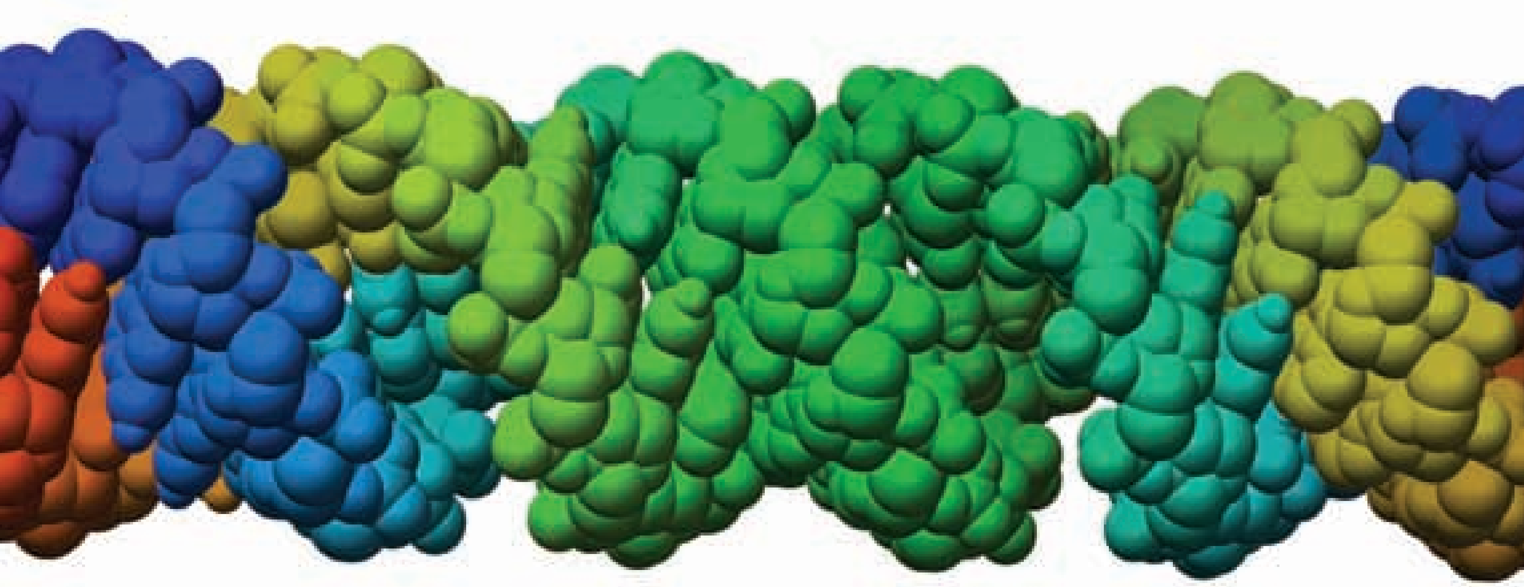
BIOESTADÍSTICA



PEARSON
Prentice Hall

R. Clifford Blair • Richard A. Taylor

Bioestadística

R. Clifford Blair
Universidad del Sur de Florida

Richard A. Taylor

TRADUCCIÓN
Anaid González Sarmiento
Verónica del Carmen Alba Ramírez
Traductoras profesionales

REVISIÓN TÉCNICA
Horacio Mauricio Rodríguez Morales
Departamento de Biofísica
Escuela Nacional de Ciencias Biológicas
Instituto Politécnico Nacional.

Ana Gabriela Prior Mier y Terán
División de Ciencias de la Salud
Instituto Tecnológico de Estudios Superiores de Monterrey
Campus Ciudad de México.

Mario Enrique Rendón Macías
División de Ciencias de la Salud
Instituto Tecnológico de Estudios Superiores de Monterrey
Campus Ciudad de México.
Instituto Mexicano del Seguro Social
Hospital de Pediatría
Centro Médico Nacional Siglo XXI

PEARSON
Educación ®

Datos de catalogación bibliográfica

BLAIR R. CLIFFORD y RICHARD A. TAYLOR

Bioestadística

PEARSON EDUCACIÓN, México, 2008

ISBN: 978-970-26-1196-7
Área: Matemáticas

Formato 20 x 25.5 cm — Páginas: 552

ISBN 0-13-117660-9

Edición en español
Editor: Rubén Fuerte Rivera
e-mail: ruben.fuerte@pearsoned.com
Editor de desarrollo: Felipe Hernández Carrasco
Supervisor de producción: Gustavo Rivas Romero

Edición en inglés
Executive Acquisitions Editor: *Petra Recter*
Editor-in-Chief: *Chris Hoag*
Project Manager: *Michael Bell*
Production Editor: *Raegan Keida Heerema*
Assistant Managing Editor: *Bayani Mendoza de Leon*
Senior Managing Editor: *Linda Mihatov Behrens*
Executive Managing Editor: *Kathleen Schiaparelli*
Manufacturing Buyer: *Maura Zaldivar*
Manufacturing Manager: *Alexis Heydt-Long*
Director of Marketing: *Patrice Jones*
Marketing Manager: *Wayne Parkins*
Marketing Assistant: *Jennifer de Leeuwerk*
Editorial Assistant: *Joanne Wendelken*
Art Director/Cover Designer: *Jayne Conte*
Creative Director: *Juan R. López*
Director of Creative Services: *Paul Belfanti*
Manager, Cover Visual Research & Permissions: *Karen Sanatar*
Cover Image: *Istockphoto.com*

PRIMERA EDICIÓN, 2008

Cámara Nacional de la Industria Editorial Mexicana. Reg. Núm. 1031

PEARSON Educación ®

ISBN: 978-970-26-1196-7

Impreso en México. *Printed in Mexico.*

1 2 3 4 5 6 7 8 9 0 - 11 10 09 08

Para Pal

Contenido

Prefacio

Como indica el título, este libro brinda una introducción a la bioestadística para estudiantes y profesionales de esta disciplina. Por lo general, se usa en cursos de nivel de licenciatura y posgrado diseñados para estudiantes que se especializan en disciplinas relacionadas con la salud, no propiamente en bioestadística. Es decir, este libro no ofrece la teoría o el rigor matemático que caracterizan a un primer curso para estudiantes de bioestadística, pero es apropiado para casi cualquier otra disciplina relacionada con la salud que requiera conocimientos básicos de bioestadística. Los requisitos esenciales de carácter matemático son mínimos; la mayoría de los estudiantes sólo necesitarán un nivel básico de álgebra.

¿Qué llevó a los autores a escribir un libro más de introducción a la bioestadística? Nuestra motivación tiene tres vertientes. La primera, elaborar un texto que funcione con igual efectividad tanto en un salón de clases tradicional como en ambientes de aprendizaje no tradicionales. Con esto último nos referimos a entornos en los que el contacto alumno-instructor está limitado por la proximidad física y/o por el método de enseñanza; entre ellos destacan los cursos de enseñanza a distancia basados en Internet o en sistemas satelitales, así como los cursos dirigidos por el propio estudiante. Nuestro deseo es presentar un libro tan claramente escrito que permita que el estudiante adquiera conocimientos estadísticos bajo la dirección de un instructor cuya interacción cara a cara con los alumnos puede ser continua o limitada.

Con este fin, se incluyen explicaciones mucho más detalladas de las que se encontrarían en la mayoría de los textos sobre el tema. El resultado es un libro que algunos calificarían de "prolijo", pero que en realidad pretende ofrecer claridad expositiva. Además, hemos incluido soluciones paso a paso para los ejercicios, en vez de sólo presentar respuestas satisfactorias. Tales soluciones incorporan referencias a las páginas en el texto, así como a ecuaciones específicas. De esta forma, los estudiantes que no logren obtener las respuestas de los ejercicios tienen la opción de remitirse a la sección de soluciones del libro y así obtener una explicación paso a paso que los conduzca a la resolución del problema.

Un segundo factor motivador fue introducir a los estudiantes a las pruebas de equivalencia. Las pruebas de equivalencia se usan comúnmente en ensayos clínicos y otros contextos, por lo que los estudiantes deben adquirir cierta familiaridad con su uso. Además, como afirman Hoenig y Heisey [23] en un artículo en el *American Statistician*, introducir a los estudiantes a las pruebas de equivalencia en las primeras etapas de su formación ayuda a cimentar su comprensión de las pruebas de hipótesis al dejar en claro que el hecho de no rechazar una hipótesis nula no constituye una evidencia de su validez. Los estudiantes aprenden que si se desea establecer la condición nula, se requieren pruebas (de equivalencia) especiales.

La tercera razón fue presentar los métodos no paramétricos bajo una nueva luz y desechar ciertos mitos relacionados con ellos. En los últimos años, con el advenimiento de las poderosas computadoras y los rápidos algoritmos, los métodos basados en la permutación se han utilizado con mayor frecuencia. Por esta razón, renunciamos al enfoque tradicional (y en nuestra opinión) anticuado, para presentar las técnicas no paramétricas en el más vasto contexto de los métodos basados en la permutación. Este enfoque tiene varias ventajas: **1.** Incluso el alumno que tiene escasa formación matemática puede discernir con claridad la lógica que hay detrás de la construcción de distribuciones muestrales relacionadas con estos métodos. **2.** El alumno puede ver que ciertos métodos no paramétricos tradicionales (por ejemplo, la prueba de Wilcoxon-

Mann-Whitney) son análogos basados en rangos de pruebas efectuadas con puntuaciones originales. **3.** Al presentar los métodos no paramétricos como versiones transformadas [8] de los estadísticos paramétricos con los cuales ya están familiarizados se elimina gran parte del misterio de esas pruebas. En otras palabras, al estudiante no se le presentan métodos nuevos y desconocidos que hacen referencia a extrañas tablas de valores críticos (que a veces van en contra de la intuición), sino que se le demuestra que las pruebas libres de distribución a menudo pueden realizarse aplicando los métodos paramétricos que aprendieron en capítulos anteriores a los rangos. Esto permite también que las tablas de valores críticos tomen formas con las que el estudiante ya está familiarizado. En resumen, se desmitifican los métodos no paramétricos.

Este libro está dividido en tres partes. Los primeros cuatro capítulos sientan las bases de todo lo que aparece después. El capítulo 1 establece el marco conceptual para el resto del libro. El capítulo 2 toma un enfoque un tanto tradicional en cuanto a la estadística descriptiva, pero incluye algunas opiniones innovadoras sobre la mediana, los percentiles y los rangos percentiles. El capítulo 3 presenta una opinión no teórica sobre la probabilidad y sienta las bases para los modelos basados en probabilidades que sustentan los métodos inferenciales siguientes. El capítulo 4 cubre las bases de la inferencia y es la piedra angular del libro. Este capítulo presenta también la lógica y el método de las pruebas de equivalencia.

Los capítulos 5 al 9 presentan técnicas específicas usadas con datos continuos y binarios. Además de los métodos tradicionales de muestras apareadas, de dos grupos, de k grupos, de correlación y regresión, se presentan métodos de equivalencia asociados con métodos de muestras apareadas y de dos grupos.

El capítulo 10 nos da una perspectiva general de los métodos basados en permutaciones y se aboca a pruebas específicas relacionadas. Éste es un capítulo largo que, con algo de complementación, podría usarse como base de un curso corto sobre métodos no paramétricos.

Hay más material en este libro de lo que puede cubrirse en un solo semestre. Sin embargo, el instructor no debe dudar en asignar partes del libro para las cuales no se puede dedicar un tiempo de clase/conferencia. Como se indicó antes, aunque las asignaciones del texto, de manera ideal, serán sustentadas con conferencias u otros medios, los detallados niveles de exposición permiten que se cubra el material que, de otra forma, tendría que omitirse debido al limitado tiempo de clase. En resumen, la intención de este libro es enseñar y no sólo instruir.

Se puede encontrar material complementario para este texto en:

`http://www.biostats-hs.com`

Estos materiales incluyen capítulos cortos relacionados con cálculos de Kaplan-Meier y pruebas log-rango, regresión logística y ANOVA factorial. También existen varios manuales descargables de software que pueden usarse como la base de un laboratorio de computación para acompañar el curso. Además, es posible encontrar muchos otros materiales útiles. Los lectores con preguntas o comentarios pueden ponerse en contacto con los autores a través de un vínculo de correo electrónico que aparece en ese sitio.

Además, permítanos decir que este texto se ha beneficiado de muchas pruebas de campo. Diversos instructores lo han utilizado tanto en aulas tradicionales como en escenarios de enseñanza a distancia. Se realizó un gran esfuerzo para recabar aportaciones de alumnos y maestros, las cuales dieron lugar a modificaciones y a un estilo general de presentación que difiere de la mayoría de los textos sobre el tema.

Finalmente, este libro fue revisado y analizado por los siguientes expertos, quienes se desempeñan en diversas disciplinas relacionadas con la salud: Sara Vesely, Centro de Ciencias de la

Salud de la Universidad de Oklahoma; Jessica L. Thomson, Centro de Ciencias de la Salud de la Universidad Estatal de Louisiana, Escuela de Salud Pública; Lynn E. Eberly, Universidad de Minnesota, Escuela de Salud Pública; Lisa M. Sullivan, Universidad de Boston; Hua Yun Chen, Universidad de Illinois en Chicago; Stephen C. Alder, Universidad de Utah; Kenneth R. Hess, Universidad Rice; Heather A. Young, Universidad George Washington; Bonnie Davis, Universidad de Nueva Inglaterra; Margaret Louis, Universidad de Nevada, Las Vegas; Reg Arthur Williams, Universidad de Michigan; Sudipto Banjeree, Universidad de Minnesota. Sus sugerencias, correcciones y opiniones se reflejan en todo el texto, aunque no se les puede responsabilizar por la forma final de éste. Expresamos nuestro más sincero aprecio a todos aquellos que contribuyeron con este esfuerzo, especialmente a los doctores James Mortimer y Lakshminarayan Rajaram de la Facultad de Salud Pública de la Universidad del Sur de Florida, quienes dirigieron el trabajo de las pruebas de campo.

Dr. R. Clifford Blair

Dr. Richard A. Taylor

CAPÍTULO 1

Fundamentos de bioestadística

1.1 INTRODUCCIÓN

Los investigadores de las disciplinas relacionadas con la salud utilizan una amplia variedad de herramientas para alcanzar el entendimiento de los fenómenos estudiados. Quizás el más importante de dichos estudios es la bioestadística. Ésta desempeña un papel fundamental en la recolección de análisis de datos en el contexto de experimentos clínicos, así como de estudios en otras áreas como epidemiología, política sanitaria, salud comunitaria y familiar, y salud ambiental y ocupacional.

Entonces, ¿qué es la bioestadística? Primero debemos decir que la bioestadística es una de las ramas del extenso campo de la estadística. La **estadística** es la disciplina interesada en (1) la organización y el resumen de datos, y (2) la obtención de conclusiones acerca de las características de algún conjunto de personas o cosas, cuando sólo una porción de estas características está disponible para su estudio. **Bioestadística**, por lo tanto, es una de las áreas de la estadística que trata principalmente con las ciencias biológicas y las disciplinas relacionadas con la medicina y la salud. De tal forma, este libro se interesa por el estudio de la estadística poniendo énfasis en su aplicación a las ciencias de la salud.

Cuando nos aproximamos al estudio de cualquier cuerpo de conocimiento organizado, en especial uno tan diverso y complejo como la bioestadística, es importante identificar un marco de referencia a partir del cual sea posible observar el material. Sin tal estructura organizada, los conceptos que serán aprendidos parecerían al estudiante como temas sin relación alguna, cuyos propósitos se perciben sólo vagamente. Esta situación es tolerable hasta cierto grado. Muchos de los elementos importantes de la bioestadística no pueden ser apreciados en su totalidad sino hasta que se yuxtaponen con otros elementos. Por lo tanto, su papel y utilidad en el gran esquema de la disciplina llegan a ser claros únicamente cuando se ven como parte de un todo.

Por fortuna, la bioestadística tiene un marco de estudio bastante natural que ayuda a aligerar este problema de cierta forma. En un curso de introducción a la bioestadística, la mayor parte de este material se puede estructurar alrededor de los conceptos de poblaciones y muestras. Estos conceptos constituyen el fundamento sobre el cual está organizado este libro.

1.2 POBLACIONES Y MUESTRAS

Quizás usted piense que debido al papel fundamental de las poblaciones en la estadística y la bioestadística, es posible que haya consensos en esta definición. Por desgracia, éste no es el caso. Compare las siguientes aseveraciones respecto de poblaciones tomadas de dos diferentes textos de estadística.

> Una población es un conjunto de personas (u objetos) que tienen una característica observable en común [29].

> Observe que la palabra *población* se refiere a datos, no a personas [36].

Estas dos aseveraciones son claramente desiguales y reflejan la imprecisión con la que el término se utiliza a menudo. Mucha de la confusión respecto de las poblaciones parte del hecho de que los especialistas en estadística utilizan el término en dos sentidos diferentes. El primero se refiere a que se puede hablar de **poblaciones populares,** y el segundo de **poblaciones estadísticas**. Las poblaciones populares están formadas por personas o cosas. Así, es habitual referirnos a la población de personas que habitan en Florida cuya prueba de hepatitis C resultó positiva, o a la población de venados en una provincia en particular en Michigan, que porta la garrapata responsable de la enfermedad de Lyme. Estas poblaciones están claramente conformadas por personas o cosas.

En contraste, las poblaciones estadísticas están conformadas por *características* de personas o cosas. Para comprender la distinción considere lo siguiente. Una población popular podría estar compuesta de los estudiantes de alguna universidad. Una población estadística, entonces, podría consistir en las presiones sanguíneas de estos mismos estudiantes. Asimismo, la población estadística posiblemente esté conformada por un indicador para cada estudiante sobre si ha experimentado alguna forma de abuso sexual en su vida o sobre su opinión respecto de la calidad de la educación que ha recibido.

Parece, entonces, que el primer autor citado arriba estaba intentando definir una población popular, mientras que la segunda aseveración estaba orientada a una población estadística. Hay otro problema sobre la segunda definición que debe ser aclarado. Las poblaciones estadísticas consisten en características de personas o cosas, independientemente de si han sido medidas o no. La palabra **datos** se refiere al registro de mediciones hechas sobre características. De modo que si las presiones sanguíneas de algunos de los estudiantes o de todos ellos se miden y registran de alguna manera, el resultado son datos. La distinción que se hará aquí es que las poblaciones estadísticas están conformadas por las características mismas y no por el registro de esas características.

Cuando tales características toman diferentes valores se conocen como **variables**. Aunque es posible que una población esté formada por una característica que no varía (es decir, una constante), esto sería de poco interés en un contexto estadístico y, por lo tanto, no será tratado en este libro. Para nuestros propósitos, los términos “característica” y “variable” se utilizarán de forma indistinta.

Obviamente, los tamaños de las poblaciones pueden variar. En la disciplina de la estadística es útil distinguir entre poblaciones finitas e infinitas, puesto que los métodos usados para tratar cada una difieren un poco. Las **poblaciones infinitas** pueden ser pensadas como poblaciones grandes, mientras que las **poblaciones finitas** son más pequeñas. Es evidente que la distinción es arbitraria. Los métodos descritos en este libro son generalmente apropiados para utilizarse con poblaciones infinitas.

Una **muestra** es un subconjunto de una población. Por ejemplo, las presiones sanguíneas de los estudiantes de un determinado grupo en la universidad antes mencionada constituirían una muestra (aunque no una escogida al azar).

El concepto de población es a menudo mucho más abstracto de lo que implica la discusión anterior. Por ejemplo, en un ensayo clínico la población podría estar constituida por las presiones sanguíneas de todos los varones con más de 65 años de edad, quienes alguna vez tomarán un nuevo medicamento contra la hipertensión. En estas circunstancias sería imposible enumerar la población, debido a que nadie conoce con exactitud quién tomará el nuevo medicamento y quién no. Por el contrario, la muestra casi siempre se define mejor. En un estudio sobre la eficacia del fármaco, el medicamento podría administrarse a 50 hombres con más de 65 años, quienes seguirán el protocolo del estudio. En este caso, la muestra se define con facilidad, ya que es posible identificar a las personas que están o no en la muestra.

En un entorno típico para este estudio, los investigadores medirían u observarían las características que conforman la muestra y tendrían que registrarlas como datos. Sin embargo, no sucedería lo mismo con la población. En el caso de una universidad grande, sería impráctico medir las presiones sanguíneas del cuerpo estudiantil entero, pero es absolutamente factible tomar medidas de una muestra de 50 presiones sanguíneas.

En esta sección se ha hecho una clara distinción entre la palabra población cuando se usa en un sentido popular y cuando se usa en un sentido estadístico. En los libros de estadística esta distinción comúnmente desaparece. No es raro leer: “Se utilizó una muestra de 50 sujetos en el estudio”. Es obvio que esta muestra fue tomada de una población de personas, lo cual implica que el término se utiliza en sentido popular. Usted también encontrará expresiones como: “La media de la muestra es 121”. Aquí, la muestra se refiere a una población estadística. Al igual que en la generalidad de los textos de estadística, en este libro usted encontrará ambos usos de la palabra. En la mayoría de los casos, el contexto dejará en claro el significado. Una vez comprendida la diferencia entre los significados estadístico y popular de la palabra “población” se evita una fuente potencial de confusión para los estudiantes novatos de estadística.

1.3 PARÁMETROS Y ESTADÍSTICOS

Los conceptos de parámetros y estadísticos están relacionados de manera muy estrecha con los de poblaciones y muestras. Un **parámetro** se define como cualquier resumen de los elementos de una población, mientras que el resumen de los elementos de una muestra se conoce como **estadístico**. (No hay que confundir la palabra “estadístico” cuando se emplea en este sentido, con “estadística”, que se utiliza para referirse a la disciplina de estudio. De nuevo, el contexto generalmente aclarará el significado). De acuerdo con estas definiciones, entonces, el promedio de las presiones sanguíneas de todos los estudiantes de la universidad mencionada sería un parámetro, mientras que el promedio de las presiones sanguíneas de los estudiantes de un grupo en particular de esa universidad sería un estadístico. Asimismo, la mediana de las presiones sanguíneas de todos los hombres por arriba de 65 años de edad, que alguna vez tomarán el medicamento contra la hipertensión, sería un parámetro; mientras que la mediana de las presiones sanguíneas de los 50 hombres que participaron en el estudio sería un estadístico. Observe que para obtener el valor de un parámetro o de un estadístico, se deben medir u observar los elementos de la población o muestra correspondiente, registrar estas medidas y observaciones en forma de datos, y después realizar el resumen de tales datos.

Un punto importante que se deduce de lo mencionado arriba es que los valores de los parámetros generalmente no están disponibles para el investigador, mientras que los valores de los estadísticos son fácilmente localizables. Gran parte del material de este libro se relaciona con este hecho.

La distinción entre parámetros y estadísticos es tan fundamental para el pensamiento estadístico, que generalmente se utilizan dos convenciones diferentes para su representación. En la primera, los parámetros se representan con letras griegas, mientras que los estadísticos se representan con el alfabeto romano o algunos de sus caracteres. Por ejemplo, el promedio (o media) de una población a menudo se designa con la letra griega μ (pronunciada "mu") mientras que el mismo resumen de datos de una muestra se representa mediante $\overline{x}$ ("x barra"). Una segunda convención representa los parámetros con letras mayúsculas del alfabeto romano y coloca un carácter, llamado "sombrero", sobre la(s) misma(s) letra(s) para representar estadísticos. Un ejemplo de esta convención es el uso de RR para representar el parámetro de la razón de riesgo (la cual se analizará en los capítulos 3, 5 y 6) y $\widehat{RR}$ —que se lee "RR sombrero"— para representar el estadístico. Algunas veces estas dos convenciones se combinan de manera que el parámetro quede representado por una letra griega y el estadístico por una letra romana con sombrero.

En este libro se combinarán ambas convenciones para adoptar las prácticas comunes. De cualquier forma, cuando hubiere riesgo de confusión, se indicará cuál convención se está usando.

1.4 ESTADÍSTICA DESCRIPTIVA E INFERENCIAL

Una vez asentados los principios básicos, resulta conceptualmente conveniente describir las disciplinas de la estadística y la bioestadística conformadas por dos componentes. El primer componente se refiere a la estadística descriptiva, mientras que el segundo se llama estadística inferencial. La **estadística descriptiva** está formada por varias técnicas utilizadas para resumir la información contenida en un conjunto de datos. Considere el siguiente problema.

Suponga que se realiza un estudio para determinar los niveles séricos de plomo de 150 niños que viven en las casas más viejas de un vecindario urbano en particular. Si uno investigara los hallazgos de este estudio, obtendría una lista de los resultados de las pruebas individuales. De tal forma, se reportaría que el primer resultado mostró un nivel de 20 mcg/dl (microgramos por decilitro, también designados como g/dl o como ug/dl), mientras que el segundo arrojó un nivel de 25 mcg/dl, y así sucesivamente. Después de incluir en la lista los 150 resultados de la prueba, es probable que el investigador alcanzara a comprender un poco de la información obtenida. Tal información no resumida abrumaría la habilidad del oyente para llegar a una conclusión significativa. Una respuesta más útil podría ser: "El *promedio* de los niveles séricos de plomo encontrados en los niños incluidos en el estudio fue de 30 mcg/dl." Otros resúmenes podrían incluir los valores más altos y los más bajos, así como varias representaciones gráficas de los datos. De esta manera, la estadística descriptiva se ocupa exactamente de lo que implica el término: descripción de datos. Para reiterar, los investigadores podrán obtener fácilmente los resúmenes de los datos relacionados con los elementos de una muestra (estadísticos), a diferencia de lo que sucede con los datos relacionados con la población (parámetros).

