5 000, con desviación estándar de \$300. Haciendo una comparación absoluta, resultó ser superior la variabilidad en el precio de la acción *A* debido a que muestra una mayor desviación estándar. Pero, con respecto al nivel de precios, deben compararse los respectivos coeficientes de variación:

$$CV(A) = \frac{\sigma}{\mu} = \frac{500}{1\,500} = 0.033 \quad \text{y} \quad CV(B) = \frac{300}{5\,000} = 0.060$$

Por ello, puede concluirse que el precio de la acción *B* ha sido casi 2 veces más variable de la acción *A* (con respecto al precio promedio para cada una de las dos.)

4.9 COEFICIENTE DE ASIMETRÍA DE PEARSON

El *coeficiente de asimetría* de Pearson mide la desviación de la simetría, expresando la diferencia entre la media y la mediana con respecto a la desviación estándar del grupo de mediciones. Las fórmulas son:

$$\text{Asimetría poblácional} = \frac{3(\mu - \text{Med})}{\sigma} \tag{4.14}$$

$$\text{Asimetría de la muestra} = \frac{3(\bar{X} - \text{Med})}{s} \tag{4.15}$$

Para una distribución simétrica, el valor del coeficiente de asimetría es siempre 0, porque la media y la mediana son iguales. Para una distribución con asimetría positiva, la media es siempre mayor que la mediana y, por ello, el valor del coeficiente es positivo. Para una distribución con asimetría negativa, la media es siempre menor que la mediana y, por ello, el valor del coeficiente es negativo.

EJEMPLO 9. Para los datos de ventas de aparatos eléctricos del ejemplo 2, la media es 10.5 unidades, la mediana es 11.0 unidades (de las secciones 3.2 y 3.4) y la desviación estándar es 3.3 unidades. El coeficiente de asimetría es;

$$\text{Asimetría} = \frac{3(\mu - \text{Med})}{\sigma} = \frac{3(10.5 - 11.0)}{3.3} = -0.45$$

Así, la distribución de las unidades vendidas tiene una ligera asimetría negativa, es decir, "está sesgada hacia la izquierda".

4.10 EL RANGO Y LOS RANGOS MODIFICADOS PARA DATOS AGRUPADOS

Para datos agrupados en una distribución de frecuencias, por lo general se define el rango como la diferencia entre el limite exacto superior de la clase más alta, $L_S(A)$, y el limite inferior de la clase más baja, L,(B). Asi, el rango para los datos agrupados es

$$R = L_S(A) - L_I(B) \tag{4.16}$$

EJEMPLO 10. Los datos agrupados de la Tabla 4.4 provienen de la distribución de frecuencias de los salarios diarios de 100 trabajadores no calificados que se presentaron en las secciones 2.1 y 3.8. El rango es

$$R = L_S(A) - L_I(B)$$

Tabla 4.4 Salarios diarios para 100 trabajadores no calificados

Salario diario	Límites exactos de clase	Número de trabajadores (f)	Frecuencia acumulada *(fa)*
$ 2400-2599	$ 2399.50-2599.50	7	7
2600-2799	2599.50-2799.50	20	27
2800-2999	2799.50-2999.50	33	60
3000-3199	2999.50-3199.50	25	85
3200-3399	3199.50-3399.50	11	96
3400-3599	3399.50-3599.50	4	100
		Total 100	

EJEMPLO 11. El rango central del 90% para la distribución de frecuencias de los salarios dados en la Tabla 4.4 es:

$$R\,90\%\text{ central} = P_{95} - P_{05} = 3381.32 - 2542.36 = \$838.96$$

donde

$$P_{95} = L_I + \left(\frac{\frac{95\,n}{100} - fa_A}{f_c}\right) i = 3199.50 + \left(\frac{95 - 85}{11}\right) 200 = \$3381.32$$

$$P_{05} = L_I + \left(\frac{\frac{5\,n}{100} - fa_A}{f_c}\right) i = 2399.50 + \left(\frac{5 - 0}{7}\right) 200 = \$2542.36$$

4.11 LA DESVIACIÓN MEDIA PARA DATOS AGRUPADOS

Para los datos agrupados en una distribución de frecuencias, se asume que el punto medio de cada clase representa a todas las mediciones incluidas en esa clase. Éste es el mismo enfoque que se utilizó al determinar la media aritmética para datos agrupados, según se describió en la sección 3.8.

Así,

$$\text{DM poblacional} = \frac{\Sigma(f|X-\mu|)}{N} \tag{4.17}$$

$$\text{DM muestral} = \frac{\Sigma(f|X-\bar{X}|)}{n} \tag{4.18}$$

EJEMPLO 12. Para los datos de salarios diarios que se presentan en la Tabla 4.4, la media aritmética es $2 949.50 (véase la sección 3.8).

La desviación media se determina de la siguiente manera, a partir de los cálculos de la Tabla 4.5

$$DM = \frac{\Sigma(f|X-\bar{X}|)}{n} = \frac{19\,600}{100} = \$196$$

Tabla 4.5 Hoja de trabajo para calcular la desviación media para datos agrupados

Salario diario	Punto medio de clase (X)	Número de trabajadores (f)	$\lvert X-\bar{X}\rvert$	$f\lvert X-\bar{X}\rvert$
$ 2400-2599	$ 2499.50	7	$450	$3150
2600-2799	2699.50	20	250	5000
2800-2999	2899.50	33	50	1650
3000-3199	3099.50	25	150	3750
3200-3399	3299.50	11	350	3850
3400-3599	3499.50	4	550	2200
		Total 100		Total $ 19 600

4.12 LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR PARA DATOS AGRUPADOS

Para datos agrupados en una distribución de frecuencias, se asume que el punto medio de cada clase representa a todas las mediciones incluidas en esa dase. Éste es el mismo enfoque que se utilizó al calcular la desviación media en la sección 4.11. Por ello, las fórmulas para datos de muestras y de poblaciones agrupados, son:

Varianza poblacional:

$$\sigma^2 = \frac{\Sigma[f(X-\mu)^2]}{N} \tag{4.19}$$

Varianza muestral:

$$s^2 = \frac{\Sigma[f(X-\bar{X})^2]}{n-1} \tag{4.20}$$

Las fórmulas para la desviación estándar de datos muestrales y poblacionales agrupados son:

Desviación estándar poblacional:

$$\sigma = \sqrt{\frac{\Sigma[f(X-\mu)^2]}{N}} \tag{4.21}$$

Desviación estándar muestral:

$$s = \sqrt{\frac{\Sigma[f(X-\bar{X})^2]}{n-1}} \tag{4.22}$$

EJEMPLO 13. Para los datos de salarios diarios que se presentan en el ejemplo 10, la media muestral es $2 949.50, tal como se determinó en la sección 3.8. De la Tabla 4.6, la desviación estándar muestral para estos datos agrupados se determina de la siguiente manera:

Tabla 4.6 Hoja de trabajo para calcular la desviación estándar muestral para datos agrupados

Salario diario	Punto medio de clase *(X)*	Número de trabajadores (f)	$X-\overline{X}$	$(X-\overline{X})^2$	$f(X-\overline{X})^2$
$2400-2599	$ 2499.50	7	$-450	$ 202 500	$ 1 417 500
2600-2799	2699.50	20	-250	62 500	1 250 000
2800-2999	2899.50	33	-50	2 500	82 500
3000-3199	3099.50	25	150	22 500	562 500
3200-3399	3299.50	11	350	122 500	1 347 500
3400-3599	3499.50	4	550	302 500	1210000
		Total 100			Total 5 870 000

Entonces, $$s = \sqrt{\frac{\Sigma[f(X-\overline{X})^2]}{n-1}} = \sqrt{\frac{5\,870\,000}{99}} = \sqrt{59\,292.93} = \$\,243.50$$

Para datos agrupados, las fórmulas de cálculo abreviadas son:

Varianza poblacional $$\sigma^2 = \frac{\Sigma(fX^2) - N\mu^2}{N} \qquad (4.23)$$

Desviación estándar poblacional $$\sigma = \sqrt{\frac{\Sigma(fX^2) - N\mu^2}{N}} \qquad (4.24)$$

Varianza muestral $$s^2 = \frac{\Sigma(fX^2) - n\bar{X}^2}{n-1} \qquad (4.25)$$

Desviación estándar muestral $$s = \sqrt{\frac{\Sigma(fX^2) - n\bar{X}^2}{n-1}} \qquad (4.26)$$

EJEMPLO 14. Para los datos salariales diarios que se presentaron en el ejemplo, se calcula enseguida la desviación estándar muestral, utilizando la fórmula de cálculo abreviada de la Tabla 4.7 para probar que la respuesta es la misma que se obtuvo con la fórmula de desviación en el ejemplo 13. La media muestral para estos datos es $2949.50.

$$s = \sqrt{\frac{\Sigma(fX^2) - n\bar{X}^2}{n-1}} = \sqrt{\frac{875825\,025 - 100(2949.50)^2}{100-1}} = \sqrt{59\,292.93} = \$\,243.50$$

Tabla 4.7 Hoja de trabajo para calcular la desviación estándar muestral para datos agrupados

Salario diario	Punto medio de clase (X)	Número de trabajadores *(f)*	X^2	fX^2
$ 2400-2599	$ 2499.50	7	$ 6 247 500.25	$ 43 732 501
2600-2799	2699.50	20	7 287 300.25	145 746 005.00
2800-2999	2899.50	33	8 407 100.25	277 434 308.25
3000-3199	3099.50	25	9 606 900.25	240 172 506.25
3200-3399	3299.50	11	10 886 700.25	119 753 702.75
3400-3599	3499.50	4	12 246 500.25	48 986 001.00
		Total 100		Total 875 825 025.00

4.13 RESULTADOS POR COMPUTADORA

Como se explicó en la sección 3.12, aunque las fórmulas de datos agrupados son útiles si los datos obtenidos de fuentes secundarias están disponibles sólo en forma de distribuciones de frecuencias, la agrupación de datos para simplificar cálculos para estos conjuntos grandes de datos ya no es necesaria debido a la disponibilidad de computadoras.

Existen programas de cómputo que permiten calcular diversas medidas de variabilidad. En el problema 4.25 se ilustra el cálculo del rango y la desviación estándar utilizando esta clase de programas.

Problemas resueltos

LOS RANGOS, LA DESVIACIÓN MEDIA Y LA DESVIACIÓN ESTÁNDAR

4.1 Para una muestra de 15 clientes en un restaurante se determinaron las siguientes cuentas, en orden de magnitud: $1,000, 1,000, 2,500, 2,500, 2,500, 3,500, 4,000, 5,300, 9,000, 12,500, 13,500, 24,500, 27,100, 30,900 y 41,000. Determínese, (a) el rango y *(b)* el rango del 50% central para estos datos muestrales.

(a) $R = My - Mn = \$\,41\,000 - 1000 - \$\,40\,000$

(b) R central 50% $- P_{75} - P_{25} = 21\,750 - 2500 = \$\,19\,250$
en donde $P_{75} = X_{[(75n/100)+(1/2)]} = X_{(11.25+0.50)} = X_{11.75} = 13\,500 + 8\,250 = \$\,21\,750$

(Nota: éste es el valor interpolado entre *tres cuartos* de la distancia entre los montos de cuenta que ocupan los lugares 11 y 12.)

$$P_{25} = X_{[(25n/100)+(1/2)]} = X_{(3.75+0.50)} = X_{4.25} = \$\,2500$$

4.2 Calcule la desviación media para los datos del problema 4.1. La media muestral para este grupo de valores fue $12 053.53, según se calculó en el problema 3.1.

Utilizando la tabla 4.8, se calcula la desviación estándar:

$$DM = \frac{\Sigma|X - \bar{X}|}{n} = \frac{\$154\,359.99}{15} = \$10\,290.67$$

Tabla 4.8 Hoja de trabajo para calcular la desviación media para los datos del restaurante

X	$X-\bar{X}$	$\lvert X-\bar{X}\rvert$
$1000	$-11053.3	$ 11053.3
1000	-11 053.3	11053.3
2500	-9553.33	9553.:'3
2500	-9553.33	9553 33
2500	-9533.33	9553.33
3500	-8553.33	8552.33
4000	-8053.33	8053.33
5300	-6753.33	6753.33
9000	-3053.33	3053.33
12 500	446.67	446.67
13 500	1446.67	1446.67
24 500	12 446.67	12 446.67
27 100	15 046.67	15 046.67
30 900	18 846.67	18 846.67
41000	28 946.67	28 946.67
		Total $111 115.45

Tabla 4.9 Hoja de trabajo para calcular la desviación estándar muestral para los datos del restaurante

X	$X-\bar{X}$	$(X-\bar{X})^2$	X^2
$1000	$-11053.3	$ 122 175 440.00	1000 000
1000	-11053.3	122 175 440.00	1000 000
2500	-9553.33	91266114.09	6 250 000
2500	-9553.33	91266 114.09	6 250 000
2500	-9533.33	91266 114.09	6 250 000
3500	-8553.33	73 159 454.09	12 250 000
4000	-8053.33	64 856 124.00	16000 000
5300	-6753.33	45 607 466.09	28 090 000
9000	-3053.33	9 322 824.09	81000 000
12 500	446.67	199 514.09	156 250 000
13 500	1446.67	2 092 854.09	182 250 000
24 500	12 446.67	155 417 860.09	600 250 000
27 100	15 046.67	226 402 278.09	734 410 000
30 900	18 846.67	355 196970.09	954 810 000
41000	28 946.67	837 909 704.09	1681000 000
		Total 2 287 817 333.3	Total 4 467 060000

4.3 Determine la desviación estándar muestral para los datos de los problemas 4.1 y 4.2 utilizando (a) la fórmula de desviaciones y (b) la fórmula abreviada alternativa, y demuestre que las respuestas son iguales.

(De la tabla 4.9.)

(a) $$s=\sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}=\sqrt{\frac{2\,187\,817\,333.3}{15-1}}=\sqrt{163\,415\,523.8}=\$12\,783.41$$

(b) $$s=\sqrt{\frac{\Sigma X^2-n\bar{X}^2}{n-1}}=\sqrt{\frac{4\,467\,060\,000-15(12\,053.33)^2}{15-1}}=\sqrt{163\,415\,609.9}\approx\$12\,783.41$$

Las respuesta :-, son ligeramente diferentes debido a error de redondeo, junto con el hecho de que la media muestral se redondeó a dos cifras decimales.

4.4 Una muestra de 20 trabajadores calificados de una compañía pequeña obtuvieron los siguientes salarios en un día determinado (cifras redondeadas al peso más cercano y ordenadas): $2400, 2400, 2400, 2400, 2400, 2400, 2400, 2400, 2550, 2550, 2650, 2650, 2800 2800, 2900, 3000, 3050, 3250, 3300, y 3400. Determine: (a) el rango y (b) el rango central del 80% para esta muestra.

(a) $R = M_y - M_n$ = $340 000 − $ 240 000 = $100 000

(b) R central 80% = $P_{90} - P_{10}$ = $ 327 500 − $ 240 000 = 87 500

donde P_{90} $X_{[(90n/100)+(1/2)]} = X_{[18+(1/2)]} = X_{18.5} =$ 325 000 + $ 2 500 =$ 327 005

$P_{10} = X_{[(10n/100)+(1/2)]} = X_{[2+(1/2)]} = X_{2.5} =$ $ 240 000

4.5 Calcule la desviación media de los salarios del problema 4.4. En el problema 3.4 se determinó que la media de esos salarios sea de $2705. De la Tabla 4.10, la desviación media es

$$DM=\frac{\Sigma\lvert X-\bar{X}\rvert}{n}=\frac{\$572.00}{20}=\$28.60$$

Tabla 4.10 Hoja de trabajo para calcular la desviación media para los datos de salarios

X	$X-\bar{X}$	$\lvert X-\bar{X}\rvert$
$2 40000	$-30500	$-30500
240 000	-30500	-30500
240000	-30 500	-30 500
240 000	-30 500	-30 500
240000	-30 500	-30 500
240 000	-30 500	-30 500
240000	-30 500	-30 500
240 000	-30 500	-30500
255 000	-15 500	-15 500
255 000	-15500	-15500
265 000	-5 500	-5 500
265 000	-5 500	-5 500
280 000	9 500	9 500
280 000	9 500	9 500
290 000	19 500	19 500
300 000	29 500	29 500
305000	34 500	34 500
325 000	54 500	54 500
330 000	59 500	59 500
340000	69 500	69 500
		Total $572 000

Tabla 4.11 Hoja de trabajo para calcular la varianza y la desviación estándar muestrales para los datos de salarios

X	$X-\bar{X}$	$(X-\bar{X})^2$
$240 000	$-30 500	$ 930 250
240000	-30 500	930 250
240000	-30 500	930 250
240000	-30 500	930 250
240000	-30 500	930 250
240000	-30 500	930 250
240 000	-30 500	930 250
240000	-30 500	930 250
255 000	-15 500	240 250
255 000	-15500	240 250
265 000	-5 500	30 250
265 000	-5 500	30250
280 000	9 500	90 250
280000	9 500	90 250
290000	19 500	380 250
300000	29 500	870 250
305000	34 500	1,190250
325 000	54 500	2,970 250
330000	59 500	3,540 250
340000	69 500	4,830250
		Total $21,945 000

4.6 Determine (a) la varianza muestral y (b) la desviación estándar muestral para los datos de los problemas 4.4 y 4.5 utilizando las fórmulas de desviación.

Con referencia a la tabla 4.11,

(a) $$\frac{21{,}945.00}{20-1} = 1{,}155.00$$

(b)

4.7. Un experto en estándares laborales observa el tiempo que se requiere para preparar una muestra de 10 cartas de negocios en una oficina, obteniendo los siguientes resultados (en orden ascendente y redondeados al minuto más cercano: 5, 5, 5, 7, 9.14.15,15,16,18. Determine: (*a)* el rango v *(b)* el rango central del 70% para la muestra.

(a) $R = M_y - M_n =$ 18-5=13min

(*b*) R central 70% $= P_{85} - P_{15} = 16.0 - 5.0 =$ 11.0 min

donde $P_{85} = X_{[85n/100)+(1/2)]} = X_{(8.5+0.5)} = X_9 +$ 16.0 min

$P_{15} = X_{[(15n/100)+(1/2)'} = X_{(1.5+0.5)} = X_2 =$ 5.0 min

4.8 Calcule la desviación media para el tiempo de preparación del problema 4.7. En el problema 3.7 se calculó la media muestral y resultó ser de 10.9 minutos.

De la tabla 4.12,

$$DM = \frac{\Sigma |X - \bar{X}|}{n} = \frac{47.0}{10} = 4.7 \text{ min}$$

Tabla 4.12 Hoja de trabajo para calcular la desviación media para los datos de tiempo de preparación

X	X-X	$\lvert X - \bar{X} \rvert$
5	-5.9	5.9
5	-5.9	5.9
5	-5.9	5.9
7	-3.9	3.9
9	-1.9	1.9
14	3.1	3.1
15	4.1	4.1
15	4.1	4.1
16	5.1	5.1
18	7.1	7.1
		Total 47.0

Tabla 4.13 Hoja de trabajo para calcular la desviación estándar y la varianza muéstrales para los datos de tiempo de preparación

X	X^1
5	25
5	25
5	25
7	49
9	81
14	196
15	225
15	225
16	256
18	324
	Total 1,431

4.9 Determine (a) la varianza muestral y (b) la desviación estándar muestral para los tiempos de mecanografía dados en los problemas 4.7 y 4.8, utilizando las fórmulas abreviadas alternativas.

Con referencia a la tabla 4.13,

(a) $s^2 = \frac{\Sigma X^2 - n\bar{X}^2}{n-1} = \frac{1{,}431 - 10(10.9)^2}{10-1} = 26.99$

(b) $s = \sqrt{\frac{\Sigma X^2 - n\bar{X}^2}{n-1}} = \sqrt{26.99} = 5.2 \text{ min}$

EL COEFICIENTE DE VARIACIÓN

4.10 Determine el coeficiente de variación para los datos de salarios analizados en los problemas 4.4 a 4.6.

Como $\bar{X}$ = \$270.50 y s - \$33.99,

$$CV = \frac{s}{\bar{X}} = \frac{33.99}{270.50} = 0.126$$

4.11 Para la misma empresa industrial del problema 4.10, el salario diario promedio de una muestra de supervisores es $\bar{X}$= \$7307.50, con s » \$455.20. Determine el coeficiente de variación para estos salarios.

$$CV = \frac{s}{\bar{X}} = \frac{45.52}{730.75} = 0.062$$

4.12 Compare la variabilidad de los salarios de los trabajadores especializados del problema 4.10 con los salarios de los supervisores del 4.11 (a) en términos absolutos y (b) con respecto al nivel promedio de ingresos diarios para los dos grupos de empleados.

(a) En términos absolutos, existe mayor variabilidad entre los sueldos de los supervisores (s - $455.20) que entre los salarios de los obreros (s = $ 339.90)

(6) En relación con las medias respectivas, si se compararan los dos coeficientes de variación de acuerdo con las soluciones obtenidas para los problemas 4.10 y 4.11, puede observarse que el coeficiente de variación para los datos de salarios por hora ($126.00) es dos veces mayor que el coeficiente correspondiente para los salarios semanales ($63.00), lo cual indica una mayor variabilidad *relativa* para los datos de salarios.

COEFICIENTE DE ASIMETRÍA DE PEARSON

4.13 Calcule el coeficiente de asimetría para los datos del restaurante que se analizaron en los problemas 4.1 a 4.3.

Como $\bar{X}$= $12 053.33, Med (del problema 3.1) y s = $12783.41

$$\text{Asimetría} = \frac{3(\bar{X} - \text{Med})}{s} = \frac{3(12\,053.53 - 5300)}{12\,783.41} = 1.58$$

Por lo tanto, la distribución de los montos de los importes tiene asimetría positiva o, lo que es lo mismo, está sesgada a la derecha.

4.14 Calcule el coeficiente de asimetría para los datos de los salarios analizados en los problemas 4.4 a 4.7.

Como $\bar{X}$ - $ 70.50, Med - $260.00 (del problema 3.4) y s -$33.99,

$$\text{Asimetría} = \frac{3(\bar{X} - \text{Med})}{s} = \frac{3(270.50 - 260.00)}{33.99} = 0.93$$

Por lo tanto, la distribución de los salarios tiene un ligero sesgo positivo.

LOS RANGOS, LA DESVIACIÓN MEDIA Y LA DESVIACIÓN ESTÁNDAR PARA DATOS AGRUPADOS.

4.15 Determine (a) el rango y (b) el rango central del 50% para las rentas de la Tabla 2.5 (página 17) Supóngase que esos son todos los departamentos de un área geográfica dada.

(a) $R = L_s(A) - L_1(B)$ = 649.50 - 349.50 = $300.00

(b) R central 50% = $P_{75} - P_{25}$ = 567.50 - 484.05 = $83.45

en donde

$$P_{75} = L_I + \left(\frac{\frac{75\,N}{100} - fa_B}{f_c}\right) i = 559.50 + \left(\frac{150 - 142}{30}\right) 30 = \$567.50$$

$$P_{25} = L_I + \left(\frac{\frac{25\,N}{100} - fa_B}{f_c}\right) i = 469.50 + \left(\frac{50 - 34}{33}\right) 30 = \$484.05$$

4.16 Calcule la desviación media para los datos de las rentas de la Tabla 2.5. Se determinó que la media de la población era de $ 522.75 en el problema 3.13.

Al utilizar la Tabla 4.14

$$DM = \frac{\Sigma(f|X-\mu|)}{N} = \frac{\$9{,}925.50}{200} = \$49.63$$

Tabla 4.14 Hoja de trabajo para calcular la desviación media para datos agrupados

Renta mensual (en miles de pesos)	Punto medio de clase (X)	Número de departamentos	$\|X-\mu\|$	$f\|X-\mu\|$
\$350-379	\$364.50	3	\$158.25	\$474.75
380-409	394.50	8	128.25	1,026.00
410-439	424.50	10	98.25	982.50
440-469	454.50	13	68.25	887.25
470-499	484.50	33	38.25	1,262.25
500-529	514.50	40	8.25	330.00
530-559	544.50	35	21.75	761.25
560-589	574.50	30	51.75	1,552.50
590-619	604.50	16	81.75	1,308.00
620-649	634.50	12	111.75	1,341.00
		Total 200		Total \$9,925.50

4.17 A partir del problema 4.14, determine la desviación estándar de la población utilizando (a) la fórmula de desviaciones y (b) la fórmula abreviada, y demuestre que las respuestas son equivalentes,

(a) De la tabla 4.15

$$\sigma = \sqrt{\frac{\Sigma[f(X-\mu)^2]}{N}} = \sqrt{\frac{768{,}487.50}{200}} = \sqrt{3{,}842.4375} \cong \$61.99$$

Tabla 4.15 Hoja de trabajo para calcular la desviación estándar mediante la fórmula de desviaciones

Renta mensual (en miles de pesos)	Punto medio de clase (X)	Número de departamentos (f)	$X-\mu$	$(X-\mu)^2$	$f(X-\mu)^2$
\$350-379	\$364.50	3	\$158.25	\$25,043.0625	\$ 75,129.1875
380-409	394.50	8	128.25	16,448.0625	131,584.5000
410-439	424.50	10	98.25	9,653.0625	96,530.6250
440-69	454.50	13	68.25	4,658.0625	60,554.8125
470-499	484.50	33	38.25	1,463.0625	48,281.0625
500-529	514.50	40	8.25	68.0625	2,722.5000
530-559	544.50	35	21.75	473.0625	16,557.1875
560-589	574.50	30	51.75	2,678.0625	80,341.8750
590-619	604.50	16	81.75	6,683.0625	106,929.0000
620-649	634.50	12	111.75	12,488.0625	149,856.7500
		Total 200			Total 768,487.5000

(b) De la Tabla 4.16,

$$\sigma = \sqrt{\frac{\Sigma(fX^2) - N\mu^2}{N}} = \sqrt{\frac{55{,}422{,}000.00 - 200(522.75)^2}{200}} = \sqrt{3{,}842.4375} \approx \$61.99$$

Tabla 4.16 Hoja de trabajo para calcular la desviación estándar utilizando la fórmula abreviada

Renta mensual (en miles de pesos)	Punto medio de clases (X)	Número de departamentos *(f)*	X^2	fX^2
\$350-379	\$364.50	3	\$132,860.25	\$ 398,580.75
380-409	394.50	8	155,630.25	1,245,042.00
410-439	424.50	10	180,200.25	1,802,002.50
440-469	454.50	13	206,570.25	2,685,413.25
470-499	484.50	33	234,740.25	7,776,428.25
500-529	514.50	40	264.710.25	10,588,410.00
530-559	544.50	35	296,480.25	10,376,808.75
560-589	574.50	30	330,050.25	9,901,507.50
590-619	604.50	16	365,420.25	5,846,724.00
620-649	634.50	12	402,590.25	4,831.00
		Total 200		Total 55,422,000.00

4.18 Junto con una auditoría anual, un despacho de contadores recopila los datos que se reportan en la Tabla 4.17. Determine *(a)* el rango y *(b)* el rango central del 80% para esta muestra de registros.

Tabla 4.17 Tiempo que se requiere para auditar saldos de cuentas

Tiempo de auditoría (en minutos)	Límites exactos de clase	Número de registros *(f)*	Frecuencia acumulada *(fa)*
10-19	9.5-19.5	3	3
20-29	19.5-29.5	5	8
30-39	29.5-39.5	10	18
40-49	39.5-49.5	12	30
50-59	49.5-59.5	20	50
		Total 50	

Utilizando la Tabla 4.17

(a) $R = L_s(A) - L_1(A) - 59.5 - 9.5 = 50.0$ min

(b) R central 80% $= P_{90} - P_{10} = 57.0 - 23.5 = 33.5$ min

en donde $P_{90} = L_I + \left(\dfrac{\frac{90\,n}{100} - fa_B}{f_c}\right) i = 49.5 + \left(\dfrac{45 - 30}{20}\right) 10 = \$57.0 \text{ min}$

y también $P_{10} = L_l + \left(\dfrac{\frac{10n}{100} - fa_B}{f_c}\right) i = 19.5 + \left(\dfrac{5-3}{5}\right) 10 = 23.5 \text{ min}$

4.19 Calcule la desviación inedia para los datos de tiempos de auditoría de la Tabla 4.17. En el problema 3.15 se determinó que la media muestral era de 3.7 minutos.

Con referencia a la Tabla 4.18,

$$DM = \frac{\Sigma(f|X - \bar{X}|)}{n} = \frac{515.2}{50} = 10.3$$

Tabla 4.18 Hoja de trabajo para calcular la desviación media

Tiempo en auditoría (en minutos)	Punto medio de clase (X)	Número de registros (f)	$\|X - \bar{X}\|$	$f\|X - \bar{X}\|$
10-19	14.5	3	28.2	84.6
20-29	24.5	5	18.2	91.0
30-39	34.5	10	8.2	82.0
40-49	44.5	12	1.8	21.6
50-59	54.5	20	11.8	236.0
		Total 50		Total 515.2

4.20 Determine (a) la varianza muestral y (b) la desviación estándar muestral para los datos de la Tabla 4.17, utilizando las fórmulas de desviaciones.

De la Tabla 4.19,

(a) $s^2 = \dfrac{\Sigma[f(X - \bar{X})^2]}{n-1} = \dfrac{7,538}{49} = 153.84$

(b) $s = \sqrt{\dfrac{\Sigma[f(X - \bar{X})^2]}{n-1}} = \sqrt{153.84} = 12.4 \text{ min}$

Tabla 4.19 Hoja de trabajo para calcular la varianza y la desviación estándar

Tiempo de auditoría (em minutos)	Punto medio de clase (X)	Número de registros (f)	$X - \bar{X}$	$(X - \bar{X})^2$	$f(X - \bar{X})^2$
10-19	14.5	3	-28.2	795.24	2,385.72
'20-29	24.5	5	-18.2	331.24	1,656.20
30-39 •	34.5	10	-8.2	67.24	672.40
40-49	44.5	12	1.8	3.24	38.88
50-59	54.5	20	11.8	139.24	2,784.80
		Total 50			Total 7,538.002

4.21 En la Tabla 4.20 se reproducen los datos sobre el número promedio de lesiones por millar horas-hombre, tal como se reportó en el problema 2.16; determine (á) el rango y *(b)* el rango central del 90% para esta muestra de 50 empresas.

Tabla 4.20 Número promedio de lesiones por millar de horas-hombre en una industria específica

Número promedio de lesiones	Límites exactos de clase	Número de empresas (f)	Frecuencia acumulada (fa)
1.5-1.7	1.45-1.75	3	3
1.8-2.0	1.75-2.05	12	15
2.1-2.3	2.05-2.35	14	29
2.4-2.6	2.35-2.65	9	38
2.7-2.9	2.65-2.95	7	45
3.0-3.2	2.95-3.25	5	50
		Total 50	

(a) $R = L_s\,(A) - L_1\,(A)$ - 3.25 - 1.45 = 1.80 lesiones
(b) R central 90% = $P_{95} - P_{05}$ = 3.10 - 1.70 = 1.40 lesionas

en donde $\pm P_{95} = L_I + \left(\dfrac{\frac{95n}{100} - fa_B}{f_c}\right) i = 2.95 + \left(\dfrac{47.5 - 45}{5}\right) 0.30 = 3.10$ lesiones

$$P_{05} = L_I + \left(\frac{\frac{5n}{100} - fa_B}{f_c}\right) 0.30 = 1.45 + \left(\frac{2.5 - 0}{3}\right) 0.30 = 1.70 \text{ lesiones}$$

4.22 Calcule la desviación media para el número de lesiones de la Tabla 4.20. En el problema 3.17 se determinó que la media muestral era de 2.32 lesiones por millar de horas-hombre.

Con referencia a la Tabla 4.21,

$$DM = \frac{\Sigma(f|X - \bar{X}|)}{n} = \frac{17.76}{50} = 0.36 \text{ lesiones}$$

Tabla 4.21 Hoja de trabajo para calcular la desviación media

Número promedio de lesiones	Punto medio de clase (X)	Número de empresas (f)	$\lvert X - \bar{X}\rvert$	$f\lvert X - \bar{X}\rvert$
1.5-1.7	1.6	3	0.72	2.16
1.8-2.0	1.9	12	0.42	5.04
2.1-2.3	2.2	14	0.12	1.68
2.4-2.6	2.5	9	0.18	1.62
2.7-2.9	2.8	7	0.48	3.36
3.0-3.2	3.1	5	0.78	3.90
				Total 17.76
		Total 50		

4.23 Determine (a) la varianza muestral y *(b)* la desviación estándar muestral para los datos de lesiones de la Tabla 4.20, utilizando las fórmulas abreviadas.

De la Tabla 4.22,

(a) $s^2 = \dfrac{\Sigma(fX^2) - n\bar{X}^2}{n-1} = \dfrac{277.94 - 50(2.32)^2}{50-1} = \dfrac{8.82}{49} = 0.1800$

(b) $s = \sqrt{\dfrac{\Sigma(fX^2) - n\bar{X}^2}{n-1}} = \sqrt{0.1800} \approx 0.42$ lesiones

Tabla 4.22 Hoja de trabajo para calcular la varianza y la desviación estándar

Número promedio de lesiones	Punto medio de clase *(X)*	Número de empresas *(f)*	X^2	fX^2
1.5-1.7	1.6	3	2.56	7.68
1.8-2.0	1.9	12	3.61	43.32
2.1-2.3	2.2	14	4.84	67.76
2.4-2.6	2.5	9	6.25	56.25
2.7-2.9	2.8	7	7.84	54.88
3.0-3.2	3.1	5	9.61	48.05
		Total 50		Total 277.94

4.24 La distribución de frecuencias de la Tabla 4.23, se reproduce del problema 2.21. Determine (a) el rango y *(b)* la desviación estándar, utilizando la fórmula abreviada. Observe que la base para determinar los limites exactos de clase y los puntos medios de clase para distribuciones de frecuencias de "y menor que" como ésta. En el problema 3.19 se determinó que la media muestral era de 10.8 minutos.

Tabla 4.23 Tiempo que se requiere para procesar y preparar órdenes de restaurante

Tiempo, minutos	Límites exactos de clase	Punto medio de clase (X)	Número de pedidos *(f)*	X^2	fX^2
5 y menor que 8	5.8-8.0	6.5	10	42.25	422.50
8 y menor que 11	8.0-11.0	9.5	17	90.25	1,534.25
11 y menor que 14	11.0-14.0	12.5	12	156.25	1,875.00
14 y menor que 17	14.0-17.0	15.5	6	240.25	1,441.50
17 y menor que 20	17.0-20.0	18.5	2	342.25	684.50
			Total 47		Total 5,957.75

(a) $R = Ls\,(A) - L_I\,(A) = 20.0 - 5.0 = 15.0$ min

(b) $s = \sqrt{\dfrac{\Sigma(fX^2) - n\bar{X}^2}{n-1}} = \sqrt{\dfrac{5{,}957.75 - 47(10.8)^2}{47-1}} = \sqrt{10.34\ 153.84} \approx 3.2$min

RESULTADOS POR COMPUTADORA

4.25 Utilice una computadora para determinar el rango y la desviación estándar para los datos de la Tabla 2.6 (página 18), sobre el tiempo de una muestra de 30 empleados requirió para terminar una tarea de ensamble. En el problema 3.24 se determinaron la media y la mediana para estos datos utilizando una computadora.

En la figura 4-2, se presentan los datos de entrada y los resultados obtenidos en la computadora. Como puede observarse, *R* - 9.0 y s = 2.4375 = 4.2 minutos. Obsérvese que en el paquete de Minitab, la instrucción DESVIACIÓN ESTÁNDAR da como resultado la desviación estándar *muestral.*

```
MTB > SET ASSEMBLY TIMES INTO Cl
DATA>   10   14   15   13   17   16   12   14   11   13   15   18   9
DATA>   14   14    9   15   11   13   11   12   10   17   16   12   11
DATA>   16   12   14   15
DATA> END
MTB > NAME FOR Cl IS 'TIME'
MTB > MAXIMUM 'TIME', PUT IN K1
   MÁXIMUM    =   18.000
MTB > MÍNIMUM 'TIME', PUT IN K2
   MÍNIMUM    =   9.0000
MTB > SUBTRACT K2 FROM Kl, PUT RANGE INTO K3
    ANSWER     =     9.0000
MTB > STANDARD DEVIATION OF 'TIME'
   ST.DEV.    =   2.4375
```

Fig. 4-2 Resultados Minitab.

PROBLEMAS COMPLEMENTARIOS

LOS RANGOS, LA DESVIACIÓN MEDIA Y LA DESVIACIÓN ESTÁNDAR PARA DATOS NO AGRUPADOS

4.26 El número de automóviles que vendió cada uno de 10 vendedores de una distribuidora en un mes específico, en orden ascendente, es: 2,4, 7,10,10,10,12,12,14,15. Determine (a) el rango y *(b)* el rango del 80% central para esos datos.

Resp. (a) 13, (b) 11.5

4.27 Calcule la desviación media para los datos de ventas para el problema 4.26. En el problema 3.25 se determinó que la media de la población para esos valores sea de 9.6.

Resp. 3.16 = 3.2

4.28 Del problema 4.26, determine la desviación estándar utilizando la fórmula de desviaciones y considerando que el grupo de valores constituye una población estadística.

Resp. 3.995 ≅ 4.0

4.29 Se determina que los pesos de una muestra de sobres que salen de una oficina postal, al gramo más cercano son: 21,18,30,12,14,17, 28,10,16 y 25. Determine (a) el rango y (b) el rango central del 50% para esos pesos.

Resp. (a) 20.0,(b) 11.0

4.30 Calcule la desviación estándar para las piezas postales del problema 4.29. En el problema 3.27 se determinó que la media muestral sea de 19.1 gramos.

Resp. 5.52 ≅ 5.5

4.31 Determine (a) la varianza muestral y (b) desviación estándar muestral para los datos del problema 4.29. Utilizando las versiones abreviadas de las fórmulas respectivas.

Resp. (a) 45.7, (b) 6.8

4.32 Veinte estudiantes inscritos en un curso de análisis de decisiones lograron las siguientes calificaciones en sus exámenes, en orden ascendente: 39, 46, 57, 65, 70, 72, 72, 75, 77, 79, 81, 81, 84, 84, 84, 87, 93, 94. 97 y 97. Determine (a) el rango y *(b)* el rango central del 90% para estos datos no agrupados.

Resp. (a) 58.0, (b) 6.8

4.33 Calcule la desviación media para las calificaciones del problema 4.32. En el problema 3.29 se determinó que la calificación promedio de los exámenes fue de 76.7

Resp. 11.76 ≅ 11.8

4.34 Considere las calificaciones de los exámenes del problema 4.32 como una población estadística y determine la desviación estándar utilizando (a) la fórmula de desviaciones y (b) la fórmula abreviada alternativa.

Resp. (a) 15.294 ≅ 15.3, (b) 15.294 ≅ 15.3

4.35 El número de accidentes que ocurrieron en un mes determinado en 13 departamentos de manufactura de una planta industrial fueron: 2, 0, 0, 3, 3, 12, 1, 0, 8, 1, 0, 5, 1. Determine (a) el rango y (b) el rango central del 50% para el número de accidentes.

Resp. (a) 12.0, (b) 3.5

4.36 Calcule la desviación media para los datos del problema 4.35. En el problema 3.31 se determinó que el número promedio de accidentes es de 2.8.

Resp. 2.646 ≅ 2.6

4.37 Al considerar los datos de los accidentes del problema 4.35 como una población estadística, calcule la desviación estándar utilizando la fórmula abreviada.

Resp. 3.465 ≅ 3.5

EL COEFICIENTE DE VARIACIÓN

4.38 Determine el coeficiente de variación para los datos de ventas de automóviles que se utilizaron en los problemas 4.26 a 4.28.

Resp. CV=0.417

4.39 Con referencia al problema 4.38, en otra distribuidora (más grande) se determinó que el número promedio de automóviles vendidos por vendedor en un mes especifico fue μ - 17.6, con una desviación estándar de σ = 6.5. Compare la variabilidad con las ventas por persona (a) en términos absolutos y (b) con respecto al nivel promedio de ventas en las dos distribuidoras.

Resp. (a) La desviación estándar para la primera distribuidora (4.0) es menor que la desviación estándar para la segunda distribuidora (6.5) (6) El coeficiente de variación para la primera distribuidora (0.417) es mayor que el coeficiente de variación para la segunda distribuidora (0.369)

COEFICIENTE DE ASIMETRÍA DE PEARSON

4.40 Calcule el coeficiente de asimetría para los datos de ventas de automóviles que se analizaron en los problemas 4.26 a 4.28. En el problema 3.25 se determinó que la mediana para esos datos fue 10.0.

Resp. Asimetría = - 0.30 (por ello, la distribución de las ventas de automóviles está ligeramente sesgada en forma negativa, es decir, hacia la izquierda).

4.41 Calcule el coeficiente de asimetría para los datos de accidentes analizados en los problemas 4.35 a 4.37. En el problema 3.31 se determinó que la mediana para estos datos es de 1.0.

Resp. Asimetría =1.54 (por ello, la distribución de accidentes tiene asimetría positiva)

LOS RANGOS, LA DESVIACIÓN MEDIA Y LA DESVIACIÓN ESTÁNDAR PARA DATOS AGRUPADOS.

4.42 Determine (a) el rango y *(b)* el rango central del 80% para los datos de kilometraje de la Tabla 2.15 (página 28)

Resp. (a)12.00,(b)6.83

4.43 Calcule la desviación media para los datos de kilometraje de la Tabla 2.15. En el problema 3.40 se determinó que el kilometraje promedio fue de 28.95.

Resp. 1.76

4.44 Al considerar que los datos de kilometraje de la Tabla 2.15 son una muestra de viajes, calcule la desviación estándar para esos datos utilizando (a) la fórmula de desviaciones y (o) la fórmula abreviada. En el problema 3.40 se determinó que el kilometraje promedio fue de 28\.95.

Resp. (a) 1.7435 = 1.74, (b) 1.7435 = 1.74

4.45 Con los datos de la Tabla 3.10 (página 48),determine (a) el rango y (b) el rango central del 50% para el monto de los préstamos personales.

Resp. (a) $2,800.00, (b) $892.31

4.46 Calcule la desviación media para los datos de los préstamos personales de la Tabla 3.10. En el problema 3.42 se determinó que la media para esa muestra sea de $ 1109.50.

Resp. $512.00

4.47 Determine la variación estándar muestral para los datos de la Tabla 3.10 utilizando la fórmula abreviada.

Resp. $627.49

4.48 Determine (a) el rango y (b) el rango central del 90% para el tiempo útil de las herramientas de corte de la Tabla 2.17 (página 29)

Resp. (a) 149.95 h, (b) 93.75 h

4.49 Calcule la desviación media para la vida útil de las herramientas de corte reportada en la Tabla 2.17. En el problema 3.44 se determinó que la vida últil promedio sea de 87.45 horas.

Resp. 18.57 h

4.50 Al considerar que los datos de la Tabla 2.17 son una población estadística, calcule (a) la varianza poblacional y *(b)* la desviación estándar poblacional utilizando las fórmulas de desviaciones. En el problema 3.44 se determinó que la vida útil promedio sea de 87.45 horas.

Resp. (a) 714.29, (b) 26.73 h

4.51 De la Tabla 2.18 (página 30), determine (a) el rango y (b) la desviación media para la edad de los solicitantes. En el problema 3.46 se determinó que la edad promedio era de 23.3 años.

Resp. (a) 14.0, (b) 2.656 ≈ 2.7 años

4.52 Calcule la desviación estándar muestral para los datos de la Tabla 2.18 utilizando la fórmula abreviada. Se determinó en el problema 3.46 que la edad promedio era de 23.3 años.

Resp. 3.4 años

RESULTADOS POR COMPUTADORA

4.53 Utilice una computadora para determinar el rango y la desviación estándar para la muestra de 50 préstamos de la Tabla 2.16 (página 28). En el problema 3.52, se calcularon la media y la mediana para esos datos utilizando una computadora. También, en los problemas 4.45 y 4.47 se calcularon el rango y la desviación estándar aproximados, con base en los datos agrupados, y resultaron ser de 4.45 y 4.47, respectivamente.

Resp. R = $2700.00 y s = $642.24. Esto se compara con las aproximaciones de $ 2800 y $ 627.49, respectivamente, que fueron las cifras que se obtuvieron a partir de los montos agrupados.

Probabilidad

5

5.1 DEFINICIONES BÁSICAS DE PROBABILIDAD

Históricamente se han desarrollado tres diferentes enfoques conceptuales para definir la probabilidad y para determinar valores de probabilidad: el clásico, el de frecuencia relativa y el subjetivo.

De acuerdo con el *enfoque clásico* de la probabilidad, si $N(A)$ resultados elementales posibles son favorables en el evento A, y existe $N(S)$ posibles resultados en el espacio muestral y todos los resultados elementales son igualmente probables y mutuamente excluyentes; entonces, la probabilidad de que ocurra el evento A es

$$P(A) = \frac{N(A)}{N(S)} \tag{5.7}$$

Obsérvese que el enfoque clásico de la probabilidad se basa en la suposición de que cada uno de los resultados es igualmente probable. Debido a que este enfoque (cuando es aplicable) permite determinar los valores de probabilidad antes de observar cualesquiera eventos muéstrales, también se le denomina *enfoque a priori.*

EJEMPLO 1. En un mazo de cartas bien barajadas que contiene 4 ases y 48 cartas de otro tipo, la probabilidad de obtener un as (A) en una sola extracción es

$$P(A) = \frac{N(A)}{N(S)} = \frac{4}{52} = \frac{1}{13}$$

Através del *enfoque de frecuencia relativa,* se determina la probabilidad con base en la proporción de veces que ocurre un resultado favorable en un determinado número de observaciones o experimentos. No hay implícita ninguna suposición previa de igualdad de probabilidades. Debido a que para determinar los valores de probabilidad se requiere de la observación y de la recopilación de datos, a este enfoque se le denomina también *enfoque empírico.* La probabilidad de que ocurra un evento A, de acuerdo con el enfoque de frecuencia relativa es

$$P(A) = \frac{\text{Número de observaciones de} A}{\text{Tamaño de la muestra}} = \frac{n(A)}{n} \tag{5.2}$$

EJEMPLO 2. Antes de incluir la cobertura para ciertos tipos de problemas dentales en pólizas de seguros médicos para adultos con empleo, una compañía de seguros desea determinar la probabilidad de ocurrencia de esa clase de problemas, para que pueda fijarse la prima de seguros de acuerdo con esas cifras. Por ello, un especialista en estadística recopila datos para 10 000 adultos que se encuentran en las categorías de edad apropiadas y encuentra que 100 de ellos han experimentado el problema dental especifico durante el año anterior.

Por ello, la probabilidad de ocurrencia es:

$$P(A) = \frac{n(A)}{n} + \frac{100}{10{,}000} = 0.01, o\ 1\%$$

Tanto el enfoque clásico como el de frecuencia relativa producen valores de probabilidad *objetivos,* en el sentido de que señalan la tasa relativa de ocurrencia del evento a largo plazo. Por el contrario, el *enfoque subjetivo* a la probabilidad es particularmente apropiado cuando sólo existe una probabilidad de que el evento ocurra, y se da el caso de que ocurra o no esa única vez. De acuerdo con el enfoque subjetivo, la probabilidad de un evento es el *grado de confianza* que una persona tiene en que el evento ocurra, con base en toda la evidencia que tiene disponible. Debido a que el valor de la probabilidad es un juicio personal, al enfoque subjetivo se le denomina también *enfoque personalista.* El desarrollo de este enfoque de la probabilidad ha sido relativamente reciente y tiene relación con el análisis bayesiano de decisión (véanse las secciones 1.3 y los capítulos 18 a 20).

EJEMPLO 3. Debido a los impuestos y a los posibles usos alternativos de sus fondos, un inversionista ha determinado que la compra de terrenos vale la pena sólo si existe una probabilidad de cuando menos 0.90 de que el terreno obtenga plusvalía por 50% o más en los próximos 4 años. Al evaluar un determinado terreno, el inversionista estudia los cambios en los precios en el área en años recientes, considera los niveles corrientes de precios, estudia el estado corriente y futuro probable de los proyectos de desarrollo inmobiliarios y revisa las estadísticas referentes al desarrollo económico del área geográfica global. Con base en esta revisión, concluye que existe una probabilidad de aproximadamente 0.75% de que se dé la plusvalía que requiere. Como esta probabilidad es menor que la mínima que requiere, (0.90), no debe llevarse a cabo la inversión.

5.2 EXPRESIÓN DE LA PROBABILIDAD

Se utiliza el símbolo *P* para designar la probabilidad de un evento. Por ello, *P (A)* denota la probabilidad de que ocurra el evento A en una sola observación o experimento.

El menor valor que puede tener una probabilidad es 0 (lo cual indica que el evento es imposible) y el mayor valor que puede tomar es 1 (lo cual indica que es seguro que ocurra). Por ello, en general,

$$0 \leq P(A) \leq 1 \tag{5.3}$$

En una observación o experimento dados, el evento debe ocurrir o no. Por ello, la suma de la probabilidad de ocurrencia más la probabilidad de no ocurrencia siempre es igual a 1. Por ello, (en donde A' indica la no ocurrencia del evento (A), se obtiene:

$$P(A) + P(A') = 1 \tag{5.4}$$

Un *diagrama de Venn* es un diagrama relacionado con la teoría de conjuntos en matemáticas, que permite ilustrar los eventos que pueden ocurrir en una observación o experimento específicos. Una figura cerrada representa un espacio muestral, y se utilizan porciones del área dentro de ese espacio para representar eventos elementales o compuestos o espacios de eventos.

EJEMPLO 4. En la figura 5.1 se representan las probabilidades de dos eventos, A y A' (que se lee "no A"). Como *P (A)* + *P* (A') = 1, se considera la totalidad del área dentro del diagrama.

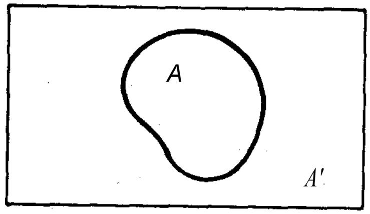


Fig. 5-1

Como una alternativa a los valores de probabilidad, pueden expresarse las probabilidades también en términos de *posibilidades,* La razón de posibilidad que favorece la ocurrencia de un evento es el cociente del número relativo de resultados, que se designa como a, que le son favorables a *A,* con respecto al número relativo de resultados, que se designan mediante *b,* que no le son favorables *a A:*

$$\text{Posibilidades} = a : b \text{ ("}a \text{ a } b\text{")} \tag{5.5}$$

Una posibilidad de 5:2 (que se lee "5 a 2") indica que para cada cinco eventos elementales que constituyen éxitos existen dos eventos elementales que constituyen fracasos. Obsérvese que, de acuerdo con el enfoque clásico de la probabilidad analizado en la sección 5.1, el valor de probabilidad equivalente a una razón de probabilidad de 5:2 es

$$P(A) = \frac{N(A)}{N(S)} = \frac{a}{a+b} = \frac{5}{5+2} = \frac{5}{7}$$

EJEMPLO 5. Suponga que se define el éxito como la extracción de cualquier naipe con figura *o* un as de un mazo bien barajado de 42 cartas. Como 16 cartas de las 52 son jotos, reinas, reyes o ases, las posibilidades asociadas con el éxito son 16:36, o 4:9. La probabilidad de éxito es 16/(16 + 36) =16/52 = 4/13.

5.3 EVENTOS MUTUAMENTE EXCLU/ENTES Y NO EXCLUYENTES

Dos o más eventos son *mutuamente excluyentes, ó disjuntos,* si no pueden ocurrir al mismo tiempo. Es decir, la ocurrencia de un evento automáticamente impide la ocurrencia del otro (u otros). Por ejemplo, supóngase que se consideran dos posibles eventos "as" y "rey" con respecto a la extracción de una carta de un mazo. Estos dos eventos son mutuamente excluyentes porque ninguna carta puede ser al mismo tiempo as y rey.

Dos o más eventos son *no excluyentes* cuando es posible que ocurran al mismo tiempo. Obsérvese que esta definición *no* indica que esos eventos deban necesariamente ocurrir en forma conjunta. Por ejemplo, supóngase que se consideran los dos posibles eventos "as" y trébol". Estos eventos no son mutuamente excluyentes porque una carta determinada puede ser al mismo tiempo as y trébol; sin embargo, esto no indica que todo as sea trébol o que todo trébol sea as.

EJEMPLO 6. En un estudio de la conducta de los consumidores, un analista clasifica a las personas que entran en una tienda de aparatos de sonido de acuerdo con su sexo ("masculino" o "femenino") y su edad ("menor de 30" o "30 o mayor"). Los dos eventos o clasificaciones, "masculino" y "femenino" son mutuamente excluyentes puesto que ninguna persona podría clasificarse en ambas categorías. De manera similar, los eventos "menor de 30" y "30 o mayor" son también mutuamente excluyentes. Sin embargo, los eventos "masculino" y "menor de 30" no son mutuamente excluyentes porque una persona elegida al azar podría estar en ambas categorías.

5.4 LAS REGLAS DE ADICIÓN

Se utilizan las reglas de la adición cuando se desea determinar la probabilidad de que ocurra un evento *u* otro (o ambos) en una sola observación. Simbólicamente, puede representarse la probabilidad de que ocurra el evento *A* o el evento *B* mediante *P (A O* B). En el lenguaje de la teoría de conjuntos, a esto se le denomina la *unión de A* y B, y se designa la probabilidad mediante *P (A* U B), que se lee: ("la probabilidad de A unión B".)

Existen dos variaciones a la regla de la adición, dependiendo de que los dos eventos sean mutuamente excluyentes. La regla de la adición para eventos mutuamente excluyentes es

$$P(A \text{ o } B) = P(A \cup B) = P(A) + P(B) \tag{5.6}$$

EJEMPLO 7. Cuando se extrae una carta de un mazo de barajas, los eventos "as" *(A*) y "rey" (R) son mutuamente excluyentes. La probabilidad de extraer ya sea un as o un rey en una sola extracción es:

$$P(A \text{ o } K) = P(A) + P(K) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}$$

(Nota: el problema 5.4 amplía la aplicación de esta regla a tres eventos.)

Para eventos que *no* son mutuamente excluyentes, se le resta a la suma de las probabilidades simples de dos eventos, la probabilidad de la ocurrencia conjunta de los dos eventos. Puede representarse la probabilidad de la ocurrencia conjunta mediante *P (A* y *B).* En el lenguaje de la teoría de conjuntos, se le denomina a esto la *intersección* de *A y B* y se designa la probabilidad como *P (A* n B) (que se lee: "la probabilidad de A intersección B"). Así, la regla de la adición para eventos que no son mutuamente excluyentes es

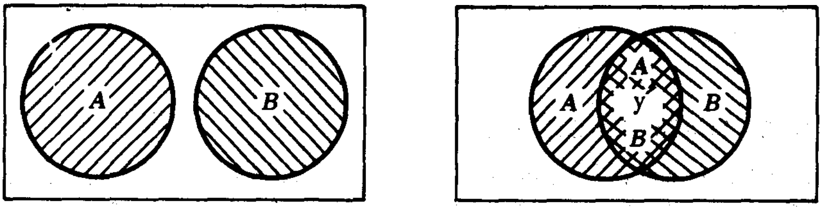
$$P(A \text{ O } B) = P(A) + P(B) - P(A \text{ y } B) \tag{5.7}$$

A la fórmula (5.7) también se le denomina con frecuencia *regla general de la adición* porque para eventos que son mutuamente excluyentes el último término es siempre igual a 0. Así, la fórmula es algebraicamente equivalente a la fórmula 5.6 para eventos mutuamente excluyentes.

EJEMPLO 8. Cuando se extrae una carta de un mazo de barajas, los eventos "as" y "trébol" no son mutuamente excluyentes. La probabilidad de obtener un as *(A)* o un trébol *(T)* (o ambos) en una sola extracción es

$$P(A \text{ o } T) = P(A) + P(T) - P(A \text{ y } T) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

Pueden utilizarse los diagramas de Venn para ilustrar el razonamiento en que se basan las dos reglas de la adición. En la figura 5-2(a), se observa que la probabilidad de que ocurran *A* o B es conceptualmente equivalente a sumar la proporción de área incluida *en A* y B. En la figura 5-2 (b), para eventos que no son mutuamente excluyentes.se incluyen algunos eventos elementales tanto en A y B; por ello, existe traslape entre estos conjuntos de eventos. Cuando se suman las áreas incluidas en *A* y B para eventos que no son mutuamente excluyentes, el área de traslape se suma dos veces. Por ello es que en la regla de la adición para eventos no excluyentes, se resta *P (A* y B) con el objeto de corregir la suma duplicada del área de intersección.



(a) (b)

Fig. 5-2

5.5 EVENTOS DEPENDIENTES, EVENTOS INDEPENDIENTES Y PROBABILIDAD CONDICIONAL

Dos eventos son *independiantes* cuando la ocurrencia o no ocurrencia de uno no tiene ningún efecto sobre la probabilidad de ocurrencia del otro. Dos eventos son *dependientes* cuando 'a ocurrencia o no ocurrencia de uno *sí* afecta la probabilidad de ocurrencia del otro evento.

EJEMPLO 9. Se considera que los resultados asociados con el lanzamiento de una moneda, dos veces seguidas son eventos independientes porque el resultado del primer lanzamiento no tiene ningún efecto sobre las respectivas probabilidades de que ocurra una cara o una cruz en el segundo lanzamiento. La extracción de dos cartas *sin reemplazo* de un mazo de barajas son eventos dependientes, porque las probabilidades asociadas con la segunda extracción dependen del resultado de la primera extracción. En forma específica, si saliera un "as" en la primera extracción, entonces la probabilidad de que salga un "as" en la segunda extracción es la razón del número de ases que sigue habiendo en el mazo con respecto al número total de cartas, o 3/51.

Cuando dos eventos son dependientes, se utiliza el concepto *de probabilidad condicional* para designar la probabilidad de ocurrencia de un evento relacionado. La expresión $P(B \mid A)$ indica la probabilidad de que ocurra el evento *B dado* que ocurre el evento A. Nótese que *"B |A" no* es una fracción.

No se requieren expresiones de probabilidad condicional para eventos independientes porque, por definición, no existe relación entre la ocurrencia de esos eventos. For lo tanto, si los eventos A y *B* son independientes, la probabilidad condicional *P(B | A)* siempre es igual a la probabilidad simple (no condicional) P (8). Por lo tanto, un enfoque que permite probar la independencia de dos eventos *A* y *B* consiste en comparar

$$P(B \mid A) \stackrel{?}{=} P(B) \qquad (5.8)$$

o

$$P(A \mid B) \stackrel{?}{=} P(A) \qquad (5.9)$$

(Véanse los problemas 5.8 a 5.10.)

Si se conocen la probabilidad simple (no condicional) de un primer evento A y la probabilidad conjunta de dos eventos A y b, entonces se puede determinar la probabilidad condicional *P (BA)* mediante:

$$P(B|A) = \frac{P(A \text{ y } B)}{P(A)} \qquad (5.10)$$

(Véanse los problemas 5.8 a 5.10.)

Con cierta frecuencia se confunde la diferencia entre eventos mutuamente excluyentes y no excluyentes, por un lado, y los conceptos de independencia y dependencia por el otro. Resulta particularmente importante observar la diferencia entre eventos que son mutuamente excluyentes y eventos que son independientes. La mutua exclusividad indica que dos eventos *no pueden* ocurrir al mismo tiempo, en tanto que la independencia señala que la probabilidad de ocurrencia de un evento no es afectada por la ocurrencia del otro. Por lo tanto, se sigue que si dos eventos son mutuamente excluyentes, se tiene un ejemplo especifico de eventos muy *dependientes,* porque la probabilidad de un evento dado que el otro ha ocurrido sería siempre igual a cero. Véase el problema 4.10.

5.6 LAS REGLAS DE LA MULTIPLICACIÓN

Las reglas de la multiplicación se refieren a la determinación de la probabilidad de la ocurrencia conjunta de *A* y B. Como se explicó en la sección 5.4, esto se refiere a la intersección de *A* y 8: *P* (A n B). Existen dos variaciones a la regla de la multiplicación de acuerdo a si los eventos son independientes o dependientes. La regla de la multiplicación para eventos independientes es

$$P(A \text{ y } B) = P(A \cap B) = P(A)\,P(B) \qquad (5.11)$$

EJEMPLO 10. De acuerdo con la fórmula 5.11, si se lanza dos veces una moneda, la probabilidad de que ambos resultados sean "cara" es un 1/2 X 1/2 = 1/4.

(Nota: en el problema 5.11 se amplía la aplicación de esa regla a tres eventos.)

Los diagramas de árbol son particularmente útiles para ilustrar los posibles eventos asociados con observaciones o ensayos secuenciales. La figura 5-3, es un ejemplo de estos diagramas para los eventos asociados con el lanzamiento de una moneda dos veces, e identifica los resultados posibles y la probabilidad en cada punto de la secuencia.

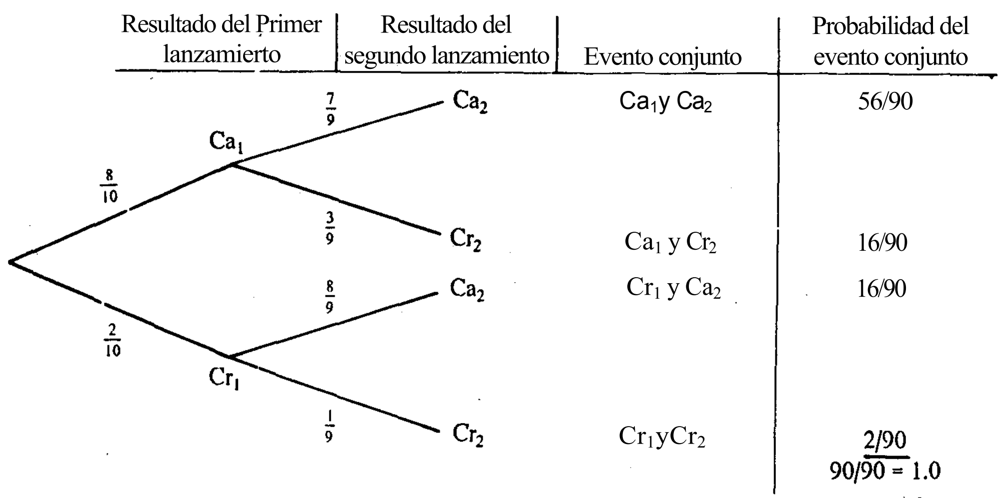


Fig. 5-3

EJEMPLO 11. Con referencia a la figura 5-3, se observa que son posibles cuatro tipos de secuencias, o eventos conjuntos: Ca y Ca, Ca y Cr, Cr y Ca, Cr y Cr. De acuerdo con la regla de la multiplicación para eventos independientes, en este caso la probabilidad de ocurrencias conjuntas para cualquiera de esas secuencias es 1/4, o 0.25. Como éstas son las únicas secuencias posibles, y como se trata de secuencias mutuamente excluyentes, de acuerdo con la regla de adición, la suma de las cuatro probabilidades conjuntas debe ser 1.0, lo cual resulta ser cierto.

Para eventos dependientes, la probabilidad de ocurrencia conjunta de *A* y B es la probabilidad de A multiplicada por la probabilidad *condicional* de *B* dado A. Se obtiene un valor equivalente si se invierte la posición de los dos eventos. Por ello, la regla de la multiplicación para eventos dependientes es:

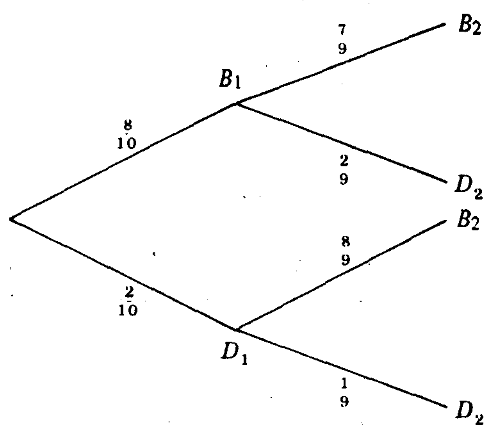
$$P(A \text{ y } B) = P(A)\, P(B \mid A) \tag{5.12}$$

$$P(A \text{ y } B) = P(B \text{ y } A) = P(B)\, P(A \mid B) \tag{5.13}$$

A la fórmula (5.*12)* o (5.*13)* se le denomina con frecuencia la *regla general de la multiplicación,* debido a que para eventos que son independientes, la probabilidad condicional *P (B* | *A)* es siempre igual al valor de la probabilidad no condicional *P* (S), lo cual da como resultado que la fórmula 5.12 sea equivalente a la 5.11 para eventos independientes.

EJEMPLO 12, Suponga que se sabe que un conjunto de 10 refacciones contiene ocho en buen estado *(B)* y dos partes defectuosas (D). Si se seleccionan al azar dos refacciones sin reemplazo, la secuencia de posibles resultados y las probabilidades correspondientes son las que se ilustran en el diagrama de árbol de la figura 5-4 (los subíndices indican la posición secuencial de los resultados.) Con base en la regla de la mutiplicación para eventos dependientes, la probabilidad de que las dos refacciones seleccionadas estén ambas en buen estado es

$$P(G_1 \text{ y } G_2) = P(G_1)P(G_2|G_1) = \left(\frac{8}{10}\right) \times \left(\frac{7}{9}\right) = \frac{56}{90} = \frac{28}{45}$$



Evento conjunto	Probabilidad
B_1 y B_2	56/90
B_1 y D_2	16/90
D_1 y B_2	16/90
D_1 y D_2	2/90
	90/90 = 1.0

Fig. 5-4

(Nota: en los problemas 5.12*(b)* y 5.13 se amplía la aplicación de esta regla a tres eventos.)

Si la probabilidad de la ocurrencia conjunta de dos eventos está disponible, sin necesidad de utilizar las reglas de la multiplicación como tales, entonces, como una alternativa a las fórmulas 5.8 y 5.9, puede probarse la independencia de dos eventos A y B comparando

$$P(A \text{ y } B) \stackrel{?}{=} P(A)\, P(B) \tag{5.14}$$

EJEMPLO 13. De acuerdo con lo que se sabe de un mazo de 52 cartas, se conoce que sólo hay una carta que sea al mismo tiempo un as (A) y trébol (7) y, por ello, P(A y B) = 1/52. También se sabe que la probabilidad de extraer cualquier as es 4/52 y que la probabilidad de extraer cualquier trébol es 13/52. Así, puede verificarse que los eventos "as" y "trébol" son eventos independientes de la siguiente manera:

$$P(A \text{ y } S) \stackrel{?}{=} P(A)\,P(S)$$

$$\frac{1}{52} \stackrel{?}{=} \frac{4}{52} \times \frac{13}{52}$$

$$\frac{1}{52} = \frac{1}{52} \text{ (por lo tanto los eventos } \textit{son} \text{ independientes)}$$

5.7 TEOREMA DE BAYES

En su forma algebraica más simple, el teorema de Bayes se refiere al cálculo de la probabilidad condicional del evento *A,* dado que ha ocurrido el evento *B.* La forma general del teorema de Bayes es

$$P(A|B) = \frac{P(A \text{ y } B)}{P(B)} \tag{5.15}$$

La fórmula 5.15 es simplemente una forma especifica de la fórmula general para la probabilidad condicional que se presentó en la sección 5.5. Sin embargo, la importancia especial del teorema de Bayes consiste en que se aplica en el contexto de eventos secuenciales y, además, en que la versión de cálculo de la fórmula proporciona la base para determinar la probabilidad condicional de un evento que ha ocurrido en la *primera* posición secuencial, dado que se ha observado un evento específico en la *segunda* posición secuencial. La forma de cálculo para el teorema de Bayes es

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A')P(B|A)'} \tag{5.16}$$

Tal como se ilustra en el problema 5.20 (c), el denominador de la expresión anterior es la probabilidad global (no condicional) del evento en la segunda posición secuencial; *P* (B) para la fórmula anterior.

EJEMPLO 14. Supóngase que existen dos urnas U_1 y U_2. La urna 1 tiene ocho bolas rojas y dos bolas verdes, en tanto que la urna 2 tiene cuatro bolas rojas y seis bolas verdes. Si se elige una urna al azar, y después se selecciona al azar una bola de esa urna, pueden representarse el proceso secuencial y las probabilidades mediante el diagrama de árbol de la figura 5-5. El diagrama de árbol indica que la probabilidad de elegir cualquiera de las urnas es 0.50 y después, las probabilidades condicionales de extraer una bola roja *(R)* o una verde *(V)* son las que se señalan y las que corresponden

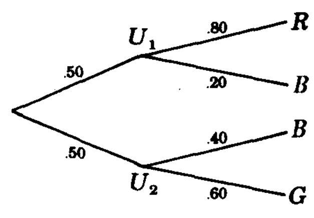


Fig. 5.5

a la urna Implicada. Ahora, supóngase que se observa una bola verde en la etapa 2 *sin* saber qué urna se seleccionó en la etapa 1. ¿Cuál es la probabilidad de que se haya seleccionado la urna 1 en la etapa 1 ? En símbolos, ¿cuál es ***P* (*U*₁/ *V*)**? Sustituyendo U_1 y V para A y B_{51}, respectivamente, en la fórmula de cálculo del teorema de Bayes:

$$P(U_1|B) = \frac{P(U_1)P(G \mid U_1)}{P(U_1)P(B|U_1) + P(U_2)P(B|U_2)} = \frac{0.50)(0.20)}{(0.50)(0.20) + (0.50)(0.60)} = \frac{0.10}{0.40} = 0.25$$

Debe observarse que en el ejemplo 14 el teorema de Bayes ofrece la base para obtener lo que podría denominarse un valor de probabilidad "condicional hacia atrás", puesto que puede determinarse la probabilidad de que se haya seleccionado una urna determinada en la etapa 1, dada la observación de un elemento muestreado de esta urna en la etapa dos. En el análisis bayesiano de decisión este teorema ofrece la base conceptual para revisar las probabilidades asociadas con diversos eventos, o estados implicados en un problema de decisión. Véase la sección 19.2

5.8 TABLAS DE PROBABILIDADES CONJUNTAS

Una *tabla de probabilidades conjuntas* es aquélla en la cual se enlistan como encabezados de renglón todos los posibles eventos (o resultados) para una variable; como encabezados de columnas se enlistan todos los posibles eventos para una segunda variable, y el valor que se anota en cada una de las celdas de la tabla es la probabilidad de su ocurrencia conjunta. Es frecuente que las probabilidades de este tipo de tablas se basen en las frecuencias de ocurrencia observadas para los diversos eventos conjuntos, más que en eventos que son a *priori* por naturaleza. La tabla de frecuencias de ocurrencia conjuntas que puede servir como base para construir una tabla de probabilidades conjuntas se denomina *tabla de contingencias.*

EJEMPLO 15. La Tabla 5.1 *(a) es* una tabla de contingencias que describe a 200 personas que entraron en una tienda de equipos de sonido de acuerdo con sexo y edad, en tanto que la Tabla 5.1 (b), es la tabla correspondiente de probabilidades conjuntas. La frecuencia que se reporta en cada una de las celdas de la tabla de contingencias se convierte en un valor de probabilidad dividiéndolo entre el número total de observaciones, que en este caso es 200.

Tabla 5.1 (a) Tabla de contingencias para los clientes de la tienda de equipos de sonido

	Sexo		
Edad	Hombre	Mujer	Total
Menor de 30	60	50	110
30 y mayor	80	10	90
Total	140	60	200

Tabla 5.1 *(b)* Tabla de probabilidad conjunta para los clientes de la tienda de equipos de sonido

	Sexo		Probabilidad
Edad	Hombre (h)	Mujer (m)	marginal
Menor de 30 (me)	0.30	0.25	0.55
30 y mayor (Ma)	0.40	0.05	0.45
Probabilidad marginal	0.70	0.30	1.00

En el contexto de las tablas de probabilidad conjunta, se denomina *probabilidad marginal* a las probabilidades que son un total marginal de renglón o de columna. Mientras que los valores de probabilidad de las celdas son probabilidad de ocurrencia conjunta, las probabilidades marginales son las probabilidades simples, no condicionales, de eventos específicos.

EJEMPLO 16. La probabilidad de 0.30 que aparece en el renglón 1 y columna 1 de la Tabla 5.1 (b) indica que existe una probabilidad de 0.30 de que una persona de ese grupo de 200, elegida al azar, sea hombre menor de 30 años. La probabilidad marginal de 0.70 de la columna 1 indica que existe una probabilidad de 0.70 de que una persona elegida al azar sea hombre.

Al reconocer que una tabla de probabilidad conjunta incluye también todos los valores de la probabilidad no condicional como totales marginales, puede utilizarse la fórmula 5.10 para determinar cualquier valor de probabilidad condicional.

EJEMPLO 17. Suponga que interesa la probabilidad de que una persona elegida al azar de la Tabla 5.1 (b) sea "menor de 30" (Me) dado que es "hombre" (H). La probabilidad, utilizando la fórmula (5.10), es

$$P(Me|H) + \frac{P(H \text{ y } Me)}{P(H)} = \frac{0.30}{0.70} = \frac{3}{7} = 0.43$$

5.9 PERMUTACIONES

De acuerdo con el enfoque clásico para determinar probabilidades que se presentó en la sección 5.1, el valor de probabilidad se basa en el cociente del número de resultados elementales igualmente probables que son favorables, con respecto al número total de resultados en el espacio muestral. Cuando los problemas son simples, es posible contar en forma directa los resultados elementales. Sin embargo, para problemas más complejos se requiere de los métodos de las permutaciones y combinaciones para determinar el número de posibles resultados elementales.

El número *de permutaciones* de n objetos es el número de formas en los que pueden acomodarse esos objetos en términos de orden:

$$\text{Permutaciones en n objetos} = n\,! = (n) \times (n-1) \times \cdots \times (2) \times (1) \tag{5./7}$$

El símbolo n! se lee *"n* factorial". En problemas de permutaciones y combinaciones, *n* es siempre positiva. Obsérvese también que, en matemáticas 0! = 1.

EJEMPLO 18.... Tres miembros de una organización social se han ofrecido a fungir, en forma voluntaria, como presidente, tesorero y secretario. El número de formas (permutaciones) en que los tres podrían asumir los puestos es:

$$n! = 3! = (3)(2)(1) = 6 \text{ formas}$$

Puede representarse este resultado mediante un diagrama secuencial. Suponga que se designan a las tres personas como *A, B* y C. El número de arreglos posibles, o permutaciones, se presenta en la figura 5-6.

Es común que se desee conocer el número de permutaciones en algún *subgrupo* de *n* objetos, en vez de todos los *n* objetos. Es decir, puede interesar el número de permutaciones de *n* objetos tomados *r* a la vez, en donde *r* es menor que n:

$$_nP_r = \frac{n!}{(n-r)!} \tag{5.18}$$

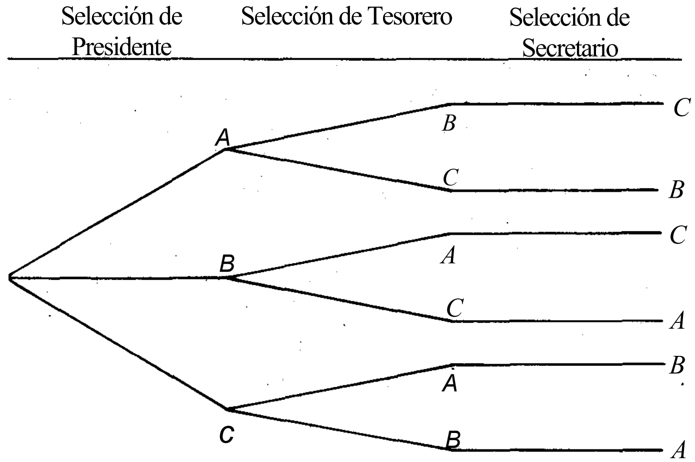


Fig. 5-6

EJEMPLO 19. En el ejemplo 18, suponga que hay 10 miembros de la organización social y que no se han otorgado aún nombramientos para presidente, tesorero y secretario. El número de arreglos diferentes de esos tres funcionarios, elegido de entre los 10 miembros de la organización es:

$$ {}_nP_r = {}_{10}P_3 = \frac{10!}{(10-3)!} = \frac{10!}{7!} = \frac{(10)(9)(8)(7!)}{7!} = (10)(9)(8) = 720 $$

5.10 COMBINACIONES

En el caso de las permutaciones, es importante el orden en el que se acomodan los objetos. En el caso de las *combinaciones,* lo importante es el número de agrupaciones diferentes de objetos que pueden ocurrir *sin* importar su orden. Por lo tanto, cuando se trabaja con combinaciones, siempre se busca el número de subgrupos diferentes que pueden tomarse a partir de *n* objetos. El número de combinaciones de *n* objetos tomados r a la vez es:

$$ {}_nC_r = \frac{n!}{r!(n-r)!} \tag{5.19} $$

En mucbos libros de texto, se representa la combinación de *n* objetos tomando *r* a la vez mediante $\binom{n}{r}$. Obsérvese que *no* es una fracción.

EJEMPLO 20. Suponga que para formar un comité se va a elegir a tres miembros de una organización social pequeña que tiene en total 10 miembros. El número de grupos diferentes de tres personas que podrían elegirse, *sin importar el orden diferente en el que cada uno de los grupos pueda ser conformado, es:*

$$ {}_nC_r = {}_{10}C_3 = \frac{10\,!}{3!(10-3)!} = \frac{(10)(9)(8)(7!)}{3!(7!)} = \frac{(10)(9)(8)}{(3)(2)} = \frac{720}{6} = 120 $$

Tal como se indica en la sección 5.9, los métodos de permutaciones y combinaciones ofrecen una base para contar los resultados posibles en situaciones relativamente complejas. En términos de combinaciones, es frecuente que resulte posible determinar la probabilidad de un evento determinando el número de combinaciones de resultados que incluyen ese evento, en comparación con el número total de combinaciones que son posibles. Por supuesto, esto representa de nuevo el enfoque clásico de la probabilidad, y se basa en la suposición de que todas las combinaciones son igualmente probables.

EJEMPLO 21. Al continuar con el ejemplo 20, si el grupo contiene seis mujeres y cuatro hombres, ¿cuál es la probabilidad de que la elección aleatoria de los miembros del comité dé como resultado una selección que incluya a dos mujeres y un hombre? El método básico consiste en determinar el número de combinaciones de resultados que contienen exactamente dos mujeres (de 6) y un hombre (de 4 en total) y después obtener el cociente de este número con el total de combinaciones posibles:

Número de comités con 2M y 1H = ${}_6C_2 \times {}_4C_1$ (véase la nota que aparece enseguida)

$$\frac{6!}{2!4!} \times \frac{4!}{1!3!} = 15 \times 4 = 60$$

Número total de combinaciones posibles = ${}_{10}C_3$

$$= \frac{10!}{3!7!} = \frac{(10)(9)(8)}{(3)(2)(1)} = \frac{720}{6} = 120$$

$$P(2M \text{ y } 1H) = \frac{{}_6C_2 \times {}_4C_1}{{}_{10}C_3} = \frac{60}{120} = 0.50$$

Nota: en el ejemplo 21 anterior, se utiliza el denominado ***método de la multiplicación.*** En general, si un evento puede ocurrir en n_1 formas y un segundo evento puede ocurrir en n_2 formas, entonces

$$\text{Número total de formas en las que pueden ocurrir dos eventos en combinación} = n_1 \times n_2 \qquad (5.20)$$

El valor de 60 en la solución al ejemplo 21 se basa en la multiplicación del número de formas en las que pueden elegirse dos mujeres de las seis disponibles (15) por el número de formas en que puede escogerse un hombre del total de cuatro disponibles (4).

Problemas resueltos

CÁLCULO DE VALORES DE PROBABILIDAD

5.1 Para cada una de las situaciones siguientes señale cuál seria el enfoque más apropiado (clásico, de frecuencia relativa o subjetivo) para determinar el valor de la probabilidad que se requiere.

(a). La probabilidad de que haya recesión el próximo año

(b). La probabilidad de que aparezca un "uno" o un "seis" en un sólo lanzamiento de un dado de seis lados.

(c). La probabilidad de que, de un envío de 20 refacciones que se sabe que contiene una defectuosa, una de ellas elegida al azar resulte ser defectuosa.

(d). La probabilidad de que una refacción elegida al azar de un envío grande de refacciones resulte defectuosa.

(e). La probabilidad de que una persona elegida al azar de las que entran en una tienda grande de departamentos realice una compra.

(f). La probabilidad de que el índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores aumente en 500 puntos en los próximos seis meses.

(a) Subjetiva, (b) clásica, (c) clásica, (d) de frecuencia relativa (puesto que no existe información respecto a la proporción global de partes defectuosas, se utilizaría la proporción de partes defectuosas en una muestra para estimar el valor de la probabilida), (e) frecuencia relativa, *(f)* subjetiva.

5.2 Determine el valor de probabilidad en cada una de las siguientes situaciones:

(a) La probabilidad de lesiones industriales en una industria específica, en términos anuales. Una muestra aleatoria de empresas que tienen en total 8 000 empleados reportó 400 lesiones de trabajo en un periodo reciente de 12 meses.

(b) La probabilidad de apostar a un número ganador en el juego de la ruleta. Los números de la ruleta incluyen un "0", "00" y de " 1 " hasta "36".

(c) La probabilidad de que un negocio concesionado de alimentos tenga éxito financiero. El prospecto de inversionista obtiene datos para otras unidades del sistema de concesiones, estudia el progreso del área residencial en la que se ubica el expendio y considera el volumen de ventas que se requiere para obtener éxito financiero, con base en la inversión de capital que se requiere y en los costos reales de operación. En términos generales, el inversionista juzga que existe un 80% de probabilidad de que el expendio obtenga éxito financiero y un 20% de probabilidad de que no lo logre.

Respuestas

(a) De acuerdo con el enfoque de frecuencia relativa, $P = 400/8000 = 0.05$. Como este valor de probabilidad se basa en una muestra, es una estimación del valor verdadero que se desconoce. También, se hace la suposición implícita de que los estándares de seguridad no han cambiado con respecto al periodo de los 12 meses en que se muestreó.

(b) De acuerdo con el enfoque clásico, P = 1/38. Este valor se basa en la suposición de que todos los números son igualmente probables, es decir, en que se tiene una ruleta bien balanceada.

(c) Con base en el enfoque subjetivo, el valor al que llega el prospecto de inversionista es P = 0.80. Observe que un juicio como éste debe basarse en toda la información disponible dentro del panorama de tiempo para el que pueda recopilarse ese tipo de información.

5.3 Para cada uno de los cocientes de posibilidad que se reportan en seguida, determine el valor de probabilidad equivalente y, para cada uno de los valores de probabilidad que se reportan, determine el correspondiente cociene de posibilidades.

(a) Un agente de compras estima que las posibilidades de que un envío llegue a tiempo son 2:1.

(b) Se determina que la probabilidad de que un componente nuevo no funcione de manera adecuada cuando se le ensambla es P = 1/5.

(c) Se estima que las probabilidades de que un producto nuevo tenga éxito son 3:1.

(d) Se determina que la probabilidad de que el equipo de casa gane en el juego inaugural es 1/3.

Respuestas

(a) La probabilidad de que el envío llegue a tiempo es $P = 2/(2+1) = 2/3 = 0.67$.

(b) Las posibilidades de que no funcione en forma apropiada son 1:4.

(c) La probabilidad de que el producto tenga éxito es P = 3/(3 + 1) = 3/4 = 0.75.

(d) las posibilidades de que gane el equipo son 1:2

APLICACION DE LAS REGLAS DE ADICION

5.4 Determine la probabilidad de obtener un as *(A),* un rey *(R)* o un dos (D) cuando se extrae un naipe de un mazo bien barajado de 52 cartas.

De la fórmula 5.6,

$$P(A \text{ o } R \text{ o } D) = P(A) + P(R) + P(D) = \frac{4}{52} + \frac{4}{52} + \frac{4}{52} = \frac{12}{52} = \frac{3}{13}$$

(Nota: los eventos son mutuamente excluyentes.)

5.5 Con referencia a la Tabla 5.2, ¿cuál es la probabilidad de que una familia elegida al azar tenga ingresos (a) entre \$18 000 y \$22 999, (o) menores de \$23 000, (c) en alguno de los dos extremos, de ser menores de \$18 000 o cuando menos \$40 000?

Tabla 5.2 Ingresos diarios de 500 familias

Categorías	Rango de ingresos	Número de familias
1	Menos de \$18 000	60
2	\$18 000-\$22 999	100
3	\$23 000-\$29 999	160
4	\$30 000-\$39 999	140
5	\$40 000 y más	40
		Total 500

(a) $P(2) = \frac{100}{500} = \frac{1}{5} = 0.20$

(b) $P(1 \text{ o } 2) = \frac{60}{500} + \frac{100}{500} = \frac{160}{500} = \frac{8}{25} = 0.32$

(c) $P(1 \text{ o } 5) = \frac{60}{500} + \frac{40}{500} = \frac{100}{500} = \frac{1}{5} = 0.20$

(Nota: los eventos son mutuamente excluyentes.)

5.6 De 300 estudiantes de negocios, 100 están inscritos en contabilidad y 80 en estadística para negocios. Estas cifras de inscripción incluyen 30 estudiantes que, de hecho, están inscritos en ambos cursos. ¿Cuál es la probabilidad deque un estudiante elegido al azar esté inscrito únicamente en contabilidad (C) o únicamente en estadística de negocios (E)?

$$P(C \text{ O } E) = P(C) + P(E) - P(C \text{ y } E) = \frac{100}{300} + \frac{80}{300} - \frac{30}{300} = \frac{150}{300} = \frac{1}{2} = 0.50$$

(a) Construya un diagrama de Venn para ilustrar estos eventos.

(b) ¿Cuál es la probabilidad de que un solicitante elegido al azar tenga experiencia de trabajo o certificado (o ambos)?

(c) ¿Cuál es la probabilidad de que un solicitante elegido al azar tenga experiencia de trabajo o certificado, *pero no ambos?*

(a) Véase la figura 5-7

(b) $P(T \text{ o } C) = P(T) + P(C) - P(T \text{ y } C) = 0.40 + 0.30 - 0.20 = 0.50$

(Nota: Los eventos son de mutualidad exclusiva.)

(c) $P(T o C$, pero no ambos$) = P(T \text{ o } C) - P(T \text{ y } C) = 0.50 - 0.20 = 0.30$

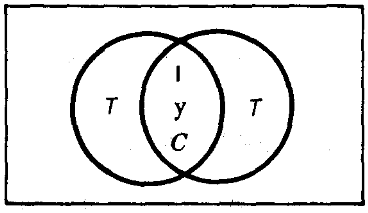


Fig. 5-7

EVENTOS INDEPENDIENTES, EVENTOS DEPENDIENTES Y PROBABILIDAD CONDICIONAL

5.8 Para el problema 5.7 (a) determine la probabilidad condicional de que un solicitante elegido al azar tenga un certificado, dado que tiene alguna experiencia de trabajo, *(b)* Aplique una prueba apropiada para determinar si la experiencia de trabajo y el certificado son eventos independientes.

(a) $P(C|T) = \dfrac{P(C \text{ y } T)}{P(T)} = \dfrac{0.20}{0.40} = 0.50$

(b) $P(C \mid T) \stackrel{?}{=} P(C)$. los eventos *T y C* son dependientes, también se podría probar la independencia aplicando la regla de la multiplicación para eventos independientes (véase el problema 5.14 (a).

5.9 Dos divisiones de productos distintos de una empresa grande son productos marinos *(M)* y equipos de oficina (O) Se estima que la probabilidad de que productos marinos tenga un margen de utilidad de cuando menos 10% en este año fiscal es de 0.30, la probabilidad de que la división de equipo de oficina tenga un margen de utilidad de cuando menos 10% es 0.20 y la probabilidad de que ambas divisiones tengan un margen de utilidad de cuando menos 10% es 0.06.

(a) Determine la probabilidad de que la división de. equipo de oficina tenga un margen de utilidad de cuando menos 10% dado que la División de productos marinos alcanza ese criterio de utilidad.

(b) Aplique una prueba apropiada para determinar si el logro de la meta de utilidades de las dos divisiones es estadísticamente independiente.

(a) $P(O|M) = \dfrac{P(O \text{ y } M)}{P(M)} = \dfrac{0.06}{0.30} = 0.20$

(b) $P(O \mid M) \stackrel{?}{=} P(O)$.

Como 0.20 ='0.20, los dos eventos son independientes. También se podría probar la independencia utilizando la regla de la multiplicación para eventos independientes [véase el problema 5.14 (b)]•

5.10 Suponga que un estudiante optimista estima que la probabilidad de obtener una calificación final de "10" en el curso de estadística de negocios es 0.60 y que la probabilidad de obtener un "8" es 0.40. Por supuesto, no puede obtener arabas calificaciones en forma final, puesto que son mutuamente excluyentes.

(a) Determine la probabilidad condicional de que obtenga un "8" dado que de hecho ha recibido la calificación final de "10", utilizando la fórmula de cálculo apropiada.

(b) Aplique una prueba apropiada para demostrar que esos eventos mutuamente excluyentes son dependientes.

Respuestas

(a) $P(8 \mid 10) = \dfrac{P(8 \text{ y } 10)}{P(10)} = \dfrac{0}{0.60} = 0$

(b) $P(8 \mid 10)$? $P(8)$. Como ($\neq$ 0.40, los eventos son dependientes. Véase la sección 5.5.

APLICACIÓN DE LAS REGLAS DE LA MULTIPLICACIÓN

5.11 En general, la probabilidad de que algún prospecto realice la compra cuando lo visita un vendedor es $P = 0.40$. Si un vendedor elige tres prospectos al azar en un archivo y los visita, ¿cuál es la probabilidad de que la totalidad de los tres prospectos realice una compra?

Como se supone que las acciones de los prospectos son independientes entre sí, se aplica la regla de la multiplicación para eventos independientes.

P(los 3 son compradores) = P(el primero es comprador) x P(el segundo es comprador) x P(el tercero es comprador)
= (0.40) x (0.40) x (0.40) = 0.064

5.12 De 12 cuentas que se tienen en un archivo, cuatro contienen un error de procedimiento en la elaboración de los saldos.

(a) Si un auditor elige al azar dos de las 12 cuentas (sin reemplazo), ¿cuál es la probabilidad de que ninguna de las cuentas contenga error de procedimiento? Construya un diagrama de árbol para presentar este proceso secuencial de muestreo.

(b) Si el auditor muestrea tres cuentas, ¿cuál es la probabilidad de que ninguna de las cuentas incluya el error?

Respuestas

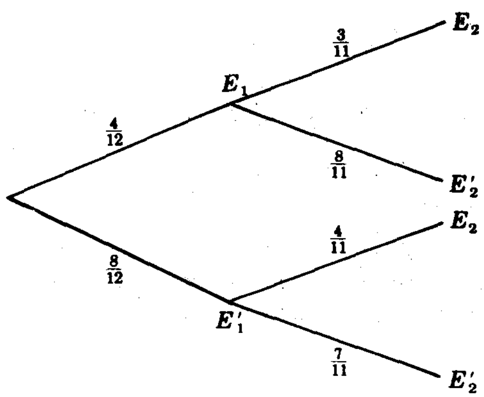
(a) En este caso los eventos son dependientes porque el resultado de la primera cuenta de la muestra afecta las probabilidades que se aplican a la segunda. En donde E_1' significa que no hay error en la primera cuenta que se muestrea y E_2' significa que no hay error en la segunda cuenta

$$P(E'_1 \text{ y } E'_2) = P(E'_1)P(E'_2|E'_1) = \frac{8}{12} \times \frac{7}{11} = \frac{56}{132}\,\frac{14}{33} \cong 0.42$$

En la figura 5-8, E representa una cuenta que tiene error de procedimiento, £' indica una cuenta que no tiene error, y el subíndice señala la posición secuencial de la cuenta muestreada.

(b) $P(E'_1 \text{ y } E'_2 \text{ y } E'_3) = P(E'_1)P(E'_2 \mid E'_1)P(E'_3 \mid E'_1 \text{ y } E'_2)$

$$= \frac{8}{12} \times \frac{7}{11} \times \frac{6}{10} = \frac{336}{1\,320} = \frac{42}{165} \cong 0.25$$



Evento conjunto	Probabilidad
2 con errores	12/132
1 con error	32/132
1 con error	32/132
0 con error	56/132
	132/132

Fig. 5-8

5.13 Cuando se muestrea *sin* reemplazo en un población finita, los valores de probabilidad asociados con los diversos eventos dependen de qué eventos (elementos muestreados) han ocurrido ya. Por otro lado, cuando se muestrea *con* reemplazo, los eventos son siempre independientes.

(a) Suponga que se eligen al azar 3 naipes, sin reemplazo, de un mazo de 52 cartas, ¿cuál es la probabilidad de que las 3 cartas sean ases?

(b) Suponga que se eligen al azar tres cartas de un mazo de 52 naipes, pero que, después de cada selección, se reemplaza el naipe y se revuelve el mazo antes de hacer la siguiente selección. ¿Cuál es la probabilidad de que las tres cartas sean ases?

Respuestas

(a) En este caso, se aplica la regla de la multiplicación para eventos dependientes.

$$P(A_1 \text{ y } A_2 \text{ y } A_3) = P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \text{ y } A_2)$$

$$= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} = \frac{24}{132\,600} = \frac{1}{5\,525} \cong 0.0002$$

(b) En este caso se aplica la regla de la multiplicación para eventos independientes.

$$P(A_1 \text{ y } A_2 \text{ y } A_3) = P(A)\, P(A)\, P(A)$$

$$= \frac{4}{52} \times \frac{4}{52} \times \frac{4}{52} = \frac{64}{140\,608} = \frac{1}{2\,197} \cong 0.0005$$

5.14 Pruebe la independencia (a) de los dos eventos que se describieron en los problemas 5.7 y 5.8, y *(b)* para los dos eventos que se describieron en el problema 5.9, utilizando la regla de la multiplicación para eventos independientes.

(a) $P(T \text{ y } C) \stackrel{?}{=} P(T)\, P(C)$

$0.20 \stackrel{?}{=} (0.40) \times (0.30)$

$0.20 \neq 0.12$

Por lo tanto, los eventos *T* y C son dependientes. Esto concuerda con la respuesta al problema 5.8 *(b)* *(b)* $P(M \text{ y } O) \stackrel{?}{=} P(M)\ P(O)$

$$0.06 \stackrel{?}{=} (0.30) \times (0.20)$$

$$0.06 = 0.06$$

Por lo tanto, los eventos *M* y O son independientes. Esto concuerda con la respuesta al problema 5.9 (b)

5.15 Con los datos del problema 5.7, ¿cuál es la probabilidad de que un solicitante elegido al azar no tenga ni experiencia de trabajo ni certificado? ¿Son independientes estos eventos?

En forma simbólica, lo que se requiere es $P(T' \text{ y } C')$ para que estos eventos no sean mutuamente excluyentes sino eventos dependientes. Sin embargo, en este caso no se tienen disponibles ni $P(T' C')$ ni $P(C' T')$ y, por lo tanto, no puede utilizarse la regla de la multiplicación para eventos dependientes. En este caso, puede obtenerse la respuesta mediante substracción, de la siguiente manera:

$$P(T' \text{ y } C') = 1 - P(T \text{ o } C) = 1 - 0.50 = 0.50$$

(La forma más fácil de comprender la lógica de este planteamiento es dibujando un diagrama de Venn.)

Puede ahora también demostrarse que los eventos son dependientes y no independientes:

$$P(T' \text{ y } C') = P(T')P(C')$$

$$0.50 \stackrel{?}{=} [1 - P(T)]\,[1 - P(C)]$$

$$0.50 \stackrel{?}{=} (0.60)\,(0.70)$$

$$0.50 \neq 0.42$$

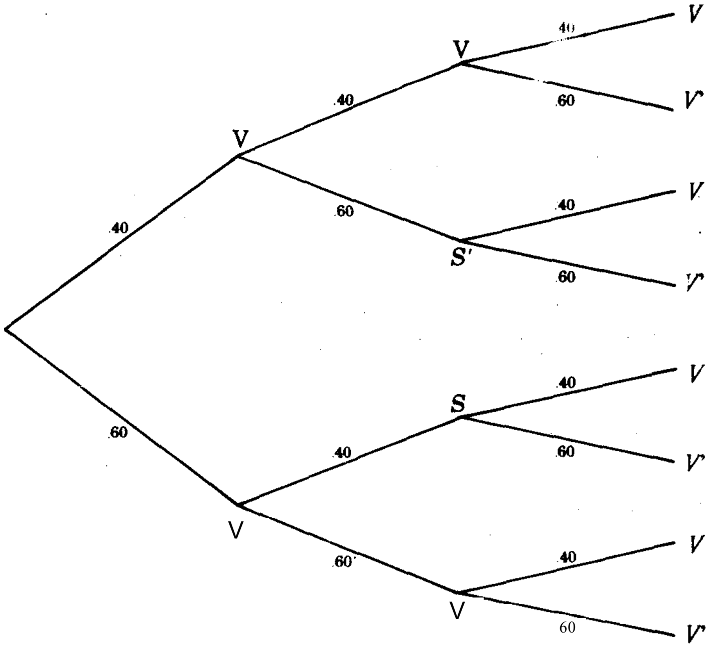
La conclusión de que los eventos son dependientes coincide con la respuesta al problema 5.14 (a), en el cual se trabajan los complementos de cada uno de estos eventos.

5.16 Con referencia al problema 5.11, *(a)* construya un diagrama de árbol para ilustrar la secuencia de las tres visitas, utilizando V para representar una venta y V para representar la no realización de una venta, *(b)* ¿Cuál es la probabilidad de que el vendedor logre *cuando menos* dos ventas? *(c)* ¿Cuál es la probabilidad de que el vendedor realice *cuando menos* una venta?

(a) Véase la figura 5-9

(b) "Cuando menos" dos ventas incluye ya sea dos o tres ventas. Además, haciendo referencia a la figura 5-9, se observa que pueden presentarse dos ventas en cualquiera de tres formas distintas. Por lo tanto, se aplica la regla de la multiplicación para eventos independientes con el objeto de determinar la probabilidad de cada frecuencia, y la regla de la adición para indicar que cualquiera de estas frecuencias constituye un "éxito".

$$\begin{aligned} \text{P(cuando menos 2 ventas)} &= P(V\,y\,V\,y\,V) + (V\,y\,V\,y\,V) \\ &\quad + P(V\,y\,V\,y\,V) + P(V\,y\,V\,y\,V) \\ &= (0.064) + (0.096) + (0.096) + (0.096) = 0.352 \end{aligned}$$



Evento conjuntó	Probabilidad
3 entas	0.064
2 ventas	0.096
2 ventas	0.096
1 venta	0.144
2 ventazs	0.096
1 venta	0.144
1 venta	0.144
0 ventas	0.216
	1.000

Fig. 5-9

(c) En vez de seguir el método de la parte *(b)* es más fácil obtener la respuesta a esta pregunta mediante resta:

$$
\begin{aligned}
P'(\text{cuando menos 1 venta}) &= 1 - P(\text{no } V) \\
&= 1 - P(V' \text{ y } V' \text{ y } V') \\
&= 1 - 0.216 = 0.784
\end{aligned}
$$

5.17 En el problema 5.12 se estableció que 4 de 12 cuentas tenían errores de procedimiento.

(a) Si un auditor muestrea una cuenta al azar, ¿cuál es la probabilidad de que tenga error?

(b) Si un auditor mmuestrea dos cuentas al azar, ¿cuál es la probabilidad de que cuando menos una de ellas tenga error?

(c) Si un auditor muestrea tres cuentas al azar, ¿cuál es la probabilidad de qué cuando menos una contenga error?

(a) $P(E) \quad \dfrac{\text{Número total de cuentas con error}}{\text{Número total de cuentas}} = \dfrac{4}{12} = \dfrac{1}{3} \approx 0.33$

(b) (cuando menos un E) $= P(E_1 \text{ y } E_2) + P(E_1 \text{ y } E_2') + P(E_1' \text{ y } E_2)$

$$
\begin{aligned}
&= P(E_1)\,P(E_2 \mid E_1) + P(E_1)\,P(E_2' \mid E_1) + P(E_1')\,P(E_2 \mid E_1') \\
&= \left(\frac{4}{12}\right)\left(\frac{3}{11}\right) + \left(\frac{4}{12}\right)\left(\frac{8}{11}\right) + \left(\frac{8}{12}\right)\left(\frac{4}{11}\right) \\
&= \frac{12}{132} + \frac{32}{132} + \frac{32}{132} = \frac{76}{132} = \frac{19}{33} \approx 0.58
\end{aligned}
$$

o

$$\begin{aligned}\text{P(cuando menos un E)} &= 1 - P(\text{no } E)\\ &= 1 - P(E_1' \text{ y } E_2')\\ &= 1 - P(E_1')\,P(E_2' \mid E_1')\\ &= 1 - \left(\frac{8}{12}\right)\left(\frac{7}{11}\right)\\ &= 1 - \frac{56}{132} = \frac{76}{132} = \frac{19}{33} \approx 0.58\end{aligned}$$

(c)
$$\begin{aligned}\text{P(cuando menos un E)} &= 1 - P(\text{no } E)\\ &= 1 - P(E_1' \text{ y } E_2' \text{ y } E_3')\\ &= 1 - P(E_1')\,P(E_2' \mid E_1')\;P(E_3' \mid E_1' \text{ y } E_2')\\ &= 1 - \left(\frac{8}{12}\right)\left(\frac{7}{11}\right)\left(\frac{6}{10}\right)\\ &= 1 - \frac{336}{1\,320} = \frac{984}{1\,320} = \frac{123}{165} \approx 0.75\end{aligned}$$

TEOREMA DE BAYES

5.18 Se sabe que la caja A contiene un centavo (C) y una peseta (P), mientras que la caja *B* contiene dos pesetas. Se elige una carta al azar y después se selecciona una moneda, también al azar, de esta caja, (a) Construya un diagrama de árbol para ilustrar esta situación que implica eventos secuenciales. *(b)* Si se elige la caja *A* en la primera etapa, ¿cuál es la probabilidad de que se seleccione una peseta (P) en la segunda etapa? *(c)* Si se selecciona una peseta (P) en la segunda etapa, ¿cuál es la probabilidad de que haya sido extraída de la caja A? *(d)* Si se selecciona un centavo (en la segunda etapa), ¿cuál es la probabilidad de que haya sido extraído de la caja A?

(a) Véase la figura 5-10,

(b) $P(D|A) = \frac{1}{2} = 0.50$

(c) $$P(A|P) = \frac{P(A \text{ y } P)}{P(P)} = \frac{P(A)P(D|A)}{P(A)P(P|A)+P(B)P(P|B)}$$

$$= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)(1)} = \frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{2}} = \frac{1}{3} \approx 0.33$$

(d) $$P(A \mid P) = \frac{P(A y C)}{P(C)} = \frac{P(A)\,P(C \mid A)}{P(A)\,P(C \mid A) + P(B)\,P(C \mid B)}$$

$$= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)(0)} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1$$

Por ello, si se obtiene un centavo, debe haber sido extraído de la caja A.

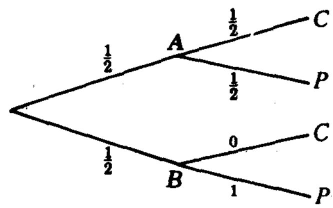


Fig. 5-10

5.19 Un analista de una empresa manufacturera estima que la probabilidad de que una empresa competidora tenga planes para comenzar a fabricar equipo nuevo en los próximos tres años es de 0.30 y de 0.70 la de que la empresa no tenga tales planes. Si la empresa de la competencia sí tiene esos planes, definitivamente se construiría una nueva instalación fabril. Si la empresa de la competencia no tiene esos planes, existe aún una probabilidad de 60% de que se construya la nueva instalación fabril por otras razones.

(a) Al utilizar *E* para la decisión de participar en el campo del equipo nuevo y *F* para la adición de una nueva instalación fabril, ilustre los eventos posibles mediante un diagrama de árbol.

(b) Suponga que se observa que la empresa de la competencia, de hecho, ha comenzado a trabajar en la nueva fábrica. Con esta información, ¿cuál es la probabilidad de que la empresa haya decidido ingresar al campo del nuevo equipo?

Respuestas

(a) Véase la figura 5-11

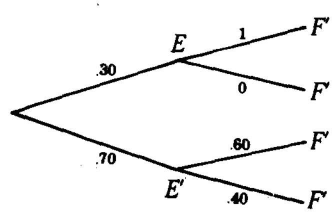


Fig. 5-11

(b) $$P(E \mid F) = \frac{P(E \text{ y } F)}{P(F)} = \frac{P(E)\,P(F \mid E)}{P(E)\,F(F \mid E) + P(E')\,P(F \mid E')}$$

$$= \frac{(0.30)(1)}{(0.30)(1) + (0.70)(0.60)} = \frac{0.30}{0.72} \approx 0.42$$

5.20 Si hay un aumento en las inversiones de capital el siguiente año, la probabilidad de que aumente el precio del acero estructural es de 0.90. Si no hay aumentos en esa clase de inversiones, la probabilidad de aumento es de 0.40. En forma global, se estima que existe una probabilidad del 60% de que aumenten el siguiente año las inversiones de capital.

(a) Al utilizar / e /' para Indicar que se dan aumentos y que no se dan aumentos en las inversiones de capital, y utilizando A y A' para representar los aumentos y los no aumentos en los precios del acero estructural, construya un diagrama de árbol para esta situación que implica eventos dependientes.

(b) ¿Cuál es la probabilidad de que no aumenten los precios del acero estructural aun cuando haya un aumento en las inversiones de capital?

(c) ¿Cuál es la probabilidad global (no condicional) de un aumento en el precio del acero estructural del siguiente año?

(d) Suponga que, de hecho, aumentan los precios del acero estructural en el siguiente año. ¿Cuál es la probabilidad de que haya habido un aumento en las inversiones de capital?

(a) Véase Fig. 5.12.

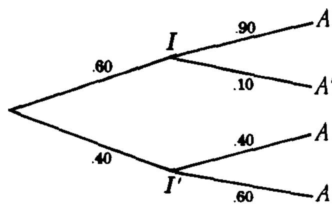


Fig. 5-12

(b) $P(A' \mid I) = 0.10$

(c) Éste es el denominador de la fórmula de Bayes:

$$P(A) = P(I \text{ y } A) \text{ o } P(I' \text{ y } A) = P(I)\,P(A \mid I) + P(I')\,P(A \mid I')$$
$$= (0.60)(0.90) + (0.40)(0.40) = 0.70$$

(d) Mediante la fórmula de Bayes

$$P(I \mid A) = \frac{P(I \text{ y } A)}{P(A)} = \frac{P(I)\,P(A \mid I)}{P(I)\,P(A \mid I) + P(I')\,P(R \mid I')}$$

$$= \frac{(0.60)(0.90)}{(0.60)(0.90) + (0.40)(0.40)} = \frac{054}{0.70} \cong 0.77$$

TABLAS DE PROBABILIDADES CONJUNTAS

5.21 La Tabla 5.3 es una tabla de contingencias que presenta las reacciones de los votantes con respecto a un nuevo plan de impuestos sobre la propiedad, de acuerdo con su afiliación partidaria, *(a)* Prepare la tabla de probabilidades conjuntas para estos datos, *(b)* Determine las probabilidades marginales e indique qué significan.

Tabla 5.3 Tabla de contingencias para las reacciones de los votantes ante un nuevo plan de impuestos sobre la propiedad

Afiliación partidaria	Reacción			Total
	A favor	Neutral	Se opone	
PAN	120	20	20	160
PRI	50	30	60	140
Los demás	50	10	40	100
Total	220	60	120	400

(a) Véase la Tabla 5.4).

Tabla 5.4 Tabla de probabilidad conjunta para las reacciones de los votantes ante un nuevo plan de impuestos sobre la propiedad

Afiliación partidaria	Reacción			Probabilidad marginal
	A favor *(F)*	Neutral (N)	Se opone *(O)*	
PAN *(A)*	0.30	0.05	0.05	0.40
PRI(R)	0.125	0.075	0.15	0.35
Los demás (D)	0.125	0.025	0.10	0.25
Probabilidad marginal	0.55	0.15	0.30	1.00

(b) Cada valor de probabilidad marginal señala la probabilidad no condicional del evento identificado según el encabezado de columna o renglón. Por ejemplo, si se elige al azar a una persona en este grupo de 400 votantes, la probabilidad de que esté a favor del plan de impuestos es *P (F)* = 0.55. Si se elige un votante al azar, la probabilidad de que pertenezca al PRI es *P (P)* - 0.35.

5.22 Con referencia a la Tabla 5.4, determine las siguientes probabilidades: (a)P(O), *(b)P(R* y O), (c) *P(D), (d)P(D* y *F), (e)* $P(O|R)$, *(f)* $P(R|O)$, *(g)* $P(R \text{ o } A)$, *(h)* $P(A \text{ o } F)$.

(a) *P(0)* = 0.30 (la probabilidad marginal)

(b) *P(R* y *O)* = 0.15 el conjunto de probabilidades en la tabla)

(c) *P(I)* = 0.25 (la probabilidad marginal)

(d) *P(I* y *F)* = 0.125 (el conjunto de probabilidad en la tabla)

(e) $P(O|R) = \frac{P(O \text{ y } R)}{P(R)} = \frac{0.15}{0.35} = \frac{3}{7} \approx 0.43$ (la probabilidad de que el votante se oponga al plan dado que pertenece al PRI)

(f) $P(R|O) = \frac{P(R \text{ y } O)}{P(O)} = \frac{0.15}{0.30} = 0.50$ (la probabilidad de que el votante sea del PRI, dado que se opone al plan)

P(R o A) = P(R) + P(A) = 0.35 + 0.40 = 0.75 (la probabilidad de que el votante sea o del PAN o del PRI, los cuales son eventos mutuamente excluyentes.)

P(A o F) = P(A) + P(F) - P(A y F) = 0.40 + 0.55 - 0.30 = 0.65 (la probabilidad de que el votante sea o del PAN o que esté a favor de la propuesta, los cuales no son eventos mutuamente excluyentes).

PERMUTACIONES Y COMBINACIONES

5.23 Van a asignarse asientos contiguos en una mesa de banquete a las cinco personas que constituyen la alta administración de una empresa manufacturera pequeña, (a) Determine el número de arreglos distintos de asientos que son posibles para las cinco personas, (b) Suponga que sólo tres de los cinco funcionarios serán invitados a representar a la compañía en el banquete. ¿Cuántos arreglos distintos son posibles en la mesa, considerando que pueden escogerse tres cualesquiera de las cinco personas?

(a) ${}_nP_n = n! = (5)(4)(3)(2)(1) = 120$

(b) $${}_nP_r = \frac{n!}{(n-r)!} = \frac{5!}{(5-3)!} = \frac{(5)(4)(3)(2)(1)}{(2)(1)} = 60$$

5.24 Para el problema 5.23 (b) suponga que no importa el número de arreglos posibles distintos de los asientos, sino que más bien, Interesa el número de agrupaciones diferentes de los tres funcionarios (de entre los cinco) que podrían asistir al banquete. ¿Cuántas agrupaciones distintas existen?

Al utilizar la fórmula (5.19),

$${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{5!}{3!(5-3)!} = \frac{(5)(4)(3)(2)(1)}{(3)(2)(1)(2)(1)} = 10$$

5.25 Un representante de ventas debe visitar seis ciudades en un viaje.

(a) Si existen 10 ciudades en el área geográfica que va a visitar, ¿cuántas agrupaciones distintas de seis ciudades existen que es posible visitar?

(b) Suponga que existen 10 ciudades en el área geográfica que va a visitar y que, además, también importa la secuencia en la que tiene programado hacer esas seis visitas. ¿Cuántas secuencias distintas existen de seis ciudades escogidas de entre el total de 10?

(c) Suponga que se han designado las seis ciudades que se visitarán, pero no se ha definido la secuencia en la que se harán las visitas. ¿Cuántas secuencias son posibles para las seis ciudades designadas?

(a) $${}_nC_r = \binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{10!}{6!(10-6)!} = \frac{(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)}{(6)(5)(4)(3)(2)(1)(4)(3)(2)(1)} = 210$$

$${}_nP_r = \frac{n!}{(n-r)!} = \frac{10!}{(10-6)!} = \frac{(10)(9)(8)(7)(6)(5)(4)(3)(2)(1)}{(4)(3)(2)(1)} = 151{,}200$$

(c) ${}_nP_n = n! = 6! = (6)(5)(4)(3)(2)(1) = 720$

5.26 De las 10 ciudades que se describieron en el problema 5.25, suponga que en realidad seis de ellas son "mercados primarios" para el producto en cuestión, mientras que las otras cuatro constituyen "mercados secundarios". Si el vendedor elige en forma aleatoria las seis ciudades que va a visitar, ¿cuál es la probabilidad de que (a) cuatro de ellas resulten ser mercados primarios y dos de ellas mercados secundarios, y (b) resulte que las seis son mercados primarios?

(a) $P = \dfrac{\text{Número de combinaciones que incluyen cuatro y dos ciudades, respectivamente.}}{\text{Número total dé combinaciones de seis ciudades.}}$

$$= \frac{{}_6C_4 \times {}_4C_2}{{}_{10}C_6} = \frac{\frac{6!}{4!2!}\,\frac{4!}{2!2!}}{\frac{10!}{6!4!}} = \frac{(15)(6)}{210} = \frac{90}{210} = \frac{3}{7} \cong 0.43$$

(b) $$p = \frac{{}_6C_6 \times {}_4C_0}{{}_{10}C_6} = \frac{\frac{6!}{6!0!}\,\frac{4!}{0!4!}}{\frac{10!}{6!4!}} = \frac{(1)(1)}{210} = \frac{1}{210} \cong 0.005$$

Para este problema, puede obtenerse también la respuesta aplicando la regla de la multiplicación para eventos dependientes. La probabilidad de elegir una ciudad que es mercado primario en la primera selección es 6/10. Siguiendo este resultado, la probabilidad de la siguiente selección es 5/9, y asi sucesivamente. Sobre esta base, la probabilidad de que las seis ciudades resulten ser mercados primarios es

5.27 Con respecto al banquete que se describió en el problema 5.23, determine la probabilidad de que el grupo de los tres funcionarios elegidos de entre los cinco incluya (a) un funcionario especifico, (b) dos funcionarios específicos y (c) tres funcionarios específicos.

(a) $P = \dfrac{\text{Número de combinaciones que incluyen al funcionario específico.}}{\text{Número de combinaciones distintas de tres funcionarios}}$

$$= \frac{{}_1C_1 \times {}_4C_2}{{}_5C_3} = \frac{\frac{1!}{1!0!}\,\frac{4!}{2!2!}}{\frac{5!}{3!2!}} = \frac{(1)(6)}{10} = \frac{6}{10} = 0.60$$

En este caso, el valor de la probabilidad es simplemente equivalente a observar que 3/5 de los funcionarios serán elegidos y, por ello, que la probabilidad de que se elija a cualquier persona especifica es de 3/5 o 0.60.

(b) $$P = \frac{{}_2C_2 \times {}_3C_1}{{}_5C_3} = \frac{\frac{2!}{2!0!}\,\frac{3!}{1!2!}}{\frac{5!}{3!2!}} = \frac{(1)(3)}{10} = \frac{3}{10} = 0.30$$

(c) $$P = \frac{{}_3C_3 \times {}_2C_0}{{}_5C_3} = \frac{\frac{3!}{3!0!}\,\frac{2!}{0!2!}}{\frac{5!}{3!2!}} = \frac{(1)(1)}{10} = \frac{1}{10} = 0.10$$

Problemas complementarios

CALCULO DE LOS VALORES DE LA PROBABILIDAD

5.28 Determine el valor de la probabilidad para cada uno de los eventos siguientes:

(a) La probabilidad de elegir al azar una cuenta por cobrar morosa, dado que 5% de las cuentas son morosas.

(b) La probabilidad de que una inversión en bienes raíces tenga éxito. En el área que se evalúa por lo general sólo la mitad de esas inversiones tienen éxito, pero los métodos de decisión de los inversionistas específicos han dado como resultado que se tenga un historial 30% mejor que para el inversionista promedio en la región.

(c) La probabilidad de que la suma de los puntos que aparecen en la cara superior de 2 dados lanzados al azar sea 7.

Resp. (a) 0.05, (b) 0.65, (c) 1/6

5.29 Para cada una de las siguientes razones de posibilidades determine el valor equivalente de probabilidad, y para cada uno de los valores de probabilidad determine la razón de posibilidad equivalente

(a) Una probabilidad de $P = 2/3$ de que se satisfaga una fecha objetivo de entrega.

(b) Una probabilidad de $P = 9/10$ de que un producto nuevo supere ei nivel de ventas de punto de equilibrio.

(c) Posibilidades de 1:2 de que un competidor logre un progreso tecnológico.

(d) Una posibilidad de 5:1 de que un producto nuevo sea redituable.

Resp. (a) 2:1 *(b)* 9:1, (c) $P = 1/3$, *(d)* $P = 5/6$

APLICACIÓN DE LAS REGLAS DE ADICIÓN

5.30 Durante una semana determinada, se estima que la probabilidad de que el precio de una acción específica aumente (A), permanezca sin cambios (S) o se reduzca *(R)* es de 0.30,0.20 y 0.50, respectivamente.

(a) ¿Cuál es la probabilidad de que el precio de ia acción aumente o permanezca sin cambios?

(b) ¿Cuál es la probabilidad de que el precio de la acción cambie durante la semana?

Resp. (a) 0.50, *(b)* 0.80

5.31 De 500 empleados, 200 participan en un plan de reparto de utilidades de la compañía (P), 400 tienen una cobertura de gastos médicos mayores *(M)* y 200 empleados participan en ambos programas. Construya un diagrama de Venn para ilustrar los eventos designados con P y *M.*

5.32 Con referencia al diagrama de Venn que se presentó en el problema 5.31, ¿cuál es la probabilidad de que un empleado elegido al azar (a) participe en cuando menos uno de los dos programas, (b) no participe en ninguno de los programas?

Resp. (a) 0.80, (b) 0.20

5.33 Se estima que la probabilidad de que un nuevo método de comercialización tenga éxito (£) es 0.60. La probabilidad de que los gastos para el desarrollo del método puedan mantenerse dentro del presupuesto original (P) es 0.50. Se estima que la probabilidad de alcanzar ambos objetivos es de 0.30. ¿Cuál es la probabilidad de que se logre cuando menos uno de los objetivos?

Resp. 0.80

EVENTOS INDEPENDIENTES, EVENTOS DEPENDIENTES Y PROBABILIDAD CONDICIONAL

5.34 Para la situación que se describió en el problema 5.31, (a) determine la probabilidad de que un empleado participe en el plan de reparto de utilidades (P) considerando que tiene seguro de gastos médicos mayores *(M),* y (b) determine si los dos eventos son independientes o dependientes, haciendo referencia al valor de la probabilidad condicional.

Resp. (a) 0.50, *(b)* dependientes

5.35 Para el problema 5.33, determine (a) la probabilidad de que el nuevo método de comercialización tenga éxito (E), dado que el costo de su desarrollo se ha mantenido dentro del presupuesto original (P), y (b) si los dos eventos son independientes o dependientes, de acuerdo al valor de la probabilidad condicional.

Resp. (a) 0.60, *(b)* dependientes

5.36 Se estima que la probabilidad de que aumenten las ventas de automóviles en el siguiente mes *(A)* es de 0.40. Se estima que la probabilidad de que aumenten las ventas de refacciones *(R)* es de 0.50. Sé estima que la probabilidad de que ambas industrias experimenten un aumento en ventas es de 0.10. ¿Cuál es la probabilidad de que (a) hayan aumentado las ventas de automóviles durante el mes, dado que existe información de que han aumentado las ventas de refacciones, (b) hayan aumentado las ventas de refacciones, dado que existe información de que aumentaron las ventas de automóviles durante ese mes?

Resp. (a) 0.20, *(b)* 0.25

5.37 Para el problema 5.36, determine si los dos eventos son independientes o dependientes, de acuerdo con uno de los valores de la probabilidad condicional.

Resp. Dependientes

APLICACIÓN DE LAS REGLAS DE LA MULTIPLICACIÓN

5.38 Durante un periodo determinado, aumentó el valor de mercado de las acciones comunes en circulación en una industria, que incluye solamente 10 compañías. Si un inversionista escoge dos de esas acciones al azar, ¿cuál es la probabilidad de que ambas hayan experimentado un aumento en su valor de mercado durante ese periodo?

Resp. $56/90 \cong 0.62$

5.39 La proporción general de artículos defectuosos en un proceso continuo de producción es 0.10. ¿Cuál es la probabilidad de que (a) de dos artículos elegidos al azar ninguno tenga defectos D'), (b) dos artículos escogidos al azar tengan defectos D), (c) cuando menos uno de los dos artículos escogidos al azar no tenga defectos D')?

Resp. (a) 0.81, (b) 0.01, (c) 0.99

5.40 Pruebe la independencia de los dos eventos que se describieron en el problema 5.31 utilizando la regla de la mutiplicación para eventos independientes. Compare su respuesta con el resultado de la prueba del problema 5.31 (b).

Resp. Dependientes

5.41 Pruebe la independencia de los dos eventos que se describieron en el problema 5.33 utilizando la regla de la multiplicación para eventos independientes. Compare su respuesta con el resultado de la prueba del problema 5.32(b).

Resp. Independientes

5.42 Del problema 5.38, suponga que un inversionista elige al azar tres de las acciones. Construya un diagrama de árbol para ilustrar los diversos resultados posibles para la secuencia de las tres acciones.

5.43 Con referencia al diagrama de árbol que se preparó en el problema 5.42, determine la probabilidad de que (a) sólo una de las tres acciones experimente un aumento en su valor de mercado, (b) aumente el valor de mercado de dos de las acciones, (c) aumente el valor de mercado de *cuando menos* dos acciones.

Resp. (a) $48/720 \cong 0.07$, *(b)* $336/72 \cong 0.47$, (c) $672/72 \cong 0.93$

6.44 Con respecto al problema 5.39, suponga que se elige al azar una muestra de cuatro artículos. Construya un diagrama de árbol para ¡lustrar los diversos resultados posibles en términos de que los artículos estén defectuosos (D) o no tengan defectos (D')

5.45 Con respecto al diagrama de árbol que se preparó en el problema 5.44, determine la probabilidad de que (a) ninguno de los cuatro artículos tenga defectos, (b) exactamente un articulo esté defectuoso, (c) estén defectuosos algunos artículos.

Resp. (a) $0.6561 \cong 0.66$, *(b)* $0.2916 \cong 0.29$, (c) $0.947 \cong 0.95$

TEOREMA DE BAYES

5.46 Suponga que existen dos urnas U_1 y U_2. La U_1 contiene dos bolas rojas y una verde, en tanto que U_2 contiene una bola roja y dos verdes.

(a) Se elige una urna ai azar y después se elige al azar una bola de esa urna. La bola es roja. ¿Cuál es la probabilidad de que la urna seleccionada haya sido la urna U_1?

(b) Se elige una urna al azar y después se seleccionan dos bolas también al azar (sin reemplazo) de esa urna. La primera bola es roja y la segunda verde. ¿Cuál es la probabilidad de que la urna seleccionada haya sido la U_1?

Resp. (a) $P(U_1) = 2/3$, (b)$P(U_1) = 1/2$

5.47 Con referencia al problema 5.46

(a) Suponga que se elige una urna al azar y después se seleccionan también al azar dos bolas (sin reemplazo) de esa urna. Ambas bolas son rojas. ¿Cuál es la probabilidad de que la urna seleccionada haya sido la U_1?

(b) Suponga que se elige al azar una urna y después se extraen de ella dos bolas también al azar, *pero la primera bola seleccionada se devuelve a la urna antes de extraer la segunda.* Ambas bolas son rojas. ¿Cuál es la probabilidad de que la urna seleccionada haya sido la U_1?

Resp. (a) $P(U_1) = 1$,(b) $P(U_1,) = 4/5$

5.48 Ochenta por ciento de material de vinil que se recibe del vendedor A es de calidad excepcional, en tanto que sólo cincuenta por ciento de material del vendedor B es de calidad excepcional. Sin embargo, la capacidad de fabricación del vendedor A es limitada y, por esa razón, sólo cuarenta por ciento del vinil que la empresa adquiere proviene de este vendedor. El sesenta por ciento restante se compra al vendedor B. Se inspecciona un embarque de vinil que acaba de llegar y se encuntra que es de excepcional calidad. ¿Cuál es la probabilidad de que provenga del vendedor A?

Resp. P(A) = 0.52

5.49 Se fabrica gasolina en tres refinerías con niveles diarios de producción de 400 000, 800 000 y 1 200 000 litros, respectivamente. La proporción de la producción que está por debajo de las especificaciones de octanaje en las tres refinerías es 0.03,0.05 y 0.04, respectivamente; se determina que una pipa transportadora de gasolina lleva gasolina que se encuentra por debajo de las especificaciones de octanaje y, por lo tanto, la gasolina se devuelve para su mejor refinación. Determine la probabilidad de que la pipa haya sido llenada en cada una de las tres refinerías (a) sin hacer referencia a la información de que el embarque está por debajo de las especificaciones de octanaje y (b) dado que se tiene la información adicional de que el envío está por debajo de las especificaciones de octanaje.

Resp. (a)

TABLAS DE PROBABILIDAD CONJUNTA

5.50 La Tabla 5.5 es una tabla de contingencias que representa la clasificación de 150 compañías muestreadas de acuerdo con cuatro grupos industriales, y respecto a si su rendimiento sobre la inversión está por encima o por debajo del rendimiento promedio en la muestra de las 150 empresas. Prepare una tabla de probabilidad conjunta con base en esos datos muestrales.

Tabla 5.5 Tabla de contingencias para el rendimiento sobre el capital, según el grupo industrial

Categoría industrial	Rendimiento sobre el capital: Superior al promedio (S)	Inferior al promedio *(D*	Total
I	20	40	60
II	10	10	20
III	20	10	30
IV	25	15	40
Total	75	75	150

5.51 Con referencia a la tabla de probabilidad conjunta que se preparó en el problema 5.50, diga cuáles son las siguientes probabilidades: (a) P(l), *(b) P(II),* (c) P(lll), (d) P(IV).

Resp. (a) 0.40, *(b)* 0.13, (c) 0.20, *(d)* 0.27

5.52 Con referencia a la tabla de probabilidad conjunta que se preparó en el problema 5.50, determine las siguientes probabilidades: (a) P(l y S), (b) P(ll o I), (c) P(S), (d) P(I o II), (e) P(l y II), (f) P(S o I). (g) P(S/I). (h).P(III / *S).*

Resp. (a) 0.13, *(b)* 0.57, (c) 0.50, (d) 0.53, (c) 0, (f) 1.0, *(g)* 0.33, (h) 0.27

PERMUTACIONES Y COMBINACIONES

5.53 Suponga que ocho obreros van a ser asignados a ocho puestos de capacitación fabril. ¿De cuántas formas pueden asignarse los ocho puestos a los ocho obreros?

Resp. 40,320

5.54 Con referencia a la situación que se describió en el problema 5.53, suponga que sólo existen disponibles seis puestos distintos para los ocho obreros calificados. ¿De cuántas formas puede asignarse a seis de las ocho personas que son, para los seis puestos disponibles?

Resp. 20,160

5.55 Con respecto a la situación que se describió en el problema 5.54, suponga que los seis puestos disponibles pueden considerarse equivalentes, o no diferentes, para propósitos prácticos. ¿De cuántas maneras puede asignarse a las seis personas de entre los ocho obreros calificados, para ocupar los seis puestos?

Resp. 28

5.56 Un grupo asignado a un proyecto está formado por dos ingenieros y tres técnicos y debe ser conformado a partir de una planta departamental que incluye a cinco ingenieros y nueve técnicos. ¿Cuántos grupos de proyectos distintos pueden formarse a partir de las 14 personas disponibles?

Resp. 840

5.57 Para la situación de asignación de personal que se describió en el problema 5.56, suponga que se asigna a las cinco personas al azar, de entre las 14 personas disponibles en el departamento, sin importar si es ingeniero o técnico. ¿Cuál es la probabilidad de que el grupo de proyecto incluya (a) exactamente dos ingenieros, (b) ningún ingeniero, (c) ningún técnico?

Resp. (a)$P \approx 0.42$, (b)$P \approx 0.06$, (c)$P \approx 0.0005$

Distribuciones de probabilidad para variables aleatorias discretas: binomial, hipergeométrica y Poisson

6

6.1 DISTRIBUCIONES DE PROBABILIDAD PARA VARIABLES ALEATORIAS

En contraste con un evento, tal como se analizó en el capitulo 5, una *variable aleatoria* es un evento *numérico* cuyo valor se determina mediante un proceso al azar. Cuando se asignan valores de probabilidad a todos los valores numéricos posibles de una variable aleatoria X, ya sea mediante un listado o a través de una función matemática, se obtiene como resultado una *distribución de probabilidad.* La suma de las probabilidades para todos los resultados numéricos posibles debe ser igual a 1.0. Pueden denotarse los valores de probabilidad individuales mediante el símbolo *f(x),* lo cual implica que hay implícita una función matemática; mediante P(x = X), el cual implica que la variable aleatoria puede asumir diversos valores específicos, o simplemente mediante P(X).

Para una *variable aleatoria discreta,* se pueden enlistar todos los valores numéricos posibles de la variable en una tabla con las probabilidades correspondientes. Existen diversas distribuciones estándar de probabilidad que pueden utilizarse como modelos para una amplia gama de variables aleatorias discretas en aplicaciones de negocios. Los modelos estándar que se describen en este capítulo son las distribuciones de probabilidad binomial, hipergeométrica y Poisson.

Para una *variable aleatoria continua* no es posible enlistar todos los posibles valores fraccionarios de la variable y, por lo tanto, las probabilidades que se determinan a través de una función matemática se ilustran en forma gráfica mediante una función de densidad de probabilidad o curva de probabilidad. En el capítulo 7 se describen diversas distribuciones estándar de probabilidad que pueden servir como modelos para variables aleatorias continuas. Véase en la sección 1.4 la explicación sobre la diferencia entre variables continuas y variables discretas.

EJEMPLO 1. En la Tabla 6.1 se muestra el número de camionetas que se han solicitado para renta en una arrendadora de automóviles, en un periodo de 50 días. En la última columna de la Tabla se incluyen las frecuencias observadas en este periodo de 50 días, convertidas en probabilidades. Así, puede observarse que la probabilidad de que se hayan solicitado exactamente siete camionetas en un día elegido al azar en ese periodo es de 0.20, y que la probabilidad de que se hayan solicitado seis o más es de 0.20 + 0.20 + 0.08 - 0.56.

Tabla 6.1 Demanda diaria de arrendamiento de camionetas durante un periodo de 50 días

Demanda posible X	Número de días	Probabilidad $[P(X)]$
3	3	0.06
4	7	0.14
5	12	0.24
6	14	0.28
7	10	0.20
8	4	0.08
	50	1.00

6.2 EL VALOR ESPERADO Y LA VARIANZA DE UNA VARIABLE ALEATORIA DISCRETA

De la misma manera en que se hace para conjuntos de datos muestrales y poblacionales, con frecuencia resulta útil describir una variable aleatoria en términos de su *media* (véase la sección 3.2) y su *varianza* (véase la sección 4.5). La media (a largo plazo) de una variable aleatoria X se denomina valor esperado y se denota mediante E(X) Para una variable aleatoria discreta, resulta ser el promedio ponderado de todos los valores numéricos posibles de la variable, utilizando las probabilidades correspondientes como pesos. Como la suma de los pesos (probabilidades) es 1.0, puede simplificarse la fórmula *3.3,* de manera que el valor esperado de una variable aleatoria discreta es

$$E(X)=\sum XP(X) \tag{6.1}$$

EJEMPLO 2. Con base en los datos de la Tabla 6.1, se presentan en la Tabla 6.2 los cálculos que conducen al valor esperado de la variable aleatoria. El valor esperado es 5.66 camionetas. Observe que el valor esperado de la variable discreta puede ser un valor fraccionario porque representa el valor promedio a largo plazo y no el valor especifico de determinada observación.

Tabla 6.2 Cálculo del valor esperado para la demanda de camionetas

Demanda posible X	Probabilidad $[P(X)]$	Valor ponderado $[XP(X)]$
3	0.06	0.18
4	0.14	0.56
5	0.24	1.20
6	0.28	1.68
7	0.20	1.40
8	0.08	0.64
	1.00	E(X) = 5.66

La varianza de una variable aleatoria X se denota mediante *V(X);* se calcula con respecto a £(X) como la media de la distribución de probabilidad. La forma general de desviaciones para la fórmula de la varianza de una variable aleatoria discreta es

$$V(X)=\sum [X-E(X)]^2P(X) \tag{6.2}$$

La forma abreviada para la fórmula de la varianza de una variable aleatoria discreta, que no requiere el cálculo de las desviaciones con respecto a la media, es

$$\begin{aligned} V(X) &= \sum X^2P(X)-[\sum XP(X)]^2 \\ &= E(X^2)-[E(X)]^2 \end{aligned} \tag{6.3}$$

EJEMPL03. En la Tabla 6.3 se presenta la hoja de trabajo utilizada para el cálculo de la varianza de la demanda de renta de camionetas, utilizando la versión abreviada de la fórmula. Tal como se señala enseguida, el valor de la varianza es de 1.74.

$$V(X) = E(X^2) - [E(X)]^2 = 33.78 - (5.66)^2 = 33.78 - 32.04 = 1.74$$

Tabla 6.3 Hola de trabajo para el cálculo de la varlanza para la demanda de camionetas

Demanda posible X	Probabilidad $[P(X)]$	Valor ponderado $[XP(X)]$	Demanda al cuadrado (X^2)	Valor ponderado al cuadrado $[X^2P(X)]$
3	0.06	0.18	9	0.54
4	0.14	0.56	16	2.24
5	0.24	1.20	25	6.00
6	0.28	1.68	36	10.08
7	0.20	1.40	49	9.80
8	0.08	0.64	64	5.12
		E(X) = 5.66		$E(X^2)$ = 33.78

6.3 LA DISTRIBUCIÓN BINOMIAL

La distribución binomial es una distribución discreta de probabilidad aplicable como modelo a diversas situaciones de toma de decisiones, siempre y cuando pueda suponerse que el proceso de muestreo se ajusta a un proceso Bernoulli. Un *proceso Bernoulll* es un proceso de muestreo en el que:

(1) Sólo son posibles dos resultados mutuamente excluyentes en cada ensayo u observación. Por conveniencia, a estos resultados se les denomina *éxito* y *fracaso.*
(2) Los resultados del conjunto de ensayos u observaciones, constituyen *eventos independientes.*
(3) La probabilidad de éxito, que se denota mediantep, permanece constante de un ensayo a otro. Es decir, el proceso es estacionario.

Puede utilizarse la distribución binomial para determinar la probabilidad de obtener un número determinado de éxitos en un proceso Bernoulli. Se requieren tres valores: el número especifico de éxitos (X), el número de ensayos u observaciones (n) y la probabilidad de éxito en cada uno de los ensayos (p). La fórmula para determinar la probabilidad de un número determinado de éxitos X para una distribución binomial, en donde q = (1-p) es:

$$P(X|n,p) = {}_nC_X p^X q^{n-X}$$

$$= \frac{n!}{X!(n-X)!} p^X q^{n-X} \qquad (6.4)$$

EJEMPLO 4. La probabilidad de que un prospecto de ventas elegido al azar realice una compra es da 0.20. Si un vendedor visita a seis prospectos, la probabilidad de que realice exactamente cuatro ventas se determina de la siguiente manera:

$$P(X=4|n=6, p=0.20) = {}_6C_4(0.20)^4(0.80)^2 = \frac{6!}{4!2!}(0.20)^4(0.80)^2$$

$$= \frac{6\times5\times4\times3\times2}{(4\times3\times2)(2)}(0.0016)(0.64) = 0.01536 \cong 0.015$$

Con frecuencia existe interés en la probabilidad acumulada de *X o más" éxitos o "X o menos" éxitos en n ensayos. En este caso, debe determinarse la probabilidad de cada uno de los resultados incluidos dentro del intervalo designado, y entonces sumar esas probabilidades.

EJEMPLO 5. En relación con el ejemplo 4, la probabilidad de que el vendedor logre cuatro o más ventas se determina de la siguiente manera:

$$P(X \geq 4 | n = 6, p = 0.20) = P(X = 4) + P(X = 5) + P(X = 6)$$
$$= 0.01536 + 0.001536 + 0.000064 = 0.016960 \cong \mathbf{0.017}$$

en donde $P(X = 4) = 0.01536$ (del ejemplo 4)

$$P(X=5)\ {}_6C_5(0.20)^5(0.80)^1 = \frac{6!}{5!\,1!}(0.20)^5(0.80) - 6(0.00032)(0.80) = 0.001536$$

$$P(X=6)\ {}_6C_6(0.20)^6(0.80)^0 = \frac{6!}{6!\,0!}(0.000064)(1) = (1)(0.000064) - 0.000064$$

(Nota: recuérdese que cualquier valor elevado a la potencia 0 es igual a 1.)

Como el uso de la fórmula binomial implica una cantidad considerable de cálculos cuando la muestra es relativamente grande, con frecuencia se utilizan tablas de probabilidades binomiales. Véase el apéndice 2.

EJEMPLO 6. Si la probabilidad de que un prospecto de ventas elegido al azar realice una compra es de 0.20, la probabilidad de que un vendedor que visita a 15 prospectos realice menos de tres ventas es

$$P(X<3 | n = 15, p = 0.20) = P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0.0352 + 0.1319 + 0.2309 \text{ (del apéndice 2)}$$
$$= 0.3980 \cong 0.40$$

Los valores de p que se incluyen en el apéndice 2 no exceden de p = 0.50. Si el valor de p en un problema especifico excede de 0.50, debe replantearse el problema para definir el evento en términos del número de "fracasos", en vez de utilizar el número de éxitos (véase el problema 6.9).

Pueden determinarse el valor esperado (media) y la varianza de una distribución binomial determinada enlistando la distribución en una tabla y aplicando las fórmulas que se presentaron en la sección 6.2. Sin embargo, es posible calcular en forma directa el número esperado de éxitos:

$$E(X) = np \qquad \textbf{(6.5)}$$

En donde q = (1 - p), la varianza del número de éxitos puede calcularse también en forma directa:

$$V(X) = npq \qquad \textbf{(6.6)}$$

EJEMPLO 7. Para el ejemplo 6, el número esperado de ventas (como promedio a largo plazo) y la varianza asociadas con la realización de visitas a 15 prospectos son:

$$E(X) = np = 15(0.20) = 3.0 \text{ ventas}$$
$$V(X) = np(q) = 15(0.20)(0.80) = 2.4$$

6.4 LA DISTRIBUCIÓN BINOMIAL EXPRESADA MEDIANTE PROPORCIONES

En vez de expresar la variable aleatoria binomial como el número de éxitos *X*, puede designarse en términos de *proporción*' de éxitos $\hat{p}$ que es el cociente del número de éxitos entre el número de ensayos:

(6.7)

En esos casos, se modifica la fórmula *(6.4)* sólo con respecto a la definición de la proporción. Por ello, la probabilidad de observar exactamente una proporción $\hat{p}$ de éxitos en *n* ensayos de Bernoulli es:

$$P\left(\hat{p}=\frac{X}{n}|n, p\right) = {}_nC_X p^X q^{n-X} \qquad (6.8)$$

o

$$P\left(\hat{p}=\frac{X}{n}|n, \pi\right) = {}_nC_X \pi^X (1-\pi)^{n-X} \qquad (6.9)$$

En la fórmula *(6.9)*, π (la letra griega "pi") *es* el equivalente de p, excepto en que indica en forma especifica que la' probabilidad de éxito en un ensayo individual es un parámetro poblacional.

