AAU

AMERICAN ANDRAGOGY
UNIVERSITY

educandis orbis

Estadística aplicada a los negocios y la economía

TERCERA EDICIÓN

ALLEN L. WEBSTER
Bradley University

Traducción

YELKA MARÍA GARCÍA
Profesional en Lenguas Modernas
Especialización en traducción
Universidad de Los Andes

Revisión Técnica

CARLOS LESMES LOZANO
Estadístico
Universidad Nacional de Colombia
Licenciado y magíster en Matemática
Universidad Pedagógica Nacional

GILBERTO PRIETO MARÍN
Profesor de Estadística y Matemáticas
Facultad de Contaduría y Administración
Universidad Nacional Autónoma de México

Irwin McGraw-Hill

Santa Fe de Bogotá • Buenos Aires • Caracas • Guatemala • Lisboa • Madrid • México
New York • Panamá • San Juan • Santiago de Chile • Sao Paulo
Auckland • Hamburgo • Londres • Milán • Montreal • Nueva Delhi • París
San Francisco • San Luis • Sidney • Singapur • Tokio • Toronto

Estadística aplicada a los negocios y la economía, tercera edición

Traducido de la tercera edición en inglés de APPLIED STATISTICS FOR BUSINESS AND ECONOMICS. An Essentials Version

ISBN: 0-256-22554-0

Editor: Lily Solano Arévalo

3124567890 213456780

ISBN: 958-41-0072-6

Impreso en Colombia Printed in Colombia

Se imprimieron 8550 ejemplares en el mes de mayo de 2001
Impreso por Quebecor World Bogotá S.A.

A mi adorable esposa Bárbara:

Por todas las promesas que la vida nos cumple.

CONTENIDO

CAPÍTULO 4

Principios de probabilidad, 74

CAPÍTULO 5

Distribuciones de probabilidad, 102

CAPÍTULO 6

Distribuciones muestrales, 142

CAPÍTULO 10

Análisis de varianza, 270

CAPÍTULO 11

Regresión simple y correlación, 322

CAPÍTULO 12

Regresión múltiple y correlación, 374

CAPÍTULO 13

Series de tiempo y números índice, 410

CAPÍTULO 14

Prueba chi-cuadrado y otras pruebas no paramétricas, 462

CAPÍTULO 15

Técnicas de control de calidad, 518

APÉNDICE I

Cómo escribir un informe de negocios, 560

APÉNDICE II

Respuestas a problemas seleccionados, 568

APÉNDICE III

Tablas estadísticas, 580

ÍNDICE, 628

PREFACIO

De la misma manera que en las ediciones anteriores de este libro, esta tercera edición continúa enfatizando la importancia de aplicar el análisis estadístico a la solución de problemas comunes en los negocios. Toda oportunidad se utiliza para demostrar la forma como la estadística puede, de manera efectiva facilitar muchas decisiones que los gerentes enfrentan casi a diario. Además, la obra se ha condensado para presentar el material de forma más concisa y manejable. Varias de las características pedagógicas descritas a continuación también han sido agregadas para mejorar las ventajas de instrucción expuestas en este texto.

Esta tercera edición de *Estadística aplicada a los negocios y la economía*, se puede utilizar de manera efectiva en un curso de estadística. A pesar de que el material ha sido comprimido para permitir una discusión típica y dinámica de un curso de un semestre, la cobertura de temas sigue siendo suficiente para retar incluso al lector que terminó su exposición inicial al análisis estadístico en una secuencia de dos semestres.

A continuación se presenta una breve descripción de las características pedagógicas contenidas en esta edición.

Aspectos que se mantuvieron de la edición anterior

• Organización del capítulo

Cada capítulo abre con un diagrama de flujos que muestra los principales temas a tratar y la forma como se relacionan entre sí. Esto permite que el lector tenga la oportunidad de organizar el proceso que puede utilizar para dominar el material en dicho capítulo.

• Ejemplos con tres partes

Cada ejemplo de las herramientas estadísticas disponibles para las personas que toman las decisiones en los negocios, presenta una situación real que requiere una solución y son las situaciones típicas que se enfrentan al gerenciar una empresa. Estos ejemplos constan de tres partes. La primera es el *planteamiento del problema* que describe el dilema que debe resolverse. La segunda parte suministra una *solución* completa y coherente a dicho problema. Quizá lo más importante para el lector, pero que se encuentra ausente en muchos otros textos de estadística, es la *interpretación* de dicha solución. No es bueno para el lector resolver un problema cuya solución no comprende.

• Un procedimiento para la elaboración de reportes

La comunicación en los negocios es fundamental. Para que un negocio funcione como una unidad, los individuos que están dentro de dicho negocio deben poder comunicar sus ideas y pensamientos. El apéndice I describe e ilustra la forma como se debe preparar un informe para comunicar los propósitos esenciales y las recomendaciones a otros que estén interesados en los resultados. Sin esta importante destreza, quienes toman decisiones están en gran desventaja en su intento por dirigir un negocio.

• Problemas resueltos

Cada capítulo concluye con problemas y soluciones elaboradas que refuerzan las herramientas estadísticas que se presentaron en dicho capítulo. Esta característica proporciona un resumen conveniente y útil de las herramientas cruciales que se espera que el lector domine.

• Lista de fórmulas

Al final de cada capítulo se proporciona una lista manual de todas las fórmulas presentadas a través del capítulo y una breve descripción de su uso.

• Aplicaciones por computador

Se proporcionan instrucciones para mostrar cómo pueden utilizarse las herramientas estadísticas en un computador. Se presentan las impresiones de pantallas de computador con énfasis en Minitab y Excel junto con la discusión de sus aspectos importantes.

• Ejercicios por computador

El texto viene con un disquete que contiene varios archivos de datos a los que se puede tener acceso en los formatos de Minitab, Excel y ASCII. Algunos capítulos proporcionan al lector un problema que debe resolver con base en los datos que aparecen en dichos archivos y con la ayuda de las herramientas estadísticas discutidas en el capítulo. Lo anterior ofrece al lector una situación real que requiere de la aplicación de las técnicas estadísticas en el conjunto de datos que hay en el computador.

• Problemas del capítulo

Al final de cada capítulo existe una amplia gama de problemas con una diversidad de grados de dificultad que brindan al lector la oportunidad de agudizar y refinar sus habilidades estadísticas. De nuevo, estos problemas son de naturaleza aplicada lo que demuestra claramente cómo la estadística puede ayudar al proceso de toma de decisiones en los negocios.

Nuevas características de la tercera edición

- **Ejercicios de la sección**

Después de cada sección, dentro de cada capítulo se encuentran varios ejercicios que el lector debe resolver y que refuerzan su comprensión del material expuesto. Esto proporciona una retroalimentación inmediata sobre si se ha comprendido en forma suficiente el material de dicha sección antes de proceder a la siguiente.

- **Escenario**

Cada capítulo inicia con un recuento corto denominado *Escenario* que presenta un caso real que se puede tratar mejor utilizando las herramientas estadísticas contenidas en dicho capítulo. Se describe la naturaleza del problema que se debe resolver y así se proporciona al lector los fundamentos sobre los cuales debe realizar sus lecturas y el análisis del material del capítulo.

- **Puesta en escena**

Es una corta sección al cierre de cada capítulo que remite al lector a la sección inicial *Escenario*. Se proporcionan datos e información adicional para resolver la situación propuesta en la sección *Escenario*, utilizando el conocimiento adquirido en el capítulo. Este ejercicio combina todo el capítulo en un solo "paquete" que mejora la comprensión total sobre el material y ata de manera efectiva los cabos sueltos.

- **Del escenario a la vida real**

El segmento final de cada capítulo proporciona sitios Web en internet que el lector puede consultar para encontrar datos e información complementaria, y la cual pueda utilizar para su estudio estadístico adicional. Estos sitios Web se seleccionan especialmente para relacionarlos con los asuntos y problemas planteados en las secciones *Escenario* y *Puesta en escena*.

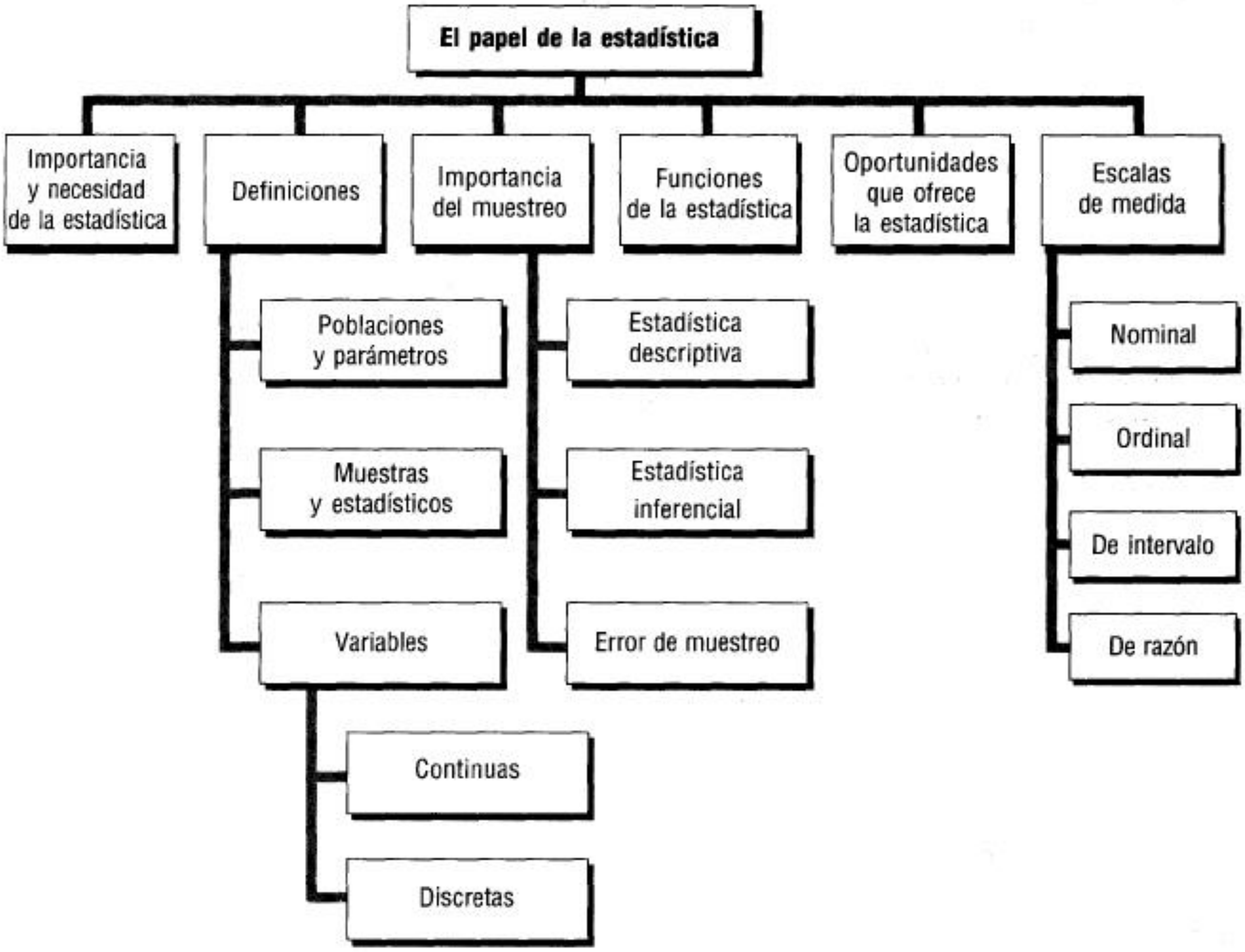
Allen L. Webster

1

El papel de la estadística

Plan del capítulo

Este capítulo introduce al concepto de estadística como un estudio organizado. Se presenta el propósito general del análisis estadístico y las muchas formas en las cuales la estadística puede ayudar a encontrar soluciones a problemas que se presentan en la vida profesional.

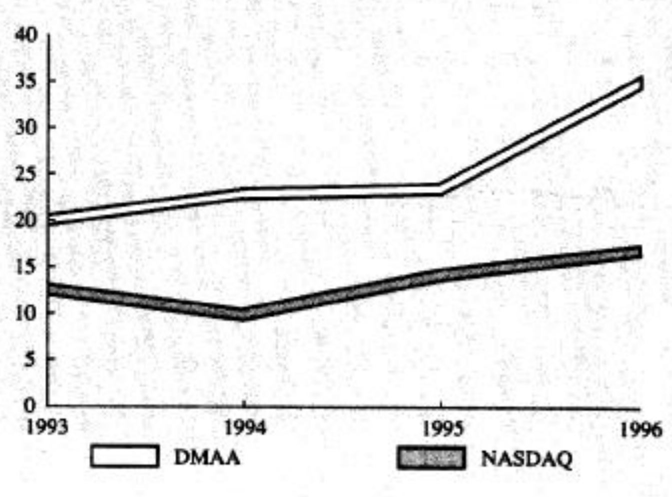


ESCENARIO

La revista *Fortune* (diciembre 9, 1996) informa que un frenesí en la contratación ha hecho que se produzca un "mercado laboral de vendedores para *nerds*". La búsqueda del talento en el sector de la información en tecnología ha generado una competencia tan intensa para "aficionados a la cuántica y a los computadores" que quienes buscan trabajo, con incluso mínimas habilidades cuantitativas, están llenos de ofertas. El talento es tan escaso y la demanda tan elevada que las compañías están contratando a estudiantes de bachillerato y otras compañías están tratando de arrebatárselos.

Pete Davis, un ingeniero senior de sistemas y redes para la TIAC, proveedor de servicios de internet en Bedford, Massachusetts, ha recibido por lo menos 15 ofertas de trabajo desde que desempeña su cargo actual –lo cual no está nada mal para alguien que tan sólo tiene 18 años–. Quienes están encargados de la contratación de personal llaman a Davis o le envían por correo electrónico opciones atractivas y ofertas de pago con incrementos lucrativos. Aunque el cargo de Davis no es del todo típico, la batalla es para los empleados que están capacitados en análisis de datos.

La estrategia parece valer la pena para negocios que buscan, de manera agresiva, talento joven que sea capaz de analizar y comprender la información cuantitativa. Como lo demuestra el siguiente diagrama, las utilidades de estas compañías han respondido de manera favorable, incrementándose mucho más rápidamente que las utilidades de otras compañías que le están dando largas a su ingreso a la era de la información. El Informe sobre la Industria Educativa (*Education Industry Report*) de St. Cloud, Minnesota, presentó un índice de utilidades de 15 compañías que están buscando, de manera agresiva, empleados que puedan manejar datos y aplicar el pensamiento analítico y estadístico básico a problemas empresariales comunes. Se comparó el índice de Evaluación de Alerta en el Manejo de Datos (*Data Management Alert Assessment*, DMAA) con el índice de la Asociación Nacional de Cotizaciones Automatizadas de Agentes de Bolsa (*National Association of Security Dealers Automated Quotations*, NASDAQ). Los resultados son evidentes. Los rendimientos de las empresas que contratan gente que posee conocimientos básicos en estadística sobrepasan los de empresas que aún no han visto tal luz. El reconocimiento, por parte de un número creciente de negocios, de la necesidad que tienen los gerentes efectivos de tener por lo menos conocimientos rudimentarios de estadística ha precipitado –según los términos de la revista *Fortune*–, una carrera por el talento, que se asemeja al frenesí agrario de la época en la que se perseguía la apropiación gratuita de la tierra en Oklahoma.



1.1 Introducción

A medida que aumenta la complejidad de nuestro mundo, se hace cada vez más difícil tomar decisiones inteligentes y bien documentadas. Con frecuencia tales decisiones deben tomarse con mucho menos que un conocimiento adecuado y experimentando una gran incertidumbre. Sin embargo, las soluciones a estos problemas son esenciales para nuestro bienestar e incluso para nuestra supervivencia final. Continuamente estamos recibiendo presiones debido a problemas económicos angustiosos como una inflación galopante, el sistema tributario engorroso y oscilaciones excesivas en el ciclo empresarial. Todo nuestro tejido económico y social está amenazado por la contaminación ambiental, la deuda pública onerosa, la tasa de criminalidad que siempre va en aumento y las impredecibles tasas de interés.

Si estas condiciones parecen ser características del estilo de vida actual, no debe olvidarse que problemas de esta naturaleza contribuyeron a la caída de la antigua Roma más que la invasión de las hordas de bárbaros provenientes del norte. Un período de éxito en este planeta, relativamente corto, no es garantía de una supervivencia futura. A menos que puedan encontrarse soluciones viables a estos apremiantes problemas, podríamos acompañar en el olvido al dinosaurio y al ave dodo, como ya lo hicieron los antiguos romanos.

Este capítulo aportará una visión general sobre lo que es la estadística y cómo puede utilizarse. Esta visión general sobre la naturaleza de la estadística y los beneficios que puede proporcionarnos se efectuará revisando:

- Las definiciones básicas de las herramientas estadísticas.
- Cómo llevar a cabo el muestreo para realizar análisis estadísticos.
- Las funciones que cumple la estadística.
- Cómo puede ayudar la estadística en la profesión.

Comencemos con una breve exposición del papel significativo que juega la estadística en el delicado proceso de toma de decisiones.

1.2 La importancia de la estadística

Virtualmente cada área de la investigación científica seria puede beneficiarse del análisis estadístico. Para quien formula las políticas económicas y para quien asesora al presidente y a otros funcionarios públicos sobre procedimientos económicos apropiados, la estadística ha demostrado ser una herramienta valiosa. Las decisiones sobre las tasas tributarias, los programas sociales, el gasto de defensa y muchos otros asuntos pueden hacerse de manera inteligente tan sólo con la ayuda del análisis estadístico. Los hombres y mujeres de negocios, en su eterna búsqueda de la rentabilidad, consideran que la estadística es esencial en el proceso de toma de decisiones. Los esfuerzos en control de calidad, minimización de costos, combinación de productos e inventarios, y una gran cantidad de otros asuntos empresariales, pueden manejarse efectivamente a través del uso de procedimientos estadísticos comprobados.

Para quienes están en el área de la investigación de mercados, la estadística es de gran ayuda en el momento de determinar qué tan probable es que un producto nuevo sea exitoso. La estadística también es muy útil para evaluar las oportunidades de inversión por parte de asesores financieros. Los contadores, los jefes de personal, y los fabricantes encuentran oportunidades ilimitadas de beneficiarse con el uso del análisis estadístico. Incluso un investigador en el campo de la medicina, interesado en la efectividad de un nuevo medicamento, considera la estadística una aliada imprescindible.

Tales aplicaciones y muchas otras se ilustran a lo largo de este texto. Se mostrará cómo utilizar la estadística en el mejoramiento del desempeño laboral y en muchos otros aspectos de la vida diaria.

1.3 Oportunidades que ofrece la estadística

Es factible que en pocos años abandone la relativa seguridad del ambiente académico y esté metido de cabeza en el mundo competitivo. Desde el punto de vista práctico, usted debe conocer la manera de utilizar los conocimientos en estadística después de graduarse. No existe duda alguna acerca de que una experiencia académica, adecuadamente relacionada con unos firmes cimientos ampliará significativamente las oportunidades de encontrar empleo y, posteriormente, le permitirá demostrar la competitividad laboral.

Cuando encuentre ese trabajo anhelado que le ponga en la rápida ruta del éxito profesional, su jefe espera que usted haga dos cosas:

1. Tomar decisiones.
2. Solucionar problemas.

Estos dos cometidos pueden lograrse a través de la aplicación de procedimientos estadísticos.

A. La aplicación universal de la estadística

Al ser capaz de solucionar problemas y tomar decisiones, se obtendrá una excelente posición en la demanda del mercado laboral. Si logra tomar decisiones incisivas cuando se está solucionando los problemas de alguien, dicha persona estará dispuesta a proporcionarle una recompensa generosa. El mundo laboral paga más a las personas que son capaces de plantear las preguntas correctas para alcanzar los objetivos fundamentales, que a quienes tienen la responsabilidad de resolverlas. Con frecuencia, las respuestas son bastante evidentes una vez que se han planteado las preguntas. El análisis estadístico probará ser de gran utilidad en la acertada formulación de estas preguntas esenciales.

Los empresarios reconocen que los problemas complejos que enfrenta el mundo actual requieren soluciones cuantitativas. Si usted no está en capacidad de aplicar la estadística y otros métodos cuantitativos a muchos de los problemas comunes que sin duda se le presentarán, estará en gran desventaja en el mercado laboral.

Casi todas las áreas del saber requieren del pensamiento estadístico. Las disciplinas de estudio que dependen ampliamente del análisis estadístico, incluyen –pero no se limitan a–, *marketing*, finanzas, economía e investigación de operaciones. Los principios aprendidos en contabilidad y gerencia administrativa también se basan en la preparación estadística.

Los analistas financieros y económicos con frecuencia se basan en sus habilidades cuantitativas para proporcionar soluciones a problemas difíciles. La comprensión de los principios financieros y económicos permitirá aplicar las técnicas estadísticas para hallar soluciones viables y tomar decisiones. Quienes aspiran a cargos en el área contable o administrativa, a ser independientes, o a desempeñar otra profesión en el sector industrial, descubrirán que comprender la estadística no sólo mejora las oportunidades de obtener un empleo, sino que también aumenta la probabilidad de promoción mediante el enriquecimiento en el desempeño laboral.

Las personas empleadas en tareas cuantitativas que trabajan con procedimientos estadísticos, con frecuencia gozan de salarios más altos y están más protegidos de los trabajos sin futuro. Además, muy al inicio de sus carreras, generalmente se encuentran en contacto cercano con personas en cargos de alto nivel. La proximidad a la élite ejecutiva es inevitable porque la alta gerencia necesita de la información y asistencia que la gente con entrenamiento en estadística puede proporcionarle. En el mercado laboral actual, los empresarios sencillamente no desean contratar o conservar a quienes no saben estadística.

Bien sea que las aspiraciones profesionales tiendan hacia la industria privada, el servicio público, el gobierno, o hacia alguna otra fuente de retribución remunerada, la experiencia académica será más completa si se adquiere una sólida formación en fundamentos de análisis estadístico.

B. Gerencia de calidad total

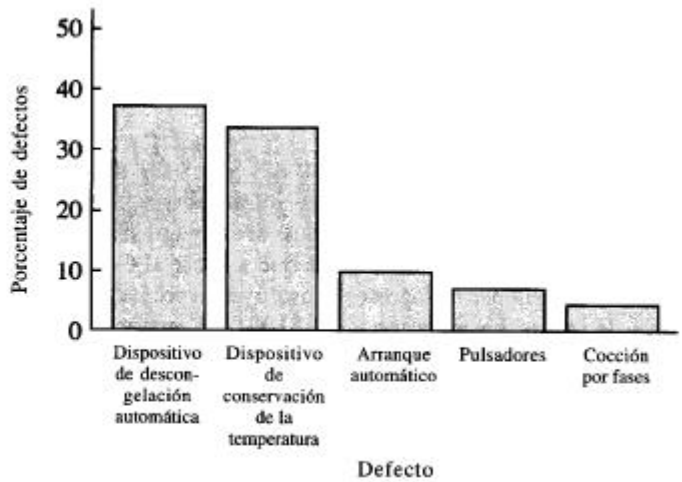
A medida que la competencia mundial se intensifica, surge, de parte de los negocios, un esfuerzo por promover la calidad de sus productos. Este esfuerzo, conocido ampliamente como *Gerencia de Calidad Total* (*Total Quality Management*, TQM), tiene como propósito central la promoción de las cualidades del producto que el consumidor considera importantes. Tales atributos van desde la ausencia de defectos hasta el servicio eficiente y la respuesta rápida a las posibles quejas del consumidor. Hoy día, la mayoría de los grandes negocios (así como también muchos negocios pequeños) tienen departamentos de *Control de Calidad* (*Quality Control*, QC) cuya función es recolectar datos sobre el desempeño y solucionar problemas de calidad. Así, la TQM representa un área creciente de oportunidades para quienes tienen conocimientos en estadística.

La TQM involucra el uso de equipos administrativos integrados conformados por ingenieros, expertos en marketing, especialistas en diseño, estadísticos, y otros profesionales que pueden contribuir a la satisfacción del cliente. La formación de estos equipos, denominada *Despliegue de la Función de Calidad* (*Quality Function Deployment*, QFD), está diseñada para reconocer y agenciar las inquietudes de los consumidores. Los especialistas actúan conjuntamente para promover la calidad del producto y para que supla de manera efectiva las necesidades y preferencias del consumidor.

Otro método común para mejorar la calidad de un producto es el uso de los círculos de *Control de Calidad (Quality Control, QC)*. Los círculos de control de calidad constan de un grupo pequeño de empleados (generalmente entre 5 y 12) que se reúnen regularmente para solucionar problemas relacionados con el trabajo. Con frecuencia se conforman tanto con trabajadores en línea como con representantes de la gerencia; los miembros de estos círculos QC son todos de la misma área de trabajo y reciben capacitación formal en control estadístico de calidad y en planeación de grupos. A través de discusiones abiertas y del análisis estadístico, los círculos QC pueden lograr mejoras significativas en diversas áreas que van desde el mejoramiento de la calidad, el diseño del producto, la productividad y los métodos de producción, hasta la reducción de costos y seguridad. Se estima que más del 90% de las 500 mejores compañías que aparecen en la revista *Fortune* utilizan de manera efectiva los círculos de control de calidad.

Uno de los elementos más importantes del TQM es un conjunto de herramientas y métodos estadísticos utilizados para promover el *Control Estadístico de Calidad* (*Statistical Quality Control*, SQC). Tales herramientas ayudan a organizar y analizar datos para efectos de solucionar problemas. Una de estas herramientas es *el diagrama de Pareto*, denominado así en honor al economista italiano Vilfredo Pareto. Este diagrama identifica los problemas de calidad que se presentan con mayor frecuencia o que han demostrado ser los más costosos. La figura 1.1 muestra un diagrama de Pareto de los defectos que afectan la producción de hornos microondas, comercializados por JC Penney.

Figura 1.1
Diagrama de Pareto



Fuente: *QC Circle*, JC Penney, 1993.

Los diagramas de Pareto con frecuencia expresan la regla 80/20: el 80% de todos los problemas se debe al 20% de las causas. Como lo demuestra la figura 1.1, aproximadamente el 75% de todos los problemas es causado por el dispositivo de descongelación automático y por el de conservación de la temperatura del horno.

Hablando en términos generales, el SQC está diseñado para asegurar que los productos cumplan con unas normas y especificaciones mínimas de producción. Este objetivo con frecuencia se promueve a través del uso del *muestreo de aceptación*, el cual es parte integral del SQC. El muestreo de aceptación implica probar una muestra aleatoria de productos existentes para determinar si se debe aceptar o rechazar todo el envío, o el lote. Esta decisión se basa en parte en un *nivel de calidad aceptable* (*Acceptable Quality Level*, AQL), o número máximo de defectos que una empresa está dispuesta a tolerar.

En las organizaciones se es cada vez más consciente de la necesidad de mantener la calidad del producto. Si una firma va a competir de manera exitosa, debe tomar todas las precauciones para garantizar que sus productos cumplan con ciertos estándares básicos. Por tanto, no es ninguna exageración insistir en la importancia de la TQM. Los principios inherentes al TQM están aumentando en popularidad; representan la dirección futura del análisis estadístico, aplicada al mundo de los negocios. En el capítulo 15 se analizará con mayor profundidad la TQM.

C. Necesidad de la formación en estadística

Se podría pensar que el tipo de trabajo a que se aspira no necesitará del análisis estadístico. O quizás podría argumentarse que el personal de estadísticos de la compañía realizará el trabajo estadístico pertinente y que no existe la necesidad de manejar los detalles del análisis estadístico.

Este no es el caso. Incluso si los estadísticos profesionales de la organización realizan el trabajo estadístico pertinente, es esencial poseer un cierto nivel de formación en este campo. Para determinar cómo puede ayudar el *staff* de estadística al desempeño del trabajo de los otros, se debe conocer qué es la estadística, qué hacen los estadísticos y cómo lo hacen. Cuando los problemas surgen se debe determinar cómo puede ayudar la estadística. Para lograrlo, es necesario comprender los procedimientos estadísticos, y poder comunicarse con los estadísticos, en un esfuerzo conjunto para diseñar soluciones adecuadas y tomar decisiones óptimas. Una vez se ha adquirido esta familiaridad con el análisis estadístico, se sorprenderá de las infinitas formas en que la estadística puede ayudar en la solución de problemas que surgen con frecuencia en un escenario empresarial.

1.4 Algunas definiciones básicas

Toda rama de la investigación científica tiene su vocabulario propio y la estadística no es la excepción. Esta sección revisa algunos de los términos comunes utilizados en el análisis estadístico. Las definiciones y expresiones que siguen son esenciales para la comprensión de cómo se realizan las pruebas estadísticas.

A. Poblaciones y parámetros

En todo estudio estadístico, el investigador está interesado en una determinada colección o conjunto de observaciones denominadas **población** (o universo). Si los ingresos de los 121 millones de asalariados de los Estados Unidos son de interés para un economista que asesore al Congreso en la formulación del plan nacional tributario, entonces los 121 millones de ingresos constituyen la población. Si se está considerando un plan tributario para los perceptores de ingresos superiores a US $100,000, entonces tales ingresos superiores a US $100,000 constituyen la población. La población es pues la colección completa de todas las observaciones de interés.

Población Es la recolección completa de todas las observaciones de interés para el investigador.

Si el director ejecutivo (*Chief Executive Officer*, CEO) de una gran empresa manufacturera desea estudiar la producción de todas las plantas de propiedad de la firma, entonces la producción de todas estas plantas es la población.

Un **parámetro** es toda medida descriptiva de una población. Algunos ejemplos son: el ingreso promedio de todos los asalariados de Estados Unidos, o la producción total de todas las plantas manufactureras. El punto clave para recordar es que un parámetro describe una población.

Parámetro Es una medida descriptiva de la población total de todas las observaciones de interés para el investigador.

B. Muestras y estadísticos

Aunque generalmente los estadísticos se interesan en algún aspecto de toda la población, generalmente descubren que las poblaciones son demasiado grandes para ser estudiadas en su totalidad. Calcular el ingreso promedio de cada uno de los 121 millones de asalariados sería una tarea abrumadora. Por consiguiente, generalmente debe ser suficiente estudiar tan sólo una pequeña porción de dicha población. A esta porción más pequeña y más manejable se le denomina **muestra**. Una muestra es un subconjunto de la población seleccionado científicamente.

Muestra Es una parte representativa de la población que se selecciona para ser estudiada ya que la población es demasiado grande como para analizarla en su totalidad.

Cada mes el Ministerio de Trabajo de Estados Unidos *(U.S. Department of Labor)* calcula el ingreso promedio de una muestra de varios miles de asalariados seleccionados entre la población total de 121 millones de trabajadores. El promedio de esta muestra se utiliza luego como una estimación del ingreso promedio para toda la población. Las muestras son necesarias porque estudiar las poblaciones completas resulta muy costoso y consume demasiado tiempo.

Un **estadístico** es una medida descriptiva de una muestra. El ingreso promedio de esos varios miles de trabajadores, calculado por el Ministerio de Trabajo, es un estadístico. El estadístico es a la muestra lo que el parámetro es a la población. El estadístico sirve como una estimación del parámetro. Aunque en realidad el interés se fija en el valor del parámetro de la población, con frecuencia debe haber conformidad con sólo calcularlo con un estadístico de la muestra que se ha seleccionado.

Estadístico Elemento que describe una muestra y sirve como una estimación del parámetro de la población correspondiente.

C. Variables

Una **variable** es la característica de la muestra o población que se está observando. Si el asesor estadístico del alcalde de San Francisco está interesado en la distancia que deben recorrer cada mañana las personas que viajan diariamente largas distancias hacia el trabajo, la variable es *millas recorridas*. En un estudio respecto al ingreso de los asalariados en Estados Unidos, la variable es *ingreso*.

Variable Es una característica de la población que se está analizando en un estudio estadístico.

Una variable puede ser (1) cuantitativa o (2) cualitativa. Si las observaciones pueden expresarse numéricamente, entonces es una variable **cuantitativa**. Los ingresos de todos los asalariados son un ejemplo de una población cuantitativa. Otros ejemplos incluyen la estatura de las personas en las que se está interesado, los puntajes que los estudiantes reciben en un examen final de estadística, y el número de millas que las personas recorren cada mañana al ir a su sitio de trabajo en San Francisco. En cada caso, las observaciones se miden numéricamente.

Una variable **cualitativa** se mide de manera no numérica. El estado civil de quienes solicitan crédito, el sexo de los estudiantes en una clase de estadística, y la raza, color de cabello, preferencia religiosa de las personas que recorren largas distancias a su trabajo en San Francisco, son ejemplos de variables cualitativas. En cada caso las observaciones se miden de manera no numérica.

Además las variables pueden ser (1) continuas o (2) discretas. Una variable **continua** es aquella que puede tomar cualquier valor dentro de un rango dado. No importa qué tan cerca puedan estar dos observaciones, si el instrumento de medida es lo suficientemente preciso, puede hallarse una tercera observación que se encuentre entre las dos primeras. Una variable continua generalmente resulta de la medición.

Una variable **discreta** está limitada a ciertos valores, generalmente números enteros. Con frecuencia son el resultado de la enumeración o del conteo. El número de estudiantes de la clase y el número de autos vendidos por General Motors son algunos ejemplos. En ningún caso se observarán valores fraccionarios.

A través del estudio de la estadística se hará referencia en repetidas ocasiones a estos conceptos y términos. Se debe ser consciente del papel que juega cada uno de ellos en el proceso de análisis estadístico. Es de especial importancia poder diferenciar entre una población y sus parámetros, y entre una muestra y sus estadísticos.

1.5 La importancia del muestreo

Gran parte del trabajo de un estadístico se realiza con muestras. Las muestras son necesarias debido a que con frecuencia las poblaciones son demasiado grandes para ser estudiadas en su totalidad. Es muy costoso y demanda mucho tiempo examinar la población total, por tanto, debe seleccionarse una muestra de la población, calcular el estadístico de la muestra, y utilizarlo para estimar el parámetro correspondiente de la población.

Este análisis sobre las muestras implica una distinción entre las dos principales ramas del análisis estadístico: (1) la estadística descriptiva y (2) la estadística inferencial. La **estadística descriptiva** es el proceso de recolectar, agrupar y presentar datos de una manera tal que describa fácil y rápidamente dichos datos. Los capítulos 2 y 3 analizan la estadística descriptiva e ilustran los diversos métodos y herramientas que pueden utilizarse para presentar y resumir grandes conjuntos de datos.

La **estadística inferencial** involucra la utilización de una muestra para sacar alguna inferencia o conclusión sobre la población de la cual hace parte la muestra. Cuando el Ministerio de Trabajo utiliza el ingreso promedio de una muestra de varios miles de trabajadores para calcular el ingreso promedio de los 121 millones de trabajadores, está utilizando una forma simple de estadística inferencial.