En contraste con la estadística descriptiva, la **estadística inferencial** está conformada por varias técnicas utilizadas para proveer información acerca de los valores de los parámetros basados en observaciones hechas sobre los valores de los estadísticos. Los sondeos de opinión son un ejemplo común de esta forma de inferencia. En un sondeo de opinión, una muestra de opiniones obtenidas de un grupo de personas relativamente pequeño es utilizada para arrojar conclusiones sobre las opiniones de alguna población. Por ejemplo, se podría preguntar a 1,000 personas si favorecen o no un determinado plan de salud que se administra a nivel federal. Si el 65% de los sondeos de opinión fueran favorables, entonces el encuestador intentaría utilizar esta información para obtener conclusiones sobre la proporción de opiniones favorables en todo el país. Note que en este caso el valor de un estadístico (la proporción de opiniones favorables en la muestra) se utiliza para comprender el valor de un parámetro no disponible (la proporción de opiniones favorables en el país). En el ejemplo anterior, los investigadores podrían calcular el promedio de las presiones sanguíneas de los 50 pacientes que recibieron el medicamento contra

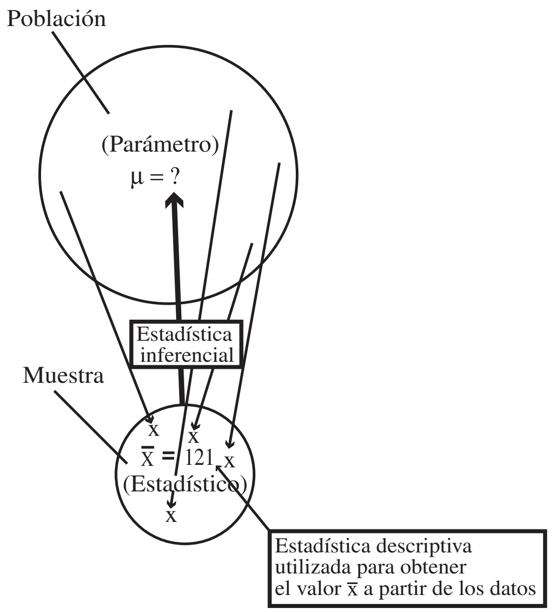


FIGURA 1.1: Esquema que muestra la relación entre poblaciones y muestras, parámetros y estadísticos, y la estadística descriptiva e inferencial.

la hipertensión (valor de un estadístico) con el propósito de estimar el promedio de la presión sanguínea de todos los hombres de más de 65 años de edad que alguna vez tomarán el medicamento (un valor de parámetro).

La relación entre poblaciones y muestras, parámetros y estadísticos, y la estadística descriptiva e inferencial se presenta en la figura 1.1. Note que la proporción de una población representada por una muestra, por lo común, es muy pequeña y ni siquiera se acerca a la proporción que esta figura podría implicar. Observe también que los elementos que conforman la muestra (representada por x en la figura) se obtienen aleatoriamente de la población.

1.5 ¿POR QUÉ POBLACIONES Y MUESTRAS?

Con anterioridad se afirmó que la estadística es una importante herramienta que los investigadores utilizan para conocer su materia de estudio. También se dijo que el fundamento de la estadística se basa en los conceptos de poblaciones y muestras.[1] Por ahora, quizá usted se pregunte cómo es que este fundamento se utiliza para ayudar a responder varias preguntas planteadas por los investigadores. Ése es el beneficio de los sondeos ya mencionados, pero, con exactitud, ¿cómo es que las inferencias que hacen de una población a partir de una muestra ayudan a determinar si un nuevo medicamento es efectivo o si la exposición a un factor de riesgo potencial se relaciona con la manifestación de alguna enfermedad?

Por desgracia, este proceso es difícil de percibir hasta que se unen más piezas de un mosaico bastante complejo. Por esta razón, usted necesitará dominar ciertos conceptos básicos, cuya utilidad no será apreciada totalmente sino hasta que pueda formar una imagen más completa. Por

[1] Una excepción es la estadística no paramétrica, de la que nos ocuparemos más adelante en este libro.

el momento, sólo tenga presente que los datos recopilados en los estudios de investigación pueden ser muy complejos e incluir numerosos elementos del azar. Se mostrará que los conceptos de poblaciones y muestras resultarán útiles para ayudar a separar tales elementos aleatorios de la realidad fundamental.

1.6 ¿QUÉ OCURRE AHORA?

Con frecuencia se critica que muchos libros de estadística se apresuran en explicar los conceptos fundamentales con el propósito de pasar a “lo interesante”. Como consecuencia, los estudiantes que están tomando el segundo o el tercer curso de estadística, a menudo no conocen bien las bases de los métodos que están estudiando. Por consiguiente, en este libro se hará un intento cuidadoso y metódico por establecer un fundamento adecuado, a partir del cual el contenido restante se desarrollará de manera lógica. Este proceso ya ha comenzado.

En el capítulo 2, después de algunos requisitos, usted podrá ocuparse del estudio de la estadística descriptiva. En el capítulo 3 se analizarán algunos fundamentos de probabilidad, el mecanismo que sirve de base a la inferencia. El capítulo 4 lo introducirá a la lógica de la inferencia, así como a algunos métodos simples de pruebas de hipótesis y a la construcción de intervalos de confianza. Los siguientes capítulos tratarán a fondo distintos métodos estadísticos comúnmente empleados en la investigación de las ciencias de la salud.

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

bioestadística 1
característica 2
datos 2
estadística 1
estadística descriptiva 4
estadística inferencial 4
estadístico 3
muestra 2
parámetro 3
población estadística 2
población finita 2
población infinita 2
población popular 2
variable 2

EJERCICIOS

1.1 ¿En cuáles tareas se enfoca principalmente la estadística?

1.2 Diferencie entre los siguientes conceptos:

a) muestras y poblaciones,
b) estadísticos y parámetros,
c) poblaciones populares y poblaciones estadísticas,
d) estadística descriptiva y estadística inferencial, y
e) poblaciones infinitas y poblaciones finitas

1.3 ¿Cuál es el significado del término “datos”?

1.4 Explique por qué las poblaciones no están formadas por datos, como algunos autores afirman.

A. Las siguientes preguntas se refieren al estudio de caso A (página 469).

1.5 Identifique la muestra en este estudio. ¿Usted diría que esta muestra se caracteriza como parte de una población estadística o de una población popular?

1.6 En este estudio, ¿la población está bien identificada? Explique su respuesta.

1.7 Describa la población de la mejor forma posible.

1.8 ¿Hay algún estadístico reportado en este estudio? De ser así, dé ejemplos.

1.9 ¿Hay algún parámetro reportado en este estudio? De ser así, dé ejemplos; de no ser así, explique por qué.

1.10 ¿Se encontrarán datos en este estudio? De ser así, dé ejemplos.

B. Las siguientes preguntas se refieren al estudio de caso B (página 470).

1.11 Identifique la muestra en este estudio. ¿Usted diría que esta muestra es característica de una población estadística o de una población popular?

1.12 ¿La población en este estudio está bien identificada? Explique su respuesta.

1.13 Describa la población de la mejor forma que le sea posible.

1.14 ¿Hay algún estadístico reportado en este estudio? De ser así, dé ejemplos.

1.15 ¿Hay algún parámetro reportado en este estudio? De ser así, dé ejemplos; de no ser así, explique por qué.

1.16 ¿Se encontrarán datos en este estudio? De ser así, dé ejemplos.

F. Las siguientes preguntas se refieren al estudio de caso F (página 473).

1.17 ¿Cree usted que los resultados de este estudio son aplicables a los casos de tuberculosis en Estados Unidos? Fundamente su respuesta.

CAPÍTULO 2

Métodos descriptivos

2.1 INTRODUCCIÓN

En el capítulo 1 precisamos que los intentos por obtener información de grandes y diversos conjuntos de datos pueden terminar frustrados a menos que se utilice alguna forma de resumen que revele aspectos relevantes de los datos. En este capítulo presentaremos algunos de los métodos más utilizados comúnmente. Los temas que se tratarán aquí se pueden dividir en distributivos, gráficos y métodos numéricos. Mientras que se aplican igualmente bien a los datos derivados de poblaciones y muestras, se aplican casi siempre a los datos de muestras.

Antes de tratar estos temas, sería de utilidad comprender primero dos temas relacionados, que a menudo se mencionan como escalas de medición y notación de sumatoria. Después de comprender estos requisitos previos, regresaremos a los tres temas mencionados arriba.

2.2 ESCALAS DE MEDICIÓN

Anteriormente mencionamos que las poblaciones y las muestras están formadas por variables que, a su vez, son características medibles y observables de personas o cosas, que toman diferentes valores. También dijimos que una vez que se toman y registran las mediciones, el resultado está constituido por datos. Pero, ¿que significa la palabra *medida*? En términos sencillos, significa que se asignan números, letras, palabras o algún otro símbolo a personas o cosas para dar a conocer la información acerca de la característica que se somete a medición. De tal manera, podemos asignar el número 220 a una persona con el fin de representar su nivel de colesterol o una “H” o una “M” para representar su género.

Con frecuencia no se reconoce el hecho de que las mediciones de variables pueden producir diferentes cantidades de información, dependiendo de la escala empleada en el proceso de medición. Esto significa que las mediciones que producen los números 1, 2 y 3 en una escala pueden transmitir una cantidad de información muy diferente sobre la variable que la que transmitirían los mismos números obtenidos con el uso de una escala diferente. Esto, a la vez, tiene consecuencias para el tratamiento estadístico de dichos datos. Todo esto quedará claro a partir de lo que sigue.

Las escalas que se explican en esta sección fueron descritas por S. S. Stevens [44]. De acuerdo con Stevens, se puede considerar que el proceso de medición existe en cuatro niveles diferentes, a los cuales él se refirió como escala nominal, ordinal, de intervalo (o intervalo igual) y de razón. A continuación se analiza cada una de ellas.

2.2.1 La escala nominal

La escala nominal es la menos elaborada de las cuatro y posee dos características principales. Primero, produce clasificaciones de personas o cosas con base en una evaluación cualitativa de la característica que se somete a consideración; segundo, su uso no brinda información con respecto a la cantidad. Considere el siguiente ejemplo referente a una clasificación por tipo de sangre.

En este caso se utiliza la escala nominal para asignar la nomenclatura del tipo de sangre A, B, AB u O a las personas, con base en un cierto criterio hematológico. Note que estas designaciones simplemente clasifican a las personas dentro de una de cuatro categorías de tipos de sangre. De esta manera, a todas las personas con el mismo tipo de sangre se les otorga la misma nomenclatura, mientras que aquellas con otro tipo de sangre reciben otra nomenclatura. Note también que no existe el concepto de "mayor que" o "menor que" implícito en estas clasificaciones. Esto significa que las mediciones a nivel nominal[1] no permiten comparaciones con personas o cosas sobre la base de *más* o *menos*, sino más bien sobre la base de *similar* o *distinto*.

Puede parecer que las nomenclaturas producidas por escalas nominales son de naturaleza numérica, pero no deben tratarse como tales. Cuando se realiza una encuesta telefónica o por correo, los hogares podrían clasificarse por área o código postal para fines de muestreo. En este caso los códigos de área 813 o 272 podrían ser simplemente indicadores de la localización geográfica de los hogares. Desde luego, no tendría sentido aseverar que los residentes con código de área 813 tienen más "código de área" que aquellos con código de área 272. Asimismo, operaciones aritméticas con dichos "números", por ejemplo, calcular un código de área promedio, producirían un resultado sin sentido. Sin embargo, sí es posible contar el número de hogares que pertenecen a cada categoría.

2.2.2 La escala ordinal

Como la escala nominal, la escala ordinal clasifica personas o cosas sobre la base de la característica evaluada. Sin embargo, a diferencia de la escala nominal, las clasificaciones producidas por esta escala incorporan los atributos muy importantes de "mayor que" o "menor que".

Por ejemplo, suponga que en el transcurso de un estudio sobre el manejo del dolor, se solicita a los pacientes que clasifiquen su percepción del dolor como "ninguno", "leve", "moderado" o "severo". Este esquema clasifica a los pacientes en una de las cuatro categorías que están ordenadas en términos de intensidad de dolor. Se ve fácilmente que la categoría "severo" representa una percepción *mayor* de dolor que la categoría "moderada" y así sucesivamente. En este sentido, se puede decir que la escala ordinal brinda más información acerca de la característica medida que la escala nominal. Otros ejemplos incluyen la clasificación de cierta patología en la etapa 1, 2, 3 o 4, o la jerarquización de pacientes en una situación de emergencia.

Note que mientras que este sistema permite ordenar categorías dependiendo de si poseen más o menos de la característica que se mide, no ofrece información sobre *qué tanto más* o *menos*. Un dolor severo representa más dolor que el dolor moderado, pero, ¿cuánto más? Un

---

[1] Muchos individuos dedicados a la psicocometría se oponen al uso de la palabra "medición" en relación con las escalas nominales, puesto que muchas definiciones del término implican cantidad.

paciente clasificado en una categoría puede tener mayor necesidad de cuidados que un paciente en otra categoría, pero, ¿cuánto más?

Los datos ordinales son comunes en investigaciones relacionadas con la salud, pero tradicionalmente han causado ciertas dificultades analíticas. Una solución común es tratar estadísticamente estos datos como si estuvieran en un nivel nominal. Aunque en cierto sentido es correcta, esta práctica generalmente desperdicia información y, por consiguiente, no es enteramente satisfactoria. Regresaremos a este problema en los siguientes capítulos.

2.2.3 La escala de intervalo (o intervalo igual)

Así como la escala ordinal agrega los atributos de mayor que y menor que a los datos de la escala nominal, la escala de intervalo (también llamada de intervalo igual) agrega los atributos de *cuánto más* y *cuánto menos* a aquellos de la escala ordinal. Mientras que hay numerosos ejemplos de escalas de intervalo, su análisis es bastante complejo y podría necesitar un estudio más profundo del campo de la psicometría[2] que el que se justifica en un libro de estadística. Por esta razón, el ejemplo que se presenta con más frecuencia es el del termómetro Fahrenheit, que resulta muy sencillo.

Cuando se obtiene la temperatura con un termómetro Fahrenheit, ésta se mide en unidades semejantes, lo cual permite cuantificar las diferencias. Una lectura de 70 representa cinco grados más de temperatura que una lectura de 65. Lo mismo es cierto para lecturas de 100 y 95. Entonces, esta escala no sólo permite comparaciones del tipo mayor que y menor que, sino que también indica la magnitud de la diferencia.

Un defecto de la escala de intervalo es la falta de un punto cero verdadero. En otras palabras, el punto cero en esta escala es una designación arbitraria, lo cual significa que no representa una ausencia de la característica medida. De tal manera, es posible tener una temperatura de 0° en un día en particular y una lectura de −10° al siguiente. La lectura de 0 no significa que no hubo temperatura, sino que fue simplemente otro punto en la escala. Resulta que esta escala no permite la formación de proporciones con significado. No se puede afirmar con validez que una lectura de 80° representa dos veces más de temperatura que una lectura de 40°.

2.2.4 La escala de razón

La escala de razón es similar a la escala de intervalo, con excepción de que posee un verdadero punto cero. Mediciones físicas como la estatura y el peso son ejemplos comunes. Cuando algo tiene peso cero, el cero indica que no hay peso presente.

2.2.5 Datos continuos y discretos

Una perspectiva sencilla de los datos los divide en continuos o discretos. Una *variable* **continua** es aquella que, al menos teóricamente, puede tomar cualquier valor en un rango especificado. Por ejemplo, una persona puede pesar 72 kilogramos, mientras que otra pesa 73 kilogramos, pero es posible encontrar un peso entre esos dos, por ejemplo, 72.5 kilogramos. También podemos encontrar un peso entre 72 y 72.5 kilogramos, como 72.25 kilos. Teóricamente este proceso podría continuar por siempre, aunque encontraríamos eventualmente que no tenemos una escala lo suficientemente sensible para hacer las distinciones necesarias. Entonces, el peso es una variable continua.

---

[2] La teoría o técnica psicológica de la medición mental.

Una variable **discreta** es aquella que no es continua. Por ejemplo, el número de personas con hogares en una determinada área geográfica puede ser 1, 2, 3, 4 y así sucesivamente, pero no puede ser 2.1367. Simplemente, las variables discretas se miden en unidades discretas y no en un continuo.

Las variables discretas que únicamente pueden tomar uno de dos valores, por ejemplo, hombre o mujer, muerto o vivo, positivo o negativo, se conocen como variables **dicotómicas**. Algunos métodos estadísticos son específicamente designados para utilizarse con datos dicotómicos.

Se podría decir que todos los *datos* son discretos porque todos los métodos de medición están limitados por su nivel de precisión, por lo que producen datos en unidades discretas más que continuas. Sin embargo, los datos obtenidos de variables continuas generalmente se consideran y se tratan como continuos, mientras que los datos de variables discretas se tratan como discretos. En ocasiones los investigadores miden una variable continua, pero registran sus descubrimientos a propósito como datos discretos. Esto ocurriría, por ejemplo, si se registrara que la presión sanguínea se encuentra dentro del rango normal o fuera del rango normal. La clasificación de datos (a diferencia de las variables) como discretos o continuos tiene reconocidamente un componente arbitrario.

2.2.6 Comentarios adicionales sobre las escalas

La conceptualización de las cuatro escalas presentadas anteriormente se formuló primero en el contexto de la psicometría y no en el de la teoría estadística. Su incorporación y potencial contribución a la literatura estadística no ha carecido de controversia [33]. Como mínimo, ellas han proporcionado un marco de referencia útil para varias estrategias analíticas. Por ejemplo, algunos métodos analíticos son claramente apropiados para utilizarse con datos nominales, mientras que otros se emplean con mayor provecho con datos de intervalo o de razón. Los datos ordinales plantean otro conjunto de preguntas analíticas. Las opiniones en algunos de estos temas varían.

2.3 NOTACIÓN DE SUMATORIA

El análisis estadístico de datos a menudo requiere sumarlos de alguna manera. Un ejemplo común es el cálculo del promedio (o media) de un conjunto de datos. En este caso los datos se suman y luego se dividen entre el número de observaciones en el conjunto de datos. Pero no todas las sumatorias son tan sencillas; algunas veces se debe sumar únicamente parte de los datos, o se deben elevar al cuadrado antes de sumarlos, o tal vez sea necesario sumarlos y después elevarlos al cuadrado. La **notación de sumatoria** es la notación que se utiliza para indicar exactamente cómo se llevará a cabo la suma. Mediante el entendimiento de unas cuantas reglas simples de la sumatoria, usted comprenderá las fórmulas que se presentan más adelante.

2.3.1 Notación básica

Suponga que se escribe una lista de cinco números en un orden arbitrario. Llame al primer número x_1, al segundo x_2 y así sucesivamente. Si quisiéramos indicar que esos números deben sumarse, podríamos escribir la instrucción

$$x_1 + x_2 + x_3 + x_4 + x_5.$$

Una forma breve de esta notación puede escribirse como

$$\sum_{i=1}^{5} x_i.$$

La notación $\sum x$ indica que los valores x serán sumados, mientras que el subíndice i en la x actúa como el portador de los número 1 al 5. La notación $i = 1$ muestra que la sumatoria debe empezar con x_1, mientras que el 5 indica que la sumatoria terminará con x_5. En otras palabras, todos los números en el conjunto se deben sumar.

Suponga ahora que la sumatoria

$$\sum_{i=2}^{4} x_i$$

se llevará a cabo con los números 3, 0, 5, 9, 2 y 7. En este caso la suma empezará con x_2 y terminará con x_4 produciendo $0 + 5 + 9 = 14$ como resultado. Si usted desea indicar que la suma incluirá el último número del conjunto de datos pero no se conoce cuántos números estarán implicados, se utiliza una n en lugar del número final. Considere lo siguiente

$$\sum_{i=2}^{n} x_i^2$$

Esto indica que se sumarán los valores elevados al cuadrado y que la suma comienza con x_2 y continúa hasta el último número. Utilizando los datos del ejemplo esto sería

$$0^2 + 5^2 + 9^2 + 7^2 = 159$$

Note también que

$$\left(\sum_{i=1}^{n} x_i\right)^2 = (3+0+5+9+2+7)^2 \neq \sum_{i=1}^{n} x_i^2 = 3^2 + 0^2 + 5^2 + 9^2 + 2^2 + 7^2$$

Muchos cálculos estadísticos requieren que los datos estén ordenados y que luego se haga una suma parcial. En este libro las sumas incluirán casi siempre todos los valores, lo que nos permitirá proporcionar notación adicional. Cuando éste no sea el caso, se indicará.

2.3.2 Algunas reglas de la sumatoria

Las cuatro reglas siguientes ayudarán a comprender las fórmulas que se presentan más adelante.

1. $\displaystyle\sum_{i=1}^{n} c = nc$
2. $\displaystyle\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i$
3. $\displaystyle\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$
4. $\displaystyle\sum_{i=1}^{n} (x_i + y_i) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$

La primera regla plantea que la suma de una constante (c) es igual a n (el número de constantes) por la constante. Esto es

$$\sum_{i=1}^{n} c = \overbrace{(c + c + \ldots + c)}^{n \text{ valores}} = nc$$

Suponga que la constante a ser sumada tiene un valor de tres y que son cuatro de ellas. Entonces tenemos

$$\sum_{i=1}^{n} 3 = \overbrace{(3 + 3 + 3 + 3)}^{4 \text{ valores}} = 4 \cdot 3 = 12$$

La segunda regla establece que la suma del producto de una constante y una variable (x) es igual al producto de la constante y la suma de la variable. Esto es

$$\begin{aligned}\sum_{i=1}^{n} cx_i &= (cx_1 + cx_2 + \cdots + cx_n) \\ &= c(x_1 + x_2 + \cdots x_n) \\ &= c\sum_{1=1}^{n} x_i\end{aligned}$$

Nuevamente, dejando que la constante tome un valor de 3 y que la variable tome los valores 3, 0, 5, 9, 2 y 7 tenemos que

$$\begin{aligned}\sum_{i=1}^{6} 3x_i &= (3 \cdot 3 + 3 \cdot 0 + 3 \cdot 5 + 3 \cdot 9 + 3 \cdot 2 + 3 \cdot 7) \\ &= 3\ (3 + 0 + 5 + 9 + 2 + 7) = 3 \cdot 26 = 78\end{aligned}$$

La tercer regla establece que la suma de la suma de dos variables (x, y) puede expresarse como la suma de la primera más la suma de la segunda. Si permitimos que y represente la segunda variable, entonces tenemos que

$$\begin{aligned}\sum_{i=1}^{n} (x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + \cdots + (x_n + y_n) \\ &= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n \\ &= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i\end{aligned}$$

Note que obtenemos este resultado aun si x y/o y son constantes. Ahora suponga que las sumas de la variable $(x_i + y_i) = (3 + 4) + (0 + 2)$ y $(5 + 1)$ serán sumadas. Esto se puede hacer de la siguiente manera

$$\begin{aligned}\sum_{i=1}^{3} (x_i + y_i) &= (3 + 4) + (0 + 2) + (5 + 1) \\ &= (3 + 0 + 5) + (4 + 2 + 1) \\ &= 8 + 7 + 15\end{aligned}$$

La regla cuatro se deduce directamente de la regla tres.

Ejemplo de aplicación

Estas reglas pueden utilizarse para encontrar un resultado simple pero importante. Suponga que la media de cierto conjunto de datos se resta de cada elemento del conjunto y que el resultado se suma. ¿Cuál sería el resultado?

$$\begin{aligned}\sum_{i=1}^{3}(x_i - \overline{x}_i) &= \overbrace{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \overline{x}}^{\text{regla 4}} \\ &= \sum_{i=1}^{n} x_i - \overbrace{n\overline{x}}^{\text{regla 1}} \\ &= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0\end{aligned}$$

El último resultado proviene de la definición de $\overline{x}$. (Véase la ecuación 2.1 en la página 25.)

Con los preliminares completos ahora dirigiremos la atención al principal enfoque de este capítulo, la descripción de datos. Para este fin, examinaremos algunas técnicas de distribución, gráficas y numéricas que se utilizan comúnmente para este propósito.

2.4 DISTRIBUCIONES

La tabla 2.1 muestra las respuestas (ficticias) de 60 pacientes postoperados, a quienes se solicitó calificar su percepción del dolor en una escala ordinal de cuatro puntos, como parte de un estudio de manejo del dolor. Como puede verse, estos datos desorganizados son básicamente no informativos en lo que se refiere a los patrones de respuesta. ¿Algunos niveles de dolor dominaron? ¿Era común el dolor severo? ¿Qué proporción de pacientes no tenía dolor? ¿Qué proporción sufría de dolor leve o de menor intensidad?

Con esta pequeña cantidad de datos usted puede pasar unos minutos observando la tabla para formular respuestas aproximadas a estas preguntas. Sin embargo, esta estrategia no sería efectiva con un gran conjunto de datos. Aun con este número limitado de respuestas sería conveniente reordenar los datos para facilitar la obtención de las respuestas.

2.4.1 Distribuciones de frecuencias

La tabla 2.2 muestra estos datos ordenados en distribuciones de frecuencias, frecuencias relativas, frecuencias acumulativas y frecuencias relativas acumulativas. La primera columna lista las categorías de la escala de menor a mayor. La segunda muestra la frecuencia de respuesta para cada categoría, que se obtiene mediante el conteo del número de veces que ocurre cada respuesta en el conjunto de datos. La **frecuencia**, entonces, es el *número* de respuestas de cada tipo.

2.4.2 Distribuciones de frecuencias relativas

La tercer columna de la tabla 2.2 muestra la frecuencia relativa de respuesta, la cual se obtiene dividiendo cada frecuencia entre el número total de respuestas (en este caso 60). La **frecuencia relativa**, entonces, es la *proporción* de respuestas de cada tipo.

TABLA 2.1: Mediciones de dolor percibido de 60 pacientes.

Número de paciente	Nivel de dolor	Número de paciente	Nivel de dolor	Número de paciente	Nivel de dolor	Número de paciente	Nivel de dolor
1	moderado	16	leve	31	ninguno	46	severo
2	ninguno	17	leve	32	moderado	47	ninguno
3	leve	18	moderado	33	ninguno	48	ninguno
4	ninguno	19	ninguno	34	ninguno	49	leve
5	severo	20	ninguno	35	leve	50	leve
6	ninguno	21	leve	36	ninguno	51	leve
7	moderado	22	ninguno	37	moderado	52	ninguno
8	ninguno	23	ninguno	38	leve	53	leve
9	ninguno	24	leve	39	ninguno	54	severo
10	leve	25	moderado	40	ninguno	55	moderado
11	leve	26	moderado	41	ninguno	56	ninguno
12	ninguno	27	ninguno	42	ninguno	57	ninguno
13	leve	28	ninguno	43	ninguno	58	ninguno
14	leve	29	leve	44	ninguno	59	leve
15	ninguno	30	severo	45	ninguno	60	ninguno

TABLA 2.2: Distribuciones de mediciones de dolor percibido.