EJEMPLO 8. La probabilidad de que un empleado elegido al azar esté participando en un programa de inversión en acciones de la compañía es 0.40. Si se eligen al a?ar cinco empleados, la probabilidad de que la proporción de participantes: sea exactamente 0.60, o tres de los cinco empleados muestreados es:

$$P(\hat{p}=0.60)=P(\hat{p}=\frac{3}{5}|n=5, p=0.40)={}_5C_3(0.40)^3(0.60)^2=\frac{5!}{3!2!}(0.064)(0.36)=0.2304\cong 0.23$$

Cuando la variable binomial se expresa en forma de proporción, la distribución sigue siendo discreta y no continua. Sólo pueden ocurrir las proporciones para las cuales el número de éxitos X es un número entero. Por lo tanto, en el ejemplo 8 no es posible que exista una participación de 0.50 participantes en una muestra de cinco. El uso de la tabla binomial cuando se utilizan proporciones simplemente requiere que se convierta la proporción designada $\hat{p}$ al número de éxitos X, para el tamaño dado de la muestra *n.*

EJEMPLO 9. La probabilidad de que un empleado elegido al azar esté participando en un programa de inversión en acciones es 0.40. Si se eligen 10 empleados al azar, la probabilidad de que la proporción de participantes en el programa sea de cuando menos 0.70 es

$$P(\hat{p}\geq 0.70)=P(X\geq 7|n=10, p=0.10) = P(X=7)+P(X=8)+P(X=9)+P(X=10)$$
$$= 0.0425+0.0106+0.0016+0.0001=0.0548$$

El valor esperado de una distribución de probabilidad binomial, expresado mediante una proporción, es igual a la proporción de la población, que puede designarse mediante p o π :

$$E(\hat{p})=p \qquad (6.10)$$

o

$$E(\hat{p}) = \pi \tag{6.11}$$

La varianza de la proporción de éxitos para una distribución binomial de probabilidad, en donde q = (1-p), es

$$V(\hat{p}) = \frac{pq}{n} \tag{6.12}$$

o

$$V(\hat{p}) = \frac{\pi(1-\pi)}{n} \tag{6.13}$$

6.5 LA DISTRIBUCIÓN HIPERGEOMÉTRICA

Cuando el muestreo se realiza *sin reemplazo* para cada uno de los elementos que se toman de una población finita de elementos, no se puede aplicar el proceso Bernoulli debido a que existe un cambio sistemático en la probabilidad de éxitos al ir extrayendo elementos de la población. Cuando se utiliza el muestreo sin reemplazo en alguna situación en la que, de no ser por el no reemplazo, se le pudiera calificar como proceso de Bernoulli, la distribución discreta de probabilidad apropiada resulta ser la distribución hipergeométrica.

Si X es el número designado de éxitos, *N* es el número de elementos de la población, *T* es el número total de "éxitos" incluidos en la población y *n* es el número de elementos de la muestra, la fórmula para determinar las probabilidades hipergeométricas es

$$P(X|N, Tn) = \frac{\binom{N-T}{n-X}\binom{T}{X}}{\binom{N}{n}} \tag{6.14}$$

EJEMPLO 10. De seis empleados, tres han estado con la compañía durante cinco o más años, si se eligen cuatro empleados al azar de ese grupo la probabilidad de que exactamente dos de ellos tengan una antigüedad de cinco años o más es:

$$P(X=2|N=6, T=3, n=4) = \frac{\binom{6-3}{4-2}\binom{3}{2}}{\binom{6}{4}} = \frac{\binom{3}{2}\binom{3}{2}}{\binom{6}{4}} = \frac{\frac{3!}{2!1!}\,\frac{3!}{2!1!}}{\frac{6!}{4!2!}} = \frac{(3)(3)}{15} = 0.60$$

Nótese que en el ejemplo 10, el valor que se requiere de la probabilidad se calcula determinando el número de combinaciones diferentes que incluirían a dos empleados con antigüedad suficiente y dos con menor antigüedad como cociente del número total de combinaciones de cuatro empleados, tomados de entre los seis. Por ello, la fórmula hipergeométrica es una aplicación directa de las reglas de análisis combinatorio que se describieron en la sección 5.10.

Cuando la población es grande y la muestra es relativamente pequeña, el hecho de que se realice el muestreo sin reemplazo tiene poco efecto sobre la probabilidad de éxito en cada ensayo. Una regla práctica conveniente consiste en utilizar la distribución blnomial como aproximación a la hipergeométrica cuando $n < 0.05\ N$. Es decir, el tamaño de la muestra debe ser cuando menos del 5% del tamaño de la población. En diferentes textos pueden encontrarse reglas un tanto distintas para determinar los casos en ios que una aproximación como ésta es apropiada.

6.6 LA DISTRIBUCIÓN POISSON

Puede utilizarse la *distribución Poisson* para determinar la probabilidad de que ocurra un número designado de eventos, cuando éstos ocurren en un continuo de tiempo o espacio. A un proceso como éste se le denomina *proceso Poisson;* es similar el proceso Bernoulli (véase la sección 6.3) excepto en que los eventos ocurren en un continuo (por ejemplo, en un intervalo de tiempo) en vez de ocurrir en ensayos u observaciones fijas. Un ejemplo es la entrada de llamadas en un conmutador telefónico. Al igual que en el caso del proceso de Bernoulli, se supone que los eventos son independientes y que el proceso es estacionario.

Sólo se requiere un valor para determinar la probabilidad de que ocurra un número designado de eventos en un proceso de Poisson: el número promedio a largo plazo de eventos para el tiempo o dimensión especifico de interés. Por lo general, esta media se representa mediante λ (la letra griega "lambda") o, es posible, mediante μ. La fórmula para determinar la probabilidad de un número determinado de éxitos *N* en una distribución Poisson es

$$P(X|\lambda) = \frac{\lambda^{X} e^{-\lambda}}{X!} \tag{6.15}$$

Aquí, e es la constante 2.7183 que es la base de los logaritmos naturales, y los valores de $e^{-\lambda}$ pueden obtenerse en el apéndice 3.

EJEMPLO 11. Un departamento de reparación de maquinaría recibe un promedio de cinco solicitudes de servicio por hora. La probabilidad de que se reciban exactamente tres solicitudes en una hora seleccionada al azar es

$$P(X = 3|\lambda = 5.0) = \frac{(5)^3 e^{-5}}{3!} = \frac{(125)(0.00674)}{6} = 0.1404$$

En forma alternativa, puede utilizarse una tabla de probabilidades Poisson. En el apéndice 4 se incluyen las probabilidades de diversos números específicos de éxitos para diversos valores de λ.

EJEMPLO 12. Puede determinarse la respuesta al ejemplo 11 utilizando el apéndice 4 de probabilidades Poisson, de la siguiente manera:

$$P(X = 3|\lambda = 5.0) = 0.1404$$

Cuando lo que interesa es la probabilidad de "X o más" o "X o menos", se aplica la regla de adición para eventos mutuamente excluyente;.

EJEMPLO 13. Si en un departamento de reparación de maquinaria se recibe un promedio de cinco solicitudes de servicio por hora, la probabilidad de que se reciban menos de tres llamadas en una hora elegida al azar se determina de la siguiente manera

$$\begin{aligned} P(X < 3|\lambda = 5.0) = P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= 0.0067 + 0.0337 + 0.0842 = 0.1246 \end{aligned}$$

en donde $P(X = 0 | \lambda = 5.0) = 0.0067$ (del apéndice 4)
$P(X = 1 | \lambda = 5.0) = 0.0337$
$P(X = 2 | \lambda = 5.0) = 0.0842$

Como se supone que un proceso Poisson es estacionario, se concluye que la media del proceso es siempre proporcional a la longitud del continuo de tiempo o espacio. Por lo tanto, si se tiene disponible una media para una longitud de tiempo, puede determinarse la media para cualquier otro periodo de tiempo que se requiera.

Esto es importante porque el valor de X que se utiliza debe aplicarse al periodo de tiempo pertinente.

EJEMPLO 14. En promedio, 12 personas hacen preguntas cada hora a un consultor de decoración en una tienda de telas. La probabilidad de que tres o más personas acudan en un periodo de 10 minutos (1/6 de hora) se determina de la siguiente manera:

$$\text{Promedio por horas} = 12$$

$$\lambda = \text{promedio por 10 minutos} = \frac{12}{6} = 2.0$$

$$P(X \geq 3|\lambda = 2.0) = P(X = 3|\lambda = 2.0) + P(X = 4|\lambda = 2.0) + P(X = 5|\lambda = 2.0) + \cdots$$
$$= 0.1804 + 0.0902 + 0.0361 + 0.0120 + 0.0034 + 0.0009 + 0.0002 = 0.3232$$

en donde $P(X = 3 | \lambda = 2.0) = 0.1804$ (del apéndice 4)
$P(X = 4 | \lambda = 2.0) = 0.0902$
$P(X = 5 | \lambda = 2.0) = 0.0361$
$P(X = 6 | \lambda = 2.0) = 0.0120$
$P(X = 7 | \lambda = 2.0) = 0.0034$
$P(X = 8 | \lambda = 2.0) = 0.0009$
$P(X = 9 | \lambda = 2.0) = 0.0002$

Por definición, el valor esperado (la media a largo plazo) de una distribución de probabilidad Poisson es igual a la media de la distribución.

$$E(X) = \lambda \qquad (6.16)$$

Resulta tambien que la varianza del número de eventos de una distribución de probabilidad Poisson es igual a la media de la distribución:

$$V(X) = \lambda \qquad (6.17)$$

6.7 APROXIMACIÓN DE POISSON A PROBABILIDADES BINOMIALES

Cuando el número de observaciones o ensayos *n,* en un proceso Bernoulli es grande, los cálculos resultan ser bastante laboriosos. Además, no es común que estén disponibles probabilidades tabuladas para valores muy pequeños de *p.* Por fortuna, la distribución Poisson es apropiada como aproximación de las probabilidades binomiales cuando *n* es grande y p o *q* son pequeñas. Una regla conveniente consiste en afirmar que puede realizarse esa aproximación cuando $n > 30$ y, o $np < 5$ o $nq < 5$. Otros textos pueden utilizar reglas un tanto distintas para determinar las ocasiones en que esa aproximación es apropiada.

La media de la distribución de probabilidades Poisson que se utiliza para aproximar probabilidades binomiales es

$$\lambda = np \qquad (6.18)$$

EJEMPLO 15. Se sabe que 1 % de los artículos de un envío grande de transistores de un proveedor tiene defectos. Si se elige una muestra de 30 transistores al azar, puede determinarse la probabilidad de que dos o más de ellos tengan defectos utilizando las probabilidades binomiales del apéndice 2:

$$P(X \geq 2 | n = 30, p = 0.01) = P(X = 2) + P(X = 3) + \cdots = 0.0328 + 0.0031 + 0.0002 = 0.0361$$

En donde $\lambda = np$ = 30 (0.01) = 0.3, la aproximación Poisson del anterior valor de probabilidad es

$$P(X \geq 2 | \lambda = 0.3) = P(X = 2) + P(X = 3) + \cdots = 0.0333 + 0.0033 + 0.0002 = 0.0368$$

Por ello, la diferencia entre la aproximación Poisson y el valor real de probabilidad binomial es solamente 0.0007.

Cuando *n* es grande, pero *np* y *nq* son mayores a 5.0, pueden aproximarse las probabilidades binomiales utilizando la distribución de probabilidad normal (véase la sección 7.4).

En términos generales, la disponibilidad de computadoras ha hecho que la aproximación de probabilidades de un modelo con base en otro sea menos necesario.

6.8 APLICACIONES EN COMPUTADORA

Los paquetes de computación para análisis estadístico incluyen con frecuencia, tablas de probabilidad para las distribuciones de probabilidad discretas comunes que se utilizan como modelos en situaciones de toma de decisiones. Esta disponibilidad es particularmente útil cuando no existen las probabilidades especificas en tablas estándar. En el problema 6.21 se revisa la obtención de una tabla de probabilidades binomiales seleccionadas utilizando una computadora.

Problemas resueltos

VARIABLES ALEATORIAS DISCRETAS

6.1 Se ha determinado que el número de camiones que llegan cada hora a un almacén tiene la distribución de probabilidad que se muestra en la tabla 6.4. Calcule (a) el número esperado de llegadas X por hora, y (b) la varíanza de esta distribución de probabilidad para la variable aleatoria discreta.

Tabla 6.4 Llegada de camiones a un almacén cada hora

Número de camiones (X)	0	1	2	3	4	5	6
Probabilidad [P (X)]	0.05	0.10	0.15	0.25	0.30	0.10	0.05

De la Tabla 6.5,

(a) E(X) = 3.15

(b) $V(X) = E(X^2) - [E(X)]^2 = 12.05 - (3.15)^2 = 12.05 - 9.9225 = 2.1275 = 2.13$

Tabla 6.5 Hoja de trabajo para el cálculo del valor esperado y de la varianza de la llegada de camiones

Número de camiones (X)	Probabilidad $[P(X)]$	Valor ponderado $[XP(X)$	Número al cuadrado (X^2)	Valor ponderado al cuadrado $[X^2P(X)]$
0	0.05	0	0	0
1	0.10	0.10	1	0.10
2	0.15	0.30	4	0.60
3	0.25	0.75	9	2.25
4	0.30	1.20	16	4.80
5	0.10	0.50	25	2.50
6	0.05	0.30	36	1.80
		$E(X) = 3.15$		$E(X^2)= 12.05$

6.2 En la Tabla 6.6, se identifica la probabilidad de que en un sistema de computación "se caiga" el número señalado de periodos por semana, durante la fase inicial de instalación del sistema. Calcule (a) el número esperado de veces por semana que la computadora no está trabajando y (b) la varianza de esta distribución de probabilidad.

Tabla 6.6 Número de periodos por semana en los que no trabaja un sistema nuevo de computación

Número de periodos (X)	4	5	6	7	8	9
Probabilidad $[P(X)]$	0.01	0.08	0.29	0.42	0.14	0.06

Utilizando la Tabla 6.7,

(a) $E(X) = 6.78$

(b) $V(X) = E(X^2) - [E(X)]^2 = 47.00 - (6.78)^2 = 47.00 - 45.9684 = 1.03$

Tabla 6.7 Hoja de trabajo para el cálculo de! valor esperado y de la varianza del mal funcionamiento de una computadora

Número de periodos (X)	Probabilidad $[P(X)]$	Valor ponderado $[XP(X)]$	Número al cuadrado (X^2)	Valor ponderado al cuadrado $[X^2P(X)]$
4	0.01	0.04	16	0.16
5	0.08	0.40	25	2.00
6	0.29	1.74	36	10.44
7	0.42	2.94	49	20.58
8	0.14	1.12	64	8.96
9	0.06	0.54	81	4.86
	1.00	$E(X) = 6.78$		$E(X^2) = 47.00$

6.3 En la Tabla 6.8 se enlistan los posibles resultados asociados con el lanzamiento de dos dados de seis lados y la probabilidad asociada con cada resultado. Se determinaron estas probabilidades utilizando las reglas de la adición y la multiplicación que se revisaron en las secciones 5.4 y 5.6. Por ejemplo, puede obtenerse un "tres* mediante la combinación de un "uno" y un "dos", o mediante la combinación de "dos" y "uno". Cada secuencia tiene una probabilidad de ocurrencia de (1/6) X (1/6) - 1/36, y como las dos secuencias son mutuamente excluyentes, P(X = 3) - 1/36 + 1/36 - 2/36. Determine (a) el número esperado en el lanzamiento de dos dados y (b) la desviación estándar de esta distribución.

Tabla 6.8 Posibles resultados en el lanzamiento de dos dados

Número de los dos dados (X)	2	3	4	5	6	7	8	9	10	11	12
Probabilidad [P(X)]	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

De la Tabla 6.9,

E(X)=7

(a) $V(X) = E(X^2) - [E(X)]^2 = 54.83 - (7)^2 = 54.83 - 49 = 5.83$

(b) $\sigma = \sqrt{V(X)} = \sqrt{5.83} \cong 2.41$

Tabla 6.9 Hoja de trabajo para el calculo del valor esperado y de la varlanza asociados con el lanzamiento de dos dados

Número *(X)*	Probabilidad	Valor ponderado *[XP(X)]*	Número al cuadrado (X^2)	Valor ponderado al cuadrado *[$X^2P(X)$]*
2	1/36	2/36	4	4/36
3	2/36	6/36	9	18/36
4	3/36	12/36	16	18/36
5	4/36	20/36	25	100/36
6	5/36	30/36	36	180/36
7	6/36	42/36	49	294/36
8	5/36	40/36	64	320/36
9	4/36	36/36	81	324/36
10	3/36	30/36	100	300/36
11	2/36	22/36	121	242/36
12	1/36	12/36	144	144/36
	36/36	*E(X)* = 252/36 = 7.0		$E(X^2)$ = 1,974/36 a 54.83

LA DISTRIBUCIÓN BINOMIAL

6.4 Debido a las elevadas tasas de interés, una empresa reporta que el 30% de sus cuentas por cobrar de otras empresas están vencidas. Si un contador toma una muestra aleatoria de cinco de esas cuentas, determine la probabilidad de cada uno de los siguientes eventos, utilizando la formula de la probabilidad binomial: (a) ninguna de las cuentas está vencida, *(b)* exactamente dos cuentas están vencidas, (c) la mayor parte de las cuentas están vencidas, (d) exactamente el 20% de las cuentas están vencidas.

(a) $P(X=0|n=5, p=0.30) = {}_5C_0(0.30)^0(0.70)^5 = \frac{5!}{0!5!}(0.30)^0(0.70)^5 = (1)(1)(0.16807) = 0.16807$

(b) $P(X=2|n=5, p=0.30) = {}_5C_2(0.30)^2(0.70)^3 = \frac{5!}{2!3!}(0.30)^2(0.70)^3 = (10)(0.09)(0.343) = 0.3087$

(c) $P(X \geq 3|n=5, p=0.30) = P(X=3) + P(X=4) + P(X=5) = 0.1323 + 0.02835 + 0.00243 = 0.16308$

en donde $P(X=3) = \frac{5!}{3!2!}(0.30)^3(0.70)^2 = (10)(0.027)(0.49) = 0.1323$

$$P(X=4) = \frac{5!}{4!1!}(0.30)^4(0.70)^1 = (5)(0.0081)(0.70) = 0.02835$$

$$P(X=5) = \frac{5!}{5!0!}(0.30)^5(0.70)^0 = (1)(0.00243)(1) = 0.00243$$

(d) $P\left(\frac{X}{n} = 0.20|n=5, p=0.30\right) = P(X=1|n=5, p=0.30) = {}_5C_1(0.30)^1(0.70)^4 = \frac{5!}{1!4!}(0.30)^1(0.70)^4$

$$= (5)(0.30)(0.2401) = 0.36015$$

6.5 Una empresa de comercialización por correo tiene una circular que produce una tasa de respuestas de 10%. Suponga que se envían por correo 20 de esas circulares en calidad de prueba de mercado, en un área geográfica nueva. Suponiendo que se aplica la tasa de respuesta del 10% en la nueva área, determine las probabilidades de los siguientes eventos utilizando el apéndice 2: (a) nadie responde, (b) exactamente dos personas responden, (c) la mayoría de las personas responde, *(d)* cuando menos el 20% de las personas responde.

(a) $P(X=0|n=20, p=0.10) = 0.1216$

(b) $P(X=2|n=20, p=0.10) = 0.2852$

(c) $P(X=11|n=20, p=0.10) = P(X=11) + P(X=12) + \cdots = 0.0000 \cong 0$

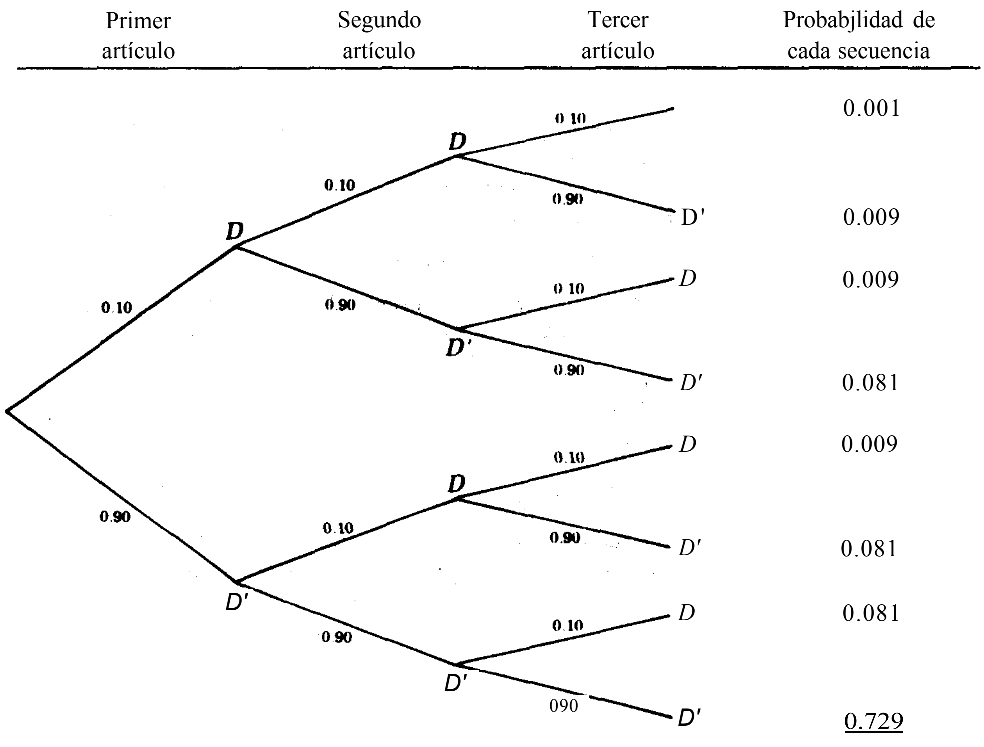
(d) $P\left(\frac{X}{n} < 0.20|n=20, p=0.10\right) = P(X \leq 3|n=20, p=0.10)$

$$= P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$= 0.1216 + 0.2702 + 0.2852 + 0.1901 = 0.8671$$

6.6 Puede considerarse que la fórmula binomial está compuesta de dos partes: una fórmula de combinaciones que determina el número de formas distintas en las que puede ocurrir el evento designado, y la regla de multiplicación para determinar la probabilidad de cada secuencia. Supóngase que se eligen al azar tres artículos de un proceso que se sabe produce 10% de artículos defectuosos. Construya un diagrama de árbol de tres etapas que ilustre la selección de los tres artículos y utilice D para indicar que se selecciona un artículo defectuoso y D' para identificar la selección de un artículo sin defectos. También, anote los valores de probabilidad adecuados en el diagrama y utilice la regla de la multiplicación para eventos independientes con el objeto de determinar la probabilidad de que ocurra cada una de las posibles secuencias de tres eventos.

Véase la figura 6-1,



Flg. 6-1

6.7 Con los datos del problema 6.6, determine la probabilidad de que exactamente uno de los tres artículos muestreados esté defectuoso, haciendo referencia a la figura 6-1 y utilizando la regla de la adición para eventos mutuamente excluyentes.

Al comenzar desde la parte superior del diagrama, las secuencias cuarta, sexta y séptima incluyen exactamente un artículo defectuoso. Por ello,

$$P(X=l) = (D \text{ y } D' \text{y } D') + (D' \text{y } D \text{ y } D') + (D' \text{ y } D' \text{ y } D)$$
$$-0.081 + 0.081 + 0.081 - 0.243$$

6.8 De los problemas 6.6 y 6.7, determine la probabilidad de obtener exactamente un articulo defectuoso utilizando la fórmula binomial, y observe la correspondencia entre los valores obtenidos mediante la fórmula y los que se obtuvieron del diagrama de árbol.

Al utilizar la fórmula *(6.4),*

$$P(X=1|n=3, p=0.10) = {}_3C_1(0.10)^1(0.90)^2 = \frac{3!}{1!2!}(0.10)(0.81)$$
$$= 3(0.081) = 0.243$$

Por ello, la primera parte de la fórmula binomial señala el número de grupos distintos de posiciones que pueden incluir el número designado de éxitos (en este caso existen tres formas en las que puede incluirse un articulo defectuoso en el conjunto de tres). La segunda parte de la fórmula representa la regla de la multiplicación para los eventos independientes especificados.

6.9 En un año especifico el 70% de las acciones que se negociaron en la Bolsa Mexicana de Valores aumentaron de precio, en tanto que el 30% restante permanecieron sin cambios o experimentaron una reducción en su precio. Al principio del año, un asesor de inversiones eligió 10 de las acciones y las calificó como "especialmente recomendables". Si las acciones de estas 10 empresas representan una selección aleatoria, ¿cuál es la probabilidad de que (a) la totalidad de las 10 y (b) cuando menos 10 de las acciones aumenten de valor?

(a) $P(X=10|n=10, p=0.70)=P(X'=0|n=10, q=0.30)=0.0282$

(Atora: cuando p es mayor que 0.50, el problema debe replantearse en términos de X' (léase complemento de X) y se concluye que $X' = n - X$. Por ello, el evento "aumenta el precio de las 10" es igual al de "no se reduce el precio de ninguna").

(b) $$P(X\geq 8|n=10, p=0.70)=P(X'\leq 2|n=10, q=0.30)$$
$$=P(X'=0)+P(X'=1)+P(X'=2)$$
$$=0.0282+0.1211+0.2335=0.3828$$

(Nota: cuando se replantea un enunciado de probabilidad en términos deX', en vez de X, y cuando hay implicita una desigualdad, debe revertirse el símbolo de la desigualdad del enunciado original.)

6.10 Al utilizar el apéndice 10, determine:

(a) $P(X=5|n=9, p=0.50)$
(b) $P(X=7|n=15, p=0.60)$
(c) $P(X\leq 3|n=20, p=0.05)$
(d) $P(X\geq 18|n=20, p=0.90)$
(e) $P(X>8|n=10, p=0.70)$

(a) $P(X=5|n=9, p=0.50)=0.2461$
(b) $P(X=7|n=15, p=0.60)=P(X'=8|n=15, q=0.40)=0.1181$
(c) $$P(X\leq 3|n=20, p=0.05)=P(X=0)+P(X=1)+P(X=2)+P(X=3)$$
$$=0.3585+0.3774+0.1887+0.0596=0.9842$$
(d) $$P(X\geq 18|n=20, p=0.90)=P(X'\leq 2|n=20, q=0.10)$$
$$=P(X'=0)+P(X'=1)+P(X'=2)$$
$$=0.1216+0.2702+0.2852=0.6770$$
(e) $$P(X>8|n=10, p=0.70)=P(X'<2|n=10, q=0.30)$$
$$=P(X'=0)+P(X'=1)=0.0282+0.1211=0.1493$$

6.11 Si se lanza una moneda cinco veces, la distribución de probabilidad con respecto al número de caras que ocurren se basa en la distribución binomial, con n = 5 yp = 0.50 (véase la Tabla 6.10) Determine (a) el número esperado de caras y (b) la varianza de la distribución de probabilidad utilizando las fórmulas *generales* para variables aleatorias discretas.

Tabla 6.10 Distribución binomial de probabilidad del número de caras que ocurren en cinco lanzamientos de una moneda

Número de caras (X)	0	1	2	3	4	5
Probabilidad [$P(X)$]	0.0312	0.1562	0.3125	0.3125	0.1562	0.0312

Al utilizar la Tabla 6.11,

(a) $E(X) = 2.4995 \cong 2.5$

(b) $V(X) = E(X^2) - [E(X)]^2 = 7.4979 - (2.4995)^2 = 7.4979 - 6.2475 = 1.2504 \cong 1.25$

Tabla 6.11 Hoja de trabajo para el cálculo del valor esperado y de la varianza para el problema 6.11

Número de caras (X)	Probabilidad	Valor ponderado [$XP(X)$]	Número al cuadrado (X^2)	Valor ponderado al cuadrado [$X^2P(X)$]
0	0.0312	0	0	0
1	0.1562	0.1562	1	0.1562
2	0.3125	0.6250	4	1.2500
3	0.3125	0.9375	9	2.8125
4	0.1562	0.6248	16	2.4992
5	0.0312	0.1560	25	0.7800
		$E(X) = 2.4995$		$E(X^2) = 7.4979$

6.12 Con referencia al problema 6.11, determine (a) el número esperado de caras y (o) la varianza de la distribución de probabilidad, utilizando las fórmulas *especiales* aplicables para distribuciones binomjales de probabilidad, y (c) compare sus respuestas con las que se obtuvieron en el problema 6.11.

(a) $E(X)=np = 5(0.50) = 2.5$

(b) $V(X) = npq = (5)(0.50)(0.50) = 1.25$

(c) Las respuestas que se obtuvieron con las fórmulas especiales aplicables a las distribuciones binomiales corresponden a las respuestas que se obtuvieron mediante las fórmulas generales más laboriosas que son aplicables para cualquier variable aleatoria discreta.

LA DISTRIBUCIÓN HIPERGEOMÉTRICA

6.13 Un gerente selecciona al azar a n = 3 personas del conjunto de 10 empleados de su departamento, para asignarlos a un estudio de clasificación de sueldos. Suponiendo que anteriormente se asignó a cuatro de los empleados a un proyecto similar, construya un diagrama de árbol de tres etapas que ilustre la selección de las tres personas en términos de si cada uno de ellos tiene experiencia previa E, o no tiene experiencia previa E', en esa clase de estudios. Además, anote los valores correspondientes de probabilidad en el diagrama y utilice la regla de la multiplicación para eventos dependientes con el objeto de determinar la probabilidad de que ocurra cada una de las posibles secuencias de tres eventos.

Véase la figura 6-2.

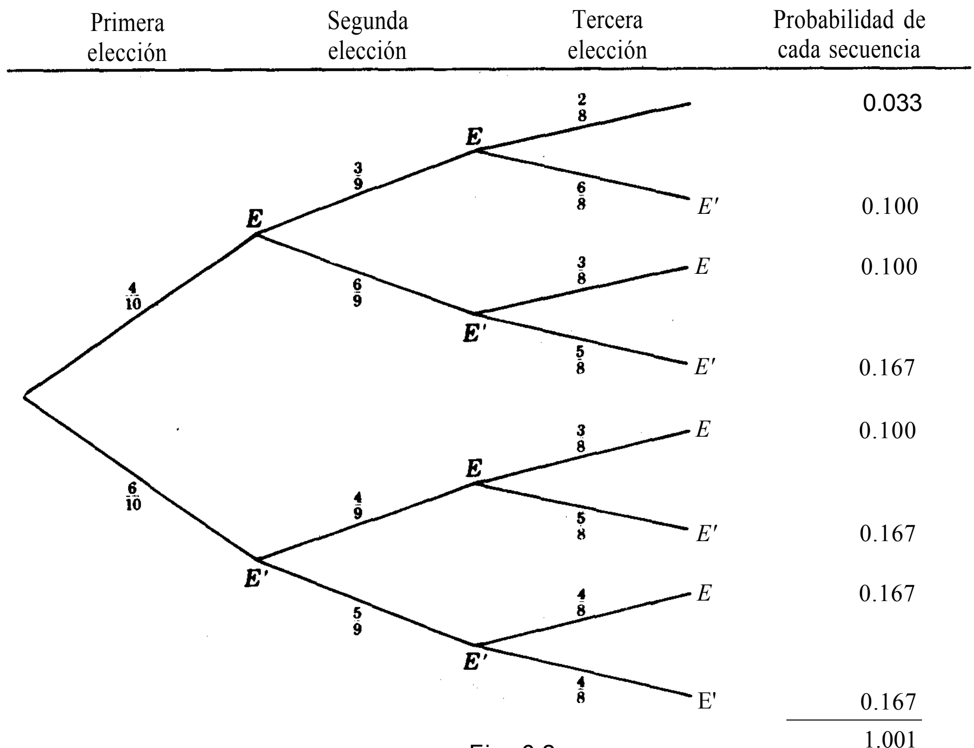


Fig. 6-2

6.14 Con referencia al problema 6.13, determine la probabilidad de que exactamente dos de los tres empleados seleccionados hayan tenido experiencia previa en los estudios de clasificación de sueldos, haciendo referencia a la figura 6-2 y utilice la regla de la adición para eventos mutuamente excluyentes.

Al comenzar de la parte superior del diagrama de árbol, las secuencias segunda, tercera y quinta incluyen exactamente dos empleados con experiencia. Por ello, mediante la regla de la adición para estas secuencias mutuamente excluyentes:

$$
\begin{aligned}
P(X=2) &= (E y E y\ E') + (E y E' y E) + (E' y E y E) \\
&= 0.100 + 0.100 + 0.100 - 0.30
\end{aligned}
$$

6.15 Con referencia al problema 6.13, determine la probabilidad de que exactamente dos de los tres empleados hayan tenido experiencia previa, utilizando la fórmula para determinar las probabilidades hipergeométricas.

De la fórmula *(6.14)*,

$$
P(X|N, T, n) = \frac{\binom{N-T}{n-X}\binom{T}{X}}{\binom{N}{n}}
$$

$$
P(X=2|N=10,\ T=4,\ n=3) = \frac{\binom{10-4}{3-2}\binom{4}{2}}{\binom{10}{3}} = \frac{\binom{6}{1}\binom{4}{2}}{\binom{10}{3}} = \frac{\left(\frac{6!}{1!5!}\right)\left(\frac{4!}{2!2!}\right)}{\left(\frac{10!}{3!7!}\right)} = \frac{(6)(6)}{120} = 0.30
$$

6.16 La sección 6.5 establece que la fórmula hipergeométrica es una aplicación directa-de las reglas del análisis combinatorio que se describieron en le sección 5.9. Para demostrar, esto, aplique la fórmula hipergeométrica al problema 5.26 (a).

$$P(X=4|N=10, T=6, n=6)=\frac{\binom{10-6}{6-4}\binom{6}{4}}{\binom{10}{6}}=\frac{\frac{4!}{2!2!}\,\frac{6!}{4!2!}}{\frac{10!}{6!4!}}$$

$$=\frac{(6)(15)}{210}=\frac{90}{210}=\frac{3}{7}\cong 0.43$$

(Nota: este resultado es equivalente a utilizar la fórmula del análisis combinatorio en la solución para el problema 5.26 (a).)

LA DISTRIBUCIÓN DE PROBABILIDAD POISSON

6.17 En promedio, cada hora cinco personas realizan transacciones en el mostrador de "servicios especiales" de un banco. Suponiendo que la llegada de esas personas tiene una distribución independiente e igualmente probable en todo el periodo de interés, ¿cuál es la probabilidad de que más de 10 personas deseen realizar transacciones en el mostrador de servicios especiales en una hora especifica?

Al utilizar el apéndice 4,

$$P(X>10|\lambda=5.0)=P(X\geq 11|\lambda=5.0)= P(X=11)+P(X=12)+\ldots$$
$$=0.0082+0.0034+0.0013+0.0005+0.0002=0.0136$$

6.18 En promedio, un barco llega a cierto muelle cada dos dias; ¿cuál es la probabilidad de que lleguen dos o más barcos en un día seleccionado al azar?

Como el promedio por dos dfas = 1.0, entonces $\lambda=$ promedio por dia = 1.0 × (1/2) = 0.5. Sustituyendo del apéndice 4,

6.19 En promedio, cada rollo de 500 metros de acero laminado tiene dos defectos. Un defecto es una raspadura o alguna otra irregularidad que afectarla el uso de ese segmento de la hoja de acero en el producto terminado. ¿Cuál es la probabilidad de que un segmento específico de 100 metros no tenga ningún defecto?

Si el promedio por rollo de 500 metros = 2.0, entonces $\lambda=$ promedio por rollo de 100 metros = 2.0 x (100/500) =0.40. Por ello, del apéndice 4.

$$P(X=0|\lambda=0.40)=0.6703$$

6.20 Una compañía de seguros está considerando la, adición de cobertura para una enfermedad relativamente rara en el campo de los seguros médicos mayores. La probabilidad de que una persona elegida al azar tenga esa enfermedad es 0.001, y en el grupo asegurado existen 3 000 personas.

(a) ¿Cuál es el número esperado de personas que tiene la enfermedad en el conjunto?
(b) ¿Cuál es la probabilidad de que ninguna persona de las 3 000 tenga la enfermedad?
(a) La distribución del número de personas qué tendrá la enfermedad seguiría la distribución binomial de probabilidad con n = 3 000 y p = 0.001.

$$E(X)=np=(3{,}000)(0.001)=3.0 \text{ personas}$$

(b) No existen valores tabulados de probabilidades binomiales paran = 3 000 y p = 0.001. Tampoco resulta atractiva la solución algebraica de la fórmula binomial debido a los números grandes implicados. Sin embargo, puede utilizarse la fórmula de la distribución Poisson para aproximar la probabilidad binomial, porque $n \geq 30$ y $np > 5$. Por lo tanto,

$$\lambda = np = (3\,000)(0.001) = 3.0$$

$$P_{\text{Binomial}}(X = 0|n = 3'000, p = 0.001) \cong P_{\text{Poisson}}(X = 0|\lambda = 3.0) = 0.0498$$

(del apéndice 4)

APLICACIONES EN COMPUTADORA

6.21 La probabilidad de que un vendedor realice una venta con un prospecto previamente evaluado, con base en un método especifico de presentación del producto, es 0.33.

(a) Al utilizar programas disponibles de computación, obtenga la tabla de probabilidades binomiales para el número de ventas, considerando que se realizan n = 10 visitas.
(b) Con referencia a la tabla, determine la probabilidad de que se realicen exactamente cinco ventas.
(c) Con referencia a la tabla, determine la probabilidad deque se realicen cinco o más ventas.
(a) En la figura 6-3, se presentan los datos de entrada y los resultados obtenidos en una computadora para la tabla que se requiere de probabilidades binomiales.
(b) Obsérvese que las probabilidades que se reportan en la figura 6-3 son *acumuladas.* Es decir, cada valor de probabilidad es la probabilidad del número correspondiente de ventas, o *menos.* Por lo tanto, la probabilidad de que ocurran justamente cinco ventas es

$$P(X = 5) = P(X < 5) - P(X < 4)$$
$$= 0.9268 - 0.7936 = 0.1332$$

(c) $P(X > 5) = 1.00 - P(X < 4)$
$= 1.00 - 0.7936 = 0.2064$

```
MTB > CDF;
SUB> BINOMIAL FOR N = 10, P = 0.33.

    BINOMIOL WITH N =  10  P = 0.330000
       K  P( X LESS OR = K)
       0          0.0188
       1          0.1080
       2          0.3070
       3          0.5684
       4          0.7936
       5          0.9268
       6          0.9815
       7          0.9988
       8          0.9997
       9          1.0000
```

Fig. 6-3 Resultados de Minitab.

Problemas complementarios

VARIABLES ALEATORIAS DISCRETAS

6.22 Se ha determinado que la llegada de clientes a un restaurante, durante intervalos elegidos al azar de 10 minutos, sigue la distribución de probabilidad que se presenta en la Tabla 6.12. Calcule el número esperado de llegadas para intervalos de 10 minutos, y la varianza de las llegadas.

Tabla 6.12 Llegada de olientes a un establecimiento en intervalos de 10 minutos

Número de clientes (X)	0	1	2	3	4	5
Probabilidad [P (X)]	0.15	0.25	0.25	0.20	0.10	0.05

Resp. *E(X) = 2.00,* V (X) = 1.90

6.23 Se ha determinado que las ventas en expendios de publicaciones de una revista mensual tiene la distribución de probabilidad de la Tabla 6.13. Calcule el valor esperado en la varianza de las ventas de la revista, en miles.

Tabla 6.13 Ventas en expendios de una revista de publicación mensual

Número de revistas (X) en miles	15	16	17	18	19	20
Probabilidad [P (X)]	0.05	0.10	0,25	0.30	0.20	0.10

Resp. *E(X)* = 17.80, *V(X)* = 1.66

6.24 Un vendedor ha determinado que la probabilidad de que realice diversos números de ventas diarias, considerando que visita 10 prospectos de clientes, es la que se presenta en la Tabla 6.14. Calcule el número esperado de ventas diarias y la varianza del número de ventas.

Tabla 6.14 Ventas diarias cuando se visita a 10 prospectos

Número de ventas (X)	1	2	3	4	5	6	7	8
Probabilidad [P (X)]	0.04	0.15	0.20	0.25	0.19	0.10	0.05	0.02

Resp. *E(X)* = 4.00, *V(X)* = 2.52

6.25 Con referencia al problema 6.24, suponga que el vendedor obtiene una comisión de $25,000 por cada venta que realiza. Determine su comisión diaria esperada (a) sustituyendo el monto de la comisión por cada uno de los números de ventas de la Tabla 6.14 y calcule la cantidad esperada por comisiones, y (b) multiplicando el número esperado de ventas que se calculó en el problema 6.24 por la comisión que obtiene por cada venta.

Resp. (a) $ 100 000, *(b)* 100 000

LA DISTRIBUCIÓN BINOMIAL

6.26 Existe una probabilidad de 90% de que un componente especifico se comporte en forma adecuada bajo condiciones de alta temperatura. Si el aparato en el que se usa el componente tiene en total cuatro de ellos, determine la probabilidad de cada uno de los siguientes eventos utilizando la fórmula de las probabilidades binomiales.
(a) Todos los componentes se comportan de forma adecuada y, por lo tanto, el aparato funciona bien.
(b) El aparato no funciona bien porque falla exactamente uno de los cuatro componentes.
(c) El aparato no funciona porque falla uno o *más* de los componentes.

Resp. (a) 0.6561, (b) 0.2916, (c) 0.3439

6.27 Verifique las respuestas del problema 6.26 construyendo un diagrama de árbol y calculando las probabilidades mediante el uso de las reglas apropiadas de multiplicación y de adición.

6.28 Verifique las respuestas al problema 6.26 utilizando el apéndice 2.

6.29 Al utilizar la tabla de probabilidades binomiales, determine:
(a) $P(X = 8|n = 20, p = 0.30)$ (d) $P(X = 5|n = 10, p = 0.40)$
(b) $P(X \geq 10|n = 20, p = 0.30)$ (e) $P(X > 5|n = 10, p = 0.40)$
(c) $P(X \leq 5|n = 20, p = 0.30)$ (f) $P(X < 5|n = 10, p = 0.40)$

Resp. (a) 0.1144, *(b)* 0.0479, (c) 0.4165, *(d)* 0.2007, (e) 0.1663, *(f)* 0.6330

6.30 Al utilizar la tabla de probabilidades binomiales, determine:
(a) $P(X = 4|n = 12, p = 0.70)$ (d) $P(X < 3|n = 8, p = 0.60)$
(b) $P(X \geq 9|n = 12, p = 0.70)$ (e) $P(X = 5|n = 10, p = 0.90)$
(c) $P(X \leq 3|n = 8, p = 0.60)$ (f) $P(X > 7|n = 10, p = 0.90)$

Resp.(a) 0.0078, (b) 0.4925, (c) 0.1738, (d) 0.0499, (e) 0.0015, *(f)* 0.9298

6.31 Suponga que el 40% de los empleados a destajo de una empresa grande están a favor de tener representación sindical y que se entrevista a una muestra aleatoria de 10 de ellos y se les solicita una respuesta anónima. ¿Cuál es la probabilidad de que (a) la mayoría de los que respondan, (b) menos de la mitad de los que respondan estarán a favor de la representación sindical?

Resp. (a) 0.1663, (b) 0.6330

6.32 Determine las probabilidades del problema 6.31, si el 60% de los empleados a destajo de la empresa están a favor de la representación sindical.

Resp. (a) 0.6330, (b) 0.1663

6.33 Con referencia a la distribución de la probabilidad del problema 6.24, ¿parecería que esta distribución de probabilidad tiene forma binomial? {*Sugerencia:* convierta la *E* (X) que se encontró en el problema 6.24 en una proporción y utilice ese valor como valor *p* para compararlo con la distribución binomial con *n* =10.)

Resp. Las dos distribuciones de probabilidad corresponden en forma cercana.

LA DISTRIBUCIÓN HIPERGEOMÉTRICA

6.34 En una clase en la que hay 20 estudiantes, 15 están insatisfechos con el texto que se utiliza. Si se le preguntara acerca del texto a una muestra aleatoria de cuatro estudiantes, determine la probabilidad de que (a) exactamente tres y (6) cuando menos tres de ellos estén insatisfechos con el texto.

Resp. (a) $P \cong 0.47$, (b) $P \cong 0.75$

6.35 Verifique las respuestas al problema 6.34 construyendo un diagrama de árbol y calculando las probabilidades mediante las reglas apropiadas de la multiplicación y de la adición.

6.36 En la sección 6.5 se .sugirió que, en términos generales, puede utilizarse la distribución binomial para aproximar las probabilidades hipergeométricas cuando $n < 0.05N$. Demuestre que la aproximación binomial de los valores de probabilidad que se solicitaron en el problema 6.34 es deficiente. *(Sugerencia:* utilice *T/N* como el valor de p para la tabla binomial, la cual *no* se ajusta al requerimiento del tamaño de la muestra en este caso, porque $n = 4$ es mucho *mayor* que 0.025 (20) = 1.)

6.37 Un equipo departamental incluye a cinco ingenieros y a nueve técnicos. Si se eligen al azar a cinco personas y se les asigna a un proyecto, ¿cuál es la probabilidad de que el equipo del proyecto incluya exactamente a dos ingenieros? *(Nota:* éste es un replanteamiento del problema 5.57(a) para el cual se determinó la respuesta mediante análisis combinatorio.)

Resp. $P \cong 0.42$

LA DISTRIBUCIÓN DE PROBABILIDAD POISSON

6.38. En promedio, seis personas utilizan un cajero bancario automático cada hora, en el transcurso de las horas más concurridas en una tienda de departamentos. ¿Cuál es la probabilidad de que

(a) exactamente seis personas utilicen el cajero automático durante una hora seleccionada al azar?
(b) menos de 5 personas utilicen la caja durante una hora elegida al azar?
(c) nadie utilice la caja durante un intervalo de 10 minutos?
(d) nadie utilice la caja durante un intervalo de 5 minutos?

Resp. (a) 0.1606, (b) 0.2851, (c) 0.3679, (d) 0.6065

6.39 Suponga que el manuscrito para un libro de texto tiene en total 50 errores de mecanografía en el total de las 500 páginas que conforman el material, y que los errores están distribuidos en forma aleatoria en todo el texto. ¿Cuál es la probabilidad de que

(a) un capítulo que cubre 30 páginas tenga dos o más errores?
(b) un capitulo que tiene 50 páginas tenga dos o más errores?
(c) una página elegida al azar no tenga error?

Resp. (a) 0.8008, (b) 0.9596, (c) 0.9048

6.40 Se encuentra que sólo un generador de cada mil está defectuoso, después de ser ensamblado en una planta manufacturera, y los generadores defectuosos se distribuyen en forma aleatoria en toda la corrida de producción.

(a) ¿Cuál es la probabilidad que un embarque de 500 generadores no tenga ningún generador defectuoso?
(b) ¿Cuál es la probabilidad que un embarque de 100 generadores incluya cuando menos un generador defectuoso?

Resp. Mediante la aproximación de Poisson a las probabilidades binomiales. (a) 0.6065, *(b)* 0.952

6.41 Con referencia a la distribución de probabilidad del problema 6.22, ¿parecería que esta distribución de probabilidades de llegadas sigue una distribución de probabilidad Poisson? *(Sugerencia:* utilice la E(X) que se calculó en el problema 6.22 como la media (λ) para determinar la distribución Poisson con la que deben compararse las probabilidades.)

Resp. Las dos distribuciones de probabilidad corresponden en forma estrecha.

(Nota: en los problemas 7.16 a 7.22 se utilizan todas las distribuciones de probabilidad que se cubren en los capítulos 6 y 7.)

APLICACIONES EN COMPUTADORA

6.42 La probabilidad de que un componente electrónico tenga defectos es 0.005.

(a) Utilizando algún programa de computadora, obtenga la tabla de probabilidades binomiales para el número de componentes defectuosos si se utiliza n - 8 componentes en un aparato.
(b) Con referencia a la tabla, determine la probabilidad de que exactamente uno de los componentes tenga defectos.
(c) Con referencia a la tabla, determine la probabilidad de que uno o más de los Componentes tengan defectos.

Resp. (b) 0.0386, (c) 0.0393

Distribuciones de probabilidad para variables aleatorias continuas: normal y exponencial

7

7.1 VARIABLES ALEATORIAS CONTINUAS

A diferencia de una variable aleatoria discreta, una *variable aleatoria continua* es la que puede tomar cualquier valor fraccionario en un rango determinado de valores. (Véase la sección 1.4.) Como existe un número infinito de posibles mediciones fraccionarias, no pueden enlistarse todos los valores posibles con una probabilidad correspondiente. Más bien, se define una *función de densidad de probabilidad.* Esta expresión matemática da la función de X, y se representa mediante el símbolo f (X), para cualquier valor designado de la variable aleatoria X. A la gráfica de una función de este tipo se le denomina *curva de probabilidad* y el área entre dos puntos cualesquiera bajo la curva da la probabilidad de la ocurrencia aleatoria de un valor entre esos dos puntos.

EJEMPLO 1. Para la distribución continua de probabilidad de la figura 7-1, la probabilidad de que un embarque seleccionado al azar tenga un peso neto entre 3 000 y 4 000 kilogramos es igual a la proporción del área total bajo la curva que se encuentra en el área sombreada. Es decir, se define que el área total bajo la función de densidad de probabilidad es igual a 1, y puede determinarse la proporción de esta área que se encuentra entre dos puntos determinados aplicando el método de la integración (del cálculo diferencial e integral) junto con Ja función matemática de densidad de probabilidad para esa curva de probabilidad.

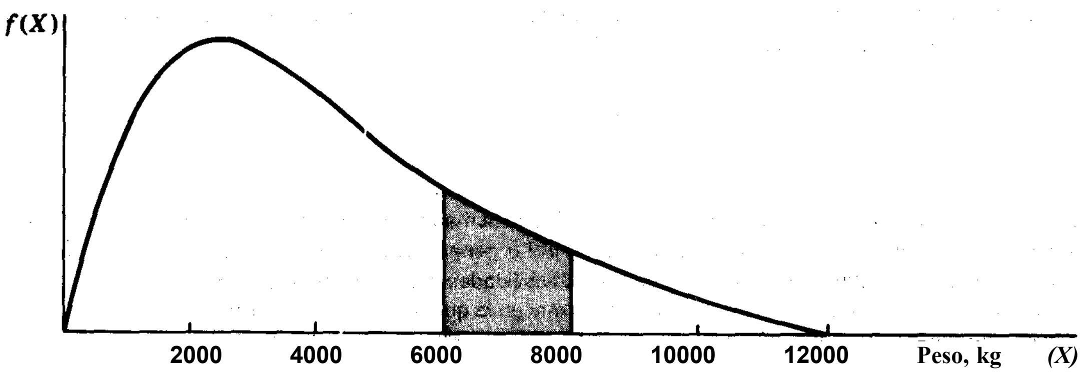


Fig. 7.1

Existen diversas distribuciones continuas de probabilidad comunes que son aplicables como modelos a una amplia gama de variables continuas en determinadas circunstancias. Existen tablas de probabilidades para esas distribuciones estándar, haciendo que resulte innecesario el método de la integración para determinar las áreas bajo la curva de probabilidad para estas distribuciones. Los modelos comunes de distribuciones de probabilidad continua que se describen en este capítulo son las distribuciones normal y la exponencial.

7.2 LA DISTRIBUCIÓN NORMAL DE PROBABILIDAD

La *distribución normal de probabilidad* es una distribución continua de probabilidad que es, al mismo tiempo, *simétrica y mesokúrtica* (tal como se definió en la sección 2.4.) Con frecuencia se describe a la curva de probabilidad que representa a la distribución normal como una campana, tal como se muestra en la curva de probabilidad de la figura 7-2.

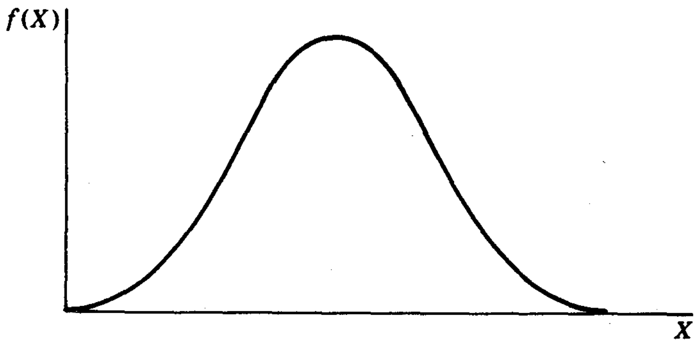


Fig. 7.2

La distribución normal de probabilidad es muy importante en inferencia estadística por tres razones principales:

(1) Se sabe que las mediciones que se obtienen en muchos procesos aleatorios tienen esta clase de distribución.

(2) Con frecuencia pueden utilizarse las probabilidades normales para aproximar otras distribuciones de probabilidad, tales como las distribuciones binomial y Poisson.

(3) Las distribuciones de estadísticas como la media muestral y la proporción muestral tienen distribución normal cuando el tamaño de la muestra es grande, sin importar la forma de la distribución de la población de origen (véase la sección 8.2).

Como se mencionó antes, en el caso de las distribuciones continuas de probabilidad sólo es posible determinar un valor de probabilidad para *un intervalo de valores.* La altura de la función de densidad, o curva de probabilidad, para una variable con distribución normal está dada por

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(X-\mu)^2/2\sigma^2]} \qquad (7.1)$$

en donde π es la constante 3.1416, e es la constante 2.7183, μ es la media de la distribución y σ es la desviación estándar de la distribución. Como cualquier combinación distinta de μ y σ genera una distribución normal de probabilidad distinta (todas ellas simétricas y mesokúrticas), las tablas de las probabilidades normales se basan en una distribución especifica: *la distribución normal estándar.* Ésta es una distribución normal en la que $\mu = 0$ y $\sigma = 1$. Cualquier valor X de una población con distribución normal puede convertirse a su valor normal estándar equivalente, z, mediante la fórmula

$$z = \frac{X - \mu}{\sigma} \qquad (7.2)$$

En el apéndice 5 pueden obtenerse las porciones de área para diversos intervalos de valores para la distribución normal estándar, en donde el limite inferior del Intervalo es siempre la media. Puede utilizarse esta tabla transformando los valores designados de la variable X en valores normales estándar, y esta posibilidad hace que resulte innecesario utilizar el método de la integración con respecto a la ecuación de la función de densidad.

EJEMPLO 2. Se sabe que el tiempo útil de un componente eléctrico tiene una distribución normal con media μ- 2 000 horas y desviación estándar σ - 200 horas. La probabilidad de que un componente elegido al azar dure entre 2 000 y 2 400 horas se determina de la siguiente manera.

En la figura 7-3 se ilustra la curva de probabilidad (función de densidad para este problema y también se señala la relación entre la escala de horas, X, y la escala normal estándarz. Además, la parte sombreada es el área bajo la curva que corresponde al intervalo "2 000 a 2 400".

El limite inferior del intervalo es la media de la distribución y, por lo tanto, se encuentra en el valor z - 0. El límite superior del intervalo designado, en términos de un valorz, es:

$$z=\frac{X-\mu}{\sigma}=\frac{2\,400-2\,000}{200}=\frac{400}{200}=+2.0$$

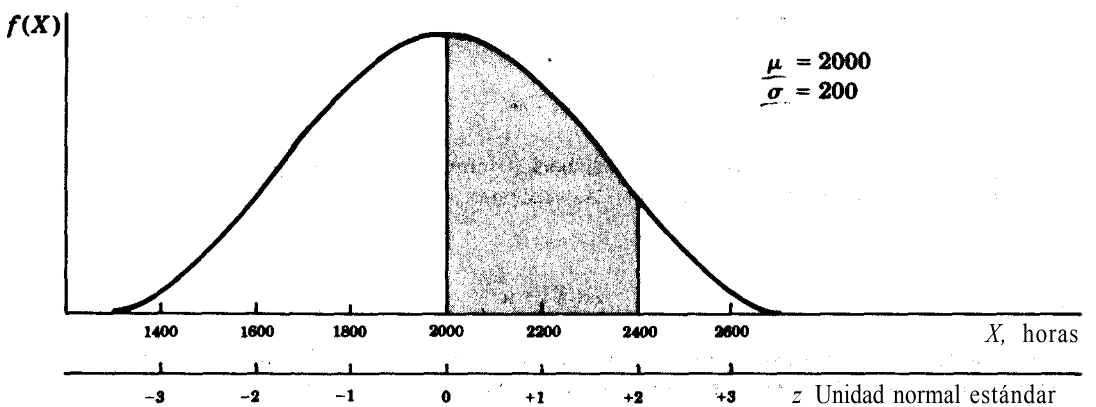


Fig. 7.3

Del apéndice 5, se encuentra que

$$P(0<z<+2.0)=0.4772$$

Por lo tanto,

$$P(2\,000<X<2\,400)=0.4772$$

Por supuesto, no todos los problemas implican un intervalo en el que la media es el limite inferior. Sin embargo, puede utilizarse el apéndice 5 para determinar el valor de probabilidad asociado con cualquier intervalo de interés, realizando la adición o substracción de áreas, según sea necesario, o utilizando el hecho de que la curva es simétrica. En el ejemplo 3, y en los problemas 7.1 a 7.8 se incluyen diversos ejemplos de esta clase de aplicaciones.

EJEMPLO 3. Con respecto a los componentes eléctricos que se describieron en el ejemplo 2, suponga que interesa la probabilidad de que un componente elegido al azar dure *más* de 2 200 horas.

Debe observarse que, por definición, la proporción total del área que se encuentra del lado derecho de la media de 2 000 en la figura 7-4, es 0.5000. Por ello, si se determina la propprción entre la media y 2 200, puede restarse este valor de 0.5000 para obtener la probabilidad de que las horas X sean mayores que 2 200, lo cual se representa gráficamente mediante la porción sombreada de la figura 7-4.

$$z=\frac{2\,200-2\,000}{200}=+1.0$$

$$P(0\leq z\leq+1.0)=0.3413 \quad \text{(del apéndice 5)}$$

$$P(z>+1.0)=0.5000-0.3413=0.1587$$

Por tanto,

$$P(X>2\,200)=0.1587$$

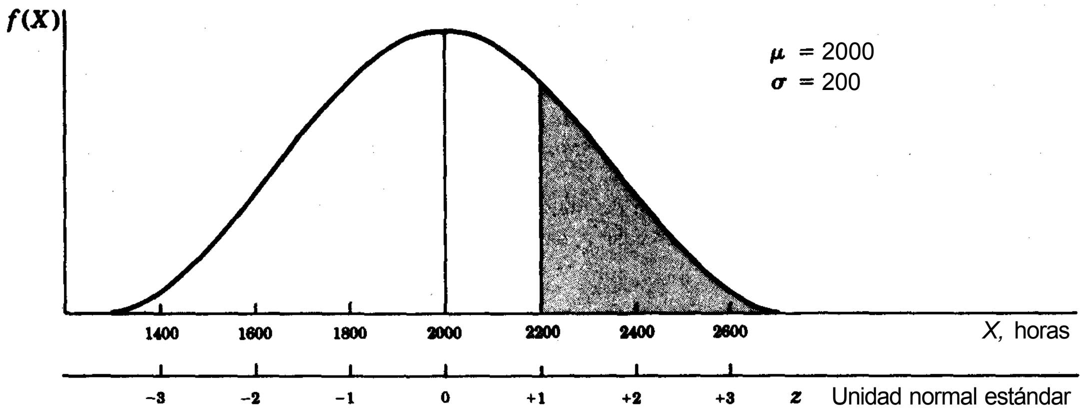


FIg. 7.4

El valor esperado de una variable aleatoria con distribución normal es (en donde puede calcularse la media μ de la población mediante las fórmulas que se presentaron en las secciones 3.2 y 3.8):

$$E(X) = \mu \tag{7.3}$$

Y, la varianza de una variable aleatoria con distribución normal es (en donde la varianza σ^2 de la población puede calcularse mediante las fórmulas que se presentaron en las secciones 4.5,4.6 y 4.12):

$$V(X) = \sigma^2 \tag{7.4}$$

7.3 PUNTOS PERCENTILES PARA VARIABLES CON DISTRIBUCIÓN NORMAL

Puede recordarse de lo visto en la sección 3.7, que el punto percentil 90 es el punto de la distribución tal que el 90% de los valores se encuentran por debajo de él y el 10% por encima. Para la distribución normal estándar, es el valor de z tal que la proporción total de área a la izquierda de ese valor, bajo la curva normal, es 0.90.

EJEMPLO 4. En la figura 7-5, se ilustra la posición del punto percentil 90 para la distribución normal estándar. Para determinar el valor requerido de z, se utiliza el apéndice 5 en el sentido *contrario* al común, porque, en este caso, el área bajo la curva entre la media y el punto de interés es 0.40, tal como se ha especificado, y se desea determinar el valor correspondiente de z. En el apéndice 5 se busca en el *cuerpo* de la tabla el valor más cercano a 0.4000. Este valor resulta ser 0.3997. Determinando los encabezados del renglón y de la columna, se encuentra que el valor de z asociado con esta área es 1.28 y, por lo tanto, $z_{0.90}$ + 1.28.

Dado el procedimiento de este ejemplo 4, que permite determinar un punto percentil para la distribución normal estándar, puede determinarse un punto percentil para una variable aleatoria con distribución normal conviniendo el valor pertinente de z al valor que se requiere de X, mediante la fórmula

$$X = \mu + z\sigma \tag{7.5}$$

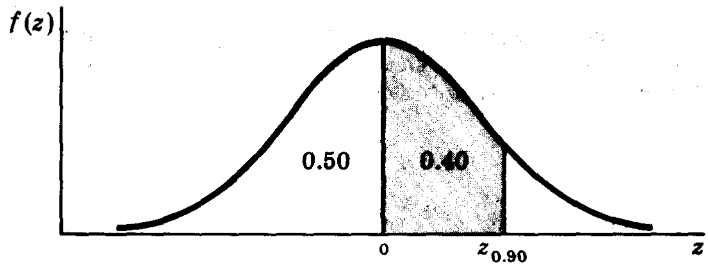


Fig. 7.5

EJEMPLO 5. Para la vida útil de los componentes eléctricos que se describieron en los ejemplos 2 y 3, y utilizando la solución del ejemplo 4, el punto percentil 90 para la vida útil de los componentes es

$$X = \mu + z\sigma = 2\ 000 + (1.28)(200) = 2\ 256 \text{ h}$$

Para los puntos percentiles que se encuentran por debajo del **50º** percentil, el valor asociado de z será negativo, puesto que ese valor se encuentra por debajo del 0 de la distribución normal estándar.

EJEMPLO 6. Al continuar con el ejemplo 5, suponga que se desea determinar la vida útil de los componentes, de manera que sólo el 10% de ellos fallen antes de ese tiempo (el punto percentil 10). En la figura 7-6, se ilustra el área correspondiente de la distribución de probabilidad normal estándar. Tal como se hizo en el ejemplo 4, se busca en el cuerpo de la tabla del apéndice 5 el área que esté más cercana a 0.4000 pero, en este caso, se toma un valor *z* negativo. La solución es

$$X = \mu + z\sigma = 2\ 000 + (-1.28)(20) = 1\ 744 \text{ h}$$

7.4 APROXIMACIÓN NORMAL A PROBABILIDADES BINOMIALES

Cuando el número de observaciones o ensayos n es relativamente grande, puede utilizarse la distribución normal de probabilidad para aproximar las posibilidades binomiales. Una regla conveniente consiste en afirmar que ésas aproximaciones son aceptables cuando $n \geq 30$, y tanto $np \geq 5$ como $nq \geq 5$. Esta regla, en combinación con la que se proporciona en la

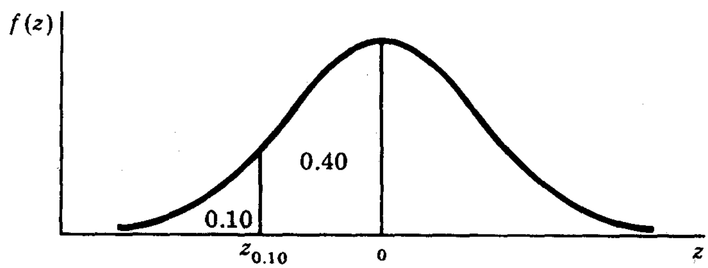


Fig. 7.6

sección 6.7 con respecto a la aproximación de Poisson a las probabilidades binomiales, significa que en ios casos en que $n > 30$, las probabilidades binomiales pueden aproximarse, ya sea mediante la distribución normal o la de Poisson, dependiendo de los valores *np y nq.* Algunos otros textos pueden utilizar reglas un tanto distintas para determinar los casos en los que esas aproximaciones son apropiadas.

Cuando se utiliza la distribución normal de probabilidad como base para aproximar un valor binomial de probabilidad, la media y la desviación estándar se basan en el valor esperado y la varianza del número de éxitos de la distribución binomial, según se vio en la sección 6.3. El número promedio de "éxitos" es

$$\mu = np \tag{7.6}$$

La desviación estándar del número de "éxitos" es

$$\sigma = \sqrt{npq} \tag{7.7}$$

EJEMPLO 7. Se ha observado que para un grupo grande de prospectos de venta, el 20% de los que un vendedor visita en forma personal realizan la compra. Si un representante de ventas visita a 30 prospectos, puede determinarse la probabilidad de que 10 o más de ellos realicen una compra utilizando las probabilidades binomiales de la apéndice 2:

$$\begin{aligned} P(X \geq 10|n = 30, p = 0.20) &= P(X = 10) + P(X = 11) + \cdots \\ &= 0.0355 + 0.0161 + 0.0064 + 0.0022 + 0.0007 + 0.0002 \\ &= 0.0611 \text{ (el valor de la probabilidad binomial)} \end{aligned}$$

Ahora, se verifica si se satisfacen los criterios para la aproximación normal

Es $n \geq 30$?	Si, $n = 30$
Es $np \geq 5$?	Si, $np = 30(0.20) = 6$
Es $nq \geq 5$?	Si, $nq = 30(0.80) = 24$

La aproximación normal del valor binomial de probabilidad es

$$\mu = np = (30)(0.20) = 6.0$$

$$\sigma = \sqrt{npq} = \sqrt{(30)(0.20)(0.80)} = \sqrt{4.8} \cong 2.19$$

$$P_{\text{Binomial}}(X \geq 10|n = 30, p = 0.20) \cong P_{\text{Normal}}(X \geq 9.5|\mu = 6.0, \sigma = 2.19)$$

(Nota: esto incluye una corrección por continuidad que se analizó anteriormente.)

$$z = \frac{X - \mu}{\sigma} = \frac{9.5 - 6.0}{2.19} = \frac{3.5}{2.19} = +1.60$$

$$\begin{aligned} P(X \geq 9.5|\mu = 6.0, \sigma = 2.19) &= P(z \geq = 1.60) \\ &= 0.5000 - P(0 \leq z \leq +1.60) = 0.5000 - 0.4452 \\ &= 0.0548 \quad \text{(la aproximación normal)} \end{aligned}$$

En el ejemplo 7, se supone que la clase de eventos "diez o más" comienza en 9.5, cuando se utiliza la aproximación normal. A este ajuste de media unidad se le denomina *corrección por continuidad.* Se requiere porque es necesario asignar toda el área bajo la distribución normal continua a algún evento numérico, aun cuando no es posible tener un número fraccionario de éxitos (por ejemplo, entre "nueve compradores" y "diez compradores"). En otras palabras, en el proceso de una aproximación normal, debe considerarse que un evento discreto tal como "diez compradores" representa un intervalo continuo de valores que varían desde un límite exacto inferior de 9.5 hasta un límite exacto superior de 10.5. Por lo tanto, si en el ejemplo 7 se hubiera solicitado la probabilidad de "más de diez compradores", la corrección por continuidad apropiada implicaría sumar 0.5 al 10, y determinar el área para el intervalo que comienza en 10.5. Siguiendo este razonamiento, el 0.5 se añade o se substrae como una corrección por continuidad, de acuerdo con la forma del planteamiento de probabilidad:

(1) Se resta 0.5 de X cuando $P(X > X_i)$ se requiere.
(2) Se resta 0.5 de X cuando $P(X < X_i)$ se requiere.
(3) Se suma 0.5 de X cuando $P(X < X_i)$ es requerido.
(4) Se suma 0.5 de X cuando $P(X > X_i)$ es requerido.

7.5 APROXIMACIÓN NORMAL A PROBABILIDADES DE POISSON

Cuando la media λ de una distribución Poisson es relativamente grande, puede utilizarse la distribución normal de probabilidad para aproximar probabilidades tipo Poisson. Una regla práctica consiste en afirmar que esa aproximación es aceptable cuando $\lambda \geq 10.0$.

La media y la desviación estándar de la distribución normal de probabilidad se basan en el valor esperado y la varianza del número de eventos de un proceso Poisson, según se identificó en la sección 6.6. Esta media es

$$\mu = \lambda \tag{7.8}$$

La desviación estándar es

$$\sigma = \sqrt{\lambda} \tag{7.9}$$

EJEMPLO 8. El número promedio de solicitudes de servicio que se reciben en un departamento de reparación de maquinaria por cada turno de 8 horas es 10.0. Puede determinarse la probabilidad de que se reciban más de 15 solicitudes en un turno de 8 horas elegido al azar utilizando el apéndice 4:

$$P(X > 15|\lambda = 10.0) = P(X=16) + P(X=17) + \cdots$$
$$- 0.0217 + 0.0128 + 0.0071 + 0.0037 + 0.0019 + 0.0009 + 0.0004 + 0.0002 + 0.0001$$
$$= 0.0488 \qquad \text{(La probabilidad Poisson)}$$

Como el valor d λ es (cuando menos) 10, la aproximación normal al valor de probabilidad Poisson resulta ser aceptable. La aproximación normal del valor de probabilidad Poisson es

$$\mu = \lambda = 10.0$$
$$\sigma = \sqrt{\lambda} = \sqrt{10.0} \cong 3.16$$
$$P_{\text{Poisson}}(X > 15 \mid \lambda = 10.0) \cong P_{\text{Normal}}(X \geq 15.5 \mid \mu = 10.0 \quad \sigma = 3.16)$$

(Nota: Esto incluye la corrección por continuidad que se analizó anteriormente.)

$$z = \frac{X - \mu}{\sigma} = \frac{15.5 - 10.0}{3.16} = \frac{5.5}{3.16} = +1.74$$

$$P(z \geq +1.74) = 0.5000 - P(0 \leq z \leq +1.74) = 0.5000 - 0.4591 = 0.0409 \qquad \text{(la aproximación normal)}$$

La corrección por continuidad que se aplicó en el ejemplo 8 es del mismo tipo que la corrección que se describió para la aproximación normal de probabilidades binomiales. Las reglas que se revisaron en la sección 7.4 con respecto al 0.5 que se suma y se resta a la X se aplican de igual manera a la situación en la que se utiliza la distribución normal para aproximar probabilidades Poisson.

7.6 LA DISTRIBUCIÓN EXPONENCIAL DE PROBABILIDAD

Si se presentan eventos en el contexto de un proceso Poisson, tal como se describió en la sección 6.6, entonces la longitud del tiempo o el espacio entre eventos sucesivos tiene una *distribución exponencial de probabilidad.* Como el tiempo o el espacio son continuos, una medición de este tipo es una variable aleatoria continua. Tal como resulta ser cierto para cualquier variable aleatoria continua, no tiene sentido preguntar "¿cuál es la probabilidad de que la primera solicitud de servicio llegue *exactamente* en un minuto?". Más bien, se debe determinar un *intervalo* dentro del cual debe ocurrir el evento, como por ejemplo preguntando "¿cuál es la probabilidad de que la primera solicitud de servicio llegue en un minuto?".

Como los procesos Poisson son estacionarios, y se tiene una probabilidad igual de que el evento ocurra a todo lo largo del periodo relevante de tiempo, la distribución exponencial se aplica: si lo que interesa es el tiempo (o espacio) hasta la ocurrencia del primer evento, o el tiempo *entre* dos eventos sucesivos, o el tiempo que transcurre hasta que se presenta el primer evento, después de cualquier punto en el tiempo elegido ai azar.

La probabilidad exponencial de que ocurra el primer evento *dentro* del intervalo designado de tiempo o espacio es (en donde λ es el número promedio de ocurrencias para *el intervalo de interés,* según la sección 6.6);

$$P(T \leq t) = 1 - e^{-\lambda} \qquad \{7.10)$$

De manera similar, la probabilidad exponencial de que el primer evento *no* ocurra dentro del intervalo designado de tiempo o espacio es

$$P(T > t) = e^{-\lambda} \qquad (7.11)$$

Para las dos fórmulas anteriores, puede obtenerse el valor de $e^{-\lambda}$ en el apéndice 3.

EJEMPLO 9. En un departamento de reparación de maquinaria se reciben 5 solicitudes por hora en promedio. Comenzando la observación en cualquier punto del tiempo, la probabilidad de que se reciba la primera solicitud de servicio dentro de un lapso de *media hora* es

Promedio por hora - 5.0
λ = Promedio por media hora » 2.5
$P = 1 - e^{-\lambda} = 1 - e^{-2.5} = 1 - 0.08208 = 0.91792$ (del apéndice 3)

El valor esperado y la varianza de una distribución exponencial de probabilidad, en donde la variable se designa como tiempo *T,* son

$$E(T) = \frac{1}{\lambda} \qquad (7.12)$$

$$V(T) = \frac{1}{\lambda^2} \qquad (7.13)$$

7.7 APLICACIONES EN COMPUTADORA

Es frecuente que los programas de computación para análisis estadístico incluyan la posibilidad de obtener probabilidades para intervalos de valores con variables con distribución normal. En los problemas 7.23 y 7.24 se ilustra el cálculo de este tipo de probabilidades y, también, el de puntos percentiles para variables aleatorias con distribución normal, utilizando una computadora.

Problemas resueltos

LA DISTRIBUCIÓN NORMAL DE PROBABILIDAD

7.1 Se ha ajustado el proceso de fabricación de un tomillo de precisión de manera que la longitud promedio de los tornillos sea μ- 13.0 cm. Por supuesto, no todos los tornillos tienen una longitud exacta de 13 centímetros, debido a fuentes aleatorias de variabilidad. La desviación estándar de la longitud de los tornillos es σ - 0.1 cm y se sabe que la distribución de las longitudes tiene una forma normal. Determine la probabilidad de que un tomillo elegido al azar tenga una longitud de entre 13.0 y 13.2 cm, e ilustre la proporción de área bajo la curva normal asociada con este valor de probabilidad.

De la figura 7.7,

$$z=\frac{X-\mu}{\sigma}=\frac{13.2-13.0}{0.1}=+2.0$$

$$P(13.0\leq X\leq 13.2)=P(0\leq z\leq +2.0)=0.4772 \qquad \text{(del apéndice 5)}$$

7.2 Para la situación que se describió en el problema 7.1, ¿cuál es la probabilidad de que la longitud del tornillo exceda de 13.25 cm? Ilustre la proporción del área bajo la curva normal correspondiente a este caso.

Con referencia a la figura 7-8,

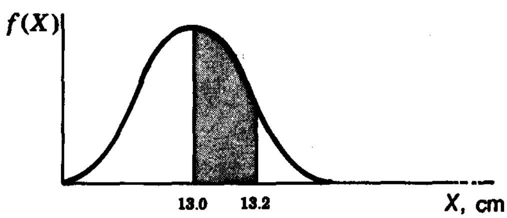


Fig. 7.7

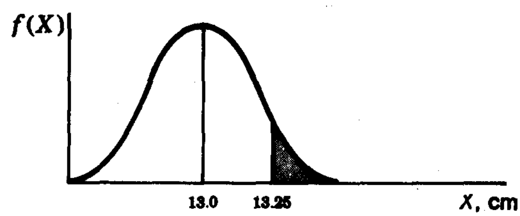


Fig. 7.8

$$z=\frac{X-\mu}{\sigma}=\frac{13.25-13.0}{0.1}=+2.5$$

$$P(X>13.25)=P(z>+2.5)=0.5000-0.4938=0.0062$$

7.3 Del problema 7.1, ¿cuál es la probabilidad de que la longitud del tornillo esté entre 12.9 y 13.1 cm? Ilustre la proporción de área bajo la curva normal correspondiente a este caso.

Con referencia a la figura 7-9,

$$z_1=\frac{X_1-\mu}{\sigma}=\frac{12.9-13.0}{0.1}=-1.0$$

$$z_2=\frac{X_2-\mu}{\sigma}=\frac{13.1-13.0}{0.1}=+1.0$$

$$P(12.9\leq X\leq 13.1)=P(-1.0\leq z\leq +1.0)=0.3413+0.3413=0.6826$$

(Nota: ésta es la proporción de área desde -1 .0z a μ, más la proporción desde μ hasta + 1 .0z. Note también que, como la distribución normal de probabilidad es simétrica, las áreas hacia la izquierda de la media para valores negativos de z son equivalentes a las áreas que se encuentran del lado derecho de la media.)

7.4 ¿Cuál es la probabilidad de que la longitud de los tornillos del problema 7.1 se encuentre entre 12.8 y 13.1 cm? Ilustre la proporción del área bajo la curva normal para este caso.

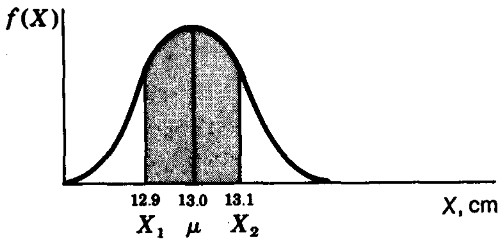


Fig. 7.9

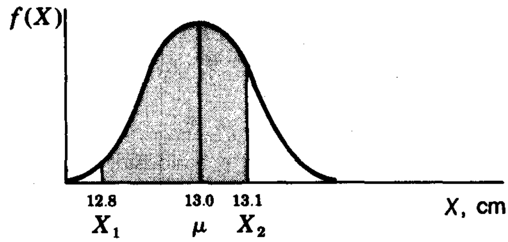


Fig. 7.10

Con referencia a la figura 7-10,

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{12.8 - 13.0}{0.1} = \frac{-0.2}{0.1} = -2.0$$

$$z_2 = \frac{X_2 - \mu}{\sigma} = \frac{13.1 - 13.0}{0.1} = \frac{0.1}{0.1} = +1.0$$

$$P(12.8 \leq X \leq 13.1) = P(-2.0 \leq z \leq 1.0) = 0.4772 + 0.3413 = 0.8185$$

7.5 En el problema 7.1, ¿cuál es la probabilidad de que la longitud del tomillo esté entre 13.1 y 13.2 cm.? Ilustre la proporción del área bajo la curva normal que es relevante en este caso.

Con referencia a la figura 7-11,

$$z_1 = \frac{X_1 - \mu}{\sigma} = \frac{13.1 - 13.0}{0.1} = +1.0$$

$$z_2 = \frac{X_2 - \mu}{\sigma} = \frac{13.2 - 13.0}{0.1} = +2.0$$

$$P(13.1 \leq X \leq 13.2) = P(+1.0 \leq z \leq +2.0) = 0.4772 - 0.3413 = 0.1359$$

(Nota: la probabilidad es igual a la proporción del área de 13.0 a 13.2, menos la proporción de área de 13.0 a 13.1.)

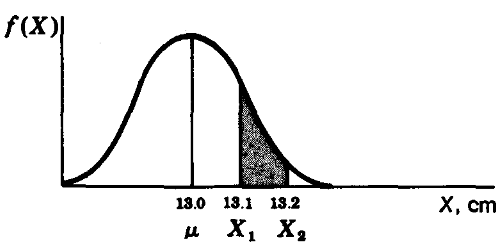


Fig. 7.11

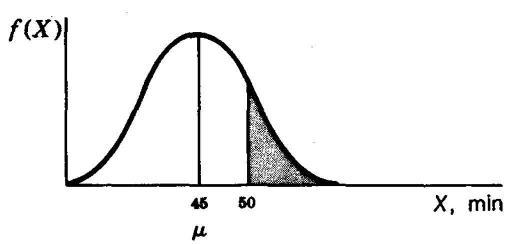


Fig. 7.12

7.6 El tiempo que se requiere para reparar cierto tipo de transmisión automotriz en un taller mecánico tiene distribución normal con media $\mu = 45$ min y desviación estándar $\sigma = 8.0$ min. El gerente de servicio planea hacer que se inicie la reparación de la transmisión de los automóviles de los clientes diez minutos después de que se recibe el vehículo, y le dice al cliente que el automóvil estará listo en una hora. ¿Cuál es la probabilidad de que el gerente esté equivocado? Ilustre la proporción de área bajo la curva normal para este caso.

De la figura 7-12,

$P(X > 50$ min), puesto que el trabajo real comienza en 10 minutos

$$z = \frac{X-\mu}{\sigma} = \frac{50-45}{8.0} = \frac{5.0}{8.0} = +0.62$$

$$P(X > 50) = P(z > +0.62) = 0.5000 - 0.2324 = 0.2676$$

PUNTOS PERCENTILES PARA VARIABLES DISTRIBUIDAS NORMALMENTE

7.7 Con referencia al problema 7.6, ¿qué asignación de tiempo de trabajo se requiere para que haya una probabilidad del 75% de que la reparación de las transmisiones se lleve a cabo dentro de ese tiempo? Ilustre la proporción de área correspondiente.

Tal como se ilustra en la figura 7-13, entre la media y el punto percentil 75 se incluye una proporción de área del 0.2500. Por lo tanto, el primer paso en la solución implica determinar el valor de z requerido encontrando el área en el cuerpo de la tabla del apéndice 5 que esté más cercano a 0.2500. El área más próxima es 0.2486, con $z_{0.75} = +0.67$. Después, se convierte este valor de z en el valor que se requiere de X, de la siguiente manera:

$$X = \mu + z\sigma = 45 + (0.67)(8.0) = 50.36 \text{ min}$$

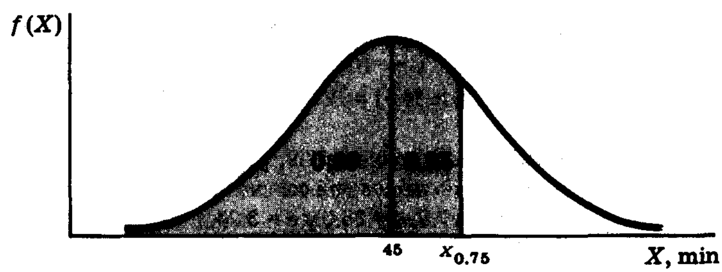


Fig. 7.13

7.8 Con referencia al problema 7.6 ¿cuál es la asignación de tiempo de trabajo que se requiere para que haya una probabilidad de sólo el 30% de que pueda terminarse el trabajo de reparación dentro de ese lapso? Ilustre la proporción de área correspondiente.

Como una proporción de área de 0.30 se encuentra a la izquierda del valor desconocido de X en la figura 7-14, se sigue que hay una proporción de 0.20 entre ese punto percentil y la media. Consultando el apéndice 5, se encuentra que la proporción de área más cercana a ese valor es 0.1985, al cual corresponde un valor de $z_{0.30} = -0.52$. El valor de z es negativo porque el punto percentil se encuentra del lado izquierdo de la media. Finalmente, se convierte el valor de z al valor que se requiere de X.

Por lo tanto,

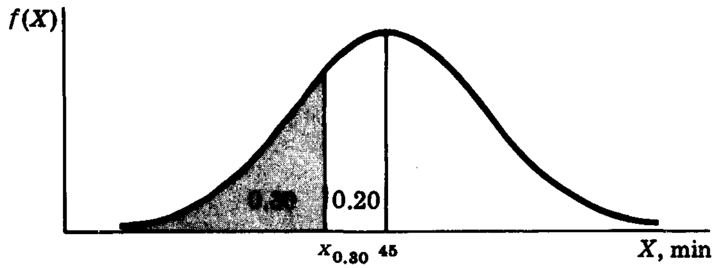


Fig. 7.14

APROXIMACIÓN NORMAL A LAS PROBABILIDADES BINOMIALES Y POISSON

7.9 Se ha encontrado que el 70% de las personas que entran a un centro comercial realizan cuando menos una compra. Para una muestra de n = 50 personas, ¿cuál es la probabilidad de que cuando menos 40 de ellas realicen una o más compras?.

Puede utilizarse la aproximación normal del valor binomial de probabilidad que se requiere porque $n > 30$, $np \geq 5$ y $nq \geq 5$

$$\mu = np = (50)(0.70) = 35.0$$

$$\sigma = \sqrt{npq} = \sqrt{(50)(0.70)(0.30)} = \sqrt{10.5} = 3.24$$

$$P_{\text{Binomial}}(X \geq 40 \mid n = 50,\ p = 0.70) \cong P_{\text{Normal}}(X \geq 39.5 \mid \mu = 35.0,\ \sigma = 3.24)$$

(*Nota:* se incluye la corrección por continuidad que se describió en la sección 7.4.)

$$z = \frac{X - \mu}{\sigma} = \frac{39.5 - 35.0}{3.24} = \frac{4.5}{3.24} = +1.39$$

$$P(X \geq 39.5) = P(z \geq +1.39) = 0.5000 - 0.4177 = 0.0823$$

7.10 Para la situación que se describe en el problema 7.9, ¿cuál es la probabilidad de que menos de 30 de entre 50 personas muestreadas realicen cuando menos una compra?.

Al considerar que, del problema **7.9, $\mu = 35.0$ y $\sigma = 3.24$,**

$$P_{\text{Binomial}}(X < 30 \mid n = 50,\ p = 0.70) \cong P_{\text{Normal}}(X \leq 29.5 \mid \mu = 35.0\quad \sigma = 3.24)$$

(*Nota:* se incluye la corrección por continuidad que se describió en la sección 7.4).

$$z = \frac{X - \mu}{\sigma} = \frac{29.5 - 35.0}{3.24} = \frac{-5.5}{3.24} = -1.70$$

$$P(X \leq 29.5) = P(z \leq -1.70) = 0.5000 - 0.4554 = 0.0446$$

7.11 Se sabe que las solicitudes de servicio llegan en forma aleatoria y en forma de proceso estacionario a un promedio de 5 solicitudes por hora. ¿Cuál es la probabilidad de que se reciban más de 50 solicitudes de servicio durante un turno de 8 horas?

Como la media del periodo dé 8 horas para este proceso de Poisson excede $\lambda = 10$, puede utilizarse la distribución nomal de probabilidad para aproximar el valor de probabilidad Poisson. Como $\mu = \lambda = 40.0$ y $\sigma = \sqrt{\lambda} = \sqrt{40.0} = 6.32$,

$$P_{\text{Poisson}}(X > 50 | \lambda = 40.0) \cong P_{\text{Normal}}(X \geq 50.5 | \mu = 40.0 \quad \sigma = 6.32)$$

(*Nota:* se incluye la corrección por continuidad.)

$$z = \frac{X - \mu}{\sigma} = \frac{50.5 - 40.0}{6.32} = \frac{10.5}{6.32} = +1.66$$

$$P(X \geq 50.5) = P(z \geq +1.66) = 0.5000 - 0.4515 = 0.0485$$

7.12 Con referencia al problema 7.11, ¿cuál es la probabilidad de que se reciban en un turno de 8 horas, 35 o menos solicitudes de servicio?

Como $\mu = 40.0$ y $\sigma = 6.32$,

$$P_{\text{Poisson}}(X \leq 35 | \lambda = 40.0) \cong P_{\text{Normal}}(X \leq 35.5 | \mu = 40.0 \quad \sigma = 6.32)$$

(*Nota:* Se incluye la corrección por continuidad.)

$$z = \frac{X - \mu}{\sigma} = \frac{35.5 - 40.0}{6.32} = \frac{-4.5}{6.32} = -0.71$$

$$P(X \leq 35.5) = P(z \leq -0.71) = 0.5000 - 0.2612 = 0.2388$$

LA DISTRIBUCIÓN EXPONENCIAL DE PROBABILIDAD

7.13 En promedio, cada dos días llega un barco a determinado muelle. ¿Cuál es la probabilidad de que, después de la salida de un barco, pasen cuatros días antes de la llegada del siguiente?

$$\text{Promedio por dos días} = 1.0$$
$$\text{Promedio por día} = 0.5$$
$$\lambda = \text{promedio de periodo de cuatro días} = 4 \times 0.5 = 2.0$$
$$P(T > 4) = e^{-\lambda} = e^{-2.0} = 0.13534 \quad \text{(del apéndice 3)}$$

7.14 Cada rollo de 500 metros de lámina de acero tiene dos defectos en promedio. ¿Cuál es la probabilidad de que, al desenrollar la lámina de acero, se encuentre el primer defecto en el primer segmento de 50 metros?

$$\text{Promedio por rollo de 500 metros} = 2.0$$
$$\lambda = \text{Promedio por segmento de 50 metros} = \frac{2.}{10} = 0.20$$
$$P(T \leq 50) = 1 - e^{-\lambda} = 1 - e^{-0.20} = 1 - 0.81873 = 0.18127 \quad \text{(del apéndice 3)}$$

7.15 Puede transformarse una aplicación que implique el uso de la distribución exponencial a una forma de distribución Poisson, y viceversa. Para ilustrar este tipo de transformación, supóngase que llega un promedio de 4 aviones a un hangar para su reparación, por cada periodo de 8 horas, (a) ¿Cuál es la probabilidad de que la primera llegada *no* ocurra durante la primera hora de trabajo? (b) Demuestre que el problema equivalente, usando la distribución Poisson es equivalente a la probabilidad de que *no haya* llegadas en el periodo de una hora, (c) ¿Cuál es la probabilidad de que la primera llegada ocurra dentro de la primera hora? (d) Demuestre que el problema equivalente, usando el enfoque Poisson corresponde a la probabilidad de que haya una o más llegadas durante el primer periodo.

(a) $\lambda = 0.5$ (por hora)
$P(T > 1) = e^{-\lambda} = e^{-0.5} = 0.60653$ (del apéndice 3)
(b) $P(X = 0|\lambda = 0.5) = 0.6065$ (del apéndice 4)
(c) $\lambda = 0.5$ (por hora)
$P(T \leq 1) = 1 - e^{-\lambda} = 1 - e^{-0.5} = 1 - 0.60653 = 0.39347$ (del apéndice 3)
(d) $P(X \geq 1|\lambda = 0.5) = 1 - P(X = 0) = 1.0000 - 0.6065 = 0.3935$ (del apéndice 4)

Así, se observa que en ambas transformaciones las respuestas son idénticas, excepto por el número de dígitos que se incluyen en las dos tablas.

PROBLEMAS DIVERSOS SOBRE DISTRIBUCIONES DE PROBABILIDAD

(Nota: en los problemas 7.16 a 7.22 se ilustra el uso de las distribuciones de probabilidad que se cubrieron en los capítulos 6 y 7.)

7.16 Un embarque de 10 máquinas Incluye una defectuosa. Si se eligen 7 máquinas al azar de ese embarque, ¿cuál es la probabilidad de que ninguna de las 7 esté defectuosa?

Al utilizar la distribución hipergeométrica (veáse la sección 6.5),

$$P(X \mid N, T, n) = \frac{\binom{N-T}{n-X}\binom{T}{X}}{\binom{N}{n}}$$

$$P(X = 0 \mid N = 10, T = 1, n = 7) = \frac{\binom{10-1}{7-0}\binom{1}{0}}{\binom{10}{7}} = \frac{\left(\frac{9!}{7!2!}\right)\left(\frac{1!}{0!1!}\right)}{\left(\frac{10!}{7!3!}\right)} = \frac{(36)(1)}{120} = 0.30$$

7.17 Suponga que en el problema 7.16, la proporción global de máquinas con defectos es de 0.10, pero que se ensambla un número grande de ellas en la planta. ¿Cuál es la probabilidad de que una muestra aleatoria de 7 máquinas no incluya ninguna defectuosa?

Al utilizar la distribución binomial (véase la sección 6.3),

$$P(X = 0 \mid n = 7, p = 0.10) = 0.4783 \qquad \text{(del apéndice 2)}$$

7.18 Suponga que la proporción de máquinas que tienen defectos en una operación de ensamble es 0.10, y que se incluye una muestra de 200 máquinas en un embarque especifico. ¿Cuál es la probabilidad de que cuando menos 30 de las 200 máquinas estén defectuosas?

En este caso, resulta aceptable el uso de la aproximación normal a la distribución binomial de probabilidad que se describió en la sección 7.4, porque $n \geq 30$, $np \geq 5$ y $nq \geq 5$.

$$\mu = np = (200)(0.10) = 20.0$$
$$\sigma = \sqrt{np(q)} = \sqrt{(200)(0.10)(0.90)} = \sqrt{18.00} \cong 4.24$$
$$P_{\text{Binomial}}(X \geq 30 \mid n = 200, p = 0.10) \cong P_{\text{Normal}}(X \geq 29.5 \mid \mu = 20.0, \sigma = 4.24)$$

(*Nota:* se Incluye la corrección por continuidad.)

$$z = \frac{X - \mu}{\sigma} = \frac{29.5 - 20.0}{4.24} = \frac{9.5}{4.24} = +2.24$$
$$P(X \geq 29.5) = P(z \geq +2.24) = 0.5000 - 0.4875 = 0.0125 \quad \text{(del apéndice 5)}$$

7.19 Suponga que la proporción de máquinas defectuosas en una operación de ensamble es de 0.01, y que se incluye una muestra de 200 de ellas en un embarque especifico. ¿Cuál es la probabilidad de que 3 o menos máquinas estén defectuosas?

En este caso, resulta aceptable el uso de la aproximación de Poisson a la distribución binomial de probabilidad (sección 6.7) porque $n \geq 30$ y $np \geq 5$.

$$\lambda = np = (200)(0.01) = 2.0$$
$$P_{\text{Binomial}}(X \leq 3 \mid n = 200, p = 0.01) = P_{\text{Poisson}}(X \leq 3 \mid \lambda = 2.0)$$
$$= 0.1353 + 0.2707 + 0.2707 + 0.1804 = 0.8571$$
(del apéndice 4)

7.20 Un promedio de 0.5 clientes por minuto lleca a una caja de salida. Después de que el cajero inicia sus operaciones, ¿cuál es la probabilidad de que tenga que esperar, cuando menos 3 minutos, antes de que llegue el primer cliente?

Al utilizar la distribución exponencial le probabilidad que se describió en la sección 7.6,

Promedio por minuto = 0.5
λ = promedio por 3 minutos -0.5 × 3 - 1.5
$P(T > 3) = e^{-\lambda} = e^{-1.5} = 0.22313$ (del apéndice 3)

7.21 Un promedio de 0.5 clientes por minuto llega a una caja de salida en un almacén. ¿Cuál es la probabilidad de que lleguen 5 o más clientes en un intervalo dado de 5 minutos?

De la sección 6.6, de la distribución de probabilidad Poisson

Promedio por minuto =0.5
λ = promedio por 5 minutos = 0.× 5 - 2.5
$P(X \geq 5 \mid \lambda = 2.5) = 0.0668 + 0.0278 + 0.0099 + 0.0031 + 0.0009 + 0.0002 - 0.1087$
(del apéndice 4)

7.22 Un promedio de 0.5 clientes por minuto llega a una caja de salida de un almacén ¿Cuál es la probabilidad de que lleguen más de 20 clientes a la caja en un intervalo especifico de media hora?

Aquí resulta aceptable el uso de la aproximación normal a la distribución de probabilidad Poisson que se describió en la sección 7.5, porque $\lambda \geq 10.0$

Promedio por minuto - 0.5
λ = promedio por 30 minutos - 0.5 × 30 - 15.0
$\mu = \lambda = 15.0$
$\sigma = \sqrt{\lambda} = \sqrt{15.0} = 3.87$
$$P_{\text{Poisson}}(X > 20 \mid \lambda = 15.0) = P_{\text{Normal}}(X \geq 20.5 \mid \mu = 15.0, \sigma = 3.87)$$

(*Nota:* se incluye la corrección por continuidad.)

$$z = \frac{X - \mu}{\sigma} = \frac{20.5 - 15.0}{3.87} = \frac{5.50}{3.87} = +1.42$$

$$P(X \geq 20.5) = P(z \geq +1.42) = 0.5000 - 0.4222 - 0.0778 \quad \text{(del apéndice 5)}$$

APLICACIONES EN COMPUTADORA

7.23 Del problema 7.1, la longitud promedio de los tornillos es $\mu = 13.0$ cm, con $\sigma = 0.1$ cm. Las longitudes tienen una distribución normal. Utilizando un paquete disponible de computación, determine la probabilidad de que la longitud de un tornillo elegido al azar (a) exceda 13.25 cm, y (b) se encuentre entre 12.9 y 13.1 cm.

(a) Con referencia a la figura 7-15, obsérvese que la probabilidad que se reportó inicialmente, 0.9938, es la probabilidad acumulada entre infinito negativo y el valor designado, 13.25. Por lo tanto, la lógica que se sigue en la parte restante del resultado, consiste en obtener la respuesta que se requiere mediante sustracción de la siguiente manera:

$$\begin{aligned} P(X > 13.25) &= 1.0000 - P(X < 13.25) \\ &= 1.0000 - 0.9938 = 0.0062 \end{aligned}$$

(Esta solución corresponde a la respuesta que se obtuvo en forma manual en el problema 7.2.)

```
MTB > CDF AT 13.25;
SUB > NORMAL MU = 13.0, SIGMA = 0.1.
     2.50    0.9936
MTB > SUBTRACT .0.9938 FROM 1.0000, PUT ANSWER IN Kl
    ANSWER =         0.0068
```

Fig. 7.15 Resultado de Minitab para el problema 7.23 (a).

(b) Con referencia a la figura 7-16, y restando de nueva cuenta las probabilidades que se incluyen en los resultados:

$$\begin{aligned} P(12.9 < X \leq 13.1) &= P(X < 13.1) - P(X < 12.9) \\ &= 0.8413 - 0.1587 = 0.6826 \end{aligned}$$

(Esta solución corresponde a la respuesta que se obtuvo en forma manual en el problema 7.3.)

```
MTB > CDF AT 12.9;
SUBC> NORMAL MU = 13.0, SIGMA = 0.1.
    -1.00    0.1587
MTB > CDF AT 13. 1;
SUBC> NORMAL MU = 13.0, SIGMA = 0.1.
     1.00    0.8413
MTB > SUBTRACT 0.1587 FROM 0.8413, PUT ANSWER IN Kl
    ANSWER =         0.6826
```

Fig. 7.16 Resultado de Minitab para el problema 7.23 (b).

7.24 Del problema 7.6, el tiempo que se requiere para la reparación de las transmisiones tiene distribución normal, con μ - 45 min y σ = 8.0 min. Utilizando algún programa de computadora, determine el tiempo al (a) punto percentil 75, y (b) punto percentil 30.

(a) Con referencia a la figura 7-17, la respuesta es 50.3959 min. (Esta respuesta corresponde a la que se obtuvo en forma manual en el problema 7.7, excepto por la ligera diferencia que se debe al redondeo).

```
MTB >  INVCDF AT 0.75;
SUBC > NORMAL MU = 45.0, SIGMA = 8.0.
        0.75    50.3959
```

F g. 7.17 Resultado de Minitab para el problema 7.24 (a).

(b) Con referencia a la figura 7-l8, la respuesta es 40.8048 min. (Esta respuesta corresponde a la que se obtuvo en forma manual en el problema 7.8, excepto por la ligera diferencia debida al redondeo.)

```
MTB >  INVCDF AT 0.30;
SUBC> NORMAL MU = 45.0, SIGMA = 8.0.
        0.30    40.3048
```

Fig. 1.18 Resultado de Minitab para el problema 7.24 (b).

Problemas complementarios

LA DISTRIBUCIÓN NORMAL DE PROBABILIDAD

7.25 El promedio de estudiantes inscritos en jardines de niños es de μ = 500, con desviación estándar σ = 100. El número de alumnos inscritos tiene una distribución aproximadamente normal. ¿Cuál es la probabilidad de que el número de alumnos inscritos en una escuela elegida al azar esté (a) entre 500 y 650?, (b) ¿Entre 450 y 600?

Resp. (a) 0.4332, *(b)* 0.5328

7.26 Para los datos de las escuelas correspondientes al problema 7.25, ¿cuál es la probabilidad de que un jardín de niños elegido al azar tenga (a) menos de 300 alumnos inscritos? (b) ¿Más de 650?

Resp. (a) 0.0228, *(b)* 0.0668

7.27 Se ha determinado que la vida útil de cierta marca de llantas radiales tiene una distribución normal con μ = 38 000 kilómetros y σ = 3 000 kilómetros, (a) ¿Cuál es la probabilidad de que una llanta elegida al azar tenga una vida útil de cuando menos 35 000 kilómetros? *(b)* ¿Cuál es la probabilidad de que dure más de 45, kilómetros?

Resp. (a) 0.8413. *(b)* 0.0099

7.28 Un distribuidor hace un pedido de 500 de las llantas especificadas en el problema 7.27. ¿Aproximadamente cuántas llantas durarán (a) entre 40 000 y 45 000 kilómetros? *(b)* ¿40 000 kilómetros o más?

Resp. (a) 121, (b)126

7.29 Una persona compra 4 de las llantas que se describieron en el problema 7.27. ¿Cuál es la probabilidad de que las 4 llantas duren (a) cuando menos 38 000 kilómetros?, (b) ¿Cuando menos 35 000 kilómetros? *(Sugerencia:* después de obtener la probabilidad para una llanta, utilice la regla de la multiplicación para eventos independientes que se vio en la sección 5.6 para determinar la probabilidad para las 4 llantas,

Resp. (a) 0.0625, (6) 0.5010

7.30 Se ha encontrado que el tiempo de servicio que se requiere por persona en una caja bancaria tiene una distribución aproximadamente normal con media μ = 130 segundos y σ - 45 segundos. ¿Cuál es la probabilidad de que una persona elegida al azar (a) requiera menos de 100 segundos para terminar sus transacciones? (b) ¿Pase entre 2.0 y 3.0 minutos en la caja bancaria?

Resp. (a) 0.2514, (b) 0.4538

PUNTOS PERCENTILES PARA VARIABLES DISTRIBUIDAS NORMALMENTE

7.31 El número de alumnos inscritos en un conjunto de jardines de niños de determinada zona tiene media μ = 500 y desviación estándar σ 100. El número de alumnos inscritos tiene una distribución normal. ¿Qué número de alumnos inscritos se encuentra en el (a) punto percentil 50, (b) punto percentil 30 y (c) punto percentil 90?

Resp. (a) 500, *(b)* 448, (c) 628

7.32 Bajo las condiciones que se especificaron en el problema 7.30, (a) ¿qué tiempo necesitan el 20% de las personas con las transacciones más simples para terminar sus operaciones en la caja?, *(b)* ¿cuando menos qué tiempo se requiere para las personas que se encuentran en el 5% más alto del tiempo?

Resp. (a) 92 *sec, (b)* 204 se c

APROXIMACIÓN NORMAL A LAS PROBABILIDADES BINOMIAL Y POISSON

7.33 Para los varios millares de artículos que se mantienen en existencia en una empresa, existe una probabilidad global del 0.08 de que un artículo específico (incluyendo tamaño y color determinados, etcétera) no se encuentre en existencia. Si un embarque cubre los pedidos para 120 artículos distintos, ¿cuál es la probabilidad de que 15 o más de ellos no se encuentren en existencia?

Resp. 0.0495

7.34 Para el embarque que se describe en el problema 7.33, ¿cuál es la probabilidad de que haya entre 10 y 15 artículos que no se encuentren en existencia?

Resp. 0.4887

7.35 En el periodo más ocupado, entre las 4 PM y 6 PM, un automóvil entra a una gasolinería cada 3 minutos, en promedio. ¿Cuál es la probabilidad de que cuando menos 25 automóviles entren a la gasolinería entre las 4 y las 5 PM.

Resp. 0.1562

7.36 Para las llegadas de automóviles a la gasolinería que se describieron en el problema 7.35, ¿cuál es la probabilidad de que entren menos de 30 automóviles entre las 4 y las 6 PM en un día elegido al azar?

Resp. 0.0485

LA DISTRIBUCIÓN EXPONENCIAL DE PROBABILIDAD

7.37 En promedio, 6 personas utilizan un cajero bancario automático cada hora, durante las principales horas comerciales de una tienda de departamentos.

(a) ¿Cuál es la probabilidad de que cuando menos pasen 10 minutos entre las llegadas de dos clientes?

(b) ¿Cuál es la probabilidad de que después de que salga un cliente, no llegue otro por cuando menos 20 minutos?

(c) ¿Cuál es la probabilidad de que llegue un segundo cliente antes de que pase un minuto después de que el primer cliente comienza su transacción bancaria?

Resp. (a) 0.36788, (b) 0.13534, (c) 0.09516

7.38 Suponga que el manuscrito para un libro de texto tiene un total de 50 errores de mecanografía incluidos en las 500 páginas del material, y que los errores están distribuidos en forma aleatoria a lo largo del texto. Al comenzar a hacer la revisión ortográfica de un capitulo específico, ¿cuál es la probabilidad de que el primer error en ese capítulo (a) esté incluido dentro de las primeras cinco páginas?, (o) ¿ocurra después de las primeras 15 páginas?

Resp. (a) 0.39347 *(b)* 0.22313

PROBLEMAS DIVERSOS

(Nota: En los problemas 7.39 a 7.46 se utilizan todas las distribuciones de probabilidad que se cubrieron en los capítulos 6 y 7.)

7.39 Se ha encontrado que la distribución de frecuencias de las estancias en un hospital es aproximadamente simétrica y mesokúrtica, con μ - 8.4 y σ - 2.6 días (en donde se han medido las fracciones de día). ¿Cuál es la probabilidad de que una persona elegida al azar permanezca en el hospital (a) menos de 5.0 días?, (b) ¿Más de 8.0 días?

Resp. (a) 0.0951, (b) 0.5596

7.40 Una empresa que fabrica y comercializa una amplia gama de juguetes novedosos de precio bajo (tales como una pelota que rebota en dirección inesperada) ha encontrado que, a largo plazo, el 40% de ios juguetes que fabrica tiene cuando menos un éxito moderado de mercado. Si se han desarrollado 6 nuevos juguetes para su introducción en el mercado el siguiente verano, ¿cuál es la probabilidad de que cuando menos 3 de ellos tenga un éxito moderado en el mercado?

Resp. 0.4557

7.41 La empresa del problema 7.40 tiene 60 ideas para juguetes en proceso de desarrollo para su introducción en los siguientes años. Si en algún momento dado se comercializan todos los 60 juguetes, ¿cuál es la probabilidad de que cuando menos 30 de ellos tengan un éxito moderado en el mercado?

Resp. 0.0735

7.42 De los problemas 7.40 y 7.41, suponga que el 5% de los juguetes que se comercializan resultan ser éxitos definitivos de venta. Si se introducen en el mercado 60 nuevos juguetes en los siguientes años, ¿cuál es la probabilidad de que ninguno de ellos resulte ser un éxito definitivo de ventas?

Resp. 0.0498

7.43 En promedio, llegan dos clientes por minuto a un puesto de refrescos de un estadio deportivo. ¿Cuál es la probabilidad de que 5 o más clientes lleguen al puesto de refrescos en un minuto elegido al azar?

Resp. 0.0526

7.44 Del problema 7.43, ¿cuál es la probabilidad de que después de abrir el puesto de refrescos pasen 2 minutos completos antes de que llegue el primer cliente?

Resp. 0.01832

7.45 Para la situación que se describe en el problema 7.43, ¿Cuál es la probabilidad de que lleguen al puesto más de 5 personas durante un periodo de media hora?

Resp. 0.8907

7.46 De los 8 hoteles ubicados en una zona recreativa, puede decirse que tres de ellos son mediocres en términos de servicios al cliente. Un agente de viajes elige al azar dos hoteles para dos clientes que están planeando vacacionar en esa región. ¿Cuál es la probabilidad de que cuando menos uno de los clientes se hospede en uno de los hoteles mediocres?

Resp. 0.6429

APLICACIONES EN COMPUTADORA

7.47 Del problema 7.26, el número de alumnos inscritos en los jardines de niños de una región determinada tiene una distribución normal, con $\mu = 500$ y $\sigma = 100$. Utilizando algún programa de computación disponible, determine la probabilidad de que una escuela elegida al azar tenga (a) menos de 300 y (b) más de 650 alumnos inscritos.

Resp. (a) 0.0228, (b) 0.0668 (lo cual corresponde a las soluciones manuales del problema 7.26)

7.48 Del problema 7.31, el número de alumnos inscritos en jardines de niños tienen una distribución normal con $\mu = 500$ y $\sigma = 100$. Utilizando un programa de computación, determine el valor en el (a) punto percentil 30 y (b) punto percentil 90.

Resp. (a) 448, (b) 628 (que corresponden a las soluciones manuales del problema 7.31)

Distribuciones de muestreo e intervalos de confianza para la media

8

8.1 ESTIMACIÓN PUNTUAL

Con frecuencia se estiman los parámetros de una población con base en estadísticas muestrales debido a factores como tiempo y costo.Tal como se definió en la sección 1.2, un *parámetro poblacional* es una medida de resumen de una población, en tanto que a una medida de resumen de una muestra se le denomina *estadística muestra!.* Con el objeto de utilizar los datos muestrales para realizar inferencias estadísticas, incluyendo estimaciones, la muestra que se tome debe ser una muestra *aleatoria,* tal como se describió en la sección 1.6.

EJEMPLO 1. La media μ y la desviación estándar σ de una población de mediciones son parámetros poblacionales. La media $\overline{X}$ y la desviación estándar s de una muestra de mediciones son estadísticas muestrales.

Un *estimador puntual* es un solo valor numérico basado en datos de una muestra aleatoria que se utiliza para estimar el valor de un parámetro poblacional. Una de las características más importantes de las estadísticas muestrales que se utilizan como estimadores es que no son sesgados. Un *estimador no sesgado* es un estadístico muestral cuyo valor esperado es igual al parámetro que se estima. Tal como se explicó en la sección 6.2, un *valor esperado* es el promedio a largo plazo de la estadística muestral.

En la Tabla 8.1 se presentan algunos estimadores puntuales de parámetros poblacionales que se utilizan con frecuencia. En todos los casos, el estimador apropiado de un parámetro poblacional es simplemente el correspondiente estadístico muestral. Sin embargo, se debe notar que, en la sección 4.5, la fórmula 4.5 de la varianza muestral incluye "un factor de corrección*. Sin esta corrección, la varianza muestral serla un estimador sesgado de la varianza poblacional.

Tabla 8.1 Estimadores puntuales utilizados con frecuencia

Parámetro de la población	Estimador
Media, μ	$\overline{X}$
Diferencia entre las medias de dos poblaciones; $\mu_1 - \mu_2$	$\overline{X}_1 - \overline{X}_2$
Proporción, π	$\hat{p}$
Diferencia entre las proporciones de dos poblaciones, $\pi_1 - \pi_2$	$\hat{p}_1 - \hat{p}_2$
Varianza, σ^2	s^2
Desviación estándar, σ	s

8.2 DISTRIBUCIÓN MUESTRAL DE LA MEDIA

Una distribución poblacional representa la distribución de valores de una población y una distribución muestral representa la distribución de los valores de una muestra. En contraste con las distribuciones de mediciones individuales, una *distribución* *muestral* es una distribución de probabilidad que se aplica a los valores posibles de una estadística muestral. Así, la *distribución muestral de la media* es la distribución de probabilidad de los valores posibles de la media muestı $\bar{X}$, con base en un determinado tamaño de muestra.

Para cualquier tamaño de muestra dado *n,* tomado de una población con media μ, los valores de la media muestral $\bar{X}$ varían de una muestra a otra. Esta variabilidad sirve de base para la distribución muestral. La distribución muestral de la media se describe determinando el valor esperado E $(\bar{X})$ o media, de la distribución y la desviación estándar de la distribución de las medias, $\sigma_{\bar{x}}$. Como esta desviación estándar indica la precisión de la media muestral como estimador puntual, por lo general se le denomina *error estándar de la media.* En general, se define el valor esperado de la media y el error estándar de la media de la siguiente manera:

$$E(\bar{X}) = \mu \tag{8.1}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{8.2}$$

EJEMPLO 2. Suponga que la media de una población muy grande es μ= 50.0 y que la desviación estándar es σ =12.0. Se determina la distribución muestral de las medias para una muestra de tamaño *n* = 36, en términos del valor esperado y del error estándar de la distribución de la siguiente manera

$$E(\bar{X}) = \mu = 50.0$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12.0}{\sqrt{36}} = \frac{12.0}{6} = 2.0$$

Cuando se muestrea a partir de una población finita, se debe incluir un *factor de corrección por población finita* en la fórmula para el error estándar de la media. Como regla general, la corrección es despreciable y puede omitirse cuando $n < 0.05N$, es decir, cuando el tamaño de la muestra es menos del 5% del tamaño de la población. Muchos textos y programas de computación no incluyen esta corrección porque suponen que la población siempre es muy grande, o quizá de tamaño infinito. La fórmula para el error estándar de la media, incluyendo el factor de corrección por población finita, es:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \tag{8.3}$$

Si no se conoce la desviación estándar de la población, puede estimarse el error estándar de la media utilizando la desviación estándar muestral como estimador de la desviación estándar de la población. Para diferenciar este error estándar del que se basa en una σ conocida, se le designa mediante el símbolo $s_{\bar{x}}$ (o mediante $\hat{\sigma}_{\bar{x}}$ en algunos textos). La fórmula del error estándar estimado de la media es:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \tag{8.4}$$

La fórmula del error estándar estimado de la media, incluyendo el factor de corrección por población finita es:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \tag{8.5}$$

Ejemplo 3. Un auditor toma una muestra aleatoria de tamaño $n = 16$ de un conjunto de $N = 100$ cuentas por cobrar. No se conoce la desviación estándar de los montos de las cuentas por cobrar para el total de las 100 cuentas. Sin embargo, la desviación estándar de la muestra es s = \$57.00. Se determina el valor del error estándar para la distribución muestral de la media de la siguiente manera:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{57.00}{\sqrt{16}} \sqrt{\frac{100-16}{100-1}} = \frac{57}{4} \sqrt{\frac{84}{99}}$$

$$= 14.25\sqrt{0.8484} = 14.25(0.9211) = 13.126 \cong \$13.13$$

En este ejemplo se estima el error estándar de la media con base en la desviación estándar muestral, y se requiere utilizar el factor de corrección por población finita porque *no* es cierto que $n < 0.05\,N$, es decir, $16 > 0.05(100)$.

El error estándar de la media ofrece la base principal para la inferencia estadística con respecto a la media de una población que se desconoce, tal como se presenta en este capítulo y los siguientes. Un teorema de la estadística que conduce a la utilidad del error estándar de la media es

Teorema del límite central: Al aumentar el tamaño de la muestra, la distribución muestral de la media se aproxima a la forma de la distribución normal *sin importar la forma de la distribución de las mediciones individuales de la población.* Para propósitos prácticos, puede suponerse que la distribución muestral de la media es aproximadamente normal cuando el tamaño de la muestra es $n \geq 30$.

Por ello, si se tiene una muestra "grande" de ≥ 30, puede utilizarse siempre la distribución normal de probabilidad junto con el error estándar de la media. Además, si la población tiene distribución normal y se conoce σ, puede utilizarse la distribución normal para hacer inferencias estadísticas a partir de muestras pequeñas. En la sección 8.5 se explica por qué se requiere conocer σ.

EJEMPLO 4. Un auditor toma una muestra aleatoria de tamaño $n = 36$ de una población de 1 000 cuentas por cobrar. El valor promedio de las cuentas por cobrar de la población es $\mu = \$\,2\,600$ con una desviación estándar poblacional de $\sigma = \$\,450$. ¿Cuál es la probabilidad de que la media muestral sea inferior a \$ 2 500?

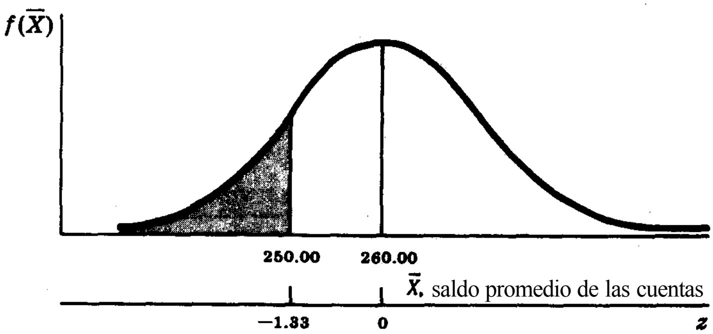


Fig. 8-1

En la figura 8-1 se ilustra la correspondiente curva de probabilidad. Se describe la distribución muestral mediante la media y el error estándar

$$E(\bar{X}) = \mu = 260.00 \qquad \text{(dado)}$$
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{45.00}{\sqrt{36}} = \frac{45.00}{6} = 7.50$$

[Note: no se requiere el factor de corrección por población finita porque $n < 0.05\,N$, es decir, 36 es menor que 0.05 (1 000).]

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{250.00 - 260.00}{7.50} = \frac{-10.00}{7.50} = -1.33$$

Por lo tanto,

$$P(\bar{X} < 250.00 \mid \mu = 260.00, \sigma_{\bar{x}} = 7.50) = P(z < -1.33)$$
$$P(z < -1.33) = 0.5000 - P(-1.33 \le z \le 0)$$
$$= 0.5000 - 0.4082 = 0.0918$$

EJEMPLO 5. Con referencia al ejemplo 4, ¿cuál es la probabilidad de que la media muestral se encuentre a no más de $ 150 de la media de la población?

En la figura 8-2 se ilustra la curva de probabilidad para la distribución muestral.

$$P(245.00 \le \bar{X} \le 275.00 \mid \mu = 260.00, \sigma_{\bar{x}} = 7.50)$$
$$z_1 = \frac{245.00 - 260.00}{7.50} = -2.00$$

donde

$$z_2 = \frac{275.00 - 260.00}{7.50} = +2.00$$
$$P(-2.00 \le z \le +2.00) = 0.4772 + 0.4772 = 0.9544 \approx 95\%$$

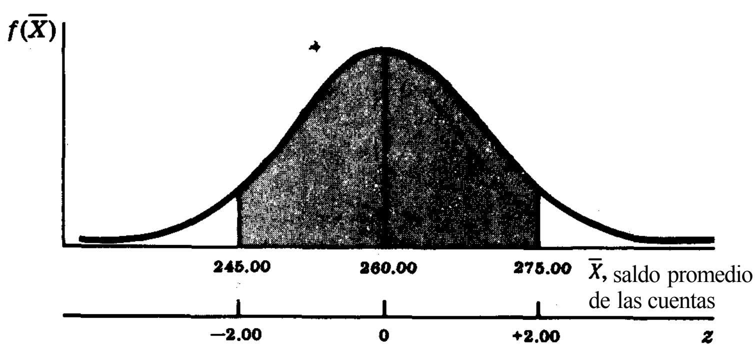


Fig. 8-2

8.3 INTERVALOS DE CONFIANZA PARA LA MEDIA UTILIZANDO LA DISTRIBUCIÓN NORMAL

Los ejemplos 4 y 5 ilustran la determinación de la probabilidad de que la media muestral tenga diversos valores cuando se conocen la media y la desviación estándar de la población. Lo que está implícito aquí, es *el razonamiento deductivo* (que parte de una verdad general y llega a un caso especifico) con respecto al resultado muestral y con base en parámetros poblacionales conocidos. Se pasa ahora a revisar el *razonamiento inductivo* (que parte de un caso particular para concluir sobre una verdad general) utilizando datos muestrales para hacer afirmaciones acerca del valor de la media poblacional.

Los métodos de estimación por intervalo que se revisan en esta sección se basan en el supuesto de que puede utilizarse la distribución normal de probabilidad. Tal como se analiza en las secciones 8.2 y 8.5, esta suposición es válida (1) cuando $n \geq 30$, debido al teorema del limite central, o (2) cuando $n < 30$, pero la población tiene distribución normal y se conoce σ.

Aunque la media muestral es útil como estimador no sesgado de la media de la población, no hay forma de expresar el grado de precisión de un estimador puntual. De hecho, en términos matemáticos, la probabilidad de que la media muestral sea *exactamente* correcta como estimador es $P = 0$. *Un intervalo de confianza* para la media es un estimador de intervalo que se construye con respecto a la media muestral y que permite especificar la probabilidad de que incluya el valor de la media poblacional. El *grado de confianza* asociado con un intervalo de confianza señala el porcentaje a largo plazo de esa clase de intervalos que incluirían el parámetro que se estima.

Por lo general, se construyen los intervalos de confianza utilizando el estimador no sesgado $\bar{X}$ como punto medio del intervalo. Sin embargo, en los problemas 8.14 y 8.15 se ilustra la construcción de un intervalo de confianza que se denomina "en un sentido", y para el cual la media muestral no es el punto medio, del intervalo. Cuando puede utilizarse la distribución normal de probabilidad, el intervalo de confianza para la media se determina mediante:

$$\bar{X} \pm z\sigma_{\bar{x}} \qquad (8.6)$$

$$\bar{X} \pm zs_{\bar{x}} \qquad (8.7)$$

Los intervalos de confianza que se utilizan con mayor frecuencia son los de 90,95 y 99%. En la Tabla 8.2 se presentan los valores de z que se requieren para esos intervalos.

Tabla 8.2 Proporciones seleccionadas de áreas bajo la curva normal

z (número de unidades de desviación estándar desde la media)	Proporción de área en el intervalo $\mu \pm z\sigma$
1.645	0.90
1.96	0.95
2.58	0.99

EJEMPLO 6. En una semana determinada, se elige al azar una muestra de 300 empleados de un número muy grande de ellos que trabajan en una empresa manufacturera. Los trabajadores realizan una labor a destajo y se encuentra que el promedio de pago por pieza trabajada es de $\bar{X} = \$1800$, con una desviación estándar muestral de s = $140. Se estima que el pago promedio a destajo para todos los empleados de la empresa, con una estimación por intervalo que permita tener una confianza del 95% de que ese intervalo incluye el valor de la media poblacional, es:

$$\bar{X} \pm 1.96s_{\bar{x}} = 180.000 \pm 1.96(2.56) = \$174.98 \text{ a } \$185.02$$

en donde $\bar{X} = \$180.00$ (dado)

$s_{\bar{x}} = s/\sqrt{n} = 14.00/\sqrt{30} = 2.56$

(Nota: se utiliza s como estimador de o.)

Por ello, puede afirmarse que el pago promedio a destajo para todos los empleados se encuentra entre $174 980 y $185 020 con un grado de confianza del 95% en esa estimación.

Además de estimar el valor de la media poblacional como tal, en ocasiones interesa estimar la cantidad o monto total de la población . Véase el problema 8.11 (b).

8.4 DETERMINACIÓN DEL TAMAÑO DE MUESTRA NECESARIO PARA ESTIMAR LA MEDIA

Supóngase que se especifica el tamaño que se desea en un intervalo de confianza y el grado de confianza correspondiente. Si se conoce σ o puede estimarse, como por ejemplo a partir de resultados de estudios similares, el tamaño de la muestra que se requiere, con base en la distribución normal, es:

$$n = \left(\frac{z\sigma}{E}\right)^2 \tag{8.8}$$

En la fórmula *(8.8), z* es el valor que se utiliza para el grado de confianza especificado, σ es la desviación estándar de la población (o un estimador) y *E* es un factor de error "más o menos" que se permite en el intervalo (que siempre es la mitad del intervalo total de confianza). *(Nota*: cuando se determina el tamaño de la muestra, cualquier resultado fraccionario siempre se redondea hacia arriba. Además, si el tamaño de la muestra que se calcula está por debajo de 30, debe incrementarse a esta cantidad, porque la fórmula *(8.8)* se basa en el uso de la distribución normal (esta regla se aplica a menos que se conozca σ y la población tenga una distribución normal).)

EJEMPLO 7. Un empleado de un departamento de personal desea estimar el número promedio de horas de capacitación que se dan a los supervisores de una división de la compañía, con un error de (más o menos) 3.0 horas y con una confianza del 90%. Con base en los datos de otras divisiones, estima que la desviación estándar de las horas de capacitación es σ = 20.0 horas. El tamaño mínimo de la muestra que se requiere es

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left[\frac{(1.645)(20.0)}{3.0}\right]^2 = \left(\frac{32.9}{3.0}\right)^2 = 120.27 \approx 121$$

8.5 LA DISTRIBUCIÓN *t* DE STUDENT Y LOS INTERVALOS DE CONFIANZA PARA LA MEDIA

En la sección 8.3 se señaló que el uso de la distribución normal en la estimación de una media poblacional es válida para cualquier muestra grande *(n* ≥ 30), y para una muestra pequeña *(n* < 30) sólo si la población tiene distribución normal *y* se conoce o . En esta sección, se revisa el caso en el que la muestra es pequeña y la población tiene una distribución normal, pero se desconoce σ.

Si una población tiene distribución normal, la distribución muestral de la media para cualquier tamaño de muestra tiene también distribución normal; esto es cierto independientemente de si se conoce σ o no. Sin embargo, en el proceso de inferir, se convierte cada valor de la media en un valor normal estándar, *y aquí radica el problema.* Si se desconoce σ, la fórmula de conversión $(\bar{X} - \mu)/s_{\bar{x}}$ incluye una variable en el denominador porque s y, por lo tanto, $s_{\bar{x}}$, diferirá de una muestra a otra. Esto da como resultado que, al incluir la variable $s_{\bar{x}}$ en vez de la constante $\sigma_{\bar{x}}$ en el denominador, se obtienen variables convertidas que *no* se distribuyen como los valores z. En cambio, los valores se distribuyen de acuerdo con la distribución t de Student, que es platikúrtica (aplastada) en comparación con la distribución normal. En el apéndice 6 se incluyen las proporciones de área bajo la distribución r, en donde la distribución específica se basa en los grados de libertad *(gl)* correspondientes a la situación. Para el caso de una sola muestra, *gl* = *n* - 1 .

La distribución es apropiada para realizar inferencias sobre la media cuando se desconoce o y la población tiene una distribución normal, sin importar el tamaño de la muestra. Sin embargo, al aumentar el tamaño de la muestra (y los grados de libertad), la distribución f se aproxima a la forma de la distribución normal. Una regla general es que puede aproximarse una distribución f mediante la distribución normal cuando n ≥ 30 (o los *gl* ≥ 29) para una sola muestra. Esta sustitución es distinta de la que se especificó mediante el teorema de limite central, y es una conveniencia que en ambos casos se requiera que la muestra sea cuando menos de tamaño n < 30.

Debe observarse que los valores de f que se reportan en el apéndice 6, indican la proporción del "extremo" superior de la distribución, y no la proporción entre la media y un punto dado, como fue el caso del apéndice 5 de la distribución normal. Cuando $gl = n - 1$, el intervalo de confianza para estimar la media de la población cuando se desconoce o, n < 30 y la población tiene una distribución normal es:

$$\bar{X} \pm t_{gl}\, s_{\bar{x}} \tag{8.9}$$

EJEMPLO 8. La vida útil promedio de una muestra aleatoria de n = 10 focos es X =4000 horas, con una desviación estándar muestral s = 200 horas. Se supone que la vida útil de los focos tiene una distribución aproximadamente normal. Se estima la vida útil promedio de la población de focos de la cual se tomó la muestra, utilizando un intervalo de confianza del 95% de la siguiente manera:

$$\begin{aligned} \text{Int. } 95\% &= \bar{X} \pm t_{gl}\, s_{\bar{x}} \\ &= 4000 \pm (2.262)(63.3) \\ &= 3856.8 \text{ a } 4143.2 = 3857 \text{ a } 4\,143 \text{ hr} \end{aligned}$$

en donde

$$\begin{aligned} \bar{X} &= 4000 \text{ (dado)} \\ t_{gl} &= t_{n-1} = t_9 = 2.262 \\ s_{\bar{x}} &= \frac{s}{\sqrt{n}} = \frac{200}{\sqrt{10}} = \frac{200}{3.16} = 63.3 \end{aligned}$$

8.6 TABLA RESUMEN PARA LA ESTIMACIÓN POR INTERVALO DE LA MEDIA DE POBLACIÓN

Tabla 8.3 Estimación por Intervalo de la media poblacional

Población	Tamaño de la muestra	Se conoce σ	Se desconoce σ
Con distribución normal	Grande ($n \geq 30$)	$\bar{X} \pm z\sigma_{\bar{x}}$	$\bar{X} \pm zs_{\bar{x}}$**
	Pequeña ($n < 30$)	$\bar{X} \pm z\sigma_{\bar{x}}$	$\bar{X} \pm ts_{\bar{x}}$
Sin distribución normal	Grande ($n \geq 30$)	$\bar{X} \pm z\sigma_{\bar{x}}$*	$\bar{X} \pm zs_{\bar{x}}$†
	Pequeña ($n < 30$)	Por lo general, se utilizarían procedimientos no paramétricos basados en la mediana (véase el capítulo 21).	

* Se utiliza el teorema del límite central.

** Se utiliza z como aproximación de t

† Se utiliza el teorema del límite central y se utiliza z como aproximación de t.

8.7 RESULTADOS POR COMPUTADORA

Los programas y paquetes de computación que se utilizan para realizar análisis estadísticos, por lo general permiten al usuario especificar el porcentaje de confianza que se desea para el intervalo de la media, con base en datos muestrales aleatorios. En los problemas 8.18 y 8.19 se ilustra el uso de la computadora para determinar intervalos de confianza para la media de la población.

Problemas resueltos

DISTRIBUCIÓN MUESTRAL DE LA MEDIA

8.1 Se sabe que la vida útil promedio de los cinescopios de una marca específica de televisores es μ= 9000 horas, con una desviación estándar de σ = 500 horas. Determine el valor esperado y el valor estándar de la distribución muestral de la media, con un tamaño de muestra de n = 25. Interprete el significado de los valores calculados.

$$E(\bar{X}) = \mu = 9000$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{25}} = \frac{500}{5} = 100$$

Estos cálculos indican que, a largo plazo, la media de un conjunto grande de medias muestrales, todas ellas basadas en un tamaño de muestra de n = 25, será igual a 9000 horas. Además, la variabilidad de esas medias muestrales con respecto al valor esperado de 9000 horas se expresa a través de una desviación estándar de 100 horas.

8.2 Para una población grande de saldos de cuentas que tienen distribución normal, se tiene un saldo promedio de μ = \$150 000, con desviación estándar σ = \$35 000. ¿Cuál es la probabilidad de que una cuenta muestreada al azar tenga un saldo que excede de \$160 000?

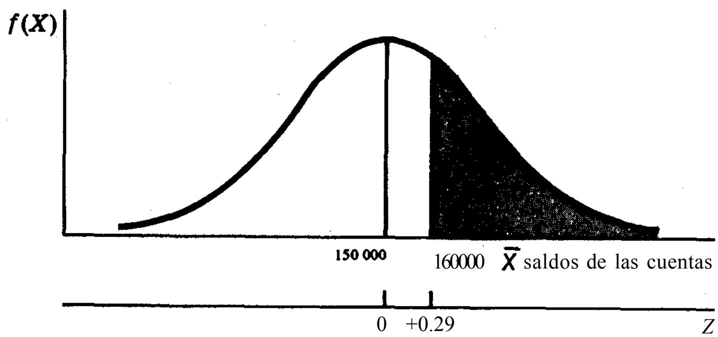


Fig. 8-3

En la figura 8-3 se ilustra la curva de probabilidad para la variable.

$$z = \frac{X - \mu}{\sigma} = \frac{160\,000 - 150\,000}{35\,000} = +0.29$$

$$P(X > 160\,000 \mid \mu = 150\,000, \sigma = 35\,000) = P(z > +0.29)$$

$$= 0.5000 - P(0 \leq z \leq +0.29) = 0.5000 - 0.1141 = 0.3859$$

8.3 Con referencia al problema 8.2, ¿cuál es la probabilidad de que la *media* de una muestra aleatoria de $n = 40$ cuentas exceda de $160 000?

En la figura 8.4 se ilustra la curva de probabilidad para la distribución muestral $E(\bar{X}) = \mu = \$150\ 000$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{35\ 000}{\sqrt{40}} = \$5\ 533$$

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{160\ 000 - 150\ 000}{5\ 533} = \frac{10\ 000}{5\ 533} = +1.81$$

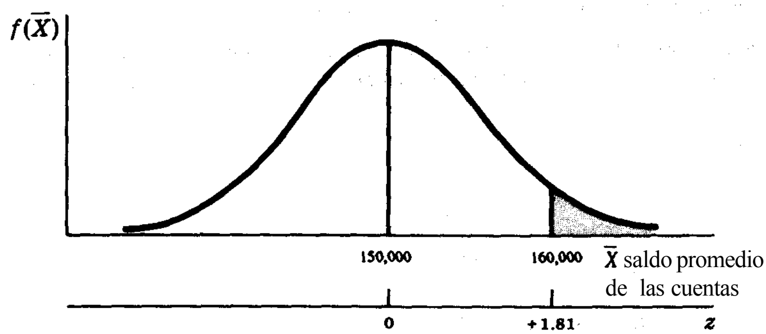


Fig. 8-4

Por lo tanto,

$$\begin{aligned} P(\bar{X} > 160\ 000 \mid \mu = 150\ 000, \sigma_{\bar{x}} = 5\ 533 &= P(z > +1.81) \\ &= 0.5000 - P(0 \leq z \leq +1.81) \\ &= 0.5000 - 0.4649 = 0.0351 \end{aligned}$$

8.4 Este problema y los dos siguientes sirven para ilustrar el significado de la distribución muestral de la media, haciendo referencia a una población altamente simplificada. Suponga que una población consta de solamente los cuatro valores 3, 5, 7 y 8. Calcule (a) la media de la población μ, y *(b)* la desviación estándar de la población σ.

Con referencia a la Tabla 8.4.

(a) $\mu = \frac{\Sigma X}{N} = \frac{23}{4} = 5.75$

(b) $\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2} = \sqrt{\frac{147}{4} - \left(\frac{23}{4}\right)^2}$

$= \sqrt{36.75 - (5.75)^2} = \sqrt{36.75 - 33.0625} = 1.92$

Tabla 8.4 Hoja de trabajo para el problema 8.4

X	X^2
3	9
5	25
7	49
8	64
$\Sigma X = 23$	$\Sigma X^2 = 147$

8.5 Para la población que se describe en el problema 8.4, suponga que se toman muestras aleatorias simples de tamaño n -=2, de esa población. En cada una de las muestras, antes de elegir el segundo elemento muestral, *no* se reemplaza el primer elemento escogido

(a) Enliste todos los pares posibles de valores que puede constituir una muestra.

(b) Para cada uno de los pares identificados en (a), calcule la media muestral $\bar{X}$ y demuestre que la media de todas las medias muestrales posibles $\mu_{\bar{X}}$ es igual a la media de la población de donde se seleccionaron las muestras.

(a) y (b) de la Tabla 8.5,

$$\mu_{\bar{X}} = \frac{\Sigma \bar{X}}{N_{\text{muestras}}} = \frac{34.5}{6} = 5.75$$

[que es igual a la μ que se calculó en el problema 8.4 (a)]

Tabla 8.5 Muestras posibles y medias muestrales para el problema 8.5

Muestras posibles	$\bar{X}$
3,5	4.0
3,7	5.0
3,8	5.5
5,7	6.0
5,8	6.5
7,8	7.5
	$\Sigma \bar{X} = 34.5$

8.6 Para la situación de muestreo que se describió en los problemas 8.4 y 8.5, calcule el error estándar de la media determinando la desviación estándar de las seis medias muestrales posibles que se identificaron en el problema 8.5, con respecto a la media poblacional μ. Después, calcule el error estándar de la media con base en la o que se conoce y, tratándose de un muestreo en una población finita, utilice la fórmula apropiada de las que se revisaron en este capítulo. Verifique que los dos valores del error estándar sean iguales.

Con referencia a la Tabla 8.5,

Tabla 8.6 Hoja de trabajo para el problema 8.6

$\bar{X}$	$\bar{X}^2$
4.0	16.00
5.0	25.00
5.5	30.25
6.0	36.00
6.5	42.25
7.5	56.25
$\Sigma \bar{X} = 34.5$	$\Sigma \bar{X}^2 = 205.75$

Primer método:

$$\sigma_{\bar{x}} = \sqrt{\frac{\Sigma \bar{X}^2}{N_S} - \left(\frac{\Sigma \bar{X}}{N_S}\right)^2} = \sqrt{\frac{205.75}{6} - \left(\frac{34.5}{6}\right)^2}$$
$$= \sqrt{34.2917 - (5.75)^2} = \sqrt{34.2917 - 33.0625} = 1.11$$

Segundo método:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{1.92}{\sqrt{2}} \sqrt{\frac{4-2}{4-1}}$$
$$= \frac{1.92}{1.414} \sqrt{0.6666} = 1.358(0.816) = 1.11$$

Por supuesto, el segundo método es el que comúnmente se utiliza para determinar el error estándar de la media cuando se trabaja con datos reales; pero conceptualmente, el primer método ilustra en forma más directa el significado del error estándar de la media.

INTERVALOS DE CONFIANZA PARA LA MEDIA UTILIZANDO LA DISTRIBUCIÓN NORMAL

8.7 Suponga que se sabe que la desviación estándar de la vida útil de los cinescopios de una marca específica de televisores es $\sigma = 500$, p aro que no se conoce el promedio de vida útil. En términos generales, se supone que la vida útil de los cinescopios tiene una distribución aproximadamente normal. Para una muestra de n = 15, la vida útil promedio es $\bar{X} = 8\ 900$ horas. Construya intervalos de confianza para estimar la media de la población (a) con el 95% y (b) el 90% de confianza.

En este caso puede utilizarse la distribución normal porque la población tiene distribución normal y se conoce o.

(a) $\bar{X} \pm z\sigma_{\bar{x}} = 8\ 900 \pm 1.96 \dfrac{\sigma}{\sqrt{n}}$

$$= 8\ 900 \pm 1.96 \frac{500}{\sqrt{15}} = 8{,}900 \pm 1.96 \left(\frac{500}{3.87}\right)$$
$$= 8\ 900 \pm 1.96(129.20) = 8\ 647 \text{ a } 9\ 153$$

(b) $\bar{X} \pm z\sigma_{\bar{x}} = 8\ 900 \pm 1.645(129.20) = 8\ 687$ a $9\ 113$ hr

8.8 Con respecto al problema 8.7, suponga que no puede asumirse que la vida útil de la población de cinescopios tiene distribución normal. Sin embargo, la media muestral de $\bar{X}$ = 8 900 se basa en una muestra de n = 35. Construya el intervalo de confianza del 95% para estimar la media de la población.

En este caso, puede utilizarse la distribución normal de probabilidad utilizando el teorema del límite central, que señala que cuando $n \geq 30$ puede asumirse que la distribución muestral tiene una distribución normal, aun cuando la población no tenga distribución normal. Por ello,

$$\bar{X} \pm z\sigma_{\bar{x}} = 8\,900 \pm 1.96\frac{\sigma}{\sqrt{n}} = 8\,900 \pm 1.96\frac{500}{\sqrt{35}}$$

$$= 8\,900 \pm 1.96\left(\frac{500}{5.92}\right) = 8\,900 \pm 1.96(84.46) = 8\,734 \text{ a } 9\,066 \text{ hr}$$

8.9 Con respecto al problema 8.8., suponga que puede asumirse que la población tiene distribución normal, pero que se desconoce la desviación estándar de la población. Más bien, se sabe que la desviación estándar muestral es s = 500 y $\bar{X}$ = 8 900. Estime la media de la población utilizando un intervalo de confianza del 90%.

Como $n \geq 30$, puede utilizarse la distribución normal para aproximar la distribución f. Sin embargo, como la población tiene distribuciór normal, no es necesario utilizar el teorema de límite central. Por lo tanto,

$$\bar{X} \pm zs_{\bar{x}} = 8\,900 \pm 1.645\left(\frac{500}{\sqrt{35}}\right) = 8\,900 \pm 1.645(84.46) = 8\,761 \text{ a } 9\,039 \text{ hr}$$

8.10 Con respecto a los problemas 8.8 y 8.9, suponga que no *se puede* asumir que la población tiene distribución normal y, además, que tampoco se conoce a. Al igual que antes, n = 35, s =500 y $\bar{X}$ =8, 00. Estime la media de la población Utilizando un intervalo de confianza del 99%.

En este caso, se utiliza el teorema del límite central, al igual que en el problema 8.8, y se utiliza z como aproximación de t, como en el problema 8.9.

$$\bar{X} \pm zs_{\bar{x}} = 8\,900 \pm 2.58\left(\frac{500}{\sqrt{35}}\right) = 8\,900 \pm 2.58(84.46) = 8\,682 \text{ a } 9\,118 \text{ hr}$$

8.11 Un analista de investigación de mercados recopila datos de una muestra aleatoria de 100 clientes, de un conjunto de 400 clientes que adquirieron un "cupón especial". Las 100 personas gastaron un promedio de $\bar{X}$ = \$24 570 en la tienda, con una desviación estándar de s = \$6 600. Utilizando un intervalo de confianza del 95%, estime (a) el monto promedio de las compras para los 400 clientes, y *(b)* el monto total en pesos para las compras realizadas por los 400 clientes.