La exactitud de toda estimación es de enorme importancia. Esta exactitud depende en gran parte de la forma como se tomó la muestra, y del cuidado que se tenga para garantizar que la muestra proporcione una imagen confiable de la población. Sin embargo, con mucha frecuencia se comprueba que la muestra no es del todo representativa de la población y resultará un **error de muestreo.** El **error de muestreo** es la diferencia entre el estadístico de la muestra utilizada para calcular el parámetro de la población y el valor real pero desconocido del parámetro.

Error de muestreo Es la diferencia entre el parámetro desconocido de la población y el estadístico de la muestra utilizado para calcular el parámetro.

Existen dos causas posibles del error de muestreo. La primera fuente del error de muestreo es el azar en el proceso de muestreo. Debido al factor azar en la selección de elementos de la muestra, es posible seleccionar sin darse cuenta, elementos atípicos que no representan la población. Por ejemplo, en el esfuerzo por estimar la media poblacional es factible que se seleccionen elementos en la muestra que sean anormalmente grandes, produciendo así una sobreestimación de la media poblacional. Por otro lado, el azar puede producir un gran número de elementos de muestra que sean inusualmente pequeños, produciendo una subestimación del parámetro. En cualquiera de los dos casos, ha ocurrido un error de muestreo.

Una forma más seria de error de muestreo es el **sesgo muestral**. El sesgo muestral ocurre cuando hay alguna tendencia a seleccionar determinados elementos de muestra en lugar de otros. Si el proceso de muestreo se diseña de manera incorrecta y tiende a promover la selección de demasiadas unidades con una característica en especial, a expensas de las unidades que no tienen dicha característica, se dice que la muestra está sesgada. Por ejemplo, el proceso de muestreo puede favorecer de manera inherente la selección de hombres excluyendo a las mujeres, o de personas casadas excluyendo a las solteras.

Sesgo muestral Es la tendencia a favorecer la selección de ciertos elementos de muestra en lugar de otros.

Un manejo más riguroso del sesgo en el muestreo se presenta en un capítulo posterior. Aunque el error de muestreo nunca puede medirse debido a que el parámetro permanece desconocido, se debe ser consciente de que existe la probabilidad que esto ocurra.

1.6 Las funciones de la estadística

En repetidas ocasiones se ha enfatizado la utilidad de la estadística y la amplia variedad de problemas que puede resolver. Para ilustrar de manera más completa esta amplia aplicabilidad, es necesario analizar las diversas funciones de la estadística. La **estadística** es la ciencia que tiene que ver con la (1) recolección, (2) organización, (3) presentación, (4) análisis, e (5) interpretación de datos.

Aunque en todo estudio estadístico el primer paso es la recolección de datos, es usual en un curso básico de estadística asumir que los datos ya han sido recolectados y que ahora están disponibles. Por consiguiente, el trabajo comienza con el esfuerzo por organizar y presentar estos datos de manera significativa y descriptiva. Los datos deben colocarse en un orden lógico que revele rápida y fácilmente el mensaje que contienen. Este procedimiento constituye el proceso de la estadística descriptiva, tal como se define y se discute en los capítulos siguientes. Luego de que los datos se han organizado y se han presentado para su revisión, el estadístico debe analizarlos e interpretarlos. Estos procedimientos se basan en la estadística inferencial y constituyen un importante beneficio para el análisis estadístico, mediante la ayuda en el proceso de toma de decisiones y solución de problemas.

Se descubrirá que a través de la aplicación de procedimientos estadísticos precisos, es posible predecir el futuro con cierto grado de exactitud. Toda empresa que se enfrenta a las presiones competitivas puede beneficiarse considerablemente de la capacidad para anticipar las condiciones del negocio, antes que éstas ocurran. Si una empresa sabe cómo van a estar sus ventas en cierto momento en el futuro cercano, la gerencia puede hacer planes más exactos y efectivos respecto a las operaciones actuales. Si se calculan las ventas futuras con un grado de exactitud confiable, la gerencia puede tomar fácilmente decisiones importantes respecto a los niveles de inventario, pedidos de materia prima, contrataciones de empleados y, virtualmente, sobre cada aspecto de las operaciones del negocio.

1.7 Escalas de medida

Las variables pueden clasificarse con base en su escala de medida. La manera en que se clasifican las variables afecta en gran parte la forma como se utilizan en el análisis. Las variables pueden ser (1) nominales, (2) ordinales, (3) de intervalo, o (4) de razón.

Una medida **nominal** se crea cuando se utilizan nombres para establecer categorías dentro de las cuales las variables pueden registrarse exclusivamente. Por ejemplo, el sexo puede clasificarse como "hombre" o "mujer". Se podría codificar también con un "1" o un "2", pero los números servirían tan sólo para indicar las categorías y no tendrían significado numérico; los cálculos matemáticos que utilizan estos códigos no tendrían sentido. Las bebidas colas pueden clasificarse como Coca-Cola, Pepsi, 7-Up, o Ale 8. Cada bebida podría registrarse en una de estas categorías excluyendo las otras.

Mediciones en escala nominales Nombres o clasificaciones que se utilizan para datos en categorías distintas y separadas.

La tabla 1.1 ilustra la forma como la revista *Money* clasificó diferentes fondos de inversión. Vale la pena destacar que cada fondo está ubicado en una categoría en particular, con base en su comportamiento financiero.

Tabla 1.1
Medidas en escala nominales de los fondos de inversión

Categoría / Nombre del Fondo	
Crecimiento Agresivo	**Fondos del Mercado Monetario**
Twenty Century Growth	Alger Portfolio
Janus	Mariner Government
Reembolso Total	**Exento de Impuestos**
Scudder	Kemper Municipals
Vanguard Star	
PaxWorld	
USAA Cornerstone	
Bonos	
Strong Short-Term	
Scudder Short-Term	

Fuente: *Money*, julio, 1992.

Es importante recordar que una medida en escala nominal no indica ningún orden de preferencia, sino que simplemente establece una disposición categórica en la cual se puede ubicar cada observación.

A diferencia de una medida en escala nominal, una medida en escala ordinal sí muestra un ordenamiento o secuencia de los datos. Es decir, que las observaciones se clasifican con base en algunos criterios. Sears Roebuck, uno de los vendedores minoristas más grandes de Estados Unidos, clasifica muchos de sus productos como "buenos", "mejores" y "los mejores". Las encuestas de opinión con frecuencia utilizan una medida en escala ordinal como "totalmente de acuerdo", "de acuerdo", "sin opinión", "en desacuerdo", y "en total desacuerdo".

Al igual que con los datos nominales, los números pueden utilizarse para ordenar los rangos. Y al igual que con los datos nominales, la magnitud de los números no es importante; el rango depende sólo del *orden* de los valores. Sears hubiera podido utilizar los rangos de "1", "2", y "3", o "1", "3,", y "12" para este asunto. Las diferencias aritméticas entre los valores carecen de sentido. Un producto con rango "2" no es dos veces mejor que uno con rango "1".

Medidas en escala ordinales Son las que clasifican las observaciones en categorías con un orden significativo.

El mismo tema que la revista *Money* mencionó anteriormente, presentó los rangos para las inversiones con base en los niveles de riesgo mostrados en la tabla 1.2. Vale la pena destacar que los rangos de riesgo "Muy alto", "Alto" , y "Bajo" se hubiera podido basar en los valores "1", "2", y "3", o "A", "B", y "C". Pero las diferencias reales en los niveles de riesgo no pueden medirse de manera significativa. Se sabe solamente que una inversión de rango "Alto" tiene mayor riesgo que una con rango "Bajo".

Tabla 1.2
Rango ordinal del riesgo de inversión

Inversión	Factor de riesgo
Oro	Muy alto
Compañías de poco crecimiento	Muy alto
Máxima ganancia de capital	Alto
Internacional	Alto
Ingreso de opciones	Bajo
Equilibrado	Bajo

En una escala de **intervalo** las variables se miden de manera numérica, y al igual que los datos ordinales, llevan inherente un rango u ordenamiento. Sin embargo, a diferencia de los rangos ordinales, la diferencia entre los valores es importante. Por eso, las operaciones aritméticas de suma y resta, son significativas. La escala Fahrenheit para las temperaturas es un ejemplo de una escala de intervalo. 70 grados no es sólo una temperatura más alta que 60 grados, sino que la misma diferencia de 10 grados existe entre 90 y 100 grados Farenheit.

En una escala de intervalos el valor de cero se selecciona arbitrariamente. No existe nada concreto que haya obligado a fijar la temperatura cero grados, sino que, simplemente, es un punto de referencia arbitrario. Hubiera podido crearse la escala Fahrenheit de manera que cero se hubiera establecido en una temperatura más caliente (o más fría). No se da a cero un significado distinto al decir que es 10 grados más frío que 10 grados Fahrenheit. Así, 80 grados no es dos veces más caliente que 40 grados y la relación 80/40 no tiene significado.

Medidas en escala de intervalo Medidas en una escala numérica en la cual el valor de cero es arbitrario pero la diferencia entre valores es importante.

De todos los cuatro niveles de medida, sólo la escala de **razón** se basa en un sistema numérico en el cual el cero es significativo. Por tanto, las operaciones de multiplicación y división también toman una interpretación racional. Una escala de razón se utiliza para medir muchos tipos de datos que se encuentran en el análisis empresarial. Variables tales como costos, rentabilidad y niveles de inventario se expresan como medidas de razón. El valor de cero dólares para medir los ingresos, puede interpretarse de manera lógica para explicar que no ha habido ventas. Además, una firma con una participación en el mercado del 40% tiene dos veces más participación que una firma con participación en el mercado del 20%. Las medidas tales como peso, tiempo y distancia también se miden en una escala de razón, ya que cero es significativo y un artículo que pesa 100 libras tiene la mitad del peso de un artículo que pesa 200 libras.

Medidas en escala de razón Medidas numéricas en las cuales cero es un valor fijo en cualquier escala y la diferencia entre valores es importante.

Vale la pena destacar que las cuatro escalas de medida incrementan en complejidad, progresando desde la medida en escala nominal rudimentaria, a una medida en escala de razón más refinada. Cada medida ofrece más información sobre la variable que la anterior. Esta distinción entre los diversos grados de refinamiento es importante, debido a que las diferentes técnicas estadísticas requieren de distintas escalas de medida. A pesar de que la mayoría de las pruebas estadísticas requieren medidas en escala de intervalo o de razón, otras pruebas denominadas pruebas no paramétricas (que se analizarán más tarde en este texto), están diseñadas para utilizar datos ordinales o nominales.

Ejercicios del capítulo

1. La directora de producción para la planta de *Ford Motor Company*, en Cleveland, debe informar a su superior sobre el número de días promedio que los empleados de la planta se ausentan del trabajo. Sin embargo, la planta emplea más de dos mil trabajadores, y la directora de producción no tiene tiempo de revisar los registros personales de cada empleado. Como asistente usted debe decidir cómo puede ella obtener la información necesaria. ¿Qué consejo podría darle?
2. Describa en sus propias palabras cómo puede utilizarse la estadística para solucionar problemas en varias disciplinas y ocupaciones.
3. ¿Qué ocupación específica planea seguir después de su grado? Si no está seguro, seleccione el área en la cual está interesado. Utilizando ejemplos específicos analice detalladamente los tipos de problemas que pueden surgir y las decisiones que tendrá que tomar cuando el análisis estadístico pruebe ser de utilidad.
4. ¿De qué forma utilizará los servicios del estadístico profesional en su organización una vez que usted encuentre empleo? ¿Por qué es poco probable escaparse de la necesidad de tener un conocimiento básico en estadística?
5. Describa en sus propios términos la diferencia entre una población y una muestra; entre un parámetro y un estadístico.
6. ¿Cuál es la diferencia entre una variable cuantitativa y una variable cualitativa? Dé ejemplos.
7. Diferencie entre una variable continua y una variable discreta. Dé ejemplos de cada una.
8. Un informe reciente en la revista *Fortune* reveló que los japoneses pronto controlarán hasta un 35% de las ventas de autos en los Estados Unidos; comparado con el 28% de finales de los años 80 está apenas un 8% por encima de lo ocurrido en 1970. ¿Esta información contiene estadística descriptiva, inferencial, o ambas? Explique.

9. ¿Cuál es la diferencia entre la estadística descriptiva y la estadística inferencial? ¿Cuál cree usted que constituye una forma más elevada de análisis estadístico y por qué?

10. ¿En qué usos o funciones se puede aplicar la estadística? ¿Cómo cree usted que pueda utilizarse para solucionar problemas comerciales en el mundo real? Dé ejemplos de problemas específicos que puedan surgir y explique cómo podría utilizarse la estadística para desarrollar soluciones y respuestas.

11. Seleccione una población cualquiera que sea de su interés. Identifique variables cuantitativas y cualitativas de esa población que puedan seleccionarse para ser estudiadas.

12. Si los estadísticos están interesados realmente en poblaciones, ¿por qué generalmente trabajan con muestras?

13. Analice si las siguientes variables son discretas o continuas:

a. Número de cursos que los estudiantes de su colegio están cursando este semestre.

b. Número de pases atrapados por el beisbolista Tim Brown, receptor de los LA Raiders.

c. Peso de los compañeros de equipo de Tim Brown.

d. Peso del contenido de las cajas de cereal.

e. Número de libros que usted leyó el año pasado.

14. Defina el error de muestreo y explique qué lo causa.

15. La revista *Forbes* (febrero de 1997) reportó datos sobre las condiciones y estilos de vida en varias ciudades de Estados Unidos. Algunos de esos datos se reproducen aquí.

Ciudad	Población en millones	Mediana del ingreso por hogar	Mejor negocio hotelero	Atracción más visitada	Tasa de criminalidad por cada 100,000
Atlanta	3.5	US$43,249	Ritz-Carlton Buckhead	Stone Mountain Park	846.2
Baltimore	2.5	43,291	Harbor Count	Harborplace	1,296.5
St. Louis	2.5	39,079	Hyatt Regency	Gateway Arch	263.4
Philadelphia	5.0	43,576	Bellevue	Liberty Bell	693.1
Raleigh-Durham	1.0	40,990	Radisson Plaza	State Fair	634.9

a. Identifique las variables cualitativas y cuantitativas.

b. ¿Cuáles variables son discretas y cuáles continuas?

c. Identifique cada variable como nominal, ordinal, o de razón.

d. ¿Cuáles son descriptivas y cuáles inferenciales?

16. El presidente de una fraternidad en el campus desea tomar una muestra de las opiniones de 112 miembros respecto a las actividades urgentes para el otoño.

a. ¿Cuál es la población?

b. ¿Cuál es la mejor forma como debe tomarse una muestra?

17. El productor de televisión del programa de drama diurno *All My Children* va a tomar una muestra entre los televidentes, para saber sus sentimientos sobre los planes de asesinar a un personaje popular. ¿Cuáles problemas pueden surgir en este esfuerzo? ¿Qué recomendaría usted y por qué?

18. La General Mills está preocupada con el peso del contenido neto de las cajas de cereal Cheerios que sale de la línea de producción de su planta en Detroit. La caja anuncia 36 onzas, y si en realidad hay menos contenido en la caja, a General Mills la podrían multar por falsa publicidad. Como miembro del equipo administrativo recién contratado en General Mills usted sugiere que se abra una muestra de las cajas y se pese su contenido. El vicepresidente de la división de control de calidad pregunta qué tipo de muestra debería tomarse. ¿Qué le respondería?

19. Debido a que la producción se ha reducido en la planta de General Mills, ¿es posible en realidad abrir todas las cajas producidas durante el período de producción más reciente? Dicho proceso evitaría el error de muestreo y produciría resultados más exactos. Ya que la población de todas las cajas no es muy grande, ¿es necesario el muestreo?

20. ¿Qué nivel de medición utilizaría usted en cada uno de los siguientes casos? Explique su respuesta.

a. Un sistema para medir las preferencias de los clientes respecto a los vehículos con base en su estilo (como convertible, van, camión, sedan, etc.).

b. Un sistema para evaluar a los empleados con base en el número de días que faltan al trabajo.

c. Un sistema para identificar las ciudades de nacimiento de los clientes.

d. Un sistema para registrar la población de las ciudades en las cuales viven los clientes.

21. ¿En cuál escala de medida puede expresarse cada una de estas variables? Explique su respuesta.

a. Los estudiantes clasifican a su profesor de estadística sobre una escala de "Terrible", "No tan Malo", "Bueno", "Maravilloso", y "Dios Griego".

b. Los estudiantes en una universidad están clasificados por profesión, tales como marketing, administración y contaduría.

c. Los estudiantes están clasificados por cursos utilizando los valores 1, 2, 3, 4, y 5.

d. Agrupar mediciones de líquidos en octavo, cuarto y galón.

e. Edades de los clientes.

22. Cite varios ejemplos de comerciales de radio o televisión que utilicen la estadística para vender sus productos. ¿Utilizan estadística descriptiva o inferencial? ¿Qué escala de medida utilizan?

PUESTA EN ESCENA

El Manual sobre Perspectivas Ocupacionales (*Occupational Outlook Handbook*, edición 1996 - 1997), publicado por el Ministerio de Trabajo de Estados Unidos, se encuentra disponible en varias bibliotecas. Contiene información extensa sobre los trabajos en cerca de 250 categorías ocupacionales que abarcan casi el 90% de los trabajos en la economía de los Estados Unidos. El manual proporciona descripciones detalladas sobre las responsabilidades laborales, las condiciones de trabajo, la compensación, la capacitación y las proyecciones de crecimiento laboral para el período comprendido entre 1995 y 2005.

Se proyecta que las vacantes para especialistas en informática y analistas de sistemas serán de cientos de miles en los años venideros, tan sólo para cargos de sustitución. Se espera que el crecimiento laboral sea el más alto hasta el año 2005. Los analistas gerenciales y consultores son otro grupo laboral que requiere de fuertes habilidades cuantitativas. En este grupo se proyecta que el crecimiento laboral sea superior al promedio. Lo mismo se aplica para los economistas y los analistas en investigación de mercados.

En el *Occupational Outlook Handbook* en la universidad o en la biblioteca pública, se buscan trabajos en las áreas de interés. ¿Cuáles son las necesidades de habilidades cuantitativas para éste? ¿Cuáles son las proyecciones de crecimiento laboral?

Del escenario a la vida real

Existe una cantidad de información laboral y de la carrera que se encuentra disponible en el Internet. Los dos sitios líder en listados laborales son Monster Board (*www.monster.com*) y America's Job Bank (*www.ajb.dni.us*). Utilizando las palabras clave de búsqueda, ensaye su interés ocupacional en las áreas de búsqueda de trabajo en estos dos sitios (las palabras clave pueden obtenerse del *Occupational Outlook Handbook* en la biblioteca). ¿A nivel nacional, o "estatal", cuántas vacantes hay en el campo de su interés? ¿Cómo se distribuyen geográficamente estos trabajos? ¿Qué información sobre salarios se proporciona en los destinos laborales?

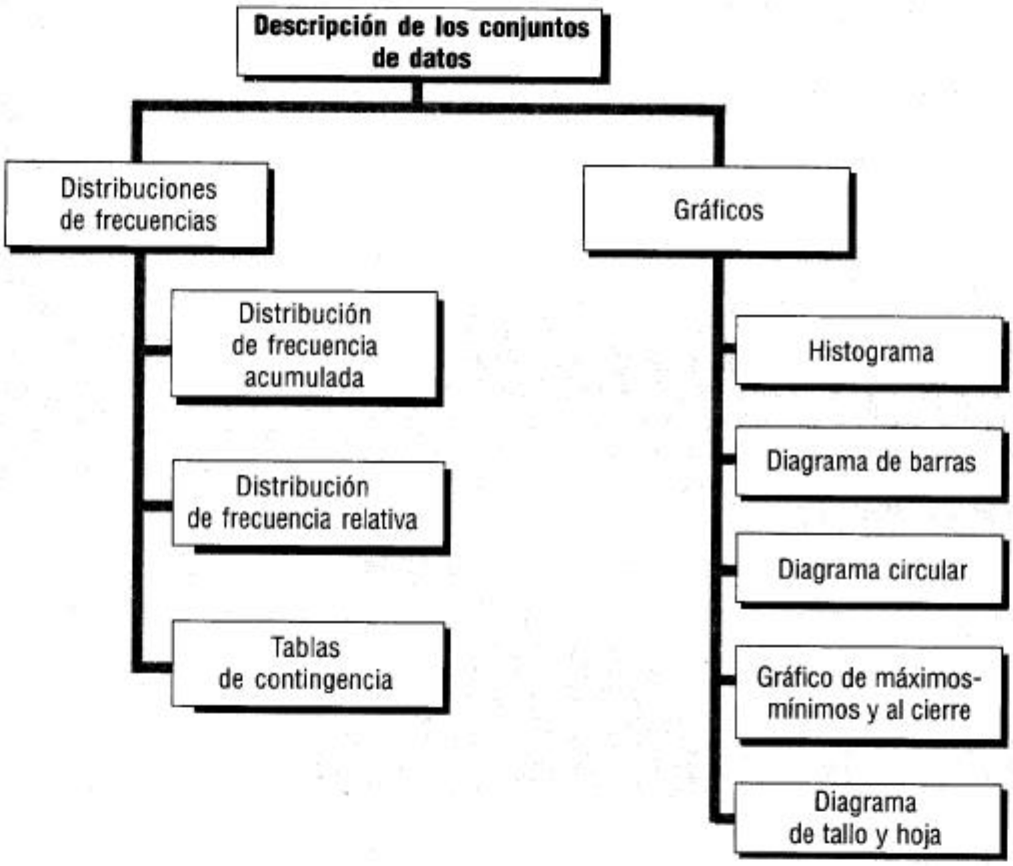
Explore las categorías de "Empleo" y "Trabajos" suministradas por los principales motores de búsqueda como Yahoo! para ponerse al corriente con otros servicios de la carrera disponibles en Internet. En especial, busque los recursos para la planeación de la carrera, hojas de vida e información salarial.

2

Descripción de los conjuntos de datos

Plan del capítulo

Este capítulo presenta ilustraciones y ejemplos de las múltiples formas en que se puede organizar y manejar un conjunto grande de datos para proporcionar una interpretación visual rápida de la información que éste contiene. Dichas herramientas estadísticas permiten describir un conjunto de datos originales de manera concisa y fácil de leer.



ESCENARIO

Los estudiantes de college que planean tomar las vacaciones de primavera del año, se enfrentan con las tarifas de alquiler de autos más altas que se hayan presentado (*Newsweek*, febrero 1997). Una encuesta realizada por *Business Travel News* reveló que en 1996 las tasas aumentaron 11.8%, y se anticipa que el incremento para 1997 estará por encima de dicho porcentaje. Los incrementos en los costos laborales, seguros y costos de parqueo, se mencionan como las principales causas de estos incrementos poco comunes en los gastos para viajes.

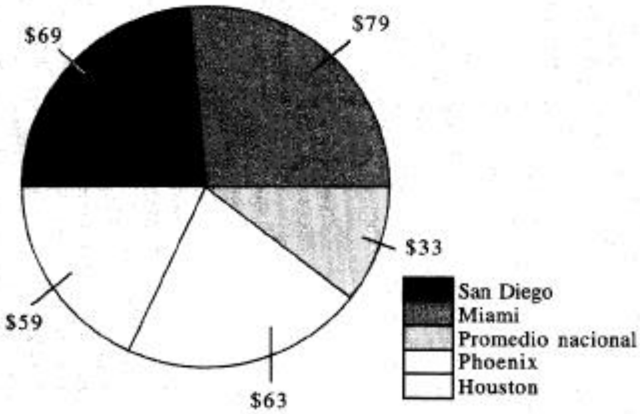
Lo más alarmante para los estudiantes que desean escaparse de los rigores de la academia es el hecho de que se esperan incrementos mucho mayores en las tasas, en las ciudades que parecen ser los lugares favoritos para pasar vacaciones, incluyendo Miami, Houston, Phoenix, las ciudades de California del sur, y otros lugares con clima cálido. Como lo muestra el diagrama circular, estas áreas populares ya exceden el promedio nacional en un monto substancial. Mientras que los viajeros de toda la nación pagan un promedio de aproximadamente $33 dólares por día, las tasas en estas áreas de mayor preferencia se aproximan a $80 dólares.

Para empeorar las cosas, para las personas que toman sus vacaciones este verano, la disponibilidad puede ser un problema. Las empresas de alquiler de autos hacen que quienes tienen plan de viajar reserven con anticipación. Pobre del estudiante quien, con sueños de pasear por Miami en un convertible deportivo, deba conformarse con un sedan cuatro puertas!

Sin embargo, muchas empresas que alquilan autos pretenden multar con un jugoso incremento en las tarjetas de crédito a quienes han reservado un carro pero no se presentan a recogerlo a la hora acordada.

Al preparar las tan esperadas y bien merecidas vacaciones de primavera se sugiere contactar a varias compañías que alquilen autos y recolectar información sobre las tarifas y tipo de autos. Para ayudar a tomar esta importante decisión respecto a qué tipo de auto alquilar, se debe preparar un resumen de estos detalles importantes para que los compañeros de viaje complementen con gráficos y otros medios de descripción de conjuntos de datos analizados en este capítulo.

Tarifas diarias de alquiler de autos



2.1 Introducción

Casi todos los trabajos que se hacen en estadística comienzan con el proceso de recolección de datos necesarios para formar con ellos un conjunto que se utilizará en el estudio. Para propósitos generales, se adoptará la suposición conveniente de que esta labor, con frecuencia tediosa, ya ha sido realizada y que los datos están disponibles.

Esta recolección de datos originales revela muy poco por sí sola. Es extremadamente difícil determinar el verdadero significado de un grupo de números que simplemente se han registrado en un papel. Nuestra labor es organizar y describir tales datos de manera concisa y significativa. Para determinar su significancia, los datos se organizan de manera que, con un simple vistazo, se pueda tener una idea de lo que pueden decirnos.

Entre las herramientas estadísticas que resultan de particular utilidad para organizar los datos se incluyen:

- Tablas de frecuencia que colocan todos los datos en clases específicas.
- Diversos gráficos que pueden proporcionar una representación visual de los datos.
- Tablas de contingencia y diagramas de "tallo y hoja", los cuales también permiten la presentación de un conjunto grande de datos de manera concisa y discernible.

2.2 Métodos de agrupación de datos

Pueden utilizarse varias herramientas básicas para describir y resumir un conjunto grande de datos. La manera más simple, pero quizás la más significativa, es la **serie ordenada**. Se asume que los puntajes de CI de cinco recién graduados de la Universidad de Podunk son 75, 73, 91, 83 y 80. Una serie ordenada simplemente enumera tales observaciones en orden ascendente o descendente. Los cinco valores pueden aparecer como 73, 75, 80, 83, 91. La serie ordenada proporciona alguna agrupación al conjunto de datos; por ejemplo, se puede ver de inmediato que los dos valores extremos son 73 y 91. Sin embargo, la utilidad de una serie ordenada es limitado. Se necesitan mejores técnicas para describir nuestro conjunto de datos. El resto de esta sección analiza algunos métodos comunes para agrupar una recolección de datos amplia, facilitando así la comprensión total de la información que contienen.

A. Distribución de frecuencias

Como estadístico residente de Pigs and People (P&P) Airlines, el director de la división de análisis estadístico le pide recolectar y agrupar los datos sobre el número de pasajeros que han decidido viajar con P&P. Tales datos correspondientes a los últimos 50 días aparecen en la tabla 2.1. Sin embargo, con estos datos en bruto, es improbable que el director pueda obtener información útil y significativa respecto a las operaciones de vuelo. Los datos no están organizados y es difícil llegar a una conclusión significativa simplemente revisando una serie de números anotados en un papel. Es preciso agrupar y presentar los datos de manera concisa y reveladora para facilitar el acceso a la información que contienen. Primero se analizará cómo puede utilizarse una distribución de frecuencia para organizar el conjunto de datos.

Tabla 2.1
Datos brutos sobre el número de pasajeros de P&P Airlines

68	71	77	83	79
72	74	57	67	69
50	60	70	66	76
70	84	59	75	94
65	72	85	79	71
83	84	74	82	97
77	73	78	93	95
78	81	79	90	83
80	84	91	101	86
93	92	102	80	69

Una **distribución de frecuencias** (o tabla de frecuencias) ordenará los datos si estos se dividen en clases y se registrará el número de observaciones en cada clase, tal y como se muestra en la tabla 2.2. Por ejemplo, ahora se puede ver fácilmente, que en 18 de los 50 días, entre 70 y 79 pasajeros volaron en P&P. En ningún momento la lista de pasajeros diarios excedió de 109. La aerolínea rara vez transportó menos de 60 pasajeros. El director puede ahora detectar características que no eran evidentes en el análisis de datos en bruto de la tabla 2.1. Estas características son útiles para la toma de decisiones inteligentes y bien informadas respecto a las operaciones de vuelo.

Tabla 2.2
Distribución de frecuencia para pasajeros de aerolíneas P&P

Clase (pasajeros)	Cuenta	Frecuencia (días)	Punto medio (M)
50 a 59	III	3	54.5
60 a 69	IIIII II	7	64.5
70 a 79	IIIII IIIII IIIII III	**18**	74.5
80 a 89	IIIII IIIII II	12	84.5
90 a 99	IIIII III	8	94.5
100 a 109	II	2	104.5
		50	

Observe que cada clase tiene un **límite inferior** y un **límite superior**. Los valores exactos de estos límites son muy importantes. Si los datos en una tabla de frecuencias son continuos, es necesario permitir valores fraccionarios. Nuestros límites de clase aparecerían como:

50 y menos de 60
60 y menos de 70
70 y menos de 80
.
.
.

Claro que P&P no puede transportar una fracción de pasajero, de manera que la naturaleza discreta del conjunto de datos reales permite el uso de los límites vistos en la tabla 2.2. El **número de clases** en una tabla de frecuencias es algo arbitrario. En general, la tabla debería tener entre 5 y 20 clases. Muy pocas clases no revelarían ningún detalle sobre los datos y demasiadas clases sería tan confuso como la misma lista de datos originales.

Se puede seguir una regla simple para aproximar el número de clases a utilizar, c, es:

Determina el número de clases
$$2^c \geq n \tag{2.1}$$

en donde n es el número de observaciones. El número de clases es la menor potencia a la cual se eleva 2, de manera que el resultado sea igual a o mayor que el número de observaciones. En el ejemplo de P&P, se tiene que $n = 50$ observaciones. Así,

$$2^c \geq 50$$

Despejando c, lo cual puede hacerse fácilmente con una calculadora manual, se encuentra que $2^6 = 64$. Esta regla sugiere que debería haber seis clases en la tabla de frecuencias. Por razones de conveniencia, puede utilizarse un número mayor o menor de clases.

El **punto medio de la clase**, M, se calcula como promedio de los límites superior e inferior de dicha clase. El punto medio para la primera clase en la tabla 2.2 es $50 + 59/2 = 54.5$.

El **intervalo de clase** es el rango de valores encontrados dentro de una clase. Se determina restando el límite superior (o inferior) de una clase del límite inferior (o superior) de la clase siguiente. El intervalo para la primera clase en la tabla 2.2 es $(60 - 50) = 10$. Es deseable que todos los intervalos de clase sean de igual tamaño, ya que facilita las interpretaciones estadísticas en usos subsiguientes. Sin embargo, puede ser conveniente utilizar **intervalos abiertos** que no mencionan un límite inferior para la primera clase o un límite superior para la última clase. La última clase en la tabla 2.2 puede leerse "100 y más".

En la construcción original de una tabla de frecuencia, el intervalo de clase puede determinarse como:

Intervalo de clase para una tabla de frecuencia

$$IC = \frac{\text{Valor más grande} - \text{valor más pequeño}}{\text{Número deseado de clases}} \qquad [2.2]$$

Ya que se deciden seis clases para la tabla de frecuencia, el intervalo de clase se convierte en:

$$IC = \frac{102 - 50}{5} = 8.7$$

Debido a que 8.7 es un número poco práctico, el intervalo puede ajustarse levemente hacia arriba o hacia abajo. Por razones de conveniencia, se seleccionó el intervalo de 10 para formar la tabla 2.2.

Con frecuencia se desea determinar el número de observaciones que son "mayor que" o "menor que" alguna cantidad. Esto puede lograrse con una distribución de **frecuencia acumulada "más de"** o una distribución de **frecuencia acumulada "menor que"**. Una distribución de frecuencias acumuladas "más de" se forma restando las frecuencias de las clases anteriores tal como se observa en la tabla 2.3. En los 50 días, por lo menos 50 pasajeros abordaron P&P Airlines. Por tanto, la frecuencia acumulada para la primera clase de la tabla 2.3 es 50. En tres de esos días, menos de 60 pasajeros compraron pasajes. Por tanto, la frecuencia acumulada de la segunda clase es 47 (50 – 3). Debido a que el número de pasajeros fue menor de 70 en 10 días, la frecuencia acumulada para la tercera clase es 40 (50 – 10). Las frecuencias acumuladas para las clases restantes se determinan de forma similar.

Tabla 2.3
Distribución de frecuencias acumuladas "más de" del número de pasajeros

Clase (pasajeros)	Frecuencia (días)	Frecuencia acumulativa (días)
50 o más	3	50
60 o más	7	47
70 o más	18	40
80 o más	12	22
90 o más	8	10
100 o más	2	2
110 o más	0	0

Una distribución de frecuencia acumulada "menor que" se construye sumando las frecuencias de cada clase. La tabla 2.4 muestra que en ningún momento volaron menos de 50 pasajeros en P&P. La frecuencia acumulada de la primera clase por tanto es cero. En tres días, menos de 60 pasajeros abordaron vuelos. La frecuencia acumulada de la segunda clase es tres. Debido a que hubo 10 días en los cuales menos de 70 pasajeros volaron, la tabla 2.4 muestra que la frecuencia acumulada de la tercera clase es 3 + 7 = 10. De nuevo las frecuencias acumuladas de las clases restantes se determinan de manera similar.

Tabla 2.4
Distribución de frecuencia acumulada "menor que" para el número de pasajeros

Clase (pasajeros)	Frecuencia acumulada (días)	Frecuencia (días)
Menos de 50	0	0
Menos de 60	3	3
Menos de 70	7	10
Menos de 80	18	28
Menos de 90	12	40
Menos de 100	8	48
Menos de 110	2	50

Finalmente, una **distribución de frecuencia relativa** expresa la frecuencia dentro de una clase como un porcentaje del número total de observaciones. En nuestro caso actual, la frecuencia relativa de una clase se determina como la frecuencia de dicha clase dividida por 50. Por ejemplo, la tabla 2.5 muestra que la frecuencia relativa de la tercera clase es 18/50 = 36%. Esto permite sacar conclusiones sobre el número de observaciones en una clase respecto a toda la muestra.