Categoría de dolor	Frecuencia	Frecuencia relativa	Frecuencia acumulativa	Frecuencia relativa acumulativa
Severo	4	.07	60	1.00
Moderado	8	.13	56	.93
Leve	17	.28	48	.80
Ninguno	31	.52	31	.52

Usted puede percibir rápidamente a partir de las dos primeras columnas que el mayor número de pacientes (31) indicó no haber tenido dolor. Este número representa .52 (o 52%) del total de la muestra. El dolor severo fue menos común, pues únicamente 4 personas (.07 de la muestra) eligieron esta categoría. En general, el número de respuestas en las categorías disminuyó conforme éstas representaban niveles más altos de dolor.

2.4.3 Distribuciones de frecuencias acumulativas

La columna de la **frecuencia acumulativa** muestra el *número* de pacientes que indicaron que su dolor era *menor o igual* al nivel representado. Por ejemplo, 48 pacientes (31 + 17) clasificaron su dolor como leve o menor que leve, mientras que 56 pacientes (31 + 17 + 8) percibieron su dolor como moderado o menor que moderado. La frecuencia acumulativa se obtiene mediante la suma de la frecuencia en una categoría dada con las categorías que indican un nivel menor de la variable medida.

2.4.4 Distribuciones de frecuencias relativas acumulativas

La frecuencia relativa acumulativa se calcula al dividir cada frecuencia acumulativa entre el número total de encuestados. Se puede ver que .80 de los pacientes creyeron que su dolor era leve o de menor intensidad, mientras que .93 sintieron que su dolor era moderado o de menor intensidad. La columna de la **frecuencia relativa acumulativa**, entonces, muestra la *proporción* de los pacientes que indicaron que su dolor fue *menor que o igual que* el nivel representado.

Las distribuciones de frecuencias, frecuencias relativas, frecuencias acumulativas y frecuencias relativas acumulativas que se muestran en la tabla 2.2 fueron calculadas para una variable de nivel ordinal. Las primeras dos distribuciones también pueden utilizarse para una variable de nivel nominal. Obviamente las distribuciones acumulativas no serían apropiadas en este caso puesto que no hay un orden cuantitativo para una variable de nivel nominal. Ahora veremos algunos temas de distribución relacionados con las variables continuas.

2.4.5 Distribuciones agrupadas

La tabla 2.3 presenta una distribución de frecuencias de las presiones sanguíneas sistólicas (ficticias) de 144 adolescentes moderadamente obesos. En esta tabla las frecuencias se relacionan con los valores de la presión sanguínea más que con categorías discretas, como fue el caso en la tabla 2.2. Como resultado, hay un gran número de valores y sus frecuencias. Esto puede causar dificultades de interpretación, especialmente cuando las frecuencias individuales son pequeñas e incluyen el cero. En estos casos a veces es útil reducir el número de valores mediante la formación de grupos. Entonces se pueden dar distribuciones de frecuencias, frecuencias relativas, frecuencias acumulativas y frecuencias relativas acumulativas para estos grupos de valores en lugar de valores individuales.

La tabla 2.4 presenta distribuciones agrupadas para los datos de la presión sanguínea. Como puede observarse, los valores de presión sanguínea se colocaron en intervalos que técnicamente se conocen como **intervalos de clase**. Las diversas distribuciones se basan entonces en esos intervalos. Al reducir los datos en esta forma, los patrones de respuesta se distinguen con mayor facilidad. Pero el precio que se paga por la comodidad interpretativa es la pérdida de información. Por ejemplo, mientras que es fácil ver que alrededor del 21.5% de los valores cae en el intervalo 135-139, no hay información acerca de los valores individuales en este intervalo.

Al construir tablas de este tipo se deben responder dos preguntas relacionadas. ¿En cuántos intervalos se deben agrupar los valores y qué tan grandes deberán ser los intervalos? Muy pocos intervalos provocan la pérdida de mucha información, mientras que muchos intervalos hacen fracasar el propósito de resumir los datos. El tamaño de los intervalos dependerá del número de intervalos utilizados y viceversa. No existen reglas rígidas y rápidas al respecto. En esencia, usted deseará presentar los datos dándoles el mayor significado posible. Sin embargo, hay algunas reglas generales que sirven como guía. Una sugerencia común es que no debe haber menos de seis ni más de 15 intervalos. Otra regla útil es que, cuando sea posible, se debe usar una anchura en los intervalos de clase de 5 unidades, de 10 unidades o de algún múltiplo de 10 para que el resumen de los datos sea más comprensible.

En el caso de la tabla 2.4, esta última sugerencia parece factible. Para determinar si la creación de intervalos de clase de 5 unidades producirá o no un número razonable de intervalos, el

TABLA 2.3: Distribución de frecuencias de las presiones sanguíneas de 144 adolescentes moderadamente obesos.

PS	Frec.	PS	Frec.	PS	Frec.	PS	Frec.
143	2	128	3	113	0	98	2
142	0	127	3	112	0	97	2
141	0	126	7	111	3	96	2
140	4	125	4	110	3	95	3
139	6	124	4	109	1	94	0
138	3	123	2	108	0	93	1
137	11	122	3	107	2	92	2
136	3	121	1	106	1	91	0
135	8	120	3	105	2	90	1
134	5	119	2	104	0	89	0
133	8	118	2	103	1	88	0
132	4	117	1	102	1	87	0
131	3	116	3	101	0	86	1
130	5	115	6	100	4		
129	3	114	2	99	1		

TABLA 2.4: Distribuciones agrupadas de las presiones sanguíneas sistólicas utilizando 12 intervalos.

Intervalo	Frecuencia	Frecuencia relativa	Frecuencia acumulativa	Frecuencia relativa acumulativa
140-144	6	.042	144	1.000
135-139	31	.215	138	.958
130-134	25	.174	107	.743
125-129	20	.139	82	.569
120-124	13	.090	62	.431
115-119	14	.097	49	.340
110-114	8	.056	35	.243
105-109	6	.042	27	.188
100-104	6	.042	21	.146
95-99	10	.069	15	.104
90-94	4	.028	5	.035
85-89	1	.007	1	.007

TABLA 2.5: Distribuciones agrupadas de las presiones sanguíneas sistólicas utilizando ocho intervalos.

Intervalo	Frecuencia	Frecuencia relativa	Frecuencia acumulativa	Frecuencia relativa acumulativa
142-149	2	.014	144	1.000
134-141	40	.278	142	.986
126-133	36	.250	102	.708
118-125	21	.146	66	.458
110-117	18	.125	45	.313
102-109	8	.056	27	.188
94-101	14	.097	19	.132
86-93	5	.035	5	.035

rango[3] de los valores de presión sanguínea se dividió entre cinco. Esto es

$$\frac{(143-86)}{5}=11.4$$

A partir de este resultado decidimos utilizar 12 intervalos con una anchura de cinco.

Siempre que sea posible, los intervalos de clase también deben tener la misma longitud, donde el extremo inferior del primer intervalo sea menor que o igual a la medida más pequeña del conjunto de datos, y el extremo superior del último intervalo sea mayor que o igual a la medida más grande. También debemos señalar que los intervalos deben ser contiguos, pero sin traslaparse. La tabla 2.4 sigue estas pautas.

Como se mencionó anteriormente, el número y tamaño de los intervalos es flexible. Para propósitos comparativos, la tabla 2.5 muestra los mismos datos integrados en ocho intervalos. El tamaño de los intervalos se estimó dividiendo el rango entre ocho, lo que dio un resultado de 7.125 y se redondeó a ocho. Las tablas 2.4 y 2.5 serán comparadas en el siguiente apartado.

No siempre es necesario formar distribuciones agrupadas para variables continuas. Cuando el número de valores no es muy grande, las distribuciones pueden basarse en datos no agrupados.

2.5 GRÁFICAS

Con frecuencia es más informativo representar las distribuciones como gráficas que en la forma tabular utilizada en el apartado anterior. Muchas formas gráficas están disponibles. En esta sección usted conocerá la gráfica de barras, el histograma, el polígono y gráficas de tallo y hojas.

2.5.1 Gráficas de barras

La figura 2.1 muestra la distribución de frecuencias relativas de la tabla 2.2 en la forma de una gráfica de barras. Como se observa, las categorías de respuesta están representadas a lo largo del eje horizontal (x) mientras que las frecuencias relativas aparecen a lo largo del eje vertical (y). La frecuencia relativa para cada categoría de respuesta se lee como la altura, medida con respecto al

[3] El rango, que se define como el valor más alto menos el valor más bajo, se analizará más adelante.

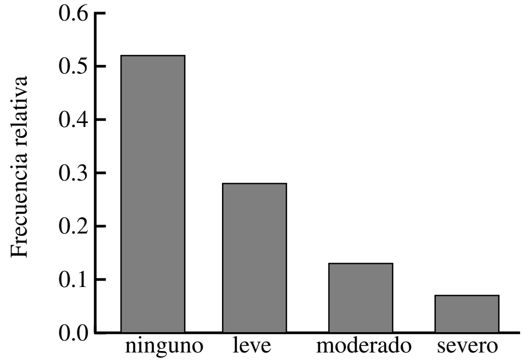


FIGURA 2.1: Gráfica de barras de las frecuencias relativas de las puntuaciones del dolor.

eje *y*, de una barra colocada por arriba de la categoría. Las gráficas de barras de frecuencias, frecuencias acumulativas y frecuencias relativas acumulativas se construyen de manera similar, donde la altura de las barras indica estas cantidades. Es fácil ver en esta figura por qué las gráficas se utilizan con mucha frecuencia para describir datos. Aun un examen precipitado revela una imagen clara de los patrones de respuesta.

2.5.2 Histogramas

La figura 2.2 es un histograma de las frecuencias relativas de las medidas de la presión sanguínea de la tabla 2.4. Se observan muchas diferencias entre las figuras 2.1 y 2.2. Primero, las barras en la figura 2.2 son contiguas, mientras que aquellas en la figura 2.1 no lo son. Esto se hizo para enfatizar el hecho de que los datos descritos en la figura 2.2 son continuos, mientras que los datos de la figura 2.1 son discretos. Ésta es la diferencia fundamental entre las gráficas de barras y los histogramas. Las primeras se emplean con datos discretos y usan barras no contiguas, mientras que los últimos representan datos continuos y utilizan barras contiguas. El rótulo del eje *x* de la figura 2.2 también requiere de una explicación.

Teóricamente al menos, el valor medido de una variable continua es indicativo del rango, y no de un punto en la escala de medición. Por ejemplo, una presión sanguínea de 120 no significa que el paciente tenga una presión sanguínea exactamente de 120, sino que su presión sanguínea cae en algún punto entre 119.5 y 120.5. El dispositivo de medición no es lo suficientemente preciso para distinguir entre los valores de 119.7 y 120.1, por ejemplo, por lo que estos valores se agrupan como 120. Esta conceptualización no está relacionada con el error de medición, sino con el nivel de precisión del dispositivo de medición. Resulta que si se determina que un intervalo de clase va de 85 a 89, las presiones sanguíneas realmente representan un rango de 84.5 a 89.5. Estos valores se llaman los **límites reales** superior e inferior del intervalo. Con propósitos comparativos, en la figura 2.3 se muestra un histograma de la distribución de frecuencias relativas con ocho intervalos de las mediciones de la presión sanguínea.

2.5.3 Polígonos

Otra gráfica usada frecuentemente es el polígono. Como sucede con el histograma, se pueden construir polígonos para cualquiera de las distribuciones aquí presentadas. La elección de una u

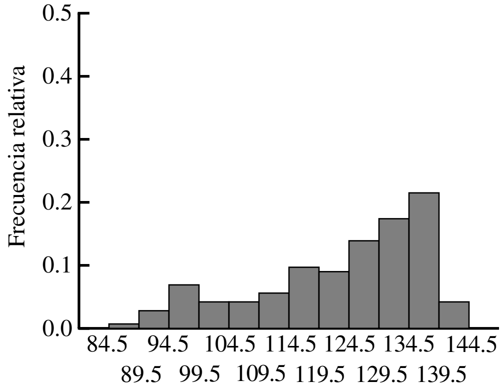


FIGURA 2.2: Histograma de frecuencias relativas de las mediciones de presión sanguínea en 12 categorías.

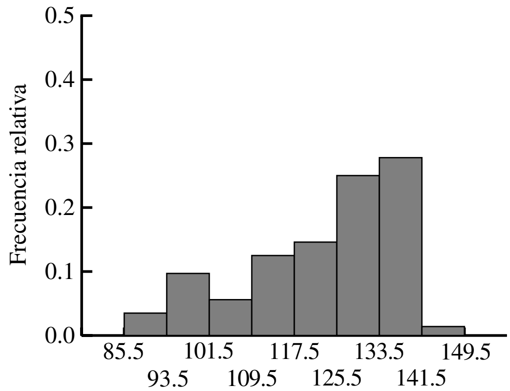


FIGURA 2.3: Histograma de frecuencias relativas de las mediciones de presión sanguínea en ocho categorías.

otra es en gran parte una cuestión de gusto y conveniencia. La figura 2.4 es un polígono de la misma distribución de la figura 2.2. Los polígonos se construyen en forma similar a los histogramas, excepto que en lugar de poner una barra sobre cada intervalo, se coloca un punto a la altura apropiada del eje *y*. En el caso de los polígonos de frecuencias y de frecuencias relativas, el punto se coloca en el punto medio del intervalo, en tanto que en las distribuciones acumulativas el punto se coloca en el límite real superior del intervalo. Estos puntos se conectan luego con líneas rectas que se unen al eje *x* en el extremo inferior, y en los polígonos de frecuencias y de frecuencias relativas, con los extremos superiores de la distribución. Para los polígonos de frecuencias y de frecuencias relativas, los puntos en los que la línea hace contacto con el eje *x* corresponden a los que serían los puntos medios de un intervalo adicional en cada extremo de la distribución. Esto se observa claramente en la figura 2.5, en la página 22, donde la figura 2.4 está superpuesta en la figura 2.2. Los polígonos de frecuencias acumulativas y de frecuencias relativas acumulativas no se conectan con la línea base en el extremo superior de la distribución, y se conectan en el extremo inferior, en el límite real superior de un intervalo adicional añadido al extremo inferior de la

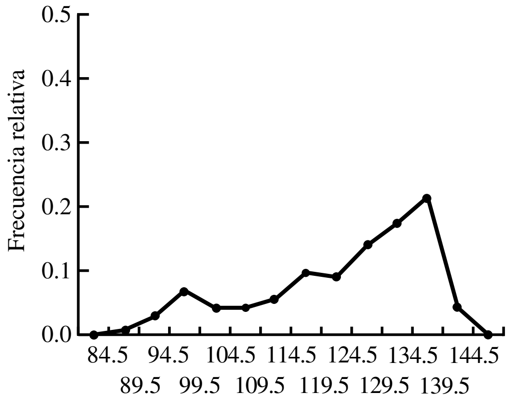


FIGURA 2.4: Polígono de frecuencias relativas de las mediciones de presión sanguínea en 12 categorías.

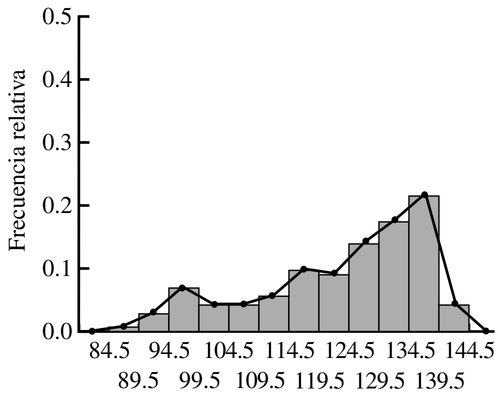


FIGURA 2.5: Polígono de frecuencias relativas superpuesto al histograma de las mediciones de presión sanguínea en 12 categorías.

distribución. Esto se puede ver en la figura 2.6. Los polígonos son particularmente convenientes cuando uno desea comparar dos o más distribuciones, por ejemplo, al incluir las presiones sanguíneas de hombres y mujeres en la misma gráfica. (Véase también la figura 2.11 en la página 33).

2.5.4 Gráficas de tallo y hojas

La gráfica de tallo y hojas es una herramienta gráfica que está relacionada con el histograma de frecuencias y algunas veces se emplea en lugar de éste. Su principal ventaja sobre el histograma es que conserva los valores de la variable mostrada. La figura 2.7 en la página 24 muestra una gráfica de tallo y hojas para los datos de la tabla 2.3. La similitud de esta figura con un histograma es evidente. Los siguientes pasos describen su construcción.

1. Divida cada observación en un componente de "tallo" y "hoja" como se describe a continuación.

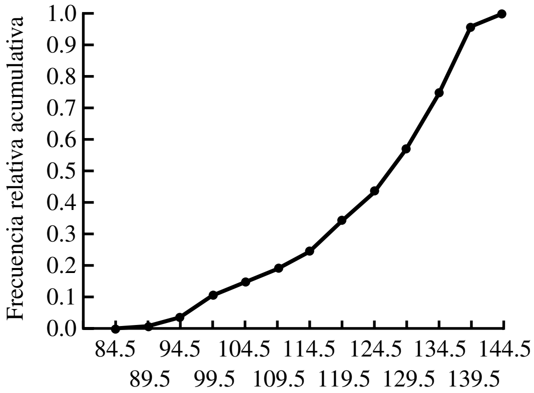


FIGURA 2.6: Polígono de frecuencias relativas acumulativas de las mediciones de la presión sanguínea en 12 categorías.

2. Elabore una lista de los componentes de tallo del valor más pequeño al más alto, como se haría en el eje x de un histograma.

3. Coloque los componentes de hoja asociados con cada tallo encima del tallo en orden ascendente.

El **tallo** de un número se define como todos los dígitos en un número excepto por el que está en el extremo derecho. La **hoja** es entonces el dígito que está en el extremo derecho. De esta manera, el tallo del valor de la presión sanguínea 86 es 8 y la hoja es 6. Asimismo, el tallo de 113 es 11, con una hoja de 3. Los tallos para las mediciones de la presión sanguínea aparecen horizontalmente bajo la línea en la figura 2.7. Las hojas que están asociadas con cada tallo se colocan arriba del tallo en orden ascendente. El efecto es similar a un histograma de frecuencias, pero los valores individuales de la presión sanguínea aún pueden recuperarse a partir del gráfico.

Es pertinente hacer algunas observaciones antes de dejar este tema. Primero, las gráficas de tallo y hojas son más efectivas con conjuntos de datos relativamente pequeños. Las 144 observaciones representadas en la figura 2.7 están probablemente cerca de un máximo para este tipo de gráfica. Segundo, en general, las gráficas de este tipo no se utilizan para mensajes difundidos masivamente, como la publicación de informes de investigación. Más bien, los investigadores los emplean de manera informal para comprender sus datos. Tercero, las gráficas de tallo y hojas pueden ser más complejas que la mostrada aquí. Por ejemplo, quizá las hojas consistan en dos o más dígitos y los tallos podrían estar agrupados en forma similar a los histogramas agrupados. Por último, debe señalarse que los tallos por lo general se colocan verticalmente, con las hojas formando renglones. Esta norma no se siguió aquí para enfatizar la semejanza del gráfico con el histograma.

2.6 MÉTODOS NUMÉRICOS

Los métodos gráficos y de distribución son herramientas excelentes para dar una descripción general de un conjunto de datos. Sin embargo, a menudo es deseable describir algunas características numéricas específicas de un conjunto de datos. Tal vez la medida más conocida de este tipo es lo que comúnmente se denomina el "promedio", o dicho con más precisión, la *media aritmética* de un conjunto de datos. En esta sección examinaremos cuatro categorías distintas de

8	6
9	0 2 2 3 5 5 5 6 6 7 7 8 8 9
10	0 0 0 0 2 3 5 5 6 7 7 9
11	0 0 0 1 1 1 4 4 5 5 5 5 5 5 6 6 6 7 8 8 9 9
12	0 0 0 1 2 2 2 3 3 4 4 4 4 5 5 5 5 6 6 6 6 6 6 6 7 7 7 8 8 8 9 9 9
13	0 0 0 0 0 1 1 1 2 2 2 2 3 3 3 3 3 3 3 3 4 4 4 4 4 5 5 5 5 5 5 5 5 6 6 6 7 7 7 7 7 7 7 7 7 7 7 8 8 8 9 9 9 9 9 9
14	0 0 0 0 3 3

FIGURA 2.7: Gráfica de tallo y hojas de las mediciones de la presión sanguínea.

tales medidas que son las medidas de tendencia central, las medidas de variabilidad, las medidas de posición relativa y las medidas asociadas con la forma de distribución.

2.6.1 Medidas de tendencia central

Las medidas de tendencia central ofrecen información acerca de valores típicos o promedio de un conjunto de datos. Existen varias medidas de este tipo, pero nosotros únicamente consideraremos la media, mediana y moda puesto que son las más usadas.

Media aritmética. La media aritmética es la medida de tendencia central más conocida y es a lo que mucha gente se refiere como el "promedio". Se calcula sumando todas las observaciones en el conjunto de datos y dividiendo esta suma entre el número de observaciones. Se agrega el adjetivo "aritmética" para distinguirla de otras medias menos conocidas. Nosotros prescindiremos de este término en el libro en tanto que es la única forma de media que se estudiará aquí.

Más formalmente, la **media de una muestra** se define como[4]

$$\boxed{\bar{x} = \frac{\sum x}{n}} \tag{2.1}$$

mientras que su equivalente para la población está dada por

$$\boxed{\mu = \frac{\sum x}{N}} \tag{2.1}$$

La notación para estas expresiones es similar a lo que se explicó en el apartado 2.3. Las formas de los estadísticos y de los parámetros difieren únicamente en el símbolo que se utiliza en la parte izquierda de la ecuación, y el uso de n minúscula y mayúscula, que es una norma común para denotar el número de observaciones en una muestra y en una población, respectivamente.

EJEMPLO 2.1

Calcule la media de los números 3, 5, 4, 8, 7.

Solución

$$\bar{x} = \frac{3+5+4+8+7}{5} = 5.4$$

■

Entre las muchas propiedades de la media se encuentran las siguientes.

1. Se define de modo inequívoco en tanto que su método de cálculo es reconocido en forma general.
2. Es única, ya que un conjunto de datos tiene una y sólo una media.
3. Su valor está influido por todas las observaciones en el conjunto de datos.

---

[4]Como se indicó anteriormente, se omitirán los subíndices de las fórmulas, excepto en situaciones potencialmente ambiguas.

Mediana. A diferencia de lo que sucede con la media aritmética, hay varias formas de definir y calcular la mediana. La definición más común sostiene que la **mediana** es el valor que divide a un conjunto de datos en dos partes iguales, de manera que el número de valores mayores que o iguales a la mediana es el mismo que el número de valores menores que o iguales a la mediana.

La forma más común de calcular la mediana cuando el número de valores es impar es ordenar las observaciones en términos de su magnitud y luego elegir el valor intermedio como la mediana. Una manera más formal de expresar lo anterior está dado por[5]

$$\boxed{\text{Mediana } (n \text{ impar}) = x_{\frac{n+1}{2}}} \tag{2.3}$$

donde n es el número de observaciones y $\frac{n+1}{2}$ es el subíndice de x.

Cuando el número de observaciones en un conjunto de datos es par, no hay un valor medio a elegir como la mediana. En este caso, la mediana se calcula como la media de los dos valores intermedios. Una manera más formal de expresar esto está dado por

$$\boxed{\text{Mediana } (n \text{ par}) = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}} \tag{2.4}$$

donde $\frac{n}{2}$ y $\frac{n}{2}+1$ son los subíndices para identificar los dos valores intermedios.

EJEMPLO 2.2

Calcule la mediana de los números 3, 5, 4, 8, 7, 0, 12.

Solución Primero se ordenan los valores de acuerdo con su magnitud como

$$\overbrace{0\ \ 3\ \ 4}^{3 \text{ valores}}\quad \overbrace{3}^{\text{mediana}}\quad \overbrace{7\ \ 8\ \ 12}^{3 \text{ valores}}$$

■

Puesto que el número de observaciones es impar, la mediana será el valor intermedio, que es cinco. Note que hay cuatro números que son mayores que o iguales a cinco y cuatro números que son menores que o iguales a cinco, por lo que se satisface la definición de mediana enunciada arriba.

La aplicación de la fórmula 2.3 con n igual a 7 produce $x_4 = 5$, que es el mismo valor obtenido arriba por inspección.

EJEMPLO 2.3

Calcule la mediana de los números 14, 8, 3, −1, 0, 12, 12 y 11.

Solución Si ordenamos los datos y notamos que el número de observaciones es impar, promediamos los dos valores intermedios (esto es, x_4 y x_5) y obtenemos $\frac{8+11}{2} = 9.5$

$$1\ 0\ 3\quad \overbrace{8\ 11}^{\substack{\text{valores}\\ \text{intermedios}}}\quad 12\ 12\ 14$$

■

[5]Nosotros no marcaremos la diferencia entre el estadístico y el parámetro de la mediana y la moda debido a que no se emplearán en un contexto inferencial en este libro. Se seguirá la misma política para los estadísticos en secciones posteriores.

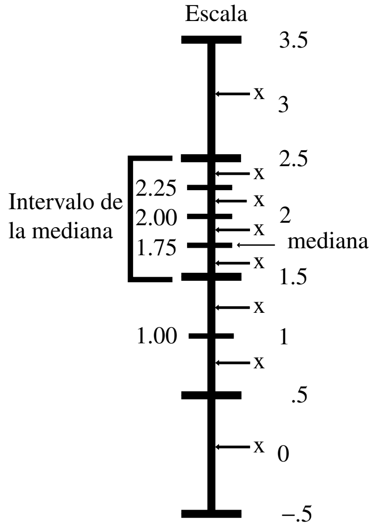


FIGURA 2.8: Conceptualización de la mediana de una variable continua.