(a) $$s_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{6.60}{\sqrt{100}}\sqrt{\frac{400-100}{400-1}}$$

$$= \frac{6.60}{10}\sqrt{0.7519} = 0.660(0.867) = 0.57$$

$$\bar{X} \pm zs_{\bar{x}} = 24.57 \pm 1.96(0.57) = \$23.45 \text{ to } \$25.69$$

(b) $$N(\bar{X} \pm zs_{\bar{x}}) = 400(\$23.45 \text{ to } \$25.69) = \$9\,380 \text{ a } \$10\,276$$

o

$$N\bar{X} \pm N(zs_{\bar{x}}) = 400(24.57) \pm 400(1.12)$$

$$= 9\,828 \pm 448 = \$9\,380 \text{ to } \$10\,276$$

(Nota: el intervalo de confianza para el monto total en pesos de las compras realizadas es simplemente el número total de clientes de *la población* multiplicado por los límites de confianza para la cantidad promedio de compra por cliente. A un valor poblacional como éste se le denomina *cantidad total* en algunos textos.)

DETERMINACIÓN DEL TAMAÑO NECESARIO DE LA MUESTRA PARA ESTIMAR LA MEDIA

8.12 Un prospecto de comprador desea estimar el promedio de ventas por cliente (en pesos), en una tienda de juguetes ubicada en un aeropuerto. Con base en datos de otras tiendas similares, se estima que la desviación estándar de ese tipo de ventas es de aproximadamente $\sigma = \$3200$. ¿Oué tamaño de muestra aleatoria se debe utilizar, como mínimo, si desea estimar las ventas promedio con un margen de error de $1000 y un intervalo de confianza del 99%?

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left[\frac{(2.58)(3200)}{1.00}\right]^2 = (8.256)^2 = 68.16 = 69$$

8.13 Con referencia al problema 8.12, ¿cuál es el tamaño mínimo que se requiere para la muestra si se supone que la distribución de los montos de ventas no es normal y el comprador desea estimar las ventas promedio con un margen de error de $2000 y un intervalo de confianza del 99%?

$$n = \left(\frac{z\sigma}{E}\right)^2 = \left[\frac{(2.58)(3200)}{2.00}\right]^2 = (4.128)^2 = 17.04 = 18$$

Sin embargo, como no se supone que la población tiene distribución normal, el tamaño mínimo de la muestra es n - 30, lo cual permite utilizar el teorema de límite central como base para aplicar la distribución normal de probabilidad para construir el intervalo de confianza.

INTERVALOS DE CONFIANZA DE UN EXTREMO PARA LA MEDIA DE LA POBLACIÓN

8.14 En ocasiones, es posible que se tenga más interés en un *intervalo de confianza de un extremo,* y no en los intervalos comunes de un extremo. Se presentar,a un caso así, si sólo se tuviera interés en el valor mayor (o solamente en el valor menor) de la media en el grado especificado de confianza. Un "intervalo de confianza del 95% superior" se extiende a partir de un límite inferior calculado hasta el infinito positivo, teniendo una proporción del 0.05 del área bajo la curva normal hacia la izquierda del límite inferior. De manera similar, un "intervalo de confianza del 95% inferior" se extiende desde el infinito negativo hasta un límite superior calculado, teniendo una proporción del 0.05 del área bajo la curva normal a la derecha del límite superior.

Supóngase que un cliente potencial de una tienda de juguetes que se encuentra en un aeropuerto, observa una muestra aleatoria de n = 64 ventas y encuentra que la media de la muestra es $\bar{X} = 14\ 630$ con una desviación estándar de s = $2 400. Determine el intervalo de confianza del 95% superior, de manera que se identifique el valor *mínimo* de la media poblacional con un grado de confianza del 95%.

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2.40}{\sqrt{64}} = \frac{2.40}{8} = 0.30$$

$$\text{Int. del 95\% superior} = \bar{X} - zs_{\bar{x}} = 14.63 - 1.645(0.30) = \$14.14 \text{ o mayor}$$

Por ello, puede afirmarse, con un grado de confianza del 95%, que el promedio de ventas para la población de todos los clientes es igual o mayor que $14 140.

8.15 Con un intervalo de confianza del 99%, ¿cuál es la estimación del valor máximo de las ventas promedio del problema 8.14?

Como $\overline{X} = \$14.63$ y $s_{\bar{x}} = 0.30$,

$$\text{Int. del 99\% inferior} = \overline{X} - zs_{\bar{x}} = 14.63 + 2.33(0.30) = \$15.33 \text{ o menos}$$

Por ello, puede afirmarse, con un grado de confianza del 99%, que la venta promedio no es mayor de $15 530.

INTERVALOS DE CONFIANZA PARA LA MEDIA UTILIZANDO LA DISTRIBUCIÓN *t*

8.16 En el problema 8.7 se construyeron intervalos de confianza para estimar la vida útil promedio de una marca especifica de cinescopios de televisión, suponiendo que la vida útil de todos los cinescopios tiene distribución aproximadamente normal y $\sigma = 500$, y suponiendo una muestra de $n = 15$, con $\overline{X} = 8\ 900$ horas. Supóngase que se desconoce a, pero que la desviación estándar de la muestra es s = 500.

(a) Construya el intervalo de confianza del 95% para estimar la media de la población y compare ese intervalo con la respuesta que obtuvo en el problema 8.7 (a).

(b) Construya el intervalo de confianza del 90% para estimar la media de la población y compare este intervalo con la respuesta dada al problema 8.7 (b).

(Nota: en este caso resulta apropiado utilizar la distribución f porque se supone que la población tiene distribución normal, se desconoce σ y la muestra es pequeña *(n* < 30).)

(a) $\overline{X} \pm t_{gl} s_{\bar{x}} = 8\ 900 \pm 2.145 \dfrac{s}{\sqrt{n}} = 8\ 900 \pm 2.145 \dfrac{500}{\sqrt{15}}$

$$= 8\ 900 \pm 2.145\left(\frac{500}{3.87}\right) = 8\ 900 \pm 2.145(129.199) = 8\ 623 \text{ a } 9\ 177 \text{ hr}$$

El intervalo de confianza es más amplio que el del problema 8.7 (a) lo cual refleja la diferencia entre una distribución r con g/ = 15 — 1 = 14, y la distribución normal de probabilidad.

$$\overline{X} \pm t_{gl}\, s_{\bar{x}} = 8\ 900 \pm 1.761(129.199) = 8\ 672 \text{ a } 9\ 128 \text{ hr}$$

De nueva cuenta, el intervalo de confianza es más amplio que el del problema 8.7 (b).

8.17 Como supervisor del proceso de empacado de café en sobres, suponga que se toma una muestra aleatoria de 12 de los sobres en la planta empacadora. El peso neto de los sobres de café es el que se reporta en la Tabla 8.7. Determine (a) el peso neto del café que se empaca en cada sobre, y (b) la desviación estándar muestral. (c). Suponiendo que el peso del café empacado tiene distribución normal, estime el peso promedio por sobre de café utilizando un intervalo de confianza del 95%.

Tabla 8.7 Peso neto del café empacado en 12 sobres

Gramos por sobre	15.7	15.8	15.9	16.0	16.1	16.2
Número de sobres	1	2	2	3	3	1

Tabla 8.8 Hoja de trabajo para él problema 8.17

X por sobre	No. de sobres	Total X	X^2 por sobre	X^2 total
15.7	1	15.7	246.49	246.49
15.8	2	31.6	249.64	499.28
15.9	2	31.8	252.81	505.62
16.0	3	48.0	256.00	768.00
16.1	3	48.3	259.21	777.63
16.2	1	16.2	262.44	262.44
	n=12	$\Sigma X = 191.6$		$\Sigma X^2 = 3{,}059.46$

Con referencia a la Tabla 8.8,

(a) $\bar{X} = \dfrac{\Sigma X}{n} = \dfrac{191.6}{12} = 15.97$ gr

(b) $s = \sqrt{\dfrac{n\,\Sigma X^2 - (\Sigma X)^2}{n(n-1)}} = \sqrt{\dfrac{12(3\,059.46) - (191.6)^2}{12(11)}} = \sqrt{0.0224} = 0.15$

(c) $\bar{X} \pm t_{gl}\, s_{\bar{x}} = 15.97 \pm t_{11}\dfrac{s}{\sqrt{n}} = 15.97 \pm 2.201\left(\dfrac{0.15}{\sqrt{12}}\right)$

$= 15.97 \pm 2.201\left(\dfrac{0.15}{3.46}\right) = 15.97 \pm 2.201(0.043) = 15.88$ a 16.06 gr

RESULTADOS POR COMPUTADORA

8.18 Con referencia a los datos de la Tabla 2.6 (página 16) que se refieren al tiempo necesario para terminar una tarea de ensamble, y suponiendo que los datos se basan en una muestra aleatoria, utilice algún programa de computación para obtener el intervalo de confianza del 90% para la media de la población.

En la figura 8-5 se presentan los datos de entrada y los resultados. Los resultados incluyen el tamaño de la muestra, la media y la desviación estándar muestrales, y el error estándar de la media. Se incluye, asimismo, el intervalo de confianza solicitado. Tal como se observa, el intervalo de confianza del 90% para el tiempo promedio de ensamble en la población es de 12.544 a 14.056 minutos.

```
MTB > SET ASSEMBLY TIMES INTO Cl
DATA>   10   14   15   13   17   16   12   14   11   13   15   18    9
DATA>   14   14    9   15   11   13   11   12   10   17   16   12   11
DATA>   16   12   14   15
DATA> END
MTB > NAME FOR Cl IS 'TIME'
MTB > TINTERVAL WITH 90 PERCENT CONFIDENCE FOR 'TIME'

            N      MEAN    STDEV   SE MEAN   90.0 PERCENT C.I.
TIME       30    13.300    2. 437    0.445   (  12.544,   14.056)
```

Fig. 8-5 Resultados de Minitab del problema 8.18

8.19 Con referencia a los datos que se reportan en la Tabla 8.7, con respecto al peso neto del café que se empaca en una muestra aleatoria de n = 12 sobres, utilizando algún programa de computación, (a) determine el intervalo de confianza del 95% para la media de la población, y (b) compare las respuestas que se incluyen en los resultados de la computadora con los que se obtuvieron mediante los cálculos manuales del problema 8.l7.

Con referencia a la figura 8-6,

(a) Intervalo de confianza del 95% = 15.8715 a 16.0618 gr

(b) Al comparar con las soluciones del problema 8.17 los valores de la media, la desviación estándar, el error estándar de la media y los límites de confianza difieren sólo ligeramente debido a las diferencias de redondeo que se incluyeron en los cálculos manuales.

```
MTB > SET WEIGHTS INTO Cl
DATA>  15.7  15.8  15.8  15.9  15.9  16.0  16.0  16.1  16.1  16.1  16.1
DATO>   16.2
DATA> END
MTB > NAME FOR Cl IS 'WEIGHT'
MTB > TINTERVAL WITH 95 PERCENT CONFIDENCE FOR 'WEIGHT'

              N     MEAN    STDEV  SE MEAN   95.0 PERCENT C.I.
WEIGHT       12   15.9667   0.1497   0.0432   ( 15.8715,  16.0618)
```

Fig. 8-6 Resultados de Minitab para el problema 8.19

Problemas complementarios

DISTRIBUCIÓN MUESTRAL DE LA MEDIA

8.20 Se sabe que, el año anterior, el promedio de ventas por tienda de un producto determinado de consumo popular tuvo una distribución normal con μ - \$3 400 000, con desviación esándar de σ = \$200 000. Si son muy numerosas las tiendas que manejan ese producto, determine el error estándar de la media para una muestra de tamaño n = 25.

Resp. \$40 000

8.21 Con referencia al problema 8.20, ¿cuál es la probabilidad de que las ventas de *una* tienda elegida al azar sean (a) mayores de \$3 500 000? (b) entre \$3 350 000 y \$3 450 000?

Resp. (a) 0.3085, (b) 0.1974

8.22 Con referencia al problema 8.20, ¿cuál es la probabilidad de que la *media* (a) sea mayor de \$ 3 500 000? (b) esté entre \$3 350 000 y \$3 450 000?. Compare sus respuestas con las que se obtuvieron en el problema 8.21.

Resp. (a) 0.0062, (b) 0.7888

8.23 Con referencia al problema 8.20, suponga que sólo hay 100 tiendas que manejan ese producto. Determine el error estándar de la media para la muestra con n = 25 en este caso, y compare su respuesta con la que se obtuvo en el problema 8.21.

Resp. \$34 800

INTERVALOS DE CONFIANZA PARA LA MEDIA

8.24 Suponga que desea estimar el promedio de ventas por tienda para un producto determinado de consumo popular durante el año anterior, y que el número de tiendas es grande. Determine el intervalo de confianza del 95%, considerando que las ventas se distribuyen en forma normal, $\bar{X}$ = \$3 425 000, σ = \$200 000 y n = 25.

Resp. \$3 346 600 a \$3 503 400

8.25 Con referencia al problema 8.24, determine el intervalo de confianza del 95%, asumiendo que la población tiene una distribución normal, $\bar{X}$ = \$3 425 000, s =\$200 000 y n = 25.

Resp. \$3 342 444 a \$3 507 555

8.26 Para el problema 8.24, determine el intervalo de confianza del 95% suponiendo que la población *no* tiene una distribución normal, $\bar{X}$ = \$3 425 000, s = \$200 000 y n = 50.

Resp. \$3 369.55 a \$3 480.45

8.27 Para una muestra de 50 empresas tomadas de una industria determinada, se encuentra que el número promedio de trabajadores por empresa es de 420.4, con una desviación estándar muestral de 55.7. Existe un total de 380 empresas en esa rama industrial. Determine el error estándar de la media que debe usarse para estimar la media de la población mediante un intervalo de confianza.

Resp. 7.33

8.28 Para el problema 8.27, determine el intervalo de confianza del 90% para estimar el número promedio de trabajadores por empresa en esa rama industrial.

Resp. 408.3 a 432.5

8.29 Para las situaciones que se describieron en los problemas 8.27 y 8.28, determine el Intervalo de confianza del 90% para estimar el número total de trabajadores empleados en esa industria.

Resp. 155 154 a 164 350

8.30 Un analista de un departamento de personal elige al azar los expedientes de 16 trabajadores a destajo y encuentra que el salario promedio por pieza es de \$950. Se supone que los salarios de esa empresa tienen una distribución normal. Si se sabe que la desviación estándar de los salarios es de \$100, estime la tasa promedio de salarios en la empresa utilizando un intervalo de confianza del 80%.

Resp. \$918 a \$982

8.31 Con referencia al problema \$.30, suponga que no se conoce la desviación estándar de la población, pero que sí se sabe que la de la muestra es de \$100. Estime el promedio de salarios en la empresa utilizando un intervalo de confianza del 80%.

Resp. \$916 a \$984

8.32 El diámetro promedio de una muestra de n =12 varillas incluidas en un embarque es 2.350 mm, con una desviación estándar de 0.050 mm. Se supone que la distribución de los diámetros de todas las varillas incluidas en el embarque tiene una distribución aproximadamente normal. Determine el intervalo de confianza del 99% para estimar el diámetro promedio de todas las varillas.

Resp. 2.307 a 2.393 mm

8.33 El diámetro promedio de una muestra de n = 100 varillas Incluidas en un embarque es 2.350 mm con desviación estándar de 0.050 mm. Estime el diámetro promedio de todas las varillas incluidas en el embarque, si éste contiene 500 varillas, utilizando un intervalo de confianza del 99%.

Resp. 2.338 a 2.362 mm

8.34 El peso promedio de la muestra de 100 varillas del problema 8.33 es 8.45 gr, con desviación estándar de 0.25 gr. Estime el peso total del embarque (excluyendo los materiales de empaque) utilizando un intervalo de confianza del 99%.

Resp. 4195 kg a 4 255 kg

DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA NECESARIO PARA ESTIMAR LA MEDIA

8.35 Se sabe, por registros históricos, que la desviación estándar del nivel de ventas por tienda de un producto de consumo popular es σ = \$200 000 y se supone que la población de la totalidad de ventas por tienda tiene una distribución normal. ¿Cuál es el tamaño mínimo de muestra que se requiere para estimar el promedio de ventas por tienda, con un margen de error de \$100 000 y con una confianza del 95%.

Resp. 15.37 $\cong$ 16

8.36 Un analista desea estimar el salario diario promedio de los trabajadores de una compañía determinada, con un margen de error de \$250 y una confianza del 90%. Se estima que la desviación estándar de los salarios no es mayor de \$1 000. ¿Cuál es el número de expedientes que deben muestrearse, como mínimo, para satisfacer este objetivo de investigación?

Resp. 43.56 $\cong$ 44

INTERVALOS DE CONFIANZA DE UN EXTREMO PARA LA MEDIA DE LA POBLACIÓN

8.37 En vez del Intervalo de confianza de dos extremos que se construyó en el problema 8.24, suponga que se desea estimar el *valor mínimo* por tienda para ese producto durante el año anterior. Al igual que antes, se supone que la distribución de las ventas es aproximadamente normal. Determine el valor mínimo de la media utilizando un intervalo de confianza del 95%, y considerando $\bar{X}$ = \$3 425 000, σ = \$200 000 y n = 25. Compare su intervalo de confianza con el que se construyó en el problema 8.24.

Resp. μ estimada ≥ \$3 359 000

8.38 Al utilizar los datos del problema 8.32, determine el intervalo de confianza del 99% inferior para estimar el diámetro promedio de todas las varillas incluidas en el embarque. Compare el intervalo con el que se construyó en el problema 8.32.

Resp. μ estimada ≤ 2.388 mm

RESULTADOS POR COMPUTADORA

8.39 Con referencia a la Tabla 2.16 (página 26), que contiene los montos de 40 préstamos personales, y suponiendo que se trata de datos muestrales aleatorios, utilice un paquete de computación para determinar el intervalo de confianza del 99% para el monto promedio de los préstamos en la población.

Resp. \$8 222 000 a \$1 372 000

Otros intervalos de confianza

9

9.1 INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS MEDIAS POBLACIONALES UTILIZANDO LA DISTRIBUCIÓN NORMAL

Con frecuencia es necesario estimar la diferencia entre dos medias poblacionales, tal como la diferencia entre los niveles salariales en dos empresas. Como se indica en la sección 8.1, el estimador puntual no sesgado de $(\mu_1 - \mu_2)$ es $(\bar{X}_1 - \bar{X}_2)$. El intervalo de confianza se construye de manera similar a como se estima el de la media, excepto que el error estándar de la distribución muestral que corresponde en este caso es el de la *diferencia* entre medias. El uso de la distribución normal se basa en las mismas condiciones que para el caso de la distribución muestral de la media (véase la sección 8.2), excepto que en este caso hay dos medias. La fórmula que se utiliza para estimar la diferencia entre las medias de dos poblaciones, con intervalos de confianza es

$$(\bar{X}_1 - \bar{X}_2) \pm z\sigma_{\bar{x}_1 - \bar{x}_2} \tag{9.1}$$

o

$$(\bar{X}_1 - \bar{X}_2) \pm z s_{\bar{x}_1 - \bar{x}_2} \tag{9.2}$$

Cuando se conocen las desviaciones estándar de las dos poblaciones, el error estándar de la diferencia entre dos medias es

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}} \tag{9.3}$$

Cuando no se conocen las desviaciones estándar de las poblaciones, el error estándar estimado de la diferencia entre dos medias, suponiendo que resulta apropiado el uso de la distribución normal es

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}} \tag{9.4}$$

Los valores de los errores estándar de las medias respectivas que se incluyen en estas fórmulas se calculan a través de las fórmulas dadas en la sección 8.2, incluyendo la posibilidad de utilizar los factores de corrección por población finita cuando sea apropiado.

EJEMPLO 1. El salario diario promedio para una muestra de n = 30 empleados de una empresa manufacturera grande es $\bar{X}$ = \$28 000, con una desviación estándar de s = \$1400. En otra empresa grande, una muestra aleatoria de n = 40 empleados tiene un salario promedio diario de \$27 000, con desviación estándar muestral de s = \$1000. El intervalo dé confianza del 99% para estimar la diferencia entre los niveles diarios de salarios en las dos empresas es

$$\text{Int. del }99\% = (\bar{X}_1 - \bar{X}_2) \pm z s_{\bar{x}_1-\bar{x}_2}$$
$$= \$10.00 \pm 2.58(3.01)$$
$$= \$10.00 \pm 7.77$$
$$= \$2.23 \text{ a } \$17.77$$

en donde $\bar{X}_1 - \bar{X}_2 = \$280.00 - \$270.00 = \$10.00$

z=2.58

$$s_{\bar{x}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{14.00}{\sqrt{30}} = \frac{14.00}{5.477} = 2.56$$

$$s_{\bar{x}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{10.00}{\sqrt{40}} = \frac{10.00}{6.325} = 1.58$$

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} = \sqrt{(2.56)^2+(1.58)^2} = \sqrt{6.5536+2.4964} \cong 3.01$$

Por ello, puede afirmarse que el salario diario promedio de la primera empresa es mayor que el correspondiente a la segunda, en una cantidad que va de \$223 a \$1777, con una confianza del 99% en esa estimación por intervalo.

Además del intervalo de confianza de dos extremos, puede construirse también un intervalo de confianza de un extremo para la diferencia entre dos medias. (Véase problema 9.4).

9.2 LA DISTRIBUCIÓN *t* DE STUDENT Y LOS INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE LAS MEDIAS DE DOS POBLACIONES.

Tal como se explicó en la sección 8.5, resulta necesario utilizar la distribución r de Student cuando:

(1) No se conocen las desviaciones estándar σ de las poblaciones.

(2) Las muestras son pequeñas ($n < 30$). Si las muestras son grandes, entonces es posible aproximar los valores r mediante la distribución normal estándar z.

(3) Se supone que las poblaciones tienen distribuciones aproximadamente normales (debe observarse que no puede aplicarse el teorema de límite central cuando se trata de muestras pequeñas).

Sin embargo, cuando se utiliza la distribución t para definir intervalos de confianza para la diferencia entre dos medias, en vez de hacerlo para una sola, se requiere una suposición adicional.

(4) Las varianzas de tas dos poblaciones (que se desconocen) son iguales, $\sigma_1^2 = \sigma_2^2$.

Debido a la anterior suposición de igualdad de varíanzas, el primer paso para determinar el error estándar de la diferencia entre dos medias cuando se utiliza la distribución r consiste en combinar las dos varianzas muestrales:

$$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} \tag{9.5}$$

El error estándar de la diferencia entre dos medias, con base en la estimación combinada de la varianza $\hat{\sigma}^2$ es

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\hat{s}^2}{n_1}+\frac{\hat{s}^2}{n_2}} \tag{9.6}$$

En donde gl = n, + n_2 = 2, el intervalo de confianza es

$$(\bar{X}_1 - \bar{X}_2) \pm t_{gl}\,\hat{\sigma}_{\bar{x}_1-\bar{x}_2} \tag{9.7}$$

EJEMPLO 2. La vida útil promedio de una muestra aleatoria de n_1 = 10 focos es $\bar{X}_1$ = 4600 horas con si = 250 horas. Para otra marca de focos, la vida útil promedio y la desviación estándar para una muestra de n_2 = 8 focos son $\bar{X}_2$ = 4000 horas y $S_2 = 200$ horas. Se asume que la vida útil de los focos de ambas marcas tiene una distribución normal. El intervalo de confianza del 90% para estimar la diferencia entre las vidas útiles promedio de las dos marcas de focos es.

$$\text{Int. del 90\%} = (\bar{X}_1 - \bar{X}_2) \pm t_{16}\hat{\sigma}_{\bar{x}_1-\bar{x}_2}$$
$$= 600 \pm 1.746(108.847) = 600 \pm 190 = 410 \text{ a } 790 \text{ h}$$

en donde $\bar{X}_1 - \bar{X}_2 = 4600 - 4000 = 600$

$$t_{gl} = t_{n_1+n_2-2} = t_{16} = 1.746$$

$$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{9(250)^2 + 7(200)^2}{10+8-2} = 52{,}656.25$$

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2}} = \sqrt{\frac{52{,}656.25}{10} + \frac{52{,}656.25}{8}} = 108.847$$

Asi, puede afirmarse con una confianza del 90%, que la primera marca de focos tiene una vida útil promedio mayor que la de la segunda, en una cantidad de entre 410 y 790 horas.

Debe observarse que en el caso de las dos muestras, es posible que ambas sean pequeñas ($n < 30$) y que, aun asf, resulte posible utilizar la distribución normal para aproximar t, porque los $gl \geq 29$. Sin embargo, al hacer esto, debe suponerse que las poblaciones tienen una distribución aproximadamente normal, porque no puede aplicarse el teorema de límite central a cada una de las muestras pequeñas por separado.

9.3 INTERVALOS DE CONFIANZA PARA LA PROPORCIÓN UTILIZANDO LA DISTRIBUCIÓN NORMAL

Tal como se explica en la sección 6.4, la distribución de probabilidad que resulta aplicable a las proporciones es la binomial. Sin embargo, los cálculos asociados con la construcción de intervalos de confianza para una proporción poblacional desconocida, con base en el proceso de Bernoulli que se describió en la sección 6.3 son bastante laboriosos. Por ello, la mayor parte de libros de texto utilizan la distribución normal para aproximar la binomial al construir intervalos de confianza para proporciones. Como se explicó en la sección 7.4, esta aproximación es apropiada cuando $n \geq 30$ y tanto $np \geq 5$, cómo $nq \geq 5$ (en donde $q = 1 - p$). *Sin embargo, cuando se desconoce la proporción de la población p* (o π), *la mayor parte de bs especialistas en estadística sugieren que se tome una muestra de* $n \geq 100$. Debe observarse que, en el contexto de la estimación estadística, no se conoce π, sino que se estima mediante $\hat{p}$.

La varianza de la distribución de las proporciones (sección 6.4) sirve de base para el error estándar. Dada una proporción muestral observada, $\hat{p}$, el error estimado de la proporción es

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{9.8}$$

En el contexto de la estimación estadística, no se conocería la proporción de la población p o π, porque es precisamente el valor que se estima. Si la población es finita, entonces resulta apropiado utilizar el factor de corrección por población finita (véase la sección 8.2). Al igual que en el caso del error estándar de la media, por lo general no se considera necesario el uso de ese factor de corrección si $n < 0.05\,N$. La fórmula para el error estándar de la población que incluye el factor de corrección por población finita es

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\sqrt{\frac{N-n}{N-1}} \tag{9.9}$$

Finalmente, el intervalo de confianza aproximado para una proporción poblacional es

$$\hat{p} \pm z s_{\hat{p}} \tag{9.10}$$

Además de los intervalos de confianza de dos extremos, también es posible construir intervalos de confianza de un extremo para la proporción de una población (véase el problema 9.12).

EJEMPLO 3. Una empresa de investigación de mercados entrevista a una muestra aleatoria de 100 hombres de una comunidad grande y encuentra que una proporción muestral de 0.40 de ellos prefieren las hojas de rasurar fabricadas por la empresa cliente de los investigadores, y no las demás marcas. El intervalo de confianza del 95% para la proporción de todos los hombres de esa comunidad que prefieren las hojas de rasurar de la empresa cliente de los investigadores se determina de la siguiente manera:

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.40)(0.60)}{100}} = \sqrt{\frac{0.24}{100}} = \sqrt{0.0024} \cong 0.05$$

$$\hat{p} \pm z s_{\hat{p}} = 0.40 \pm 1.96(0.05)$$

$$= 0.40 \pm 0.098 = 0.40 \pm 0.10 = 0.30 \text{ a } 0.50$$

Por lo tanto, puede estimarse con una confianza del 95% que la proporción de hombres de esa comunidad que prefieren las hojas de rasurar de la empresa está entre 0.30 y 0.50.

9.4 DETERMINACIÓN DEL TAMAÑO DE MUESTRA NECESARIO PARA ESTIMAR LA PROPORCIÓN

Antes de obtener realmente una muestra, puede determinarse el tamaño mínimo que se requiere especificando el grado de confianza necesario, el error que resulta aceptable, y haciendo una estimación inicial de π, la proporción desconocida de la población:

$$n = \frac{z^2 \pi(1-\pi)}{E^2} \tag{9.11}$$

En (9.11), z es el valor que se utiliza para el intervalo de confianza especificado, π es la estimación inicial de la proporción poblacional y E es el factor de error de "más y menos" que se permite en el intervalo (siempre es la mitad del intervalo total de confianza).

Si no resulta posible realizar una estimación inicial de π, entonces debe estimarse que es de 0.50. Esta estimación es "conservadora" porque es el valor para el cual se requeriría el tamaño de muestra más grande. Bajo esa suposición, la fórmula general para el tamaño de la muestra se simplifica de la siguiente manera:

$$n = \left(\frac{z}{2E}\right)^2 \tag{9.12}$$

(Atora; Cuando se determina el tamaño de la muestra, todos los resultados fraccionarios deben redondearse hacia el número superior. Además, cualquier tamaño de muestra calculado que resulte ser menor de 100, debe aumentarse hasta 100, porque las fórmulas 9.11 y 9.12 se basan en el uso de la distribución normal.)

EJEMPLO 4. Para el estudio del ejemplo 3, suponga que antes de recopilar los datos se especificó que la estimación por intervalo del 95% deberla estar dentro del ± 0.05%, y no se hizo ninguna evaluación previa acerca del probable valor de π. El tamaño mínimo de muestra que se debe estudiar es

$$n = \left(\frac{z}{2E}\right)^2 = \left(\frac{1.96}{2(0.05)}\right)^2 = \left(\frac{1.96}{0.10}\right)^2 = (19.6)^2 = 384.16 = 385$$

Además de estimar la proporción poblacional, también puede estimarse el *número total* de elementos en la categoría de la población [véase el problema 9.7 (b)].

9.5 INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS PROPORCIONES POBLACIONALES

Para estimar la diferencia entre las proporciones de dos poblaciones, el estimador puntual sesgado de $(\pi_1 - \pi_2)$ es $(\hat{p}_1 - \hat{p}_2)$ (véase la sección 8.1). El intervalo de confianza implica utilizar el error estándar de la *diferencia entre las proporciones.* El uso de la distribución normal se basa en las mismas condiciones que para la distribución muestral de la proporción que se vio en la sección 9.3, excepto que se trata de dos muestras y se aplican los requerimientos a cada una de ellas. El intervalo de confianza para estimar la diferencia entre las proporciones de dos poblaciones es

$$(\hat{p}_1 - \hat{p}_2) \pm z s_{\hat{p}_1 - \hat{p}_2} \tag{9.13}$$

El error estándar de la diferencia entre dos proporciones se determina mediante la fórmula 9.14, en donde el valor de cada uno de los errores estándar de la proporción se calcula según se describió en la sección 9.3:

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{s^2_{\hat{p}_1} + s^2_{\hat{p}_2}} \tag{9.14}$$

EJEMPLO 5. En el ejemplo 3 se reportó que una proporción de 0.40 hombres de una muestra aleatoria de 100 tomada de una comunidad grande, manifestó preferir las hojas de rasurar de la empresa cliente de los investigadores y no las demás marcas. En otra comunidad grande, 60 hombres de una muestra aleatoria de 200 prefirieron las hojas de rasurar de la empresa cliente. El intervalo de confianza del 90% para la proporción de hombres de las dos comunidades que prefieren las hojas de rasurar de la empresa cliente es:

$$\begin{aligned}\text{Int. del 90\%} &= (\hat{p}_1 - \hat{p}_2) \pm z s_{\hat{p}_1 - \hat{p}_2} \\ &= 0.100 \pm 1.645(0.059) \\ &= 0.100 \pm 0.097 \\ &= 0.003 \text{ a } 0.197\end{aligned}$$

en donde $\bar{p}_1 - \bar{p}_2 = 0.40 - 0.30 = 0.10$

$$z = 1.645$$

$$s^2_{\hat{p}_1} = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} = \frac{(0.40)(0.60)}{100} = 0.0024$$

$$s^2_{\hat{p}_2} = \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} = \frac{(0.30)(0.70)}{200} = 0.00105$$

$$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{s^2_{\hat{p}_1} + s^2_{\hat{p}_2}} = \sqrt{0.0024 + 0.00105} = \sqrt{0.00345} \cong 0.059$$

9.6 LA DISTRIBUCIÓN x^2 (JI-CUADRADA) Y LOS INTERVALOS DE CONFIANZA PARA LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR

Si se tiene una población de valores con distribución normal, puede mostrarse que las distribuciones x^2 (ji-cuadrada) son las distribuciones de probabilidad apropiadas para la razón $(n-1)s^2/\sigma^2$. Existe una distribución ji cuadrada distinta para los diferentes valores de n - 1 , que representan los grados de libertad *(gl).* Asi,

$$X^2_{gl} = \frac{(n-1)s^2}{\sigma^2} \tag{9.15}$$

Como la varianza muestral es el estimador insesgado de la varianza poblacional, el valor esperado a largo plazo del cociente anterior es igual a los grados de libertad, o n - 1 . Sin embargo, en el casó de la mayoría de las muestras, la varianza muestral *no* es idéntica a la varianza poblacional. Como se sabe que el cociente anterior tiene una distribución ji-cuadrada, puede utilizarse esta distribución para realizar inferencias estadísticas sobre varianzas o desviaciones estándar desconocidas.

Las distribuciones ji-cuadrada no son simétricas. Por ello, un intervalo de confianza de dos extremos para una varianza o una desviación estándar requiere del uso de dos valores distintos de x^2, en voz del método de "más-menos" que se utilizó en los intervalos de confianza que se basan en la distribución normal. La fórmula para construir un intervalo de confianza par la varianza de la población es

$$\frac{(n-1)s^2}{\chi^2_{gl,\ \text{superior}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{gl,\ \text{inferior}}} \tag{9.16}$$

El intervalo de confianza para la desviación estándar de la población es

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{gl,\ \text{superior}}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{gl,\ \text{inferior}}}} \tag{9.17}$$

En el apéndice 7 se enlistan las proporciones de área bajo las distribuciones ji-cuadrada, de acuerdo con diversos grados de libertad, o *gl.* En la fórmula general anterior, los subíndices "superior" e "inferior" identifican los puntos percentiles de la distribución x^2 especifica que se utiliza para construir el intervalo de confianza. Por ejemplo, para un intervalo de confianza del 90%, el "superior" es $\chi^2_{0.95}$ y el "inferior" es $\chi^2_{0.05}$" Al excluir el 5% mayor y el 5% menor de la distribución ji-cuadrada, lo que resta es el 90% "central".

EJEMPLO 6. El salario diario promedio para una muestra de 30 trabajadores de una empresa grande es $\bar{X} = 2800$, cor desviación estándar muestral de s =140. Se supone que los salarios diarios de la empresa tienen una distribuciór aproximadamente normal. El intervalo de confianza del 95% para estimar la desviación estándar de los salarios es

$$\sqrt{\frac{(n-1)s^2}{X^2_{gl,\text{superior}}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{X^2_{gl,\ \text{inferior}}}}$$

$$\sqrt{\frac{(29)(196.00)}{X^2_{29,0.975}}} \leq \sigma \leq \sqrt{\frac{(29)(196.00)}{X^2_{29,0.025}}}$$

$$\sqrt{\frac{5,684.00}{45.72}} \leq \sigma \leq \sqrt{\frac{5,684.00}{16.05}}$$

$$\sqrt{124.3220} \leq \sigma \leq \sqrt{354.1433}$$

$$11.15 \leq \sigma \leq 18.82$$

En el ejemplo anterior debe observarse que, como los encabezados de las columnas del apéndice 7 son probabilidades del lado derecho, y no valores percentiles, los encabezados de columna que se utilizan en la tabla son los valores complementarios de los valores percentiles "superior" e "inferior* que se requieren.

Como alternativa para los intervalos de confianza de dos extremos, también pueden construirse intervalos de confianza de un extremo para la varianza y la desviación estándar (véase el problema 9.14).

9.7 RESULTADOS POR COMPUTADORA

Por lo general, los programas de computadora que realizan análisis estadísticos tienen la capacidad de determinar diversos tipos de intervalos de confianza, además de los intervalos de confianza para la media poblacional. En el problema 9.15 se ilustra el uso de una computadora para obtener un intervalo de confianza para la diferencia entre dos medias poblacionales.

Problemas resueltos

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS MEDIAS POBLACIONALES UTILIZANDO LA DISTRIBUCIÓN NORMAL

9.1 Para el estudio que se reportó en el problema 8.11, suponga que hubo 900 clientes que no adquirieron el "cupón especial" pero que sí realizaron otras compras en la tienda durante el período de estudio. Para una muestra de 200 de estos clientes, el promedio de compras fue de $\bar{X}$ = \$19 600, con una desviación estándar muestral de s = \$8400.

(a) Estime el monto promedio de compras para los clientes que no compraron los cupones, utilizando un intervalo de confianza del 95%.

(b) Estime la diferencia entre el monto promedio de compra para los clientes con y sin cupón, utilizando un intervalo de confianza del 90%.

(a) $$s_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{8.40}{\sqrt{200}}\sqrt{\frac{900-200}{900-1}} = \frac{8.40}{14.142}\sqrt{0.7786} = (0.594)(0.882) = 0.52$$

$$\bar{X} \pm z s_{\bar{x}} = 19.60 \pm 1.96(0.52) = \$19.60 \pm 1.02 = \$18.58 \text{ a } \$20.62$$

(b) $$s_{\bar{x}_1-\bar{x}_2} = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}} = \sqrt{(0.57)^2 + (0.52)^2} = \sqrt{0.3249 + 0.2704} = \sqrt{0.5953} = 0.772$$

$$(\bar{X}_1 - \bar{X}_2) \pm z s_{\bar{x}_1-\bar{x}_2} = (24.57 - 19.60) \pm 1.645(0.772) = \$4.97 \pm 1.27 = \$3.70 \text{ a } \$6.24$$

Asf, puede afirmarse con una confianza del 90% que el nivel promedio de ventas para los clientes con cupón excede a los que no lo tenían, en una cantidad que se encuentra entre \$3700 y \$6240.

9.2 Una muestra aleatoria de 50 hogares de la comunidad *A* tiene ingresos diarios promedio de $\bar{X}$ = 34 600, con desviación estándar s = \$2200. Una muestra aleatoria de 50 hogares de la comunidad *B* tiene un promedio de $\bar{X}$ = \$33 800, con desviación estándar de s = \$2800. Estime la diferencia en los ingresos diarios promedio de los hogares en las dos comunidades utilizando un intervalo de confianza del 95%.

$$s_{\bar{x}_1} = \frac{s_1}{\sqrt{n_1}} = \frac{2\,200}{\sqrt{50}} = \frac{2\,200}{7.07} = \$311.17$$

$$s_{\bar{x}_2} = \frac{s_2}{\sqrt{n_2}} = \frac{2\,800}{\sqrt{50}} = \frac{2\,800}{7.07} = \$396.04$$

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{s^2_{\bar{x}_1} + s^2_{\bar{x}_2}} = \sqrt{(311.17)^2 + (396.04)^2} = \sqrt{96\,826.77 + 156\,847.68} = \$503.66$$

$$(\bar{X}_1 - \bar{X}_2) \pm z s_{\bar{x}_1-\bar{x}_2} = (34\,600 - 33\,800) \pm 1.96(503.66)$$
$$= 800 \pm 987.17 = -\$187.17 \text{ a } \$1\,787.17$$

Con un grado de confianza del 95%, los límites del intervalo de confianza señalan que la media de la primera comunidad podría ser inferior a la media de la segunda comunidad en \$187.17, mientras que en el otro extremo, la media de la primera comunidad podría exceder a la segunda hasta en \$1787.17. Debe observarse que la posibilidad de que no exista diferencia entre las medias de las dos poblaciones se encuentra incluida en este intervalo de confianza del 95%.

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS MEDIAS UTILIZANDO LA DISTRIBUCIÓN *t*

9.3 En una planta empacadora, el peso neto promedio del café envasado en sobres, para una muestra de *n* = 12 sobres, es $\bar{X}_1$ = 15.97 gramos, con s_1 = 0.15 gr. En otra planta empacadora, el peso neto promedio del café que se empaca en sobres, para una muestra de $n_2 = 15$ es $\bar{X}_2 = 16.14$ gr., con desviación estándar de $s_2 = 0.09$ gr. Se supone que las distribuciones de las cantidades empacadas tienen una distribución aproximadamente normal. Estime la diferencia del peso promedio de café que se empaca en sobres en las dos plantas, utilizando un intervalo de confianza del 90%.

$$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = \frac{11(0.15)^2+14(0.09)^2}{12+15-2} = 0.014436$$

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\hat{\sigma}^2}{n_1}+\frac{\hat{\sigma}^2}{n_2}} = \sqrt{\frac{0.014436}{12}+\frac{0.014436}{15}} = 0.047$$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{gl}\,\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = (15.97 - 16.14) \pm t_{25}(0.047)$$

$$= (-0.17) \pm 1.708(0.047) = (-0.17) \pm 0.08 = -0.25 \text{ a } -0.09$$

En otras palabras, puede afirmarse con una confianza del 90% que el peso neto promedio del café que se empaca en la *segunda* planta se encuentra entre 0.09 y 0.25 gr de más que en la *primera* planta.

INTERVALOS DE CONFIANZA DE UN EXTREMO PARA LA DIFERENCIA ENTRE DOS MEDIAS

9.4 Al igual que se hizo para la media, tal como se explicó en el problema 8.14, puede estimarse la diferencia entre dos medias utilizando un intervalo de confianza de un extremo. Con referencia a los datos del problema 9.1 (b), estime la diferencia entre las compras promedio para los clientes "con cupón* y "sin cupón", construyendo un intervalo de confianza del 90% superior.

Dado que, del problema 9.1 (*b*), $\bar{X}_1 - \bar{X}_2 = \$4.97$ y $s_{\bar{x}_1-\bar{x}_2} = 0.772$,

$$\begin{aligned} \text{Est. } (\mu_1 - \mu_2) &\geq (\bar{X}_1 - \bar{X}_2) - z s_{\bar{x}_1-\bar{x}_2} \\ &\geq \$4.97 - (1.28)(0.772) \\ &\geq \$3.98 \end{aligned}$$

9.5 Para los datos de ingresos que se reportaron en el problema 9.2, estime la diferencia máxima entre los niveles promedio de ingresos de la primera y segunda comunidades construyendo un intervalo de confianza del 95% inferior.

Dado que, del problema 9.2 $\bar{X}_1 - \bar{X}_2 = \$800$ y $s_{\bar{x}_1-\bar{x}_2} = \$503.66$,

$$\begin{aligned} \text{Est. } (\mu_1 - \mu_2) &\leq (\bar{X}_1 - \bar{X}_2) + z s_{\bar{x}_1-\bar{x}_2} \\ &\leq \$800 + 1.645(503.66) \\ &\leq \$1\,628.52 \end{aligned}$$

INTERVALOS DE CONFIANZA PARA ESTIMAR LA PROPORCIÓN DE UNA POBLACIÓN

9.6 Un administrador universitario recopila datos sobre una muestra aleatoria nacional de 230 estudiantes inscritos en programas de posgrado en administración de empresas y encuentra que 54 de ellos tienen licenciaturas en administración o contaduría. Estime la proporción de esos estudiantes a nivel nacional, que tienen licenciaturas en administración de empresas o contaduría, utilizando un intervalo de confianza del 90%.

$$\hat{p} = \frac{54}{230} = 0.235$$

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.235)(0.765)}{230}} = \sqrt{\frac{0.179775}{230}} = \sqrt{0.0007816} = 0.028$$

$$\hat{p} \pm z s_{\hat{p}} = 0.235 \pm 1.645(0.028)$$

$$= 0.235 \pm 0.046 \cong 0.19 \text{ a } 0.28$$

Aquí se supone que, a nivel nacional, el número de estudiantes es lo suficientemente grande como para que no se requiera el factor de corrección por población finita.

9.7 En una área metropolitana grande en la que existen 800 expendios de gasolina, para una muestra aleatoria de *n* - 36 expendios, 20 de ellos venden una marca determinada de aceite. Utilizando un intervalo de confianza del 95%, estime (a) la proporción de todos los expendios de gasolina del área que venden ese aceite, y (b) el número total de expendios del área en los que se vende el aceite.

(a) $\hat{p} = \frac{20}{36} = 0.5555 \cong 0.56$

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{(0.56)(0.44)}{36}} = \sqrt{\frac{0.2464}{36}} = \sqrt{0.006844} = 0.083$$

$$\hat{p} \pm zs_{\hat{p}} = 0.56 \pm 1.96(0.083) = 0.40 \text{ a } 0.72$$

(b) (Nota: al igual que en el caso de la solución del problema 8.11 (b) para la media y la cantidad total, *el número total* en una categoría de la población se determina multiplicando los limites de confianza de la proporción por el número total de elementos en la población.)

$$N(\hat{p} \pm zs_{\hat{p}}) = 800(0.40 \text{ a } 0.72) = 320 \text{ a } 576 \text{ estaciones}$$

o

$$N(\hat{p}) \pm N(zs_{\hat{p}}) = 800(0.56) \pm 800(0.16) = 320 \text{ a } 576 \text{ estaciones}$$

DETERMINACIÓN DEL TAMAÑO DE MUESTRA NECESARIO PARA ESTIMAR LA PROPORCIÓN

9.8 Un administrador universitario desea estimar la proporción de estudiantes inscritos en programas de posgrado en administración de empresas, que también tienen licenciaturas en la misma área, con un margen de error del 0.05 y una confianza del 90%. ¿Qué tamaño de muestra debe utilizarse, como mínimo, si no existe ninguna base para estimar el valor apropiado de la proporción antes de tomar la muestra?

Al utilizar la fórmula *(9.12),*

$$n = \left(\frac{z}{2E}\right)^2 = \left(\frac{1.645}{2(0.05)}\right)^2 = (16.45)^2 = 270.60 \cong 271$$

9.9 Con respecto al problema 9.8, ¿cuál es el tamaño mínimo de la muestra que se requiere si la información previa señala que la proporción no es mayor de 0.30?

De la fórmula (9.11)

$$n = \frac{z^2\pi(1-\pi)}{E^2} = \frac{(1.645)^2(0.30)(0.70)}{(0.05)^2} = 227.31 \cong 228$$

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS PROPORCIONES POBLACIONALES

9.10 Al intentar medir la opinión de los profesores de un sistema escolar con respecto a un nuevo plan de estudios, un supervisor escolar recopila muestras aleatorias de 100 padres de familia en cada una de dos regiones residenciales importantes incluidas en el sistema escolar. En la primera región, 70 de los 100 padres de familia señalaron que tenían intención de votar a favor de la proposición, mientras que en la segunda región, 50 de los 100 votantes muestreados indicaron su intención de hacerlo. Estime la diferencia entre las proporciones reales de padres de familia de las dos áreas que tienen la intención de votar a favor de la proposición, utilizando límites de confianza del 95%.

$$s_{\hat{p}_1}^2 = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} = \frac{(0.70)(0.30)}{100} = \frac{0.21}{100} = 0.0021$$

$$s_{\hat{p}_2}^2 = \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} = \frac{(0.50)(0.50)}{100} = \frac{0.25}{100} = 0.0025$$

Por lo tanto,

$$s_{\hat{p}_1-\hat{p}_2} = \sqrt{s_{\hat{p}_1}^2 + s_{\hat{p}_2}^2} = \sqrt{0.0021+0.0025} = \sqrt{0.0046} = 0.068$$

$$(\hat{p}_1 - \hat{p}_2) \pm z s_{\hat{p}_1-\hat{p}_2} = (0.70-0.50) \pm 1.96(0.068) = 0.20 \pm 0.13 = 0.07 \text{ a } 0.33$$

Asi, la diferencia en las proporciones poblacionales se encuentra entre 0.07 y 0.33 (o, lo que es lo mismo, entre 7 y 33%). En esta solución, se supone que el número de padres de familia en cada área es lo suficientemente grande y resulta innecesario utilizar el factor de corrección por población finita.

INTERVALOS DE CONFIANZA DE UN EXTREMO PARA PROPORCIONES

9.11 Al igual que se hizo para la media y para la diferencia entre dos medias (véanse los problemas 8.14. 9.4 y 9.5), puede estimarse una proporción o una diferencia entre dos proporciones utilizando intervalos de confianza de un extremo. Para los datos del problema 9.6, encontrar la proporción *mínima* de estudiantes de posgrado que tienen una licenciatura en administración de empresas o contaduría, utilizando un intervalo de confianza del 90%

Dado que, el problema 9.6, $\hat{p}$ = 0.235 y $s_{\hat{p}}$ - 0.028,

$$\hat{p} - z s_{\hat{p}} = 0.235 - 1.28(0.028) = 0.199 \text{ o mayor}$$

9.12 Para los datos del problema 9.10, ¿cuál es el intervalo de confianza del 95% superior para la diferencia entre las proporciones de personas de la primera y la segunda comunidad que tienen intención de votar a favor del nuevo plan de estudios?

$$(\hat{p}_1 - \hat{p}_2) - z s_{\hat{p}_1-\hat{p}_2} = (0.70-0.50) - 1.645(0.068) = 0.088 \text{ o mayor}$$

Asi, puede afirmarse con un grado de confianza del 95% que la diferencia *mínima* entre las proporciones de votantes en las dos regiones que tienen intenciones de votar a favor de la proposición es de 0.088, u 8.8%

INTERVALOS DE CONFIANZA PARA LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR

9.13 Para la muestra aleatoria de n =12 sobres de café del problema 8.17, la media era de $\bar{X}$=15.97 gr, la varianza fue de s^2 = 0.0224, y la desviación estándar s = 0.15. Estime (a) la varianza y *(b)* la desviación estándar para todos los sobres de café que se empacan en la planta, utilizando intervalos de confianza del 90%.

(a)

$$\frac{(n-1)s^2}{\chi^2_{gl,\,\text{superior}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{gl,\,\text{inferior}}}$$

$$\frac{(11)(0.0224)}{c^2_{11,0.95}} \leq \sigma^2 \leq \frac{(11)(0.0224)}{\chi^2_{11,0.05}}$$

$$\frac{0.2464}{19.68} \leq \sigma^2 \leq \frac{0.2464}{4.57}$$

$$0.0125 \leq \sigma^2 \leq 0.0539$$

(b) $\sqrt{0.0125} \leq \sigma \leq \sqrt{0.0539}$

$$0.11 \leq \sigma \leq 0.23$$

9.14 Al igual que para otros intervalos de confianza (véanse los problemas 8.14, 9.4, 9.5 y 9.11), puede estimarse una varianza o una desviación estándar poblacionales utilizando intervalos de confianza de un extremo. Por lo general, lo que interesa es el límite superior de la varianza o la desviación estándar y, por ello, el intervalo de confianza del extremo inferior es ei intervalo de ese tipo que se utiliza con mayor frecuencia. Para los datos del problema 9.13, ¿cuál es el intervalo de confianza del 90% inferior para estimar la desviación estándar de la población?

$$\text{Est. } \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{gl,\ \text{inferior}}}} \leq \sqrt{\frac{(11)(0.0224)}{\chi^2_{11,0.90}}}$$

$$\leq \sqrt{\frac{0.2464}{5.58}} \leq \sqrt{0.04416} \leq 0.21$$

Asf, puede decirse, con un grado de confianza del 90%, que la desviación estándar de la población no es superior a 0.21.

RESULTADOS POR COMPUTADORA

9.15 En la Tabla 9.1 se presentan las cantidades correspondientes a los pagos de seguro de daños de automóviles para cada una de dos áreas geográficas (en miles de pesos), obtenidos de un número grande de esas indemnizaciones, según los registros de una compañía aseguradora. Se supone que los montos de las indemnizaciones en cada área tienen una distribución aproximadamente normal, y se supone que la varianza es más o menos igual en las dos áreas. Utilizando algún paquete de computación, obtenga el intervalo de confianza del 95% para la diferencia entre el monto promedio de las indemnizaciones en las dos áreas.

Tabla 9.1 Reclamaciones de daños de automóviles en dos ¿reas geográficas (en miles de pesos)

Área 1		Área 2	
\$1033	\$1069	\$1177	\$1146
1274	1121	258	1096
1114	1269	715	742
924	1150	1027	796
1421	921	871	905

Como puede observarse en la parte correspondiente del listado de la figura 9-1, puede concluirse con una confianza del 95% que el monto total de las indemnizaciones en el Área 1 es mayor que en el Área 2, en una cantidad que está entre \$47 000 y \$466 000. La amplitud del intervalo de confianza es muy grande debido a la gran variabilidad de los datos, así como también a lo reducido de la muestra. Debe observarse que, en los paquetes de computación como el que se ilustra en la figura 9-1, cuando no se conocen las varianzas poblacionales, entonces se utiliza la distribución t aun para muestras grandes, porque no es necesario utilizar la aproximación normal cuando se pueden determinar, mediante la computadora, los valores exactos de í para números grandes de grados de libertad. (La parte del listado de la figura 9-1 que está después del intervalo de confianza se analiza en el problema 11.20, que se refiere a pruebas de hipótesis.)

```
MTB > SET CLAIMS FOR AREA-ONE INTO Cl
DOTO> 1033  1274  1114   924  1421  1069  1121  1269  1150   921
DATA> END
MTB > SET CLAIMS FOR AREA-TWO INTO C2
DOTA) 1177   258   715  1027   871  1146  1096   742   796   905
DATA) END
MTB ) TWOSAMPLE TTEST 95% CONFIDENCE INTERVAL.FOR Cl AND C2;
SUBO POOLED.

TWOSAMPLE T FOR Cl VS C2
      N      MEAN     STDEV   SE MEAN
Cl   10      1130       158        50
C2   10       873       272        86

95 PCT Cl FOR MU Cl - MU C2: (47, 466)
TTEST MU Cl - MU C2 (VS NE): T=2.57 P=0.019 DF=18.0
```

Flg.9-1 Resultado de Minitab.

Problemas complementarios

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS MEDIAS POBLACIONALES

9.16 Para un determinado producto de consumo popular, el promedio de ventas por tienda fue, el año anterior, en una muestra de $n_1=10$ tienc $\bar{X}_1 =$ \$3 425 000, con $S_1 =$ \$200 000. Para un segundo producto, el promedio de ventas por tienda de una muestra $n_2 = 12$ tiendas fue de $\bar{X}_2 =$ \$3 250 000, con $s_2 =$\$175 000. Se supone que los montos de las ventas por tienda tienen distribución normal, para ambos productos. Estime la diferencia entre el nivel promedio de ventas por tienda del año anterior utilizando un intervalo de confianza del 95%.

Resp. \$8.28 a \$341.72

9.17 De los datos del problema 9.16, suponga que los tamaños de las muestras fueron $n_1 = 20$ y $n_2 = 24$. Determine el intervalo de confianza del 95% para la diferencia entre las dos medias suponiendo que las varianzas poblacionales no son iguales.

Resp. \$62.83 a \$287.17

9.18 Al utilizar los datos del problema 9.16, suponga que lo que interesa es solamente determinar la diferencia mínima entre los niveles de venta de los dos productos. Determine el límite inferior de ese intervalo de estimación con un grado de confianza del 95%.

Resp. \$37.13 o más

9.19 Para una muestra de 50 empresas de una determinada rama industrial, se encuentra que el número promedio de empleados por empresa es $\bar{X}_1$ - 420.4 con $s_1 =$ 55.7. Existe un total de 380 empresas en esa rama. En una segunda rama industrial que cuenta con 200 empresas, el número promedio de empleados de una muestra de 50 de ellas es $\bar{X}_2 =$ 392. 5 empleados con $s_2 =$ 87.9. Estime la diferencia del número promedio de empleados por empresa en ese ramo industrial, utilizando un intervalo de confianza del 95%.

Resp. 2.3 a 53.5 empleados

9.20 Construya el intervalo de confianza del 99% para la diferencia entre las medias del problema 9.19.

Resp. -5.8 a 61.6 empleados

9.21 **Para** una muestra de 30 empleados de una empresa grande, se encuentra que el salario promedio diario es de $\bar{X}_1 = \$7500$, con $s_1 = \$1000$. En una segunda empresa se determina que el salario promedio diario para una muestra de 40 empleados es $\bar{X}_2 = \$9050$, con $s_2 = \$1200$. Estime la diferencia entre el salario diario promedio de las dos empresas utilizando un intervalo de confianza del 90%.

Resp. \$0.02 a \$0.88 por hora

9.22 Para los datos del problema 9.21, suponga que se desea determinar la diferencia máxima entre los salarios promedio, utilizando un intervalo de confianza del 90%. Construya ese intervalo de confianza del extremo inferior.

Resp. Est. $(\mu_1 - \mu_2) \leq \$0.78$ por hora

INTERVALOS DE CONFIANZA PARA ESTIMAR LA PROPORCIÓN DE UNA POBLACIÓN

9.23 Para una muestra aleatoria de 100 hogares de un área metropolitana grande, el número de hogares en los que cuando menos un adulto está en esos momentos desempleado es 12. Estime el porcentaje de hogares de esa área en los que cuando menos un adulto está desempleado, utilizando un intervalo de confianza del 95%. *(Nota:* pueden obtenerse los límites porcentuales determinando en primer lugar el intervalo de confianza para la proporción y después mutiplicando sus limites por cien.)

Resp. 5.7% a 18.3%

9.24 Suponga que el intervalo de confianza que se obtuvo en el problema 9.23 resulta ser demasiado amplio para propósitos prácticos (es decir, carece de precisión). Se desea, más bien, que el intervalo de confianza del 95% se encuentre a no más de dos puntos porcentuales del porcentaje real de hogares en los que existe cuando menos un adulto desempleado. ¿Cuál es el tamaño mínimo de la muestra que se requiere para satisfacer esta especificación (a) si no se hacen suposiciones acerca de porcentaje real antes de recolectar una muestra de mayor tamaño, y (6) si, con base en la muestra utilizada en el problema 9.22, se supone que el porcentaje real no es superior a 18%?

Resp. (a) 2401, (b) 1418

9.25 Un pequeño fabricante adquiere un lote de 200 partes electrónicas del "exceso de inventario" de una empresa grande. Se encuentra que, para una muestra aleatoria de 50 de las refacciones, 5 de ellas tienen defectos. Estime la proporción de todas las partes del embarque que tienen defectos, utilizando un intervalo de confianza del 95%.

Resp. 0.03 a 0.17

9.26 Para el problema 9.25, estime el número total de refacciones del embarque que tienen defectos, utilizando un intervalo de confianza del 90%.

Resp. 8 a 32

9.27 Para la situación del problema 9.25, suponga que el precio de las refacciones sería satisfactorio para el fabricante si la proporción real de los artículos defectuosos no excede del 0.20. Construya un intervalo de confianza del 95% de un extremo y observe si el límite superior de ese intervalo excede la proporción del 0.20.

Resp. Est. $\pi \leq 0.16$

INTERVALOS DE CONFIANZA PARA LA DIFERENCIA ENTRE DOS PROPORCIONES POBLACIONALES

9.28 En contraste con los datos del problema 9.23, en una segunda área metropolitana, una muestra aleatoria de 1000 hogares arroja que sólo en 6 de ellos existe cuando menos un adulto desempleado que busca un empleo de tiempo completo. Estime la diferencia en el porcentaje de hogares de las dos áreas que incluyen un adulto desempleado, utilizando un intervalo de confianza del 90%.

Resp. -0.6 a +12.6%

9.29 Con referencia al problema 9.28, ¿cuál es el porcentaje máximo en el que el desempleo por hogares en la primera área metropolitana excede al correspondiente a la segunda área, utilizando un intervalo de confianza de un extremo del 90%?

Hesp.Est.Dif. ≤11.1%

INTERVALOS DE CONFIANZA PARA LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR

9.30 Para un determinado producto de consumo popular, el promedio de ventas por expendio, el año pasado, de acuerdo con una muestra de n=10 tiendas,fue $\bar{X}$ = \$3 425 000, con s = 200 000. Se supone que las ventas por expendio tienen una distribución normal. Estime (a) la varianza y (b) la desviación estándar de las ventas de ese producto en todas las tiendas, el año anterior, utilizando un intervalo de confianza del 90%.

Resp. (a) $21\ 278 \leq \sigma^2 \leq 108\ 271$, (b) $145.9 \leq \sigma \leq 329.0$

9.31 Con referencia al problema 9.30, existe preocupación con respecto a *qué tan grande* puede ser la desviación estándar de las ventas. Construya un intervalo de confianza de un extremo del 90% que identifique este valor.

Resp. $\sigma \leq 293.9$

RESULTADOS POR COMPUTADORA

9.32 Una empresa que procesa un gran número de sus pedidos por teléfono tiene dos tipos de clientes: generales y comerciales. En la Tabla 9.2 se presentan los pedidos solicitados por teléfono para una muestra aleatoria de 12 clientes generales y 10 comerciales. Se supone que las cantidades de tiempo que se requieren para atender cada tipo de llamada tienen una distribución aproximadamente normal. Utilice un programa de computación para obtener el intervalo de confianza del 95% para la diferencia en la cantidad promedio de tiempo por llamada que se requiere para cada uno de los dos tipos.

Resp. -23 a 52 seg.

Tabla 9.2 Tiempo requerido (en segundos) para procesar pedidos

Clientes generales	Clientes comerciales
48	81
66	137
106	107
84	110
146	107
139	40
154	154
150	142
177	34
156	165
122	
121	

Pruebas de hipótesis sobre la media de una población

10

10.1 ETAPAS BÁSICAS EN PRUEBAS DE HIPÓTESIS

Al realizar pruebas de hipótesis, se parte de un valor supuesto (hipotético) de un parámetro poblacional. Después de recolectar una muestra aleatoria, se compara la estadística muestral, así como la media ($\bar{X}$), con el parámetro hipotético, se compara con una supuesta media poblacional (M). Después, se *acepta* o se *rechaza* el valor hipotético, según proceda. Se rechaza el valor hipotético sólo si el resultado muestral resulta muy poco probable cuando la hipótesis es cierta.

Etapa 1: *plantear la hipótesis nula y la hipótesis alternativa.* La *hipótesis nula* (H_0) es el valor hiotético del parámetro que se compara con el resultado muestral. Se rechaza *sólo si* el resultado muestral es muy poco probable en caso de que la hipótesis sea cierta. Se acepta la *hipótesis alternativa* (H_1) sólo si se rechaza la hipótesis nula.

EJEMPLO 1. Un auditor desea probar el supuesto de que el valor promedio de todas las cuentas por cobrar en una empresa determinada es $260 000, tomando una muestra de $n = 36$ y calculando la media muestral. Desea rechazar el valor supuesto de $260 000 sólo si la media muestral lo contradice en forma clara, por lo que debe "darse el beneficio de la duda" al valor hipotético en el procedimiento de prueba. Las hipótesis nula y alternativa para esta prueba son H_0: $\mu = \$\ 260\,000$ y H_1: $\mu \neq 260\,000$.

Etapa 2: *Especificar el nivel de significancia que se va a utilizar.* El nivel de significancia es el estándar estadístico que se especifica para rechazar la hipótesis nula. Si se especifica un nivel de significancia del 5%, entonces se rechaza la hipótesis nula solamente si el resultado muestral es tan diferente del valor hipotético que una diferencia de esa magnitud o mayor, pudiera ocurrir aleatoriamente con una probabilidad de 0.05 o menos.

Debe observarse que si se utiliza el nivel de significancia del 5%, existe una probabilidad del 0.05 de rechazar la hipótesis nula cuando, de hecho, es cierta. A esto se le denomina error *tipo* I. La probabilidad del error tipo 1 es siempre igual al nivel de significancia que se utiliza como criterio para rechazar la hipótesis nula; se le designa mediante la letra griega a ("alfa") y , por ello, a designa el nivel de significancia. Los niveles de significancia que se utilizan con mayor frecuencia en las pruebas de hipótesis son el 5 y el 1 %.

Ocurre un *error tipo* II si se acepta la hipótesis nula cuando, de hecho, es falsa. En la sección 10.3 se explica la forma en que se determina la probabilidad del error tipo II. En la Tabla 10.1 se resumen los tipos de decisiones y las consecuencias posibles, al realizar pruebas de hipótesis.

Etapa 3: Elegir la estadística de prueba. La estadística de prueba puede ser la estadística muestral (el estimador no sesgado del parámetro que se prueba) o una versión transformada de esa estadística muestral. Por ejemplo, para probar el valor hipotético de una media poblacional, se toma la media de una muestra aleatoria de esa población para utilizarla como estadística de prueba. Sin embargo, si la distribución de muestreo de la media tiene distribución normal, entonces es común que se transforme la media muestral en un valor z el cual, a su vez, sirve como estadística de prueba.

Tabla 10.1 Consecuencias de las decisiones en pruebas de hipótesis

Decisiones posibles	Situaciones posibles: La hipótesis nula es verdadera	Situaciones posibles: La hipótesis nula es falsa
Aceptar la hipótesis nula	Se acepta correctamente	Error tipo II
Rechazar la hipótesis nula	Error tipo I	Se rechaza correctamente

Etapa 4: Establecer el valor o valores críticos de la estadística de prueba. Habiendo especificado la hipótesis nula, el nivel de significancia y la estadística de prueba que se van a utilizar, se procede a establecer el o los valores críticos de estadística de prueba. Puede haber uno o más de esos valores, dependiendo de si se va a realizar una prueba de uno o dos extremos (véase la sección 10.2). En cualquier caso, un *valor critico* identifica el valor de estadística de prueba que se requiere para rechazar la hipótesis nula.

Efapa *5: Determinar el valor real de la estadística de prueba.* Por ejemplo, al probar un valor hipotético de la media poblacional, se toma una muestra aleatoria y se determina el valor de la media muestral. Si el valor crítico que se establece es un valor de *z,* entonces se transforma la media muestral en un valor de *z.*

Etapa 6: Tomar la decisión. Se compara el valor observado de la estadística muestral con el valor (o valores) críticos de la estadística de prueba. Después, se acepta o se rechaza la hipótesis nula. Si se rechaza ésta, se acepta la alternativa; a su vez, esta decisión tendrá efecto sobre otras decisiones de los administradores operativos, como por ejemplo, mantener o no un estándar de desempeño o cuál de dos estrategias de mercadotecnia utilizar.

10.2 PRUEBA DE UN VALOR HIPOTÉTICO DE LA MEDIA UTILIZANDO LA DISTRIBUCIÓN NORMAL

Puede utilizarse la distribución normal para probar un valor hipotético de la media poblacional: (1) cuando $n > 30$, utilizando el teorema del límite central, o (2) cuando $n < 30$, pero la distribución de la población es normal y se conoce a (véase la sección 8.2).

Se utiliza *una prueba de dos extremos* cuando lo que interesa es una posible desviación en cualquier dirección, a partir del valor hipotético de la media. La fórmula que se utiliza para establecer los valores críticos de la media muestral es similar a la que se utiliza para determinar los límites de confianza para estimar la media de una población (véase la sección 8.3), excepto que el valor hipotético de la media poblacional μ_0 es el punto de referencia, y no la media muestral. Los valores críticos de la media muestral para una prueba de dos extremos, dependiendo de si se conoce a, son:

$$\bar{X}_{CR} = \mu_0 \pm z\sigma_{\bar{x}} \tag{10.1}$$

o

$$\bar{X}_{CR} = \mu_0 \pm zs_{\bar{x}} \tag{10.2}$$

EJEMPLO 2. Para la hipótesis nula que se planteó en el ejemplo 1, determine los valores críticos de la media muestral para probar la hipótesis con un nivel de significancia del 5%. Como se sabe que la desviación estándar de las cuentas por cobrar es o - 43 000, los valores críticos son:

Hipótesis: H_0: μ - \$260 000; H_1: $\mu \neq$ **\$260 000.**
Nivel de significancia: $\alpha = 0.05$
Estadística de prueba $\bar{X}$ con base en una muestra de n = 36, y con σ = 43 000.
$\bar{X}_{CR}$ - valores críticos de la media muestral

$$\bar{X}_{CR} = \mu_0 \pm z\sigma_{\bar{x}} = 260\,000 \pm 1.96 \frac{\sigma}{\sqrt{n}} = 260\,000 \pm 1.96 \frac{43\,000}{\sqrt{36}}$$

$$= 260\,000 \pm 1.96\ 7166.67 = 260\,000 \pm 14\,046.67 = \$245\,953.33 \text{ y } 274\,046.6$$

Por lo tanto, para rechazar la hipótesis nula, la media muestral debe tener un valor inferior a \$245 950 o mayor de \$274 050. Asi, existen dos regiones de rechazo en el caso de una prueba de dos extremos (véase la figura 10.1). Se utilizan los valores de z de ±1.96 para establecer los limites críticos porque para la distribución normal estándar se tiene 0.05 de proporción del área en los dos extremos (0.025 en cada extremo), lo cual corresponde al valor de a « 0.05 que se especifica.

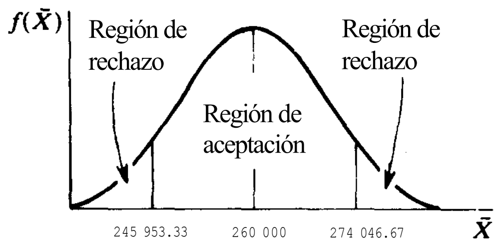


Fig. 10-1

En vez de establecer valores críticos en términos de la media muestral como tal, es común que se especifiquen los valores críticos en las pruebas de hipótesis en términos de valores *z*. Para el nivel de significancia del 5%, los valores críticos z para una prueba de dos extremos son -1.96 y +1.96, por ejemplo. Cuando se determine el valor de la media muestral, se le transforma en un valor z para que pueda compararse con los valores críticos de z. La fórmula de transformación, dependiendo de si se conoce o o no, es

$$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} \tag{10.3}$$

o

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} \tag{10.4}$$

EJEMPLO 3. Para el problema de prueba de hipótesis de los ejemplos 1 y 2, suponga que la media muestral es X = \$240 000. Se determina si se debe aceptar o rechazar la hipótesis nula transformando esa media a un valor z y comparando éste con los valores críticos de ±1.96, de la siguiente manera:

$$\sigma_{\bar{x}} = 7166.67 \qquad \text{(del ejemplo)}$$

$$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{240\,000 - 260\,000}{7166.67} = \frac{-20\,000}{7166.67} = -2.79$$

Este valor de z se encuentra en la región de rechazo del extremo izquierdo del modelo dé prueba de hipótesis que se ilustra en la figura 10-2. Por ello, se rechaza la hipótesis nula y se acepta la alternativa $\mu \neq 60\,000$. Puede observarse que se llega a la misma conclusión comparando la media muestral de $\bar{X} = \$240\,000$ con los limites críticos de la media que se identificaron en la figura 10-1.

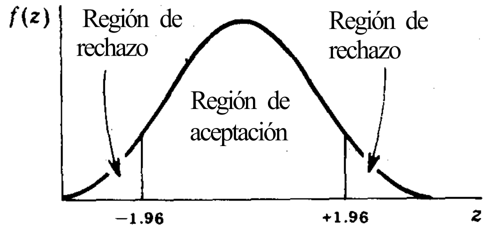


Fig. 10-2

Una *prueba de un extremo* resulta apropiada cuando únicamente interesan las desviaciones en un sólo sentido con respecto al valor hipotético de la media. Al auditor del ejemplo 1 puede no interesarle que el promedio real de todas las cuentas por cobrar exceda \$260 000, sino sólo que sea inferior a esa cantidad. Por ello, si le otorga el beneficio de la duda a la afirmación de que la media verdadera es *cuando menos* \$260 000, la hipótesis nula y la alternativa son:

$$H_0: \mu \geq \$260\,000 \text{ y } H_1: \mu < \$260\,000$$

En las pruebas de un extremo existe una sola región de rechazo y, para el ejemplo anterior, la prueba es del extremo inferior. La región de rechazo para una prueba de un extremo siempre se encuentra en el extremo que representa apoyo para la hipótesis *alternativa.* Al igual que para las pruebas de dos extremos, puede determinarse el valor crítico para la media como tal, o en términos de un valor de z. Sin embargo, los valores críticos para las pruebas de un extremo difieren de los que se utilizan para las pruebas de dos extremos, porque la proporción dada de área total se encuentra en ese extremo de la distribución. En la Tabla 10.2, se presentan los valores de z necesarios para pruebas de uno y dos extremos. La fórmula general para establecer el valor crítico de la media muestral para una prueba de un extremo, dependiendo de si se conoce o o no, es

$$\bar{X}_{CR} = \mu_0 + z\sigma_{\bar{x}} \qquad (10.5)$$

o

$$\bar{X}_{CR} = \mu_0 + zs_{\bar{x}} \qquad (10.6)$$

En las fórmulas 10.5 y 10.6 anteriores, puede observarse que es *posible que z* sea negativa, dando como resultado una substracción del segundo término en cada fórmula.

Tabla 10.2 Valores críticos de z en pruebas de hipótesis

	Tipo de prueba	
Nivel de significancia	De un extremo	De dos extremos
5%	+1.645 (0-1.645)	±1.96
1%	+2.33 (0-2.33)	±2.58

EJEMPLO 4. Suponga que el auditor de los ejemplos 1 a 3, parte de la hipótesis nula de que la media de todas las cuentas por cobrar es de cuando menos $260 000. Si la media muestral es $240 000, se prueba esta hipótesis con un nivel de significancia del 5%, mediante los dos siguientes procedimientos.

(1) Se determina el valor crítico de la media muestral, en donde $H_0 = \mu \geq \$260\ 000$ y $H_1 = \mu < \$260\ 000$

$$\bar{X}_{CR} = \mu_0 + z\sigma_{\bar{x}} = 260\ 000 + (-1.645)(7\ 166.67 = \$248\ 210.82$$

Como $\bar{X} = \$240\ 000$, se encuentra en la sección de rechazo. Por ello, se rechaza la hipótesis nula y se acepta la alternativa de que $\mu < \$\ 260\ 000$.

(2) Se especifica el valor crítico en términos de z, donde el valor critico de z ($\alpha = 0.05$) es -1.645:

$$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{240\ 000 - 260\ 000}{7\ 166.67} = -2.79$$

Como z = -2.79 se encuentra en la sección de rechazo (a la izquierda del valor crítico de -1.645), se rechaza la hipótesis nula. En la figura 10-3 se ilustra el valor crítico para esta prueba de un extremo, en términos de $\bar{X}$ y z.

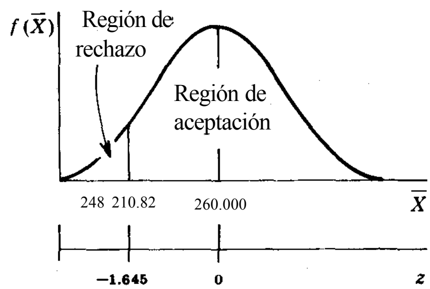


Fig. 10-3

10.3 ERRORES TIPO I Y TIPO II EN PRUEBAS DE HIPÓTESIS

En esta sección se analizan en forma completa los errores tipo I y tipo II (definidos en la sección 10.1), con respecto a las pruebas de un extremo sobre una media hipotética. Sin embargo, los conceptos que se ilustran aquí son aplicables también a otros modelos de pruebas de hipótesis.

La probabilidad del error tipo I es siempre igual al nivel de significancia que se utiliza al probar hipótesis nulas. Esto es así porque, por definición, la proporción de área en la región de rechazo es igual a la proporción de resultados muestrales que ocurrirían en esa región, cuando la hipótesis es verdadera.

Por lo general, a la probabilidad del error del tipo II se le designa mediante la letra griega |3 ("beta"). La única forma en que se puede determinar es con respecto a un valor *específico* incluido dentro del rango de la hipótesis *alternativa.*

EJEMPLO 5. Al igual que en el ejemplo 4, la hipótesis nula que se va a probar es que la media de todas las cuentas por cobrar es cuando menos $260 000, y esta prueba se llevará a cabo con un nivel de significancia del 5%. Además, el auditor señala que consideraría que una media real de $240 000 (o menos) constituye una diferencia material importante con respecto al valor hipotético de la media. Al igual que antes, σ = **\$43 000** y el tamaño de la muestra es n - 36 cuentas. Para determinar la probabilidad del error tipo II, se requiere:

(1) plantear las hipótesis nula y alternativa para esta prueba,
(2) determinar el valor crítico de la media muestral que debe utilizarse para probar la hipótesis nula con un nivel de significancia del 5%,
(3) identificar la probabilidad del error tipo I correspondiente al valor critico que se calculó antes, como base para la regla de decisión,
(4) identificar la probabilidad del error tipo II correspondiente a la regla de decisión, dada una media alternativa especifica de \$240 000.

La solución completa es

(1) $H_0\colon \mu \geq \$260\,000 \qquad H_1\colon \mu < \$260\,000$

(2) $\bar{X}_{CR} = \mu_0 + z\sigma_{\bar{x}} = 260\,000 + (-1.645)(7166.67) = \$248\,210.82$

donde $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{43\,000}{\sqrt{36}} = \dfrac{43\,000}{6} = 7166.67$

(3) la probabilidad del error tipo I es igual a 0.05 (el nivel de significancia que se utiliza para probar la hipótesis nula)
(4) la probabilidad del error tipo II es la probabilidad de que la media de la muestra aleatoria sea igual o superior a \$248 210, dado que la media de todas las cuentas es en realidad \$240 000.

$$z = \frac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{x}}} = \frac{248.21 - 240\ 00}{7166.67} = \frac{8210.83}{7166.67} = +1.15$$

$$P(\text{error tipo II}) = P(z \geq +1.15) = 0.5000 - 0.3749 = 0.1251 \cong 0.13$$

En la figura 10-4 se ilustra el método que se siguió en el ejemplo 5. En general, el valor crítico de la media que se determina con respecto a la hipótesis nula se "reduce" y se utiliza como valor crítico con respecto a la hipótesis alternativa específica. En el problema 10.13 se ilustra la forma de determinar la probabilidad del error tipo II para una prueba de dos extremos.

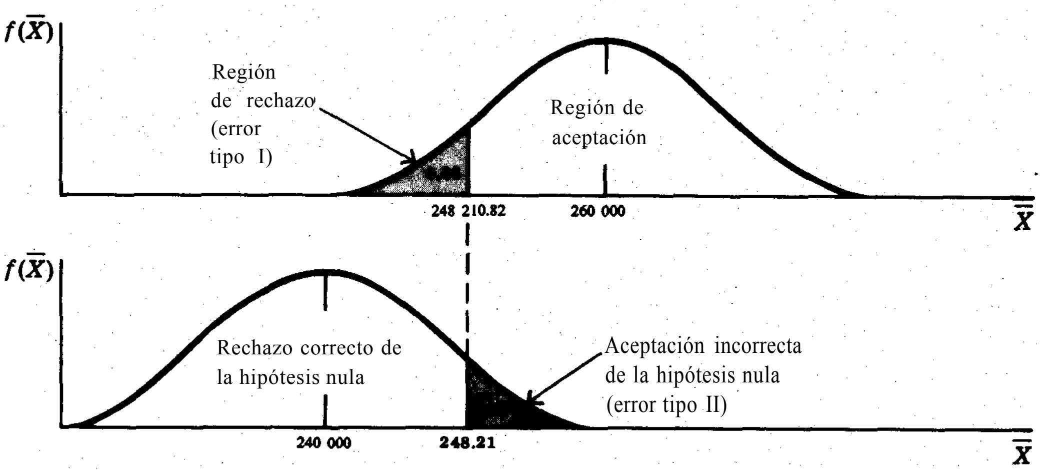


Fig. 10-4

Al mantener constantes el nivel de significancia y el tamaño de la muestra, la probabilidad del error tipo II disminuye conforme el valor de la media alternativa se aleja del valor de la hipótesis nula. Esa probabilidad aumenta al acercarse el valor alternativo al valor de la hipótesis nula. Una curva *característica operativa* (CO) ilustra la probabilidad de aceptar la hipótesis nula dados diversos valores alternativos de la media verdadera. La figura 10.5, es la curva *CO* aplicable a cualquier *prueba del extremo inferior* de una media hipotética, con un nivel de significancia del 5%, y con base en el uso de la distribución normal de probabilidad. Debe observarse que es aplicable a *cualquiera* de esas pruebas, porque los valores sobre el eje horizontal están planteados en unidades del error estándar de la media. Para cualesquiera valores que se encuentren a la izquierda de μ_0, la probabilidad de aceptación señala la probabilidad del error tipo I I. A la derecha de μ_0.las probabilidades señalan la aceptación correcta de la hipótesis nula. Tal como se señala con las líneas punteadas, cuando, de hecho, $\mu = \mu_0$ entonces la probabilidad de aceptar la hipótesis nula es 1 - a, o en este caso, 1 - 0.05 = 0.95.

EJEMPLO 6. Puede verificarse la probabilidad del error tipo II que se determinó en el ejemplo 5, haciendo referencia a la figura 10-5, de la siguiente manera:

Tal como se determinó en el ejemplo 5, μ_0 = \$260 000,μ_1 = \$240 000 y $\sigma_{\bar{x}}$ =7166.67. Por lo tanto, la diferencia entre los dos valores designados de la media *en unidades del error estándar* es

$$z = \frac{\mu_1 - \mu_0}{\sigma_{\bar{x}}} = \frac{240\,000 - 260\,000}{7166.67} = -2.8$$

Con referencia a la figura 10-5, la altura de la curva en el valor de $\mu_0 - 2.8\ \sigma_{\bar{x}}$ sobre el eje horizontal está justamente sobre el 0.10, tal como se muestra en las lineas punteadas. El valor real calculado en el ejemplo 5 es 0.13.

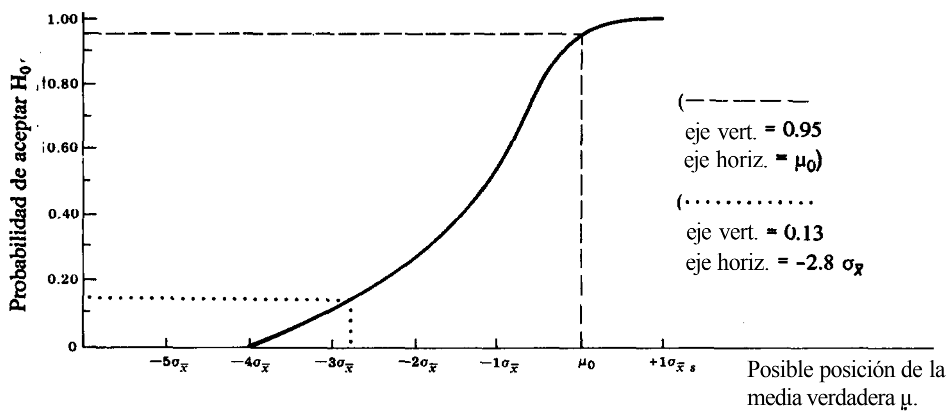


Fig. 10-5

Al realizar pruebas de hipótesis, el concepto de *potencia* se refiere a la probabilidad de rechazar una hipótesis nula dado un valor alternativo específico para el parámetro (en los ejemplos que se han revisado, es la media de la población). Cuando se designa mediante p la probabilidad del error tipo I I, se sigue que la potencia de una prueba es 1 - *ß*. Con referencia a la figura 10-5, puede observarse que la potencia para valores alternativos de la media es la diferencia entre el valor señalado por la curva CO y 1.0 y, por ello, puede obtenerse una *curva de potencia* mediante "substracción" utilizando la curva *CO.*

EJEMPLO 7. Con referencia al ejemplo 5, puede determinarse la potencia de la prueba con el valor alternativo específico de la media de \$240 000, de la siguiente manera:

Como β = *P* (error tipo II) - 0.13 (del ejemplo 5),

$$\text{Potencia} = 1 - \beta = 1.00 - 0.13 = 0.87$$

(Nota: Esta es la probabilidad de rechazar la hipótesis nula, en forma correcta, cuando μ= \$240 000.)

10.4 DETERMINACIÓN DEL TAMAÑO NECESARIO DE LA MUESTRA PARA LA MEDIA

Antes de extraer una muestra, puede determinarse el tamaño que se requiere especificando (1) el valor hipotético de la media; (2) un valor alternativo específico para la media, de manera que la diferencia con respecto al valor hipotético resulte considerable; (3) el nivel de significancia que debe utilizarse en la prueba; (4) la probabilidad del error tipo II que se permite; y (5) el valor de la desviación estándar para la población, σ. La fórmula para determinar el tamaño mínimo que se requiere para la muestra, a fin de probar un valor hipotético de la media con base en la distribución normal es

$$n = \frac{(z_0 - z_1)^2\sigma^2}{(\mu_1 - \mu_0)^2} \tag{10.7}$$

En (10.7), z_0 es el valor crítico de z que se utiliza para el nivel de significación especificado (nivel a), en tanto que z_1 es el valor de z correspondiente a la probabilidad especificada del error tipo II (nivel β). El valor d e σ e b e ser conocido o estimado de alguna manera. Puede utilizarse la fórmula 10.7 para pruebas de uno o dos extremos. El único valor que difiere para los dos tipos de pruebas es el valor de z_0 que se utiliza (véanse los ejemplos 8 y 9).

(Nota: Cuando se está determinando el tamaño mínimo de muestra, siempre se redondean hacia arriba los resultados fraccionarios. Además, si no se conoce σ, o la población no tiene una distribución normal, cualquier tamaño de muestra que se calcule debe aumentarse cuando menos a este valor, porque la fórmula 10.7 se basa en el uso de la distribución normal.)

EJEMPLO 8. Un auditor desea probar la hipótesis nula de que el valor promedio de todas las cuentas por cobrar es de cuando menos \$260 000. Considera que la diferencia entre este valor hipotético y un valor específico alternativo de \$240 000 (o menos) sería considerable. Los niveles aceptables de los errores tipo I (α) y tipo II (β) son 0.05 y 0.10, respectivamente. Se sabe que la desviación estándar de los montos de las cuentas por cobrar es de σ = \$43 000. El tamaño de la muestra que debe extraerse, como mínimo, para llevar a cabo esta prueba es

$$n = \frac{(z_0 - z_1)^2\sigma^2}{(\mu_1 - \mu_0)^2} = \frac{(-1.645 - 1.28)^2(43\,000)^2}{(240\,000 - 260\,000)^2} = \frac{(8.5556)(1\,849\,000\,000)}{400\,000\,000} = 39.55 \cong 40$$

(Nota: Como z_0 y z_1 siempre tienen signos algebraicos contrarios, se tiene que los dos valores zeta siempre se acumulan en el numerador. Si el valor acumulado es negativo, el proceso de elevar al cuadrado da como resultado valores positivos.)

EJEMPLO 9. Suponga que el auditor del ejemplo 8 está preocupado por una discrepancia en *cualquier dirección* con respecto al valor nulo hipotético de \$260 000, y que consideraría que una discrepancia de \$20 000 en cualquier dirección sería importante. Considerando la otra información y las especificaciones del ejemplo 8, el tamaño mínimo de la muestra que debe analizarse es

$$n = \frac{(z_0 - z_2)^2\sigma^2}{(\mu_1 - \mu_0)^2} = \frac{(-1.96 - 1.28)^2\sigma^2}{(240\,000 - 260\,000)^2} \quad \text{o} \quad \frac{[1.96 - (-1.28)]^2\sigma^2}{(280\,000 - 260\,000)^2}$$

$$= \frac{(-3.24)^2(43\,000)^2}{(-20)^2} \quad \text{o} \quad \frac{(3.24)^2(43\,000)^2}{(20)^2}$$

$$= \frac{(10.4976)(1\,849\,000\,000)}{400\,000\,000} = 48.53 \cong 49$$

(Nota: Como las desviaciones con respecto al valor hipotético sólo pueden darse en una dirección, se utiliza el valor de +1.96 o -1.96 como valor de z_0, con el correspondiente valor de z_1. Al igual que en el ejemplo 8, los dos valores de z se acumulan siempre antes de elevarlos al cuadrado.)

10.5 PRUEBA DE UN VALOR HIPOTÉTICO DE LA MEDIA UTILIZANDO LA DISTRIBUCIÓN *t* DE STUDENT

La distribución t (véase la sección 8.5) es apropiada como estadística de prueba cuando la muestra es pequeña (n < 30), la población tiene distribución normal, y se desconoce σ. El procedimiento que se utiliza para probar un valor hipotético de una media poblacional es idéntico al que se describió en la sección 10.2, excepto por el uso del como estadística de prueba. La estadística de prueba que se utiliza es

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} \tag{10.8}$$

EJEMPLO 10. Se ha planteado la hipótesis nula de que la vida útil promedio de los focos de una marca específica es cuando menos de 4200 horas. La vida útil promedio para una muestra aleatoria de *n* = 10 focos es X = 4000 horas,con desviación estándar muestral de s = 200 horas. En términos generales, se supone que la vida útil de los focos tiene una distribución normal. Se prueba la hipótesis nula con un nivel de significancia del 5%, de la siguiente manera:

$$H_0: \mu \geq 4\,200 \qquad H_1: \mu < 4\,200$$

$$t \text{ crítica } (gl = 9, \alpha = 0.05) = -1.833$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{200}{\sqrt{10}} = \frac{200}{3.16} = 63.3 \text{ hrs.}$$

$$= \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{4\,000 - 4\,200}{63.3} = \frac{-200}{63.3} = -3.16$$

Como -3.16 se encuentra en la región de rechazo del extremo izquierdo (a la izquierda del valor critico -1.833), la hipótesis nula se rechaza y se acepta la alternativa de que la vida útil promedio es inferior a 4200 horas.

10.6 EL MÉTODO DEL VALOR *P* PARA PROBAR HIPÓTESIS NULAS REFERENTES A UNA MEDIA POBLACIONAL

Al seguir el método del valor P, en vez de comparar el valor observado de un estadístico de prueba con un valor crítico, se determina la probabilidad de ocurrencia del estadístico de prueba, suponiendo que la hipótesis nula es cierta, y se le compara con el nivel de significancia α. Se rechaza la hipótesis nula si el valor *P* es *inferior* al nivel designado α.

EJEMPLO 11. Con referencia al ejemplo 4, en el que Ho: $\mu \geq$ \$260 000, Hi: $\mu <$ \$260 000α = 0.05, $\bar{X}$ - \$240 000, y la estadística de prueba calculada es z = -2.79, como la estadística z se encuentra en dirección de la región de rechazo, se determina la probabilidad de que ocurra un valor de z de esa magnitud en forma aleatoria:

$$P(z \leq -2.79) = 0.500 - 0.4974 = 0.0026$$

Como el valor de P de 0.0026 es inferior al nivel de significancia especificado, α - 0.05, se rechaza la hipótesis nula.

Para pruebas de dos extremos, se *duplica* el valor P que se calcula con respecto a un extremo de la distribución, y se le compara con el nivel de significancia **a** (véase el problema 10.19).

Cuando se utiliza t como estadística de prueba y se utiliza también una tabla para la distribución t, no es posible determinar un valor exacto de P. Más bien, se identifica un rango de valores de probabilidad (véase el problema 10.21).

10.7 EL MÉTODO DEL INTERVALO DE CONFIANZA PARA PROBAR HIPÓTESIS NULAS REFERENTES A MEDIAS POBLACIONALES

De acuerdo con este método, se construye un intervalo de confianza para la media de la población con base en los resultados muestrales y después se observa si el valor hipotético de la media pobiacíonal queda incluido dentro de este intervalo. Si el valor hipotético está dentro del intervalo, entonces no es posible rechazar la hipótesis nula. Si el valor hipotético no está dentro del intervalo, entonces se rechaza la hipótesis nula. Se construye el intervalo 1 - α, en donde a es el nivel de significancia que se utiliza para la prueba.

EJEMPLO 12. Con referencia al ejemplo 3, en el que Ho: μ -\$260 000, H_1: $\mu \neq$ \$260 000 α =\$0.05, $\bar{X}$ =\$240 000, y $\sigma_{\bar{x}}$ = 7166.67, puede probarse la hipótesis nula con un nivel de significancia del 5%, construyendo el intervalo de confianza del 95%:

$$\bar{X} \pm z\sigma_{\bar{x}} = 240\,000 \pm 1.97\,(7166.67) = 240\,000 \pm 14\,046.87$$
$$= \$225\,953.13 \text{ a } \$254\,046.87$$

Como el valor hipotético de \$260 000 no está dentro del intervalo de confianza del 95%, la hipótesis nula se rechaza con un nivel de significancia del 5%.

Para pruebas de un extremo, se construyen intervalos de confianza de un extremo (véase el problema 10.23).

10.8 TABLA RESUMEN PARA PROBAR UN VALOR HIPOTÉTICO DE UNA MEDIA

Tabla 10.3 Tabla de resumen de la prueba de un valor hipotético para la media

Población	Tamaño de la muestra	a conocida	σ desconocida
Con distribución normal	Grande ($n \geq 30$)	$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}}$	$z = \frac{\bar{X} - \mu_0}{s_{\bar{x}}}$**
	Pequeña ($n < 30$)	$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}}$	$t = \frac{\bar{X} - \mu_0}{s_{\bar{x}}}$
Sin distribución normal	Grande ($n \geq 30$)	$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}}$*	$z = \frac{\bar{X} - \mu_0}{s_{\bar{x}}}$†
	Pequeña ($n < 30$)	Por lo general, se utilizarían pruebas no paramétricas relacionadas con la media (véase capítulo 21).	

* Se aplica el teorema central del límite
**Se utiliza z como aproximación de t
+ Se utiliza el teorema central del límite y se aplica z como aproximación de *t*

10.9 RESULTADOS POR COMPUTADORA

Los programas de computación para probar hipótesis sobre medias permiten al usuario especificar el valor hipotético, así como solicitar pruebas de uno o dos extremos. Por lo general, se utiliza el método de *P* para las pruebas de hipótesis, según se describió en la sección 10.6. En los problemas 10.24 y 10.25 se ilustra el uso de esos programas para pruebas de dos o un extremo, respectivamente.

Problemas resueltos

PRUEBA DE UN VALOR HIPOTÉTICO DE LA MEDIA UTILIZANDO LA DISTRIBUCIÓN NORMAL

10.1 El representante de una comunidad informa a una empresa que se propone construir un centro comercial cuyo ingreso promedio por hogar del área es de \$25 000 diarios. Suponga que, para ese tipo de área, puede suponerse que los ingresos por hogar tienen una distribución aproximadamente normal y que puede suponerse que la desviación estándar es σ - \$2000, con base en un estudio previo. Se encuentra que el ingreso promedio por hogar para una muestra aleatoria de n = 15 hogares es $\bar{X}$= \$24 000. Pruebe la hipótesis nula de que μ = \$25 000, estableciendo limites críticos para la media muestral en términos de pesos y utilizando el nivel de significancia del 5%.

(Nofa; Puede utilizarse la distribución normal de probabilidad aun cuando la muestra es pequeña, porque se supone que la población tiene distribución normal y se conoce o.)

Como H_0 : μ = \$25 000, y H_1 : $\mu \neq$ \$25 000, los límites críticos de $\bar{X}$ (α 0.05) son

$$\bar{X}_{CR} = \mu_0 \pm z\sigma_{\bar{x}} = \mu_0 \pm z\left(\frac{\sigma}{\sqrt{n}}\right) = 25\ 000 \pm 1.96\left(\frac{2\ 000}{\sqrt{15}}\right)$$

$$= 25\ 000 \pm 1.96\left(\frac{2\ 000}{3.87}\right) = 25\ 000 \pm 1.96(516.80) = \$23\ 987 \quad \text{y} \quad \$26\ 013$$

Como la media muestral de $\bar{X}$ - \$24 000 se encuentra entre los dos límites críticos y dentro de la región de aceptación de la hipótesis nula, no puede rechazarse la afirmación del representante de la comunidad con un nivel de significación del 5%.

10.2 Pruebe la hipótesis del problema 10.1, utilizando la variable normal estándar z como estadística de prueba.

$$H_0: \mu = \$25\ 000 \qquad H_1: \mu \neq \$25\ 000$$

$$z \text{ crítica } (\alpha = 0.05) = \pm 1.96$$

Por ello

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2\ 000}{\sqrt{15}} = \frac{2\ 000}{3.87} = \$516.80$$

$$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{24\ 000 - 25\ 000}{516.80} = \frac{-1\ 000}{516.80} = -1.93$$

Como el valor calculado de z de -1.93 se encuentra en la región de aceptación de la hipótesis nula, no puede rechazarse la afirmación del representante de la comunidad con un nivel de significancia del 5%.

10.3 Con referencia a los problemas 10.1 y 10.2, el constructor no está en realidad preocupado por la posibilidad de que el ingreso promedio por hogar sea superior a los \$25 000 que se han supuesto, sino que sólo le preocupa que pueda ser inferior. Por ello, replantee las hipótesis nula y alternativa y realice la prueba de hipótesis apropiada, concediendo el beneficio de la duda a la afirmación del representante de la comunidad.

$$H_0: \mu \geq \$25\ 000 \qquad H_1: \mu < \$25\ 000$$

$$z \text{ crítica } (\alpha = 0.05) = -1.645$$

$$z = -1.93 \qquad (\text{del problema } 10.2)$$

El valor calculado de z (-1.93) es inferior al valor crítico de *z* (-1.645) para esta prueba del extremo izquierdo. Por ello, se rechaza la hipótesis nula con un nivel de significancia del 5% y se acepta la hipótesis alternativa, que afirma que el ingreso promedio por hogar es inferior a $25 000. La razón del cambio en la decisión con respecto ai problema 10.2 se debe a que la región del rechazo del 5% se localiza en su totalidad en uno de los extremos de la distribución.

10.4 Para el problema 10.3, suponga que no se conoce la desviación estándar de la población, lo cual es común, y que no se supone que la población de ingresos tiene distribución normal. Se toma una muestra de $n = 30$ hogares, y se encuentra que la desviación estándar es s = $2 000 y la media muestral sigue siend($\bar{X}$= $ 24 000. Pruebe la hipótesis nula de que el ingreso promedio por hogar es de cuando menos $25 000 en la población utilizando un nivel de significancia del 5%.

(Nota: Puede utilizarse la distribución normal de probabilidad con base en el teorema del limite central y debido a que puede utilizarse z como aproximación de f cuando $n \geq 30$.)

$$H_0: \mu \geq \$25\,000 \qquad H_1: \mu < \$25\,000$$

$$z \text{ crítica } (\alpha = 0.05) = -1.645$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{2\,000}{\sqrt{30}} = \frac{2\,000}{5.48} = \$364.96$$

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{24\,000 - 25\,000}{364.96} = \frac{-1\,000}{364.96} = -2.74$$

El valor calculado de *z* de -2.74 es inferior al valor critico de -1.645 para esta prueba del extremo inferior. Por ello, se rechaza la hipótesis nula con un nivel de significancia del 5%. Se debe observar que, en este caso, el valor calculado de z es menor y se encuentra en forma más clara en la región de rechazo, en comparación con el problema 10.3. Esto se debe al aumento en el tamaño de la muestra de n = 15 a n = 30, lo cual da como resultado un valor menor para el error estándar de la media.

10.5 Un fabricante está considerando la adquisición de un nuevo equipo para la fabricación de herramientas y especifica que, en promedio, el equipo no debe requerir más de diez minutos de tiempo de preparación por hora de operación. Un agente de compras visita una compañia en donde existe instalado otro equipo como el que se está considerando. De los registros que se tienen ahí, observa que una muestra aleatoria de 40 horas de operación incluye un total de 7 horas y 30 minutos de tiempo de preparación, con una desviación estándar del tiempo de preparación por hora de 3.0 minutos. Con base en este resultado muestral, ¿puede rechazarse la suposición de que el equipo satisface las especificaciones de tiempo de preparación, con un nivel de significancia de 1%?

$$H_0: \mu \leq 10.0 \text{ min (por hora)} \qquad H_1: \mu > 10.0 \text{ min (por hora)}$$

$$z \text{ crítica } (\alpha = 0.01) = +2.33$$

$$\bar{X} = \frac{\Sigma X}{n} = \frac{450 \text{ min}}{40} = 11.25\text{min}$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{3.0}{\sqrt{40}} = \frac{3.0}{6.32} = 0.47 \text{ min}$$

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{11.25 - 10.0}{0.47} = \frac{1.25}{0.47} = +2.66$$

El valor calculado de z de +2.66 es mayor que el valor crítico de +2.33 para esta prueba del extremo superior. Por ello, se rechaza la hipótesis nula en el nivel de significancia del 1 %, y se acepta la hipótesis alternativa de que el tiempo promedio de preparación para ese equipo es mayor de 10 minutos por hora de operación.

10.6 Se sabe que la desviación estándar de la vida útil de una marca determinada de tubos de luz ultravioleta es de σ 500 horas, y que la vida útil de los tubos tiene distribución normal. El fabricante afirma que la vida útil promedio de los tubos es de cuando menos nueve mil horas. Pruebe esta afirmación con un nivel de significancia del 5%, considerándola como hipótesis nula, y suponiendo que la vida útil promedio para una muestra de n = 15 tubos fue de $\bar{X}=$ 8800 horas.

$$H_0: \mu \geq 9\,000 \qquad H_1: \mu < 9\,000$$

$$z \text{ crítica } (\alpha = 0.05) = -1.645$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{15}} = \frac{500}{3.87} = 129.20$$

$$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{8\,800 - 9\,000}{129.20} = \frac{-200}{129.20} = -1.55$$

El valor calculado de z de -1.45 no es *inferior* al valor crítico de z = -1.645 para esta prueba del extremo inferior. Por ello, no puede rechazarse la hipótesis nula a un nivel de significancia del 5%.

10.7 Con respecto al problema 10.6, suponga que los datos muestrales se obtuvieron de una muestra de n = 35 tubos. Pruebe la afirmación con un nivel de significancia del 5%.

$$H_0: \mu \geq 9\,000 \qquad H_1: \mu < 9\,000$$

$$z \text{ crítica } (\alpha = 0.05) = -1.645$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{500}{\sqrt{35}} = \frac{500}{5.92} = 84.46$$

$$z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{8\,800 - 9\,000}{84.46} = \frac{-200}{84.46} = -2.37$$

El valor calculado de z de -2.37 es inferior al valor crítico de -1.645 para esta prueba del extremo inferior. Por ello, no es posible rechazar la hipótesis nula con un nivel de significancia del 5%.

10.8 Un analista de investigación de mercados recolecta datos para una muestra aleatoria de 100 clientes que adquirieron un "cupón especiar. Las 100 personas gastaron un promedio de $\bar{X}$ = \$24 570 en la tienda, con desviación estándar de s = \$10 600. Antes de obtener estos resultados muestrales, el gerente de mercadotecnia había afirmado que la compra promedio para los clientes que habían aceptado la oferta de los cupones sería de cuando menos \$25 000. ¿Puede rechazarse esta afirmación utilizando un nivel de significancia del 5%?

$$H_0: \mu \geq \$25\,000 \qquad H_1: \mu < \$25\,000$$

$$z \text{ crítica } (\alpha = 0.05) = -1.645$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{6\,600}{\sqrt{100}}\sqrt{\frac{400-100}{400-1}} = \frac{6\,600}{10}\sqrt{0.7519} = 660(0.867) = 572.22$$

(Se requiere el factor de corrección de población finita porque $n \geqslant 0.05$ N.)

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{24.57 - 25.00}{572.22} = \frac{-430}{572.22} = -0.75$$

La z calculada de -.75 *no es inferior* al valor crítico de -1.645 para esta prueba del extremo inferior. Por ello, no puede rechazarse la afirmación con un nivel de significancia del 5%.

ERRORES TIPO I Y TIPO II EN PRUEBAS DE HIPÓTESIS

10.9 Para el problema 10.3, suponga que el constructor consideraría como una discrepancia importante que el promedio de los ingresos diarios por hogar fuera de $23 500, o menos, en contraste con los $25 000 que se afirma como dato. Identifique (a) la probabilidad del error tipo I, (b) la probabilidad del error tipo II, y (c) la potencia asociada con esta prueba del extremo inferior.

(a) P (error tipo I) = 0.05 (el nivel α o nivel de significancia)
(b) P (error tipo II) =*P* (de que se exceda el límite crítico de $\bar{X}$, dado μ = $23 500).

Límite inferior crítico de $\bar{X} = \mu_0 + z\sigma_{\bar{x}} = 25\ 000 + (-1.645)(516.80) = \$24\ 149.86$

en donde $\mu_0 = \$25\ 000$

$z = -1.645$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2\ 000}{\sqrt{15}} = \frac{2\ 000}{3.87} = \$516.80$$

$$\text{P(del error tipo II)} = P(\bar{X} \geq 24\ 149.86 \mid \mu_1 = 23\ 500,\ \sigma_{\bar{x}} = 516.80)$$

$$z_1 = \frac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{x}}} = \frac{24\ 149.86 - 23\ 500}{516.80} = \frac{649.86}{516.80} = 1.26$$

$$\text{P(del error tipo II)} = P(z_1 \geq +1.26) = 0.5000 - 0.3962 = 0.1038 \cong 0.10$$

(c) Potencia - 1 - P(del error tipo II) = 1.00 - 0.10 = 0.90

10.10 Con respecto a los problemas 10.3 y 10.9, el valor nulo hipotético específico de la media es μ_0 - $25 000 y el valor alternativo específico es μ_T $23 500. Con base en la muestra de *n* =15, $\sigma_{\bar{x}}$ 516.80 y el valor crítico de la media muestra $\bar{X}_{CR}$ $24149.86, (a) Explique qué se quiere decir cuando se señala que este último valor está por encima de un "valor crítico", y (b) si el verdadero valor de la población es μ= $23 000, ¿cuál es la potencia de la prueba estadística?

(a) El valor de $24149.86 es el valor crítico" de la media muestral porque si la media muestral es igual a este valor o lo excede, se aceptará la hipótesis nula (*H_o*: $\mu \geq$ $25 000), en tanto que si la media muestral es inferior a este valor, se rechazará la hipótesis nula y se aceptará la alternativa (H_1 : μ < $25 000).

(b) Potencia = P (Rechazar una hipótesis nula falsa dado un valor especifico de μ). En este caso, potencia $P(\bar{X}) < 24149.86|\mu_1 = 23000, \sigma_{\bar{x}} = 516.80$

$$z = \frac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{x}}} = \frac{24\ 149.86 - 23\ 000}{516.80} = \frac{1\ 149.86}{516.80} = +2.22$$

$$\text{Potencia} = P(z < +2.22) = 0.5000 + 0.4868 = 0.9868 \cong 0.99$$

En otras palabras, dada una media poblacional de μ- $23 000, existe una probabilidad de aproximadamente 99% de que se rechace la hipótesis nula (lo cual seria correcto) y aproximadamente una probabilidad del 1 % de que sea aceptada (lo cual es incorrecto) y constituye un error de tipo II).

10.11 Con referencia al problema 10.10

(a) ¿Cuál es la potencia de la prueba estadística si la media poblacional es μ= $24 000? Compare esta respuesta con el valor que se obtuvo para μ= $23 000 en la respuesta del problema 10.10 (b).

(b) ¿Cuál es la potencia de la prueba estadística si la media poblacional es μ= $24 500? Compare esta respuesta con los valores que se obtuvieron para μ=$23 000 del problema 10.10 (b) y de μ= $24 000 del inciso anterior.

(a) Potencia $= P(\bar{X} < 24\ 149.86 | \mu = 24\ 000,\ \sigma_{\bar{x}} = 516.80) = P(z < 0.29)$

$= 0.5000 + 0.1141 = 0.6141 \cong 0.61$

en donde $z = \frac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{x}}} = \frac{24\ 149.86 - 24\ 000}{516.80} = \frac{149.86}{516.80} = +0.29$

En este caso, la potencia de la prueba es considerablemente menor que la potencia de 0.99 del problema 10.10 (b). Esto seria de esperar, puesto que la media actual de μ= $24 000, es un valor más cercano a la media hipotética nula μ ≥ $25 000.

(b) Potencia $= P(\bar{X} < 24\ 149.86 | \mu = 24\ 500,\ \sigma_{\bar{x}} = 516.80) = P(z < -0.68)$

$= 0.5000 - 0.2518 = 0.2482 \cong 0.25$

en donde $z = \frac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{x}}} = \frac{24\ 149.86 - 24\ 500}{516.80} = \frac{-350.14}{516.80} = -0.68$

La potencia de la prueba es considerablemente inferior que los valores previos de 0.99 y 0.61.

10.12 En los problemas 10.9, 10.10 y 10.11, se determinaron las potencias de pruebas estadísticas de un extremo correspondientes a valores alternativos de la media poblacional de $24 500, $24 000, $23 500 y $23 000. Estas potencias resultaron ser 0.25, 0.61. 0.89 y 0.99, respectivamente. Determine los valores de las CO asociados con estos valores alternativos de la media poblacional, y también para el caso en que μ= $25 000 (con H_0: μ ≥ $25 000).

Como el valor CO = P (aceptar la hipótesis nula dado un valor especifico de μ), el valor CO = 1 - potencia (para cada valor específico de μ) tal como se reporta en la Tabla 10.4.

Tabla 10.4 Potencia y valores de la *CO*

Valor de μ	Potencia	Valor de la *CO*
$25 000	0.05	0.95
24 500	0.25	0.75
24 000	0.61	0.39
23 500	0.90	0.10
23 000	0.99	0.01

En la Tabla 10.4, como $25 000 es el punto en el que comienza el rango de valores hipotéticos-nulos, la "potencia" es, de hecho, la probabilidad del rechazo incorrecto de la hipótesis nula y es, también, igual al nivel de significancia y a la probabilidad del error tipo I. De manera similar, cuando μ es inferior a $25 000 en la tabla, el valor *CO* es la probabilidad de aceptar incorrectamente la hipótesis nula y, por ello, es la probabilidad del error tipo II.

10.13 Para los problemas 10.1 y 10.2, suponga que al constructor le resultarla excesivo que los ingresos promedio diario por hogar difirieran de los supuestos $25 000 en $1 500 o más, en *cualquier* dirección. Como la hipótesis nula se está probando a un nivel de significancia del 5%, identifique la probabilidad del error tipo I y del error tipo II.

(Nota: En esta solución se amplia la explicación que se dio en la sección 10.3, porque se trata de una prueba de dos extremos, en vez de la que se vio antes de un extremo.)

P(del error tipo I) = 0.05 (el nivel α, o nivel de significancia)

P(del error tipo II) = P(de que se exceda el límite crítico inferior de $\bar{X}$, dado μ= $23 500) o

= P(de que $\bar{X}$ esté por debajo del límite superior crítico, dado μ = $26 500)

(Nota: Cualquiera de los cálculos producirla la solución para la probabilidad del error tipo II. Estos dos valores *no* se acumulan porque el valor alternativo especifico sólo puede encontrarse en un punto en alguna situación determinada. El cálculo que aparece enseguida corresponde al primero de los dos métodos alternativos.)

Como, del problema 10.1, el límite crítico inferior de $\bar{X}$ = $23 987 y $\sigma_{\bar{x}}$ = $516.80,

$$P(\text{del error tipo II}) = P(\bar{X} \geq 23\,987 \mid \mu_1 = 23\,500,\ \sigma_{\bar{x}} = 516.80) = P(z_1 \geq +0.94)$$
$$= 0.5000 - 0.3264 = 0.1736 \cong 0.17$$

en donde $z_1 = \dfrac{\bar{X}_{CR} - \mu_1}{\sigma_{\bar{x}}} = \dfrac{23\,987 - 23\,500}{516.80} = \dfrac{487}{516.80} = +0.94$

DETERMINACIÓN DEL TAMAÑO DE MUESTRA NECESARIO PARA PROBAR LA MEDIA

10.14 Suponga que el posible constructor del centro comercial del problema 10.3 desea probar la hipótesis nula $H_0 : \mu \geq$ $25 000 con un nivel de significancia del 5%, y que considera que sería una diferencia importante que el nivel promedio de ingresos de los hogares se encontrara en $23 500 (o menos). Como le preocupa en particular el error de construir un centro comercial en una área en la que no pueda tener éxito, fija el nivel deseado del error tipo II en β = 0.01. Si se supone que la desviación estándar de esos datos es $2 000, determine el tamaño de la muestra que se requiere para lograr los objetivos del constructor con respecto a los errores tipo I y tipo II, si no se hacen suposiciones de normalidad en la población.

$$n=\frac{(z_0-z_1)^2\sigma^2}{(\mu_1-\mu_0)^2}=\frac{[-1.645-(+2.33)]^2(2\ 000)^2}{(23\ 500-25\ 000)^2}=\frac{(-3.975)^2(4\ 000\ 000)}{(-1\ 500)^2}$$

$$=\frac{63\ 202\ 500}{2\ 250\ 000}=28.09=29 \text{ hogares}$$

Sin embargo, como no se supone que la población tenga distribución normal, se aumenta el tamaño de muestra a n = 30, para que pueda aplicarse el teorema del limite central, y de esa manera, poder utilizar la distribución normal de probabilidad.

10.15 El fabricante que está considerando la adquisición del equipo de fabricación de herramientas del problema 10.5, ha especificado en la hipótesis nula que el tiempo promedio de preparación para el equipo es igual o menor a 10 minutos por hora, ($H_0: \mu \leq 10.0$). Suponga que le preocupa en particular que el tiempo promedio de preparación por hora sea de 12.0 minutos o más. Estima que la desviación estándar es de aproximadamente σ = 3.0 minutos, y supone que una variable como la del tiempo de preparación tiene probablemente distribución normal. Debido a las implicaciones de su decisión a largo plazo, señala que la probabilidad de los errores tipo I y tipo II debe mantenerse en, cuando mucho, 0.01. ¿Cuántas horas seleccionadas al azar deben muestrearse para el equipo en preparación, como mínimo, para satisfacer los objetivos de esta prueba?

$$n=\frac{(z_0-z_1)^2\sigma^2}{(\mu_1-\mu_0)^2}=\frac{[2.33-(-2.33)]^2(3.0)^2}{(12.0-10.0)^2}$$

$$=\frac{(4.66)^2(9)}{(2.0)^2}=\frac{21.7156(9)}{4.0}=\frac{195.4404}{4.0}=48.86=49 \text{ hrs.}$$

PRUEBA DE UN VALOR HIPOTÉTICO DE LA MEDIA UTILIZANDO LA DISTRIBUCIÓN *t* DE ESTUDIANTE.

10.16 Como modificación del problema 10.3, un representante de la comunidad le informa al posible constructor del centro comercial que el ingreso promedio por hogar de la comunidad es de *cuando menos* μ = \$25 000, y se le da a esta afirmación el beneficio de la duda en el procedimiento de la prueba de hipótesis. Al igual que antes, se supone que las cifras de ingresos de la población tienen distribución normal. Para una muestra aleatoria de n = 15 hogares, se encuentra que la media muestral es $\bar{X}$= \$24 000, y que la desviación estándar muestral es s = \$2000. Pruebe la hipótesis nula con un nivel de significancia del 5%.

(Nota: La distribución í es apropiada porque la muestra es pequeña *(n* < 30), se desconoce σ y se supone que la población tiene distribución normal.)

$$H_0: \mu \geq \$25\ 000 \qquad H_1: \mu < \$25\ 000$$

$$\text{t critica } (gl = 14, \alpha = 0.05) = -1.761$$

$$s_{\bar{x}}=\frac{s}{\sqrt{n}}=\frac{2\ 000}{\sqrt{15}}=\frac{2\ 000}{3.87}=516.80$$

$$t=\frac{\bar{X}-\mu_0}{s_{\bar{x}}}=\frac{24\ 000-25\ 000}{516.80}=\frac{-1\ 000}{516.80}=-1.935$$

El valor calculado de t de -1.935 es inferior al valor crítico de -1.761 para esta prueba del extremo inferior. Por lo tanto, se rechaza la hipótesis nula a un nivel de significancia del 5%. La decisión es la misma que la del problema

10.3, en el que se conocía σ, excepto que en aquel caso se utilizó z como estadística de prueba y el valor crítico fue z = -1.645.

10.17 De 100 alumnos de contaduría de una escuela de administración de empresas, una muestra aleatoria de n = 12 estudiantes tiene un promedio global de calificaciones de 7 (en donde la calificación más alta es 10) con una desviación estándar muestral de s = 1. Se supone que las calificaciones de estos estudiantes tienen distribución normal. Pruebe la hipótesis de que la calificación promedio global para todos los estudiantes de contaduría de esta escuela es de cuando menos 8.0, utilizando un nivel de significancia del 1 %.

$$H: \mu \geq 8 \qquad H_1: \mu < 8$$

$$t \text{ crítica } (gl = 11, \alpha = 0.01) = -2.718$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{1}{\sqrt{12}}\sqrt{\frac{100-12}{100-1}}$$

$$= \frac{1}{3.46}\sqrt{\frac{88}{99}} = 0.089(0.943) = 0.273$$

(Se requiere el factor de corrección por población finita porque $n >$ 0.05 N.)

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{7-8}{0.273} = \frac{-1}{0.273} = -3.669$$

El valor calculado de t de -2.752 es menor que el valor crítico de -2.718 para esta prueba del extremo inferior. Por ello, se rechaza la hipótesis nula con un nivel de significancia del 1 %. Si se hubiera rechazado el factor de corrección finita, el valor f calculado hubiera sido -2.586, y se hubiera aceptado erróneamente la hipótesis nula.

10.18 En su calidad de comprador comercial para una marca privada de un supermercado, suponga que toma una muestra aleatoria de 12 sobres de café de una empacadora. Se encuentra que el peso promedio del contenido de café de cada sobre es $\bar{X}$= 15.97 gr., con s = 0.15. Los empacadores afirman que el peso neto promedio mínimo del café es de 16.0 gr. por sobre. ¿Puede rechazar esa afirmación con un nivel de significancia del 10%?

$$H_0: \mu \geq 16.0 \qquad H_1: \mu < 16.0$$

$$t \text{ crítica } (gl = 11, \alpha = 0.10) = -1.363$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.15}{\sqrt{12}} = \frac{0.15}{3.46} = 0.043 \text{ gr.}$$

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{x}}} = \frac{15.97 - 16.00}{0.043} = \frac{-0.03}{0.043} = -0.698$$

El valor calculado de t de - 0.698 *no es inferior* al valor crítico de -1.363 para esta prueba del extremo inferior. Por ello, no puede rechazarse la afirmación con un nivel de significancia del 10%.

EL MÉTODO DEL VALOR DE *P* PARA PROBAR HIPÓTESIS NULAS SOBRE LA MEDIA DE LA POBLACIÓN

10.19 Al utilizar el método del valor P, pruebe la hipótesis nula del problema 10.2 con un nivel de significancia del 5%.

$$H_0: \mu = \$25\,000 \qquad H_1: \mu \neq \$25\,000$$

$$z = -1.93 \qquad \text{(del problema 10.2)}$$

$$P(z \leq -1.93) = 0.5000 - 0.4732 = 0.0268$$

Finalmente, como se trata de una prueba de dos extremos:

$$P = 2(0.0268) = 0.0536$$

Como el valor de P de 0.0536 es mayor que el nivel de significancia 0.05, *no es posible* rechazar la hipótesis nula a este nivel. Por ello, no puede rechazarse la afirmación de que el promedio de ingresos por hogar en la población es $25 000.

10.20 Al utilizar el método del valor P, pruebe la hipótesis nula del problema 10.3, a un nivel de significancia del 5%.

$$H_0: \mu \geq \$25\,000 \qquad H_1: \mu < \$\,25\,000$$

$$z = -1.93 \qquad \text{(del problema 10. 3)}$$

(Observe que la estadística *z* está en la dirección de la región de rechazo para esta prueba del extremo izquierdo.)

$$P(z \leq -1.93) = 0.5000\text{-}0.4732 = 0.0268$$

Como el valor de P = 0.268 es menor que el nivel de significancia de 0.05, se rechaza la hipótesis nula. Por ello, se acepta la hipótesis alternativa de que el promedio de ingresos por hogar en la población es inferior a $25 000.

10.21 Al utilizar el método del valor P, pruebe la hipótesis nula del problema 10.16 a un nivel de significancia del 5%.

$$H_0: \mu \geq \$25\,000 \qquad H_1: \mu < \$25\,000$$

$$t = -1.93 \qquad \text{(del problema 10.16)}$$

Para esta prueba del extremo Izquierdo, y con *gl* = 14:

$$P(t \leq -1.93) = (0.025 < P < 0.05)$$

La probabilidad de que ocurra la estadística í con un valor de -1.93 en forma aleatoria cuando la hipótesis nula es cierta y se encuentra entre 0.025 y 0.05, dado que se requiere un valor de t de -2.145 para una probabilidad de 0.025 en el extremo izquierdo y un valor de t de -1.761 para la probabilidad de 0.05 en el extremo derecho. Por ello, la hipótesis nula se rechaza a un nivel de significancia del 5% , y se concluye que el promedio de ingresos por hogar en la población es inferior a $25 000.

EL MÉTODO DEL INTERVALO DE CONFIANZA PARA PROBAR HIPÓTESIS NULAS SOBRE LA MEDIA DE LA POBLACIÓN

10.22 Aplique el método del intervalo de confianza para probar la hipótesis nula del problema 10.2, utilizando un nivel de significancia del 5%.

$$H_0: \mu = \$25\,000 \qquad H_1: \mu \neq \$25\,000$$

$$\bar{X} = \$24\,000$$

$$\sigma_{\bar{x}} = \$516.80 \qquad \text{(del problema 10.2)}$$

$$\text{Int. de conf. del 95\%} = \bar{X} \pm z\sigma_{\bar{x}} = 24{,}000 \pm 1.96(516.80)$$

$$= 24\,000 \pm 1\,012.93 = \$22\,987.07 \text{ a } \$25\,012.93$$

Como el intervalo de confianza del 95% incluye el valor hipotético de $25 000, no puede rechazarse la hipótesis nula con un nivel de significancia del 5%.

10.23 Aplique el método del intervalo de confianza para probar la hipótesis nula del problema 10.3, utilizando un nivel de significancia del 5%.

$$H_0: \mu \geq \$25\,000 \qquad H_1: \mu < \$25\,000$$

$$\bar{X} = \$24\,000$$

$$\sigma_{\bar{x}} = \$516.80 \qquad \text{(del problema 10.3)}$$

$$\text{Límite sup. de conf.} = \bar{X} + z\sigma_{\bar{x}} = 24\,000 + 1.645(516.80)$$

$$= 24\,000 + 850.14 = \$24\,850.14$$

Con un 95% de confianza, se concluye que la media de la población puede ser de hasta $24 850.14. Como este intervalo de confianza de un extremo no incluye el valor hipotético de $25 000 (o mayor), se rechaza la hipótesis

RESULTADOS POR COMPUTADORA

10.24 En la Tabla 10.5 se presentan las cantidades reclamadas por daños a automóviles, para una muestra aleatoria de 10 aseguradoras implicadas en accidentes automovilísticos en una área geográfica determinada (los datos están dados en miles de pesos). Utilice algún programa de computación para probar la hipótesis nula de que las reclamaciones en la población muestreada tienen un promedio de $1 000 000.

Tabla 10.5 Reclamaciones por daños de automóviles (en 000)

$1033	$1069
1 274	1 121
1 114	1 269
924	1 150
1 421	921

$$H_0: \mu = \$1\,000 \qquad H_1: \mu \neq \$1.000 \qquad \alpha = 0.05$$

Observe la figura 10-6. Se reporta que el valor P es 0.029. Como esta probabilidad es inferior al nivel de significancia de 0.05 que se especificó, se rechaza la hipótesis nula y se concluye que el promedio de reclamaciones en la población difiere de $1 000 000, con base en esta prueba de dos extremos.

```
MTB >  SET  DAMA6E  CLAIMS  INTO  Cl
DATA>  1033   1274   1114    924   1421   1069   1121   1269   1150    921
DATA>  END
MTB >  NAME  FOR  Cl  IS  'CLAIM'
MTB >  TTEST  OF  MU  =  1000  FOR  'CLAIM'

TEST  OF  MU  =  1000.0  VS  MU  N.E.  1000.0

                 N       MEAN     STDEV    SE MEAN          T      P VALUE
CLAIM           10     1129.6     158.0       50.0       2.59        0.029
```

Fig. 10-6 Resultados de Minitab para el problema 10.24

10.25 Con referencia a los datos de la Tabla 10.5, utilice algún programa de computación para probar la hipótesis nula de que el promedio de reclamaciones en la población objetivo *no es mayor* que $1 000 000, utilizando el nivel de significancia de 5%.

$$H_0: \mu \leq \$1\ 000 \qquad H_1: \mu > \$1\ 000 \qquad \alpha = 0.05$$

Con referencia a la figura 10-7, se reporta que el valor P es 0.015. Como esta probabilidad es inferior al nivel designado de significancia del 0.05, se rechaza la hipótesis nula y se concluye que el promedio de reclamaciones en la población es mayor que $1 000 000, con base en la prueba del extremo superior. *(Nota:* en Minitab, se especifican las pruebas del extremo superior mediante el *subcomando* ALTERNATIVE = 1, en tanto que las pruebas del extremo inferior se especifican con "ALTERNATIVE = -1".)

```
MTB >  SET  DAMAGE  CLAIMS  INTO  Cl
DATA>  1033   1274   1114   9S4   1421   1069   1121   1269   1150   921
DATA>  END
MTB >  NAME  FOR  Cl  IS  'CLAIM'
MTB >  TTEST  OF  MU  =  1000  FOR  'CLAIM'-;
SUBC>  ALTERNATIVE  =  1.

TEST  (OF  MU  =  1000.0  VS  MU  G.T.  1000.0

                 N       MEAN     STDEV    SE MEAN          T      P VALUE
CLAIM           10     1129.6     158.0       50.0       E.59        0.015
```

Fig. 10-7 Resultados de Minitab para el problema 10.25

Problemas complementarios

PRUEBA DE UN VALOR HIPOTÉTICO DE UNA MEDIA

10.26 Una cadena de restaurantes de comida rápida planea construir un nuevo expendio, si cuando menos 200 automóviles pasan por el lugar propuesto cada hora, durante determinadas horas. Para 20 horas muestreadas al azar, dentro del intervalo especificado, se encuentra que el número promedio de automóviles que pasan por ese lugar es $\overline{X}$= 208.5, con s = 30.0. Se supone que la población estadística es aproximadamente normal. Los admlnistradores de la cadena de restaurantes adoptan, conservadoramente, la hipótesis nula de que el volumen de tráfico *no* satisface sus requerimientos, es decir, H_0: $\mu \leq 200.0$. ¿Puede rechazarse esta hipótesis con un nivel de significancia del 5%?

Resp. No

10.27 Suponga que los resultados muestrales del problema 10.26 se basan en una muestra de n = 50 horas. ¿Puede rechazarse la hipótesis nula con un 5% de nivel de significancia?

Resp. Sí

10.28 Suponga que se determina que el valor promedio de ventas por tienda, para un producto determinado de consumo popular, durante el año anterior, es $\bar{X}$ = \$3 425 000, en una muestra de n = 25 tiendas. Con base en datos de ventas de otros productos similares, se concluye que la distribución de las ventas es normal y que la desviación estándar de la población es σ = \$200 000. Suponga además, que se ha afirmado que el verdadero monto de ventas por ienda es de cuando menos \$3 500 000. Pruebe esta afirmación con los niveles de significancia (a) 5%, y (b) 1 %

Resp. (a) Rechazar H_0, *(b)* aceptar H_0.

10.29 En el problema 10.28, suponga que no se hace ninguna afirmación con respecto a la desviación estándar de la población, pero que s = \$200 000. Pruebe esa afirmación con los niveles de significancia del (a) 5%, y *(b)* 1 %

Resp. (a) Rechazar H_0 (b) aceptar H_0.

10.30 Se encuentra que el número promedio de empleados para una muestra de 50 empresas de una industria específica es de 420.4, con una desviación estándar muestral de 55.7. Existe un total de 380 empresas en ese ramo industrial. Antes de recolectar los datos, se planteó la hipótesis de que el número promedio de empleados por empresa en esa industria no era superior a 408. Pruebe esa hipótesis con el 5% de nivel de significancia.

Resp. Rechazar H_0

10.31 Suponga que el analista del problema 10.30 olvida utilizar el factor de corrección por población finita al determinar el valor del error estándar de la media. ¿Cuál seria el resultado de la prueba si se sigue utilizando el 5% del nivel de significancia?

Resp. Aceptar H_Q

10.32 El fabricante de un automóvil compacto afirma que el automóvil rendirá cuando menos 9 kilómetros por litro, en promedio, manejando en carretera. En 40 ensayos de prueba, el promedio de rendimiento de gasolina de los automóviles fue de 8.85 kilómetros por litro, con una desviación estándar de 0.059 kilómetros por litro. ¿Puede rechazarse la afirmación del fabricante con un nivel de significación del 5%?

Resp. No

10.33 Con referencia al problema 10.32, antes de realizar las pruebas en carretera, una agencia de defensa del consumidor afirma que ese automóvil compacto *no* rinde los nueve kilómetros por litro en las condiciones especificadas. ¿Puede rechazarse esa afirmación con un nivel de significancia del 5%? Considere usted las implicaciones de su respuesta a esta pregunta y a la del problema 10.32 con respecto a cuál hipótesis se designa como hipótesis nula.

Resp. No

10.34 Un analista de un departamento de personal selecciona al azar ios expedientes de 16 empleados que trabajan a destajo, y encuentra que el salario promedio es de $\bar{X}$ = \$75 000 semanales, con desviación estándar de s = \$10 000.

Se supone que los salarios de la empresa tienen una distribución normal. Pruebe la hipótesis nula H_0: $\mu = \$80\,000$, utilizando un nivel de significancia del 10%.

Resp. Rechazar H_0

10.35 A una muestra aleatoria de 30 empleados de nivel secretarial II de una organización grande se le aplica una prueba estándar de mecanografía. Los resultados muestrales son $\bar{X} = 63.0$ ppm (palabras por minuto) con s = 5.0 ppm. Pruebe la hipótesis nula de que las secretarias en general *no* exceden la velocidad de mecanografía de 60 ppm, utilizando un nivel de significancia del 1 %.

Resp. Rechazar H_0

10.36 Se ha ajustado una máquina llenadora de botellas para que introduzca 4.00 litros de liquido en cada recipiente. Se encuentra que, para una muestra de n = 10 botellas, la cantidad promedio de liquido introducido es $\bar{X} = 4.05$ litros, con s = 0.10 litros. Se supone que las cantidades con las que se llenan las botellas tienen una distribución normal. Basando la hipótesis nula en la suposición de que este proceso está "bajo control", ¿debe reajustarse la máquina llenadora como resultado de una prueba con un nivel de significancia del 5%?

Resp. No

10.37 En un departamento de reparación de maquinaria se recibe un embarque de 100 máquinas defectuosas. Se encuentra que, para una muestra aleatoria de 10 de ellas, el tiempo promedio de reparación que se requiere es $\bar{X} =$ 85.0 minutos, con s =15.0 minutos. Pruebe la hipótesis nula H_0: $\mu = 100.0$ minutos, utilizando un nivel de significancia del 10%, y suponiendo que la distribución del tiempo de reparación es aproximadamente normal.

Resp. Rechazar H_0

ERRORES TIPO I Y TIPO II EN PRUEBAS DE HIPÓTESIS

10.38 Con referencia al problema 10.35, suponga que se considera que, si la velocidad de mecanografía promedio es de cuando menos 64.0 ppm, se tiene una diferencia considerable con respecto al valor hipotético. Determine la probabilidad del error (a) tipo I y (b) tipo II.

Resp. (a) $\alpha = 0.01$,(6) $\beta = 0.0192 \cong 0.02$.

10.39 Para el problema 10.38, determine la probabilidad del error tipo II (a) si se cambia el nivel de significancia al 5%, y (b) si se mantiene el nivel de significancia del 1 %, pero se cambia el tamaño de muestra a $n = 60$, y en lugar de n -30.

Resp. (a) $\beta = 0.003$, (b) $\beta < 0.001$.

10.40 Para el procedimiento de prueba que se describió en el problema 10.32, suponga que se considera como discrepancia importante que el promedio de rendimiento de gasolina sea de 8.72 kilómetros por litro o menos. Con esta información adicional, determine (a) *el* valor *CO* mínimo si la hipótesis nula es cierta, y (b) la potencia mínima asociada con la prueba estadística, si la discrepancia con respecto a la afirmación del fabricante resulta ser importante.

Resp. (a) *CO* = 0.95, (b) potencia = 0.8729 $\cong$ 0.87.

10.41 Determine la potencia mínima de la prueba del problema 10.40, si "se considera como discrepancia importante un rendimiento real de (a) 0.128 kilómetros por litro, y (b) 0.0256 kilómetros por litro menos que el rendimiento hipotético. Comparando diversos valores de potencia, considere la implicación del estándar utilizado para definir una "discrepancia importante".

Resp. (a) Potencia = 0.4013 $\cong$ 0.40, (b) potencia = 0.0869 $\cong$ 0.09.

DETERMINACIÓN DEL TAMAÑO DE MUESTRA NECESARIO PARA PROBAR LA MEDIA

10.42 Antes de recolectar datos muestrales, el inversionista potencial del problema 10.26 estipula que el nivel del error tipo I no debe ser mayor que $\alpha = 0.01$, y que, si el número de automóviles que pasan por el lugar es de $\mu = 210$ automóviles por hora o mayor, entonces el nivel del error tipo II tampoco debe exceder de $\beta = 0.01$. Estima que la desviación estándar de la población no es superior a $\sigma = 40$. ¿Qué tamaño de muestra se requiere para lograr esos objetivos?

Resp. n = 347.45 = 348.

10.43 Para la situación de prueba que se describió en el problema 10.28, se considera como discrepancia importante que el promedio de ventas por tienda sea de $100 000 menos de lo que se pretende. ¿Qué tamaño de muestra se requiere para que la prueba se lleve a cabo con un nivel de significancia del 1 %, y permitiendo una probabilidad máxima de error tipo II de $\beta = 0.05$?

Resp. n -=63.20 = 64.

EL MÉTODO DEL VALOR DE *P* PARA PROBAR HIPÓTESIS NULAS SOBRE LA MEDIA DE UNA POBLACIÓN

10.44 Al utilizar el método del valor *P*, pruebe la hipótesis nula del problema 10.27 a un nivel de significancia del 5%.

Resp. Rechazar H_0 (P = 0.0228).

10.45 Al utilizar el método del valor *P*, pruebe la hipótesis nula del problema 10.28 a un nivel de significancia del 5%.

Resp. Rechazar H_0 (P = 0.0307).

EL MÉTODO DEL INTERVALO DE CONFIANZA PARA PROBAR HIPÓTESIS NULAS SOBRE LA MEDIA

10.46 Aplique el método del intervalo de confianza para probar la hipótesis nula del problema 10.27, utilizando un nivel de significancia del 5%.

Resp. Rechazar H_0

10.47 Aplique el método del intervalo de confianza para probar la hipótesis nula del problema 10.34, utilizando un nivel de significancia del 10%.

Resp. Rechazar H_0

RESULTADOS POR COMPUTADORA

10.48 Consulte la Tabla 2.26 (página 26) para los montos de 40 préstamos personales. Suponiendo que son datos muestrales aleatorios, utilice algún paquete de computación para probar la hipótesis nula de que el promedio de los préstamos en la población no es mayor de \$1 000 000, utilizando un nivel de significancia del 1 %.

Resp. Aceptar H_0 ($P = 0.17$).

Otras pruebas de hipótesis

11

11.1 PRUEBA DE LA DIFERENCIA ENTRE DOS MEDIAS UTILIZANDO LA DISTRIBUCIÓN NORMAL

El procedimiento que se utiliza para probar la diferencia entre dos medias es similar al de la prueba para el valor hipotético de una media (véanse las secciones 10.1 y 10.2), excepto que se utiliza el error estándar de la *diferencia* entre las medias para determinar el valor z correspondiente al resultado muestral. El uso de la distribución normal se basa en las mismas condiciones que en el caso de una muestra, excepto que ahora se tienen dos muestras independientes. La fórmula general para determinar el valor de z para probar la diferencia entre dos medias, dependiendo de si se conocen los valores σ para las dos poblaciones, es

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{\sigma_{\bar{x}_1 - \bar{x}_2}} \tag{11.1}$$

o

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)_0}{s_{\bar{x}_1 - \bar{x}_2}} \tag{11.2}$$

Tal como se observa en (11.1) y (11.2), puede precederse a probar cualquier diferencia $(\mu_1 - \mu_2)_0$. Sin embargo, la hipótesis nula que generalmente se prueba consiste en que las dos muestras se obtienen de poblaciones con medias iguales. En este caso, $(\mu_1 - \mu_2)_0 = 0$, y las fórmulas anteriores se vuelven más simples:

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sigma_{\bar{x}_1 - \bar{x}_2}} \tag{11.3}$$

o

$$z = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{x}_1 - \bar{x}_2}} \tag{11.4}$$

En general, el error estándar de la diferencia entre las medias se calcula como se describió en la sección 9.1 (fórmulas *9.3* y *9.4).* Sin embargo, al probar la diferencia entre dos medias, la hipótesis nula de Interés, por lo general, no sólo se refiere a que las medias muestrales se obtuvieron de poblaciones con medias iguales, sino que, de hecho, las dos muestras se obtuvieron de la *misma* población de valores. Esto significa que $\sigma_1 = \sigma_2$, que puede simplemente designarse con σ. Por ello, es frecuente que se estime la varianza supuestamente común, combinando las dos varianzas muestrales, y se utiliza después el valor estimado de σ^2 como base para el error estándar de la diferencia. La estimación combinada de la varianza poblacional es

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \tag{11.5}$$

El error estándar estimado de la diferencia, con base en la suposición de que las desviaciones estándar de las poblaciones son iguales, es

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2}=\sqrt{\frac{\hat{\sigma}^2}{n_1}+\frac{\hat{\sigma}^2}{n_2}} \qquad (11..6)$$

Puede probarse como hipótesis nula la suposición de que las dos varianzas mustrales se obtuvieron de poblaciones con varianzas iguales (véase la sección 11.9).

En los siguientes ejemplos se ilustran pruebas con respecto a la diferencia entre medias, para uno y dos extremos.

EJEMPLO 1. El salario promedio mensual para una muestra de n_1 = 30 empleados de una empresa manufacturera grande $\bar{X}_1$ = \$280 000, con desviación estándar muestral de s, = \$14,000. En otra empresa grande, una muestra aleatoria de n_z= 40 empleados tiene un salario promedio de $\bar{X}_2$ = \$270 000, con una desviación estándar muestral de s_2 = \$10 000. No se supone que las desviaciones estándar de las dos poblaciones de salarios sean iguales. Se prueba la hipótesis de que no existe diferencia entre los salarios promedio mensuales de las dos empresas, utilizando un nivel de significancia del 5%, de la siguiente manera:

$H_0: (\mu_1 - \mu_2) = 0$ o, de manera equivalente, $\mu_1 = \mu_2$ $\quad \bar{X}_1 = \$280\,000 \quad \bar{X}_2 = \$270\,000$

$H_1: (\mu_1 - \mu_2) \neq 0$ o, de manera equivalente, $\mu_1 \neq \mu_2$ $\quad s_1 = \$14\,000 \quad s_2 = \$10\,000$

$n_1 = 30 \quad n_2 = 40$

Z Crítica α = 0.05 = ±1.96

$$z=\frac{\bar{X}_1-\bar{X}_2}{s_{\bar{x}_1-\bar{x}_2}}$$

$$=\frac{\$280\,000-\$270\,000}{3\,005.58}=\frac{10\,000}{3\,005.58}=+3.33$$

a donde $s_{\bar{x}_1}=\dfrac{s_1}{\sqrt{n_1}}=\dfrac{14.000}{\sqrt{30}}=\dfrac{14.000}{\sqrt{5.477}}=2556.14$

$$s_{\bar{x}_2}=\frac{s_2}{\sqrt{n_2}}=\frac{10.000}{\sqrt{40}}=\frac{10.000}{6.325}=\$1581.03$$

$$s_{\bar{x}_1-\bar{x}_2}=\sqrt{s^2_{\bar{x}_1}+s^2_{\bar{x}_2}}=\sqrt{(2556.14)^2+(1581.03)^2}=\sqrt{6\,333\,851.7+2\,499\,655.86}=3005.58$$

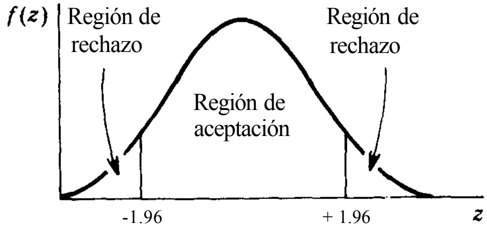


Fig. 11-1

El valor calculado de z, +3.32, se encuentra en la región de rechazo de la hipótesis, tal como puede observarse en el modelo que se ilustra en la figura 11-1. Por ello, se rechaza la hipótesis nula, y se acepta la hipótesis alternativa de que el salario promedio mensual de las dos empresas es diferente.