Tabla 2.5
Distribución de frecuencia relativa para pasajeros

Clase (pasajeros)	Frecuencia (días)	Frecuencia relativa
50-59	3	3 ÷ 50 = 6%
60-69	7	7 ÷ 50 = 14%
70-79	18	18 ÷ 50 = **36%**
80-89	12	12 ÷ 50 = 24%
90-99	8	8 ÷ 50 = 16%
100-109	2	2 ÷ 50 = 4%
	50	100%

Casi todo el trabajo estadístico que se encuentra puede realizarse fácil y rápidamente con la ayuda de los computadores modernos. La pantalla 2.1 contiene la tabla de frecuencias para los datos de los pasajeros de P&P, creado en Microsoft Excel. Se puede especificar el límite superior de cada clase. Después, las observaciones se colocan en las clases correspondientes. La mayoría de los paquetes de computador crearán un resultado similar.

Pantalla 2.1
Tabla de frecuencias para datos de pasajeros de P&P

BIN	FREQUENCY	CUMULATIVE %
59	3	6.00%
69	7	20.00%
79	18	56.00%
89	12	80.00%
99	8	96.00%
109	2	100.00%

BIN = Límite superior
FREQUENCY = Frecuencia
CUMULATIVE = Acumulado

B. Tablas de contingencia

Las tablas de frecuencia pueden organizar datos de sólo una variable a la vez. Si se desea examinar o comparar dos variables, una **tabla de contingencia** resulta de mucha utilidad.

Se supone que además de recolectar información sobre el número de pasajeros de P&P, también se obtuvieron datos sobre las edades de los pasajeros y el número de vuelos en los que se registraron cada año. Ambas variables pueden verse en detalle mediante una tabla de contingencia que enumera el número de vuelos en la parte superior y las edades debajo, tal como se muestra en la tabla 2.6. Al dividir la edad en cuatro categorías y los vuelos en tres categorías, se han creado 12 celdas en la tabla. Ahora se puede determinar si estas dos características están relacionadas. Por ejemplo, en la última columna se puede ver que el número más grande de pasajeros, 15 o 30%, está en la categoría de edad comprendida entre los 40 y los 65 años y vuelan más de cinco veces al año. El número más pequeño de pasajeros, tan sólo 4 personas, son menores de 25 años o tienen 65 años o más. Una mayoría, 28 personas, toma de forma típica más de 5 vuelos cada año.

Tabla 2.6

Tabla de contingencia para P&P Airlines

	Número de vuelos por año			
Edad	**1-2**	**3-5**	**Mayor de 5**	**Total**
Menor de 25	1 (0.02)	1 (0.02)	2 (0.04)	**4 (0.08)**
25-40	2 (0.04)	8 (0.16)	10 (0.20)	20 (0.40)
40-65	1 (0.02)	6 (0.12)	**15 (0.30)**	22 (0.44)
65 y más	1 (0.02)	2 (0.04)	1 (0.02)	**4 (0.08)**
Total	5 (0.10)	17 (0.34)	**28 (0.56)**	50 (1.00)

También se podría mostrar la información en cada celda según el género. De las 15 personas en el nivel de edad entre los 40 y los 65 años, que vuelan más de 5 veces al año, ocho pueden ser hombres y siete mujeres. Entonces la entrada en dicha celda podría aparecer como 8,7 (0.30). Todas las otras características relevantes de los pasajeros también podrían incorporarse en la tabla. Es fácil imaginar que dicho perfil descriptivo podría ser muy útil en la identificación del pasajero típico y en el establecimiento de políticas de marketing efectivas.

Ejercicios de la sección

1. Un conjunto de datos contiene 100 observaciones; la más grande es 315 y la más pequeña es 56.

a. ¿Cuántas clases debería tener la tabla de frecuencias?

b. ¿Cuál es el intervalo de clase?

c. ¿Cuáles son los límites y puntos medios de cada clase?

2. En un estudio reciente sobre 500 graduados en administración de negocios, el salario inicial más alto que se reportó fue de $27,500 dólares y el más bajo fue de $19,900 dólares. Usted desea crear la tabla de frecuencias para analizar y comparar estos datos con las ofertas de trabajo que usted ha recibido.

a. ¿Cuántas clases pondrá en su tabla de frecuencia?

b. ¿Cuál es el intervalo de clase?

c. ¿Cuáles son los límites y puntos medios de cada clase?

3. Los siguientes datos son los ingresos de 60 ejecutivos de marketing para empresas de Estados Unidos. Los datos están expresados en miles de dólares.

58	76	89	45	67	34
64	76	34	65	45	39
79	74	56	71	85	87
74	38	69	79	61	71
69	62	56	38	69	79
71	54	31	69	62	39
65	79	47	46	77	66
55	75	62	57	77	36
73	72	64	69	51	50
40	50	74	61	69	73

a. Construya una tabla de frecuencia para los datos. Tenga mucho cuidado en la selección de sus intervalos de clase. Muestre las frecuencias acumulativas y relativas para cada clase. ¿Qué conclusión puede sacar de la tabla?

b. Presente y explique una distribución de frecuencia acumulada "más que" y una distribución de frecuencia acumulada "menor que".

4. De los datos que se presentan a continuación, prepare una tabla de contingencia para evaluar a 45 empleados respecto a su nivel de educación expresado en años y el nivel administrativo que poseen actualmente. Divida la educación en tres grupos: el grupo 1 entre los 10 y los 12 años de educación, el grupo 2 entre los 13 y los 15 años, y el grupo 3 para 16 años y más. ¿Qué patrones, si los hay, observa y a qué conclusiones puede llegar?

Nivel gerencial	Años de educación	Nivel gerencial	Años de educación
1	14	4	16
2	13	4	18
3	16	4	14
2	16	2	15
1	12	3	17
4	16	2	12
1	12	1	12
2	12	2	15
3	14	3	16
3	14	1	10
1	13	2	14
2	12	4	16
3	20	2	14
4	17	4	16
2	14	1	10
1	13	1	12
3	16	4	13
2	11	1	10
4	16	2	13
4	16	4	17
2	10	2	15
3	11	3	14
1	14		

5. Un proceso para producción de materiales de construcción está diseñado para generar contenedores de tres pesos diferentes: 10 libras, 11 libras y 12 libras. Un examen de 40 de los contenedores registra sus pesos reales y sus pesos deseados. Se considera que un contenedor es "defectuoso" si su peso real difiere en más de 0.5

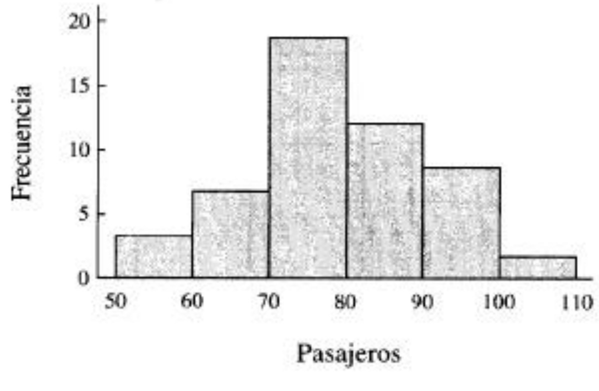
libras de su peso deseado. Realice una tabla de contingencia con estos datos en la cual indique cuántos contenedores, en cada uno de los tres grupos, están dentro de la diferencia permisible. Registre las observaciones como "1" si es defectuoso y "2" si no es defectuoso. ¿Puede detectar algún patrón? ¿Parece que un grupo de peso tiene una mayor proporción de defectos?

Peso real	Peso deseado	Peso real	Peso deseado
9.5	10	12.3	11
9.6	10	10.4	12
12.1	11	12.1	10
11.2	12	10.0	11
11.6	11	11.2	10
12.3	12	9.9	12
9.6	10	9.6	11
10.6	12	12.4	10
11.0	11	11.2	12
11.2	10	11.6	11
9.8	11	12.3	10
10.5	10	9.6	12
11.9	12	10.6	12
11.0	10	11.2	11
9.8	10	10.5	12
11.9	10	12.3	10
10.4	12	12.1	11
10.0	12	11.2	10
9.9	12	9.6	11
11.5	10	9.5	12

2.3 Gráficos

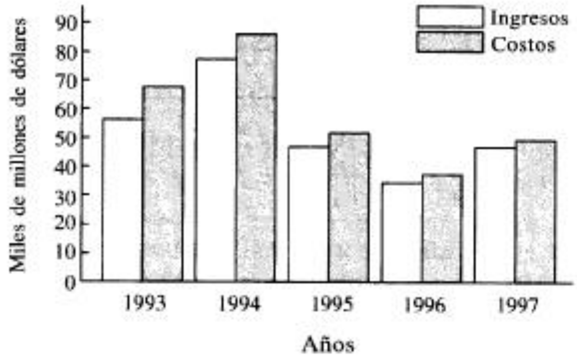
Los gráficos también son métodos útiles para describir conjuntos de datos. Un **histograma** coloca las clases de una distribución de frecuencia en el eje horizontal y las frecuencias en el eje vertical. La figura 2.1 muestra el histograma para la distribución de frecuencias de la tabla 2.2. Revela detalles y patrones que no se pueden discernir fácilmente de los datos originales. La frecuencia absoluta, así como las frecuencias relativas de cada clase, están ilustradas claramente.

Figura 2.1 Histograma para los pasajeros de P&P



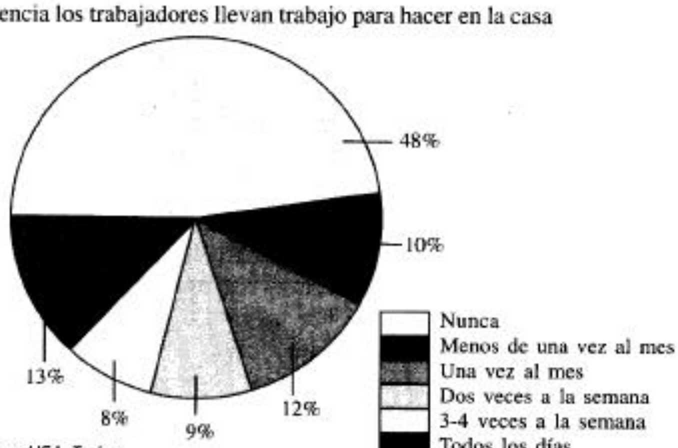
Parecido a un histograma, el **diagrama de barras** puede mostrar cantidades o porcentajes para dos o más valores sobre el eje vertical. La figura 2.2 demuestra los costos e ingresos para P&P Airlines.

Figura 2.2
Desempeño de P&P



Un **diagrama circular** es de especial utilidad para mostrar proporciones (porcentajes) relativas de una variable. Se crea marcando una porción del círculo correspondiente a cada categoría de la variable. La figura 2.3 muestra en porcentajes con qué frecuencia los empleados llevan trabajo de la oficina a la casa.

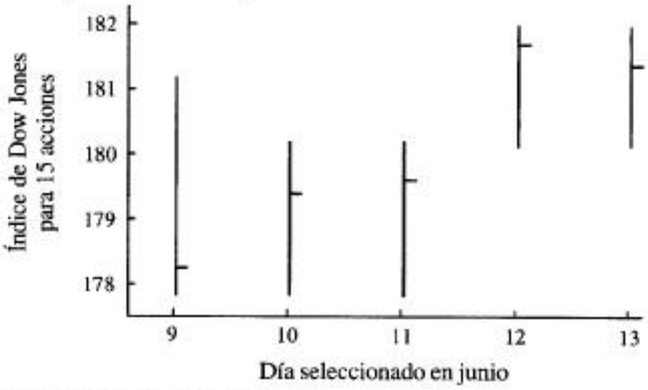
Figura 2.3
Diagrama circular



Fuente: *USA Today.*

Los datos financieros con frecuencia están contenidos en un **gráfico de máximos - mínimos y al cierre**. Como su nombre lo indica, muestra el valor más alto, el valor más bajo y el valor al cierre de los instrumentos financieros como por ejemplo las acciones. La figura 2.4 es un gráfico basado en los datos tomados de *The Wall Street*

Figura 2.4
Gráfico de máximos-mínimos y cierre para 15 acciones



Fuente: *The Wall Street Journal.*

Journal para el índice Dow Jones respecto a 15 acciones durante un período de cinco días con base en los siguientes datos:

	Máximos	Mínimos	Cierre
Junio 9	181.07	178.17	178.88
10	180.65	178.28	179.11
11	180.24	178.17	179.35
12	182.79	179.82	181.37
13	182.14	179.53	181.31

Algunas veces se denominan máximos y mínimos al extremo superior de la recta vertical, o máximo, que marca el valor más alto que alcanzó el índice en ese día; el extremo inferior indica el valor más bajo que alcanzó ese día. El valor de cierre se muestra mediante la línea pequeña que se encuentra entre las dos anteriores. Podrían hacerse presentaciones similares para productos y monedas transadas en las bolsas organizadas del mundo.

John Tukey, un estadístico notable, creó el **diagrama de tallo y hoja** como una alternativa para que el histograma proporcionara una impresión visual rápida sobre el número de observaciones en cada clase. Cada observación se divide en dos partes: un tallo y una hoja, separadas por una recta vertical. El diseño preciso puede adaptarse para que se ajuste a cualquier conjunto de datos, identificando un punto conveniente en el cual las observaciones puedan separarse de tal manera que formen un tallo y una hoja. Los valores fraccionarios tales como 34.5, 34.6, 45.7, 45.8 y 56.2 pueden separarse en el punto decimal produciendo un tallo y una hoja de la siguiente manera:

Tallo	Hoja
34	5.6
45	7.8
56	2

Vale la pena destacar que el tallo y la hoja están colocados en series ordenadas.

Si un tallo único contiene un gran número de observaciones en su hoja, es común dividirlo en dos tallos independientes, exactamente en el punto medio. La pantalla 2.2 muestra un diagrama de tallo y hoja para nuestros datos sobre pasajeros, proporcionado por Minitab. La pantalla contiene tres columnas. La segunda y la tercera muestran el tallo y la hoja, respectivamente. Hay tres observaciones en los cincuentas: 50, 57 y 59. La primera columna muestra las profundidades, indicando la suma total de observaciones desde la parte superior de la distribución, para los valores menores que la mediana (lo cual se discutirá más adelante en este capítulo) o hasta la parte inferior de las distribuciones, para valores mayores que la mediana. La profundidad en paréntesis, (9), muestra el número de observaciones en el tallo que contiene la mediana. Por ejemplo, hay 19 observaciones desde 50 hasta 74, y 22 observaciones desde 80 hasta la observación máxima de 102. Se nota que hay dos tallos para los setentas separando las observaciones en el punto medio entre 74 y 75.

Pantalla 2.2
Stem-and-leaf display (Tallo y hoja para P&P)

```
Character Stem-and-Leaf Display
Stem-and-leaf of pass  N = 50
Leaf Unit = 1.0
1    5  :  0
3    5  :  79
4    6  :  0
10   6  :  567899
```

```
Character Stem-and-Leaf Display = Visualización
                                  de caracteres
           Stem-and-leaf of pass = Tallo y hoja
                                  de pasajeros
                       Leaf Unit = Unidad de hoja
```

Continúa

```
19    7  :  001122344
(9)   7  :  567788999
22    8  :  0012333444
12    8  :  56
10    9  :  012334
4     9  :  57
2    10  :  12
```

La *unidad de hoja* expresa en dónde colocar el punto decimal. Con la unidad de hoja = 1.0, la primera observación es 50. La unidad de hoja = 0.1 significaría que la primera observación es 5.0 y la unidad de hoja = 10 significaría que la primera observación es 500.

Ejercicios de la sección

6. Construya un diagrama de tallo y hoja sobre las tasas de desempleo en 15 países industrializados: 5.4%, 4.2%, 4.7%, 5.5%, 3.2%, 4.6%, 5.5%, 6.9%, 6.7%, 3.7%, 4.7%, 6.8%, 6.2%, 3.6% y 4.8%.

7. Desarrolle e interprete un histograma de la tabla de frecuencias que usted construyó para el ejercicio 3.

8. *Investor's Report* (julio de 1996) afirmó que el mes anterior la gente había invertido, en millones de dólares, las siguientes cantidades, en tipos de fondos mutuos: 16.7 en fondos de desarrollo, 12.5 en fondos de ingreso, 28.2 en fondos internacionales, 15.9 en mercado monetario, y 13.9 en "otros". Haga un diagrama circular que represente estos datos, complete con los porcentajes correspondientes.

9. Los cambios del mes anterior para las inversiones hechas en cada uno de los fondos del problema anterior, fueron respectivamente, 2.3, 1.5, -3.6, 4.5 y 2.9. Haga un diagrama de barras que refleje estos cambios.

Problemas resueltos

1. Una organización estudiantil está por revisar el monto que los estudiantes invierten en textos cada semestre. Cincuenta estudiantes reportaron las siguientes cantidades, aproximadas al dólar más próximo.

US$125	US$157	US$113	US$127	US$201
165	145	119	148	158
148	168	117	105	136
136	125	148	108	178
179	191	225	204	104
205	197	119	209	157
209	205	221	178	247
235	217	222	224	187
265	148	165	228	239
245	152	148	115	150

a. Debido a que 2^c ³ 50 produce seis clases, el intervalo para la distribución de frecuencias se encuentra como (más alto – más bajo / 6 = 265 – 104/6 = 26.8. Por conveniencia se utiliza un intervalo de 25. Esto resulta en realidad en siete clases en lugar de las seis propuestas. No hay problema. Si usted fija el límite inferior para la primera clase en 100 (de nuevo, por conveniencia usted podría utilizar 104), entonces la tabla sería:

Intervalo de clase	Frecuencia	M	Frecuencia acumulada	Frecuencia relativa
100-124	8	112	8	0.16
125-149	11	137	19	0.22
150-174	8	162	27	0.16
175-199	6	187	33	0.12
200-224	10	212	43	0.20
225-249	6	237	49	0.12
250-274	1	262	50	0.02

b. Las visualizaciones del histograma y del diagrama del tallo y hoja proporcionados por Minitab son:

```
Histogram of C1          N= 50                    Histogram of C1 = Histograma de C1
Midpoint                 Count                    Midpoint = Punto medio
   112.0                   8     ********         Count = Conteo
   137.0                  11     ***********
   162.0                   8     ********
   187.0                   6     ******
   212.0                  10     **********
   237.0                   6     ******
   262.0                   1     *
```

```
Stem-and-leaf of C1        N = 50        Stem-and-leaf of C1 = Tallo y hoja de C1
Leaf Unit = 1.0                          Leaf Unit = Unidad de hoja

   3     10     458
   8     11     35799
  11     12     557
  13     13     66
  19     14     588888
  24     15     02778
  (3)    16     558
  23     17     889
  20     18     7
  19     19     17
  17     20     145599
  11     21     7
  10     22     12458
   5     23     59
   3     24     57
   1     25
   1     26     5
```

Las tres cantidades más bajas son US$104, US$105, y US$108. Hay ocho estudiantes que pagaron entre US$104 y US$119. Esto corresponde a la frecuencia de la primera clase en la tabla anterior. La siguiente cantidad más alta es US$125. La mediana está en el tallo para US$160.

2. En una escala de 1 a 4, siendo 4 el mejor, un grupo de consumidores clasifica la "conciencia social" de 50 organizaciones clasificadas como públicas (indicadas con un "1" en los siguientes datos), privadas (indicadas con un "2"), o controladas por el gobierno (indicadas con un "3").

Tipo	Clasificación	Tipo	Clasificación
1	1	2	2
2	2	3	3
2	3	1	1
3	2	2	4
1	4	3	4
2	2	1	2
3	3	2	3
2	2	3	2
1	1	1	1
2	2	3	4
3	3	2	2
1	4	1	3
1	2	3	1
2	3	2	4
3	1	3	2
3	2	1	1
2	3	2	3
1	2	3	2
2	1	1	1
3	4	2	4
2	4	1	1
3	1	2	2
1	2	3	3
3	4	1	2
2	1	2	1

a. Prepare una tabla de contingencia completa con totales y porcentajes en columnas y filas. ¿Qué conclusiones puede sacar? Minitab produce la siguiente tabla. Observe que tan sólo el 2 o el 4%, de las 15 organizaciones públicas recibieron una clasificación de "4". Exactamente dos veces el porcentaje de las organizaciones privadas y del gobierno en la clasificación más alta. De las 50 organizaciones, 14% de las que recibieron un "1" eran unidades públicas. De las que recibieron un "1", 53.85% eran públicas. Muchos otros hechos se hacen evidentes al examinar la tabla.

ROWS: type	COLUMNS:Rating				
	1	2	3	4	ALL
(Public) 1	7	5	1	**2**	**15**
	46.67	33.33	6.67	13.33	100.00
	53.85	29.41	10.00	20.00	30.00
	14.00	10.00	2.00	**4.00**	30.00
(Private) 2	3	7	5	4	19
	15.79	36.84	26.32	21.05	100.00
	23.08	41.18	50.00	40.00	38.00
	6.00	14.00	10.00	8.00	38.00

ROWS:type - FILAS:tipo
COLUMNS:Rating = COLUMNAS: Clasificación
ALL = Todos
Public = Públicas
Private = Privadas

Continúa

```
(Government) 3      3       5       4       4      16      Government = Del gobierno
                18.75   31.25   25.00   25.00  100.00      ALL = Todos
                23.08   29.41   40.00   40.00   32.00      CELL CONTENTS = CONTENIDO
                 6.00   10.00    8.00    8.00   32.00                      DE LAS
                                                                           CELDAS

ALL                13      17      10      10      50
                26.00   34.00   20.00   20.00  100.00
               100.00  100.00  100.00  100.00  100.00
                26.00   34.00   20.00   20.00  100.00

CELL CONTENTS --
                COUNT                                   COUNT      = CONTEO
                % OF ROW                                % OF ROW   = % DE FILA
                % OF COL.                               % OF COL.  = % DE COLUMNA
                % OF TBL                                % OF TBL   = % TOTAL
```

Lista de fórmulas

[2.1] $$2^c \geq n$$ Determina el número de clases en una tabla de frecuencias

[2.2] $$IC = \frac{\text{Valor más grande} - \text{valor más pequeño}}{\text{Número deseado de clases}}$$ Intervalo de clase para una tabla de frecuencias

Ejercicios del capítulo

10. Bill Bissey, vicepresidente de Bank One en Indianapolis, controla la aprobación de créditos para el desarrollo de negocios locales. Durante los últimos cinco años el crédito más grande fue de US$1.2 millones, y el más pequeño fue de US$10,000. Él desea crear una tabla de frecuencias con 10 clases. ¿Cuáles serían los límites de las clases? ¿Cuál sería el intervalo de clase?

11. El Sr. Bissey también guarda registros de las cuentas personales de ahorro. De las 40 nuevas cuentas abiertas el mes anterior, los saldos corrientes son:

US$179.80	US$890.00	US$712.10	US$415.00
112.17	1,200.00	293.00	602.02
1,150.00	1,482.00	579.00	1,312.52
100.00	695.15	287.00	1,175.00
1,009.10	952.51	1,112.52	783.00
1,212.43	510.52	1,394.05	1,390.00
470.53	783.00	1,101.00	666.66
780.00	793.10	501.01	1,555.10
352.00	937.01	711.11	1,422.03
1,595.10	217.00	1,202.00	1,273.01

Haga una tabla de frecuencias con siete clases. ¿Está trabajando con datos continuos o datos discretos? Explique.

12. Utilizando los datos del ejercicio 11, realice e interprete una tabla de frecuencias relativa y una tabla de frecuencias acumulada.

13. Utilizando los datos del ejercicio 11, construya un histograma.

14. Utilizando los datos del ejercicio 11, construya un diagrama de barras que muestre los porcentajes en cada clase.

15. Las pérdidas y ganancias para las 50 firmas más grandes (por concepto de ventas) en la lista de las 500 mejores empresas de la revista Fortune para el año de 1992 aparecen a continuación en millones de dólares. El valor más bajo es una pérdida de US$4,453 millones (-4453), y el más alto es una ganancia de US$5,600 millones. Haga una tabla de frecuencias con el número apropiado de clases.

US$-4,453	US$-795	US$423	US$184	US$97
5,600	1,567	454	258	939
-2,258	1,773	709	535	460
-2,827	1,484	-578	1,461	-387
2,636	20	368	601	-404
1,920	-1,021	755	-273	63
3,006	1,080	-732	1,681	308
1,403	17	-617	-142	73
1,294	311	1,154	454	97
1,293	942	-1,086	2,056	505

16. Como consultor económico privado, usted considera necesario leer fielmente *The Wall Street Journal* para estar al día en su campo profesional. Un reporte reciente en *The Wall Street Journal* mostró los siguientes datos sobre el porcentaje de ejecutivos que están en las 42 mejores corporaciones de Estados Unidos que sufren de problemas de abuso de drogas.

5.9	8.8	14.3	8.3	9.1	5.1	15.3
17.5	17.3	15.0	9.3	9.9	7.0	16.7
10.3	11.5	17.0	8.5	7.2	13.7	16.3
12.7	8.7	6.5	6.8	13.4	5.5	15.2
8.4	9.8	7.3	10.0	11.0	13.2	16.3
9.1	12.3	8.5	16.0	10.2	11.7	14.2

a. Construya un diagrama de tallo y hoja.

b. Construya el histograma correspondiente

c. Construya la distribución de frecuencias y halle los puntos medios de clase.

d. Construya la distribución de frecuencias relativa.

17. Los siguientes datos respecto a dónde pasan las vacaciones los norteamericanos, se publicaron recientemente en *Travel and Leisure*: en ciudad 31%; en océano 26%; en lago 5%; en montañas 10%; en parques estatales y nacionales 6%; en pequeñas poblaciones rurales 22%. Como director de la junta de turismo de su estado, es su función presentar estos datos en un diagrama circular.

18. Big Al, el agiotista local, actualmente tiene 120 cuentas pendientes por cobrar. El contador de Big Al le informa que de las 25 cuentas que están en el intervalo comprendido entre US$0 a US$4,999, 10 están vencidas, 5 atrasadas y el resto son morosas, colocando al deudor en peligro de ser visitado por el ejecutor de Big Al. De las 37 cuentas comprendidas entre el intervalo de US$5,000 y US$9,999, 15 vencen, 10 han vencido y el resto son morosas. Hay 39 cuentas en el intervalo comprendido entre US$10,000 y US$14,999 de las cuales 11 vencen, 10 han vencido y el resto son morosas. De las cuentas restantes que están en el intervalo comprendido entre US$15,000 y más, 5 vencen, 7 han vencido y el resto son morosas. Big Al desea ver una tabla de contingencia de estas cuentas. Interprete su significancia citando unas cuantas estadísticas que usted considere que sean las más importantes y reveladoras.

19. Los promedios del índice Dow Jones para 30 acciones industriales reportaron los siguientes valores en el mes de junio de 1997. Haga un diagrama de máximos–mínimos y cierre con base en estos datos.

	Máximos	Mínimos	Al cierre
Junio 2	6119.31	6081.79	6093.35
3	6123.58	6084.82	6101.71
4	6144.15	6084.64	6099.40
5	6148.12	6111.13	6124.47

20. *Newsweek* (agosto 26, 1996) reportó porcentajes de preferencia de varias drogas de los estudiantes de octavo grado que habían utilizado drogas, y el porcentaje de uso de tales drogas cambió desde 1991. Utilice los datos sobre las tasas de 1996 sobre el uso de drogas para crear un diagrama circular y los cambios porcentuales para hacer un diagrama de barras. Comente los resultados.

Droga de preferencia	Tasa de uso en 1996	Cambio en porcentajes desde 1991
Crack	2.7%	108%
Heroína	2.3	92
Marihuana	19.5	95
Cocaína	4.2	83
Cigarrillos	46.0	5
Alcohol	25.3	-5

21. Como proyecto de clase, un estudiante junior de mercadeo hizo una encuesta entre 20 negocios locales en Peoria, Illinois, respecto a su preferencia por un producto nuevo. Sus respuestas se registraron como un "1" si les gustaba el producto, un "2" si les disgustaba el producto, y un "3" si no opinaban. Los niveles de ventas anuales para las tiendas también se registraron así

1	Si las ventas fueron menores que US$50,000
2	Si las ventas fueron de US$50,000 pero menores de US$100,000
3	Si las ventas fueron de US$100,000 pero menores de US$200,000
4	Si fueron de US$200,000 o más

Construya una tabla de contingencia con base en los datos mostrados. ¿A qué conclusiones llega?

Opinión	Ventas	Opinión	Ventas
1.00	4.00	3.00	1.00
1.00	4.00	2.00	1.00
3.00	3.00	3.00	2.00
1.00	4.00	3.00	4.00
3.00	1.00	1.00	4.00
3.00	1.00	1.00	4.00
3.00	1.00	1.00	4.00
1.00	2.00	3.00	4.00
2.00	3.00	2.00	4.00
1.00	4.00	3.00	1.00

Ejercicio en computador

Elizabeth Dunham es la directora ejecutiva de operaciones del personal para Minor Maturity Fashions, una fábrica de ropa cuyo objetivo es la mujer joven. Ella ha recolectado datos sobre las ventas semanales en cientos de dólares durante las 100 últimas semanas y sobre los costos de nómina en miles de dólares para el mismo período. Sin embargo, la señorita Dunham está algo sorprendida por la cantidad de números que tiene a su disposición y necesita ayuda para organizar estos datos en una forma significativa. Ella debe sacar alguna impresión sobre la naturaleza de las ventas y los costos de nómina de manera que pueda obtenerse la información importante para la próxima reunión de junta directiva. También necesita algún consejo sobre cuáles características de estas variables importantes son las más significativas.

Ingrese al archivo MINOR y proporcione a la señorita Dunham todas las distribuciones de frecuencia relevantes y gráficos que su software pueda generar. Usted puede tener que crear a mano algunas de dichas herramientas descriptivas si su software no es lo suficientemente flexible. Prepare un reporte estadístico como el que aparece en el apéndice I. Incluya todas las partes importantes del reporte estadístico incluyendo el resumen ejecutivo y sus conclusiones y recomendaciones.

PUESTA EN ESCENA

En el caso presentado al inicio del capítulo usted debía describir y caracterizar los datos que recolectó sobre la tarifas de alquiler de autos para sus tan esperadas vacaciones de primavera. Con base en esos datos sobre los costos diarios del alquiler provenientes de varias compañías en cada ciudad, prepare un reporte detallado describiendo sus descubrimientos. Incluya tantas herramientas de las que se trataron en este capítulo como le sea posible. Proporcione gráficos que muestren los costos promedios para cada ciudad como un porcentaje del costo promedio de todas las ciudades.

Haga una tabla de contingencia que muestre cada una de las ciudades que aparecen en las columnas. Agrupe los datos sobre tarifas en 3 a 5 grupos para las filas. Incluya los porcentajes de la columna y de la fila y los totales, así como se hizo en el problema resuelto número 2.

Ciudad	Tarifa	Ciudad	Tarifa	Ciudad	Tarifa
Miami	85	Houston	58	San Diego	89
Houston	68	Phoenix	65	New Orleans	79
New Orleans	65	Houston	63	Miami	62
Miami	50	Phoenix	52	New Orleans	96
Miami	102	Houston	87	Miami	98
New Orleans	84	Houston	69	New Orleans	87
Miami	110	Miami	85	Houston	69
Phoenix	85	Phoenix	64	New Orleans	81
Houston	95	New Orleans	97	Houston	62
Phoenix	58	Phoenix	69	Phoenix	59
Houston	59	New Orleans	74	Houston	63
Phoenix	51	Phoenix	87	New Orleans	62
New Orleans	76	Houston	67	New Orleans	52

Del escenario a la vida real

En internet se pueden visitar varias compañías de alquiler de autos y "Guías para compradores" para comparar los precios de alquiler y hacer reservaciones. La siguiente es una selección de tales sitios:

Servicios de compras:

Breeze Net's Guide to Airport Cars *www.bmn.com*

Car Rental Shopper's guide *www.rent-cars.com*

Compañías nacionales de alquiler de autos:

Alamo	*www.goalamo.com*
Avis	*www.avis.com*
Budget	*www.budgetrentacar.com*
Hertz	*www.hertz.com*
Thrifty	*www.thrifty.com*

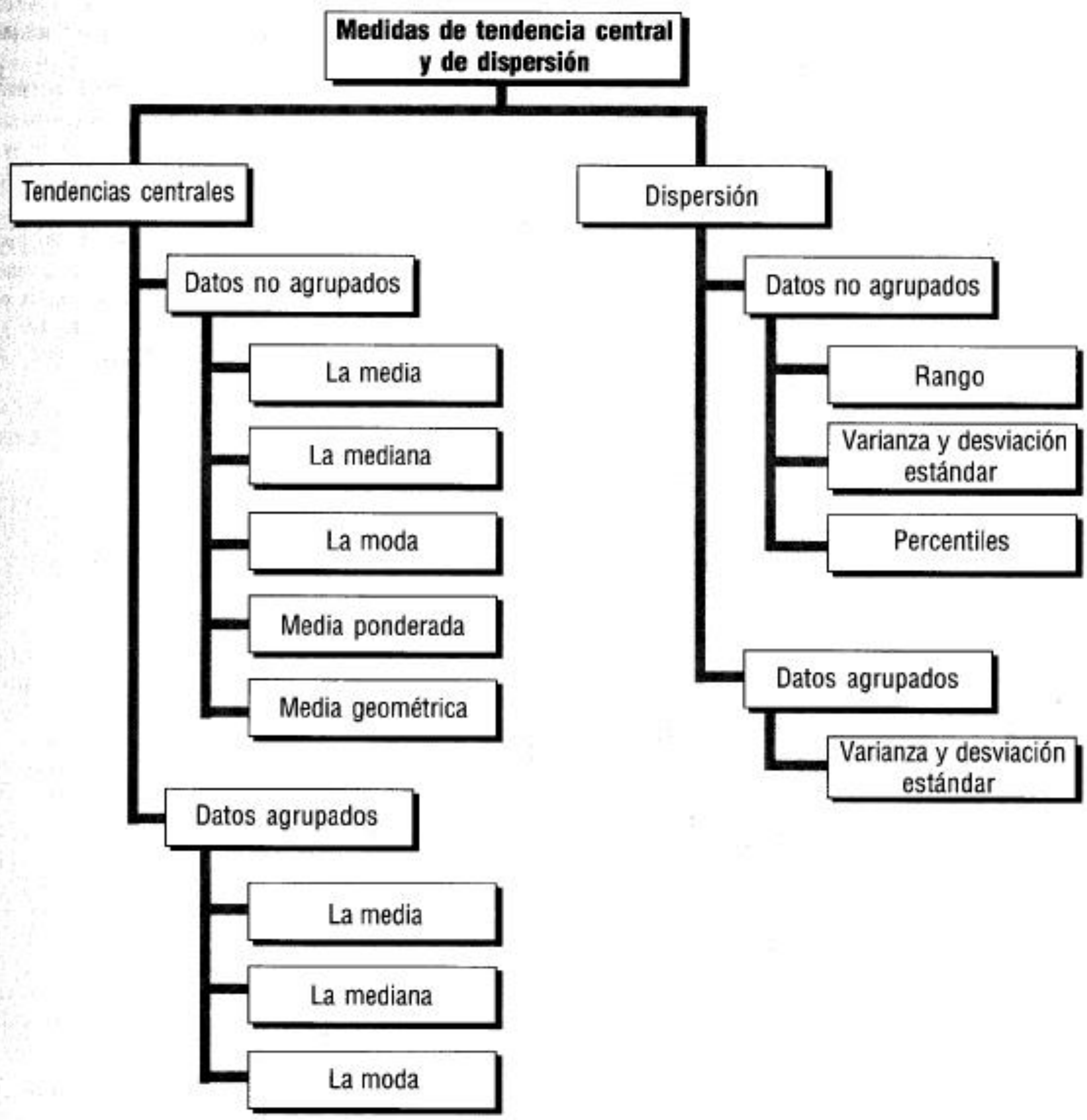
Utilizando la información sobre precios reales obtenida de estos sitios en internet, elabore nuevamente la tabla de contingencia que elaboró en el ejercicio de *Puesta en Escena*. ¿Cuáles de las ciudades de Miami, Houston, Phoenix, o New Orleans tiene el número más grande de compañías de alquiler de autos? ¿Cuál ofrece el rango de precios más amplio?