EJEMPLO 2.4

Calcule la mediana de los números 3, 2, 0, 2, 1, 1, 2, 2.

Solución Nuevamente, notando que el número de observaciones es par, promediamos los valores intermedios (esto es, x_4 y x_5) para obtener $\frac{2+2}{2} = 2$.

$$0\ 1\ 1\ \overbrace{2\ 2}^{\text{valores intermedios}}\ 2\ 2\ 3$$

Note que este valor no satisface la definición de mediana enunciada arriba, puesto que hay siete valores menores que o iguales a dos, y cinco que son mayores que o iguales a dos. Claramente, el dos no divide el conjunto de datos en dos partes iguales. Esta dificultad a menudo se origina cuando hay observaciones empatadas cerca de la mitad de la distribución. ■

Una definición de **mediana** menos conocida pero potencialmente más útil plantea que hay un punto en la escala de medición ubicado de tal manera que deja a la mitad de las observaciones por arriba de él y a la otra mitad por debajo.[6] Esta definición no da un valor único para la mediana, pero cuando se combina con el método de cálculo descrito a continuación sí arroja tal resultado.

La escala conceptualizada que se muestra en la figura 2.8 resulta útil para demostrar esta definición de la mediana, así como su cálculo. Usted recordará del apartado 2.5.2 que las puntuaciones obtenidas de las mediciones de variables continuas se consideran valores indicativos que se sitúan dentro de intervalos específicos de la característica subyacente, más que en puntos específicos. De tal manera, si tenemos intervalos de longitud 1, una puntuación de 0 indica un valor de la característica que se sitúa entre –.5 y .5. Estos valores se conocen, respectivamente, como los límites reales inferior y superior de un intervalo. Una puntuación de 1 indica un valor entre el límite real inferior de .5 y el límite real superior de 1.5, y así sucesivamente. Así pues, el límite real superior de un intervalo es también el límite real inferior del siguiente valor más grande. La

[6] Véase la explicación sobre los percentiles en la página 38 para una definición de la mediana como percentil.

figura 2.8 muestra estos intervalos para los datos en el último ejemplo. Para los actuales propósitos se supone que las observaciones empatadas están uniformemente distribuidas a lo largo del intervalo que representan.[7] De ese modo, puesto que hay cuatro 2 en el conjunto de datos, se supone que el intervalo de 1.5 a 2.5 puede dividirse en cuatro partes iguales, donde los cuatro 2 están distribuidos de igual forma a lo largo de estos subintervalos, como se muestra en la figura. Asimismo, como hay dos 1, el intervalo de .5 a 1.5 se divide en dos subintervalos, donde cada una de las dos puntuaciones ocupa cada subintervalo. Las ubicaciones específicas de los valores, marcadas como "X" en la figura, son desconocidas. Se supone que sólo se conoce el intervalo o subintervalo dentro del cual caen los valores.

El problema es encontrar un punto en esta escala debajo del cual y sobre el cual caiga la mitad de las observaciones. Puesto que hay un total de 8 observaciones, la mitad de las observaciones sería (.5)(8) = 4. Existe una observación menor que .5, 3 observaciones menores que 1.5, y 7 menores que 2.5. La mediana, por lo tanto, debe caer en el intervalo de 1.5 a 2.5. A este intervalo se le conoce como el **intervalo de la mediana**. Como hay tres observaciones que son menores que el límite real inferior del intervalo de la mediana, esta última debe caer en un punto en el intervalo de la mediana que sea mayor que una de las cuatro observaciones en el intervalo. Como se aprecia en la figura, este punto estaría en 1.75. Note que dada la suposición de una dispersión similar a lo largo del intervalo, hay cuatro observaciones por debajo y por arriba de 1.75, por lo que se satisface la definición enunciada antes. Este método para calcular la mediana puede formalizarse al ordenar los datos en una distribución de frecuencias acumulativas y aplicando

$$\boxed{\text{Mediana} = LRL + (w)\,\frac{(.5)(n)\quad cf}{f}} \tag{2.5}$$

donde LRL es el límite real inferior (por las siglas de *lower real limit*) del intervalo de la mediana, *w* es la anchura del intervalo de la mediana, calculado como la diferencia entre los límites reales superior e inferior del intervalo, *n* es el número total de observaciones, *cf* es la frecuencia acumulativa (*cumulative frequency*) *hasta* el intervalo de la mediana y *f* es la frecuencia del intervalo de la mediana.

Aplicando la ecuación 2.5 al problema se produce

$$1.5 + (1.0)\,\frac{(.5)(8)\quad 3}{4} = 1.75$$

que es el resultado obtenido por inspección.

Es conveniente señalar dos excepciones al uso de la fórmula 2.5. La primera de ellas se conceptualiza en la figura 2.9. En este caso hay ocho observaciones, cuatro de las cuales están por debajo del límite superior de 1.5 y las otras cuatro están por arriba de ese punto. Entonces no hay intervalo de mediana debido a que el punto que divide (de manera inequívoca) los datos en dos partes iguales se ubica en el límite superior (o inferior) de un intervalo. En situaciones de este tipo la mediana se toma como el valor en el límite real, como se muestra en la figura.

La segunda excepción está representada en la figura 2.10. En este caso hay cuatro observaciones, dos de las cuales caen por debajo y las otras dos caen por arriba del límite real superior de .5. El problema es que esta afirmación es cierta para cualquier punto entre .5 y 2.5 en la escala. Esto resulta de las frecuencias cero en intervalos, cerca del centro de la distribución. En este caso

[7] Esto se conoce algunas veces como la suposición de isomorfismo.

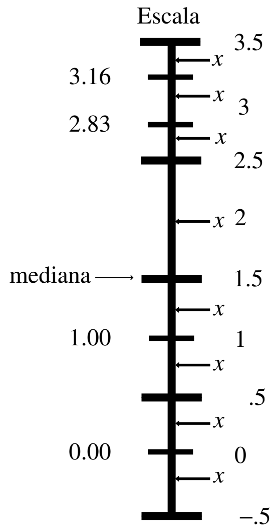


FIGURA 2.9: Conceptualización de la mediana de una variable continua sin intervalo de mediana.

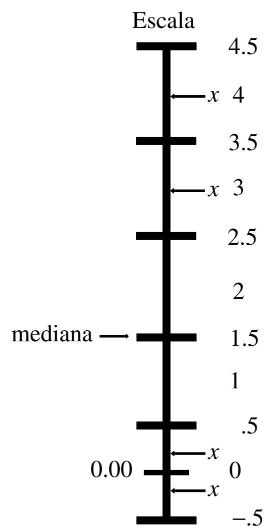


FIGURA 2.10: Conceptualización de la mediana de una variable continua con frecuencias cero en algunos intervalos clave.

el punto medio del intervalo (o intervalos) se toma como la mediana, lo que da por resultado una mediana de 1.5.

EJEMPLO 2.5

Utilice el método presentado aquí para calcular las medianas de los valores en los ejemplos 2.2 y 2.3.

Solución Aplicando la fórmula 2.5 al primer ejemplo se obtiene

$$4.5 + (1.0)\frac{(.5)(7) - 3}{1} = 5.0$$ ■

que es la respuesta obtenida antes.

Para el segundo ejemplo, notamos que hay cuatro observaciones por debajo de 8.5 y cuatro observaciones por arriba de 10.5. Como cualquier punto en el intervalo de 8.5 a 10.5 satisface la definición, el punto medio del intervalo, que es 9.5, se toma como la mediana. Éste es el resultado obtenido previamente.

EJEMPLO 2.6

Calcule la mediana de las puntuaciones de la tabla 2.3 en la página 18.

Solución Ordenando los datos en una distribución de frecuencias simple y luego aplicando la ecuación 2.5 obtenemos

$$125.5 + (1.0)\frac{(.5)(144) - 66}{7} = 126.36$$ ■

EJEMPLO 2.7

Encuentre la mediana de los siguientes datos.

Puntuación	Frecuencia	Frecuencia acumulativa
2.4	3	145
2.3	40	142
2.2	36	102
2.1	21	66
2.0	18	45
1.9	8	27
1.8	14	19
1.7	5	5

Solución Al observar que $(.5)(n) = 72.5$ y que los límites reales inferior y superior del intervalo 2.2 tienen frecuencias acumulativas de 66 y 102 respectivamente, identificamos que el intervalo de la mediana va de 2.15 a 2.25. Aplicando la fórmula 2.5 obtenemos

$$2.15 + (.10)\frac{(.5)(145) - 66}{36} = 2.17$$ ■

Para resumir, cuando se calcula la mediana por el método aquí descrito, tres diferentes escenarios son posibles. 1) Cuando el intervalo de la mediana es identificado, se aplica la fórmula 2.5. 2) Cuando la mitad de las observaciones caen por debajo y la mitad por arriba del límite real, y el intervalo que está por arriba del límite tiene una frecuencia diferente de cero, el límite real se toma como la mediana. 3) Cuando la mitad de las observaciones caen por debajo y la mitad por arriba del límite real, y el intervalo que está por arriba del límite tiene una frecuencia de cero, el punto medio del intervalo (o intervalos) de la frecuencia cero se toma como la mediana.

Entre las muchas propiedades de la mediana están las siguientes:

1. Se puede definir y calcular en diversas formas.
2. Es única de acuerdo con una definición específica y una forma de cálculo, ya que un conjunto de datos tiene una y solo una mediana.
3. Es insensible a observaciones extremas.

Moda. La **moda** de un conjunto de datos es la puntuación o puntuaciones que ocurren con mayor frecuencia. Si todos los puntos en un conjunto ocurren con la misma frecuencia, no hay moda. Por otro lado, si dos o más puntuaciones ocurren con igual frecuencia y esa frecuencia es mayor que la de las otras puntuaciones en el conjunto, entonces habrá más de una moda.

En el caso de datos de nivel nominal u ordinal se puede calcular la categoría modal. La categoría modal es aquella que tiene mayor frecuencia. Si dos o más categorías tienen la misma frecuencia y ésta es mayor que la de todas las demás categorías, entonces hay más de una categoría modal.

EJEMPLO 2.8

Calcule la moda de los datos que se muestran en la tabla 2.3 en la página 18.

Solución Al revisar la tabla 2.3 vemos que el valor de presión sanguínea de 137 ocurre 11 veces en el conjunto de datos, más que cualquier otro valor. La moda es entonces 137.

EJEMPLO 2.9

Calcule la(s) moda(s) de los números 7, 8, 9, 7, 7, 4, 9, 5, 9, 3, 1, 9, 7 y 8.

Solución Las modas pueden identificarse fácilmente una vez que los datos se han ordenado en una distribución de frecuencias, como se muestra más adelante. Puesto que tanto el 9 como el 7 tienen una frecuencia de 4, que es mayor que la frecuencia de cualquier otro valor, 9 y 7 son las modas de los datos.

Puntuación	Frecuencia
1	1
2	0
3	1
4	1
5	1
6	0
7	4
8	2
9	4

■

EJEMPLO 2.10

Calcule la categoría modal de los datos descritos en la tabla 2.2 en la página 16.

Solución La categoría modal es “ninguno” porque tiene una frecuencia de 31, que es mayor que la frecuencia de cualquier otra categoría. ■

Comparación de las propiedades de la media, mediana y moda.

1. La media y la mediana son medidas que localizan la “mitad” de una distribución en cierto sentido. Esto no necesariamente es cierto para la moda.
2. La moda puede ser inestable en muestras pequeñas. Por ejemplo, la moda de los números 1, 1, 1, 2, 3, 4, 5 y 5 es 1. Pero si uno de los 1 se cambia por 0, el conjunto se vuelve bimodal. Por otra parte, si uno de los 1 se cambia por 5 la moda será 5.
3. La mediana *no* es afectada por el *tamaño* de las puntuaciones en los extremos superior e inferior de la distribución para la que se calcula. Por ejemplo, la mediana de 1, 2, 3, 4 y 5 es 3. Si el 5 se cambia por 490, la mediana aún es 3.
4. La media *es* afectada por el tamaño de cada valor en un conjunto de datos. Por ejemplo, la media de un conjunto original de números dado anteriormente fue 3. Después de cambiar el 5 por 490, la media fue 100.
5. La media juega un papel importante en la estadística inferencial, mientras que la mediana juega un papel insignificante y la moda prácticamente ninguno.

2.6.2 Mediciones de variabilidad

La figura 2.11 muestra las distribuciones de frecuencias de dos conjuntos de datos que tienen una media, una mediana y una moda igual a cuatro. A pesar de sus medidas de tendencia central similares, estos conjuntos de datos difieren en un aspecto importante. Las puntuaciones de la distribución A están menos dispersas o separadas que las de la distribución B. En otras palabras, los puntos de A son más homogéneos que los de B. Como usted aprenderá en capítulos posteriores, es importante poder cuantificar el grado de dispersión en un conjunto de datos. Las medidas de este tipo se conocen como medidas de variabilidad o dispersión.

Como sucede con las medidas de tendencia central, hay muchas medidas de variabilidad. En este apartado usted conocerá cuatro de ellas: el rango, la desviación media, la varianza y la desviación estándar.[8] En tanto que todas ellas son útiles en ciertas circunstancias, las últimas dos son las más importantes y formarán la base de muchos métodos que usted estudiará en capítulos posteriores.

Rango. El rango está en función únicamente de las puntuaciones más grande y más pequeña de un conjunto de datos. A menudo se identifican dos formas de rango. El **rango exclusivo** se define como la diferencia entre las puntuaciones más grande y más pequeña de un conjunto de datos, o más formalmente

$$\boxed{\text{Rango (exclusivo)} = x_L - x_S} \tag{2.6}$$

donde x_L y x_S son las puntuaciones más grande y más pequeña del conjunto de datos, respectivamente.

[8] Una medida adicional de variabilidad, el rango semi-intercuartilar, se analiza en relación con los percentiles en el apartado 2.6.3.

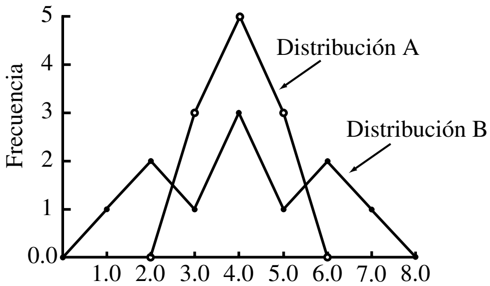


FIGURA 2.11: Dos distribuciones con media, mediana y moda en común.

EJEMPLO 2.11

Calcule los rangos exclusivos de los datos en la figura 2.11

Solución El rango exclusivo para la distribución A es 5 – 3 = 2, mientras que para la distribución B es 7 – 1 = 6. ■

El **rango inclusivo** toma en cuenta los límites reales superior e inferior (véase el apartado 2.5.2 en la página 20 y la explicación al inicio de la página 27) de la puntuación más alta y de la puntuación más baja, y se expresa como

$$\boxed{\text{Rango (inclusivo)} = URL_L - LRL_S} \tag{2.7}$$

donde URL_L y LRL_S son el límite real superior (*upper real limit*) de la puntuación más grande y el límite real inferior (*lower real limit*) de la puntuación más pequeña, respectivamente, en un conjunto de datos.

EJEMPLO 2.12

Calcule los rangos inclusivos de los datos en la figura 2.11.

Solución El rango inclusivo para la distribución A es 5.5 − 2.5 = 3, mientras que para la distribución B es 7.5 − .5 = 7. ■

El rango es una medida de variabilidad inestable por el hecho de que se basa en dos valores únicamente. Un cambio en una de estas puntuaciones puede afectar drásticamente el rango.

Desviación media. Al igual que el rango, la desviación media es una medida de variabilidad altamente intuitiva. Sin embargo, a diferencia del rango, la desviación media toma en cuenta todos los datos para el cálculo de la variabilidad, por lo que se trata de un estadístico más estable.

Diversas medidas de variabilidad están basadas en las diferencias entre los valores de una distribución y algún punto central en la distribución. Por ejemplo, suponga que se calcula la diferencia $x - \bar{x}$ para cada puntuación de un conjunto de datos. Este valor, llamado **puntuación de desviación** o simplemente una **desviación**,[9] indica el número de unidades entre la puntuación y la media. Cuando los datos están "agrupados" de forma estrecha alrededor de la media, las desviaciones tienden a ser pequeñas. Para datos que están más dispersos, las desviaciones son más grandes. Es posible que una representación razonable de la variabilidad se base en el promedio

[9] Las desviaciones también pueden tomarse a partir de la mediana o de otros puntos de una distribución, pero usaremos el término para referirnos a las desviaciones con respecto a la media.

TABLA 2.6: Cálculo de la desviación media y la varianza de la distribución A.

	(1) x	(2) x^2	(3) $x-\bar{x}$	(4) $\lvert x = \bar{x}\rvert$	(5) $\lvert x = \bar{x}\rvert^2$
	3	9	−1	1	1
	3	9	−1	1	1
	3	9	−1	1	1
	4	16	0	0	0
	4	16	0	0	0
	4	16	0	0	0
	4	16	0	0	0
	4	16	0	0	0
	5	25	1	1	1
	5	25	1	1	1
	5	25	1	1	1
Σ	44	182	0	6	6

de estas desviaciones. Cuando los datos se encuentran más esparcidos, el promedio de las desviaciones es mayor que para los datos con menor dispersión. La dificultad del uso de las desviaciones de esta forma es que siempre suman cero. (Véase el apartado 2.3.2 en la página 13 para más detalles). Este problema puede ser eliminado tomando los valores absolutos de las desviaciones. Esto, entonces, es el fundamento de la desviación media. La **desviación media** (MD, por las siglas de *mean deviation*) es el promedio de los valores absolutos de las desviaciones en un conjunto de puntuaciones. La expresión formal es

$$\boxed{MD = \frac{\sum |x - \bar{x}|}{n}} \tag{2.8}$$

EJEMPLO 2.13

Calcule la desviación media de los datos descritos en la figura 2.11

Solución La columna (1) de la tabla 2.6 muestra las puntuaciones que forman la distribución A en la figura 2.11. Las columnas (3) y (4) de esta tabla muestran, respectivamente, las puntuaciones de desviación y los valores absolutos de las puntuaciones de desviación para los datos de la columna (1). Utilizando la suma de la columna (4) obtenemos

$$MD = \frac{6}{11} = .55$$

■

La tabla 2.7 presenta los resultados para la distribución B. Utilizando la suma de la columna (4) de esta tabla obtenemos

$$MD = \frac{16}{11} = 1.45$$

De esta manera, el promedio de la desviación de las puntuaciones de la distribución A fue de .55 unidades, mientras que en la distribución B fue de 1.45 unidades; por lo tanto, se confirma que la distribución B tiene mayor variabilidad que la distribución A.

TABLA 2.7: Cálculo de la desviación media y la varianza de la distribución B.

	(1) x	(2) x^2	(3) $x - \bar{x}$	(4) $\lvert x = \bar{x} \rvert$	(5) $\lvert x = \bar{x} \rvert^2$
	1	1	–3	3	9
	2	4	–2	2	4
	2	4	–2	2	4
	3	9	–1	1	1
	4	16	0	0	0
	4	16	0	0	0
	4	16	0	0	0
	5	25	1	1	1
	6	36	2	2	4
	6	36	2	2	4
	7	49	3	3	9
Σ	44	212	0	16	36

Varianza. La varianza es una medida de variabilidad menos intuitiva pero generalmente más útil que el rango o la desviación media. Como estadístico descriptivo, la varianza es menos interesante que la desviación media, pero en general resulta más útil en virtud de su papel en la inferencia, como se verá en capítulos posteriores. Igual que la desviación media, la varianza utiliza desviaciones como base, pero las eleva al cuadrado en lugar de utilizar los valores absolutos. El parámetro está dado por

$$\boxed{\sigma^2 = \frac{\sum (x - \mu)^2}{N}} \tag{2.9}$$

De esta manera, el parámetro de la **varianza** es el promedio del *cuadrado* de las desviaciones de las puntuaciones que conforman la población. El estadístico es

$$\boxed{s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}} \tag{2.10}$$

Usted notará que el divisor para el estadístico es $n - 1$, mientras que para el parámetro es N. Esto se deriva del hecho de que s^2 se utiliza comúnmente como un estimado de σ^2 en escenarios inferenciales, como se verá más adelante. Se puede ver que si el divisor de s^2 fuera n, la estimación resultante estaría sesgada. Esto es, en promedio el valor de s^2 sería más pequeño que σ^2. Dividiendo entre $n - 1$ se elimina este sesgo, haciendo que s^2 sea un mejor estimado de σ^2. Esto es más bien teórico y está más allá del alcance de este libro, por lo que no necesitaremos abundar más al respecto.

EJEMPLO 2.14

Calcule la varianza de los datos presentados en la figura 2.11 de la página 33.

Solución La columna (5) de la tabla 2.6 muestra el cuadrado de las desviaciones de los datos de la distribución A. Utilizando la suma de esta columna obtenemos

$$s^2 = \frac{6}{10} = .60$$

Al utilizar la suma de la misma columna en la tabla 2.7 obtenemos

$$s^2 = \frac{36}{10} = 3.60$$

para la distribución B. ■

Las ecuaciones presentadas arriba para la varianza se llaman **ecuaciones conceptuales** porque reflejan el concepto de varianza. Esto es, transmiten la idea de que la varianza está basada en desviaciones al cuadrado. Algunas veces es más conveniente emplear ecuaciones computacionales. Las **ecuaciones computacionales** son útiles para calcular la varianza, pero no dan a conocer el concepto. Podemos utilizar las reglas dadas en el apartado 2.3.2 para encontrar una ecuación computacional comúnmente utilizada para la varianza. Utilizando el numerador de la fórmula 2.10, que se denomina **suma de cuadrados**, notamos que

$$\begin{aligned}
\sum (x-\bar{x})^2 &= \sum \left(x^2 - 2x\bar{x} + \bar{x}^2\right) \\
&= \overbrace{\sum x^2 - \sum 2x\bar{x} + \sum \bar{x}^2}^{\text{regla 4}} \\
&= \sum x^2 - \overbrace{2\bar{x}\sum x}^{\text{regla 2}} + \overbrace{n\bar{x}^2}^{\text{regla 1}} \\
&= \sum x^2 - 2\frac{\left(\sum x\right)^2}{n} + \frac{\left(\sum x\right)^2}{n} \\
&= \sum x^2 - \frac{\left(\sum x\right)^2}{n}
\end{aligned}$$

Esta forma de la suma de cuadrados, así como su contraparte poblacional, pueden dividirse entre $n-1$ o N para producir las siguientes ecuaciones computacionales para la varianza.

$$\boxed{\sigma^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}} \tag{2.11}$$

$$\boxed{s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} \tag{2.12}$$

EJEMPLO 2.15

Utilice la ecuación 2.12 para calcular la varianza de los datos mostrados en la figura 2.11.

Solución Utilizando las sumas de las columnas (1) y (2) de la tabla 2.6 se obtiene

$$s^2 = \frac{212 - \frac{(44)^2}{11}}{10} = .60$$

que es el mismo resultado obtenido con la ecuación conceptual. Asimismo, tomando las mismas sumas de la tabla 2.7 obtenemos

$$s^2 = \frac{212 - \frac{(44)^2}{11}}{10} = 3.60$$

que es nuevamente el mismo valor obtenido por medio de la ecuación conceptual. ■

Tal vez la varianza no sea muy atractiva como estadístico puramente descriptivo, ya que se expresa en unidades al cuadrado, pero como veremos, es uno de los conceptos más importantes en estadística.

Desviación estándar. La **desviación estándar** se define como la raíz cuadrada de la varianza. De los resultados anteriores se deduce que el parámetro está dado por

$$\boxed{\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}} \tag{2.13}$$

y

$$\boxed{\sigma = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N}}} \tag{2.14}$$

El estadístico es

$$\boxed{s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}} \tag{2.15}$$

y

$$\boxed{s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}} \tag{2.16}$$

EJEMPLO 2.16

Calcule la desviación estándar de los datos mostrados en la figura 2.11.

Solución La desviación estándar de la distribución A es $\sqrt{.60} = .77$ y la de la distribución B es $\sqrt{3.60} = 1.90$. ■

La varianza está expresada en términos del *cuadrado* de las unidades originales de medición. De esta manera, si las mediciones están en gramos, la varianza se expresa en gramos al cuadrado. En contraste, la desviación estándar calcula el valor de la variabilidad en términos de las unidades originales, gramos en este caso. Tendremos la oportunidad de utilizar más la varianza y la desviación estándar en los capítulos siguientes.

2.6.3 Medidas de posición relativa

En este apartado usted aprenderá sobre los métodos que localizan las posiciones relativas de las observaciones en una distribución. Por ejemplo, tal vez desee encontrar un punto en la escala de medición debajo del cual se localiza el 25% de las observaciones de la distribución. Los puntos caracterizados en esta forma se llaman percentiles. Por otra parte, quizás desee calcular el porcentaje de observaciones que se localizan por debajo de un punto específico en la escala. Estos porcentajes se llaman rangos percentiles. Por último, tal vez desee determinar la posición de una observación en relación con la media de una distribución por medio de las llamadas puntuaciones z. Así que ahora examinaremos los percentiles, los rangos percentiles y las puntuaciones z.

Percentiles. Para calcular la mediana usted aprendió a encontrar un punto en la escala de medición debajo del cual se localiza la mitad (o 50%) de las observaciones. Este punto es la mediana. Este concepto puede ampliarse para encontrar un punto en la escala debajo del cual se localiza un porcentaje arbitrario de observaciones. Por ejemplo, tal vez usted desee encontrar el punto debajo del cual cae el 25 o el 75% de las observaciones. A estos puntos se les llama percentiles 25 y 75, respectivamente. Un **percentil** es un punto en la escala de medición debajo del cual se localiza un porcentaje específico de las observaciones. Con base en esta definición, la **mediana** se define como el percentil 50.