EJEMPLO 2. Antes de observar los resultados del ejemplo 1, un analista de sueldos consideraba que el salario promedio de la primera empresa era mayor que en la segunda empresa. Con el objeto de someter su suposición a una prueba crítica, le da el beneficio de la duda a la posibilidad contraria y plantea la hipótesis nula de que el salario promedio de la primera empresa es igual o menor que el de la segunda. Se prueba la hipótesis, con nivel de significancia del 1 %, de nueva cuenta sin suponer que las desviaciones estándar de las dos poblaciones son iguales:

$$H_0: (\mu_1 - \mu_2) \leq 0 \quad \text{o, de manera equivalente,} \quad \mu_1 \leq \mu_2$$
$$H_0: (\mu_1 - \mu_2) > 0 \quad \text{o, de manera equivalente,} \quad \mu_1 > \mu_2$$
$$z \text{ crítica } (\alpha = 0.001) = +\,2.33$$
$$Z \text{ calculada} = +\,3.32 \text{ (del ejemplo 1)}$$

El valor calculado de z de + 3.32 es mayor que el valor crítico de +2.32 para esta prueba del extremo superior, tal como se ilustra en la figura 11 -2. Por ello, se rechaza la hipótesis nula y se acepta la alternativa de que el salario promedio de la primera empresa es *mayor que* el salario promedio de la segunda empresa.

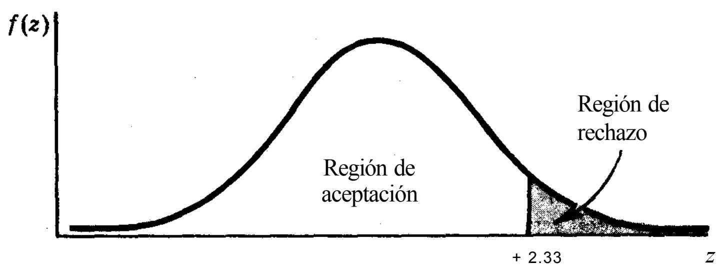


Fig. 11-2

11.2 PRUEBA DE LA DIFERENCIA ENTRE DOS MEDIAS UTILIZANDO LA DISTRIBUCIÓN *t* DE STUDENT

Cuando se prueba la diferencia entre dos medias utilizando la distribución í, se requiere la suposición de que las varianzas de las dos poblaciones son iguales. Por ello, en una prueba como ésta, el error estándar estimado de la media se calcula con las fórmulas *(11.5)* y (77.6). En las secciones 8.5 y 10.5 se describen los diversos requerimientos asociados con el uso apropiado de la distribución /.

EJEMPLO 3. Para una muestra aleatoria de n_1 = 10 focos, se encuentra que la vida promedio es $\overline{X}_1$ = 4000 horas con s,= 200. Para otra marca de focos, para los cuales se supone también que tienen una vida útil con distribución normal, una muestra aleatoria de n_2 = 8 tiene una media muestral de $\overline{X}_2$ = 4300 y una desviación estándar muestral de s = 250. Se prueba la hipótesis de que no existe diferencia entre la vida útil promedio de las dos marcas de focos, utilizando un nivel de significancia del 1%:

$$H_0\text{: } (\mu_1 - \mu_2) = 0 \qquad \bar{X}_1 = 4000 \text{ hr} \qquad \bar{X}_2 = 4\,300 \text{ hr}$$

$$H_1\text{: } (\mu_1 - \mu_2) \neq 0 \qquad s_1 = 200 \text{ hr} \qquad s_2 = 250\text{hr}$$

$$n_1 = 10 \qquad n_2 = 8$$

$$gl = n_1 + n_2 - 2 = 10 + 8 - 2 = 16$$

$$t \text{ crítica } (gl = 16,\ \alpha = 0.01) = \pm 2.921$$

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(9)(200)^2 + (7)(250)^2}{10 + 8 - 2} = \frac{360\,000 + 437\,500}{16} = 49\,843.75$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2}} = \sqrt{\frac{49\,843.75}{10} + \frac{49\,843.75}{8}} = \sqrt{11\,214.843} = 105.9$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{4\,000 - 4\,300}{105.9} = \frac{-300}{105.9} = -2.833$$

El valor calculado de í de -2.833 cae en la región de aceptación de la hipótesis nula. Por ello, no es posible rechazar esta hipótesis a un nivel de significancia del 1 %.

11.3 PRUEBA PARA LA DIFERENCIA ENTRE DOS MEDIAS, CON BASE EN OBSERVACIONES APAREADAS

Los procedimientos de las secciones 11.1 y 11.2 se basan en la suposición de que las dos muestras se obtienen como muestras aleatorias independientes. Sin embargo, en muchas situaciones las muestras se extraen como pares de valores, tal como cuando se determina el nivel de productividad de los trabajadores, antes y después de un programa de capacitación. A esta clase de datos se les denomina *observaciones apareadas* o *pares asociados.* También, a diferencia de las muestras independientes, a dos muestras que contienen observaciones apareadas se les denomina *muestras dependientes.*

El método apropiado para probar la diferencia entre las medias de dos muestras, es decir, para observaciones apareadas, consiste primero en determinar la diferencia d entre cada par de valores, y después probar la hipótesis nula de que *la diferencia* poblacional promedio es 0. Por ello, desde el punto de vista de los cálculos, se aplica una prueba a *una* muestra de valores *d.*

La media y la desviación estándar de la muestra de valores d se obtienen utilizando las fórmulas básicas de los capítulos 3 y 4, excepto que se sustituye *d* por X. La diferencia promedio para el conjunto de observaciones apareadas es

$$\bar{d} = \frac{\Sigma d}{n} \tag{11.7}$$

La fórmula de desviaciones y la fórmula abreviada para la distribución estándar y las diferencias entre observaciones apareadas son, respectivamente:

$$s_d = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} \tag{11.8}$$

$$s_d = \sqrt{\frac{\Sigma d^2 - n\bar{d}^2}{n-1}} \tag{11.9}$$

El error estándar del promedio de las diferencias entre observaciones apareadas se obtiene mediante la fórmula *(8.4)* del error estándar de la media, excepto que, de nueva cuenta, se sustituye *d* por *X:*

$$t = \frac{\bar{d}}{s_d} \tag{11.11}$$

Como el error estándar del promedio de las diferencias se calcula con base en las diferencias observadas en las muestras apareadas (es decir, se desconoce el valor poblacional σ_d), y como, por lo general, se supone que los valores *d* tienen una distribución normal, la distribución *t* resulta apropiada para probar la hipótesis nula de que $\mu_d = 0$.

Los grados de libertad son el número *de pares* de valores observados, menos uno, o *n - 1.* Tal como se analizó en la sección 8.5, puede utilizarse la distribución normal estándar z en lugar de las distribuciones t cuando $n \geq 30$. En el ejemplo 4 se ilustra una prueba de dos extremos, y en el 11.5 se ilustra una prueba de un extremo. Así, la estadística de prueba que se utiliza para probar la hipótesis de que no existe diferencia entre las medias de un conjunto de observaciones apareadas es

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} \tag{11.10}$$

EJEMPLO 4. Un fabricante de automóviles obtiene datos de rendimiento de gasolina para una muestra de *n* = 10 automóviles en diversas categorías de peso utilizando gasolina común, con y sin un determinado aditivo. Por supuesto, se afinan las máquinas de acuerdo con las mismas especificaciones antes de realizar cada prueba y se utilizan los mismos conductores para las dos condiciones (de hecho, el conductor no sabe qué tipo de gasolina se utiliza en las pruebas). Con los datos de rendimiento de la Tabla 11.1, se prueba la hipótesis de que no existe diferencia entre el kilometraje promedio que se obtiene con y sin el aditivo, utilizando un nivel de significancia del 5%:

Tabla 11.1 Rendimiento de kilometraje de automóviles y hoja de trabajo para calcular la diferencia promedio y la desviación estándar de la diferencia

Automóvil	Kilometraje con aditivo	Kilometraje sin aditivo	d	d^2
1	12.66	12.48	0.17	0.03
2	12.34	12.31	0.03	0.00
3	11.00	11.14	-0.14	0.02
4	10.10	10.21	-0.10	0.01
5	9.79	9.69	0.10	0.01
6	8.86	8.90	0.03	0.00
7	8.84	8.24	0.10	0.01
8	7.79	7.59	0.21	0.04
9	7.55 '	7.41	0.14	0.02
10	7.0	6.90	0.10	0.01
Total	95.45	24.86	0.58.62	0.9

$$\text{Distancia con aditivo} = \frac{95.45}{10} = 9.55 \text{ kpl}$$

$$\text{Distancia sin aditivo} = \frac{99\backslash 4,86}{10} = 9.49 \text{ kpl}$$

$$H_0: \mu_d = 0$$

$$H_1: \mu_d \neq 0$$

$$t \text{ crítica } (gl = 9, \alpha = 0.05 = \pm\ 2.262$$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{0.5862}{10} = 0.059$$

$$s_d = \sqrt{\frac{\Sigma d^2 - n\bar{d}^2}{n-1}} = \sqrt{\frac{0.1558 - 10(0.059)^2}{10-1}} = \sqrt{\frac{1.31 - 10(0.003481)}{9}} = \sqrt{0.0169} = 0.13$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{0.13}{\sqrt{10}} = \frac{0.13}{3.16} = 0.0412$$

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{0.059}{0.0412} = +\ 1.43$$

El valor calculado de t de + 1.59, se encuentra en la región de aceptación de la hipótesis nula. Por ello, se acepta la hipótesis nula de que no existe diferencia en el rendimiento que se obtiene por litro de gasolina, con y sin el aditivo.

11.4 PRUEBA DE UN VALOR HIPOTÉTICO DE LA PROPORCIÓN POBLACIONAL UTILIZANDO LA DISTRIBUCIÓN BINOMIAL

Cuando puede suponerse que un proceso de muestreo se ajusta a un proceso Bernoulli (sección 6.3), puede utilizarse la distribución binomial para realizar pruebas de hipótesis sobre proporciones poblacionales.

Por lo general, las pruebas de proporciones que se basan en la distribución binomial son de un extremo. Dado el valor hipotético de la proporción poblacional, la "región" de rechazo es el conjunto de observaciones muestrales que se desvia del valor hipotético y para el cual la probabilidad de ocurrencia aleatoria no excede el nivel especificado de significancia. Es decir, en esencia, se utiliza el método del valor *P* para pruebas de hipótesis (véase la sección 11.10). El procedimiento de prueba para un extremo se ilustra en el ejemplo 5. En el ejemplo 11.8 se ilustra una prueba de dos extremos utilizando una distribución binomial.

EJEMPLO 5. El director de la bolsa de trabajo de una universidad afirma que hacia el primero de marzo, cuando menos el 50% de los recién egresados tendrá empleo de tiempo completo. El primero de marzo se entrevista a una muestra aleatoria de 10 egresados, y sólo dos afirman haber obtenido empleo. ¿Puede rechazarse la afirmación del director utilizando un nivel de significancia del **5%?** H_0: $\pi \geq 0.50$ y $H1$: $\pi < 0.50$.

Con base en la distribución binomial, los valores de probabilidad correspondientes al hecho de que menos de 5 estudiantes hayan obtenido empleo, dada una proporción poblacional de 0.50, son los que se presentan en la Tabla 11.2 (del Apéndice 2 con n = 10 y π = 0.50).

Valores críticos de la estadística de prueba: La estadística de prueba es el número de estudiantes de la muestra de n = 10 que ya ha obtenido empleo. Con el objeto de rechazar la hipótesis nula en el nivel de significancia del 5%, sólo "0" o " 1 " estudiantes deberían haber obtenido empleo. Esto es así porque se acumulan las probabilidades en el "extremo inferior* de esta distribución binomial para determinar la región de rechazo. Si se intenta incluir "dos" estudiantes en la región de rechazo, se obtiene una probabilidad acumulada (para "cero, uno o dos") de 0.0010 + 0.0098 + 0.0439 - 0.0547, lo cual supera el nivel especificado para la prueba de 0.05.

Resultado de la prueba. Con base en los valores críticos que se identifican en el párrafo anterior, el que sólo dos estudiantes de la muestra de 10 hayan obtenido empleos no es lo suficientemente reducido para rechazar la afirmación del director en un nivel de significancia del 5%.

Tabla 11.2 Valores de probabilidad correspondientes al hecho de que menos de cinco de diez estudiantes hayan obtenido empleo

Número de estudiantes	Probabilidad
0	0.0010
1	0.0098
2	0.0439
3	0.1172
4	0.2051

(Nota: Con una muestra de mayor tamaño, la misma diferencia relativa con respecto al valor hipotético podría, de hecho, conducir al rechazo de la hipótesis nula. Véase el problema 11.7.)

11.5 PRUEBA DEL VALOR HIPOTÉTICO DE UNA PROPORCIÓN POBLACIONAL UTILIZANDO LA DISTRIBUCIÓN NORMAL

Como se explicaba en la sección 7.4, puede utilizarse la distribución normal como aproximación de la binomial, cuando $n \geq 30$ y, tanto $np \geq 5$ *como n (* ≥ 5, en donde $q = 1 - p$. Fue sobre esta base que se construyeron intervalos de confianza para la proporción en la sección 9.3, en donde se analizó también el error estándar de la proporción. Sin embargo, en el caso de intervalos de confianza, se requiere un tamaño de muestra de cuando menos n = 100, tal como se explicó en la sección 9.3.

En pruebas de hipótesis, el valor del error estándar de la proporción que se utiliza se basa en el valor hipotético π_0:

$$\sigma_{\hat{p}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} \tag{11.12}$$

La fórmula del error estándar de la proporción que incluye el factor de corrección por población finita es:

$$\sigma_{\hat{p}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}}\sqrt{\frac{N-n}{N-1}} \tag{11.13}$$

The procedure associatEedJ with testing a hypothesize EdM value of the population proportion is identical to that described in Section 10.2, except that a value of the proportion rather than of the mean is being tested. Thus, the formula for the z statistic for testing a hypothesized value of the proportion is

$$z = \frac{\hat{p} - \pi_0}{\sigma_{\hat{p}}} \tag{11.14}$$

EJEMPLO 6 En el ejemplo 5, el director de la bolsa de trabajo afirmaba que cuando menos, el 50% de los egresados habría obtenido empleo hacia el primero de marzo. Suponga que se entrevista a una muestra de *n* - 30 egresados, en vez de los 10 del ejemplo 5, y que sólo 10 de ellos señalan haber obtenido empleo hacia el primero de marzo. ¿Puede rechazarse la afirmación del director con un nivel de significancia del 5%? Se utilizaz como estadística de prueba, de la siguiente manera:

$$H_0\colon \pi \geq 0.50 \qquad H_1\colon \pi < 0.50$$

$$z \text{ crítica } (\alpha = 0.05) = -1.645$$

[Se justifica el uso de la distribución normal porque $n \geq 30$, $n\pi_0 \geq 5$, y $n(1 - \pi_0) \geq 5$.]

$$\sigma_{\hat{p}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{(0.50)(0.50)}{30}} = \sqrt{\frac{0.25}{30}} = \sqrt{0.0083} = 0.09$$

(Se supone que la muestra es menos del 5% del tamaño de la población, y por ello no se utiliza el factor de corrección por población finita.)

$$z = \frac{\hat{p} - \pi_0}{\sigma_{\hat{p}}} = \frac{0.33 - 0.50}{0.09} = \frac{-0.17}{0.09} = -1.88$$

El valor calculado de z de -1.88 es menor que el valor crítico de -1.645 para esta prueba del extremo inferior. Por ello, se rechaza la afirmación del director en un nivel de significancia del 5%.

11.6 DETERMINACIÓN DEL TAMAÑO DE MUESTRA NECESARIO PARA PROBAR LA PROPORCIÓN

Es posible determinar el tamaño que se requiere para una muestra, para probar el valor hipotético de una proporción (antes de extraerla) especificando (1) el valor hipotético de la proporción, (2) un valor alternativo específico de la proporción, de manera que la diferencia con respecto al valor hipotético-nulo resulte considerable, (3) el nivel de significancia que debe utilizarse en la prueba, y (4) la probabilidad del error tipo II que se permite. La fórmula para determinar el tamaño mínimo de la muestra que se requiere para probar un valor hipotético de la proporción es:

$$n = \left[\frac{z_0\sqrt{\pi_0(1-\pi_0)} - z_1\sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0}\right]^2 \qquad (11.15)$$

En la fórmula *(11.15),* z_0 es el valor crítico de z que se utiliza con el nivel especificado de significancia (nivel a), en tanto que z_1 es el valor que corresponde a la probabilidad designada del error tipo II (nivel B). Al igual que en la sección 10.4, cuando se determina el tamaño de la muestra para probar la media, z_0 y z_1 siempre tienen signos algebraicos opuestos. Ésto da como resultado que los dos productos del numerador siempre se acumulen. También, puede utilizarse la fórmula *(11.15)* en pruebas de uno o dos extremos, y cualquier tamaño de muestra que resulte ser fraccionario se redondea al entero superior. Además, el tamaño de la muestra debe ser suficientemente grande para permitir el uso de la distribución normal de probabilidad, junto con π_0 y π_1, tal como se vio en la sección 11.5.

EJEMPLO 7 Un legislador desea probar la hipótesis de que, cuando menos, 60% de sus representados están a favor de cierta legislación laboral que se está presentando en el Congreso, utilizando el 5% como nivel de significancia. Considera que una discrepancia importante con respecto a su hipótesis consistiría en que sólo el 50% (o menos) de las personas estuvieran a favor de la legislación, y está dispuesto a aceptar un riesgo del error del tipo II de ß = 0.05. El tamaño de la muestra que debe extraer, como mínimo, para satisfacer esas especificaciones es

$$n = \left[\frac{z_0\sqrt{\pi_0(1-\pi_0)} - z_1\sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0}\right]^2 = \left[\frac{-1.645\sqrt{(0.60)(0.40)} - (+1.645)\sqrt{(0.50)(0.50)}}{0.50 - 0.60}\right]^2$$

$$= \left[\frac{-1.645(0.49) - 1.645(0.50)}{-0.10}\right]^2 = \left(\frac{-0.806 - 0.822}{-0.10}\right)^2 = (16.28)^2 = 265.04 = 266$$

11.7 PRUEBA PARA LA DIFERENCIA ENTRE DOS PROPORCIONES POBLACIONALES

Cuando se desea probar la hipótesis de que las proporciones de dos poblaciones no son distintas, se combinan las dos proporciones muestrales para proceder a determinar el error estándar de la diferencia de las proporciones. Debe observarse que esto es distinto del procedimiento que se utilizó en la sección 9.5 (sobre estimación estadística), en la que *no* se hizo la suposición de que no hay diferencia. Además, este procedimiento es conceptualmente similar al que se presentó en la sección 11.1, en la que se combinaron dos varíanzas muestrales para calcular el error estándar de la diferencia entre dos medias. La estimación combinada de la proporción poblacional, con base en las proporciones obtenidas en dos muestras independientes es

$$\hat{\pi} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 - n_2} \tag{11.16}$$

El error estándar de la diferencia entre dos proporciones que se utiliza para probar el supuesto de que no existe diferencia es

$$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}} \tag{11.17}$$

La fórmula de la estadística z para probar la diferencia entre dos proporciones es

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}} \tag{11.18}$$

Las pruebas sobre la diferencia entre dos proporciones pueden llevarse a cabo como pruebas de un extremo (véase problema 11.11) o de dos extremos (véase ejemplo 8).

EJEMPLO 8. Una muestra de 50 hogares de cierta comunidad arroja que 10 de ellos se encuentran viendo un programa especial de televisión. En una segunda comunidad, 15 hogares de una muestra aleatoria de 50 se encuentran observando el programa especial. Se prueba la hipótesis de que la proporción global de televidentes en las dos comunidades no difiere, utilizando el nivel de significancia del 1 %, de la siguiente manera:

$$H_0\colon (\pi_1 - \pi_2) = 0 \quad \text{o, de manera equivalente} \quad \pi_1 = \pi_2$$
$$H_1\colon (\pi_1 - \pi_2) \neq 0 \quad \text{o, de manera equivalente} \quad \pi_1 \neq \pi_2$$

$$z \text{ crítica } (\alpha = 0.01) = \pm 2.58$$

$$\hat{\pi} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{50(0.20) + 50(0.30)}{50 + 50} = \frac{10 + 15}{100} = 0.25$$

$$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}} = \sqrt{\frac{(0.25)(0.75)}{50} + \frac{(0.25)(0.75)}{50}} = \sqrt{0.00375 + 0.00375} = 0.087$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}} = \frac{0.20 - 0.30}{0.087} = \frac{-0.10}{0.087} = -1.15$$

El valor calculado de z de -1.15 se encuentra en la región de aceptación. Por ello, no es posible rechazar la hipótesis de que no existe diferencia en la proporción de televidentes en las dos comunidades.

11.8 PRUEBA PARA EL VALOR HIPOTÉTICO DE LA VARIANZA UTILIZANDO LA DISTRIBUCIÓN JI CUADRADA

Tal como se explicó en la sección 9.6, para una población con distribución normal, el cocient $(n-1)s^2 / \sigma^2$ tiene una distribución de probabilidad de x^2, y se tiene una distribución ji-cuadrada distinta para los diferentes grados de libertad $(n-1)$. Por ello, la estadística que se utiliza para probar el valor hipotético de una varianza poblacional es

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \tag{11.19}$$

Las pruebas que se basan en la fórmula *(11.19),* pueden ser de uno o de dos extremos, aunque las hipótesis más comunes con respecto a las varianzas poblacionales se refieren a pruebas de un extremo. Puede utilizarse el apéndice 7 para determinar el o los valores críticos de la estadística ji-cuadrada para diversos niveles de significancia.

EJEMPLO 9. La vida útil promedio de una muestra aleatoria de $n = 10$ focos es X = 4000 horas, con desviación estándar de s = 200 horas. En general, se asume que la vida útil de los focos tiene una distribución normal. Suponga que, antes de obtener la muestra, se plantea la hipótesis de que la desviación estándar de la población no es superior a $\sigma = 150$. Con base en los resultados muestrales, se prueba esa hipótesis, con un nivel de significancia del 1 %, de la siguiente manera:

$$H_0\text{: } \sigma^2 \leq 22\,500 \text{ [porque } \sigma_0^2 = (150)^2 = 22\,500] \qquad H_1\text{: } \sigma^2 > 22\,500$$

$$\chi^2 \text{ crítica } (gl = 9 \;\; \alpha = 0.01) = 21.67$$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(9)(40\,000)}{22\,500} = \frac{360\,000}{22\,500} = 16.0$$

Como la estadística de prueba calculada de 16.0 *no* excede el valor critico de 21.67 para esta prueba del extremo superior, no puede rechazarse la hipótesis nula de que $\sigma \leq 150$, en el nivel de significancia del 1 %.

11.9 LA DISTRIBUCIÓN *F* Y LA PRUEBA PARA LA DIFERENCIA ENTRE DOS VARIANZAS

Puede probarse que la distribución F es el modelo de probabilidad apropiado para el cociente de las varianzas de las muestras tomadas en forma independiente de la misma población con distribución normal, y que existe una distribución *F* diferente para cada combinación de grados de libertad *gl,* correspondientes a cada muestra. Para todas las muestras, $gl = n - 1$. Por ello, la estadística que se utiliza para probar la hipótesis nula de que no existe diferencia entre las dos varianzas es:

$$F_{gl_1, gl_2} = \frac{s_1^2}{s_2^2} \tag{11.20}$$

Como todas las varianzas muestrales son estimadores ¡nsesgados de la varianza poblacional, el valor esperado a largo plazo del cociente anterior es aproximadamente 1.0. *(Nota:* el valor esperado no es exactamente 1.0, sino $gl_2/(gl_2 - 2)$, por razones matemáticas que están fuera del alcance de esta introducción.) Sin embargo, no es probable que las varianzas muestrales de cualquier par de muestras sean idénticas, aun cuando la hipótesis nula sea cierta. Como se sabe que ese cociente tiene una distribución *F,* puede utilizarse esta distribución de probabilidad para probar la diferencia entre dos varianzas. Aunque se tiene una suposición matemática necesaria en el sentido de que las dos poblaciones tienen distribución normal, se ha demostrado que la prueba F es relativamente insensible a desviaciones con respecto a la normalidad cuando las poblaciones son, cuando menos, unimodales y los tamaños de las muestras son más o menos iguales.

En el apéndice 8, se señalan los valores de *F* que son excedidos por proporciones de 0.05 y 0.01 de la distribución de valores *F.* Los grados de libertad *gl* asociados con el numerador del cociente calculado de *F* son los encabezados de la columna de la tabla, y los grados de libertad para el denominador son los encabezados de los renglones. En la tabla no se identifica ningún valor crítico de *F* para el extremo inferior de la distribución, en parte debido a que la distribución *F* se utiliza, por lo general, en pruebas de un extremo. Esto es cierto en especial cuando se usa la distribución *F* en el análisis de varianza (capítulo 13). Otra razón para ofrecer sólo los valores de *F* del extremo superior es que los valores del extremo inferior pueden calcularse mediante la denominada *propiedad recíproca* de la distribución F, de la siguiente manera:

$$F_{gl_1,\, gl_2 \text{ inferior}} = \frac{1}{F_{gl_2,\, gl_1, \text{ superior}}} \qquad (11.21)$$

Al aplicar la fórmula *(11.21),* se determina el valor de *F* del 5% inferior anotando un valor del extremo superior en el punto del 5% del denominador. Sin embargo, debe observarse que los dos valores de *gl* en el denominador son el *inverso* del orden en el valor *F* que se requiere.

EJEMPLO 10. Para los datos del ejemplo 3, se supuso que la vida útil de los focos tiene distribución normal. Se prueba la hipótesis nula de que las muestras se obtuvieron de poblaciones con varianzas iguales, utilizando el nivel de significancia del 10%, y mediante la distribución F:

H_0: $\sigma_1^2 = \sigma_2^2$	$s_1 = 200$ hr	$s_2 = 250$ hr
H_1: $\sigma_1^2 \neq \sigma_2^2$	$s_1^2 = 40\,000$	$s_2^2 = 62\,500$
	$n_1 = 10$	$n_2 = 8$

Para la prueba con el nivel de significancia del 10%, los valores críticos de F para el 5% superior y el 5% inferior son:

$$F_{9,7} \text{ crítica (5\% superior)} = 3.68$$

$$F_{9,7} \text{ crítica (5\% inferior)} = \frac{1}{F_{7,9}(5\% \text{ superior})} = \frac{1}{3.29} = 0.304$$

$$F_{gl_1,\, gl_2} = \frac{s_1^2}{s_2^2}$$

$$F_{9,7} = \frac{40\,000}{62\,500} = 0.64$$

Como el cociente F calculado no es menor que 0.304, ni mayor que 3.68, se encuentra en la región de aceptación de la hipótesis nula. Por ello, no es posible rechazar la suposición de que las varianzas de las dos poblaciones son iguales en el nivel de significancia del 10%.

11.10 MÉTODOS ALTERNATIVOS PARA PRUEBAS DE HIPÓTESIS NULAS

Tal como se describió en las secciones 10.6 y 10.7, los métodos del valor P y el del intervalo de confianza son alternativas para el método del valor crítico en las pruebas de hipótesis que se utilizaron en las secciones anteriores de este capítulo.

Mediante el método del valor P, en vez de comparar el valor observado de una estadística de prueba con un valor crítico, se determina la probabilidad de ocurrencia de la estadística de prueba, considerando que la hipótesis nula es cierta, y se le compara con el nivel de significancia a. Se rechaza la hipótesis nula si el valor *P* es *inferior* que la a designada. En los problemas 11.16 y 11.17 se ¡lustra la aplicación de este método a pruebas de dos y de un extremo, respectivamente, para la diferencia entre dos medias.

Con el método del intervalo de confianza, se construye el intervalo de confianza de 1-a para el valor del parámetro pertinente. Si el valor hipotético del parámetro no está incluido en el intervalo, entonces se rechaza la hipótesis nula. En los problemas 11.18 y 11.19 se ilustra la aplicación de este método a pruebas de dos y de un extremo, respectivamente, para la diferencia entre dos medias.

11.11 RESULTADOS POR COMPUTADORA

Existen paquetes de computación disponibles para el cálculo de la mayoría de las pruebas de hipótesis que se describieron en este capítulo. Por lo general, se utiliza el método del valor P para probar las hipótesis, tal como se describió en la sección 11.10. En el problema 11.20, se ilustra el uso de paquetes de computación para probar la diferencia entre las medias de dos muestras independientes, y en el problema 11.21, se ejemplifica el uso de esos programas con el diseño de observaciones apareadas para probar la diferencia entre dos medias.

Problemas resueltos

PRUEBAS PARA LA DIFERENCIA ENTRE DOS MEDIAS UTILIZANDO LA DISTRIBUCIÓN NORMAL

11.1 Un constructor está considerando dos lugares alternativos para un centro comercial regional. Como los ingresos de los hogares de la comunidad son una consideración importante en esa selección, desea probar la hipótesis nula de que no existe diferencia entre el ingreso promedio por hogar en las dos comunidades. Consistente con esta hipótesis, supone que la desviación estándar del ingreso por hogar es también igual en las dos comunidades. Para una muestra de m=30 hogares de la primera comunidad, encuentra que el ingreso diario promedio es $\bar{X}_1$ = \$35 500, con desviación estándar muestral de s_1 = \$1 800. Para una muestra de *m* = 40 hogares de la segunda comunidad, $\bar{X}_2$= \$34 600, y S2 = \$2400. Probar la hipótesis nula en el nivel de significancia del 5%.

$$H_0: (\mu_1 - \mu_2) = 0 \qquad H_1: (\mu_1 - \mu_2) \neq 0$$

$$\bar{X}_1 = \$35\ 500 \qquad \bar{X}_2 = \$34\ 600$$

$$s_1 = \$1\ 800 \qquad s_2 = \$2\ 400$$

$$n_1 = 30 \qquad n_2 = 40$$

$$z \text{ crítica } (\alpha = 0.05) = \pm 1.96$$

$$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{29(1\ 800)^2 + 39(2\ 400)^2}{30+40-2} = \frac{318\ 600\ 000}{68} = \$4\ 685\ 294$$

(Se combinan las varianzas debido a la suposición de que los valores de las desviaciones estándar de las poblaciones son iguales.)

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2}=\sqrt{\frac{\hat{\sigma}^2}{n_1}+\frac{\hat{\sigma}^2}{n_2}}=\sqrt{\frac{4\,685\,294}{30}+\frac{4\,685\,294}{40}}=\sqrt{156.176.46+117.132.35}=\$522.79$$

$$z=\frac{\bar{X}_1-\bar{X}_2}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}}=\frac{35\,500-34\,600}{522.79}=\frac{900}{522.79}=+1.72$$

El valor calculado de z de +1.72 se encuentra en la región de aceptación de la hipótesis nula. Por ello, no es posible rechazar la hipótesis nula al nivel de significancia del 5%, y se acepta la hipótesis de que el ingreso promedio por hogar de las dos comunidades no es diferente.

11.2 Con referencia al problema 11.1, antes de recolectar los datos, el constructor consideró que el ingreso de la primera comunidad pudiera ser superior. Con el objeto de someter esta evaluación a una prueba crítica, le otorgó el beneficio de la duda a la otra posibilidad y planteó la hipótesis nula $(\mu_1 - \mu_2) \leq 0$. Pruebe esta hipótesis con un nivel de significancia del 5%, con la suposición adicional de que los valores de la desviación estándar para las dos poblaciones *no* son necesariamente iguales.

$$H_0\text{: } (\mu_1-\mu_2)\leq 0 \qquad H_1\text{: } (\mu_1-\mu_2)>0$$

$$z \text{ crítica } (\alpha=0.05)=+1.645$$

$$s_{\bar{x}_1}=\frac{s_1}{\sqrt{n_1}}=\frac{1\,800}{\sqrt{30}}=\frac{1\,800}{5.48}=\$328.47$$

$$s_{\bar{x}_2}=\frac{s_2}{\sqrt{n_2}}=\frac{2\,400}{\sqrt{40}}=\frac{2\,400}{6.32}=\$379.75$$

$$s_{\bar{x}_1-\bar{x}_2}=\sqrt{s_{\bar{x}_1}^2+s_{\bar{x}_2}^2}=\sqrt{(328.47)^2+(379.75)^2}=\sqrt{252\,102.60}=\$502.10$$

$$z=\frac{\bar{X}_1-\bar{X}_2}{s_{\bar{x}_1-\bar{x}_2}}=\frac{35\,500-34\,600}{502.10}=\frac{900}{502.10}=+1.79$$

El valor calculado z de +1.79 es mayor que el valor crítico de +1.645 para esta prueba del extremo superior. Por ello, se rechaza la hipótesis nula a un nivel de significancia del 5%, y se acepta la hipótesis alternativa de que el ingreso promedio por hogar es mayor en la primera comunidad que en la segunda.

11.3 Con respecto a los problemas 11.1 y 11.2, antes de recolectar los datos, el constructor consideró que el ingreso promedio de la primera comunidad excede al promedio de la segunda comunidad en cuando menos $1500 diarios. En este caso, concediendo a esta evaluación el beneficio de la duda, pruebe esa suposición como hipótesis nula utilizando un nivel de significancia del 5%. No se supone que las desviaciones estándar de las poblaciones son iguales.

$$H_0\text{: } (\mu_1-\mu_2)\geq 1\,500 \qquad H_1\text{: } (\mu_1-\mu_2)<1\,500$$

$$z \text{ crítica } (\alpha=0.05)=-1.645$$

$$s_{\bar{x}_1-\bar{x}_2}=\$502.10 \quad \text{(del problema 11.2)}$$

$$z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)_0}{s_{\bar{x}_1-\bar{x}_2}}=\frac{(35\,500-34\,600)-1\,500}{502.10}=\frac{-600}{502.10}=-1.19$$

El valor calculado de z de -1.19 *no* es menor que el valor crítico de -1.645 para esta prueba del extremo inferior. Por ello, no es posible rechazar la hipótesis a un nivel de significancia del 5%. Aunque la diferencia de la muestra ($900) no equivale a la diferencia de $1500 que el constructor supuso, no es lo suficientemente distinta cuando se le otorga a esa suposición el beneficio de la duda y se le considera como la hipótesis nula.

PRUEBA PARA LA DIFERENCIA ENTRE DOS MEDIAS UTILIZANDO LA DISTRIBUCIÓN *t*

11.4 De 100 recién graduados en contaduría de una escuela superior de administración de empresas, una muestra aleatoria de $n_1 = 12$ estudiantes tiene un promedio de calificación de 2.70 (en donde la calificación más alta es de 4), con una desviación estándar muestral de 0.40. Para los 50 recién egresados de sistemas de información computarizada, una muestra aleatoria de $n_2 = 10$ estudiantes tiene un promedio de calificación de $\bar{X}_2 = 2.90$, con desviación estándar de 0.30. Se supone que las calificaciones tienen distribución normal. Pruebe la hipótesis nula de que la calificación promedio para las dos categorías de estudiantes es distinta, utilizando el nivel de significancia del 5%.

$$H_0: (\mu_1 - \mu_2) = 0 \qquad H_1: (\mu_1 - \mu_2) \neq 0$$

$$\bar{X}_1 = 2.70 \qquad \bar{X}_2 = 2.90$$

$$s_1 = 0.40 \qquad s_2 = 0.30$$

$$n_1 = 12 \qquad n_2 = 10$$

$$N_1 = 100 \qquad N_2 = 50$$

$$t \text{ crítica } (gl = 20, \alpha = 0.05) = \pm 2.086$$

(Nota: En la sección 11.2 se especifica que una suposición necesaria cuando se utiliza la distribución t para probar la diferencia entre dos medias es que las varianzas sean iguales.) Por ello, se combinan las dos varianzas muestrales:

$$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(11)(0.40)^2 + (9)(0.30)^2}{12+10-2} = \frac{1.76+0.81}{20} = \frac{2.57}{20} = 0.128$$

$$\hat{\sigma}_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\hat{\sigma}^2}{n_1}\left(\frac{N_1-n_1}{N_1-1}\right) + \frac{\hat{\sigma}^2}{n_2}\left(\frac{N_2-n_2}{N_2-1}\right)} = \sqrt{\frac{0.128}{12}\left(\frac{100-12}{100-1}\right) + \frac{0.128}{10}\left(\frac{50-10}{50-1}\right)}$$

$$= \sqrt{0.011\left(\frac{88}{99}\right) + 0.013\left(\frac{40}{49}\right)} = \sqrt{0.011(0.889) + 0.013(0.816)} = \sqrt{0.020387} = 0.143$$

(Para cada muestra $n < 0.05\,N$ y, por ello, se requiere utilizar el factor de corrección por población finita.)

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{x}_1-\bar{x}_2}} = \frac{2.70-2.90}{0.143} = \frac{-0.20}{0.143} = -1.399$$

El valor calculado de t, -1.399, se encuentra en la región de aceptación de la hipótesis nula. Por ello, no es posible rechazar la hipótesis nula de que no existe diferencia entre los promedios de calificaciones para las dos poblaciones de estudiantes, a un nivel de significancia del 5%.

PRUEBA DE LA DIFERENCIA ENTRE DOS MEDIAS CON BASE EN OBSERVACIONES APAREADAS

11.5 La directora de capacitación de una compañía desea comparar un nuevo método de capacitación técnica, que implica una combinación de discos de computación de tutoría y resolución de problemas en laboratorio, junto con el método tradicional de análisis de casos. Ella designa doce pares de entrenandos, de acuerdo con sus antecedentes y a su desempeño académico, y asigna a un miembro de cada par a la clase tradicional y el otro al nuevo método. Al final del curso, se determina el nivel de aprendizaje mediante un examen que abarca información básica, al igual que la habilidad para aplicar esa información. Como la directora de capacitación desea otorgar el beneficio de la duda al sistema establecido de enseñanza, plantea la hipótesis nula de que el desempeño promedio para el sistema establecido es igual o mayor que el nivel promedio del desempeño para el nuevo sistema. Pruebe esta hipótesis con el nivel de significancia del 5%. En las primeras 3 columnas de la tabla 11.3 se presentan los datos del desempeño de esta muestra.

Tabla 11.3 Datos del programa de capacitación y hoja de trabajo para calcular la diferencia promedio y la desviación estándar de la diferencia.

Paren capacitación	Método tradicional (X_1)	Método nuevo (X_2)	d $(X_1 - X_2)$	d^2
1	89	94	-5	25
2	87	91	-4	16
3	70	68	2	4
4	83	88	-5	25
5	67	75	-8	64
6	71	66	5	25
7	92	94	-2	4
8	81	88	-7	49
9	97	96	1	1
10	78	88	-10	100
11	94	95	-1	1
12	79	87	-8	64
Total	988	1,030	-42	378

$$\text{Desempeño promedio (método tradicional)} = \frac{988}{12} = 82.33$$

$$\text{Desempeño promedio (método nuevo)} = \frac{1030}{12} = 85.83$$

$$H_0: \mu_d \geq 0 \qquad H_1: \mu_d < 0$$

$$t \text{ crítica } (gl = 11, \alpha = 0.05) = -1.796$$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{-42}{12} = -3.5$$

$$s_d = \sqrt{\frac{\Sigma d^2 - n\bar{d}^2}{n-1}} = \sqrt{\frac{378 - 12(-3.5)^2}{11}} = \sqrt{\frac{378-147}{11}} = \sqrt{\frac{231}{11}} = \sqrt{21} = 4.58$$

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \frac{4.58}{\sqrt{12}} = \frac{4.58}{3.46} = 1.32$$

$$t = \frac{\bar{d}}{s_{\bar{d}}} = \frac{-3.5}{1.32} = -2.652$$

El valor calculado de f de -2.652 es menor que el valor crítico de -1.796 para esta prueba del extremo inferior. Por ello, se rechaza la hipótesis nula a un nivel de significancia del 5% y se concluye que el nivel promedio de desempeño para las personas que fueron capacitadas con el nuevo método es superior al de quienes fueron capacitados con el método tradicional.

PRUEBA PARA EL VALOR HIPOTÉTICO DE PROPORCIÓN UTILIZANDO LA DISTRIBUCIÓN BINOMIAL

11.6 Cuando un proceso de producción se encuentra bajo control, el porcentaje de artículos defectuosos que se tienen que eliminar en el proceso de inspección no supera el 1 %. Para una muestra aleatoria de n =10 artículos, se encuentra uno defectuoso. Con base en este resultado muestral, ¿puede rechazarse la hipótesis nula de que el proceso está bajo control a un nivel de significancia del 5%?

Para las hipótesis H_0: $\pi \leq 0.01$ y H_1: $\pi > 0.01$, con base en la distribución binomial, la probabilidad de obtener uno o más artículos defectuosos por efectos del azar, dado que $\pi = 0.01$ es 1.0 menos la probabilidad de obtener cero defectuosos (del apéndice 1, con *n* = 10, p = 0.01):

$$P(X \geq 1 \mid n = 10, p = 0.01) = 1.000 - 0.9044 = 0.0956$$

Como esta probabilidad es mayor que 0.05, no puede rechazarse la hipótesis nula. Para este problema, se tendrían que encontrar dos o más artículos defectuosos para poder rechazar la hipótesis nula, porque la probabilidad correspondiente a este "extremo" de la distribución es inferior al 0.05. Además, la probabilidad de que dos o más artículos estén defectuosos es también inferior a 0.0l:

$$P(X \geq 2 \mid n = 10, p = 0.01) = 0.0042 + 0.0001 + 0.0000 + \cdots = 0.0043$$

11.7 Con referencia al ejemplo 5 (página 182), suponga que una muestra de n = 20 estudiantes señala en una encuesta que sólo cuatro tienen empleos hacia el primero de marzo (la misma proporción muestral del ejemplo 5). ¿Puede rechazarse la afirmación del director en este caso, utilizando un nivel de significancia del 5%?

Para las hipótesis H_0: $\pi \geq 0.05$ y H_1:1 $\pi < 0.50$, con base en la distribución binomial, las probabilidades de que los resultados muestrales difieran de la afirmación del director, y que no excedan una probabilidad de 0.05, son las siguientes (del apéndice 2, con *n* = 20 y p = 0.50):

Número de estudiantes que obtiene empleo	Probabilidad	Probabilidad acumulada
0	0.0000	
1	0.0000	
2	0.0002	
3	0.0011	0.0207
4	0.0046	
5	0.0148	
6	0.0370	

Por lo tanto, para una prueba de un extremo al 5% de nivel de significancia (de hecho, a un nivel de 2.07%) el número critico para el rechazo es 5 o menos. Incluir la categoría "6" daría como resultado una probabilidad mayor que 0.05.

Dado el resultado muestral de que sólo *cuatro* estudiantes reportaron tener empleos, se rechaza la hipótesis nula. Se debe observar que, aun cuando la proporción muestral es la misma que la del ejemplo 5, el mayor tamaño de muestra está asociado con un menor error muestral, lo cual conduce a una prueba más sensible para detectar la diferencia.

11.8 Se supone que el 40% de los votantes en una elección votará por el candidato, y que el otro 60% de los votos estará distribuido entre otros tres candidatos. De una muestra aleatoria de 20 votantes registrados que tienen intenciones de votar en esas elecciones, 12 señalan que votarán por el candidato. Pruebe la hipótesis de que la proporción global de votantes que estará a favor del candidato es $\pi = 0.40$, utilizando un nivel de significancia del 5%.

Como $H_0: \pi = 0.40$ y $H_1: \pi \neq 0.40$, con base en la distribución binomial, las probabilidades de las observaciones extremas en cualquiera de los "extremos" de la distribución, y que no exceden una probabilidad acumulada de 0.025 en cada extremo, son las siguientes (del apéndice 1, con n = 20, p = 0.40).

Número de candidato	Probabilidad	Probabilidad acumulada
0	**0.0000**	
1	**0.0005**	
2	**0.0031**	0.0159
3	**0.0123**	
4	**0.0350**	
12	**0.0355**	
13	**0.0146**	
14	**0.0049**	
15	**0.0013**	
16	**0.0003**	0.0211
17	**0.0000**	
18	**0.0000**	
19	**0.0000**	
20	**0.0000**	

Por lo tanto, para la prueba de dos extremos, el nivel global de significancia que no excede 0.025 en cada extremo es, de hecho, $\alpha = 0.037$. El número critico de elementos de la muestra que implicaría el rechazo es "tres o menos* o "trece o más". Como 12 votantes señalaron que tenían intenciones de votar por el candidato, no es posible rechazar la hipótesis'nula.

PRUEBA DE UN VALOR HIPOTÉTICO DE LA PROPORCIÓN, UTILIZANDO LA DISTRIBUCIÓN NORMAL

11.9 Se plantea la hipótesis de que no más del 5% de las refacciones que se fabrican en un proceso de manufactura tienen defectos. Para una muestra aleatoria de n = 100 refacciones, se encuentra que 10 están defectuosas. Pruebe la hipótesis nula al 5% del nivel de significancia.

$$H_0\text{: } \pi \leq 0.05 \qquad H_1\text{: } \pi > 0.05$$
$$\text{z crítica } (\alpha = 0.05) = +1.645$$

(Se justifica el uso de la distribución normal porque $n \geq 30$, $n\pi_0 \geq 5$, y $n(1 - \pi_0) \geq 5$.)

$$\sigma_{\hat{p}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{(0.05)(0.95)}{100}} = \sqrt{\frac{0.0475}{100}} = \sqrt{0.000475} = 0.022$$

$$z = \frac{\hat{p} - \pi_0}{\sigma_{\hat{p}}} = \frac{0.10 - 0.05}{0.022} = \frac{0.05}{0.022} = +2.27$$

El valor calculado de z de + 2.27 es mayor que el valor crítico de +1.645 para esta prueba del extremo superior. Por lo tanto, como se encuentran 10 refacciones defectuosas en el lote de 100, se rechaza la hipótesis de que la proporción de artículos defectuosos en la población es de 0.05 o menor, utilizando un nivel de significancia del 5% en la prueba.

11.10 Para el problema 11.9, el administrador estipula que la probabilidad de detener el proceso para ajustarlo, cuando de hecho no es necesario, debe ser a un nivel de sólo el 1 %, mientras que la probabilidad de *no* detener el proceso cuando la proporción verdadera de defectuosos es de $\pi = 0.10$ puede fijarse en el 5%. ¿Qué tamaño de muestra debe obtenerse, como mínimo, para satisfacer esos objetivos de prueba?

$$n = \left[\frac{z_0\sqrt{\pi_0(1-\pi_0)} - z_1\sqrt{\pi_1(1-\pi_1)}}{\pi_1 - \pi_0}\right]^2 = \left[\frac{2.33\sqrt{(0.05)(0.95)} - (-1.645)\sqrt{(0.10)(0.90)}}{0.10 - 0.05}\right]^2$$
$$= \left[\frac{2.33(0.218) + 1.645(0.300)}{0.05}\right]^2 = \left(\frac{1.0014}{0.05}\right)^2 = (20.03)^2 = 401.2 = 402 \text{ refacciones}$$

Se trata de una muestra un tanto grande para efectos de muestreo industrial, por lo que el administrador podría reconsiderar los objetivos de la prueba con respecto a la P (error tipo I) de 0.01 y la *P* (error tipo II) de 0.05.

PRUEBA PARA LA DIFERENCIA ENTRE DOS PROPORCIONES POBLACIONALES

11.11 Un fabricante está evaluando dos tipos de equipo para fabricar un artículo. Se obtiene una muestra aleatoria de $n_1 = 50$ para la primera marca de equipo y se encuentra que 5 de ellos tienen defectos. Se obtiene una muestra aleatoria de $n_2 = 80$ para la segunda marca y se encuentra que 6 de ellos tienen defectos. La tasa de fabricación es la misma para las dos marcas. Sin embargo, como la primera cuesta bastante menos, el fabricante le otorga a esa marca el beneficio de la duda y plantea la hipótesis H_0: $\pi_1 \leq \pi_2$. Pruebe la hipótesis en el nivel de significancia del 5%.

$$H_0\text{: } (\mu_1 - \mu_2) \leq 0 \qquad H_1\text{: } (\mu_1 - \mu_2) > 0$$
$$\text{z crítica } (\alpha = 0.05) = +1.645$$

$$\hat{\pi} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2} = \frac{50(0.10) + 80(0.075)}{50 + 80} = \frac{5 + 6}{130} = 0.085$$

$$\hat{\sigma}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n_1} + \frac{\hat{\pi}(1-\hat{\pi})}{n_2}} = \sqrt{\frac{(0.085)(0.915)}{50} + \frac{(0.085)(0.915)}{80}}$$

$$= \sqrt{\frac{0.0778}{50} + \frac{0.0778}{80}} = \sqrt{0.0016 + 0.0010} = 0.051$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}_{\hat{p}_1-\hat{p}_2}} = \frac{0.10 - 0.075}{0.051} = \frac{0.025}{0.051} = +0.49$$

El valor calculado de z de +0.49 no es mayor que +1.645 para esta prueba del extremo superior. Por ello, no puede rechazarse la hipótesis nula en el nivel de significancia del 5%.

PRUEBA DEL VALOR HIPOTÉTICO DE UNA VARIANZA UTILIZANDO LA DISTRIBUCIÓN JI-CUADRADA

11.12 Suponga que se plantea la hipótesis de que la desviación estándar del salario por hora de los trabajadores a destajo en una determinada industria es \$3000. Para una muestra de n = 15 trabajadores elegidos al azar, se encuentra que la desviación estándar es s = \$2000. Se supone que las cifras de ingresos de los trabajadores de la población tienen una distribución normal. Con base en este resultado muestral, ¿puede rechazarse la hipótesis nula utilizando el nivel de significancia del 5%?

$$H_0 : \sigma^2 = (\$3000)^2 = \$9\,000\,000 \qquad H_1 : \sigma^2 \neq \$9\,000\,000$$

X^2 critica **(gl = 14, α = 0.05) = 5.63 y 26.12** (respectivamente, para la prueba de dos extremos)

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{(14)(2.000)^2}{(3.000)^2} = \frac{14(4\,000\,000)}{9\,000{,}000} = \frac{56\,000\,000}{9\,000{,}000} = 6.22$$

El valor calculado de 6.22 es mayor que el valor crítico de 5.63 y menor que el valor crítico superior de 26.12 para esta prueba de dos extremos. Por ello, no se rechaza la hipótesis nula de que σ = \$ 3000 a nivel de significancia del 5%.

11.13 Suponga que, en el problema 11.12, la hipótesis nula consistía en que la desviación estándar de la población es de *cuando menos* \$3000. Pruebe esa hipótesis a un nivel de significancia del 5%.

$$H_0: \sigma^2 \geq \$9\ 000\ 000 \qquad H_1: \sigma^2 < \$9\ 000\ 000$$

x^2 **crítica (gl = 14, α = 0.05) = 6.57** (valor crítico del extremo inferior)

x^2 **= 6.22** (del problema 11.12)

La estadística de prueba calculada de 6.22 es apenas menor que el valor crítico de 6.57 para esta prueba del extremo inferior. Por ello, se rechaza la hipótesis nula a un nivel de significancia del 5%, y se acepta la hipótesis alternativa de que σ^2 < **\$9 000 000 (que σ < \$3000).**

PRUEBA DE LA DIFERENCIA ENTRE DOS VARIANZAS

11.14 En el problema 11.4, que se refería a la prueba de la diferencia entre dos medias muestrales utilizando una distribución f, se requirió la suposición de que las dos varianzas poblacionales eran iguales. Las dos varianzas muestrales eran $s_1^2 = 0.40^2 = 0.16$ y $s_2^2 = 0.30^2 = 0.09$, con $n_1 = 12$ y $n_2 = 10$, respectivamente. Pruebe la hipótesis nuia de que las dos varianzas poblacionales son iguales, utilizando un nivel de significancia del 10%.

$$H_0\colon \sigma_1^2 = \sigma_2^2 \qquad s_1^2 = 0.16 \qquad s_2^2 = 0.09$$
$$H_1\colon \sigma_1^2 \neq \sigma_2^2 \qquad n_1 = 12 \qquad n_2 = 10$$

$$F_{11,9}\text{ crítica (5\% superior)} = 3.10 \qquad \text{(del apéndice 8)}$$

$$F_{11,9}\text{ crítica(5\% inferior)} = \frac{1}{F_{9,11}(5\%\text{ superior})} = \frac{1}{2.90} = 0.345$$

$$F_{gl_1,\,gl_2} = \frac{s_1^2}{s_2^2} = \frac{0.16}{0.09} = 1.78$$

El valor calculado de F de 1.78 es mayor que el valor crítico inferior de 0.345 y menor que el valor crítico superior de 3.10. Por lo tanto, no es posible rechazar la hipótesis de que no existe diferencia entre las varianzas.

11.15 En el problema 11.1, se supuso que la varianza de los ingresos por hogar no era diferente en las dos comunidades. Pruebe la hipótesis nula de que las dos varianzas son iguales, utilizando un nivel de significancia del 10%

$$H_0\colon \sigma_1^2 = \sigma_2^2 \qquad s_1^2 = (1\ 800)^2 = 3\ 240\ 000 \qquad s_2^2 = (2\ 400)^2 = 5\ 760\ 000$$
$$H_1\colon \sigma_1^2 \neq \sigma_2^2 \qquad n_1 = 30 \qquad n_2 = 40$$

(Nota: Debido a las limitaciones del apéndice 8, no es posible determinar los valores específicos de F para 29 y 39 grados de libertad. Por ello, se determinan valores F aproximados utilizando los grados de libertad más cercanos, que son 30 y 40, respectivamente.)

$$F_{30,40}\text{ crítica (5\% superior)} = 1.74$$

$$F_{3040}\text{ crítica (5\% inferior)} = \frac{1}{F_{40,30}(5\%\text{ superior})} = \frac{1}{1.79} =$$

$$F_{gl_1,\,gl_2} = \frac{s_1^2}{s_2^2} = \frac{3\ 240\ 000}{5\ 760\ 000} = 0.562$$

El valor calculado de la estadística F es mayor que el valor crítico inferior de 0.559, y menor que el valor crítico superior de 1.74. Por ello, la estadística F se encuentra apenas dentro de la región de aceptación de la hipótesis nula a un nivel de significancia del 10%.

MÉTODOS ALTERNATIVOS PARA PRUEBAS DE HIPÓTESIS NULAS

11.16 Al utilizar el método del valor *P,* pruebe la hipótesis nula del problema 11.1 a un nivel de significancia del 5%.

$$H_0: (\mu_1 - \mu_2) = 0 \qquad H_1: (\mu_1 - \mu_2) \neq 0$$

$$Z = +1.72 \qquad \text{(del problema 11.1)}$$

$$P(z \geq +1.72) = 0.5000 - 0.4573 = 0.0427$$

Finalmente, como se trata de una prueba de dos extremos.

$$P = 2(0.0427) = 0.0854$$

Como el valor *P* de 0.0854 es mayor que el nivel de significancia de 0.05, *no es posible* rechazar la hipótesis nula a ese nivel. Por ello, se concluye que no existe diferencia entre los promedios de ingresos en los hogares en las dos comunidades.

11.17 Al utilizar el método del valor P, pruebe la hipótesis nula del problema 11.2 con un nivel de significancia del 5%.

$$H_0: (\mu_1 - \mu_2) \leq 0 \qquad H_1: (\mu_1 - \mu_2) > 0$$

$$z = +1.79 \qquad \text{(del problema 11.2)}$$

(Debe observarse que la estadística z se encuentra en la dirección de la región de rechazo para esta prueba del extremo superior.)

$$P(z \geq +1.79) = 0.5000 - 0.4633 = 0.0367$$

Como el valor de P de 0.0367 es menor que el nivel de significancia de 0.05, se rechaza la hipótesis nula y se concluye que el nivel promedio de ingresos en la primera comunidad es mayor que en la segunda.

11.18 Aplique el método del intervalo de confianza para probar la hipótesis nula del problema 11.1, utilizando un nivel de significancia del 5%.

$$H_0: (\mu_1 - \mu_2) = 0 \qquad H_1: (\mu_1 - \mu_2) \neq 0$$

$$\bar{X}_1 = \$35\,500 \qquad \bar{X}_2 = \$34\,600$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \$522.79 \qquad \text{(del problema 11.1)}$$

$$\begin{aligned} \text{lnt.de conf.del95\%} &= (\bar{X}_1 - \bar{X}_2) \pm z\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} \\ &= (35\,500 - 34\,600) \pm 1.96(522.79) \\ &= 900 \pm 1\,024.67 = -\$124.67 \text{ a } \$1\,924.67 \end{aligned}$$

Como el intervalo de confianza del 95% incluye la diferencia hipotética de \$0, no es posible rechazar la hipótesis a un nivel de significancia del 5%.

11.19 Aplique el método del intervalo de confianza para probar la hipótesis nula del problema 11.2.

$$H_0: (\mu_1 - \mu_2) \leq 0 \qquad H_1: (\mu_1 - \mu_2) > 0$$

$$\bar{X}_1 = \$35\,500 \qquad \bar{X}_2 = \$34\,600$$

$$s_{\bar{x}_1 - \bar{x}_2} = \$502.10 \qquad \text{(del problema 11.2)}$$

$$\text{Límite inferior del intervalo de confianza} = (\bar{X}_1 - \bar{X}_2) - z s_{\bar{x}_1 - \bar{x}_2}$$
$$= (35\,500 - 34\,600) - 1.645(502.10)$$
$$= 900 - 825.95 = \$74.05$$

Con una confianza del 95%, se concluye que la diferencia entre las medias poblacionales puede ser hasta de \$74.05. Como este intervalo de confianza de un extremo no incluye el valor hipotético de \$0 (o menos), se rechaza la hipótesis nula a un nivel de significancia del 5% y se concluye que el ingreso promedio por hogar en la primera comunidad es mayor que para la segunda.

RESULTADOS POR COMPUTADORA

11.20 Con referencia a los datos de entrada y a los resultados que se obtuvieron en el problema 9.15 (página 153), se tienen los datos de dos muestras aleatorias de reclamaciones de daños de automóviles, habiendo obtenido las muestras en áreas geográficas distintas. La primera parte de los resultados que se analizaron en el problema 9.15 incluyen el intervalo de confianza del 95% para la diferencia entre las dos medias poblacionales. Con referencia a la última parte del listado de la computadora, pruebe la hipótesis nula de que no existe diferencia entre la media de las dos poblaciones, utilizando un nivel de significancia del 5% para la prueba.

$$H_0: (\mu_1 - \mu_2) = 0 \qquad H_1: (\mu_1 - \mu_2) \neq 0 \qquad \alpha = 0.05$$

En la última parte del listado de la figura 9-1, se observa que el valor reportado de *P* para la prueba de dos extremos es de 0.019. Como esta probabilidad es menor que el nivel de significancia especificado de 0.05, se rechaza la hipótesis nula. Se concluye que sí existe una diferencia entre el nivel promedio de reclamaciones de daños en las dos áreas. Debe observarse que, en este listado, puede utilizarse también el método del intervalo de confianza para probar la hipótesis nula. Como el intervalo de confianza del 95% *no* incluye la diferencia de 0, se rechaza la hipótesis nula de que la diferencia entre las medias de las poblaciones es 0, a un nivel de significancia del 5%.

11.21 Con referencia a los datos de la Tabla 11.3, que se referían a la prueba de la diferencia entre medias de observaciones apareadas del problema 11.5, pruebe la hipótesis nula que se presentó en ese problema a un nivel de significancia del 5%, utilizando algún paquete de computación.

$$H_0: \mu_d \geq 0 \qquad H_1: \mu_d < 0 \qquad \alpha = 0.05$$

Observe la figura 11-3. Se reporta que el valor P es de 0.011. Como esta probabilidad es menor que el nivel especificado de significancia de 0.05, se rechaza la hipótesis nula a ese nivel de significancia y se concluye que el nivel promedio de desempeño para las personas capacitadas con el nuevo método es superior al desempeño obtenido con el método tradicional. Este resultado coincide con la solución manual que se obtuvo utilizando el método del valor critico en el problema 11.5. *(Nota:* El "subcomando" ALTERNATIVE = -1 especifica una prueba del extremo inferior en Minitab.)

```
MTB >  READ OLD METHOD AND NEW METHOD  INTO C1  AND C2
DATA>    89     94
DATA>    87     91
DATA>    70     68
DATA>    83     88
DATA>    67     75
DATA>    71     66
DATA>    92     94
DATA>    81     88
DATA>    97     96
DATA>    78     88
DATA>    94     95
DATA>    79     87
DATA> END
MTB >  NAME C1 - 'OLD', C2 = 'NEW, C3 = 'CHANGE'
MTB >  LET 'CHANGE' = 'OLD'  =  'NEW
MTB >  TTEST OF MU = 0  FOR  'CHANGE';
SUBC> ALTERNATIVE = -1.

TEST (DF MU = 0. 00 VS  MU L.T.  0.00

                N       MEAN     STDEV    SE MEAN         T     P VALUE
CHANGE         12      -3.50      4.58       1.32     -2.65       0.011
```

Problemas suplementarios

PRUEBA DE LA DIFERENCIA ENTRE DOS MEDIAS

11.22 Tal como se reportó en el problema 9.16, el promedio de ventas por tienda de un articulo el año anterior fue en una muestra de n_1 =10 tiendas, $\bar{X}_1$ = \$3 425 000, con s = \$200 000. Para un segundo producto, el promedio de ventas por tienda en una muestra de n =12 tiendas fue $\bar{X}_2$ = \$3 250 000, con s_2 = \$175 000. Se supuso que las cantidades por tienda tienen una distribución normal para ambos productos. Pruebe la hipótesis nula de que no existe diferencia entre el promedio de ventas para los dos productos utilizando el nivel de significancia del 1 %.

Resp. Aceptar H_0.

11.23 Para los datos del problema 11.22, suponga que los tamaños de las dos muestras eran n_1 = 20 y n_z = 24. Pruebe la diferencia entre las dos medias a nivel de significancia del 1 %.

Resp. Rechazar H_0.

11.24 Para una muestra de 30 empleados de una empresa grande, el salario promedio por semana es de $\overline{X}_1=$ \$95 000, con s = \$10 000. En una segunda empresa grande, el salario promedio por hora para una muestra de 40 empleados es $\overline{X}_2=$ \$90 500, con $s_2=$ \$12 000. Pruebe la hipótesis de que no existe diferencia entre el salario promedio que se paga en las dos empresas, utilizando un nivel de significancia del 5%, y suponiendo que las varianzas de las dos poblaciones no son necesariamente iguales.

Resp. Aceptar H_0.

11.25 En el problema 11.24, suponga que la hipótesis nula que se desea probar es que el salario promedio de la segunda empresa es igual o mayor que el salario promedio de la primera empresa. ¿Puede rechazarse la hipótesis a un nivel de significancia del 5%?

Resp. SI

11.26 Una muestra aleatoria de n_1 = 10 vendedores se inscribe en un programa de incentivos, en tanto que una muestra aleatoria de n_2 = 10 vendedores distintos se inscriben en un segundo sistema de incentivos. Durante el periodo de comparación, los vendedores que se encuentran en el primer sistema tienen comisiones promedio por venta de $\overline{X}_1=$ \$5 000, con una desviación estándar de s = \$1 200, en tanto que los vendedores que participan en el segundo sistema tienen comisiones promedio por artículo de $\overline{X}_2$ = \$4 600, con una desviación estándar de s_1 = \$1 000. Pruebe la hipótesis nula de que no existe diferencia entre las comisiones por artículo para los dos sistemas de incentivos, utilizando el nivel de significancia del 5%.

Resp. Aceptar H_0.

11.27 Con el objeto de comparar dos paquetes de computación, un administrador hace que 10 personas utilicen cada uno de los paquetes para llevar a cabo un conjunto estándar de tareas comunes en la oficina. Por supuesto, al llevar a cabo la comparación, el administrador tiene el cuidado de utilizar personas que no tengan una preferencia o una capacidad distinguible en ninguno de los dos paquetes, y se seleccionan 5 personas para utilizar el paquete A en primer lugar, en tanto que los otros 5 utilizan en primer lugar el paquete B. El tiempo que se requiere para llevar a cabo el conjunto de tareas, al minuto más cercano, es el que se reporta en la tabla 11.4. Pruebe la hipótesis nula de que no existe diferencia entre el tiempo promedio que se requiere para realizar los trabajos estándar utilizando los dos paquetes de computación, con un nivel de significancia del 5%.

Resp. Rechazar H_0

Tabla 11.4 Tiempo requerido para realizar un conjunto estándar de labores utilizando dos paquetes de computación (redondeado al minuto más cercano)

Persona	1	2	3	4	5	6	7	8	9	10
Paquete A	12	16	15	13	16	10	15	17	14	12
Paquete B	10	17	18	16	19	12	17	15	17	14

PRUEBA DE UNA PROPORCIÓN HIPOTÉTICA UTILIZANDO LA PROPORCIÓN BINOMIAL

11.28 Suponga que se plantea la hipótesis de que una moneda es "justa", no habiendo la oportunidad de examinarla en forma directa. Se lanza la moneda, y se obtiene como resultado 5 "caras" en las 5 ocasiones. Pruebe la hipótesis nula con los niveles de significancia de (a) 5%, y (b) 10%.

Resp. (a) Aceptar H_0, (b) Rechazar H_0

11.29 Un vendedor afirma que, en promedio, obtiene pedidos de cuando menos el 30% de sus prospectos. Para una muestra aleatoria de 10 prospectos, obtiene solamente un pedido. ¿Puede rechazarse su afirmación con base en el resultado muestral y a un nivel de significancia del 5%?

Resp. No

11.30 El patrocinador de un programa "especial" de televisión esperaba que cuando menos el 40% del auditorio observaría el programa en una área metropolitana especifica. Para una muestra aleatoria de 20 hogares que tenían sus televisores prendidos, sólo en 4 de ellos se estaba observando el programa. Con base en este tamaño limitado de muestra, pruebe la hipótesis nula de que cuando menos el 40% de los televidentes estaban observando el programa, utilizando un nivel de significancia del 10%.

Resp. Rechazar H_0

PRUEBAS PARA LAS PROPORCIONES UTILIZANDO LA DISTRIBUCIÓN NORMAL

11.31 Con referencia al problema 11.29, suponga que el vendedor logra obtener 20 pedidos de 100 prospectos seleccionados al azar. ¿Puede rechazarse su afirmación a un nivel de significancia de (a) 5%, y (b)1 %?

Resp. (a) Sí, (b) no

11.32 Con referencia al problema 11.30, se amplía el tamaño de la muestra, de manera que se visitan 100 hogares con los televisores prendidos. De estos 100 hogares, 30 tenían sintonizado el programa especial. ¿Puede rechazarse la suposición del patrocinador de que cuando menos el 40% de los hogares estarían observando el programa a un nivel de significancia de (a) 10%, y *(b)* 5%?

Resp. (a) Sí, (b) sí.

11,33 Para los problemas 11.30 y 11.32, suponga que el patrocinador especifica que, como resultado del estudio, la probabilidad de rechazar una afirmación verdadera no debe ser mayor que $P = 0.02$, y que la probabilidad de aceptar la afirmación dado que el porcentaje de personas que observan el programa es realmente de 30% o menos, no debe ser menor que $P = 0.05$. ¿Qué tamaño de muestra se requiere en el estudio, como mínimo, para satisfacer estos requerimientos?

Resp. 311 hogares

11.34 Se sugiere, en los problemas 11.30 y 11.32, que el programa pudiera representar un atractivo distinto para los televidentes urbanos y los de los suburbios, pero existe una diferencia de opinión entre el personal de producción con respecto al sentido de la diferencia. Para una muestra aleatoria de 50 hogares urbanos, 20 reportaron haber estado observando el programa. Para una muestra aleatoria de 50 hogares de los suburbios, 30 reportaron estar observando el programa. ¿Puede considerarse significativa la diferencia a un nivel del (a) 10%, y (b) 5%?

Resp. (a) Sí, (b) sí

PRUEBAS SOBRE EL VALOR HIPOTÉTICO DE LA VARIANZA Y DE LA DIFERENCIA ENTRE DOS VARI ANZAS

11.35 Con base en las especificaciones dadas por un ingeniero de proceso, se plantea la hipótesis de que la desviación estándar de los diámetros de ciertas piezas no es mayor de 3.0 mm. Para una muestra de n =12 piezas, se encuentra una desviación estándar muestral de s = 4.2 mm. Se supone que la distribución de los diámetros es aproximadamente normal. ¿Puede rechazarse la hipótesis nula de que la desviación estándar verdadera no es mayor de 3.0 mm a un nivel de significancia del (a) 5% y (b) 1 %?

Resp. (a) Si, (b) no

11.36 En el problema 11.22, bajo la suposición necesaria de que las varianzas de las dos poblaciones eran iguales, no se pudo rechazar la hipótesis nula de que las medias eran iguales utilizando la prueba t a un nivel de significancia del 5%. En un nivel de significancia del 10%, ¿se justifica la suposición de que las dos varianzas no son diferentes?

Resp. Si

11.37 Se diseña un nuevo proceso de moldeo para reducir la variabilidad en el diámetro de las piezas. Para probar el nuevo proceso, se plantea, conservadoramente, la hipótesis de que la varianza de los diámetros de las piezas con el nuevo proceso es igual o mayor que la varianza para el proceso antiguo. Rechazar esta hipótesis nula permitiría aceptar la alternativa de que la varianza del proceso nuevo es menor que para el proceso antiguo. Para una muestra de n_1 = 8 piezas producidas con el nuevo proceso, s,= 4.2 mm. Para una muestra de n_2= 10 piezas fabricadas con el proceso antiguo s_2 = 5.8 mm. ¿Puede rechazarse la hipótesis nula a un nivel de significancia del 5%?

Resp. No

MÉTODOS ALTERNATIVOS PARA PROBAR HIPÓTESIS NULAS

11.38 Al utilizar el método del valor P, pruebe la hipótesis nula del problema 11.24 a un nivel de significancia del 5%.

Resp. Aceptar H_0 (P= 0.0836).

11.39 Al utilizar el método del valor P, pruebe la hipótesis nula del problema 11.25 a un nivel de significancia del 5%.

Resp. Rechazar H_0 (P= 0.0418).

11.40 Aplique el método del Intervalo de confianza para probar la hipótesis nula del problema 11.24, utilizando un nivel de significancia del 5%.

Resp. Aceptar H_0.

11.41 Aplique el método del intervalo de confianza para probar la hipótesis nula del problema 11.25 utilizando un nivel de significancia del 5%.

Resp. Rechazar H_0.

RESULTADOS POR COMPUTADORA

11.42 Con referencia a los datos de la Tabla 9.2, en los que se utiliza un paquete de computación para construir el intervalo de confianza del 95% de la diferencia entre las medias del problema 9.32, y utilizando algún programa de computación disponible, pruebe la hipótesis nula de que la cantidad promedio de tiempo por artículo no difiere para los dos tipos de llamadas, utilizando un nivel de significancia del 5%.

Resp. Aceptar H_0 (P = 0.42).

11.43 En la Tabla 11.4 se tienen los datos del tiempo que se requiere para realizar un conjunto estándar de trabajos utilizando dos paquetes distintos de computación. En el problema 11.27, se probó la hipótesis nula de que no existe diferencia entre el tiempo promedio que se requiere para los dos paquetes. Utilizando algún paquete de computación, pruebe de nuevo esa hipótesis a un nivel de significancia del 5%.

Resp. Rechazar H_0 ($P = 0.038$)

La prueba de ji-cuadrada

12

12.1 LA PRUEBA DE JI-CUADRADA COMO PROCEDIMIENTO PARA PRUEBA DE HIPÓTESIS.

Los procedimientos que se describen en este capítulo, implican la comparación de frecuencias muestrales clasificadas en categorías definidas de datos, teniendo en todos los casos el patrón esperado de frecuencias que se basan en una hipótesis nula específica. Por ello, los procedimientos son todos de pruebas de hipótesis y en los análisis se utilizan datos de muestras aleatorias.

La distribución de probabilidad X^2 (ji-cuadrada) se describe en las secciones 9.6 y 11.8. La estadística de prueba que se presenta en la sección siguiente se distribuye como el modelo de probabilidad de ji-cuadrada y, como se trata de pruebas de hipótesis, se aplican también en este capítulo las etapas básicas que se describieron en la sección 10.1.

En este capítulo se cubre el uso de la distribución ji-cuadrada para *pruebas de bondad del ajuste, pruebas de la independencia de dos variables* y *pruebas para hipótesis sobre proporciones.* Una de las pruebas de proporciones consiste en probar las diferencias entre varias proporciones, lo cual es una extensión de la prueba para la diferencia de dos proporciones que se describió en la sección 11.7.

12.2 PRUEBAS DE BONDAD DEL AJUSTE

La hipótesis nula en una prueba de bondad del ajuste es una afirmación sobre el patrón esperado de las frecuencias en un conjunto de categorías. El patrón esperado puede ajustarse a la suposición de igualdad de probabilidades y puede, por ello, ser uniforme. O, por otro lado, el patrón esperado puede ajustarse a distribuciones de probabilidad como la binomial, la Poisson o la normal.

EJEMPLO 1. Un distribuidor regional de sistemas de aire acondicionado ha subdividido su región en cuatro territorios. A un posible comprador de una distribuidora se le dice que las instalaciones de equipos se distribuyen de manera aproximadamente igual en los cuatro territorios. El prospecto de comprador toma una muestra aleatoria de 40 instalaciones colocadas el año anterior, de los archivos de la compañía, y encuentra que el número de instalaciones en cada uno de los cuatro territorios son los que se enlistan en el primer renglón de la Tabla 12.1, (en donde f_o significa "frecuencia observada"). Con base en la hipótesis de que las instalaciones están distribuidas en forma equitativa, en el segundo renglón de la tabla 12.1 se presenta la distribución uniforme esperada de las instalaciones (en donde f_e significa "frecuencia esperada").

Tabla 12.1 Número de instalaciones de sistemas de aire acondicionado por territorio

	Territorio				Total
	A	B	C	D	
Número instalado en la muestra f_o	6	12	14	8	40
Número esperado de instalaciones, f_e	10	10	10	10	40

Para aceptar la hipótesis nula, debe ser posible atribuir las diferencias entre las frecuencias observadas y las esperada? a la variabilidad del muestreo y al nivel especificado de significancia. Asi, la estadística de prueba ji-cuadrada se basa en la magnitud de esta diferencia para cada una de las categorías de la distribución de frecuencias. El valor de ji-cuadrada que se utiliza para probar la diferencia entre un patrón de frecuencias observado y otro esperado es:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} \qquad (12.1)$$

En la fórmula *(12.1)* anterior se observa que, si las frecuencias observadas son muy cercanas a las frecuencias esperadas, entonces el valor calculado de la ji-cuadrada estará cercano a 0. Conforme las frecuencias observadas se alejan de las frecuencias esperadas, el valor de ji-cuadrada se vuelve mayor. Por ello, se concluye que las pruebas de ji-cuadrada implican el uso de solamente el extremo superior, con el objeto de determinar si un patrón observado de frecuencias es diferente de un patrón esperado.

EJEMPLO 2. El cálculo de la estadística de prueba ji-cuadrada para el patrón de frecuencias observadas y esperadas de la tabla 12.1 es:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(6-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(14-10)^2}{10} + \frac{(8-10)^2}{10} = \frac{40}{10} = 4.00$$

El valor que se requiere de la estadística de prueba ji-cuadrada para rechazar la hipótesis nula depende del nivel de significancia que se especifique y de los grados de libertad. En pruebas de bondad del ajuste, los grados de libertad gl son iguales al número de categorías menos el número de estimadores de parámetros y menos 1. Los grados de libertad para una prueba de bondad del ajuste con ji-cuadrada son (en donde k - número de categoría de datos y m = número de parámetros estimados con base en la muestra):

$$gl = k - m - 1 \qquad 12.2)$$

Cuando la hipótesis nula consiste en afirmar que las frecuencias tienen una distribución igual, no se incluye ninguna estimación de parámetros y $m = 0$. (En los problemas 12.6 y 12.8 se presentan ejemplos en los que m es mayor que cero. Siempre se incluye la substracción del " 1 ", porque, dado un número total de observaciones, una vez que se han anotado las frecuencias observadas en k-1 categorías de la tabla de frecuencias, la última celda en realidad no tiene "libertad" para variar. Por ejemplo, con las tres primeras categorías de la Tabla 12.1 se observan las frecuencias 6,12 y 14, respectivamente, y se sigue que la cuarta categoría debe tener una frecuencia de 8 con el objeto de poder acumular el tamaño designado de la muestra de $n = 40$.

EJEMPLO 3. Enseguida se presenta un ejemplo completo del procedimiento de prueba de hipótesis para los datos de la Tabla 12.1, probando la hipótesis nula a un nivel de significancia del 5%.

H_0: El número de instalaciones están distribuidas de manera uniforme en los cuatro territorios.
H_1 El número de instalaciones no está distribuido de manera uniforme en los cuatro territorios.

$$gl = k - m - 1 = 4 - 0 - 1 = 3$$

X^2 crítica $(gl = 3, \alpha = 0.05) = 7.81$ (del apéndice 7)

X^2 calculada = 4.00 (del ejemplo 2)

Los valores calculados de la estadística de prueba ji-cuadrada se basan en conteos discretos, en tanto que la distribución ji-cuadrada es una distribución continua. Cuando las frecuencias esperadas f_e de las celdas no son pequeñas, este factor carece de importancia en términos de la medida en la que la distribución de la estadística de prueba puede ser aproximada mediante la distribución ji-cuadrada. *Una regla práctica que se utiliza ampliamente es qué la frecuencia esperada f_e para cada celda, o categoría, debe ser de cuando menos* 5. Las celdas que no satisfacen este criterio deben combinarse con categorías adyacentes, cuando sea posible. Entonces, el número reducido de categorías constituye la base para determinar los grados de libertad *gl* aplicables a la prueba.Véanse los problemas 12.6 y 12.8. También pueden aumentarse las frecuencias esperadas de todas las celdas de una tabla de datos aumentando el tamaño global de la muestra. Compare las frecuencias esperadas de los datos de los problemas 12.1 y 12.3.

Las frecuencias esperadas pueden basarse en cualquier suposición sobre la forma de la distribución de frecuencias de la población. Si la suposición se basa simplemente en el patrón histórico de las frecuencias, entonces, al igual que en el caso de la hipótesis de igualdad de probabilidad, no se incluye ninguna estimación de parámetros y gl = k - m -1= k - 0 -1 = k-1.

EJEMPLO 4. Durante mucho tiempo, un fabricante de aparatos de televisión ha tenido el 40% de sus ventas en aparatos de pantalla pequeña (de menos de 14 pulgadas), 40% de tamaño mediano (de 14 a 19 pulgadas) y el 20% en la categoría de pantalla grande (de 21 pulgadas y más). Para fijar los programas adecuados de producción para el mes siguiente, se toma una muestra aleatoria de 100 ventas durante el periodo y se encuentra que 55 de los aparatos eran pequeños, 35 medianos y 10 grandes. Enseguida, se prueba la hipótesis nula de que el patrón histórico de ventas sigue siendo igual, utilizando el nivel de significancia del 1 %.

H_o : Los porcentajes de compras de aparatos de televisión de pantalla pequeña, mediana y grande son 40%, 40% y 20%, respectivamente.

H_1 : El patrón actual de ventas de televisores es diferente del patrón histórico planteado en Ho.

$$gl = k - m - 1 = 3 - 0 - 1 = 2$$
$$\chi^2 \text{ crítica } (gl = 2, \alpha = 0.01) = 9.21$$

La x^2 calculada (en la Tabla 12.2 se encuentran las frecuencias observadas y esperadas) es:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(55-40)^2}{40} + \frac{(35-40)^2}{40} + \frac{(10-20)^2}{20} = 11.25$$

La estadística ji-cuadrada calculada de 11.25 es mayor que el valor crítico de 9.21. Por ello, se rechaza la hipótesis nula a un nivel de significancia del 1 %. Comparando las frecuencias observadas y esperadas de la Tabla 12.2, se encuentra que el cambio principal consiste en que se venden más aparatos pequeños y menos grandes, con cierta reducción en las ventas de los aparatos de tamaño mediano.

Tabla 12.2 Compras observadas y esperadas de aparatos de televisión, de acuerdo con el tamaño de la pantalla

	Tamaño de la pantalla			Total
	Pequeña	Mediana	Grande	
Frecuencia observada, f_0	55	35	10	100
Patrón histórico, f_e	40	40	20	100

12.3 PRUEBAS PARA LA INDEPENDENCIA DE DOS VARIABLES CATEGÓRICAS (PRUEBAS PARA TABLAS DE CONTINGENCIAS)

En el caso de las pruebas de bondad del ajuste, existe sólo una variable categórica, tal como el tamaño de las pantallas de los televisores y lo que se prueba es el patrón hipotético de las frecuencias, o distribución, de la variable. Las frecuencias observadas pueden enlistarse en un solo renglón, o en una sola columna de categorías. Las "pruebas de independencia" implican dos variables categóricas y lo que se prueba es la suposición de que las dos variables son estadísticamente independientes. La independencia implica que el saber la categoría en la que se clasifica una observación con respecto a una variable, no tiene ningún efecto sobre la probabilidad de caer también en alguna de las diversas categorías de las otras variables. Como se trabaja con dos variables, se anotan las frecuencias observadas en una tabla de clasificación doble o *tabla de contingencia* (véase la sección 5.8). Mediante la expresión $r \times k$ se definen las dimensiones de este tipo de tablas, en donde r indica el número de renglones y k el número de columnas.

EJEMPLO 5. La Tabla 12.3 es una reproducción de la sección 5.8 y es un ejemplo del formato más simple posible de una tabla de contingencias, ya que las dos variables (sexo y edad) tienen sólo dos niveles de clasificación, o categorías.Por ello, se trata de una tabla de contingencias de 2 × 2.

Tabla 12.3 Tabla de contingencia para los clientes de la tienda de aparatos de sonido

Edad	Sexo		Total
	Hombre	Mujer	
Menor de 30	60	50	110
30 y más	80	10	90
Total	140	60	200

Si se rechaza la hipótesis nula de independencia para datos clasificados como los de la Tabla 12.3, es señal de que las dos variables son *dependientes* y que existe una *relación* entre ellas. Por ejemplo, para la Tabla 12.3, esto indicaría que existe una relación entre la edad y el sexo para los clientes de la tienda de aparatos de sonido.

Dada la hipótesis de independencia de las dos variables, la frecuencia esperada correspondiente a cada una de las celdas de la tabla de contingencia debe ser proporcional al total de frecuencias observadas de columna y de renglón. Si f_r es la frecuencia total de un renglón determinado y f_k es la frecuencia total de una columna determinada, entonces una fórmula conveniente para determinar la frecuencia esperada para la celda de la tabla de contingencia que se encuentra en ese renglón y columna es

$$f_e = \frac{f_r f_k}{n} \tag{12.3}$$

La fórmula general para los grados de libertad correspondientes a una prueba de independencia es

$$gl = (r - 1)(k - 1) \tag{12.4}$$

EJEMPLO 6. En la Tabla 12.4 se presentan las frecuencias esperadas para los datos de la Tabla 12.3. Por ejemplo, para la celda del renglón 1 y columna 1, el cálculo de la frecuencia esperada es

$$f_e = \frac{f_r f_k}{n} = \frac{(110)(140)}{200} = \frac{15\,400}{200} = 77$$

En este caso, las tres frecuencias esperadas restantes pueden obtenerse mediante substracción de los totales de renglón y de columna, como alternativa al uso de la fórmula *(12.3).* Ésta es una indicación directa de que existe un grado de libertad para una tabla de contingencia de 2 × 2 y que sólo la frecuencia de una ceida "tiene libertad" para variar.

Tabla 12.4 Tabla de frecuencias esperadas para las frecuencias observadas que se reportan en la Tabla 12.3

Edad	Sexo		Total
	Hombre	Mujer	
Menor de 30	77	33	110
30 y más	63	27	90
Total	140	60	200

La estadística de prueba ji-cuadrada para tablas de contingencia se calcula exactamente de la misma manera que para las pruebas de bondad del ajuste (sección 12.2).

EJEMPLO 7. Enseguida se realiza la prueba de la hipótesis nula de independencia para los datos de la Tabla 12.3, utilizando un nivel de significancia del 1 %.

H_0: El sexo y la edad de los clientes de la tienda es independiente.
H_1. El sexo y la edad son variables dependientes (existe una relación entre las variables sexo y edad).

$$gl = (r-1)(k-1) = (2-1)(2-1) = 1$$
$$\chi^2 \text{ crítica } (gl = 1,\ \alpha = 0.01) = 6.63$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(60-77)^2}{77} + \frac{(50-33)^2}{33} + \frac{(80-63)^2}{63} + \frac{(10-27)^2}{27} = 27.80$$

La estadística de prueba calculada de 27.8 excede el valor crítico de 6.63. Por ello, se rechaza la hipótesis nula de independencia a un nivel de significancia del 1 %. Con referencia a la Tabla 12.3, se observa que es más probable que los clientes de sexo masculino tengan más de 30 años de edad, al tiempo que es más probable que las mujeres tengan menos de 30 años. El resultado de la prueba de ji-cuadrada arroja que no puede pensarse que esa relación observada en la muestra se debe al azar, a un nivel de significancia del 1 %.

12.4 PRUEBAS DE HIPÓTESIS SOBRE PROPORCIONES

Prueba de un valor hipotético de la proporción. Dada una proporción poblacional hipotética y la proporción observada para una muestra aleatoria tomada de esa población, en la sección 11.5 se utilizó la distribución normal de probabilidad como aproximación del proceso binomial, con el objeto de probar ese valor hipotético. Puede probarse matemáticamente que una prueba de dos extremos como esa es equivalente a la prueba ji-cuadrada de bondad del ajuste que implica un renglón de frecuencias y dos categorías (una tabla de 2 × 1). Como la prueba de ji-cuadrada implica analizar las diferencias entre frecuencias observadas y esperadas, sin tomar en cuenta la dirección de las diferencias, no existe procedimiento de prueba ji-cuadrada que sea equivalente para una prueba de un extremo sobre la proporción de una población.

EJEMPLO 8. El gerente de un departamento de personal estima que una proporción de $\pi = 0.40$ de los empleados de una empresa grande participará en un nuevo programa de inversión en acciones. Se entrevista a una muestra aleatoria de $n = 50$ empleados y 10 de ellos manifiestan su intención de participar. Se podría probar el valor hipotético de la proporción poblacional utilizando la distribución normal de probabilidad, tal como se describe en la sección 11.5. Enseguida, se ilustra el uso de la prueba de ji-cuadrada para lograr ese mismo objetivo utilizando un nivel de significancia del 5%:

$$H_0: \pi = 0.40 \qquad H_1: \pi \neq 0.40$$
$$gl = k - m - 1 = 2 - 0 - 1 = 1$$

(Existen dos categorías de frecuencias observadas, tal como se muestra en la Tabla 12.5).

$$\chi^2 \text{ crítica } (gl = 1, \alpha = 0.05) = 3.84$$

La x^2 calculada (en la Tabla 12.5 se muestran las frecuencias observadas y las esperadas) es:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(10-20)^2}{20} + \frac{(40-30)^2}{30} = 8.33$$

Tabla 12.5 Frecuencias observadas y esperadas para el ejemplo 8

	Participación en los programas		
	Sí	No	Total
Número observado en la muestra, f_0	10	40	50
Número esperado en la muestra, f_e	20	30	50

La estadística de prueba calculada de 8.33 excede el valor crítico de 3.84. Por ello se rechaza la hipótesis nula a un nivel de significancia de 5% y se concluye que la proporción de participantes en el programa en toda la empresa no es de 0.40.

Prueba de la diferencia entre dos proporciones. En la sección 11.7 se presenta un procedimiento para probar la diferencia entre dos proporciones, con base en la distribución normal. Puede probarse matemáticamente, que una prueba de dos extremos como esa es equivalente a una prueba ji-cuadrada para tablas de contingencias en la que las frecuencias observadas se condensan en una tabla de 2 × 2. De nueva cuenta, no existe prueba de ji-cuadrada equivalente para una prueba de un extremo con base en la distribución normal.

El procedimiento de muestreo que se utiliza con las pruebas de la diferencia entre dos proporciones consiste en obtener *dos muestras aleatorias,* una para cada una de las dos *(k)* categorías. Esto difiere del uso de una tabla de 2 × 2 para probar la independencia de dos variables (sección 12.3) en que en este caso sólo se obtiene *una* muestra aleatoria para el análisis global.

EJEMPLO 9. En el ejemplo 8 de la sección 11.7 se reporta que en 10 de 50 hogares de una comunidad se estaba viendo un programa especial de televisión, y que 15 de 50 hogares en una segunda comunidad estaban viendo el mismo programa. En ese ejemplo, se probó la hipótesis nula H_0: $(\pi_1 - \pi_2) = 0$ o, de manera equivalente H_0: $\pi_1 = \pi_2$ en el nivel de significancia del 5%. Enseguida se presenta la prueba correspondiente utilizando la estadística de prueba ji-cuadrada.

$$H_0: \pi_1 = \pi_2 \qquad H_1: \pi_1 \neq \pi_1$$
$$gl = (r - 1)(k - 1) = (2 - 1)(2 - 1) = 1$$

(Tal como se muestra en la Tabla 12.6, las frecuencias observadas se condensan en una tabla de 2 × 2.)

Tabla 12.6 Audiencia del programa de televisión en dos comunidades

Número de televidentes	10	15	25
Número de no televidentes	40	35	75
Total	50	50	100

$$\chi^2 \text{ crítica } (gl = 1, \alpha = 0.01) = 6.63$$

La x^2 calculada (en la tabla 12.6 se presentan las frecuencias observadas, en tanto que las frecuencias esperadas que se calcularon con la fórmula *(12.3)* se presentan en la Tabla 12.7) es

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(10-12.5)^2}{12.5} + \frac{(15-12.5)^2}{12.5} + \frac{(40-37.5)^2}{37.5} + \frac{(35-37.5)^2}{37.5} = 1.34$$

Tabla 12.7 Frecuencias esperadas para los datos de la Tabla 12.6

	Comunidades		
	Comunidad 1	Comunidad 2	Total
Número de televidentes	12.5	12.5	25
Número de no televidentes	37.5	37.5	75
Total	50	50	100

La estadística de prueba calculada de 1.34 *no* es mayor que el valor crítico de 6.63. Por ello, no es posible rechazar la hipótesis nula a un nivel de significancia del 1 %, y se concluye que la proporción de televidentes de las dos comunidades no es distinta.

Prueba de la diferencia entre diversas proporciones poblacionales. Siguiendo el método básico del ejemplo 9, puede utilizarse la prueba de ji-cuadrada para probar la diferencia entre diversas *(k)* proporciones poblacionales, utilizando una tabla de 2 × *k* para el análisis de las frecuencias. En este caso, no existe procedimiento matemático equivalente basado en la distribución de probabilidad normal. En este caso, la hipótesis nula afirma que no existe diferencia entre las diversas proporciones poblacionales (o que las diversas proporciones muestrales distintas pudieron haber sido obtenidas al azar de la misma población). El procedimiento de muestreo que se sigue consiste en recolectar varias muestras aleatorias independientes, una para cada una de las *k* categorías de datos.

EJEMPLO 10. Al seguir con el ejemplo 9, suponga que se muestrean los hogares de cuatro comunidades y se investiga el número en los que se estaba viendo el programa especial de televisión. En la Tabla 12.8 se presentan los datos muestrales observados, y en la Tabla 12.9 se presentan las frecuencias esperadas, calculadas con la fórmula *(12.3).* Enseguida se realiza la prueba de la hipótesis nula de que no existen diferencias entre las proporciones poblacionales.

(Nota: El rechazo de la hipótesis nula no indica que todas las igualdades son falsas, sino sólo que cuando menos una es falsa.)

$$gl = (r - 1)(k - 1) = (2 - 1)(4 - 1) = 3$$
$$\chi^2 \text{ crítica } (gl = 3, \alpha = 0.01) = 11.35$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(10-12)^2}{12} + \frac{(15-12)^2}{12} + \frac{(5-12)^2}{12} + \frac{(18-12)^2}{12} + \frac{(40-38)^2}{38} + \frac{(35-38)^2}{38} + \frac{(45-38)^2}{38} + \frac{(32-38)^2}{38}$$
$$= 0.33 + 0.75 + 4.08 + 3.0 + 0.11 + 0.24 + 1.29 + 0.95 = 10.75$$

Tabla 12.8 Audiencia del programa de televisión en cuatro comunidades

	Comunidades				
	1	2	3	4	Total
Número de televidentes	10	15	5	18	48
Número de no televidentes	40	35	45	32	152
Total	50	50	50	50	200

Tabla 12.9 Frecuencias esperadas para los datos de la Tabla 12.8

	Comunidades				
	1	2	3	4	Total
Número de televidentes	12.0	12.0	12.0	12.0	48
Número de no televidentes	38.0	38.0	38.0	38.0	152
Total	50	50	50	50	200

El valor calculado de la estadística ji-cuadrada, 10.75, *no* es mayor que el valor crítico de 11.35. Por ello, las diferencias en las proporciones de televidentes en las cuatro comunidades muestreadas no son lo suficientemente grandes para rechazar la hipótesis nula a un nivel de significancia del 5%.

12.5 RESULTADOS POR COMPUTADORA

Por lo general, los paquetes de computación para análisis estadístico tienen la capacidad de llevar a cabo pruebas de bondad del ajuste y pruebas de tablas de contingencia. Sin embargo, como tanto las frecuencias esperadas como las observadas se obtienen en forma directa de los datos muestrales en las pruebas de tablas de contingencia, estas pruebas requieren sólo la captura de las frecuencias observadas para realizar la prueba. Como se explica en las secciones 12.3 y 12.4, pueden utilizarse pruebas de tablas de contingencia para probar la independencia de dos variables categóricas o para probar hipótesis sobre la igualdad de dos o más proporciones poblacionales. En el problema 12.15 se ilustra el uso de un programa de computación para una prueba de tabla de contingencia.

Problemas resueltos

PRUEBAS DE BONDAD DEL AJUSTE

12.1 Alguien afirma que los clientes de una tienda de pantalones vaqueros son hombres y mujeres, en proporciones iguales. Se observa una muestra aleatoria de 40 clientes y 25 resultan ser hombres y 15 mujeres. Pruebe la hipótesis nula de que el número global de hombres y mujeres que son clientes en esa tienda es igual, aplicando la prueba de ji-cuadrada, y utilizando el nivel de significancia del 5%.

Tabla 12.10 Frecuencias observadas y esperadas para el problema 12.1

	Clientes		
	Hombres	Mujeres	Total
Número en la muestra (f_0)	25	15	40
Número esperado (f_e)	20	20	40

De la Tabla 12.10

H_0: El número de clientes hombres y mujeres es igual.
H_1: El número de clientes hombres y mujeres no es igual.

$$gl = k - m - 1 = 2 - 0 - 1 = 1$$
$$\chi^2 \text{ crítica } (gl = 1, \alpha = 0.05) = 3.84$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(25-20)^2}{20} + \frac{(15-20)^2}{20}$$
$$= \frac{(5)^2}{20} + \frac{(-5)^2}{20} = 2.50$$

La estadística de prueba calculada, 2.50, no es mayor que el valor crítico de 3.84. Por lo tanto, no es posible rechazar la hipótesis nula a un nivel de significancia del 5%.

12.2 Con referencia al problema 12.1, suponga que la afirmación dice, más bien, que los clientes de la tienda son dos terceras partes de hombres y una tercera parte de mujeres. Utilizando los datos observados de la Tabla 12.1, pruebe esta hipótesis utilizando el nivel de significancia del 5%

De la Tabla 12.11

H_0: Existe el doble de clientes hombres que mujeres.
H_1: No existe el doble de clientes hombres que mujeres.

$$gl = k - m - 1 = 2 - 0 - 1 = 1$$
$$\chi^2 \text{ crítica } (gl = 1, \alpha = 0.05) = 3.84$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(25-26.67)^2}{26.67} + \frac{(15-13.33)^2}{13.33}$$
$$= \frac{(-1.67)^2}{26.67} + \frac{(1.67)^2}{13.33} = 0.10 + 0.21 = 0.31$$

Resulta evidente que la estadística calculada de ji-cuadrada, 0.31, no excede el valor crítico de 3.84. Por ello, no se rechaza la hipótesis nula a un nivel de significancia del 5%. El hecho de que no se hayan podido rechazar las hipótesis nulas de los problemas 12.1 y 12.2 ilustra el "beneficio de la duda" que se otorgó a cada una de esas hipótesis. Sin embargo, el tamaño de la muestra también afecta la probabilidad de los resultados (véase el problema 12.3).

Tabla 12.11 Frecuencias esperadas y observadas para el problema 12.2

	Clientes		Total
	Hombres	Mujeres	
Número en la muestra (f_0)	25	15	40
Número esperado (f_e)	26.67	13.33	40

12.3 Para la situación que se describe en el problema 12.1, suponga que se prueba la misma hipótesis nula, pero que se duplican las frecuencias muestrales en cada una de las categorías. Es decir, de una muestra aleatoria de 80 clientes, 50 son hombres y 30 son mujeres. Pruebe la hipótesis nula a un nivel de significancia del 5% y compare la decisión con la que se tomó en el problema 12.1.

De la Tabla 12.12

H_0: El número de clientes hombres y mujeres es igual.
H_1: El número de clientes hombres y mujeres no es igual.

$$gl = k - m - 1 = 2 - 0 - 1 = 1$$
$$\chi^2 \text{ crítica } (gl = 1, \alpha = 0.05) = 3.84$$
$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(50-40)^2}{40} + \frac{(30-40)^2}{40} = 5.00$$

El valor calculado de ji-cuadrada de 5.00 es mayor que el valor crítico de 3.84. Por ello, se rechaza la hipótesis nula, a un nivel de significancia del 5%. Aun cuando los datos muestrales son proporcionales en estos dos problemas, en este caso la decisión es "rechazar H_0" en lugar de "aceptar H_0." Esto demuestra la mayor sensibilidad de las pruebas estadísticas que se realizan con tamaños grandes de muestra.

Tabla 12.12 Frecuencias esperadas y observadas para el problema 12.3

	Clientes		Total
	Hombres	Mujeres	
Número en la muestra (f_0)	50	30	80
Número esperado (f_e)	40	40	80

12.4 Un fabricante de refrigeradores ofrece tres líneas básicas de su producto que pueden describirse, en términos comparativos de su precio, como "bajo", "intermedio" y "alto". Antes de llevar a cabo una campaña de promoción para resaltar las virtudes de los refrigeradores de precio alto, los porcentajes de ventas en las tres categorías eran de 45, 30 y 25, respectivamente. De una muestra aleatoria de 50 refrigeradores que se vendieron después de la promoción, el número de los que se vendieron en las categorías de precio bajo, intermedio y alto fueron 15,15 y 20, respectivamente. Pruebe la hipótesis nula de que el patrón actual de ventas no difiere del patrón histórico, utilizando un nivel de significancia del 5%.

Con referencia a la Tabla 12.13,

H_0: El patrón actual de las frecuencias de ventas sigue el patrón histórico.
H_1: patrón actual de las frecuencias de ventas es diferente del patrón histórico.

$$gl = k - m - 1 = 3 - 0 - 1 = 2$$
$$X^2 \text{ crítica } (gl = 2, \alpha = 0.05) = 5.99$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(15-22.5)^2}{22.5} + \frac{(15-15)^2}{15} + \frac{(20-12.5)^2}{12.5}$$
$$= \frac{(-7.5)^2}{22.5} + \frac{(0)^2}{15} + \frac{(7.5)^2}{12.5} = 7.00$$

El valor calculado de la estadística de prueba de 7.00 es mayor que el valor crítico de 5.99. Por ello, se rechaza la hipótesis nula a un nivel de significancia del 5%. Aunque este rechazo no prueba en sí mismo en qué difieren los patrones de ventas actual e histórico, un repaso de la Tabla 12.13 muestra que se vendieron más refrigeradores de precio alto y menos de precio bajo de lo que se hubiera esperado de acuerdo con el patrón histórico de ventas.

Tabla 12.13 Frecuencias esperadas y observadas para el problema 12.4

	Categorías de precio del refrigerador			
	Bajo	Intermedio	Alto	Total
Número vendido (f_0)	15	15	20	50
Número esperado de ventas (f_e)	22.5	15	12.5	50

12.5 Cualquier distribución de probabilidad puede servir de base para determinar las frecuencias esperadas correspondientes a una prueba de bondad del ajuste (véase la sección 12.2). Suponga que se plantea la hipótesis de que la distribución de descomposturas de maquinaria por hora en una planta de ensamble se ajusta a una distribución de probabilidad Poisson, como se describe en la sección 6.6. Sin embargo, no se identifica la distribución especifica de Poisson utilizando la media de la distribución, X. En la Tabla 12.14 se presenta el número de descomposturas observadas durante 40 horas que se incluyeron en la muestra.

(a) Determine el valor de X que debe utilizarse para probar la hipótesis de que el número de descomposturas de máquinas se ajusta a una distribución de probabilidad Poisson.
(b) Construya la tabla de las frecuencias esperadas con base en la distribución Poisson que se identificó en (a) para una muestra de n = 40 horas

Tabla 12.14 Número observado de descomposturas dé maquinaria durante 40 horas muestreadas y hoja de trabajo para el cálculo del número promedio de descomposturas por hora

Número de descomposturas (X)	Frecuencia observada (f_o)	$f_o(X)$
0	0	0
1	6	6
2	8	16
3	11	33
4	7	28
5	4	20
6	3	18
7	1	7
	$\Sigma f_o = 40$	$\Sigma[f_o(X)] = 128$

(a) $\bar{X} = \dfrac{\Sigma[f_o(X)]}{\Sigma f_o} = \dfrac{128}{40} = 3.2$ descomposturas por hora

Por lo tanto, se fija la media de la distribución Poisson en $\lambda = 3.2$.

(o) Las frecuencias esperadas se determinan por medio de la distribución Poisson en el apéndice 4. Véase la Tabla 12.15.

Tabla 12.15 Cálculo de las frecuencias esperadas para el problema de las descomposturas de maquinaria, de acuerdo con la distribución Poisson, con $\lambda = 3.2$ y $n = 40$

Número de descomposturas (X)	Probabilidad (P)	Frecuencia esperada $f_e(= nP)$
0	0.0408	1.6
1	0.1304	5.2
2	0.2087	8.3
3	0.2226	8.9
4	0.1781	7.1
5	0.1140	4.6
6	0.0608	2.4
7	0.0278	1.1
8	0.0111	0.4
9	0.0040	0.2
10	0.0013	0.1
11	0.0004	0.0
12	0.0001	0.0
13	0.0000	0.0
Total	1.0001	39.9

12.6 Con la información de las Tablas 12.14 y 12.15, pruebe la hipótesis nula de que la distribución de las descomposturas de maquinaría por hora se ajusta a una distribución Poisson a un nivel de significancia del 5%.

H_0: La distribución de las descomposturas observadas en la maquinaria cada hora se ajusta a una variable con distribución Poisson.

H_1: La distribución de las descomposturas de maquinaria no se ajusta a una variable con distribución Poisson.

X^2 crítica: En la Tabla 12.16 se muestran las frecuencias observadas y esperadas que se comparan. Observe que, para satisfacer el requerimiento de que cada f_e sea de cuando menos 5, se tuvieron que combinar varias categorías en cada uno de los extremos de la distribución de frecuencias. Además, se estima un parámetro, λ, con base en la muestra. Por ello, $gl = k - m - 1 = 5 - 1-1 = 3$, y la x^2 crítica ($gl = 3$, $\alpha =$ 0.05) = 7.81.

X^2 calculada. Tal como se indica en la Tabla 12.16, la x^2 calculada = 7.81.

Tabla 12.16 Frecuencias esperadas y observadas para el problema de descomposturas de maquinaria y cálculo de ji-cuadrada

Número de descomposturas	Frecuencia observada (f_0)	Frecuencia esperada (f_e)	$\frac{(f_o - f_e)^2}{f_e}$
0	0 } 6	1.6 } 6.8	0.094
1	6	5.2	
2	8	8.3	0.011
3	11	8.9	0.496
4	7	7.1	0.001
5	4 } 8	4.6 } 8.8	0.073
6	3	2.4	
7	1	1.1	
8	0	0.4	
9	0	0.2	
10	0	0.1	
			$X^2 = 0.675$

Como resulta evidente que la estadística de prueba calculada, 0.675, no excede el valor crítico de 7.81, no puede rechazarse la hipótesis nula de que el número de descomposturas de maquinaria por hora es una variable con distribución Poisson, a un nivel de significancia del 5%.

12.7 Con respecto a los datos muestrales que se presentan en la Tabla 12.5, suponga que, en una planta de ensamble similar, las descomposturas de maquinaria por hora tienen distribución Poisson, con $\lambda = 2.5$. Determine si las descomposturas en esta planta difieren significativamente de las descomposturas en aquélla, utilizando un nivel de significancia del 5%.

En este caso, no se estima ningún parámetro base en la muestra y las frecuencias esperadas se determinan utilizando la distribución Poisson con media - 2.5

Ho-. La distribución observada de las descomposturas de maquinaria por hora se ajusta a una variable con distribución Poisson, con $\lambda = 2.5$.

H_1. La distribución observada de descomposturas de maquinaria no se ajusta a una variable con distribución Poisson y $\lambda = 2.5$.

En la Tabla 12.17 se ilustra el cálculo de las frecuencias esperadas.

Tabla 12.17 Cálculo de las frecuencias esperadas para el problema de las descomposturas de maquinaria, de acuerdo con la distribución Poisson con $\lambda = 2.5$ y $n = 40$

Número de descomposturas (X)	Probabilidad (P)	Frecuencia esperada $f_e\ (= nP)$
0	0.0821	3.3
1	0.2052	8.2
2	0.2565	10.3
3	0.2138	8.6
4	0.1336	5.3
5	0.0668	2.7
6	0.0278	1.1
7	0.0099	0.4
8	0.0031	0.1
9	0.0009	0.0
10	0.0002	0.0
Total	0.9999	40.0

X^2 crítica: En la Tabla 12.18 se muestra la comparación de las frecuencias observadas y esperadas. Con el reducido número de categorías, K = 4, y sin estimar ningún parámetro con base en la muestra, **$gl = K - m - 1 = 4 - 0 - 1 = 3$, y la χ^2 crítica ($gl = 3$, $\alpha = 0.05$) = 7.81.**

X^2 calculada: Tal como se muestra en la Tabla 12.18, la x^2 calculada = 6.85.

Tabla 12.18 Frecuencias observadas y esperadas para el problema de las descomposturas de maquinaria y cálculo del valor de ji-cuadrada

Número de descomposturas (X)	Frecuencia observada (f_0)	Frecuencia esperada (f_e)	$\frac{(f_o - f_e)^2}{f_e}$
0	0 } 6	3.3 } 11.5	2.63
1	6	8.2	
2	8	10.3	0.51
3	11	8.6	0.67
4	7	5.3	
5	4	2.7	
6	3 } 15	1:1 } 9.6	3.04
7	1	0.4	
8	0	0.1	
			$x^2 = 6.85$

La estadística calculada de ji-cuadrada, 6.85, no excede el valor critico de 7.81. Por ello, no se rechaza la hipótesis nula de que el número de descomposturas de maquinaria por hora tenga distribución de variable Poisson, con $\lambda = 2.5$, y a un nivel de significancia del 5%. Tal como era de esperarse, la estadística de prueba x^2 es mayor en este problema que en el problema 12.6, en donde X se basó en la misma media muestral. Sin embargo, la estadística de prueba sigue cayendo en la región de aceptación de la hipótesis nula.

12.8 En la Tabla 12.19, tomada del problema 2.16, se muestra el número de lesiones por millar de horas-hombre en una muestra de 50 empresas tomadas de una rama industrial determinada. La media para esta distribución se calculó en el problema 3.17 y es $\bar{X}=2.32$; la desviación estándar muestral es s = 0.42, según se determinó en el problema 4.23. Pruebe la hipótesis nula de que el patrón de frecuencias de la población se distribuye en forma normal, utilizando el nivel de significancia del 5%.

Número promedio de lesiones por millar de horas-hombre	Número de empresas
1.5-1.7	3
1.8-2.0	12
2.1-2.3	14
2.4-2.6	9
2.7-2.9	7
3.0-3.2	5
	50

H_o: La distribución de frecuencias tiene distribución normal.
H_1: La distribución de frecuencias no sigue una distribución normal.

En la Tabla 12.20 se muestran las frecuencias esperadas, utilizando el apéndice 5 de la distribución normal de probabilidad y utilizando la media y la desviación estándar muestrales como estimadores de los respectivos parámetros de la población. En la Tabla 12.21 se muestran las frecuencias observadas y las esperadas que se comparan.

Tabla 12.20 Cálculo de las frecuencias esperadas para las lesiones industriales en 50 empresas

Número promedio de lesiones por millar de horas-hombre (límites exactos de clase)	Límites exactos de clase en unidades normal estándar (z)*	Probabilidad pertenencia para cada categoría (P)	Frecuencia esperada (=50xP)
1.45-1.75	-2.07 to -1.36	0.09	4.5
1.75-2.05	-1.36 to -0.64	0.17	8.5
2.05-2.35	-0.64 to 0.07	0.27	13.5
2.35-2.65	0.07 to 0.79	0.26	13.0
2.65-2.95	0.79 to 1.50	0.15	7.5
2.95-3.25	1.50 to 2.21	0.07	3.5
		1.01	50.5

*Con base en $\hat{\mu}=\bar{X}=2.32$ y $\hat{\sigma}$ =s- 0.42; por ejemplo, para X = 1.45, $z=(X-\hat{\mu})/\hat{\sigma}$ = (1.45 - 2.32)/ 0.42 = - 2.07. †El primer valor de probabilidad de 0.09 es la proporción de área que se encuentra en el "extremo" total a la izquierda de z =-1.36 y el último valor de probabilidad de 0.07 es la proporción de área que se encuentra en el extremo completo a la derecha de z = 1.50. Este procedimiento es necesario para las clases de los extremos, con el objeto de asignar la totalidad del área bajo la curva normal en la distribución de frecuencias.

Tabla 12.21 Frecuencias observadas y esperadas para os datos de lesiones-industriales y cálculo de x^2

Número promedio de lesiones por millar de horas-hombre	Frecuencia observada (f_0)	Frecuencia esperada (f_e)	$\frac{(f_o - f_e)^2}{f_e}$
1.5-1.7 1.8-2.0	3 12 } 15	4.5 8.5 } 13.0	0.31
2.1-2.3	14	13.5	0.02
2.4-2.6	9	13.0	1.23
2.7-2.9 3.0-3.2	7 5 } 12	7.5 3.5 } 11.0	0.09
			x^2 =1.65

$$gl = k - m - 1 = 4 - 2 - 1 = 1$$
$$X^2 \text{ crítica } (gl = 1, \alpha = 0.05) = 3.84$$
$$X^2 \text{ calculada} = 1.65 \text{ (como se muestra en la Tabla 12.21)}$$

La estadística ji-cuadrada calculada, 1.65, no es mayor que el valor crítico de 3.84. Por ello, no se rechaza la hipótesis nula de que la distribución de frecuencias de la tasa de accidentes sigue una distribución normal, a un nivel de significancia del 5%, y, por ello, se acepta la hipótesis.

PRUEBAS PARA LA INDEPENDENCIA DE DOS VARIABLES (PRUEBAS DE TABLAS DE CONTINGENCIAS)

12.9 En la Tabla 12.22 (una tabla de contingencia tomada del problema 5.21) se presentan las reacciones de los votantes ante un nuevo plan de impuestos sobre bienes raíces, de acuerdo con su afiliación política partidaria. Con esos datos, construya una tabla de frecuencias esperadas suponiendo que no existe relación entre afiliación partidaria y reacción ante el plan fiscal.

Tabla 12.22 Tabla de contingencia para las reacciones de los votantes ante un nuevo plan de impuestos

Afiliación partidista	Reacción			Total
	A favor	Neutral	Se opone	
PAN	120	20	20	160
PRI	50	30'	60	140
Otro	50	10	40	100
Total	220	60	120	400

En la Tabla 12.23 se presentan las frecuencias esperadas por celda, calculadas con la fórmula $f_e = (f_r\ f_k)/n$ (véase la sección 12.3).

Tabla 12.23 Tabla de las frecuencias esperadas para las frecuencias observadas que se reportan en la Tabla 12.22

Afiliación partidista	Reacción			
	A favor	Neutral	Se opone	
PAN	88	24	48	160
PRI	77	21	42	140
Otro	55	15	30	100
Total	220	60	120	400

12.10 Con referencia a las Tablas 12.22 y 12.23, pruebe la hipótesis nula de que no existe relación entre la afiliación partidaria y la reacción de los votantes, utilizando el nivel de significancia del 1 %.

H_0: La afiliación partidaria y la reacción de los votantes son independientes (no existe relación).
H_1: La afiliación partidaria y la reacción de los votantes no son independientes.

$$gl = (r-1)(k-1) = (3-1)(3-1) = 4$$

$$X^2 \text{ crítica } (gl = 4,\ \alpha = 0.01) = 13.28$$

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(120-88)^2}{88} + \frac{(20-24)^2}{24} + \frac{(20-48)^2}{48} + \frac{(50-77)^2}{77}$$

$$+ \frac{(30-21)^2}{21} + \frac{(60-42)^2}{42} + \frac{(50-55)^2}{55} + \frac{(10-15)^2}{15} + \frac{(40-30)^2}{30}$$

$$= 11.64+0.67+16.33+9.47+3.86+7.71+0.45+1.67+3.33-55.13$$

La estadística de prueba calculada, 55.13, claramente excede el valor crítico de 13.28. Por ello, se rechaza la hipótesis nula a un nivel de significancia del 1 %, y se concluye que existe una relación entre la afiliación partidaria y la reacción ante el nuevo plan de impuestos.

12.11 En la Tabla 12.24, se presentan los datos relacionados con la reacción de los estudiantes ante la ampliación de un programa deportivo colegial de acuerdo con la clase a la que pertenecen, en donde "división menor" indica que se trata de un alumno de nuevo ingreso o que se encuentra en el segundo año, y la "división superior" señala que los alumnos se encuentran en el tercero o cuarto año. Pruebe la hipótesis nula de que la posición de clase y la reacción ante el programa deportivo son variables independientes, utilizando el nivel de significancia del 5%.

Tabla 12.24 Reacción de los estudiantes ante el plan deportivo, de acuerdo con su generación

Reacción	Generación		Total
	División inferior	División superior	
A favor	20	19	39
En contra	10	16	26
Total	30	35	65

H_0: La posición de clase y la reacción ante el programa deportivo son independientes.
H_1: La posición de clase y la reacción ante el programa deportivo no son independientes.

$$gl = (r - 1)(k - 1) = (2 - 1)(2 - 1) = 1$$
$$\chi^2 \text{ crítica } (gl = 1, \alpha = 0.05) = 3.84$$

X^2 calculada (las frecuencias esperadas por celda se presentan en la Tabla 12.25).

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(20-18)^2}{18} + \frac{(19-21)^2}{21} + \frac{(10-12)^2}{12} + \frac{(16-14)^2}{14} = 1.03$$

El valor calculado de la estadística de prueba, 1.03, *no* es mayor que el valor crítico de 3.84. Por ello, no puede rechazarse la hipótesis nula a un nivel de significancia del 5%, y se acepta la hipótesis de que las dos variables son independientes.

Tabla 12.25 Tabla de las frecuencias esperadas para las frecuencias observadas que se reportan en la tabla 12.24

Reacción	Generación		Total
	División inferior	División superior	
A favor	18	21	39
En contra	12	14	26
Total	30	35	65

PRUEBAS DE HIPÓTESIS SOBRE PROPORCIONES

12.12 Como el problema 12.1 implica una tabla de frecuencias observadas de 1 × 2, el procedimiento es equivalente a probar una proporción poblacional hipotética, tal como se explicó en la sección 12.4.

(a) Plantee la hipótesis nula como una proporción hipotética e interprete el resultado de la prueba que se llevó a cabo en el problema 12.1, desde este punto de vista.
(b) Pruebe la proporción hipotética utilizando la distribución normal de probabilidad, y demuestre que el resultado es equivalente a la misma prueba realizada con la distribución ji-cuadrada.

(a) H_0: La proporción de clientes hombres $\pi = 0.50$; H_1: $\pi \neq 0.50$

$$\chi^2 \text{ crítica } (gl = 1, \alpha = 0.05) = 3.84$$
$$\chi^2 = 2.50 \text{ (del problema 12.1)}$$

El valor calculado de ji-cuadrada, 2.50, no es mayor que el valor critico de 3.84. Por ello, no puede rechazarse la hipótesis de que $\pi = 0.50$, al nivel de significancia del 5%.

(b) H_0: $\pi = 0.50$ $\quad H_1$: $\pi \neq 0.50$

$$z \text{ crítica } (\alpha = 0.05) = \pm 1.96$$

Utilizando la fórmula *(11.12),*

$$\sigma_{\hat{p}} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{(0.50)(0.50)}{40}} = \sqrt{\frac{0.25}{40}} = \sqrt{0.00625} = 0.079$$

De la fórmula (11.*14),*

$$z = \frac{\hat{p}-\pi_0}{\sigma_{\hat{p}}} = \frac{0.375-0.50}{0.079} = \frac{-0.125}{0.079} = -1.58$$

El valor z calculado, -1.58, no se encuentra en ninguna de las reglones de rechazo de la hipótesis nula. Por ello, no puede rechazarse la hipótesis nula a un nivel de significancia del 5%, y se concluye que $\pi = 0.50$.

12.13 Como el problema 12.11 implica una tabla de frecuencias observadas de 2 × 2, el procedimiento resulta equivalente a probar la diferencia entre dos proporciones muestrales. Plantee las hipótesis nula y alternativa desde este punto de vista e interprete la prueba que se realizó en el problema 12.11. En este caso, se supone que se tienen dos muestras aleatorias independientes, una para cada condición de clase de los estudiantes.

$$H_0\colon\ \pi_1 = \pi_2 \qquad H_1\colon\ \pi_1 \neq \pi_2$$

en donde π_1 = proporción de estudiantes de la división inferior que están a favor del programa deportivo.
π_2 = proporción de estudiantes de la división superior que están a favor del programa deportivo.

χ^2 crítica (gl = 1, α = 0.05) = 3.84
χ^2 = 1.03 (del problema 12.11)

El valor calculado de la estadística de prueba, 1.03, es menor que el valor crítico de 3.84. Por ello, no puede rechazarse la hipótesis nula a un nivel de significancia del 5%, y se acepta la hipótesis de que la proporción de estudiantes de la división inferior que está a favor del programa deportivo es igual a la proporción de estudiantes de la división superior que también están de acuerdo con el programa deportivo.

12.14 En la Tabla 12.26, se representa una ampliación del estudio que se analizó en los problemas 12.11 y 12.13. Plantee la hipótesis nula desde el punto de vista de una tabla de frecuencias de 2 × *k* que pueda utilizarse para probar la diferencia entre *k* proporciones, utilizando el nivel de significancia del 5%. En este caso se supone que se obtuvo una muestra aleatoria independiente para cada una de las tres posiciones de clase.

Tabla 12.26 Reacción de los estudiantes a la ampliación del programa deportivo, de acuerdo con su generación

Reacción	Generación			Total
	División inferior	División superior	Egresado	
A favor	20	19	15	54
En contra	10	16	35	61
Total	30	35	50	115

$$H_0\colon \pi_1 = \pi_2 = \pi_3 \qquad H_1\colon \text{No todas } \pi_1 = \pi_2 = \pi_3$$

en donde π_1 = proporción de estudiantes de la división inferior que están a favor del programa deportivo.
π_2 = proporción de estudiantes de la división superior que están a favor del programa deportivo.
π_3 = Proporción de estudiantes graduados que están a favor del programa deportivo.

$$gl = (r - 1)(k - 1) = (2 - 1)(3 - 1) = 2$$
$$X^2 \text{ crítica } (gl = 2, \alpha = 0.05)\ 5.99$$

X^2 calculada (las frecuencias esperadas por celda se presentan en la Tabla 12.27):

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(20-14.1)^2}{14.1} + \frac{(19-16.4)^2}{16.4} + \frac{(15-23.5)^2}{23.5} + \frac{(10-15.9)^2}{15.9} + \frac{(16-18.6)^2}{18.6} + \frac{(35-26.5)^2}{26.5} = 11.23$$

El valor calculado de la estadística de prueba ji-cuadrada, 11.23, es mayor que el valor crítico de 5.99. Por ello, se rechaza la hipótesis nula a un nivel de significancia del 5% y se concluye que no todas las tres proporciones poblacionales son iguales.

Tabla 12.27 Tabla de las frecuencias esperadas para las frecuencias observadas de la Tabla 12.26

Reacción	Generación			Total
	División inferior	División superior	Egresado	
A favor	14.1	16.4	23.5	54
En contra	15.9	18.6	26.5	61
Total	30	35	50	115

RESULTADOS POR COMPUTADORA

12.15 Con referencia a los datos de la tabla de contingencia (12.22), utilizando algún paquete de computación, pruebe la hipótesis nula de que no existe relación entre la afiliación partidaria y la reacción de los votantes, utilizando un nivel de significancia del 1 %.

H_0: La afiliación partidaria y la reacción de los votantes son independientes (no existe relación).
H_1: La afiliación partidaria y la reacción de los votantes no son independientes.

Del listado de computadora que aparece en la figura 12-1:

$$gl = 4$$
$$X^2 \text{ calculada } = 55.13$$
$$X^2 \text{ crítica } (gl = 4, \alpha = 0.01) = 13.28 \text{ (del apéndice 7)}$$

Como la estadística de prueba calculada, 55.13, es mayor que el valor crítico de 13.28, se rechaza la hipótesis nula a un nivel de significancia del 1 %, y se concluye que existe una relación entre la afiliación partidaria y la reacción al nuevo plan impositivo. Este resultado es igual al que se obtuvo mediante cálculos manuales en el problema 12.10.

```
MTB >  READ CONTINGENCY-TOBLE  INTO C1-C3
DOTO)  120    20    20
DOTO)   50    30    60
DOTO)   50    10    40
DOTO)  END
MTB >  CHISQUARE Cl--C3
Expected counts are printed below observed counts

              Cl        C2        C3     Total
    1        120        20        20      160
            88.0      24.0      48.0

    2         50        30        60      140
            77.0      21.0      42.0

    3         50        10        40      100
            55.0      15.0      30.0

 Total       220        60       120      400

 ChiSq =   11.64 +    0.67 +   16.33 +
            9.47 +    3.86 +    7.71 +
            0.45 +    1.67 +    3.33 = 55.13
 df = 4
```

Fig. 12-1 Listado de Minitab

Problemas complementarios

PRUEBAS DE BONDAD DEL AJUSTE

12.16 Con los datos de la Tabla 12.28, pruebe la hipótesis nula de que las preferencias de los consumidores para las cuatro marcas de vino son iguales, utilizando el nivel de significancia del 1 %.

Resp. Rechazar H_0

12.17 Al utilizar la Tabla 12.28, pruebe la hipótesis de que los consumidores que prefieren la marca C son tantos como los que prefieren las otras tres marcas combinadas, a un nivel de significancia del 1 %.

Tabla 12.28 Preferencia de grupos de consumidores para cuatro marcas de vino

Marca				Total
A	B	C	D	
30	20	40	10	100

Resp. Aceptar H_0

12.18 En la Tabla 12.29 se reporta la característica de seguridad más importante que prefiere una muestra aleatoria de compradores de automóviles. Pruebe la hipótesis nula de que la población global de compradores de automóviles tiene una distribución similar, en términos de la preferencia por estas características de seguridad, utilizando el nivel de significancia del (a) 5% y (b) 1 %.

Tabla 12.29 Identificación de las características de seguridad más importantes para compradores de automóviles nuevos

Dispositivo de seguridad					
Frenos de disco	Suspensión modificada	Bolsas de aire	Seguros automáticos	Control de crucero	Total
20	10	30	25	15	100

Resp. (a) Rechazar H_0, (b) Aceptar H_0

12.19 En un curso universitario de estadística para negocios, la distribución histórica de las calificaciones NP (no presentado) NA (no acreditado), S (suficiente), B (bien) y MB (muy bien), ha sido del 10, 30, 40 , 10 y 10%, respectivamente. En el grupo de un profesor nuevo se termina el semestre con ocho estudiantes de calificación NP, 17 con NA, 20 con S, 3 con B y 2 con MB. Pruebe la hipótesis nula de que esta muestra difiere en forma significativa del patrón histórico, utilizando el nivel de significancia del 5%.

Resp. Aceptar H_0

12.20 En la tabla 12.30 se presenta el número de transistores que no satisfacen un riguroso requerimiento de calidad, en 20 muestras de n =10 cada una. Pruebe la hipótesis nula de que esa distribución no difiere en forma significativa de la distribución binomial con n =10 y p = 30, utilizando un nivel de significancia del 5%.

Tabla 12.30 Número de transistores defectuosos en 20 muestras de tamaño de n = 10 cada una

Número de defectuosos por muestra	0	1	2	3	4	5	6	7	8	9	10
Número de muestras	0	1	2	4	5	5	2	1	0	0	0

Resp. Rechazar H_0

12.21 Con referencia a la Tabla 2.15 (página 25), se determinó que $\bar{X}$**= 28.95** y s = 2.52 (véanse los problemas 3.40 y 4.44). Pruebe la hipótesis nula de que esa distribución de frecuencias se ajusta a una distribución normal de probabilidad con un nivel de significancia del 5%.

Resp. No puede probarse porque *gl* = 0.

12.22 Para los datos de la Tabla 2.18 (página 27), se encontró que, para esa muestra de datos agrupados, $\bar{X}$ **= 23.3** años y s =3.4 años (véanse los problemas 3.46 y 4.52). Pruebe la hipótesis nula de que esa distribución de frecuencias se ajusta a una distribución normal de probabilidad utilizando el nivel de significancia del 1 %.

Resp. Rechazar H_0

PRUEBAS PARA LA INDEPENDENCIA DE DOS VARIABLES (PRUEBAS PARA TABLAS DE CONTINGENCIA)

12.23 Como extensión del problema 12.18, se clasificaron en forma separada las opiniones de hombres y mujeres (véase la Tabla 12.31). Pruebe la hipótesis de que no existe relación entre el sexo y el dispositivo de seguridad que se prefiere, utilizando el nivel de significancia del 1 %.

Tabla 12.31 Identificación de las características de seguridad más importantes para compradores de automóviles, según su sexo

Entrevistados	Frenos de disco	Suspensión modificada	Bolsas de aire	Seguros automáticos	Control de crucero	Total
Hombres	15	5	20	5	5	50
Mujeres	5	5	10	20	10	50
Total	20	10	30	25	15	100

Resp. Rechazar H_0

12.24 Para estudiar la relación entre la condición de empleo al momento de un préstamo y el hecho de si después el préstamo se vuelve moroso o no, un gerente bancario elige al azar 100 cuentas, y obtiene los resultados que se muestran en la Tabla 12.32. Pruebe la hipótesis nula de que la condición de empleo y la condición del préstamo son variables independientes, utilizando para la prueba un nivel de significancia del 5%.

Tabla 12.32 Condición de empleo y condición de préstamo para una muestra de 100 cuentas

Condición actual del préstamo	Condición de empleo al momento del préstamo		Total
	Con empleo	Desempleado	
Moroso	10	8	18
No moroso	60	22	82
Total	70	30	100

Resp. Aceptar H_0

12.25 El director de una escuela primaria divide a los padres de familia en tres categorías de ingresos, de acuerdo con el rumbo en donde viven y de acuerdo con tres niveles de participación en los programas escolares. Con los datos de la Tabla 12.33, pruebe la hipótesis de que no existe relación entre los ingresos y la participación de los programas escolares, utilizando un nivel de significancia del 5%. Evalúe el significado de los resultados de la prueba.

Resp. Rechazar H_0

Tabla 12.33 Nivel de ingresos y participación en los programas escolares de los padres de estudiantes de una escuela primaria

Participación en los programas	Nivel de ingresos			Total
	Bajo	Medio	Alto	
Nunca	28	48	16	92
Ocasional	22	65	14	101
Regular	17	74	3	94
Total	67	187	33	287

PRUEBAS DE HIPÓTESIS SOBRE PROPORCIONES

12.26 Con referencia a los problemas 12.16 y 12.25, identifique las aplicaciones de la prueba ji-cuadrada que son equivalentes a probar la proporción hipotética de la población. Para cada problema que identifique, plantee la hipótesis nula correspondiente.

Resp. Problema 12.17

12.27 Con referencia a los problemas 12.16 a 12.25, identifique las aplicaciones que son equivalentes a probar la diferencia entre dos proporciones muestrales, identificando las dos muestras aleatorias distintas que se requieren. Para cada problema que identifique, plantee la hipótesis nula correspondiente.

Resp. Problema 12.24

12.28 Con referencia a los problemas 12.16 a 12.25, identifique las aplicaciones que son equivalentes a probar las diferencias entre tres o más proporciones, identificando las *k* muestras aleatorias distintas que se requieren, y plantee la hipótesis nula correspondiente.

Resp. Problema 12.23

RESULTADOS POR COMPUTADORA

12.29 Con los datos de la tabla de contingencia que aparece en la Tabla 12.33, utilice algún paquete de computación para probar la hipótesis nula de que no existe relación entre el nivel de ingresos y la participación en los programas, utilizando un nivel de significancia del 5%.

Resp. Rechazar H_0

Análisis de varianza

13

13.1 FUNDAMENTOS DE LAS PRUEBAS PARA LA DIFERENCIA ENTRE VARIAS MEDIAS

La prueba de ji-cuadrada se utiliza para probar las diferencias entre diversas proporciones, y el análisis de varianza (ANOVA, por sus iniciales en inglés) se utiliza para probar las diferencias entre diversas medias. Una suposición fundamental en la que se basa el análisis de varianza consiste en que las diversas medias muestrales se obtienen a partir de poblaciones con distribución normal y con la misma varianza σ^2. Sin embargo, se ha encontrado que el procedimiento de prueba es bastante insensible a las violaciones de la suposición de normalidad cuando las poblaciones son unimodales y los tamaños de muestra son aproximadamente iguales. Como la hipótesis nula consiste en que las medias poblacionales son iguales, la suposición de igualdad de varianzas *(homogeneidad de la varianza),* también implica que, para propósitos prácticos, la prueba se ocupa de la hipótesis de que las medias provienen de la misma población. Esto es así porque cualquier población distribuida normalmente queda definida por sus dos parámetros, la media y la varianza (o desviación estándar). En la sección 7.2 puede encontrarse una descripción general de la distribución normal de probabilidad. Todos los procedimientos de cálculo que se presentan en este capítulo son para modelos de efectos fijos, en contraste con los modelos de efectos aleatorios. Esta diferencia se explica en la sección 13.6.

Los fundamentos en los que se basa el análisis de varianza fueron desarrollados inicialmente por el estadístico británico Ronald A. Fisher, y la distribución *F* se denomina, en su honor. El razonamiento conceptual es el siguiente:

(1) Calcúlese la media para cada grupo muestral y, después, determínese el error estándar de la media $s_{\bar{X}}$ *con base sólo en las diversas medias muestrales.* En términos de cálculo, esta es la desviación estándar de esos diversos valores promedio.

(2) Ahora, dada la fórmula $s_{\bar{X}} = s/\sqrt{n}$, se sigue que, $s = \sqrt{n}\, s_{\bar{X}}$, y que $s^2 = n s_{\bar{X}}^2$. Por ello, el error estándar de la media calculado en (1) puede utilizarse para estimar la varianza (común) de la población, de la que fueron obtenidas las diversas muestras. A esta estimación de la varianza poblacional se le denomina *cuadro medio entre tratamientos (CMET).* Fisher denominaba a cualquier estimación de la varianza un "cuadrado medio" porque, en términos de cálculo, una varianza es el promedio de las desviaciones con respecto a la media grupa!, elevadas al cuadrado, (véase la sección 4.5).

(3) Calcúlese la varianza de cada grupo muestral por separado y con respecto a la media de cada uno. Después se combinan estas varianzas ponderándolas con el *n* -1 correspondiente a cada muestra. Este procedimiento de ponderación de la varianza es una extensión del que se utiliza para combinar y ponderar dos varianzas muestrales (sección 11.1). La estimación resultante de la varianza poblacional se denomina *cuadrado medio del error (CME)* y se basa sólo en las diferencias *intergrupales.* De nueva cuenta, se le denomina "cuadrado medio" porque es una estimación de varianza. Se le denomina "error" porque las desviaciones dentro de cada uno de los pocos grupos muestrales pueden deberse solamente a errores de muestreo aleatorio, y no pueden deberse a diferencias entre las medias de los diferentes grupos de la población.

(4) Si la hipótesis nula de que $\mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$ es cierta, entonces se sigue que cada uno de los dos cuadrados medios que se obtienen en (2) y (3) son un estimador insesgado e independiente de la misma varianza poblacional σ^2. Sin embargo, si la hipótesis nula es falsa, entonces el valor esperado del *CMET* es mayor que el *CME.* En

esencia, cualesquiera diferencias entre las medias poblacionales incrementarían el *CMET,* al tiempo que no tendrían ningún efecto sobre el *CME,* que se basa sólo en las diferencias *íntergrúpales.*

(5) Con base en la observación de (4), se puede utilizar la distribución *F* para probar la diferencia entre las dos varianzas, como se describe en la sección 11.9. Se trata de una prueba de un extremo, y la forma general de la prueba *F* en análisis de varianza es

$$F_{gl_1,gl_2} = \frac{CMET}{CME}$$

Si el cociente *F* se encuentra en la región de rechazo para el nivel de significancia especificado, entonces se rechaza la hipótesis de que las diversas medias muestrales provienen de la misma población.

En el problema 13.1 se ilustra la aplicación de estas cinco etapas para probar una hipótesis que implica la diferencia entre tres medias.

Aunque esas etapas anteriores son útiles para describir el enfoque conceptual que subyace al análisis de varianza, la extensión de este procedimiento para diseños que son más complejos que la simple comparación de *k* medias muestrales, resulta laborioso. Por esta razón, en las secciones siguientes se describe cada uno de los diseños en términos de modelo lineal que identifica los componentes que influyen sobre la variable aleatoria. También, se presenta una tabla de análisis de varianza que muestra el cálculo de los valores de los cuadrados medios que se requieren, para cada tipo de diseño muestral.

13.2 DISEÑO COMPLETAMENTE ALEATORIZADO DE UN FACTOR (ANOVA CON UN CRITERIO DE CLASIFICACIÓN)

El procedimiento del análisis de varianza con un criterio de clasificación se ocupa de probar la diferencia entre *k* medias muestrales cuando se asignan los elementos en forma aleatoria a cada uno de los diversos grupos de tratamiento. Por ello, la explicación general de la sección 13.1 se refiere al modelo de clasificación con un criterio.

La ecuación lineal, o modelo, que representa este diseño completamente aleatorizado de un factor es

$$X_{ik} = \mu + \alpha_k + \varepsilon_{ik} \qquad (13.2)$$

en donde μ - media global de todas las poblaciones sometidas al tratamiento *k*

α_k - efecto del tratamiento de un grupo particular K de donde el valor se obtuvo por muestreo

ε_{jk} - el error aleatorio asociado al proceso de muestreo (e es la letra griega épsilon)

La Tabla 13.1 es la tabla resumen para el diseño completamente aleatorizado de un factor de análisis de varíanza, incluyendo todas las fórmulas de cálculo. En los problemas 13.2 a 13.5 se ilustra la aplicación de estas fórmulas a datos muestrales. El sistema de símbolos que se utiliza en esta tabla es un tanto distinto del que se empleó en la sección 13.1, debido a la necesidad de un sistema que pueda extenderse lógicamente al análisis de varianza de dos sentidos. Por ello, el *CMET* se convierte en *cuadrado medio entre los A grupos* de *tratamientos (PCA).* Además, debe observarse que la definición de símbolos en el contexto del análisis de varianza no necesariamente es consistente con el uso de esos símbolos en el análisis estadístico general. Por ejemp α_k en *(13.2)* se refiere al efecto sobre un valor muestreado al azar, proveniente del grupo de tratamiento en el que se encuentra ubicado; no tiene nada que ver con el concepto de a en los procedimientos generales de pruebas de hipótesis que se definieron en la sección 10.1 De manera similar, *N* en la Tabla 13.1 designa el tamaño total de la muestra para la totalidad de los grupos de tratamiento juntos, y no el tamaño de la población. Otros símbolos nuevos que se incluyen en la Tabla 13.1 son $T_{k'}$ que representa la suma (total) de los valores en un determinado grupo de tratamiento, y *T,* que representa la suma de los valores muestreados en todos los grupos combinados.

Tabla 13.1 Tabla resumen para el análisis de varianza con un criterio de calificación (no es necesario que los grupos de tratamientos sean iguales)

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados *(SC)*	Cuadrado medio *(CM)*	Cociente *F*
Entre grupos de tratamientos *(A)*	*K-l*	$SCA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N}$	$CMA = \frac{SCA}{K-1}$	$F = \frac{CMA}{CME}$
Error de muestreo (E)	*N-K*	*SCE=SCT-SCA*	$CME = \frac{SCE}{N-K}$	
Total (7)	*N-l*	$SCT = \sum_{i=1}^{n} \sum_{K=1}^{K} X^2 - \frac{T^2}{N}$		

A diferencia de la forma de la hipótesis nula que se describe en la sección 13.1, la forma general de la hipótesis nula en el análisis de varianza hace referencia al componente relevante del modelo lineal. Por ello, para el análisis de varianza con un criterio de clasificación pueden plantearse las hipótesis nula y alternativa de la siguiente manera:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ o, de manera equivalente H_0: $\alpha_k = 0$ para todos los tratamientos (niveles de factor)

H_1: no todas $\mu_1 = \mu_2 = \cdots = \mu_k$ H_1: $\alpha_k \neq 0$ para algunos tratamientos

13.3 ANÁLISIS DE VARIANZA CON DOS CRITERIOS DE CLASIFICACIÓN

El análisis de varianza con dos criterios de clasificación se basa en dos dimensiones de clasificación, o tratamientos. Por ejemplo, al analizar el nivel de aprovechamiento en un programa de capacitación, podría considerarse tanto el efecto de un método de instrucción como el efecto de la escolaridad previa. De manera similar, podría investigarse el rendimiento de gasolina de acuerdo con la categoría del automóvil y de acuerdo con el octanaje de la gasolina. En tablas de datos, a los tratamientos que encabezan las columnas normalmente se les denomina tratamientos A, y a los que encabezan los renglones se les denomina *B.*

La *interacción* en un experimento de dos factores significa que los dos tratamientos no son independientes, y que el efecto de un tratamiento determinado sobre otro difiere según los niveles del otro factor. Por ejemplo, al estudiar el rendimiento de la gasolina en automóviles, un octanaje elevado podría mejorar el rendimiento en cierto tipo de automóviles, pero no en otros. De manera similar, puede diferir la efectividad de diversos métodos de instrucción según los niveles de habilidad de los estudiantes. Para probar la interacción, debe incluirse en cada una de las celdas de una tabla de datos de dos sentidos más de una observación o medición muestreada (es decir, *réplicas).* En la sección 13.4 se presenta el procedimiento analítico apropiado cuando existe sólo una observación por celda, y en el cual no es posible probar la interacción entre los dos factores. El procedimiento analítico se amplia, para incluir la replicación y el análisis del efecto de la interacción en la sección 13.5.