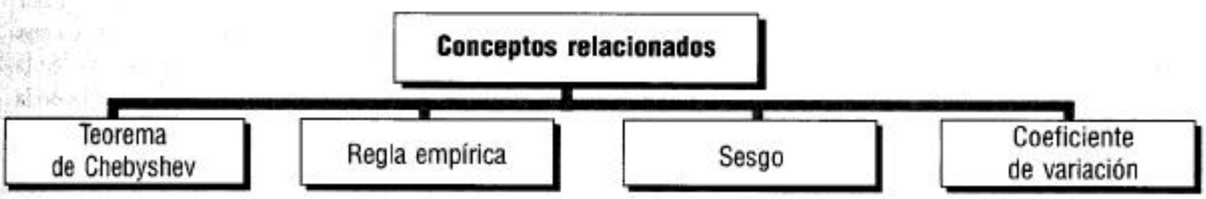
3

Medidas de tendencia central y de dispersión

Plan del capítulo

Este capítulo ilustra cómo puede describirse con unos cuantos números un conjunto de datos completo. Se observará lo útil que pueden ser la tendencia central y las medidas de dispersión en el análisis estadístico.





ESCENARIO

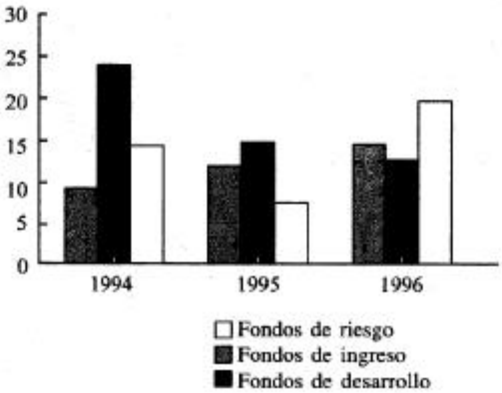
La revista *Fortune* (febrero 17 de 1997), reportó el rendimiento promedio anual de tres fondos (fondos de riesgo, de títulos y de desarrollo) que ofrecía Janus Funds, una gran compañía de fondos mutuos de inversión. Como analista de inversiones se le asigna la tarea de analizar las operaciones de Janus. Para efectuar esta asignación laboral se debe preparar un estudio descriptivo del desempeño de los tres fondos que aparecen en el gráfico.

Como lo demuestra el gráfico, los rendimientos en cada uno de estos fondos han variado considerablemente con los años. Hay muchas consideraciones que se deben tener en cuenta al realizar el análisis, y el supervisor inmediato ha enfatizado la importancia de una presentación minuciosa y precisa.

Surgen muchas preguntas sobre la forma como se debe proceder. Como lo enseña este capítulo, existen muchas herramientas descriptivas del análisis estadístico que se pueden utilizar para desarrollar el análisis. Cada herramienta juega un papel importante en el suministro de valiosa información. Los estadísticos descriptivos tales como rendimientos promedio y medidas de dispersión, serán muy útiles.

En este capítulo se analizarán muchas de las herramientas descriptivas que se utilizan usualmente para transmitir información valiosa de manera concisa e informativa. Tales prácticas reducen la incertidumbre y facilitan ampliamente el proceso de toma de decisiones.

Rendimientos porcentuales – Janus Funds



3.1 Introducción

Los datos, al igual que los estudiantes, se congregan alrededor de sus puntos de encuentro favoritos. Parece que los estudiantes acuden en masa a sitios tales como partidos de fútbol, fraternidades, bares populares y otros sitios de reunión y en raras ocasiones hasta la biblioteca. De igual forma, los números parecen disfrutar de la compañía de otros números y están propensos a reunirse alrededor de un punto central denominado **medida de la tendencia central** o, más comúnmente, *media*. Una medida de tendencia central ubica e identifica el punto alrededor del cual se centran los datos.

Un conjunto grande de datos puede ser rápidamente descrito de manera sucinta con un solo número. Si el profesor dice que el promedio de la clase en el último examen de estadística fue de 95, esto indica algo. Si se dice que el promedio fue de 35, esto indica algo totalmente diferente.

Además, las **medidas de dispersión** indican el punto hasta el cual las observaciones individuales se esparcen alrededor de su punto central. Miden la dispersión o la variabilidad de los datos y reflejan la tendencia de las observaciones individuales a desviarse de dicho punto central.

3.2 Medidas de la tendencia central a partir de datos no agrupados

Existen tres métodos comunes para identificar el centro de un conjunto de datos: la **media**, la **mediana** y la **moda**. Su cálculo e interpretación dependerá de la definición de "centro". En cada caso, se ubican en el punto alrededor del cual se aglomeran los datos.

A. La media

La media, o media aritmética, es la medida de la tendencia central que usualmente se le llamaba promedio. Por ejemplo, para calcular la media de los últimos 10 exámenes de estadística, simplemente se suman y se divide por 10.

> **Media aritmética** La medida de la tendencia central que normalmente era considerada como el promedio.

La media de una población es el parámetro μ (que se pronuncia miu). Si hay N observaciones en el conjunto de datos de la población, la media se calcula así:

Media poblacional

$$\mu = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N} \qquad [3.1]$$

La letra griega mayúscula Σ es el signo de sumatoria que indica que se suman todas las observaciones de 1 a N. Las X_i denotan las observaciones individuales. En aras de la simplicidad, el subíndice y el superíndice se reducirán y aparecerá el signo de la sumatoria como simplemente Σ.

La media de una muestra es un estadístico $\overline{X}$ (que se lee X – barra). Con n observaciones en el conjunto de datos de la muestra, la media se determina así:

Media muestral

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n} \qquad [3.2]$$

Se supone que una muestra de los ingresos por ventas mensuales en miles de dólares para cinco meses es de 56, 67, 52, 45, y 67. La media se calcula así:

$$\overline{X} = \frac{56 + 67 + 52 + 45 + 67}{5} = 57.4$$

B. La mediana

La mediana algunas veces es llamada *media posicional*, porque queda exactamente en la mitad del conjunto de datos después de que las observaciones se han colocado en serie ordenada. La mitad de las observaciones estará por encima de la mediana, la otra mitad estará por *debajo* de ella.

La mediana La observación de la mitad después de que se han colocado los datos en una serie ordenada.

Si el conjunto de datos tiene un número impar de observaciones, la *posición* de la mediana es

$$\text{Posición de la mediana} = \frac{n+1}{2} \qquad [3.3]$$

Dado el ejemplo anterior sobre los ingresos por ventas, se debe colocar primero los datos en serie ordenada:

45, 52, 56, 67, 67

La posición del valor de la mediana se vuelve

$$\text{Posición de la mediana} = \frac{5+1}{2} = 3$$

o la tercera posición. De la serie ordenada anterior se puede observar que este valor es 56.

Con un conjunto de datos que contiene un número par de observaciones es necesario promediar los dos valores medios. Si los ingresos para un sexto mes son de 35 y se adiciona al conjunto de datos, entonces la serie ordenada se vuelve 35, 45, 52, 56, 67, 67. La posición de la mediana es

$$\text{Posición de la mediana} = \frac{6+1}{2} = 3.5$$

es decir, la posición tres y medio. Los dos valores de las posiciones tercera y cuarta se promedian para producir una mediana de 52 + 56 = 54. Esto significa que en la mitad de los meses las ventas estuvieron por debajo de US$54,000, y en la mitad de los meses los ingresos excedieron dicha suma.

C. La moda

La observación modal es la observación que ocurre con mayor frecuencia. Utilizando las seis observaciones anteriores de 35, 45, 52, 56, 67, 67, la moda es 67. Si la séptima observación de 56 se agregara, el conjunto de datos sería *bimodal*, con modas de 56 y 67.

Ejemplo 3.1 Medidas de tendencia central

La emisión de la revista *Fortune* del 17 de febrero de 1997 reportó que en 1996 las utilidades en millones de dólares de varias de las 500 mejores compañías que aparecen en la revista incluían:

Exxon	US$7,510	General Electric	US$7,280
Philip Morris	6,246	IBM	5,429
Intel	5,157	General Motors	4,289

Calcule las tres medidas de tendencia central.

Solución

Media: Tratando los datos como una muestra se tiene que

$$\bar{X} = \frac{7{,}510 + 7{,}280 + 6{,}246 + 5{,}429 + 5{,}157 + 4{,}289}{6} = 5{,}985$$

Mediana: Primero deben colocarse los datos en una serie ordenada. Si no están ordenados, la posición que se encuentre utilizando la fórmula 3.3 no tiene sentido.

4,289, 5,157, 5,429, 6,246, 7,280, 7,510

La posición de la mediana es $(n + 1)/2 = 3.5$. La mediana es el promedio de los valores tercero y cuarto: $5{,}429 + 6{,}246 = 5{,}837.5$

Moda: Este conjunto de datos no tiene moda debido a que todas las observaciones ocurrieron con igual frecuencia.

Interpretación

La media y la mediana identifican el centro del conjunto de datos alrededor del cual se encuentran las observaciones.

D. La media ponderada

En la discusión sobre la media, se asume que cada observación era de igual importancia. Sin embargo, en ciertos casos, puede querer darse mayor peso a algunas de las observaciones. Por ejemplo, si el profesor de estadística amenaza que el examen final valdrá el doble de los otros exámenes para determinar la nota final, entonces al puntaje que se obtenga en el examen final debe dársele el doble de peso. Es decir, que debe contarse doble al calcular la nota. Esto es exactamente lo que hace la media ponderada al utilizar la fórmula 3.4

$$\bar{X}_W = \frac{\Sigma XW}{\Sigma W} \qquad [3.4]$$

en donde $\bar{X}_w$ es la media ponderada

X es la observación individual

W es el peso o ponderación asignada a cada observación.

Se asume que se obtuvo un puntaje de 89, 92 y 79 en los exámenes parciales y 94 en el examen final. Estos puntajes y sus respectivas ponderaciones están reflejados en la tabla 3.1 La fórmula (3.4) da

$$\bar{X}_W = \frac{\Sigma XW}{\Sigma W} = \frac{448}{5} = 89.6$$

Tabla 3.1
Cálculo de la media ponderada

Nota (*X*)	Peso (*W*)	*XW*
89	1	89
92	1	92
79	1	79
94	2	188
	5	448

Este método es igual que sumar la nota del examen final dos veces al calcular la media:

$$\bar{X}_w = \frac{89 + 92 + 79 + 94 + 94}{5} = 89.6$$

> **Media ponderada** La media ponderada toma en cuenta la importancia relativa de las observaciones.

Ejemplo 3.2 Cálculo de la media ponderada

Paul el fontanero vende cinco tipos de limpiadores para desagües. En la tabla se muestra cada tipo junto con la utilidad por lata y el número de latas vendidas.

Limpiador	Utilidad por lata (X)	Volumen de ventas en latas (W)	XW
Glunk Out	US$ 2.00	3	US$ 6.00
Bubble Up	3.50	7	24.50
Dream Drain	5.00	15	75.00
Clear More	7.50	12	90.00
Main Drain	6.00	15	90.00
	US$ 24.00	52	US$ 285.50

Se puede calcular la media aritmética simple de la utilidad de Paul como US$24/5 = US$4.80 por lata.

Solución

Sin embargo, probablemente este no sea un buen estimado de la utilidad promedio de Paul, debido a que vende más de algunos tipos de limpiadores que de otros. Para obtener un estado financiero más representativo del desempeño real de su negocio, Paul debe dar más peso a los tipos más populares de limpiadores. Por tanto, el cálculo apropiado sería la media ponderada. La medida de peso apropiada sería las cantidades vendidas. La media ponderada es entonces

$$\bar{X}_w = \frac{\Sigma XW}{\Sigma W} = \frac{\text{US\$}285.50}{52} = \text{US\$}5.49$$

Interpretación

La media ponderada es mayor que la media aritmética simple porque Paul vende más limpiadesagües de los tipos que tienen un margen de utilidad mayor.

E. La media geométrica

La media geométrica puede utilizarse para mostrar los cambios porcentuales en una serie de números positivos. Como tal, tiene una amplia aplicación en los negocios y en la economía, debido a que con frecuencia se está interesado en establecer el cambio porcentual en las ventas, en el producto nacional bruto o en cualquier serie económica.

> **Media geométrica** La media geométrica proporciona una medida precisa de un cambio porcentual promedio en una serie de números.

La media geométrica (MG) se halla tomando la raíz enésima del producto de n números. Así,

$$MG = \sqrt[n]{X_1 X_2 X_3 \cdots X_n} \qquad [3.5]$$

La mayoría de las calculadoras manuales pueden calcular la raíz enésima de cualquier número.

La media geométrica se utiliza con más frecuencia para calcular la tasa de crecimiento porcentual promedio de algunas series dadas, a través del tiempo. Para ilustrar esta aplicación en un escenario de negocios, se consideran las cifras sobre ingresos del ejemplo 3.3 para White-Knuckle Airlines, la principal competencia de P&P durante los últimos cinco años.

Ejemplo 3.3 La media geométrica

El CEO (director ejecutivo) de White-Knuckle Airlines desea determinar la tasa de crecimiento promedio en los ingresos con base en las cifras dadas en la tabla. Si la tasa de crecimiento promedio es menor que el promedio industrial del 10%, se asumirá una nueva campaña publicitaria.

Ingresos para White-Knuckle Airlines		
Año	Ingreso	Porcentaje del año anterior
1992	US$ 50,000	-- --
1993	55,000	55/50 = 1.10
1994	66,000	66/55 = 1.20
1995	60,000	60/66 = 0.91
1996	78,000	78/60 = 1.30

Solución

Primero es necesario determinar el porcentaje que los ingresos de cada año representan respecto de los obtenidos el año anterior. En otras palabras, ¿qué porcentaje del ingreso de 1992 es el ingreso en 1993? Esto se encuentra dividiendo los ingresos de 1992 entre los de 1993. El resultado, 1.10 revela que los ingresos de 1993 son 110% de los ingresos de 1992. También se calculan los porcentajes para los tres años restantes. Tomando la Media Geométrica (MG) de estos porcentajes da

$$\text{MG} = \sqrt[4]{(1.10)(1.2)(0.91)(1.3)} = 1.1179$$

Restando 1 para convertirlo a un incremento anual promedio da 0.1179, o un incremento promedio de 11.79% para el período de cinco años.

Por otro lado, la media aritmética simple es

$$\bar{X} = \frac{1.1 + 1.2 + 0.91 + 1.3}{4} = \frac{4.51}{4} = 1.1275$$

o un cambio promedio de 12.75%. Se divide por 4 ya que se presentaron cuatro cambios durante el período de cinco años.

Sin embargo, si un incremento promedio de 12.75%, basado en la media aritmética simple, se aplica a la serie que comienza con US$50,000, los resultados son

US$ 50,000 × 1.1275 = US$ 56,375

US$ 56,375 × 1.1275 = US$ 63,563

US$ 63,563 × 1.1275 = US$ 71,667

US$ 71,667 × 1.1275 = US$ 80,805

Ya que US$80,805 excede los US$78,000 que White-Knuckle Airlines en realidad ganó, el incremento del 12.75% es obviamente muy alto. Si se utiliza la tasa de crecimiento de la media geométrica del 11.79%, se obtiene

$$\text{US\$ }50{,}000 \times 1.1179 = \text{US\$ }55{,}895$$
$$\text{US\$ }55{,}895 \times 1.1179 = \text{US\$ }62{,}485$$
$$\text{US\$ }62{,}485 \times 1.1179 = \text{US\$ }69{,}852$$
$$\text{US\$ }69{,}852 \times 1.1179 = \text{US\$ }78{,}088 \approx \text{US\$}78{,}000$$

Esto da un valor de US$78,088, lo que está mucho más cerca al ingreso real de US$78,000.

Interpretación

La media geométrica representa el cambio promedio con el tiempo. Debido a que la tasa de crecimiento supera el promedio de la industria del 10%, la nueva campaña publicitaria no se llevará a cabo.

La media geométrica siempre será menor que la media aritmética salvo en el extraño caso en el que todos los incrementos porcentuales sean iguales. Entonces las dos medias serán iguales.

3.3 Comparación entre media, mediana y moda

La media es la medida más común de tendencia central. Se presta para mayor manipulación e interpretación algebraica. Desafortunadamente, la media se ve afectada por valores extremos, o valores atípicos, y a diferencia de la mediana, puede ser sesgada por las observaciones que están muy por encima o muy por debajo de ésta.

Por ejemplo, para los datos 4, 5, 6, 6,7, 8 la media y la mediana son ambas 6 y representa una excelente medida del punto central de los datos. Si la última observación fuera 80 en lugar de 8, la media sería 18, pero la mediana todavía sería 6. Debido a que la mediana no se ve afectada por este valor extremo, representa mejor las seis observaciones.

La moda también es menos afectada por unas pocas observaciones atípicas y sigue siendo 6 aun cuando el último valor sea 80. Sin embargo, si no hay moda, o si el conjunto de datos es bimodal, su uso puede ser confuso.

Esto no implica que una medida sea necesariamente mejor que las otras. La medida que se seleccione depende de la naturaleza de los datos o de la forma como se utilicen los datos. Por ejemplo Land's End un vendedor minorista popular de equipos para acampar, se beneficiaría muy poco del hecho de saber que la talla promedio de las botas de excursionismo que vendió fue de 7.3492. De mayor utilidad para las decisiones futuras del negocio sería conocer el tamaño *modal* - reconociendo que vendió más botas de talla 8 que de cualquier otra talla.

Sin embargo, se asume que Land's End desea comercializar una nueva tienda de acampar. Las dimensiones de la tienda dependerán, entre otras cosas, de la estatura promedio de los adultos. La experiencia ha demostrado que la media sirve muy bien como medida de tendencia central cuando se trata de productos que están hechos para acomodarse a la estatura de las personas. El tamaño de los marcos de las puertas de las entradas de los hogares y negocios minoristas, y gran parte del mobiliario se fabrica con base en la estatura promedio.

Ejercicios de la sección

1. Su firma está introduciendo un nuevo chip de computador del cual se promociona que realiza cálculos estadísticos mucho más rápidamente que los que actualmente se encuentran en el mercado. Se hacen veinte cálculos diferentes, produciendo los tiempos en segundos que se ven más adelante. Aunque usted no puede

tergiversar su producto, usted desea presentar los resultados de la manera más favorable para su empresa. Determine la media, la mediana y la moda. Comente los beneficios relativos de utilizar cada estadístico.

3.2	4.1	6.3	1.9	0.6
5.4	5.2	3.2	4.9	6.2
1.8	1.7	3.6	1.5	2.6
4.3	6.1	2.4	2.2	3.3

2. Como gerente de ventas de Acme, Inc., usted desea calcular las medidas de tendencia central para los niveles de utilidad de su firma, durante los últimos doce meses, ya que las siguientes utilidades mensuales están dadas en miles de dólares:

US$12.3	US$14.3	US$25.7
21.6	21.6	−12.8
22.3	18.0	23.1
−3.4	17.9	22.3

3. El director de la planta de Intel desea que usted compare los salarios promedio en su planta de Palo Alto con las de la competencia que queda ubicada en las cercanías de San José. De los 6,012 empleados que supervisa 1,212 ganan US$12.30 la hora; a 650 se les paga US$15.50; 3,098 ganan US$23.50, y al resto se les paga US$17.12. De los 5,634 empleados que hay en la otra planta 1,654 ganan US$12.75; 815 reciben US$17.80 y los otros US$20.10. Escriba un breve informe para que el director obtenga detalladamente la información que desea.

4. Una empresa grande de equipos deportivos está probando el efecto de dos planes publicitarios sobre las ventas de los últimos 4 meses. Dadas las ventas que se ven aquí, ¿cuál programa de publicidad parece producir el crecimiento promedio más alto en ventas mensuales?

Mes	Plan 1	Plan 2
Enero	US$ 1,657	US$ 4,735
Febrero	1,998	5,012
Marzo	2,267	5,479
Abril	3,432	5,589

3.4 Medidas de dispersión

En nuestro esfuerzo por describir un conjunto de números hemos visto que es de utilidad ubicar el centro del conjunto de datos. Pero identificar una medida de tendencia central rara vez es suficiente. Una descripción más completa del conjunto de datos puede obtenerse si se mide qué tan dispersos están los datos alrededor de dicho punto central. Esto es precisamente lo que hacen las medidas de dispersión. Indican cuánto se desvían las observaciones alrededor de su media.

Medidas de dispersión Miden qué tanto se dispersan las observaciones alrededor de su media.

Se toman por ejemplo los tres conjuntos pequeños de datos que se observan aquí.

Conjunto de datos 1	Conjunto de datos 2	Conjunto de datos 3
0,5,10	4,5,6	5,5,5

Los tres tienen una media de cinco. ¿Se debe por tanto concluir que los conjuntos de datos son similares? Claro que no. Sin embargo, si se informa sólo sus medias, sin ver las observaciones, se puede concluir que hay similitud. Una imprecisión más notoria de los conjuntos de datos resultaría si se compara el grado en el cual se dispersaron las observaciones individuales en cada conjunto de datos o se expandieron alrededor de la media cinco. Las observaciones en el primer conjunto de datos están muy dispersas por encima y por debajo de la media, mientras que aquellas del segundo grupo de datos están comparativamente cerca de ésta. El primer conjunto de datos tiene una medida de dispersión mayor que la segunda. El tercer conjunto de datos no tiene dispersión, todas las observaciones son iguales a la media. Sabiendo esto, sería poco probable asumir de manera errónea cualquier similitud en los conjuntos de datos simplemente con base en su media. En este sentido, las medidas de dispersión son muy útiles e informativas.

A. El rango

La medida de dispersión más simple (y menos útil) es el rango o recorrido. El **rango** es simplemente la diferencia entre la observación más alta y la más baja. Su ventaja es que es fácil de calcular. Su desventaja es que considera sólo dos de los cientos de observaciones que hay en un conjunto de datos. El resto de las observaciones se ignoran. Los rangos de los tres conjuntos de datos anteriores son 10, 2 y 0 respectivamente.

B. Varianza y desviación estándar de una población

La varianza y su raíz cuadrada, y la desviación estándar son medidas de dispersión mucho más útiles. Proporcionan una medida más significativa sobre el punto hasta el cual se dispersan las observaciones alrededor de su media.

La **varianza** es el "promedio de las desviaciones respecto a su media elevadas al cuadrado". ¿Qué significa esto? Significa que (1) se encuentra la cantidad por la cual cada observación se desvía de la media, (2) se elevan al cuadrado tales desviaciones, y (3) se halla la media de tales desviaciones elevadas al cuadrado. Así, se tiene el promedio de las desviaciones de la media elevadas al cuadrado.

Varianza El promedio de las observaciones respecto a su media elevadas al cuadrado.

La varianza para una población σ^2 (se lee sigma al cuadrado) es

Varianza poblacional

$$\sigma^2 = \frac{(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 - \mu)^2 + \cdots + (X_N - \mu)^2}{N} = \frac{\Sigma(X_i - \mu)^2}{N} \qquad [3.6]$$

En donde	$X_1, X_2, X_3, \ldots, X_N$	son las observaciones individuales
	μ	es la media poblacional
	N	es el número de observaciones

La desviación estándar σ es

Desviación estándar poblacional

$$\sigma = \sqrt{\sigma^2} \qquad [3.7]$$

Vale la pena notar que debido a que se está trabajando con una población, la media es μ, no $\bar{X}$, como para una muestra, y el número de observaciones es N y no n, como para una muestra.

Desviación estándar Es la raíz cuadrada de la varianza. Es una medida importante de la dispersión de los datos.

A manera de ilustración, Chuckie Chambers vende cinco pólizas de seguro diferentes del baúl de su Plymouth modelo 1973. Sus respectivas primas mensuales son de US$110, US$145, US$125, US$95 y US$150. La prima promedio es

$$\mu = \frac{110 + 145 + 125 + 95 + 150}{5} = \text{US\$}125$$

La varianza se halla: (1) restando la media de US$125 de cada una de las observaciones, (2) elevando al cuadrado estas desviaciones, y (3) hallando el promedio de estas desviaciones al cuadrado. Al seguir estos tres pasos resulta:

$$\sigma^2 = \frac{(110 - 125)^2 + (145 - 125)^2 + (125 - 125)^2 + (95 - 125)^2 + (150 - 125)^2}{5}$$

$$= 430$$

A pesar del uso común de la varianza, ésta presenta dos problemas: es un número muy grande con respecto a las observaciones. Como se puede ver, es varias veces mayor incluso que la observación más grande. Debido a su gran tamaño, con frecuencia la varianza se vuelve difícil para trabajar.

Un problema aún más angustioso es que debido a que las desviaciones son elevadas al cuadrado, la varianza siempre se expresa en términos de los datos originales elevados al cuadrado. En el caso de Chuckie, debido a que elevó al cuadrado las desviaciones de la media, entonces se convierte en 430 dólares al cuadrado – una unidad de medida que no tiene sentido. En la mayoría de los casos la varianza se expresa en términos que no tienen significado o interpretación lógica.

Sin embargo, ambas complicaciones pueden resolverse rápidamente. Tan sólo con hallar la desviación estándar σ, sacando la raíz cuadrada de la varianza:

$$\sigma = \sqrt{430} = \text{US\$}20.74$$

Así de fácil, se solucionan ambos problemas. Ahora se tiene un número más pequeño con el cual es más fácil trabajar, y más importante aún, ahora está expresado en dólares ya que se tomó la raíz cuadrada de los dólares elevados al cuadrado.

El concepto de desviación estándar es muy importante en los negocios y en la economía. Por ejemplo, en finanzas la desviación estándar se utiliza como medida de riesgo relacionada con varias oportunidades de inversión. Mediante el uso de la desviación estándar para medir la variabilidad en las tasas de rendimiento ofrecidas por diferentes inversiones, el analista financiero puede medir el nivel de riesgo que tiene cada activo financiero. Generalmente, entre mayor sea la desviación estándar de la tasa de rendimiento de una inversión en particular, mayor será el grado de riesgo. Veamos el siguiente ejemplo:

Ejemplo 3.4 Varianza y desviación estándar de una población

Markus Boggs es gerente de Nest Egg Investments. Recientemente, Markus estaba interesado en las tasas de rendimiento de los últimos cinco años de dos diferentes fondos mutuos. Megabucks, Inc. mostró, durante un período de cinco años, tasas de rendimiento del 12, 10, 13, 9 y 11%, mientras que Dynamics Corporation arrojó 13, 12, 14, 10 y 6%. Un cliente se acercó a Boggs y expresó su interés en uno de estos fondos mutuos. ¿Cuál debería escoger Boggs para su cliente?

Solución

Vale la pena destacar que ambos fondos ofrecen un rendimiento promedio del 11%. Debido a que ambos ofrecen el mismo rendimiento en promedio, una inversión más segura es la que tiene un grado menor de riesgo, tal como se midió mediante la desviación estándar. Para Megabucks, Boggs halla

$$\sigma^2 = \frac{(12-11)^2 + (10-11)^2 + (13-11)^2 + (9-11)^2 + (11-11)^2}{5} = 2$$

La desviación estándar es

$$\sigma = \sqrt{2} = 1.41\%$$

Para Dynamics,

$$\sigma^2 = \frac{(13-11)^2 + (12-11)^2 + (14-11)^2 + (10-11)^2 + (6-11)^2}{5} = 8$$

La desviación estándar por tanto es

$$\sigma = \sqrt{8} = 2.83\%$$

Interpretación

Debido a que Megabucks presenta menos variabilidad en sus rendimientos y ofrece la misma tasa de rendimiento promedio que ofrece Dynamics, Megabucks representa la más segura de las dos inversiones y por ende es la oportunidad de inversión preferida.

C. Varianza y desviación estándar para una muestra

Tome nota: Los ejemplos anteriores se relacionan con la varianza y la desviación estándar para una *población*. Los símbolos σ^2 y σ son letras griegas típicas de los parámetros.

Rara vez se pueden calcular parámetros. En la mayoría de los casos más bien se estimarán tomando una muestra y calculando los estadísticos correspondientes. Teniendo esto presente, esta sección analiza la forma como se calculan estas importantes medidas de dispersión en la medida en que se relacionan con las muestras.

La varianza y la desviación estándar para una muestra representan medidas de dispersión alrededor de la media. Se calculan de manera parecida a aquellas para una población. La varianza de la muestra s^2 es

Varianza de la muestra	$s^2 = \dfrac{\Sigma(X_i - \bar{X})^2}{n-1}$	[3.8]

y la desviación estándar de la muestra es

Desviación estándar de la muestra	$s = \sqrt{s^2}$	[3.9]

Vale la pena destacar que la media en la fórmula (3.8) se expresa como $\bar{X}$, y no μ, ya que se está trabajando con muestras. Además, se divide por $n - 1$ en lugar de N, debido a que se tiene $n - 1$ *grados de libertad*, $gl = n - 1$. El número de grados de libertad en toda operación estadística es igual al número de observaciones menos toda restricción impuesta en tales observaciones. Una restricción es cualquier valor que deba calcularse de dichas observaciones.

Por ejemplo, se asume que se deben seleccionar $n = 4$ observaciones que deben promediar $\bar{X} = 10$. Bajo estas condiciones, se está en libertad de escoger tres números que se deseen – digamos el tamaño del sombrero, la edad o el coeficiente intelectual-CI. Aunque una vez que esos tres números se seleccionan, el cuarto está predeterminado si los anteriores deben promediar $\bar{X} = 10$. Como lo ilustra la fórmula (3.8), la varianza utiliza el valor para $\bar{X}$ que funciona como una restricción y por tanto reduce los grados de libertad en 1. Por consiguiente, se disminuye el número de observaciones n, en 1.

Otra razón por la cual se toma $n - 1$ es que una muestra generalmente está un poco menos dispersa que la población de la cual se tomó. Por tanto, existe la tendencia a que la desviación estándar de la muestra s sea un poco menor que la desviación estándar de la población σ. Este es un hecho infortunado. Vale la pena recordar que se intenta utilizar el valor de s como una estimación de σ. Sin embargo, s subestimará σ de manera consistente. Se debe compensar esta condición inflando artificialmente s dividiendo por un número levemente menor n-1, en lugar de n.

Para ilustrar la técnica para establecer estas medidas de dispersión para una muestra, se considera otro problema que tiene Boggs en su esfuerzo por ayudar a sus clientes a tomar decisiones sobre inversión.

Ejemplo 3.5 Varianza y desviación estándar para una muestra

El Sr. Boggs desea determinar la estabilidad del precio de una acción en particular. Decide basar su juicio en la estabilidad de la desviación estándar del precio de cierre diario de dicha acción. Al revisar las páginas financieras, Boggs sabe que la acción ha sido transada en la bolsa durante algún tiempo y que hay muchos precios de cierre desde hace varios meses. En lugar de utilizar todos estos precios, Boggs decide simplificar su aritmética y seleccionar una muestra aleatoria de $n = 7$ días. (Aunque 7 probablemente es una muestra muy pequeña, servirá por el momento para los propósitos que se tienen.) Boggs nota que los precios de cierre son de

US\$87, US\$120, US\$54, US\$92, US\$73, US\$80, y US\$63

Solución

Entonces, $\bar{X} =$ US\$81.29 y

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

$$= \frac{(87 - 81.29)^2 + (120 - 81.29)^2 + \ldots + (63 - 81.29)^2}{7 - 1}$$

$$s^2 = 465.9 \text{ cuadrado de dólares}$$

$$s = \sqrt{465.9} = \text{US\$}21.58$$

Interpretación

Boggs ha estimado que la media del precio de cierre de la acción es de US\$81.29, con una tendencia a variar por encima o por debajo de dicho precio en US\$21.58. Una explicación adicional del uso e interpretación de la desviación estándar se proporciona más tarde en este capítulo. Sin embargo, se tiene en mente que Boggs puede interpretar siempre la desviación estándar de US\$21.58 como una medida de la tendencia de los precios de cierre que tienden a fluctuar alrededor de su media de $81.29 dólares.

Todos estos estadísticos importantes pueden obtenerse en los computadores utilizando varios paquetes. La pantalla 3.1 muestra una copia impresa en Excel para los datos de P&P. Es de destacar que las medidas de tendencia central y de dispersión se dan junto con otros estadísticos que se discutirán más adelante.

Pantalla 3.1

Estadísticos descriptivos para P&P

Mean (Media)	78.36
Standard Error (Error estándar)	1.599938774
Median (Mediana)	78.5
Mode (Moda)	83
Standard Deviation (Desviación estándar)	11.31327557
Sample Variance (Varianza de la muestra)	127.9902041
Kurtosis (Curtosis)	-0.036918787
Skewness (Sesgo)	-0.05225955
Range (Rango)	52
Minimum (Mínimo)	50
Maximum (Máximo)	102
Sum (Suma)	3918
Count (Conteo)	50

Ejercicios de la sección

5. Se utilizan dos procesos para producir discos de computador. Han surgido problemas respecto a las variaciones en los tamaños de tales discos. Con base en los datos de muestra aquí observados, de ocho tamaños de discos en pulgadas para cada proceso, explique cuál proceso aconsejaría usted si su objetivo es minimizar la desviación en el tamaño alrededor de la media.

Proceso 1		Proceso 2	
3.41	3.22	3.81	3.26
3.74	3.06	3.26	3.79
3.89	3.65	3.07	3.14
3.65	3.33	3.35	3.51

6. Explique con sus propias palabras qué miden la varianza y la desviación estándar. ¿Por qué su cálculo es algo diferente para las poblaciones y las muestras?

7. Un analista de inversiones sugiere que usted invierta en Boomer Securities en lugar de Reliable Stocks. Dadas las tasas anuales de rendimiento que se muestran a continuación para una muestra de cada inversión, ¿qué le dice al analista si usted desea minimizar su exposición al riesgo?

Boomer		Reliable	
15.5%	3.6%	4.5%	6.2%
21.7	27.2	5.5	7.2
−7.8	2.2	3.5	4.2
−5.0	12.2	4.1	

8. Curly, Moe, y Larry venden seguros de vida para la Shemp Insurance Company. El Sr. Shemp ascenderá a uno de sus vendedores a un cargo administrativo con base en su desempeño en ventas. Su decisión depende de cuál miembro de su equipo de ventas tiene (1) el promedio más alto en ventas y (2) el registro de ventas más consistente. Dados los siguientes datos de muestra semanales en ventas, ¿cuál vendedor obtendrá el ascenso?