Dado que la mediana es sólo un ejemplo de un percentil, a usted no le sorprenderá saber que el método que aprendió para el cálculo de la mediana puede generalizarse para obtener cualquier percentil. Éste es un buen momento para que usted repase los conceptos relacionados con la mediana presentados en la página 26, tal como se aplican directamente al problema en cuestión. Una pequeña modificación de la ecuación 2.5 produce

$$\boxed{P_p = LRL + (w)\left[\frac{(pr)(n) - cf}{f}\right]} \tag{2.17}$$

donde P_p representa el *p-ésimo* percentil, *LRL* es el límite real inferior del intervalo que contiene el *p-ésimo* percentil, w es la anchura del intervalo, calculada como la diferencia entre los límites reales superior e inferior de ese intervalo, pr es p expresada como una proporción (esto es, $p/100$), n es el número total de observaciones, cf es la frecuencia acumulativa *hasta* el intervalo del percentil y f es la frecuencia de ese intervalo.

Note que la única diferencia entre las ecuaciones 2.5 y 2.17 es la sustitución de pr por la constante .5. Esto permite la flexibilidad para calcular cualquier percentil más que restringir el cálculo al percentil 50% (mediana).

EJEMPLO 2.17

Calcule los percentiles 25, 60 y 75 de los datos en la tabla 2.3 de la página 18.

Solución Mediante la construcción de una distribución de frecuencias acumulativas y observando que $(.25)(144) = 36$, el intervalo del percentil 25 es 114.5 a 115.5. Esto se deduce del hecho de que la frecuencia acumulativa hasta el límite real inferior de 114.5 es 35 y para 115.5 es 41. El punto en la escala debajo del cual hay 36 observaciones debe estar entre estos dos límites. Utilizando esta información con la ecuación 2.17 se obtiene

$$P_{25} = 114.5 + (1.0)\left[\frac{(.25)(144) - 35}{6}\right] = 114.67$$

Dado que (.6)(144) = 86.4, y que las frecuencias acumulativas hasta 129.5 y 130.5 son 82 y 87 respectivamente, P_{60} debe localizarse en este intervalo. Nuevamente, aplicando la ecuación 2.17 se obtiene

$$P_{60} = 129.5 + (1.0)\left[\frac{(.6)(144) - 82}{5}\right] = 130.38$$

Utilizando el mismo método, P_{75} se obtiene de la siguiente manera

$$P_{75} = 134.5 + (1.0)\left[\frac{(.75)(144) - 107}{8}\right] = 134.63$$ ■

EJEMPLO 2.18

Calcule P_{15}, P_{40} y P_{69} de los siguientes datos.

Puntuación	Frecuencia	Frecuencia acumulativa
.7	22	80
.6	26	58
.5	0	32
.4	0	32
.3	20	32
.2	7	12
.1	4	5
.0	1	1

Solución P_{15} es el punto en la escala de medición debajo del cual se localiza el 15% o (.15)(80) = 12 de las observaciones. Puesto que la frecuencia acumulativa del límite real superior de .25 es 12, .25 es el percentil 15. (Repase la explicación sobre la mediana en la página 26 si usted no sigue esta lógica.)

Dado que (.4)(80) = 32 de las observaciones se localizan debajo de cualquier punto entre .35 y .55, el punto medio de .45 se toma como el percentil 40. (Nuevamente, repase la discusión de la mediana si la respuesta se le escapa.)

Puesto que las frecuencias acumulativas hasta .55 y .65 son 32 y 58 respectivamente, el punto debajo del cual se localizan (.69)(80) = 55.2 observaciones debe estar en el intervalo de .55 a .65. Aplicando la ecuación 2.17 se obtiene

$$P_{69} = .55 + (.10)\left[\frac{(.69)(80) - 32}{26}\right] = .64.$$ ■

Los percentiles a menudo se utilizan para comparar la puntuación de un individuo con las de un grupo más grande. Por ejemplo, un pediatra puede estar preocupado porque el peso de un bebé está por debajo del quinto percentil en una tabla de crecimiento de bebés de la misma edad, ya que pocos bebés de esa edad tienen un peso tan bajo. Asimismo, las personas que obtienen puntuaciones por arriba de algún percentil específico en un examen podrían ser seleccionadas para recibir honores especiales. Observe que en ninguno de los casos se hacen comparaciones con algún estándar absoluto, sino con otras puntuaciones.

Los percentiles también son útiles para describir distribuciones. Mediante el uso de percentiles uno puede divulgar el hecho de que el 5% de las observaciones caen debajo de un punto dado

(P_5) o que el 50% de las observaciones se ubican entre dos puntos, P_{25} y P_{75}. Como los percentiles se utilizan tan comúnmente para este propósito, algunos de ellos tienen designaciones específicas. Por ejemplo, P_{10}, P_{20}, P_{30} … P_{90} se llaman **deciles** (del primero al noveno) porque dividen la distribución en 10 componentes con 10% de las observaciones en cada uno. Asimismo, P_{20}, P_{40}, P_{60} y P_{80} se llaman primero, segundo, tercero y cuarto **quintiles** (término derivado de cinco), mientras que P_{25}, P_{50} y P_{75} son el primero, segundo y tercer **cuartiles**. (Note que el quinto decil, el segundo cuartil y la mediana son términos diferentes que designan el mismo punto). En conjunto, los percentiles, deciles, quintiles y cuartiles se conocen como **cuantiles**.

Finalmente, algunos estadísticos descriptivos se basan en percentiles. Por ejemplo, una medida de variabilidad que se utiliza algunas veces, llamada **rango semiintercuartil** o **semiintercuartílico** (o Q) se calcula por

$$\boxed{Q = \left[\frac{P_{75} - P_{25}}{2}\right]} \tag{2.18}$$

Rangos de percentiles. Como usted sabe, un percentil es un punto en la escala de medición debajo del cual se localiza cierto porcentaje de las observaciones. En contraste, un **rango de percentil** (PR, por las siglas de *percentile rank*) es el porcentaje de las observaciones que cae debajo de un punto dado en la escala. De esta manera, los percentiles son puntos y los rangos de percentiles son porcentajes. Debido a la estrecha relación entre estos dos conceptos, a menudo se confunden o se utilizan de forma indistinta. Con base en las mismas suposiciones utilizadas para el cálculo de los percentiles, podemos usar la siguiente ecuación para obtener los rangos de percentiles.

$$\boxed{RP_P = \frac{100\left[\frac{f(P-LRL)}{w} + cf\right]}{n}} \tag{2.19}$$

donde P es el punto en la escala para el cual se calculará el rango de percentil, LRL es el límite real inferior del intervalo que contiene a P, w es la anchura del intervalo, calculado como la diferencia entre los límites real superior e inferior, n es el número total de observaciones, cf es la frecuencia acumulativa *hasta* el intervalo que contiene a P, y f es la frecuencia de ese intervalo.

EJEMPLO 2.19

Utilice los datos de la tabla 2.3 en la página 18 para calcular los rangos de percentiles de los puntos de escala 114.67, 130.38 y 134.63.

Solución Si observamos que 114.67 cae en el intervalo de 114.5 a 115.5 y utilizamos otra información de la tabla 2.3 en la ecuación 2.19 se obtiene

$$RP_{114.67} = \frac{100\left[\frac{6(114.67-114.50)}{1.0} + 35.0\right]}{144} = 25$$

Utilizando el mismo método para los últimos valores se produce

$$RP_{130.38} = \frac{100\left[\frac{5(130.38-129.50)}{1.0} + 82.0\right]}{144} = 60$$

y

$$RP_{134.63} = \frac{100\left[\frac{8(134.63-134.50)}{1.0} + 107.0\right]}{144} = 75 \qquad \blacksquare$$

Ninguno de estos resultados debe sorprenderlo ya que estos tres puntos se identificaron antes como los percentiles 25, 60 y 75.

Por desgracia, los analistas de datos generalmente no están interesados en encontrar el rango percentil de un *punto* en la escala de medición, sino que más bien quieren conocer el rango de percentil de una *puntuación* u *observación*. Esto presenta una dificultad ya que, como usted sabe, las puntuaciones no son puntos en la escala, sino que representan intervalos. ¿Cómo se calcula entonces el rango de percentil de una puntuación? En esencia, usted debe elegir un punto en la escala para representar la puntuación. Para ello se utilizan tres alternativas comunes.

1. El límite real inferior del intervalo de la puntuación.
2. El punto medio del intervalo de la puntuación.
3. El límite real superior del intervalo de la puntuación.

Cuando se usa el límite real inferior, la ecuación 2.19 se simplifica a

$$\boxed{RP_P = 100\left[\frac{cf}{n}\right]} \tag{2.20}$$

Para los métodos del punto medio y límite real superior las simplificaciones son

$$\boxed{RP_P = 100\left[\frac{(.5)(f) + cf}{n}\right]} \tag{2.21}$$

y

$$\boxed{RP_P = 100\left[\frac{f + cf}{n}\right]} \tag{2.22}$$

donde todos los términos se definieron previamente.

EJEMPLO 2.20

Utilice los métodos del límite real inferior, del punto medio y del límite real superior para calcular el rango de percentil de una *puntuación* de 134 en la distribución presentada en la tabla 2.3.

Solución Los métodos del límite real inferior, del punto medio y del límite real superior producen respectivamente

$$RP_{133.5} = 100\left[\frac{102}{144}\right] = 71$$

$$RP_{134} = 100\left[\frac{(.5)(5) + 102}{144}\right] = 73$$

y

$$RP_{134.5} = 100\left[\frac{5+102}{144}\right] = 74$$

Redondeamos los resultados porque generalmente no se reportan fracciones de los rangos de percentiles. ■

Antes de dejar este tema queremos enfatizar la aseveración previa de que existen varias definiciones y métodos de cálculo para los percentiles y rangos de percentiles. Los autores generalmente no aclaran esto, lo cual algunas veces provoca confusión. Esto también explica por qué varios programas de cómputo parecen producir resultados conflictivos. Nuestro objetivo aquí ha sido no sólo ofrecerle los métodos útiles para el manejo de estos estadísticos, sino también darle una base intuitiva que le permitan comprender los conceptos implicados.

Puntuaciones z. Los percentiles localizan puntos relativos a todas las observaciones en un conjunto de datos. Por ejemplo, el percentil 25 es el punto debajo del cual se localiza el 25% de los datos. En contraste, las puntuaciones z localizan puntos relativos a la media de los datos. Usted ya ha visto una forma de hacer esto. Las puntuaciones de desviación (véase la página 33), calculadas como $x-\bar{x}$ muestran qué tan lejos está una observación por arriba o por debajo de la media de los datos. Una puntuación de desviación de menos seis indica que el punto en cuestión se localiza a seis unidades por debajo de la media, mientras que una desviación de uno indica que el punto está una unidad por arriba de la media. Un problema que surge con las puntuaciones de desviación es que son dependientes de la escala. Suponga que las puntuaciones de desviación de la presión sanguínea y el peso de un paciente son 12 y 16 respectivamente, cuando se comparan con otros pacientes de la misma edad. Es difícil comparar estas dos puntuaciones porque están en escalas diferentes. Obviamente, el paciente está por arriba de la media de cada medida pero, en términos relativos, ¿qué tan lejos? Sería útil que la presión sanguínea y el peso estuvieran en una escala común.

Una **puntuación z** indica la distancia y dirección de un punto a partir de la media en términos de unidades estándar. De manera más precisa, una puntuación z de una muestra está dada por

$$\boxed{z = \frac{x-\bar{x}}{s}} \tag{2.23}$$

El equivalente poblacional es

$$\boxed{Z = \frac{x-\mu}{\sigma}} \tag{2.24}$$

Note que las puntuaciones z son simplemente puntuaciones de desviación divididas entre la desviación estándar de la distribución. Esta división se hace para *estandarizar* las desviaciones o colocarlas en una escala común. En vez de expresar las desviaciones en términos de sus unidades originales (presión sanguínea y peso en este ejemplo), ahora se expresan en unidades de desviación estándar. Cuando todas las puntuaciones en una distribución son convertidas en puntuaciones z, éstas tienen una media de cero y una desviación estándar de uno, independientemente de la escala original de los datos.

Suponga ahora que la desviación de la presión sanguínea de 12 tiene una puntuación z de 1.5, mientras que la del peso es de .6. Esto indica que el paciente está 1.5 desviaciones estándar por arriba de la media de un grupo en términos de la presión sanguínea, pero que únicamente está .6 desviaciones estándar por arriba de la media en términos del peso. Nosotros no hablaremos más acerca de las puntuaciones z en este momento porque las analizaremos con mayor detalle cuando comencemos nuestro estudio de la inferencia.

EJEMPLO 2.21

Convierta el conjunto de puntuaciones 1, 3, 3 y 9 a puntuaciones *z*. Luego calcule la media y la desviación estándar de las puntuaciones *z*.

Solución La media y la desviación estándar de los datos originales son 4 y 3.46, respectivamente. Utilizando estos valores en la ecuación 2.23 se obtiene

$$z_1 = \frac{1-4}{3.46} = .867$$
$$z_3 = \frac{3-4}{3.46} = .289$$
$$z_3 = \frac{3-4}{3.46} = .289$$
$$z_9 = \frac{9-4}{3.46} = 1.445$$

Como la suma de estas puntuaciones es cero, la media también es cero. Con una media de cero, la ecuación 2.15 se simplifica a

$$s = \sqrt{\frac{\sum z^2}{n-1}} = \sqrt{\frac{3}{3}} = 1$$ ■

EJEMPLO 2.22

Convierta el conjunto de puntuaciones 11, 32, 13, 9 y 10 a puntuaciones z. Luego encuentre la media y la desviación estándar de las puntuaciones *z*. Al comparar los resultados del ejemplo 2.21 con los resultados aquí obtenidos, ¿cuál fue la observación más extrema en términos de distancia desde la media?

Solución $\bar{x} = \frac{75}{5} = 15, \qquad s = \sqrt{\frac{1495 - \frac{75^2}{5}}{4}} = 9.62$

Utilizando estos resultados

$$z_{11} = \frac{11-15}{9.62} = -.416$$
$$z_{32} = \frac{32-15}{9.62} = 1.767$$
$$z_{13} = \frac{13-15}{9.62} = -.208$$
$$z_9 = \frac{9-15}{9.62} = -.624$$
$$z_{10} = \frac{10-15}{9.62} = -.520$$

Nuevamente, la media es cero (con el redondeo) y la desviación estándar (con el redondeo) es

$$\sqrt{\frac{4}{4}} = 1$$

La puntuación 32 es la más extrema en los dos conjuntos de datos, ya que se encuentra a 1.767 desviaciones estándar por arriba de la media de su distribución. Esto es más que cualquier otro valor. ■

Mientras que las puntuaciones *z* son útiles para la descripción de datos, su verdadera importancia se descubre en la inferencia.

2.6.4 Medidas de la forma de una distribución

Ciertos aspectos de la forma de una distribución pueden ser caracterizados numéricamente. Dos de los más comunes, el sesgo y la curtosis, se analizarán en este apartado. Por el momento simplemente los trataremos como dos más en la lista de los métodos descriptivos que usted tiene que aprender. En capítulos posteriores los utilizaremos cuando discutamos las violaciones a los supuestos que subyacen en ciertos métodos inferenciales.

Sesgo. Considere la forma de los tres polígonos de frecuencia mostrados en la figura 2.12. La distribución A es asimétrica en el sentido de que una mitad de la distribución es una imagen en espejo de la otra. La distribución B tiene la mayoría de sus observaciones en el extremo superior de la escala, pero con una "cola" extendida a la izquierda. La distribución C tampoco es simétrica, pero tiene sus observaciones más frecuentes en la parte baja de la escala y su cola se extiende hacia la derecha. Las distribuciones de estas tres formas generales se denominan respectivamente **simétrica, con asimetría negativa** y **con asimetría positiva**.

Se han desarrollado varios métodos para describir numéricamente la cantidad de asimetría (o falta de ella) que caracteriza a una distribución. El **sesgo** se define generalmente como el grado de asimetría de una distribución. La medida más común del sesgo está dada por

$$\boxed{\text{Sesgo}^{10} = \frac{\sum z^3}{n}} \qquad (2.25)$$

donde z es la desviación estandarizada, como se describió por medio de la ecuación 2.23 y n es el tamaño de la muestra. Dicho de otra forma, esta expresión de la asimetría es simplemente el promedio del cubo de las puntuaciones z.

Considere la distribución B de la figura 2.12. La media de esta distribución es aproximadamente 4.83. Note que las desviaciones de las puntuaciones que son mayores que la media tendrán un signo positivo, mientras que las que están por debajo de la media tendrán un signo negativo. Cuando estas desviaciones son estandarizadas y elevadas al cubo conservan su signo y, debido a la cola extendida por debajo de la media, su suma será negativa. De esta forma, el valor del sesgo también será negativo. La misma lógica indica que el sesgo de la distribución C es positivo, mientras que el de A es cero. En general, cuanto más se aleje de cero, más grande es el sesgo.

EJEMPLO 2.23

Calcule el sesgo de cada una de las tres distribuciones en la figura 2.12.

Solución La tabla 2.8 muestra los valores de los datos que forman la figura 2.12, junto con las puntuaciones z correspondientes y sus cubos. Como se esperaba, la suma de los valores z de cada distribución es cero. La suma del cubo de los valores z también es cero para la distribución simétrica, por lo que el valor del sesgo es cero. Esto ocurre siempre con las distribuciones simétricas. El sesgo de la distribución B es $\frac{-11.670}{18} = -.648$ mientras que el de C es $\frac{-11.670}{18} = .648$. ■

[10] Una estimación sin sesgo de la asimetría de la población está dada por la expresión $\frac{n\sum z^3}{(n-1)(n-2)}$.

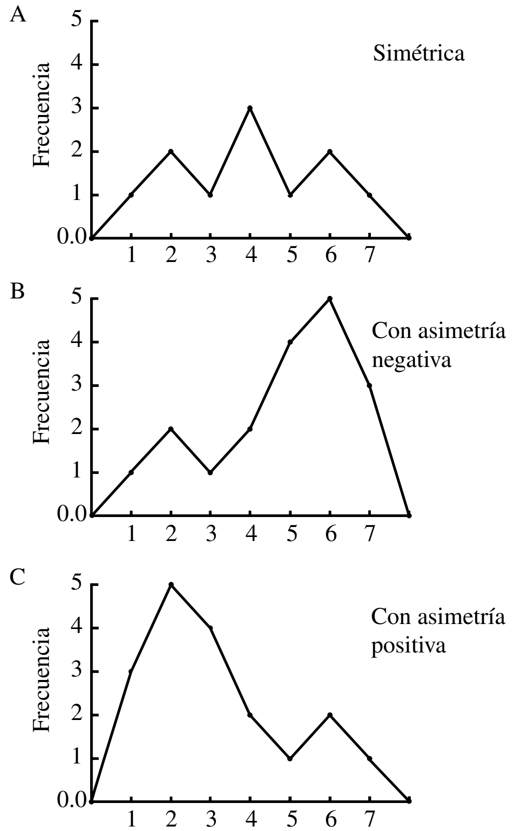


FIGURA 2.12: Polígonos con varias formas.

Curtosis. Considere las dos distribuciones mostradas en la figura 2.13. Cada una tiene una media y una mediana igual a cuatro, con una varianza y desviación estándar de .73 y .85 respectivamente. Cada una también es simétrica. A pesar de tener estos estadísticos en común, las dos distribuciones tienen formas muy diferentes. La distribución A es puntiaguda en la mitad y tiene colas que se extienden más allá de las de la distribución B. En contraste, la distribución B es aplanada en la mitad y tiene colas más cortas.

La “forma de pico” de una curva puede expresarse por su curtosis. De manera más precisa, la **curtosis** se refiere al pico de una distribución con respecto a la longitud y el tamaño de sus colas. Las distribuciones muy puntiagudas, como la distribución A, se conocen como **leptocúrticas**, mientras que aquellas que están aplanadas a la mitad se conocen como **platicúrticas**.[11]

La fórmula para la curtosis es

$$\boxed{\text{Curtosis}^{12} = \frac{\sum z^4}{n}} \tag{2.26}$$

[11] El prefijo *lepto* significa “alto y delgado” o “estrecho”, mientras que *plati* significa “plano” o “ancho”.

[12] Una estimación sin sesgo de la curtosis de la población está dada por la expresión $\frac{n\sum z^4}{(n-1)(n-2)}$.

TABLA 2.8: Datos, puntuaciones *z* y puntuaciones *z* al cubo de la figura 2.12.

Distribución A			Distribución B			Distribución C		
x	z	z^3	x	z	z^3	x	z	z^3
1	–1.581	–3.952	1	–2.103	–9.301	1	–1.188	–1.677
2	–1.054	–1.171	2	–1.554	–3.753	1	–1.188	–1.677
2	–1.054	–1.171	2	–1.554	–3.753	1	–1.188	–1.677
3	–.527	–.146	3	–1.006	–1.018	2	–.640	–.262
4	.000	.000	4	–.457	–.095	2	–.640	–.262
4	.000	.000	4	–.457	–.095	2	–.640	–.262
4	.000	.000	5	.091	.001	2	–.640	–.262
5	.527	.146	5	.091	.001	2	–.640	–.262
6	1.054	1.171	5	.091	.001	3	–.091	–.001
6	1.054	1.171	5	.091	.001	3	–.091	–.001
7	1.581	3.952	6	.640	.262	3	–.091	–.001
			6	.640	.262	3	–.091	–.001
			6	.640	.262	4	.457	.095
			6	.640	.262	4	.457	.095
			6	.640	.262	5	1.006	1.018
			7	1.188	–1.677	6	1.554	3.753
			7	1.188	–1.677	6	1.554	3.753
			7	1.188	–1.677	7	2.103	9.301
Σ	0.000	0.000		.000[a]	– 11.670		0.000	11.670

[a] Este valor es cero cuando los cálculos se realizan con un número suficiente de decimales.

Usted notará que mientras la asimetría se expresó como el promedio de las puntuaciones *z* al cubo en una distribución, la curtosis es el promedio de las puntuaciones *z* elevadas a la cuarta potencia. En general, los valores grandes de curtosis reflejan formas más puntiagudas que los valores pequeños. Usted también deberá notar que la curtosis se aplica a distribuciones con no más de una moda.

EJEMPO 2.24

Encuentre la curtosis de las distribuciones mostradas en la figura 2.13.

Solución La tabla 2.9 muestra los datos que forman las distribuciones de la figura 2.13, junto con sus puntuaciones *z* y las puntuaciones *z* elevadas a la cuarta potencia. Utilizando las sumas de las puntuaciones *z* elevadas a la cuarta potencia, las curtosis de la distribución A y B son, respectivamente

$$\frac{60.5}{12} = 5.04 \text{ y } \frac{15.125}{12} = 1.26$$

Estos resultados confirman la aseveración anterior de que si todo se mantiene igual, las distribuciones más puntiagudas tienen mayores valores de curtosis que las distribuciones menos puntiagudas. ■

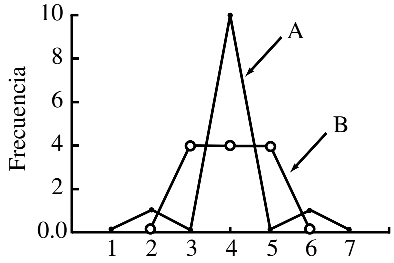


FIGURA 2.13: Dos distribuciones con media y varianza en común pero diferente curtosis.

TABLA 2.9: Datos, puntuaciones *z* y puntuaciones *z* elevadas a la cuarta potencia de la figura 2.13.

	Distribución A			Distribución B		
	x	z	z^4	x	z	z^4
	2	−2.345	30.250	3	−1.173	1.891
	4	0.000	0.000	3	−1.173	1.891
	4	0.000	0.000	3	−1.173	1.891
	4	0.000	0.000	3	−1.173	1.891
	4	0.000	0.000	4	0.000	0.000
	4	0.000	0.000	4	0.000	0.000
	4	0.000	0.000	4	0.000	0.000
	4	0.000	0.000	4	0.000	0.000
	4	0.000	0.000	5	1.173	1.891
	4	0.000	0.000	5	1.173	1.891
	4	0.000	0.000	5	1.173	1.891
	6	2.345	30.250	5	1.173	1.891
Σ		0.000	60.500		0.000	15.125

2.7 UNA REORIENTACIÓN

Usted acaba de terminar un capítulo largo y, en cierto grado, tedioso que comenzó con las escalas de medición y la notación de sumatoria, y continuó con un grupo de estadísticos descriptivos. Con un estudio tan intenso de “los árboles” es difícil mantener una perspectiva adecuada del “bosque”. Para comprender la bioestadística es importante que usted mantenga una amplia perspectiva mientras estudia los detalles. Con este fin, tal vez desee dar un repaso al capítulo 1, antes de continuar. Debe prestar especial atención a la figura 1.1.

El enfoque de este libro ahora pasa de la estadística descriptiva a la inferencial. Sin embargo, antes de estudiar la estadística inferencial, usted debe aprender algunos fundamentos de probabilidad. Éste será el tema del capítulo 3.

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

curtosis 45
desviación estándar 37
desviación media 34
distribuciones agrupadas 17
distribuciones de frecuencias 15
distribuciones de frecuencias acumulativas 16
distribuciones de frecuencias relativas 15
distribuciones de frecuencias relativas acumulativas 17
escalas de medición 9
gráfica de barras 19
gráficas de tallo y hojas 22
histogramas 20
intervalos de clase 17
límite real inferior 20
límite real superior 20
media 25
mediana 26
moda 31
notación de sumatoria 12
percentiles 38
polígonos 20
puntuaciones z 42
rango 32
rango semiintercuartil 40
rangos de percentiles 40
sesgo 44
tendencia central 25
variabilidad 32
variables continuas 11
variables dicotómicas 12
variables discretas 12
varianza 35

EJERCICIOS

2.1 La tabla 2.10 en la siguiente página muestra el número de días al año que las enfermeras de un hospital urbano tuvieron incapacidad por enfermedad en 2003. Las enfermeras son listadas por antigüedad, esto es, la enfermera número 1 tiene la menor antigüedad, mientras que la enfermera número 21 tiene la mayor antigüedad.

a) ¿Qué nivel de medición representa la variable del número de enfermera? Justifique su respuesta.

b) ¿La variable de los días de enfermedad es continua o discreta? Justifique su respuesta.

c) Permita que x_i represente el número de días al año de incapacidad por enfermedad de la *i*-ésima enfermera donde el índice *i* es el número de enfermera. Calcule lo siguiente.

i. x_3, x_9, x_{21}

ii. $\sum_{i=1}^{10} x_i$

iii. $\sum_{i=11}^{n} x_i$

iv. $\sum_{i=1}^{n} x_i$

v. $\sum_{i=1}^{n} x_i^2$

d) Suponga que cada enfermera tomó exactamente dos días más por enfermedad de lo reportado en la tabla. Utilice la notación de sumatoria para expresar la suma en el inciso iv de la pregunta 1*c*), de forma que refleje los dos días adicionales que tomó cada enfermera por enfermedad.

e) Utilice los datos de los días de incapacidad por enfermedad al año de las enfermeras para construir distribuciones de frecuencias, de frecuencias acumulativas, de frecuencias relativas y de frecuencias relativas acumulativas.

f) Utilice los datos de los días de incapacidad por enfermedad al año de las enfermeras para construir un histograma de frecuencias relativas, un polígono de frecuencias relativas y un polígono de frecuencias acumulativas.

g) Utilice los datos de los días de incapacidad por enfermedad al año de las enfermeras para calcular lo siguiente.

i. media, mediana y moda
ii. varianza y desviación estándar
iii. rangos exclusivo e inclusivo
iv. puntuaciones *z* de cada una de las x_i
v. los percentiles 15, 50 y 80
vi. rangos de percentiles de 2, 5 y 8 días de enfermedad
vii. sesgo y curtosis

TABLA 2.10: Tabla para el ejercicio 2.1.