13.4 EL DISEÑO ALEATORIZADO EN BLOQUES (ANOVA CON DOS CRITERIOS DE CLASIFICACIÓN, UNA OBSERVACIÓN POR CELDA)

El modelo de análisis de varianza con dos criterios de clasificación, en el cual existe sólo una observación por celda, se denomina, por lo general, *diseño aleatorizado en bloquea,* porque es el principal uso para el modelo. ¿Qué tal si se extiende la idea de utilizar observaciones apareadas para comparar dos medias muestrales (sección 11.3) al modelo básico de análisis de varianza con un criterio de clasificación, y se asignan aleatoriamente a cada nivel de tratamiento los grupos *áek* elementos *apareados?* En análisis de varianza, a esos grupos asociados se les denomina *bloques,* y como los elementos se asignan en forma aleatoria con base en su pertenencia a los bloques, a ese diseño se le denomina aleatorizado en bloques. En este tipo de diseño, la dimensión de 'bloque" no es una dimensión de tratamiento como tal. El propósito específico de este diseño no es probar el efecto de "bloques". Más bien, al ser posible asignar parte de la variabilidad entre los elementos a logros anteriores, por ejemplo, puede reducirse el *CME,* con lo que la prueba resultante para los *A* tratamientos resulta ser más sensible.

El modelo lineal para el análisis de varianza con dos criterios de clasificación con una observación por celda (sin replicación), es

$$X_{jk} = \mu + \beta_j + \alpha_k + \varepsilon_{jk} \tag{133}$$

en donde μ = la media global, sin importar el tratamiento.
β_j = efecto del tratamiento/ en la dimensión (renglón) *B*.
α_k = efecto del tratamiento *k* en la dimensión (columna) A,
ε_{jk} = error aleatorio asociado al proceso de muestreo.

La Tabla 13.2 presenta el resumen del análisis de varianza con dos criterios de clasificación sin replicación. En comparación con la Tabla 13.1, para un análisis de varianza con un criterio de clasificación, el único símbolo nuevo en esta tabla es T_j^2, que indica que se eleva al cuadrado el total de cada grupo/ (para los tratamientos *B,* o bloques). En los problemas 13.6 y 13.8 se ilustra la aplicación de estas fórmulas.

Tabla 13.2 Tabla resumen para el análisis de varianza con dos criterios de calificación con una observación por celda (diseño aleatorizado en bloques)

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados *(SC)*	Cuadrado medio (CM)	Cociente *F*
Entre grupos de tratamiento *(A)*	*K-l*	$SCA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N}$	$CMA = \frac{SCA}{K-1}$	$F = \frac{CMA}{CME}$
Entre grupos de tratamiento o bloques *(B)*	*J-l*	$SCB = \frac{1}{K}\sum_{j=1}^{J} T_j^2 - \frac{T^2}{N}$	$CMB = \frac{SCB}{J-1}$	$F = \frac{CMB}{CME}$
Error de muestreo (E)	*(J-1)(K-l)*	*SCE = SCT-SCA-SCB*	$CME = \frac{SCE}{(J-1)(K-1)}$	
Total (7)	*N-*	$SCT = \sum_{j=1}^{J} \sum_{k=1}^{K} X^2 - \frac{T^2}{N}$		

13.5 DISEÑO COMPLETAMENTE ALEATORIZADO DE DOS FACTORES (ANOVA CON DOS CRITERIOS DE CLASIFICACIÓN, *n* OBSERVACIONES POR CELDA)

Tal como se explicó en la sección 13.3, cuando se incluye la replicación en un diseño con dos criterios de clasificación, si es posible probar la interacción entre los dos factores. Por ello, cuando se utiliza este tipo de diseño, pueden probarse, con análisis de varianza, tres hipótesis nulas distintas: que no existen efectos por columna (los promedios por columna no difieren en forma significativa), que no existen efectos por renglón (las medias por renglón no difieren en forma significativa) y que no existe interacción entre los dos factores (los dos factores son independientes). Un efecto de interacción significativo indica que el efecto de los tratamientos de un factor varía de acuerdo con los niveles del otro factor. En ese caso, la existencia de efectos por renglón y/o columna puede no ser significativo desde el punto de vista de la aplicación de los resultados.

El modelo lineal para el análisis de varianza con dos criterios de clasificación, con replicación, es

(13.4)

en donde μ = la media global sin importar el tratamiento
β_j = efecto del tratamiento j en la dimensión (renglón) B
α_k = efecto del tratamiento *k* en la dimensión (columna) A
ι_{jk} = efecto de la Interacción entre el tratamiento y (del factor B) y el tratamiento *k* (del factor A) en donde , (es la letra griega "iota")
ε_{ijk} = el error aleatorio asociado al proceso de muestreo

La Tabla 13.3, presenta el resumen para el análisis de varianza con dos criterios de clasificación con replicación. Las fórmulas que se incluyen en la tabla se basan en la suposición de que existe un número igual de observaciones en todas as celdas. En el problema 13.9 se ilustra la aplicación de estas fórmulas.

Tabla 13.3 Tabla resumen para el análisis de varianza con un criterio de calificación (con más de una observación por celda)

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados *(SC)*	Cuadrado medio (CAÍ)	Cociente *F*
Entre grupos de tratamientos *(A)*	*K-1*	$SCA = \sum_{k=1}^{K} \frac{T_k^2}{nJ} - \frac{T^2}{N}$	$SCT = \frac{CME}{K-1}$	$F = \frac{SCE}{CMA}$
Entre grupos de tratamientos o bloques (B)	*J-1*	$SCB = \sum_{j=1}^{J} \frac{T_j^2}{nK} - \frac{T^2}{N}$	$CMA = \frac{SCA}{J-1}$	$F = \frac{CME}{CMB}$
Interacción entre los factores *A* y *B(I)*	*(J-1) (K-1)*	$SCI = \frac{1}{n}\sum_{j=1}^{J} \sum_{k=1}^{K} \left(\sum_{i=1}^{n} X \right)^2 - SCA - SCB - \frac{T^2}{N}$	$CMB = \frac{SCB}{(J-1)(K-1)}$	$F = \frac{CMB}{PCI}$
Error de muestreo (E)	*JK (n-l)*	*SCE = SCT-SCA - SCB - SCI*	$CMT = \frac{SCI}{JK(n-1)}$	
Total (7)	*N-l*	$SCT = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{K} X^2 - \frac{T^2}{N}$		

13.6 CONSIDERACIONES ADICIONALES

Todos los procedimientos de cálculo que se presentan en este capitulo son para modelos de efectos fijos en análisis de varianza. En un *modelo de efectos fijos,* se incluyen en el experimento todos los tratamientos de interés para un factor determinado. Por ejemplo, en el problema 13.1 se supone que los únicos métodos de instrucción que interesan son los tres métodos que se incluyen en el diseño. Por otro lado, en un *modelo de efectos aleatorios* se incluye sólo una muestra aleatoria de todos los posibles tratamientos, para el factor del experimento. Por ejemplo, de diez métodos distintos de instrucción, pueden elegirse tres de ellos al azar. Se requiere un método de cálculo distinto para este último caso, porque la hipótesis nula afirma que no existen diferencias entre los diversos métodos de instrucción en general, y no sólo entre los métodos específicos de instrucción que se incluyen en el experimento. En la mayoría de los experimentos, el modelo de efectos fijos resulta apropiado y, por lo tanto, en este capítulo se presentan sólo este tipo de modelos.

Los conceptos que se muestran en este capítulo pueden ampliarse a más de dos factores. A los diseños que implican tres o más factores se les denomina *diseños factoriales* y, de hecho, la mayor parte de los especialistas en estadística incluyen en esta categoría al análisis de varianza con dos criterios de clasificación. Aunque es posible probar diversos números de hipótesis nulas con el mismo conjunto de datos, utilizando diseños factoriales, la extensión de esos diseños puede conducir a un número extremadamente grande de categorías (celdas) en la tabla, con los correspondientes problemas de muestreo. Debido a estas dificultades, se han elaborado diseños que no requieren la inclusión de todas las combinaciones posibles de los niveles de tratamientos para cada factor. Esos diseños, tales como el de *cuadrados latinos* y el de *bloques incompletos,* son ejemplos de esos desarrollos, y se describen en textos especializados en análisis de varianza.

Sin importar cuál diseño de experimento se use, es común que el rechazo de una hipótesis nula en análisis de varianza no proporcione al analista base alguna para tomar decisiones finales, porque esos rechazos no sirven para identificar las diferencias exactas entre los tratamientos o los niveles de los factores. Por ejemplo, si existe una diferencia significativa en el logro de los estudiantes para tres métodos distintos de instrucción, lo siguiente sería intentar determinar qué pares de métodos difieren entre sí. Se han desarrollado diversos métodos para realizar esas comparaciones por pares que se llevan a cabo junto con el análisis de varianza.

13.7 APLICACIONES EN COMPUTADORA

Existen programas de computación disponibles para todos los diseños de análisis de varianza que se describen en este capítulo, así como también para diseños más compiejos, que salen del alcance de este libro. En los problemas 13.3,13.7 y 13.10, se analiza el uso de programas de computación para el diseño completamente aleatorizado de un factor, para el diseño aleatorizado en bloques y para el diseño completamente aleatorizado de dos factores, respectivamente.

Problemas resueltos

DISEÑO COMPLETAMENTE ALEATORIZADO DE UN FACTOR

13.1 Se asignan en forma aleatoria 15 participantes de un programa técnico a tres tipos distintos de métodos de instrucción, todos los cuales pretenden desarrollar un nivel determinado de habilidad en diseño auxiliado por computadora. En la Tabla 13.4, se presentan las calificaciones del avance al término de la unidad de instrucción, y se presentan también las calificaciones promedio correspondientes. Utilice el procedimiento de análisis de varianza de la sección 13.1 para probar la hipótesis nula de que las tres medias muestrales se obtienen de la misma población, utilizando un nivel de significancia del 5% para la prueba.

Tabla 13.4 Calificaciones de las personas en capacitación con tres métodos de instrucción distintos

Método de instrucción	Calificación de las pruebas					Calificaciones totales	Calificaciones promedio
A_1	86	79	81	70	84	400	80
A_2	90	76	88	82	89	425	85
A_3	82	68	73	71	81	375	75

H_0: $\mu_1 = \mu_2 = \mu_3$ o, de manera equivalente, H_0: $\alpha_k = 0$ Para todos los tratamientos

H_1: no todas $\mu_1 = \mu_2 = \mu_3$ H_1: $\alpha_k \neq 0$ Para algunos tratamientos

(1) La media global de las 15 calificaciones es

$$\bar{X}_T = \frac{\Sigma X}{n} = \frac{1200}{15} = 80.0$$

El error estándar de la media, con base en las tres medias reportadas es

$$s_{\bar{x}} = \sqrt{\frac{\Sigma(\bar{X} - \bar{X}_T)^2}{\text{Núm. de medias} - 1}} = \sqrt{\frac{(80-80)^2 + (85-80)^2 + (75-80)^2}{3-1}} = \sqrt{\frac{50}{2}} = 5.0$$

(2) $CMTR = ns_{\bar{x}}^2 = 5(5.0)^2 = 5(25.0) = 125.0$

(3) De la fórmula general:

$$s^2 = \frac{\Sigma(X - \bar{X})^2}{n-1}$$

la varianza para cada una de las tres muestras es

$$s_1^2 = \frac{(86-80)^2 + (79-80)^2 + (81-80)^2 + (70-80)^2 + (84-80)^2}{5-1} = \frac{154}{4} = 38.5$$

$$s_2^2 = \frac{(90-85)^2 + (76-85)^2 + (88-85)^2 + (82-85)^2 + (89-85)^2}{5-1} = \frac{140}{4} = 35.0$$

$$s_3^2 = \frac{(82-75)^2 + (68-75)^2 + (73-75)^2 + (71-75)^2 + (81-75)^2}{5-1} = \frac{154}{4} = 38.5$$

Entonces

$$\hat{\sigma}^2 \text{ (combinada)} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2}{n_1+n_2+n_3-3} = \frac{(4)(38.5) + 4(35.0) + 4(38.5)}{5+5+5-3} = \frac{448.0}{12} = 37.3$$

Por ello, *CME*- 37.3

(4) Como el *CMET* es mayor que el *CME,* resulta apropiada la prueba de la hipótesis nula.

$$F\text{crítica}(gl = k-l, kn-k, \alpha = 0.05) = F(2,12; \alpha = 0.05) = 3.88$$

(5) $F = CMTR / CME = 125 / 37.3 = 3.35$

Como la estadística calculada *F,* 3.35, no es mayor que el valor critico, 3.88, no es posible rechazar, al nivel de significancia del 5%, la hipótesis nula de que las calificaciones promedio para los tres métodos de instrucción de la población son iguales entre sí.

13.2 Repita el análisis de varianza del problema 13.1 (con los datos de la Tabla 13.14) utilizando el procedimiento general que se describe en la sección 13.2, junto con las fórmulas de la Tabla 13.1.

Las diversas cantidades que se requiere sustituir en las fórmulas de la Tabla 13.1 son:

$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$N = 15$
$T_1 = 400$	$T_2 = 425$	$T_3 = 375$	$T=1\ 200$
$T_1^2 = 160\ 000$	$T_2^2 = 180\ 625$	$T_3^2 = 140\ 625$	$T^2 = 1440000$

$$\frac{T^2}{N} = \frac{1\ 440\ 000}{15} = 96\ 000$$

$$\sum_{i=1}^{n} \sum_{k=1}^{K} X^2 = 86^2 + 79^2 + \cdots + 81^2 = 96.698$$

$$SCT = \sum_{i=1}^{n} \sum_{k=1}^{K} X^2 - \frac{T^2}{N} = 96\ 698 - 96\ 000 = 698$$

$$SCA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N} = \frac{160\ 000}{5} + \frac{180\ 625}{5} + \frac{140\ 625}{5} - 96\ 000 = 250$$

$$SCE = SCI - SCA = 698 - 250 - 448$$

En la Tabla 13.5 se presenta el análisis de varianza (ANOVA) para los datos de la Tabla 13.4. Tal como se esperaba, el cociente *F* es idéntico al que se calculó en el problema 13.1, y con base en 2 y 12 grados de libertad, es inferior al valor critico *F,* 3.88 que se requiere para un nivel de significancia del 5%. Por ello, se concluye que no existe efecto entre los niveles de tratamientos (métodos de instrucción) y, por ello, se concluye también que las diferencias entre las medias no son significativas, a niveles del 5%.

Tabla 13.5 Tabla de ANOVA para el análisis de tres métodos de instrucción (datos de la Tabla 13.4)

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados	Cuadrado medio *(CM)*	Cociente *F*
Entre grupos de tratamientos *(A)*	3 - 1 - 2	250	$\frac{250}{2} = 125$	$\frac{125}{37.33} = 3.35$
Error de muestreo (E)	15 - 3 - 12	448	$\frac{448}{12} = 37.33$	
Total (7)	15 - 1 - 14	698		

```
MTB >  REOD TEST SCORES BY METHOD INTO C1-C3
DATA>   8&     98     8E
DATA>   79     76     68
DATA>   81     88     73
DATA>   70     82     71
DATA>   64     89     81
DATA>  END
MTB >  NOME C1 = '0-1',  C2 = 'A-2',  C3 = '0-3'
MTB >  AOVONEWAY FOR DOTO IN C1-C3

ANALYSIS OF VARIANCE
SOURCE      DF        SS         MS         F
FOCTOR       2      E50.0      125.0      3. 35
ERROR       12      448.0       37.3
TOTOL       14      698.0
                                      INDIVIDUAL 95 PCT CI'S FOR MEON
                                      BOSED ON POOLED STDEV
LEVEL        N       MEON      STDEV  -----+---------+---------+---------+-
0-1          5     80.000      6. E05         (---------*---------)
0-2          5     85.000      5.916              (---------*---------)
0-3          5     75.000      6. E05  (---------*---------)
                                      -----+---------+---------+---------+-
POOLED STDEV =     6. 110                 7E.0      78.0      84.0      90.0
```

Fig. 13-1 Listado de Minitab para el problema 13.3

ANÁLISIS DE VARIANZA CON UN CRITERIO DE CALIFICACIÓN CON GRUPOS DESIGUALES

13.4 En ocasiones, los datos que se utilizan para realizar análisis de varianza con un criterio de calificación, no incluyen grupos de igual tamaño para los diversos niveles de tratamiento. En la Tabla 13.6, se reportan las palabras por minuto que se mecanografiaron en diferentes marcas de máquinas eléctricas de escribir. Las palabras por minuto corresponden a personas asignadas en forma aleatoria, y sin experiencia previa en estos equipos, después de haberles dado cierta capacitación. Pruebe la hipótesis nula de que el promedio de palabras por minuto para cada una de las tres máquinas no es distinto, utilizando un nivel de significación del 5%.

Tabla 13.6 Promedio de palabras por minuto para tres marcas de máquinas de escribir, con base en un periodo de prueba de 15 minutos

Marca de máquina de escribir	Número de palabras por minuto					Total de PPM	Promedio de PPM
A_1	79	83	62	51	77	352	70.4
A_2	74	85	72	—	—	231	77.0
A_3	81	65	79	55	—	280	70.0

H_0: $\mu_1 = \mu_2 = \mu_3$ *o,* de manera equivalente, H_0: $\alpha_k = 0$ para todos los tratamientos (marcas de máquinas de escribir).

H_1. no todas $\mu_1 = \mu_2 = \mu_3$ H_1: $\alpha_k \neq 0$ para algunos tratamientos, Las diversas cantidades que es necesario substituir en las fórmulas comunes del análisis de varianza con un criterio de calificación son:

$n_1 = 5$ $\quad n_2 = 3$ $\quad n_3=4$ $\quad N = 12$

$T_1=352$ $\quad T_2 = 231$ $\quad T_3 = 280$ $\quad T = 863$

$T_1^2 = 123\,904$ $\quad T_2^2 = 53\,361$ $\quad T_3^2 = 78\,400$ $\quad T^2 = 744\,769$

$$\frac{T^2}{N} = \frac{744\,769}{12} = 62\,064.1$$

$$\sum_{i=1}^{n} \sum_{k=1}^{K} X^2 = 79^2 + 83^2 + \cdots + 55^2 = 63\,441$$

$$SCT = \sum_{i=1}^{n} \sum_{k=1}^{K} X^2 - \frac{T^2}{N} = 63\,441 - 62\,064.1 = 1376.9$$

$$SCA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N} = \frac{123\,904}{5} + \frac{53\,361}{3} + \frac{78\,400}{4} - 62\,064.1 = 103.7$$

$$SCE = SCT - SCA = 1\,376.9 - 103.7 = 1\,273.2$$

En la Tabla 13.7 se presenta el análisis de varianza para esta prueba. Con un nivel de significancia del 5%, el valor critico de *F* (*gl* = 2,9) es 4.26 y el cociente calculado *F*, 0.37, se encuentra entonces en la región de aceptación de la hipótesis nula y se concluye, con base en esos resultados muestrales, que no existen diferencias entre las tres marcas de máquinas de escribir, en términos de su velocidad de escritura. De hecho, como el *CMET* es menor que ei *CME,* puede observarse que la variabilidad entre las tres máquinas de escribir es menor que la variabilidad esperada, suponiendo que no existen diferencias entre las marcas de las máquinas.

Tabla 13.7 Tabla de ANOVA para el análisis de la velocidad en mecanografía en tres marcas de máquinas de escribir

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados *(SC)*	Cuadrado medio (CM)	Cociente *F*
Entre grupos de tratamientos *(A)* (marca de máquina de escribir)	3 - 1 - 2	103.7	$\frac{103.7}{2} = 51.8$	$\frac{51.8}{141.5} = 0.37$
Error de muestreo (E)	1 2 - 3 - 9	1273.2	$\frac{1273.2}{9} = 141.5$	
Total (7)	1 2 - 1 - 1 1	1376.9		

RELACIÓN DEL DISEÑO COMPLETAMENTE ALEATORIZADO DE UN FACTOR CON LA PRUEBA *t* PARA PROBAR LA DIFERENCIA ENTRE LAS MEDIAS DE DOS MUESTRAS INDEPENDIENTES

13.5 Puede considerarse que el diseño completamente aleatorizado de un factor es una extensión a *k* grupos de la prueba para la diferencia entre las medias de dos muestras independientes utilizando la distribución *t* de Student. En las dos aplicaciones se requiere la suposición de que las muestras se obtienen de la misma población con distribución normal, para la cual se desconoce la varianza o^2 (véase la sección 11.2). Por ello, se sigue que, cuando se aplica el análisis de varianza con un criterio de calificación a un diseño en el que k - 2, el resultado es directamente equivalente a utilizar la distribución t para probar la diferencia. De hecho, para el análisis en el que $k = 2$, $F = t^2$, tanto en términos de los valores críticos que se requieren para la significancia, como con respecto a los correspondientes valores calculados de la estadística de prueba para los datos muestrales. Para demostrar estas observaciones, se lleva a cabo (a) la prueba f y (b) el análisis de varianza con un criterio de calificación para las diferencias entre las medias de las primeras dos marcas de máquinas de escribir que se reportan en la Tabla 13.6. Utilice el 5% del nivel de significancia para probar la hipótesis nula de que no existe diferencia entre las dos medias.

(a) Al utilizar la prueba f para grupos independientes:

$$H_0: \mu_0 - \mu_2 = 0 \quad (\text{o } \mu_1 = \mu_2) \quad \bar{X}_1 = 70.4 \quad \bar{X}_2 = 77.0$$
$$H_1: \mu_1 - \mu_2 \neq 0 \quad (\text{o } \mu_1 \neq \mu_2) \quad n_1 = 5 \quad n_2 = 3$$

$$t \text{ crítica } (gl = 6, \alpha = 0.05) = \pm 2.447$$

$$s_1^2 = \frac{\Sigma(X - \bar{X}_1)^2}{n_1 - 1} = \frac{(79-70.4)^2 + (83-70.4)^2 + (62-70.4)^2 + (51-70.4)^2 + (77-70.4)^2}{5-1}$$
$$= \frac{723.20}{4} = 180.8$$

$$s_2^2 = \frac{\Sigma(X - \bar{X}_2)^2}{n_2 - 1} = \frac{(74-77.0)^2 + (85-77.0)^2 + (72-77.0)^2}{3-1} = \frac{98.0}{2} = 49.0$$

$$\hat{\sigma}^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(4)180.8 + (2)49.0}{5+3-2} = \frac{821.2}{6} = 136.8667$$

$$\hat{\sigma}_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\hat{\sigma}^2}{n_1} + \frac{\hat{\sigma}^2}{n_2}} = \sqrt{\frac{136.8667}{5} + \frac{136.8667}{3}} = \sqrt{72.9955} = 8.54$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\hat{\sigma}_{\bar{x}_1 - \bar{x}_2}} = \frac{70.4 - 77.0}{8.54} = \frac{-6.6}{8.54} = -0.77$$

El estadístico r calculado, - 0.77, se encuentra en la región de aceptación de la hipótesis nula, y, por lo tanto, no es posible rechazarlo a un nivel de significancia del 5%.

(b) Para los dos niveles de tratamientos (marcas de máquinas de escribir, marcas de vino) y con base en el análisis de varianza con un criterio de calificación:

$$H_0: \mu_1 = \mu_2 \quad (\text{o } \alpha_k = 0) \quad n_1 = 5 \quad n_2 = 3$$
$$H_1: \mu_1 \neq \mu_2 \quad (\text{o } \alpha_k \neq 0) \quad T_1 = 352 \quad T_2 = 231$$
$$T_1^2 = 123\ 904 \quad T_2^2 = 53\ 361$$

$$\frac{T^2}{N} = \frac{339\ 889}{8} = 42\ 486.1$$

$$\sum_{i=1}^{n} \sum_{k=1}^{K} X^2 = 79^2 + 83^2 + \cdots + 72^2 = 43\ 389$$

$$SCT = \sum_{i=1}^{n} \sum_{k=1}^{K} X^2 - \frac{T^2}{N} = 43\ 389 - 42\ 486.1 = 902.9$$

$$SCA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N} = \frac{123.904}{5} + \frac{53\,361}{3} - 42\,486.1 = 81.7$$

$$SCE = SC7 - SCA = 902.9 - 81.7 = 821.2$$

En la Tabla 13.8 se presenta el análisis de varianza para esta prueba. Al nivel de significancia del 5%, el valor crítico de *F* $(gl = 1,6)$ es 5.99. Como la estadística *F,* 0.60, se encuentra en la región de aceptación de la hipótesis nula, se acepta y se concluye que no existe efecto de tratamientos.

Tabla 13.8 Tabla de ANOVA para el análisis de la velocidad en mecanografía en dos marcas de máquinas de escribir

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados *(SC)*	Cuadrado medio (CAM)	Cociente *F*
Entre grupos de tratamientos *(A)* (marca de máquina de escribir)	2 - 1 = 1	81.7	$\frac{81.7}{1} = 81.7$	$\frac{81.7}{136.9} = 0.60$
Error de muestreo *(E)*	8 - 2 = 6	821.2	$\frac{821.2}{6} = 136.9$	
Total (T)	8 - 1 = 7	902.9		

Por ello, se sustenta la observación acerca de la comparabilidad de los dos procedimientos de prueba. En términos de los valores críticos, la $F_{crítica} = 5.99$ y la $t_{critica} \pm 2.44$. Por ello, $(\pm 2.447)^2 = 5.99$. Al igual que para los valores calculados de las estadísticas de prueba, *F y t* .$F = 0.60$ y $t = -0.77$; $(-0.77)^2 = 0.59 \equiv 0.60$, en donde la ligera diferencia se debe simplemente a errores de redondeo.

DISEÑO ALEATORIZADO EN BLOQUES (ANÁLISIS EN DOS SENTIDOS SIN INTERACCIÓN)

13.6 Para los datos de la Tabla 13.4, suponga que, de hecho, se utilizó un diseño aleatorizado en bloques y que se asoció a los entrenandos antes del experimento, asignando a un entrenando de cada uno de los grupos de habilidad (con base en los logros de cursos anteriores) a cada uno de los métodos de instrucción. La Tabla 13.9 es la Tabla 13.4 modificada, porque los valores que se reportan se reorganizaron para reflejar el diseño aleatorizado en bloques. Sin embargo, puede observarse que se incluyen los mismos valores de la Tabla 13.4 en los grupos de tratamientos A, excepto que se les reporta de acuerdo con los grupos de habilidad *B* y, por lo tanto, se les ordenó de manera diferente. Pruebe la hipótesis nula de que no existe diferencia en el desempeño promedio entre los tres métodos de instrucción, utilizando un nivel de significancia del 5%.

Tabla 13.9 Calificaciones de las personas en capacitación con tres métodos de instrucción, de acuerdo con su nivel de habilidad

Nivel de habilidad (bloque)	Método de instrucción			Total (T_j)	Promedio ($\bar{X}_j$)
	A_1	A_2	A_3		
	86	90	82	258	86.0
B_2	84	89	81	254	84.7
B_3	81	88	73	242	80.7
B_4	79	76	68	223	74.3
B_5	70	82	71	223	74.3
Total (T_K)	400	425	375	Gran total T =1200	
Promedio ($\bar{X}_k$)	80.0	85.0	75.0		Gran media $\bar{X}_T = 80.0$

Las diversas cantidades que se requieren para la tabla de análisis de varianza son:

Paraj :

$T_1 = 258$ $T_2 = 254$ $T_3 = 242$ $T_4 = 223$ $T_5 = 223$

$T_1^2 = 66\ 564$ $T_2^2 = 64\ 516$ $T_3^2 = 58\ 564$ $T_4^2 = 49\ 729$ $T_5^2 = 49\ 729$

Para *k:*

$T_1 = 400$ $T_2 = 425$ $T_3 = 375$

$T_1^2 = 160\ 000$ $T_2^2 = 180\ 625$ $T_3^2 = 140\ 625$

$n_1 = 5$ $n_2 = 5$ $n_3 = 5$

Global:

$T = 1\ 200$ $T^2 = 1\ 440\ 000$ $N = 15$

$$\frac{T^2}{N} = \frac{1\ 440\ 000}{15} = 96\ 000$$

$$\sum_{j=1}^{J}\sum_{k=1}^{K} X^2 = 86^2 + 84^2 + \cdots + 71^2 = 96\ 698$$

$$SCT = \sum_{j=1}^{J}\sum_{k=1}^{K} X^2 - \frac{T^2}{N} = 96\ 698 - 96\ 000 = 698$$

$$SCA = \sum_{k=1}^{K} \frac{T_k^2}{n_k} - \frac{T^2}{N} = \frac{160\ 000}{5} + \frac{180\ 625}{5} + \frac{140\ 625}{5} - 96\ 000 = 250$$

$$SCB = \frac{1}{K}\sum_{j=1}^{J} T_j^2 - \frac{T^2}{N} = \frac{1}{3}(66\ 564 + 64\ 516 + 58\ 564 + 49\ 729 + 49\ 729) - 96{,}000 = 367.3$$

$$SCE = SCT - SCA - SCB = 698 - 250 - 367.3 = 80.7$$

En la Tabla 13.10 se presenta el análisis de varianza (ANOVA) para esos datos. Con respecto a la ecuación lineal que representa este modelo, los dos cocientes F de la Tabla 13.10 se refieren a las pruebas de las siguientes hipótesis nula y alternativa:

H_0 $\alpha_k = 0$ para todas las columnas $\qquad$ H_a: $\beta_j = 0$ para tcxios los renglones

H_1: $\alpha_k \neq 0$ para algunas columnas $\qquad$ H_1: $\beta_j \neq 0$ para algunos renglones

Tabla 13.10 Tabla de ANOVA para el análisis de tres métodos de instrucción, de acuerdo en su nivel de habilidad

Fuente de variación	Suma de cuadrados (SC)	Grados de libertad *(gl)*	Cuadrado medio (CM)	Cociente *(F)*
Entre grupos de tratamientos *(A)* (método)	250.0	3 - 1 = 2	$\frac{250.0}{2} = 125.0$	$\frac{125}{10.1} = 12.4$
Entre bloques (b) (nivel de habilidad)	367.3	5 - 1 = 4 ,	$\frac{367.3}{4} = 91.8$	$\frac{91.8}{10.1} = 9.1$
Error de muestreo (E)	80.7	(5-1)(3-1) = 8	$\frac{80.7}{8} = 10.1$	
Total (T)	698.0	15 - 1 = 14		

En términos de las implicaciones prácticas, la primera hipótesis nula se refiere a la prueba de la diferencia entre las medias por columna, que es el propósito básico del análisis. La segunda hipótesis nula se refiere a la prueba de la diferencia entre las medias por renglón, que reflejan la disposición en bloques que se realizó de acuerdo a un nivel de habilidad.

Al utilizar el nivel de significancia del 5%, la razón F que se requiere para rechazar la primera hipótesis nula *(gl* - 2,8) es 4.46, en tanto que la F que se requiere para la segunda hipótesis nula *(gl* =4,8) es 3.84. Por ello, los dos cocientes *F* calculados en la Tabla 13.10 se encuentran en la región de rechazo de la hipótesis nula. Se concluye que existe una diferencia significativa en las calificaciones para los diferentes métodos de instrucción, y que también existe una diferencia significativa en las calificaciones para los diferentes niveles de habilidad. Note que esto es distinto a lo que se obtuvo en el problema 13.2, en donde se consideró que los datos eran tres muestras independientes. El *CME* del presente análisis es sustancialmente menor que el que se encontró en el problema 13.2, porque una buena parte de esa variabilidad pudo asignarse a las diferencias en el nivel de habilidad.

13.7 Utilice algún paquete de computación para realizar el análisis del problema 13.6 (con los datos de la Tabla 13.9) para el diseño aleatorizado en bloques.

El listado de computadora que se presenta en la figura 13-2, corresponde a los resultados que se obtuvieron mediante los cálculos manuales de la Tabla 13.10. Sin embargo, debe observarse que, para este programa especifico, no se reportan los dos cocientes F, sino que es necesario calcularlos con base en los cuadrados medios que se incluyen en el listado. Al igual que en el problema 13.6, se concluye que existe una diferencia entre las medias, de acuerdo a un método de instrucción, y que existe también una diferencia entre las medias según el factor de bloque del nivel de habilidad, ambos a un nivel de significación del 5%.

```
MTB >  READ METHOD  IN Cl, BLOCK  IN C2,  SCORE  IN C3
DATA>    1      1     86
DATA>    1      2     84
DATA>    1      3     81
DATA>    1      4     79
DATA>    1      5     70
DATA>    2      1     90
DATA>    2      2     89
DATA>    2      3     88
DATA>    2      4     76
DATA>    2      5     82
DATA>    3      1     82
DATA>    3      2     81
DATA>    3      3     73
DATA>    3      4     68
DATA>    3      5     71
DATA> END
MTB >  NAME Cl = 'METHOD', C2 = 'BLOCK', C3 = 'SCORE'
MTB >  TWOWAY AOV FOR 'SCORE', DIMENSIONS 'METHOD', 'BLOCK'

ANALYSIS OF VARIANCE   SCORE

SOURCE          DF        SS          MS        F
METHOD           2     250.0       125.0      12.4
BLOCK            4     367.3        91.8     91.8
ERROR            8      80.7        10.1
TOTAL           14     698.0
```

Fig. 13.2 Listado de Minitab para el problema 13.7

RELACIÓN DEL DISEÑO ALEATORIZADO EN BLOQUES CON LA PRUEBA *t* PARA PROBAR LA DIFERENCIA ENTRE DOS MEDIAS UTILIZANDO OBSERVACIONES APAREADAS

13.8 Cuando $k = 2$, el diseño aleatorizado en bloques es equivalente a la prueba t para la diferencia entre las medias de observaciones apareadas y $F = t^2$. Esto es similar al caso del problema 13.5 para dos muestras independientes. Para ilustrar estos puntos, se aplica el análisis de varianza a los datos de la Tabla 13.11. que se tomó del ejemplo 4 del capítulo 11. En dicho ejemplo, la t critica ($gl = 9$, $\alpha = 0.05$) es ± 2.262, la estadística t calculada es + 1.59 y, por ello, no es posible rechazar la hipótesis nula. Realice la prueba a un nivel de significancia del 5%.

Tabla 13.11 Kilometraje obtenido en automóviles con y sin aditivo para la gasolina, para 10 automóviles muestreados

Automóvil	aditivo (por litro)	aditivo (por litro)	(Tj) total	$(\bar{X}_j)$ promedio
1	9.41	9.28	18.69	9.345
2	9.18	9.15	18.33	9.165
3	8.18	8.28	16.46	8.23
4	7.51	7.59	15.10	7.55
5	7.28	7.21	14.49	7.245
6	6.59	6.62	13.21	6.605
7	6.21	6.13	12.33	6.17
8	5.79	5.64	11.44	5.715
9	5.62	5.51	11.13	5.565
10	5.21	5.13	10.33	5.17
Total (T_k)	60.97	70.54	Gran total T=141.52	
Promedio $(\bar{X}_k)$	7.10	7.05		Gran media $\bar{X}_T = 7.0076$

Las diversas cantidades que se requieren para la tabla de análisis de varianza son

Para j:

$T_1 = 18.69$ $T_2 = 18.33$ $T_3 = 16.$ $T_4 = 15.10$ $T_5 = 14.49$

$T_1^2 = 349.3161$ $T_2^2 = 335.9889$ $T_3^2 = 270.9316$ $T_4^2 = 228.01$ $T_5^2 = 209.9601$

$T_6 = 13.21$ $T_7 = 12.34$ $T_8 = 11.43$ $T_9 = 11.13$ $T_{10} = 10.34$

$T_6^2 = 174.5041$ $T_7^2 = 152.2756$ $T_8^2 = 130.6449$ $T_9^2 = 123.8769$ $T_9^2 = 106.9156$

Para K:

$T_1 = 70.98$ $T_2 = 70.54$

$T_1^2 = 5\,038.1604$ $T_2^2 = 4\,975.8916$

$n_1 = 10$ $n_2 = 10$

Global:

T= 141.52 $T^2 = 20027.91$ N- 20

$$\frac{T^2}{N} = \frac{20\,027.91}{20} = 1\,001.3955$$

$$\sum_{j=1}^{j} \sum_{k=1}^{k} X^2 = 9.41^2 + 9.18^2 + \cdots + 5.13^2 = 1\,041.2561$$

$$SCT = \sum_{j=1}^{j} \sum_{k=1}^{k} X^2 - \frac{T^2}{N} = 1\,041.2561 - 1\,001.3955 = 39.86$$

$$SCA = \sum_{k=1}^{k} \frac{T_k^2}{n} - \frac{T^2}{N} = \frac{5\,038.1604}{10} + \frac{4\,975.8916}{10} - 1\,001.3955 = .0097$$

$$SCP = \frac{1}{K} \sum_{j=1} T_j^2 - \frac{T^2}{N} = \frac{1}{2}(349.3161 + \cdots + 106.9156 - 1\,001.39.55 = 39.8164$$

$$SCE = SCT\text{-}SCA - SCB = 39.86 - 0.0097 - 39.8164 = 0.0339$$

En la Tabla 13.12 se presenta el análisis de varianza para los datos de la Tabla 13.11. La hipótesis nula de Interés, que es conceptualmente Igual a la que se probó en el ejemplo 4 del capítulo 11, es, para los grupos de tratamientos H_0: $\alpha_k = 0$ y H_1: $\alpha_k \neq 0$. La F crítica (gl -=1, !α =0.05)= 5.12.

Como el valor calculado de la estadística F es 2.54 y no es mayor que el valor crítico de 5.12, no es posible rechazar la hipótesis de que no existe efecto de tratamientos (de no diferencia entre las medias) al nivel de significancia del 5%.

Note el valor muy grande de la F calculada para la dimensión de bloques. Con referencia a los datos de la Tabla 13.11, este resultado no resulta sorprendente porque, en apariencia, los 10 automóviles se encontraban en diferentes categorías de peso y existen diferencias sustanciales en el kilometraje para las diversas categorías.

Tabla 13.12 Tabla de ANOVA para el análisis de kilometraje de automóviles con y sin aditivo para la gasolina

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados (SC)	Cuadrado medio (CM)	Cociente *F*
Entre grupos de tratamientos *(A)*	2 - 1 = 1	.0097	$\frac{.0097}{1} = .0097$	$\frac{.0097}{0.0038} = 2.55$
Entre bloques *(B)* (diferentes automóviles)	10 - 1 = 9	39.8164	$\frac{39.8164}{9} = 4.424$	$\frac{4.424}{0.0038} = 1164.21$
Error de muestreo (E)	(2 - 1)(10 - 1) = 9	0.0339	$\frac{.0339}{9} = 0.0038$	
Total *(T)*	20 - 1 = 19	39.86		

Tal como se esperaba, el valor crítico de *F* en este problema es Igual al cuadrado del valor crítico que se utilizó en la prueba t: 5.12 = $(\pm\,2.262)^2$. También, excepto por los errores debidos al redondeo, el valor t^2 calculado del ejemplo 4 del capítulo 11, es igual al valor calculado de *F,* 2.54, con $t^2 = (1.59)^2 = 2.53$.

DISEÑO COMPLETAMENTE ALEATORIZADO DE DOS FACTORES

13.9 Se está capacitando a nueve personas en cuatro materias distintas y se les asigna, en forma aleatoria, a tres métodos diferentes de instrucción. A cada método de instrucción se le asignaron tres estudiantes. Con referencia a la Tabla 13.13, pruebe las diversas hipótesis nulas de interés con respecto a ese diseño, con un nivel de significancia del 5%.

Las diversas cantidades que se requieren para la tabla de análisis de varianza son

Para.J:

$T_1 = 717$ $T_2 = 709$ $T_3 = 722$ $T_4 = 732$

$T_1^2 = 514\ 089$ $T_2^2 = 502\ 681$ $T_3^2 = 521\ 284$ $T_4^2 = 535\ 824$

Para k:

$T_1 = 960$ $T_2 = 1\ 020$ $T_3 = 900$

$T_1^2 = 921\ 600$ $T_2^2 = 1\ 040\ 400$ $T_3^2 = 810\ 000$

Tabla 13.13 Calificaciones de las personas en capacitación con tres métodos de instrucción, y para cuatro áreas de estudio

Materia	Método de instrucción			Total	Promedio
	A_1	A_2	A_3	*(Tj)*	$(\bar{X}_j)$
	70 79 72	83 89 78	81 86 79	717	79.7
B_2	77 81 79	77 87 88	74 69 77	709	78.8
B_3	82 78 80	94 83 79	72 79 75	722	80.2
B4	85 90 87	84 90 88	68 71 69	732	81.3
Total (T_k)	960	1020	900	Gran total T=2880	
Promedio $(\bar{X}_k)$	80	85	75		Gran media $\bar{X}_T = 80$

Global:

$T = 2\ 880$ $T^2 = 8\ 294\ 400$ $N = 36$

$$\frac{T^2}{N} = \frac{8\ 294\ 400}{36} = 230\ 400$$

$$\sum_{=1}^{J}\sum_{k=1}^{K}\left(\sum_{i=1}^{n} X\right)^2 = 70 + 79 + 72)^2 + (77 + 81 + 79)^2 + \bullet\bullet\bullet + (68 + 71 + 69)^2 = 694\ 694$$

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K} X^2 = 70^2 + 79^2 + 72^2 + 77^2 + \bullet\bullet\bullet + 69^2 = 232\ 000$$

$$SCT = \sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K} X^2 - \frac{T^2}{N} = 232\ 000\text{-}230\ 400 = 1\ 600$$

$$SCA = \sum_{k=1}^{K} \frac{T_k^2}{nJ} - \frac{T^2}{N} = \frac{921\ 600}{(3)(4)} + \frac{1\ 040\ 400}{(3)(4)} + \frac{810\ 000}{(3)(4)} - 230\ 400 = 600$$

$$SCB = \sum_{j=1}^{J} \frac{T_j^2}{nK} - \frac{T^2}{N} = \frac{514\ 089}{(3)(3)} + \frac{502\ 681}{(3)(3)} + \frac{521\ 284}{(3)(3)} + \frac{535\ 824}{(3)(3)} - 230\ 400 = 30.8$$

$$SCI = \frac{1}{n} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\sum_{i=1}^{n} X \right)^2 - SCA - SCB \quad - \frac{T^2}{N} = \frac{1}{3}(694\ 694) - 600 - 30.8 - 230\ 400 = 533.9$$

$$SCE - SCT - SCA - SCB - SCI = l\ 600.0 - 600.0 - 30.8 - 533.9 - 435.3$$

Tabla 13.14 Tabla de ANOVA para el análisis de los tres métodos de instrucción que se aplicaron en cuatro materias

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados *(SC)*	Cuadrado medio *(CM)*	Cociente *F*
Entre grupos de tratamientos *(A)* (método)	3 - 1 = 2	600.0	$\frac{600.0}{2} = 300.0$	$\frac{300.0}{18.1} = 16.57$
Entre grupos de tratamientos *(B)* (materia)	4 - 1 = 3	30.8	$\frac{30.8}{3} = 10.3$	$\frac{10.3}{18.1} = 0.57$
Interacción entre el método y la materia (I)	(4-1)(3-1) = 6	533.9	$\frac{533.9}{6} = 89.0$	$\frac{89.0}{18.1} = 4.92$
Error de muestreo (E)	(4)(3)(3-1) = 24	435.3	$\frac{435.3}{24} = 18.1$	
Total (T)	36-1=35	1600.0		

En la Tabla 13.14 se presenta el análisis de varianza (ANOVA) para los datos de la Tabla 13.13. Con respecto al modelo lineal para el análisis de varianza con dos criterios de calificación con replicación, en la Tabla 13.14 se presentan los tres cocientes *F* que se refieren a las siguientes hipótesis nulas y alternativas:

$H_0 : \alpha_k = 0$ para todas las columnas; $H_1 : \alpha_k \neq 0$ para algunas columnas
$H_0 : \beta_j = 0$ para todos los renglones; $H_1 : \beta_j \neq 0$ para algunos renglones
$H_0 : \iota_{jk} = 0$ para todas las celdas; $H_1 : \iota\varphi_k \neq 0$ para algunas celdas

Al utilizar el nivel de significancia del 5%, el cociente *F* que se requiere para rechazar la primera hipótesis nula *(gl* = 2,24) es 3.40; para la segunda, la F que se requiere *(gl* = 3,24) es 3.01; y para la tercera, la *F* que se requiere *(gl* = 6,26), es 2.51. Por ello, se concluye que existe una diferencia significativa en las calificaciones para los diferentes métodos de instrucción, que no existe diferencia significativa entre las diversas materias, y que existe una interaccicn significativa entre los dos factores. Esta última conclusión indica que la efectividad de los tres métodos de instrucción varía para las diferentes materias. Por ejemplo, puede observarse en la Tabla 13.13, que para la materia B_1 el método A_1 fue el menos efectivo, en tanto que para la materia B_4, el método A_3 fue el menos efectivo. Sin embargo, al revisar la tabla, se observa que el método A_2 es cuando menos igual a los otros métodos para todas las materias.

Por ello, en este caso no parece que la mejor decisión sea la posibilidad de utilizar diferentes métodos de instrucción para diferentes materias aun cuando existe un importante efecto de interacción.

13.10 Al utilizar algún programa de computación, realice el análisis del problema 13.9 (con los datos de la Tabla 13.13), para el diseño completamente aleatorízado de dos factores.

El listado de computadora que aparece en la figura 13-3, corresponde a los resultados que se obtuvieron mediante los cálculos manuales de la Tabla 13.14. Sin embargo, para este programa especifico no se reportan los tres cocientes *F,* sino que es necesario calcularlos con base en los cuadrados medios que se incluyen en el listado. Los resultados de las pruebas de significancia son los mismos que en el problema 13.9.

Problemas complementarios

DISEÑO COMPLETAMENTE ALEATORIZADO DE UN FACTOR

13.11 En doce tiendas se colocaron cuatro tipos de publicidad, y se asignaron tres de esas tiendas al azar a cada uno de los distintos tipos de publicidad, con el propósito de estudiar el impacto de punto de venta de los carteles. Con referencia a la Tabla 13.15, pruebe

Tabla 13.15 Ventas de productos, de acuerdo con el cartel publicitario que se utilizó

Tipo de cartel	Ventas			Ventas totales	Ventas promedio
A_1	40	44	43	127	42.3
A_2	53	54	59	166	55.3
A_3	48	38	46	132	44.0
A_4	48	61	47	156	52.0

la hipótesis nula de que no existen diferencias entre los valores promedio de ventas para los cuatro tipos de carteles, utilizando el nivel de significancia del 5%.

Resp. La *F* critica (*gl*= 3,8) +4.07. La *F* calculada = 4.53. Por ello, se rechaza la hipótesis nula de que $\alpha_k =$ 0 para todos los niveles de tratamiento.

13.12 Utilice algún programa de computación para realizar el análisis del problema 13.11 (con los datos de la Tabla 13.15) para el diseño completamente aleatorízado de un factor.

Resp. Igual que para el problema 13.11

```
MTB > READ METHOD IN Cl, SUBJECT IN C2, SCORE IN C3
DATA>   1   1   70
DATA>   1   1   79
DATA>   1   1   72
DATA>   1   2   77
DATA>   1   2   81
DATA>   1   2   79
DATA>   1   3   82
DATA>   1   3   78
DATA>   1   3   80
DATA>   1   4   85
DATA>   1   4   90
DATA>   1   4   87
DATA>   2   1   83
DATA>   2   1   89
DATA>   2   1   78
DATA>   2   2   77
DATA>   2   2   87
DATA>   2   2   88
DATA>   2   3   94
DATA>   2   3   83
DATA>   2   3   79
DATA>   2   4   84
DATA>   2   4   90
DATA>   2   4   88
DATA>   3   1   81
DATA>   3   1   86
DATA>   3   1   79
DATA>   3   2   74
DATA>   3   2   69
DATA>   3   2   77
DATA>   3   3   72
DATA>   3   3   79
DATA>   3   3   75
DATA>   3   4   68
DATA>   3   4   71
DATA>   3   4   69
DATA > END
MTB > NAME Cl = 'METHOD', C2 = 'SUBJECT', C3 = 'SCORE'
MTB > TWOWAY AOV FOR 'SCORE', DIMENSIONS 'METHOD', 'SUBJECT'

ANALYSIS OF  VARIANCE  SCORE

SOURCE         DF        SS         MS       F
METHOD          2     600.0      300.0     16.57
SUBJECT         3      30.9       10.3      O.57
INTERACTION     6     533.8       89.0      4.92
ERROR          24     435.3       18.1
TOTAL          35    1600.0
```

Fig. 13.3 Resultados de Minitab para el problema 13.10

EL DISEÑO ALEATORIZADO EN BLOQUES (ANÁLISIS CON DOS CRITERIOS DE CALIFICACIÓN, SIN INTERACCIÓN)

13.13 Tres gerentes de producto evaluaron los diseños elaborados por cuatro diseñadores de automóviles, según se reporta en la Tabla 13.16. Pruebe la hipótesis nula de que las calificaciones promedio dadas a los diseñadores no difieren, utilizando un nivel de significancia del 1 %.

Evaluador	Diseñador				Total (7})	Promedio ($\bar{X}_j$)
	1	2	3	4		
A	87	79	83	**92**	341	85.25
B	83	73	85	89	330	82.50
C	91	85	**90**	**92**	358	89.50
Total *(T_k)*	261	237	**258**	273	Gran total T=**1029**	
Calificación promedio **($\bar{X}_k$)**	87.0	79.0	86.0	**91.0**		Gran media $\bar{X}_T$ = 85.75

Resp. La *F (gl* - 3,6) - 9.78. La F calculada =12.29. Por ello, se rechaza la hipótesis nula de que $\alpha_k = 0$ para todos los efectos de tratamiento (columna).

13.14 Utilice algún programa disponible de computación para realizar el análisis del problema 13.13 (con los datos de la Tabla 13.16) para el diseño aleatorizado en bloques.

Resp. Igual que para el problema 13.13

DISEÑO COMPLETAMENTE ALEATORIZADO DE DOS FACTORES

13.15 La tabla de datos de menor tamaño en la cual puede aplicarse el análisis de varianza con un criterio de calificación con interacción, es una tabla de 2 X 2, con dos observaciones por celda. En la Tabla 13.17, que es de este tipo, se presentan los datos de ventas para un producto de consumo popular, en ocho regiones asignadas al azar. Pruebe el efecto de los dos factores y de la interacción entre ellos, para los niveles de ventas, utilizando el nivel de significancia del 1 %. Comente el significado de los resultados de la prueba.

Resp. Al nivel de significancia del 1 %, los valores críticos de F correspondientes a los efectos de columna, renglón e interacción, son, respectivamente, *F (gl* - 1, 4) = 21.20; F *(gl* = 1, 4) = 21.20; y F *(gl* = 1, 4) = 21.20. Los correspondientes cocientes F calculados son, respectivamente, *F* = 96.00, F = 49.52 y F = 7.66. Por lo tanto, existen efectos significativos por columna y por renglón, pero no existen efectos significativos de interacción.

Tabla 13.17 Ventas semanales, en millones de pesos, con y sin publicidad, y con y sin descuento en los precios

Descuento en el precio	Con publicidad	Sin publicidad	Total (T_j)	Promedio (X_j)
Con descuento	9.8 10.6	6.0 5.3	31.7	7.925
Sin descuento	6.2 7.1	4.3 3.9	21.5	5.375
Total (T_k)	33.7	19.5	Gran total T=53.2	
Promedio ($\bar{X}_k$)	8.425	4.875		Gran media $\bar{X}_T$ = 6.650

13.16 Utilice algún programa de computación para realizar el análisis del problema 13.15 (con los datos de la Tabla 13.17) para el diseño completamente aleatorizado de dos factores.

Resp. Igual que para el problema 13.15, excepto por ligeras diferencias debidas al redondeo.

Análisis de regresión y correlación lineal

14

14.1 OBJETIVOS Y SUPOSICIONES DEL ANÁLISIS DE REGRESIÓN

El principal objetivo del análisis de regresión es estimar el valor de una variable aleatoria (la *variable dependiente)* conociendo el valor de una variable asociada (la *variable independiente).* La *ecuación de regresión* es la fórmula algebraica mediante la cual se estima el valor de la variable dependiente (sección 14.3).

El término de *análisis de regresión simple* indica que se estima el valor de la variable dependiente con base en una variable independiente, en tanto que el *análisis de regresión múltiple* (que se analiza en el capitulo 15) se ocupa de la estimación del valor de la variable dependiente con base en dos o más variables independientes.

Las suposiciones generales en las que se basa el modelo de la regresión que se presenta en este capítulo son: (1) la variable dependiente es una variable aleatoria; (2) las variables dependiente e independiente tienen una relación lineal; y (3) las varianzas de las distribuciones condicionales de la variable dependiente, para diversos valores de la variable independiente, son iguales *(homoscedasticidad).* La primera suposición indica que, aunque puedan controlarse los valores de la variable independiente, los valores de la variable dependiente se deben obtener a través del proceso de muestreo.

Si se utiliza la estimación por intervalos en el análisis de regresión, se requiere una suposición adicional: (4) las distribuciones condicionales de la variable dependiente, para valores diferentes de la variable dependiente, son todas distribuciones normales para la población de valores.

EJEMPLO 1 Un analista desea estimar el tiempo de entrega de refacciones industriales embarcadas por camión. Desea utilizar el tiempo de entrega como variable dependiente y la distancia como variable independiente. Suponga que elige diez embarques recientes de los registros de la compañía, de manera que las distancias por carretera correspondientes están más o menos equitativamente dispersas entre 100 y 1000 kilómetros de distancia, y registra el tiempo de entrega para cada embarque. Como se va a utilizar la distancia por carretera como variable independiente, esa selección de viajes con distancias especificas resulta aceptable. Por otro lado, la variable dependiente (el tiempo de entrega) es una variable aleatoria en su estudio, lo cual se ajusta a los supuestos del análisis de regresión. El que las variables tengan o no una relación lineal, por lo general se determina construyendo un diagrama de dispersión (véase la sección 14.2) o una gráfica de residuales (véase la sección 14.4) Estos diagramas se utilizan también para observar si la dispersión vertical (varianza) es más o menos igual a lo largo de la linea de regresión.

14.2 EL DIAGRAMA DE DISPERSIÓN

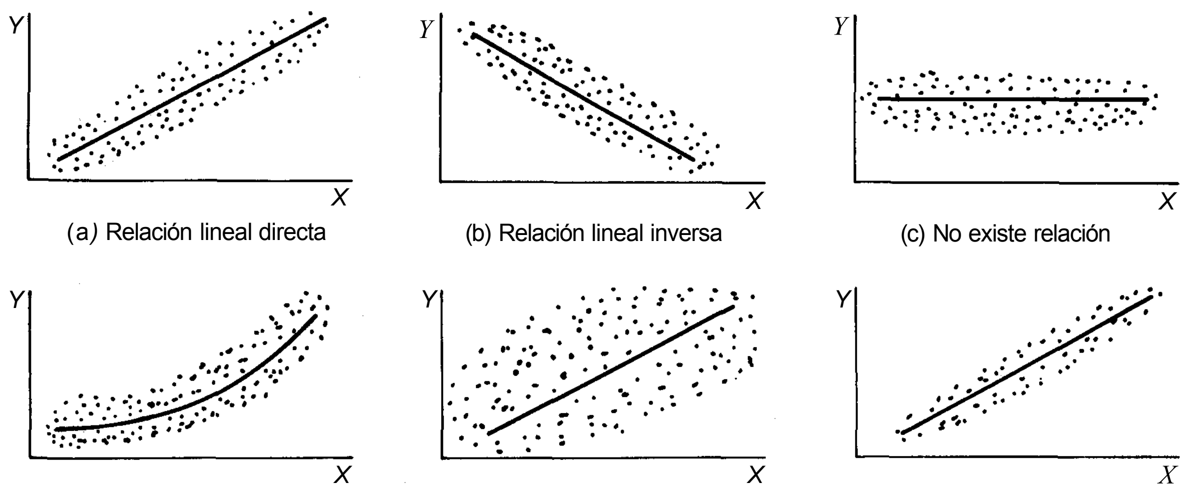
Un *diagrama de dispersión* es una gráfica en la que se traza cada uno de los puntos que representan un par de valores observados para las variables independiente y dependiente. El valor de la variable independiente se gráfica con respecto al eje horizontal, y el valor de la variable dependiente *Y* se traza con respecto al eje vertical (véase el problema 14.1).

La forma de la relación representada mediante el diagrama de dispersión puede ser *curvilínea* y no lineal. Aunque el análisis de regresión para las relaciones curvilíneas está fuera del alcance de este libro, en la sección 16.2 se revisa en forma limitada el análisis de tendencias curvilíneas. Para relaciones que no son lineales, un enfoque utilizado con frecuencia consiste en determinar algún método para transformar los valores de una o ambas variables, de manera que la relación de los valores transformados sí sea lineal. Después, puede aplicarse el análisis de regresión a los valores transformados y pueden transformarse los valores estimados de la variable dependiente, de vuelta a la escala original de medición.

EJEMPLO 2. Un ejemplo de una relación curvilínea, seria la relación entre los años transcurridos desde la constitución de una compañía y su nivel de ventas, en el caso de que las ventas hayan aumentado cada año en el mismo porcentaje sobre el año anterior. La curva resultante, con una pendiente creciente, sería un ejemplo de la denominada relación exponencial.

Si la gráfica de dispersión indica una relación que, en términos generales es lineal, entonces se ajusta una linea recta a esos datos. La ubicación precisa de esa línea se determina mediante el método de mínimos cuadrados (sección 14.3). Tal como se ilustra en el ejemplo 3, una linea de regresión con pendiente positiva indica una relación directa entre las variables; una pendiente negativa indica una relación inversa; y una pendiente de cero indica que las variables no tienen relación entre sí. Además, la medida de la dispersión vertical de los puntos trazados con respecto a la línea de regresión es señal del grado de relación entre las dos variables.

EJEMPLO 3. En la figura 14-1 se incluyen varias gráficas de dispersión y las lineas de regresión asociadas, que ilustran diversos tipos de relación entre variables.



(*a*) Relación lineal directa
(b) Relación lineal inversa
(c) No existe relación
(d) Relación curvilínea directa
(e) Relación lineal directa, con menor grado de relación que en a.
(*f*) Relación lineal directa, con mayor grado de relación que en a.

Fig. 14.1

14.3 EL MÉTODO DE MÍNIMOS CUADRADOS PARA AJUSTAR UNA LÍNEA DE REGRESIÓN

El modelo lineal que representa el modelo de regresión lineal simple es:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{14.1}$$

en donde Y_i = Valor de la variable dependiente en el i-ésimo ensayo u observación.
β_0 = Primer parámetro de la ecuación de regresión, que indica el valor de Y cuando $X = 0$.
β_1 = Segundo parámetro de la ecuación de regresión, que indica la pendiente de la línea de regresión.
X_i = El valor especificado de la variable independiente en el i-ésimo ensayo, u observación.
ε_i = Error aleatorio de muestreo en el i-ésimo ensayo, u observación (ε **es** el griego "épsilon")

Los parámetros β_0 y β_1 del modelo de regresión lineal se estiman mediante los valores b_0 y b_1 con base en datos muestrales. Asi, la ecuación de regresión lineal, con base en datos muestrales, es

$$\hat{Y} = b_0 + b_1 X \qquad (14.2)$$

Al depender del criterio matemático que se utilice, pueden elaborarse diversas ecuaciones lineales distintas para una gráfica de dispersión determinada. De acuerdo *con el criterio de mínimos cuadrados,* la linea de regresión que mejor se ajusta (y su correspondiente ecuación) es aquella para la cual se minimiza la suma de los cuadrados de las desviaciones entre los valores estimados y los valores observados de la variable dependiente. Las fórmulas de cálculo para determinar los valores de b_0 y b_1 de la ecuación de regresión lineal que satisface el criterio de mínimos cuadrados son

$$b_1 = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} \qquad (14.3)$$

$$b_0 = \bar{Y} = b\bar{X} \qquad (14.4)$$

Una vez que se plantea la ecuación de regresión, puede utilizársele para estimar el valor de la variable dependiente para un determinado valor de la variable independiente. Sin embargo, esas estimaciones deben hacerse sólo en el rango de valores dentro de los que se muestreó originalmente la variable independiente, porque no existe base estadística para suponer que la linea de regresión es apropiada fuera de sus limites. Además, debe determinarse si la relación que expresa la ecuación de regresión es real o pudiera haber ocurrido en los datos muestrales debido simplemente al azar (véase la sección 14.6). En el problema 14.2 se ilustra la forma de obtener la ecuación de regresión a partir de datos muestrales.

14.4 RESIDUALES Y GRÁFICAS DE RESIDUALES

Para un valor X dado de la variable independiente, al valor $\hat{Y}$ frecuentemente se le denomina el *valor ajustado* de la variable dependiente. A la diferencia entre el valor observado *Y* y el valor ajustado $\hat{Y}$ se le denomina *residual* para esa observación, y se le denota mediante e:

$$e = Y - \hat{Y} \qquad (14.5)$$

En el problema 14.3 se reporta el conjunto completo de residuales para determinados datos muestrales. Se obtiene una *gráfica de residuales* trazando los residuales e, con respecto a la variable independiente X o, de manera alternativa, con respecto a los valores ajustados de la linea de regresión $\hat{Y}$. Estas gráficas son útiles como alternativa al uso del diagrama de dispersión, para investigar si parecen satisfacerse las suposiciones sobre linealidad e igualdad de varianzas condicionales. En el problema 14.3 se incluye una gráfica de residuales para los datos. Estas gráficas son especialmente importantes en el análisis de regresión múltiple, según se describe en el capitulo 15.

El conjunto de residuales para los datos muestrales sirve también para calcular el error estándar de la estimación, tal como se describe en la siguiente sección.

14.5 EL ERROR ESTÁNDAR DEL ESTIMADOR

El *error estándar del estimador* es la desviación estándar condicional de la variable dependiente *Y,* dado un valor de la variable independiente X. Para datos poblacionales, el error estándar del estimador se representa mediante el símbolo $\sigma_{Y.X}$. La fórmula de desviaciones que permite estimar este valor con base en datos muestrales es

$$s_{Y.X} = \sqrt{\frac{\Sigma(Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{\Sigma e^2}{n-2}} \qquad (14.6)$$

Nótese que el numerador de la fórmula *(14.6)* es la suma de los cuadrados de los residuales que se analizaron en la sección anterior. Aunque la fórmula *(14.6)* refleja en forma clara la idea de que el error estándar del estimador es la desviación estándar con respecto a la línea de regresión (es decir, es la desviación estándar de la "dispersión" vertical con respecto a la linea), en términos de cálculos, la fórmula implica calcular todos los valores ajustados $\hat{Y}$ para los datos muestrales. La fórmula alternativa para cálculos que no requiere determinar cada uno de los valores ajustados y es, por ello y en términos generales, más fácil de utilizar,

$$s_{Y.X} = \sqrt{\frac{\Sigma Y^2 - b_0 \Sigma Y - b_1 \Sigma XY}{n-2}} \qquad (14.7)$$

Tal como se verá en las siguientes secciones, el error estándar del estimador sirve de base para los diversos errores estándar que se utilizan en los procedimientos de prueba de hipótesis y de estimación por intervalos en el análisis de regresión. En los problemas 14.5 y 14.6 se ilustra el cálculo estándar del estimador, utilizando las fórmulas de cálculo y desviaciones, respectivamente.

14.6 INFERENCIAS SOBRE LA PENDIENTE

Antes de utilizar la ecuación de regresión para realizar estimaciones o predicciones, debe determinarse en primer lugar si, de hecho, existe una relación entre las dos variables de la población o, por otro lado, si pudiera ser que la relación que se observa en la muestra haya ocurrido por azar. Si no existe relación en la población, la pendiente de la línea de regresión poblacional seria cero, por definición: $\beta_1 = 0$. Por ello, la hipótesis que generalmente se prueba es H_0: $\beta_1 = 0$. También puede plantearse la hipótesis nula, como prueba con un criterio de calificación, en cuyo caso la hipótesis alternativa no es simplemente que las dos variables están relacionadas, sino que la relación es de algún tipo específico (directa o inversa).

Se prueba el valor hipotético de una pendiente calculando la estadística t y utilizando n - 2 grados de libertad. Se pierden dos grados de libertad en el proceso de la inferencia porque se incluyen en el análisis de regresión *dos* estimaciones de parámetros, b_0 y b_1. La fórmula general es

$$t = \frac{b_1 - (\beta_1)_0}{s_{b_1}} \qquad (14.8)$$

en donde

$$s_{b_1} = \frac{s_{Y.X}}{\sqrt{\Sigma X^2 - n\bar{X}^2}} \qquad (14.9)$$

Sin embargo, cuando la hipótesis nula dice que la pendiente es cero, lo cual generalmente es el caso, se simplifica la fórmula *(14.8)* y se plantea de la siguiente manera:

$$t = \frac{b_1}{s_{b_1}} \qquad (14.10)$$

En los problemas 14.7 y 14.8 se ilustran pruebas para la pendiente con uno y dos criterios de calificación, respectivamente.

El Intervalo de confianza para la pendiente de la población, β_1, se construye de la siguiente manera (y los grados de libertad asociados con la t son n - 2).

$$b_1 \pm t s_{b_1}$$

14.7 INTERVALOS DE CONFIANZA PARA LA MEDIA CONDICIONAL

La estimación por punto de la *media condicional* de la variable dependiente, dado un valor específico de X, es el valor de la línea de regresión $\hat{Y}$. Cuando se utiliza la ecuación de regresión para estimar la media condicional, el símbolo apropiado para representarla es $\hat{\mu}_Y$

$$\hat{\mu}_Y = b_0 + b_1 X \qquad (14.12)$$

Con base en los datos muestrales, el error estándar de la media condicional varía de acuerdo con el valor designado de X, y es

$$s_{\bar{Y}.X} = s_{Y.X}\sqrt{\frac{1}{n}+\frac{(X-\bar{X})^2}{\Sigma X^2-[(\Sigma X)^2/n]}} \qquad (14.13)$$

Dada la estimación puntual de la media condicional y el error estándar de la media condicional, el intervalo de confianza para la media condicional, utilizando n - 2 grados de libertad, es

$$\hat{\mu}_Y \pm t s_{\bar{Y}.X} \qquad (14.14)$$

De nueva cuenta, se tienen n - 2 grados de libertad por los dos estimadores de parámetros b_0 y b_1 que se requieren para la ecuación de regresión. En el problema 14.10 se ilustra la construcción de un intervalo de confianza para la media condicional.

14.8 INTERVALOS DE PREDICCIÓN PARA VALORES INDIVIDUALES DE LA VARIABLE DEPENDIENTE

En contraste con los intervalos de confianza, que son estimaciones de parámetros de poblaciones, en un *intervalo de predicción* se estima un valor *individual* y es, por lo tanto, un intervalo de probabilidad. Podría parecer posible construir un intervalo de predicción utilizando el error estándar del estimador que se define en las fórmulas *(14.6)* y *(14.7).* Sin embargo, ese intervalo estaría incompleto, porque el error estándar del estimador no incluye la ¡ncertidumbre asociada con el hecho de que la línea de regresión basada en datos muestrales incluye también error muestral y, por lo general, no es idéntica a la línea de regresión para la población. El error estándar completo para un intervalo de predicción se denomina *error estándar del pronóstico,* e incluye la ¡ncertidumbre asociada con la "dispersión" vertical con respecto a la línea de regresión *y además,* la ¡ncertidumbre asociada con la posición del valor mismo en la línea de regresión. La fórmula básica para el error estándar del pronóstico es

$$s_{Y(\text{siguiente})} = \sqrt{s^2_{Y.X} + s^2_{\bar{Y}.X}} \qquad (14.15)$$

La versión de cálculo para la fórmula del error estándar del pronóstico es

$$s_{Y(\text{siguiente})} = s_{Y.X}\sqrt{1+\frac{1}{n}+\frac{(X-\bar{X})^2}{\Sigma X^2-[(\Sigma X)^2/n]}} \qquad (14.16)$$

Finalmente, el intervalo de predicción para un valor individual de la variable dependiente, utilizando n - 2 grados de libertad, es

$$\hat{Y} \pm t s_{Y\,(\text{siguiente})} \qquad (14.17)$$

En el problema 14.11 se ilustra el cálculo del error estándar del pronóstico y la construcción de un intervalo de predicción.

14.9 OBJETIVOS Y SUPOSICIONES DEL ANÁLISIS DE CORRELACIÓN

En contraste con el análisis de regresión, en el análisis de correlación se mide el grado de relación entre las variables. Al igual que para el análisis de regresión, en este capítulo sólo se cubre el *análisis de correlación simple,* que se ocupa de la medición de la relación entre sólo una variable independiente y la variable dependiente. En el capítulo 15 se describe el análisis de correlación múltiple.

Las suposiciones sobre la población en las que se basa el análisis de correlación simple son: (1) la relación entre las dos variables es lineal, (2) ambas variables son variables aleatorias, (3) para cada una de las variables, las varianzas condicionales para diferentes valores de la otra variable son iguales (homoscedasticidad), y (4) para cada variable las distribuciones condicionales, dados diferentes valores de la otra variable, son todas ellas distribuciones normales. La última suposición es la de una *distribución normal bivariada.* Debe observarse que estas suposiciones son similares a las que se manejaron para el análisis de regresión excepto que en el análisis de correlación los supuestos se aplican a ambas variables, en tanto que en el análisis de regresión puede fijarse la variable independiente en diversos valores específicos, y no es necesario que sea una variable aleatoria.

EJEMPLO 4 Con referencia al procedimiento de recolección de datos del ejemplo 1, que se refiere a las variables de distancia por carretera y tiempo de entrega para una muestra de 10 embarques recientes de refacciones industriales enviadas por camión, en vez de que el analista escoja los 10 embarques para que estén equitativamente dispersos entre las distancias de 100 y 1000 kilómetros, se eligen esos 10 embarques enteramente al azar, sin tomar en consideración la distancia por carretera o el tiempo de entrega para cada observación. A diferencia del ejemplo 1, en donde sólo el tiempo de entrega es una variable aleatoria, en este plan modificado de muestreo, ambas variables son aleatorias y, por ello, sí son apropiadas para el análisis de correlación.

14.10 EL COEFICIENTE DE DETERMINACIÓN

Considérese, si se estimara un valor individual de la variable independiente Y sin conocer el valor de ninguna otra variable, la incertidumbre asociada con esta estimación, y la base para construir un intervalo de predicción, sería la varianza σ_Y^2. Sin embargo, si se tiene un valor dado *X,* la incertidumbre asociada con la estimación se representa mediante $\sigma_{Y.X}^2$ (o $s_{Y.X}^2$ para los datos muestrales), tal como se describe en la sección 14.5. Si existe relación entre las dos variables, entonces $\sigma_{Y.X}^2$ será menor que σ_Y^2, para una relación perfecta, en la que todos los valores de la variable dependiente son iguales a los valores correspondientes ajustados de la línea de regresión, $\sigma_{Y.X}^2 = 0$. Por ello, si no existe la relación perfecta, el valor de $\sigma_{Y.X}^2$ indica la incertidumbre que permanece *después* de considerar el valor de la variable dependiente. O, puede decirse que el cociente de $\sigma_{Y.X}^2$ con respecto a σ_Y^2 indica la proporción de varianza (incertidumbre) de la variable dependiente que sigue sin explicarse después de haberse dado un valor especifico de la variable dependiente:

$$\frac{\sigma_{Y.X}^2}{\sigma_Y^2} = \frac{\text{varianza no explicada restante en } Y}{\text{varianza total en } Y} \tag{14.18}$$

En la fórmula *(14.18),* se determina $\sigma_{Y.X}^2$ mediante el procedimiento que se describió en la sección 14.5 (excepto que se utilizan los datos de la población) y σ_Y^2 mediante las fórmulas generales que se presentaron en las secciones 4.5 y 4.6.

Dada la proporción de varianza no explicada, una medida útil de la relación es el *coeficiente de determinación;* el complemento del cociente anterior que indica la proporción de la varianza en la variable dependiente que queda estadísticamente *explicada* mediante la ecuación de regresión (es decir, conociendo la variable Independiente asociada, X).

Para datos poblacionales, el coeficiente de determinación se representa mediante la letra griega ρ^2 ("ro cuadrada"), y se calcula de la siguiente manera:

$$\rho^2 = 1 - \frac{\sigma^2_{Y.X}}{\sigma^2_Y} \tag{14.19}$$

Para datos muestrales, puede obtenerse el valor estimado del coeficiente de determinación mediante la fórmula correspondiente:

$$r^2 = 1 - \frac{s^2_{Y.X}}{s^2_Y} \tag{14.20}$$

Para propósitos de cálculo resulta conveniente la siguiente fórmula para el coeficiente de determinación muestral:

$$r^2 = \frac{b_0 \Sigma Y + b_1 \Sigma XY - n\bar{Y}^2}{\Sigma Y^2 - n\bar{Y}^2} \tag{14.21}$$

En el problema 14.12 se ilustra la aplicación de la fórmula *(14.21).* Aunque esta es una fórmula que se utiliza con frecuencia para calcular el coeficiente de determinación de datos muestrales, no incluye ninguna corrección para el sesgo y tiene un ligero sesgo positivo. En la sección siguiente puede verse el factor de corrección que se utiliza.

14.11 EL COEFICIENTE DE CORRELACIÓN

Aunque el coeficiente de determinación r^2 es relativamente fácil de interpretar, no se prueba muy bien en pruebas estadísticas. Sin embargo, la raíz cuadrada del coeficiente de determinación, que se denomina el *coeficiente de correlación r* si se presta para las pruebas estadísticas, porque puede utilizarse para definir una estadística de prueba que tiene distribución í cuando la correlación en la población ρ es igual a 0. El valor del coeficiente puede variar de -1.00 a + 1.00. El signo aritmético asociado con el coeficiente de correlación, que es siempre igual al signo de β_1 de la ecuación de regresión, indica la dirección de la relación entreX y Y (positiva - directa; negativa - inversa). El coeficiente de correlación poblacional, teniendo el mismo signo aritmético que β_1 de la ecuación de regresión es:

$$\rho = \sqrt{\rho^2} \tag{14.22}$$

El coeficiente de correlación muestral es

$$r = \sqrt{r^2} \tag{14.23}$$

En resumen, el signo del coeficiente de correlación indica la dirección de la relación entre las variables X y *Y,* en tanto que el valor absoluto del coeficiente muestra la medida de la relación. El coeficiente de correlación elevado al cuadrado es el coeficiente de determinación e indica la proporción de la varianza de *Y* que queda explicada por el conocimiento de X (y viceversa).

EJEMPLO 5. En la figura 14-2 se ilustra la apariencia general de los diagramas de dispersión asociados con diversos valores de correlación.

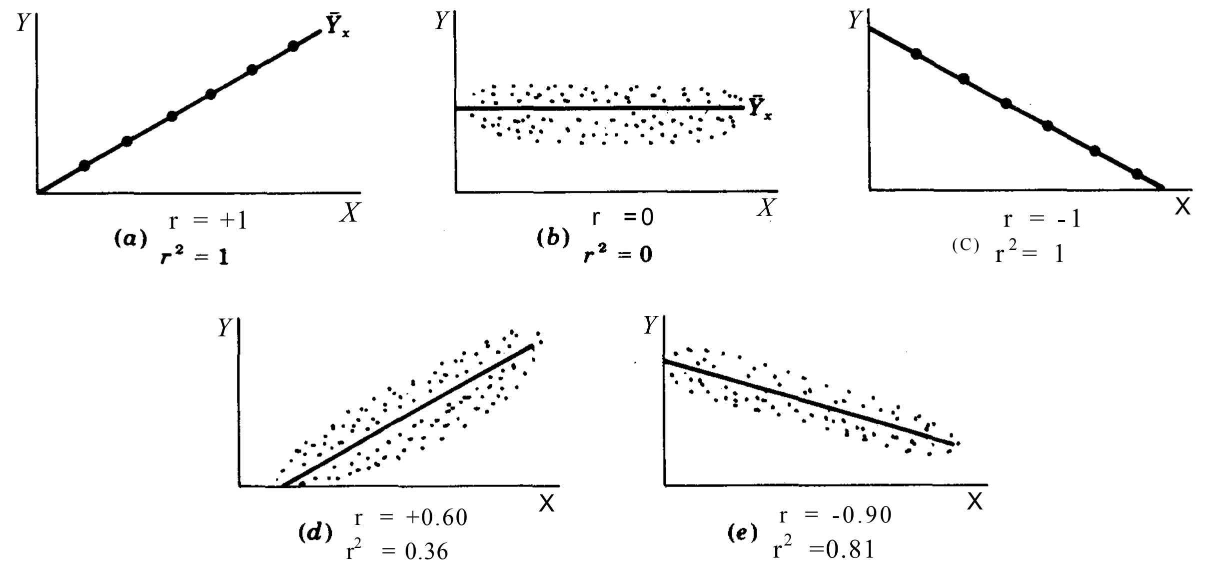


Fig. 14.2

Como una alternativa a la fórmula (*14.21),* la siguiente fórmula no requiere del cálculo previo de los valores de regresión de b_0 y b_1. Se utilizarla esta fórmula cuando el propósito del análisis es determinar el grado y el tipo de la relación entre dos variables, pero sin un interés simultáneo en estimar Y, dada X. Cuando se utiliza esta fórmula se determina en forma automática el signo de coeficiente de correlación, sin necesidad de observar o calcular la pendiente de la línea de regresión. La fórmula es

$$r = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}} \qquad (14.24)$$

En el problema 14.13 se ilustra la aplicación de la fórmula (*14.24).*

El coeficiente de correlación muestral *r* es un estimador un tanto sesgado de p, con un valor absoluto que es demasiado grande. Este factor no se menciona en muchos libros de texto porque la magnitud del sesgo es reducida, excepto para muestras muy pequeñas. Un estimador insesgado para el coeficiente de determinación de la población puede obtenerse de la siguiente manera:

$$\hat{\rho}^2 = 1 - (1 - r^2)\left(\frac{n-1}{n-2}\right) \qquad (14.25)$$

14.12 EL MÉTODO DE LA COVARIANZA PARA COMPRENDER EL COEFICIENTE DE CORRELACIÓN

Otra medida que puede utilizarse para expresar la relación entre dos variables aleatorias es la *covaríanza* muestral. La covarianza mide la medida en la que dos variables "varían juntas" y se utiliza en análisis financieros para determinar el riesgo total asociado con inversiones interrelacionadas. Al igual que para el coeficiente de correlación, un signo positivo indica una relación *directa,* en tanto que un signo negativo indica una relación *inversa, la* fórmula para las covarianzas muestrales es

$$\mathrm{cov}(X, Y) = \frac{\Sigma[(X - \bar{X})(Y - \bar{Y})]}{n-1} \tag{14.26}$$

Al estudiar la fórmula *(14.26)* puede concluirse, por ejemplo, que cuando el valor observado de *Y* tiende a variar en la misma dirección con respecto a su media, que el valor observado de X con respecto a su propia media, entonces los productos de esas desviaciones tienden a ser valores positivos. Por ello, la suma de esos productos sería positiva, indicando una relación directa, y no inversa. Mientras que el coeficiente de correlación puede variar solamente entre -1.00 y + 1.00 y, por lo general, se le considera como una medida de la relación, la covarianza no tiene esos límites de valor y no es una medida generalizada. La fórmula que permite transformar la covarianza en el coeficiente de correlación es

$$r = \frac{\mathrm{cov}(X, Y)}{s_X s_Y} \tag{14.27}$$

En el problema 14.14 se ilustra el cálculo de una covarianza muestral y su transformación en el coeficiente de correlación correspondiente.

14.13 SIGNIFICACIÓN DEL COEFICIENTE DE CORRELACIÓN

Es común que la hipótesis nula de interés sea que la correlación en la población $\rho = 0$, porque si se rechaza esta hipótesis a un nivel especificado α, se concluiría que existe una relación real entre las variables. También puede plantearse la hipótesis como prueba con un criterio de calificación. Considerando que se satisfacen las suposiciones que se mencionaron en la sección 14.9, la siguiente estadística muestral que incluye *a r* se distribuye como la distribución r, con *gl* = *n-2,* cuando p-0:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \tag{14.28}$$

Probar la hipótesis nula de que $\rho = 0$ es equivalente a probar la hipótesis nula de que $\beta = 0$ en la ecuación de regresión (véase el problema 14.15).

14.14 ESCOLLOS Y LIMITACIONES ASOCIADAS CON LOS ANÁLISIS DE REGRESIÓN

(1) No es legítimo estimar un valor de *Y* en análisis de regresión si el valor de X está fuera del rango de valores que sirviera como base para la ecuación de regresión.
(2) Si el estimador de *Y* implica la predicción de un resultado que no ha ocurrido aún, los datos históricos que sirvieron como base para la ecuación de regresión podrían no ser relevantes para eventos futuros.
(3) El uso de un intervalo de confianza de predicción se basa en el supuesto de que las distribuciones condicionales de y son normales y tienen igual varianza.
(4) Un coeficiente de correlación significativo no necesariamente indica causalidad, sino que puede simplemente indicar una asociación común con otros eventos.
(5) Una correlación "significativa" no es necesariamente una correlación importante. Dada una muestra grande, una correlación de, por ejemplo, *r* = + 0.10 puede ser significativamente diferente de 0 al nivel de $\alpha = 0.05$. Aun así, un coeficiente de determinación de $r^2 = 0.01$ para este ejemplo, indica que sólo el 1 % de la varianza de Y queda estadísticamente explicada conociendo X.
(6) La interpretación de los coeficientes de correlación y de determinación se basa en el supuesto de una distribución normal bivariada para la población y, para cada variable, de igualdad de varianzas condicionales.

(7) Tanto para el análisis de regresión como para el de correlación, se supone un modelo lineal. Para una relación curvilínea, es posible que pueda aplicarse una transformación que permita lograr la linealidad. Otra posibilidad consiste en restringir el análisis del rango de valores dentro de los cuales la relación es esencialmente lineal.

14.15 RESULTADOS POR COMPUTADORA

Existen programas de computación para ejecutar los diversos procedimientos de los análisis de regresión y correlación que se cubren en este capítulo. En el problema 14.17 se ilustra el uso de programas y se comparan diversas partes de los listados con los resultados que se obtienen mediante cálculos manuales en los problemas de final del capitulo.

Problemas resueltos

ANÁLISIS DE REGRESIÓN LINEAL

14.1 Suponga que un analista toma una muestra aleatoria de 10 embarques recientemente enviados por camión de una compañía y registra la distancia en kilómetros y el tiempo de entrega, al mediodía más cercano, y a partir del momento en que el embarque estuvo listo para su transportación. Construya la gráfica de dispersión para los datos de la Tabla 14.1 y determine si un análisis de regresión lineal es apropiado.

Tabla 14.1 Observaciones muestrales de distancia de transporte y tiempo de entrega para 10 embarques elegidos al azar

Embarque muestreado	1	2	3	4	5	6	7	8	9	10
Distancia (X), kilómetros	825	215	1070	550	480	920	1350	325	670	1215
Tiempo de entrega *(Y)*, días	3.5	1.0	4.0	2.0	1.0	3.0	4.5	1.5	3.0	5.0

En la figura 14-3 se muestra el diagrama de dispersión para estos datos. El primer par de valores que se reporta en la tabla se representa mediante el punto trazado por encima de 825 del eje X y alineado con el 3.5 del eje Y.

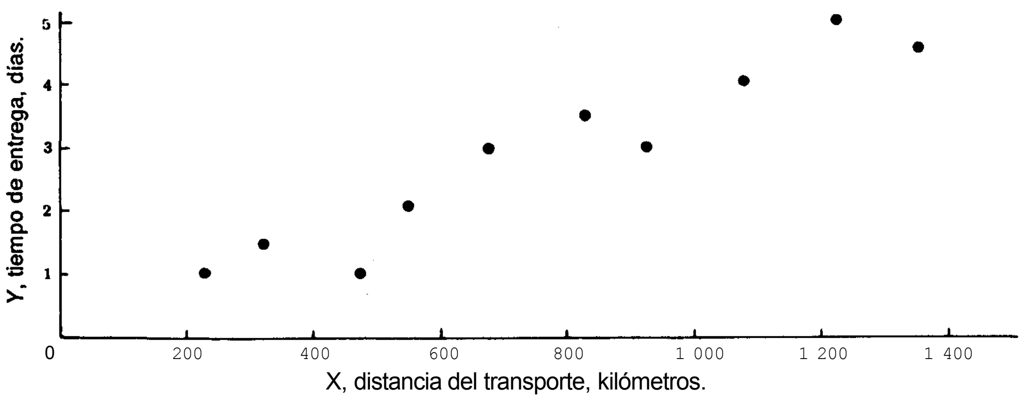


Fig. 14.3

Los otros puntos del diagrama de dispersión se trazaron en forma similar. Observando el diagrama, parece ser que, en términos generales, los puntos trazados siguen una relación lineal, y la dispersión vertical con respecto a la linea es más o menos igual para los valores bajos que para los valores altos de X. Por eso, el análisis de regresión lineal parece ser apropiado.

14.2 Determine la ecuación de regresión por mínimos cuadrados para los datos del problema 14.1, y trace la línea de regresión sobre la gráfica de dispersión para estos datos.

Con referencia a la Tabla 14.2,

$$b_1 = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} = \frac{(26\,370)-(10)(762)(2.85)}{7\,104\,300-(10)(762)^2} = \frac{4\,653}{1\,297\,860} = 0.0035851 \cong 0.0036$$

$$b_0 = \bar{Y} - b\bar{X} = 2.85 - (0.0036)(762) = 0.1068 \cong 0.11$$

Por lo tanto,

$$\hat{Y} = b_0 + b_1 X = 0.11 + 0.0036X$$

Tabla 14.2 Cálculos para la ecuación de regresión lineal, para estimar el tiempo de entrega, con base en la distancia de transporte

Embarque muestreado	Distancia *(X)*, kilómetros	Tiempo de entrega *(Y)*, días	*XY*	X^2	Y^2
1	825	3.5	2 887.5	680 625	12.25
2	215	1.0	215.0	46 225	1.00
3	1070	4.0	4 280.0	1 144 900	16.00
4	550	2.0	1 100.0	302 500	4.00
5	480	1.0	480.0	230 400	1.00
6	920	3.0	2 760.0	846 400	9.00
7	1 350	4.5	6 075.0	1822 500	20.25
8	325	1.5	487.5	105 625	2.25
9	670	3.0	2 010.0	448 900	9.00
10	1 215	5.0	6 075.0	1 476 225	25.00
Totales	7 620	28.5	26 370.0	7 104 300	99.75
Media	$\bar{X} = \frac{\Sigma X}{n} = \frac{7\,620}{10} = 762$	$\bar{Y} = \frac{\Sigma Y}{n} = \frac{28.5}{10} = 2.85$			

Esta línea estimada de regresión, con base en datos muestrales, se traza en el diagrama de dispersión para los datos, y es la que se presenta en la figura 14-4. Deben observarse las lineas punteadas que indican la cantidad de desviación entre cada uno de los valores muestreados de Y y el valor estimado correspondiente, $\hat{Y}$. Esla suma de estas desviaciones al cuadrado lo que se minimiza mediante la linea de regresión, línea que se determinó mediante el procedimiento anterior.

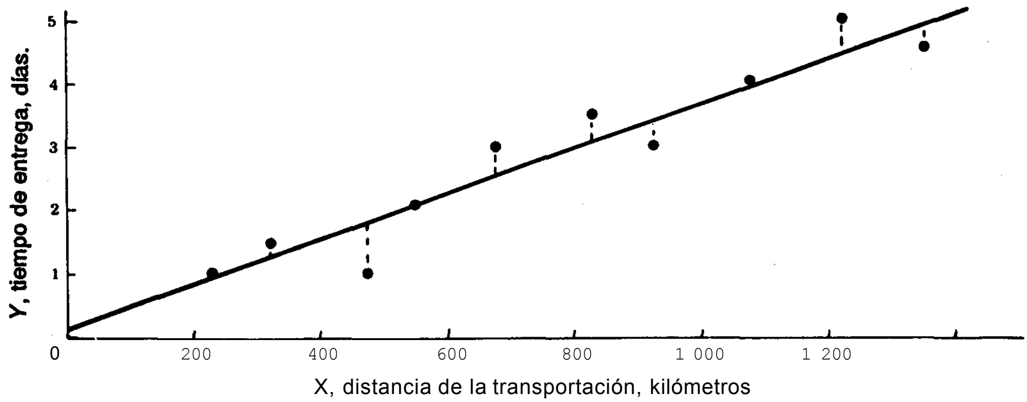


Fig. 14.4

14.3 Determine los residuales y construya una gráfica de residuales con respecto a los valores ajustados para los datos de la Tabla 14.2, utilizando la ecuación de regresión que se elaboró en el problema 14.2. Compare la gráfica de residuales con el diagrama de dispersión de la figura 14-4.

Tabla 14.3 Cálculo de los residuales para el problema del tiempo de entrega

Embarque muestreado	Distancia *(X)*, kilómetros	Tiempo de entrega *(Y)*, días	Valor ajustado ($\hat{Y}$)	Residual ($e = Y - \hat{Y}$)
1	825	3.50	3.08	0.42
2	215	1.00	0.88	0.12
3	1,070	4.00	3.96	0.04
4	550	2.00	2.09	-0.09
5	480	1.00	1.84	-0.84
6	920	3.00	3.42	-0.42
7	1,350	4.50	4.97	-0.47
8	325	1.50	1.28	0.22
9	670	3.00	2.52	0.48
10	1,215	5.00	4.48	0.52

En la Tabla 14.3 se presenta el cálculo de los residuales, en tanto que en la figura 14-5 se muestra la gráfica correspondiente. Observe que la forma general de la dispersión de los valores de la variable dependiente *Y,* con respecto a la línea de regresión de la gráfica de esparcimiento, y la dispersión de los residuales e con respecto a la linea de desviación "0", de la gráfica de residuales, son similares, exceptuando por la escala vertical "alargada" de la gráfica de residuales.

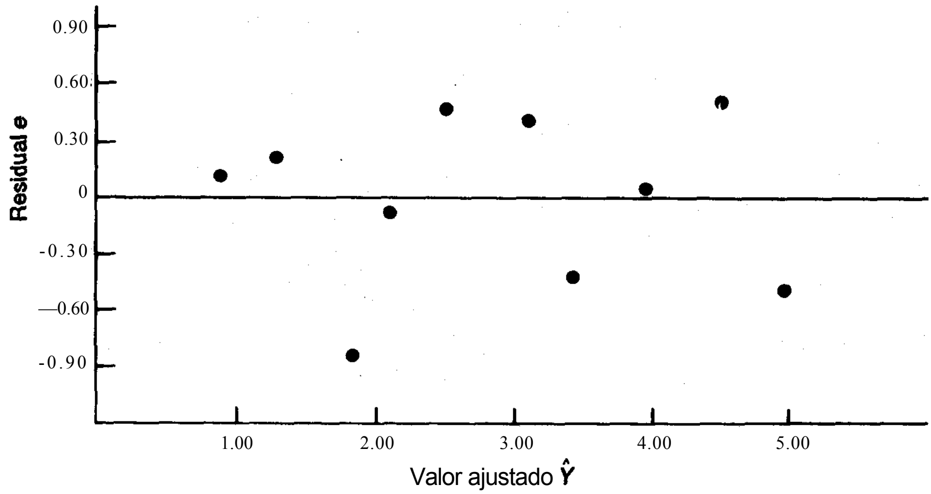


FIg. 14.5 Diagrama residual

14.4 Al utilizar la ecuación de regresión que se elaboró en el problema 14.2, estime el tiempo de entrega, desde el momento en que el embarque está disponible para un viaje de 1000 kilómetros. ¿Podría utilizarse esta ecuación de regresión para estimar el tiempo de entrega para un embarque de 2500 kilómetros?.

$$\hat{Y} = 0.11 + 0.0036\,x = 0.11 + 0.0036(1000) = 3.71 \text{ días}$$

No es apropiado utilizar la ecuación anterior para un viaje de 2500 kilómetros, porque los datos muestrales para esa ecuación de regresión lineal estimada incluyeron sólo datos de distancias de hasta 1350 kilómetros.

14.5 Calcule el error estándar del estimador para el problema del análisis de tiempo de entrega, y con referencia a los valores que se determinaron en la solución al problema 14.2.

$$s_{Y.X} = \sqrt{\frac{\Sigma Y^2 - b_0\Sigma Y - b_1\Sigma XY}{n-2}} = \sqrt{\frac{99.75-(0.11)(28.5)-(0.0036)(26\,370)}{10-2}}$$
$$= \sqrt{\frac{1.683}{8}} = \sqrt{0.2104} = 0.4587 \cong 0.46$$

14.6 Calcule el error estándar del estimador mediante la fórmula básica que utiliza los residuales y que se determinó en el problema 14.3. Compare su respuesta con la que se obtuvo mediante la fórmula abreviada del problema 14.5 anterior.

$$s_{Y.X} = \sqrt{\frac{\Sigma e^2}{n-2}} = \sqrt{\frac{1.8526}{10-2}} = 0.4812 \cong 0.48$$

Excepto por una ligera diferencia de redondeo, esta respuesta coincide con la que se obtuvo en el problema 14.5.

14.7 Al utilizar el error estándar del estimador del problema 14.5, pruebe la hipótesis nula H_0: $\beta_1 = 0$ para la distancia del viaje y el tiempo de entrega de la Tabla 14.2, utilizando un nivel de significancia del 5%.

Dado $s_{Y.X}$ 0.46 y los valores de la Tabla 14.2,

$$s_{b_1} = \frac{s_{Y.X}}{\sqrt{\Sigma X^2 - n\bar{X}^2}} = \frac{0.46}{\sqrt{7\,104\,300 - 10(762)^2}} = \frac{0.46}{1\,139.24} = 0.0004$$

$$H_0\colon \beta_1 = 0 \qquad H_1\colon \beta_1 \neq 0$$

$$t \text{ crítica } (gl = 8, \alpha = 0.05) = \pm 2.306$$

$$t = \frac{b_1}{s_{b_1}} = \frac{0.0036}{0.0004} = +9.00$$

Como el valor calculado de f de + 9.00 se encuentra en una de las regiones de rechazo para esta prueba con dos criterios de calificación, se concluye que existe una relación significativa entre la distancia y el tiempo de entrega.

14.8 Con referencia al problema 14.7 anterior, pruebe la hipótesis nula H_0: $\beta_1 \leq 0$ a un nivel de significancia del 5%.

$$H_0 : \beta_1 \leq 0 \qquad H_1 : \beta_1 > 0$$

$$t \text{ crítica } (gl = 8, \alpha = 0.05) = +1.860$$

$$t = +9.00 \quad \text{(del problema 14.7)}$$

Como la estadística (calculada de + 9.00 excede el valor crítico de + 1.860, se rechaza la hipótesis nula. Para esta prueba con un criterio de calificación, se concluye que la pendiente de la línea de regresión poblacional es positiva y que, por ello, existe una relación directa (y no inversa) entre la distancia del embarque y el tiempo de entrega.

14.9 Determine el intervalo de confianza del 95% para β_1, con los datos de distancia y tiempo de entrega que se analizaron en los problemas anteriores.

Como $b_1 = 0.0036$ (del problema 14.2) y $gl = n - 2 = 10 - 2 = 8$, el intervalo de confianza del 95% para β_1 es

$$b_1 \pm ts_{b_1} = 0.0036 \pm (2.306)\,(0.0004) = 0.0036 \pm 0.0009 = 0.0027 \text{ a } 0.0045$$

14.10 Al utilizar los valores que se determinaron en los problemas anteriores, construya el intervalo de confianza del 95% para *media* del tiempo de entrega para una distancia de 1000 kilómetros.

Dado $\hat{\mu}_Y$ (para X= 1000) = 3.71 días, $s_{Y.X} = 0.46$, y los valores de la Tabla 14.2,

$$s_{\bar{Y}.X} = s_{Y.X}\sqrt{\frac{1}{n} + \frac{(X - \bar{X})^2}{\Sigma X^2 - (\Sigma X)^2/n}} = 0.46\sqrt{\frac{1}{10} + \frac{(1000 - 762)^2}{7\,104\,300 - (7\,620)^2/10}} = 0.1748 \cong 0.17$$

El intervalo de confianza del 95% para la media condicional (en donde $gl = 10 - 2 = 8$) es

$$\hat{\mu}_Y \pm ts_{\bar{Y}.X} = 3.71 \pm (2.306)\,(0.17) = 3.71 \pm 0.39 = 3.32 \text{ a } 4.10 \text{ días}$$

Por ello, se estima que el tiempo promedio de entrega para embarques a 1000 kilómetros, desde el momento en que el embarque está disponible, se encuentra entre 3.32 y 4.10 días, con una confianza del 95% en ese intervalo de estimación.

14.11 Al utilizar los valores que se determinaron en los problemas anteriores, determine el intervalo de predicción del 95% para el tiempo de entrega de los embarques que se envían a una distancia de 1000 kilómetros. Compare ese intervalo con el que se construyó en el problema 14.10.

Como $\hat{Y}$ (para $X = 1000$) = 3.71 días, $s_{Y.X} = 0.46$, y $s_{\hat{Y}.X} = 0.17$

$$s_{Y\,\text{siguiente}} = \sqrt{s_{Y.X}^2 + s_{\hat{Y}.X}^2} = \sqrt{(0.46)^2 + (0.17)^2} = \sqrt{0.2405} = 0.4904 \cong 0.49$$

En donde $gl = 10-2=8$, el intervalo de predicción del 95% es

$$\hat{Y} \pm ts_{Y\,\text{siguiente}} = 3.71 \pm 2.306(0.49) = 3.71 \pm 1.13 = 2.58 \text{ a } 4.84 \text{ días}$$

Tal como se esperaba, este intervalo de predicción es un tanto más amplio que el que se construyó para la media condicional en el problema 14.10, puesto que en esta aplicación el intervalo se refiere a un valor *individual,* y no a una media.

ANÁLISIS DE CORRELACIÓN

14.12 Para los datos de distancia y tiempo de entrega, el procedimiento de muestreo que se describe en el problema 14.1 indica que ambas variables son, de hecho, variables aleatorias. Si, además, se asume una distribución normal bivariada para la población e iguales varianzas condicionales para cada variable, entonces se puede aplicar el análisis de correlación a los datos muestrales. Utilizando los datos que se calcularon en el problema 14.2, calcule el coeficiente de determinación para los datos muestrales (ignore el sesgo de este coeficiente).

$$r^2 = \frac{b_0\Sigma Y + b_1\Sigma XY - n\bar{Y}^2}{\Sigma Y^2 - n\bar{Y}^2}$$

$$= \frac{(0.11)(28.5) + (0.0036)(26\ 370) - (10)(2.85)^2}{99.75 - (10)(2.85)^2}$$

$$= \frac{16.842}{18.525} = 0.9091 \cong 0.91$$

Por ello, como estimador puntual, puede concluirse que aproximadamente el 91 % de la varianza del tiempo de entrega queda estadísticamente explicada por la distancia del viaje. Además, considerando la distancia de los viajes, puede concluirse también que sólo el 9% de la varianza permanece sin explicación.

14.13 Para los datos de distancia y de tiempo de entrega (a) calcule el coeficiente de correlación, con respecto al coeficiente de determinación del problema 14.12, y (b) determine el coeficiente de correlación utilizando la fórmula abreviada alternativa para r.

(a) $r = \sqrt{r^2} = \sqrt{0.9091} = +0.9535 \cong +0.95$

El valor positivo de la correlación se basa en que la pendiente b_1 de la línea de regresión es positiva, tal como se calculó en el problema 14.2.

(b) $$r = \frac{n\Sigma XY - \Sigma X\Sigma Y}{\sqrt{n\Sigma X^2 - (\Sigma X)^2}\sqrt{n\Sigma Y^2 - (\Sigma Y)^2}} = \frac{(10)(26\ 370) - (7\ 620)(28.5)}{\sqrt{(10)(7\ 104\ 300) - (7\ 620)^2}\sqrt{(10)(99.75) - (28.5)^2}}$$

$$= \frac{46\ 530}{(3\ 602.5824)(13.6107)} = \frac{46\ 530}{49\ 033.668} = +0.9489 \cong +0.95$$

Los dos valores son iguales, excepto por una ligera diferencia debida a redondeo.

14.14 Para los datos de distancia y de tiempo de entrega, (a) calcule la covarianza muestral, y (b) convierta el valor de la covarianza en el coeficiente de variación. Compare su respuesta al inciso (b) con las respuestas que se obtuvieron para el coeficiente de correlación en el problema 14.13.

(a) $$\text{cov}(X, Y) = \frac{\Sigma[(X - \bar{X})(Y - \bar{Y})]}{n - 1}$$

$$= \frac{4\,653}{9} = +517.00 \qquad \text{(de la Tabla 14.4)}$$

(b) $$r = \frac{\text{cov}(X, Y)}{s_X s_Y}$$

donde $$s_X = \sqrt{\frac{\Sigma X^2 - n\bar{X}^2}{n-1}} \qquad \text{[de la fórmula } \mathit{(4.11)}\text{]}$$

$$= \sqrt{\frac{7\,104.300 - 10(762)^2}{10 - 1}} \qquad \text{(de la Tabla 14.2)}$$

$$= \sqrt{144\,206.66} = 379.7455$$

$$s_Y = \sqrt{\frac{\Sigma Y^2 - n\bar{Y}^2}{n-1}} \qquad \text{[de la fórmula } \mathit{(4.11)}\text{]}$$

$$= \sqrt{\frac{99.75 - 10(2.85)^2}{10 - 1}} \qquad \text{(de la Tabla 14.2)}$$

$$= \sqrt{2.058333} = 1.4347$$

$$r = \frac{517.00}{(379.7455)(1.4347)} = +0.9489 \cong +0.95$$

El coeficiente de correlación es igual al que se calculó en el problema 14.13.

Tabla 14.4 Cálculo de la covarianza

Embarque muestreado	Distancia, *(X)*, kilómetros	Tiempo de entrega *(Y)*, días	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X})(Y - \bar{Y})$
1	825	3.5	63	0.65	40.95
2	215	1.0	-547	-1.85	1011.95
3	1 070	4.0	308	1.15	354.20
4	550	2.0	-212	-0.85	180.20
5	480	1.0	-282	-1.85	521.70
6	920	3.0	158	0.15	23.70
7	1 350	4.5	588	1.65	970.20
8	325	1.5	-437	-1.35	589.95
9	670	3.0	-92	0.15	-13.80
10	1215	5.0	453	2.15	973.95
	$\bar{X} = 762$	$\bar{Y} = 2.85$			$\Sigma = 4\,653.00$

14.15 Determine si el valor de la correlación que se calculó en el problema 14.14 (b) es significativamente distinto de 0, a un nivel de significación del 5%.

$$\rho = 0 \qquad\qquad \rho \neq 0$$

$$t \text{ crítica } (gl = 8, \alpha = 0.05) = \pm 2.306$$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.9489}{\sqrt{\frac{1-0.9004}{10-2}}} = \frac{0.9489}{0.1116} = +8.50$$

Como la estadística de prueba t = + 8.50, se encuentra en una región de rechazo, se rechaza la hipótesis nula y se concluye que sf existe una relación significativa entre la distancia y el tiempo de entrega. Observe que esta conclusión coincide con la prueba de la hipótesis nula de que $\beta_1 = 0$, del problema 14.7. La diferencia en los valores calculados de f en estas dos pruebas se debe al redondeo de valores en cálculos previos.

14.16 Se encuentra que el coeficiente de correlación entre los ingresos y los montos de la deudas pendientes a corto plazo es de r = +0.50, para una muestra de n=10 personas a las que una institución financiera otorgó un préstamo.

(a) Pruebe la hipótesis de que no existe correlación entre esas dos variables, en toda la población de personas que obtienen préstamos, utilizando el nivel de significación del 5%.

(b) Interprete el significado del coeficiente de correlación que se calculó.

(a)

$$H_0\colon \rho = 0 \qquad\qquad H_1\colon \rho \neq 0$$

$$t \text{ crítica } (gl = 10 - 2 = 8, \alpha = 0.05) = \pm 2.306$$

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.50}{\sqrt{\frac{1-(0.50)^2}{8}}} = \frac{0.50}{0.306} = +1.634$$

Como la estadística f calculada de +1.634 no se encuentra en una región de rechazo, no es posible rechazar la hipótesis nula y se continúa aceptando la suposición de que no existe relación entre las dos variables. Puede pensarse que la relación que se observa en la muestra se debe al azar, con un nivel de significancia del 5%.

(b) Con base en el coeficiente de correlación de r = + 0.50, podría tenerse la tentación de concluir que, como r^2- 0.25, aproximadamente de 25% de ta varianza de las deudas a corto plazo se explica estadísticamente por la cantidad de ingreso por hogar. *Sin embargo,* como no se rechazó la hipótesis nula en la parte (a), una interpretación más apropiada diría que nada de la varianza de Y está asociada con cambios en X. De esta manera, es apropiado considerar la interpretación de r^2 sólo si se rechaza la hipótesis nula de que no existe relación.

RESULTADOS POR COMPUTADORA

14.17 Utilice algún programa de computación para realizar los análisis de regresión y correlación para los datos de distancia y tiempo de entrega de la Tabla 14.2. Compare los resultados específicos con los que se obtuvieron en los cálculos manuales de los ejercicios anteriores.

Los datos de entrada y los resultados de la computadora se presentan en la figura 14-6.
Se identifican los siguientes elementos en los resultados:

(1) El diagrama de dispersión, que corresponde al diagrama que se preparó en forma manual en la gráfica 14-3,

② La ecuación de regresión, que corresponde a la solución del problema 14.2, excepto por el efecto del redondeo de valores en los cálculos manuales.
③ La gráfica de residuales, que corresponde a la gráfica que se preparó en forma manual en la figura 14-5.
④ El tiempo estimado de entrega para un embarque de 1000 kilómetros que corresponde a la solución del problema 14.4.
⑤ El error estándar del estimador, que corresponde a la solución del problema 14.6.
⑥ La prueba t para la hipótesis nula, H_0: $\beta_1 = 0$, que corresponde a la solución del problema 14.7
⑦ El intervalo de confianza del 95% para el tiempo de entrega promedio de todos los embarques de 1000 kilómetros que corresponde a la solución del problema 14.10.
⑧ El intervalo de predicción del 95% para el tiempo de entrega de un embarque especifico a 1000 kilómetros, que corresponde a la solución del problema 14.11.
⑨ El coeficiente de determinación entre el tiempo de entrega y la distancia, que corresponde a la solución del problema 14.12.
⑩ El coeficiente de correlación entre el tiempo de entrega y la distancia, que corresponde a la solución del problema 14.13.
⑪ La covarianza de la distancia y el tiempo de entrega, que corresponde a la solución del problema 14.14(a).

PROBLEMAS COMPLEMENTARIOS

ANÁLISIS DE REGRESIÓN LINEAL

14.18 En la Tabla 14.5 se presentan datos muestrales sobre el número de horas de estudio invertidas por los estudiantes fuera de clase, durante un periodo de tres semanas, para un curso de estadística de negocios, junto con las calificaciones que obtuvieron en un examen aplicado al final de ese periodo. Prepare un diagrama de dispersión para estos datos, y observe si parecen satisfacerse las suposiciones de linealidad y de igualdad de las varianzas condicionales.

Tabla 14.5 Horas invertidas en estudio, en un curso de estadística y calificaciones del examen para una muestra *de n =8* estudiantes

Estudiante muestreado	1	2	3	4	5	6	7	8
Horas de estudio (X)	20	16	34	23	27	32	18	22
Calificación en el examen (Y)	64	61	84	70	88	92	72	77

Resp. Con base en el diagrama de dispersión, parece que las suposiciones de linealidad y de igualdad de varianzas condicionales se satisfacen razonablemente.

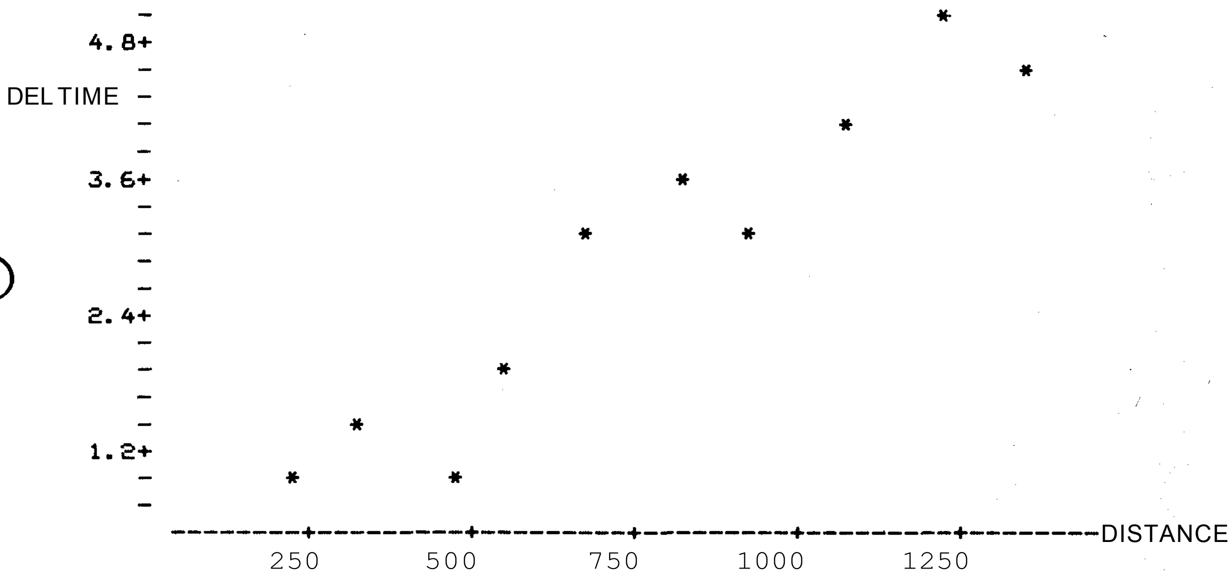
14.19 Determine la recta de regresión por mínimos cuadrados para los datos de la Tabla 14.5, y trace la recta sobre la gráfica que se construyó en el problema 14.18.

Resp. $\hat{Y} = 40 + 1.5X$

14.20 Determine los residuales y construya una gráfica de residuales con los valores ajustados de los datos de la tabla 14.5, utilizando la ecuación de regresión que se elaboró en el problema 14.19. Compare la gráfica de residuales con el diagrama de dispersión que se elaboró en el problema 14.18.

```
MTB > NAME C1 = 'DISTANCE', C2 = 'DELTIME'
MTB > READ 'DISTANCE', 'DELTIME'
DATA >     825       3.5
DATA >      215      1.0
DATA >    1070       4.0
DATA >     550       2.0
DATA >     480       1.0
DATA >      920      3.0
DATA >    1350       4.5
DATA >      325      1.5
DATA >     670       3.0
DATA >     1215      5.0
DATA > END
MTB > PLOT 'DELTIME' VS. 'DISTANCE'
```

1




```
MTB > REGRESS 'DELTIME' USING 1 PREDICTOR 'DISTANCE', C3, PUT FITS IN C4;
SUBO RESIDUPLS IN C5;
SUBO PREDICT 1000.
```

2

```
The regression equation is
DELTIME = 0.118 + 0.00353 DISTANCE

Predictor        Coef         Stdev      t-ratio
Constant       0.1181        0.3551         0.33
DISTANCE   0.00935651     0.0004214         8.51
```

6

```
s = 0.4800        R-sq = 90.0%       R-sq(adj) = 88.8%
```

5 9

```
Analysis of Variance

SOURCE        DF          SS          MS
Regression     1      16.882      16.682
Error          8       1.843       0.230
Total          9      18.525
```

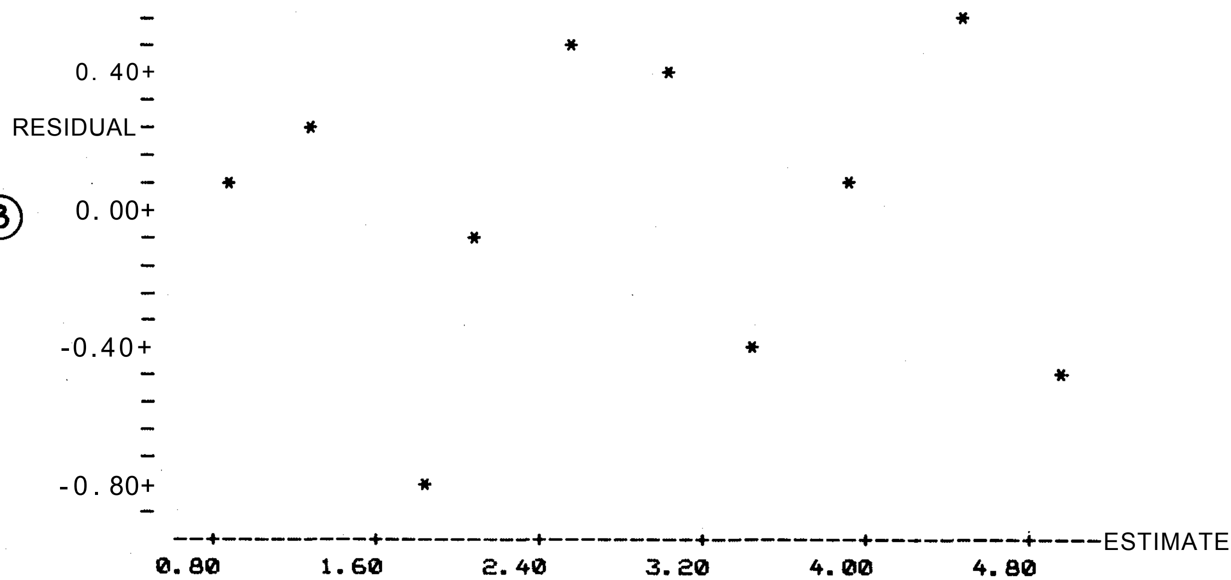
Fig. (14.6) Resultado de Minitab (las partes numeradas se definen en la solución para el problema 14.17).

(Continúa en la p. 296)

```
    Fit   Stdev.Fit          95% C.I.             95% P.I.
  3.703       0.182      (  3.284,   4.123)    (  2.519,   4.887)
(4)                   (7)                   (8)

MTB > NAME C4 = 'ESTIMATE', C5 = 'RESIDUAL'
MTB > PLOT 'RESIDUAL' VS. 'ESTIMATE'
```

(3)




```
MTB > CORRELATION OF 'DISTANCE' AND 'DELTIME'

Correlation of DISTANCE and DELTIME = 0.949   (10)

MTB > COVARIANCE OF 'DISTANCE' AND 'DELTIME'

             DISTANCE     DELTIME
DISTANCE    144806.62
DELTIME  (11) 517.00         2.06
```

Fig. (14.6) *(Continuación).*

14.21 Calcule el error estándar del estimador para los datos de la Tabla 14.5, utilizando los residuales que se determinaron en el problema 14.20.

Resp. $S_{YX} = 6.16$

14.22 Para la información muestral que se presentó en el problema 14.18, (a) pruebe la hipótesis nula de que la pendiente de la recta de regresión es 0, utilizando el nivel de significancia del 1 %, e interprete el resultado de esta prueba; *(b)* repita la prueba para la hipótesis nula de que el verdadero coeficiente de regresión es igual *o menor que* cero, utilizando el nivel de significancia del 1 %.

Resp. (a) Rechazar H_0 y concluir que existe una relación significativa, (b) Rechazar H_0 y concluir que existe una relación positiva significativa.

14.23 Utilice la ecuación de regresión que se determinó en el problema 14.19 para estimar la calificación de un estudiante que dedicó 30 horas al estudio de la materia.

Resp. $Y = 85$

14.24 Con referencia a los problemas 14.21 y 14.23, construya el intervalo de confianza del 90% para estimar la calificación promedio de los estudiantes que dedicaron 30 horas a preparar el curso.

Resp. 79.02 a 90.98.

14.25 Con referencia a los problemas anteriores, construya el intervalo de predicción del 90% para las calificaciones de un estudiante individual que dedicó 30 horas a la preparación del curso.

Resp. 71.59 a 98.41

14.26 En la Tabla 14.6 se presentan datos que relacionan el número de semanas de experiencia en un trabajo de instalación de cables de componentes electrónicos en miniatura, y el número de componentes que se rechazaron la semana anterior, para 12 trabajadores seleccionados al azar. Trace estos datos muestrales en un diagrama de dispersión.

Tabla 14.6 Semanas de experiencia y número de componentes rechazados durante una semana muestreada para 12 trabajadores de ensamble

Trabajador muestreado	1	2	3	4	5	6	7	8	9	10	11	12
Semanas de experiencia *(X)*	7	9	6	14	8	12	10	4	2	11	1	8
Número de rechazos *(Y)*	26	20	28	16	23	18	24	26	38	22	32	25

14.27 De la Tabla 14.6, determine la ecuación de regresión para justificar el número de componentes rechazados, dado un número específico de semanas de experiencia, y grafique la recta de regresión sobre el diagrama de dispersión. Comente la naturaleza de la relación, según señala la ecuación de regresión.

Resp. $\hat{Y} = 35.57 - 1.40X$

14.28 Pruebe la hipótesis nula de que no existe relación entre las variables de la Tabla 14.6, y que la pendiente de la recta poblacional de regresión es cero, utilizando un nivel de significancia del 5% para esta prueba.

Resp. Rechazar H_0 y concluir que existe una relación.

14.29 Para la información muestra! de los problemas 14.26 y 14.27, construya un Intervalo de confianza del 95% para estimar el valor del coeficiente poblacional de regresión β_1, e interprete el valor de este coeficiente.

Resp. -1.85 a -0.95

14.30 Al utilizar la ecuación de regresión que se elaboró en el problema 14.27, estime el número de componentes rechazados para un empleado con tres semanas de experiencia en el empleo.

Resp. $\hat{Y} = 31.37$

14.31 Al continuar con el problema 14.30, construya el intervalo de confianza del 95% para estimar el número promedio de rechazos para empleados que tienen tres semanas de experiencia en la operación.

Resp. 28.74 a 34.00

14.32 Con referencia a los problemas 14.30 y 14.31, construya el intervalo de predicción del 95% para el número de componentes rechazados correspondiente a un empleado con tres semanas de experiencia en el empleo.

Resp. 25.09 a 37.65

ANÁLISIS DE CORRELACIÓN

14.33 Calcule el coeficiente de determinación y el coeficiente de correlación para los datos de la Tabla 14.5, que se analizaron en los problemas 14.18 a 14.25, aprovechando el hecho de que en el problema 14.19 se calcularon los valores de b_0 y b_1 para la ecuación de regresión. Interprete los coeficientes dados.

Resp. $r^2 = 0.7449 \cong 0.74$, $r = +0.863 \cong +0.86$

14.34 Para el valor muestral que se determinó en el problema 14.33, pruebe la hipótesis nula de que (a) $p = 0$ y (b) $p \leq 0$, utilizando el nivel de significancia del 1 %, con respecto a cada prueba. Interprete sus resultados.

Resp. (a) Rechazar H_0 y concluir que existe una relación significativa, *(b)* Rechazar H_1 y concluir que existe una relación positiva significativa.

14.35 Para los datos de la Tabla 14.5, (a) calcule la covarianza muestral, y (b) convierta esta covarlanza en coeficiente de correlación. Compare su respuesta a (b) con el coeficiente de correlación que se obtuvo en el problema 14.33.

Resp. (a) $\text{cov}(x, y) = +62.86$, (b) $r = +0.8622 \cong +0.86$

14.36 Para los datos muestrales que se reportan en la Tabla 14.6, y que se analizan en los problemas 14.26 a 14.32, determine el valor del coeficiente de correlación utilizando la fórmula que no se basa en el uso de los valores b_0 y b_1 de la recta estimada de regresión. Interprete el significado de este valor calculando el coeficiente de determinación.

Resp. $r = -0.908 \cong -0.91$, $r^2 = 0.8245 \cong 0.82$

RESULTADOS POR COMPUTADORA

14.37 Utilice algún programa de computación para realizar los análisis de regresión y correlación de los datos que aparecen en la Tabla 14.6. Como parte de los resultados, incluya un diagrama de dispersión, el cálculo de la ecuación de regresión, el intervalo de confianza del 95% para el número promedio de rechazos correspondiente a empleados con tres semanas de experiencia, el intervalo de predicción del 95% para un empleado individual que tiene 3 semanas de experiencia y el coeficiente de correlación muestral. Compare los resultados de la computadora con los que se obtuvieron mediante los cálculos manuales en los ejercicios anteriores.

Resp. Excepto por diferencias pequeñas debidas a redondeo, los resultados son iguales a los que se obtuvieron mediante los cálculos manuales.

Regresión y correlación múltiple

15

15.1 OBJETIVOS Y SUPOSICIONES DEL ANÁLISIS DE REGRESIÓN LINEAL MÚLTIPLE

El análisis de regresión múltiple es una extensión del análisis de regresión simple, según se describe en el capítulo 14, a aplicaciones que implican dos o más variables independientes para estimar el valor de la variable dependiente. En el caso de dos variables independientes, que se denotan con X_1 y X_2, el modelo algebraico lineal es

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \varepsilon_i \tag{15.1}$$

Las definiciones de los términos anteriores son equivalentes a las definiciones que se dieron en la sección 14.3 para el análisis de regresión simple, excepto que en este caso existe más de una variable dependiente. Con base en los datos muestrales, la ecuación de regresión lineal para el caso de dos variables independientes es

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 \tag{15.2}$$

(Nota: en algunos libros de texto y en algunos programas de computación se designa a la variable dependiente X_1; entonces, las diversas variables independientes se identifican en forma secuencial comenzando con X_2.)

La ecuación de regresión simple identifica la mejor línea de ajuste con base en el método de mínimos cuadrados, según se describe en la sección 14.3. En el caso del análisis de la regresión múltiple, la línea del mejor ajuste es un hiperplano en el espacio n-dimensional (espacio tridimensional en el caso de dos variables independientes). Los cálculos que se requieren para determinar los valores de los estimadores de los parámetros para la ecuación de regresión múltiple y los correspondientes valores del error estándar son complicados y, por lo general, implican álgebra de matrices. Sin embargo, hay muchos paquetes de computación disponibles para llevar a cabo esos cálculos, y los problemas resueltos del final de este capítulo utilizan esos programas. Algunos libros de texto especializados en el análisis de regresión y correlación incluyen descripciones matemáticas completas de los análisis que se llevan a cabo.

Las suposiciones para el análisis de regresión lineal múltiple son similares a las del caso simple que implica una sola variable independiente. Para las estimaciones puntuales, las suposiciones principales son que: (1) la variable dependiente es una variable aleatoria, pero no es necesario que las variables independientes sean variables aleatorias; (2) la relación entre las diversas variables independientes y la variable dependiente es lineal; y (3) las varianzas de las distribuciones condicionales de la variable dependiente, dadas diversas combinaciones de valores de las variables independientes, son todas iguales (homoscedasticidad). Para la estimación por intervalos, una suposición adicional es que (4) las distribuciones condicionales para la variable dependiente tienen distribución de probabilidad normal.

15.2 CONCEPTOS ADICIONALES EN EL ANÁLISIS DE REGRESIÓN MÚLTIPLE

La Constante (en la ecuación de regresión): aunque tanto b_0 como los diversos valores b_i son todos estimaciones de parámetros en la ecuación de regresión, en la mayoría de los listados de computación, el término "constante" se refiere al valor de la ordenada al origen, b_0. En análisis de regresión múltiple, éste es el valor de la ecuación de regresión que corresponde a la variable dependiente Y, cuando todas las variables independientes son iguales a cero.

Coeficiente de regresión parcial (o *coeficiente neto de regresión):* de hecho, cada uno de los coeficientes de regresión b_i es un coeficiente de regresión parcial. Un coeficiente de regresión parcial es el coeficiente condicional dado que se incluyen también en la ecuación una o más de las otras variables independientes (y sus coeficientes). En términos conceptuales, un coeficiente parcial de regresión representa la pendiente de la linea de regresión entre la variable independiente de interés, y la variable dependiente *dado que* las otras variables independientes se incluyen en el modelo y "se mantienen constantes". El símbolo $b_{Y1.2}$ (o $b_{12.3}$ cuando la variable dependiente se designa mediante X_1), es el coeficiente parcial de regresión para la primera variable independiente, dado que también se incluye en la ecuación de regresión una segunda variable independiente. Por sencillez, cuando se presenta la ecuación de regresión completa, a este coeficiente se le designa por lo general como b_1.

Uso de la prueba F. Como se describe en la sección 15.5, el análisis de varianza se utiliza en análisis de regresión para probar la significancia del modelo global, antes de considerar la significación de variables independientes individuales.

Uso de pruebas t: Tal como se muestra en la sección de problemas resueltos al final del capítulo, las pruebas t se utilizan para determinar si el coeficiente parcial de regresión de cada variable independiente representa una contribución significativa al modelo global.

Intervalo de confianza para la media condicional: En el problema 15.9 se ilustra la determinación de estos intervalos de confianza. La fórmula para el intervalo de confianza es (en donde el error estándar de la media condicional para dos variables independientes se designa mediante $s_{\bar{Y}.12}$).

$$\hat{\mu}_Y \pm t s_{\bar{Y}.12} \tag{15.3}$$

Intervalos de predicción: El intervalo de predicción para estimar el valor de una observación individual de la variable dependiente, dados los valores de diversas variables independientes, es similar al intervalo de predicción en análisis de regresión simple, según se describe en la sección 14.8. En el problema 15.10 se ilustra el cálculo de uno de esos intervalos. La fórmula general para el intervalo de predicción es

$$\hat{Y} \pm t s_{Y\,(\text{siguiente})} \tag{15.4}$$

Análisis de regresión por pasos: En el análisis de regresión por pasos hacia adelante, se añade al modelo una variable independiente en cada uno de los pasos o etapas de selección de esas variables. En el análisis de regresión por pasos hacia atrás, se comienza incluyendo en el modelo todas las variables pertinentes y, después (posiblemente), se elimina una variable en cada paso. Existen dos enfoques, de entre varios disponibles, para elegir los modelos y el análisis de regresión múltiple. Existen varios programas de computación para realizar el análisis de regresión hacia adelante o hacia atrás. En los problemas resueltos del final del capitulo se utiliza un procedimiento por pasos hacia atrás.

15.3 EL USO DE VARIABLES INDICADORAS (FICTICIAS)

Aunque el modelo de regresión lineal se basa en que las variables independientes se encuentran en una escala cuantitativa de medición, es posible incluir variables cualitativas (categóricas) en los modelos de regresión múltiple. Algunos ejemplos de esas variables son el sexo de un empleado en un estudio de niveles de sueldo y los códigos de ubicación en un modelo de evaluación de bienes raíces.

Las variables indicadoras utilizan el código binario, 0, 1. Se utiliza k para designar el número de categorías que existen para la variable cualitativa, y se requieren k -1 variables indicadoras para codificar la variable cualitativa. Por ello, el sexo de un empleado puede codificarse mediante una variable indicadora, porque k = 2 en este caso y 2 - 1 = 1. El sistema de codificación puede ser, entonces, 0 = mujer y 1 = hombre (o viceversa). Para un modelo de evaluación de bienes raíces en el que existen tres tipos de ubicaciones, identificadas como A, B y C, $k = 3$ y, por lo tanto, se requieren 3 - 1 = 2 variables indicadoras. Estas podrían codificarse de la siguiente manera, suponiendo que existen otras dos variables independientes incluidas en el modelo, X_1 y X_2.

Ubicación	X_3	X_4
A	0	0
B	1	0
C	0	1

(Nota: En el esquema binario de codificación que se ilustra en el párrafo anterior, no puede haber más de un " 1 " en cada renglón, y debe haber exactamente un " 1 " en cada columna.)

La dificultad que se presenta con las variables cualitativas que tienen más de dos categorías, es que se utiliza más de una variable indicadora para representar la misma variable cualitativa. En específico, las variables indicadoras X_3 y X_4 presentadas antes, si se toman yuntas, representan una variable de ubicación. En la regresión por pasos surge la cuestión de si se deben, y cómo, seleccionar en forma apropiada esas variables indicadoras, de manera individual, para incluirlas en el modelo, si resulta que son parte de un conjunto de ellas que se utilizaron para codificar una variable cualitativa. Este tema se trata en libros avanzados sobre análisis de regresión. Los problemas resueltos dei final del capítulo incluyen la consideración de una variable cualitativa con dos categorías.

15.4 RESIDUALES Y GRÁFICAS DE RESIDUALES

Tal como se definió en el capitulo 14.4, en análisis de regresión un residual es

$$e = Y - \hat{Y} \qquad (15.5)$$

En la sección 14.4, se definió la *gráfica de residuales* como una gráfica de los residuales e con respecto a la variable independiente X, o con respecto al valor ajustado $\hat{Y}$. En el análisis de regresión simple, puede utilizarse un diagrama de dispersión o una gráfica de residuales para observar si parecen satisfacerse las suposiciones de linealidad y de igualdad de varianzas condicionales. Sin embargo, en análisis de regresión múltiple, el único tipo de gráfica que permite abordar este análisis para el modelo global es la gráfica de residuales con respecto al valor ajustado $\hat{Y}$, porque ésta es la única gráfica bidimensional que puede incluir el uso de *varias* variables independientes. Si se observa en una de esas gráficas que existe algún problema con respecto a la linealidad o la igualdad de varianzas condicionales, entonces pueden elaborarse gráficas individuales de residuales o de dispersión para cada variable independiente del modelo, con el objeto de buscar la fuente del problema. Los listados de computadora correspondientes a los problemas resueltos incluyen la gráfica de residuales para el modelo global, asi como también las gráficas de residuales para las variables independientes individuales.

15.5 ANÁLISIS DE VARIANZA EN REGRESIÓN LINEAL

Tal como se indicaba en la sección 15.2, se utilizan las pruebas *F* para probar la significancia del modelo global. Es decir, se les utiliza para probar la hipótesis nula de que no existe relación en la población, entre las (diversas) variables independientes consideradas como grupo y la variable dependiente. Si existe sólo una variable dependiente en el modelo de regresión, entonces la prueba *F* es equivalente a una prueba *t* de dos extremos, realizada sobre la pendiente b_1. Por ello, en el análisis de regresión simple no se requiere la prueba *F.* Por claridad, sin embargo, se hace hincapié en el modelo de regresión simple para explicar el razonamiento que subyace al uso del análisis de varianza. La explicación se aplica de manera similar al caso de la regresión múltiple, excepto que los valores estimados $\hat{Y}$ se basan en diversas variables independientes, en vez de sólo una.

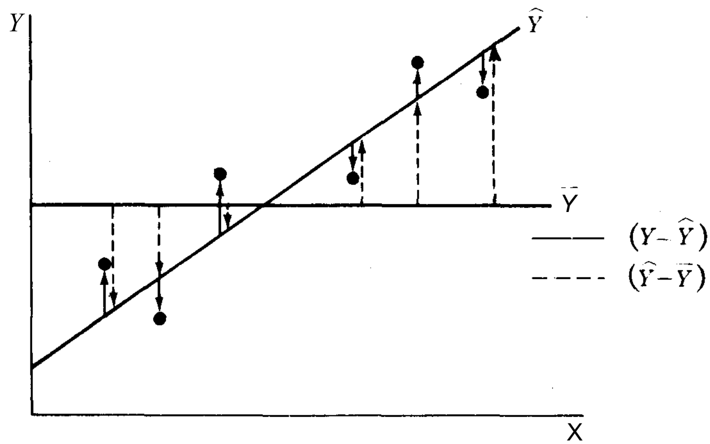


Fig. 15-1

Observe el diagrama de dispersión de la figura 15-t. Si no existe efecto de regresión en la población, entonces la linea $\hat{Y}$ difiere de la linea Y simplemente por azar. Se sigue que el estimador de la varianza basado en las diferencias $(\hat{Y}-\bar{Y})$ que se denomina *cuadrado medio de regresión, (CMR)* sería diferente de la varianza estimada con base en los residuales (Y - $\hat{Y}$) que se denomina *cuadrado medio del error (CME),* sólo por cuestiones aleatorias. Por otro lado, si existe efecto de regresión, entonces el cuadrado medio de regresión se ve aumentado, en comparación con el cuadrado medio del error. En la Tabla 15.1 se presenta el formato estándar de las tablas de análisis de varianza que se utilizan para probar la significancia a un efecto global de regresión. Los grados de libertad *k* asociados con el *CMR* de la tabla son el número de variables independientes que aparecen en la ecuación de regresión múltiple. Tal como se muestra en la tabla, la estadística de prueba es:

$$F=\frac{CMR}{CME} \qquad (15.6)$$

En los problemas 15.1 y 15.3 se ilustra el uso de la prueba *F* en análisis de regresión múltiple.

Tabla 15.1 Tabla de análisis de varianza para comprobar la significancia del efecto de regresión

Fuente de variación	Grados de libertad *(gl)*	Suma de cuadrados (SC)	Cuadrado medio (CM)	Cociente *F*
Regresión *(R)*	*k*	*SCR*	$CMR=\frac{SCR}{k}$	$F=\frac{CMR}{CME}$
Error de muestreo (E)	*n-k-l*	*SCE*	$CME=\frac{SCE}{n-k-1}$	
Total (T)	*n-l*	*SCT*		

15.6 OBJETIVOS Y SUPOSICIONES DEL ANÁLISIS DE CORRELACIÓN MÚLTIPLE

El análisis de correlación múltiple es una extensión del análisis de correlación simple (que se describió en el capítulo 14), a situaciones que implican dos o más variables independientes y su grado de asociación con la variable dependiente. Al igual que en el análisis de regresión múltiple que se describe en la sección 15.1, la variable dependiente se designa mediante Y, en tanto que las diversas variables independientes se designan en forma secuencial comenzando con X_1. *(Nota:* en algunos libros de texto y programas de computación, a la variable dependiente se le designa mediante X_1, en cuyo caso las variables independientes se designan en forma secuencial comenzando con X_2.)

El coeficiente de correlación múltiple, que se designa mediante $R_{Y.12}$ para el caso de dos variables independientes, muestra la medida de la relación entre dos variables independientes consideradas como grupo, y la variable dependiente. Como es posible que una de las variables Independientes tenga una relación positiva con la dependiente, al mismo tiempo que la otra puede tener una relación negativa con la variable dependiente, todos los valores de R que se reportan son valores absolutos, sin signo algebraico. El problema 15.12 incluye el cálculo de un coeficiente de correlación múltiple y la prueba de significancia de ese coeficiente.

El *coeficiente de determinación múltiple* se representa mediante $R^2_{Y.12}$ para el caso de dos variables independientes. Siendo su interpretación similar a la del coeficiente de determinación simple (sección 14.10), éste indica la proporción de varianza de la variable dependiente que queda estadísticamente explicada al conocer las dos (o más) variables independientes. El coeficiente de determinación múltiple muestra! para el caso de dos variables independientes es

$$R^2_{Y.12} = 1 - \frac{s^2_{Y.12}}{s^2_Y} \tag{15.7}$$

La fórmula *(15.7)* es equivalente a la fórmula *(14.20)* del análisis de regresión simple, y se presenta con propósitos conceptuales, más que para su aplicación en cálculos. Como en este capítulo se utilizan programas de computación para los análisis de regresión y correlación múltiples, no se incluyen los procedimientos de cálculo. En el problema 15.12 se considera un coeficiente de determinación múltiple incluido en los listados de la computadora, y su interpretación.

Las suposiciones del análisis de correlación múltiple son similares a las que se aplican al caso simple con una sola variable independiente. Estas suposiciones son: (1) todas las variables implicadas en el análisis son variables aleatorias, (2) todas las relaciones son lineales, (3) las varianzas condicionales son todas iguales (homoscedasticidad) y (4) las distribuciones condicionales son todas normales. Estos requerimientos son bastante exigentes y rara vez se satisfacen por completo en situaciones reales. Sin embargo, el análisis de correlación múltiple es bastante robusto porque permite violar algunas de estas suposiciones, en particular la que se refiere a que todas las distribuciones condicionales son normales, sin que se tengan consecuencias serias sobre la validez de los resultados.

15.7 CONCEPTOS ADICIONALES EN EL ANÁLISIS DE CORRELACIÓN MÚLTIPLE

Además del coeficiente de correlación múltiple y del coeficiente de determinación múltiple que se describieron en la sección anterior, los siguientes conceptos o procedimientos son únicos para el análisis de correlación múltiple.

Coeficiente de correlación parcial: Indica la correlación entre una de las variables independientes en el análisis de correlación múltiple, y la variable dependiente, manteniendo la otra u otras variables independientes estadísticamente constantes. La correlación parcial con la primera de las dos variables independientes se designaría mediante $r_{Y1.2}$, en tanto que la correlación parcial con la segunda de las variables independientes se designaría mediante $r_{Y2.1}$. (Si la variable dependiente se designa mediante X_1, entonces estos dos coeficientes se designarían mediante $r_{12.3}$ y $r_{13.2}$, respectivamente.) El valor de la correlación parcial difiere del valor de la correlación simple porque en este último caso, no se controlan estadísticamente las otras variables independientes.

Coeficiente de determinación parcial: Este es el cuadrado del valor del coeficiente de correlación parcial que se describió en el párrafo anterior. Este coeficiente indica la proporción de varianza que queda estadísticamente explicada por una variable independiente especifica, manteniendo la otra o las otras variables independientes constantes en términos estadísticos.

15.8 FALLAS Y LIMITACIONES ASOCIADOS CON LOS ANÁLISIS MÚLTIPLES DE REGRESIÓN Y CORRELACIÓN

Dos áreas importantes de dificultad son las que se refieren a la colinealidad y a la autocorrelación. Enseguida se describen en forma breve. En libros especializados en análisis de regresión y correlación pueden encontrarse revisiones detalladas de estos problemas y de lo que puede hacerse para resolverlos.

Colinealidad o (Multicolinealidad): Cuando las variables independientes de un análisis de regresión múltiple están muy correlacionadas entre sí, los coeficientes de regresión parcial (o neta) son poco confiables en términos de significado. De manera similar, puede resultar cuestionable el significado práctico de los coeficientes de la correlación parcial. Por ejemplo, es posible que la correlación parcial para una variable independiente determinada sea extremadamente negativa, aun cuando la correlación simple sea altamente positiva. Por ello, en general, debe tenerse cuidado al interpretar los coeficientes de regresión parcial y los de correlación parcial, cuando existen variables independientes que tienen una correlación positiva o negativa extrema entre si.

Autocorrelación: Se refiere a la falta de independencia de la variable dependiente *Y* en el muestreo. Es probable que se presente esta situación cuando los valores y correspondan a series de tiempo, en cuyo caso el valor de la variable dependiente de un periodo casi siempre está relacionada con los valores de periodos subsecuentes. En estos casos, se subestima el error estándar de cada uno de los coeficientes de regresión parcial b_j al igual que se subestima el error estándar del estimador. Esto da como resultado que las predicciones o intervalos de confianza sean más estrechos (más precisos) de lo que deberían, y con demasiada frecuencia se rechazan las hipótesis nulas sobre la ausencia de relación.

15.9 RESULTADOS POR COMPUTADORA

Los problemas resueltos al final de este capitulo hacen todos ellos referencia al análisis de Minitab que se presenta en la figura 15-2 y que se aplicó a los datos de la Tabla 15.2. Tal como se ilustra en los problemas, se utiliza un procedimiento por pasos hacia atrás (véase la sección 15.2). Sin embargo, en vez de utilizar el procedimiento automático por pasos hacia atrás disponible en Minitab, se utilizan "comandos" de regresión en cada paso con el objeto de hacer resaltar la lógica del procedimiento, y el criterio que se utiliza para decidir la inclusión de las variables independientes en el modelo de regresión múltiple.

```
MTB > NAME Cl = 'SALARY' , C2 = 'EXPER', C3 = 'EDUC , C4 = 'SEX'
MTB > READ 'SALARY', 'EXPER', 'EDUC, 'SEX'
DATA > 34900   5.5   4.0   0
DATA > 40500   9.0   4.0   1
DATA > 38900   4.0   5.0   0
DATA > 39000   8.0   4.0   1
DATA > 37500   9.5   5.0   1
DATA > 35500   3.0   4.0   0
DATA > 36000   7.0   3.0   0
DATA > 32700   1.5   4.5   0
DATA > 45000   8.5   5.0   1
DATA > 40000   7.5   6.0   0
DATA > 36000   9.5   2.0   1
DATA > 33600   6.0   2.0   0
DATA > 35000   2.5   4.0   1
DATA > 32500   1.5   4.5   1
DATA > END
MTB > REGRESS Cl ON 3 PREDICTORS C2-C4

1 The regression equation is
SALARY = 25495 + 802 EXPER + 1596 EDUC + 383 SEX
```

Fig. 15-2 Listado de Minitab (las etiquetas numeradas son las que se identifican en las soluciones.)