Curly		Moe		Larry	
US$ 986	US$ 1,265	US$645	US$893	US$534	US$534
1,337	734	645	230	534	534
2,745	245	734	415	534	534
2,645	5,344	822	723	534	534
3,658	4,867				

3.5 Medidas de tendencia central y de dispersión para datos agrupados

Al trabajar con datos que han sido agrupados en una distribución de frecuencia, no se conoce cuáles son las observaciones individuales. En una tabla de frecuencia para los pasajeros de P&P, la tabla 2.2, se sabe solamente que en tres días, entre 50 y 59 pasajeros abordaron un avión. No se tienen las cifras exactas de esos tres días. Sin los valores específicos, los procedimientos mostrados anteriormente para calcular las medidas descriptivas, simplemente no se aplican. Deben encontrarse métodos alternativos. Debe tenerse en mente que los cálculos hechos utilizando datos agrupados son sólo aproximaciones. Por tanto, las observaciones individuales no agrupadas deberían utilizarse cuando sea posible.

A. La media

Al calcular la media de datos agrupados, se supone que las observaciones en cada clase son iguales al punto medio de la clase. Aunque esto puede ser una suposición heroica, probablemente se compensa debido a que posiblemente algunas de las observaciones superen el punto medio, mientras que otros queden por debajo de él. Dada esta suposición, se debe tener en cuenta la frecuencia y los puntos medios de cada clase cuando se calcule la media utilizando datos agrupados. La fórmula 3.10 hace justo esto.

Media con datos agrupados
$$\bar{X}_g = \frac{\Sigma fM}{n} = \frac{\Sigma fM}{\Sigma f} \qquad [3.10]$$

en donde f es la frecuencia o número de observaciones en cada clase
M es el punto medio de cada clase
n es el tamaño de la muestra y es igual a las frecuencias sumadas en todas las clases.

La tabla de frecuencia para Pigs & People Airlines del capítulo 2 se repite en la tabla 3.2 por conveniencia, junto con los puntos medios para cada clase, los cuales, como se debe recordar, se determinaron promediando los límites superior e inferior.

Utilizando la fórmula (3.10), se puede ver que P&P transportó un promedio diario de 78.7 pasajeros.

$$\bar{X}_g = \frac{\Sigma fM}{n} = \frac{3935}{50} = 78.7 \text{ pasajeros}$$

Tabla 3.2
Distribución de frecuencia para pasajeros

Clase (pasajeros)	Frecuencia (*f*) (días)	*M*	*fM*
50-59	3	54.5	163.5
60-69	7	64.5	451.5
70-79	18	74.5	1341.0
80-89	12	84.5	1014.0
90-99	8	94.5	756.0
100-109	2	104.5	209.0
	50		3935.0

B. La mediana

Si se han registrado datos en una tabla de frecuencia, no pueden colocarse en un arreglo ordenado para calcular la mediana. A manera de ilustración, se suministra la tabla de frecuencia de P&P Airlines, en la tabla 3.3.

Tabla 3.3
Distribución de frecuencia para pasajeros

Clase	*f*	Frecuencia acumulada
50-59	3	3
60-69	7	10
70-79	18	28
80-89	12	40
90-99	8	48
100-109	2	50

Primero se debe hallar la clase de la mediana de la distribución de frecuencia. La **clase mediana** es la clase cuya frecuencia acumulada es mayor que o igual a $n/2$.

Debido a que n es 50, se necesita localizar la primera clase con una frecuencia acumulada de 25 o más. La tercera clase tiene una frecuencia acumulada de 28. La mediana puede determinarse entonces como

Mediana para datos agrupados $$\text{Mediana} = L_{md} + \left[\frac{n/2 - F}{f_{md}}\right](C) \qquad [3.11]$$

en donde L_{md} es el límite inferior de la clase de la mediana (70)
F es la frecuencia acumulada de la clase que antecede a la clase de la mediana (10)
f_{md} es la frecuencia de la clase de la mediana (18)
C es el intervalo de clase de la clase de la mediana (10)

Utilizando la fórmula (3.11), se obtiene la mediana:

$$\text{Mediana} = 70 + \left[\frac{50/2 - 10}{18}\right]10 = 78.33 \text{ pasajeros}$$

Se puede concluir que en 25 días – la mitad de los 50 días estudiados – menos de 78.33 pasajeros volaron en P&P Airlines, y en los otros 25 días, más de 78.33 pasajeros volaron los agradables cielos de P&P.

C. La moda

Ya que por definición la moda es la observación que ocurre con mayor frecuencia, se hallará en la clase que tenga la frecuencia más alta, llamada la **clase modal**. Para estimar la moda en el caso de datos agrupados, se utiliza la fórmula (3.12).

Moda para datos agrupados	$\text{Moda} = L_{mo} + \left[\frac{D_a}{D_b + D_a}\right](C)$	[3.12]

en donde
- L_{mo} es el límite inferior de la clase modal
- D_a es la diferencia entre la frecuencia de la clase modal y la clase que la antecede
- D_b es la diferencia entre la frecuencia de la clase modal y la clase que le sigue
- C es el intervalo de clase de la clase modal

De la tabla 3.3, la moda es

$$\text{Moda} = 70 + \left[\frac{18 - 7}{(18 - 12) + (18 - 7)}\right](10) = 76.47 \text{ pasajeros}$$

D. Varianza y desviación estándar

Si los datos están agrupados en una tabla de frecuencia, la varianza y la desviación estándar pueden calcularse como

Varianza de la muestra de datos agrupados	$s^2 = \frac{\Sigma fM^2 - n\overline{X}^2}{n - 1}$	[3.13]

y

Desviación estándar muestral para datos agrupados	$s = \sqrt{s^2}$	[3.14]

Los datos de pasajeros para P&P Airlines se utilizarán como ilustración.

Ejemplo 3.6 Medidas de dispersión con datos agrupados

El director de vuelo de P&P requiere información respecto a la dispersión del número de pasajeros. Las decisiones que se tomen respecto a la programación y al tamaño más eficiente de los aviones, dependerá de la fluctuación en el transporte de pasajeros. Si esta variación en número de pasajeros es grande, se pueden necesitar aviones más grandes para evitar el sobrecupo en los días en los que el transporte de pasajeros es más solicitado. La tabla de frecuencia para P&P es la siguiente:

Clase (pasajeros)	f (días)	M	fM	M^2	fM^2
50-59	3	54.5	163.5	2,970.25	8,910.75
60-69	7	64.5	451.5	4,160.25	29,121.75
70-79	18	74.5	1,341.0	5,550.25	99,904.50
80-89	12	84.5	1,014.0	7,140.25	85,683.00
90-99	8	94.5	756.0	8,930.25	71,442.00
100-109	2	104.5	209.0	10,920.25	21,840.50
	$n = 50$		$\Sigma fM = 3{,}935.0$		$\Sigma fM^2 = 316{,}902.50$

Solución

Se calculó $\bar{X}_g$ en un ejemplo anterior

$$\bar{X}_g = \frac{\Sigma fM}{n} = \frac{3935}{50} = 78.7$$

Por tanto, la fórmula (3.13) da

$$s^2 = \frac{\Sigma fM^2 - n\bar{X}^2}{n-1}$$

$$s^2 = \frac{316{,}902.50 - 50(78.7)^2}{49} = 147.31 \text{ pasajeros elevado al cuadrado}$$

$$s = \sqrt{147.31} = 12.14 \text{ pasajeros}$$

Interpretación

El director de vuelo ahora puede decidir si los aviones que se están utilizando actualmente pueden acomodar fluctuaciones en los niveles de pasajeros tal y como lo mide una desviación estándar de 12.14. Si no, quizá se utilizarán aviones más grandes para acomodar cualquier excedente que pueda ocurrir en esos días de tráfico pesado.

Ejercicios de la sección

9. El ausentismo diario en su oficina parece ir en aumento. El año pasado un promedio de 47.8 empleados estuvo ausente algunos días, con una desviación estándar de 14.7. Se recolectó una muestra de datos para el año en curso y se ubicaron en la tabla de frecuencias que se muestra a continuación. Calcule la media, la mediana, la moda, y la desviación estándar para estos datos y compárelos con los del año anterior. ¿A qué conclusiones llega?

Número de empleados ausentes	Días en los que ese número estuvo ausente
20-29	5
30-39	9
40-49	8
50-59	10
60-69	12
70-79	11
80-89	8
90-99	3

10. Anteriormente, el tiempo para completar un trabajo determinado en las oficinas de Harmon electronics había arrojado las siguientes estadísticas en horas: una media de 12.2, una mediana de 13.2, y una moda de 14.5. La varianza fue de 8.21. Se reflejan datos más recientes en la siguiente tabla de frecuencias. El Sr. Harmon lo contrata como consultor externo para evaluar los cambios en la eficiencia de los empleados. Calcule los estadísticos correspondientes con base en estos datos y prepare un breve informe. ¿Qué conclusiones obtiene?

Horas tomadas hasta la finalización del trabajo	Número de veces que dicho trabajo se tomó este tiempo
5 y menos de 7	4
7 y menos de 9	8
9 y menos de 11	12
11 y menos de 13	8
13 y menos de 15	5
15 y menos de 17	2

3.6 Otras medidas de dispersión

Aunque la varianza y la desviación estándar son las medidas de dispersión más útiles en análisis estadístico, existen otras técnicas con las cuales puede medirse la dispersión de un conjunto de datos. Estas medidas adicionales de dispersión son los **cuartiles**, los **deciles** y los **percentiles**.

Cada conjunto de datos tiene tres cuartiles que lo dividen en cuatro partes iguales. El primer cuartil es ese valor debajo del cual clasifica el 25% de las observaciones, y sobre el cual puede encontrarse el 75% restante. El segundo cuartil es justo la mitad. La mitad de las observaciones están por debajo y la mitad por encima; en este sentido, es lo mismo que la mediana. El tercer cuartil es el valor debajo del cual está el 75% de las observaciones y encima del cual puede encontrarse el 25% restante.

La determinación de cuartiles con frecuencia es de utilidad. Por ejemplo muchas escuelas de postgrados admitirán sólo a aquellos estudiantes que estén en el 25% superior (tercer cuartil) de los candidatos. Las empresas, con frecuencia, desean señalar las plantas cuyos deficientes registros de producción los colocan por debajo del cuartil inferior. Con un poco de imaginación es posible prever numerosos ejemplos en los cuales la determinación de cuartiles puede ser de gran beneficio.

Los deciles separan un conjunto de datos en 10 subconjuntos iguales, y los percentiles en 100 partes. El primer decil es la observación debajo de la cual se encuentra el 10% de las observaciones, mientras que el 90% restante se encuentra encima de éste. El primer percentil es el valor debajo del cual se encuentra el 1% de las observaciones, y el resto están encima de éste. Puede aplicarse una interpretación similar al resto de deciles y percentiles. Todo conjunto de datos tiene 9 deciles y 99 percentiles.

Un percentil y su ubicación en un arreglo ordenado se identifica mediante los subíndices. Por ejemplo, el decimoquinto percentil se indica como P_{15}, y su ubicación en la serie ordenada es L_{15}.

Para ilustrar el cálculo de percentiles, se asume que se tienen observaciones para el número de acciones correspondientes a 50 acciones transados en la Bolsa de Valores de Nueva York, como se muestra en la tabla 3.4. Vale la pena destacar que los datos han sido puestos en una serie ordenada. El lugar del P ésimo percentil se halla

Ubicación de un percentil
$$Lp = (n + 1)\frac{P}{100} \qquad [3.15]$$

en donde Lp es el sitio del percentil deseado en una serie ordenada
n es el número de observaciones
P es el percentil deseado

Se asume que se desea calcular el percentil 25, P_{25}, para las acciones de la tabla 3.4. Se debe hallar primero su ubicación en la serie ordenada.

$$L_{25} = (50 + 1)\frac{25}{100}$$
$$= 12.75$$

Tabla 3.4
Números de acciones transadas en la Bolsa de Valores de Nueva York (en 100's)

3	10	19	27	34	38	48	56	67	74
4	12	**20**	29	34	39	48	59	67	74
7	14	**21**	31	36	43	52	62	69	76
9	15	25	31	37	45	53	63	72	79
10	17	27	34	38	47	56	64	73	80

El valor resultante de 12.75 dice que el percentil 25 está ubicado al 75% del trayecto comprendido entre la doceava observación, que es 20 y la treceava observación que es 21, es decir, $P_{25} = 20 + 0.75\ (21\text{-}20) = 20.75$.

Si se desea calcular el percentil 35, se halla

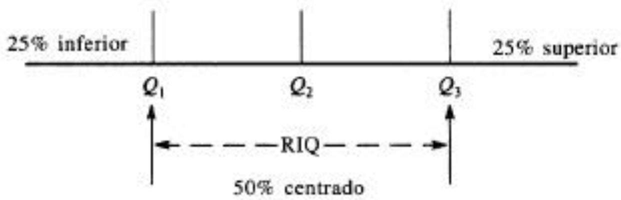
$$L_{35} = (50 + 1)\frac{35}{100}$$
$$= 17.85$$

El percentil 35 está al 85% del trayecto comprendido entre la observación 17, que es 29 y la observación 18 que es 31, es decir, $P_{35} = 29 + (0.85)(31 - 29) = 30.7$ Por tanto el 35% de las observaciones está por debajo de 30.7 y el 65% restante por encima de 30.7

Regresando a los deciles y cuartiles por un momento, se nota que el primer decil es igual a P_{10}, el segundo decil es igual a P_{20}, y así sucesivamente. Adicionalmente, el primer cuartil es igual a P_{25}, el segundo cuartil es igual a P_{50}, y P_{75} se encuentra en el tercer cuartil. Teniendo esto en mente, el cálculo de deciles y cuartiles se vuelve simplemente un asunto de determinación de los percentiles apropiados de acuerdo con las reglas que se acaban de establecer.

Una medida única de dispersión es el **rango o recorrido intercuartílico** (*Interquartile range* - RIQ). El RIQ es la diferencia entre el tercer cuartil y el primer cuartil. Es decir, $P_{75} - P_{25}$. La mitad de las observaciones se clasifican dentro de este rango. Consta del 50% de la mitad de las observaciones y corta el 25% inferior y el 25% superior de los puntos de datos. Como resultado, el RIQ proporciona una medida de dispersión que no está muy influenciada por unas cuantas observaciones extremas. El rango intercuartil se ilustra en la figura 3.1.

Figura 3.1
Recorrido intercuartílico



Ejercicios de la sección

11. Defina en sus propias palabras:
a. El percentil 80
b. El cuarto decil
c. El tercer cuartil

12. Utilizando los datos de la tabla 3.4, calcule e interprete:
a. El rango o recorrido intercuartílico.
b. La mediana
c. El percentil 50
d. El percentil 70
e. El tercer decil

3.7 Usos frecuentes de la desviación estándar

Como se ha enfatizado, la desviación estándar es útil para describir un conjunto de datos midiendo el grado de dispersión de las observaciones individuales alrededor de su media. Existen dos aplicaciones adicionales para la desviación estándar: (1) Teorema de Chebyshev y (2) la Regla empírica.

A. Teorema de Chebyshev

El teorema de Chebyshev (algunas veces escrito teorema de Tchebysheff) fue formulado por el matemático ruso P.L. Chebyshev (1821 – 1894). Establece que para *todo* conjunto de datos, por lo menos $1 - 1/K^2$ % de las observaciones están dentro de K desviaciones estándar de la media, en donde K es cualquier número mayor que 1. El teorema de Chebyshev se expresa como

Teorema de Chebyshev $$1 - \left[\frac{1}{K^2}\right] \qquad [3.16]$$

Así, por ejemplo, si se forma un intervalo de K = tres desviaciones estándar por encima de la media hasta tres desviaciones estándar por debajo de la media, entonces por lo menos

$$1 - \frac{1}{3^2} = 88.89\%$$

de todas las observaciones estarán dentro de dicho intervalo.

Ejemplo 3.7 Teorema de Chebyshev aplicado a Pigs & People Airlines

El trabajo anterior en este capítulo con datos agrupados para P&P Airlines, reveló una media de 78.7 pasajeros por día, con una desviación estándar de 12.14. Para programar los tiempos para una nueva ruta que abrió P&P, la gerencia desea saber con qué frecuencia los pasajeros están dentro de K = dos desviaciones estándar de la media, y cuál es dicho intervalo.

Solución

Si se transportan dos desviaciones estándar $(2 \times 12.14) = 24.28$ pasajeros por encima y por debajo de la media de 78.7, se tendrá un intervalo de $(78.7 - 24.28) = 54.42$ a $(78.7 + 24.28) = 102.98$ pasajeros. Se puede estar seguro de que por lo menos

$$1 - \frac{1}{2^2} = 75\%$$

del tiempo, el número de pasajeros diarios estuvo entre 54 y 103.

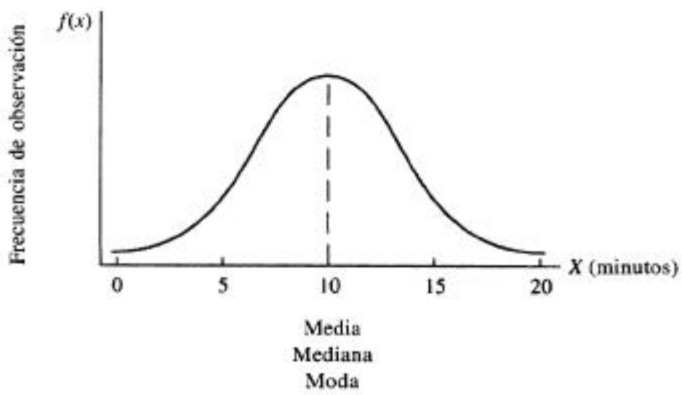
Interpretación

En por lo menos el 75% de los días (es decir, 75% de 50 es igual a 37 días), el número de pasajeros estuvo entre 54 y 103. Esto proporciona a la gerencia de P&P una valiosa información sobre para cuántos pasajeros deben prepararse en términos de operaciones en vuelo.

B. La distribución normal y la regla empírica

La desviación estándar puede utilizarse para sacar ciertas conclusiones si el conjunto de datos en cuestión está *distribuido normalmente*. El concepto de una distribución normal se encuentra usualmente en análisis estadístico y es de importancia considerable. Una discusión minuciosa de la distribución normal se presenta en capítulos posteriores. Sin embargo, una introducción a todo este concepto importante permitirá demostrar un uso práctico para la desviación estándar, y establecerá la base para una futura investigación más completa. Una **distribución normal** es una distribución de datos continuos (no discretos) que produce una curva simétrica en forma de campana, como la que muestra la figura 3.2

Figura 3.2
Una distribución normal



Se asume que se tiene un número grande de observaciones para el tiempo, en minutos, que le toma a los esquiadores terminar un trayecto en particular. Si los datos están distribuidos normalmente, una gráfica de la frecuencia con la cual ocurre cada observación tomará la forma de la figura 3.2. Las observaciones en cada extremo ocurrirán relativamente de forma poco frecuente, pero las observaciones que están más cerca de la mitad ocurrirán con una frecuencia alta, por tanto se produce la curva simétrica en forma de campana. La observación modal, 10 en este caso, es la que ocurre con la mayor frecuencia y por tanto está en el pico de la distribución. En una distribución normal, la media, la mediana y la moda son todas iguales.

Es de importancia que la mitad de las observaciones está por encima de la media y la mitad está por debajo. Esto significa que la mitad del área que está bajo la curva está a la izquierda de la media y la otra mitad del área que está debajo de la curva está a la derecha de la media.

Para ilustrar cómo se aplica la desviación estándar en la distribución normal, se asume que 1,000 esquiadores de slalom bajan una pendiente empinada en Vail. Los tiempos para todos los esquiadores parecen estar distribuidos normalmente, con una media de $\mu = 10$ minutos y una desviación estándar de $\sigma = 2$ minutos. La **regla empírica** dice que si se incluyen todas las observaciones que están a una desviación estándar de la media (una desviación estándar por encima de la media y una desviación estándar por debajo de la media) estas serán el 68.3% de todas las observaciones. Es decir, que no importa cuál es la media ni cuál es la desviación estándar, se puede estar seguro de que el 68.3% de las observaciones quedan a una desviación estándar de la media si las observaciones están distribuidas normalmente.

Debido a que el promedio de los esquiadores se toma 10 minutos para completar el trayecto, mover una desviación estándar (es decir, 2 minutos) por encima y por debajo de esta media de 10 produce un rango de 8 a 12 minutos. Así, de acuerdo con la regla empírica, 683 (68.3% de 1,000) esquiadores se tomaron entre 8 y 12 minutos para bajar la montaña.

Claro que si se mueve más de una desviación estándar por encima y por debajo de la media, se comprenderá un porcentaje más grande de observaciones. La regla empírica especifica que:

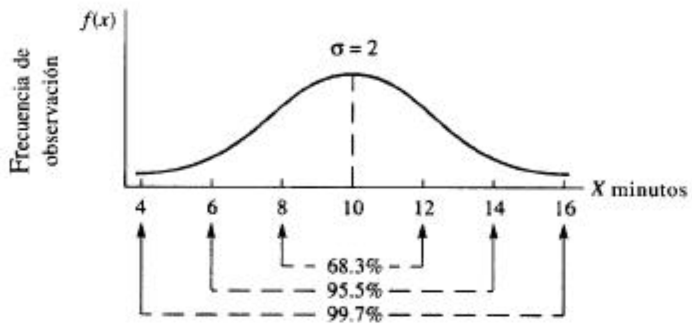
68.3% de las observaciones están dentro de más o menos una desviación estándar de la media

95.5% de las observaciones están dentro de más o menos dos desviaciones estándar de la media

99.7% de las observaciones están dentro de más o menos tres desviaciones estándar de la media.

Dados los tiempos de los esquiadores, una desviación estándar (2 minutos) por encima y por debajo de la media de 10 da un rango de 8 a 12 minutos. Dos desviaciones estándar (4 minutos) por encima y por debajo de la media de 10 da un rango de 6 a 14 minutos. Tres desviaciones estándar (6 minutos) da un rango de 4 a 16 minutos. Esto se muestra en la figura 3.3.

Figura 3.3
Tiempos distribuidos normalmente de los 1,000 esquiadores

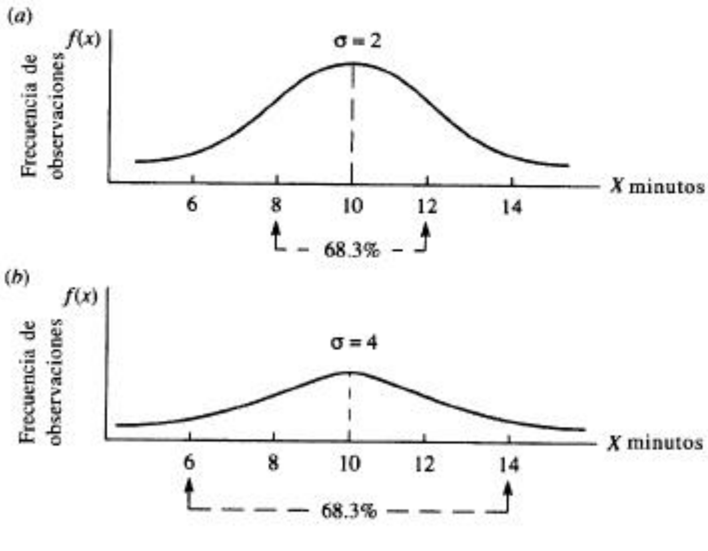


De acuerdo con la regla empírica, 997 de los 1,000 esquiadores se tomaron entre 4 y 16 minutos para terminar el trayecto. Así, sólo 3 de los 1,000 esquiadores fueron o muy buenos esquiadores y tomaron menos de 4 minutos o eran muy malos y se tomaron más de 16 minutos. Una observación de más de tres desviaciones estándar de la media (por encima o por debajo de ésta) es raro que ocurra y se da menos de 1% del tiempo si los datos están distribuidos normalmente.

También es importante recordar que la regla empírica describe el área total bajo la curva normal que se encuentra dentro de un rango dado. No sólo el 68.3% de todos los esquiadores se toman entre 8 y 12 minutos para bajar de forma segura la montaña, sino que, además, el 68.3% de toda el área que está bajo la curva normal está dentro del mismo rango de 8 a 12 minutos.

Si las observaciones están altamente dispersas, la curva en forma de campana se aplanará y se esparcirá. Se asume que un segundo grupo de esquiadores también hizo un promedio de 10 minutos, pero tuvo una desviación estándar de 4 minutos. Los tiempos del segundo grupo están más dispersos que los del primero. Los tiempos más rápidos en esquí estaban por debajo de 10, y los más lentos estaban muy por encima de 10, comparados con los del primer grupo. Esta dispersión mayor se reflejaría en una curva de distribución normal más extensa, tal y como se muestra en la figura 3.4

Figura 3.4
Dos distribuciones normales con medias iguales pero con desviaciones estándar diferentes

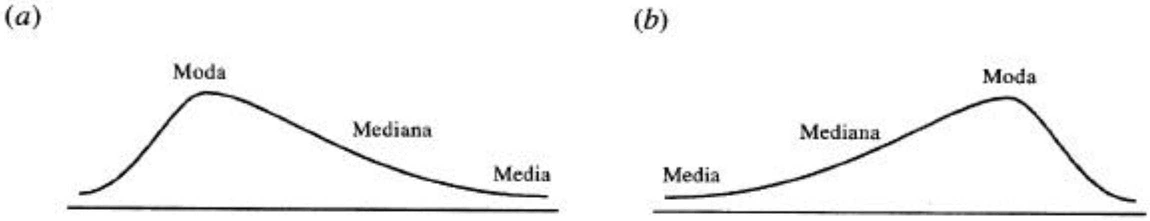


Ambas distribuciones están centradas en la media de $\mu = 10$ minutos, pero la que tiene la distribución con mayor $\sigma = 4$ minutos está más dispersa que el conjunto de observaciones con menos dispersión. Para abarcar el 68.3% de las observaciones en este grupo más disperso, es necesario incluir todas las que están dentro del intervalo de 6 a 14.

C. Sesgo

No todas las distribuciones son normales. Algunas están **sesgadas** a la izquierda o a la derecha. En la figura 3.5, se encuentran las curvas de distribución para el peso de las personas. En la figura 3.5 (a) se dice que la distribución está sesgada a la derecha. Parecería que las pocas personas más pesadas que están en el extremo superior en la escala de peso (quizá algunos hombres más grandes) halan la cola de la distribución hacia la derecha. En una segunda distribución de pesos que se muestra en la figura 3.5 (b), unas cuantas mujeres diminutas halan la distribución hacia el extremo inferior, haciendo que se desvíe hacia la izquierda.

Figura 3.5
Distribución sesgadadelpesode las personas



En ambos casos, la moda es por definición la observación que ocurre con mayor frecuencia. Por tanto, está en el pico de la distribución. Sin embargo, como se dijo anteriormente, por su sola naturaleza, la media se ve más afectada por las observaciones extremas. Por tanto, es halada en la dirección del sesgo, más de lo que está la mediana, la cual está en algún sitio entre la media y la moda.

El sesgo puede medirse mediante el **coeficiente de sesgo de Pearson**.

Coeficiente de sesgo	$P = \frac{3(\bar{X} - \text{mediana})}{s}$	[3.17]

Si $P < 0$, los datos están sesgados a la izquierda, si $P > 0$, entonces están sesgados a la derecha; si $P = 0$ están distribuidos normalmente.

Ejemplo 3.8 Sesgo para los pasajeros de P&P

Utilizando los datos agrupados de la lista de pasajeros de P&P, se calcula $\bar{X} = 78.7$, $s = 12.14$, y la mediana = 78.33. Dada esta información, el director ejecutivo de P&P puede ver claramente que los datos están sesgados a la derecha, debido a que la media excede a la mediana. Además, también desea una medida del grado de sesgo.

Solución

Se tiene

$$P = \frac{3(78.7 - 78.33)}{12.14} = 0.03$$

Interpretación

Debido a que P > 0, los datos para P&P están, como se presumió, sesgados a la derecha. El grado hasta el cual están sesgados se refleja en el valor del coeficiente de Pearson. Si se fuera a hacer la gráfica de los datos, aparecerían como en la figura 3.5 (*a*).

D. Coeficiente de variación

Como se ha enfatizado, un uso importante de la desviación estándar es servir como medida de dispersión. Sin embargo, se aplican ciertas limitaciones. Cuando se consideran dos o más distribuciones que tienen medias significativamente diferentes, o que están medidas en unidades distintas, es peligroso sacar conclusiones respecto a la dispersión sólo con base en la desviación estándar. Es como violar el viejo adagio sobre la comparación entre manzanas y naranjas.

Por tanto, con frecuencia debemos considerar el **Coeficiente de Variación** (CV), el cual sirve como medida relativa de dispersión. El coeficiente de variación determina el grado de dispersión de un conjunto de datos relativo a su media. Se calcula dividiendo la desviación estándar de una distribución por su media y multiplicando por 100.

Coeficiente de variación	$CV = \frac{s}{\bar{X}}(100)$	[3.18]

Los datos agrupados para P&P reportaron una media de 78.7 pasajeros por día, con una desviación estándar de 12.14 pasajeros. Se supone que P&P también recolecta datos sobre el mismo período para el número de millas que la aerolínea voló y dicha media y desviación estándar ha probado ser de 1,267.5 y 152.7, respectivamente. La desviación estándar más alta para las millas voladas puede sugerir que estos datos presentan una variación mucho mayor.

Sin embargo, si se calcula el coeficiente de variación para pasajeros, se encuentra que es

$$CV = \frac{12.14}{78.70}(100) = 15.43$$

mientras que para las millas es solamente

$$CV = \frac{152.7}{1267.5}(100) = 12.05$$

Es claro que al comparar la variación en dos conjuntos de datos muy diferentes, es sabio utilizar el coeficiente de variación y no sólo la desviación estándar.

Ejercicios de la sección

13. Un conjunto de datos sobre el peso del contenido de 1000 bolsas de comida para perros marca Happy Pooch tiene una media de 50 libras y una desviación estándar de 2.3 libras. No se sabe si los datos están distribuidos normalmente. Los fabricantes de Happy Pooch esperan que por lo menos 750 de tales bolsas pesen entre 45.4 y 54.6 libras. ¿Qué seguridad puede darles?

14. Un conjunto de datos distribuidos normalmente tiene una media de 5,000 y una desviación estándar de 450. Qué porcentaje de las observaciones están:

a. ¿Entre 4550 y 5450?

b. ¿Entre 4100 y 5900?

c. ¿Entre 3650 y 6350?

d. ¿Por encima de 6350?

e. ¿Por debajo de 4550?

15. Shard Lumber corta troncos a una longitud media de 20 pies, con una desviación estándar de 3.5 pies. Si los cortes están distribuidos normalmente, qué porcentaje de troncos tienen menos de:

a. ¿16.5 pies?

b. ¿13 pies?

16. Datos sobre las edades de los 100 mejores ejecutivos de las mejores 500 firmas de la revista Fortune revelan una edad media de 56.2 años y una desviación estándar de 12.7 años. Su ingreso medio es US$89,432, con s= US$16,097. ¿Cuál variable, edad o ingreso, presenta la mayor variación?

17. Si el ingreso medio del ejercicio 16 es de $87,567 dólares,

a. ¿Los datos están sesgados a la derecha, a la izquierda o distribuidos normalmente?

b. ¿Cuál es el coeficiente de sesgo de Pearson?

Problemas resueltos

1. La información paga Una emisión del año 1996 de *Inc. Technology* reportó que *Information Please*, una nueva firma que vende información por internet, registró el siguiente número de "topes" que tuvo cada hora 47, 52, 69, 72, 112, 36, 28, 68, 69 y 41. Suponga que el gerente desea evaluar estos datos con base en sus medidas de tendencia central y de dispersión.

La media es

$$\bar{X} = \frac{\Sigma X_i}{n} = \frac{594}{10} = 59.4$$

Dada la serie ordenada 28, 36, 41, 47, 52, 68, 69, 69, 72, 112, la posición de la mediana es

$$\text{Posición de la mediana} = \frac{n+1}{2} = 5.5$$

y la mediana es el promedio de las observaciones quinta y sexta o 52 + 68 / 2 = 60. La moda es la observación que ocurrió más frecuentemente, 69. La varianza es

$$s^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1} = \frac{5264.4}{9} = 584.933$$

y la desviación estándar es $\sqrt{584.933} = 24.185$.

2. **Pesaje** Los miembros de Heart-to-Heart Health Club deben pagar cuotas con base en su peso promedio. De los 60 miembros, 12 pesaron 110 libras, 25 pesaron 120 libras, 18 hicieron girar la balanza hasta 150 y el resto registraron 180 libras. Si los miembros deben pagar US$5 por cada libra que pesan en promedio, ¿cuánto debe desembolsar cada uno?

Debido a que cada categoría de peso contiene un número diferente de miembros, la media ponderada es necesaria.

Libras *(X)*	Número de miembros *(W)*	*XW*
110	12	1320
120	25	3000
150	18	2700
180	5	900
	60 = ΣW	7920 = ΣXW

$$\bar{X}_w = \frac{\Sigma XW}{\Sigma W} = 132 \text{ libras}$$

Así, cada miembro debe pagar 132 3 5 = US$660

3. **Dificultades iniciales** El descontento de los empleados de Bates Electronics se refleja en el número de quejas oficiales durante lo últimos cuatro meses: 23, 41, 37 y 49. Con base en estos datos, ¿cuál es el incremento promedio mensual en las quejas?

La media geométrica, cuyo trabajo es determinar el cambio porcentual promedio en el tiempo, requiere que primero se determine qué porcentaje de cada número es del mes anterior:

Mes	Quejas	Porcentaje con respecto al mes anterior
1	23	--
2	41	1.78
3	37	0.90
4	49	1.32

Entonces,

$$\text{MG} = \sqrt[n]{(X_1)(X_2)\ldots(X_n)} = \sqrt[3]{(1.78)(0.9)(1.32)} = 128$$

Entonces, 1.28 - 100 = 28% de incremento mensual medio en quejas.

4. **Personas que cuidan** Los estadísticos del programa de Meals on Wheels (comida sobre ruedas), el cual lleva comidas calientes a enfermos confinados en casa, desean evaluar sus servicios. El número de comidas diarias que suministran aparece en la siguiente tabla de frecuencia.