Número de enfermera	Días de enfermedad	Número de enfermera	Días de enfermedad	Número de enfermera	Días de enfermedad
1	2	8	7	15	9
2	9	9	8	16	2
3	1	10	8	17	8
4	0	11	6	18	9
5	5	12	3	19	6
6	4	13	7	20	8
7	6	14	8	21	5

TABLA 2.11: Tabla para el ejercicio 2.2.

Núm. de sol.	Género	Ps	Pr	Núm. de sol.	Género	Ps	Pr
1	m	165	167	16	m	220	225
2	m	215	210	17	f	135	137
3	m	190	186	18	f	180	201
4	f	115	111	19	m	210	205
5	m	158	160	20	f	145	144
6	f	165	160	21	f	131	133
7	f	120	118	22	m	177	180
8	m	173	170	23	m	135	135
9	m	188	195	24	m	183	180
10	m	180	195	25	m	165	166
11	f	135	137	26	f	160	166
12	m	155	155	27	m	178	180
13	m	190	195	28	m	155	152
14	m	187	185	29	f	155	154
15	f	154	156	30	f	130	128

2.2 Los datos en la tabla 2.11 de esta página fueron obtenidos de solicitantes de empleo en el Servicio de Salud Pública. Está registrado el número del solicitante, el género, el peso reportado por el solicitante (Ps), tomado de la solicitud de empleo, y el peso registrado en el momento del análisis físico (Pr) del solicitante. Los pesos se registraron en libras.

a) ¿Qué nivel de medición representa la variable género? ¿De qué otra forma se podría caracterizar a esta variable? Justifique sus respuestas.

b) Construya una gráfica de barras de frecuencias relativas para la variable género. ¿Será posible construir un polígono de frecuencias acumulativas para la variable género? Justifique su respuesta.

c) Utilizando el número de solicitante como el índice de sumatoria (i), calcule lo siguiente:

i. $\sum_{i=1}^{15} \mathrm{W}s_i$ y $\sum_{i=1}^{15} \mathrm{W}p_i$

ii. $\sum_{i=1}^{15} (\mathrm{W}s_i - \mathrm{W}p_i)$

(¿Este cálculo es realmente necesario para obtener el resultado deseado? ¿Por qué?)

d) Utilizando los intervalos de clase 110-119, 120-129, … , 220-229, use la variable Ps para construir un histograma de frecuencias relativas agrupadas, un polígono de frecuencias relativas agrupadas y un polígono de frecuencias relativas acumulativas agrupadas.

e) Utilice los datos de Pr para calcular lo siguiente:

i. media, mediana y moda
ii. varianza y desviación estándar
iii. rangos exclusivo e inclusivo
iv. puntuaciones z de cada una de las x_i
v. el primero, segundo y tercer cuartiles
vi. rangos de percentiles de 111, 137 y 180
vii. sesgo y curtosis

A. Las siguientes preguntas se refieren al estudio de caso A (página 469).

2.3 ¿Hay alguna variable dicotómica mencionada en este estudio? De ser así, ¿puede usted determinar si pertenecen a una de las cuatro escalas de medición?

2.4 ¿Hay alguna variable continua mencionada en este estudio? De ser así, ¿puede usted determinar si pertenecen a una de las cuatro escalas de medición?

2.5 ¿La edad promedio de los sujetos reportada en este estudio está representada adecuadamente por x o μ? ¿Por qué? ¿Cuál símbolo debería utilizarse para la desviación estándar de la variable edad?

2.6 ¿Cuál es la categoría modal si/no para los datos A.M. (mañana) reportados en la tabla?

2.7 ¿Cuál será la puntuación z de un sujeto que tiene 18 años de edad? ¿Y de uno de 75 años de edad?

D. Las siguientes preguntas se refieren al estudio de caso D (página 471).

2.8 Construya una distribución de frecuencia relativa con las puntuaciones de la batería NPZ-8.

2.9 Construya un histograma de frecuencias relativas y un polígono con las puntuaciones de la batería NPZ-8.

2.10 Calcule la media, mediana y moda de los datos de la batería NPZ-8.

2.11 Calcule la media, mediana y moda de los datos de PBV.

2.12 Calcule la varianza y la desviación estándar de los datos de la batería NPZ-8.

2.13 Obtenga el primero, segundo y tercer cuartiles de los datos de la batería NPZ-8.

2.14 Inspeccione el histograma de frecuencias relativas y el polígono que construyó con los datos de la batería NPZ-8 en el ejercicio 2.9. A partir de su inspección, ¿cómo podría definir esta distribución, con asimetría positiva, con asimetría negativa o simétrica?

2.15 Calcule el sesgo para los datos de la batería NPZ-8. ¿Este cálculo confirma su observación hecha en el ejercicio 2.14?

2.16 Calcule el sesgo para los datos de la batería NPZ-8 utilizando únicamente datos de 15 sujetos infectados. En su opinión, ¿la inclusión de los cinco sujetos sanos en el conjunto de datos incrementa la asimetría de manera notoria?

O. Las siguientes preguntas se refieren al estudio de caso O (página 477).

2.17 ¿Cuáles escalas describen mejor los niveles de medición representados por las variables mencionadas en el punto *a*)?

CAPÍTULO 3

Probabilidad

3.1 INTRODUCCIÓN

En el capítulo 1 señalamos que la probabilidad es la base de la estadística inferencial. Dicho de otra manera, es el mecanismo por medio del cual se hacen inferencias. Por eso es necesario que usted adquiera conocimientos básicos de probabilidad antes de comenzar sus estudios sobre inferencias. Este capítulo le dará tales conocimientos.

La probabilidad es una compleja rama de las matemáticas y tiene sus raíces firmemente arraigadas en los juegos de azar. En efecto, gran parte del trabajo inicial en esta área se llevó a cabo en un intento por obtener ventaja en esos juegos. Por fortuna, sus estudios en bioestadística requerirán solamente los conocimientos más básicos de probabilidad.

La mayoría de las introducciones a la probabilidad, por lo menos las que se encuentran en los textos de estadística, se basan en los antecedentes de la disciplina. Así, se utilizan muchos ejemplos que tienen que ver con lanzamientos de dados y juegos de naipes para impartir los conceptos fundamentales. Éste es un enfoque razonable del estudio de la probabilidad. Pero nuestro objetivo en este capítulo no es hacer que usted aprenda probabilidad, sino probabilidad en relación con la estadística inferencial. En nuestra experiencia, los estudiantes a menudo tienen dificultades para relacionar estos ejemplos con el estudio posterior de la inferencia. Por tal motivo, dejaremos de lado la tradición y en su lugar presentaremos la probabilidad en contextos específicos más relacionados con la inferencia. Específicamente, después de definir algunos términos, usted estudiará la probabilidad y su relación con las tablas de contingencia y la curva normal. A lo largo de este estudio, también lo conduciremos a importantes métodos descriptivos que no fueron cubiertos en el capítulo 2, tales como tasas de riesgo, razones de probabilidad, sensibilidad, especificidad y valores predictivos positivos y negativos. Empezaremos con una definición de probabilidad.

3.2 UNA DEFINICIÓN DE PROBABILIDAD

Las definiciones de probabilidad que son suficientes para el estudio riguroso de la disciplina pueden ser complejas y controvertidas. Sin embargo, para el propósito de este texto bastará con una definición sencilla e intuitiva. Comenzaremos con un ejemplo.

Suponga que hay cuatro canicas en un balde, tres son negras y una es blanca. Sacudimos vigorosamente el balde y luego, con los ojos cerrados, sustraemos una canica. (Sacudir el balde y cerrar los ojos sirven para asegurar que la selección se ha hecho *al azar*.) Ahora nos hacemos la pregunta, "¿cuál es la probabilidad de que la canica sustraída sea negra?". Sospechamos que su respuesta será algo así como "tres de cuatro" o "tres cuartos" o "punto setenta y cinco". Concordamos con estas respuestas, pero, ¿qué definición de probabilidad utilizó usted para llegar a ellas? Si reflexiona un poco, se dará cuenta de que definió la probabilidad de sustraer una canica negra como el número de canicas negras dividido entre el número total de canicas. ¿Cuál es la probabilidad de que la canica sea blanca? De nuevo, la probabilidad es simplemente el número de canicas blancas dividido entre el número total de canicas, o .25. En otras palabras, la probabilidad de una canica negra era sólo la *proporción* de canicas negras, y es igual para la probabilidad de una canica blanca. Formalmente, podemos definir la **probabilidad** de algún suceso,[1] llamémosle A, como

$$P(A) = \frac{N_A}{N} \tag{3.1}$$

El símbolo P () se lee "La probabilidad de..." en la que se usa A para representar cualquier suceso de interés, por ejemplo, "sustraer una canica negra". Se utilizará el símbolo $\overline{A}$ para referirse al complemento de A o lo que en este caso significaría "no sustraer una canica negra". Así, en el caso que nos ocupa, $P(\overline{A})$ se leería, "La probabilidad de no sustraer una canica negra". N_A es el número de eventos que cumplen el criterio especificado (por ejemplo, canicas negras) y N es el número total de eventos.

Se pueden deducir varias propiedades importantes a partir de la definición en la expresión 3.1.

1. $P(A) \geq 0$. Esto se debe al hecho de que N_A es una cantidad y, por lo tanto, no puede ser menor a cero.

2. $P(A) \leq 1$. Esto se debe al hecho de que N_A nunca puede exceder a N.

3. $P(A) + P(\overline{A}) = 1$ o $P(\overline{A}) = 1 - P(A)$, lo que se debe al hecho de que los resultados de $N - N_A$ no cumplen con el criterio declarado A. Por lo tanto, $P(\overline{A}) = \frac{(N - N_A)}{N} = 1 - \frac{N_A}{N} = 1 - P(A)$.

Con esta definición en mente, comencemos nuestro estudio sobre probabilidad en contextos específicos. El primero de ellos se relacionará con las tablas de contingencia.

3.3 TABLAS DE CONTINGENCIA

3.3.1 Muestreo de la población

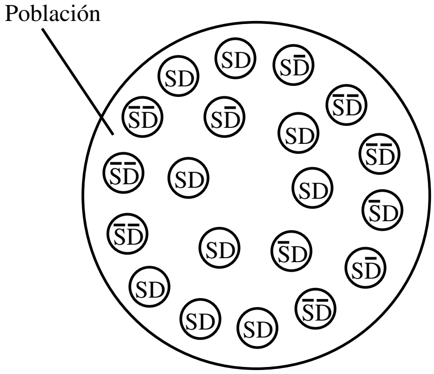
La figura 3.1 describe una población conceptualizada que consta de dos características asociadas con cada una de 20 personas.[2] Cada una de las 20 personas se caracteriza por ser fumador (S), por no ser fumador $\overline{S}$, por tener alguna enfermedad en particular (D), o por no tener la enfermedad $\overline{D}$.

Ahora, planteemos la pregunta, "¿cuál es la probabilidad de seleccionar al azar en esta población a una persona que es fumadora?". Ésta es sólo una variación del ejemplo de las canicas

---

[1] Esta definición supone eventos igualmente probables.

[2] En el capítulo 1 señalamos que las poblaciones consideradas en este libro son muy grandes. Por razones pedagógicas, la presente descripción es pequeña.



S = fumador, $\bar{S}$ = no fumador, D = enfermedad, $\bar{D}$ = no enfermedad.

FIGURA 3.1: Población de 20 personas con dos características.

en el balde que vimos anteriormente y se resuelve de la misma manera. Si contamos el número de fumadores y aplicamos la ecuación 3.1, tenemos que $P\ (S) = \frac{12}{20} = .60$. De la misma forma, la probabilidad de seleccionar a una persona que no tiene la enfermedad es $P\ (\bar{D}) = \frac{9}{20} = .45$.

También se pueden calcular probabilidades que impliquen ambas características. Por ejemplo, ¿cuál es la probabilidad de seleccionar a una persona que fuma *y* tiene la enfermedad? Ya que nueve personas cumplen con este criterio, la probabilidad es $P\ (SD) = \frac{9}{20} = .45$. Note que $P\ (SD)$ se lee: "La probabilidad de ser fumador *y* estar enfermo." Ahora, calcule $P\ (\bar{S}D)$. ¿Qué representa? Ésta es la probabilidad de seleccionar a una persona que no fuma y tiene la enfermedad y se calcula como $P\ (\bar{S}D) = \frac{2}{20} = .10$.

¿Cuál es la probabilidad de encontrar que la persona seleccionada fuma *o* tiene la enfermedad? En este caso el criterio se cumple con cualquiera que fume *o* tenga la enfermedad. Puesto que hay 14 personas asociadas con una S o una D (o ambas), la probabilidad es $P\ (S \cup D) = \frac{14}{20} = .70$. Observe que la notación $P\ (S \cup D)$ se lee: "La probabilidad de ser fumador o estar enfermo."

En muchas aplicaciones estadísticas el interés se centra sólo en una porción especificada de la población. Por ejemplo, la pregunta, "¿cuál es la probabilidad de enfermedad, dado que la selección se hace *entre los fumadores*?", implica que sólo los fumadores deben ser considerados. El símbolo usado para este tipo de probabilidad es $P\ (D \mid S)$, que se lee: "La probabilidad de enfermedad dado un fumador." Esta forma general se conoce como **probabilidad condicional**.

La figura 3.2 muestra la población redefinida, es decir, la población original quitando a todos los no fumadores. La probabilidad de seleccionar a una persona enferma entre los fumadores será entonces $P\ (D \mid S) = \frac{9}{12} = .75$.

3.3.2 Tablas de frecuencias

Las sumas utilizadas para los cálculos en la sección 3.3.1 pueden resumirse fácilmente en una tabla de contingencia, como se muestra en la tabla 3.1.

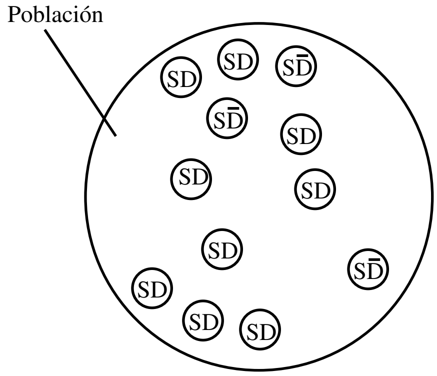


FIGURA 3.2: Población de enfermos sólo para fumadores.

TABLA 3.1: Tabla de contingencia que muestra las frecuencias de población.

	D	$\overline{D}$	
S	9	3	12
$\overline{S}$	2	6	8
	11	9	

Los números en las cuatro celdas son recuentos de las personas que cumplen con los dos criterios indicados. Eso es, había nueve personas que fumaban y tenían la enfermedad, tres que fumaban y no tenían la enfermedad, dos que no fumaban y tenían la enfermedad y seis que no fumaban y no tenían la enfermedad. Los valores en los márgenes de la tabla dan el conteo total de la característica indicada. Así, 12 personas fumaban, ocho no fumaban, 11 tenían la enfermedad y nueve no tenían la enfermedad.

EJEMPLO 3.1

Utilice los datos de la tabla de contingencia para calcular lo siguiente. También determine el significado de cada probabilidad. Calcule $P\ (\overline{S})$, $P\ (D)$, $P\ (S\overline{D})$, $P\ (\overline{S}\,\overline{D})$, $P\ (S \cup \overline{D})$, $P\ (\overline{S} \cup D)$, $P\ (\overline{D} \mid S)$, y $P\ (\overline{S} \mid \overline{D})$.

Solución $P\ (\overline{S})$ es la probabilidad de seleccionar a una persona que no fuma. Si utilizamos el conteo apropiado de la tabla de contingencia nos da $\frac{8}{20} = .40$. De igual forma, $P\ (D)$ es la probabi lidad de seleccionar a alguien con la enfermedad y es $\frac{11}{20} = .55$.

$P\ (S\overline{D})$ es la probabilidad de seleccionar a alguien que fuma *y* no tiene la enfermedad, en tanto que $P\ (\overline{S}\,\overline{D})$ es la probabilidad de seleccionar a un no fumador que no tiene la enfermedad. Éstas son, respectivamente, $\frac{3}{20} = .15$ y $\frac{6}{20} = .30$.

$P\ (S \cup \overline{D})$ es la probabilidad de seleccionar a alguien que fuma *o* que no tiene la enfermedad. El número de personas que cumplen con este requerimiento es $9 + 3 + 6 = 18$, lo que nos da

TABLA 3.2: Tabla de contingencia que muestra las probabilidades de la población.

	D	$\overline{D}$	
S	.45	.15	.60
$\overline{S}$	.10	.30	.40
	.55	.45	

una probabilidad de $\frac{18}{20} = .90$. Ya que 9 + 2 + 6 = 17 personas son no fumadores *o* tienen la enfermedad, $P\,(\overline{S} \cup D) = \frac{17}{20} = .85$.

La probabilidad condicional $P\,(\overline{D} \mid S)$ es la probabilidad de seleccionar a alguien que no tiene la enfermedad, siendo que él o ella es un fumador. Dicho de otra forma, es la probabilidad de seleccionar a alguien que no tiene la enfermedad si la selección se hace deliberadamente entre los fumadores. Puesto que hay 12 fumadores, tres de los cuales no tienen la enfermedad, la probabilidad de seleccionar a una persona no enferma entre los fumadores es $\frac{3}{12} = .25$.

Finalmente, $P\,(\overline{S} \mid \overline{D})$ es la probabilidad de un no fumador, de aquellos que no tienen la enfermedad. Otra vez, si utilizamos los datos de la tabla, hay un total de nueve personas sin enfermedad. Ya que seis de ellas son no fumadores, la probabilidad de seleccionar a un no fumador entre las personas sin enfermedad es $\frac{6}{9} = .67$. ■

3.3.3 Tablas de probabilidad

Otra forma común de tabla de contingencia se obtiene al dividir cada conteo en la tabla de frecuencia entre N para calcular probabilidades. La tabla de probabilidad representada por la tabla 3.2 se construyó a partir de la tabla de frecuencias que se muestra en la tabla 3.1 de la página opuesta.

Observe que las entradas en las celdas de esta tabla representan $P\,(SD)$, $P\,(S\overline{D})$, $P\,(\overline{S}D)$ y $P\,(\overline{S}\,\overline{D})$, de forma que estos valores pueden leerse directamente de la tabla. Lo mismo ocurre con $P\,(S)$ y $P\,(\overline{S})$, que se encuentran en los márgenes de renglón y $P(D)$ y $P\,(\overline{D})$, que se encuentran en los márgenes de columna. Las probabilidades de la forma $P\,(S \cup \overline{D})$ y $P\,(\overline{S} \cup D)$ se obtienen sumando las entradas de las celdas apropiadas. Estos valores son, respectivamente, .45 + .15 + .30 = .90 y .45 + .10 + .30 = .85, que son las mismas respuestas obtenidas previamente de la tabla de frecuencias.

Las probabilidades condicionales como $P\,(\overline{D} \mid S)$ y $P\,(\overline{S} \mid \overline{D})$ se calculan de la misma forma que con las tablas de frecuencias. Debido a que $P\,(\overline{D} \mid S)$ es la probabilidad de no enfermedad, dado un fumador o, de forma equivalente, la proporción de fumadores que no tienen la enfermedad, la fracción es $\frac{.15}{.60} = .25$, que es el resultado obtenido con anterioridad. De igual forma, la probabilidad de seleccionar a un no fumador entre las personas sin la enfermedad es simplemente la proporción de personas sin la enfermedad que son no fumadores o $\frac{.30}{.45} = .67$, otra vez, la respuesta obtenida de la tabla de frecuencias.

En general, si dejamos que A, $\overline{A}$, B, y $\overline{B}$, representen características arbitrarias, entonces las entradas de la tabla de probabilidad correspondiente serán

	B	$\overline{B}$	
A	$P\,(AB)$	$P\,(A\overline{B})$	$P\,(A)$
$\overline{A}$	$P\,(\overline{A}B)$	$P\,(\overline{A}\,\overline{B})$	$P\,(\overline{A})$
	$P\,(B)$	$P\,(\overline{B})$	

Las entradas en las celdas de una tabla de probabilidad se conocen como probabilidades **conjuntas** mientras que aquéllas en los márgenes se denominan probabilidades **marginales**. Usted también debe tomar en cuenta que el orden en el que se listan las características en las expresiones de probabilidad conjunta no es importante. Esto es, $P(AB)$ es equivalente a $P(BA)$. Lo mismo es cierto para las expresiones de la forma $P\,(A \cup B)$, ya que ésta significa lo mismo que $P\,(B \cup A)$. Es importante comprender que lo mismo *no* es cierto para la probabilidad condicional. Así, $P\,(A \mid B)$ *no* es lo mismo que $P\,(B \mid A)$.

Finalmente, presentamos dos fórmulas útiles para calcular y comprender las probabilidades relacionadas con tablas de contingencia. Éstas son

$$\boxed{P\,(A \cup B) = P\,(A) + P\,(B) - P\,(AB)} \tag{3.2}$$

y

$$\boxed{P\,(A \mid B) = \frac{P\,(AB)}{P\,(B)}} \tag{3.3}$$

3.3.4 Independencia

La independencia juega un papel fundamental en muchos de los métodos inferenciales que usted estudiará en capítulos posteriores, así como en la metodología general de la investigación relacionada con la salud. Se dice que dos eventos, A y B, son **independientes** si

$$\boxed{P\,(A \mid B) = P(A)} \tag{3.4}$$

Si no se mantiene esta igualdad, se dice que los eventos son *no independientes* o *dependientes*. El significado de este enunciado se comprenderá si se consideran las probabilidades en la tabla 3.2. La probabilidad de enfermedad en la población general (es decir, $P(D)$) es .55. Pero la probabilidad de enfermedad entre los fumadores (es decir, $P\,(D \mid S)$) es $\frac{.45}{.60} = .75$. Como los fumadores tienen una probabilidad más alta de enfermedad que la población general, parece razonable concluir que existe alguna forma de relación entre el tabaquismo y esta enfermedad.[3]

Por otra parte, supongamos que la probabilidad de enfermedad entre los fumadores fuera la misma que la de la población general. Probablemente usted concluiría que los fumadores no corren un riesgo mucho mayor de tener la enfermedad que la población general. En esta circunstancia, el tabaquismo y la enfermedad serían eventos independientes.

La definición de independencia dada con anterioridad se presenta, a menudo, de manera ligeramente diferente. Si la parte derecha de la ecuación 3.3 se sustituye por la parte izquierda de la ecuación 3.4 y ambas partes se multiplican por $P(B)$, el resultado es

$$\boxed{P\,(AB) = P\,(A)\,P\,(B)} \tag{3.5}$$

---

[3] Sin embargo, esto no prueba que el tabaquismo *causa* la enfermedad.

De esta forma, la definición de independencia estipula que A y B son independientes si la probabilidad conjunta de A y B es igual al producto de sus probabilidades marginales. En referencia a la tabla 3.2, vemos que $.45 \neq (.60)\,(.55)$, y así, otra vez vemos que el tabaquismo y la enfermedad no son independientes.

EJEMPLO 3.2

Utilice la siguiente tabla de probabilidad para calcular $P(\overline{B})$, $P(A)$, $P(A\overline{B})$, $P(\overline{A}\overline{B})$, $P(\overline{A} \cup \overline{B})$, $P(\overline{A} \cup B)$, $P(\overline{B} \mid A)$ y $P(A \mid \overline{B})$. ¿Son independientes A y B? ¿Cuál es la justificación de su respuesta?

	B	$\overline{B}$	
A	.28	.12	.40
$\overline{A}$	.42	.18	.60
	.70	.30	

Solución Si leemos los márgenes de la tabla, vemos que $P(\overline{B}) = .30$ y $P(A) = .40$. De las celdas de la tabla tenemos que $P(A\overline{B}) = .12$ y $P(\overline{A}\overline{B}) = .18$. $P(A \cup \overline{B}) = .12 + .42 + .18 = .72$ en tanto que $P(\overline{A} \cup B) = .28 + .42 + .18 = .88$. Si se utiliza la ecuación 3.3 para calcular las probabilidades condicionales se obtiene $P(B \mid A) = \frac{.12}{.40} = .30$ y $P(A \mid \overline{B}) = \frac{.12}{.30} = .40$.

Como $P(A \mid B) = \frac{.28}{.70} = .40$ y $P(A) = .40$, la ecuación 3.4 implica que A y B son independientes. Se llega a la misma conclusión al aplicar la ecuación 3.5 y al observar que $P(AB) = .28$, que es igual a $P(A)\,P(B)$. ■

3.3.5 Sensibilidad, especificidad y conceptos relacionados

Las pruebas diseñadas para establecer la presencia o ausencia de alguna enfermedad rara vez son perfectas. Por ejemplo, nos gustaría que una prueba médica para establecer la presencia o ausencia de alguna enfermedad en particular fuera positiva para quienes tengan la enfermedad y negativa para quienes no la tengan. Por desgracia, en algunas ocasiones una persona enferma recibe un resultado negativo o una persona sana obtiene un resultado positivo.