Predictor	Coef	Stdev	t-ratio
Constant	25495	2810	9.07
EXPER	801.6	228.5	3.51
EDUC	1595.7	560.6	2.85
SEX	383	1287	0.30 ③

s = 2252   R-sq = 67.5% ⑫   R-sq(adj) = 57.8%

Analysis of Variance

SOURCE	DF	SS	MS	F
Regression	3	105309888	35103896	6.92 ②
Error	10	50702240	5070224	
Total	13	156012128		

SOURCE	DF	SEQ SS
EXPER	1	6328147E
EDUC	1	41580688
SEX	1	447733

Unusual Observations

Obs.	EXPER	SALARY	Fit	Stdev.Fit	Residual	St.Resid
5	9.50	37500	41472	1193	-3972	-2.08R
9	8.50	45000	40670	1084	4330	2.19R

R denotes an obs. with a large st. resid.

```
MTB > REGRESS Cl ON 2 PREDICTORS C2-C3, C5, PUT FITS IN C6;
SUBC> RESIDUALS IN C7;
SUBC> PREDICT 5.5 4.0.
```

④ The regression equation is
SALARY = 25511 + 826 EXPER + 1604 EDUC

Predictor	Coef	Stdev	t-ratio
Constant	25511	2690	9.48
EXPER	825.7	204.6	4.04
EDUC	1603.7	536.3	2.99 ⑥

s = 2156   R-sq = 67.2%   R-sq(adj) = 61.3%

Analysis of Variance

SOURCE	DF	SS	MS
Regression	2	104862160	5E431072
Error	11	51149968	4649997
Total	13	156012128	

SOURCE	DF	SEQ SS
EXPER	1	63281472
EDUC	1	41580688

Unusual Observations

Obs.	EXPER	SALARY	Fit	Stdev.Fit	Residual	St.Resid
5	9.50	37500	41374	1098	-3874	-2.09R
9	8.50	45000	40548	961	4452	2.31R

R denotes an obs. with a large st. resid.

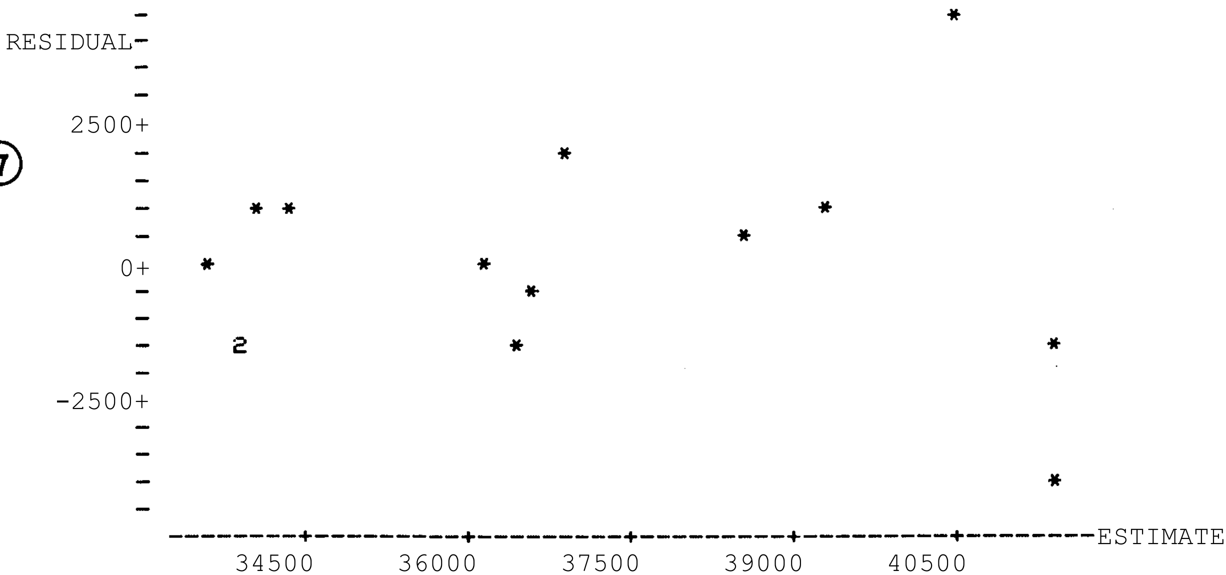
FIg. 15-2 (*continuación*)

```
        Fit   Stdev.Fit          95% C.I.            95% P.I.
      36467         585   (  35180,   37755)   (  31548,   41386)
```

8 9 10

```
MTB > NAME C6 = 'ESTIMATE', C7 = 'RESIDUAL'
MTB > PLOT 'RESIDUAL' VS. 'ESTIMATE'
```

7




```
MTB > PLOT 'RESIDUAL' VS. 'EXPER'
```

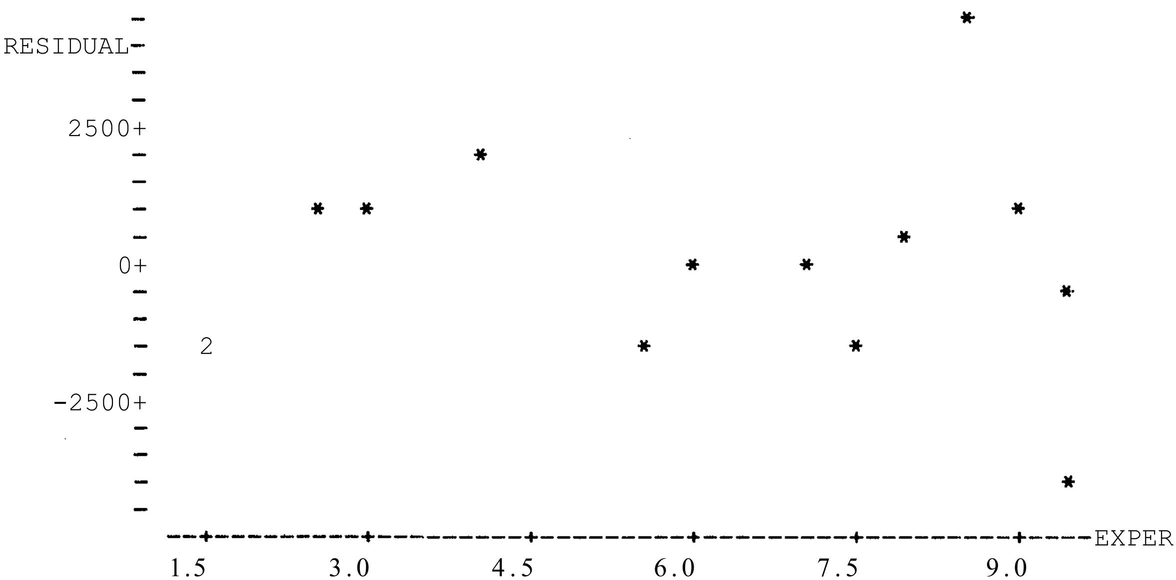
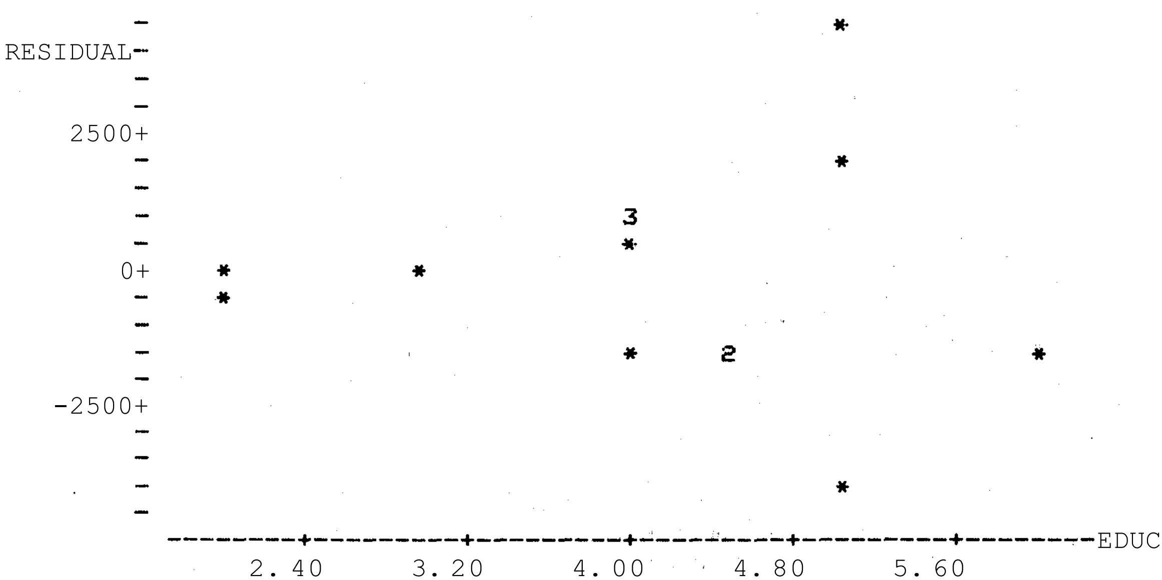


Fig. 15-2 *(continuación)*

```
MTB > PLOT 'RESIDUAL' VS. 'EDUC
```




```
MTB > CORRELATION FOR C1-C4
```

11

```
           SALARY     EXPER      EDUC
EXPER       0.637
EDUC        0.432    -0.126
SEX         0.297     0.352     0.000
```

Ffg. 15-2 *(continuación)*

Tabla 15.2 Salarios diarios para una muestra de supervisores

Persona muestreada	Salario diario *(Y)*	Años de experiencia (X_1)	Años de educación después de la secundaria (X_2)	Sexo (X_3)
1	$34 900	5.5	4.0	F
2	40 500	9.0	4.0	M
3	38 900	4.0	5.0	F
4	39 900	8.0	4.0	M
5	37 500	9.5	5.0	M
6	35 500	3.0	4.0	F
7	36 000	7.0	3.0	F
8	32 700	1.5	4.5	F
9	45 000	8.5	5.0	M
10	40 000	7.5	6.0	F
11	36 000	9.5	2.0	M
12	33 600	6.0	2.0	F
13	35 000	2.5	4.0	M
14	32 500	1.5	4.5	M

Problemas resueltos

En la Tabla 15.2 se reportan datos de sueldos diarios para una muestra aleatoria de n = 14 supervisores, de una población grande. En la figura 15-2 se presenta el análisis de Minitab que sirve de base para las soluciones de los problemas resueltos que aparecen enseguida.

Con los resultados correspondientes al primer comando de regresión de la figura 15-2 (identificado **con ①✗ y para** el caso en el que se incluyen en el modelo las tres variables independientes), pruebe la hipótesis nula de que existe un efecto significativo de regresión, utilizando el 5% como nivel de significancia.

La F crítica (3,10, 0.05) = 3.71. Como la estadística de prueba calculada es F = 6.92 (identificada con ② en el listado), se rechaza la hipótesis nula de que no existe efecto de regresión. Se concluye que ***sí** existe* una relación en la población, entre las tres variables independientes en conjunto y la variable dependiente de salarios.

Debe observarse que con este programa de computación es necesario calcular la estadística F en forma manual.

15.2 Al continuar con el problema 15.1, identifique el coeficiente de regresión parcial que tiene la estadística t menor y determine si la contribución de esa variable al modelo es significativa, a un nivel del 5%.

Debe observarse que los grados de libertad para la prueba t son los mismos que para el CME del análisis de varianza y son iguales al tamaño de la muestra menos el número de parámetros estimados en la ecuación de regresión múltiple o, en este caso, gl = 14 - 4 =10.

La variable para la cual la estadística t es menor es la variable indicadora del sexo (X_3), con t = 0.30 (que se Identifica con ③ en la figura 15-2). Como la t critica (10,0.05) = ± 2.228, no es posible rechazar la hipótesis nula de que $\beta_3 = 0$. Por ello, se concluye que el sexo no contribuye en forma significativa al modelo y se le debe eliminar de la ecuación de regresión múltiple. Como se analizó en la sección 15.2, a ese procedimiento de eliminar una variable a la vez se le denomina *análisis de regresión por pasos hacia atrás.*

15.3

Observe el resultado del segundo comando de regresión de la figura 15-2, que se identifica con el ④, y que se basa en el uso de los años de experiencia y en los años de educación como variables independientes, pero sin incluir la variable indicadora de sexo. Pruebe la hipótesis nula de que no existe efecto de regresión, utilizando un nivel de significancia del 5%.

La F critica (2,11,0.05) = 3.98. Como la estadística F calculada es 11.28 (que se identifica con ⑤ en el listado), se rechaza la hipótesis nula y se concluye que existe un efecto significativo de regresión.

15.4 Al continuar con el problema 15.3, observe cuál de los coeficientes parciales de regresión tiene la menor estadística t y determine si la contribución de esa variable al modelo de regresión múltiple es significativa a un nivel del 5%.

La variable que tiene la menor estadística t es educación (X_2), con t =2.99 (se identifica con ⑥ en el listado). Como la f crítica (11, 0.05) = ±2.201, se rechaza la hipótesis de que $\beta_2 = 0$ a un nivel de significancia del 5%, y se concluye que esa variable *sí* contribuye en forma significativa al modelo de regresión múltiple.

El hecho de que no se elimine ninguna variable en esta etapa del procedimiento por pasos hacia atrás significa que ya se han determinado las variables Independientes que se incluirán en el modelo de regresión múltiple.

Con referencia a la gráfica de residuales identificada con un ⑦ en la figura 15-2, en la que se trazan los residuales con respecto a los valores ajustados $\hat{Y}$, determine si los requerimientos de linealidad y de igualdad de varianzas condicionales parecen satisfacerse.

Parece que si se satisface la suposición de linealidad. Sin embargo, con respecto a la igualdad de las varianzas condicionales, parece que éstas tienen valores un tanto mayores que el salario estimado, cuando se rebasa el valor de $40 000 sobre la escala horizontal de la gráfica de residuales.

15.6 Continuando con el problema 15.5, y observando las dos siguientes gráficas de residuales de la figura 15-2, comente sobre el proposito de estudiarlas y sobre los resultados de sus observaciones.

Pueden utilizarse esas gráficas individuales para tratar de identificar la fuente de cualquier problema observado en la gráfica de residuales para el modelo global. Las gráficas de residuales individuales para años de experiencia y años de educación son lineales, pero ambas muestran varíanzas condicionales mayores cuando se eleva el número de años, lo cual identifica la fuente de la posible dificultad que se identificó en la solución del problema 15.5. Aunque se procede a realizar cálculos adicionales en estos problemas resueltos, la posible desigualdad de las varíanzas condicionales serla una consideración importante al interpretar los intervalos de confianza y de predicción.

15.7 Con referencia a la ecuación de regresión múltiple que se basa en las dos variables independientes de la figura 15-2 (que se identifican con ④ en el listado, estime el sueldo anual de una persona que tiene 5.5 años de experiencia y 4.0 años de educación posterior a la secundaria.

$$\begin{aligned}\hat{Y} &= b_0 + b_1X_1 + b_2X_2\\ &= 25\,511 + 826\,X_1 + 1604\,X_2\\ &= 25\,511 + 826(5.5) + 1604(4.0)\\ &= \$\,36\,470\end{aligned}$$

(o, del listado de computadora con base en valores menos redondeados durante los cálculos, e identificados con ⑧ en la figura 15-2, $36 467).

15.8 Interprete el significado de los diversos valores b_j que se incluyen en la ecuación de regresión del problema anterior.

La cantidad b_0 = 25 511 es la ordenada al origen, es decir, el punto sobre el eje *Y,* para el cual $X_1 = 0$ y $X_2 = 0$. Literalmente, sería el salario diario estimado para una persona que no tiene ni experiencia ni educación posterior a la secundaria. Sin embargo, no existe ese significado práctico, porque no se incluyeron esos valores de las variables independientes en el rango de los valores muestreados.

La cantidad b_1 = 826 indica que, en promedio, a un aumento de un año en los años de experiencia le corresponde un aumento de $826 de salario diario, *dado que la variable años de educación también se incluye en el modelo de regresión* y, por lo tanto, queda estadísticamente controlada o "constante".

La cantidad b_2 = 1 604 indica que, en promedio, a un aumento de un año en los años de educación le corresponde un aumento de $1 604 de salario diario, dado que también se incluye en el modelo de regresión la variable años de experiencia.

15.9 Con referencia a la porción apropiada del listado, determine el intervalo de confianza del 95% para el sueldo diario promedio de todos los elementos de la población con 5.5 años de experiencia y 4.0 años de educación posterior a la secundaria.

95% C.I. = $35180 a $37 775 (identificado con ⑨ en el listado)

15.10 Con referencia a la porción pertinente del listado de la figura 15-2, determine el intervalo de predicción del 95% para una persona especifica que tiene 5.5 años de educación y 4.0 años de educación posterior a la secundaria. Compare este intervalo con el que se obtuvo en el problema 15.9 e interprete el significado de este intervalo de predicción.

95% P.I. = $31 548 a $41 386 (identificado con ⑩ **en** el listado)

Tal como se esperaba, este intervalo es más amplio que el intervalo de confianza del problema 15.9. Un intervalo de predicción es un intervalo de probabilidad para un valor *individual* de Y, en vez de serlo para una media. El intervalo de predicción indica que existe una probabilidad de 0.95 que una persona con 5.5 años de experiencia y 4.0 años de educación posterior a la secundaría obtenga un sueldo diario de entre $31 548 y $41 386.

15.11 Con referencia a la tabla de correlación que aparece al final del listado, y que se identifica con (11) (a) ¿qué variable tiene la mayor correlación con el sueldo diario, y cuál es esa correlación? (b) ¿Qué otras dos variables, aparte del salario, tienen la mayor correlación entre si, y cuál es esa correlación?

(Nota: la significancia de estos coeficientes simples de correlación se determinaría usando la prueba t que se describió en la sección 14.13. De hecho, el único valor de correlación simple de la tabla que es significativo a un nivel del 5% es el coeficiente de 0.637.)

(a) La mayor correlación se da entre la experiencia y el sueldo, con $r = 0.637$.

(b) La mayor correlación se da entre la experiencia y el sexo, con $r = 0.352$. Puede observarse en los datos de entrada para la computadora que el código que se utilizó en la variable indicadora de sexo es: mujer = 0 y hombre =1. El hecho de que parezca positiva la correlación entre la experiencia y el sexo indica que los años de experiencia para el código de sexo " 1 " (hombre) tiende a ser mayor que los años de experiencia para el código de sexo "0" (mujer).

15.12 Al utilizar el listado de la figura 15-2, determine el coeficiente de determinación múltiple y el coeficiente de correlación múltiple para el modelo reducido con dos variables independientes. Interprete el coeficiente de determinación y señale la forma en que podría probarse la significación del coeficiente de correlación.

El coeficiente de determinación múltiple es $R^2 = 0.675$ (sin ajuste para los grados de libertad), y se identifica con (12) en el listado. Esto indica que aproximadamente el 67.5% de la varianza de los sueldos de los analistas queda estadísticamente explicado por sus años de experiencia y sus años de educación posterior a la secundaria.

El coeficiente de correlación múltiple es $R = \sqrt{0.675} = 0.82$, que se reporta como valor absoluto. La hipótesis nula de que este coeficiente poblacional de correlación múltiple es igual a 0 se prueba aplicando la prueba *F* al modelo correspondiente de regresión múltiple del problema 15.3, en el que se rechazó la hipótesis nula de que no existe relación entre las variables independientes de experiencia y educación, consideradas en conjunto, con la variable dependiente de salario.

Problemas complementarios

En la Tabla 15.3 se reportan los precios de Instalaciones fabriles, con una muestra aleatoria de 30 instalaciones elegidas al azar, 10 de cada una de 3 subdivisiones. Tal como se observa, además de la subdivisión y del precio, se incluyen los datos de superficie del terreno y la superficie construida. Siendo el precio la variable dependiente, realice un análisis de regresión por pasos hacia atrás utilizando algún programa de computación. También obtenga una gráfica de residuales para el modelo final de regresión y una matriz de correlación de todos los coeficientes de correlación simple, como base para las soluciones de los problemas complementarios que aparecen enseguida. Utilice el esquema del código binario de la sección 15.3 para las dos variables indicadoras que se requieren para codificar la subdivisión a la que corresponde cada una de las instalaciones.

15.13 Obtenga la ecuación de regresión múltiple para estimar el precio de una fábrica con base en las tres variables de tamaño del terreno, superficie construida y ubicación. Pruebe la significancia del modelo de regresión múltiple a un nivel del 5%.

Resp. $\hat{Y} = 40\,462 + 36.0X_1 + 0.94\,X_2 + 4267\,X_3 + 26\,648X_4$. El modelo de regresión es significativo al nivel del 5%.

15.14 Observe cuál de las variables tiene el coeficiente de regresión parcial con el menor, valor de la estadística r, y determine si la contribución de esa variable es significativa a un nivel del 5%. *(Nota:* para la solución, se adopta la posición de que, dado que las dos variables indicadoras representan *una* variable cualitativa de ubicación, o se eliminan ambas o ninguna. Por ello, sólo se considerarla el *mayor* de los dos cocientes í asociados con las variables indicadoras.)

Resp. El tamaño del lote tiene el menor valor de la estadística de prueba (f - 0.44) y se le debe eliminar del modelo.

Tabla 15.3 Precios de instalaciones fabriles en tres subdivisiones

Instalación fabril muestreada	Precio (en miles de millones de pesos)	Área construida (metros cuadrados)	Superficie del terreno (metros cuadrados)	Subdivisión
1	102 200	1500	A	12 000
2	103 950	1200	A	10 000
3	87 900	1200	A	10 000
4	110 000	1600	A	15 000
5	97 000	1400	A	12 000
6	95 700	1200	A	10 000
7	113 600	1600	A	15 000
8	109 600	1500	A	12 000
9	110 800	1500	A	12 000
10	90 600	1300	A	12 000
11	109 000	1600	B	13 000
12	133 000	1900	B	15 000
13	134 000	1800	B	15 000
14	120 300	2000	B	17 000
15	137 000	2000	B	17 000
16	122 400	1700	B	15 000
17	121 700	1800	B	15 000
18	126 000	1900	B	16 000
19	128 000	2000	B	16 000
20	117 500	1600	B	13 000
21	158 700	2400	C	18 000
22	186 800	2600	C	18 000
23	172 400	2300	C	16 000
24	151 200	2200	C	16 000
25	179 100	2800	C	20 000
26	182 300	2700	C	20 000
27	195 850	3000	C	22 000
28	168 000	2400	C	18 000
29	199 400	2500	C	20 000
30	163 000	2400	C	18 000

15.15 Al continuar con el problema 15.14, obtenga la ecuación de regresión múltiple para el modelo reducido y pruebe la significancia del modelo a un nivel del 5%.

Resp. $\hat{Y} = 41\ 153 + 43.6X_1 + 4025X_3 + 24\ 319X_4$. El modelo de regresión es significativo al nivel del 5%.

15.16 Al igual que en el problema 15.14, observe cuál de las variables tiene el coeficiente de regresión parcial con el menor valor reportado de la estadística f, y determine si la contribución de esa variable al modelo de regresión múltiple es significativa a un nivel del 5%.

Resp. El menor valor de la estadística de prueba es para la primera variable indicadora (f = 0.79), el cual se ignora dada la posición que se describió en el problema 15.14 con respecto a las variables indicadoras. La siguiente estadística de prueba de menor magnitud es la correspondiente a la segunda variable indicadora (t = 2.44) y es significativa a un nivel del 5%. Por ello, esta variable no debe eliminarse del modelo.

15.17 Observe la gráfica de residuales para el modelo de regresión múltiple del problema 15.15. ¿Se satisfacen las suposiciones de linealidad e igualdad de varianzas condicionales?

Resp. Las suposiciones se satisfacen en forma razonable. Existe una observación "aberrante" en la porción superior derecha de la gráfica, y debe verificarse su precisión en los datos originales para esa instalación fabril (fábrica número 29).

15.18 Utilice el modelo de regresión del problema 15.15 y estime el precio de una fábrica con (a) 1200 metros cuadrados, en la subdivisión A; (b) 1800 metros cuadrados, en la subdivisión B; y (c) 2400 metros cuadrados en la subdivisión C.

Resp. (a) \$93 423, (b) \$123 583, (c) \$ 170 012

15.19 Utilice el modelo de regresión del problema 15.15 para determinar el intervalo de confianza del 95% para el precio promedio condicional de todas las fábricas con (a) 1200 metros cuadrados, en la subdivisión A, (b) 1800 metros cuadrados, en la subdivisión B, y (c) 2400 metros cuadrados en la subdivisión C.

Resp. (a) \$87118 a\$99 728, (b) \$118 229 a\$128 937, (c) \$164250a\$175 774

15.20 Al continuar con el problema 15.19, determine el Intervalo de predicción del 95% para el precio de una fábrica individual en cada una de las tres categorías que se Identificaron en el problema anterior.

Resp. (a) \$75 428 a \$111 419, (b) \$105 898 a \$141 268, (c) \$152199 a \$187 825

15.21 Con referencia a la matriz de coeficientes de correlación simple para todas las variables que se incluyeron en el estudio, (a) ¿Qué variable tiene la mayor correlación con el precio? (b) ¿Cuál de las dos variables "candidatas" a ser las variables independientes tienen la mayor correlación entre si?

Resp. (a) La superficie de construcción con $r = 0.962$, (b) la superficie de construcción y de lote, con r = 0.961.

15.22 Determine (a) el coeficiente de determinación múltiple (sin corrección por grados de libertad), y (b) el coeficiente de correlación múltiple para el modelo final de regresión múltiple.

Resp. *(a)* $R^2 = 0.945$, *(b)* $R = \sqrt{0.945} \approx \mathbf{0.972}$

Análisis de series de tiempo y pronósticos de negocios

16

16.1 EL MODELO CLÁSICO DE SERIES DE TIEMPO

Una *serie de tiempo* es un conjunto de valores observados, tales como datos de producción o ventas, para series ordenadas secuencialmente de periodos de tiempo. Algunos ejemplos de estos datos son las ventas de un producto determinado para un conjunto de meses, y el número de trabajadores que laboran en una industria determinada durante varios años. Las series de tiempo se ilustran mediante gráficas de lineas (sección 2.8), y los periodos de tiempo se representan sobre el eje horizontal, mientras que en el vertical se representa la serie de valores.

EJEMPLO 1. En la figura 16-1 se presenta una gráfica de linea que ilustra las ventas anuales de una compañía (ficticia) que comenzó a operar en 1980. Como puede observarse, se alcanzó una *cumbre* en las ventas anuales en 1985, la cual fue seguida por dos años de menores ventas, que culminaron en el valle de 1987, la cual fue de nuevo seguida por niveles crecientes de ventas durante los últimos tres años de la serie de tiempo que se reporta.

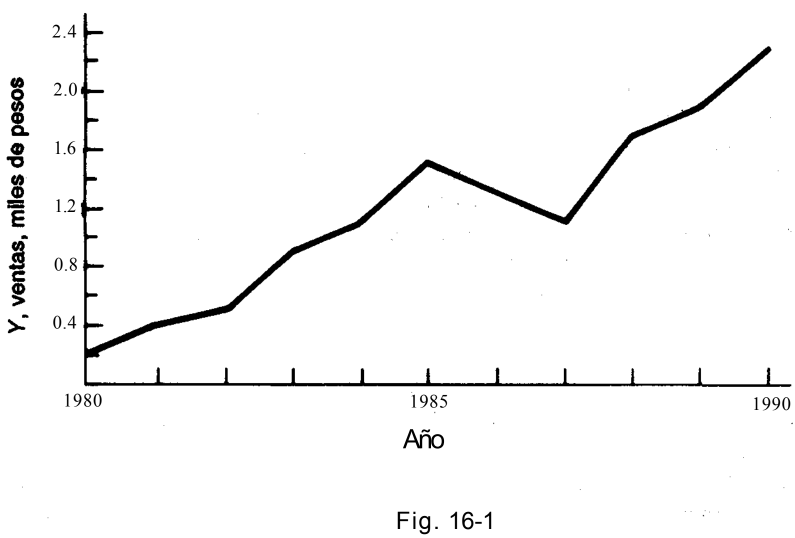


Fig. 16-1

El *análisis de series de tiempo* es el procedimiento mediante el cual se identifican y separan los factores relacionados con el tiempo que influyen sobre los valores observados de la serie. Una vez que se identifican esos valores, se les puede utilizar para mejorar la interpretación de los valores históricos de la serie de tiempo y para pronosticar valores futuros. El enfoque clásico al análisis de series de tiempo identifica cuatro de esos efectos, o *componentes:*

(1) *Tendencia* (T): el movimiento global a largo plazo de los valores de la serie de tiempo *(Y)* durante un número prolongado de años.

(2) *Fluctuaciones cíclicas:* (C): movimientos recurrentes hacia arriba y hacia abajo con respecto a la tendencia y que tienen duración de varios años.

(3) *Variaciones estacionales:* (E): movimientos hacia arriba y hacia abajo con respecto a la tendencia y que no duran más de un año y que, además, se presentan todos los años. Es común que se identifiquen esas variaciones con base en datos mensuales o trimestrales.

(4) *Variaciones irregulares (I)*: las variaciones erráticas con respecto a la tendencia, que no pueden adjudicarse a efectos estacionales o cíclicos.

El modelo en el que se basa el análisis clásico de seríes de tiempo se apoya en la suposición de que, para cualquier periodo de la serie de tiempo, el valor de la variable está determinado por el efecto de los cuatro componentes que se definieron antes y, además, que los componentes tienen una relación multiplicativa. Por ello, *Y* representa el valor observado de la serie de tiempo,

$$Y = T \times C \times E \times I \tag{16.1}$$

El modelo que se representa mediante la fórmula *(16.1)* se utiliza como base para separar los efectos de los diversos componentes que influyen sobre los valores de la serie de tiempo, según se describe en las secciones siguientes de este capitulo.

16.2 ANÁLISIS DE TENDENCIA

Como el análisis de tendencia se ocupa de la dirección del movimiento de la serie de tiempo a largo plazo, es común que esos análisis se lleven a cabo analizando datos anuales. Por lo general, se deben utilizar datos de cuando menos 15 o 20 años, para no Incluir como señal de la tendencia global de la serie de tiempo los movimientos cíclicos que implican pocos años de duración.

El método de mínimos cuadrados (sección 14.3) es la base más común que se utiliza para identificar el componente de tendencia de la serie de tiempo, determinando la ecuación que mejor se ajuste a la línea de tendencia. Debe observarse que, en términos estadísticos, una línea de tendencia no es una linea de regresión porque la variable dependiente *Y* no es una variable aleatoria sino que, más bien, es un valor histórico acumulado. Además, sólo puede haber un valor histórico para cualquier periodo de tiempo determinado (no una distribución de valores) y los valores asociados con períodos de tiempo adyacentes son dependientes. No obstante, el método de mínimos cuadrados es una base conveniente para determinar el componente de la tendencia de una serie de tiempo. Cuando parece que el aumento o la disminución a largo plazo sigue una tendencia lineal, la ecuación para los valores de la linea de tendencia, utilizando X para representar el año, es

$$Y_T = b_0 + b_1 X \tag{16.2}$$

Como se explicaba en la sección 14.3, el componente b_0 de la fórmula *(16.2)* representa el punto de intersección de la línea de tendencia con el eje *Y,* mientras que b_1 representa la pendiente de la línea de tendencia. Utilizando X para representar el año Y como el valor observado de la serie de tiempo, las fórmulas para determinar los valores de b_0 y b_1, para la ecuación de la linea de tendencia son

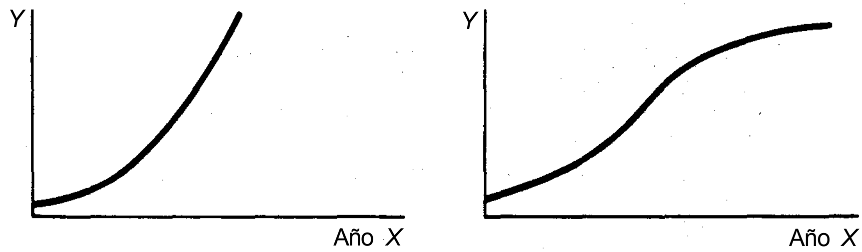
$$b_1 = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} \tag{16.3}$$

$$b_0 = \bar{Y} - b_1\bar{X}$$

En el problema 16.1 puede observarse el cálculo de la ecuación de tendencia lineal.

En el caso de tendencias no lineales, con frecuencia se utilizan dos tipos de curvas de tendencia para realizar el análisis: la curva de tendencia exponencial y la curva de tendencia parabólica. Una curva de *tendencia exponencial* típica es aquella que refleja una tasa constante de crecimiento durante un periodo de años, y que podría aplicarse a las ventas de computadoras personales durante 1980. Véase la figura 16-2 (a). A las curvas exponenciales se les denomina de esa manera porque la variable independiente X es el exponente de b_1 en la ecuación general:

$$Y_T = b_0 b_1^X \tag{16.5}$$



(a) Tendencia exponencial *(b)* Curva de Gompertz

Fig. 16-2

Al aplicar el logaritmo en ambos lados de *(16.5)* se obtiene una ecuación lineal de tendencia logarítmica,

$$\log Y_T = \log b_0 + X \log b_1 \tag{16.6}$$

La ventaja de la transformación en logaritmos es que la ecuación lineal para el análisis de tendencia puede aplicarse a los logaritmos de los valores cuando la serie de tiempo sigue una curva exponencial. Después, es posible reconvertir los valores pronosticados en logaritmos de Y_T, a sus unidades originales de medición, sacando el antilogaritmo de los valores. No se ilustra este tipo de análisis en este libro.

Puede observarse que muchas series de tiempo que representan las ventas de productos presentan tres etapas: una *etapa introductoria* de crecimiento lento en las ventas, una *etapa intermedia* de rápidos aumentos en las ventas y una *etapa final* de crecimiento lento al saturarse el mercado. Este conjunto de tres etapas puede abarcar muchos años para algunos productos, tales como acero estructural. Para otros productos, tales como radios de banda civil, la etapa de saturación puede alcanzarse con relativa rapidez. La curva de tendencia específica que incluye las tres etapas que se acaban de describir es la *curva de Gompertz,* que se ilustra en la figura 16-2 (b). La ecuación para la curva de tendencia de Gompertz es

$$Y_T = b_0 b_1^{(b_2)^X} \tag{16.7}$$

Los valores de b_0, b_1 y b_2 se determinan obteniendo en primer lugar el logaritmo de ambos lados de la ecuación, de la siguiente manera

$$\log Y_T = \log b_0 + (\log b_1) b_2^X \tag{16.8}$$

Finalmente, los valores de la curva de tendencia se calculan obteniendo los antilogaritmos de ios valores calculados mediante la fórmula *(16.8).* Los detalles de esos cálculos se incluyen en libros especializados en análisis de series de tiempo.

16.3 ANÁLISIS DE VARIACIONES CÍCLICAS

Los valores anuales de una serie de tiempo representan únicamente los efectos de los componentes de tendencia y cíclicos, porque se definen los componentes estacional e irregular como efectos a corto plazo. Por ello, puede identificarse el componente cíclico en datos anuales dividiendo los valores observados entre el valor correspondiente de la tendencia, de la siguiente manera:

$$\frac{Y}{Y_T} = \frac{T \times C}{T} = C \qquad (16.9)$$

El cociente que se obtiene de *(16.9)* se multiplica por 100 para que la media cíclica relativa sea de 100.0. Una media cíclica relativa de 100 indicaría la ausencia de efectos cíclicos sobre el valor anual de la serie de tiempo. Véase el problema 16.2.

Para auxiliar en la interpretación de relativos cíclicos, con frecuencia se prepara una *"tabla de ciclos"* que ilustra los relativos cíclicos de acuerdo con el año. Construyendo una tabla como esta, pueden realzarse las cumbres y los valles correspondientes al componente cíclico de la serie de tiempo. Véase el problema 16.3.

16.4 MEDICIÓN DE VARIACIONES ESTACIONALES

El efecto del componente estacional sobre los valores de la serie de tiempo se identifica determinando el número índice estacional correspondiente a cada mes (o trimestre) del año. La media aritmética de los doce números índices mensuales (o cuatro números trimestrales) es 100. La identificación de influencias estacionales positivas y negativas es importante para la planeación de la producción y los inventarios.

EJEMPLO 2. Un número índice de 110 para un mes determinado indica que los valores de la serie de tiempo para ese mes han promediado 10% por encima de los otros meses, debido a algún factor estacional positivo. Por ejemplo, pudieran haber aumentado las ventas de junio de rasuradoras eléctricas en 10%, en comparación con otros meses, debido al Día del Padre.

El procedimiento que se utiliza con mayor frecuencia para determinar los números índices estacionales es el *método del cociente del promedio móvil.* Con este método se determina, en primer lugar, el cociente que resulta de dividir el valor de cada mes entre el promedio móvil centrado en ese mes. Como un promedio móvil que se basa en datos mensuales (o trimestrales) de un año completo, "equilibraría" las fluctuaciones irregulares y estacionales, pero no la tendencia a largo plazo ni los efectos cíclicos, el cociente del valor mensual (o trimestral) con respecto a un promedio móvil puede representarse en símbolos de la siguiente manera:

$$\frac{Y}{\text{Promedio móvil}} = \frac{T \times C \times E \times I}{T \times C} = S \times I \qquad (16.10)$$

El segundo paso del método del cociente del promedio móvil, consiste en promediar el componente irregular. Por lo general, esto se realiza enlistando los diversos cocientes aplicables al mismo mes (o trimestre) para los diversos años, eliminando los valores mayor y menor, y calculando el promedio de los cocientes restantes. A esta media se le denomina *media modificada,* debido a la eliminación de los dos valores extremos.

El paso final en el método del cociente del promedio móvil consiste en ajustar las razones promedio modificadas mediante un factor de corrección, para que la suma de los 12 cocientes mensuales sea 1200 o (400 para cocientes trimestrales). Véase el problema 16.4.

16.5 APLICACIÓN DE AJUSTES ESTACIONALES

Una aplicación frecuente de los Índices estacionales consiste en ajustar datos observados de seríes de tiempo, eliminando la influencia del componente estacional de los datos. A esos datos estacionales se les denomina *datos ajustados estacionalmente, o datos desestadonalizados.* Los ajustes estacionales son de particular importancia cuando se desea comparar datos de diferentes meses con el objeto de determinar si se ha presentado un aumento (o disminución) con respecto a las expectativas estacionales.

EJEMPLO 3. Un aumento del 10% de abril a mayo en el precio de fertilizante doméstico en un año determinado representa una *disminución* relativa si el número índice estacional para mayo está 20% encima del número Índice correspondiente a abril. En otras palabras, si se presenta un aumento, pero no es tan grande como se esperaba de acuerdo con los datos históricos, entonces se presenta una disminución en la demanda con respecto a esas expectativas.

Se ajustan los efectos estacionales de los valores mensuales (o trimestrales) observados de una serie de tiempo dividiendo a cada uno de los valores entre el índice mensual (o trimestral) correspondiente a ese mes. Después, se multiplica el resultado por 100 para mantener la posición decimal de los datos originales. El proceso de ajustar las variaciones estacionales de los datos puede representarse de la siguiente manera:

$$\frac{Y}{S} = \frac{T \times C \times E \times I}{S} = T \times C \times I \tag{16.ll}$$

Aunque los valores resultantes después de aplicar la fórmula *(16.11)* se encuentran en las mismas unidades de medición que los datos originales, no representan ocurrencias reales. Más bien, son valores relativos y sólo tienen sentido para propósitos de comparación. Véase el problema 16.5.

16.6 PRONÓSTICOS CON BASE EN FACTORES DE TENDENCIA Y ESTACIONALES

La ecuación de la linea de tendencia ofrece un punto de partida para realizar pronósticos a largo plazo sobre valores anuales. Sin embargo, una consideración de particular importancia en los pronósticos a largo plazo es el componente cíclico de la serie de tiempo. No existe método estándar para pronosticar el componente cíclico con base únicamente en los valores históricos de la serie de tiempo, pero existen diversos indicadores económicos (sección 16.7) que son útiles para anticipar puntos cíclicos de cambio.

Para los pronósticos a corto plazo, el punto de partida es el valor de tendencia que se proyecta y que después se ajusta en su componente estacional. Debido a que, por lo general, la ecuación de la linea de tendencia se basa en el análisis de valores anuales, el primer paso que se requiere es "simplificar" esa ecuación para que quede expresada en términos de meses (o trimestres). Se modifica una ecuación de tendencia, con fechas anuales, para obtener valores proyectados mensuales, de la siguiente manera:

$$Y_T = \frac{b_0}{12} + \left(\frac{b_1}{12}\right)\left(\frac{X}{12}\right) = \frac{b_0}{12} + \frac{b_1}{144}X \tag{16.12}$$

Para obtener valores trimestrales proyectados, se modifica una ecuación de tendencia con datos anuales, de la siguiente manera:

$$Y_T = \frac{b_0}{4} + \left(\frac{b_1}{4}\right)\left(\frac{X}{4}\right) = \frac{b_0}{4} + \frac{b_1}{16}X \tag{16.13}$$

La base de las modificaciones anteriores no resulta evidente si se pasa por alto el hecho de que los valores de tendencia no están asociados con puntos en el tiempo, sino más bien, con periodos de tiempo. Debido a esta consideración, es necesario separar los *tres* elementos de la ecuación de la tendencia anual (b_0, b_1 y X).

Después de la transformación para los datos mensuales de la fórmula *(16.12*), el punto base del año que antes se identificaba como X = 0, se encontraría a mediados del año, o sea el primero de julio. Como es necesario que el punto base se encuentre a la mitad del primer mes del año base, 15 de enero, se modifica la ordenada al origen $b_0/12$, en la ecuación modificada, y después se reduce en 5.5 veces la pendiente modificada. Se lleva a cabo un ajuste similar para los datos trimestrales. Por ello, una ecuación de tendencia modificada para obtener valores mensuales *y* con X = 0, teniendo el 15 de enero como año base, es

$$Y_T = \frac{b_0}{12} - (5.5)\left(\frac{b_1}{144}\right) + \frac{b_1}{144}X \qquad (16.14)$$

De manera similar, una ecuación de tendencia modificada para obtener valores proyectados trimestrales *y* con X = *0,* enmedio del primer trimestre del año base, es:

$$Y_T = \frac{b_0}{4} - (1.5)\left(\frac{b_1}{16}\right) + \frac{b_1}{16}X \qquad (16.15)$$

En el problema 16.6 se ilustra el proceso de simplificar una ecuación de tendencia. Después que se determinan los valores mensuales (o trimestrales) de la tendencia, puede multiplicarse cada uno de esos valores por el índice estacional apropiado (y dividir entre 100 para conservar la ubicación decimal de los valores) para establecer un punto inicial que permita realizar pronósticos a corto plazo. Véase el problema 16.7.

16.7 PRONÓSTICOS CÍCLICOS E INDICADORES DE NEGOCIOS

Tal como se indicaba en la sección 16.6, los pronósticos que se basan en los componentes de tendencia o estacional de una serie de tiempo se consideran sólo como el punto inicial de los pronósticos económicos. Una razón de esto es la necesidad de considerar el efecto probable del componente cíclico durante el periodo pronosticado, en tanto que una segunda razón es la importancia de identificar los factores causales específicos que han influido sobre las variables de la serie de tiempo.

Para los pronósticos a corto plazo, con frecuencia se supone que el efecto del componente cíclico es igual que el que se ha dado en los valores recientes de la serie de tiempo. Sin embargo, para periodos más prolongados, o aun para períodos breves de inestabilidad económica, resulta importante identificar los *puntos cíclicos de cambio* de la economía nacional. Por supuesto, las variaciones cíclicas asociadas con un producto determinado pueden o no coincidir con el ciclo general de negocios.

EJEMPLO 4. Históricamente, las ventas de automóviles han coincidido en forma estrecha con el ciclo general de negocios de la economía nacional. Por otro lado, las ventas de refacciones para automóviles tienden a ser contracíclícas, con respecto al ciclo global de negocios.

Se han identificado diversas seríes de tiempo que, históricamente, han resultado ser indicadores de surgimientos y recesiones cíclicas con respecto al ciclo global de negocios. Un grupo de estos, a los que se denomina *indicadores líder,* por lo general llegan a los puntos cíclicos de cambio antes del cambio correspondiente en la actividad económica general. Los indicadores líder incluyen medidas como la tasa de desempleo en manufactura, el valor de los pedidos nuevos en las industrias de bienes duraderos, y un índice de precios y cotizaciones del mercado bursátil. Un segundo grupo, al que se denomina *Indicadores coincidentes,* son series de tiempo que por lo general han tenido puntos de cambio que coinciden con el ciclo general de negocios. Los indicadores coincidentes incluyen medidas como la tasa de desempleo y el índice de producción industrial. El tercer grupo, al que se denomina *indicadores rezagados,* son las seríes de tiempo para tas cuales las cumbres y los valles generalmente se retrasan con respecto al ciclo general de negocios. Los indicadores rezagados incluyen medidas como la manufactura y los inventarios comerciales, y las tasas preferenciales promedio que indican los bancos.

Además de considerar el efecto de las fluctuaciones cíclicas y de pronosticar esas fluctuaciones, también deben estudiarse las variables causales específicas que históricamente han influido sobre los valores de la serie de tiempo.

El análisis de regresión y de correlación (capítulos 14 y 15) son particularmente aplicables a estudios como la relación entre la estrategia del precio y el volumen de ventas. Aparte de los análisis históricos, otras áreas que requieren atención son las posibles implicaciones de productos nuevos y los cambios en el ambiente del mercado.

16.8 LA SUAVIZACIÓN EXPONENCIAL COMO MÉTODO DE PRONÓSTICO

La *suavización exponencial* es un método de pronóstico que no se basa en ei análisis de los componentes históricos de la serie de tiempo como tales. Más bien, utiliza como pronóstico un promedio móvil ponderado, en donde los pesos que se asignan se reducen en forma exponencial conforme mayor es la antigüedad de los periodos. Existen, de hecho, diversos tipos de modelos de suavización exponencial, tal como se describe en los libros especializados en pronósticos de negocios. El método que se presenta en esta sección es el de *suavización exponencial simple.*

El siguiente modelo algebraico sirve para representar la forma en que se determinan los pesos exponencialmente decrecientes. En especifico, en donde a es una constante de suavización que se analiza más adelante, el valor más reciente de la serle de tiempo se pondera con a, el siguiente valor más reciente se pondera con $\alpha(1-\alpha)$, el siguiente valor con $\alpha(1-\alpha)^2$, etc; y después se suman todos los valores ponderados para determinar el pronóstico:

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha) Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \cdots + \alpha(1-\alpha)^k Y_{t-k} \qquad (16.16)$$

en donde $\hat{Y}_{t+1}$ = pronóstico para el siguiente periodo
α = constante de suavización $(0 \leq \alpha \leq 1)$
Yt = valor real para el periodo más reciente
Y t =1 = valor real para el periodo que precede al más reciente
Yt=k = valor real para los *k* periodos precedentes al periodo más reciente

Aunque la tabla anterior sirve para presentar el razonamiento que subyace a la suavización exponencial, su uso es bastante laborioso. Por ello, se utiliza un procedimiento simplificado que requiere de un pronóstico "semilla" inicial, pero que no exige la determinación de los pesos. La fórmula para determinar el pronóstico mediante el método simplificado de suavización exponencial es

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) \qquad (16.17)$$

en donde $\hat{Y}_{t+1}$ = pronóstico para el siguiente periodo
$\hat{Y}_t$ = pronóstico para el periodo más reciente
α = constante de suavización $(0 \leq \alpha \leq 1)$
Y_t = valor real para el periodo más reciente

Como el valor más reciente de la serie de tiempo debe estar disponible para determinar un pronóstico para el siguiente periodo, sólo puede utilizarse la suavización exponencial para pronosticar el valor del *siguiente* periodo de la serie y no para diversos periodos a futuro. Conforme más cercano a 1.0 se fije el valor de la constante de suavización, más ponderado estará el pronóstico con los resultados más recientes. Véase el problema 16.8.

La suavización exponencial simple es más efectiva como método de pronóstico cuando las influencias cíclica y regular representan los principales efectos sobre los valores de una serie de tiempo. Cuando existe una tendencia lineal definida, la suavizadón exponencial doble produce un mejor pronóstico. La suavización exponencial triple es apropiada en el caso de una tendencia no lineal definida, al tiempo que pueden utilizarse otros métodos que se describen en textos avanzados de pronósticos, para incorporar la existencia de una influencia estacional poderosa.

16.9 RESULTADOS POR COMPUTADORA

Existe una amplia gama de programas de computación para realizar análisis de series de tiempo y para efectuar pronósticos de negocios, y muchos de ellos incluyen técnicas más elaboradas de las que pueden incluirse en la explicación limitada que se da en este libro sobre este tema especializado. En el problema 16.9 se ilustra el uso de Minitab para determinar la ecuación de tendencia lineal, mientras que en el problema 16.10 se ilustra el uso del programa denominado SEASON para determinar limites estacionales y para ajustar estacionalmente los datos de entrada.

Problemas resueltos

ANÁLISIS DE TENDENCIA

16.1 En la Tabla 16.1 se presentan los datos de ventas para un periodo de 11 años de una compañía ficticia, que empezó a operar en 1980, según se describió en el ejemplo 1. Los datos de esta serie de tiempo se ilustran mediante la gráfica de línea de la figura 16-1. También se incluyen los cálculos necesarios para determinar la ecuación de la tendencia lineal. (a) Determine la ecuación de tendencia lineal para esos datos mediante el método de mínimos cuadrados, codificando 1980 como 0, y llevando todos los valores con dos cifras decimales. *(b)* Dibuje la línea de tendencia sobre la gráfica de línea de la figura 16.1.

Tabla 16.1 Ventas anuales (en miles de pesos) de una empresa que vende equipo de computación, incluida la hoja de trabajo para determinar la ecuación de la tendencia lineal

Año	Año codificado (X)	Ventas, en miles (Y)	XY	X^2
1980	0	\$ 0.2	0	0
1981	1	0.4	0.4	1
1982	2	0.5	1.0	4
1983	3	0.9	2.7	9
1984	4	1.1	4.4	16
1985	5	1.5	7.5	25
1986	6	1.3	7.8	36
1987	7	1.1	7.7	49
1988	8	1.7	13.6	64
1989	9	1.9	17.1	81
1990	10	2.3	23.0	100
Totales	55	12.9	85.2	385

(a) $$Y_T = b_0 + b_1 X$$

en donde $\bar{X} = \frac{\Sigma X}{n} = \frac{55}{11} = 5.00$

$$\bar{Y} = \frac{\Sigma Y}{n} = \frac{12.9}{11} = 1.17$$

$$b_1 = \frac{\Sigma XY - n\bar{X}\bar{Y}}{\Sigma X^2 - n\bar{X}^2} = \frac{85.2 - 11(5.00)(1.17)}{385 - 11(5.00)^2}$$

$$= \frac{20.85}{110} = 0.19$$

$$b_0 = \bar{Y} - b_1\bar{X} = 1.17 - 0.19(5.00) = 0.22$$

$$Y_T = 0.22 + 0.19X \text{ (con } X = 0 \text{ en 1980)}$$

Puede utilizarse esta ecuación como punto de partida para los pronósticos, como se describió en la sección 16.6. La pendiente de 0.19 indica que en los 11 años de existencia de la compañía, ha habido un aumento promedio en ventas de 0.19 millones de pesos ($190 000) anuales.

(b) La figura 16-3 representala gráfica de linea de la figura 16-1, pero habiendo dibujado también la línea de tendencia. La cumbre y el valle de la serie de tiempo que se analizaron brevemente en el ejemplo 1 resultan ahora visibles más claramente.

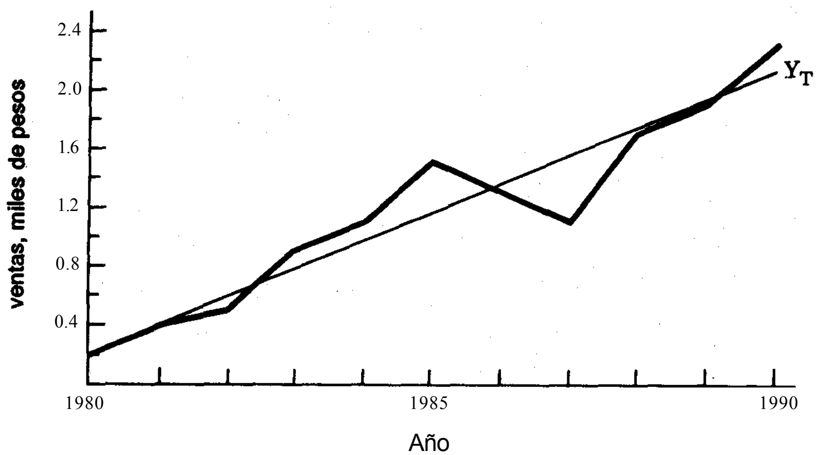


Fig. 16-3

ANÁLISIS DE VARIACIONES CÍCLICAS

16.2 Determine el componente cíclico de cada uno de los valores de la serie de tiempo que se reporta en la Tabla 16.1, utilizando la ecuación de tendencia del problema 16.1.

En la Tabla 16.2 se presentan los cálculos de los relativos cíclicos. Como se indica en la última columna de la tabla, se determina cada uno de los relativos cíclicos multiplicando el valor observado en la serie de tiempo por 100 y dividiendo entre el valor de la tendencia. Por ello, se calculó el relativo cíclico de 90.00 para 1980, evaluando: 100(0.20)/0.22.

Tabla 16.2 Cálculo de los relativos cíclicos

Año	Año codificado (X)	Ventas, en miles de pesos Real (Y)	Esperada (X)	Relativo cíclico 100 $Y/\ Y$
1980	0	0.20	0.22	90.9
1981	1	0.40	0.41	97.6
1982	2	0.50	0.60	83.3
1983	3	0.90	0.79	113.9
1984	4	1.10	0.98	112.2
1985	5	1.50	1.17	128.2
1986	6	1.30	1.36	95.6
1987	7	1.10	1.55	71.0
1988	8	1.70	1.74	97.7
1989	9	1.90	1.93	98.4
1990	10	2.30	2.12	108.5

16.3 Construya una tabla de ciclos para los datos de ventas que se reportan en la Tabla 16.1, y que se basan en los relativos cíclicos determinados en la Tabla 16.2.

La tabla de ciclos se presenta en la figura 16-4 e incluye la cumbre y el valle que se observaron previamente en la figura 16-3.

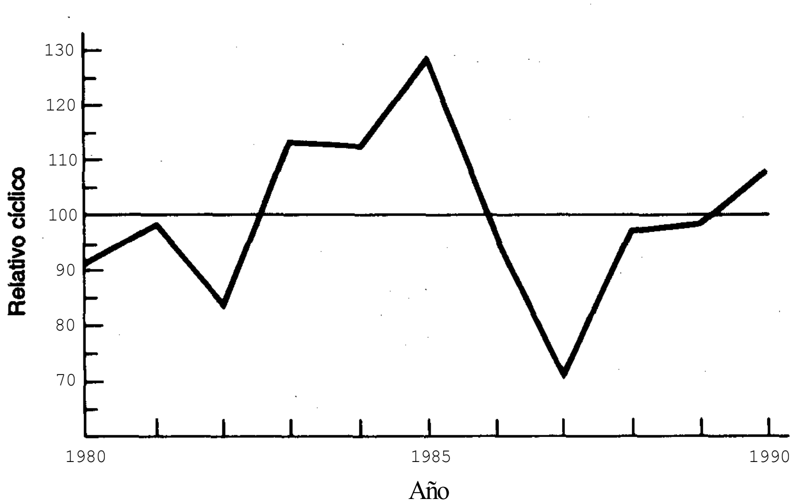


Fig. 16-4

MEDICIÓN DE VARIACIONES ESTACIONALES

16.4 En la Tabla 16.3 se presentan los datos trimestrales de ventas para la compañía ficticia cuyos datos anuales se reportan en la Tabla 16.1. Determine los índices estacionales mediante el método del cociente a promedio móvil.

Tabla 16.3 Ventas trimestrales para la empresa de computación

Trimestre	Ventas trimestrales, en millones de pesos					
	1985	1986	1987	1988	1989	1990
1	500	450	350	550	550	750
2	350	350	200	350	400	500
3	250	200	150	250	350	400
4	400	300	400	550	600	650

En la Tabla 16.4 se presenta el primer paso del método del cociente de promedios móviles, y que consiste en calcular el cociente de cada valor trimestral con respecto al promedio móvil de cuatro trimestres, centrado en ese trimestre.

Los totales móviles de cuatro trimestres se centran entre los trimestres de la tabla, porque, en su calidad de total móvil para un número par de trimestres, el total caerla siempre entre los dos trimestres. Por ejemplo, el primer total que se enlista, 1500 (en miles de pesos) es el monto total de las ventas del primero al cuarto trimestre de 1985. Como se trata de trimestres, el total se centra en forma vertical en la tabla, entre los trimestres segundo y tercero.

Como se desea que el promedio móvil esté centrado en cada trimestre, en vez de hacerlo entre trimestres, los totales móviles de cuatro trimestres que se encuentran juntos se combinan para formar los totales móviles centrados a dos años. Debe observarse que este tipo de total *no* incluye dos años de datos como tales. Más bien, se incluyen en el total dos periodos de cuatro trimestres que se traslapan. Por ejemplo, el primer total móvil centrado de 2950 (en miles) incluye el total de los cuatro trimestres de 1985 (1500) más el total de 1450 correspondiente al segundo trimestre de 1985 hasta el primer trimestre de 1986.

El promedio móvil centrado de cuatro trimestres es simplemente el Total móvil centrado de dos años dividido entre 8.

Finalmente, la razón a promedio móvil de la última columna de la Tabla 16.4 es el cociente de cada valor trimestral de ventas dividido entre el promedio móvil centrado de ese trimestre. Este cociente se multiplica por 100 y, de hecho, se reporta como porcentaje.

En la Tabla 16.5 se incluyen los pasos segundo y tercero necesarios para determinar los índices estacionales. La media modificada para cada uno de los trimestres es el promedio de los tres cocientes restantes para cada trimestre, *después* de eliminar los cocientes mayor y menor. Por ejemplo, para el trimestre 1, los dos cocientes extremos (122.2 y 146.7) se eliminan del análisis, y se tiene que el promedio de los tres cocientes restantes es 132.6.

Finalmente, se multiplican las medias modificadas por un factor de ajuste, para que la suma de los índices sea aproximadamente 400 (1200 para Índices mensuales). El factor de ajuste que se utiliza con los índices trimestrales promedio modificados, con el objeto de obtener índices trimestrales, es

$$\text{factor trimestral de ajuste} = \frac{400}{\text{suma de medias trimestrales}} \qquad (16.18)$$

El factor de ajuste que se usa con los índices mensuales promedio modificados, para obtener índices mensuales, es

Tabla 16.4 Tabla de trabajo para determinar los cocientes de promedios móviles para los datos trimestrales de ventas (en millones de pesos)

Año	Trimestre	Ventas	Total móvil de cuatro trimestres	Total móvil centrado de cuatro trimestres	Promedio móvil centrado de cuatro trimestres	Cociente del promedio móvil (porcentaje)
1985	I	500				
	II	350				
			1500			
	III	250		2 950	368.75	67.8
			1 450			
	IV	400		2 900	362.50	110.3
			1450			
1986	I	450		2 850	356.25	126.3
			1 400			
	II	350		2 700	337.50	103.7
			1300			
	III	200		2 500	312.50	64.0
			1200			
	IV	300		2 250	281.25	106.7
			1050			
1987	I	350		2 050	256.25	136.6
			1 000			
	II	200		2 100	262.50	76.2
			1 100			
	II	150		2 400	300.00	50.0
			1300			
	IV	400		2 750	343.75	116.4
			1450			
1988	I	550		3000	375.00	146.7
			1550			
	II	350		3 250	406.25	86.2
			1 700			
	III	250		3 400	425.00	58.8
			1700			
	IV	550		3 450	431.25	127.5
			1750			
1989	I	550		3 600	450.00	122.2
			1850			
	II	400		3 750	468.75	85.3
			1900			
	III	350		4000	500.00	70.0
			2 100			
	IV	600		4 300	537.50	111.6
			2 200			
1990	I	750		4 450	556.25	134.8
			2 250			
	II	500		4 550	568.75	87.9
			2 300			
	II	400				
	IV	650				

$$\text{factor mensual de ajuste} = \frac{1200}{\text{suma de medias mensuales}} \tag{16.19}$$

Para los cocientes de la Tabla 16.5, el factor de ajuste es 400/395.4 o 1.0116. Cada una de las medias modificadas se multiplica por este factor de ajuste para obtener índices trimestrales. En términos generales, puede observarse que el trimestre que tiene el mayor efecto estacional "positivo" es el trimestre 1, en el que las ventas están por lo general 34.1% arriba del trimestre típico. Por otro lado, por lo general las ventas del trimestre 3 son solamente de 64.2% de las ventas del trimestre típico. Al tener conciencia de esas variaciones estacionales, las empresas las pueden tomar en consideración para planear programas de trabajo y para realizar pronósticos de ventas.

Tabla 16.5 Cálculos de los índices estacionales para los datos trimestrales

Trimestre	1985	1986	1987	1988	1989	1990	Media modificada	índice estacional: Media x 1.0116*
1		126.3	136.6	146.7	122.2	134.8	132.6	134.1
2		103.7	76.2	86.2	85.3	87.9	86.5	87.5
3	67.8	64.0	50.0	58.8	70.0		63.5	64.2
4	110.3	106.7	116.4	127.5	111.6		112.8	114.1
							395.4	399.9

*Factor de ajuste = 400/395.4 = 1.0116

APLICACIÓN DE AJUSTES ESTACIONALES

16.5 Con referencia a los índices estacionales que se determinaron en la Tabla 16.5,

(a) Calcule el valor ajustado estacionalmente para cada trimestre, redondeando los valores a millares de pesos.

(b) Compare los resultados de los trimestres tercero y cuarto de 1990, con base en *(1)* el nivel reportado de las ventas reales, y (2) los valores con ajuste estacional.

(a) Los valores con ajuste estacional se reportan en la Tabla 16.6. Cada uno de los valores desestacionalizados se determinó dividiendo el valor trimestral de la Tabla 16.3 entre el índice estacional para ese trimestre (de la Tabla 16.5) y multiplicando por 100 para conservar la ubicación del punto decimal. Por ejemplo, el valor con ajuste estacional de 373 (en miles) para el trimestre I de 1985, se obtuvo dividiendo 500 (de la Tabla 16.3) entre 134.1 (de la Tabla 16.5) y multiplicando por 100.

(b) Los niveles reportados de ventas para los trimestres tercero y cuarto indican un aumento considerable en las ventas entre esos dos trimestres; de $400 000 a $ 650 000. Sin embargo, con base en los valores ajustados estacionalmente, existe una reducción de 8.5%, de $623 000 a $570 000. Por ello, aunque las ventas aumentaron en el cuarto trimestre, el aumento no fue tan grande como debiera haber sido, de acuerdo con las diferentes influencias históricas de esos dos trimestres.

Tabla 16.6 Valores ajustados estacionalmente para los datos trimestrales

Trimestre	1985	1986	1987	1988	1989	1990
1	\$373	\$336	\$261	\$410	\$410	\$559
2	400	400	229	400	457	571
3	389	312	234	389	545	623
4	351	263	351	482	526	570

PRONÓSTICOS CON BASE EN FACTORES ESTACIONALES Y DE TENDENCIA

16.6 Simplifique la ecuación de tendencia que se elaboró en el problema 16.1 para que quede expresada en términos de trimestres y no de años. También, ajuste la ecuación para que los valores de la tendencia queden en millares de pesos y no en millones. Trabaje los valores finales con una cifra decimal.

$$\begin{aligned} Y_T \text{ (trimestral)} &= \frac{b_0}{4} - 1.5\left(\frac{b_1}{16}\right) + \left(\frac{b_1}{16}\right) X \\ &= \frac{0.22}{4} - 1.5\left(\frac{0.19}{16}\right) + \left(\frac{0.19}{16}\right) X \\ &= 0.0550 - 0.0178 + 0.0119X \\ &= 0.0372 + 0.0119X \end{aligned}$$

La ecuación trimestral de tendencia, en unidades de miles de pesos, es:

$$Y_T \text{ (trimestral)} = 1000\,(0.0372 + 0.0119X) = 37.2 + 11.9\,X$$

Para la ecuación anterior, X - 0 en el primer trimestre de 1990.

16.7 Pronostique el nivel de las ventas trimestrales para cada uno de los trimestres de 1991, con base en la ecuación de tendencia trimestral que se determinó en el problema 16.6, y con ios Índices estacionales calculados en la Tabla 16.5.

Los valores pronosticados con base en la ecuación trimestral de tendencia, y después ajustados con los índices estacionales trimestrales, son:

$$\text{Primer trimestre, 1991} = [37.2 + 11.9(44)] \times \frac{134.1}{100} = 752.0$$

$$\text{Segundo trimestre, 1991} = [37.2 + 11.9(45)] \times \frac{87.5}{100} = 501.1$$

$$\text{Tercer trimestre, 1991} = [37.2 + 11.9(46)] \times \frac{64.2}{100} = 375.3$$

$$\text{Cuarto trimestre, 1991} = [37.2 + 11.9(47)] \times \frac{114.1}{100} = 680.6$$

LA SUAVIZACIÓN EXPONENCIAL COMO MÉTODO DE PRONÓSTICO

16.8 Con referencia a los datos anuales de la serie de tiempo de la Tabla 16.1, y utilizando el nivel real de ventas de 1984 (1.1 millón de pesos) como el pronóstico "semilla" de 1985, determine el pronóstico de las ventas anuales utilizando el método de la suavización exponencial simple, y redondeando el pronóstico a una cifra decimal. En primer lugar, utilice una constante de suavización de $\alpha = 0.80$, después utilice una constante de suavización de $\alpha = 0.20$, y entonces compare los dos conjuntos de pronósticos.

La Tabla 16.7 es una tabla de trabajo en la que se reportan los dos conjuntos de pronósticos. Por ejemplo, el pronóstico para 1986, con $\alpha = 0.20$, se determinó de la siguiente manera:

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$$

$$\hat{Y}_{1986} = \hat{Y}_{1985} + \alpha(Y_{1985} - \hat{Y}_{1985})$$

$$= \$1.1+0.20(0.4) = 1.1+0.08 = 1.18 \cong \$1.2$$

Por lo general, los errores de pronóstico son menores para α - 0.80. Por ello, el mayor peso que se da a los errores de pronóstico produce mejores pronósticos para esos datos.

Tabla 16.7 Pronósticos, año por año, con el método de suavización exponencial

		α=0.20		α=0.80	
Año (t)	Ventas en millones (Y_t)	Pronóstico ($\hat{Y}_t$)	Error del pronóstico ($Y_t - \hat{Y}_t$)	Pronóstico ($\hat{Y}_t$)	Error del pronóstico ($Y_t - \hat{Y}_t$)
1985	\$1.5	\$1.1	\$0.4	\$1.1	\$0.4
1986	1.3	1.2	0.1	1.4	-0.1
1987	1.1	1.2	-0.1	1.3	-0.2
1988	1.7	1.2	0.5	1.1	0.6
1989	1.9	1.3	0.6	1.6	0.3
1990	2.3	1.4	0.9	1.8	0.5
1991		1.6		2.2	

RESULTADOS POR COMPUTADORA

16.9 Aplique algún paquete de computación para obtener la ecuación de tendencia lineal para los datos de ventas anuales de la Tabla 16.1. Utilice los años codificados en la Tabla 16.1.

En la figura 16-5 se incluye la ecuación lineal, que difiere ligeramente de la que se determinó en forma manual en el problema 16.1 debido al redondeo de los valores. Como se explicaba en la sección 16.2, debe observarse que *no es* una ecuación de regresión como tal, aun cuando se utilizó el comando de "regresión" para obtener la ecuación lineal.

16.10 Utilice algún paquete de computación para determinar los índices estacionales con base en los datos trimestrales de la Tabla 16.3, y también haga el ajuste estacional de los datos.

En la figura 16-6 se incluyen los índices trimestrales y los datos dβsestacionalizados. Excepto por ligeras diferencias debidas a redondeo, los resultados son iguales a las soluciones de los problemas 16.4 y 16.5.

```
MTB > NAME Cl = 'YEAR', CS = 'SALES'
MTB > READ 'YEAR', 'SALES'
DATA)   0    0.2
DATA)   1    0.4
DATA)   2    0.5
DATA)   3    0.9
DATA)   4    1.1
DATA)   5    1.5
DATA)   6    1.3
DATA)   7    1.1
DATA)   8    1.7
DATA)   9    1.9
DATA)  10    2.3
DATA) END
MTB >  REBRESS 'SALES' USING 1 PREDICTOR 'YEAR'

 The regression equation is
 SALES = 0. 232 + 0.188 YEAR

 Predictor        Coef          Stdev      t-ratio
 Constant       0.2318         0.1169         1.98
 YEAR          0.18818        0.01976         9.52

 s = 0.2072      R-sq   91.0%     R-sq(adj) = 90.0%

 Analysis of Variance

 SOURCE        DF          SS           MS
 Regression     1      3.8954       3.8954
 Error          9      0.3865       0.0429
 Total         10      4.2818

 Unusual Observations
 Obs.     YEAR     SALES       Fit Stdev. Fit   Residual    St.Resid
   8       7.0    1.1000    1.5491     0.0739    -0.4491      -2.32R

 R denotes an obs. with a large st. resid.
```

Fig. 16-5 Resultado de Minitab para el problema 16.9

INPUT VALUES

PERIOD	1985	1986	1987	1988	1989	1990	1991	1992
I	500.00	450.00	350.00	550.00	550.00	750.00	0.00	0.00
II	350.00	350.00	200.00	350.00	400.00	500.00	0.00	0.00
III	250.00	200.00	150.00	250.00	350.00	400.00	0.00	0.00
IV	400.00	300.00	400.00	550.00	600.00	650.00	0.00	0.00

SEASONAL INDEXES

(USING PERCENT OF FOUR-QUARTER CENTERED MOVING AVERAGES)

PERIOD	1985	1986	1987	1988	1989	1990	1991	1992	SEASONAL INDEX
I	0.00	126.32	136.59	146.67	122.22	134.83	0.00	0.00	134.13
II	0.00	103.70	76.19	86.15	85.33	87.91	0.00	0.00	87.48
III	67.80	64.00	50.00	58.82	70.00	0.00	0.00	0.00	64.29
IV	110.34	106.67	116.36	127.54	111.63	0.00	0.00	0.00	114.10
									400.00

SEASONALLY ADJUSTED DATA

PERIOD	1985	1986	1987	1988	1989	1990	1991	1992
I	372.77	335.49	260.94	410.04	410.04	559.15	0.00	0.00
II	400.09	400.09	228.62	400.09	457.24	571.55	0.00	0.00
III	388.89	311.11	233.33	388.89	544.45	622.23	0.00	0.00
IV	350.56	262.92	350.56	482.03	525.85	569.67	0.00	0.00

Fig. 16-6 Resultado de computadora para el problema 16.10 (SEASON, derechos reservados por *Leonard J. Kazmier).*

Problemas complementarios

ANÁLISIS DE TENDENCIA

16.11 Con los datos de la Tabla 16.8, determine la ecuación de tendencia lineal para el valor total del producto de la industria petrolera mexicana, designando 1975 como el año base, con el propósito de codificar los años.

Resp. $Y_T =$ 11 969.32 + 2 704.98 X (en millones de pesos)

16.12 Construya la gráfica de línea para los datos de la Tabla 16.8, y dibuje la linea de tendencia en esa gráfica.

ANÁLISIS DE VARIACIONES CÍCLICAS

16.13 Determine los relativos cíclicos para los datos anuales de la serie de tiempo que se reportan en la Tabla 16.8.

Tabla 16.8 Valor total de la producción de la industria petrolera mexicana (millones de pesos)

Año	Valor (en millones de pesos de 1970)
1970	13 042.9
1971	14 081.6
1972	15 505.9
1973	18 241.3
1974	21 601.8
1975	26776.4
1976	31247.3
1977	34 073.8
1978	34 665.1
1979	35 701.2
1980	35 499.2

Fuente: Indicadores Económicos del Banco de México S.N.C.

16.14 Elabore una tabla de ciclos para ilustrar los relativos cíclicos que se determinan en el problema 16.13 anterior.

Resp. Los relativos cíclicos muestran una cumbre en 1981-1982 y un valle en 1979.

MEDICIÓN DE VARIACIONES ESTACIONALES

16.15 En la Tabla 16.9 se presentan los datos trimestrales para el número de pedidos que se recibieron en una empresa ficticia. Determine los índices estacionales con base en esos datos, utilizando tres cifras decimales durante los cálculos, y redondeando los índices a dos decimales.

Resp. I = 68.19, II = 93.76, III = 113.42, IV = 124.64

APLICACIÓN DE AJUSTES ESTACIONALES

16.16 Al continuar con el problema 16.15, realice el ajuste estacional de los datos y comente la comparación de los valores resultantes para los trimestres tercero y cuarto de 1990.

Resp. Aunque hubo un aumento en los pedidos del tercero al cuarto trimestres en los datos originales, con base en los datos ajustados estacionalmente se observa una reducción en los pedidos.

Tabla 16.9 Número de pedidos en una empresa (en millares)

Trimestre	Año 1986	1987	1988	1989	1990
I	9	10	13	11	14
II	16	15	22	17	18
III	18	18	17	25	25
IV	21	20	24	21	26

PRONÓSTICOS CON BASE EN FACTORES DE TENDENCIA Y ESTACIONALES

16.17 Suponga que la ecuación de tendencia para los pedidos trimestrales (en millares) es $Y_T = 13.91 + 0.43X$, en donde X = 0, en el primer trimestre de 1986. Pronostique el número de pedidos para cada uno de los cuatro trimestres de 1991, con base en los componentes estacional y de tendencia de la serie de tiempo.

Resp. I = 15.34,II = 21.51, III = 26.51, IV = 29.66

LA SUAVIZACIÓN EXPONENCIAL COMO MÉTODO DE PRONÓSTICO

16.18 Con referencia a los datos de la Tabla 16.8, utilizando la cantidad de 1982 ($34 073.8, millones de pesos) y el pronóstico "semilla" de 1983, siga el método de la suavización exponencial simple con una constante de suavización de α= 0.70 para determinar los pronósticos, año por año, de 1983 a 1986.

Resp. 36 292.53, 35 404.25, 35 272.14. 33 035.25

RESULTADOS POR COMPUTADORA

16.19 Utilice algún paquete de computación para determinar la ecuación de tendencia lineal para los datos de valor de la producción petrolera, que aparecen en la Tabla 16.8, y designando a 1975 como el año codificado X = 0.

16.20 Utilice algún paquete de computación para (*a*) determinar los índices estacionales trimestrales, y (*b*) desestacionalizar los datos sobre el número de pedidos de clientes de la Tabla 16.9.

Números índice para economía y negocios

17

17.1 INTRODUCCIÓN

Un *número índice* es un relativo porcentual que expresa una medición en un *periodo determinado* como el cociente con respecto a un *periodo base* determinado. Las mediciones pueden estar relacionadas con *cantidad, precio o valor.*

EJEMPLO 1. El índice de Precios al Consumidor (IPC) que elabora el Banco de México, S.N.C., es un ejemplo de un índice de precios, mientras que el índice de Producción Industrial es un ejemplo de un índice de cantidad.

Cuando el número índice representa una comparación para un producto o articulo *individual,* se trata de un *número índice simple.* Por otro lado, cuando el número índice se construye para un *conjunto* de productos o artículos, se trata de un *número índice agregado* o *número índice compuesto.*

EJEMPLO 2. El índice de precios de, por ejemplo, 130 para mantequilla es un índice de precios simple que indica que el precio de la mantequilla en un periodo dado fue 30% superior al precio del periodo base, para el cual por definición el índice de precios es 100. El mismo índice de precios de 130 para el índice de Precios al Consumidor (un índice de precios agregado) indicaría que el precio promedio de la "canasta básica" de aproximadamente 120 bienes y servicios resultó ser 30% superior en el periodo dado, en comparación con el periodo base.

17.2 CONSTRUCCIÓN DE ÍNDICES SIMPLES

Se utiliza p_n para representar el precio de un artículo básico en un periodo determinado, y p_0 indica el precio en el periodo base. La fórmula general para el índice de precios simple, o *relativo de precios,* es

$$I_q = \frac{p_n}{p_o} \times 100 \quad \text{(Véase el problema 17.1)} \tag{17.1}$$

De manera similar, utilizando q_n para indicar la cantidad de un producto que se fabrica o se vende en un periodo determinado, y q_0 para la cantidad del periodo base, la fórmula general para el índice simple de cantidad, o *relativo de cantidad,* es

$$I_q = \frac{q_n}{q_o} \times 100 \quad \text{(Véase el problema 17.2)} \tag{17.2}$$

Finalmente, el valor de un artículo en un periodo designado es igual a su precio multiplicado por la cantidad vendida (o fabricada). Por ello, p_nq_n indica el valor de un artículo en un periodo determinado, p_0q_0 indica el valor de un artículo en el periodo base. La fórmula general para el índice simple de valor, o *relativo de valor,* es

$$I_p = \frac{p_n q_n}{p_0 q_0} \times 100 \quad \text{(Véase el problema 17.3)} \tag{17.3}$$

17.3 CONSTRUCCIÓN DE ÍNDICES AGREGADOS DE PRECIOS

Para obtener un índice agregado de precios se podrían simplemente sumar los precios de diversos artículos o productos para el periodo dado y para el periodo base, respectivamente, para después compararlos. Un índice como éste sería un índice de precios agregado *no ponderado.* Por lo general, un índice no ponderado no es muy útil porque el peso implícito de cada articulo en el índice depende de las unidades sobre las cuales se basan los precios.

EJEMPLO 3. Si se reporta el precio de la leche "por galón", y no "por litro", entonces el precio de la leche haría una mayor contribución en un índice de precios no ponderado para un conjunto de artículos que lo incluye.

Debido a la dificultad que se describe en el ejemplo anterior, por lo general los índices de precios agregados se ponderan de acuerdo con las cantidades q de los artículos. La pregunta referente a qué tamaño de periodo debe utilizarse es lo que permite diferenciar los distintos tipos de relativos de precios agregados. Uno de los índices de precios agregados más popular es el *índice de Laspeyres,* en el cual los precios se ponderan con las cantidades correspondientes del año *base* antes de hacer la suma. La fórmula es

$$I(L) = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times 100 \quad \text{(Véase el problema 17.4)} \tag{17.4}$$

En vez de utilizar las cantidades del año base como pesos, se podrían utilizar las cantidades del *año dado.* En este caso, se construye lo que se denomina *índice de Paasche.* La fórmula es

$$I(P) = \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} \times 100 \quad \text{(Véase el problema 17.5)} \tag{17.5}$$

Puede decirse que tanto el método de Laspeyres como el de Paasche para construir índices de precios agregados utilizan el método de la *suma ponderada de precios.* Un método alternativo es el del promedio ponderado de relativos de precios, mediante el cual el índice simple de precios se pondera para cada artículo individual mediante un valor pq. Los valores que se utilizan pueden ser para el año base, p_0q_0 o para el año dado, p_nq_n. Por lo general, se utilizan como pesos los valores del año base, lo cual da como resultado la siguiente fórmula para los relativos de precios ponderados:

$$I_p = \frac{\Sigma[(p_o q_o)(p_n/p_o) \times 100)]}{\Sigma p_o q_o} \quad \text{(Véase el problema 17.6)} \tag{17.6}$$

Algebraicamente, *(17.6)* es equivalente al índice de Laspeyres, mientras que si se utilizan los valores del periodo dado como pesos, se obtiene un índice equivalente al de Paasche. La razón de la popularidad del método del promedio ponderado de relativos de precios es que a través de este procedimiento se calculan los índices para cada artículo por separado, y con frecuencia se requieren esos índices simples para propósitos de análisis, además del índice de precios agregado mismo.

17.4 RELATIVOS EN CADENA

Los *relativos en cadena* son índices para los cuales siempre se utiliza como base el periodo anterior. Por ello, para un conjunto de relativos en cadena referentes al valor anual de ventas, cada número índice representa una comparación con respecto al año anterior. Esos relativos son útiles para resaltar comparaciones de un año a otro, pero no son convenientes para hacer comparaciones a largo plazo. Véase el problema 17.7.

17.5 CAMBIO DEL PERIODO BASE

Con frecuencia la base de un número índice establecido se cambia a un año más reciente, para que las comparaciones actuales resulten más significativas. Suponiendo que las cantidades originales con las que se calculó una serie de números Índice, no están disponibles, se puede cambiar el periodo base para un número índice dividiendo cada uno de los índices originales entre el índice del nuevo año base y multiplicando el resultado por 100:

$$I_{n(\text{cambiado})} = \frac{I_{n(\text{antiguo})}}{\text{índice antiguo de la nueva base}} \times 100 \quad \text{(Véase el problema 17.9)} \tag{17.7}$$

17.6 FUSIÓN DE DOS SERIES DE NÚMEROS ÍNDICE

Con frecuencia, los números índice sufren cambios por la adición de ciertos productos o por la exclusión de otros, así como también por cambios en el año base. Aun así, para mantener la continuidad histórica, es deseable tener series unificadas de números índice. Con el objeto de *fusionar* las dos series de tiempo distintas para formar una sola serie continua de números índice, debe haber un año de traslape para las dos series, de manera que se hayan calculado ambos tipos de números índice para ese año. Por lo general, el año del traslape es también el nuevo año base, porque resulta ser el año en el que se añadieron o eliminaron productos del índice agregado. Los números índice que deben cambiarse en el proceso de combinación son los índices de la serie antigua. Se logra el cambio dividiendo el número índice nuevo para el año del traslape (100.0, si ésta es la nueva base) entre el índice antiguo para ese año y después multiplicando cada uno de los números índice de la serie vieja por este cociente. Véase el problema 17.10.

17.7 EL ÍNDICE DE PRECIOS AL CONSUMIDOR (IPC)

El *índice de Precios al Consumidor* es el índice más ampliamente conocido debido a que se le utiliza como indicador del costo de vida. Lo publica mensualmente el Banco de México, y es un índice de precios agregado para una "canasta básica" de aproximadamente 120 artículos y servicios. Se publican dos tipos de índices: el índice de precios al consumidor nacional, así como también el índice para las principales ciudades. En las publicaciones actualos el periodo de base del IPC es 1978.

17.8 EL PODER DE COMPRA Y LA DEFLACIÓN DE LOS VALORES DE UNA SERIE DE TIEMPO

Mientras que el índice de Precios al Consumidor indica el precio de una canasta básica de artículos y servicios, en comparación con el año base de 1978, el recíproco de IPC indica el *poder de compra* del peso con respecto al año base (es decir, en pesos de 1978):

$$\text{Valor del peso} = \frac{1}{\text{IPC}} \times 100 \quad \text{(Véase el problema 17.11)} \tag{17.8}$$

La deflación de una serie de tiempo es el proceso mediante el cual una serie de valores a precios corrientes se convierte a su equivalente en valores a pesos constantes. Utilizando la fórmula *(17.8),* el replanteamiento de esos valores a pesos constantes del año base se lleva a cabo mediante:

$$\text{Cantidad defiacionada} = \frac{\text{cantidad reportada}}{\text{IPC}} \times 100 \qquad (17.9)$$

Véase el problema 17.12.

17.9 OTROS ÍNDICES PUBLICADOS

El Banco de México publica también el índice de Precios al Productor, el índice de Precios de Materias Primas Consumidas por Rama de Actividad Económica, el índice del Costo de Edificación de la Vivienda de Interés Social, el índice de Salarios, Sueldos y Prestaciones Medias, el índice del Total de los Salarios, Sueldos y Prestaciones Pagadas, el índice del Costo Medio de las Horas-Hombre Trabajadas por el Personal Ocupado, y otros.

Problemas resueltos

ÍNDICES SIMPLES

17.1 Con referencia a la Tabla 17.1, determine los índices de precios simples para los tres artículos, en 1990, utilizando 1985 como año base.

Tabla 17.1 Precios y consumo de tres artículos básicos en una área metropolitana, 1985 y 1990

Artículo	Unidades	Precio promedio		Consumo mensual *per cápita*	
		1985 (p_o)	1990 (p_n)	1985 (q_0)	1990 (q_n)
Leche	Litro	$100	1500	5	6
Pan	Pieza 0.5	88	1300	3.8	3.7
Huevo	Docena	168	2500	1.0	1.2

Para la leche: $$I_p = \frac{p_n}{p_0} \times 100 = \frac{1500}{100} \times 100 = 1500$$

Para el pan: $$I_p = \frac{p_n}{p_0} \times 100 = \frac{1300}{88} \times 100 = 1477.27$$

Para el huevo: $$I_p = \frac{p_n}{p_0} \times 100 = \frac{2500}{168} \times 100 = 1488.1$$

17.2 Con referencia a la Tabla 17.1, determine los índices de cantidad simples para los tres artículos en 1990, utilizando 1985 como año base.

Para la leche: $I_q = \frac{q_n}{q_0} \times 100 = \frac{6}{5} \times 100 = 120.0$

Para el pan: $I_q = \frac{q_n}{q_0} \times 100 = \frac{3.7}{3.8} \times 100 = 97.4$

Para el huevo: $I_q = \frac{q_n}{q_0} \times 100 = \frac{1.2}{1.1} \times 100 = 109.1$

17.3 Calcule los relativos simples de valor de 1990, para los tres artículos de la Tabla 17.1. Utilice el año de 1985 como año base.

Para la leche: $I_v = \frac{p_n q_n}{p_0 q_0} \times 100 = \frac{(1500)(6)}{(100)(5)} \times 100 = 1800$

Para el pan: $I_v = \frac{p_n q_n}{p_0 q_0} \times 100 = \frac{(1300)(3.7)}{(88)(3.8)} \times 100 = 1438.40$

Para el huevo: $I_v = \frac{p_n q_n}{p_0 q_0} \times 100 = \frac{(2500)(1.2)}{(168)(1.0)} \times 100 = 1785.71$

ÍNDICES AGREGADOS DE PRECIOS

17.4 Calcule el índice agregado de precios de Laspeyres para los tres artículos de la Tabla 17.1, para 1990, utilizando 1985 como año base.

Con referencia a la Tabla 17.2, el índice se calcula de la siguiente manera:

$$I(L) = \frac{\Sigma p_n q_0}{\Sigma p_0 q_0} \times 100 = \frac{14\,940}{1002.4} \times 100 = 1490.42$$

Tabla 17.2 Hoja de trabajo para determinar el índice de Laspeyres para los datos de la Tabla 17.1

Artículo	$p_n q_n$	$p_0 q_0$
Leche	\$7 500	\$500
Pan de caja	4 940	334.4
Huevos	2 500	168
Total	$\Sigma p_n p_0 = \$14\,940$	$\Sigma p_0 q_0 = \$1002.4$

17.5 Calcule el índice agregado de precios de Paasche para los tres artículos de la Tabla 17.1, para el año 1990, utilizando el año 1985 como año base.

Al utilizar la Tabla 17.3, se calcula el índice de la siguiente manera:

$$I(P) = \frac{\Sigma p_n q_n}{\Sigma p_0 q_n} \times 100 = \frac{16\,810}{1127.2} \times 100 = 1491.31$$

Tabla 17.3 Hoja de trabajo para determinar el índice de Paasche para loa datos de la Tabla 17.1

Articulo	$p_n q_0$	$p_0 q_n$
Leche	$9000	$600
Pan de caja	4810	325.6
Huevos	3000	201.6
Total	$\Sigma p_n p_n = \$16\,810$	$\Sigma p_0 q_n = \$1127.2$

17.6 Calcule el indios de precios para los tres artículos de la Tabla 17.1, mediante el método del promedio ponderado de relativos de precios, utilizando 1985 como año base.

Con referencia a la Tabla 17.4,

$$I_p = \frac{\Sigma[(p_o q_o)(p_n/p_o) \times 100)]}{\Sigma p_o q_o} = \frac{1\,493\,999.89}{1002.4} = 1490.42$$

Esta respuesta coincide con el Índice de Laspeyres que se calculó en el problema 17.4.

Tabla 17.4 Hoja de trabajo para el cálculo del promedio ponderado del relativo de precios para loa dato de la Tabla 17.1

Artículo	Relativo de precios (p_n/q_0x100)	Ponderación de valor (p_0q_0)	Relativo ponderado [(p_0q_0)p_n/q_0x100)]
Leche	1500	500	750000
Pan de caja	1477.27	334.4	493 999.09
Huevos	1488.10	168	250000.8
Total		1002.4	1 493 999.89

RELATIVOS EN CADENA

17.7 En la Tabla 17.5 se presentan los datos de ventas de la industria farmacéutica mexicana (en millones de pesos), para los años de 1983 a 1988. Determine los relativos en cadena para esos datos.

En la misma Tabla 17.5 se muestran los relativos en cadena. Por ejemplo, el relativo en cadena de 449.70 indica que las ventas de 1984 fueron 339.70% mayores que las ventas del año anterior, 1983.

Tabla 17.S Ventas de la industria farmacéutica (millones de pesos)

Aflo	1983	1984	1985	1986	1987	1988
Ventas (millones de pesos)	8 931	40163	86 053	150 810	368 294	737 897
Relativo en cadena	___	449.70	214.25	175.25	244.21	200.36

Fuente: Revista Expansión Vol. XXI, Num. 531, 20 de diciembre de 1989.

17.8 Con referencia a los montos de ventas de la Tabla 17.5, calcule los índices de valor para esos seis años, utilizando 1983 como año base.

En la Tabla 17.6 se reportan los índices de valor.

Tabla 17.6 índices de valor para las ventas de la industria farmacéutica, 1983-1988

Año	1983	1984	1985	1986	1987	1988
índice de valor (1983 =100)	100.0	449.70	963.53	1 688.61	4123.77	8262.20

Fuente: Tabla 17.5

CAMBIO DEL PERIODO BASE

17.9 Cambie los índices de valor que se reportan en la Tabla 17.6, del año base de 1983 a 1988 como el año base común para los índices, utilizando solamente los índices que se reportan en esa tabla.

En la Tabla 17.7 se reportan los índices calculados utilizando el año de 1988 como base. A manera de ejemplo, el nuevo índice de valor para 1983 se. determina de la siguiente manera:

$$I_{N(\text{modificado})} = \frac{I_{n(\text{antiguo})}}{\text{índice antiguo del nuevo año base}} \times 100 = \frac{I_{1980(\text{antiguo})}}{\text{índice para 1985}} \times 100$$

$$= \frac{100.0}{8262.2} \times 100 = 0.012$$

Tabla 17.7 Índices de valor utilizando los años base de 1983 y 1988, para las ventas de la industria farmacéutica

Año	1983	1984	1985	1986	1987	1988
índice de valor (1983=100)	100.0	449.70	963.53	. 1688.61	4 123.77	8 262.20
índice de valor (1988= 100)	0.012	0.054	0.117	0.204	0.499	100

Fuente: Tabla 17.6

FUSIÓN DE DOS SERIES DE NÚMEROS ÍNDICE

17.10 En la Tabla 17.8 se presentan dos series hipotéticas de índices de precios, una calculada para un conjunto de artículos básicos para los años 1975 a 1980, utilizando como año base 1975, y la otra calculada comenzando en 1980 para un conjunto modificado de artículos básicos. Así, 1980 es el año de traslape para el que se calculan ambos índices. Combine estas dos seríes para conformar una serie continua de números índice, utilizando 1980 como año base.

Las seríes fusionadas de índices se reportan en la última columna de la Tabla 17.8. El cociente que resulta de dividir el nuevo número índice de 1980 (100.0) entre el húmero índice para 1980 de la serie antigua (119.2) es 0.839, y se utiliza este valor como factor de multiplicación para convertir cada uno de los números índice de la serie antigua.

Tabla 17.8 Fusión de dos series hipotéticas de números índice

Año	índice antiguo de precios (1975 - 100)	índice de precios modificado (1980 - 100)	índice de precios combinado (1980 = 100)
1975	100.0		83.9
1976	103.1		86.5
1977	106.9		89.7
1978	110.0		92.3
1979	114.1		95.7
1980	119.2	100.0	100.0 *
1981		105.2	105.2
1982		111.3	111.3
1983		117.5	117.5
1984		124.8	124.8
1985		129.9	129.9
1986		137.7	137.7

EL ÍNDICE DE PRECIOS AL CONSUMIDOR Y LA DEFLACIÓN DE VALORES DE SERIES DE TIEMPO

17.11 En la Tabla 17.9 se reportan los valores del índice de Precios al Consumidor de 1980 a 1986. Determine el poder de compra del peso para cada uno de esos años, en términos del valor del peso en el año base de 1978.

En la última columna de la Tabla 17.9 se reporta el valor del peso en cada año. Por ejemplo, el valor para 1980 se determinó de la siguiente manera:

$$\text{Valor del peso} = \frac{1}{\text{IPC}} \times 100 = \frac{1}{149.3} \times 100 = \$0.67$$

Así, en 1980, el valor del peso era 67 centavos en términos de pesos de 1978, en promedio.

Tabla 17.9 índices de precios al consumidor y valor del peso, 1980-1986 (año base - 1978)

Año	Índice de precios al consumidor (IPC)	Valor del peso
1980	149.3	0.670
1981	191.1	0.523
1982	303.6	0.329
1983	612.9	0.163
1984	1014.1	0.099
198S	1599.7	0.063
1986	2979.2	0.034

Fuente: Banco de México, *Indicadores Económicos*

17.12 En la Tabla 17.10 se reportan los sueldos anuales de un profesor universitario que comenzó a trabajar en 1980. (a) Deflacione esta serie de tiempo, de manera que las cantidades queden expresadas en pesos de 1978. (b) Replantee los salarios en pesos de 1980. (c) ¿Cuál fue el aumento salarial porcentual entre 1983 y 1989, en términos de pesos nominales (corrientes)? (d) ¿Cuál fue el aumento porcentual del sueldo entre 1983 y 1989 en términos de pesos constantes?

Tabla 17.10 Sueldos anuales

IPC	Año	Sueldo (en millones de pesos)	Salario deflacionado (pesos de 1978)	Salario replanteado (en pesos de 1983)
766.1491	1983	1	130 522	1000 000
1219.3764	1984	3	246027	1 884 934
1999.6229	1985	5	250410	1918 514
4 108.2000	1986	8	194 732	1491937
10 647.2	1987	13	122 098	935 453
16147.3	1988	20	123 860	948 952
19327.9	1989	26	134 521	1030 631

(a) En la tercera columna de la Tabla 17.10 se reportan los montos deflacionados. Por ejemplo, la cantidad deflacionada para 1983 se determinó de la siguiente manera:

$$\text{Cantidad deflacionada} = \frac{\text{Cantidad reportada}}{\text{IPC}} \times 100$$

$$= \frac{\$1\,000\,000}{766.1491} \times 100 = \$1305.22$$

(b) En la última columna de la Tabla 17.10, se reportan los salarios en términos de pesos de 1983. Pueden determinarse esas cantidades multiplicando cada uno de los montos corrientes por el cociente del IPC de 1983 y el IPC de cada uno de los años respectivos. Sin embargo, como ya se han determinado las cantidades defladonadas en pesos de 1978, un procedimiento más sencillo consiste en multiplicar cada una de las cantidades deflacionadas por el IPC de 1983. Por ejemplo, el salario de 1984, planteado en pesos de 1983, es

$$\text{Sueldo de 1984 (pesos de 1983)} = \text{salario de 1981 (pesos de 1978)}\ \frac{\text{IPC(1983)}}{100}$$

$$= \$\,246\,027 \times \frac{766.1491}{100} = \$\,246\,034.66$$

(c) $$\text{Aumento (pesos nominales)} = \frac{\text{Sueldo de 1989 - sueldo de 1983}}{\text{Sueldo de 1983}} \text{x100}$$

$$= \frac{26000000\text{-}1000000}{\$1000000} \text{ x } 100 = 2\ 500\%$$

(d) $$\text{Aumento (pesos de 1978)} = \frac{\$134\ 521\text{-}\$130\ 522}{\$130\ 522} \text{ x } 100 = 3.06\%$$

Problemas complementarios

ÍNDICES SIMPLES

17.13 Con referencia a la Tabla 17.11, y utilizando 1988 como año base, calcule los relativos simples de precio para (a) el papel tamaño carta, *(b)* los blocks de taquigrafía, (c) los marcadores y (d) los clips.

Resp. (a) 111.7, *(b)* 121.6, (c) 109.5, (d) 109.1

Tabla 17.11 Precios promedio y consumo mensual de una muestra seleccionada de artículos de oficina en determinada empresa, 1988 y 1989

Artículo	Unidades	Precio promedio		Consumo mensual	
		1988	1989	1988	1989
Papel tamaño carta	Millar	$26400	$29 500	4.5	8.0
Blocks para taquigrafia	Pieza	3 700	4 500	10.0	16.0
Marcadores	Pieza	2100	2300	8.0	6.0
Clips	Caja	1 100	1200	2.0	2.0

17.14 Con referencia a la Tabla 17.11, y utilizando 1988 como año base, calcule los relativos simples de cantidad para (a) el papel tamaño carta, (6) los blocks de taquigrafía, (c) ios marcadores y (d) los clips.

Resp. (a) 177.8, *(b)* 160.0, (c) 75.0, (d) 100.0

17.15 Con referencia a la Tabla 17.11, y utilizando 1988 como año base, calcule los relativos simples de valor para (a) el papel tamaño carta, *(b)* los blocks de taquigrafía, (c) los marcadores y *(d)* los clips.

Resp. (a) 198.7, (b) 194.6, (c) 82.1, (d) 109.1

ÍNDICES AGREGADOS DE PRECIOS

17.16 Determine el índice de precios de Laspeyres para los precios de 1989, correspondientes a los artículos de oficina que aparecen en la Tabla 17.11.

Resp. $I(L)=$ 113.6

17.17 Determine el índice de precios de Paasche para los precios de los artículos de oficina, correspondientes a 1989, que aparecen en la Tabla 17.11.

Resp. $I(P)=113.7$

17.18 Determine el índice agregado de precios para los precios de 1989 que aparecen en la Tabla 17.11, utilizando el método del promedio ponderado de relativos.

Resp. $I_p=$ 113.6

RELATIVOS EN CADENA

17.19 En la Tabla 17.12 se reporta la cantidad de energía eléctrica generada en México durante los años 1983 a 1988. Determine los relativos en cadena para esos datos.

Resp. Comenzando con 1983:108.1,107.89,102.89,106.80,105.61.

Tabla 17.12 Cantidad de energía eléctrica generada a nivel nacional 1983-1988 (en kilowatt-horas)

Año	1983	1984	1985	1986	1987	1988
Gigawatt - hora	77 521	83 804	90 412	93 024	99 349	104 924

Fuente: Primer Informe de Gobierno 1989.

17.20 Con referencia a las cifras de la cantidad de energía eléctrica generada que aparecen en la Tabla 17.12, calcule los relativos de cantidad de 1983 a 1988, utilizando 1983 como año base.

Resp. Comenzando con 1983:100.0,108.1,116.63,120.00,128.16,135.35.

CAMBIO DEL PERIODO BASE

17.21 Cambie el año base de 1983 que se utilizó para los índices de cantidad que se calcularon en el problema 17.20, al año base de 1988, utilizando los índices que se determinaron en ese problema, y no las cantidades originales de la Tabla 17.12.

Resp. Comenzando con 1983: 77.88, 79.87, 86.17, 88.66, 94.69,100.00.

FUSIÓN DE DOS SERIES DE NÚMEROS ÍNDICE

17.22 El periodo base que se utilizó para el índice de Precios al Consumidor antes de que se utilizara 1978 como año base fue 1970. En la Tabla 17.13 se reporta el índice de Precios al Consumidor para los años 1971 a 1980, utilizando el periodo base de 1970, y el índice de Precios al Consumidor revisado para 1978 a 1986, utilizando el año base 1978. Un economista desea elaborar una serie unificada de Índices utilizando como año base común el de 1978. Combine las dos series históricas de índices para satisfacer este objetivo.

Tabla 17.13 índices de precios al consumidor para la serie antigua (1970 - 100) y la serie nueva (1978 -100)

Año	índice antiguo de precios (1970 - 100)	Nuevo índice de precios (1978 = 100)
1971	103.1	
1972	104.2	
1973	105.4	
1974	106.7	
1975	108.1	
1976	109.9	
1977	113.1	
1978	116.3	100.0
1979	121.2	118.2
1980	127.7	149.3
1981		191.1
1982		303.6
1983		612.9
1984		1 014.1
1985		1 599.7
1986		2 979.2

Fuente: Banco de México S.N.C. *Indicadores Económicos*

EL ÍNDICE DE PRECIOS AL CONSUMIDOR Y LA DEFLACIÓN DE SERIES DE TIEMPO

17.23 En la Tabla 17.4 se reportan los salarios iniciales que se les pagaban a los empleados de determinada compañía entre 1983 y 1989. Utilizando los índices de Precios al Consumidor que se reportan en la Tabla 17.10, (a) Def lacione los salarios mensuales y expréselos en pesos de 1978, y (b) replantee los sueldos mensuales en términos de pesos constantes de 1983.

Resp. Comenzando con 1983: (a) $53 583.39,138 822,138 340,109 316,108 471,104 395, 72 838; *(b)* 79 999, 207 261,106 542,163 209,161 947,155 862,108 747.

Tabla 17.14 Promedio de salarios mensuales en una empresa, 1983 -1986

Año	1983	1984	1985	1986	1987	1988	1989
Sueldo mensual	$80 000	$250000	$420000	$670000	$1100000	$1 670 000	$2670 000

Análisis bayesiano de decisiones: Tablas de pagos y árboles de decisión

18

18.1 LA ESTRUCTURA DE LAS TABLAS DE PAGOS

Desde el punto de vista de la teoría estadística de las decisiones, una situación de decisiones bajo condiciones de incertldumbre puede representarse mediante ciertos factores comunes que se incluyen en la estructura de la *tabla de pagos pesa* la situación. En esencia, una tabla de pagos identifica la ganancia (o pérdida) condicional correspondiente a todas las combinaciones posibles de actos y eventos de decisión; también, típicamente, indica la probabilidad de ocurrencia para cada uno de los eventos mutuamente excluyentes.

Tabla 18.1 Estructura general de una tabla de pagos

Eventos	Probabilidad	Acciones				
		A_1	A_2	A_3	...	A_n
E_1	P_1	X_{11}	X_{12}	X_{13}	...	X_{1n}
E_2	P_2	X_{21}	X_{22}	X_{23}	...	X_{2n}
E_3	P_3	X_{31}	X_{32}	X_{33}	...	X_{3n}
...	...	...	...	...	...	...
E_m	P_m	X_{m1}	X_{m2}	X_{m3}	...	X_{mn}

En la Tabla 18.1, *las acciones* son los cursos alternativos de acción, o estrategias, que están disponibles para quien toma ias decisiones. Como resultado del análisis, se elige alguna de esas acciones como la mejor. La base para la elección es el tema de este capítulo. Como mínimo, debe haber cuando menos dos acciones posibles, de manera que exista de hecho la oportunidad de selección. Un ejemplo de una acción es el número de unidades de un artículo específico que deben ordenarse para el almacén.

Los *eventos* identifican las ocurrencias que están fuera del control de quien toma las decisiones y que determinan el nivel de éxito de una acción determinada. A estos eventos se les denomina con frecuencia "estados de la naturaleza", "estados" o "resultados". En este capítulo se revisan sólo los eventos discretos. Un ejemplo de un evento es el nivel de la demanda del mercado para un artículo determinado en algún periodo de tiempo estipulado.

Se incluye la *probabilidad* de cada evento como parte del formato general de una tabla de decisiones cuando esas probabilidades están disponibles. Sin embargo, una característica del análisis bayesiano de decisiones es que esas probabilidades deben estar siempre disponibles, puesto que pueden basarse en datos objetivos o pueden determinarse de manera subjetiva con base en juicios personales. Como los eventos de la tabla de pagos son mutuamente excluyentes y <haustivos, la suma de los valores de probabilidad debe ser 1.0.

Finalmente, las anotaciones en las celdas son los valores condicionales, o consecuencias económicas condicionales. Por lo general, a estos valores se les denomina pagos en la literatura, y son condicionales en el sentido de que el resultado económico que se experimenta depende del acto de decisión que se elige y del evento que ocurre.

EJEMPLO 1. Un contratista de aire acondicionado y calefacción debe comprometerse a la adquisición de unidades de aire acondicionado central para el primero de abril, las cuales va a revender e instalar durante la siguiente temporada veraniega. Con base en la demanda del verano anterior, las condiciones económicas actuales y los factores competitivos del mercado, estima que existe una probabilidad de 0.10 de vender solamente 5 unidades, una probabilidad de 0.30 de vender 10 unidades, una probabilidad de 0.40 de vender 15 unidades, y una probabilidad de 0.20 de vender 20 unidades. Los aparatos de aire acondicionado pueden ordenarse en conjuntos de cinco, y se tiene un precio unitario de $2 000 000, y el precio de reventa al menudeo es de $2 600 000 (más los cargos de instalación). Las unidades que no se venden al final de la temporada se devuelven al fabricante, el cual otorga un crédito neto de $1 600 000, después de deducir los cargos por embarque.

La Tabla 18.2 es la tabla de pagos para esta situación. Debe observarse que, como se estima que habrá una demanda de mercado de cuando menos cinco unidades, pero no más de veinte, éstos son lógicamente los límites para las posibles acciones (unidades ordenadas). La estipulación de unidades puede darse sólo en conjuntos de 5, y sirve para simplificar el problema y reducir el tamaño de la tabla de pagos. Los pagos de la Tabla 18.2 se basan en un margen de utilidades de $600 000 por unidad vendida, y una pérdida de $400 000 por cada unidad que no se venda. Así, por ejemplo, si se ordenan 15 unidades para las existencias y sólo se presenta una demanda de 10, el resultado económico es una ganancia de $6 000 000 para las 10 unidades vendidas, menos una pérdida de $2 000 000 para las 5 unidades que se devuelven al fabricante, lo cual da una ganancia resultante de $4 000 000 en A_3, E_2 en la tabla.

Tabla 18.2 Tabla de ganancias para el número de unidades de aire acondicionado que deben ordenarse

Demanda del mercado	Probabilidad	Cantidad del pedido			
		A_1:5	A_2:10	A_3:15	A_4:20
E_1:5	0.10	$3000000	$1000000	-$1000000	-$3 000 000
E_2:10	0.30	3000000	6000000	4 000000	2000 000
E_3:15	0.40	3000000	6000000	9000000	7000 000
E_4:20	0.20	3000000	6000000	9000000	12000 000
	1.00				

En las secciones siguientes se hace referencia al ejemplo 1 para ilustrar la aplicación de los diferentes criterios (o estándares de decisión) que pueden utilizarse para identificar la decisión que se considera mejor. Los métodos de este capitulo que implican el uso de valores de probabilidad asociados con cada evento, se refieren sólo a las probabilidades planteadas durante la estructuración inicial de la tabla de pagos. Como estos valores se plantean antes de la recopilación de cualquier información adicional, se les denomina probabilidades *a priori* en análisis bayesiano de decisiones. Véanse en los capítulos 19 y 20 otros conceptos del análisis bayesiano de decisión.

18.2 TOMA DE DECISIONES CON BASE ÚNICAMENTE EN PROBABILIDADES

El enfoque del análisis bayesiano de decisiones implica utilizar toda la información que se incluye en la tabla de pagos. Sin embargo, en esta sección se consideran brevemente los criterios que se utilizarían si se ignoraran las consecuencias económicas (o si no se determinaran), y si la decisión se basara exclusivamente en las probabilidades de ios eventos posibles.

En esos casos, un criterio de decisión que puede utilizarse consiste en identificar el evento que tiene la *máxima probabilidad* de ocurrencia, y elegir la decisión que corresponda a ese evento. Otra base para elegir la mejor acción consistiría en calcular la *esperanza* del evento y elegir la acción correspondiente. Sin embargo, como ninguno de estos criterios hace referencia a las consecuencias económicas de las diversas decisiones y eventos, representan una base incompleta para elegir la mejor decisión.

EJEMPLO 2. En la Tabla 18.3 se presenta la distribución de probabilidad para la demanda del mercado de las unidades de aire acondicionado central del ejemplo 1. El evento con la máxima probabilidad es E_3 = 15, para el cual P - 0.40. Con base en el criterio de máxima probabilidad, el número de unidades ordenadas sería 15.

En la Tabla 18.3 se incluye el cálculo de la demanda esperada E (D) (véase también la sección 6.2). Como las unidades de aire acondicionado pueden ordenarse sólo como unidades completas, y como, además, sólo se les puede ordenar en conjuntos de cinco, no puede ordenarse el nivel esperado de demanda de 13.5 unidades. Se ordenarían diez unidades con una esperanza de no tener disponibles 3.5 unidades (como promedio a largo plazo), o se pedirían 15 unidades con una esperanza de tener un exceso de 1.5 unidades (en promedio).

Tabla 18.3 Distribución de probabilidad de la demanda de mercado para las unidades de aire acondicionado y cálculo de la demanda esperada

Demanda del mercado (D)	Probabilidad *[P(D)]*	(D)P(D)
E_1:.5	0.10	0.5
E_2:10	0.30	3.0
E_3:15	0.40	6.0
E_4:20	0.20	4.0
	1.00	E(D) = 13.5

Una dificultad asociada con los dos criterios que se describieron en el ejemplo 2, consiste en que en realidad no es posible evaluar el éxito a largo plazo sin hacer ninguna referencia a las consecuencias económicas. Por ejemplo, supóngase que el contratista puede ordenar aparatos de aire acondicionado para sus existencias sin realizar ningún pago previo, y con la oportunidad de devolver las unidades que no se venden, a costa del fabricante. En circunstancias como éstas, no habría riesgos asociados con el exceso de inventarios, y la mejor decisión sería ordenar 20 unidades, para que el inventario resultara adecuado para el mayor nivel posible de la demanda.

18.3 TOMA DE DECISIONES CON BASE ÚNICAMENTE EN LAS CONSECUENCIAS ECONÓMICAS

La matriz de pagos que se utiliza para la toma de decisiones con base únicamente en las consecuencias económicas, es similar a la Tabla 18.1, excepto por la ausencia de la distribución de probabilidad de los eventos posibles. Se han definido y utilizado tres criterios para la toma de decisiones en estas circunstancias, y son los criterios maximín, maximax, y el del arrepentimiento minimax.

El *criterio maximín* es el estándar que señala que la mejor acción es aquella para la cual el valor mínimo es mayor que el mínimo de cualquier otra decisión. Utilizar este criterio conduce a una estrategia de decisión muy conservadora, ya que quien toma las decisiones se preocupa especialmente por "lo peor que puede pasar" con respecto a cada acción. En términos de cálculos, se determina el valor mínimo de cada columna de la tabla de pagos, y la mejor acción es aquella cuyo valor resultante es mayor.

EJEMPLO 3. En la Tabla 18.4 se presentan las consecuencias económicas de las diversas acciones y eventos para el problema que se describe en el ejemplo 1. El valor mínimo de decisión se enlista en el último renglón de la tabla. De estos valores, el mayor resultado económico (el máximo de los cuatro mínimos) es $3 000 000. Como la acción *"A_1:* ordenar cinco unidades de aire acondicionado" es la que corresponde a este resultado, se tiene entonces que es la mejor decisión desde el punto de vista del criterio maximín.

Tabla 18.4 Número de unidades de aire acondicionado que deben ordenarse de acuerdo con el criterio maximín

Demanda del mercado	A_1:5.	A_2:10	A_3:15	A_4:20
E_1:5	$3 000000	$1000 000	-$1000000	-$3000 000
E_2:10	3000000	6000000	4000000	2 000 000
E_3:15	3000000	6000000	9 000000	7000 000
E_4:20	3000 000	6000000	9000000	12 000 000
Mínimo	$1000000	$1000000	-$1000000	-$3 000 000

El *criterio maximax* es el estándar que define que la mejor acción es aquella cuyo valor máximo es mayor que el máximo para cualquier otra decisión. Este criterio es filosóficamente opuesto al criterio maximín, puesto que quien toma las decisiones se orienta específicamente hacia "lo mejor que puede suceder" con respecto a cada acción. En términos de cálculos, se determina el valor máximo de cada una de las columnas de la tabla de pagos, y la mejor acción es aquella cuyo valor resultante es mayor.

EJEMPLO 4. En la Tabla 18.5 se enlista, en el último renglón de la tabla, el valor máximo de decisión. El mayor de estos valores máximos (el máximo de los máximos) es $12 000 000 y, por ello, la acción correspondiente A_4: ordenar 20 unidades de aire acondicionado" sería la mejor acción desde el punto de vista del criterio maximax. En vez de llevar a cabo los dos pasos para identificar los máximos de columna y después el máximo de estos diversos valores, puede utilizarse un procedimiento abreviado para ubicar el valor máximo de la tabla.

Tabla 18.5 Número de unidades de aire acondicionado que deben ordenarse de acuerdo con el criterio maximax

Demanda del mercado	A_1:5	A_2:10	A_3: 15	A_4:20
E_1:5	$3 000000	$1000000	-$1000000	-$3 000 000
E_2:10	3000000	6000000	4000000	2 000 000
E_3:15	3000000	6000000	9000000	7 000 000
E_4:20	3000000	6000000	9000000	12 000000
Máximo	$3 000000	$6000000	$9 000 000	$12 000 000

El análisis mediante el criterio *del arrepentimiento minimax* se basa en lo que se denominan "arrepentimientos", y no en valores convencionales como tales. Un *arrepentimiento,* o *pérdida condicional de oportunidad* para cada acción, es la diferencia entre el resultado económico para esa acción, y el resultado económico para la mejor acción *dado que un evento especifico ha ocurrido.* Por ello, el valor de arrepentimiento más deseable, o "mejor", es "cero", que indica que el acto corresponde perfectamente al evento dado. También puede observarse que aun cuando exista una ganancia económica para un evento y una acción específicos, también podría haber una pérdida de oportunidad debido a que algún otro acto pudiera dar como resultado un pago mayor que el evento considerado.

En el ejemplo 5 se ilustra la construcción de la tabla de pérdidas de oportunidad, o arrepentimientos. Se identifica como la mejor acción aquella para la cual el arrepentimiento máximo posible es menor. Filosóficamente, el criterio del arrepentimiento minimax es similar al criterio maximín en términos de "asumir lo peor". Sin embargo, el uso del concepto de pérdida de oportunidad da como resultado un criterio más amplio, ya que se considera que no poder mejorar una ganancia es ciertamente una forma de pérdida.

EJEMPLO 5. La Tabla 18.6 es la tabla de pérdidas de oportunidad para los valores condicionales de la Tabla 18.5. En términos de cálculos, los valores de arrepentimiento se determinan asumiendo cada uno de los eventos en turno, es decir, de acuerdo con los renglones. Por ejemplo, dado que ocurre E1, la mejor acción (con referencia a la matriz de pagos de la Tabla 18.2).

Tabla 18.6 Tabla de pérdidas de oportunidad para el número de unidades de aire acondicionado que se ordenan, y aplicación del criterio del arrepentimiento minimax

Demanda del mercado	**A_1: 5**	**A_2: 10**	**A_3: 15**	**A_4: 20**
E_1: 5	$ 0	$2 000 000 (=$3000000-1000000)	$4 000 000 [-$3 000 000 - (-1 000 000)]	$6 000 000 [- $3 000 000 - (-3 000 000)]
E_2: 10	3000000 (=6 000000-3000000)	0	2000 000 (=6 000000-4 000 000)	4 000 000 (-6 000 000-2 000 000)
E_3: 15	6000000 (=9 000000-3000000)	3000 000 (-9 000 000-6 000 000)	0	2 000 000 (-9 000 000-7 000 000)
E_4: 20	9000 000 (=12000000-3000000)	6000000 (-12000000-6000 000)	3000 000 (-12000000-9 000 000)	0
Máximo de arrepentimiento	$9000 000	$6 000 000	$4000000	$6 000 000

es A_1, con un valor de $3 000 000. Si se elige A_2, el resultado es $1 000 000, lo cual difiere de la mejor acción en $3 000 000, que es entonces el valor del arrepentimiento para A_2. Si se elige A_3, el arrepentimiento = $3 000 000 - (-1 000 000) - $4 000 000. Si se elige A_4, el arrepentimiento = $3 000 000 - (-3 000 000) = 6 000 000. Los valores restantes de las pérdidas de oportunidad se determinan de manera similar, y el mejor valor condicional en cada renglón sirve como base para determinar los valores de arrepentimiento en ese renglón. Por ejemplo, en el renglón 2, para E_2:10 la mejor acción es A_2:10; con un pago de $6 000 000, de acuerdo con la Tabla 18.5.

En el último renglón de la Tabla 18.6, se enlista el máximo arrepentimiento que puede ocurrir de acuerdo con cada acto de decisión. El menor de estos máximos (el mínimo de los arrepentimientos máximos) es $4 000 000; *"$A_3$: ordenar 15 unidades"* se elegirla, entonces, como la mejor acción desde el punto de vista del criterio de arrepentimiento minimax.

18.4 TOMA PE DECISIONES CON BASE EN LAS PROBABILIDADES Y EN LAS CONSECUENCIAS ECONÓMICAS: EL CRITERIO DEL PAGO ESPERADO

Los métodos que se presentan en esta sección utilizan toda la información contenida en la tabla básica de pagos (sección 18.1). Así, se consideran tanto las probabilidades de los posibles eventos, como las consecuencias económicas para todas las combinaciones de acciones y eventos.

El criterio del *pago esperado (PE),* es el estándar de acuerdo con el cual la mejor acción es la que tiene el mayor resultado económico esperado, en calidad de promedio a largo plazo. Debe observarse que, en el presente caso, lo que interesa es el resultado económico promedio a largo plazo, y no simplemente el valor promedio del evento a largo plazo (el nivel de demanda), que se analizó en la sección 18.2. En términos de cálculo, se determina el pago esperado para cada acción multiplicando el pago condicional correspondiente a cada combinación de evento y acción por la probabilidad del evento, y sumando estos productos para cada acción.

EJEMPLO 6. En la Tabla 18.7 se repite la información de la Tabla 18.2, excepto que se han colocado en el último renglón los pagos esperados. En la Tabla 18.8 se ilustra el procedimiento que se utiliza para calcular esos pagos esperados. En la Tabla 18.7, el mayor pago esperado es $6 500 000 y, por ello, la acción correspondiente *"A3:* ordenar 15 unidades" es la mejor acción desde el punto de vista del criterio del pago esperado.

Tabla 18.7 Tabla de ganancias para el número de unidades de aire acondicionado que deben ordenarse, e identificación de la mejor decisión, de acuerdo con el criterio de la ganancia esperada

Demanda del mercado	Probabilidad	Cantidad del pedido			
		A_1: 5	A_2: 10	A_3: 15	A_4: 20
E_1: 5	0.10	$3000000	$1000000	-$1000000	-$ 3 000000
E_2: 10	0.30	3000000	6000000	4000000	2000000
E_3: 15	0.40	3000000	6000000	9000000	7000000
E_4: 20	0.20	3000000	6000000	9000000	12 000000
Pago esperado (PE)		$3000000	$5500000	$6 500 000	$5 500000

Tabla 18.8 Cálculo de la ganancia esperada para la decisión A_A en la Tabla 18.7

Demanda del mercado	Pago pata A_4: X	$P(X)$	$XP(X)$
E_1: 5	-$ 3 000000	0.10	-$ 300000
E_2: 10	2000000	0.30	600000
E_3: 15	7000000	0.40	2 800000
E_4: 20	12000000	0.20	2 400000
			$\Sigma XP(X)$ = $5 500 000

Con frecuencia se denomina *criterio bayeslano* al criterio del pago esperado. El uso del adjetivo "bayesiano" aquí difiere de su uso en el teorema de Bayes para revisar valores de probabilidad a *priori* (véanse las secciones 5.7 y 19.2). Por ello, una referencia a "procedimientos bayesianos" en análisis de decisiones puede Implicar el uso del criterio del pago esperado, la revisión de valores de probabilidad a *priori,* o ambas cosas.

La mejor acción que se identifica mediante el criterio del pago esperado puede determinarse también identificando la acción que tiene la mínima pérdida esperada de oportunidad *(PEO),* o arrepentimiento esperado. Esto es as!, porque es lógico que la acción que tiene la mayor ganancia esperada tiene, al mismo tiempo, el menor arrepentimiento esperado. En los libros en los que se hace referencia a las pérdidas de oportunidad en términos simples como *pérdidas* o *pérdidas esperadas,* se sobreentiende que se trata de pérdidas de oportunidad.

EJEMPLO 7. En la Tabla 18.9 se repiten los valores de pérdida de oportunidad de la Tabla 18.6. Se muestra también la probabilidad de cada uno de los eventos y la pérdida esperada de oportunidad para cada decisión. En la Tabla 18.10 se ilustra el procedimiento utilizado para calcular las pérdidas esperadas de oportunidad. Como se indica en la Tabla 18.9, la mínima pérdida de oportunidad esperada es $1 800, y, por ello, el correspondiente acto "ordenar 15 unidades" es el mejor desde el punto de vista de minimizacion de la pérdida esperada de oportunidad. Debe observarse que este mismo acto es el que se identificó utilizando el criterio del pago esperado en el ejemplo 6.

Tabla 18.9 Tabla de pérdidas de oportunidad para el número de unidades de aire acondicionado que deben ordenarse, y cálculo de las pérdidas de oportunidad esperadas

Demanda del mercado	Probabilidad	Cantidad del pedido			
		A_1:5	A_2:10	A_3:15	A_4: 20
E_1:5	0.10	$ 0	$2 000000	$4 000 000	$6 000 000
E_2:10	0.30	3000 000	0	2 000 000	4 000 000
E_3:15	0.40	6000 000	3000 000	0	2 000 000
E_4:20	0.20	9000000	6000 000	3 000 000	0
Pérdida de oportunidad esperada (POP)		$5 100 000	$2 600 000	$1600000	$2 600 000

Tabla 18.10 Cálculo de la pérdida de oportunidad esperada para la decisión A_4 en la Tabla 18.9

Demanda del mercado	Pérdida de oportunidad condicional para A_4: PO	P(PO)	(PO)P(PO)
E_1:5	$6000 000	0.10	$ 600 000
E_2:10	4000000	0.30	1200 000
E_3:15	2000000	0.40	800 000
E_4:20	0	0.20	0
			POE = $2 600 000

18.5 ANÁLISIS DE ÁRBOLES DE DECISIÓN

Con frecuencia los problemas de decisiones se complican por el hecho de que los pagos no están solamente asociados con una decisión inicial, sino que los eventos subsecuentes implican la necesidad de decisiones adicionales en cada una de las etapas de un proceso secuencial. Necesariamente, la evaluación de las acciones alternativas de decisión en la primera etapa de un proceso secuencial como esos debe basarse en la evaluación de los eventos y las decisiones del proceso en su conjunto. El análisis de árboles de decisión es el método que puede utilizarse para identificar la mejor acción inicial, así como también las mejores acciones subsecuentes. El criterio de decisión que se satisface es el criterio bayesiano del pago esperado (véase la sección 18.4).

El primer paso en el análisis de árboles de decisión consiste en construir el árbol que corresponda a una situación de decisiones secuenciales. El árbol se construye de izquierda a derecha, identificando en forma apropiada los puntos de decisión (puntos secuenciales en los que debe tomarse una decisión) y los eventos aleatorios (puntos secuenciales en los que ocurre algún evento probabilístico). La construcción de un árbol de decisión válido para una situación determinada secuencial depende en especial de un análisis apropiado de la situación global. Véase el problema 18.12.

Después de que se construye el árbol de decisiones, se anotan en el diagrama los valores de probabilidad de los eventos aleatorios y los pagos que pueden ocurrir. Muchos de los pagos están a diversas etapas de distancia del punto inicial. Con el objeto de determinar los pagos esperados de las acciones alternativas en el punto inicial de decisión, los pagos esperados se calculan sistemáticamente, de derecha a izquierda, en el árbol. En ocasiones, a este proceso se le denomina "reversión" (véase el problema 18.13). Como resultado de la aplicación de este proceso analítico, puede identificarse el mejor acto en el punto inicial de decisión.

18.6 LA UTILIDAD ESPERADA COMO CRITERIO DE DECISIÓN

Por lo general, el criterio del pago esperado se utiliza junto con el análisis de la tabla de pagos, y con el análisis de árboles de decisión (véase la sección 18.4). Sin embargo, cuando la persona que toma las decisiones percibe que alguna o más de las consecuencias económicas son demasiado grandes o pequeñas, el criterio del pago esperado no necesariamente ofrece una base para identificar la "mejor" decisión. Esto es particularmente probable para situaciones únicas, más que para situaciones repetitivas.

EJEMPLO 8. Considere la situación de la Tabla 18.11, y suponga que la decisión con respecto a cada par se toma solamente una vez. Aunque puede demostrarse con facilidad que el valor esperado de A_2 es mayor que el de A_1 en todos los casos, la mayoría de las personas elegiría A_1, y no A_2 para cada uno de estos pares. Como se supone que cada elección es una acción única, la elección de A_1 es racional aunque no se maximice el pago esperado. La implicación de estas conclusiones es que los valores monetarios pueden no representar en forma apropiada los valores reales, de acuerdo con quien toma las decisiones, en situaciones que incluyen la posibilidad de pérdidas o ganancias excepcionales.

Tabla 18.11 Tres pares de decisiones alternativas con sus consecuencias correspondientes

A_1: Recibir $ 1 000 000 con seguridad	A_2: Recibir $2 000 000 con probabilidad de 0.50 0 recibir $100 con probabilidad de 0.50
A_1: Pagar $10	A_2: Sufrir una pérdida de $8 000 con probabilidad de 0.001,0 no sufrir ninguna pérdida con probabilidad de 0.999
A_1: Recibir $15 000, con probabilidad de 0.50, ó recibir $5 000, con probabilidad de 0.50.	A_2: Recibir $50 000 con probabilidad de 0.50 0 perder $25 000 con probabilidad de 0.50

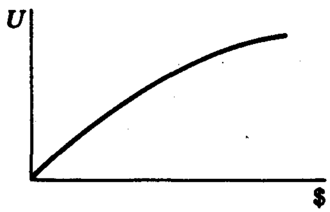
La utilidad es una medida que expresa el verdadero valor relativo de diversos resultados, incluyendo las consecuencias económicas, para una persona que toma decisiones. En este libro se maneja sólo la utilidad de las consecuencias económicas. Cualquier escala dada de utilidad puede comenzar en un valor mínimo arbitrario y tener un valor máximo arbitrario también. Sin embargo, es conveniente tener valores de utilidad que comiencen en un mínimo de 0 y que lleguen hasta un máximo de 1.00, y ésta es la escala que se utiliza con mayor frecuencia. Con una escala como ésta, se entiende que un resultado con utilidad de 0.60 es doblemente deseable que uno con una utilidad de 0.30. Por otro lado, debe observarse que un resultado económico de $60 000 no es necesariamente el doble de deseable que un resultado de $30 000 para un tomador de decisiones con recursos limitados.

Al utilizar un contrato de referencia, pueden determinarse los valores de utilidad de una persona para diferentes valores monetarios. Siguiendo este método, se le pide a la persona que señale una cantidad cierta que aceptaría, o que pagaría, y que considera equivalente a cada una de un conjunto de situaciones inciertas que implican riesgo. La primera situación riesgosa que se ilustra siempre incluye los dos limites extremos del rango de valores monetarios de interés, es decir, los límites superior e inferior que tienen utilidades de 0 y 1.0, respectivamente. Véase el problema 18.14.

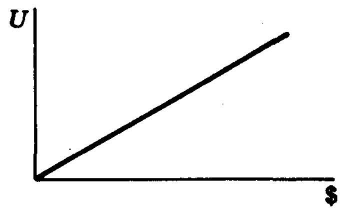
Una vez que se ha establecido el procedimiento para el contrato de referencia, puede continuarse cambiando ya sea las probabilidades designadas o una o más de las consecuencias económicas en la situación riesgosa. De esta manera, es posible determinar el conjunto de valores de utilidad que corresponde a un rango de valores monetarios. Véase el problema 18.15.

Después de determinar los valores de utilidad, es posible trazar los valores apareados en una gráfica. Puede trazarse una línea suave ajustada sobre los puntos trazados, a manera de aproximación de la función de utilidad de quien toma las decisiones, para varios pagos. La convención usual consiste en trazar los valores monetarios en el eje horizontal y los valores de utilidad en el eje vertical. Puede utilizarse esta gráfica como base para estimar el valor de utilidad de cualquier resultado monetario entre los limites designados de la función. A su vez, esto hace posible substituir valores de utilidad por valores monetarios en la tabla de pagos, y permite también determinar la mejor acción identificando la que tiene la máxima utilidad esperada (UE). Véanse los problemas 18.16 y 18.17.

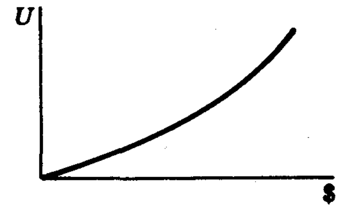
La forma de la función de utilidad indica si un tomador de decisiones es aficionado al riesgo o si lo evita (véase la figura 18-1). Para quien evita los riesgos, a cada peso adicional sobre el eje horizontal le corresponde una pendiente cada vez menor de la función de utilidad. Es decir, podría decirse que cada incremento adicional tiene un valor positivo para quien toma las decisiones, pero no tanto como los incrementos anteriores (la curva es cóncava). Por el contrario, la figura 18-1 (c) indica que, para la persona que busca los riesgos, cada incremento monetario adicional tiene un valor mayor (la curva es convexa). También puede probarse que las personas que evitan los riesgos designan una cantidad cierta que es consistentemente *menor* que el pago esperado para la situación de riesgo, en tanto que las personas que buscan los riesgos designan una cantidad cierta que es consistentemente *mayor* que el pago esperado para la situación. Véase el problema 18.16.



(a) Evita asumir riesgos



(b) Neutral ante el riesgo



(c) Asume riesgos

Fig. 18-1

En términos generales, la forma de la curva de utilidad refleja la actitud que tiene hacia el riesgo la persona que toma las decisiones y, por lo tanto, es un factor importante para determinar qué acción es mejor para *quien toma las decisiones* en un momento especifico. Para una empresa que se encuentra en peligro financiero, un contrato de oportunidad especifico que incluye la posibilidad de una perdida grande y el consecuente fracaso del negocio puede no representar una buena oportunidad, aun cuando el pago esperado del contrato resulte positivo y represente un rendimiento importante. El concepto de utilidad ofrece una base para demostrar por qué las dos partes que Intervienen en un contrato (tal como la aseguradora y el asegurado) pueden experimentar una utilidad positiva, y por qué una situación de riesgo que es "apropiada" para una empresa puede no serio para otra. Sin embargo, si la función de utilidad es lineal, o aproximadamente, como en la figura 18-1(b), entonces el criterio de pago esperado es equivalente al criterio de la utilidad esperada para ese tomador de decisiones.

Problemas resueltos

TABLAS DE PAGOS

18.1 Con base en un nuevo método tecnológico, un fabricante ha desarrollado un aparato de televisión a color con una pantalla de 45'. El propietario de una tienda pequeña estima que al precio de venta de $2 800 000, los valores de probabilidad asociados con la venta de 2, 3,4 ó 5 aparatos durante los tres meses de interés son 0.30,0.40, 0.20 y 0.10, respectivamente. Con base sólo en estos valores de probabilidad, ¿qué número de aparatos debe ordenar el encargado de la tienda para sus existencias, suponiendo que no es posible reordenar pedidos durante el periodo?

Con base en el criterio de máxima probabilidad, deben ordenarse 3 aparatos, puesto que la probabilidad de 0.40 para la venta de 3 aparatos es mayor que la probabilidad de cualquier otro evento.

Por otro lado, la expectativa del nivel de demanda es

$$2(0.30) + 3(0.40) + 4(0.20) + 5(0.10) = 3.1$$

Con base en esta expectativa para el evento, la decisión más parecida consiste en ordenar 3 aparatos.

18.2 Para la decisión de inventarios del problema 18.1, el margen de utilidad para cada aparato que se vende es $200 000. Si no se venden aparatos durante los tres meses, la pérdida total para la tienda es de $300 000 por aparato. Con base sólo en estas consecuencias económicas, e ignorando los valores de probabilidad que se identificaron en el problema 18.1, determine las mejores decisiones desde el punto de vista de los criterios maximln y maximax.

Con referencia a la Tabla 18.12, la mejor acción para el criterio maximln es A_1: ordenar dos aparatos. Para el criterio maximax, la mejor acción es A_4: ordenar 5 aparatos.

Tabla 18.12 Número de aparatos de televisión que deben ordenarse, de acuerdo con los criterios maxlmín y maximax

	Cantidad del pedido			
Demanda del mercado	A,: 2	A_2:3	A_3:4	A_4=5
E_1:2	$ 400000	$ 100000	-$ 200000	-$ 500 000
E_2:3	400000	600000	300000	0
E_3:4	400000	600000	800000	500000
E_4:5	400000	600000	800000	1000000
Mínimo	$ 400 000	$ 100000	-$ 200000	$ 500 000
Máximo	$ 400000	$ 600000	$ 800000	$1 000 000

18.3 Determine la mejor decisión desde el punto de vista del criterio del arrepentimiento minimax, para la situación que se describe en los problemas 18.1 y 18.2.

Tabla 18.13 Tabla de pérdidas de oportunidad para el número de aparatos de televisión que deben ordenarse, y aplicación del criterio del arrepentimiento minimax

	Cantidad del pedido			
Demanda del mercado	A_1: 2	A_2:3	A_3:4	A_4:5
E_1:2	$ 0	$ 300000	$ 600000	$ 900 000
E_2:3	200000	0	300000	600000
E_3:4	400000	200000	0	300000
E_4:5	600000	400000	200000	0
Arrepentimiento máximo	$ 600000	$ 400 000	$ 600000	$ 900000

En la Tabla 18.13 se incluyen las pérdidas de oportunidad (arrepentimientos) para esta situación. Desde el punto de vista del criterio del arrepentimiento minimax, la mejor decisión es A_2. ordenar 3 aparatos.

18.4 Con referencia a los problemas 18.1 y 18.2, determine la mejor decisión desde el punto de vista del criterio del pago esperado.

En la Tabla 18.14 se presenta la tabla completa de decisión para este problema, y se reportan los pagos esperados para los posibles actos de decisión. Como se indica en la tabla, la mejor acción, desde el punto de vista del criterio del pago esperado es A_2: ordenar 3 aparatos.

18.5 Al utilizar la Tabla 18.14 determine la mejor decisión de acuerdo con el criterio bayesiano, identificando la que tiene la menor pérdida de oportunidad esperada.

En la Tabla 18.15 se repiten los valores de las pérdidas de oportunidad de la Tabla 18.13, y se identifica la pérdida de oportunidad esperada correspondiente a cada decisión. Como se indica en la tabla, la mejor acción desde el punto de vista de la minimización de la pérdida de oportunidad esperada *es A_2:* ordenar 3 aparatos. Este resultado era de esperarse, puesto que la acción que satisface el criterio bayesiano de maximizar el pago esperado también resulta ser la decisión que tiene la menor pérdida de oportunidad esperada.

Tabla 18.14 Tabla de pagos para el número de aparatos de televisión que deben ordenarse, y aplicación del criterio del pago esperado

Demanda del mercado	Probabilidad	Cantidad del pedido			
		A_1: 2	A_2:3	A_3:4	A_4:5
E_1:2	0.30	\$ 400000	\$ 100 000	-\$ 200 000	-\$ 500 000
E_2:3	0.40	400000	600000	300 000	0
E_3:4	0.20	400000	600000	800 000	500 000
E_4:5	0.10	400000	600000	800 000	1 000 000
Pago esperado *(PE)*		\$ 400000	\$ 450000	\$ 300000	-\$ 50 000

Tabla 18.15 Tabla de pérdidas de oportunidad para el número de aparatos de televisión que deben ordenarse, y cálculo de las pérdidas de oportunidad esperadas

Demanda del mercado	Probabilidad	Cantidad del pedido			
		A_1: 2	A_2:3	A_3:4	A_4: 5
E_1:2	0.30	\$ 0	\$ 300 000	\$ 600 000	\$ 900 000
E_2:3	0.40	200000	0	300 000	600 000
E_3:4	0.20	400 000	200 000	0	300 000
E_4:5	0.10	600 000	400 000	200 000	0
Pérdida de oportunidad esperada (*POE*)		\$ 220 000	\$ 170000	\$ 320 000	\$ 570 000

18.6 En la Tabla 18.16 se presentan los pagos (rendimientos) correspondientes a cinco tipos alternativos de decisiones de inversión para un periodo de un año. Como no se tienen disponibles las probabilidades correspondientes a los posibles estados, determine las mejores decisiones desde el punto de vista del criterio maximín y maximax.

Tabla 18.16 Rendimientos monetarios asociados con diversas alternativas de inversión, para un fondo de $10 000 000

Estado de la economía	Decisión de inversión				
	A_1 Cuenta de ahorros	Bonos	A_3 Acciones AAA	A_4 Acciones especulativas	A_5 Acciones de riesgo
E_1: Recesión	$600	$500	-$2 500	-$5 000	-$10 000
E_2: Estable	600	900	800	400	5 000
E_3: Expansión	600	900	4000	10000	20 000

En la Tabla 18.17 se identifican los valores máximo y mínimo para cada uno de los actos de la Tabla 18.16. Como sería de esperar, el criterio maximín conduce a tomar la muy conservadora decisión A_1: invertir en una cuenta de ahorros. Por otro lado, el criterio maximax conduce a la decisión del otro extremo, A_5: invertir en acciones de riesgo.

Tabla 18.17 Decisión de inversión que debe tomarse de acuerdo con los criterios maximín y maximax

Consecuencia económica	Decisión de inversión				
	Cuenta de ahorros	A_2 Bonos	A_3 Acciones AAA	A_4 Acciones especulativas	A_5 Acciones de riesgo
Mínima	$600	$500	-$2 500	- $ 5 000	-$10 000
Máxima	$600	$900	$4 000	$10 000	$20 000

18.7 Determine la mejor decisión desde el punto de vista del criterio del arrepentimiento minimax para la situación que se describe en el problema 18.6.

En la Tabla 18.18 se muestra que la mejor acción en este caso es A_4: invertir en acciones especulativas. Debe observarse la medida en que los valores del arrepentimiento minimax para los primeros cuatro actos se ven afectados por la posibilidad de una gran ganancia con A_5, y que no se consideran los valores de probabilidad utilizando este criterio del arrepentimiento minimax.

Tabla 18.18 Tabla de pérdida de oportunidad para el problema de decisión sobre inversiones, y aplicación del criterio del arrepentimiento minimax

Estado de la economía	Decisión de inversión				
	A_1 Cuenta de ahorros	A_2 Bonos	A_3 Acciones AAA	A_4 Acciones especulativas	A_5 Acciones de riesgo
E_1: Recesión	$ 0	$ 100	$ 3 100	$ 5 600	$10 600
E_2. Estable	300	0	100	500	5 900
E_3: Expansión	19 400	19 100	16 000	10 000	0
Arrepentimiento máximo	$19 400	$19 100	$16 000	$10000	$10 600

18.8 Para el problema 18.6, suponga que las probabilidades correspondientes a condiciones de recesión, estabilidad económica y expansión son 0.30, 0.50 y 0.20, respectivamente. Determine la mejor decisión desde el punto de vista del criterio bayesiano de maximizar el pago esperado.

En la Tabla 18.19 se presenta la tabla completa de pagos para este problema, y también se señala el pago esperado correspondiente a cada decisión. La mejor acción desde el punto de vista del criterio del pago esperado es A_2: invertir en bonos.