Número de comidas por día	Número de días	M	fM	fM^2
0-5	3	2.5	7.5	18,75
6-11	6	8.5	51.0	433.50
12-17	5	14.5	72.5	1,051.25
18-23	8	20.5	164.0	3,362.00
24-29	2	26.5	53.0	1,404.50
30-35	3	32.5	97.5	3,168.75
	27		445.5	9,438.75

La media, la mediana y el número modal de comidas son

$$\bar{X}_g = \frac{\Sigma fM}{n} = \frac{445.5}{27} = 16.5 \text{ comidas por día}$$

$$\begin{aligned}
\text{Mediana} &= L_{md} + \left[\frac{n/2 - F}{f_{md}}\right](c) \\
&= 12 + \left[\frac{27/2 - 9}{5}\right](6) \\
&= 17.4 \text{ comidas}
\end{aligned}$$

$$\begin{aligned}
\text{Moda} &= L_{mo} + \left[\frac{D_a}{D_a + D_b}\right](c) \\
&= 18 + \left[\frac{3}{3+6}\right](6) \\
&= 20 \text{ comidas}
\end{aligned}$$

La varianza y la desviación estándar son

$$\begin{aligned}
s^2 &= \frac{\Sigma fM^2 - n\bar{X}^2}{n-1} \\
&= \frac{9438.75 - 27(16.5)^2}{27-1} \\
&= 80.31
\end{aligned}$$

$$s = \sqrt{80.31} = 8.96 \text{ comidas}$$

Lista de fórmulas

[3.1] $$\mu = \frac{X_1 + X_2 + X_3 + \cdots + X_N}{N} = \frac{\sum_{i=1}^{N} X_i}{N}$$ Media poblacional

[3.2] $$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$ Media muestral

[3.3] $$\text{Posición de la mediana} = \frac{n+1}{2}$$ Determina la posición de la mediana en una serie ordenada

[3.4]	$\bar{X}_w = \dfrac{\Sigma XW}{\Sigma W}$	Determina la media ponderada
[3.5]	$MG = \sqrt[n]{(X_1)(X_2)\ldots(X_n)}$	Calcula el cambio porcentual promedio
[3.6]	$\sigma^2 = \dfrac{\Sigma(X_i - \mu)^2}{N}$	Varianza poblacional
[3.7]	$\sigma = \sqrt{\sigma^2}$	Desviación estándar poblacional
[3.8]	$s^2 = \dfrac{\Sigma(X_i - \bar{X})^2}{n}$	Varianza muestral
[3.9]	$s = \sqrt{s^2}$	Desviación estándar muestral
[3.10]	$\bar{X}_g = \dfrac{\Sigma fM}{n}$	Media para datos agrupados
[3.11]	$\text{Mediana} = L_{md} + \left[\dfrac{n/2 - F}{f_{md}}\right](C)$	Mediana para datos agrupados
[3.12]	$\text{Moda} = L_{mo} + \left[\dfrac{D_a}{D_a - D_b}\right](C)$	Moda para datos agrupados
[3.13]	$s^2 = \dfrac{\Sigma fM^2 - n\bar{X}^2}{n - 1}$	Varianza muestral para datos agrupados
[3.14]	$s = \sqrt{s^2}$	Desviación estándar para datos agrupados
[3.15]	$L_p = (n + 1)\dfrac{P}{100}$	Ubicación de un percentil
[3.16]	$1 - \left[\dfrac{1}{K^2}\right]$	Teorema de Chebyshev
[3.17]	$P = \dfrac{3(\bar{X} - \text{mediana})}{s}$	Coeficiente de sesgo
[3.18]	$CV = \dfrac{s}{\bar{X}}(100)$	Coeficiente de variación

Ejercicios del capítulo

18. Los precios de las acciones están cotizados en octavos de dólar de manera que por ejemplo, 5 1/8 es US$5,125, 5 ¼ es US$5.25, 5 3/8 es US$5,375 y así sucesivamente hasta 5 7/8, que es US$5,875. A continuación se da una muestra de siete precios de cierre de las acciones tomadas de *The Wall Street Journal* de octubre 8 de 1997.

Walt Mart	27 3/8	General Mills	69 7/8
Disney	42 5/8	Toys R Us	38 5/8
Mobil	69 7/8	Dow Jones	29 1/4
General Motors	39 1/2		

a. Calcule la media, la mediana, y la moda. Interprete cada estadístico. ¿Qué le dice cada una? ¿Por qué son diferentes si todas son promedios?

b. Calcule e interprete la varianza y la desviación estándar.

c. Calcule e interprete el rango o recorrido intercuartílico.

d. Calcule e interprete el percentil cuarenta.

19. The Snowflake comercializa botas para esquiar en San Luis Obispo, California. De los últimos 100 pares vendidos, 4 eran talla 9, 33 talla 9 ½, 26 talla 10, 29 talla 10 ½ y 8 eran talla 13. Haga comentarios sobre el uso de la media, la mediana y la moda como medidas de tendencia central y el uso de cada una en la toma de decisiones sobre los tamaños que se deben tener en inventario. Calcule cada medida.

20. Debido a que las tasas de interés cayeron a comienzos de 1997, se encontró que una muestra de las tasas hipotecarias para hipotecas a 15 años de las instituciones de crédito en Peoria, Illinois era

7.1%, 7.3%, 7.0%, 6.9%, 6.6%, 6.9%, 6.5%, 7.3%, 6.85%

a. Calcule e interprete la media, la mediana y la moda.

b. ¿Estos datos están sesgados a la izquierda, a la derecha, o están distribuidos normalmente? Calcule el coeficiente de Pearson como medida de sesgo.

c. Calcule e interprete la varianza y la desviación estándar.

21. Una encuesta de instituciones de crédito en un centro urbano cerca de Peoria (ver problema anterior) reveló tasas de crédito hipotecario de

7.1%, 7.3%, 6.3%, 6.7%, 6.8%, 6.85%, 7.5%

a. ¿Las tasas de crédito hipotecario son más altas en Peoria o en otros centros urbanos?

b. ¿Cuál ciudad parece tener las tasas de interés más consistentes entre las instituciones?

c. Calcule e interprete el coeficiente de sesgo de Pearson.

22. Alan Munday fabrica una pintura sellante para automóviles en el área de Denver. Él utiliza cuatro químicos diferentes en el proceso de producción. Para hacer su producto, Munday debe utilizar 2 galones de calcimina que cuesta US$2.50 el galón, ½ galón de kalsolita a US$1.25 por galón, 1 galón de aglutinante que cuesta US$0.75 el galón, y 3 galones de aceite secante a US$2.00 por galón. Calcule el costo de un galón de sellante.

23. La emisión de la revista *Business Week* del 31 de mayo de 1997 reportó que el número de transacciones en miles de millones de dólares realizadas en las instalaciones bancarias de la nación ATM fueron

1991	3.9	1994	4.5
1992	4.1	1995	6.5
1993	4.3	1996	6.5

La industria bancaria intenta prepararse para 8 miles de millones de transacciones para el año de 1998. ¿Será suficiente para manejar el nivel de actividad que usted pronostica para ese año?

24. The Noah Fence Company vende cuatro tipos de cercas a los barrios residenciales de las afueras de la ciudad. El grado A le cuesta a Noah US$5.00 por pie lineal de instalación, el grado B cuesta US$3.50, el grado C cuesta US$2.50, y el grado D cuesta US$2.00. Ayer, Noah instaló 100 yardas del grado A, 150 del grado B, 75 yardas del grado C y 200 yardas del grado D. ¿Cuál fue el costo promedio de instalación por pie lineal?

25. Una muestra de los recibos de ventas semanales para Pig-In-A-Poke Bar-B-Q está en cientos de dólares,

43.3, 54.2, 34.8, 42.9, 49.2, 29.5, 28.6

Se implementa un programa publicitario diseñado para emparejar las ventas. Una muestra subsiguiente de ventas es

45.5, 39.5, 35.7, 36.7, 42.6, 42.14

¿La campaña publicitaria logró su meta de suavizar las ventas semanales?

26. Bill Karl compró 20 acciones a US$15 cada una, 50 acciones a US$20 cada una, 100 acciones a US$30 cada una y 75 acciones a US$35 cada una.

a. ¿Cuál es el monto total de su inversión?

b. ¿Cuál es el precio promedio por acción?

27. Las edades de cincuenta de los directores ejecutivos de las mejores corporaciones de la nación reportadas en la edición de la revista *Forbes* de la edición del 24 de mayo de 1997 aparecen en la siguiente tabla de frecuencias.

a. Calcule e interprete la media, la mediana y la moda.

b. Calcule e interprete la varianza y la desviación estándar.

Edades	Frecuencias
50 y menos de 55	8
55 y menos de 60	13
60 y menos de 65	15
65 y menos de 70	10
70 y menos de 75	3
75 y menos de 80	1

28. La misma edición de la revista *Forbes* (que se vio en el problema anterior) también proporcionó datos sobre los salarios en miles de dólares. Resultó la siguiente tabla de frecuencias:

Salario (en miles de dólares)	Frecuencias
90 y menos de 440	9
440 y menos de 790	11
790 y menos de 1140	10
1140 y menos de 1490	8
1490 y menos de 1840	4
1840 y menos de 2190	3
2190 y menos de 2540	5

a. Calcule la media, la mediana y la moda. Interprete sus respuestas.

b. ¿Los salarios están tan dispersos como las edades del problema anterior?

29. Janna Vice utiliza dos máquinas diferentes para producir papeleras para las fotocopiadoras Kodak. Una muestra de las papeleras de la primera máquina midieron 12.2, 11.9, 11.8, 12.1, 11.9, 12.4, 11.3 y 12.3 pulgadas. Las bandejas elaboradas con la segunda máquina midieron 12.2, 11.9, 11.5, 12.1, 12.2, 11.9 y 11.8 pulgadas. Janna debe utilizar la máquina con la mayor consistencia en los tamaños de las papeleras. ¿Cuál máquina debe utilizar?

30. Los puntajes de las dos primeras pruebas de estadística que usted presentó estaban distribuidos normalmente y reportaron medias de 90 para la prueba A y 50 para la prueba B. ¿Usted espera una desviación estándar más alta o más baja para la prueba A? ¿Le gustaría una desviación estándar más alta o más baja para la prueba B si siente que le fue bien en la prueba? ¿Por qué? Haga una gráfica que ilustre la lógica de sus respuestas.

31. Los siguientes datos de muestras se han obtenido para el número de clientes diarios en Rosie's Flower Shoppe:

34, 45, 23, 34, 26, 32, 31, 41

Calcule la varianza, la desviación estándar y el rango o recorrido intercuartílico.

32. La siguiente es una muestra de las ganancias por acción en dólares, para las acciones cotizadas en la Bolsa de Valores de Nueva York:

1.12, 1.43, 2.17, -1.19, 2.87, -1.49

Calcule la varianza, la desviación estándar y el rango o recorrido intercuartílico. Interprete cada uno.

33. Las horas trabajadas por Ronnie cada semana durante los últimos dos meses son

52 48 37 54 48 15 42 12

Asumiendo que estos son datos muestrales, calcule:

a. La media
b. La mediana
c. La moda
d. ¿Cuál es probablemente una mejor medida para el punto central?

34. Utilizando las horas de trabajo de Ronnie del problema anterior, calcule e interprete:

a. El rango
b. La varianza
c. La desviación estándar
d. El primer cuartil
e. El percentil 25
f. El rango o recorrido intercuartílico

35. Quienes ponen los discos en KAYS claman que ponen más canciones cada hora que sus rivales de la KROC del otro pueblo. Durante las últimas 24 horas se recolectaron y tabularon los datos sobre el número de canciones puestas por ambas estaciones. Utilice los datos para preparar un reporte que compare las dos estaciones. Su reporte terminado debe presentarse a la Comisión Federal de Comunicaciones, y debe contener referencias respecto a las medidas de tendencia central y de dispersión.

Número de canciones por hora	KAYS	KROC
5-10	2	4
11-16	4	5
17-22	6	7
23-28	8	5
29-34	2	2
35-40	2	1

36. *The Wall Street Journal* describió una disputa entre la gerencia y el sindicato de trabajo local respecto a la eficiencia y productividad de los trabajadores. La gerencia argumentaba que a los empleados les tomaba más de 20 minutos terminar cierto trabajo. Si se mide el tiempo de 85 empleados, arrojando los resultados tabulados, con base en esta muestra, ¿la gerencia está en lo correcto? Calcule las tres medidas de tendencia central.

Clase (número de minutos)	Frecuencia (número de empleados)
5 y menos de 7	2
7 y menos de 9	8
9 y menos de 11	10
11 y menos de 13	15
13 y menos de 15	17
15 y menos de 17	14
17 y menos de 19	7
19 y menos de 21	9
21 y menos de 23	3

37. En el ejercicio anterior, la gerencia también se encuentra preocupada porque el desempeño de los empleados es demasiado errático; no existe mucha variación en la cantidad de tiempo que toma a los trabajadores completar un trabajo. Identifique y calcule el estadístico apropiado de acuerdo con la preocupación de la gerencia.

38. Dados los siguientes puntajes de 9 pruebas para la clase de economía del profesor Pundit, calcule el coeficiente de sesgo de Pearson. Asuma que estos son datos muestrales.

80 83 87 85 90 86 84 82 88

39. Los sindicalistas de la planta de Ford Motor Company en Toledo argumentan que, en contravención del contrato laboral, los trabajadores de la línea de producción hacen un promedio salarial por hora menor con una mayor variabilidad que los trabajadores de oficina. Una muestra de $n = 10$ se toma de cada clase de trabajadores, arrojando los siguiente valores. ¿Tales valores apoyan a los sindicalistas?

Trabajadores	Trabajadores de producción	Trabajadores de oficina
1	12.15	15.12
2	18.17	18.42
3	19.42	17.12
4	15.17	16.92
5	18.63	18.15
6	16.42	15.81
7	15.49	19.12
8	18.73	19.15
9	19.12	18.73
10	18.36	19.66

40. Dos marcas de zapatos para correr fueron evaluados en cuanto a uso y desgaste. Cada una reportó los siguientes números de horas de uso antes que se detectara un desgaste significativo.

Marca A	Marca B
97	78
83	56
75	87
82	54
98	89
65	65
75	

a. ¿Cuál zapato parece presentar mayor desgaste?

b. ¿Cuál zapato parece tener un programa de control de calidad que produzca la mejor consistencia en su desgaste?

41. Manly Bankford trabaja como corredor para E. F. Hutton. Sus registros muestran que las tasas de rendimiento (en porcentaje) sobre dos valores para 10 meses seleccionados fueron de

Valor 1	5.6	7.2	6.3	6.3	7.1
	8.2	7.9	5.3	6.2	6.2
Valor 2	7.5	7.3	6.2	8.3	8.2
	8.0	8.1	7.3	5.9	5.3

a. ¿Cuál valor puede ser mejor para los clientes que están interesados en un rendimiento más alto?

b. ¿Cuál valor debería aconsejar Manly a sus clientes que prefieren menos riesgo?

42 Aquí se muestran las relaciones precio-ganancia para 30 acciones diferentes transadas en la Bolsa de Valores de Nueva York (*New York Stock Exchange* - NYSE)

4.8	5.2	7.6	5.7	6.2	6.6	7.5	8.0	9.0	7.7
3.7	7.3	6.7	7.7	8.2	9.2	8.3	7.3	8.2	6.5
5.4	9.3	10.0	7.3	8.2	9.7	8.4	4.7	7.4	8.3

a. Calcule la media y la desviación estándar.

b. De acuerdo con el teorema de Chebyshev, por lo menos ¿cuántas relaciones precio-ganancia están dentro de dos desviaciones estándar de la media?

c. ¿Cuántas están realmente a dos desviaciones estándar de la media?

43. La mecánica local en Vinney's Auto Shop y Charm School le dice que las reparaciones de su carro le costarán US$714.12. Los datos de la industria muestran que la cuenta promedio por reparaciones parecidas a las suyas es de US$615, con una desviación estándar de US$31. ¿Qué puede concluir sobre las tasas de Vinney's si usted asume que las reparaciones están distribuidas normalmente?

44. Dada la siguiente tabla de frecuencias de las ventas mensuales en dólares para equipos de paracaidismo en el mercado del sur de California (las cifras están en cientos).

Clase (en US$100's)	Número de meses
5 y menos de 10	5
10 y menos de 15	7
15 y menos de 20	9
20 y menos de 25	10
25 y menos de 30	8
30 y menos de 35	3
35 y menos de 40	2

a. Usted es estadístico jefe para Bounce Twice Parachute Company, y su gerente solicita un desglose de la frecuencia de ventas. Él está interesado en el valor debajo del cual están a lo sumo el 60% de las observaciones, junto con un desglose completo en cuartiles.

b. Además, usted siente que sería de utilidad determinar los valores de los percentiles 10 y 90.

45. Una supervisora en una planta ensambladora recibió las siguientes clasificaciones de eficiencia durante 12 meses

$$56, 69, 48, 75, 65, 72, 81, 43, 61, 42, 36, 52,$$

a. Si ella desea crear la impresión más favorable, ¿debería reportar la media, la mediana o la moda en su autoevaluación anual?

b. ¿Qué tan consistentes han sido sus medidas?

Ejercicios por computador

Entre al archivo OUTPUT en su disco de datos. Contiene 100 observaciones de la producción semanal de Leisure Sports, un fabricante de productos deportivos con sede en Cleveland. Ernst Rawls, Director de Operaciones Corporativas está preocupado por cumplir con el itinerario de producción establecido en la asamblea de noviembre pasado que busca una producción semanal de por lo menos 1000 unidades. El señor Rawls también debe librarse de grandes variaciones en la producción de una semana a otra.

Utilizando el software de su computador, prepare un informe para el Sr. Rawls tal y como se describe en el apéndice 1. Incluya todos los estadísticos descriptivos, análisis de descubrimientos y sus conclusiones y recomendaciones.

PUESTA EN ESCENA

En el planteamiento del Escenario a comienzos de este capítulo usted debía funcionar como analista de inversiones para evaluar los tres fondos ofrecidos por Janus. Asuma que los fondos dieron los rendimientos aquí mostrados.

Año	Riesgo	Ingresos	Desarrollo
1	14.2%	9.2%	22.2%
2	9.2	10.5	15.1
3	19.9	11.5	10.5
4	21.5	12.4	10.8
5	22.8	15.8	11.8
6	25.1	17.2	12.8

Utilizando las herramientas descriptivas presentadas en este capítulo, proporcione una descripción comparativa de cada fondo. Incluya todas las medidas de tendencia central y las medidas de dispersión. ¿Qué conclusiones puede sacar respecto al desempeño de cada fondo?

Si su cliente invirtió el 30% de su cartera en fondos de riesgo, el 50% en fondos de ingreso, y el resto en fondos de desarrollo, ¿cuál sería el rendimiento anual promedio? ¿Qué rendimientos proyectaría para cada fondo en el año 7?

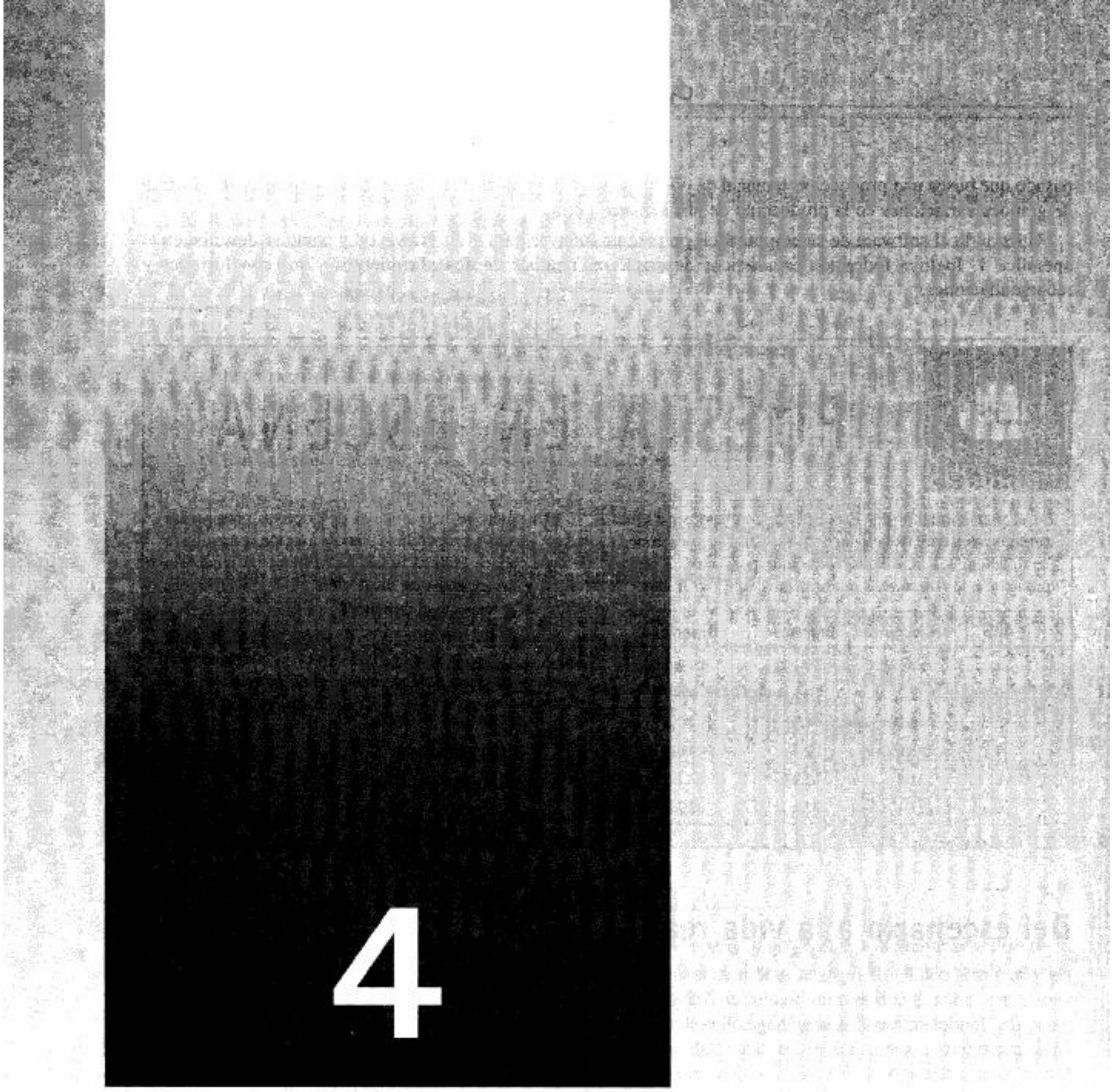
Del escenario a la vida real

La forma más común de reportar rendimientos de fondos mutuos es con rendimientos a un año, tres años, cinco años y diez años. Se observarán fondos en Vanguard Group (*www.vanguard.com*). En Home Page, seleccione el icono de "fondos mutuos". Luego haga clic sobre "fondos por categoría". Encontrará rendimientos a un año, tres años, cinco años y diez años para fondos individuales junto con el índice de Standard & Poor's. The Vanguard Funds están agrupados en Desarrollo, Desarrollo e Ingreso, y otras categorías. Seleccione 5 fondos específicos de cada una de las categorías de Desarrollo y de Desarrollo e Ingreso y los datos de Standard & Poor's en este sitio. Utilizando las herramientas descriptivas de este capítulo y del capítulo 2, proporcione descripciones comparativas de cada fondo con los resultados de Standard & Poor's.

Después de que usted haya terminado sus análisis, usted puede querer comparar sus técnicas con los análisis proporcionados por algunos sitios populares de la "Guía de Inversionistas":

Net Worth Mutual Fund Market Manager	*http://networth.quicken.com/investments*
Mutual Fund Investor's Center	*www.mfea.com*
Tradeline Mutual Fund Center	*http://nestegg.iddis.com*

Estos sitios tienen enlaces con los home pages de importantes empresas de fondos. Usted puede querer comparar Vanguard Funds con los de otra compañía como Fidelity o Dreyfus.

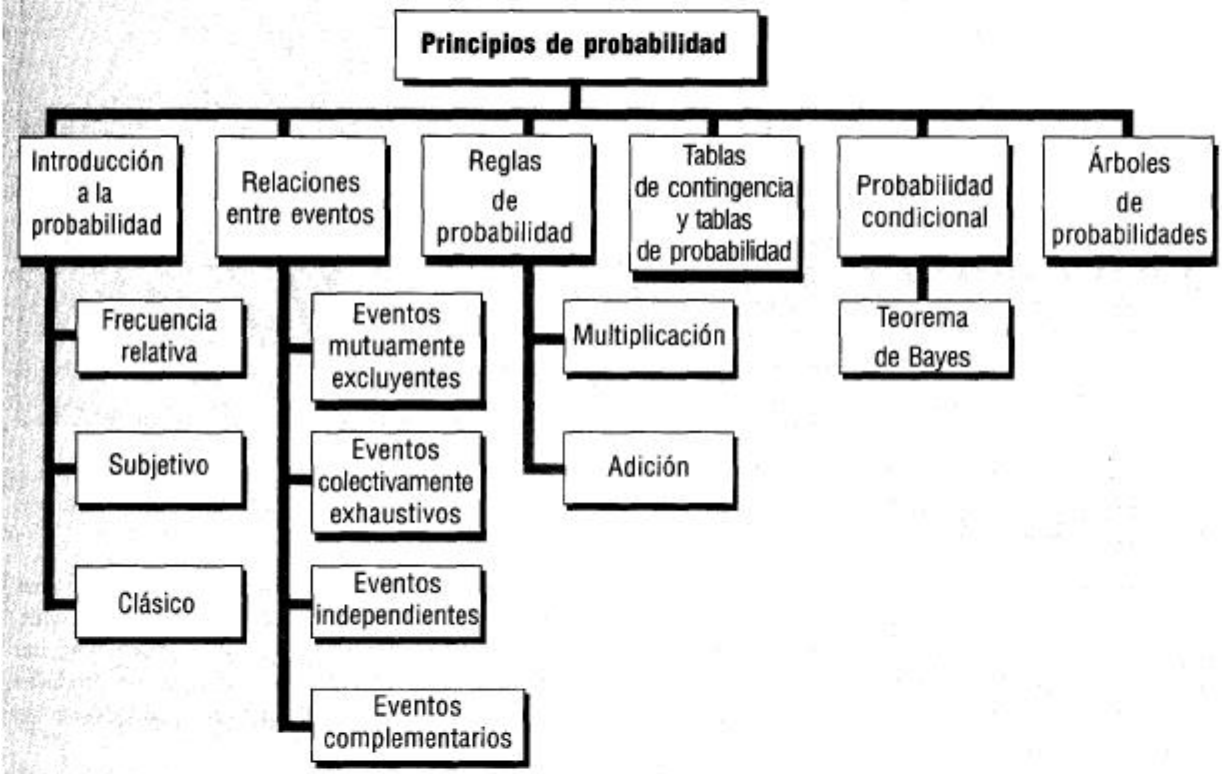


4

Principios de probabilidad

Plan del capítulo

Muchas decisiones administrativas se apoyan en los posibles resultados de tales decisiones. Este capítulo analiza las formas como se puede determinar la probabilidad relacionada con tales resultados. Al establecer la posibilidad o probabilidad de eventos futuros, se puede reducir ampliamente el riesgo en el proceso de toma de decisiones.



ESCENARIO

La Asociación Nacional de Esquí estudió el impacto financiero de la ubicación de los 850 centros de esquí de Estados Unidos (*Forbes*, mayo 1996). El propósito era determinar si un centro ubicado cerca de un centro urbano atraía más esquiadores o tenía más utilidades que un centro más aislado. La comparación incluía también los centros de esquí situados muy cerca de otros centros–similares llamados concentración de centros de esquí. Michael Berry, presidente de la Asociación Nacional de Áreas de Esquí expresó: "Muchas zonas para esquiar enfrentan una alta probabilidad de quedar en bancarrota en las estaciones venideras".

Con base en este estudio, es posible identificar los centros cuyas posiciones financieras tienen más probabilidad de presentar un descenso, y permitirles tomar las acciones correctivas que permitan aumentar sus ingresos económicos.

Como consultor contratado por Berry usted tiene la labor de proporcionar una evaluación comparativa sobre el futuro de tales centros y su potencial de éxito. Esta labor necesitará de toda su experiencia en principios de probabilidad.

4.1 Introducción

Sin tener en cuenta la profesión que se haya elegido, algo sí es seguro: en algún momento se han de tomar decisiones. Con mucha frecuencia esto tendrá que hacerse sin conocer todas las consecuencias de tales decisiones. Por ejemplo, los inversionistas deben decidir sobre la conveniencia de invertir en una acción en particular, con base en sus expectativas sobre rendimientos futuros. Los empresarios, al decidir comercializar un producto enfrentan la incertidumbre sobre la posibilidad de éxito. En cada caso, como sucede con la mayoría de los asuntos comerciales, se han de tomar decisiones sin toda la información pertinente.

Todo esfuerzo por reducir el nivel de incertidumbre en el proceso de toma de decisiones incrementará enormemente la probabilidad de que se tomen decisiones más inteligentes y bien informadas. El propósito de este capítulo es ilustrar las formas en las cuales puede medirse la posibilidad o probabilidad de ocurrencia de eventos futuros. Al mejorar la habilidad para juzgar la ocurrencia de eventos futuros, se puede minimizar el riesgo y la especulación arriesgada relacionadas con el proceso de toma de decisiones.

4.2 Experimentos, resultados y conjuntos

La probabilidad es la posibilidad numérica de que ocurra un evento. La probabilidad de un evento es medida por valores comprendidos entre 0 y 1. Entre mayor sea la probabilidad de que ocurra un evento, su probabilidad asignada estará más próxima a 1. La probabilidad de certeza es 1. La probabilidad de una imposibilidad es 0. Esto se podría expresar así:

$$P\text{ (evento cierto)} = 1$$

$$P\text{ (evento imposible)} = 0$$

$$\text{Por tanto, } 0 \leq P(E_i) \leq 1$$

en donde E_i es algún evento.

La probabilidad de que el sol salga mañana es muy alta – muy cercana a 1. La probabilidad de que se apruebe este curso sin estudiar, estando en el otro extremo, está muy cerca de cero.

Probabilidad Es la posibilidad numérica, medida entre 0 y 1, de que ocurra un evento.

El proceso que produce un evento es denominado **experimento**. Un experimento es toda acción bien definida que conlleva a un **resultado** único bien definido. Lanzar un dado (mitad de un par de dados) es un experimento. Un resultado bien definido es un número de 1 a 6. Un experimento puede consistir en revisar un producto para determinar si cumple con ciertas especificaciones de fabricación. El resultado es o (1) defectuoso o (2) no defectuoso.

El conjunto de todos los posibles resultados para un experimento es el **espacio muestral**. El espacio muestral para lanzar un dado es

$$SS = (1, 2, 3, 4, 5, 6)$$

mientras que el espacio muestral para el experimento de lanzar una moneda al aire es:

$$SS = (\text{caras, sellos})$$

La probabilidad de que al menos uno de los eventos que están en el espacio muestral ocurra es igual a 1. Si se lanza un dado, el resultado debe ser un número entre 1 y 6. Debido a que esto es una certeza puede decirse que

$$\sum P(E_i) = 1$$

4.3 Introducción a la probabilidad

En la actualidad vemos que la teoría de la probabilidad ocupa un lugar importante en muchos asuntos de negocios. Los seguros y prácticas actuariales se basan firmemente en los principios de la teoría de la probabilidad. Las pólizas de seguros de vida dependen de las tablas de mortalidad, las cuales a su vez se basan en las probabilidades de muerte en edades específicas. Otras tasas de seguros tales como seguro de bienes raíces y de automóviles se determinan de manera similar. La probabilidad también juega un papel importante en la estimación del número de unidades defectuosas en un proceso de fabricación, la probabilidad de recibir pagos sobre cuentas por cobrar y las ventas potenciales de un nuevo producto. Incluso los apostadores profesionales en eventos deportivos deben tener una comprensión sólida de la teoría de la probabilidad.

A pesar de la difundida aplicación de los principios de la probabilidad, existen sólo tres formas generalmente aceptadas para enfocar: (1) el modelo de frecuencia relativa (*a posteriori*), (2) el modelo subjetivo y (3) el modelo clásico (*o a priori*).

El **modelo de frecuencia relativa** utiliza datos que se han observado empíricamente, registra la frecuencia con que ha ocurrido algún evento en el pasado y estima la probabilidad de que el evento ocurra nuevamente con base en estos datos históricos. La probabilidad de un evento con base en el modelo de frecuencia relativa se determina mediante

Frecuencia relativa
$$P(E) = \frac{\text{Número de veces que ha ocurrido el evento en el pasado}}{\text{Número total de observaciones}} \qquad [4.1]$$

Por ejemplo, asumiendo que durante el año anterior hubo 50 nacimientos en un hospital local, de los cuales 32 de los recién nacidos eran niñas. El modelo de frecuencia relativa revela que la probabilidad de que el siguiente nacimiento (o un nacimiento seleccionado aleatoriamente) sea una niña es

$$\text{P(niña)} = \frac{\text{Número de niñas que nació el año anterior}}{\text{Número total de nacimientos}} = \frac{32}{50}$$

Otro ejemplo. Un importador de cristal irlandés de Nueva York recibe envíos de cajas de tres artículos. Los datos para las últimas 100 cajas indicaron el número de artículos dañados que había en cada caja. Por ejemplo, mostraba que en 40 de las cajas ningún artículo se había dañado mientras que en 12 de las cajas todos los tres artículos se habían roto.

Resultado (E_i) (número de defectos)	Número de cajas	$P(E_i)$
0	40	40/100 = 0.40
1	27	27/100 = 0.27
2	21	21/100 = 0.21
3	12	12/100 = 0.12
	100	1.00

En el pasado, 21 de las 100 cajas totales contenían exactamente dos artículos dañados. Entonces el modelo de frecuencia relativa asignaría una probabilidad de que dos artículos en cualquier caja dada estuvieran dañados así $P(2) = 21 / 100 = 0.21$. La probabilidad para cada resultado individual se muestra en la última columna, la cual suma 1.

Un problema común con el modelo de frecuencia relativa resulta cuando se hacen estimaciones con un número insuficiente de observaciones. Por ejemplo, se asume que en los dos vuelos de una aerolínea en los que se hicieron registros el año pasado estuvieron retrasados al llegar a sus destinos. Por tanto, se concluye que el vuelo que se abordará el próximo mes de la misma aerolínea, también estará retrasado. Aunque tales inferencias son comunes, no existen suficientes datos como para sacar tal conclusión y se deben evitar las decisiones basadas en tales inferencias.

En muchos casos los datos pasados no se encuentran disponibles. Por tanto, no es posible calcular la probabilidad a partir del desempeño anterior. La única alternativa es estimar la probabilidad con base en nuestro mejor criterio. El modelo subjetivo requiere establecer la probabilidad de algún evento con base en la mejor evidencia disponible. En muchos casos esto puede ser apenas una conjetura hecha sobre cierta base. El **modelo subjetivo** se utiliza cuando se desea asignar probabilidad a un evento que nunca ha ocurrido. La probabilidad de que una mujer sea elegida como presidente de los Estados Unidos es un ejemplo. Debido a que no hay datos sobre los cuales confiar, se deben analizar las opiniones y creencias para obtener una estimación subjetiva.