El buen o mal desempeño de una prueba en este aspecto puede evaluarse a través del cálculo de su sensibilidad, especificidad, valor predictivo positivo y valor predictivo negativo. Usaremos la siguiente tabla de probabilidad para explicar cada uno de estos conceptos. En esta tabla un "+" indica un resultado positivo de la prueba y un "–" indica un resultado negativo. Como en las tablas anteriores, se utiliza D para indicar enfermedad y $\overline{D}$ para representar la ausencia de enfermedad. En esta tabla se puede ver que .015 de la población tiene la enfermedad y da positivo en la prueba, mientras que .970 de la población no tiene la enfermedad y da negativo.

	D	$\overline{D}$	
+	.015	.010	.025
–	.005	.970	.975
	.020	.980	

La **sensibilidad** de una prueba es la probabilidad de que una persona con la enfermedad dé un resultado positivo o

$$\boxed{\text{Sensibilidad} = P\,(+ \mid D)} \tag{3.6}$$

Para la tabla anterior, la sensibilidad es $\frac{.015}{.020} = .75$. Esto quizá no sea muy reconfortante, ya que sólo el 75% de las personas con la enfermedad están debidamente identificadas.

La **especificidad** de una prueba es la probabilidad de que una persona que no tiene la enfermedad (sana) dé un resultado negativo o

$$\boxed{\text{Especificidad} = P\,(- \mid \overline{D})} \tag{3.7}$$

Para la tabla anterior la especificidad es $\frac{.97}{.98} = .99$. Esto significa que si usted no tiene la enfermedad, es casi seguro (pero no totalmente) que la prueba dé negativo.

El **valor predictivo positivo** (*positive predictive value*, PPV) es la probabilidad de que una persona que da positivo tenga la enfermedad o

$$\boxed{PPV = P\,(D \mid +)} \tag{3.8}$$

El valor predictivo positivo para la tabla es $\frac{.015}{.025} = .60$. Esto significa que si una persona da positivo en la prueba de la enfermedad, la probabilidad de que tenga la enfermedad es de sólo .60.

El **valor predictivo negativo** (*negative predictive value*, NPV) es la probabilidad de que una persona que da negativo no tenga la enfermedad o

$$\boxed{NPV = P\,(\overline{D} \mid -)} \tag{3.9}$$

El valor predictivo negativo para la tabla es $\frac{.970}{.975} = .99$. Esto significa que si una persona da negativo en la prueba de enfermedad, la probabilidad de que no tenga la enfermedad es de .99.

Finalmente, la **prevalencia** es simplemente la probabilidad de enfermedad o

$$\boxed{\text{Prevalencia} = P(D)} \tag{3.10}$$

En este caso, este valor es .02.

EJEMPLO 3.3

Utilice la siguiente tabla para calcular la sensibilidad, la especificidad, el valor predictivo positivo, el valor predictivo negativo y la prevalencia.

	D	$\overline{D}$	
+	.008	.011	.019
−	.001	.980	.981
	.009	.991	

Solución La aplicación de las ecuaciones 3.6 a 3.10 da como resultados Sensibilidad = $\frac{.008}{.009} = .89$, Especificidad = $\frac{.980}{.991} = .99$, $PPV = \frac{.008}{.019} = .42$, $NPV = \frac{.980}{.981} = .999$, y Prevalencia = .009. ■

EJEMPLO 3.4

Suponga que le informaron a usted que su prueba médica es positiva. ¿*Cuál* de las características anteriores sería de mayor interés para usted?

Solución Con seguridad usted querría conocer la probabilidad que tiene una persona con un resultado positivo de tener la enfermedad. Por lo tanto, querría conocer el valor predictivo positivo. ■

3.3.6 Tasa de riesgo y razón de probabilidad

El capítulo 2 no cubrió todos los estadísticos descriptivos de interés. Dos de los más importantes, en lo que se refiere a su uso común en la investigación relacionada con la salud, son la tasa de riesgo y la razón de probabilidad. Ambos se explicarán en esta sección.

La tasa de riesgo. Las preguntas de investigación por lo común giran en torno a si las personas expuestas a algún factor de riesgo potencial tienen más o menos probabilidades de desarrollar una enfermedad que las personas no expuestas. Se podría pensar que el tabaquismo o trabajar con asbestos incrementa la probabilidad de enfermedad, en tanto que la exposición a cierta vacuna reduce la probabilidad de enfermedad. Un método común para comparar las probabilidades de enfermedad para personas expuestas y no expuestas es traducirlas en un cociente denominado **tasa de riesgo** (*risk ratio*, RR), que formalmente se expresa como

$$\boxed{RR = \frac{P(D \mid E)}{P(D \mid \bar{E})}} \tag{3.11}$$

en donde D se ha definido previamente, y E está expuesto, mientras que $\overline{E}$ no lo está.[4] De la ecuación 3.11 se deduce que una tasa de riesgo de 2 significaría que la probabilidad de enfermedad para las personas expuestas es dos veces mayor que para las personas no expuestas. De igual manera, un valor de .5 significaría que la probabilidad de enfermedad para el grupo expuesto es sólo la mitad de la del grupo no expuesto. Cuando RR es menor que 1.0, se dice que la exposición es **protectora**. Note también que una tasa de riesgo de 1.0 significaría que la probabilidad de enfermedad en ambos grupos es la misma.

EJEMPLO 3.5

Utilice la siguiente tabla de probabilidad para calcular la tasa de riesgo.

	E	$\overline{E}$	
D	.15	.10	.25
$\overline{D}$	.05	.70	.75
	.20	.80	

[4]Lo normal para la tasa de riesgo es que RR represente el parámetro y $\widehat{RR}$ represente el estadístico. Vea la explicación en la sección 1.3 de la página 3.

Solución Dado que $P\ (D \mid E) = \frac{.15}{.20} = .75$ y $P\ (D \mid \bar{E}) = \frac{.10}{.80} = .125$, $RR = \frac{.75}{.125} = 6$. Esto significa que la probabilidad de enfermedad para las personas expuestas es seis veces mayor que la de las personas no expuestas. ■

Razón de probabilidad. En algunos entornos de investigación, de los cuales usted aprenderá más adelante, la tasa de riesgo no aporta una comparación significativa del grupo expuesto y del no expuesto. En tales casos, se utiliza la razón de probabilidad para hacer comparaciones. Para comprender la razón de probabilidad, primero es necesario entender las posibilidades.

Las posibilidades de que ocurra un evento están indicadas por el cociente de la probabilidad de que el evento suceda y la probabilidad de que el evento no ocurra. Por lo tanto, las posibilidades de enfermedad en un cierto grupo serían $P\ (D)\ /\ P\ (\bar{D})$. Una posibilidad de 2.0 significaría que la probabilidad de enfermedad es dos veces mayor que la de no tenerla. Si se calculan las posibilidades de dos grupos y se convierten en un cociente, el resultado es, naturalmente, una razón de probabilidad. Si, como en el caso anterior, usted está comparando un grupo expuesto y un grupo no expuesto, la **razón de probabilidad** (*odds ratio*, OR) sería

$$OR = \frac{\dfrac{P(D|E)}{P(\bar{D}|E)}}{\dfrac{P(D|\bar{E})}{P(\bar{D}|\bar{E})}}$$

que se simplifica a

$$\boxed{OR = \frac{P\ (D \mid E)\ P\ (\bar{D} \mid \bar{E})}{P\ (\bar{D} \mid E)\ P\ (D \mid \bar{E})}} \tag{3.12}$$

donde E y $\bar{E}$ representan la exposición y la no exposición, respectivamente.

La razón de probabilidad no es tan intuitiva como la tasa de riesgo, pero tiene la ventaja de ser aplicable en un rango más amplio de diseños de estudio. Como ocurre con RR, una razón de probabilidad con valor de uno implica que ambos grupos están al mismo nivel de riesgo. Cuando la prevalencia de enfermedad es relativamente pequeña, la razón de probabilidad provee un buen cálculo de la tasa de riesgo. Retomaremos este tema en un capítulo posterior.[5]

EJEMPLO 3.6

Calcule la razón de probabilidad para la tabla del ejemplo 3.5 de la página 59.

Solución Dado que $P\ (D \mid E) = .15/.20 = .75$, $P\ (\bar{D} \mid \bar{E}) = .70/.80 = .875$, $P\ (\bar{D} \mid E) = .05/.20 = .25$ y $P\ (D \mid \bar{E}) = .10/80 = .125$, $OR = (.75)(.875)\ /\ [(.25)(.125)] = 21$. Esto significa que las posibilidades de enfermedad del grupo expuesto son 21 veces más altas que las del grupo no expuesto. ■

3.3.7 Regla de Bayes

Aunque la regla de Bayes es importante en el esquema general de la estadística, nosotros no ahondaremos tanto en su aplicación; simplemente, la presentaremos porque con seguridad usted se

[5] La norma para la razón de probabilidad es que OR represente el parámetro y $\widehat{OR}$ represente el estadístico.

encontrará con ella en sus futuros estudios de estadística. En su forma simple, la regla de Bayes le permite a usted usar $P(A \mid B)$ para calcular $P(B \mid A)$. La regla se expresa como

$$\boxed{P(B \mid A) = \frac{P(A \mid B)\,P(B)}{P(A \mid B)\,P(B) + P(A \mid \bar{B})\,P(\bar{B})}} \tag{3.13}$$

Note que

$$P(A \mid B)\,P(B) = \frac{P(AB)\,P(B)}{P(B)} = P(AB)$$

y que

$$P(A \mid \bar{B})\,P(\bar{B}) = \frac{P(A\bar{B})\,P(\bar{B})}{P(\bar{B})} = P(A\bar{B}).$$

También

$$P(AB) + P(A\bar{B}) = P(A).$$

Al sustituir éstas en la ecuación 3.13 tenemos que

$$P(B \mid A) = \frac{P(AB)}{P(A)}$$

lo que satisface la definición dada en 3.3.

EJEMPLO 3.7

Dada una sensibilidad de .83, una especificidad de .89 y una prevalencia de .05, calcule el valor predictivo positivo.

Solución Utilizando la definición del valor predictivo positivo dado en la ecuación 3.8 y la regla de Bayes obtenemos

$$P(D \mid +) = \frac{P(+ \mid D)\,P(D)}{P(+ \mid D)\,P(D) + P(+ \mid \bar{D})\,P(\bar{D})}$$

que puede reescribirse como

$$PPV = \frac{\text{(Sensibilidad) (Prevalencia)}}{\text{(Sensibilidad) (Prevalencia)} + (1 - \text{Especificidad})\,(1 - \text{Prevalencia})}. \tag{3.14}$$

Los términos en el numerador y en la parte izquierda del denominador de esta expresión se deducen directamente de las definiciones de sensibilidad y prevalencia. Para comprender el término en la parte derecha del denominador, observe que uno menos la especificidad es igual a

$$1 - \frac{P(-\bar{D})}{P(\bar{D})} = \frac{P(+\bar{D}) + P(-\bar{D})}{P(\bar{D})} - \frac{P(-\bar{D})}{P(\bar{D})} = \frac{P(+\bar{D})}{P(\bar{D})} = P(+ \mid \bar{D})$$

y que

$$P(\bar{D}) = 1 - P(D) = 1 - \text{Prevalencia}.$$

Sustituyendo los valores apropiados en la ecuación 3.14, tenemos que

$$PPV = \frac{(.83)(.05)}{(.83)(.05) + (1 - .89)\,(1 - .05)} = .28. \qquad \blacksquare$$

Como podrá ver, mientras que la sensibilidad y la especificidad para esta prueba parecen adecuadas, el valor predictivo positivo no es muy alto.

EJEMPLO 3.8

Utilice la información en el ejemplo 3.7 para calcular el valor predictivo negativo.

Solución Por medio de la regla de Bayes, el valor predictivo negativo se expresa como

$$P\ (\bar{D}\ |\ -) = \frac{P\ (-\ |\ \bar{D})\ P\ (\bar{D})}{P\ (-\ |\ \bar{D})\ P\ (\bar{D}) + P\ (-\ |\ D)\ P\ (D)}$$

que puede escribirse como

$$NPV = \frac{\text{(Especificidad) (Prevalencia)}}{\text{(Especificidad) (1 − Prevalencia) + (1 − Sensibilidad) (Prevalencia).}} \tag{3.15}$$

Sustituyendo en esta ecuación, tenemos que

$$NPV = \frac{(.89)\ (1 - .5)}{(.89)\ (1 - .5) + (1 - .83)\ (.5)} = .99$$

■

3.4 LA CURVA NORMAL

3.4.1 Muestreo de la población

En la sección 3.3 planteamos y contestamos preguntas relacionadas con la probabilidad de seleccionar una sola muestra de una población, observando las características dicótomas asociadas con esa muestra y encontrando que esas características satisfacían algunos criterios ya especificados. Las soluciones a estas preguntas se obtuvieron al observar la proporción de características en la población que cumplían con el criterio de interés.

Los problemas a tratar en esta sección son similares, pero difieren en dos aspectos importantes. Primero, la característica a observar es continua,[6] en lugar de dicotómica, y segundo, supondremos que las proporciones de las observaciones en la población necesarias para calcular las probabilidades deseadas no están disponibles, por lo cual se necesita usar un modelo matemático para calcular las soluciones. La figura 3.3 describe una población de este tipo que consta de presiones arteriales sistólicas. La notación “...” indica que el resto de la población consta de un número desconocido de tipos de valores similares.

El modelo que usaremos para calcular las probabilidades asociadas con esta población es la conocida curva normal. Antes de demostrar cómo se utiliza la curva para este propósito, describiremos brevemente algunas de sus características.

3.4.2 Algunas características de la curva normal

La **curva normal** es una función matemática utilizada por lo general como un modelo de la realidad, cuando esa realidad no puede ser tratada de forma directa. En secciones posteriores discutiremos y demostraremos el uso de la curva normal como un modelo. Entre tanto, sólo describiremos algunas de sus características más notables.

[6] Vea el comentario sobre variables continuas y dicotómicas en el apartado 2.2.5 de la página 11.

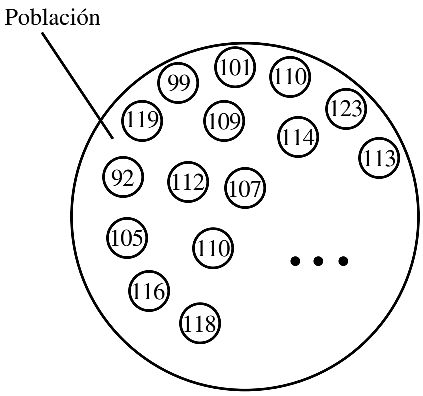


FIGURA 3.3: Población de un número desconocido de presiones arteriales sistólicas.

La forma funcional de la curva normal está dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{3.16}$$

donde *e* es una constante aproximadamente igual a 2.7183, μ y σ son constantes que determinan, respectivamente, la media y la desviación estándar de la distribución, y *x* es la variable cuya función debe determinarse.

La conocida "curva de campana" se logra al igualar μ y σ a las constantes deseadas y al evaluar 3.16 para los valores contiguos apropiados de *x*. En la figura 3.4 se muestran gráficas de este tipo para los valores seleccionados de μ y σ.

Note que la parte superior de la figura 3.4 describe dos curvas normales con valores idénticos de σ, pero valores diferentes de μ. El efecto de tener valores medios diferentes es localizar las distribuciones en puntos diferentes de la línea numérica. Por esta razón, μ se conoce a veces con el nombre de **parámetro de localización**.

En contraste, la parte inferior de la figura 3.4 muestra dos distribuciones normales con valores comunes de μ, pero valores diferentes de σ. Puesto que comparten una media común, se localizan en el mismo punto de la línea numérica, pero difieren en el grado de dispersión porque una tiene una σ más grande que la otra. En este contexto, σ se conoce como **parámetro de escala**.

En virtud de que las curvas normales pueden diferir en localización y escala, es quizá más apropiado pensar en una *familia* de curvas normales en vez de *la* curva normal. Usaremos esta flexibilidad en una sección posterior para obtener los cálculos de probabilidad. Otras características de las curvas normales son:

1. Tanto la media, la mediana y la moda se localizan en el centro de la distribución.

2. La distribución es simétrica con respecto a su media, mediana y moda.

3. Se definen para todos los valores de *x* entre $-\infty$ y ∞. Esto significa que las descripciones como las que aparecen en la figura 3.4 muestran sólo un segmento de la curva, ya que ésta se extiende infinitamente en cualquier dirección.

4. El área comprendida por la curva es igual a uno, sin importar los valores de μ y σ.

3.4.3 Cómo calcular áreas bajo la curva normal

Antes de utilizar la curva normal para calcular probabilidades, usted debe aprender a calcular áreas bajo la curva. Para lograrlo, debe aprender a utilizar una tabla de curva normal. Usted aprenderá ambas técnicas en este apartado.

Para comenzar, suponga que, por alguna razón, desea descubrir la proporción del área total de la curva normal que cae entre algún punto del eje *x* y la media de la curva. Esta área se representa con el sombreado gris claro en la figura 3.5.[7] Por inspección, usted podría pensar que esta área representa alrededor del 20% o .20 del área total de la curva. Por otra parte, quizá desee calcular el área representada por el sombreado más oscuro. Nuevamente, podría calcular esta área por inspección visual. Se pueden obtener respuestas precisas utilizando la tabla de la curva normal en el apéndice A. En esta tabla, la columna 1 nos da varios puntos a lo largo del eje *x*, la columna 2 nos da las áreas entre estos puntos y la media de la curva, y la columna 3 nos muestra el área en la cola, representada con el sombreado oscuro. Ponga atención en que las áreas en las columnas 2 y 3 suman .5, ya que juntas representan la mitad de la curva. El siguiente ejemplo sirve para demostrar cómo se utiliza la tabla.

Dada una curva normal con una media (μ) igual a 100 y una desviación estándar (σ) igual a 5, calcule la proporción de la curva que se encuentra entre 100 y 105. El área a determinar se representa por el sombreado gris en la figura 3.6. Se puede usar la tabla de la curva normal para responder esta pregunta, ya que nos da las áreas entre puntos del eje *x*, por ejemplo, 105, y la media de la curva, que en este caso es 100. Sin embargo, no podemos buscar el punto 105 en la columna 1 de la tabla porque es un valor dependiente de la escala. Obviamente, la tabla no puede manejar unidades en pulgadas, pies, libras, puntajes de cociente intelectual (CI) y todos los demás valores que pudieran interesarnos. En vez de ello, la columna 1 expresa puntos a lo largo del eje *x* en términos de puntuaciones *Z*. Usted recordará de la página 42[8] que las puntuaciones *Z* indican el número de desviaciones estándar de distancia a las que se encuentra un punto en el eje *x* desde la media de la distribución.

Aplicar la ecuación 2.24 da como resultado $Z = \frac{105 - 100}{5} = 1.00$. Así, 105 tiene una puntuación *Z* de 1.00, lo que significa que 105 está a una desviación estándar por arriba de la media de la distribución. Si localizamos 1.00 en la columna 1 del Apéndice A y leemos el valor adyacente en la columna 2, encontramos que el área entre 105 y 100 constituye .3413 del área total de la curva. Observe que .3413 de la curva normal se extiende entre la media y un punto que está a una desviación estándar de la media, sin importar qué valores tomen la media y la desviación estándar. Así, si se construye una curva normal con una media de 10 y una desviación estándar de 2, el .3413 de la curva cae entre 10 y 12. Si el problema original hubiera sido encontrar el área *por arriba de* 105, hubiéramos procedido de la misma manera, excepto que la respuesta, .1587, se habría encontrado en la columna 3.

EJEMPLO 3.9

Dada una curva normal con una media de 250 y una desviación estándar de 25, ¿qué porción de la curva cae *por debajo* de 220?

Solución Por lo regular es útil esbozar el problema como lo hicimos en la parte A de la figura 3.7. Observe que la puntuación *Z* asociada con 220 es $Z = \frac{220 - 250}{25} = -1.20$. Esto

[7] Como el área total por debajo de la curva es igual a 1, usted puede considerar que la porción sombreada representa una área o una proporción del área total.

[8] Quizá quiera repasar esta sección antes de continuar.

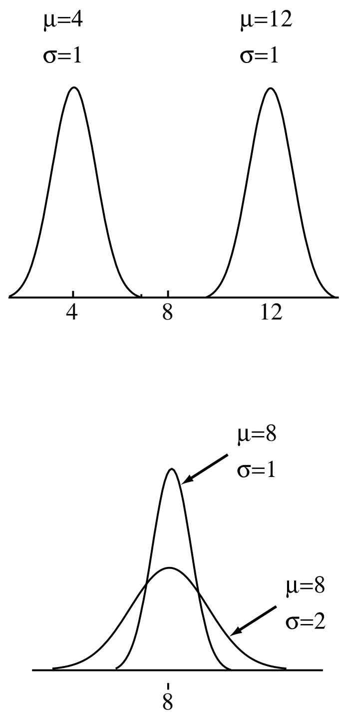


FIGURA 3.4: Distribuciones normales con valores seleccionados de μ y σ.

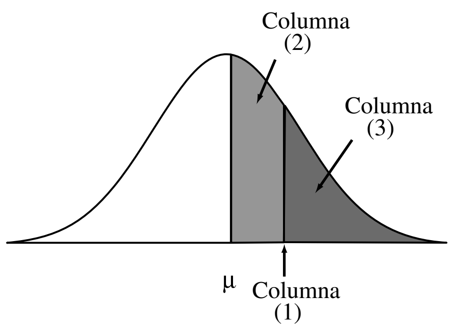


FIGURA 3.5: Áreas de la curva normal dadas en el Apéndice A.

indica que 220 está a 1.2 desviaciones estándar por *debajo* de la media de la distribución. Aunque el Apéndice A no nos da puntuaciones Z negativas, se puede utilizar la simetría de la curva normal para encontrar el área en cuestión. Es decir, las áreas asociadas con una puntuación Z de −1.20 son las mismas que las asociadas con una puntuación Z de 1.20.

El Apéndice A muestra que el área en la cola de la curva asociada con una puntuación Z de 1.20 es .1151, que es la respuesta a la pregunta planteada. Para ayudarle a familiarizarse con la curva, también hemos etiquetado las otras áreas de la curva. ■

EJEMPLO 3.10

Dada una curva normal con una media de 80 y una desviación estándar de 10, calcule el área entre 65 y 85.

Solución Hemos representado el problema en el panel B de la figura 3.7. Fíjese en que el área sombreada no se puede leer de forma directa en la tabla porque ésta no nos da áreas entre dos puntos arbitrarios cualesquiera, sino entre un punto y una media o el área de la cola, como se describió previamente. Aún así, usted puede obtener el área requerida si calcula cada una de las dos áreas componentes y las suma.

La puntuación Z para 65 es $Z = \frac{65 - 80}{10} = -1.50$. La columna 2 muestra que el área entre 65 y 80 es .4332. De igual manera, la puntuación Z para 85 es $Z = \frac{85 - 80}{10} = .50$, que tiene una área de .1915. El área entre 65 y 85 es, entonces, .4332 + .1915 = .6247. ■

EJEMPLO 3.11

Dada una curva normal con una media de 500 y una desviación estándar de 50, calcule el área entre 555 y 600.

Solución Como en el problema anterior, la solución no puede leerse directamente de la tabla. Sin embargo, note que la puntuación Z para 600 es $Z = \frac{600 - 500}{50} = 2.00$. La columna 2 muestra que el área entre 600 y 500 es .4772. Pero ésta no es la respuesta que buscamos, porque el área entre 500 y 555 está incluida y no forma parte del área a calcular. Ahora bien, podemos encontrar la puntuación Z para 555, que es 1.10, y el área (no deseada) entre 500 y 555, que es .3643. El área entre 600 y 555 es, entonces, .4772 − .3643 = .1129. El problema se muestra en el panel C de la figura 3.7.

EJEMPLO 3.12

Dada una curva normal con una media de .05 y una desviación estándar de .01, calcule el área por debajo de .0722.

Solución El área a localizar está descrita en el panel D de la figura 3.7. Como se observa en el bosquejo, el área consta de dos partes. La primera es el área debajo de .05. Como .05 es la media de la distribución, tiene una puntuación Z de .00, que consta de una área asociada en la cola de .5. La segunda área está entre .0722 y .0500. La puntuación Z para .0722 es $Z = \frac{.0722 - .0500}{.01} = 2.22$. La columna 2 muestra que el área entre .0722 y .0500 es .4868. El área debajo de .0722 es, entonces, .5000 + .4868 = .9868. ■

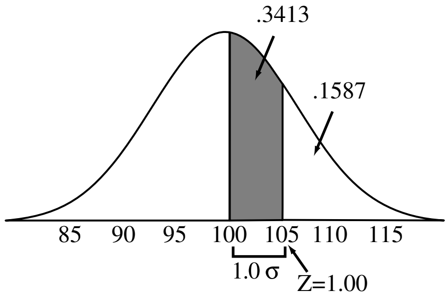


FIGURA 3.6: Cálculo de una área bajo la curva normal.

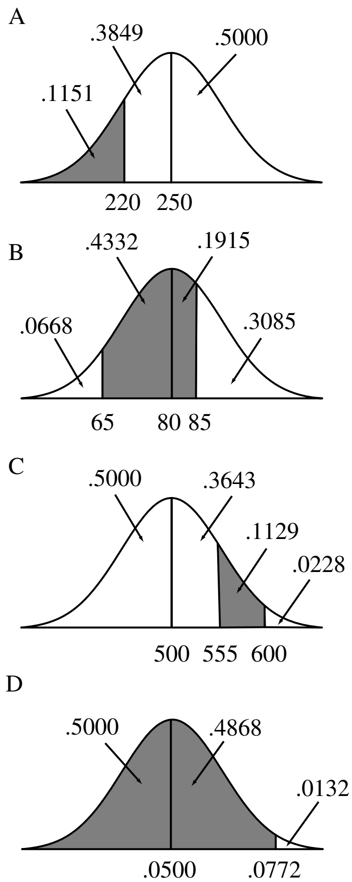


FIGURA 3.7: Cálculo de áreas bajo la curva normal.