Tabla 18.19 Tabla de pagos para el problema de inversiones, y determinación de la mejor acción, de acuerdo con el criterio de la ganancia esperada

Estado de la economía	Probabilidad	Decisión de inversión				
		A_1 Cuenta de ahorros	A_2 Bonos	A_3 Acciones AAA	A_4 Acciones especulativas	A_5 Acciones de riesgo
E_1: Recesión	0.30	$600	$500	-$2 500	- $ 5 000	-$10 000
E_2: Estable	0.50	600	900	800	400	-5 000
E_3: Expansión	0.20	600	900	4 000	10 000	20 000
Pago esperado (PE)		$600	$780	$ 450	$ 700	-$ 1 500

18.9 Hasta aqui, se han presentado situaciones en las que pueden ocurrir consecuencias económicas tanto positivas como negativas. Cuando las consecuencias son todas ellas valores de costos, pueden aplicarse las mismas técnicas suponiendo que todas las cifras de costos se identifican como valores negativos (pagos). Sin embargo, con frecuencia se identifican las cifras de costos como "costos" pero se reportan como valores absolutos, sin el signo negativo. En estos casos, debe tenerse presente que el costo mínimo, y no el máximo, es el costo "mejor", y debe ajustarse la aplicación de los diversos criterios de decisión en forma correspondiente. Utilizando la Tabla 18.20, determine las mejores decisiones desde el punto de vista de los criterios maximín y maximax, modificando en forma apropiada la aplicación de esos criterios.

Tabla 18.20 Tabla de costos asociados con tres planes de inspección de productos

Proporción de productos defectuosos en el embarque	Plan de inspección		
	A_1 Inspección al 100%	A_2 Inspección al 5%	Ninguna inspección
E_1:0.01	$100	$ 20	$ 0
E_2: 0.05	100	150	400

Como el menor costo es el mejor y el mayor costo el peor, cuando se aplica el criterio maximín, el objetivo consiste en seleccionar la acción cuyo costo máximo sea el menor. En este contexto, al criterio maximín se le podría denominar el criterio "minimax". Tal como se señala en la Tabla 18.21, la decisión que satisface ese criterio es A_1: inspección al 100%. Utilizando este criterio, se minimiza el costo máximo que puede ocurrir.

En el contexto de una tabla de costos, utilizar el criterio maximax implica elegir la decisión cuyo valor mínimo sea el menor. Por ello, al criterio se le podría denominar "minimin" en este caso. Como se muestra en la Tabla 18.21, el criterio que satisface ia orientación optimista que representa el criterio maximax (es decir, el que minimiza el costo mínimo) es A_3: ninguna inspección.

Tabla 18.21 Plan de Inspección de calidad que debe elegirse de acuerdo con tos criterios maxlmín y maximax

Proporción de productos defectuosos en el embarque	Plan de inspección		
	A_1 Inspección al 100%	A_2 Inspección al 5%	A_3 Ninguna inspección
E_1: 0.01 E_2:0.05	\$100 100	\$ 20 150	\$ 0 400
Costo máximo	**\$100**	\$150	\$400
Costo mínimo	\$100	\$ 20	**\$ 0**

18.10 Determine la mejor decisión desde el punto de vista del criterio del arrepentimiento minimax para la situación del problema 18.9.

Observe la Tabla 18.22. Como la mejor decisión para un estado determinado es aquella que da como resultado el costo menor, a esa decisión se le asigna una pérdida de utilidad de cero. Por ello, dado E_1, A_3 es la mejor decisión ya que tiene el menor costo y, por ello, un arrepentimiento de \$0. La decisión que minimiza el arrepentimiento máximo que puede ocurrir es A_2: inspección del 5%.

Tabla 18.22 Tabla de pérdidas de oportunidad para el problema del plan de inspección de calidad, y aplicación del criterio de arrepentimiento minimax

Proporción de productos defectuosos en el embarque	Plan de inspección		
	A_1 Inspección al 100%	A_2 Inspección al 5%	A_3 Ninguna inspección
E_1: 0.01 E_2:0.05	\$100 0	\$20 50	\$ 0 300
Arrepentimiento máximo	\$100		\$300

18.11 Del problema 18.9, determine la mejor decisión desde el punto de vista del criterio bayesiano general de maximizar el pago esperado, dada una probabilidad de 0.80 de que el embarque contenga una proporción de 0.01 defectuosos, y dada una probabilidad del 0.20 de que el embarque tenga una proporción de 0.05 defectuosos.

Tabla 18.23 Tabla de costos para el problema del plan de inspección de calidad, y aplicación del criterio bayesiano

Proporción de productos defectuosos en el embarque	Probabilidad	Plan de inspección		
		A_1 Inspección al 100%	A_2 Inspección al 5%	A_3 Ninguna inspección
$\boldsymbol{E_1}$**: 0.01** $\boldsymbol{E_2}$**: 0.05**	0.80 0.20	\$100 100	\$ 20 150	\$ 0 400
Costo esperado		\$100		\$ 80

Cuando se manejan cifras de costos, se satisface el criterio bayesiano eligiendo la decisión que tenga el mínimo costo esperado. La Tabla 18.23 es una tabla completa de decisión para este problema, y muestra que la mejor decisión desde este punto de vista es A_2: inspección del 5%. Debe observarse que, en términos del costo esperado (a largo plazo), la acción menos preferida es la estrategia "conservadora" de inspeccionar el 100%.

ANÁLISIS DE ÁRBOLES DE DECISIÓN

18.12 A un fabricante se le presentó una proposición para un producto nuevo y debe decidir desarrollarlo o no. El costo del desarrollo del proyecto es $200 000 000; la probabilidad de éxito es 0.70. Si el desarrollo no tiene éxito, se termina el proyecto. Si tiene éxito, el fabricante debe entonces decidir si el nivel de producción ha de ser alto o bajo. Si la demanda es alta, el aumento en la utilidad, dado un nivel elevado de producción, es de $700 000 000; dado un nivel bajo, es de $150 000 000. Si la demanda es baja, el incremento en la utilidad, dado un nivel elevado de producción, es $100 000 000; dado un nivel bajo, es $150 000 000. Todos estos incrementos en utilidades son cifras brutas (es decir, *antes* de restar los $200 000 000 del costo de desarrollar el producto). Se estima que la probabilidad de una demanda elevada es P - 0.40, y para la demanda baja es P - 0.60. Elabore el árbol de decisiones para esta situación.

La figura 18-2 muestra el árbol de decisión para este problema. Como es usual en estos diagramas, los puntos de decisión se identifican mediante un cuadro, y los eventos aleatorios se identifican mediante círculos.

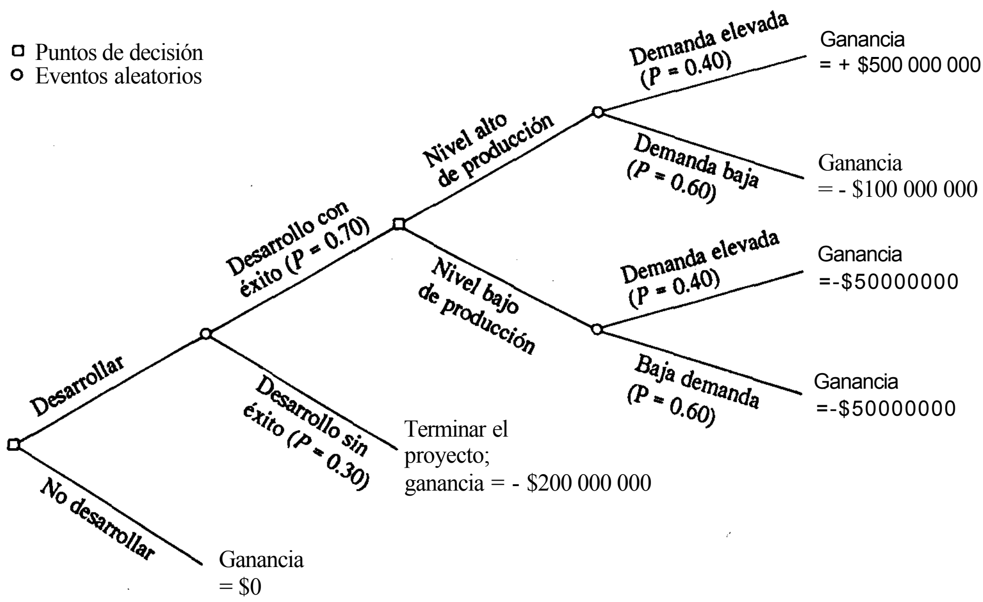


Fig. 18-2

18.13 Con referencia a la figura 18-2, determine si el fabricante debe intentar el desarrollo de este producto, o no, determinando la ganancia esperada correspondiente a las acciones alternativas "desarrollar" y "no desarrollar".

En la figura 18-3 se repite el árbol de decisión presentado en la figura 18-2, pero ahora se incluyen las ganancias esperadas correspondientes a cada una de las decisiones posibles del proceso secuencial, y se eliminan las acciones no preferidas de cada decisión cruzando con dos rayas la rama correspondiente. Trabajando de derecha a izquierda, se determina la ganancia esperada de la decisión "nivel alto de producción", y que es \$140 000 000, de la siguiente manera:

$$PE \text{ (nivel alto de fabricación) - (0.40) (500 000 000) + (0.60) (-100 000 000) - \$140 000 000.}$$

De manera similar,

$$PE \text{ (nivel bajo de fabricación) - (0.40) (-50 000 000) + (0.60) (-50 000 000) = \$50 000 000.}$$

Al comparar las dos ganancias esperadas, la mejor decisión es "nivel alto de fabricación"; se elimina la posibilidad de la otra acción. Pasando a la izquierda hacia el siguiente punto de decisión (que es también el punto inicial de decisión en este caso), las ganancias esperadas para las dos posibles decisiones son

$$\text{EP(Desarrollar) - (0.70)(140 000) + (0.30)(-200 000) - \$38 000}$$
$$\text{EP(No desarrollar) - \$0}$$

Al comparar las dos ganancias esperadas, la mejor acción en el punto inicial es "desarrollar". Al calcular la ganancia esperada de "desarrollar", debe observarse que la probabilidad de 0.70 (para "desarrollo con éxito") se multiplica por \$140 000 000, y se ignora la rama adyacente de -\$50 000 000, debido a que corresponde a una decisión eliminada en la etapa previa de análisis.

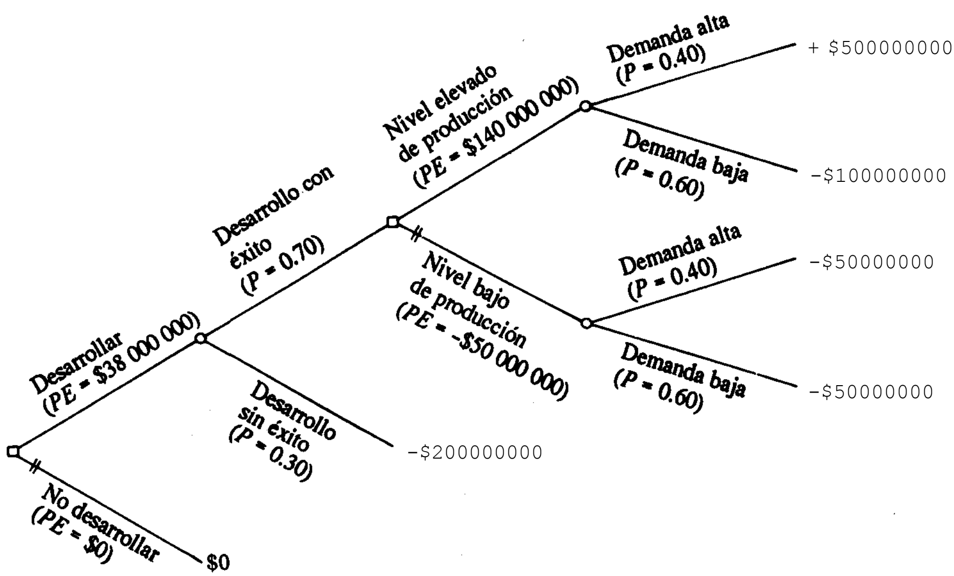


Fig. 18-3

FUNCIONES DE UTILIDAD Y UTILIDAD ESPERADA

18.14 En la Tabla 18.2 se observa que los dos resultados monetarios extremos son=$3 000 000 y +$12 000 000. Supóngase que quien toma las decisiones indica que le seria indiferente recibir una cantidad cierta de $2 400 000 o participar en una situación de riesgo en la que existe una oportunidad del 50% de ganar $12 000 000, y una probabilidad del 50% de perder $3 000 000. Siguiendo la convención de asignar a los valores monetarios extremos inferior y superior valores de utilidad de "0" y "1.0", respectivamente, determine los valores de utilidad asociados con=$3 000 000, $2 400 000 y $12 000 000.

Se comienza asignando valores de utilidad a las dos consecuencias monetarias extremas:

U(-$10 000 000)-0 U($12 000 000) =1.0

La utilidad de la cantidad cierta de $2 400 000 se determina de la siguiente manera:

U(cantidad cierta) = *P(U* de un resultado elevado) + (1= P) (U de un resultado bajo)
U($2 400 000)= 0.50(1.0) + 0.50(0)
U($2 400 000)=0.50

18.15 Al continuar con el problema 18.14, suponga que se le presentan cuatro contratos de referencia adicionales a la persona que toma las decisiones (véase la Tabla 18.24). Determine los valores de utilidad para cada una de las cantidades ciertas.

Tabla 18.24 Cuatro contratos de referencia con sus correspondientes cantidades ciertas

Número del contrato	Probabilidad de ganar $12 000 000	Probabilidad de perder $3 000 000	Cantidad cierta equivalente
1	0.10	0.90	-$2 000000
2	0.30	0.70	0
3	0.70	0.30	6000000
4	0.90	0.10	10000000

Contrato Núm. 1:

Cantidad cierta = 0.10(+$12 000 000) vs. 0.90(-$3 000 000)
-$1000 000 = 0.10(+$12 000 000) vs. 0.90(-$3 000 000)
U(-$2 000 000) =0.10(1.0)+0.90(0)-0.10

Contrato Núm. 2:

Cantidad cierta = 0.30(+$12 000 000) vs. 0.70(-$3 000 000)
$0 = 0.30(+$12 000 000) vs. 0.70(-$3 000 000)
U($0) = 0.30(1.0)+0.70(0)= 0.30

Contrato Núm. 3:

Cantidad cierta = 0.70(+$12 000 000) vs. 0.30(-$3 000 000)
$6 000 000 = 0.70(+$12 000 000) vs. 0.30(-$3 000 000)
U($6 000 000) = 0.70(1.0)+0.30(0)-0.70

Contrato Núm. 4:

$$\text{Cantidad cierta} = 0.90(+\$12\ 000\ 000) \text{ vs. } 0.10(-\$3\ 000\ 000)$$
$$\$10\ 000\ 000 = 0.90(+\$12\ 000\ 000) \text{ vs. } 0.10(-\$3\ 000\ 000)$$
$$U(\$10\ 000\ 000) = 090(1.0)+0.10(0)-0.90$$

En la Tabla 18.25 se resumen los valores de utilidad equivalentes a los diversos valores monetarios determinados antes y en el problema 18.14. Observe que, cuando la situación de riesgo del contrato de referencia implica los dos resultados extremos con valores de utilidad de "1.0" y "0", los valores de utilidad de la cantidad cierta fijada por quien toma las decisiones es siempre igual a la probabilidad del resultado que tiene la utilidad asignada de 1.0.

Tabla 18.25 Valores de utilidad y valores monetarios equivalentes

Valor monetario	-$1500000	-1000000	0	1200000	3000000	5000000	6000 000
Valor de utilidad	0	0.10	0.30	0.50	0.70	0.90	1.00

18.16 Con referencia a la Tabla 18.25, construya una gráfica para ilustrar la función de utilidad de este tomador de decisiones. Describa su actitud con respecto al riesgo que se manifiesta en los contratos de referencia que le presentaron.

En la figura 18-4 se observa que este tomador de decisiones evita los riesgos en el rango de consecuencias monetarias manejadas en este problema. Observe la figura 18-1. Básicamente, muestra que la cantidad cierta que considera equivalente a la situación de riesgo designada en el contrato de referencia siempre es inferior al pago

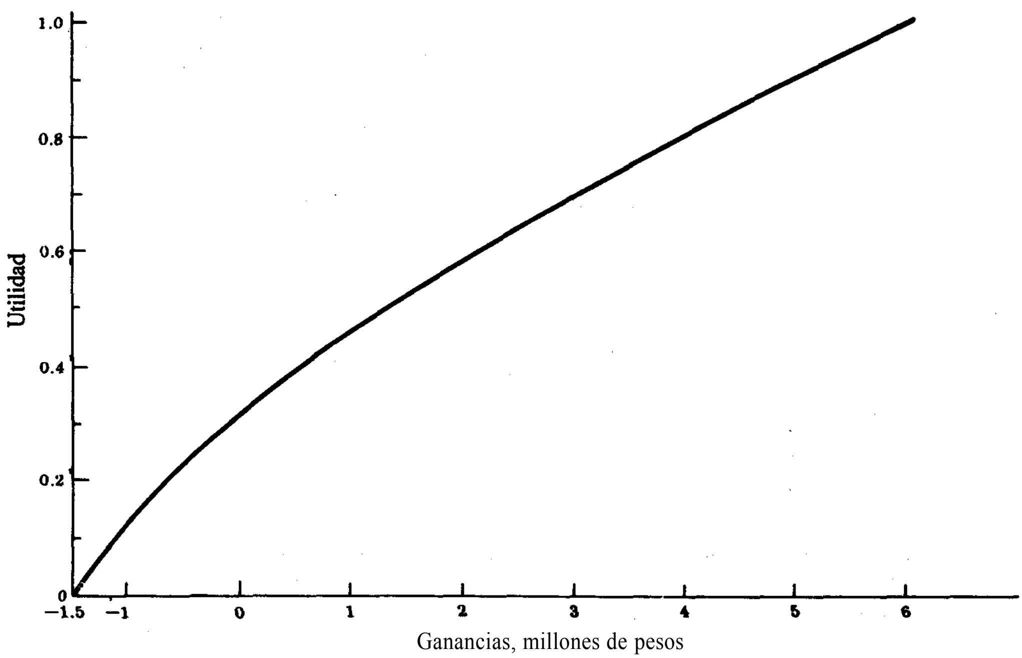


Fig. 18-4

esperado correspondiente, excepto por los dos puntos extremos de la curva de utilidad. Por ejemplo, en el problema 18.14, esta persona planteó que, para ella, una cantidad cierta de $1 200 000 era equivalente a una probabilidad de 0.50 de ganar $12 000 000 y una probabilidad de 0.50 de perder $3 000 000. Sin embargo, el pago de esta situación de riesgo es 0.50 ($12 000 000) + 0.50 (-$3 000 000) - $6 000 000 + (-$1 500 000) = $4 500 000.

18.17 Determine los valores de utilidad aproximados para cada uno de los pagos de la Tabla 18-2, con referencia a la función de utilidad de la figura 18-4, y determine la mejor decisión desde el punto de vista de la maximización de la utilidad esperada.

En la Tabla 18.26 se identifican los valores de utilidad equivalentes a los pagos de la Tabla 18.2. Utilizando el criterio de maximización de la utilidad esperada, la mejor acción *es* A_3, al igual que para el criterio del pago esperado (ejemplo 6). Sin embargo, observe que mientras que los pagos esperados para las acciones A_3 y A_A eran iguales en la Tabla 18.7, los valores de utilidad esperados de estos dos actos no son iguales.

Tabla 18.26 Tabla de utilidad para el número de unidades de aire acondicionado que deben ordenarse, y selección de la mejor acción, de acuerdo con el criterio de la utilidad esperada

Demanda del mercado	Probabilidad	Cantidad del pedido			
		A_1: 5	A_2:10	A_3:15	A_1: 20
E_1:5	0.10	0.55	0.38	0.21	0
E_2:10	0.30	0.55	0.70	0.61	0.47
E_3:15	0.40	0.55	0.70	0.86	0.78
E_4:20	0.20	0.55	0.70	0.86	1.00
Utilidad esperada *(UE)*		0.550	0.668	0.720	0.653

Problemas complementarios

TABLAS DE PAGOS

18.18 Un analista de inversiones estima que existe una probabilidad de aproximadamente 50% de un "auge" en la industria química durante el primer trimestre del año, y que las probabilidades de "ningún cambio" y de "recesión" son más o menos iguales. Un cliente está considerando invertir $10 000 000 en un fondo de inversión que se especializa en acciones comunes de la industria química, o invertir ese dinero en bonos AAA, que producen 28% anual. Si la industria química experimenta un auge durante el primer trimestre, el valor de las acciones de la sociedad de inversión (incluyendo dividendos) aumentará en 35.0% en los próximos doce meses. Si no existe cambio, el valor aumentará en 3.0%. Si ocurre una recesión, el valor *se reducirá* en 30.0%. Ignorando los costos por comisiones, construya una gráfica de pagos para este problema de inversión.

18.19 Para la decisión sobre inversiones que se describió en el problema 18.18, determine las mejores decisiones desde el punto de vista de los criterios (a) maximín y (o) maximax.

Resp. (a) A_2*:* nvertir en bonos AAA, (b) A_1: invertir en un fondo de inversión.

18.20 Determine la mejor decisión, desde el punto de vista del criterio del arrepentimiento minimax, para el problema 18.18.

Resp. A_2: invertir en bonos AAA

18.21 Para el problema 18.18, determine la mejor decisión desde el punto de vista del criterio del pago esperado.

Resp. A_2: invertir en bonos AAA

18.22 Un vendedor al menudeo adquiere cierto producto en $3 000 por caja y lo vende en $5 000 por caja. El elevado margen de utilidad refleja que los productos son perecederos, puesto que no tienen ningún valor después de cinco días. Con base en experiencias con productos similares, el vendedor confía en que la demanda para el artículo está entre 9 y 12 cajas, inclusive.

Construya la tabla de ganancias apropiada. Determine la mejor decisión desde el punto de vista del criterio (a) maximín y (b) maximax.

Resp. (a)A_1: ordenar 9 cajas, (b) A_4: ordenar 12 cajas

18.23 Construya la tabla de pérdidas de oportunidad (arrepentimientos) para el problema 18.22, y determine la mejor decisión desde el punto de vista del criterio de arrepentimiento minimax.

Resp. A_2: ordenar 10 cajas

18.24 Al continuar con el problema 18.22, el vendedor estima que los valores de probabilidad asociados con la venta de 9 a 12 cajas del producto son 0.30,0.40, 0.20 y 0.10, respectivamente. Identifique las mejores decisiones desde el punto de vista de (a) el criterio de probabilidad máxima y (b) la esperanza del evento.

Resp. (a) A_2: ordenar 10 cajas, (b) A_2: ordenar 10 cajas

18.25 Con referencia a los problemas 18.22 y 18.23, determine la mejor cantidad del pedido, desde el punto de vista de (a) el criterio del pago esperado y (b) el criterio de minimización de la pérdida de oportunidad esperada. Demuestre que esos criterios son equivalentes identificando también las acciones que son mejores en segundo y tercer lugar, con respecto a cada criterio.

Resp. (a)A_2: ordenar 10 cajas, (b) A_2: ordenar 10 cajas

18.26 Junto con la instalación de una nueva máquina, el propietario de un barco que hace viajes especializados tiene la oportunidad de comprar hélices de refacción, a $600 000 cada una, para utilizarlas durante la siguiente temporada de trabajo. Las hélices se ajustan especialmente para el barco, y como el propietario ya ha firmado un contrato para vender la nave después de la temporada, las hélices extra que se tengan después de haber terminado la temporada carecen de valor. Si no se tiene una hélice de repuesto en forma inmediata cuando se le requiere, el costo de comprarla, incluyendo el tiempo muerto de navegación, es $1 200 000. Con base en su experiencia, el propietario del barco estima que las probabilidades correspondientes a necesitar de 0 a 3 hélices en la próxima temporada son 0.30, 0.30, 0.30 y 0.10, respectivamente. Construya una tabla de costos para este problema, y determine el número de hélices que deben ordenarse desde el punto de vista del criterio (a) maximín, y (b) maximax.

Resp. (a) A_4: ordenar tres hélices, (b) A,: no ordenar ninguna hélice extra

18.27 Para la situación del problema 18.26, determine la mejor decisión desde el punto de vista de (a) el criterio de la probabilidad máxima y (b) la esperanza del evento.

Resp. (a) indiferente entre 0,1 y 2 hélices; (b) A_2: ordenar una hélice

18.28 Para el problema 18.26, determine la mejor decisión desde el punto de vista del criterio de la ganancia esperada (en este caso, costo esperado).

Resp. A_2: ordenar una hélice

18.29 Varias semanas antes de la temporada de vacaciones, un grupo de comerciantes debe decidir si ordena colchones y sombrillas de playa, o nada, para venderlos en un centro de recreo. Como se indica en la Tabla 18.27, el éxito monetario de la decisión de vender esos artículos depende de las condiciones del clima. Determine las mejores acciones desde el punto de vista de los criterios (a) maximín y (b) maximax.

Resp. (a) A_3: nada, (B) A_2: sombrillas de playa

Tabla 18.27 Pagos para el problema de decisión sobre la temporada vacacional

Clima	A_1: Colchones de playa	A_2: Sombrillas de playa	A_3: Ninguna
Frío	$100	-$ 80	$0
Caliente	-50	150	0

18.30 Determine la mejor acción desde el punto de vista del criterio del arrepentimiento minimax del problema 18.29.

Resp. A_3: nada

18.31 Continuando con el problema 18.29, uno de los comerciantes visita la oficina del servicio meteorológico y se entera de que, en los últimos diez años, el clima ha sido "caliente" en seis de los diez dfas que dura la temporada vacacional. Utilice esta información para determinar las probabilidades correspondientes a las dos condiciones del clima. Después, determine las mejores acciones desde el punto de vista (a) del criterio de probabilidad máxima y (b) del criterio de la ganancia esperada.

Resp. (a) A_2: sombrillas de playa, (b) A_2: sombrillas de playa

ANÁLISIS DE ARBOLES DE DECISIÓN

18.32 Un inversionista está considerando colocar un depósito de $10 000 000 para reservar una oportunidad de concesión en una nueva área residencial durante un año. Existen dos áreas de incertidumbre asociadas con esa situación de decisiones secuenciales: en primer lugar, es posible que un competidor importante de otra concesionaria decida abrir un expendio en la misma área y, por otro lado, no se sabe si esa área residencial crecerá para convertirse en un mercado moderado o grande. El inversionista estima que existe una probabilidad de .50 de que el sistema competidor de concesiones establezca una tienda. Por ello, lo primero que el inversionista debe decidir es si realiza el pago inicial de $10 000 000. Después de que se sepa la decisión del sistema competidor, el inversionista debe decidir si ha de proceder o no a construir su propio expendio. Si existe competencia y el mercado es grande, la ganancia neta para el periodo de Interés se estima en $15 000 000; si el mercado es moderado habrá una pérdida neta de $10 000 000. Si no existe competencia y el mercado es grande, la ganancia neta será de $30 000 000; si el mercado es moderado, habrá una ganancia de $10 000 000. El inversionista estima que existe una probabilidad de aproximadamente .4% de que el mercado sea grande. Utilizando el análisis de árbol de decisiones, determine si debe realizarse o no el depósito inicial de $10 000 000.

Resp. Realizar el depósito *(PE* - $9 000 000).

FUNCIONES DE UTILIDAD Y UTILIDAD ESPERADA

18.33 Para los problemas 18.6 y 18.8, los puntos extremos de las posibles consecuencias monetarias son -$10 000 000 y $20 000 000. Describa lo que usted haría para desarrollar una función de utilidad para la persona que debe tomar esta decisión de inversión.

18.34 Con base en su respuesta al problema 18.33, suponga que se ha obtenido el conjunto de valores de utilidad que se reporta en la Tabla 18.28. Construya la curva de utilidad correspondiente y utilícela para determinar los términos en los que puede describirse al inversionista con respecto al riesgo: lo evita, lo busca o es neutral. Ilustre el significado de su conclusión en el contexto del problema.

Resp. Al tomador de decisiones le gusta asumir riesgos.

Tabla 18.28 Valores de utilidad equivalentes a los valores monetarios, para el problema de decisiones sobre inversión

Valor monetario	-10 000	-2 500	2000	10 000	14 000	18 000	20 000
Valor de utilidad	0	0.1	0.2	0.4	0.6	0.8	1.0

18.35 Determine la mejor decisión para la situación que se describe en los problemas 18.6 a 18.8, con base en el criterio de la utilidad esperada, y utilizando la función de utilidad que se construyó en el problema 18.34.

Resp. A_5: Invertir en acciones de riesgo (UE = 0.230)

Análisis bayesiano de decisiones: El uso de información muestral

19

19.1 EL VALOR ESPERADO DE LA INFORMACIÓN PERFECTA *(VEIP)*

Como los diversos eventos posibles, o estados, son ciertos y están asociados con una distribución de probabilidad, la disponibilidad de *Información perfecta* indica que quien toma las decisiones sabe qué evento ocurrirá para cada una de las alternativas individuales de decisión. El *valor esperado de la Información perfecta (VEIP)* es la diferencia entre la ganancia esperada (a largo plazo) dada esa información, menos la ganancia esperada correspondiente a la mejor decisión bajo condiciones de incertidumbre. Aunque la información perfecta como tal raras veces está disponible, el *VEIP* sirve como indicación del valor máximo que cualquier muestra puede tener para quien toma las decisiones. Al determinar el *VEIP,* debe observarse que, bajo condiciones de incertidumbre, la "mejor acción" se identifica como la estrategia consistente para cada una de las alternativas de acción. Por otro lado, la disponibilidad de información perfecta conduce a una estrategia mixta, en la que la decisión está perfectamente asociada con el evento, o estado, para cada una de las alternativas. La fórmula para determinar el valor esperado de la información perfecta, en donde *GEIP* designa la *ganancia esperada con información perfecta,* es

$$VEIP = GEIP - PE(bajo \text{ condiciones de incertidumbre}) \tag{19.1}$$

EJEMPLO 1. Con referencia a la Tabla 18.7, de la página 306, en donde la mejor acción bajo condiciones de incertidumbre *es A_3:* ordenar 15 unidades, con una ganancia esperada de $6 500 000. Si la información perfecta se tuviera disponible, se elegiría la acción A_1 el 10% de las veces, se elegiría la acción A_2 el 30 % de las veces, etc., de acuerdo con las frecuencias relativas de los correspondientes niveles de demanda. Observe la Tabla 19.1. Como se señala, la ganancia esperada con información perfecta es $8 100 000. Por ello, el *VEIP* = 8100 000 - 6 500 000 - $1 600 000. Esto es lo más que valdría cualquier información muestral, en promedio, porque es la cantidad máxima en que se podría aumentar la ganancia esperada, eliminando toda la incertidumbre.

Tabla 19.1 Determinación del pago esperado con información perfecta para la situación de decisión de la Tabla 18.7

Demanda del mercado	Probabilidad	Mejor acción	Valor condicional de la mejor acción	Pago esperado
E_1: 5	0.10	A_1	$ 3 000 000	$ 300 000
E_2:10	0.30	A_2	6 000 000	1 800 000
E_3:15	0.40	A_3	9000 000	3 600 000
E_4:20	0.20	A_4	12 000 000	2 400 000
				$8 100 000

Considérese ahora la relación entre la pérdida de oportunidad y el valor de la información perfecta. Por definición, la *POE* de la mejor acción es la cantidad promedio de arrepentimiento para la acción óptima, en calidad de promedio a largo plazo. Véase la sección 18.4. Como la *POE* de la mejor acción es la cantidad promedio que se pierde sobre las mejores ganancias posibles dada una estrategia perfecta (combinada), se sigue que debe ser también el valor de la información perfecta, para una situación determinada de decisiones. Por ello, una base alternativa para determinar el valor esperado de la información perfecta es

$$VEIP = POE \text{ (mejor acción)} \tag{19.2}$$

EJEMPLO 2. En el ejemplo 7 del capítulo 18 (página 307), se determinó que la mejor decisión (A_3) era la misma, independientemente de si se identificaba la acción con la *PE* más alta o con la *POE* menor. Ahora, puede observarse además que la *POE* de la mejor acción ($1 600 000) indica también el *VEIP* para esta situación. Este valor corresponde al valor que se obtuvo en el ejemplo 1 anterior.

Deben observarse en los ejemplos anteriores la diferencia entre la ganancia esperada *con* información perfecta, y el valor esperado de la información perfecta. La ganancia esperada *con* información perfecta (GE/P) indica el resultado a largo plazo cuando se elimina toda la incertidumbre, mientras que el valor esperado de la información perfecta (*VEIP)* es el promedio de *incrementos* de valor al eliminar la incertidumbre en las decisiones.

Al igual que en el capitulo 18, en éste se restringe el análisis a los eventos que se ajustan a una distribución discreta de probabilidad. En el capitulo 20 se analiza la forma de determinar el valor de la información perfecta y el valor de la información muestral para eventos que tienen una distribución de probabilidad normal.

19.2 DISTRIBUCIONES DE PROBABILIDAD *A PRIORI* Y *A POSTERIORI*

En la sección 5.7 se utiliza el teorema de Bayes para determinar la probabilidad de un primer evento, dado que ya ha ocurrido un segundo evento especifico. Esta sección presenta la aplicación del teorema de Bayes para modificar las *diversas* probabilidades asociadas con el conjunto total de posibles eventos, o estados, en una situación de decisiones. En este contexto, la *distribución da probabilidades a priori* es la distribución de probabilidad aplicable antes de recolectar información muestral. Con frecuencia, esas distribuciones de probabilidad son subjetivas en el análisis bayesiano de decisiones, puesto que se basan en juicios, aunque también pudieran basarse en información histórica. La *distribución de probabilidad a posteriori* es la distribución de probabilidad que se tiene después de observar la información muestral, y después de utilizarla para modificar la distribución de probabilidad a *priori,* aplicando el teorema de Bayes.

Como se ilustra en la sección 5.7, para poder aplicar el teorema de Bayes, deben conocerse la probabilidad a *priori* del evento incierto, y la probabilidad condicional del resultado muestral. Típicamente, se determinan las probabilidades condicionales utilizando alguna de las distribuciones estándar de probabilidad, de acuerdo con la naturaleza de la situación de muestreo.

EJEMPLO 3. Cuando un proceso de manufactura se encuentra bajo control, la proporción de productos defectuosos es de sólo 0.01; cuando está fuera de control, la proporción es 0.10. Históricamente, se ha encontrado que el proceso está fuera de control el 5% de las veces, y sobre esta base se presentan los valores de probabilidad de la Tabla 19.2. Se inspecciona una muestra de n = 10 productos y se encuentra que uno de ellos está defectuoso. Para utilizar la fórmula de

Tabla 19.2 Valores de probabilidad previos correspondientes a los dos niveles posibles de fracción de productos defectuosos en un proceso de manufactura

Fracción de productos defectuosos	Probabilidad previa
0.01	0.95
0.10	0.05

Bayes con el propósito de modificar los valores de la probabilidad *a priori,* la distribución binomial de probabilidad (sección 6.3) sirve de base para determinar la probabilidad condicional del resultado muestral, dada cada fracción de productos defectuosos. Por ejemplo, la probabilidad de que x - 1 producto esté defectuoso, dada una muestra *n* = 10, y una fracción de p = 0.01, es 0.0914, de acuerdo con el apéndice 2. Debe observarse que en este contexto, el valor de *"p"* de la tabla binomial es la fracción de productos defectuosos y *no* el valor de la probabilidad *a priori* asociado con esa fracción de productos defectuosos. El diagrama de árbol para esta situación de muestreo (Fig. 19-1) incluye todas las probabilidades condicionales, asi como también las a *priorí.*

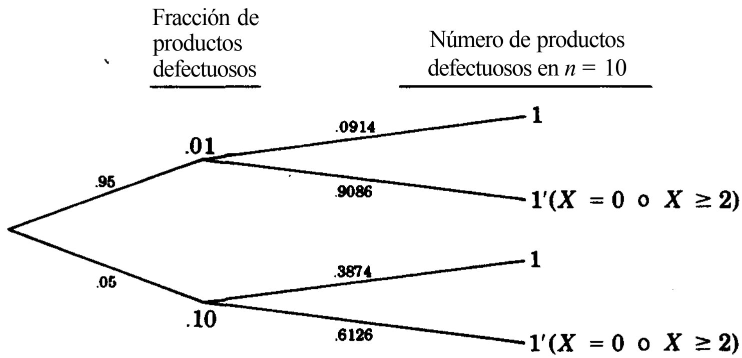


Fig. 19-1

Al utilizar la fórmula (5.76), se determina la probabilidad a *posteriori* de que la verdadera fracción de artículos defectuosos sea 0.01, dado que se ha encontrado que uno de los elementos de la muestra n = 10 es defectuoso.

$$P(A\mid B)=\frac{P(A)P(B\mid A)}{P(A)P(B\mid A)+P(A')P(B\mid A')}$$

$$P(p=0.01\mid 1\text{ def.})=\frac{P(p=0.01)\times P(1\text{ def}\mid p=0.01)}{P(p=0.01)\times P(1\text{ def}\mid p=0.01)+P(p=0.10)\times P(1\text{ def}\mid p=0.10)}$$

$$=\frac{(0.95)(0.0914)}{(0.95)(0.0914)+(0.05)(0.3874)}=\frac{0.08683}{0.10620}=0.81761\cong 0.82$$

Ahora podría aplicarse también la fórmula de Bayes para determinar la probabilidad de que la verdadera fracción de productos defectuosos sea 0.10, dado el resultado muestral obtenido. Sin embargo, como sólo son posibles dos niveles de fracción de productos defectuosos, este valor es evidentemente el complemento del 0.82 que se determinó antes, o también 0.18. En la Tabla 19.3 se resumen los valores de probabilidad a *priori* y a *posteriori* para este ejemplo. Debe observarse que el cambio en los valores de probabilidad después de la muestra refleja el hecho de que observar un producto defectuoso en una muestra aleatoria de *n* = 10 es más probable con la fracción de defectuosos de 0.10 que con la de 0.01.

Tabla 19.3 Distribuciones de probabilidad *a priori* y *a posteriori*

Fracción de productos defectuosos	Probabilidad *a priori*	Probabilidad *a posteriori*
0.01 0.10	0.95 0.05	0.82 0.18

En el ejemplo 3 se presenta una aplicación directa de la metodología de la sección 5.7. Sin embargo, cuando se trata de modificar un conjunto completo de valores de probabilidad, resulta más conveniente un *enfoque tabular* para elaborar la distribución de probabilidades a *posterior!,* y no la aplicación repetida de la fórmula de Bayes para determinar cada uno de los valores de probabilidad posteriores.

Tabla 19.4 Modificación de la probabilidad para el problema de fracción de productos defectuosos

(1) Evento	(2) *P a priori*	(3) Probabilidad condicional del resultado muestral	(4) Probabilidad conjunta Col. (2) x Col. (3)	(5) *P aposteriori* Col. (4) / suma 6
0.01	0.95	0.0914	0.08683	0.81761 - 0.82
0.10	0.05	0.3874	0.01937	0.18239 = 0.18
Total	1.00		0.10620	1.00

EJEMPLO 4. En la Tabla 19.4 se ilustra el uso del enfoque tabular para elaborar la distribución de probabilidad a *posteriori,* para la situación que se describe en el ejemplo 3. Para cada uno de los eventos, se determina el valor de la probabilidad a *posteriori,* dividiendo la probabilidad conjunta de la celda correspondiente entre la suma de la columna (4). Mediante este procedimiento, cada una de las anotaciones de la columna (4) es equivalente al numerador de la fórmula de Bayes, en tanto que la suma de la columna (4) es equivalente al denominador. Tal como se señala, se obtienen los mismos

19.3 ANÁLISIS BAYESIANO *A POSTERIORi* Y EL VALOR DE LA INFORMACIÓN MUESTRAL (DESPUÉS DEL MUESTREO)

Puede aplicarse ahora el procedimiento mediante el cual se utiliza la información muestral para revisar una distribución de probabilidad a *priori* a un problema de decisión que se refiere a si se ha de aceptar o no un embarque de un proveedor. Primero se determina la mejor acción con base en la distribución a *priori* y, después, con base en una distribución *a posteriori.* Finalmente, se determina el valor del resultado muestral especifico con base en si la identificación de la mejor acción resultó modificada al revisar los valores de las probabilidades a *priori.*

EJEMPLOS. La Tabla 19.5 muestra que la fracción (proporción) de productos defectuosos, en un embarque de 1 000 de ellos, puede darse en uno de cuatro niveles: 0.01,0.05,0.10 y 0.20. Se incluyen en la tabla los valores de la probabilidad a *priori* con base en la experiencia histórica de ese proveedor. Los costos para A_1: Aceptar, se basan en el hecho de que la identificación y eliminación de un producto defectuoso que se convierte en parte de un componente ensamblado cuesta $1 000. Por ello, para una proporción de productos defectuosos de 0.01, existen 0.01 X1 000 = 10 productos defectuosos, que implican un costo de 10 X $1 000 = $10 000 por su eliminación a *posteriori.* Los costos asociados con A_2*:* Rechazar, se basan en la información de que si la verdadera proporción de productos defectuosos es 0.05 o menos, la compañía está contractualmente obligada a aceptar el embarque y, por ello, debe pagar los gastos adicionales del embarque cuando el proveedor lo devuelve a la compañia. Con base en la distribución de probabilidad a *priori,* la acción que tiene el menor costo esperado es A_2*:* Aceptar y, por ello, en ausencia de información muestral, deberla aceptarse el embarque en forma rutinaria. De hecho, si la mejor acción fuera otra, serfa indicación de que se ha contratado al proveedor equivocado

Tabla 19.5 Tabla de decisión para el problema del embarque, utilizando la distribución *a priori*

Fracción de productos defectuosos (p_i)	*(F) a priori*	Costo condicional de A_1: Aceptar	Costo condicional de A_2: Rechazar	Costo esperado de A_1: Aceptar	Costo esperado de A_2: Rechazar
0.01	0.50	$10000	$200000	$ 5000	$100 000
0.05	0.30	50000	200000	15 000	60 000
0.10	0.10	100000	0	10 000	0
0.20	0.10	200000	0	20000	0
Total	1.00			$50000	$160000

Mientras que el análisis del ejemplo 5 se basa en la distribución de probabilidad a *priori,* en el ejemplo 6 se supone que existe información muestral disponible que se utiliza para modificar la distribución de probabilidad *a priori.*

EJEMPLO 6. Para la situación que se describe en el ejemplo 5, suponga que se selecciona una muestra aleatoria de $n = 10$ productos y que se encuentra que dos de ellos están defectuosos. Utilizando el procedimiento tabular que se describió en la sección 19.2, se modifica la distribución de probabilidad a *priori* en la Tabla 19.6. Como era de esperarse, los valores de probabilidad *a posteriori* correspondientes a las proporciones de productos defectuosos de 0.10 y 0.20 son mayores que los respectivos valores de probabilidad a *priori,* debido al resultado muestral. Las probabilidades condicionales de la Tabla 19.6 se basan en la distribución binomial de probabilidad. Al igual que en la sección 19.2, observe que, cuando se utiliza la tabla binomial, los valores de *"p"* son cada una de las proporciones correspondientes de productos defectuosos, y *no* el valor de la probabilidad a *priori.*

Tabla 19.6 Modificación de probabilidades para el problema de los embarques

Fracción de productos defectuosos	*(F) a priori*	Probabilidad condicional del resultado muestral $[P(X = 2 \mid n = 10, p_i)]$	Probabilidad conjunta	*(P) a posteriori*
0.01	0.50	0.0042	0.00210	0.0284 = 0.03
0.05	0.30	0.0746	0.02238	0.3022 = 0.30
0.10	0.10	0.1937	0.01937	0.2616 = 0.26
0.20	0.10	0.3020	0.03020	0.4078 = 0.41
Total	1.00		0.07405	1.00

El ***análisis bayesiano a posteriori*** es el proceso que consiste en determinar la mejor acción revisando la distribución de probabilidad a *priori,* con base en datos muestrales, para después utilizar la distribución de probabilidad a *posteriori* que se obtiene para determinar la mejor decisión. Teniendo disponible la distribución de probabilidad a *posteriori* que se obtuvo en el ejemplo 6, en el ejemplo 7 se termina el análisis bayesiano a *posteriori.*

EJEMPLO 7. En la Tabla 19.7, el costo esperado para A_2: Rechazar ($66 000) es inferior al costo esperado correspondiente *a* A_1: Aceptar. Por ello, dado que la muestra de n= 10 contenia dos productos defectuosos, la mejor acción, desde el punto de vista del criterio bayesiano de minimizar los costos esperados, consiste en rechazar el embarque y devolverlo, en vez de introducir los productos al proceso de manufactura.

Tabla 19.7 Tabla de costos para el problema de los embarques, utilizando la distribución de probabilidad a *posteriori*

Fracción de productos defectuosos (P_i)	*(P) a posteriori*	Costo condicional de A_1: Aceptar	Costo condicional de A_2: Rechazar	Costo esperado de A_1: Aceptar	Costo esperado de A_2:Rechazar
0.01	0.03	$ 10 000	$200000	$ 300	$ 6000
0.05	0.30	50000	200000	15 000	60000
0.10	0.26	100000	0	26 000	0
0.20	0.41	200000	0	82 000	0
Total	1.00			$123,300	$66000

Al haber ilustrado en los ejemplos 5 a 7 el proceso del análisis bayesiano a *posteriorí,* ahora es posible considerar el valor asociado con una muestra que ya se ha tomado. El *valor de la información muestral (VIM)* estimado para una muestra que ya ha sido observada, se basa en la diferencia entre las *ganancias esperadas a posteriorí* (o costos) de las mejores acciones que se identifican antes y después del muestreo, o

$$\begin{aligned} VIM = &\text{(pago esperado posterior de la mejor acción posterior)} \\ &- \text{(pago esperado posterior de la mejor acción previa)} \end{aligned} \tag{19.3}$$

Por ello, si identificar la mejor acción no cambia el resultado de un análisis a *posteriori,* entonces el valor de la información muestral es $0, aun cuando la ganancia esperada correspondiente a esa decisión hubiera cambiado en términos generales.

EJEMPLO 8. Para estimar el valor de la información muestral del ejemplo 7, observe en primer lugar que, como resultado del análisis a *posteriori,* se modifica la mejor decisión para que sea ahora *A¿* Rechazar, y ya *no* A_1. Aceptar, la cual fue aplicable sin la muestra. El valor estimado de la información muestral que se obtiene es

$$VIM = -66\,000 - (-123\,300) = -66\,000 + 123\,300 - \$57\,300$$

(Nota: Como en este caso los dos valores esperados son costos, se anotan como valores negativos para que el *VIM* resulte, como debe ser, positivo. A largo plazo, la información muestral sirvió para modificar las probabilidades a *priori* y obtener una evaluación más precisa. Aun cuando se pensaba que el costo esperado del A_1: Aceptar era $50 000, con base en el análisis a *priori,* considerando la información muestral se reconoce que una evaluación más confiable de ese costo esperado es $123 300. La información muestral obligó a cambiar la decisión, de esta última, a la decisión A_2: Rechazar, con un costo esperado (posterior) de $66 000, reduciendo de esta manera el costo esperado en $57 300, que es lo que se hubiera experimentado de haber elegido la decisión A_1.*)*

19.4 ANÁLISIS PREPOSTERIOR: EL VALOR ESPERADO DE LA INFORMACIÓN MUESTRAL *(VEIM)* ANTES DEL MUESTREO

El *análisis preposterior* es el proceso mediante el cual se estima el valor de la información muestral antes de recolectar la muestra. El procedimiento básico consiste en considerar todos los resultados posibles muestrales, determinar el valor estimado de cada uno de los resultados muestrales para el proceso de decisión y, después, determinar los *valores esperados de la información muestra/ (VEIM)* ponderando cada uno de esos distintos valores con la probabilidad de que ocurra el resultado muestral correspondiente.

En el ejemplo 9 se ilustra el proceso del análisis preposterior para una muestra de solamente *n* =1. Como es necesario considerar todos los posibles resultados muestrales, el cálculo del *VEIM* para muestras que no son pequeñas resulta muy laborioso. Sin embargo, el procedimiento que se ilustra enseguida puede aplicarse a muestras más grandes utilizando una computadora.

EJEMPLO 9. Suponga que desea determinar el valor esperado de la Información muestral correspondiente a una muestra de *n* = 1, para el problema de decisión que se describe en el ejemplo 5. La Tabla 19.5 es la tabla de costos para este problema, y con base en la distribución a *priori,* la mejor acción *es* A_1: Aceptar el embarque, con un costo esperado de $50 000 (en contraste *con* A_2: Rechazar el embarque, con un costo esperado de $160 000). Se procede determinando el *VIM* correspondiente a cada uno de los posibles resultados muestrales, y después se combinan esos valores para calcular el *VEIM* para la muestra de *n* =1.

Paso 1: Determinar la distribución a *posteriori* si se encuentran 0 productos defectuosos en la muestra de *n* = 1. En la Tabla 19.8 se presentan el procedimiento de revisión y la distribución de probabilidad a *posteriori.* Observe que los valores de probabilidad han "cambiado" un tanto en la dirección esperada. Observe también que la suma de la columna de "probabilidad conjunta" se identifica indicando la probabilidad global (incondicional) del resultado muestral que se observa. Este valor de probabilidad se utiliza en la etapa final de determinación del *VEIM.*

Tabla 19.8 Modificación de la probabilidad para el problema de los embarques, dado que no hay productos defectuosos en una muestra de *n* - 1

Fracción de productos defectuosos (P_i)	(P) *a priori*	Probabilidad condicional de resultado muestral $[P(X=0\|n=1,P_i)]$	Probabilidad conjunta	*(P) a posteriori*
0.01	0.50	0.99	0.495	0.5211=0.521
0.05	0.30	0.95	0.285	0.3000 = 0.300
0.10	0.10	0.90	0.090	0.0947 = 0.095
0.20	0.10	0.80	0.080	0.0842 = 0.084
Total	1.00		$P(X=0)$ = 0.950	1.000

Paso 2: Determinar la mejor acción si se encuentran cero productos defectuosos en la muestra de *n* = 1, y determinar los pagos (costos) esperados correspondientes a las acciones alternativas. Como se muestra en la Tabla 19.9, la mejor acción es A_1: Aceptar el embarque, que es la misma acción que se consideró mejor con base en el análisis previo (antes de utilizar la información muestral) en el ejemplo 5.

Tabla 19.9 Determinación de la mejor acción para el problema de los embarques, dado que no existen productos defectuosos en una muestra de n=1

Fracción de productos defectuosos (P_i)	*(P) a posteriori*	Costo condicional de A_1: Aceptar	Costo condicional de A_2: Rechazar	Costo esperado de A_1:Aceptar	Costo esperado de A_2: Rechazar
0.01	0.521	$ 10000	$ 200000	$ 5 210	$104 200
0.05	0.300	50000	200000	15 000	60 000
0.10	0.095	100000	0	9 500	0
0.20	0.084	200000	0	16 800	0
				Total: $46 510	Total: $164 200

Paso 3: Determinar el valor estimado de la información de que se encuentren cero productos defectuosos en la muestra de n = 1. Como no ha cambiado la mejor acción al observar los resultados muestrales, el valor de la muestra es $0. O, utilizando la fórmula *(19.3),*

VIM = (ganancia esperada posterior de la mejor acción posterior) - (ganancia esperada posterior de la mejor acción previa)
= -46 510 - (-46 510 = -46 510 + 46 510 = $0

Paso 4: Determinar la distribución *a posteriori* si se encuentra un producto defectuoso en la muestra de n = 1. En la Tabla 19.10 se presentan el proceso de revisión y la distribución de probabilidad *a posteriori.* De nueva cuenta, observe que los valores de probabilidad se han modificado en la dirección esperada. También, de nueva cuente la suma de la columna de "probabilidad conjunta* señala la probabilidad global del resultado muestral que se considera, en este caso el de X = 1, dado n =1, y dada la distribución de probabilidad *a priori.*

Tabla 19.10 Modificación de las probabilidades para el problema de los embarques, dado un producto defectuoso en una muestra de n - 1

Fracción de productos defectuosos (p_i)	*(P) a priori*	Probabilidad condicional del resultado muestral $[P(X = 1 \mid n = 1, p_i)]$	Probabilidad conjunta	*(P) a posteriori*
0.01	0.50	0.01	0.005	0.10
0.05	0.30	0.05	0.015	0.30
0.10	0.10	0.10	0.010	0.20
0.20	0.10	0.20	0.020	0.40
			P Total (X = 1) = 0.050	Total: 1.00

Paso 5 : Determine la mejor acción si se encuentra un producto defectuoso en la muestra de n - 1 y las ganancias (costos) esperados correspondientes a las acciones alternativas. Como se muestra en la Tabla 19.11, la mejor acción es A_2: Rechazar el embarque, que difiere de la mejor acción, de acuerdo con la distribución de probabilidad *a priori.*

Tabla 19.11 Determinación de la mejor acción para el problema de los embarques, dado un producto defectuoso en una muestra de n - 1

Fracción de productos defectuosos *(Pd*	*(P)a posteriori*	Costo condicional de A_1: Aceptar	Costo condicional de A_2: Rechazar	Costo esperado de A_1: Aceptar	Costo esperado de A_2: Rechazar
0.01	0.10	$ 10000	$ 200000	$ 1000	$ 20 000
0.05	0.30	50000	200000	15 000	60000
0.10	0.20	100000	0	20000	0
0.20	0.40	200000	0	80000	0
				Total: $ 116000	Total: $ 80 000

Paso 6: Determine el valor estimado de la información de que un producto está defectuoso en la muestra *de* n = 1. Como existen cambios en la mejor acción si se incluyen los resultados observados en la muestra, ese valor se determina utilizando la fórmula *(19.3):*

VIM = (Ganancia esperada posterior de la mejor acción posterior) - (ganancia esperada posterior de la mejor acción previa)
= -80 000 - (-116 000 = -80 000 + 116 000 = $36 000

Paso 7: Como ya se ha determinado el *VIM* correspondiente a cada uno de los posibles resultados muestrales, se combinan estos valores para obtener el valor esperado de la información muestral, a través de la siguiente fórmula:

$$VEIM = \Sigma\,(VIM_i)\,P(X_i) \qquad (19.4)$$

Al aplicar esta fórmula:

$$VEIM = 0(0.950) + 36\,000(0.50) = 0 + 1\,800 = \$1\,800$$

En el cálculo anterior, se multiplica cada uno de los posibles *VIM* por la probabilidad de que ocurra el correspondiente resultado muestral, y se suman estos productos. Como se muestra en las etapas 1 y 4 anteriores, las probabilidades de que ocurran los dos diferentes resultados muestrales para este problema son las que se indican mediante las sumas de las columnas de "probabilidad conjunta" de las Tablas 19.8 y 19.10, respectivamente.

19.5 GANANCIA NETA ESPERADA DEL MUESTREO *(GNEM)* Y EL TAMAÑO ÓPTIMO DE LA MUESTRA

Para cada uno de los tamaños de muestra que se consideran, se define la *ganancia neta esperada del muestreo (GNEM)* como la diferencia entre el valor esperado de la información muestral y el costo de la muestra *(CM):*

$$GNEM = VEIM - CM \qquad (19.5)$$

Si la *GNEM* de la muestra que se considera es superior a 0, entonces debe tomarse la muestra a menos que otra muestra tenga otra *GNEM* mayor. Este tema se contempla en una parte posterior de esta sección, con respecto al tamaño óptimo de la muestra.

EJEMPLO 10. Para el tamaño de muestra de $n = 1$ que se consideró en el ejemplo 9, el *VEIM* = \$1 800. Suponga que el costo de la inspección es de \$1 200 por producto. Se determina la ganancia neta esperada al obtener una muestra de $n = 1$ en esta situación, de la manera siguiente:

$$GNEM = VEIM - CM = 1\,800 - 1\,200 = \$600$$

Por ello, es mejor obtener una muestra de un artículo antes de decidir aceptar o rechazar el embarque, en vez de tomar la decisión sin la información muestral.

Al aumentar el tamaño de la muestra que se considera en el análisis preposterior, el *VEIM* se vuelve cada vez mayor y se aproxima al valor del *VEIP,* para esa situación. El *tamaño óptimo de la muestra* es aquel que permite maximizar la *GNEM.* En la figura 19-2 se ilustra la relación entre el *VEIP,* el *VEIM,* y el *CM,* conforme se aumenta el tamaño de la muestra. En la figura 19-3 se indica la relación general entre el tamaño de la muestra y la *GNEM.*

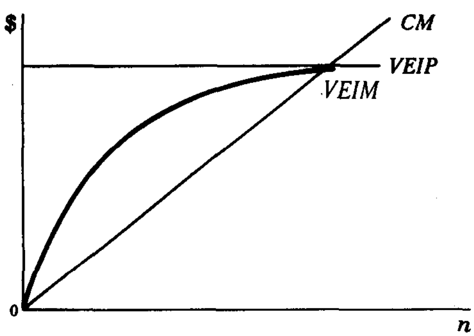


Fig. 19-2

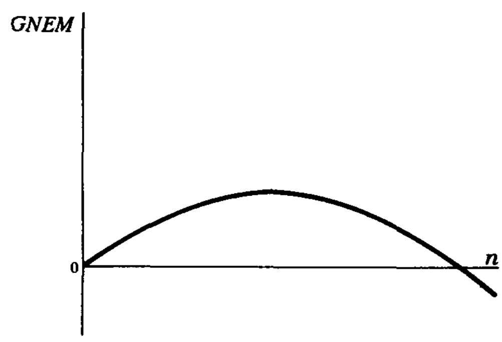


Fig. 19-3

EJEMPLO 11. En la Tabla 19.12, y de acuerdo con el orden del tamaño de la muestra,-se presentan los valores de *VEIM, CM* y la *GNEM* para el problema de la inspección de embarques que se revisa en los ejemplos 9 y 10. Los valores correspondientes a n - 0 se basan en la ausencia de una muestra, en tanto que los valores correspondientes a n=1 se determinaron en los ejemplos 9 y 10. Los valores para $n = 2$ se determinaron mediante un procedimiento similar al que se siguió para $n = 1$, excepto que fue necesario considerar tres posibles resultados muestrales (X = 0, X = 1 y X = 2). No se presentan en este capítulo los cálculos correspondientes. Como la *GNEM* siempre tiene una "cumbre" en algún tamaño muestral (incluyendo posiblementen = 0), y después declina, se concluye de la información de la Tabla 19.12 que el tamaño óptimo de la muestra para este problema simplificado es $n = 1$. Para problemas más realistas que implican la consideración de tamaños de muestra más grandes, sería imprescindible el uso de una computadora.

Tabla 19.12 *VEIM, CM* y *GNEM* para el problema de los embarques, de acuerdo con el tamaño de la muestra

Tamaño de la muestra (n)	*VEIM*	*CM*	*GNEM*
0	\$0	\$0	\$0
1	1800	1200	600
2	2 830	2 400	430

En algunos textos en los que se incluye el análisis bayesiano de decisiones, se revisa la determinación de ***la ganancia esperada terminal global (GETG)*** y de la ***ganancia esperada terminal global neta (GETGN)*** junto con el análisis preposterior. Para cada uno de los posibles tamaños de la muestra, la ganancia esperada terminal global es el valor esperado de la mejor acción antes del muestreo, más el valor esperado de la información muestral. Se reduce esta suma mediante el costo del muestreo, con el objeto de determinar la ganancia esperada global neta. Así, la ganancia esperada terminal global es

$$GETG = PE\,(\text{mejor acción previa}) + VEIM \qquad (19.6)$$

La ganancia esperada terminal global neta es

$$GETGN = GETG - CM \qquad (19.7)$$

Finalmente, debe observarse que si es necesario tomar la muestra, entonces la mejor decisión dependerá del resultado muestral observado. Por ello, una regla de decisión bayesiana completa que implique el muestreo exige que se identifique el tamaño óptimo de la muestra y que, además, se especifique la mejor acción de acuerdo con diversos resultados muestrales.

EJEMPLO 12. La Tabla 19.13 es un resumen del problema de inspección de embarques. Además de los valores que se reportaron antes, se indican las *GETG* y *GETGN* para cada uno de los tres tamaños de la muestra que se consideran. Los costos se expresan como valores negativos, para que la substracción que conduce a determinar los valores *GETGN* resulte ser algebraicamente consistente con la fórmula general. Por ello, el tamaño óptimo de la muestra para la Tabla 19.3 es el que tiene el mayor valor (algebraico) de *GETGN,* o $n = 1$. Observe que esta conclusión es consistente con el resultado que se señala en el ejemplo 11 y en la Tabla 19.12. Dado que debe tomarse una muestra de n = 1, se observa también en la Tabla 19.13 que, si el artículo muestreado no está defectuoso (X = 0), la mejor acción es A_1: Aceptar. Si el producto muestreado está defectuoso (X = 1), la mejor acción es A_2: Rechazar.

Tabla 19.13 Resumen del análisis preposterior para el problema de los embarques y la designación de la mejor regla bayesiana de decisión (los costos de inspección se expresan como números negativos)

Tamaño de la muestra (n)	*VEIM*	*GETG*	*CM*	*GNEM*	*GETGN*	Regla de decisión	
						A_1 : Aceptar	A_2: Rechazar
0	$ 0	-$ 50 000*	$ 0	$0	-$50 000	Siempre	Nunca
1 †	1800	-48 200	1200	600	-49 400	$X = 0$	$X = 1$
2	2 830	-47 170	2 400	430	-49 570	$X = 0$	$X > 1$

*Este es el pago (costo) esperado de la mejor acción, con base en la distribución de probabilidad previa y sin muestreo. Vea la Tabla 19.7.
† El tamaño óptimo de la muestra, ya sea maximizando *GNEM* o *GETGN.*

Problemas resueltos

EL VALOR ESPERADO DE LA INFORMACIÓN PERFECTA.

19.1 Determine el *VEIP* para la situación de decisiones sobre inversión de los problemas 18.6 a 18.8, determinando la GOE correspondiente a la mejor acción, con base en el criterio bayesiano de la ganancia esperada.

Como se muestra en la Tabla 18.19, la mejor acción desde el punto de vista del criterio de la ganancia esperada es A_2: Bonos. Las pérdidas de oportunidad para todas las acciones se identifican en la Tabla 18.18. Por ello,

$$VEIP = GOE(A_2) = 100(0.30) + 0(0.50) + 19\ 100(0.20) = \$3\ 850$$

19.2 Determine el *VEIP* para el problema del embarque que se describe en el ejemplo 5, determinando en primer lugar el costo esperado con información perfecta.

Como este problema de decisión se refiere a costos, debe observarse que en este caso se determina el *VEIP* restando el costo esperado con información perfecta del costo esperado de la mejor acción, bajo condiciones de incertidumbre. Con respecto a la Tabla 19.5, con información perfecta, se aceptaría el embarque en los casos en los que la fracción de productos defectuosos sea de 0.01 o 0.05, y se rechazaría en los casos en que sea 0.10 o 0.20. Con base en esta regla de decisión, el costo esperado, y después el *VEIP,* se calcula de la siguiente manera:

$$CE \text{ (con información perfecta)} = 10\ 000(0.50) + 50\ 000(0.30) + 0(0.10) + 0(0.10) = \$20\ 000$$
$$CE \text{ (bajo condiciones de incertidumbre)} = \$50\ 000 \text{ (del ejemplo 5).}$$
$$VEIP = CE \text{ (bajo condiciones de incertidumbre)} - CE \text{ (con información perfecta).}$$
$$=50\ 000 - 20\ 000 = \$30\ 000$$

19.3 En el problema 18.5, la mejor acción con respecto al número de televisiones con pantalla grande que deben ordenarse se determina identificando la acción que tenga la GOE menor. La mejor acción es A_2: Ordenar tres aparatos, y se tiene que GOE (A_2) = $170 000. Con referencia a la Tabla 18.14, demuestre que esta *GOE* es igual al *VEIP* que puede calcularse determinando primero el valor esperado con la información perfecta.

$$PE1P = 400\ 000(0.30) + 600\ 000(0.40) + 800\ 000(0.20) + 1\ 000\ 000(0.10) = \$620\ 000$$
$$PE \text{ (bajo condiciones de incertidumbre)} = \$450\ 000 \text{ (del problema 18.4)}$$
$$VEIP = PE1P - PE \text{ (bajo condiciones de incertidumbre)}$$
$$= 620\ 000 - 450\ 000 = \$170\ 000$$

ANÁLISIS DE DECISIONES PREVIO Y POSTERIOR

19.4 Un producto nuevo que se está considerando en la empresa tendrá un nivel de ventas bajo, moderado o alto. La elección consiste en comercializar o no el producto. La Tabla 19.14 es la tabla de pagos que indica las probabilidades correspondientes a los tres niveles posibles de ventas, con base en el mejor juicio de los administradores, y las consecuencias económicas de esta situación.

(a) Determine la mejor acción desde el punto de vista del criterio bayesiano de maximizar la ganancia esperada.

(b) Determine el valor máximo que puede tener cualquier información adicional.

Tabla 19.14 Tabla de pagos para la comercialización de un producto nuevo (en miles de pesos)

Evento	Probabilidad	Acción de decisión	
		A_1: Comercializar	A_2: No comercializar
E_1: Ventas bajas	0.40	-$30 000	$0
E_2: Ventas moderadas	0.40	10 000	0
E_3: Ventas altas	0.20	50 000	0

(a)
$$GE(A_1) = -30000(0.40) + 10\,000(0.40) + 50\,000(0.20) = \$2\,000$$
$$GE(A_2) = \$0$$

Por lo tanto, la mejor acción es A_1: Comercializar el producto.

(b) En la Tabla 19.15 se ilustra la tabla de pérdidas de oportunidad, o arrepentimientos, para esta situación de decisiones. Como el VEIP es igual a la GOE de la mejor acción,

$$\text{VEIG} = \text{GOE}(A_1) = 30\,000(0.40) + 0(0.40) + 0(0.20) = \$12\,000$$

De manera alternativa, puede determinarse el VEIP calculando primero la ganancia esperada con la información perfecta, y después restando el valor esperado de la mejor acción (bajo incertidumbre) de ese valor:

$$\text{GEIP} = 0(0.40) + 10\,000(0.40) + 50\,000(0.20) = \$14\,000$$
$$\text{VEIG} = \text{GEIP} - \text{GE (bajo condiciones de incertidumbre)}$$
$$= 14\,000 - 2\,000 = \$12\,000$$

Tabla 19.15 Tabla de pérdida de oportunidad para la comercialización de un producto nuevo (en miles de pesos)

Evento	Probabilidad	Acción de decisión	
		A_1: Comercializar	A_2: No comercializar
E_1: Ventas bajas	0.40	$ 30 000	$ 0
E_2: Ventas moderadas	0.40	0	10 000
E_3: Ventas altas	0.20	0	50 000

19.5 Para la situación que se describe en el problema 19.4, una prueba de mercado de producto en una región da como resultado un nivel moderado de ventas. Una revisión de los patrones históricos de ventas en esa región señala que, para productos con ventas bajas a nivel nacional, las ventas en esa región fueron moderadas el 30% de las veces. Para productos con un nivel moderado de ventas a nivel nacional, las ventas en esa región fueron moderadas el 70% de las veces. Para productos con un nivel elevado de ventas nacional, las ventas en la región fueron moderadas sólo el 10% de las veces.

(a) Modifique las probabilidades a *priori* que se utilizaron en el problema 19.4, con base en el resultado muestral de que las ventas fueron moderadas en la región de prueba de mercado.

(b) Determine la mejor decisión con base en el análisis de decisiones *a posteriori.*

(a) En la Tabla 19.16 se determina la distribución de probabilidad a *posteriori.* Tal como sería de esperar, la probabilidad de que el verdadero nivel de la demanda sea moderado aumenta después de tomar en consideración los resultados de la muestra.

(b) La mejor acción después de la muestra (la acción bayesiana posterior) se determina utilizando las probabilidades a *posteriori* que se determinan en la Tabla 19.16, junto con los valores condicionales que se presentaron en la Tabla 19.14. Las ganancias esperadas correspondientes a las acciones A_1 (Comercializar) y A_2 (No comercializar) son:

$$PE(A_1) = -30\,000(0.29) + 10\,000(0.67) + 50\,000(0.05) = \$500$$
$$PE(A_2) = \$0$$

Por lo tanto, al igual que en el caso del análisis previo, la mejor acción es A_1: Comercializar el producto.

Tabla 19.16 Modificación de las probabilidades para la comercialización de un producto nuevo, con base en un nivel moderado de ventas en el mercado de prueba

Evento	*(P) a priori*	Probabilidad condicional del resultado muestral	Probabilidad conjunta	*(P) a posteriori*
E_1: Ventas bajas	0.40	0.30	0.12	0.2857 = 0.29
E_2: Ventas moderadas	0.40	0.70	0.28	0.6666 = 0.67
E_3: Ventas altas	0.20	0.10	0.02	0.0476 = 0.05
Total	1.00		0.42	1.01

19.6 ¿Cuál es el valor estimado de la información muestral que se obtuvo en el problema 19.5? ¿Cómo se reconcilia este valor con la diferencia observada en el pago esperado, en esta situación de riesgo, después y antes de la muestra?

Como la acción óptima no resultó modificada por el resultado muestral, la muestra no tenía valor como tal. En forma más específica,

$$\begin{aligned} VIM &= (\text{ganancia esperada posterior de la mejor acción posterior}) \\ &\quad - (\text{pago esperado posterior de la mejor acción previa}) \\ &= 500 - 500 = \$0 \end{aligned}$$

La diferencia entre los valores posterior y previo de la acción óptima es 500 000 - 2 000 000 = - $1 500 000. En otras palabras, la ganancia esperada posterior es $1 500 000 menos que la ganancia esperada previa para la misma acción. Sin embargo, este cambio se presentó porque la muestra ofreció mayor información con respecto a los estados inciertos, y en este sentido la información muestral no ocasionó ningún cambio como tal. Sin la muestra, se determinó que la ganancia esperada de la mejor acción era $2 000 000, pero después de la muestra se reconoce que, para esta situación específica, debe cambiar la expectativa previa. Sin embargo, como todavía se elegiría la misma acción como la óptima, este conocimiento no ocasiona cambios en el resultado monetario esperado eventual, con respecto a lo que hubiera ocurrido sin la información muestral.

19.7 El propietario de una pequeña planta manufacturera tiene la oportunidad de adquirir cinco tornos usados, en conjunto, del equipo de desecho de una gran empresa de la región. El factor incierto es el número de máquinas que requerirán un ajuste general antes de que puedan estar en posibilidades de usarse. La decisión de compra debe tomar en consideración el costo de las máquinas, el valor que tienen si estuvieran en buenas condiciones, y el costo del ajuste general, las diversas consecuencias económicas de esta decisión se dan en la Tabla 19.17. Determine la mejor acción en este caso, y la ganancia esperada correspondiente a esta acción.

Tabla 19.17 Tabla de decisión para la compra de tomos usados (en miles de pesos)

Número de tornos defectuosos	Probabilidad	Acto de decisión	
		A_1= Comprar	A_2: No comprar
0	0.05	\$5 000	\$0
1	0.10	3 000	0
2	0.20	1000	0
3	0.30	-1000	0
4	0.30	-3 000	0
5	0.05	-5 000	0

$$PE(A_1) = 5\,000(0.05) + 3\,000(0.10) + 1\,000(0.20) + (-1\,000)(0.30) + (-3\,000)(0.30) + (-5\,000)(0.05)$$
$$= -\$700$$
$$PE(A2) = \$0$$

Por ello, la mejor acción es A_2: No comprar las máquinas con una ganancia esperada de \$0.

19.8 Determine el valor esperado de la información perfecta para la situación de decisión que se describe en el problema 19.7.

$$PEIP = 5\,000(0.05) + 3\,000(0.10) + 1000(0.20) + 0(0.30) + 0.(0.30) + 0(0.05) = \$750$$
$$PE \text{ (bajo condiciones de incertidumbre)} \quad = \quad \$0 \quad \text{(del problema 19.10)}$$
$$VEIP = PEIP - PE \text{ (bajo condiciones de incertidumbre)} - 750 - 0 = \$750$$

19.9 Con referencia al problema 19.7, suponga que se permite realizar una inspección minuciosa de cada una de las máquinas, pero que esa inspección cuesta \$50 000 cada una. El propietario de la compañía manufacturera elige dos máquinas al azar, hace que las inspeccionen, y se entera de que ninguna de las dos necesita el ajuste general. Determine la mejor acción con base en un análisis bayesiano posterior.

Como la población (cinco máquinas) es finita y de tamaño relativamente reducido, la distribución de probabilidad que se utiliza para determinar las probabilidades condicionales de los resultados muestrales debe tomar en consideración el hecho de que se trata de una situación de muestreo sin reemplazo. Pueden determinarse las probabilidades necesarias utilizando diagramas de árbol, o mediante la distribución hipergeométrica (sección 6.5). Para la población de cinco máquinas que incluye dos defectuosas, por ejemplo, puede utilizarse la figura 19-4, junto con la regla de multiplicación para eventos dependientes, para obtener la probabilidad condicional que se requiere (véase la sección 5.6). En esta figura, D representa una máquina defectuosa observada, y D' representa una máquina observada que no está defectuosa. Tal como se señala, la probabilidad de no observar ningún producto defectuoso en una muestra de n = 2 para la población dada es 0.30. Pueden determinarse de manera similar las probabilidades condicionales para otras poblaciones.

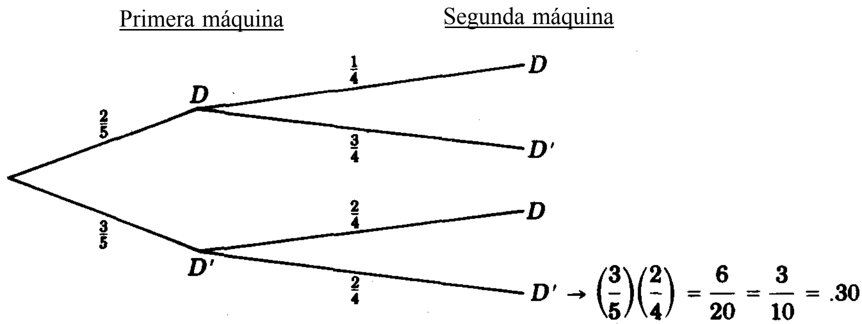


Fig. 19-4

La modificación de la distribución de probabilidad a *priori,* con base en el resultado que se encontró que de dos máquinas muestreadas ninguna era defectuosa, se presenta en la Tabla 19.18 (los valores de la probabilidad condicional de la tercera columna se determinaron por alguno de los dos procedimientos que se describieron antes). Utilizando la distribución de probabilidad a *posteriori,* se calculan los pagos esperados posteriores, de la siguiente manera:

$PE\ (A_1)=5\ 000(0.25) + 3\ 000(0.30) + 1\ 000(0.30) + (-1000)(0.15) + (-3\ 000)(0.00) + (-5\ 000)(0.00)$
$-\ \$2\ 300$

$PE(A_2)\ =\ \$0$

Por lo tanto, la mejor acción bayesiana posterior es A_1: Comprar las máquinas con un pago esperado de $2 300 000

Tabla 19.18 Modificación de las probabilidades para la compra de los tornos usados con base en un resultado de ninguna máquina defectuosa en una muestra *de n = 2*

Número de tornos defectuosos	(P) *a priori*	Probabilidad condicional del resultado muestral	Probabilidad conjunta	(P) *a posteriori*
0	0.05	1.00	0.05	0.25
1	0.10	0.60	0.06	0.30
2	0.20	0.30	0.06	0.30
3	0.30	0.10	0.03	0.15
4	0.30	0	0	0
5	0.05	0	0	0
Total	1.00		P(X=0) = 0.20	1.00

19.10 ¿Cuál es el valor neto estimado de la información muestral que se obtuvo en el problema 19.9?

VIM = (valor esperado posterior de la mejor acción posterior)
- (valor esperado posterior de la mejor acción previa)
= 2 3 0 0 - 0 = $2 300
WM neto = VIM - CM = 2 300 - 100 = $2 200

Observe que el valor neto calculado arriba no es la *GNEM* que se describe en la sección 19.5, porque el valor del párrafo anterior es la ganancia neta estimada correspondiente a una muestra específica que ya se ha tomado, y no el valor esperado de una muestra de un tamaño determinado, previo al muestreo.

ANÁLISIS PREPOSTERIOR

19.11 En la sección 19.5 se indicó que es posible determinar la ganancia esperada total global *(GETG)* sumando el *VEIM para* el tamaño dado de la muestra al *PE* (previo). Como es lógico, también es posible determinar la *GETG* identificando la *PE* (posterior) correspondiente a cada uno de los posibles resultados muestrales y después ponderando esas ganancias con sus respectivas probabilidades, para determinar la ganancia esperada terminal global. Aplique este procedimiento al análisis preposterior del ejemplo 9, y demuestre que el valor resultante de la GETG es el mismo que se reportó en la Tabla 19.13.

La fórmula alternativa para la ganancia esperada terminal global es

$$GETG = \Sigma\, (PE_i \text{ posterior})P(X_i) \qquad (19.8)$$

Las cantidades *GE* posteriores posibles para la muestra de *n* - 1 se reportan en las Tablas 19.9 y 19.11, en tanto que las correspondientes probabilidades de los resultados muestrales se reportan en las Tablas 19.8 y 19.10. Sustituyendo estos valores en la ecuación anterior,

$$GETG = (-46\,510)(0.95) + (-80\,000)(0.05) = -\$48\,180$$

Excepto por los errores de redondeo, este es el mismo pago (costo) negativo, que los - $48 200 que se reportaron en la Tabla 19.13.

Problemas complementarios

EL VALOR ESPERADO DE LA INFORMACIÓN PERFECTA

19.12 Para los problemas 18.22 a 18.25, determine el valor esperado de la información perfecta, restando las ganancias esperadas de la mejor acción bajo incertidumbre, de la ganancia esperada con información perfecta. Compare su respuesta con la pérdida de oportunidad esperada de la mejor acción, según se determinó en el problema 18.25.

Resp. VEIP = 20.20 - 18.50 = $1.70

19.13 Para los problemas 18.26 a 18.28, determine el valor esperado de la información perfecta concerniente al número de hélices extra que se requieren.

Resp. VEIP = $480 000

ANÁLISIS DE DECISIONES PREVIO Y POSTERIOR

19.14 Un comerciante tiene la oportunidad de adquirir un embarque de diez sistemas estereofónicos de alta calidad fabricados en el extranjero en $1 000 000. Sin embargo, el equipo ha estado en tránsito en el barco carguero durante algún tiempo, y existe alguna posibilidad real de que los aparatos hayan sido dañados por la humedad. Con base en su experiencia con la compañía transportista, el comerciante estima que existe una probabilidad del 20% de que el embarque esté dañado. Si el embarque está dañado, el comerciante puede vender los aparatos estereofónicos por solamente $500 000. Si no se han dañado los aparatos, puede revender el embarque para obtener una utilidad neta de $300 000. Determine si esta persona debe o no comprar el embarque, desde el punto de vista del criterio de la ganancia esperada.

Resp. Comprar, con *GE* = $140 000

19.15 Con respecto a la decisión del comerciante del problema 19.14, para propósitos prácticos, "dañado" significa que la mitad de los aparatos se han visto afectados, porque un daño más amplio sería evidente. El comerciante paga \$10 000 para hacer que se seleccione al azar un aparato, se desempaque y se pruebe, y encuentra que funciona de manera adecuada. Determine la mejor acción dada esta información muestral, y la ganancia esperada correspondiente.

Resp. Comprar. *GE* = \$212 000

19.16 ¿Cuál es el valor estimado de la información muestral que se obtuvo en el problema 19.15?

Resp. \$0, o considerando el costo del muestreo, -\$10 000.

19.17 Con referencia a las decisiones sobre inversión que se describieron en los problemas 18.18 a 18.21, determine *(a)* la ganancia esperada con información perfecta *(GEIP)* y (b) el *VEIP* para el problema.

Resp. (a)\$1 150, (b) \$350

19.18 Para los problemas 18.18 a 18.21, a través de algún enfoque simplificado, el analista de inversiones define "aumento" el que (cuando menos) el 70% de los compradores regulares de productos químicos aumenten el monto de sus pedidos; "sin cambio" el que (aproximadamente) el 50% de los compradores aumenten el monto de sus pedidos, y "reducción" el que el 30% (o menos) de los compradores aumenten el monto de sus pedidos. Entrevista a una muestra de 20 compradores de productos químicos y se entera de que 14 de ellos están aumentando el monto de sus pedidos, con respecto a pedidos anteriores.

(a) Modifique la distribución de probabilidad previa correspondiente a los tres estados posibles de la industria química, según se dan en el problema 18.18, tomando en consideración este resultado muestral.

(b) Utilizando la distribución de probabilidad a *posteriori,* determine la mejor acción desde el punto de vista del criterio de la ganancia esperada, y compárela con la respuesta al problema 18.21.

Resp. (a) P (Auge) = 0.91, *P* (sin cambio) = 0.09, *P* (depresión) = 0.00;
(o) Invertir en un fondo de inversión *(PE* = \$1 392 000)

19.19 Estime el valor de la información muestral que se obtuvo en el problema 19.18.

Resp. \$592 000

19.20 Al utilizar la distribución de probabilidad a *posteriori* que se determinó en el problema 19.18, determine (a) el pago esperado con información perfecta y (b) el *VEIP* después de que se ha tomado la muestra. Compare sus resultados con los que se obtuvieron en el problema 19.17.

Resp. (a) \$1 437 000, (b) \$45 000

19.21 Con referencia al problema 18.29, suponga que el club no necesita comprometerse a comprar los colchones o las sombrillas de playa sino hasta 3 días antes del evento. Por ello, los miembros deciden completar el pronóstico del meteorólogo con información adicional. Observando el historial del meteorólogo, los días que han estado de hecho frescos, él ha pronosticado en forma correcta el tiempo el 90% de las veces. Para los días que son de hecho calurosos, ha pronosticado de forma correcta el clima el 70% de las veces. Para el día del carnaval, el meteorólogo pronostica clima fresco. Modifique la distribución de probabilidad previa que se dio en el problema 18.31, con base en este pronóstico.

Resp. P (Fresco) - 0.67, P (Caluroso) - 0.33

19.22 Al tomar en consideración el pronóstico del clima que se da en el problema 19.21, determine la mejor decisión para la situación que se describe en el problema 18.29.

Resp. Av. Ordenar colchones de playa.

19.23 Estime el valor del pronóstico del tiempo que se da en el problema 19.21.

Resp. \$54 600

ANÁLISIS PREPOSTERIOR

19.24 En el ejemplo 9 se determinó que el *VEIM* para una muestra de n - 1 fuera de \$1800. Utilizando el procedimiento de cálculo que se ilustró en ese ejemplo 9, determine el *VIM* correspondiente a una muestra de $n = 2$, y siendo que se encuentra que el número de productos defectuosos es $X = 0$, $X = 1$ y $X = 2$, respectivamente.

Resp. \$0, \$23 220 y \$134 140

19.25 Con referencia a las cifras condicionales del *VEIM* que se calcularon en el problema 19.24, determine el *VEIM* correspondiente a una muestra *de* $n = 2$.

Resp. \$2 830

Análisis bayesiano de decisiones:Aplicación de la distribución normal

20

20.1 INTRODUCCIÓN

En este capitulo se revisa el análisis bayesiano de decisiones para un evento (estado) que es una variable aleatoria continua con distribución normal, y no una variable aleatoria discreta. En la sección 20.2 se ilustra la metodología para determinar la media y la desviación estándar para una distribución de probabilidad a *priori.* En este caso, las técnicas de análisis se basan además en el requerimiento de que sólo se evalúen o comparen dos decisiones, y que las funciones de pago correspondientes a esas acciones sean funciones lineales de pago (sección 20.3). Aunque puede parecer que estos requerimientos son demasiado restrictivos, es posible analizar una amplia gama de decisiones mediante estas técnicas.

EJEMPLO 1. Una empresa que fabrica aparatos electrónicos ha decidido ensamblar un aparato de localización por radio para distribuirlo después. La decisión específica que debe tomarse ahora es si se habrá de instalar en la planta de ensamble el equipo Tipo 1 o Tipo 2. Se requiere una mayor inversión de capital para el equipo Tipo 2, pero el costo variable de manufactura es menor, lo cual, a su vez, arroja una mayor utilidad, después de haber cubierto el costo fijo (de capital) (véase el ejemplo 5). Sin embargo, el equipo Tipo 2 sería más redituable que el equipo Tipo 1, sólo con niveles de venta relativamente mayores, lo cual es el evento incierto en este problema. Se supone que quien toma las decisiones estima que el nivel de ventas incierto sigue una distribución normal. Como la estimación del nivel de ventas es el factor clave para determinar el tipo de equipo que producirá el mejor rendimiento, resulta ser el área de interés principal en el análisis previo. En este ejemplo, se señala que (1) la incertidumbre con respecto al evento tiene una distribución normal; (2) se consideran dos acciones, comprar el equipo Tipo 1 o el equipo Tipo 2; y (3) existe un aumento constante en las utilidades por cada unidad vendida, lo cual señala la existencia de funciones lineales de pago- para las dos acciones.

20.2 DETERMINACIÓN DE LOS PARÁMETROS DE LA DISTRIBUCIÓN DE PROBABILIDAD NORMAL *A PRIORI*

La distribución de probabilidad a *priori* describe la incertidumbre asociada con la estimación de ese evento aleatorio por parte de quien toma las decisiones. No es el evento el que sigue la distribución de probabilidad, sino más bien, la estimación de la ocurrencia del evento. Como la estimación se basa en un juicio, no existe teorema matemático que justifique el uso de la distribución normal para describir ese juicio en casos específicos. Sin embargo, en las situaciones de juicio en las que la persona que toma las decisiones está consciente de varios factores inciertos que podrían influir sobre el valor del resultado, se ha encontrado que el uso de la distribución normal es una aproximación satisfactoria.

Como se describe en la sección 7.2, se define una distribución normal identificando su media y su desviación estándar. Puede obtenerse la media de la distribución *a priori* pidiendo al tomador de decisiones que identifique el valor "más probable" del evento aleatorio, o pidiéndole un valor tal que exista un 0.5% de probabilidad de que el valor real sea menor y un 0.5% de probabilidad de que sea mayor. Obsérvese que, de hecho, el primer método implica solicitar para la moda de la distribución de probabilidad, en tanto que en el segundo caso se pide la mediana de la distribución. Como se señala en la sección 3.6, la media, la mediana y la moda, se encuentran todas en el mismo punto para una variable con distribución normal. En este texto se designa la media de la distribución de probabilidad previa mediante M_0

EJEMPLO 2. Para el problema de decisión que se presenta en el ejemplo 1, suponga que quien toma las decisiones indica que el nivel nacional más probable de ventas para el aparato de localización por teléfono es de 40 000 unidades durante el periodo de interés. Con base en este juicio, se designa el valor de la media de la distribución previa como M_0 = 40 000 unidades.

La media previa se obtiene describiendo ya sea la moda o la mediana, porque es fácil describir y conceptualizar estas medidas de manera no matemática. En el caso de la desviación estándar, seria aún más difícil obtener en forma directa esa estimación previa. En cambio, quien toma las decisiones proporciona los limites para un intervalo establecido de probabilidad, y se determina la desviación estándar de la distribución previa mediante el tamaño observado del intervalo. En contraste con los intervalos de confianza clásicos que se basan en información muestral, a los intervalos que se basan en juicios se les denomina con frecuencia *Intervalos* de *credibilidad.* El intervalo del "50% central" es el que más se utiliza, puesto que es posible que este rango de resultados sea el más fácil de conceptualizar. Después, se determina la desviación estándar observando que el 50% central de la distribución normal se encuentra contenido dentro de aproximadamente ± 2/3 σ unidades (el valor especific**0.67σ).** Utilizando S_0 para designar la desviación estándar previa, la fórmula de cálculo para determinar la desviación estándar previa utilizando el intervalo de credibilidad del 50% central es

$$\frac{2}{3} S_0 = \frac{\text{intervalo del. 50\% central}}{2} \tag{20.1}$$

o

$$S_0 = \frac{3\,(\text{intervalo del 50\% central})}{4} \tag{20.2}$$

EJEMPLO 3. Si en el problema de decisión analizado en los ejemplos 1 y 2, el tomador de decisiones establece que tiene una confianza del 50%, y el nivel real de ventas se encuentra entre 35 000 y 45 000 unidades, la desviación estándar de la distribución de probabilidad previa es

$$\frac{2}{3} S_0 = \frac{45\,000 - 35\,000}{2} = 5000$$

$$S_0 = 7\,500 \text{ unidades}$$

Con los procedimientos que se ilustran en los ejemplos 2 y 3 se determinan los valores de M_0 y S_0 para la demanda total del mercado. Sin embargo, con frecuencia se analiza la estimación para eventos como ventas con base en los resultados promedio por tienda, en vez de utilizar el mercado global. Esto es asi por dos razones. En primer lugar, un administrador puede encontrar que esto es más fácil. En segundo término, y lo más Importante desde el punto de vista del análisis, si se va a revisar la distribución de probabilidad previa con base en Información muestral, entonces la información muestra! se obtendría de tiendas elegidas al azar. Para utilizar esa información muestral, la distribución previa misma debe estar planteada con base en tiendas individuales.

EJEMPLO 4. Si se piensa recolectar información muestral, al mismo tiempo que se utiliza el juicio de los administradores, pueden convertirse las estimaciones que se obtuvieron en los ejemplos 2 y 3 a una base de tiendas individuales. De manera alternativa, pudiera haberse pedido en primer lugar las estimaciones por tienda. Suponga que el dispositivo de localización telefónica se va a distribuir en un total de 500 tiendas en todo el país. Si se le pide a quien toma las decisiones que diga cuál es el nivel más probable de ventas por tienda, la respuesta deberla ser M_0 = 80, para que fuera consistente con la estimación hecha de 40 000 unidades para el mercado total en el ejemplo 2. De manera similar, debería establecer que existe una probabilidad del 0.5% de que el promedio de ventas por tienda se encuentre entre 70 y 90 unidades, para que sea consistente con el intervalo dado para las ventas totales en el ejemplo 3. Por ello, con base en las tiendas, la desviación estándar previa es

$$\frac{2}{3}S_0 = \frac{90-70}{2} = 10$$

$$S_0 = 15$$

Se ha ilustrado la forma de obtener los parámetros de una distribución de probabilidad previa, suponiendo que la distribución es normal. El estudiante debe prestar atención especial al hecho de que la media previa y la desviación estándar previa describen la estimación de quien toma las decisiones del evento aleatorio, y de su incertidumbre con respecto al evento. Como se supone que la estimación es insesgada, se sigue que M_0 es una estimación de la media poblacional global μ (por ejemplo, el nivel promedio real de ventas de las 500 tiendas). Sin embargo, *no* puede concluirse que S_0 es un estimador de σ. Con respecto a esto, resulta más apropiado pensar que M_0 y S_0 son similares a $\bar{X}$ **y** $\sigma_{\bar{X}}$, respectivamente, pero que se basan en juicios y no en valores muestrales.

20.3 DEFINICIÓN DE LAS FUNCIONES DE PAGO LINEALES Y DETERMINACIÓN DE LA MEJOR ACCIÓN

La existencia de una función lineal de pago indica que la ganancia esperada correspondiente a las acciones es una función lineal con respecto al nivel de incertidumbre. En términos algebraicos,

$$\text{Pago}\,(A_1) = k_1 + b_1X \tag{20.3}$$

$$\text{Pago}\,(A_2) = k_2 + b_2X \tag{20.4}$$

en donde *k* es el factor constante (una erogación de capital seria una constante negativa), *b* es el aumento en los ingresos (o costos) por unidad de la variable aleatoria, y, por ello, indica la pendiente de la recta, y X indica el valor de la variable' aleatoria.

EJEMPLO 5. Para la situación que se describe en el ejemplo 1, suponga que el gasto de capital que se requiere es $350 000 000 para el equipo Tipo 1, y de $450 000 000 para el Tipo 2. Debido al menor costo variable de manufactura correspondiente al equipo Tipo 2, el aumento en las utilidades (con respecto al costo variable) es de $12 000 por unidad cuando se utiliza este equipo, en comparación con $10 000 de aumento en las utilidades por unidad para el Tipo 1. En seguida se representan en forma algebraica las funciones de la ganancia, y se representan gráficamente en la figura 20-1

$$\text{Pago}\,(A_1) = 350\,000 + 10\,X$$

$$\text{Pago}\,(A_2) = 450\,000 + 12X$$

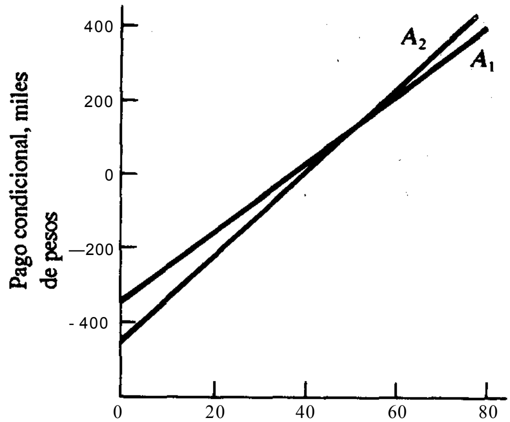


Fig. 20-1

El punto en el que se cruzan las dos funciones lineales de ganancia indica el punto de indiferencia con respecto a elegir entre las dos acciones, porque las ganancias condicionales son iguales en ese punto. A este punto se le denomina con frecuencia *punto de equilibrio* en el análisis estadístico de decisiones. Puede determinarse en forma algebraica el valor específico de la variable aleatoria en el punto de equilibrio resolviendo simultáneamente las dos ecuaciones:

$$k_1 + b_1 X_b = k_2 + b_2 X_b$$

$$b_1 X_b - b_2 X_b = k_2 - k_1$$

$$X_b (b_1 - b_2) = k_2 - k_1$$

$$X_b = \frac{k_2 - k_1}{b_1 - b_2} \tag{20.5}$$

Al haber establecido el punto de equilibrio, puede determinarse gráficamente la mejor acción observando si la media previa se encuentra por arriba o por debajo del punto de equilibrio, y seleccionando la acción que resulte óptima de ese lado del punto de equilibrio. También puede identificarse la mejor acción, sustituyendo el valor de la media previa en cada una de las funciones de pago y eligiendo la acción que tenga el mayor pago esperado. Las ganancias esperadas para las dos decisiones que tienen funciones lineales de pago son:

$$PE(A_1) = k_1 + b_1 M_0 \tag{20.6}$$

$$PE(A_2) = k_2 + b_2 M_0 \tag{20.7}$$

EJEMPLO 6. Para las acciones y las funciones de pago que se describen en el ejemplo 5, el punto de equilibrio es (en donde las cantidades se expresan en miles):

$$X_b = \frac{k_2 - k_1}{b_1 - b_2} = \frac{-450\,000 - (-350\,000)}{10 - 12} = \frac{-100\,000}{-2} = 50\,000 \text{ unidades}$$

Se determinó que la media previa es M_0 = 40 000 unidades en el ejemplo 2. Observando la figura 20-1, se aprecia qué este valor se encuentra a la izquierda del punto de equilibrio, y que la mejor acción es A_1 (comprar el equipo Tipo 1). En forma alternativa, puede determinarse la ganancia esperada asociada con cada acción:

$$PE(A_1) = k_1 + b_1 M_0 = -350\,000 + 10(40\,000) = \$50\,000$$
$$PE(A_2) = k_2 + b_2 M_0 = -450\,000 + 12(40\,000) = \$30\,000$$

Por lo tanto, la mejor acción es A_1 (adquirir el equipo Tipo 1), con una ganancia esperada de $50 000 000. Debe observar que, al determinar la mejor acción, sólo se requiere M_0 para la distribución de probabilidad previa.

El análisis presentado hasta aquí se refiere a la demanda total de mercado. Sin embargo, como se señalaba en la sección 20.2, si se piensa recolectar información muestral, resulta más conveniente un análisis por tienda para el uso subsecuente de los resultados muestrales. En el análisis por tienda, las funciones lineales de pago se basan en las ventas promedio por tienda, en calidad de variable:

$$\text{Pago}(A_1) = k_1 + Nb_1\,\mu \tag{20.8}$$

$$\text{Pago}(A_2) = k_2 + Nb_2\,\mu \tag{20.9}$$

En las ecuaciones anteriores, N es el número de tiendas, **y** μ es el nivel promedio de ventas por tienda. De manera similar, el punto de equilibrio en términos de ventas promedio por tienda debe incluir el número de tiendas:

$$\mu_b = \frac{k_2 - k_1}{Nb_1 - Nb_2} \tag{20.10}$$

Finalmente, las ganancias esperadas correspondientes a las dos acciones, cuando las funciones de pago son lineales, se basan en las ventas promedio por tienda, y se determinan de la siguiente manera:

$$PE(A_1) = k_1 + Nb_1 M_0 \tag{20.11}$$

$$PE(A_2) = k_2 + Nb_2 M_0 \tag{20.12}$$

EJEMPLO 7. Como existen 500 tiendas, cada unidad adicional de ventas promedio por tienda representa 500 unidades adicionales en total. En lugar de las funciones de pago que se determinaron en el ejemplo 5, se tiene.

$$\text{Pago}(A_1) = k_1 + Nb_1\,\mu = -350\,000 + (500)(10)\,\mu = -350\,000 + 5\,000\mu$$

$$\text{Pago}(A_2) = k_2 + Nb_2\,\mu = -450\,000 + (500)(12)\,\mu = -450\,000 + 6\,000\mu$$

En lugar del punto de equilibrio para las ventas totales que se reportan en el ejemplo 6, el punto de equilibrio promedio de ventas por tienda es

$$\mu_b = \frac{k_2 - k_1}{Nb_1 - Nb_2} = \frac{-450\,000 - (-350\,000)}{500(10) - 500(12)} = 100 \text{ unidades por tienda}$$

Del ejemplo 4, M_0 = 80 por tienda. Por ello, la ganancia esperada correspondiente a cada acción, con base en un análisis por tienda, es

$$PE(A_1) = k_1 + Nb_1M_0 = -350\,000 + (500)(10)(80) = \$50\,000$$

$$PE(A_2) = k_2 + Nb_2M_0 = -450\,000 + (500X12X80) = \$30\,000$$

Estos valores son idénticos a los que se obtuvieron mediante el análisis del mercado total en el ejemplo 6.

Si las funciones lineales son costos y no ingresos, la mejor acción es la que tiene el menor costo esperado. Véanse los problemas 20.5 a 20.8.

20.4 FUNCIONES LINEALES DE PERDIDA POR PARTES Y EL VALOR ESPERADO DE LA INFORMACIÓN PERFECTA *(VEIP)*

En el contexto del análisis estadístico de decisiones, el término *función de pérdida* siempre se refiere a una función de pérdida de oportunidad. Para el problema de dos acciones con funciones lineales de pago, se tiene que la función de pérdida correspondiente a cada acción es una función lineal por partes, conformada por dos elementos lineales, o segmentos. Esto es asi porque de un lado del punto de equilibrio la pérdida de oportunidad asociada a una acción especifica será cero, en tanto que del otro lado la pérdida de oportunidad aumenta de manera lineal con cada unidad adicional, a partir del punto de equilibrio. La mejor manera de presentar el proceso para determinar las funciones lineales de pérdida por partes es a través de un ejemplo.

EJEMPLO 8. En los ejemplos 4 a 6 se determinaron las siguientes funciones de pago y punto de equilibrio:

$$\text{Pago}(A_1) = k_1 + b_1X = -\$350\,000 + 10X$$

$$\text{Pago}(A_2) = k_2 - b_2X = -\$450\,000 + 12X$$

$$X_b = 50\,000$$

Se observó en el ejemplo 6 que la acción A_1 es la óptima (sin pérdida de oportunidad) cuando el nivel de ventas se encuentra por debajo del punto de equilibrio, y que la acción A_2 es la óptima cuando el nivel de ventas se encuentra por encima del punto de equilibrio. Para la acción $A_{1(}$ si el nivel real de ventas es superior al punto de equilibrio, entonces la pérdida de oportunidad es la diferencia entre los dos aumentos en los ingresos (\$12 - \$10) multiplicado por el número de unidades en las que se excede el punto de equilibrio. Específicamente,

$$PO(A_1, X)\begin{cases} = \$2.00(X - 50\,000) & \text{para } X > X_b(=50\,000) \\ = 0 & \text{para } X \le X_b(=50\,000) \end{cases}$$

Por otro lado, existe un incremento de \$2, para la acción A_2, en la pérdida de oportunidad, por cada una de las unidades en las que el nivel real de ventas está por debajo del punto de equilibrio:

$$PO(A_2, X)\begin{cases} = 0 & \text{para } X \ge X_b(=50\,000) \\ = \$2.00(50\,000 - X) & \text{para } X < X_b(=50\,000) \end{cases}$$

Las funciones lineales de pérdida por partes que se elaboraron antes para A_1 y A_2 se ilustran gráficamente en las figuras 20-2 y 20-3, respectivamente.

En el ejemplo 8 se observa que la pérdida de oportunidad unitaria para la acción A_1 es igual a la pérdida de oportunidad unitaria para la acción A_2, cuando ocurre cada tipo de pérdida. En otras palabras, la pendiente de la función de pérdida de oportunidad, en términos de valor absoluto es la misma para ambas acciones, en la porción por partes que no es igual a cero. Resulta conveniente designar el valor absoluto de la pendiente de la función de pérdida como 6. En términos de cálculos,

$$b = |b_1 - b_2| \tag{20.13}$$

Como se señalaba en la sección 19.1, el valor esperado de la información perfecta es la pérdida de oportunidad esperada para la mejor acción. Así, para la situación de dos acciones con funciones de pérdida lineales por partes, sólo resulta Importante la función de pérdida de oportunidad para la mejor acción, al calcular la *POE.* La fórmula para determinar

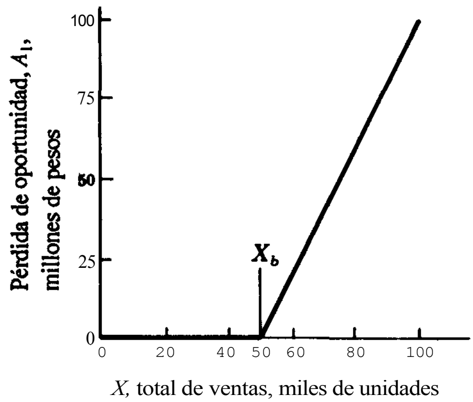


Fig. 20-2

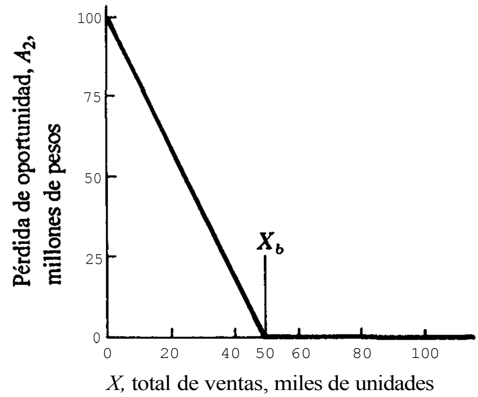


Fig. 20-3

el valor esperado de la información perfecta, para un problema de dos acciones con funciones lineales de pérdida por partes •' una distribución normal de probabilidad previa es

$$VEIP = bS_0L(D) \tag{20.14}$$

en donde $b = |b_1 - b_2|$

S_0 = Desviación estándar de la distribución de probabilidad previa

$$D = \left|\frac{X_b - M_0}{S_0}\right|$$

L(D) - La función de pérdida normal unitaria para *D* (Véase el apéndice 9)

EJEMPLO 9. En el ejemplo 6 se determinó que la mejor acción para el problema de compra de equipo era la acción A_1 Comprar equipo Tipo 1, con un pago esperado de $50 000. Utilizando la información de los ejemplos anteriores, se puede determinar el *VEIP* de la siguiente manera:

$$VEIP = bS_0\,L(D) = (2)(7\ 500)(0.4270) = \$640.50$$

en donde $b = |b_1 - b_2| = |10 - 12| = 2$

$S_0 = 7\ 500$ (del ejemplo 3)

$$D = \left|\frac{X_b - M_0}{S_0}\right| = \left|\frac{50\ 000 - 40\ 000}{7\ 500}\right| = 1.33$$

L(D) = 0.04270 (del apéndice 9)

La cantidad máxima que la ganancia esperada puede incrementar a largo plazo, suponiendo que se elimina la incertidumbre, es $640.50. Por lo tanto, ninguna información muestral podría valer más que esa cantidad, como promedio a largo plazo.

Con el objeto de comprender el procedimiento de cálculo de la fórmula para el *VEIP,* resulta útil sobreponer la función de probabilidad normal a la función de pérdida de oportunidad para la mejor acción. En la figura 20-4, para el cálculo del *VEIP* del ejemplo 9, las pérdidas de oportunidad ocurren a la derecha de $X_b = 50$. Observe que conforme mayor es la diferencia entre M_0 y X_b, menor es la probabilidad de sufrir una pérdida de oportunidad, según lo señalaría la proporción de la distribución de probabilidad que cae a la derecha de X_b. Por k> tanto, conforme se incrementa el valor de *D* en la fórmula del *VEIP,* menor es el valor de *L* (D). Conceptualmente *L(D)* representa el producto de la función de pérdida de oportunidad y de la función de probabilidad a la derecha del punto de equilibrio de la figura 20-4, y puede considerarse como la *POE* de la mejor acción, ***dada una función de pérdida con pendiente de 1.0 y una distribución normal con desviación estándar de 1.0.*** Por lo tanto, el producto de b y S_0 en la fórmula del *VEIP* sirve para transformar el valor de *L(D)* dado en el apéndice 9 para la función de pérdida normal unitaria, en el valor apropiado para la aplicación especifica.

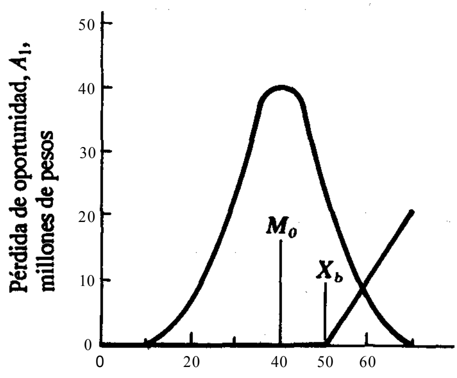


Fig. 20-4

Como se señalaba en las secciones 20.2 y 20.3, resulta más apropiado un análisis de los datos por tienda, que el análisis total del mercado, si se sabe con anticipación que habrán de recolectarse datos muestrales, posteriormente, para tiendas elegidas al azar. En un análisis por tienda como éste, el cálculo del *VEIP* debe considerar que la pérdida de oportunidad por unidad debe representar el cambio de pérdida de oportunidad unitario en el valor *promedio* por tienda. La pendiente de las funciones de pago para las dos acciones, con respecto a cambios en las ventas promedio por tienda, se designa mediante Nb_1 y Nb_2 . Por ello, el valor, absoluto de la pendiente de la función de pérdida de oportunidad, para el análisis por tienda, se determina de la siguiente manera:

$$b = |Nb_1 - Nb_2| \tag{20.15}$$

EJEMPLO 10. En el ejemplo 7 se identificaron las funciones de pago para las dos acciones, en términos de las ventas promedio por tienda, y fueron:

$$\text{Pago } (A_1) = -350\,000+5\,000\ \mu$$

$$\text{Pago } (A_2) = -450\,000+6\,000\ \mu$$

Se identificó que el punto de equilibrio, en términos del promedio de ventas por unidad, fue $\mu_b = 100$ unidades por tienda.

En el ejemplo 4 se determinaron la media y la desviación estándar previas, con base en el análisis por tienda, y resultaron ser $M_0 = 80$ y $S_0 = 15$, respectivamente.

Con la información que se resume anteriormente, se determina el *VEIP* para este problema de decisión de la siguiente manera:

$$VEIP = bS_0UD) = (1\,000)(15)(0.04270) = \$640.50$$

en donde $b = |Nb_1 - Nb_2| = |5\,000 - 6\,000| = 1\,000$

$S_0 = 15$

$$D = \left|\frac{\mu_b - M_0}{S_0}\right| = \left|\frac{100-80}{15}\right| = 1.33$$

$L(D) = 0.04270$ (del apéndice 9)

Por ello, el *VEIP* que se determinó con base en el promedio de ventas por tienda coincide con el resultado del ejemplo 9, que se basó en el análisis del mercado total.

20.5 ANÁLISIS BAYESIANO POSTERIOR

Cuando se recopila información muestral después de haber planteado una distribución de probabilidad previa, es posible determinar la distribución de probabilidad a *posteriori* calculando los valores de la media posterior y de la varianza posterior. La media posterior se designa como M_1 y se calcula a través de alguna de las siguientes fórmulas:

$$M_1 = \frac{(1/S_0^2)M_0 + (1/\sigma_{\bar{x}}^2)\bar{X}}{(1/S_0^2) + (1/\sigma_{\bar{x}}^2)} \qquad (20.16)$$

o

$$M_1 = \frac{M_0\sigma_{\bar{x}}^2 + \bar{X}S_0^2}{S_0^2 + \sigma_{\bar{x}}^2} \qquad (20.17)$$

La fórmula *(20.16)* muestra la base conceptual para determinar la media posterior. En esencia, esta media posterior es un promedio ponderado, en el cual se ponderan las medias previa y muestral mediante el inverso de sus respectivas varianzas. Sin embargo, la fórmula *(20.17)* resulta más conveniente para realizar cálculos. Si se desconoce la desviación estándar de la población σ, entonces se considera, en términos generales, aceptable el uso de la desviación estándar muestral, s, y se sustituye $s_{\bar{x}}$ **por** $\sigma_{\bar{x}}$ en las fórmulas anteriores.

EJEMPLO 11. En el ejemplo 4 se determinaron la media previa y la desviación estándar previa para el problema de decisión de compra de equipo, y fueron $M_0 = 80$ y $S_0 = 15$. Suponga que se fabrica el dispositivo electrónico para realizar una prueba piloto y se distribuye a través de nueve tiendas elegidas al azar. Se encuentra que el nivel promedio de ventas por tienda $\bar{X} = 110$ unidades, con desviación estándar de s =18 unidades. Como se desconoce σ, se usa el estimador de $\sigma_{\bar{x}}^2 : s_{\bar{x}}^2 = s^2/n = 324/9 = 36$ y la fórmula *(20.17)* para calcular la media posterior.

$$M_1 = \frac{M_0\sigma_{\bar{x}}^2 + \bar{X}S_0^2}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(80)(36) + (110)(225)}{225 + 36} = \frac{27\,630}{261} = 105.9 \text{ unidades}$$

Se designa mediante S_1^2 la varianza de la distribución posterior, y la raíz cuadrada de ese valor es la desviación estándar posterior S_1. La base conceptual que se utiliza para determinar la varianza posterior queda indicada mediante la ecuación:

$$\frac{1}{S_1^2}=\frac{1}{S_0^2}+\frac{1}{\sigma_{\bar{x}}^2} \qquad (20.18)$$

En la fórmula *(20.18)* se muestra que el inverso de la varianza posterior es igual al inverso de la varianza previa más el inverso de la varianza de la media. Debe observarse que, desde el punto de vista del análisis de decisiones, una varianza más grande indica mayor incertidumbre, y que, conforme mayor es la varianza, menor es su inverso.

La fórmula abreviada que se deriva de la fórmula *(20.18)* es:

$$S_1^2=\frac{S_0^2\sigma_{\bar{x}}^2}{S_0^2+\sigma_{\bar{x}}^2} \qquad (20.19)$$

Al igual que cuando se determinó la media posterior, si la desviación estándar poblacional o es desconocida, por lo general se utiliza la desviación estándar muestral *s* para calcular $s_{\bar{x}}^2$ como estimador de $\sigma_{\bar{x}}^2$.

EJEMPLO 12. Dada la información del ejemplo 11, se calcula la varianza y la desviación estándar de la distribución posterior de la siguiente manera:

$$S_1^2=\frac{S_0^2\sigma_{\bar{x}}^2}{S_0^2+\sigma_{\bar{x}}^2}\cong\frac{S_0^2 s_{\bar{x}}^2}{S_0^2+s_{\bar{x}}^2}=\frac{(225)(36)}{225+36}=\frac{8\,100}{261}=31.03$$

$$S_1=5.57\cong 5.6$$

Después de plantear la distribución de probabilidad posterior, se determina la mejor acción para esas decisiones, con base en esa distribución de probabilidad modificada, y con base también en las funciones de pago establecidas previamente. Como se explicaba en la sección 19.3, la información muestral tiene valor sólo si se da un cambio en la identificación de la mejor acción, como resultado de revisar la distribución previa. Al igual que en el capitulo 19, el valor estimado de la información muestral, para una muestra que ya se ha seleccionado es

VIM = (pago esperado posterior de la mejor acción posterior) (20.20
- (pago esperado posterior de la mejor acción previa)

EJEMPLO 13 En el ejemplo 7 se determinaron las funciones de pago para el problema de decisiones sobre equipo, con base en las ventas promedio por tienda, y resultaron ser:

$$\text{Pago }(A_1) = -350\,000 + 5000\mu \qquad \text{Pago }(A_2) = -450\,000 + 6000\mu$$

Al utilizar el valor de la media posterior, M_1 = 105.9, se determina la ganancia esperada correspondiente a cada acción:

Pago (A_1) = -350 000 + 5 000(105.9) = \$179 500 Pago (A_2) = -450 000 + 60 00(105.9) = \$185 400

Por lo tanto, se elige la acción A_2, con base en el análisis bayesiano posterior, y no la acción A_1 que fue la que se eligió con base en el análisis previo. El valor estimado de la muestra que se obtuvo en este caso es

VM = (pago esperado posterior de la mejor acción posterior) - (pago esperado posterior de la mejor acción previa)
=185 400 - 179 500 = \$5 900

20.6 ANÁLISIS PREPOSTERIOR Y EL VALOR ESPERADO DE LA INFORMACIÓN MUESTRAL *(VEIM)*

Como se considera que la media previa es un estimador insesgado de la media poblacional, no hay razón para anticipar una diferencia entre los valores de la media previa y la media posterior, antes de tomar una muestra. Sin embargo, puede anticiparse que la varianza posterior será menor que la varíanza previa. Debe recordarse de la sección 20.4, que la pérdida de oportunidad esperada de la mejor acción depende de la diferencia entre la media de la distribución de probabilidad y el punto de equilibrio. Cuando se reduce la varianza, se reduce la *POE* (y, por ello, el *VEIP),* porque la diferencia entre la media y el punto de equilibrio es entonces mayor, en unidades de la desviación estándar. Conceptualmente, el ***valor esperado de la información muestral*** es la diferencia esperada entre el *VEIP,* antes de la muestra, y el *VEIP* que se anticipa después de la muestra. Por ello, el *VEIM* es la reducción que se espera en la pérdida de oportunidad esperada correspondiente a la mejor acción, como resultado de tomar una muestra del tamaño especificado.

La reducción en la varianza entre las distribuciones de probabilidad previa y posterior se designa mediante S_*^2, y puede determinarse mediante alguna de las siguientes fórmulas:

$$S_*^2 = S_0^2 - S_1^2 \qquad (20.21)$$

$$S_*^2 = \frac{S_0^4}{S_0^2 + \sigma_{\bar{x}}^2} \qquad (20.22)$$

En la fórmula abreviada *(20.22),* se requiere el valor de la desviación estándar poblacional para determinar la varianza de la media $\sigma_{\bar{x}}^2$. Sin embargo, por lo general ese valor es desconocido. Además, no está disponible un estimador muestral porque no se ha tomado aún la muestra. Por ello, es necesario estimar el valor de σ mediante el análisis de situaciones similares.

Una vez que se determina el valor de S_*^2, se calcula el valor esperado de la información muestral mediante la fórmula:

$$VEIM = bS_* L(D_*) \qquad (20.23)$$

en donde $b = |b_1 - b_2|$

$$D_* = \left|\frac{\mu_b - M_0}{S_*}\right|$$

$L(D_*)$ = función de pérdida normal unitaria para D_* (véase el apéndice 9)

En *(20.23), b* es la pendiente de la función de pérdida correspondiente a la mejor acción (véase la sección 20.4) y D_* es similar a D en la fórmula del *VEIP* que se presentó en la sección 20.4, excepto que en el denominador de la fórmula se encuentra S_* y no S_0.

EJEMPLO 14. En el ejemplo 4 se calcularon la media previa y la desviación estándar previa para el problema de decisiones sobre equipo, en un análisis por tienda, y resultaron ser $M_0 = 80$ y $S_0 = 15$. Se tiene información de los dispositivos electrónicos que fabrican otras compañías, en el sentido de que la desviación estándar de las ventas por tienda para el periodo de tiempo relevante es aproximadamente **σ = 20.** Es posible determinar el valor esperado de la información proveniente de un estudio de mercado en el que se analizaron nueve tiendas al detalle, de la siguiente manera:

$$S_*^2 = \frac{S_0^4}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(15)^4}{(15)^2 + 44.44} = \frac{50\,625}{225 + 44.44} = \frac{50\,625}{269.44} = 187.89$$

en donde $\sigma_{\bar{x}}^2 \text{ est.} = \dfrac{\sigma^2 \text{ est.}}{n} = \dfrac{(20)^2}{9} = \dfrac{400}{9} = 44.44$

$$VEIM = bS_* L(D_*) = (1\,000)(13.71)(0.03208) = \$439.82$$

donde $b = |Nb_1 - Nb_2| = |5\,000 - 6\,000| = 1\,000$ (del ejemplo 10)

$$D_* = \left|\frac{\mu_b - M_0}{S_*}\right| = \left|\frac{100 - 80}{13.71}\right| = 1.46$$

$L(D_*) = 0.03208$

Debe recordarse que, en el ejemplo 13, el valor estimado de la información muestral para una prueba específica de n -=9, *que ya se había tomado,* fue de $5 900. El *VEIM* de $439.82 que se calculó en el ejemplo 14 señala que éste sería el valor promedio a largo plazo para una muestra de tamaño $n = 9$. En muchos casos específicos, la muestra tendrá un valor de $0, porque la selección de la mejor acción no habrá resultado afectada por la información muestral.

20.7 GANANCIA NETA ESPERADA DEL MUESTREO *(GNEM)* Y EL TAMAÑO ÓPTIMO DE MUESTRA

En el ejemplo 15 se determinan la ganancia neta esperada del muestreo y el tamaño óptimo de la muestra para el problema de decisión sobre equipo. En la sección 19.5 puede revisarse un análisis de estos conceptos y procedimientos.

EJEMPLO 15. Se encontró en el ejemplo 14 que el *VEIM,* para una muestra de n = 9, es $439.82. Suponga que el costo de diseñar el estudio de mercado es $300 y que el costo de obtener los datos de ventas de cada una de las tiendas muestreadas es $10. Dada la distribución previa de probabilidad, con $M_0 = 80$, $S_0 = 15$ y una σ estimada de 20, puede determinarse la *GNEM* correspondiente a tamaños de muestra alternativos de $n_1 = 9$, $n_2 = 12$ y $n_3 = 15$, de la siguiente manera:

Para $n = 9$:

$$VEIM = \$439.82 \qquad \text{(del ejemplo 14)}$$

$$GNEM = VEIM - CS = 439.82 - 390.00 = \$49.82$$

Para n = 12

$$S_*^2 = \frac{S_0^4}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(15)^4}{(15)^2 + 33.33} = \frac{50\,625}{258.33} = 195.97$$

en donde $\sigma_{\bar{x}}^2 \text{ est.} = \dfrac{\sigma^2 \text{ est.}}{n} = \dfrac{(20)^2}{12} = \dfrac{400}{12} = 33.33$

$$VEIM = bX_* L(D_*) = (1\,000)(14.00)(0.03431) = \$480.34$$

en donde $b = |b_1 - b_2| = |5\,000 - 6\,000| = 1\,000$

$$D_* = \left|\frac{\mu_b - M_0}{S_*}\right| = \left|\frac{100 - 80}{14.00}\right| = 1.43$$

$$GNEM = VEIM - CM = 480.34 - 420.00 = \$60.34$$

para $n = 15$:

$$S_*^2 = \frac{S_0^4}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(15)^4}{(15)^2 + 26.67} = \frac{50\,625}{251.67} = 201.16$$

donde $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2 \text{ est.}}{n} = \dfrac{(20)^2}{12} = \dfrac{400}{12} = 26.67$

$$VEIM = bS_* L(D_*) = (1\,000)(14.18)(0.03587) = \$508.64$$

where $b = |b_1 - b_2| = |5\,000 - 6\,000| = 1\,000$

$$D_* = \left|\frac{\mu_b - M_0}{S_*}\right| = \left|\frac{100 - 80}{14.18}\right| = 1.41$$

$$GNEM = VEIM - CM = 508.64 - 450.00 = \$58.64$$

Observe la información resumida que se presenta en la Tabla 20.1. De los tamaños de muestra que se consideran, el tamaño óptimo de muestra es n =12, con *una GNEM* de $60.34. Aunque no se consideraron todos ios tamaños de muestra posibles, como las *GNEM* correspondientes a n=9 y n=15 son menores que la GNEM de $n = 12$, resulta evidente que el tamaño de muestra de $n = 15$ ha rebasado el punto del tamaño óptimo de la muestra, y que el óptimo se encuentra en algún punto entre n = 9 *y* $n = 15$.

Tabla 20.1 *VEIM, CM* y *GNEM* para la decisión sobre la compra de equipo, de acuerdo con el tamaño de la muestra

Tamaño de la muestra (n)	*VE1M*	*CM*	*GNEM*
0	$ 0	$ 0	$ 0
9	439.82	390.00	49.82
12	480.34	420.00	60.34
15	508.64	450.00	58.64

Finalmente, es posible determinar el *pago esperado terminal global (PETG)* y *e\pago esperado terminal global neto (PETGN)* junto con el análisis preposteríor. Estos conceptos se explican en la sección 19.5, y los cálculos correspondientes se ilustran en el ejemplo 12.

20.8 ANÁLISIS BAYESIANO DE DECISIONES *VERSUS* PROCEDIMIENTOS CLÁSICOS DE DECISIÓN

En este capitulo y los dos anteriores se ha revisado el análisis bayesiano de decisión, y se le ha comparado con los procedimientos clásicos que se utilizaron en la mayor parte del texto. Las principales técnicas de la inferencia clásica son la estimación por intervalos y las pruebas de hipótesis. El análisis bayesiano de decisión se ocupa principalmente de la selección de una decisión. Aunque las técnicas clásicas se refieren en forma directa a la estimación o a las pruebas de hipótesis sobre parámetros poblacionales, los resultados de estos procedimientos se refieren a cursos alternativos de acción, o decisiones. Por ejemplo, aceptar la hipótesis nula de que el nivel promedio de ventas para un producto se encuentra por debajo del punto de equilibrio correspondería a la decisión de no comercializar el producto. Así, tanto los procedimientos clásicos como los bayesianos se ocupan del proceso de seleccionar las mejores decisiones.

La principal diferencia entre los procedimientos clásicos y los bayesianos es el uso de la información subjetiva (previa) en el análisis bayesiano, y la evaluación de decisiones alternativas en términos de consecuencias económicas (o, posiblemente, utilidades). Las consecuencias económicas pueden plantearse en términos ya sea de pagos o pérdidas de oportunidad condicionales (arrepentimientos). En esencia, la selección de los niveles α y β para las probabilidades de los errores Tipo I y Tipo II en pruebas de hipótesis es la base para evaluar la importancia de los dos tipos alternativos de errores. El uso de pérdidas de oportunidad en el análisis bayesiano representa una evaluación similar, sólo que de manera más explícita. Mientras los procedimientos clásicos de decisión se basan totalmente en el análisis de datos recopilados mediante muestreo aleatorio, los procedimientos bayesianos pueden también incluir el análisis de datos muestrales (mediante el análisis posterior), pero no dependen de la disponibilidad de esos datos.

Desde el punto de vista práctico, un factor importante en la ejecución de análisis bayesiano de decisiones es que éstos identifican los juicios de los administradores, para incluirlos en el análisis. Esto significa que se requiere que el especialista en estadística trabaje junto con el personal de administración. Por el contrario, la orientación exclusiva hacia los datos muestrales que se da en los procedimientos clásicos de decisión puede no dar oportunidad a los administradores de incluir sus conocimientos en el análisis de las decisiones, y éstos pueden considerar que sus juicios son importantes para el proceso.

Problemas resueltos

ANÁLISIS DE DECISIONES ANTES DEL MUESTREO

20.1 El propietario de una pequeña compañía manufacturera está considerando añadir generadores eléctricos a la línea de equipo eléctrico automotriz que fabrica. La inversión de capital que se requiere para fabricar los generadores es $150 000 000, y se da una utilidad de $2 000 por generador que se vende a través del sistema de distribución, que cuenta con 100 tiendas. Durante el periodo total de importancia para este caso, el propietario estima que el promedio de ventas de generadores eléctricos será de 700 unidades por tienda, y que existe una probabilidad del 0.5% de que el nivel promedio de ventas por tienda se encuentre entre 600 y 800 unidades.

(a) Al suponer una distribución de probabilidad previa normal, determine la media y la desviación estándar de esta distribución.

(b) Las dos posibles decisiones son A_1 : Fabricar los generadores, y A_2: No fabricarlos. Plantee las funciones de pago lineales correspondientes a esas acciones, e ilústrelas en una sola gráfica.

(a)

$$M_0 = 700 \text{ unidades}$$

$$\frac{2}{3} S_o = \frac{\text{intervalo del 50\% central}}{2} = \frac{800 - 600}{2} = 100$$

$$S_0 = 150$$

(b)

$$\text{Pago}(A_1) = k_1 + Nb_1\,\mu = -150\,000\,000 + (100)(2\,000)\,\mu = -150\,000\,000 + 200\,000\,\mu$$

$$\text{Pago}(A_2) = k_2 + Nb_2\,\mu = 0 + (100)(0)\,\mu = 0$$

En la figura 20-5 se ilustran esas dos funciones de pago lineales.

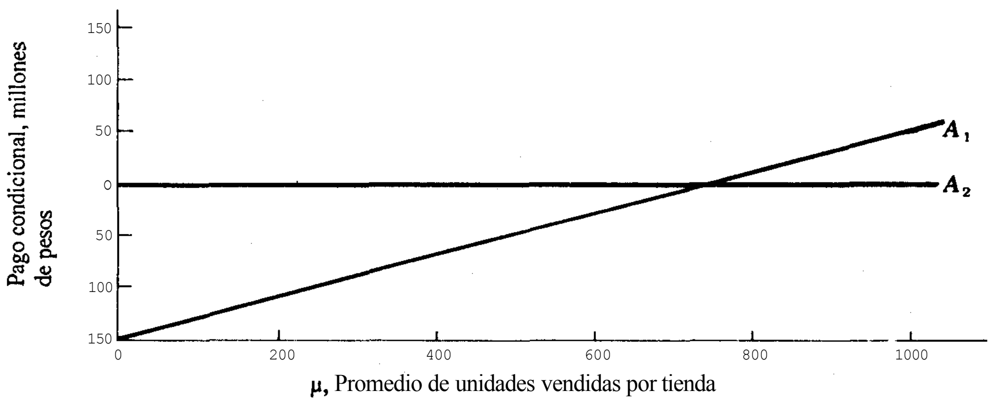


Fig. 20-5

20.2 Determine la mejor acción para la situación de decisiones que se plantea en el problema 20.1, calculando los valores esperados correspondientes a las dos posibles acciones.

$$PE(A_1) = k_1 + Nb_1 M_0 = -150\,000\,000 + (100)(2\,000)(700) = -\$10\,000\,000$$

$$PE(A_2) = k_2 + Nb_2 M_0 = 0 + 100(0)700 = 0$$

Por lo tanto, la mejor acción es A_2: No fabricar los generadores.

20.3 Con referencia al problema 20.1, (a) determine el valor del punto de equilibrio en términos de las ventas promedio por tienda, en unidades, y (b) plantee las funciones de pérdida lineales por partes para las dos acciones posibles.

Ilustre la función de pérdida para la acción A_2: No fabricar, en una gráfica en la que se ilustre también la distribución de probabilidad previa normal.

(a)
$$\mu_b = \frac{k_2 - k_1}{Nb_1 - Nb_2} = \frac{0 - (-150\,000\,000)}{100\,(2\,000) - 100(0)} = \frac{150\,000\,000}{200\,000} = \$750\,000$$

Observe que este es el volumen promedio de ventas en el que se cruzan las dos funciones lineales de pago, de la figura 20-5.

(b)
$$PO(A_1, \mu)\begin{cases} = 0 & \text{para } \mu \geq \mu_b \ (=750) \\ = 200(750 - \mu) & \text{para } \mu < \mu_b \ (=750) \end{cases}$$

$$PO(A_2, \mu)\begin{cases} = 200(\mu - 750) & \text{para } \mu > \mu_b \ (=750) \\ = 0 & \text{para } \mu \leq \mu_b \ (=750) \end{cases}$$

Observe la figura 20-6

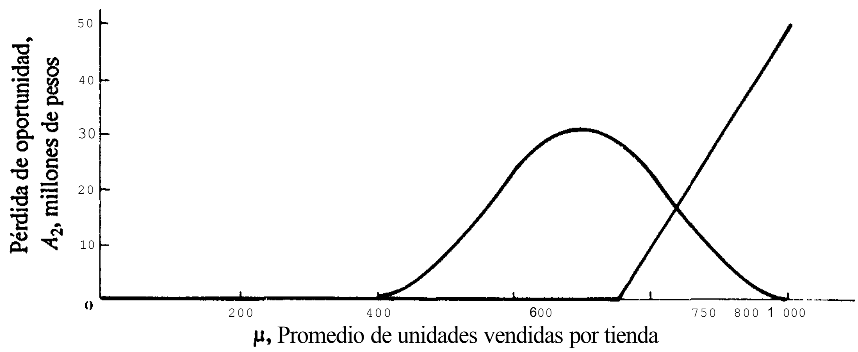


Fig. 20-6

20.4 Para los problemas 20.1 a 20.3, (a) determine el valor esperado de la información perfecta, (b) Suponga que el propietario de la empresa manufacturera elige la mejor acción, de acuerdo con el análisis bayesiano. ¿Cuál es la probabilidad de que su decisión resulte equivocada?

(a)
$$VEIP = bS_0\,L\,(\text{D}) = (200\,000)\,(150)\,(0.2555) = \$\ 7\,665\,000$$

en donde $b = |Nb_1 - Nb_2| = |200\,000 - 0| = 200\,000$

$S_0 = 150$

$$D = \left|\frac{\mu_b - \mu_0}{S_0}\right| = \left|\frac{750 - 700}{150}\right| = \frac{50}{150} = 0.33$$

L (D) = 0.2555 (del apéndice 9)

(b) Con referencia a la figura 20-6, la probabilidad de que la decisión de no fabricar los generadores resulte equivocada es equivalente a la probabilidad de que el promedio de ventas por tienda supere el punto de equilibrio de 750 unidades. Convirtiendo el valor del punto de equilibrio en un valor z normal unitario, se tiene

$$z = \frac{\mu_b - M_0}{S_0} = \frac{750 - 700}{150} = 0.33$$

Por lo tanto,

$$P(\mu > 750) = P(z > 0.33) = 0.5000 - P(0 \le z \le 0.33)$$
$$= 0.5000 - 0.1293 = 0.3707 \cong 0.37$$

20.5 Un gerente puede adquirir una máquina fotocopiadora con dos proveedores. La máquina de marca *Y* cuesta \$8 000 000, y tiene un costo variable de \$5 por copia. La marca Z cuesta \$9 000 000 pero tiene un costo variable de \$4.50 por copia.

(a) Plantee la función de costo lineal correspondiente a la acción A_1 : Comprar la marca Y, y la acción A_2: Comprar la marca Z.

(b) El gerente estima que la vida útil de las máquinas es de 500 semanas. Replantee las funciones de costo lineales que se determinaron en (a) en términos del promedio de copias por semana.

(a)
$$C(A_1) = k_1 + b_1 X = 8\,000\,000 + 50\,X$$
$$C(A_2) = k_2 + b_2 X = 9\,000\,000 + 45\,X$$

Nota: Estas funciones de costos se podrían expresar como funciones de pago utilizando signos negativos en las ecuaciones anteriores para la inversión de capital y para el incremento en los costos. Sin embargo, por lo general, en estudios de costos comparativos se considera más conveniente utilizar valores positivos, y considerar que representan costos.

(b)
$$C(A_1) = k_1 + Nb_1\,\mu = 8\,000 + (500)(50)\,\mu = 8\,000 + 25\,000\,\mu$$
$$C(A_2) = k_2 + Nb_2\,\mu = 9\,000 + (500)(45)\,\mu = 9\,000 + 22\,500\,\mu$$

20.6 (a) No se esperan cambios en el volumen de trabajo para el periodo de tiempo en el que se planea utilizar la máquina fotocopiadora del problema 20.5 (¡una suposición evidentemente simplificadora!). Ei gerente estima que se obtendrá un promedio de 500 copias por semana en la máquina. Determine qué marca de máquina debe comprarse con base en ei costo esperado global correspondiente a cada una de las dos marcas de máquinas.

(b) Determine el volumen del punto de equilibrio en términos del número promedio de copias que se fabrica por semana, y que daría como resultado que el gerente no prefiriera ninguna de las dos máquinas, en términos del costo esperado global.

(a)
$$EC(A_1) = k_1 + Nb_1\,M_0 = 8\,000\,000 + 25\,000\,(500) = \$20\,500\,000$$
$$EC(A_2) = k_2 + Nb_2\,M_0 = 9\,000\,000 + 22\,500\,(500) = \$20\,250\,000$$

Por lo tanto, la mejor acción es A_2: Comprar la marca Z, porque ésta tiene un menor costo esperado. *(Nota: Si el análisis se hubiera llevado a cabo en términos de los valores esperados, los dos valores anteriores hubieran sido negativos y, por ello, el mayor valor sería el correspondiente a la acción A_2.)*

(b)
$$\mu_b = \frac{k_2 - k_1}{Nb_1 - Nb_2} = \frac{9\,000\,000 - 8\,000\,000}{500\,(50) - 500\,(45)} = \frac{1\,000\,000}{25\,000 - 22\,500} = 400 \text{ copias por semana}$$

20.7 Repase los problemas 20.5 y 20.6: *(a)* Suponiendo una distribución de probabilidad previa normal, es posible determinar su desviación estándar con base en alguna estimación del gerente sobre dos puntos percentiles de la distribución, y no necesariamente identificando un intervalo de credibilidad del 50% central. Por supuesto, la media previa de 500 copias por semana se encuentra en el percentil 50 de la distribución. Además, suponga que eferente estima que existe una probabilidad de sólo el 0.10% de que el número de copias promedio por semana sea superior a 700. Determine la desviación estándar de la distribución de probabilidad previa, *(b)* Plantee las funciones de pérdida lineal por partes para las dos acciones, con base en el número promedio de copias por semana.

(a) Con referencia a la distribución normal estándar (apéndice 5), se observa que el valor correspondiente al punto percentil 90 es aproximadamente z = +1.28. Puede utilizarse una fórmula general de z y despejar la incógnita de S_0, de la siguiente manera:

$$z = \frac{\mu - M_0}{S_0} \qquad S_0 = \frac{\mu - M_0}{z} = \frac{700-500}{1.28} = 156.25$$

(b)

$$PO(A_1, \mu)\begin{cases} = 2.5(400-\mu) & \text{para } \mu > \mu_b \ (=400) \\ = 0 & \text{para } \mu \leq \mu_b \ (=400) \end{cases}$$

$$PO(A_2, \mu)\begin{cases} = 0 & \text{para } \mu \geq \mu_b \ (=400) \\ = 2.5(\mu - 400) & \text{para } \mu < \mu_b \ (=400) \end{cases}$$

20.8 Determine el *VEIP* para la decisión sobre la máquina fotocopiadora que se describe en los problemas 20.5 a 20.7.

$VEIP = bS_0 L(D)$ = (2 500) (156.25) (0.1580) = \$61,718.75

En donde $b = |Nb_1 - Nb_2| = |25\ 000 - 22\ 500| = 2\ 500$

S_0= 156.25

$$D = \left|\frac{\mu_b - M_0}{S_0}\right| = \left|\frac{400-500}{156.25}\right| = \frac{100}{156.25} = 0.64$$

L (D) 0.1580 (del apéndice 9)

El reducido valor del *VEIP* indica que no existe riesgo financiero considerable en esa situación de toma de decisiones, ni tampoco una gran posibilidad de ganancia económica esperada para el muestreo.

ANÁLISIS BAYESIANO POSTERIOR

20.9 Con referencia a los problemas 20.1 a 20.4, se realiza una prueba de mercado con un generador similar al que se va a a fabricar en 10 tiendas elegidas al azar. El promedio de ventas por tienda es de $\bar{X}$ = 800, con una desviación estándar muestral de s = 110. Determine la media y la desviación estándar de la distribución posterior.

$$M_1 = \frac{M_0\sigma_{\bar{x}}^2 + \bar{X}S_0^2}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(700)(1\ 100) + (800)(22\ 500)}{22\ 500 + 1\ 100} = \frac{18\ 770\ 000}{23\ 600} = 795.3 \text{ unidades}$$

en donde $\sigma_{\bar{x}}^2 \text{ est.} = \frac{s^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{12\ 100}{10}\left(\frac{100-10}{100-1}\right) = (1\ 210)(0.909) = 1\ 099\ 999 \cong 1\ 100$

(En este caso se utiliza el factor de corrección finita porque $n > 0.05 N$. Véase la sección 8.2)

$$S_1^2 = \frac{S_0^2\sigma_{\bar{x}}^2}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(22\ 500)(1\ 100)}{22\ 500 + 1\ 100} = \frac{24\ 750\ 000}{23\ 600} = 1\ 048.7288$$

$$S_1 = 32.38$$

20.10 Con referencia al problema 20.9, (a) determine la mejor acción (A_1,: Fabricar los generadores, o A_2: No fabricar los generadores) *(b)* ¿Cuál es el valor estimado de la información muestral?

(a) $PE(A_1) = k_1 + Nb_1 M_1$ = - 150 000 000 + (100) (2 000) (795.3) = \$9 060 000

$PE(A_2) = k_2 + Nb_2 M_1$ = 0 + (100) (0) (795.3) = 0

Por lo tanto, la mejor acción *es* A_1: Fabricar los generadores.

(b) VIM = (pago esperado posterior de la mejor acción posterior) - (pago esperado posterior de la mejor acción previa)
=9060000 - 0 = $9060000

20.11 En el problema 20.4, se determinó que el *VEIP,* antes de tomar cualquier muestra, era de $7 665. (a)¿Cómo es posible que el valor de la información muestral que se determinó en el problema 20.10 exceda el *VEIP? (b)* Determine el *VEIP* (posterior) para la decisión sobre la fabricación de generadores. Es decir, calcule el valor esperado de la información perfecta, después de haber tomado la muestra que se describe en el problema 20.9, y después de haber incorporado los resultados al análisis.

(a) *VEIP* es el valor *esperado* de la información perfecta. Por lo tanto, es posible que un resultado muestral específico tenga un valor estimado que supere ese valor promedio a largo plazo para la información perfecta.

(b) $VEIP\text{ (posterior)} = bS_1\,L(D_1) = (200\,000)(32.38)(0.03667) = \$237.47$

donde $b = 200\,000$

$S_1 = 32.38$

$$D_1 = \left|\frac{\mu_b - M_1}{S_1}\right| = \left|\frac{750 - 795.3}{32.38}\right| = \frac{45.3}{32.38} = 1.40$$

$L(D_1) = 0.03667$ (del apéndice 9)

20.12 El gerente de los problemas 20.5 a 20.8 decide contar el número de oopias que se requieren en el departamento durante un periodo de 5 semanas, antes de comprar la fotocopiadpra. Considera que esas cinco semanas son una muestra aleatoria del periodo de 500 semanas (de nueva cuenta, una suposición simplificadoral). El número promedio de copias en ese periodo de muestra es $\bar{X} = 450$, con desviación estándar s =100. Determine la desviación estándar y la media de la distribución posterior.

$$M_1 = \frac{M_0\sigma_{\bar{x}}^2 + \bar{X}S_0^2}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(500)(2\,000) + (450)(156.25)^2}{(156.25)^2 + 2\,000} = \frac{11\,986\,327}{26\,414.062} = 453.8$$

en donde la $\sigma_{\bar{x}}^2$ **est.** $= \dfrac{s^2}{n} = \dfrac{10\,000}{5} = 2\,000$

$$S_1^2 = \frac{S_0^2\sigma_{\bar{x}}^2}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(24\,414.062)(2\,000)}{24\,414.062 + 2\,000} = 1\,848.5655$$

$S_1 = 42.99$

20.13 Con referencia al problema 20.12, determine (a) la mejor acción, (A_1: Comprar la marca *Y* o A_2: Comprar la marca Z) con base en la distribución posterior, y (b) el valor estimado de esta información muestral.