De los tres métodos para medir la probabilidad, el **modelo clásico** es el que se relaciona con mayor frecuencia con las apuestas y juegos de azar. La probabilidad clásica de un evento E se determina mediante:

Modelo clásico $$P(E) = \frac{\text{Número de formas en las que puede ocurrir un evento}}{\text{Número total de posibles resultados}} \quad [4.2]$$

Aun sin conocer a fondo la probabilidad clásica, se puede estar consciente de que la probabilidad de obtener una cara en un solo lanzamiento de una moneda es de la mitad. Esto puede ilustrarse utilizando la fórmula (4.2) así:

$$P(\text{cara}) = \frac{\text{Número de formas en las que el evento puede ocurrir}}{\text{Número total de posibles resultados}} = \frac{1}{2}$$

Existe sólo una forma en que puede ocurrir el evento (se obtiene una cara), y existen dos posibles resultados (una cara y un sello). De igual forma, la probabilidad de sacar un 3 con un dado de seis caras es:

$$P(3) = \frac{\text{Número de formas en las que el evento puede ocurrir}}{\text{Número total de posibles resultados}} = \frac{1}{6}$$

Existe sólo una forma en que puede ocurrir el evento (se saca un 3), y seis posibles resultados.

La probabilidad clásica implica la determinación de la probabilidad de algún evento *a priori* (antes del hecho). Por tanto, *antes* de sacar una carta de una baraja de 52 cartas, se puede determinar que la probabilidad de sacar un as es:

$$P(\text{as}) = \frac{\text{Número de formas en las que el evento puede ocurrir}}{\text{Número total de posibles resultados}} = \frac{4}{52}$$

Ejercicios de la sección

1. ¿Cuál modelo de probabilidad es apropiado para cada uno de los experimentos enumerados a continuación? Explique el porqué de su respuesta.

a. El Índice Dow Jones del precio de las acciones hoy cerrará alto.

b. Una unidad de producción será defectuosa.

c. Sacar un 6 con un dado.

d. El sol será nova.

2. Cite tres ejemplos de negocios para cada uno de los tres modelos de probabilidad.

3. La siguiente tabla muestra el número de computadores vendidos diariamente por una tienda minorista.

Número de computadores vendidos	Número de días
0	12
1	43
2	18
3	20
4	25

Determine la probabilidad de que el número de computadores que se vendan hoy sea:

a. 2

b. Menos de 3

c. Más de 1

d. Por lo menos 1

4. Durante los últimos cuatro campeonatos de fútbol norteamericano, el lanzamiento de la moneda cayó cara todas las veces. Su entrenador le dice que pedir sello esta vez aumentará la probabilidad de que usted gane el lanzamiento. ¿Está en lo cierto o está equivocado? Explique su respuesta completamente.

5. ¿Cuál modelo de probabilidad utilizó en el problema anterior? Explique.

6. Durante el año anterior, las ventas semanales en Petunia's Pet Shoppe han sido "bajas" durante 16 semanas, "considerables" durante 27 semanas y "altas" el resto de las semanas. Cuál es la probabilidad de que las ventas de esta semana sean:

a. Considerables

b. Bajas

c. Altas

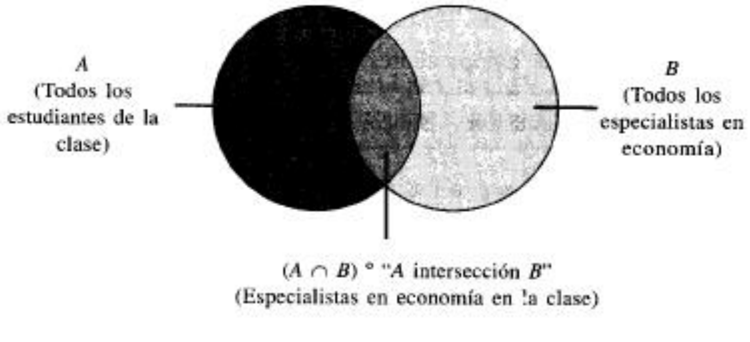
d. Por lo menos considerables

4.4 Uniones, intersecciones y relaciones entre eventos

Un **conjunto** es toda reunión de objetos. Con frecuencia es de utilidad identificar cómo pueden relacionarse los conjuntos entre sí. Se asume que se han identificado dos conjuntos A y B. Cada uno contiene numerosos elementos. Es completamente posible que algunos elementos estén en ambos conjuntos. Por ejemplo, se asume que el conjunto A consta de todos los estudiantes de la clase de estadística, y el conjunto B consta de todos los estudiantes de la universidad que están especializándose en economía. Aquellos elementos (estudiantes) que están en ambos conjuntos son los especialistas en economía de la clase de estadística. Tales estudiantes constituyen la intersección entre A y B. La **intersección** entre A y B, que se escribe $A \cap B$ y se lee como "A intersección B", consta de los elementos que son comunes tanto a A como a B.

Un diagrama de Venn es una herramienta útil para mostrar la relación entre conjuntos. Fue desarrollado por John Venn (1834 – 1923), un matemático inglés, esta representación gráfica se muestra en la figura 4.1. En el diagrama pueden verse los dos conjuntos A y B. Estos son los estudiantes que están tanto en el conjunto A (la clase) como en el conjunto B (especialistas en economía).

Figura 4.1
Diagrama de Venn



Intersección entre A y B Es el conjunto de todos los elementos que están tanto en A como en B.

Para que ocurra $A \cap B$, **tanto** "A como B" deben ocurrir. El estudiante debe estar en la clase de estadística y en la especialización en economía. Los eventos A y B se les denomina eventos **no disyuntos.** Ambos deben ocurrir antes que ocurra el evento $A \cap B$ ("A y B").

La **unión** de A y B, que se escribe $A \cup B$ y se lee "A unión B", consta de tales elementos que están o en A o en B o en ambos. Como se ve en la figura 4.1, todos los estudiantes que están en la clase (conjunto A), sin tener en cuenta su especialización, y todos los especialistas en economía (conjunto B), sin tener en cuenta si están en la clase de estadística, son elementos de $A \cup B$.

La unión de A y B	Es el conjunto de todos los elementos que están en A o en B.

Para que un elemento esté en $A \cup B$, sólo necesita estar en el conjunto A o en el conjunto B o en ambos.

Una comprensión total de la probabilidad no puede adquirirse sin un entendimiento de las formas como pueden relacionarse los eventos. Se dice que dos eventos son **mutuamente excluyentes** si la ocurrencia de uno prohíbe la ocurrencia del otro. Un ejemplo clásico de eventos mutuamente excluyentes es el de sacar una cara o un sello al lanzar una moneda una vez. Si se obtiene una cara, no puede ocurrir un sello. Seleccionar una unidad de producción y encontrarla defectuosa o no defectuosa son eventos mutuamente excluyentes.

Sacar una carta de una baraja bien sea una reina o un as también son eventos mutuamente excluyentes. Sin embargo, sacar una reina y un corazón no lo son, debido a que ambos se presentarían si se sacara la reina de corazones. En el caso inicial a comienzos de este capítulo, la Asociación Nacional de Esquí va a analizar la ubicación de los centros de esquí. Seleccionar un centro que esté (1) en un área aislada y (2) que se encuentre ubicado cerca de un centro urbano son eventos mutuamente excluyentes. Si está aislado, no está cerca de una gran ciudad. Por otro lado, un centro de esquí ubicado (1) cerca de una ciudad y (2) cerca de otros centros similares de esquí no son eventos mutuamente excluyentes debido a que hay muchos centros en el área de Chicago.

Los eventos **colectivamente exhaustivos** constan de todos los posibles resultados de un experimento y constituyen su espacio muestral. Los eventos colectivamente exhaustivos de lanzar un dado son 1, 2, 3, 4, 5 y 6. Además, debido a que existe la *certeza* de que uno de estos eventos ocurrirá, su probabilidad combinada será igual a uno:

$$P(1 \text{ o } 2 \text{ o } 3 \text{ o } 4 \text{ o } 5 \text{ o } 6) = 1$$

De los 500 empleados de King Dynamics, Inc. 170 están clasificados como miembros del personal administrativo, 290 como trabajadores de línea, y los 40 trabajadores restantes son empleados auxiliares. Los eventos colectivamente exhaustivos son S, L y A. Si un empleado se selecciona al azar,

$$P(S) = 170/500 = 0.34$$
$$P(L) = 290/500 = 0.58$$
$$P(A) = 40/500 = 0.08$$

Debido a que ocurre la certeza de que el empleado seleccionado provenga de una de estas tres categorías colectivamente exhaustivas, $P(S \text{ o } L \text{ o } A) = 0.34 + 0.58 + 0.08 = 1.00$.

Eventos independientes son eventos en los que la ocurrencia de uno no tiene nada que ver con la ocurrencia del otro. Algunos ejemplos incluyen el resultado de un lanzamiento de una moneda y el de un dado. El resultado del lanzamiento de una moneda no afecta al dado. Dos lanzamientos de una moneda también son eventos independientes.

¿Los resultados de sacar dos cartas de una baraja son eventos independientes? Es decir, ¿el resultado de sacar la primera carta afecta la probabilidad del segundo resultado? Depende de si se reemplaza o no la primera carta antes de sacar la segunda. Sea el primer evento sacar una reina y el segundo evento sacar un as. De acuerdo con el modelo clásico, la probabilidad de sacar una reina en el primer intento es $P(Q) = 4/52$.

La probabilidad de sacar un as en el segundo intento depende de si la primera carta fue reemplazada antes de sacar la segunda. Se asume que una reina, o cualquier otra carta distinta a un as se saca la primera vez. Esa carta se mantiene fuera de la baraja en el segundo intento, entonces la probabilidad de sacar un as es $P(A) = 4/51$ debido a que 4 de las 51 cartas restantes son ases. Si la carta se regresa a la baraja antes del segundo intento, la probabilidad de sacar un as en el segundo intento es $P(A) = 4/52$.

Cuando se saca de un conjunto finito, como por ejemplo una baraja de cartas, dos eventos son independientes si y sólo si se realiza el ***reemplazo***. Sin embargo, si el primer elemento no se reemplaza antes de sacar el segundo elemento, los dos eventos son dependientes.

Si se seleccionan *dos* trabajadores de King Dynamics, la probabilidad de que el primero sea un miembro del personal administrativo es $P(S) = 170/500 = 0.34$. Si esta selección no se reemplaza, la probabilidad de que la segunda sea un trabajador de línea es $P(L) = 290/499$, y no $290 / 500$.

Eventos complementarios son los eventos en los que si un evento no ocurre, el otro debe ocurrir. Si un evento *A* es lanzar un número par con un dado (2, 4 o 6), el complemento es lanzar un número impar (1, 3, o 5). Si no se obtiene un número par, se debe obtener un número impar. El complemento de *A* se escribe como *A*, y se denomina "no *A*".

Claro que los eventos complementarios también son colectivamente exhaustivos, porque si *A* no ocurre, *A* debe ocurrir. Por tanto,

$$P(A) + P(\bar{A}) = 1$$

y

$$P(A) = 1 - P(\bar{A})$$

Si no se selecciona un miembro del personal administrativo de King Dyamics, entonces debe ser o uno de línea o uno auxiliar. La probabilidad de que sea un miembro del personal administrativo es $P(S)$ y la probabilidad de que sea un trabajador de línea o uno auxiliar es $P(S)$. Entonces, $P(S) + P(S) = 1$

Ejercicios de la sección

7. Dada una baraja de 52 cartas, el conjunto *A* consta de los 13 corazones y el conjunto *B* son los cuatro ases. Identifique cuáles cartas están incluidas en $(A \cup B)$ y $(A \cap B)$.

8. Haga un diagrama de Venn para el ejercicio anterior.

9. Algunos de los trabajadores hombres y mujeres de una planta grande tienen educación secundaria. El conjunto *A* consta de los trabajadores hombres, el conjunto *B* de las trabajadoras mujeres, el conjunto *C* es el conjunto con educación secundaria, y el conjunto *D* es el conjunto de los trabajadores que no tienen educación secundaria. Identifique y explique $(A \cup C)$, $(B \cup D)$ y $(A \cap C)$.

10. Para el problema anterior, ¿cuál es la diferencia entre $(B \cup D)$ y $(B \cap D)$?

11. Dadas las condiciones del ejercicio 9, identifique los eventos que son:

a. Mutuamente excluyentes.

b. Colectivamente exhaustivos con respecto al género.

c. Si 300 de los 1000 trabajadores son hombres, ¿cuál es la probabilidad de que un trabajador sea mujer $P(F)$? ¿Qué papel jugó la regla de la complementariedad en su respuesta?

d. ¿Cuál es la diferencia entre $P(F)$ y $P(\bar{M})$?

12. Describa tres ejemplos relacionados con los negocios en los cuales los eventos sean independientes.

4.5 Tablas de contingencia y tablas de probabilidad

Las tablas de contingencia y las tablas de probabilidad son útiles al calcular la probabilidad de eventos. Regresando al ejemplo de King Dynamics, se supone que la tabla de contingencia para todos los 500 empleados aparece en la tabla 4.1.

Tabla 4.1
Tabla de contingencia para King Dynamics

	Clasificación de los Empleados			
Género	**Personal**	**Línea**	**Auxiliar**	**Total**
Hombres	120	150	30	300
Mujeres	50	140	10	200
Total	170	290	40	500

La tabla muestra por ejemplo que de los 170 miembros del personal administrativo, 120 son hombres y 50 son mujeres. Una tabla de probabilidad puede crearse dividiendo cada una de las entradas de la tabla 4.1 por el total, 500 trabajadores. Los resultados se ven en la tabla 4.2

Tabla 4.2
Tabla de probabilidad para King Dynamics

	Clasificación de Empleados			
Género	**Personal (S)**	**Línea (L)**	**Auxiliar (A)**	**Total**
Hombres (M)	120/500=**0.24**	150/500=0.30	30/500=0.06	300/500=**0.60**
Mujeres	50/500 =0.10	140/500=0.28	10/500=0.02	200/500=0.40
Total	170/500=0.34	290/500=**0.58**	40/500=0.08	500/500=1.00

Los valores en las márgenes de la tabla se llaman *probabilidades marginales*. Por ejemplo, la probabilidad de seleccionar un trabajador de línea de manera aleatoria es $P(L) = 0.58$ y la probabilidad de seleccionar un hombre es $P(M) = 0.60$. Las *probabilidades conjuntas* en las celdas de la estructura principal de la tabla muestran la probabilidad de la intersección entre dos eventos. Por ejemplo, la probabilidad de seleccionar un miembro del personal administrativo hombre, es decir, un trabajador que sea parte del personal administrativo y que sea hombre, es $P(M \cap S) = 0.24$. Una probabilidad marginal se encuentra como la suma de las probabilidades conjuntas correspondientes. Por tanto $P(M) = P(M \cap S) + P(M \cap L) + P(M \cap A) = 0.24 + 0.30 + 0.06 = 0.60$.

Ejercicios de la sección

13. Usted recolectó datos sobre 500 economistas en la academia, la industria privada, y el gobierno respecto a sus opiniones sobre si la economía podría ser estable, podría expandirse o podría entrar en un período de contracción en el futuro próximo. Sin embargo, parte de la información se perdió, resultando la siguiente tabla de contingencia parcial. Con base en los datos restantes, cree una tabla de probabilidad.

	Economía			
Economistas	**Estable (S)**	**Expansión (E)**	**Contracción (C)**	**Total**
Academia (A)	125		100	
Industria privada (I)		35		110
Gobierno (G)	25	40		65
Total	200			

De la tabla de probabilidad halle:

a. $P(A)$

b. $P(G)$.

c. $P(A \cap S)$

d. $P(A \cap E)$

e. $P(G \cap C)$.

14. La revista *Forbes* (febrero de 1997) clasificó las 120 ciudades de Estados Unidos de acuerdo con la calidad de vida, con base en parte del porcentaje de empleados que tenían título universitario. Los resultados se ven en la siguiente tabla de contingencia parcial, en donde A es menos del 15% con título universitario, B es del 15 al 20% con título universitario y C es más del 20% con título universitario. Realice una tabla de probabilidad y responda las preguntas que se presentan en la siguiente tabla.

	Calidad de Vida			
Porcentajes con Título Universitario	**Pobre (P)**	**Bueno (G)**	**Excelente (E)**	**Total**
A	10	20		40
B			20	
C		10		20
Total	20	60		

a. ¿$P(A)$?

b. ¿$P(P \cap B)$?

c. ¿$P(E \cap C)$?

d. ¿$P(A \cap G)$?

15. Con base en su tabla de probabilidad del ejercicio anterior, escriba un informe al presidente del comité para seleccionar un nuevo sitio para las oficinas principales de la compañía. Incluya y evalúe todas las comparaciones relevantes, con base en los factores porcentuales /de calidad. ¿Qué conclusión saca?

4.6 Probabilidad condicional

Con frecuencia se desea determinar la probabilidad de algún evento, dado que antes otro evento ya haya ocurrido. Lógicamente, esta es llamada **probabilidad condicional**. Se denota como $P(A \mid B)$ y se lee la "probabilidad de A dado B".

Probabilidad condicional Es la probabilidad de que el evento A ocurra, dado que o a condición de que el evento B ya haya ocurrido.

La probabilidad condicional se utiliza comúnmente en el planteamiento de un negocio para *revisar* la probabilidad de algún evento dado sobre el cual se ha recolectado información adicional. Por ejemplo, se puede estimar la probabilidad de que se haga una venta (S) a un cliente antiguo de $P(S) = 0.80$. Sin embargo, si se sabe posteriormente que este cliente ahora está comprando a alguien de la competencia, se puede revisar la probabilidad de que se haga una venta, dado que el competidor (C) ha presentado una oferta $P(S \mid C) = 0.30$.

Para ilustrar con un ejemplo simple, ya se sabe que la probabilidad de sacar una jota de una baraja de 52 cartas es $P(J) = 4 / 52$ debido a que hay 4 jotas en una baraja. Sin embargo, es de suponer que se desea saber la probabilidad de que la carta sacada fuese una jota, dada la información adicional de que es una figura (F). Es decir, $P(J \mid F)$. Ya que 4 de las 12 figuras de una baraja son jotas, $P(J \mid F). = 4/12$, y no 4/52. Esta es la fórmula general para la probabilidad condicional del evento A, dado que se conoce que el evento B ya ha ocurrido:

Probabilidad condicional de A dado B	$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{P(A)P(B \mid A)}{P(B)}$	[4.3]

Por tanto,

$$P(J \mid F) = \frac{P(J \cap F)}{P(F)} = \frac{P(J)P(F \mid J)}{P(F)} = \frac{(4/52)(1)}{12/52} = 4/12$$

$P(F \mid J)$ es 1 debido a que todas las jotas son figuras. Aunque en este ejemplo simple fue posible determinar $P(J \mid F)$ sin el uso de la fórmula (4.3), existen muchos ejemplos en los cuales la fórmula es necesaria. Para ilustrar el valor de la fórmula (4.3), se retoma por un momento la tabla 4.2 para King Dynamics. Se puede observar que la probabilidad de que un trabajador tomado aleatoriamente sea hombre es $P(M) = 0.60$. Sin embargo, si se desea calcular la probabilidad de que el trabajador sea hombre dado que es un miembro del personal administrativo $P(M \mid S)$ se puede hallar así

$$P(M \mid S) = \frac{P(M \cap S)}{P(S)} = \frac{0.24}{0.34} = 0.71$$

Ejercicios de la sección

16. De la tabla de probabilidad que usted creó en el ejercicio 13:

a. Halle $P(S \mid A)$.

b. Si usted es un economista académico, ¿es más probable que pronostique una economía estable que si trabaja con el gobierno?

c. Dado que usted trabaja en la industria privada, ¿es más probable que usted pronostique una contracción en la economía que un académico?

d. Si usted trabaja para el gobierno, ¿cuál de los tres pronósticos es más probable que usted haga?

17. Con base en la tabla de probabilidad que usted creó en el ejercicio 14,

a. Dando un rango de Excelente, ¿cuál de las tres categorías porcentuales es más probable que ocurra?

b. Si el 19% de los empleados de una ciudad tienen título universitario, ¿es más probable que la calidad de vida se clasifique en Pobre, Buena o Excelente?

c. Si más del 20% de los empleados de una ciudad tienen título, ¿qué tan probable es que la ciudad sea clasificada como Excelente?

d. Si una ciudad es clasificada como Excelente, ¿qué tan probable es que más del 20% de sus empleados tengan títulos universitarios?

4.7 Las dos reglas de la probabilidad

Existen dos reglas básicas que deben seguirse para calcular la probabilidad de eventos más complejos: la regla de la multiplicación y la regla de la adición. Cada una se utiliza para propósitos específicos. La regla de la multiplicación se utiliza para determinar la probabilidad de que "A y B", $P(A \cap B)$, y la regla de la adición se utiliza para calcular la probabilidad de "A o B", $P(A \cup B)$.

A. Regla de la multiplicación

El propósito de la **regla de la multiplicación** es determinar la probabilidad del evento conjunto $P(A \cap B)$. Es decir, que para encontrar la probabilidad de "A y B", simplemente se multiplican sus respectivas probabilidades. El procedimiento exacto depende de si A y B son dependientes o independientes. Los eventos A y B son independientes si $P(A) = P(A \mid B)$. Es decir, la probabilidad de A es la misma bien se considere o no el evento B. De igual forma, si A y B son independientes, si $P(B) = P(B \mid A)$.

Para eventos *independientes* la probabilidad de dos eventos se vuelve:

Probabilidad de eventos independientes	$P(A \cap B) = P(A) \text{ X } P(B)$	[4.4]

La probabilidad de los dos eventos independientes de sacar un 3 con un dado y una cara con una moneda es

$$P(3 \cap H) = P(3) \times P(H) = 1/6) \times (1/2) = 1/12$$

Debido a que estos dos eventos son independientes, simplemente se multiplican sus probabilidades individuales. De igual forma, la probabilidad de sacar una carta de las 13 cartas de corazones de una baraja de 52 cartas y de sacar un número par con un dado es $P(H \cap E) = P(H) \text{ X } P(E) = 13/52 \text{ X } 3/6 = 39/312$.

Para atraer a los clientes, Norman, propietario de Bates Motel, ha modernizado sus instalaciones. Observa que el 20% de todos los autos que pasan por allí, se detienen para alquilar un cuarto. ¿Cuál es la probabilidad de que los próximos dos carros se detengan? Asumiendo que estos eventos son independientes, $P(S_1 \cap S_2) = P(S_1) \text{ x } P(S_2) = (0.20)(0.20) = 0.04$ La probabilidad de que los dos próximos autos alquilen un cuarto a Norman es del 4%. ¿Cuál es la probabilidad de que el primer auto pare y que el segundo no lo haga? $P(S_1 \cap S_2) = P(S) \text{ x } P(S_2) = (0.20) \text{ X } (0.80) = 0.16$.

Si los eventos son dependientes, entonces, por definición, se debe considerar el primer evento al determinar la probabilidad del segundo. Es decir, la probabilidad del evento B depende de la condición que A ya haya ocurrido. Se necesita del principio de probabilidad condicional. La probabilidad de los eventos conjuntos A y B es

Probabilidad de eventos dependientes	$P(A \cap B) = P(A) \text{ X } P(B \mid A)$	[4.5]

Regresando a la tabla de probabilidad para King Dynamics, tabla 4.2, se observa que la probabilidad marginal de la primera fila muestra claramente que $P(M) = 0.60$ sin considerar si el trabajador es miembro administrativo, de línea o auxiliar. Sin embargo, la probabilidad conjunta de que sea hombre *y* miembro administrativo se observa como $P(M \cap S) = 0.24$.

También se puede calcular esta probabilidad utilizando la fórmula (4.5); $P(M \cap S) = P(M) \text{ X } P(S \mid M)$. El último término es probabilidad condicional, la cual se determinó anteriormente como

$$P(S \mid M) = \frac{P(S \cap M)}{P(M)} = \frac{0.24}{0.60} = 0.40$$

Entonces,

$$P(M \cap S) = P(M) \times P(S \mid M) = (0.60) \times (0.40) = 0.24$$

Aunque el uso de una tabla puede simplificar los cálculos de probabilidad, existen ejemplos en los cuales es muy difícil la creación de una tabla, por tanto se requiere el uso de fórmulas. El ejemplo 4.1 ilustra este caso.

Ejemplo 4.1

El gerente de créditos de Dollar-Wise Department Store recolecta datos sobre 100 de sus clientes. De los 60 hombres, 40 tienen tarjetas de crédito (C). De las 40 mujeres, 30 tienen tarjeta de crédito (C). Diez de los hombres tienen saldos vencidos (B), mientras que 15 de las mujeres tienen saldos vencidos (B). El gerente de crédito desea determinar la probabilidad de que un cliente seleccionado al azar sea:

a. Una mujer con tarjeta de crédito.

b. Una mujer con un saldo.

c. Un hombre sin un saldo.

d. Un hombre con un saldo.

Solución

Crear una tabla de probabilidad es difícil ya que existen tres factores: género, tarjeta de crédito y saldo en la tarjeta. El uso de la fórmula (4.5) es quizá el modelo preferido.

a. $P(W \cap C) = P(W) \times P(C \mid W)$. Claramente $P(W) = 40/100$. Además, de las 40 mujeres 30 tienen tarjetas de crédito. Por tanto, dado que el cliente es una mujer, la probabilidad de que tenga una tarjeta de crédito es $P(C \mid W) = 30/40$. Entonces, $P(W \cap C) = P(W) \times P(C \mid W) = (40/100) \text{ X } (30/40) = 0.30$

b. $P(W \cap B) = P(W) \times P(B \mid W)$. De las 40 mujeres, 15 tienen saldos. Dado que el cliente es una mujer, la probabilidad de que tenga un saldo es de 15/40. De manera que $P(W \cap B) = P(W) \times P(B \mid W) = (40/100)(15/40) = 0.15$.

c. $P(M \cap \bar{B}) = P(M) \times P(\bar{B} \mid M)$. Debido a que 50 de los 60 hombres no tienen saldos, $P(\bar{B} \mid M) = 50/60$. Entonces, $P(M \cap \bar{B}) = P(M) \times P(\bar{B} \mid M) = (60/100)(50/60) = 0.50$.

d. $P(M \cap B) = P(M) \times P(B \mid M)$. De los 60 hombres, 10 tienen saldos. $P(\bar{B} \mid M) = 10/60$. Así, $P(M \cap B) = P(M) \times P(B \mid M) = (60/100)(10/60) = 0.10$.

Interpretación

Pueden determinarse las probabilidades de otros eventos conjuntos que ayudarían al gerente de crédito a determinar las políticas de la tienda y a incrementar las ventas.

B. Regla de la adición

La **regla de la adición** se utiliza para determinar la probabilidad de A o B, $P(A \cup B)$.

La probabilidad del evento A o B (cuando los eventos no son mutuamente excluyentes)	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	[4.6]

Vale la pena recordar que los eventos A y B no son mutuamente excluyentes si ambos pueden ocurrir al mismo tiempo. En este caso, la fórmula requiere que se reste la probabilidad del evento conjunto A y B. La probabilidad de sacar un as o una de las 13 cartas de corazones de una baraja es $P(A) + P(H) - P(A \cap H)$. Los eventos A y H no son mutuamente excluyentes, debido a que ambos ocurren si se sacara el as de corazones. Por tanto, $P(A) + P(H) - P(A \cap H) = (4/52) + (13/52) - (1/52) = 16/52$.

La razón por la cual se debe restar la probabilidad conjunta cuando los eventos son mutuamente excluyentes es para evitar el *doble conteo*. Cuando se cuentan los cuatro ases, se incluye el as de corazones. Cuando se cuentan las trece cartas de corazones se incluye el as de corazones por segunda vez. Debido a que sólo hay un as de corazones, es necesario restarlo una vez.

En el ejemplo de King Dynamics, la probabilidad de que un empleado sea un trabajador hombre o un trabajador administrativo es $P(M) + P(S) - P(M \cap S) = (0.60) + (0.34) - (0.24) = 0.70$. De nuevo, se debe restar la probabilidad conjunta $P(M \cap S)$, debido a que se incluyeron trabajadores administrativos al contarse todos los hombres y se incluyeron hombres al contar todos los trabajadores administrativos. Los trabajadores administrativos hombres se contaron dos veces.

Ejemplo 4.2

La mayoría de las estaciones de servicio venden tres tipos de gasolina: corriente, super y premium. Con frecuencia, alguna de cada tipo está enriquecida con etanol. La tabla de contingencia que se muestra a continuación ilustra los porcentajes de clientes que prefieren cada tipo.

	Corriente (R)	Super (S)	Premium (P)	Total
Etanol	0.05	0.10	0.05	0.20
Sin etanol	0.15	0.40	0.25	0.80
Total	0.20	0.50	0.30	1.00

Determine la probabilidad de que el siguiente cliente prefiera:

a. Corriente o etanol; $P(R \cup E)$.

b. Super o sin etanol; $P(S \cup \bar{E})$.

c. Premium o etanol; $P(P \cup E)$.

d. Premium o sin etanol; $P(P \cup \bar{E})$.

Solución

a. $P(R \cup E) = P(R) + P(E) - P(R \cap E)$
$= 0.20 + 0.20 - 0.05 = 0.35$

b. $P(S \cup \bar{E}) = P(S) + P(\bar{E}) - P(S \cap \bar{E})$
$= 0.50 + 0.80 - 0.40 = 0.90$

c. $P(P \cup E) = P(P) + P(E) - P(P \cap E)$
$= 0.30 + 0.20 - 0.05 = 0.45$

d. $P(P \cup \bar{E}) = P(P) + (P\bar{E}) - P(P \cap \bar{E})$
$= 0.30 + 0.80 - 0.25 = 0.85$

Interpretación

Otras probabilidades podrían determinarse para comparar la popularidad de las preferencias del consumidor y así formular un programa de mercadeo diseñado para incrementar las ventas globales.

Vale la pena recordar que los ejemplos anteriores se relacionan con dos eventos que no son mutuamente excluyentes. Si A y B *son* mutuamente excluyentes, $P(A \cap B) = 0$. Por definición no pueden ocurrir simultáneamente. Debido a que no tiene sentido restar cero, la fórmula (4.6) se reduce a

Probabilidad del evento *A* o del evento *B* (cuando los eventos son mutuamente excluyentes	$P(A \cup B) = P(A) + P(B)$	[4.7]

La probabilidad de que un cliente prefiera corriente o super (eventos mutuamente excluyentes debido a que no puede preferir ambas) es $P(R \cup S) = P(R) + P(S) = 0.20 + 0.50 = 0.70$.

Algunos casos pueden necesitar el uso tanto de la regla de la multiplicación como la de la adición. Se supone que se lanza una moneda tres veces y se desea determinar la probabilidad de que se obtengan dos caras. La pregunta es ¿cómo puede suceder el evento (2 de 3 sean caras)? Se pueden obtener dos caras si:

Sólo el tercer lanzamiento es sello	$P(H_1 \text{ y } H_2 \text{ y } T_3) = 1/2 \times 1/2 \times 1/2 = 1/8$
Sólo si el segundo lanzamiento es sello	$P(H_1 \text{ y } T_2 \text{ y } H_3) = 1/2 \times 1/2 \times 1/2 = 1/8$
Sólo si el primer lanzamiento es sello	$P(T_1 \text{ y } H_2 \text{ y } H_3) = 1/2 \times 1/2 \times 1/2 = \dfrac{1/8}{3/8}$
con H: cara, T: sello	

Debido a que el evento puede suceder de la primera, de la segunda o de la tercera forma, se deben sumar sus respectivas probabilidades. Así, la probabilidad de que los tres lanzamientos de la moneda produzcan dos caras es de 3/8. El ejemplo 4.3 ilustra esto de manera más completa.

Ejemplo 4.3

De los 10 chips de un computador, 4 están defectuosos. ¿Cuál es la probabilidad de seleccionar 3 sin que haya reemplazo, de los cuales sólo uno sea defectuoso?

Solución

Usted debe preguntarse, "¿Cómo puede ocurrir el evento?" Existen tres formas en las cuales sólo uno puede ser defectuoso: (1) sólo el primero es defectuoso, (2) sólo el segundo es defectuoso, y (3) sólo el último es defectuoso. De acuerdo con la regla de la multiplicación,

La primera forma es $P(D_1 \cap \overline{D}_2 \cap \overline{D}_3) = (4/10)(6/9)(5/8) = 120/720.$

La segunda forma es $P(\overline{D}_1 \cap D_2 \cap \overline{D}_3) = (6/10)(4/9)(5/8) = 120/720.$

La última forma es $(\overline{D}_1 \cap \overline{D}_2 \cap D_3) = (6/10)(5/9)(4/8) = 120/720.$

Debido a que el evento "sólo uno es defectuoso" puede suceder de la primera forma, la segunda forma o la tercera forma, se debe adicionar las tres probabilidades, de acuerdo con la regla de la adición. P(sólo uno es defectuoso) = 120/720 + 120/720 + 120/720 = 360/720 = 0.50

Interpretación

Muchas compañías basan las garantías de sus productos en la probabilidad de que un cierto número de unidades vendidas salgan defectuosas. Si esta probabilidad es demasiado alta y costosa, las compañías se verán forzadas a ofrecer garantías menos atractivas.

Ejercicios de la sección

18. Con base en la tabla de probabilidad del ejercicio 13, halle

a. $P(A \cup S)$.
b. $P(P \cup C)$.
c. $P(E \cup G)$.
d. $P(G \cup E)$.
e. $P(A \cup G)$.

19. Wally, propietario de Wally's Workout World, desea construir un perfil de miembros para desarrollar una campaña publicitaria que atraiga a clientes potenciales típicos de quienes actualmente prefieren este gimnasio. El 30% de los miembros actuales son mujeres, el 80% de ellas es menor de 30 años. El 60% de los hombres son menores de 30 años. Cuál es la probabilidad de que un miembro seleccionado de manera aleatoria sea:

a. ¿Una mujer menor de 30 años?
b. ¿Una mujer mayor de 30 años?
c. ¿Un hombre mayor de 30 años o una mujer menor de 30?
d. ¿Mayor de 30 años?
e. ¿Un hombre o una mujer mayor de 30 años?

20. De 1,000 jóvenes de 18 años, 600 tienen empleo y 800 son bachilleres. De los 800 bachilleres, 500 tienen trabajo. Cuál es la probabilidad de que un joven de 18 años tomado aleatoriamente sea:

a. ¿Un bachiller empleado?
b. ¿Empleado pero no bachiller?
c. ¿Desempleado o un bachiller?
d. ¿Desempleado o no bachiller?

21. Sammy's Sweat Shop vende dos tipos de zapatos para correr, los Mercury y los Racer. Las probabilidades de que un cliente dado compre los Mercury es de $P(M) = 0.40$ y la probabilidad de que compre los Racer es de $P(R) = 0.30$. La probabilidad de que compre ambos es $P(B) = 0.10$. ¿Cuál es la probabilidad de que un cliente compre o M o R?

22. Un corredor de bolsa sabe por experiencias anteriores que la probabilidad de que un cliente compre acciones es del 65%. La probabilidad de que el cliente compre un bono del gobierno si ya tiene acciones es del 35%.

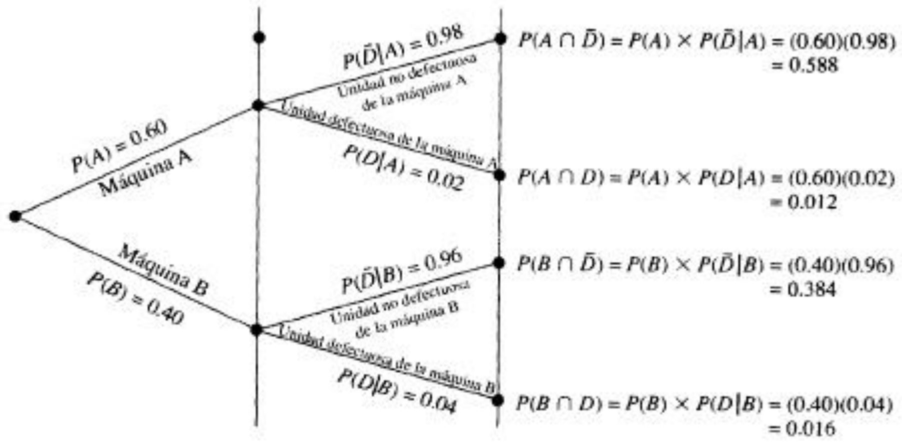
a. ¿Cuál es la probabilidad de que el cliente posea ambos?
b. ¿Son B y S independientes? Explique.

4.8 Teorema de Bayes

El reverendo Thomas Bayes (1702-1761) desarrolló un concepto útil al calcular ciertas probabilidades. Se asume que Dunham Manufacturing utiliza dos máquinas para producir su producto. La máquina A produce el 60% de la producción total, y la máquina B produce el restante 40%. El 2% de las unidades producidas por A son defectuosas, mientras que B tiene una tasa de defectos del 4%.

Esto se muestra en el *diagrama de árbol* que acompaña la figura 4.2. Se asume que una unidad de producción se selecciona aleatoriamente. El primer conjunto de "ramas" en el árbol que indica cuál máquina produjo la unidad, muestra que la probabilidad de que provenga de la máquina A es $P(A) = 0.60$ y que la probabilidad de que provenga de la máquina B es $P(B) = 40$. El segundo grupo de ramas que indica la calidad de la unidad dice que si provino de la máquina A puede ser o defectuosa o no defectuosa. Estas *probabilidades condicionales* demuestran que la probabilidad de que no sea defectuosa dado que provino de la máquina A es $P(\bar{D} \mid A) = 0.98$ y la probabilidad de que sea defectuosa dado que proviene de A es $P(D/A) = 0.02$. Las probabilidades condicionales para B revelan que la probabilidad de que no sea defectuosa dado que provino de B es $P(\bar{D} \mid B) = 0.96$ y la probabilidad de que sea defectuosa con base en la condición de que provino de B es $P(D \mid B) = 0.04$.

Figura 4.2
Diagrama de árbol para Dunham Manufacturing



Finalmente, se observa que hay cuatro posibles resultados para el experimento de seleccionar una unidad de producción. La probabilidad de cada uno se calcula multiplicando las probabilidades sobre cada una de las ramas que conllevan a ella. Para ilustrarlo, el primer posible resultado es que la unidad provenga de A y que no sea defectuosa, $P(A \cap \bar{D})$. Utilizando la regla de la multiplicación, $P(A \cap \bar{D}) = P(A) \times P(\bar{D} \mid A) = (0.60)(0.98) = 0.588$. Las probabilidades de los tres resultados restantes pueden determinarse de forma similar.

Se puede observar directamente en la figura 4.2 que $P(A) = 0.60$. Suponiendo que la unidad es defectuosa, se desea saber la probabilidad de que la unidad provino de la máquina A. Con esta información adicional, se puede revisar la probabilidad de que la unidad fue producida por la máquina A. Ahora se desea determinar $P(A \mid D$, no sólo $P(A)$.

Vale la pena recordar la regla de la probabilidad condicional:

$$P(A \mid D) = \frac{P(A \cap D)}{P(D)} = \frac{P(A) \times P(D \mid A)}{P(D)}$$

Sin embargo, $P(D)$ no es discernible de inmediato. Aquí es donde participa el teorema de Bayes. Existen dos formas en las cuales la unidad puede ser defectuosa. Puede provenir de la máquina A y ser defectuosa, o puede provenir de la máquina B y ser defectuosa. Utilizando la regla de la adición,

$$\begin{aligned} P(D) &= P(A \cap D) + P(B \cap D) \\ &= P(A) \times P(D \mid A) + P(B) \times P(D \mid B) \end{aligned}$$

Cuando se hace la sustitución en el denominador de la fórmula de la probabilidad condicional anterior para $P(D)$, el teorema de Bayes dice

Teorema de Bayes

$$P(A\,|\,D) = \frac{P(A \cap D)}{P(A \cap D) + P(B \cap D)} = \frac{P(A) \times P(D\,|\,A)}{P(A) \times P(D\,|\,A) + P(B) \times P(D\,|\,B)} \qquad [4.8]$$

Ahora se puede hallar $P(A \mid D)$

$$\begin{aligned} P(A\,|\,D) &= \frac{P(A \cap D)}{P(D)} \\ &= \frac{P(A \cap D)}{P(A \cap D) + P(B \cap D)} \\ &= \frac{(0.012)}{(0.012) + (0.016)} \\ &= 0.429 \end{aligned}$$

Mientras que $P(A) = 0.60$, $P(A \mid \mathrm{D}) = 0.429$. Se nota que $P(A \mid D) < P(A)$ debido a que la máquina A produce un porcentaje menor de defectos que la máquina B.

Ejemplo 4.4 Teorema de Bayes

El departamento de personal de una empresa grande ha descubierto que sólo el 60% de los candidatos entrevistados están realmente calificados (Q) para asumir un cargo en la compañía. Una revisión de los registros de la firma muestra que quienes estaban calificados, el 67% tuvo un entrenamiento previo en estadística (T), mientras que el 20% de quienes no estaban calificados habían recibido instrucción estadística mucho antes. Es decir,

$$P(Q) = 0.60 \qquad P(T\,|\,Q) = 0.67 \qquad P(T\,|\,\overline{Q}) = 0.20$$

El director de personal puede ver claramente que dado que usted está calificado, es más probable que usted tenga algo de capacitación en estadística que si no está calificado (0.67 > 0.20). Se perdió mucho tiempo entrevistando a los candidatos que resultaron no calificados; sin embargo, el director está considerando conceder entrevistas sólo a aquellos candidatos que tengan capacitación en estadística. Él espera incrementar la probabilidad de encontrar candidatos calificados para ocupar el cargo. La pregunta entonces sería, ¿es más probable que usted esté calificado dado que ha tenido capacitación: $P(Q \mid T)$? Si es así, el departamento de personal podría evitar demoras y costos innecesarios restringiendo las entrevista sólo a aquellos candidatos que tengan capacitación en análisis estadístico.

Solución

Utilizando la regla de probabilidad condicional

$$P(Q\,|\,T) = \frac{P(Q \cap T)}{P(T)} = \frac{P(Q) \times (T\,|\,Q)}{P(T)} = \frac{(0.60) \times (.067)}{P(T)}$$

Debido a que los registros de la compañía no proporcionan $P(T)$, se debe utilizar el teorema de Bayes para hallar el denominador. Existen dos formas en las que un candidato puede tener entrenamiento previo: (1) el candidato puede estar calificado y tener capacitación, $P(Q \cap T)$, y (2) el candidato puede no estar calificado y tener entrenamiento $P(Q \cap T)$. Por tanto,

$$\begin{aligned} P(T) &= P(Q \cap T) + P(\overline{Q} \cap T) \\ &= P(Q) \times P(T \mid Q) + P(\overline{Q}) \times P(T \mid \overline{Q}) \\ &= (0.60)(0.67) + (0.40)(0.20) \\ &= 0.482 \end{aligned}$$

Entonces

$$\begin{aligned} P(Q \mid T) &= \frac{(0.60)(0.67)}{0.482} \\ &= (0.834) \end{aligned}$$

Puesto que $P(Q) = 0.60$, $P(Q \mid T) = 0.834$

Interpretación

Para aumentar la probabilidad de entrevistar sólo candidatos calificados, el departamento de personal debería entrevistar solamente a los candidatos que tienen capacitación en análisis estadístico.

Ejercicios de la sección

23. James Business Equipment vende una amplia gama de suministros de oficina en la región central de Estados Unidos. El gerente de ventas cree que ha descubierto una tendencia que podría disparar las ventas. Los resultados de un estudio parecen sugerir que las firmas jurídicas están más inclinadas a hacer pedidos que otras con otras áreas de negocios. Sus datos muestran que la probabilidad de que la comunidad de comercio general compre sus suministros es del 20%, $P(B) = 0.20$. Entre otros clientes durante varios de los últimos meses, el 50% eran empresas jurídicas. El gerente siente que concentrarse en las oficinas jurídicas incrementará la probabilidad de que una llamada al departamento de ventas termine en una compra. ¿Cuál es su conclusión?

24. Una empresa productora de medicinas que está probando un nuevo medicamento contra la fiebre del heno descubrió que el 60% de todas las personas que sufren de este mal sienten alivio de los síntomas a las 4 semanas, hayan o no utilizado el medicamento. De quienes sienten alivio, el 40% ha tomado el medicamento, mientras que el 30% de quienes no han sentido alivio han probado el medicamento. La empresa productora de medicinas desea determinar si es aconsejable tomar el medicamento comparando la probabilidad de aliviarse de la fiebre del heno si quienes la padecen sí toman el medicamento con la probabilidad de alivio si quienes la padecen no toman el medicamento.

4.9 Técnicas de conteo

Muchas decisiones comerciales requieren que se cuente el número de subconjuntos que se pueden obtener de un conjunto. De las ventas de una línea que consta de 10 productos, ¿cuántos subconjuntos de 3 productos se pueden ofrecer a los clientes? Siete agentes de ventas están en un concurso para ver quién gana un viaje gratis a Cancún. ¿Cuántas órdenes diferentes de terminación hay? ¿Cuántos números telefónicos distintos pueden asignarse a una oficina grande dados los dígitos 0-9? Muchos otros ejemplos abundan. Se analizarán cuatro técnicas de conteo utilizadas para responder estas preguntas y otras parecidas: combinaciones, permutaciones, escogencia múltiple y multiplicación.

Al seleccionar los elementos en los subconjuntos, la distinción entre *permutaciones* y *combinaciones* depende de si el orden de las selecciones hace la diferencia. Si un orden es suficiente para constituir otro subconjunto, entonces se trata de permutaciones. Si dos subconjuntos se consideran iguales debido a que simplemente se han reordenado los mismos elementos, entonces se involucran las combinaciones.

De la lista de las primeras cinco letras del alfabeto A, B, C, D y E, ¿cuántos subconjuntos diferentes se pueden obtener? Dos posibilidades son {A, B ,C} y {A,C,B}. Ambos tienen los mismos elementos y difieren sólo en el orden en el cual se seleccionaron los elementos. Si se considera que dos subconjuntos son *diferentes* porque el orden es diferente, se consideran como permutaciones. Si se considera que dos subconjuntos son idénticos y constituyen el mismo subconjunto porque ambos tienen los mismos elementos sin considerar el orden, son denominados combinaciones. Es decir, que con las permutaciones el orden es el que hace la diferencia.

Dado un conjunto de n elementos, el número de permutaciones, cada uno de tamaño r, se determina como

El número de permutaciones de n elementos tomados r a la vez	$_nP_r = \frac{n!}{(n-r)!}$	[4.9]

en donde $n!$ se lee como "n factorial" y significa el producto de todos los números de 1 a n. Por tanto $4! = 1 \times 2 \times 3 \times 4 = 24$. Por definición $0! = 1$

El número de combinaciones de n elementos tomados r a la vez es

El número de combinaciones de n elementos tomados r a la vez	$_nC_r = \frac{n!}{r!\,(n-r)!}$	[4.10]

Antes se preguntó que dados 10 productos, ¿cuántos subconjuntos de 3 productos podrían empacarse juntos y ofrecerse a los clientes? Si se considera que el orden en el cual se ofrecen los 3 productos no influirá en los clientes, es decir, que el orden no hará diferencia alguna, se debe hallar el número de combinaciones de 10 elementos tomados 3 a la vez.

$$
\begin{aligned}
_{10}C_3 &= \frac{10!}{3!\,(10-3!)} \\
&= 120
\end{aligned}
$$

Hay 120 paquetes de 3 artículos que se pueden ofrecer a los clientes.

Las combinaciones también pueden utilizarse para calcular la probabilidad de que uno de los chips de computador del ejemplo 4.3 sea defectuoso. En lugar de determinar la probabilidad de cada forma en la cual uno sea defectuoso, simplemente se halla la probabilidad de que *una* de las formas tenga un chip defectuoso: $P(D_1 \cap D_2 \cap D_3) = 120/720$. Entonces se cuenta el número de formas en las cuales uno de los tres es defectuoso como $_3C_1 = 3$, y se multiplica: $(3)(120/720) = 0.50$.

Ejemplo 4.5

Un caso reciente en la corte del condado de Madison, Kentucky, sobre las prácticas de contratación de una compañía de teléfonos local. La compañía planeó contratar 3 nuevos empleados. Había 8 candidatos para los cargos, 6 de los cuales eran hombres. Los 3 que fueron contratados eran hombres. Un cargo por discriminación de sexo se impuso contra la compañía. ¿Cómo decidiría usted?

Solución

La decisión depende ampliamente de la probabilidad de que todos los 3 contratados sean hombres si a cada candidato se le da igual consideración sin tener preferencias de género. Por tanto, se debe determinar la probabilidad de que todos los 3 sean hombres.

$$P(\text{todos 3 hombres}) = \frac{\text{Número de formas en las cuales todos 3 pueden ser hombres}}{\text{Número total de posibles resultados}}$$

El número de formas en las cuales 3 de los 6 hombres y ninguna de las 2 mujeres pueda ser contratada es ${}_6C_3 \times {}_2C_0 = 20 \times 1 = 20$. El número total de formas en las que 3 de todos los 8 candidatos pueden ser contratado es ${}_8C_3 = 56$. Así

$$P(\text{todos 3 son hombres}) = \frac{20}{56} = 36\%$$

Interpretación

Existe el 36% de probabilidades de que los 3 empleados contratados sean hombres si no se da ninguna preferencia al género. Este evento ocurriría aproximadamente una de cada 3 veces. La corte reguló que todo evento que pudiera ocurrir con tal frecuencia es usual y no constituye prueba que vaya más allá de una duda razonable. El cargo de discriminación no se mantuvo.

Regresando al problema para determinar de cuántas formas podrían empacarse 3 de los 10 productos, si la investigación sugirió que el orden en el cual se empacaran los 3 productos afectaría las ventas, se debería determinar el número de permutaciones de los 10 elementos tomados 3 a la vez.

$$\begin{aligned} {}_{10}P_3 &= \frac{10!}{(10 - 3!)} \\ &= 720 \end{aligned}$$

Dados los valores de n y r, ${}_nP_r > {}_nC_r$ debido a que se pueden obtener otras permutaciones tan sólo cambiando el orden.

El número de órdenes terminadas de los siete agentes de ventas que compiten por el viaje a Cancún nombrados anteriormente es ${}_7P_7$ =5,040. Hay 5,040 formas distintas en las cuales esas siete personas pueden clasificarse u ordenarse de acuerdo con sus registros de ventas. Si el gerente de ventas va a recompensar a tres miembros de su equipo de ventas con un viaje y el orden de terminación es importante porque cada uno tiene un porcentaje diferente del viaje por el cual se paga, se tiene que ${}_7P_3 = 210$. Por otro lado, si los tres ganaran el mismo viaje y el orden no hiciera diferencia alguna, entonces se tendría ${}_7C_3 = 35$.

Ni las permutaciones ni las combinaciones permiten que se seleccione un elemento más de una vez. Si se permite la duplicación, se utilizaría el método de escogencia múltiple de conteo. El número de arreglos de *escogencia múltiple* de n elementos tomados r a la vez es ${}_nM_r = n^r$.

La pregunta anterior respecto al número de líneas telefónicas en una oficina grande requiere del modelo de escogencia múltiple, debido a que cada uno de los 10 dígitos puede utilizarse más de una vez. Si la oficina utiliza solo una extensión seguida por cuatro dígitos, tales como 555-*XXXX*, el número de líneas telefónicas es ${}_{10}M_4 = 10^4$ = 10,000. Si se utiliza una segunda extensión, como 556-*XXXX*, $({}_{10}M_4)(2)$, o 20,000 líneas son posibles.

En todos los casos anteriores, la selección se hizo de un solo conjunto. Los números telefónicos provinieron de un solo conjunto de 10 elementos. Si debemos escoger un elemento de dos o más conjuntos, es adecuado el proceso de *multiplicación*. Este principio requiere que simplemente se multiplique el número de elementos en cada conjunto. George intenta comprar un nuevo automóvil para sus vacaciones en Piney Flats. Puede escoger entre tres colores, dos tamaños de motor y cuatro diseños de interior distintos. ¿Cuántos modos distintos de transporte están a disposición de George? De acuerdo con el principio de multiplicación, debido a que tres conjuntos contienen tres elementos, dos elementos y cuatro elementos respectivamente, George debe decidir entre $3 \times 2 \times 4 = 24$ formas distintas de llegar a su destino.

Ejercicios de la sección

25. De los 15 miembros de la junta directiva de una gran empresa, ¿cuántos comités de 5 miembros pueden seleccionarse si el orden no importa?

26. De los 10 ejecutivos, 3 van a ser seleccionados para que sirvan como presidente, vicepresidente y tesorero. ¿Cuántas selecciones distintas son posibles?

27. Sus dos compañeros de cuarto están enfermos y a usted lo envían a un centro estudiantil para llevar comida a cada uno de ellos. Si usted debe escoger entre cinco selecciones, ¿en cuántas formas puede alimentar a sus compañeros? (Pista: ¿el orden hace la diferencia? ¿puede repetir?).

28. Como ingeniero constructor de Base Electronics, usted debe determinar ¿cuántos reproductores de CD puede ensamblar de tal forma que tengan un sistema de parlantes, un tocadiscos y un mecanismo de sintonización si usted puede escoger entre 3 sistemas distintos de parlantes, 4 de tocadiscos y 2 de sintonización?

29. De los 12 empleados de Worldwide Travel Services, 7 han tenido capacitación especial. Si 5 empleados van a ser enviados a Europa, ¿cuál es la probabilidad de que 3 estén dentro de los que han tenido entrenamiento especial?

Problemas resueltos

1. Probabilidades básicas Ortez Exports, con sede en Brasil, envía productos de carne congelada a clientes en Norteamérica (N), Europa (E) y Asia (A). Los paquetes protectores son o de lujo (D) o estándar (S). La siguiente tabla de probabilidad muestra las frecuencias relativas de muchos de los envíos más recientes.

	Europa	Norteamérica	Asia	Total
De lujo	0.083	0.167	0.167	0.417
Estándar	0.167	0.167	0.250	0.584
Total	0.250	0.334	0.417	1.000

Para finalizar la toma de decisiones de varios negocios pendientes, el Sr. Ortez determina la probabilidad de hacer envíos a tres destinos con cualquiera de las dos opciones de empaque.

a. De lujo para Europa $= P(D \cap E) = 0.083$, como se ve en la primera celda que muestra la probabilidad de los dos eventos de lujo y Europa.

b. ¿Es más probable que el Sr. Ortez envíe D a E o un S a A? $P(S \cap A) = 0.25 > 0.083$.

c. Estándar o cualquier empaque a Norteamérica:

$$P(S \cup N) = P(S) + P(N) - P(S \cap N) = 0.584 + 0.334 - 0.167 = 0.751$$

d. $P(N \cup D) = P(N) + P(D) - P(N \cap D) = 0.334 + 0.417 - 0.167 = 0.584.$

2. Uso del teorema de Bayes Sólo el 60% de los estudiantes de la clase de estadística del Profesor Harmond pasaron la primera prueba. De quienes pasaron, el 80% estudiaron, el 20% de quienes no pasaron sí estudiaron. ¿Debería usted estudiar para las pruebas de ese profesor?

Esto puede determinarse calculando la probabilidad de que usted pase, dado que usted estudió. De la información anterior: $P(P) = 0.60$, $P(S \mid P) = 0.80$, y $P(S \mid P) = 0.20$.

Entonces

$$P(P\mid S) = \frac{P(P \cap S)}{P(S)} = \frac{P(P) \times P(S\mid P)}{P(P) \times P(S\mid P) + P(\bar{P}) \times P(S\mid \bar{P})}$$

$$= \frac{(0.60)(0.80)}{(0.60)(0.80) + (0.40)(0.20)}$$

$$= \frac{0.48}{(0.48) + (0.08)}$$

$$= 0.857 > P(P) = 0.60$$

3. **¿Cómo puede ocurrir un evento?** La probabilidad de que John solucione un problema de estadística en particular es del 40%. Hay un 70% de probabilidad de que Fred lo solucione. ¿Cuál es la probabilidad de que sea resuelto? Asuma que John y Fred trabajan separadamente y los resultados son por tanto independientes.

   Puede resolverse si John lo soluciona (y Fred no lo hace), si Fred lo soluciona (y John no lo hace), o si ambos lo solucionan.

$$\begin{aligned} P(S) &= P(J \cap \bar{F}) + P(\bar{J} \cap F) + P(J \cap F) \\ &= (0.40)(0.30) + P(0.60)(0.70) + (0.40)(0.70) \\ &= 0.82 \end{aligned}$$

4. **Simplificación utilizando combinaciones** Harry vende al 30% de los clientes a quienes llama. Si él hace tres llamadas hoy ¿cuál es la probabilidad de que haga exactamente una venta?

   Harry puede vender a cualquiera de los 3 clientes:

$$\begin{aligned} P(1) &= S \cap \bar{S} \cap \bar{S} = (0.30)(0.70)(0.70) = 0.147 \\ &+ \bar{S} \cap S \cap \bar{S} = (0.70)(0.30)(0.70) = 0.147 \\ &+ \bar{S} \cap \bar{S} \cap S = \underline{(0.70)(0.70)(0.30) = 0.147} \\ &= \qquad\qquad\qquad\qquad\qquad 0.441 \end{aligned}$$

   Como alternativa, determine la probabilidad de que se haga una venta a cualquiera de los 3, y luego multiplique por el número de formas en las cuales puede hacerse una venta.

$$\begin{aligned} P(1) &= [(S \cap \bar{S} \cap \bar{S})] \times {}_3C_1 \\ &= 0.147 \times 3 \\ &= 0.441 \end{aligned}$$

Lista de fórmulas

[4.1]	$P(E) = \dfrac{\text{Número de veces en que el evento ha ocurrido en el pasado}}{\text{Número total de observaciones}}$	Frecuencia relativa
[4.2]	$P(E) = \dfrac{\text{Número de formas en las que puede ocurrir un evento}}{\text{Número total de posibles resultados}}$	Modelo clásico

[4.3]	$P(A\mid B) = \dfrac{P(A \cap B)}{P(B)}$	Probabilidad condicional
[4.4]	$P(A \cap B) = P(A) \times P(B)$	Probabilidad de eventos independientes
[4.5]	$P(A \cap B) = P(A) \times P(B\mid A)$	Probabilidad de eventos dependientes
[4.6]	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$	Probabilidad de eventos que no son mutuamente excluyentes
[4.7]	$P(A \cup B) = P(A) + P(B)$	Probabilidad de eventos mutuamente excluyentes
[4.8]	$P(A\mid D) = \dfrac{P(A \cap D)}{P(A \cap D) + P(B \cap D)} = \dfrac{P(A) \times P(D\mid A)}{P(A) \times P(D\mid A) + P(B) \times P(D\mid B)}$	Teorema de Bayes
[4.9]	${}_nP_r = \dfrac{n!}{(n-r)!}$	Permutaciones
[4.10]	${}_nC_r = \dfrac{n!}{r!\,(n-r)!}$	Combinaciones

Ejercicios del capítulo

30. Dell Publishing tiene 75 títulos distintos de libros, clasificados por tipo y costo de la siguiente manera:

	Costo		
Tipo	**US\$10**	**US\$15**	**US\$20**
Ficción	10	8	3
Biografías	12	10	9
Histórico	4	17	2

Halle la probabilidad de que un libro seleccionado aleatoriamente sea:

a. Ficción o cueste US\$10.

b. Histórico y cueste US\$20.

c. Histórico y cueste o US\$10 o US\$15.

d. Ficción y cueste menos de US\$20.

e. Biográfico o cueste US\$15.

f. Biográfico o cueste más de US\$10.

31. El departamento administrativo de State University tiene acceso a tres máquinas de fax. La probabilidad de que cada una esté fuera de servicio es 20/100, 25/100, y 30/100, respectivamente. Asumiendo independencia entre ellas, encuentre la probabilidad de que:

a. La primera y la segunda estén fuera de servicio.

b. La primera y la tercera estén fuera de servicio.

c. Todas estén fuera de servicio.

d. Ninguna esté fuera de servicio.

e. Una esté fuera de servicio

f. Dos estén fuera de servicio.

g. Dos o más estén fuera de servicio.

32. Mark compra tres acciones diferentes. La probabilidad de que la primera aumente su valor es 1/3, la probabilidad de que la segunda aumente es de 3 /4 y la probabilidad de que la tercera aumente su valor es de 1/10. Determine la probabilidad de que:

a. Todas aumenten de valor.

b. Ninguna aumente su valor.

c. Una aumente su valor

d. Dos aumenten su valor.

e. Por lo menos dos aumenten su valor.

f. Por lo menos una aumente su valor.

33. Una empresa de construcción local descubrió que sólo el 20% de todos los trabajos se terminaban a tiempo, mientras que el 30% sufrían sobrecostos. Además, los sobrecostos se presentaban el 75% de las veces en las que se terminaba el trabajo a tiempo. El propietario de la empresa desea conocer la probabilidad de que un trabajo tenga sobrecostos:

a. Y se realice a tiempo.

b. O se realice a tiempo.

34. Del problema anterior, ¿cómo puede probar que los sobre costos y la probabilidad de que un trabajo se haga a tiempo no son eventos mutuamente excluyentes?

35. La revista *Fortune* descubrió que el 10% de los trabajadores en cargos ejecutivos de alto nivel eran mujeres, y que el 3% de quienes estaban en un alto nivel eran mujeres con MBA (Master en Administración de Negocios). La junta directiva de una empresa grande, cuyo perfil ejecutivo se ajusta a esta descripción, desea seleccionar una de las mujeres ejecutivas aleatoriamente. ¿Qué probabilidad hay de que seleccionen a una mujer con MBA?

36. Diez unidades de producción se seleccionan de una línea de producción. Tres de estas 10 son defectuosas. Si se deben sacar 5 de las 10, ¿cuál es la probabilidad de que 2 sean defectuosas?

37. Biggie Burger ofrece sus hamburguesas con una selección de 5 condimentos diferentes: mostaza, pepinillos, salsa de tomate, cebolla y tomate. ¿Cuántas hamburguesas diferentes puede comprar?

38. En Illinois las placas de los autos constan de 3 letras seguidas de 3 números. ¿Cuántas placas distintas pueden hacerse?

39. Randy Rusty, propietario de Rusty Cars, Inc., ofrece a sus clientes automóviles con 8 opciones de color, 4 paquetes de interior, y 3 diseños diferentes de techo corredizo. ¿Entre cuántos automóviles pueden escoger los clientes de Randy?

40. Estudios hechos por la Asociación Nacional de Educación demuestran que el 30% de los profesores de la nación dejan la profesión después de 10 años. Además, entre quienes la abandonan, el 60% tienen un título avanzado, mientras que entre los que no dejan la profesión el 20% tienen un título avanzado. El Sr. Chips, el profesor favorito de los estudiantes, acaba de obtener un título avanzado. ¿Cuál es la probabilidad de que deje a los estudiantes y consiga un trabajo distinto?

41. Una empresa manufacturera tiene plantas en Chicago y Houston. La planta de Chicago produce el 40% de la producción total, con un 10% en la tasa de defectos. La planta de Houston tiene una tasa de defectos del 20%. Si sólo se encuentra que una unidad es defectuosa, ¿es más probable que provenga de Chicago o de Houston?

42. El presidente debe seleccionar 5 miembros de una lista de 12 senadores, de los cuales 7 lo apoyan y 5 le hacen oposición. Si él selecciona al azar, ¿cuál es la probabilidad de que la mayoría del comité apoye al presidente?

43. Kelly Katz vende teléfonos móviles ofreciendo 5 estilos, 4 colores y 7 opciones de servicio. ¿Cuántos teléfonos diferentes puede ofrecer el Sr. Katz a sus clientes?

PUESTA EN ESCENA

En el caso presentado en la sección Planteamiento del Escenario al inicio de este capítulo, el Sr. Michael Berry, presidente de la Asociación Nacional de Áreas de Esquí para analizar las condiciones financieras de 850 centros de esquí en la nación, Berry le proporciona la siguiente información que él ha recolectado: ninguno de los centros urbanos está en bancarrota, 635 de los centros de esquí son urbanos, 17 de los centros de esquí que están aislados están en bancarrota, 765 de los centros no están en bancarrota y 60 de todos los centros están ubicados en zonas aisladas.

Utilizando estos datos limitados, proporcione una respuesta a Berry que contenga todos los hechos relevantes que usted pueda aportar a sus esfuerzos por mejorar las condiciones financieras de los centros de esquí. El reporte debe contener una tabla de contingencia y una tabla de probabilidad. ¿Qué conclusiones puede sacar y qué resultados específicos considera usted que serían de mayor utilidad para Berry? Berry está interesado especialmente en cuál localidad es la que tiene mayor probabilidad de quedar en bancarrota.

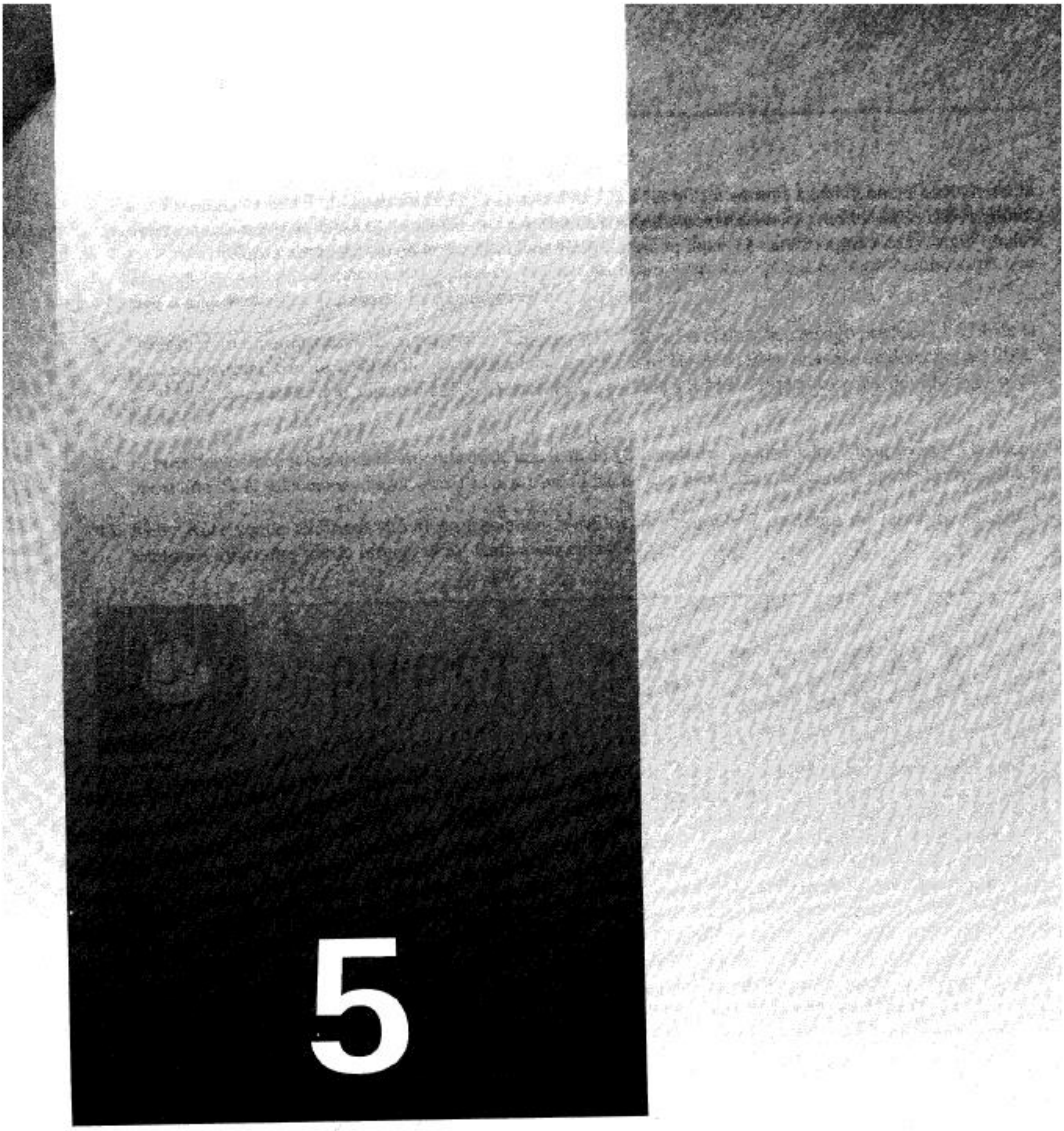
Formalice sus resultados en forma de un reporte de negocios tal y como se describe en el apéndice 1.

Del escenario a la vida real

En el análisis de la sección *Puesta en Escena* sobre los centros de esquí, se vio que la ubicación en un área urbana importante era un factor determinante en su éxito financiero. ¿Puede pensar en los motivos de esto? ¿Puede pensar en factores adicionales que contribuirían al éxito financiero de los centros de esquí?

La Asociación Mundial de Esquí (*www.worldski.com*) es una organización deportiva que organiza paquetes de viajes para esquiadores a centros de esquí de todo el mundo. En este sitio, visite "Directorio de Beneficios" y revise algunos de los paquetes disponibles para los miembros. ¿Cuál es el elemento común en todos estos convenios? ¿Cree usted que los convenios de descuentos serían un factor importante en el éxito de un centro de esquí?

La Asociación Nacional de Áreas de Esquí mantiene la Guía Oficial de Centros de Esquí (*The Official Ski Resort Guide*) (*www.skifund.com*). En las descripciones de los centros que se encuentran en este sitio, ¿qué características le atraen hacia centros distintos? La Asociación Colegiada de Esquí de Estados Unidos (*The U.S. Collegiate Ski Association*) (*www.beater.com/uscsa*) enumera las competencias de esquí programadas por equipos, fecha y lugar. ¿Las competencias de esquí podrían influir en su selección de un centro de esquí?



5

Distribuciones de probabilidad