(a)
$$EC(A_1) = k_1 + Nb_1 M_1 = 8\,000\,000 + 25\,000\,(453.8) = \$19\,345\,000$$
$$EC(A_2) = k_2 + Nb_1 M_1 = 9\,000\,000 + 22\,500\,(453.8) = \$19\,210\,500$$

Por lo tanto, la mejor acción es A_2: Comprar la máquina Z, porque a ésta le corresponde el menor costo esperado.

(b) Como los resultados de la información muestral no cambiaron la identificación de la mejor acción, la muestra no tiene valor. O, analizando la fórmula general:

VIM = (pago esperado posterior de la mejor acción posterior) - (pago esperado posterior de la mejor acción previa)
= -19 210 500 - (- 19 210 500) = $ 0

EL VALOR ESPERADO DE LA INFORMACIÓN MUESTRAL *(VEIM)* Y LA GANANCIA NETA ESPERADA DEL MUESTREO *(GNEM)*

20.14 Con referencia a los problemas 20.1 y 20.4, determine el valor esperado de la información muestral *(VEIM)* correspondiente a una muestra de n = 10 tiendas, si se estima que la desviación estándar es $\sigma = 300$.

Dada la media previa M_0 =700, la desviación estándar previa S_0 =150, y la desviación estándar estimada de las ventas por tienda, $\sigma = 300$,

$$S_*^2 = \frac{S_0^4}{S^2 + \sigma_{\bar{x}}^2} = \frac{(150)^4}{(150)^2 + 8\,181} = \frac{506\,250\,000}{30\,681} \cong 16\,500$$

$$S_* = 128$$

en donde $\sigma_{\bar{x}}^2 \text{ est.} = \frac{\sigma^2 \text{ est.}}{n}\left(\frac{N-n}{N-1}\right) = \frac{90\,000}{10}\left(\frac{100-10}{100-1}\right) = (9\,000)(0.909) \cong 8\,181$

[Se utiliza el factor de corrección por población finita porque $n > 0.05\,N$. Véase la sección 8.2.)

$$VEIM = bS_*\,L(D_*) = (200)(128)(0.2339) = \$5\,988$$

en donde $b = |Nb_1 - Nb_2| = |200 - 0| = 200$

$$D_* = \left|\frac{\mu_b - M_0}{S_*}\right| = \left|\frac{750-700}{128}\right| = \frac{50}{128} = 0.39$$

(del apéndice 9)

20.15 En los problemas 20.6 y 20.7, la media previa en términos de copias promedio por semana fue de M_0= 500, con $S_0 \cong$ 156. Determine el *VEIM* correspondiente a una muestra de n = 5, si la desviación estándar estimada es $\sigma = 250$.

$$S_*^2 = \frac{S_0^4}{S_0^2 + \sigma_{\bar{x}}^2} = \frac{(156)^4}{(156)^2 + 12\,500} = \frac{592\,240\,896}{36\,836} = 16\,077.77$$

$$S_* = 126.80$$

en donde $\sigma_{\bar{x}}^2 \text{ est.} = \frac{\sigma^2 \text{ est.}}{n} = \frac{62\,500}{5} = 12\,500$

$$VEIM = bS_*\,L(D_*) = (2500)(126.80)(0.1223) = 3\,876.91$$

en donde $b = |Nb_1 - Nb_2| = |25\,000 - 22\,500| = 2\,500$

$$D_* = \left|\frac{\mu_b - M_0}{S_*}\right| = \left|\frac{400-500}{126.80}\right| = \frac{100}{126.80} = 0.79$$

$L(D_*) = 0.1223$

En el problema 20.8 se determinó que el *VEIP* era de $61.72. Por ello, el valor esperado de una muestra de sólo 5 semanas serviría para reducir la incertidumbre económica en esta situación de decisiones, en $3 876 91, en promedio.

20.16 Con referencia al problema 20.15, suponga que el costo de contar el número de copias que se producen en esos momentos en el departamento se estima en $5 000 por semana. Determine la ganancia neta esperada correspondiente a una muestra n = 5 semanas.

$$GNEM = VEIM - CM = 3\,876.91 - 2\,500 = 1\,376.91$$

Problemas complementarios

ANÁLISIS DE DECISIONES ANTES DEL MUESTREO

20.17 Un distribuidor tiene la opción de manejar una nueva línea de muebles de oficina. El manejo de esa línea requiere de un gasto adicional de capital de $80 000 000 para un nuevo almacén. El aumento en las utilidades, por encima de los costos variables por el manejo de los muebles de oficina, es del 10% del volumen de ventas. En el periodo durante el cual se va a amortizar el gasto de capital, el propietario de la empresa estima que el volumen de ventas más probable para los muebles será de $900 000 000, y que existe una probabilidad del 0.50% de que el volumen esté entre $750 000 000 y 1 050 000 000.

(a) Al suponer una distribución de probabilidad previa normal, determine su media y su desviación estándar.

(b) Las dos posibles decisiones son A_1 Distribuir y A_2. No distribuir. Plantee las funciones lineales de pago correspondientes a esas dos acciones e ilústrelas en una sola gráfica.

Resp. (a) M_0: $900 000 000 y S_0 = $225 000 000.

20.18 Determine la mejor acción para las condiciones que se describen en el problema 20.17.

Resp. A_1 Distribuir, con *PE* = $10 000 000

20.19 Para el problema 20.17, (a) determine el punto de equilibrio en términos del volumen total de ventas que se requieren.

(b) Plantee las funciones de pérdidas lineales por partes de las dos posibles acciones. Ilustre la función de pérdida para la acción A_1: Distribuir, en una gráfica que también incluya la distribución de probabilidad previa normal.

Resp. (a) X_b = $800 000 000

20.20 Para los problemas 20.17 a 20.19, (a) ¿cuál es el valor promedio máximo de la información sobre el mercado? (b) Suponga que el propietario de esa distribuidora elige la mejor acción, de acuerdo con el análisis bayesiano. ¿Cuál es la probabilidad de que su decisión sea correcta?

Resp. (a) $4 880 250, (b) G = 0.67

20.21 El fabricante de un producto nuevo requiere de una inversión de capital de $100 000 000. Para ese producto, los precios variables de fabricación serán de $2 000 por unidad, el precio de venta es de $5 000 por unidad, y el gerente de mercadotecnia estima que el valor promedio de ventas más probable será de 80 unidades por tienda. Existen 50 tiendas. Suponiendo una distribución de probabilidad previa normal, determine si la mejor acción es A_1: Fabricar, o A_1 No fabricar, determinando los pagos esperados correspondientes a esas acciones.

Resp. A1 Fabricar con *GE* = $20 000 000.

20.22 Al continuar con el problema 20.21, el gerente de mercadotecnia afirma que existe una probabilidad de 0.70% de que las ventas del producto superen el promedio de 75 unidades por tienda. Determine (a) la desviación estándar de la distribución de probabilidad normal previa, y *(b)* el valor esperado de la información perfecta.

Resp. (a) S_0= 9 600, *(b) VEIP* = $540 000

20.23 Se supone que la incertidumbre inherente en la estimación de las ventas para la nueva línea de productos tiene distribución normal con M_0 = 7 000 unidades y S_0 = 1 000. La inversión de capital que se requiere para incluir la nueva línea de productos es \$80 000 000, y la ganancia (sobre el costo variable) para cada uno de los artículos vendidos es \$10 000. Señale cuál es la mejor decisión en estas condiciones (A_1: Comercializar o A_2:No comercializar) y la correspondiente ganancia esperada.

Resp. A_2: No comercializar, con GE = 0

20.24 Con referencia al problema 20.23, si la empresa lleva a cabo un estudio de mercado para probar la línea de productos en una reglón de muestra, ¿cuál es el máximo del valor promedio de esa información adicional?

Resp. \$833 000.

ANÁLISIS BAYESIANO POSTERIOR

20.25 Una vez que se ha determinado la distribución de probabilidad previa con base en un análisis del mercado total, pueden determinarse los valores de la media previa, la desviación estándar previa y el punto de equilibrio por tienda, simplemente dividiendo cada uno de los valores correspondientes del mercado total entre el número de tiendas que se tienen. Para los problemas 20.17 a 20.20, suponga que existe un total de 800 tiendas a nivel nacional. Determine los valores por tienda para (a) la media previa, (b) la desviación estándar previa y (c) el punto de equilibrio.

Resp. (a) M_0 = \$1 125 000, (b) S_0 = \$281 000, (c) μ_b = \$1 000 000

20.26 Con respecto a la transformación de los valores por tienda que se realizó en el problema 20.25, suponga que se elige al azar una muestra de n = 4 tiendas para probar los resultados de mercado para la linea de productos. Para esta muestra, el nivel promedio de ventas de la línea de muebles de oficina es X = \$900 000, con s - \$300 000. Determine (a) la media y (b) la desviación estándar de la distribución posterior del nivel estimado de ventas por tienda.

Resp. (a) M_1 = \$950 000, (b) S_1 = \$132 000

20.27 (a) Con base en la distribución posterior que se determinó en el problema 20.26, determine la mejor acción (A,: Distribuir, o A_2: No distribuir) para el problema de decisiones que se describe en los problemas 20.17 a 20.20.

(b) ¿Cuál es el valor estimado de la información muestral que se obtuvo en el problema 20.26?

Resp. (a) A_2: No distribuir, (b) \$4 000 000

20.28 Con referencia a los problemas 20.21 y 20.22, suponga que el producto se ofrece a través de 10 tiendas elegidas al azar. El nivel promedio de ventas del producto en esas tiendas es X - 60.0 unidades, con s = 25.0 unidades. Determine (a) la media y (b) la desviación estándar de la distribución posterior del nivel estimado de ventas por tienda.

Resp. (a) M_1 = 68.1, (b) S_1 = 6.1

20.29 Al continuar con el problema 20.28, identifique (a) la mejor acción y (b) el valor estimado de la información muestral queso recolectó.

Resp. (a) A_1: Fabricar, con *PE* = \$2150 000, (b) \$0

20.30 (a) Determine el *VEIP* para los problemas 20.21 y 20.22 después de haber tomado la muestra que se describe en el problema 20.28.

(b) Compare el valor del *VEIP* posterior que se determinó en el inciso (a) anterior, con el *VEIP* de \$540 000 que se determinó antes de tomar la muestra en el problema 20.22. Explique el significado del cambio que ocurre.

Resp. (a) *VEIP* posterior= \$2 694 000

EL VALOR ESPERADO DE LA INFORMACIÓN MUESTRAL *(VEIM)* Y LA GANANCIA NETA ESPERADA DEL MUESTREO *(GNEM)*

20.31 En los problemas 20.21 y 20.22 se calculó la distribución de probabilidad previa por tienda, y fueron $M_0 = 80.0$ y $S_0 = 9.6$. Existe un total de 500 tiendas. Si se fabrica el producto, se estima que la desviación estándar del nivel de ventas en cada una de las tiendas será de aproximadamente o = 30. Determine el *VEIM* correspondiente a una muestra de $n = 10$ tiendas.

Resp. VEIM = \$96 340

20.32 Explique la razón de la diferencia entre el *VEIP* de \$96 340 para una muestra de $n = 10$, que se calculó en el problema 20.31, y el valor estimado de \$0 para la muestra de $n = 10$ que se determinó en el problema 20.29.

20.33 Con respecto a la muestra que se considera en el problema 20.31, suponga que el costo de que todas las tiendas participen en el estudio es de \$15 000. Determine el *GNEM* para una muestra de $n = 10$.

Resp. GNEM = - \$53 660

Pruebas estadísticas no paramétricas

21

21.1 ESCALAS DE MEDICIÓN

Antes de considerar las diferencias entre los métodos no paramétricos con respecto a los procedimientos paramétricos que constituyen la mayor parte de este libro, resulta útil definir cuatro tipos de escalas de medición en términos de la precisión de los valores reportados.

En la escala nominal se utilizan números sólo para identificar categorías. No representan ninguna cantidad o monto como tales.

EJEMPLO 1. Si a cuatro regiones de ventas se les asignan los números 1 a4, sólo a manera de identificación, entonces se trata de una escala nominal, puesto que los números sirven simplemente como nombres de categorías.

En la escala ordinal los números representan rangos. Los números indican magnitud relativa, pero no se supone que las diferencias de los rangos son iguales.

EJEMPLO 2. Un analista financiero jerarquiza cinco acciones, del 1 al 5, en términos de su potencial de aumento de precio. Por lo general, la diferencia entre el potencial de crecimiento del precio entre las acciones con los rangos 1 y 2 no seria igual que, por ejemplo, la diferencia entre las acciones con los rangos 3 y 4.

En la escala de intervalo se representan diferencias entre valores que se miden. Sin embargo, el punto cero es arbitrario y no un cero "absoluto". Por ello, no es posible comparar los números mediante razones o cocientes.

EJEMPLO 3. En las escalas de temperatura Fahrenheit o Celsius, una diferencia de 5° de, por ejemplo, 70°F a 75°F, es la misma diferencia en temperatura que de 80°F a 85T. Sin embargo, no puede decirse que 60°F es el doble de caliente que 30°F, porque el punto 0°F no es un cero absoluto (la ausencia completa de calor).

En la escala de razón, existe un verdadero punto cero, y, por ello, es posible comparar las mediciones en forma de razones.

EJEMPLO 4. No sólo es cierto que una diferencia en el valor de un inventario de $5 000 000 es la misma cantidad que, por ejemplo, entre $50 000 000 y $55 000 000, que entre $60 000 000 y $65 000 000; también es cierto que un valor de inventarío de $100 000 000 es el doble de grande que un valor de inventario de $50 000 000.

21.2 COMPARACIÓN DE LOS MÉTODOS ESTADÍSTICOS PARAMÉTRICOS VERSUS NO PARAMÉTRICOS

La mayor parte de métodos estadísticos que se describen en este libro se denominan métodos paramétricos. El punto focal del análisis paramétrico es algún parámetro poblacional, cuya estadística muestral tiene una distribución conocida. Además, las mediciones se realizan en el nivel de escala de intervalo o de razón. Cuando no se satisfacen uno o más de estos requerimientos o suposiciones, pueden utilizarse los denominados métodos no paramétricos. Un término alternativo son los métodos de distribución libre, que describen en particular el hecho de que no se conoce la distribución de la estadística muestral.

Si se justifica el uso de una prueba paramétrica, tal como la prueba í, entonces siempre es preferible utilizar ésta y no su equivalente paramétrico. Esto se debe a que, si se utiliza el mismo nivel de significancia en ambas pruebas, la potencia de la prueba no paramétrica siempre es inferior a la de su equivalente paramétrica. (Recuerde, de la sección 10.3, que la potencia de una prueba estadística es la probabilidad de rechazar una hipótesis nula falsa.) Con frecuencia se utilizan pruebas no paramétricas cuando se trata con muestras pequeñas, debido a que en estos casos no puede aplicarse el teorema central del límite (sección 8.2).

Las pruebas no paramétricas pueden utilizarse para probar hipótesis referentes a (a forma, la dispersión o posición (mediana) de la población. En la mayor parte de las aplicaciones, las hipótesis se refieren al valor de una mediana, a la diferencia entre dos medianas, o a las diferencias entre diversas medianas. Esto difiere de los procedimientos paramétricos, que se concentran principalmente en medias poblacionales.

De las pruebas estadísticas descritas en el texto, la prueba de ji-cuadrada que se presenta en el capítulo 12 es, de hecho, una prueba no paramétrica. Recuerde, por ejemplo, que los datos que se analizan están dados en escala nominal (datos categóricos). Se dedicó un capítulo especial a la prueba de ji-cuadrada debido a su amplia utilización y a la variedad de sus aplicaciones.

21.3 LA PRUEBA DE RACHAS CORRIDAS PARA LA ALEATORIEDAD

Una *corrida* es un conjunto de observaciones similares. Se utiliza *la prueba de corridas* para probar la aleatoriedad de un conjunto de observaciones (donde a cada una de éstas se le asigna una de dos categorías).

EJEMPLO 5. Suponga que, al clasificar según el sexo una muestra aleatoria de $n = 10$ personas, la secuencia de observaciones es M, M, M, M, F, F, F, F, M, M. Existen tres corridas en estos datos, es decir, existen tres conjuntos de observaciones iguales.

Para datos numéricos, una forma de obtener el esquema de dos categorías que se requiere, consiste en clasificar cada una de las observaciones según sea mayor o menor que la mediana del grupo. En general, se rechaza la hipótesis nula de que la secuencia de observaciones es aleatoria si se obtiene un número muy grande o muy pequeño de corridas, con respecto a lo que se esperarla en una muestra aleatoria.

Se determina el número de corridas de elementos similares para los datos muestrales, utilizando el símbolo C para designar el número de corridas observadas. Utilizando n_1 para representar el número de elementos muestreados de un tipo, y n_2 para el número de elementos muestreados del segundo tipo, la media y el error estándar correspondientes a la distribución muestral de la estadística de prueba C, cuando la secuencia es aleatoria, son:

$$\mu_R = \frac{2n_1n_2}{n_1+n_2}+1 \tag{21.1}$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} \tag{21.2}$$

Si n_1 > 20 o n_2 > 20, la distribución muestral de C se aproxima a la distribución normal. Por ello, bajo esas circunstancias, es posible convertir a la estadística C a la estadística de prueba z, de la siguiente manera:

$$z = \frac{R - \mu_R}{\sigma_R} \tag{21.3}$$

Existen tablas de la estadística de prueba C en libros de texto especializados en estadística no paramétrica, para los casos en que $n_1 \leq 20$ y $n_2 \leq 20$.

En el problema 21.1 se ilustra el uso de la prueba de corridas para aleatoriedad.

21.4 UNA MUESTRA: LA PRUEBA DEL SIGNO

Puede utilizarse la *prueba del signo* para probar una hipótesis nula sobre el valor de una mediana poblacional. Por ello, es el equivalente no paramétrico de las pruebas de hipótesis sobre el valor de una media poblacional. Se requiere que los valores de la muestra aleatoria se encuentren cuando menos en escala ordinal, y no se hacen suposiciones con respecto a la forma de la distribución poblacional.

Las hipótesis nula y alternativa pueden designar pruebas de uno o de dos criterios de calificación. Utilizando Med_0 para representar la mediana de la población, y Med_0 para representar el valor hipotético, las hipótesis nula y alternativa para una prueba de dos extremos son:

$$H_0\text{: Med} = \text{Med}_0$$

$$H_1\text{: Med} \neq \text{Med}_0$$

Se asigna un signo positivo (+) a cada valor muestral observado que resulte ser mayor que el valor hipotético de la mediana, y un signo negativo(-) a los valores que son menores que ese valor hipotético de la mediana. Si un valor muestral es exactamente igual a la mediana hipotética, no se registra ningún signo, y se reduce en forma correspondiente el tamaño efectivo de la muestra. Si es verdadera la hipótesis nula con respecto al valor de la mediana, el número de signos positivos debe ser aproximadamente igual al número de signos negativos. O, en otros términos, la proporción de signos positivos (o de signos negativos) debe ser de aproximadamente 0.50. Por ello, la hipótesis de una prueba de dos criterios de calificación es H_0: π » 0.50, en donde π es la proporción poblacional de signos positivos (o negativos). Por ello, una hipótesis que se refiere al valor de la mediana se prueba, de hecho, como una hipótesis sobre π. Si el tamaño de la muestra es pequeño (n < 30), se utiliza la distribución binomial para realizar la prueba, según se describe en la sección 11.4. Si la muestra es grande, puede utilizarse la distribución normal, como se describe en la sección 11.5.

En el problema 21.2 se utiliza la prueba del signo para probar una hipótesis nula referente a una mediana poblacional.

21.5 UNA MUESTRA: LA PRUEBA DE WILCOXON

Al igual que en el caso de la prueba del signo, puede utilizarse la *prueba de Wilcoxon* para probar una hipótesis nula sobre el valor de una mediana poblacional. Como la prueba de Wilcoxon considera la magnitud de la diferencia entre cada uno de los valores muestrales y el valor hipotético de la mediana, es una prueba más sensible que la prueba del signo. Por otro lado, como se determinan diferencias, los valores deben estar dados, cuando menos, en escala de intervalo. No se requieren suposiciones con respecto a la forma de la distribución poblacional.

Las hipótesis nula y alternativa se plantean con respecto a la mediana poblacional, y pueden ser de uno o de dos criterios de calificación. Se determina la diferencia entre cada uno de los valores observados y el valor hipotético de la mediana, y esta diferencia, con signo aritmético, se designa d: d = (X - Med_0). Si alguna de las diferencias es igual a cero, se elimina del análisis la observación correspondiente, y se reduce el tamaño efectivo de la muestra. Después, se ordenan los valores absolutos de la diferencia, de menor a mayor, asignando el rango de 1 a la diferencia absoluta más pequeña. Cuando las diferencias absolutas son iguales, se asigna el rango promedio a los valores que son iguales. Finalmente, se obtiene por separado la suma de los rangos para las diferencias positiva y negativa. La menor de esas dos sumas es la estadística T de Wilcoxon para una prueba de dos criterios de calificación. En el caso de las pruebas de un criterio de calificación, la suma menor debe corresponder a la dirección de la hipótesis nula. En el apéndice 10 se identifican los valores críticos de T, de acuerdo con el tamaño de la muestra y el nivel de significancia. Para rechazar la hipótesis nula, el valor que se obtiene de T debe ser *menor* que el valor critico dado en la tabla.

Cuando $n > 25$ y la hipótesis nula es cierta, la estadística T tiene una distribución aproximadamente normal. La media y el error estándar correspondientes a esa distribución muestral son, respectivamente:

$$\mu_T = \frac{n(n+1)}{4}$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Por ello, puede llevarse a cabo la prueba para una muestra relativamente grande utilizando la distribución de probabilidad normal, y calculando la estadística de prueba, de la siguiente manera:

$$z = \frac{T - \mu_T}{\sigma_T} \tag{21.6}$$

En el problema 21.3 se ilustra el uso de la prueba de Wilcoxon para probar una hipótesis nula sobre una mediana poblacional.

21.6 DOS MUESTRAS INDEPENDIENTES: LA PRUEBA DE MANN-WHITNEY

Puede utilizarse l*a prueba de Mann-Whttney para* probar ta hipótesis nula de que las medianas de dos poblaciones son iguales. Se supone que las dos poblaciones tienen la misma forma y la misma dispersión, porque si existieran diferencias en estos parámetros, podrían conducir a rechazar la hipótesis nula. Se requiere que los valores de las dos muestras aleatorias independientes se encuentren cuando menos en escala ordinal.

Se combinan dos muestras en un arreglo ordenado, identificando los valores muestrales de acuerdo con el grupo muestral al que pertenecen. Después, se ordenan los valores de menor a mayor, asignando el rango de 1 al valor más pequeño. Cuando se encuentran valores iguales, se les asigna el promedio de sus rangos. Si la hipótesis nula es cierta, el promedio de los rangos para los dos grupos muestrales debe ser aproximadamente igual. Se designa mediante U a la estadística que se calcula para realizar esta prueba, y puede basarse en la suma de los rangos de cualquiera de las dos muestras aleatorias, mediante las siguientes fórmulas:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

o

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

en donde n_1 = tamaño de la primera muestra
n_2 = tamaño de la segunda muestra
R_1 = suma de los rangos de la primera muestra
R_2 = suma de los rangos de la segunda muestra

Si $n_1 > 10$, $n_2 > 10$, y la hipótesis nula es cierta, la distribución muestral de U es aproximadamente normal, con los siguientes parámetros:

$$\mu_U = \frac{n_1 n_2}{2} \tag{21.9}$$

$$\sigma_U = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} \tag{21.10}$$

Por ello, la estadística de prueba para la hipótesis nula de que las medianas de dos poblaciones son iguales es

$$z = \frac{U - \mu_U}{\sigma_U} \tag{21.11}$$

en donde U es igual a U_1 o U_2.

En situaciones en las que $n_1 < 10$, $n_2 < 10$, o ambas, no es posible utilizar la distribución de probabilidad normal para esta prueba. Sin embargo, existen tablas especiales de la estadística U para muestras pequeñas, en libros especializados sobre estadística no paramétrica.

En el problema 21.4 se ilustra el uso de la prueba de Mann-Whitney.

21.7 OBSERVACIONES APAREADAS: LA PRUEBA DEL SIGNO

Puede utilizarse la *prueba del signo* descrita en la sección 21.4, cuando se tienen dos muestras recolectadas como *observaciones apareadas* (sección 11.3) para probar la hipótesis nula de que las medianas poblacionales son iguales. Los valores muestrales deben estar cuando menos en escala ordinal, y no se requieren suposiciones con respecto a las formas de las dos distribuciones poblacionales.

Se asigna un signo positivo (+) a cada par de valores cuya medición en la primera muestra es mayor que la medición en la segunda, y un signo negativo (-) cuando lo contrario resulta ser cierto. Sin un par de mediciones tienen el mismo valor, se eliminan esos valores empatados del análisis, y se reduce el tamaño efectivo de la muestra. Si es cierta la hipótesis de que las dos poblaciones se encuentran en el mismo nivel de magnitud, el número de signos positivos debe ser aproximadamente igual al de signos negativos. Por ello, la hipótesis nula que se prueba es H_0: $\pi = 0.50$, en donde π es la proporción poblacional de los signos positivos (o negativos). Si el número de pares muestrales es pequeño ($n < 30$), se utiliza la distribución binomial para realizar la prueba, como se describe en la sección 11.4. Si la muestra es grande ($n > 30$) puede utilizarse la distribución normal, según se describe en la sección 11.5. Observe que, aun cuando se han recolectado dos muestras, la prueba se aplica al conjunto de signos positivos y negativos que se obtiene al comparar los pares de mediciones.

El problema 21.5 ilustra el uso de la prueba del signo para probar la diferencia entre dos medianas, correspondientes a datos recolectados como observaciones apareadas.

21.8 OBSERVACIONES APAREADAS: LA PRUEBA DE WILCOXON

Puede utilizarse la *prueba de Wilcoxon* que se describe en la sección 21.5, cuando se tienen dos muestras recolectadas como observaciones apareadas, y cuando se desea probar la hipótesis nula de que las dos medianas poblacionales son iguales. Como la prueba de Wilcoxon considera la magnitud de la diferencia entre los valores en cada uno de los pares, y no solamente la dirección, o signo, de la diferencia, resulta ser una prueba más sensible que la prueba del signo. Sin embargo, los valores muestrales deben estar cuando menos en escala de intervalo. No se requieren suposiciones con respecto a las formas de las dos distribuciones.

Se determina la diferencia entre cada par de valores, y esta diferencia, con su correspondiente signo algebraico, se designa mediante d. Cuando las diferencias son iguales a cero, se eliminan las observaciones correspondientes y se reduce el tamaño efectivo de la muestra. Después, se ordenan los valores absolutos de las diferencias, de menor a mayor, asignando el rango de 1 a la diferencia absoluta más pequeña. Cuando las diferencias absolutas son iguales, se asigna el rango promedio de los valores empatados. Finalmente, se obtiene por separado la suma de los rangos para las diferencias positivas y negativas. La menor de esas dos sumas es la estadística T de Wilcoxon para una prueba de dos criterios de calificación. Para una prueba de un criterio de calificación, la suma más pequeña debe coincidir con la dirección de la hipótesis nula, como se ilustra en la aplicación de la prueba de Wilcoxon a una sola muestra, en el problema 21.3. En el apéndice 10 se identifican los valores críticos de T, de acuerdo con el tamaño de la muestra y el nivel de significancia. Para rechazar la hipótesis nula, la estadística T debe ser *menor* que el valor crítico dado en la tabla.

Cuando $n \geq 25$ y la hipótesis nula es cierta, la distribución de la estadística T es aproximadamente normal. Las fórmulas de la media y del error estándar de la distribución muestral de T, así como también para la estadística de prueba z, se dan en la sección 21.5, que se refiere al uso de la prueba de Wilcoxon a una muestra.

En el problema 21.6 se ilustra el uso de la prueba de Wilcoxon para probar la diferencia entre dos medianas, de datos recolectados como observaciones apareadas.

21.9 VARIAS MUESTRAS INDEPENDIENTES: LA PRUEBA DE KRUSKAL-WALLIS

La *prueba de Kruskal-Wallis* se utiliza para probar la hipótesis nula de que varías poblaciones tienen la misma mediana. Como tai, es el equivalente no paramétríco del diseño completamente aleatorizado de un factor del análisis de varianza. Se supone que las diversas poblaciones tienen la misma forma y dispersión para que la hipótesis anterior resulte aplicable, ya que diferencias en la forma o la dispersión pueden conducir a rechazar la hipótesis nula. Se requiere que los valores de las diversas muestras aleatorias independientes estén cuando menos en escala ordinal.

Primero se considera a las diversas muestras como un conjunto de valores, y se ordenan de menor a mayor los valores de ese conjunto combinado. Cuando se tienen valores iguales, se les asigna su rango promedio. Si la hipótesis nula es cierta, el promedio de rangos debe ser más o menos igual para cada uno de los grupos muestrales. La estadística de prueba que se calcula se designa mediante *H,* y se basa en la suma de los rangos de cada una de las diversas muestras aleatorias, de la siguiente manera:

$$H = \left\{ \left[\frac{12}{N(N+1)} \right] \left[\sum \frac{R_j^2}{n_j} \right] \right\} - 3(N+1) \qquad (21.12)$$

en donde N - tamaño de muestra combinado para las diversas muestras (observe que N no designa tamaño de la población en este caso)

R_j= suma de los rangos para la j-ésima muestra, o grupo de tratamiento

n_j=número de observaciones en la j-ésima muestra

Si el tamaño de cada una de las muestras es de cuando menos $n_j > 5$ y la hipótesis nula es cierta, la estadística H se distribuye en forma aproximada como la distribución x^2, con $gl = K - 1$, en donde K es el número de grupos de tratamiento, o muestras. El valor de x^2 que se aproxima el valor critico de la estadística de prueba es siempre el valor del extremo superior. Este procedimiento de prueba es similar al uso que se da al extremo superior de la distribución F en análisis de varianza.

La estadística H debe corregirse cuando se tienen rangos empatados. El valor corregido de la estadística de prueba se designa mediante H_c, y se calcula de la siguiente manera:

$$H_c = \frac{H}{1 - [\Sigma(t_j^3 - t_j)/(N^3 - N)]} \qquad (21.13)$$

en donde t_j representa el número de calificaciones empatadas en la *j*-ésima muestra.

El efecto de esta corrección consiste en aumentar el valor de la estadística H calculado. Por ello, si un valor de H corregido conduce a rechazar la hipótesis nula, no es necesario corregir el valor cuando se tienen rangos empatados.

En el problema 21.7, se ilustra el uso de la prueba de Kruskal-Wallis para probar la hipótesis nula de que diversas poblaciones tienen la misma mediana.

Problemas resueltos

LA PRUEBA DE RACHAS PARA ALEATORIEDAD

21.1 Se entrevista a una muestra de 36 personas en una investigación de mercados, conformada por 22 mujeres (M) y 14 hombres (H). Las personas muestreadas se entrevistaron en el siguiente orden:
H, M, M, M, M, H, H, H, M, H, M, M, M, H, H, M, M, M, M, H, M, M, M, H, H, M, M, M, H, M, H,H, M, M, M, H. Utilice la prueba de rachas para probar la aleatoriedad de este conjunto de observaciones, utilizando un nivel de significancia del 5%.

El número de rachas de esta muestra es R = 17, según se aprecia con el subrayado. Utilizando n_1= el número de mujeres, y n_2 = el número de hombres, puede utilizarse la distribución normal para probar la hipótesis nula de que el proceso de muestreo resultó ser aleatorio, porque $n_1 > 20$. Se calcula la media y el error estándar de la distribución muestral de R de la siguiente manera:

$$\mu_R = \frac{2n_1n_2}{n_1+n_2}+1 = \frac{2(22)(14)}{22+14}+1 = \frac{616}{36}+1 = 18.1$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}} = \sqrt{\frac{2(22)(14)[2(22)(14)-22-14]}{(22+14)^2(22+14-1)}}$$

$$= \sqrt{\frac{616(616-36)}{36^2(35)}} = \sqrt{\frac{357\,280}{45\,360}} = \sqrt{7.8765} = 2.81$$

Cuando se utiliza el nivel de significancia del 5%, los valores críticos de la estadística z son z = +1.96. El valor de la estadística de prueba z para estos datos es

$$z = \frac{R-\mu_R}{\sigma_R} = \frac{17-18.1}{2.81} = \frac{-1.1}{2.81} = -0.39$$

Por ello, utilizando el nivel de significancia del 5%, no se puede rechazar la hipótesis nula de que la secuencia de hombres y mujeres ocurrió al azar.

UNA MUESTRA: LA PRUEBA DEL SIGNO

21.2 Se afirma que las unidades ensambladas en un sistema rediseñado de ensamble de productos será mayor que con el sistema antiguo, cuya mediana poblacional es de 80 unidades por turno. *No* otorgando el beneficio de la duda al sistema rediseñado, plantee la hipótesis nula y pruébela a un nivel de significancia del 5%. Los datos muestrales se reportan en la primera parte de ila Tabla 21.1.

Tabla 21.1 Número de unidades ensambladas en el sistema rediseñado

Tumo muestreado	Unidades ensambladas (X)	Signo de la diferencia $(X - 80)$
1	75	–
2	85	+
3	92	+
4	80	0
5	94	+
6	90	+
7	91	+
8	76	-
9	88	+
10	82	+
11	96	+
12	83	+

Las hipótesis nula y alternativa son

$$H_0\text{: Med} \leq 80$$

$$H_1\text{: Med} > 80$$

Resulta apropiada una prueba no paramétrica porque no se hacen suposiciones con respecto a la forma de la distribución poblacional. Utilizando la prueba del signo, las hipótesis nula y alternativa, en términos de la proporción de signos positivos para las diferencias, y en donde d - (X - 80), son

$$H_0\text{: } \pi \leq 0.50$$

$$H_1\text{: } \pi > 0.50$$

En la Tabla 21.1 se observa que en el cuarto tumo muestreado, el número de unidades ensambladas resultó ser exactamente igual que el valor hipotético de la mediana poblacional. Por ello, esta observación se omite de cualquier análisis ulterior, y se tiene una muestra de tamaño efectivo de n =11. De los once signos de las diferencias que se reportan en la Tabla 21.1,9 son positivos. La prueba debe llevarse a cabo utilizando un nivel de significancia del 5%.

Como el tamaño de la muestra es n < 30, la base apropiada para esa prueba es la distribución binomial. Utilizando el enfoque del valor *P* para pruebas de hipótesis, como se describe en la sección 10.6, se determina la probabilidad de observar 9 o más signos positivos en 11 observaciones, dado que la proporción poblacional de signos positivos es 0.50, utilizando el apéndice 2 para obtener las probabilidades binomiales:

$$P(X \geq 9 \mid n = 11, \pi = 0.50) = 0.0269 + 0.0054 + 0.0005 = 0.0328$$

Como el valor *P* correspondiente ai resultado muestral es menor que 0.05, se rechaza la hipótesis nula a un nivel de significancia del 5% para esta prueba de un criterio de calificación. Es decir, la probabilidad de observar un número tan grande (o mayor) de signos positivos cuando la hipótesis nula es cierta es menor de 0.05 y, específicamente, la probabilidad es 0.0328. Por ello, se acepta la hipótesis alternativa y se concluye que la producción mediana por tumo es mayor de 80 unidades para el nuevo sistema de ensamble.

En el problema 21.5 se ¡lustra la prueba del signo de dos criterios de calificación, y se utiliza la prueba del signo para observaciones apareadas.

UNA MUESTRA: LA PRUEBA DE WILCOXON

21.3 Utilice la prueba de Wilcoxon para la hipótesis nula y los datos del problema 21.2, y compare su solución con la que se obtuvo ahi.

$$H_0: \text{Med}<80$$

$$H_1: \text{Med}>80$$

En la Tabla 21.2 se repiten los datos muestrales y se le utiliza como tabla de trabajo para la prueba de Wilcoxon. Observe que, en esta tabla, el número de unidades ensambladas correspondiente al cuarto turno muestreado es exactamente igual al valor hipotético de la mediana de la población. Por ello, se elimina esa observación y esto da como resultado un tamaño efectivo de muestra de π - 11. Observe también que el valor absoluto de d para los tumos muestreados en primero y segundo lugar es igual, por lo que se asigna a cada uno de ellos el rango promedio de 4.5 (en vez de asignarles los rangos 4 y 5). Al siguiente rango se le asigna entonces el número 6.

Tabla 21.2 Número de unidades ensambladas con el sistema rediseñado, y cálculo de los rangos con signo

Turno muestreado	Unidades ensambladas (X)	Diferencia $[d=(X-80)]$	Rango de $\|d\|$	Rango con signo	
				(+)	(-)
1	75	-5	4.5		4.5
2	85	5	4.5	4.5	
3	92	12	9	9	
4	80	0			
5	94	14	10	10	
6	90	10	7	7	
7	91	11	8	8	
8	76	-4	3		3
9	88	8	6	6	
10	82	2	1	1	
11	96	16	11	11	
12	83	3	2	2	
Total				58.5	T = 7.5

Para rechazar la hipótesis nula H_0: Med < 80 en esta prueba de un criterio de calificación, las diferencias d = (X - Med_0) deben ser predominantemente positivas, puesto que las diferencias negativas representan un apoyo para la hipótesis nula. Por ello, para esta prueba del criterio de calificación superior, la suma de los rangos de las diferencias negativas debe ser la suma de menor tamaño. En la Tabla 21.2 se observa que éste es, de hecho, el caso, y el valor del estadístico de prueba de Wilcoxon es T - 7.5.

La prueba de hipótesis nula debe realizarse en el nivel de significancia del 5%. En el apéndice 10 puede observarse que para n =11 (el tamaño efectivo de la muestra) el valor critico de T para una prueba de un criterio de calificación a un nivel de significancia del 5% es T = 14. Como el valor que se obtiene de T es menor que el valor critico, se rechaza la hipótesis nula a un nivel de significancia del 5%, y se concluye que el número mediano de unidades ensambladas con el nuevo sistema es mayor de 80 unidades. Esta es la misma conclusión a la que se llegó con la prueba del signo en la sección anterior, cuando se encontró que el valor P para la prueba fue de $P = 0.0328$. Sin embargo, observando el apéndice 10 para la prueba de Wilcoxon se encuentra que para $n = 11$ y $\alpha = 0.025$, el valor critico es $T = 11$. Por ello, serla posible rechazar la hipótesis nula en el nivel de significancia del 2.5% utilizando la prueba de Wilcoxon, pero no utilizando la prueba de los signos. Este detalle es consistente con las observaciones anteriores en el sentido de que la prueba de Wilcoxon es la más sensibie de las dos.

DOS MUESTRAS INDEPENDIENTES: LA PRUEBA DE MANN-WHITNEY

21.4 Para evaluar y comparar dos métodos de capacitación industrial, un director de capacitación asigna a 15 entrenandos elegidos al azar a cada uno de dos métodos. Debido a la deserción normal, 14 aprendices terminan el curso mediante el método 1, y 12 lo terminan llevando el método 2. A los dos grupos de entrenandos se les aplica el mismo examen para evaluar el aprendizaje, según se reporta en la Tabla 21.3. Pruebe la hipótesis nula de que el nivel mediano de desempeño en la prueba no difiere en los dos métodos de capacitación, utilizando un nivel de significancia del 5%.

Tabla 21.3 Calificaciones de los aprendices capacitados mediante los dos métodos de instrucción

Método 1	Método 2
70	86
90	78
82	90
64	82
86	65
77	87
84	80
79	88
82	95
89	85
73	76
81	94
83	
66	

$$H_0\text{: Med}_1 = \text{Med}_2 \qquad H_1\text{: Med}_1 \neq \text{Med}_2$$

En la Tabla 21.4 se enlistan las calificaciones del examen en un arreglo, ordenando las calificaciones de menor a mayor. Los rangos se presentan en la segunda columna de la Tabla 21.4. Observe la forma en que se asignan los rangos a las calificaciones empatadas: para las tres calificaciones de 82, por ejemplo, se tienen las posiciones 12, 13 y 14, en la segunda columna de la tabla. Por ello, el promedio de esos tres rangos es 13 y se les asigna a las tres calificaciones empatadas; después, observe que el siguiente rango asignado es 15, y no 14, debido a que asignar el rango promedio de 13 a las tres calificadones de 82 colocó la calificación de 83 en la posición 15 en el arreglo. De manera similar, existen dos calificaciones de 86; los rangos para esas calificaciones son 18 y 19, y, por lo tanto, el rango promedio es 18.5 y se le asigna a cada una de esas dos calificaciones. La siguiente calificación del arreglo es 87 y se le asigna el rango 20. Las últimas dos columnas de la Tabla 21.4 incluyen los rangos de acuerdo con la muestra, que representa el método de enseñanza utilizado, y las sumas de esas dos columnas de rangos son los valores R_1 y R_2, respectivamente.

Aplicando la fórmula *(21.7),* se determina el valor de U_1, de la siguiente manera:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 14(12) + \frac{14(14+1)}{2} - 161$$
$$= 168 + \frac{210}{2} - 161 = 112$$

Como $n_1 > 10$ y $n_2 > 10$, puede utilizarse la distribución de probabilidad normal para probar la hipótesis nula de que el valor de μ_u es el mismo para ambas muestras. La prueba se lleva a cabo en el nivel de significancia del 5%. Con base en la fórmula (27.9), el valor de μ_u es

$$\mu_U = \frac{n_1 n_2}{2} = \frac{14(12)}{2} = 84$$

Por ello, las hipótesis son

$$H_0\colon \mu_U = 84$$
$$H_1\colon \mu_U \neq 84$$

Con base en la fórmula *(21.10),* el error estándar de *U* es

$$\sigma_U = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} = \sqrt{\frac{14(12)(14+12+1)}{12}} = \sqrt{\frac{4\,536}{12}} = 19.4$$

Al aplicar la fórmula *(21.11),* se determina el valor del estadístico de prueba z de la siguiente manera:

$$z = \frac{U_1 - \mu_U}{\sigma_U} = \frac{112-84}{19.4} = 1.44$$

Como los valores críticos de z para una prueba a un nivel de significancia del 5% son $z = \pm 1.96$, no es posible rechazar la hipótesis de que no existe diferencia entre las medianas de las dos distribuciones de calificaciones.

OBSERVACIONES APAREADAS: LA PRUEBA DE LOS SIGNOS

21.5 A un *grupo* de consumidores que consta de 14 personas se le pide que califique dos marcas de té, de acuerdo con un sistema de evaluación por puntos que se basa en diversos criterios. En la Tabla 21.5 se reportan los puntos asignados, y se indica también el signo de la diferencia para cada par de calificaciones. Pruebe la hipótesis nula de que no existe diferencia en el nivel de las calificaciones para las dos marcas de té, a un nivel de significancia del 5%, utilizando la prueba del signo y planteando las hipótesis nula y alternativa en términos de la proporción de signos positivos.

Tabla 21.4 Arreglo combinado de las calificaciones con sus rangos correspondientes

Calificación *	Rango	Rangos del método 1	Rangos del método 2
64	1	1	
65	2		2
66	3	3	
70	4	4	
73	5	5	
76	6		6
77	7	7	
78	8		8
79	9	9	
80	10		10
81	11	11	
82	13	13	
82	13	13	
82	13		13
83	15	15	
84	16	16	
85	17		17
86	18.5	18.5	
86	18.5		18.5
87	20		20
88	21		21
89	22	22	
90	23.5	23.5	
90	23.5		23.5
94	25		25
95	26		26
Totales		$R_1 = 161.0$	$R_2 = 190.0$

* Se subrayan las calificaciones del método 1.

$$H_0: \pi = 0.50$$

$$H_1: \pi \neq 0.50$$

Al observar la Tabla 21.5 se nota que dos de los consumidores evaluaron de la misma manera dos marcas de té y, por ello, se omiten del análisis subsiguiente, lo cual da como resultado un tamaño efectivo de muestra de n =12. De los 12 signos, 8 son positivos.

Como el número de pares de valores es $n < 30$, la distribución binomial es la base apropiada para esta prueba. Al igual que en el problema 21.2, en el que se utilizó la prueba del signo para una muestra, se utiliza el método del valor *P* para probar la hipótesis. La probabilidad de observar 8 b más signos positivos en 12 observaciones, dado que la proporción poblacional de signos positivos es 0.50, se determina consultando las probabilidades binomiales en el apéndice 2:

$$\begin{aligned} P(X \geq 8 \mid n = 12, \pi = 0.50) &= 0.1208 + 0.0537 + 0.0161 + 0.0029 + 0.0002 \\ &= 0.1937 \end{aligned}$$

Tabla 21.5 Calificaciones asignadas por grupo de consumidores a dos marcas de té

Miembro del grupo	Calificación por puntos que se asigna a cada marca		Signo de la diferencia
	Marca 1	Marca 2	
1	20	16	+
2	24	26	-
3	28	18	+
4	24	17	+
5	20	20	0
6	29	21	+
7	19	23	-
8	27	22	+
9	20	23	-
10	30	20	+
11	18	18	0
12	28	21	+
13	26	17	+
14	24	26	-

La probabilidad calculada en el párrafo anterior corresponde a un criterio de calificación *P* (X > 8). Como en la aplicación que se revisa aquí se utiliza una prueba de dos criterios de calificación ***debe duplicarse esta probabilidad*** para obtener el valor *P:* $P = 2 \times 0.1937 = 0.3874$. Es decir, 0.3874 es la probabilidad de la diferencia observada en *cualquier* dirección con respecto al número esperado de signos positivos, que son $n\pi_0 = 12(0.50) = 6$. Este valor P es superior al nivel de probabilidad de 0.05 y, por ello, resulta evidente que no es posible rechazar la hipótesis nula de que las dos marcas de té son igualmente apreciadas. Observe que, en el problema 11.8 de la página 191 se obtuvo en forma separada la probabilidad para cada uno de los criterios de calificación de la distribución binomial. En este caso, simplemente se multiplica la probabilidad de un criterio de calificación por dos, porque la distribución binomial es simétrica cuando $\pi = 0.50$.

OBSERVACIONES APAREADAS: LA PRUEBA DE WILCOXON

21.6 Aplique la prueba de Wilcoxon a los datos del problema 21.5, utilizando un nivel de significancia del 5%.

En la Tabla 21.6 se repiten las calificaciones, se determinan los valores de d, se ordenan los valores absolutos de d y, finalmente, se determina la suma de rangos positivos y negativos.

En la Tabla 21.6 se observa que dos de los miembros del grupo que calificaron de igual manera las dos marcas de té se eliminan del análisis, porque a la diferencia d - 0 no se le asigna rango. Por ello, el tamaño efectivo de la muestra es $n = 12$. De los 12 valores absolutos de d, los diversos pares empatados dan como resultado la asignación de sus rangos promedio. Por ello, existen dos diferencias absolutas empatadas para las posiciones del rango 1 y el rango 2. Asi, a cada una de las diferencias absolutas se les asigna el rango promedio 1.5. El siguiente rango asignado es el 3. Como la suma de los rangos negativos es menor que la suma de los rangos positivos, a la suma menor se le identifica como *T* en la Tabla 21.6.

Tabla 21.6 Calificaciones asignadas por un panel de consumidores a dos marcas de té, y cálculo de los rangos con signo

Miembro del grupo	Marca 1	Marca 2 (X_2)	Diferencia	Rango de $\|d\|$	Rango con signo	
					(+)	(-)
1	20	16	4	4.5	4.5	
2	24	26	-2	1.5		1.5
3	28	18	10	11.5	11.5	
4	24	17	7	7.5	7.5	
5	20	20	0			
6	29	21	8	9	9	
7	19	23	-4	4.5		4.5
8	27	22	5	6	6	
9	20	23	-3	3		3
10	30	20	10	11.5	11.5	
11	18	18	0			
12	28	21	7	7.5	7.5	
13	26	17	9	10	10	
14	24	26	-2	1.5		1.5
Total					67.5	T=10.5

En el apéndice 10 puede observarse que para n = 12 (el tamaño efectivo de la muestra) el valor crítico de la estadística T de Wilcoxon, para una prueba de dos extremos en el nivel de significancia del 5%, es T= 14. Como T es la suma menor de los rangos para la prueba de dos criterios de calificación, el valor que se obtiene de la estadística de prueba, que se calcula en la Tabla 21.6, es T = 10.5. Como este valor es inferior al valor crítico, se rechaza la hipótesis nula de que no existe diferencia en las calificaciones para las dos marcas de té.

Observe que no fue posible rechazar la hipótesis nula de que no existe diferencia en el nivel del 5%, cuando se aplicó la prueba de los signos a los mismos datos, en el problema 21.5. Como la prueba de Wilcoxon considera la magnitud de la diferencia entre cada uno de los pares, y no solamente el signo de la diferencia, la prueba de Wilcoxon para observaciones apareadas es más sensible que la prueba de los signos.

VARIAS MUESTRAS INDEPENDIENTES: LA PRUEBA DE KRUSKAL-WALLIS

21.7 En el problema 13.1 se ilustró el diseño completamente aleatorizado de un factor de análisis de varianza. Se presentaron las calificaciones de una muestra aleatoria de n = 5 entrenandos que se asignaron al azar a cada uno de tres métodos de enseñanza. En estas secciones, la hipótesis nula era que no había diferencias entre los niveles promedio de aprendizaje para los tres métodos de enseñanza. Además, como se tenían muestras pequeñas, fue necesario suponer que las tres poblaciones tenían distribución normal. Suponga que no puede hacerse esa suposición. Aplique la prueba de Kruskal-Wallis para probar la hipótesis nula de que las diversas poblaciones tienen la misma mediana, utilizando el nivel de significancia del 5%.

En la Tabla 21.7 se repiten los datos de la Tabla 13.14, y se incluyen también los rangos, de menor a mayor, considerando que todas las calificaciones son un solo grupo combinado. Utilizando la tabla, se calcula el valor de la estadística de prueba de la siguiente manera:

Tabla 21.7 Calificaciones de las personas con tres métodos distintos de instrucción

Método A_1		Método A_2		Método A_3	
Calificación	Rango	Calificación	Rango	Calificación	Rango
86	12	90	15	82	9.5
79	6	76	5	68	1
81	7.5	88	13	73	4
70	2	82	9.5	71	3
84	11	89	14	81	7.5
Total	$R_1 = 38.5$		$R_2 = 56.5$		$R_3 = 25.0$

$$H = \left\{\left[\frac{12}{N(N+1)}\right]\left[\sum \frac{R_j^2}{n_j}\right]\right\} - 3(N+1)$$

$$= \left\{\left[\frac{12}{15(15+1)}\right]\left[\frac{(38.5)^2}{5} + \frac{(56.5)^2}{5} + \frac{(25.0)^2}{5}\right]\right\} - 3(15+1)$$

$$= \frac{12}{240}\left(\frac{5\,299.5}{5}\right) - 48 = \frac{63\,594}{1\,200} - 48 = 52.995 - 48 = 4.995$$

Como hubo algunos empates en las calificaciones, resulta apropiado corregir el valor de la estadística de prueba:

$$H_c = \frac{H}{1 - [\sum(t_j^3 - t_j)/(N^3 - N)]}$$

$$= \frac{4.995}{1 - \{[(1^3 - 1) + (1^3 - 1) + (2^3 - 2)]/(15^3 - 15)\}}$$

$$= \frac{4.995}{1 - [(0+0+6)/(3\,375 - 15)]} = \frac{4.995}{1 - (6/3\,360)}$$

$$= \frac{4.995}{1 - 0.00179} = \frac{4.995}{0.99821} = 5.004$$

Asi, el valor corregido H_c = 5.004 es sólo ligeramente mayor que el valor no corregido de H = 4.995, y los empates tienen un efecto despreciable sobre el valor de la estadística de prueba. Con gl = K - I = 3 - 1 = 2, en el apéndice 7 puede observarse que el valor crítico de la estadística de prueba es x^2= 5.99. Como el valor observado de la estadística de prueba es menor que el valor crítico, no es posible rechazar la hipótesis nula de que los tres grupos muestrales se obtuvieron de poblaciones que tienen la misma mediana. Se toma esta decisión en un nivel de significancia del 5%. Esta conclusión es consistente con la conclusión obtenida en los problemas 13.1 y 13.2, en los que se aplicó el análisis de varianza a estos mismos datos.

Problemas complementarios

LAS PRUEBAS DE RACHAS PARA ALEATORIEDAD

21.8 En la Tabla 2.16 de la página 26, se reportan 40 préstamos personales, provenientes de una muestra aleatoria. La secuencia de la recopilación de los datos aparece en el renglón de la tabla (es decir, la secuencia de montos observados fue \$932 000, \$1 000 000, \$956 000, etc.). El monto mediano de los préstamos para los datos de la tabla es \$944 500. Pruebe la aleatoriedad de esa secuencia de préstamos clasificando cada uno de ellos según esté por encima o por debajo de la mediana, y utilizando un nivel de significancia del 5%.

Resp. z critica = ± 1.96, R = 28 y z = 2.56. Se rechaza la hipótesis nula de que la secuencia es aleatoria. Existe un número demasiado grande de rachas.

UNA MUESTRA: LA PRUEBA DEL SIGNO

21.9 En la siguiente tabla se reportan las ventas de una nueva herramienta en una muestra de 12 ferreterías en un mes determinado. Se desconoce la forma de la distribución y, por ello, en este caso no resulta apropiada una prueba estadística paramétrica, considerando también el tamaño pequeño de la muestra. Utilice la prueba del signo para probar la hipótesis de que la mediana de las ventas en la población es mayor de 10.0 unidades por ferretería, utilizando un nivel de significancia del 5%.

Herramientas / ferretería	8	18	9	12	10	14	16	7	14	11	10	20

Resp. P = 0.1719; no es posible rechazarla hipótesis nula.

UNA MUESTRA: LA PRUEBA DE WILCOXON

21.10 Aplique la prueba de Wilcoxon a la hipótesis nula y a los datos del problema 21.9, utilizando el nivel de significancia del 5%.
Resp. T crítica = 11, T = 10; se rechaza la hipótesis nula.

DOS MUESTRAS INDEPENDIENTES: LA PRUEBA DE MANN-WHITNEY

21.11 En un departamento de control de calidad desean comparar el tiempo que se requiere para diagnosticar fallas de equipo utilizando dos sistemas alternativos. Se asigna al azar una muestra de 30 fallas de equipo, para diagnosticarlas mediante los dos sistemas. El primer sistema se utiliza para diagnosticar 14 fallas, y al segundo sistema se le asignan 16 fallas. En la Tabla 21.8 se reporta el tiempo total, en minutos, que se requiere para diagnosticar cada una de las fallas. No puede hacerse ninguna suposición con respecto a la distribución del tiempo total que se requiere para estos diagnósticos. Utilizando el nivel de significancia del 10%, pruebe la hipótesis nula de que las dos muestras se obtuvieron de poblaciones que tienen la misma mediana.

Resp. R_1 = 173.5, R_2 = 291.5, z crítica = ± 1.645, z = 1.87; se rechaza la hipótesis nula en α = 0.10.

OBSERVACIONES APAREADAS: LA PRUEBA DE LOS SIGNOS

21.12 En vez del diseño experimental que se utilizó en el problema 21.11, se usaron los dos sistemas para diagnosticar la misma muestra aleatoria de 10 fallas de equipo. Asi, se diagnosticó dos veces cada uno de los 10 equipos, y se registra el tiempo que se requiere para diagnosticar la falla, para cada sistema. En la Tabla 21.9 se muestra el número de minutos que se requiere para cada diagnóstico. Utilizando el nivel de significancia del 10%, pruebe la hipótesis nula de que las dos muestras se obtuvieron en poblaciones que tienen la misma mediana. Es decir, la hipótesis nula

consiste en afirmar que no hay diferencia entre la cantidad mediana de tiempo que se requiere para diagnosticar fallas de equipo mediante los dos métodos. En el análisis, utilice la prueba del signo.
Resp. P = 0.3438; no es posible rechazar la hipótesis nula.

Tabla 21.8 Tiempo que se requiere para diagnosticar fallas de equipo (en minutos)

Sistema 1	Sistema 2
25	18
29	37
42	40
16	56
31	49
14	28
33	20
45	34
26	39
34	47
30	31
43	65
28	38
19	32
	24
	49

Tabla 21.9 Tiempo en minutos que se requiere para diagnosticar fallas de equipo, mediante el diseño de observaciones apareadas

Equipo muestreado	Sistema 1	Sistema 2
1	23	21
2	40	48
3	35	45
4	24	22
5	17	19
6	32	37
7	27	29
8	32	38
9	25	24
10	30	36

OBSERVACIONES APAREADAS: LA PRUEBA DE WILCOXON

21.13 En el nivel de significancia del 10%, aplique la prueba de Wilcoxon para observaciones apareadas con los datos de . la Tabla 21.9, probando de nuevo la hipótesis nula de que no existe diferencia entre la mediana de la cantidad de tiempo que se requiere con los dos métodos. Compare sus resultados con los que se obtuvieron en el problema 21.12.

Resp. T critica = 11, T = 8.0; se rechaza la hipótesis nula con a = 0.10.

INDEPENDIENTES: LA PRUEBA DE KRUSKAL-WALLIS

21.14 Además de los datos que se reportan en la Tabla 21.8, suponga que un tercer sistema de diagnóstico se evalúa asignando al azar 12 fallas a este sistema, y observando el tiempo que se requiere para el diagnóstico. Los minutos que se requieren para los diagnósticos mediante este tercer sistema son 21,36,34,19,46,25, 38, 31,20,26, 30 y 18. Utilizando el nivel de significancia del 10%, pruebe la hipótesis nula de que las tres muestras se obtuvieron en poblaciones con la misma mediana

Resp. R_1 = 260.5, R_2 = 431.0, R_3 = 211.5, x^2 critica = 4.61, H = 5.12; se rechaza la hipótesis nula con a = 0.10.

Apéndice 1

Tabla de números aleatorios

10097	85017	84532	13618	23157	86952	02438	76520	91499	38631	79430	62421	97959	67422	69992	68479
37542	16719	82789	69041	05545	44109	05403	64894	80336	49172	16332	44670	35089	17691	89246	26940
08422	65842	27672	82186	14871	22115	86529	19645	44104	89232	57327	34679	62235	79655	81336	85157
99019	76875	20684	39187	38976	94324	43204	09376	12550	02844	15026	32439	58537	48274	81330	11100
12807	93640	39160	41453	97312	41548	93137	80157	63606	40387	65406	37920	08709	60623	2237	16505
66065,	99478	70086	71265	11742	18226	29004	34072	61196	80240	44177	51171	08723	39323	05798	26457
31060	65119	26486	47353	43361	99436	42753	45571	15474	44910	99321	72173	56239	04595	10836	95270
85269	70322	21592	48233	93806	32584	21828	02051	94557	33663	86347	00926	44915	34823	51770	67897
63573	58133	41278	11697	49540	61777	67954	05325	42481	86430	19102	37420	41976	76559	24358	97344
73796	44655	81255	31133	36768	60452	38537	03529	23523	31379	68588	81675	15694	43438	36879	73208
98520	02295	13487	98662	07092	44673	61303	14905	04493	98086	32533	17767	14523	52494	24826	75246
11805	85035	54881	35587	43310	48897	48493	39808	00549	33185	04805	05431	94598	97654	16232	64051
83452	01197	86935	28021	61570	23350	65710	06288	35963	80951	68953	99634	81949	15307	00406	26898
88685	97907	19078	40646	31352	48625	44369	86507	59808	79752	02529	40200	73742	08391	49140	45427
99594	63268	96905	28797	57048	46359	74294	87517	46058	18633	99970	67348	49329	95236	32537	01390
65481	52841	59684	67411	09243	56092	84369	17468	32179	74029	74717	17674	90446	00597	45240	87379
80124	53722	71399	10916	07959	21225	13018	17727	69234	54178	10805	35635	45266	61406	41941	20117
74350	11434	51908	62171	93732	26958	02400	77402	19565	11664	77602	99817	28573	41430	96382	01758
69916	62375	99292	21177	72721	66995	07289	66252	45155	48324	32135	26803	16213	14938	71961	19476
09893	28337	20923	87929	61020	62841	31374	14225	94864	69074	45753	20505	78317	31994	98145	36168

Apéndice 2

Probabilidades binomiales*

							p					
n	x	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
1	0	.9900	.9500	.9000	.8500	.8000	.7500	.7000	.6500	.6000	.5500	.5000
	1	.0100	.0500	.1000	.1500	.2000	.2500	.3000	.3500	.4000	.4500	.5000
2	0	.9801	.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
	1	.0198	.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
	2	.0001	.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500
3	0	.9703	.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
	1	.0294	.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
	2	.0003	.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
	3	.0000	.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250
4	0	.9606	.8145	.6561	.5220	.4096	.3164	.2401	.1785	.1296	.0915	.0625
	1	.0388	.1715	.2916	.3685	.4096	.4219	.4116	.3845	.3456	.2995	.2500
	2	.0006	.0135	.0486	.0975	.1536	.2109	.2646	.3105	.3456	.3675	.3750
	3	.0000	.0005	.0036	.0115	.0256	.0469	.0756	.1115	.1536	.2005	.2500
	4	.0000	.0000	.0001	.0005	.0016	.0039	.0081	.0150	.0256	.0410	.0625
5	0	.9510	.7738	.5905	.4437	.3277	.2373	.1681	.1160	.0778	.0503	.0312
	1	.0480	.2036	.3280	.3915	.4096	.3955	.3602	.3124	.2592	.2059	.1562
	2	.0010	.0214	.0729	.1382	.2048	.2637	.3087	.3364	.3456	.3369	.3125
	3	.0000	.0011	.0081	.0244	.0512	.0879	.1323	.1811	.2304	.2757	.3125
	4	.0000	.0000	.0004	.0022	.0064	.0146	.0284	.0488	.0768	.1128	.1562
	5	.0000	.0000	.0000	.0001	.0003	.0010	.0024	.0053	.0102	.0185	.0312
6	0	.9415	.7351	.5314	.3771	.2621	.1780	.1176	.0754	.0467	.0277	.0156
	1	.0571	.2321	.3543	.3993	.3932	.3560	.3025	.2437	.1866	.1359	.0938
	2	.0014	.0305	.0984	.1762	.2458	.2966	.3241	.3280	.3110	.2780	.2344
	3	.0000	.0021	.0146	.0415	.0819	.1318	.1852	.2355	.2765	.3032	.3125
	4	.0000	.0001	.0012	.0055	.0154	.0330	.0595	.0951	.1382	.1861	.2344
	5	.0000	.0000	.0001	.0004	.0015	.0044	.0102	.0205	.0369	.0609	.0938
	6	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0018	.0041	.0083	.0156
7	0	.9321	.6983	.4783	.3206	.2097	.1335	.0824	.0490	.0280	.0152	.0078
	1	.0659	.2573	.3720	.3960	.3670	.3115	.2471	.1848	.1306	.0872	.0547
	2	.0020	.0406	.1240	.2097	.2753	.3115	.3177	.2985	.2613	.2140	.1641
	3	.0000	.0036	.0230	.0617	.1147	.1730	.2269	.2679	.2903	.2918	.2734
	4	.0000	.0002	.0026	.0109	.0287	.0577	.0972	.1442	.1935	.2388	.2734
	5	.0000	.0000	.0002	.0012	.0043	.0115	.0250	.0466	.0774	.1172	.1641
	6	.0000	.0000	.0000	.0001	.0004	.0013	.0036	.0084	.0172	.0320	.0547
	7	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0006	.0016	.0037	.0078
8	0	.9227	.6634	.4305	.2725	.1678	.1002	.0576	.0319	.0168	.0084	.0039
	1	.0746	.2793	.3826	.3847	.3355	.2670	.1977	.1373	.0896	.0548	.0312
	2	.0026	.0515	.1488	.2376	.2936	.3115	.2085	.2587	.2090	.1569	.1094
	3	.0001	.0054	.0331	.0839	.1468	.2076	.2541	.2786	.2787	.2568	.2188
	4	.0000	.0004	.0046	.0185	.0459	.0865	.1361	.1875	.2322	.2627	.2734
	5	.0000	.0000	.0004	.0026	.0092	.0231	.0467	.0808	.1239	.1719	.2188
	6	.0000	.0000	.0000	.0002	.0011	.0038	.0100	.0217	.0413	.0403	.1094
	7	.0000	.0000	.0000	.0000	.0001	.0004	.0012	.0033	.0079	.0164	.0312
	8	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0007	.0017	.0039
9	0	.9135	.6302	.3874	.2316	.1342	.0751	.0404	.0207	.0101	.0046	.0020
	1	.0830	.2985	.3874	.3679	.3020	.2253	.1556	.1004	.0605	.0339	.0176
	2	.0034	.0629	.1722	.2597	.3020	.3003	.2668	.2162	.1612	.1110	.0703
	3	.0001	.0077	.0446	.1069	.1762	.2336	.2668	.2716	.2508	.2119	.1641
	4	.0000	.0006	.0074	.0283	.0661	.1168	.1715	.2194	.2508	.2600	.2461
	5	.0000	.0000	.0008	.0050	.0165	.0389	.0735	.1181	.1672	.2128	.2461
	6	.0000	.0000	.0001	.0006	.0028	.0087	.0210	.0424	.0743	.1160	.1641
	7	.0000	.0000	.0000	.0000	.0003	.0012	.0039	.0098	.0212	.0407	.0703
	8	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0013	.0035	.0083	.0176
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0008	.0020
10	0	.9044	.5987	.3487	.1969	.1074	.0563	.0282	.0135	.0060	.0025	.0010
	1	.0914	.3151	.3874	.3474	.2684	.1877	.1211	.0725	.0403	.0207	.0098
	2	.0042	.0746	.1937	.2759	.3020	.2816	.2335	.1757	.1209	.0763	.0439
	3	.0001	.0105	.0574	.1298	.2013	.2503	.2668	.2522	.2150	.1665	.1172
	4	.0000	.0010	.0112	.0401	.0881	.1460	.2001	.2377	.2508	.2384	.2051
	5	.0000	.0001	.0015	.0085	.0264	.0584	.1029	.1536	.2007	.2340	.2461
	6	.0000	.0000	.0001	.0012	.0055	.0162	.0368	.0689	.1115	.1596	.2051
	7	.0000	.0000	.0000	.0001	.0008	.0031	.0090	.0212	.0425	.0746	.1172
	8	.0000	.0000	.0000	.0000	.0001	.0004	.0014	.0043	.0106	.0229	.0439
	9	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016	.0042	.0098
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
11	0	.8953	.5688	.3138	.1673	.0859	.0422	.0198	.0088	.0036	.0014	.0005
	1	.0995	.3293	.3835	.3248	.2362	.1549	.0932	.0518	.0266	.0125	.0054
	2	.0050	.0867	.2131	.2866	.2953	.2581	.1998	.1395	.0887	.0513	.0269
	3	.0002	.0137	.0710	.1517	.2215	.2581	.2568	.2254	.1774	.1259	.0806
	4	.0000	.0014	.0158	.0536	.1107	.1721	.2201	.2428	.2365	.2060	.1611
	5	.0000	.0001	.0025	.0132	.0388	.0803	.1321	.1830	.2207	.2360	.2256
	6	.0000	.0000	.0003	.0023	.0097	.0268	.0566	.0985	.1471	.1931	.2256
	7	.0000	.0000	.0000	.0003	.0017	.0064	.0173	.0379	.0701	.1128	.1611
	8	.0000	.0000	.0000	.0000	.0002	.0011	.0037	.0102	.0234	.0462	.0806
	9	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018	.0052	.0126	.0269
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0007	.0021	.0054
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0005
12	0	.8864	.5404	.2824	.1422	.0687	.0317	.0138	.0057	.0022	.0008	.0002
	1	.1074	.3413	.3766	.3012	.2062	.1267	.0712	.0368	.0174	.0075	.0029
	2	.0060	.0988	.2301	.2924	.2835	.2323	.1678	.1088	.0639	.0339	.0161
	3	.0002	.0173	.0852	.1720	.2362	.2581	.2397	.1954	.1419	.0923	.0537
	4	.0000	.0021	.0213	.0683	.1329	.1936	.2311	.2367	.2128	.1700	.1208
	5	.0000	.0002	.0038	.0193	.0532	.1032	.1585	.2039	.2270	.2225	.1934
	6	.0000	.0000	.0005	.0040	.0155	.0401	.0792	.1281	.1766	.2124	.2256
	7	.0000	.0000	.0000	.0006	.0033	.0115	.0291	.0591	.1009	.1489	.1934
	8	.0000	.0000	.0000	.0001	.0005	.0024	.0078	.0199	.0420	.0762	.1208
	9	.0000	.0000	.0000	.0000	.0001	.0004	.0015	.0048	.0125	.0277	.0537
	10	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0025	.0068	.0161
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0029
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
13	0	.8775	.5133	.2542	.1209	.0550	.0238	.0097	.0037	.0013	.0004	.0001
	1	.1152	.3512	.3672	.2774	.1787	.1029	.0640	.0259	.0113	.0045	.0016
	2	.0070	.1109	.2448	.2937	.2680	.2059	.1388	.0836	.0453	.0220	.0095
	3	.0003	.0214	.0997	.1900	.2457	.2517	.2181	.1651	.1107	.0660	.0349
	4	.0000	.0028	.0277	.0838	.1535	.2097	.2337	.2222	.1845	.1360	.0873

*Ejemplo P (X = 3| n = 5,p = 0.30) = 0.1323

n	x	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
13	5	.0000	.0003	.0055	.0266	.0691	.1258	.1803	.2154	.2214	.1989	.1571
	6	.0000	.0000	.0008	.0063	.0230	.0559	.1030	.1546	.1968	.2169	.2095
	7	.0000	.0000	.0001	.0011	.0058	.0186	.0442	.0833	.1312	.1775	.2095
	8	.0000	.0000	.0001	.0001	.0011	.0047	.0142	.0336	.0656	.1089	.1571
	9	.0000	.0000	.0000	.0000	.0001	.0009	.0034	.0101	.0243	.0495	.0873
	10	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0022	.0065	.0162	.0349
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0012	.0036	.0095
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0016
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
14	0	.8687	.4877	.2288	.1028	.0440	.0178	.0068	.0024	.0008	.0002	.0001
	1	.1229	.3593	.3559	.2539	.1539	.0832	.0407	.0181	.0073	.0027	.0009
	2	.0081	.1229	.2570	.2912	.2501	.1802	.1134	.0634	.0317	.0141	.0056
	3	.0003	.0259	.1142	.2056	.2501	.2402	.1943	.1366	.0845	.0462	.0222
	4	.0000	.0037	.0349	.0998	.1720	.2202	.2290	.2022	.1549	.1040	.0611
	5	.0000	.0004	.0078	.0352	.0860	.1468	.1963	.2178	.2066	.1701	.1222
	6	.0000	.0000	.0013	.0093	.0322	.0734	.1262	.1759	.2066	.2088	.1833
	7	.0000	.0000	.0002	.0019	.0092	.0280	.0618	.1082	.1574	1952	.2095
	8	.0000	.0000	.0000	.0003	.0020	.0082	.0232	.0510	.0918	.1398	.1833
	9	.0000	.0000	.0000	.0000	.0003	.0018	.0066	.0183	.0408	.0762	.1222
	10	.0000	.0000	.0000	.0000	.0000	.0003	.0014	.0049	.0136	.0312	.0611
	11	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0033	.0093	.0222
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0019	.0056
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002	.0009
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
15	0	.8601	.4633	.2059	.0874	.0352	.0134	.0047	.0016	.0005	.0001	.0000
	1	.1303	.3658	.3432	.2312	.1319	.0668	.0305	.0126	.0047	.0016	.0005
	2	.0092	.1348	.2669	.2856	.2309	.1559	.0916	.0476	.0219	.0090	.0032
	3	.0004	.0307	.1285	.2184	.2501	.2252	.1700	.1110	.0634	.0318	.0139
	4	.0000	.0049	.0428	.1156	.1876	.2252	.2186	.1792	.1268	.0780	.0417
	5	.0000	.0006	.0105	.0449	.1032	.1651	.2061	.2123	.1859	.1404	.0916
	6	.0000	.0000	.0019	.0132	.0430	.0917	.1472	.1906	.2066	.1914	.1527
	7	.0000	.0000	.0003	.0030	.0138	.0393	.0811	.1319	.1771	.2013	.1964
	8	.0000	.0000	.0000	.0005	.0035	.0131	.0348	.0710	.1181	.1647	.1964
	9	.0000	.0000	.0000	.0001	.0007	.0034	.0116	.0298	.0612	.1048	.1527
	10	.0000	.0000	.0000	.0000	.0001	.0007	.0030	.0096	.0245	.0515	.0916
	11	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0074	.0191	.0417
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052	.0139
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010	.0032
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
16	0	.8515	.4401	.1853	.0743	.0281	.0100	.0033	.0010	.0003	.0001	.0000
	1	.1376	.3706	.3294	.2097	.1126	.0535	.0228	.0087	.0030	.0009	.0002
	2	.0104	.1463	.2745	.2775	.2111	.1336	.0732	.0353	.0150	.0056	.0018
	3	.0005	.0359	.1423	.2285	.2463	.2079	.1465	.0888	.0468	.0215	.0085
	4	.0000	.0061	.0514	.1311	.2001	.2252	.2040	.1553	.1014	.0572	.0278
	5	.0000	.0008	.0137	.0555	.1201	.1802	.2099	.2008	.1623	.1123	.0667
	6	.0000	.0001	.0028	.0180	.0550	.1101	.1649	.1982	.1983	.1684	.1222

n	x	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
16	7	.0000	.0000	.0004	.0045	.0197	.0524	.1010	.1524	.1889	.1969	.1746
	8	.0000	.0000	.0001	.0009	.0055	.0197	.0487	.0923	.1417	.1812	.1964
	9	.0000	.0000	.0000	.0001	.0012	.0058	.0185	.0442	.0840	.1318	.1746
	10	.0000	.0000	.0000	.0000	.0002	.0014	.0056	.0167	.0392	.0755	.1222
	11	.0000	.0000	.0000	.0000	.0000	.0002	.0013	.0049	.0142	.0337	.0667
	12	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0040	.0115	.0278
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0008	.0029	.0085
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0002
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
17	0	.8429	.4181	.1668	.0631	.0225	.0075	.0023	.0007	.0002	.0000	.0000
	1	.1447	.3741	.3150	.1893	.0957	.0426	.0169	.0060	.0019	.0005	.0001
	2	.0117	.1575	.2800	.2673	.1914	.1136	.0581	.0260	.0102	.0035	.0010
	3	.0006	.0415	.1556	.2359	.2393	.1893	.1245	.0701	.0341	.0144	.0052
	4	.0000	.0076	.0605	.1457	.2093	.2209	.1868	.1320	.0796	.0411	.0182
	5	.0000	.0010	.0175	.0668	.1361	.1914	.2081	.1849	.1379	.0875	.0472
	6	.0000	.0001	.0039	.0236	.0680	.1276	.1784	.1991	.1839	.1432	.0944
	7	.0000	.0000	.0007	.0065	.0267	.0668	.1201	.1685	.1927	.1841	.1484
	8	.0000	.0000	.0001	.0014	.0084	.0279	.0644	.1134	.1606	.1883	.1855
	9	.0000	.0000	.0000	.0003	.0021	.0093	.0276	.0611	.1070	.1540	.1855
	10	.0000	.0000	.0000	.0000	.0004	.0025	.0095	.0263	.0571	.1008	.1484
	11	.0000	.0000	.0000	.0000	.0001	.0005	.0026	.0090	.0242	.0525	.0944
	12	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0081	.0215	.0472
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0021	.0068	.0182
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0004	.0016	.0052
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003	.0010
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
18	0	.8345	.3972	.1501	.0536	.0180	.0056	.0016	.0004	.0001	.0000	.0000
	1	.1517	.3763	.3002	.1704	.0811	.0338	.0126	.0042	.0012	.0003	.0001
	2	.0130	.1683	.2835	.2556	.1723	.0958	.0458	.0190	.0069	.0022	.0006
	3	.0007	.0473	.1680	.2406	.2297	.1704	.1046	.0547	.0246	.0095	.0031
	4	.0000	.0093	.0700	.1592	.2153	.2130	.1681	.1104	.0614	.0291	.0117
	5	.0000	.0014	.0218	.0787	.1507	.1988	.2017	.1664	.1146	.0666	.0327
	6	.0000	.0002	.0052	.0301	.0816	.1436	.1873	.1941	.1655	.1181	.0708
	7	.0000	.0000	.0010	.0091	.0350	.0820	.1376	.1792	.1892	.1657	.1214
	8	.0000	.0000	.0002	.0022	.0120	.0376	.0811	.1327	.1734	.1864	.1669
	9	.0000	.0000	.0000	.0004	.0033	.0139	.0386	.0794	.1284	.1694	.1855
	10	.0000	.0000	.0000	.0001	.0008	.0042	.0149	.0385	.0771	.1248	.1669
	11	.0000	.0000	.0000	.0000	.0001	.0010	.0046	.0151	.0374	.0742	.1214
	12	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0047	.0145	.0354	.0708
	13	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0045	.0134	.0327
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011	.0039	.0117
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0009	.0031
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000

n	x	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
19	0	.8262	.3774	.1351	.0456	.0144	.0042	.0011	.0003	.0001	.0000	.0000
	1	.1586	.3774	.2852	.1529	.0685	.0268	.0093	.0029	.0008	.0002	.0000
	2	.0144	.1787	.2852	.2428	.1540	.0803	.0358	.0138	.0046	.0013	.0003
	3	.0008	.0533	.1796	.2428	.2182	.1517	.0869	.0422	.0175	.0062	.0018
	4	.0000	.0112	.0798	.1714	.2182	.2023	.1491	.0909	.0467	.0203	.0074
	5	.0000	.0018	.0266	.0907	.1636	.2023	.1916	.1468	.0933	.0497	.0222
	6	.0000	.0002	.0069	.0374	.0955	.1574	.1916	.1844	.1451	.0949	.0518
	7	.0000	.0000	.0014	.0122	.0443	.0974	.1525	.1844	.1797	.1443	.0961
	8	.0000	.0000	.0002	.0032	.0166	.0487	.0981	.1489	.1797	.1771	.1442
	9	.0000	.0000	.0000	.0007	.0051	.0198	.0514	.0980	.1464	.1771	.1762
	10	.0000	.0000	.0000	.0001	.0013	.0066	.0220	.0528	.0976	.1449	.1762
	11	.0000	.0000	.0000	.0000	.0003	.0018	.0077	.0233	.0532	.0970	.1442
	12	.0000	.0000	.0000	.0000	.0000	.0004	.0022	.0083	.0237	.0529	.0961
	13	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0024	.0085	.0233	.0518
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0024	.0082	.0222
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0022	.0074
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0005	.0018
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0003
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
20	0	.8179	.3585	.1216	.0388	.0115	.0032	.0008	.0002	.0000	.0000	.0000
	1	.1652	.3774	.2702	.1368	.0576	.0211	.0068	.0020	.0005	.0001	.0000
	2	.0159	.1887	.2852	.2293	.1369	.0669	.0278	.0100	.0031	.0008	.0002
	3	.0010	.0596	.1901	.2428	.2054	.1339	.0716	.0323	.0123	.0040	.0011
	4	.0000	.0133	.0898	.1821	.2182	.1897	.1304	.0738	.0350	.0139	.0046
	5	.0000	.0022	.0319	.1028	.1746	.2023	.1789	.1272	.0746	.0365	.0148
	6	.0000	.0003	.0089	.0454	.1091	.1686	.1916	.1712	.1244	.0746	.0370
	7	.0000	.0000	.0020	.0160	.0545	.1124	.1643	.1844	.1659	.1221	.0739
	8	.0000	.0000	.0004	.0046	.0222	.0609	.1144	.1614	.1797	.1623	.1201
	9	.0000	.0000	.0001	.0011	.0074	.0271	.0654	.1158	.1597	.1771	.1602
	10	.0000	.0000	.0000	.0002	.0020	.0099	.0308	.0686	.1171	.1593	.1762
	11	.0000	.0000	.0000	.0000	.0005	.0030	.0120	.0336	.0710	.1185	.1602
	12	.0000	.0000	.0000	.0000	.0001	.0008	.0039	.0136	.0355	.0727	.1201
	13	.0000	.0000	.0000	.0000	.0000	.0002	.0010	.0045	.0146	.0366	.0739
	14	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0049	.0150	.0370
	15	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0049	.0148
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0013	.0046
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0011
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
25	0	.7778	.2774	.0718	.0172	.0038	.0008	.0001	.0000	.0000	.0000	.0000
	1	.1964	.3650	.1994	.0759	.0236	.0063	.0014	.0003	.0000	.0000	.0000
	2	.0238	.2305	.2659	.1607	.0708	.0251	.0074	.0018	.0004	.0001	.0000
	3	.0018	.0930	.2265	.2174	.1358	.0641	.0243	.0076	.0019	.0004	.0001
	4	.0001	.0269	.1384	.2110	.1867	.1175	.0572	.0224	.0071	.0018	.0004

n	x	.01	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50
25	5	.0000	.0060	.0646	.1564	.1960	.1645	.1030	.0506	.0199	.0063	.0016
	6	.0000	.0010	.0239	.0920	.1633	.1828	.1472	.0908	.0442	.0172	.0053
	7	.0000	.0001	.0072	.0441	.1108	.1654	.1712	.1327	.0800	.0381	.0143
	8	.0000	.0000	.0018	.0175	.0623	.1241	.1651	.1607	.1200	.0701	.0322
	9	.0000	.0000	.0004	.0058	.0294	.0781	.1336	.1635	.1511	.1084	.0609
	10	.0000	.0000	.0000	.0016	.0118	.0417	.0916	.1409	.1612	.1419	.0974
	11	.0000	.0000	.0000	.0004	.0040	.0189	.0536	.1034	.1465	.1583	.1328
	12	.0000	.0000	.0000	.0000	.0012	.0074	.0268	.0650	.1140	.1511	.1550
	13	.0000	.0000	.0000	.0000	.0003	.0025	.0115	.0350	.0760	.1236	.1550
	14	.0000	.0000	.0000	.0000	.0000	.0007	.0042	.0161	.0434	.0867	.1328
	15	.0000	.0000	.0000	.0000	.0000	.0002	.0013	.0064	.0212	.0520	.0974
	16	.0000	.0000	.0000	.0000	.0000	.0000	.0004	.0021	.0088	.0266	.0609
	17	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0031	.0115	.0322
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0009	.0042	.0143
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0013	.0053
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0016
	21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0004
	22	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001
30	0	.7397	.2146	.0424	.0076	.0012	.0002	.0000	.0000	.0000	.0000	.0000
	1	.2242	.3389	.1413	.0404	.0093	.0018	.0003	.0000	.0000	.0000	.0000
	2	.0328	.2586	.2277	.1034	.0337	.0086	.0018	.0003	.0000	.0000	.0000
	3	.0031	.1270	.2361	.1703	.0785	.0269	.0072	.0015	.0003	.0000	.0000
	4	.0002	.0451	.1771	.2028	.1325	.0604	.0208	.0056	.0012	.0002	.0000
	5	.0000	.0124	.1023	.1861	.1723	.1047	.0464	.0157	.0041	.0008	.0001
	6	.0000	.0027	.0474	.1368	.1795	.1455	.0829	.0353	.0115	.0029	.0006
	7	.0000	.0005	.0180	.0828	.1538	.1662	.1219	.0652	.0263	.0081	.0019
	8	.0000	.0001	.0058	.0420	.1106	.1593	.1501	.1009	.0505	.0191	.0055
	9	.0000	.0000	.0016	.0181	.0676	.1298	.1573	.1328	.0823	.0382	.0133
	10	.0000	.0000	.0004	.0067	.0355	.0909	.1416	.1502	.1152	.0656	.0280
	11	.0000	.0000	.0001	.0022	.0161	.0551	.1103	.1471	.1396	.0976	.0509
	12	.0000	.0000	.0000	.0006	.0064	.0291	.0749	.1254	.1474	.1265	.0806
	13	.0000	.0000	.0000	.0001	.0022	.0134	.0444	.0935	.1360	.1433	.1115
	14	.0000	.0000	.0000	.0000	.0007	.0054	.0231	.0611	.1101	.1424	.1354
	15	.0000	.0000	.0000	.0000	.0002	.0019	.0106	.0351	.0783	.1242	.1445
	16	.0000	.0000	.0000	.0000	.0000	.0006	.0042	.0177	.0489	.0953	.1354
	17	.0000	.0000	.0000	.0000	.0000	.0002	.0015	.0079	.0269	.0642	.1115
	18	.0000	.0000	.0000	.0000	.0000	.0000	.0005	.0031	.0129	.0379	.0806
	19	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0010	.0054	.0196	.0509
	20	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0020	.0088	.0280
	21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006	.0034	.0133
	22	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0002	.0012	.0055
	23	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0003	.0019
	24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0006
	25	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

Apéndice 3

$-\lambda$

λ	$e^{-\lambda}$	λ	e^{λ}
0.0	1.00000	2.5	.08208
0.1	.90484	2.6	.07427
0.2	.81873	2.7	.06721
0.3	.74082	2.8	.06081
0.4	.67032	2.9	.05502
0.5	.60653	3.0	.04979
0.6	.54881	3.2	.04076
0.7	.49659	3.4	.03337
0.8	.44933	3.6	.02732
0.9	.40657	3.8	.02237
1.0	.36788	4.0	.01832
1.1	.33287	4.2	.01500
1.2	.30119	4.4	.01228
1.3	.27253	4.6	.01005
1.4	.24660	4.8	.00823
1.5	.22313	5.0	.00674
1.6	.20190	5.5	.00409
1.7	.18268	6.0	.00248
1.8	.16530	6.5	.00150
1.9	.14957	7.0	.00091
2.0	.13534	7.5	.00055
2.1	.12246	8.0	.00034
2.2	.00180	8.5	.00020
2.3	.10026	9.0	.00012
2.4	.09072	10.0	.00005

Apéndice 4

Probabilidades Poisson *

λ

X	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0	.9048	.8187	.7408	.6703	.6065	.5488	.4966	.4493	.4066	.3679
1	.0905	.1637	.2222	.2681	.3033	.3293	.3476	.3595	.3659	.3679
2	.0045	.0164	.0333	.0536	.0758	.0988	.1217	1438	.1647	.1839
3	0002	.0011	.0033	.0072	.0126	.0198	.0284	.0383	.0494	.0613
4	.0000	.0001	.0002	.0007	.0016	.0030	.0050	.0077	.0111	.0153
5	0000	.0000	0000	0001	.0002	0004	.0007	0012	.0020	.0031
6	0000	.0000	.0000	.0000	0000	.0000	.0001	.0002	.0003	.0005
7	.0000	.0000	0000	0000	.0000	.0000	.0000	.0000	.0000	.0001

λ

X	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
0	3329	3012	2725	.2466	2231	2019	.1827	.1653	.1496	.1353
1	3662	3614	.3543	.3452	.3347	.3230	3106	.2975	.2842	.2707
2	2014	2169	2303	2417	.2510	2584	.2640	.2678	.2700	.2707
3	.0738	0867	.0998	1128	.1255	.1378	1496	.1607	.1710	.1804
4	.0203	0260	0324	.0395	0471	.0551	0636	.0723	.0812	.0902
5	0045	.0062	0084	.0111	0141	.0176	0216	.0260	.0309	.0361
6	0008	.0012	.0018	.0026	.0035	.0047	.0061	.0078	.0098	.0120
7	0001	0002	.0003	.0005	.0008	.0011	.0015	.0020	.0027	.0034
8	0000	.0000	0001	0001	.0001	.0002	.0003	.0005	.0006	.0009
9	0000	.0000	.0000	.0000	0000	.0000	0001	0001	.0001	.0002

λ

X	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	1225	.1108	.1003	.0907	.0821	.0743	.0672	.0608	.0550	.0498
1	2572	2438	2306	.2177	2052	.1931	.1815	1703	.1396	.1494
2	2700	.2681	2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	1890	1966	.2033	.2090	2138	.2176	.2205	.2225	.2237	.2240
4	0992	1082	.1169	.1254	.1336	.1414	1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	0804	.0872	.0940	.1008
6	.0146	0174	0206	.0241	.0278	0319	.0362	.0407	.0455	0504
7	.0044	0055	0068	0083	.0099	.0118	0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	0031	.0038	.0047	.0057	0068	.0081
9	0003	0004	.0005	0007	.0009	.0011	0014	0018	.0022	.0027
10	0001	0001	0001	.0002	0002	0003	0004	.0005	.0006	.0008
11	0000	0000	.0000	.0000	.0000	0001	.0001	.0001	.0002	.0002
12	.0000	.0000	.0000	.0000	0000	.0000	0000	.0000	0000	.0001

λ

X	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	1397	.1304	.1217	.1135	1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	2008	.1929	.1850	.1771	.1692	1615	1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	1734	.1781	.1823	.1858	1888	1912	1931	.1944	1951	.1954

*Ejemplo: $P(X = 5 \mid \lambda = 2.5) = 0.0668$

					λ					
X	3.?	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
5	.1075	.1140	.1203	.1264	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	0881	.0936	.0989	.1042
7	.0246	.0278	0312	.0348	.0385	.0425	.0466	0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	0215	.0241	.0269	.0298
9	.0033	.0040	.0047	.0056	.0066	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0028	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0002	.0002
14	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001

					λ					
X	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842
3	.1904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.1961	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1756
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1756
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0868	.0914	.0969	.1002	.1044
8	.0328	.0360	.0393	.0428	.0463	.0600	.0637	.0676	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0266	.0280	.0307	.0334	.0363
10	.0061	.0071	.0081	.0082	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0043	.0049	.0066	.0064	.0073	.0082
12	.0008	.0009	.0011	.0014	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0006	.0006	.0007	.0008	.0008	.0011	.0013
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0002

					λ					
X	5.1	5.2	5.3	5.4	5.5	5.5	5.7	5.8	5.9	6.0
0	.0061	.0055	.0050	.0045	.0041	.0037	.0033	.0030	.0027	.0025
1	.0311	.0287	.0265	.0244	.0225	.0207	.0191	0176	.0162	.0149
2	.0793	.0746	.0701	.0659	.0618	.0580	.0544	.0509	.0477	.0446
3	.1348	.1293	.1239	.1185	.1133	.1082	.1033	.0985	.0938	.0892
4	.1719	.1681	.1641	.1600	.1558	.1615	.1472	.1428	.1383	.1339
5	.1753	.1748	.1740	.1728	.1714	.1697	.1678	.1656	.1632	.1606
6	.1490	.1515	.1537	.1555	.1571	.1584	.1594	.1601	.1605	.1606
7	.1086	.1125	.1163	.1200	.1234	.1267	.1298	.1326	.1353	.1377
8	.0692	.0731	C771	.0810	.0849	.0887	.0925	.0962	.0998	.1033
9	.0392	.0423	.0454	.0486	.0519	.0552	.0586	.0620	.0654	.0688
10	.0200	.0220	.0241	.0262	.0285	.0309	.0334	.0359	.0386	.0413
11	.0093	.0104	.0116	.0129	.0143	.0157	.0173	.0190	.0207	.0225
12	.0039	.0045	.0051	.0058	.0065	.0073	.0032	.0092	.0102	.0113
13	.0015	.0018	.0021	.0024	.0028	.0032	.0036	.0041	.0046	.0052
14	.0006	.0007	.0008	.0009	.0011	.0013	.0015	.0017	.0019	.0022
15	.0002	.0002	.0003	.0003	.0004	.0005	.0006	.0007	.0008	.0009
16	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003
17	.0000	.0000	.0000	0000	0000	.0001	.0001	.0001	.0001	.0001

					λ					
X	*6.1*	*6.2*	*6.3*	*6.4*	*6.5*	*6.6*	*6.7*	*6.8*	*6.9*	*7.0*
0	0022	.0020	.0018	.0017	.0015	.0014	.0012	.0011	.0010	.0009
1	.0137	.0126	.0116	.0106	.0098	.0090	.0082	.0076	.0070	.0064
2	.0417	.0390	.0364	.0340	.0318	.0296	.0276	.0258	.0240	.0223
3	.0848	.0806	.0765	.0726	.0688	.0652	.0617	.0584	.0552	.0521
4	.1294	.1249	.1205	.1162	.1118	.1076	.1034	.0992	.0952	.0912
5	.1579	.1549	.1519	.1487	.1454	.1420	.1385	.1349	.1314	.1277
6	.1605	.1601	.1595	.1586	.1575	.1562	.1546	.1529	.1511	.1490
7	.1399	.1418	.1435	.1450	.1462	.1472	.1480	.1486	.1489	.1490
8	.1066	.1099	.1130	.1160	.1188	.1215	.1240	•1263	.1284	.1304
9	.0723	.0757	.0791	.0825	.0858	.0891	.0923	.0954	.0985	.1014
10	.0441	.0469	.0498	.0528	.0558	.0558	.0618	.0649	.0679	.0710
11	.0245	.0265	.0285	.0307	.0330	.0353	.0377	.0401	.0426	.0452
12	.0124	.0137	.0150	.0164	.0179	0194	.0210	.0227	.0245	.0264
13	.0058	.0065	.0073	.0081	.0089	.0098	.0108	.0119	.0130	.0142
14	.0025	.0029	.0033	.0037	.0041	.0046	.0052	.0058	.0064	.0071
15	.0010	.0012	.0014	.0016	.0018	.0020	.0023	.0026	.0029	.0033
16	.0004	.0005	.0005	.0006	.0007	.0008	.0010	.0011	.0013	.0014
17	.0001	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0006
18	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002
19	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

					λ					
X	*7.1*	*7.2*	*7.3*	*7.4*	*7.5*	*7.6*	*7.7*	*7.8*	*7.9*	*8.0*
0	.0008	.0007	.0007	.0006	.0006	.0005	.0005	.0004	.0004	.0003
1	.0059	.0054	.0049	.0045	.0041	.0038	.0035	.0032	.0029	.0027
2	.0208	.0194	0180	.0167	.0156	0145	.0134	.0125	.0116	.0107
3	.0492	.0464	.0438	.0413	.0389	.0366	.0345	.0324	.0305	.0286
4	.0874	.0836	.0799	.0764	.0729	.0696	.0663	.0632	.0602	.0573
5	.1241	.1204	.1167	.1130	.1094	.1057	.1021	.0986	.0951	.0916
6	.1468	.1445	.1420	.1394	.1367	.1339	.1311	.1282	.1252	.1221
7	.1489	.1486	.1481	.1474	.1465	.1454	.1442	.1428	.1413	.1396
8	.1321	.1337	.1351	.1363	.1373	.1382	.1388	.1392	.1395	.1396
9	.1042	.1070	.1096	.1121	.1144	.1167	.1187	.1207	.1224	.1241
10	.0740	.0770	.0800	.0829	.0858	.0887	.0914	.0941	.0967	.0993
11	.0478	.0504	.0531	.0558	.0585	.0613	.0640	.0667	.0695	.0722
12	.0283	.0303	.0323	.0344	.0366	.0388	.0411	.0434	.0457	.0481
13	.0154	.0168	.0181	.0196	.0211	.0227	.0243	.0260	.0278	.0296
14	.0078	.0086	.0095	.0104	.0113	.0123	.0134	.0145	.0157	.0169
15	.0037	.0041	.0046	.0051	.0057	.0062	0069	.0075	.0083	.0090
16	.0016	.0019	.0021	.0024	.0026	.0030	.0033	.0037	.0041	.0045
17	.0007	.0008	.0009	.0010	.0012	.0013	.0015	.0017	.0019	.0021
18	.0003	.0003	.0004	.0004	.0005	.0006	.0006	0007	.0008	.0009
19	.0001	.0001	.0001	.0002	.0002	.0002	.0003	.0003	.0003	.0004
20	.0000	.0000	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002
21	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	0001	0001

					λ					
X	*8.1*	*8.2*	*8.3*	*8.4*	*8.5*	*8.6*	*8.7*	*8.8*	*8.9*	*9.0*
0	.0003	.0003	.0002	.0002	0002	.0002	.0002	.0002	.0001	.0001
1	.0025	.0023	.0021	.0019	.0017	.0016	.0014	0013	.0012	.0011

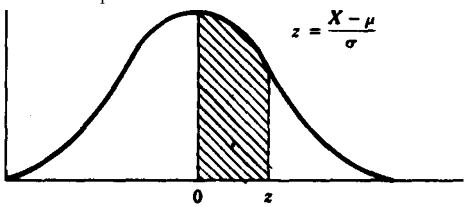
					λ					
X	***8.1***	***8.2***	***8.3***	***8.4***	***8.5***	***8.6***	***8.7***	***8.8***	***8.9***	***9.0***
2	.0100	.0092	.0086	.0079	.0074	.0068	.0063	.0058	0054	OObO
3	.0269	.0252	.0237	.0222	.0208	.0195	.0183	.0171	.0160	.0150
4	.0544	.0517	.0491	.0466	.0443	.0420	.0398	0377	0357	0337
5	.0882	.0849	.0816	.0784	.0752	.0722	.0692	.0663	.0635	.0607
6	.1191	.1160	.1128	.1097	.1066	.1034	.1003	.0972	.0941	.0911
7	.1378	.1358	.1338	.1317	.1294	.1271	.1247	1222	.1197	.1171
8	.1395	.1392	.1388	.1382	.1375	.1366	.1356	.1344	.1332	.1318
9	.1256	.1269	.1280	.1290	.1299	.1306	.1311	.1315	.1317	.1318
10	.1017	.1040	.1063	.1084	.1104	.1123	.1140	.1157	.1172	.1186
11	.0749	.0776	.0802	.0828	.0853	.0878	.0902	.0925	.0948	.0970
12	.0505	.0530	.0555	.0579	.0604	.0629	.0654	.0679	.0703	.0728
13	.0315	.0334	.0354	.0374	.0395	.0416	.0438	.0459	.0481	.0504
14	.0182	.0196	.0210	.0225	.0240	.0256	.0272	.0289	.0306	.0324
15	.0098	.0107	.0116	.0126	.0136	.0147	.0158	.0169	.0182	0194
16	.0050	.0055	.0060	.0066	.0072	.0079	.0086	.0093	.0101	.0109
17	.0024	.0026	.0029	.0033	.0036	.0040	.0044	.0048	.0053	.0058
18	.0011	.0012	.0014	.0015	.0017	.0019	.0021	.0024	.0026	.0029
19	.0005	.0005	.0006	.0007	.0008	.0009	.0010	.0011	.0012	.0014
20	.0002	.0002	.0002	.0003	.0003	.0004	.0004	.0005	.0005	.0006
21	.0001	.0001	.0001	.0001	.0001	.0002	.0002	.0002	.0002	.0003
22	.0000	.0000	.0000	.0000	.0001	.0001	.0001	.0001	.0001	0001

					λ					
X	***9.1***	***9.2***	***9.3***	***9.4***	**as**	***9.6***	***9.7***	***9.8***	***9.9***	***10.0***
0	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0000
1	.0010	.0009	.0009	.0008	.0007	.0007	.0006	.0005	.0005	.0005
2	.0046	.0043	.0040	.0037	.0034	.0031	.0029	.0027	.0025	.0023
3	.0140	.0131	.0123	.0115	.0107	.0100	.0093	.0087	.0081	.0076
4	.0319	.0302	.0285	.0269	.0254	.0240	.0226	.0213	.0201	.0189
5	.0581	.0555	.0530	.0506	.0483	.0460	.0439	.0418	.0398	.0378
6	.0881	.0851	.0822	.0793	0764	.0736	.0709	.0682	.0656	.0631
7	.1145	.1118	.1091	.1064	.1037	.1010	.0982	.0955	.0928	.0901
8	.1302	.1286	.1269	.1251	.1232	.1212	.1191	.1170	.1148	.1126
9	.1317	1315	.1311	.1306	.1300	.1293	.1284	.1274	.1263.	.1251
10	.1198	.1210	.1219	.1228	.1235	.1241	.1245	.1249	.1250	.1251
11	.0991	.1012	.1031	.1049	.1067	.1083	.1098	.1112	.1125	.1137
12	0752	.0776	.0779	.0822	.0844	.0866	.0888	.0908	.0928	.0948
13	.0526	.0549	.0572	.0594	.0617	.0640	.0662	.0685	.0707	.0729
14	.0342	.0361	.0380	.0399	.0419	0439	.0459	.0479	.0500	.0521
15	.0208	.0221	.0235	.0250	.0265	.0281	.0297	.0313	.0330	.0347
16	.0118	.0127	.0137	.0147	.0157	.0168	.0180	.0192	.0204	.0217
17	.0063	.0069	.0075	0081	.0088	.0095	.0103	.0111	.0119	.0128
18	.0032	.0035	.0039	.0042	.0046	.0051	.0055	.0060	.0065	.0071
19	.0015	.0017	.0019	.0021	.0023	0026	.0028	.0031	.0034	,0037
20	.0007	.0008	.0009	.0010	.0011	0012	.0014	.0015	.0017	.0019
21	.0003	.0003	.0004	.0004	0005	.0006	.0006	:0007	.0008	.0009
22	.0001	.0001	.0002	.0002	.0002	.0002	.0003	.0003	0004	.0004
23	.0000	.0001	.0001	0001	.0001	.0001	.0001	.0001	.0002	.0002
24	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0001	.0001	.0001

Apéndice 5

Propociones de área para la distribución normal estándar

El área reportada en la tabla:*

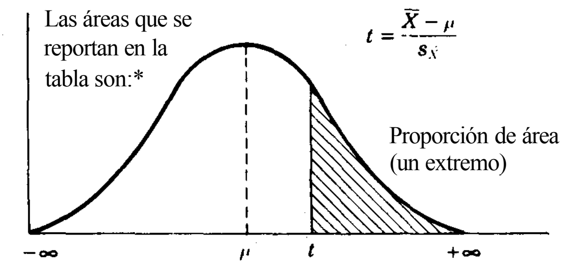


2	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0567	.0696	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	2.612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3999	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4014
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	4625	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4691	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	4911	4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	4964
2.7	.4965	.4966	.4967	.4968	4969	.4970	4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4983	.4983	.4984	.4984	.4985	.4985	4986	.4986
3.0	.4987									
3.5	.4997									
4.0	.4999									

*Ejemplo: Para z = 1.96, el área sombreada es 0.4750 del área total de 1.0000.

Apéndice 6

Proporciones de área para la distribución *t*



df	*0.10*	*0.05-*	*0.025*	*0.01*	*0.005*
1	3.078	6.314	12.7G6	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1 397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.92,1
17	1.333	1.740	2.110	2.567	2.898

df	*0.10*	*0.05*	*0.025*	*0.01*	*0.005*
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

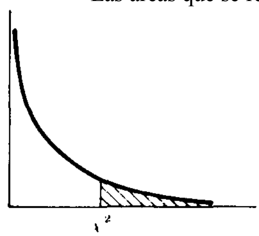
*Ejemplo: Para que el área sombreada represente el 0.05 del área total de 1.0 el valor de t con 10 grados de libertad es 1.812.

Fuente: De la tabla III de Fisher y Yates. *Stastistical Tables For Biological, Agricultural and Medical Research,* 6a. ed... 1974, publicada por Longman Group Ltd., London (Publicada previamente por Oliver & Boyd. Edinburgh). Con autorización de los autores y editores.

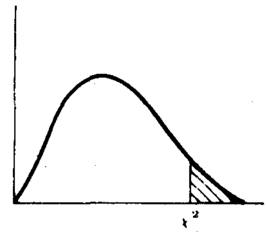
Apéndice 7

Proporciones de área para la distribución x^2

Las áreas que se reportan en la tabla son: *



Para gl - 1 . 2



Para *gl* ≤ 30

gl	Proporción del área										
	0 995	0990	0.975	0.950	0.900	0.500	0.100	0.050	0.025	0.010	0.005
1	0.00004	0.00016	0.00098	0.00393	0.0158	0.455	2.71	3.84	5.02	6 63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	1.386	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	2.366	6.25	7.81	9.35	11.34	1284
4	0.207	0.297	0.484	0.711	1.064	3.357	7.78	9.49	11.14	13.28	14 86
5	0.412	0.554	0.831	1.145	1.61	4.251	9 24	11.07	12.83	15.09	16.75
6	0.676	0.872	124	1.64	2.20	535	1064	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	635	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	273	3 49	7 34	13.36	1551	17.53	20 09	21.96
9	1.73	2 09	270	3.33	4.17	8.34	1468	1692	1902	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	934	15.99	18.31	20.48	23.21	25 19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	630	11.34	18.55	21.03	23.34	26 22	28.30
13	3.57	4.11	5.01	5 89	7 04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4 66	5.63	6.57	7.79	1334	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8 55	14.34	22 31	25.00	27.49	30.58	32.80
16	5.14	5.81	691	7.96	931	15.34	2354	26.30	28 85	32.00	34.27
17	5.70	641	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6 26	7.01	8.23	9.39	1086	17.34	25.99	28.87	31.53	34.81	37.16
19	6,84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38 58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	3141	34.17	37.57	40.00
21	8.03	890	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	1234	14.04	21.34	30.81	33 92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	3201	35 17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15 66	2334	3320	36.42	39 36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34 38	37 65	40 65	44.31	46 93
26	11 16	1220	13 84	1538	17.29	25 34	35 56	38.89	41.92	45 64	48.29
27	11.81	12.83	14.57	16.15	18 11	26.34	36.74	40.11	43.19	46.96	49.64
28	1246	13 56	15.31	16.93	1894	27.34	37 92	41.34	44.46	48.28	50.99
29	13 12	1426	16.05	17.71	19.77	28.34	3909	4256	45.72	49 59	52 34
30	1379	14 95	1679	18.49	20.60	29.34	40.26	4377	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39 34	51.81	5576	59.34	63 69	66.77
50	27.99	29.71	32.36	3476	37 69	49.33	63 17	67 50	71.42	76.15	7949
6(i	35.53	37 43	40.48	43.19	46 46	59 33	74.40	79.08	83.30	88.38	91.95
70	43 28	45 44	48.76	51 74	55 33	69 33	85 53	90 53	95.02	1004	104.2
80	51.17	5354	51.17	60 39	64 28	79 33	98.58	101.9	106.6	112.3	116 3
90	59 20	61.75	65.65	69.13	7329	89 33	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74 22	77.93	82.36	99.33	118 5	124.3	129.6	135.8	140 2

*Ejemplo: para que el área sombreada represente el 0.05 del área total de 1.0 bajo la función de densidad, el valor de x^2 es 18.31, cuando gl - 10.

Fuente: De la tabla IV de Fisher y Yates, *Stalistical Tables For Biological, Agricultural and Medical Research,* 6a. cd., 1974, publicada por Lor gman Group Ltd., London (Publicada previamente por Oliver & Boyd. Edinburgli). Con autorización de los autores *y* editores.

Apéndice 8

Valores de F excedidos con probabilidades

gl (denominador)	gl (numerador) 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254
	4,052	**4,999**	**5,403**	**5,625**	**5,764**	**5,859**	**5,928**	**5,981**	**6,022**	**6,056**	**6,082**	**6,106**	**6,142**	**6,169**	**6,208**	**6,234**	**6,261**	**6,286**	**6,302**	**6,323**	**6,334**	**6,352**	**6,361**	**6,366**
2	18.51	19.00	19.16	19.25	19.30	19.33	19.36	19.37	19.38	19.39	19.40	19.41	19.42	19.43	19.44	19.45	19.46	19.47	19.47	19.48	19.49	19.49	19.50	19.50
	98.49	**99.00**	**99.17**	**99.25**	**99.30**	**99.33**	**99.36**	**99.37**	**99.39**	**99.40**	**99.41**	**99.42**	**99.43**	**99.44**	**99.45**	**99.46**	**99.47**	**99.48**	**99.48**	**99.49**	**99.49**	**99.49**	**99.50**	**99.50**
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53
	34.12	**30.82**	**29.46**	**28.71**	**28.24**	**27.91**	**27.67**	**27.49**	**27.34**	**27.23**	**27.13**	**27.05**	**26.92**	**26.83**	**26.69**	**26.60**	**26.50**	**26.41**	**26.35**	**26.27**	**26.23**	**26.18**	**26.14**	**26.12**
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63
	21.20	**18.00**	**16.69**	**15.98**	**15.52**	**15.21**	**14.98**	**14.80**	**14.66**	**14.54**	**14.45**	**14.37**	**14.24**	**14.15**	**14.02**	**13.93**	**13.83**	**13.74**	**13.69**	**13.61**	**13.57**	**13.52**	**13.48**	**13.46**
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36
	16.26	**13.27**	**12.06**	**11.39**	**10.97**	**10.67**	**10.45**	**10.29**	**10.15**	**10.05**	**9.96**	**9.89**	**9.77**	**9.68**	**9.55**	**9.47**	**9.38**	**9.29**	**9.24**	**9.17**	**9.13**	**9.07**	**9.04**	**9.02**
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
	13.74	**10.92**	**9.78**	**9.15**	**8.75**	**8.47**	**8.26**	**8.10**	**7.98**	**7.87**	**7.79**	**7.72**	**7.60**	**7.52**	**7.39**	**7.31**	**7.23**	**7.14**	**7.09**	**7.02**	**6.99**	**6.94**	**6.90**	**6.88**
7	5.59	4.74	4.34	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
	12.25	**9.55**	**8.45**	**7.85**	**7.46**	**7.19**	**7.00**	**6.84**	**6.71**	**6.62**	**6.54**	**6.47**	**6.35**	**6.27**	**6.15**	**6.07**	**5.98**	**5.90**	**5.85**	**5.78**	**5.75**	**5.70**	**5.67**	**5.65**
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.94	2.93
	11.26	**8.65**	**7.59**	**7.01**	**6.63**	**6.37**	**6.19**	**6.03**	**5.91**	**5.82**	**5.74**	**5.67**	**5.56**	**5.48**	**5.36**	**5.28**	**5.20**	**5.11**	**5.06**	**5.00**	**4.96**	**4.91**	**4.88**	**4.86**
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71
	10.56	**8.02**	**6.99**	**6.42**	**6.06**	**5.80**	**5.62**	**5.47**	**5.35**	**5.26**	**5.18**	**5.11**	**5.00**	**4.92**	**4.80**	**4.73**	**4.64**	**4.56**	**4.51**	**4.45**	**4.41**	**4.36**	**4.33**	**4.31**
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54
	10.04	**7.56**	**6.55**	**5.99**	**5.64**	**5.39**	**5.21**	**5.06**	**4.95**	**4.85**	**4.78**	**4.71**	**4.60**	**4.52**	**4.41**	**4.33**	**4.25**	**4.17**	**4.12**	**4.05**	**4.01**	**3.96**	**3.93**	**3.91**
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40
	9.65	**7.20**	**6.22**	**5.67**	**5.32**	**5.07**	**4.88**	**4.74**	**4.63**	**4.54**	**4.46**	**4.40**	**4.29**	**4.21**	**4.10**	**4.02**	**3.94**	**3.86**	**3.80**	**3.74**	**3.70**	**3.66**	**3.62**	**3.60**
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30
	9.33	**6.93**	**5.95**	**5.41**	**5.06**	**4.82**	**4.65**	**4.50**	**4.39**	**4.30**	**4.22**	**4.16**	**4.05**	**3.98**	**3.86**	**3.78**	**3.70**	**3.61**	**3.56**	**3.49**	**3.46**	**3.41**	**3.38**	**3.36**
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21
	9.07	**6.70**	**5.74**	**5.20**	**4.86**	**4.62**	**4.44**	**4.30**	**4.19**	**4.10**	**4.02**	**3.96**	**3.85**	**3.78**	**3.67**	**3.59**	**3.51**	**3.42**	**3.37**	**3.30**	**3.27**	**3.21**	**3.18**	**3.16**
14	4.60	3.74	3.34	3.11	2.96	2.85	2.77	2.70	2.65	2.60	2.56	2.53	2.48	2.44	2.39	2.35	2.31	2.27	2.24	2.21	2.19	2.16	2.14	2.13
	8.86	**6.51**	**5.56**	**5.03**	**4.69**	**4.46**	**4.28**	**4.14**	**4.03**	**3.94**	**3.86**	**3.80**	**3.70**	**3.62**	**3.51**	**3.43**	**3.34**	**3.26**	**3.21**	**3.14**	**3.11**	**3.06**	**3.02**	**3.00**
15	4.54	3.68	3.29	3.06	2.90	2.79	2.70	2.64	2.59	2.55	2.51	2.48	2.43	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.12	2.10	2.08	2.07
	8.68	**6.36**	**5.42**	**4.89**	**4.56**	**4.32**	**4.14**	**4.00**	**3.89**	**3.80**	**3.73**	**3.67**	**3.56**	**3.48**	**3.36**	**3.29**	**3.20**	**3.12**	**3.07**	**3.00**	**2.97**	**2.92**	**2.89**	**2.87**
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45	2.42	2.37	2.33	2.28	2.24	2.20	2.16	2.13	2.09	2.07	2.04	2.02	2.01
	8.53	**6.23**	**5.29**	**4.77**	**4.44**	**4.20**	**4.03**	**3.89**	**3.78**	**3.69**	**3.61**	**3.55**	**3.45**	**3.37**	**3.25**	**3.18**	**3.10**	**3.01**	**2.96**	**2.98**	**2.86**	**2.80**	**2.77**	**2.75**
17	4.45	3.59	3.20	2.96	2.81	2.70	2.62	2.55	2.50	2.45	2.41	2.38	2.33	2.29	2.23	2.19	2.15	2.11	2.08	2.04	2.02	1.99	1.97	1.96
	8.40	**6.11**	**5.18**	**4.67**	**4.34**	**4.10**	**3.93**	**3.79**	**3.68**	**3.59**	**3.52**	**3.45**	**3.35**	**3.27**	**3.16**	**3.08**	**3.00**	**2.92**	**2.86**	**2.79**	**2.76**	**2.70**	**2.67**	**2.65**
18	4.41	3.55	3.16	2.93	2.77	3.66	2.58	2.51	2.46	2.41	2.37	2.34	2.29	2.25	2.19	2.15	2.11	2.07	2.04	2.00	1.98	1.95	1.93	1.92
	8.28	**6.01**	**5.09**	**4.58**	**4.25**	**4.01**	**3.85**	**3.71**	**3.60**	**3.51**	**3.44**	**3.37**	**3.27**	**3.19**	**3.07**	**3.00**	**2.91**	**2.83**	**2.78**	**2.71**	**2.68**	**2.62**	**2.59**	**2.57**

de cinco y uno por ciento.

19	4.38	3.52	3.13	2.90	2.74	2.63	2.55	2.48	2.43	2.38	2.34	2.31	2.26	2.21	2.15	2.11	2.07	2.02	2.00	1.96	1.94	1.91	1.90	1.88
	8.18	**5.93**	**5.01**	**4.50**	**4.17**	**3.94**	**3.77**	**3.63**	**3.52**	**3.43**	**3.36**	**3.30**	**3.19**	**3.12**	**3.00**	**2.92**	**2.84**	**2.76**	**2.70**	**2.63**	**2.60**	**2.54**	**2.51**	**2.49**
20	4.35	3.49	3.10	2.87	2.71	2.60	2.52	2.45	2.40	2.35	2.31	2.28	2.23	2.18	2.12	2.08	2.04	1.99	1.96	1.92	1.90	1.87	1.85	1.84
	8.10	**5.85**	**4.94**	**4.43**	**4.10**	**3.87**	**3.71**	**3.56**	**3.45**	**3.37**	**3.30**	**3.23**	**3.13**	**3.05**	**2.94**	**2.86**	**2.77**	**2.69**	**2.63**	**2.56**	**2.53**	**2.47**	**2.44**	**2.42**
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.20	2.15	2.09	2.05	2.00	1.96	1.93	1.89	1.87	1.84	1.82	1.81
	8.02	**5.78**	**4.87**	**4.37**	**4.04**	**3.81**	**3.65**	**3.51**	**3.40**	**3.31**	**3.24**	**3.17**	**3.07**	**2.99**	**2.88**	**2.80**	**2.72**	**2.63**	**2.58**	**2.51**	**2.47**	**2.42**	**2.38**	**2.36**
22	4.30	3.44	3.05	2.82	2.66	2.55	2.47	2.40	2.35	2.30	2.26	2.23	2.18	2.13	2.07	2.03	1.98	1.93	1.91	1.87	1.84	1.81	1.80	1.78
	7.94	**5.72**	**4.82**	**4.31**	**3.99**	**3.76**	**3.59**	**3.45**	**3.35**	**3.26**	**3.18**	**3.12**	**3.02**	**2.94**	**2.83**	**2.75**	**2.67**	**2.58**	**2.53**	**2.46**	**2.42**	**2.37**	**2.33**	**2.31**
23	4.28	3.42	3.03	2.80	2.64	2.53	2.45	2.38	2.32	2.28	2.24	2.20	2.14	2.10	2.04	2.00	1.96	1.91	1.88	1.84	1.82	1.79	1.77	1.76
	7.88	**5.66**	**4.76**	**4.26**	**3.94**	**3.71**	**3.54**	**3.41**	**3.30**	**3.21**	**3.14**	**3.07**	**2.97**	**2.89**	**2.78**	**2.70**	**2.62**	**2.53**	**2.48**	**2.41**	**2.37**	**2.32**	**2.28**	**2.26**
24	4.26	3.40	3.01	2.78	2.62	2.51	2.43	2.36	2.30	2.26	2.22	2.18	2.13	2.09	2.02	1.98	1.94	1.89	1.86	1.82	1.80	1.76	1.74	1.73
	7.82	**5.61**	**4.72**	**4.22**	**3.90**	**3.67**	**3.50**	**3.36**	**3.25**	**3.17**	**3.09**	**3.03**	**2.93**	**2.85**	**2.74**	**2.66**	**2.58**	**2.49**	**2.44**	**2.36**	**2.33**	**2.27**	**2.23**	**2.21**
25	4.24	3.38	2.99	2.76	2.60	2.49	2.41	2.34	2.28	2.24	2.20	2.16	2.11	2.06	2.00	1.96	1.92	1.87	1.84	1.80	1.77	1.74	1.72	1.71
	7.77	**5.57**	**4.68**	**4.18**	**3.86**	**3.63**	**3.46**	**3.32**	**3.21**	**3.13**	**3.05**	**2.99**	**2.89**	**2.81**	**2.70**	**2.62**	**2.54**	**2.45**	**2.40**	**2.32**	**2.29**	**2.23**	**2.19**	**2.17**
26	4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.10	2.05	1.99	1.95	1.90	1.85	1.82	1.78	1.76	1.72	1.70	1.69
	7.72	**5.53**	**4.64**	**4.14**	**3.82**	**3.59**	**3.42**	**3.29**	**3.17**	**3.09**	**3.02**	**2.96**	**2.86**	**2.77**	**2.66**	**2.58**	**2.50**	**2.41**	**2.36**	**2.28**	**2.25**	**2.19**	**2.15**	**2.13**
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.14**	**3.06**	**2.98**	**2.93**	**2.83**	**2.74**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.25**	**2.21**	**2.16**	**2.12**	**2.10**
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.11**	**3.03**	**2.95**	**2.90**	**2.80**	**2.71**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.22**	**2.18**	**2.13**	**2.09**	**2.06**
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**	**2.92**	**2.87**	**2.77**	**2.68**	**2.57**	**2.49**	**2.41**	**2.32**	**2.27**	**2.19**	**2.15**	**2.10**	**2.06**	**2.03**
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**	**2.90**	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.38**	**2.29**	**2.24**	**2.16**	**2.13**	**2.07**	**2.03**	**2.01**
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**3.01**	**2.94**	**2.86**	**2.80**	**2.70**	**2.62**	**2.51**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**	**2.08**	**2.02**	**1.98**	**1.96**
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.97**	**2.89**	**2.82**	**2.76**	**2.66**	**2.58**	**2.47**	**2.38**	**2.30**	**2.21**	**2.15**	**2.08**	**2.04**	**1.98**	**1.94**	**1.91**
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55
	7.39	**5.25**	**4.38**	**3.89**	**3.58**	**3.35**	**3.18**	**3.04**	**2.94**	**2.86**	**2.78**	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.17**	**2.12**	**2.04**	**2.00**	**1.94**	**1.90**	**1.87**
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**	**2.75**	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.22**	**2.14**	**2.08**	**2.00**	**1.97**	**1.90**	**1.86**	**1.84**
40	4.07	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**	**2.73**	**2.66**	**2.56**	**2.49**	**2.37**	**2.29**	**2.20**	**2.11**	**2.05**	**1.97**	**1.94**	**1.88**	**1.84**	**1.81**

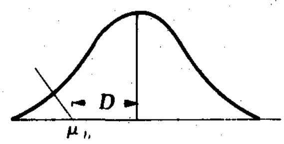
(*continúa*)

gl (denominador)	*gl* (numerador) 1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49
	7.27	5.15	4.29	3.80	3.49	3.26	3.10	2.96	2.86	2.77	2.70	2.64	2.54	2.46	2.35	2.26	2.17	2.08	2.02	1.94	1.91	1.85	1.80	1.78
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48
	7.24	5.12	4.26	3.78	3.46	3.24	3.07	2.94	2.84	2.75	2.68	2.62	2.52	2.44	2.32	2.24	2.15	2.06	2.00	1.92	1.88	1.82	1.78	1.75
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46
	7.21	5.10	4.24	3.76	3.44	3.22	3.05	2.92	2.82	2.73	2.66	2.60	2.50	2.42	2.30	2.22	2.13	2.04	1.98	1.90	1.86	1.80	1.76	1.72
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45
	7.19	5.08	4.22	3.74	3.42	3.20	3.04	2.90	2.80	2.71	2.64	2.58	2.48	2.40	2.28	2.20	2.11	2.02	1.96	1.88	1.84	1.78	1.73	1.70
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44
	7.17	5.06	4.20	3.72	3.41	3.18	3.02	2.88	2.78	2.70	2.62	2.56	2.46	2.39	2.26	2.18	2.10	2.00	1.94	1.86	1.82	1.76	1.71	1.68
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41
	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75	2.66	2.59	2.53	2.43	2.35	2.23	2.15	2.06	1.96	1.90	1.82	1.78	1.71	1.66	1.64
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39
	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.40	2.32	2.20	2.12	2.03	1.93	1.87	1.79	1.74	1.68	1.63	1.60
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37
	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.79	2.70	2.61	2.54	2.47	2.37	2.30	2.18	2.09	2.00	1.90	1.84	1.76	1.71	1.64	1.60	1.56
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35
	7.01	4.92	4.08	3.60	3.29	3.07	2.91	2.77	2.67	2.59	2.51	2.45	2.35	2.28	2.15	2.07	1.98	1.88	1.82	1.74	1.69	1.62	1.56	1.53
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32
	6.96	4.88	4.04	3.56	3.25	3.04	2.87	2.74	2.64	2.55	2.48	2.41	2.32	2.24	2.11	2.03	1.94	1.84	1.78	1.70	1.65	1.57	1.52	1.49
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28
	6.90	4.82	3.98	3.51	3.20	2.99	2.82	2.69	2.59	2.51	2.43	2.36	2.26	2.19	2.06	1.98	1.89	1.79	1.73	1.64	1.59	1.51	1.46	1.43
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25
	6.84	4.78	3.94	3.47	3.17	2.95	2.79	2.65	2.56	2.47	2.40	2.33	2.23	2.15	2.03	1.94	1.85	1.75	1.68	1.59	1.54	1.46	1.40	1.37
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22
	6.81	4.75	3.91	3.44	3.14	2.92	2.76	2.62	2.53	2.44	2.37	2.30	2.20	2.12	2.00	1.91	1.83	1.72	1.66	1.56	1.51	1.43	1.37	1.33
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19
	6.76	4.71	3.88	3.41	3.11	2.90	2.73	2.60	2.50	2.41	2.34	2.28	2.17	2.09	1.97	1.88	1.79	1.69	1.62	1.53	1.48	1.39	1.33	1.28
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.91	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13
	6.70	4.66	3.83	3.36	3.06	2.85	2.69	2.55	2.46	2.37	2.29	2.23	2.12	2.04	1.92	1.84	1.74	1.64	1.57	1.47	1.42	1.32	1.24	1.19
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08
	6.66	4.62	3.80	3.34	3.04	2.82	2.66	2.53	2.43	2.34	2.26	2.20	2.09	2.01	1.89	1.81	1.71	1.61	1.54	1.44	1.38	1.28	1.19	1.11
	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00
	6.64	4.60	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.24	2.18	2.07	1.99	1.87	1.79	1.69	1.59	1.52	1.41	1.36	1.25	1.15	1.00

Fuente: reimpreso con autorización de *Statistical Methods*, 6a. ed., por George W. Snedecor y William G. Cochran, © 1967, de la Iowa State University Press, Ames, Iowa.

Apéndice 9

Función de pérdida normal unitaria



D	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.3989	.3940	.3890	.3841	.3793	.3744	.3697	.3649	.3602	.3556
.1	.3509	.3464	.3418	.3373	.3328	.3284	.3240	.3197	.3154	.3111
.2	.3069	.3027	.2986	.2944	.2904	.2863	.2824	.2784	.2745	.2706
.3	.2668	.2630	.2592	.2555	.2518	.2481	.2445	.2409	.2374	.2339
.4	.2304	.2270	.2236	.2203	.2169	.2137	.2104	.2072	.2040	.2009
.5	.1978	.1947	.1917	.1887	.1857	.1828	.1799	.1771	.1742	.1714
.6	.1687	.1659	.1633	.1606	.1580	.1554	.1528	.1503	.1478	.1453
.7	.1429	.1405	.1381	.1358	.1334	.1312	.1289	.1267	.1245	.1223
.8	.1202	.1181	.1160	.1140	.1120	.1100	.1080	.1061	.1042	.1023
.9	.1004	.09860	.09680	.09503	.09328	.09156	.08986	.08819	.08654	.08491
1.0	.08332	.08174	.08019	.07866	.07716	.07568	.07422	.07279	.07138	.06999
1.1	.06862	.06727	.06595	.06465	.06336	.06210	.06086	.05964	.05844	.05726
1.2	.05610	.05496	.05384	.05274	.05165	.05059	.04954	.04851	.04750	.04650
1.3	.04553	.04457	.04363	.04270	.04179	.04090	.04002	.03916	.03831	.03748
1.4	.03667	.03587	.03508	.03431	.03356	.03281	.03208	.03137	.03067	.02998
1.5	.02931	.02865	.02800	.02736	.02674	.02612	.02552	.02494	.02436	.02380
1.6	.02324	.02270	.02217	.02165	.02114	.02064	.02015	.01967	.01920	.01874
1.7	.01829	.01785	.01742	.01699	.01658	.01617	.01578	.01539	.01501	.01464
1.8	.01428	.01392	.01357	.01323	.01290	.01257	.01226	.01195	.01164	.01134
1.9	.01105	.01077	.01049	.01022	$.0^{2}9957$ *	$.0^{2}9698$	$.0^{2}9445$	$.0^{2}9198$	$.0^{2}8957$	$.0^{2}8721$
2.0	$.0^{2}8491$	$.0^{2}8266$	$.0^{2}8046$	$.0^{2}7832$	$.0^{2}7623$	$.0^{2}7418$	$.0^{2}7219$	$.0^{2}7024$	$.0^{2}6835$	$.0^{2}6649$
2.1	$.0^{2}6468$	$.0^{2}6292$	$.0^{2}6120$	$.0^{2}5952$	$.0^{2}5788$	$.0^{2}5628$	$.0^{2}5472$	$.0^{2}5320$	$.0^{2}5172$	$.0^{2}5028$
2.2	$.0^{2}4887$	$.0^{2}4750$	$.0^{2}4616$	$.0^{2}4486$	$.0^{2}4358$	$.0^{2}4235$	$.0^{2}4114$	$.0^{2}3996$	$.0^{2}3882$	$.0^{2}3770$
2.3	$.0^{2}3662$	$.0^{2}3556$	$.0^{2}3453$	$.0^{2}3352$	$.0^{2}3255$	$.0^{2}3159$	$.0^{2}3067$	$.0^{2}2977$	$.0^{2}2889$	$.0^{2}2804$
2.4	$.0^{2}2720$	$.0^{2}2640$	$.0^{2}2561$	$.0^{2}2484$	$.0^{2}2410$	$.0^{2}2337$	$.0^{2}2267$	$.0^{2}2199$	$.0^{2}2132$	$.0^{2}2067$
2.5	$.0^{2}2004$	$.0^{2}1943$	$.0^{2}1883$	$.0^{2}1826$	$.0^{2}1769$	$.0^{2}1715$	$.0^{2}1662$	$.0^{2}1610$	$.0^{2}1560$	$.0^{2}1511$
2.6	$.0^{2}1464$	$.0^{2}1418$	$.0^{2}1373$	$.0^{2}1330$	$.0^{2}1288$	$.0^{2}1247$	$.0^{2}1207$	$.0^{2}1169$	$.0^{2}1132$	$.0^{2}1095$
2.7	$.0^{2}1060$	$.0^{2}1026$	$.0^{3}9928$	$.0^{3}9607$	$.0^{3}9295$	$.0^{3}8992$	$.0^{3}8699$	$.0^{3}8414$	$.0^{3}8138$	$.0^{3}7870$
2.8	$.0^{3}7611$	$.0^{3}7359$	$.0^{3}7115$	$.0^{3}6879$	$.0^{3}6650$	$.0^{3}6428$	$.0^{3}6213$	$.0^{3}6004$	$.0^{3}5802$	$.0^{3}5606$
2.9	$.0^{3}5417$	$.0^{3}5233$	$.0^{3}5055$	$.0^{3}4883$	$.0^{3}4716$	$.0^{3}4555$	$.0^{3}4398$	$.0^{3}4247$	$.0^{3}4101$	$.0^{3}3959$
3.0	$.0^{3}3822$	$.0^{3}3689$	$.0^{3}3560$	$.0^{3}3436$	$.0^{3}3316$	$.0^{3}3199$	$.0^{3}3087$	$.0^{3}2978$	$.0^{3}2873$	$.0^{3}2771$
3.1	$.0^{3}2673$	$.0^{3}2577$	$.0^{3}2485$	$.0^{3}2396$	$.0^{3}2311$	$.0^{3}2227$	$.0^{3}2147$	$.0^{3}2070$	$.0^{3}1995$	$.0^{3}1922$
3.2	$.0^{3}1852$	$.0^{3}1785$	$.0^{3}1720$	$.0^{3}1657$	$.0^{3}1596$	$.0^{3}1537$	$.0^{3}1480$	$.0^{3}1426$	$.0^{3}1373$	$.0^{3}1322$
3.3	$.0^{3}1273$	$.0^{3}1225$	$.0^{3}1179$	$.0^{3}1135$	$.0^{3}1093$	$.0^{3}1051$	$.0^{3}1012$	$.0^{4}9734$	$.0^{4}9365$	$.0^{4}9009$
3.4	$.0^{4}8666$	$.0^{4}8335$	$.0^{4}8016$	$.0^{4}7709$	$.0^{4}7413$	$.0^{4}7127$	$.0^{4}6852$	$.0^{4}6587$	$.0^{4}6331$	$.0^{4}6085$
3.5	$.0^{4}5848$	$.0^{4}5620$	$.0^{4}5400$	$.0^{4}5188$	$.0^{4}4984$	$.0^{4}4788$	$.0^{4}4599$	$.0^{4}4417$	$.0^{4}4242$	$.0^{4}4073$
3.6	$.0^{4}3911$	$.0^{4}3755$	$.0^{4}3605$	$.0^{4}3460$	$.0^{4}3321$	$.0^{4}3188$	$.0^{4}3059$	$.0^{4}2935$	$.0^{4}2816$	$.0^{4}2702$
3.7	$.0^{4}2592$	$.0^{4}2486$	$.0^{4}2385$	$.0^{4}2287$	$.0^{4}2193$	$.0^{4}2103$	$.0^{4}2016$	$.0^{4}1933$	$.0^{4}1853$	$.0^{4}1776$
3.8	$.0^{4}1702$	$.0^{4}1632$	$.0^{4}1563$	$.0^{4}1498$	$.0^{4}1435$	$.0^{4}1375$	$.0^{4}1317$	$.0^{4}1262$	$.0^{4}1208$	$.0^{4}1157$
3.9	$.0^{4}1108$	$.0^{4}1061$	$.0^{4}1016$	$.0^{5}9723$	$.0^{5}9307$	$.0^{5}8908$	$.0^{5}8525$	$.0^{5}8158$	$.0^{5}7806$	$.0^{5}7469$
4.0	$.0^{5}7145$	$.0^{5}6835$	$.0^{5}6538$	$.0^{5}6253$	$.0^{5}5980$	$.0^{5}5718$	$.0^{5}5468$	$.0^{5}5227$	$.0^{5}4997$	$.0^{5}4777$
4.1	$.0^{5}4566$	$.0^{5}4364$	$.0^{5}4170$	$.0^{5}3985$	$.0^{5}3807$	$.0^{5}3637$	$.0^{5}3475$	$.0^{5}3319$	$.0^{5}3170$	$.0^{5}3027$
4.2	$.0^{5}2891$	$.0^{5}2760$	$.0^{5}2635$	$.0^{5}2516$	$.0^{5}2402$	$.0^{5}2292$	$.0^{5}2188$	$.0^{5}2088$	$.0^{5}1992$	$.0^{5}1901$
4.3	$.0^{5}1814$	$.0^{5}1730$	$.0^{5}1650$	$.0^{5}1574$	$.0^{5}1501$	$.0^{5}1431$	$.0^{5}1365$	$.0^{5}1301$	$.0^{5}1241$	$.0^{5}1183$
4.4	$.0^{5}1127$	$.0^{5}1074$	$.0^{5}1024$	$.0^{6}9756$	$.0^{6}9296$	$.0^{6}8857$	$.0^{6}8437$	$.0^{6}8037$	$.0^{6}7655$	$.0^{6}7290$
4.5	$.0^{6}6942$	$.0^{6}6610$	$.0^{6}6294$	$.0^{6}5992$	$.0^{6}5704$	$.0^{6}5429$	$.0^{6}5167$	$.0^{6}4917$	$.0^{6}4679$	$.0^{6}4452$
4.6	$.0^{6}4236$	$.0^{6}4029$	$.0^{6}3833$	$.0^{6}3645$	$.0^{6}3467$	$.0^{6}3297$	$.0^{6}3135$	$.0^{6}2981$	$.0^{6}2834$	$.0^{6}2694$
4.7	$.0^{6}2560$	$.0^{6}2433$	$.0^{6}2313$	$.0^{6}2197$	$.0^{6}2088$	$.0^{6}1984$	$.0^{6}1884$	$.0^{6}1790$	$.0^{6}1700$	$.0^{6}1615$
4.8	$.0^{6}1533$	$.0^{6}1456$	$.0^{6}1382$	$.0^{6}1312$	$.0^{6}1246$	$.0^{6}1182$	$.0^{6}1122$	$.0^{6}1065$	$.0^{6}1011$	$.0^{7}9588$
4.9	$.0^{7}9096$	$.0^{7}8629$	$.0^{7}8185$	$.0^{7}7763$	$.0^{7}7362$	$.0^{7}6982$	$.0^{7}6620$	$.0^{7}6276$	$.0^{7}5950$	$.0^{7}5640$

*Los números pequeños que aparecen como exponentes indican el número de ceros que siguen al punto decimal. Por ejemplo, $.0^{2}9957$ es el valor .009957.

Fuente: reproducida con autorización de los propietarios de los derechos de autor. The President and Fellows of Harvard College. De la tabla integral de pérdida normal unitaria que aparece como Tabla IV en *Introduction to Statistics for Business Decisions* de.Robert Schlaifer, publicado por McGraw-Hill Book Company, New York, 1961.

Apéndice 10

Valores críticos de *T* en la prueba de Wilcoxon

De 1 extremo	De 2 extremos	$n = 5$	$n = 6$	$n=7$	$n = 8$	$n = 9$	$n = 10$
P = 0.05	P = 0.10	1	2	4	6	8	11
P = 0.025	P = 0.05		1	2	4	6	8
P = 0.01	P = 0.02			0	2	3	5
P = 0.005	P = 0.01				0	2	3

De1 extremo	De 2 extremos	$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n=15$	$n = 16$
P = 0.05	P = 0.10	14	17	21	26	30	36
P = 0.025	P = 0.05	11	14	17	21	25	30
P = 0.01	P = 0.02	7	10	13	16	20	24
P = 0.005	P = 0.01	5	7	10	13	16	19

De1 extremo	De 2 extremos	n = 17	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$
P = 0.05	P = 0.10	41	47	54	60	68	75
P = 0.025	P = 0.05	35	40	46	52	59	66
P = 0.01	P = 0.02	28	33	38	43	49	56
P = 0.005	P = 0.01	23	28	32	37	43	49

De 1 extremo	De 2 extremos	$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
P = 0.05	P = 0.10	83	92	101	110	120	130
P = 0.025	P = 0.05	73	81	90	98	107	117
P = 0.01	P = 0.02	62	69	77	85	93	102
P = 0.005	P = 0.01	55	68	68	75	84	92

De 1 extremo	De 2 extremos	$n = 29$	$n = 30$	$n = 31$	$n = 32$	$n = 33$	$n = 34$
P = 0.05	P = 0.10	141	152	163	175	188	201
P = 0.025	P = 0.05	127	137	148	159	171	183
P = 0.01	P = 0.02	111	120	130	141	151	162
P = 0.005	P = 0.01	100	109	118	128	138	149

De 1 extremo	De 2 extremos	$n = 35$	$n = 36$	$n = 37$	$n = 38$	$n = 39$	
P = 0.05	P = 0.10	214	228	242	256	271	
P = 0.025	P = 0.05	195	208	222	235	250	
P = 0.01	P = 0.02	174	186	198	211	224	
P = 0.005	P = 0.01	160	171	183	195	208	

De 1 extremo	De 2 extremos	$n = 40$	n = 40	$n = 42$	$n = 43$	$n = 44$	$n = 45$
P = 0.05	P = 0.10	287	303	319	336	353	371
P = 0.025	P = 0.05	264	279	295	311	327	344
P = 0.01	P = 0.02	238	252	267	281	297	313
P = 0.005	P = 0.01	221	234	248	262	277	292

De 1 extremo	De 2 extremos	*n = 46*	*n = 47*	*n = 48*	*n = 49*	*n = 50*
P = 0.05	P = 0.10	389	408	*427*	446	466
P = 0.025	P = 0.05	361	379	397	415	434
P = 0.01	P = 0.02	329	345	362	380	398
P = 0.005	P = 0.01	307	323	339	356	373

(Fuente: de From F. Wilcoxon y R. A. Wilcox, Some Rapid Approximate Statistical Procedures, 1964, Reproducida con autorización de Lederle Laboratories Division, American Cyanamid Company.)

Índice

Schaum

Esta nueva edición presenta cambios muy importantes respecto de la anterior. En principio, el autor ha hecho un reordenamiento de algunos de los capítulos del libro, así, ahora, se estudia primero el análisis bayesiano de decisión, antes de adentrarse en el estudio relativo a regresión lineal, análisis de correlación, regresión múltiple y análisis de correlación múltiple; se incorporan un nuevo capítulo sobre estadística no paramétrica, y un nuevo apéndice sobre valores críticos de T en la prueba de Wilcoxon.

Un punto más que importante es la participación del Lic. Alfredo Díaz Mata como coautor del libro, su amplio conocimiento y su gran experiencia en la materia, así como las investigaciones realizadas por él, permiten contar con un libro más apegado al entorno latinoamericano y en especial a México. El manejo de unidades, datos estadísticos y tabuladores de salarios se han modificado, buscando con ello que el alumno obtenga una mejor perspectiva sobre la materia, y su aplicación a casos prácticos.

Sin duda el libro constituye un material muy completo para el estudio de la estadística, tanto descriptiva como inferencial, y su magnífico desarrollo teórico-práctico le permiten cubrir dos cursos básicos de la materia.

McGraw Hill

ISBN 968-422-787-6