TABLA 3.3: Distribución de frecuencias relativas de la población.

Presión arterial	Frecuencia relativa	Presión arterial	Frecuencia relativa
95	.004	110	.078
96	.000	111	.078
97	.003	112	.070
98	.004	113	.067
99	.008	114	.059
100	.009	115	.049
101	.013	116	.035
102	.025	117	.039
103	.031	118	.024
104	.033	119	.014
105	.050	120	.011
106	.057	121	.002
107	.066	122	.004
108	.074	123	.005
109	.085	124	.002
		125	.001

3.4.4 Cómo utilizar la curva normal para aproximar probabilidades

¿Cuál es la probabilidad de escoger una muestra de la población representada por la figura 3.3 de la página 63, y encontrar que la presión arterial sistólica así obtenida es 111? Para responder esta pregunta usted tendría que saber la proporción de presiones arteriales en la población que son de 111. La tabla 3.3 muestra la distribución de frecuencias relativas de las presiones arteriales de esta población. Como muestra esta tabla, la proporción, y por ende la probabilidad, asociada con 111 es .078. ¿Cuál es la probabilidad de que la observación esté entre 112 y 114 (es decir, 112, 113 o 114)? Otra vez, usted puede usar la información de la tabla 3.3 para calcular la probabilidad como .070 + .067 + .059 = .196.

Ahora, suponga que las proporciones usadas en los cálculos anteriores no están disponibles. En caso de que (1) se conozcan la media y la desviación estándar de la población,[9] y (2) la distribución de frecuencias relativas de la población tenga una forma *relativamente* normal, se puede utilizar la curva normal para aproximar las probabilidades deseadas. Es decir, la curva normal se emplea como un modelo de la distribución de frecuencias relativas de la población.

En el presente caso la media y la desviación estándar de la población son 110.023 y 4.970, respectivamente. La figura 3.8 muestra una curva normal con esta media y esta desviación estándar impuestas sobre la distribución de frecuencias relativas de la población. Como se aprecia, se cumple con el requerimiento número 2 antes mencionado.

La aproximación se obtiene al calcular el área debajo de la curva que corresponde al evento que nos ocupa. Por ejemplo, para estimar la probabilidad de que la presión seleccionada sea 111,

[9] En el capítulo 4 veremos cómo se pueden conocer estas cantidades.

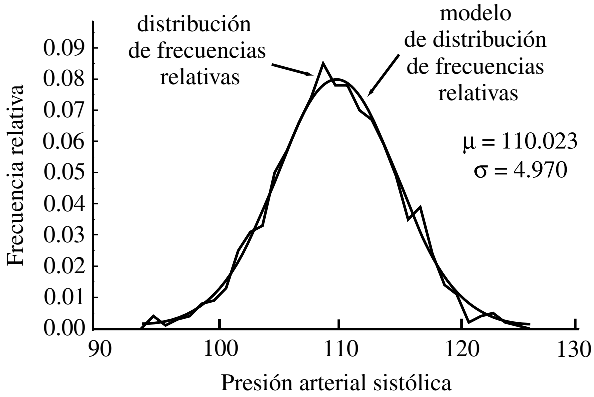


FIGURA 3.8: Comparación de una curva normal con la distribución de frecuencias relativas de la población.

usted calculará el área debajo de la curva normal con una media de 110.023 y una desviación estándar de 4.970 que corresponde a 111. El área correspondiente a 111 es el área entre el límite real inferior de 110.5 y el límite real superior de 111.5.[10] La puntuación Z para 111.5 es (aproximadamente) $Z = \frac{111.5 - 110.023}{4.970} = .30$, a la cual corresponde una área de .1179. La puntuación Z para 110.5 es (aproximadamente) $Z = \frac{110.5 - 110.023}{4.970} = .10$ y le corresponde una área de .0398. El área entre 111.5 y 110.5 es, entonces, .1179 – .0398 = .0781, que se acerca mucho al valor exacto de .078. El problema está esbozado en el panel A de la figura 3.9.

Al encontrar el área entre 99.5 y 105.5 se obtiene un estimado de la probabilidad de que la observación seleccionada aleatoriamente esté entre 100 y 105 (inclusive). La puntuación Z para 99.5 es (aproximadamente) $Z = \frac{99.5 - 110.023}{4.970} = -2.12$ con una área correspondiente de .4830. La puntuación Z y el área asociada para 105.5 son, respectivamente, –.91 y 0.3186. La probabilidad aproximada es, entonces, .4830 – .3186 = .1644, que puede compararse con el valor de .009 + .013 + .025 + .031 + .033 + .050 = .161, obtenido de la tabla 3.3. (Véase el panel B de la figura 3.9.)

EJEMPLO 3.13

Calcule la probabilidad de que la muestra aleatoria discutida con anterioridad sea mayor que 103. Compare este estimado con el valor exacto calculado de la tabla 3.3.

Solución El estimado se obtiene al calcular el área bajo la curva normal usada previamente que está por arriba de 103.5. (Véase el panel C de la figura 3.9). Para calcular esta área, tome en cuenta que la puntuación Z para 103.5 es (aproximadamente) $Z = \frac{103.5 - 110.023}{4.970} = -1.31$, que tiene una área asociada de .4049. Como ésta es el área entre 103.5 y 110.023, y el área por encima de 110.023 es .5000, el estimado está dado por .4049 + .5000 = .9049. El resultado exacto obtenido de la tabla 3.3 es la suma de las probabilidades asociadas con 104 y valores mayores, es decir, .903. ■

[10] Tal vez quiera repasar el apartado 2.5.2 de la página 20 si es que ha olvidado la lógica en los conceptos de los límites reales superior e inferior.

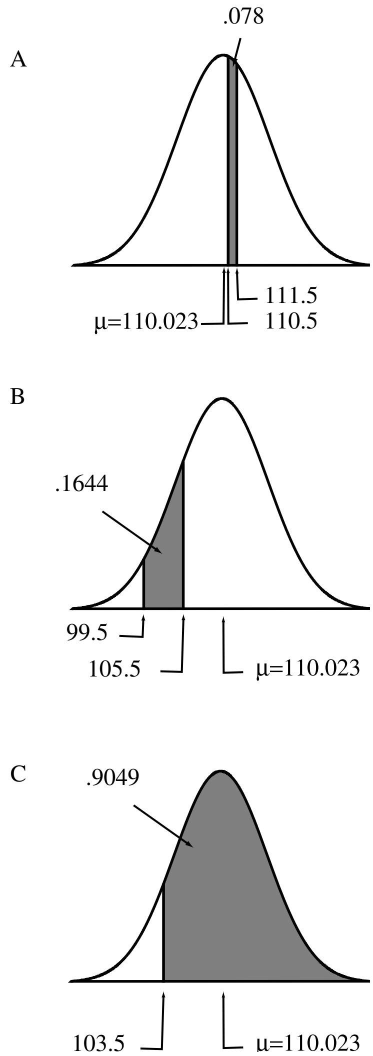


FIGURA 3.9: Uso de la curva normal para aproximar probabilidades.

No debe concluirse que la curva normal siempre ofrece aproximaciones tan buenas como las aquí presentadas. Si la población está sesgada o difiere considerablemente de la curva normal de alguna otra manera, tal vez los cálculos obtenidos no se acerquen mucho a los valores exactos. Sin embargo, en el capítulo 4 aprenderá que la curva normal puede servir para obtener buenos estimados de probabilidad cuando las poblaciones no están distribuidas normalmente, toda vez que los cálculos impliquen ciertos estadísticos y no puntajes individuales como en los casos presentados aquí.

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

área de la curva normal, 64
curva normal, 62
efecto protector, 59
especificidad, 58
independencia, 56
parámetro de escala, 63
parámetro de localización, 63
prevalencia, 58
probabilidad, 52
probabilidad condicional, 53
probabilidad conjunta, 56
probabilidad marginal, 56
razón de probabilidad, 60
regla de Bayes, 60
sensibilidad, 58
tabla de contingencia, 53
tabla de frecuencias, 53
tabla de probabilidad, 55
tasa de riesgo, 59
valor predictivo negativo, 58
valor predictivo positivo, 58

EJERCICIOS

3.1 La siguiente tabla de probabilidad presenta los resultados de un censo hecho a todos los estudiantes de una universidad grande. Cada estudiante fue clasificado por género (*H/M*) y por haber contestado, o no, afirmativamente a la pregunta de si habían estado ebrios en los últimos 30 días ($I/\bar{I}$).

	I	$\bar{I}$	
H	.22	.32	.54
M	.10	.36	.46
	.32	.68	

Utilice la notación de probabilidad de este capítulo para calificar lo siguiente. Después utilice las entradas en la tabla para encontrar la probabilidad indicada.

a) La probabilidad de que una estudiante, seleccionada al azar entre las mujeres, no haya estado ebria en los últimos 30 días.

b) La probabilidad de que un estudiante, seleccionado al azar, sea mujer.

c) La probabilidad de que un estudiante, seleccionado al azar, sea hombre o haya estado ebrio en los últimos 30 días.

d) La probabilidad de que un estudiante, seleccionado al azar, sea mujer y haya estado ebria en los últimos 30 días.

e) La probabilidad de que un estudiante, seleccionado al azar entre aquellos que reportan haber estado ebrios en los últimos 30 días, sea mujer.

f) La probabilidad de que un estudiante, seleccionado al azar, sea hombre o mujer.

3.2 ¿Son independientes las categorías género y embriaguez, como se representa en la tabla anterior? ¿Cuál es la evidencia para su conclusión?

3.3 Suponga que la mitad de los residentes de una comunidad son mujeres y que el 20% de los residentes en esa comunidad están de acuerdo en un incremento a los impuestos, con el objetivo de proveer fondos para vacunar gratuitamente a los niños. Si el 10% de los residentes de la comunidad son mujeres y partidarias del incremento, ¿puede decirse que el género y el apoyo al incremento son eventos independientes? ¿Cuál es la evidencia para sustentar su respuesta?

3.4 Utilice la información en el ejercicio 3.3 y calcule $P(\overline{T} \mid M)$, en donde T indica el apoyo al incremento a los impuestos y M representa a un residente hombre.

3.5 Suponga que para los eventos A, $\overline{A}$, B, $\overline{B}$, $P(\overline{B} \mid A) = .22$, $P(\overline{A}) = .20$ y $P(B) = .72$. Utilice esta información para construir una tabla de probabilidad que muestre todas las probabilidades marginales y conjuntas.

3.6 ¿Qué representan cada uno de los siguientes incisos?

a) La proporción de personas que dan positivo para alguna enfermedad y quienes, en verdad, tienen la enfermedad.
b) La proporción de personas con alguna enfermedad y que dan positivo para la enfermedad.
c) La proporción de personas que no tienen ninguna enfermedad o son sanas y tienen un resultado negativo para la prueba de enfermedad.
d) La proporción de personas que dan negativo a una prueba para alguna enfermedad y no están enfermas.

3.7 Utilice la siguiente tabla para calcular lo que se pide.

	D	$\overline{D}$	
+	.065	.010	.075
−	.005	.920	.925
	.070	.930	

a) sensibilidad
b) especificidad
c) valor predictivo positivo
d) valor predictivo negativo

3.8 Utilice la siguiente tabla para calcular e interpretar la tasa de riesgo y la razón de probabilidad.

	E	$\overline{E}$	
D	.18	.07	.25
$\overline{D}$	.02	.73	.75
	.20	.80	

3.9 Suponga que en cierta comunidad, en la que el 40% de la población tiene menos de 40 años, se descubre que una proporción de .72 de residentes de menos de 40 años apoya la vacunación obligatoria de los niños en edad escolar contra ciertas enfermedades. La proporción de residentes mayores de 40 años que apoya la propuesta es de .52. Utilice esta información para calcular la proporción de partidarios de la inoculación que tienen menos de 40 años.

3.10 Dada una sensibilidad de .82, una especificidad de .93 y un prevalencia de .20, calcule el valor predictivo positivo.

3.11 Dada una variable distribuida normalmente, con una media de 80 y una desviación estándar de 12, calcule las siguientes áreas de la curva:

a) El área entre 80 y 98.
b) El área por debajo de 74.
c) El área por debajo de 82.
d) El área entre 72 y 94.
e) El área entre 56 y 60.
f) El área por arriba de 104.
g) El área por debajo de 54.
h) El área entre 82 y 94.

3.12 Dada una variable con valor entero, que tiene una distribución aproximadamente normal, con una media de 100 y una desviación estándar de 8, utilice la curva normal para aproximar la probabilidad de seleccionar aleatoriamente una sola observación de la población y encontrar que cumple con cada uno de los siguientes criterios.

a) Es igual a 102.
b) Tiene un valor entre 92 y 108.
c) Tiene un valor mayor que 110.
d) Tiene un valor menor que 105.

A. Las siguientes preguntas se refieren al estudio de caso A (página 469).

3.13 ¿Cuál es la probabilidad de seleccionar aleatoriamente a un participante de este estudio y descubrir que el seleccionado es hombre?

3.14 Convierta los datos am (mañana) en una tabla de frecuencias y en una tabla de contingencia de probabilidad.

3.15 Suponiendo que T representa “escogió el lente tratado”, U “escogió el lente no tratado”, Y “respondió **sí** a la pregunta sobre duración”, y N “respondió **no** a la pregunta sobre duración", utilice la tabla de probabilidad del ejercicio 3.14 para calcular lo siguiente:

a) $P(T)$
b) $P(Y)$
c) $P(T / Y)$
d) $P(T / N)$
e) Exprese cada una de las probabilidades anteriores en forma de enunciado.
f) ¿T y Y son independientes? Dé evidencia para su respuesta.

Continúe usando los datos de la tabla del ejercicio 3.14 para los ejercicios 3.16 al 3.20.

3.16 ¿Cuál es el riesgo de seleccionar "el lente no tratado" para las personas que respondieron **sí** a la pregunta sobre duración? ¿Para las personas que respondieron **no** a la pregunta sobre duración?

3.17 Para el grupo de la mañana, calcule una tasa de riesgo que compare el riesgo de seleccionar "el lente no-tratado" para las personas del grupo que respondió **no**, en relación con el riesgo para las personas del grupo que contestó afirmativamente. ¿Qué implicaciones tiene este cociente para el estudio?

3.18 Para el grupo de la mañana, calcule una tasa de riesgo que compare el riesgo de seleccionar "el lente tratado" para las personas del grupo que respondió **no,** con el riesgo para las personas del grupo que contestó **sí**. ¿Las implicaciones de este cociente son las mismas que las calculadas en el ejercicio 3.17?

3.19 Para el grupo de la mañana, ¿cuáles son las posibilidades de seleccionar "el lente no tratado" para las personas que respondieron **sí** a la pregunta sobre duración? ¿Para las personas que contestaron **no** a la pregunta sobre duración?

3.20 Para el grupo de la mañana, calcule una razón de probabilidad que compare las posibilidades de seleccionar el "lente no tratado" para las personas del grupo que respondió **no**, con las posibilidades para las personas del grupo que contestó **sí**. ¿Por qué piensa usted que los investigadores que condujeron este estudio se interesarían en este cociente?

E. Las siguientes preguntas se refieren al estudio de caso E (página 472).

3.21 Calcule la sensibilidad, especificidad y los valores predictivos positivo y negativo para la nueva prueba.

O. Las siguientes preguntas se refieren al estudio de caso O (página 477).

3.22 Comente el inciso *d*).

P. Las siguientes preguntas se refieren al estudio de caso P (página 478).

3.23 Usted ha aprendido a calcular probabilidades relacionadas con dos variables dicotómicas presentadas en una tabla de contingencia de dos por dos. Los siguientes puntos están diseñados para evaluar su nivel de comprensión de tales probabilidades al aplicar sus conocimientos a una situación que conlleva *tres* variables, una de las cuales no es dicotómica.

Remítase a la tabla J.10 para los siguientes puntos y tome en cuenta que H = hombre, M = mujer, I = diabetes tipo I, II = diabetes tipo II, y (a – b) = rango de edades de *a* a *b*.

a) Escriba un enunciado para expresar $P\,(F\,I\,(25-44))$.

b) Calcule $P\,(\,F\,I\,(25-\ 44))$.

c) Calcule $P\,(\,F \cup I \cup (25-\ 44))$.

d) Escriba un enunciado para expresar $P\,(\,M \mid (15-24 \cup I\,I)$.

e) Calcule $P\,(\,M \mid (15-24 \cup I\,I)$.

CAPÍTULO 4

Introducción a los métodos de inferencia y de una muestra

4.1 INTRODUCCIÓN

En el capítulo 1 presentamos brevemente el concepto de inferencia e indicamos su papel fundamental en estadística. En este capítulo usted aprenderá a aplicar ciertos métodos inferenciales, así como el fundamento de su uso. Los métodos presentados aquí no se emplean muy a menudo en la investigación, pero tienen importancia pedagógica porque son relativamente simples y su dominio le abrirá el camino para comprender los métodos más complejos expuestos en los siguientes capítulos.

Las técnicas que aprenderá se dividen en dos categorías amplias: **1.** pruebas de hipótesis y **2.** intervalos de confianza. Sin embargo, antes de comenzar a estudiar tales técnicas, es necesario dejar claro el concepto de distribuciones muestrales, pues es el fundamento de las dos clases de métodos de inferencia.

Debemos advertir que éste es el capítulo más importante del libro. El material que aquí se presenta es esencial para todos los posteriores. Con el tiempo invertido aquí obtendrá grandes beneficios después. Ahora que hemos captado su atención, nos ocuparemos del tema de distribuciones muestrales.

4.2 DISTRIBUCIONES MUESTRALES

4.2.1 Definición

El concepto de distribuciones muestrales, aunque es un tanto abstracto, es relativamente simple. Una **distribución muestral** es una distribución de muestras estadísticas que se obtienen, por lo regular, de una o más poblaciones. Por ejemplo, suponga que usted escogerá al azar una muestra formada por cinco observaciones de la población representada en la figura 3.3. Luego, usted calcula y registra la media de la muestra. Ahora suponga que usted repetirá este proceso varias (en realidad, un número infinito de) veces. Si después usted ordenara esas medias de las muestras en una distribución de frecuencias relativas podría generar así la distribución muestral de la media o $\bar{x}$. Usted podrá, por supuesto, utilizar este método para generar la distribución muestral de otros estadísticos, como la desviación estándar de la muestra, la mediana, el percentil quinto o cualquiera de los muchos estadísticos.

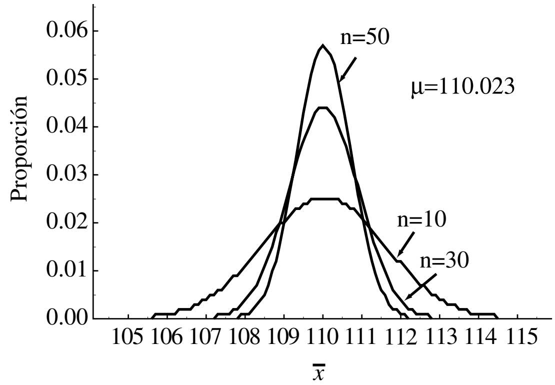


FIGURA 4.1: Distribuciones muestrales de $\bar{x}$.

4.2.2 Distribución muestral de $\bar{x}$

La figura 4.1 muestra tres distribuciones muestrales de $\bar{x}$ derivadas de la población cuya distribución de frecuencias relativas se presenta en la figura 3.8, en la página 69. El método por el cual nosotros generamos esas distribuciones le ayudará a comprender el concepto general de las distribuciones muestrales.

Con el fin de construir las distribuciones muestrales mostradas en la figura 4.1 creamos un programa computacional que (1) genera la población antes mencionada; (2) selecciona aleatoriamente 10,000,000[1] de muestras de tamaño 10 de esta población y calcula la media de cada muestra; (3) calcula la media, la varianza y la desviación estándar de las 10,000,000 medias y, finalmente, (4) genera la distribución de frecuencias relativas de las medias. Después, este proceso se repitió utilizando muestras de tamaño 30 y 50.

Muchos factores relacionados con estas distribuciones muestrales deberán observarse. Primero, la media de las tres distribuciones es 110.023, la cual es la media de la población como se reportó en el apartado 3.4.4. La media de una distribución muestral se conoce como el **valor esperado** del estadístico y se simboliza con $E[]$ donde [] contiene un identificador del estadístico. En el caso presente, el valor esperado de $\bar{x}$ está representado por $E[\bar{x}]$. En general, $E[\bar{x}] = \mu$. Usted no deberá concluir con esto que el valor esperado de todos los estadísticos es igual a su parámetro equivalente. Por ejemplo, $E[s] \neq \sigma$.

Segundo, usted notará que la desviación estándar de las tres distribuciones muestrales disminuye mientras el tamaño de la muestra se incrementa. La desviación estándar de las distribuciones muestrales generalmente se conoce como el **error estándar** de un estadístico particular. De tal manera, la desviación estándar de la distribución muestral de $\bar{x}$ se conoce como el **error estándar de la media,** y se simboliza como $\sigma_{\bar{x}}$. El error estándar de la media se expresa como

$$\boxed{\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}} \tag{4.1}$$

donde σ es la desviación estándar de la población y n es el tamaño de la muestra. Para el caso donde n es igual a 10 en la figura 4.1, el error estándar de la media es $\frac{4.970}{\sqrt{10}} = 1.572$. Usted puede ver

[1] Esto fue extenuante, pero quisimos enfatizar el hecho de que las distribuciones muestrales están basadas en un gran número de muestras.

en la ecuación 4.1 que para una desviación estándar de una población fija, el error estándar de la media disminuye mientras el tamaño de la muestra se incrementa, como se aprecia en la figura.

Dado que la varianza es el cuadrado de la desviación estándar, se infiere que la varianza de la distribución muestral de $\bar{x}$ es

$$\boxed{\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}} \tag{4.2}$$

Finalmente, usted observará que las tres distribuciones muestrales parecen tener formas casi normales, así que serán bien modeladas con curvas normales. Mientras estos modelos tengan medias comunes, necesariamente tendrán diferentes desviaciones estándar. Quizá no le sorprenda que las distribuciones muestrales de $\bar{x}$ tengan formas normales en aproximación, dado que la población de la cual se derivan también es relativamente normal, pero cuando la población fuente no es aproximadamente normal comienzan a surgir las preguntas referentes a la distribución muestral.

El panel superior de la figura 4.2 indica con contundencia una distribución de población no normal, mientras que el panel inferior presenta tres distribuciones muestrales de medias tomadas de la misma población. Como se observa, las distribuciones muestrales son casi normales a pesar de la forma no normal de la población fuente. Este fenómeno es atribuido al **teorema del límite central**.

A grandes rasgos, el **teorema del límite central** enuncia que las distribuciones muestrales de ciertas clases de estadísticos se aproximarán a la normalidad, a medida que el tamaño de la muestra (n) se incrementa independientemente de la forma de la población muestreada. Esto significa que para tamaños muy pequeños de muestras, las distribuciones muestrales de estos estadísticos podrán no ser muy normales, pero mientras el tamaño de la muestra se incremente, la forma de las distribuciones muestrales se acercará más a una curva normal. Este resultado es sumamente importante porque implica que muchas distribuciones muestrales pueden ser modeladas por la curva normal, aun cuando la población fuente no lo sea del todo. La pregunta que surge entonces es qué tan grande debe ser n para que la curva normal sea un modelo apropiado para una distribución muestral en particular. Eso depende de muchos factores, incluyendo la forma de la población muestreada y la naturaleza del estadístico cuya distribución muestral está en cuestión. Como sugiere la figura 4.2, el teorema del límite central generalmente produce distribuciones normales en aproximación para $\bar{x}$, aun cuando los tamaños de la muestra sean pequeños.

EJEMPLO 4.1

Encuentre las desviaciones estándar de las tres distribuciones muestrales expresadas en la figura 4.2.

Solución Utilizando la ecuación 4.1 y la información de la figura 4.2, los errores estándar de la media para $n = 10$, 30 y 50 son, respectivamente, $\frac{5.293}{\sqrt{10}} = 1.67$, $\frac{5.293}{\sqrt{30}} = .97$ y $\frac{5.293}{\sqrt{50}} = .75$. ■

4.2.3 Uso de la curva normal para aproximación de probabilidades asociadas con $\bar{x}$

En el apartado 3.4.4 expusimos la pregunta: "¿Cuál es la probabilidad de elegir una observación de la población representada por la figura 3.3 y encontrar que la presión arterial sistólica obtenida de tal manera es 111?". Usted aprendió que esta probabilidad podría darse aproximadamente

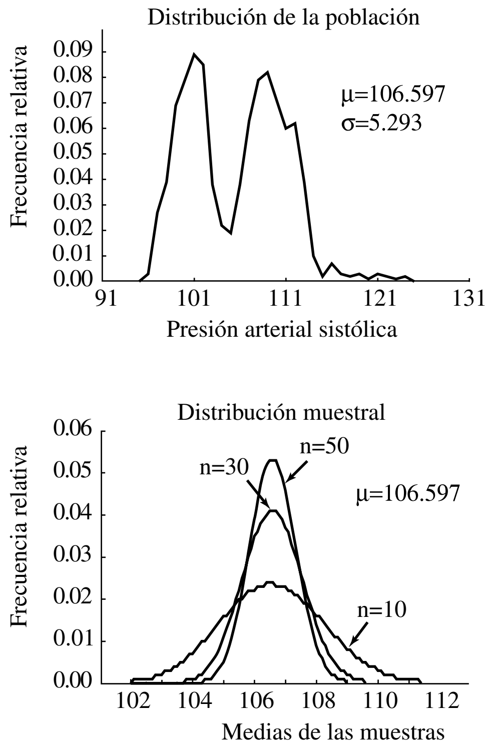


FIGURA 4.2: Población no normal con distribuciones muestrales para $\bar{x}$.

mediante la búsqueda del área apropiada bajo un modelo de la curva normal, siempre y cuando tal modelo tenga las mismas media y desviación estándar de la población.

En esta sección usted aprenderá el uso del modelo de la curva normal para responder preguntas como: "¿Cuál es la probabilidad de elegir una *muestra* de la población de tamaño 15 representada por la figura 3.3 y encontrar que la *media de la muestra* es mayor que 111?" Observe que si la distribución muestral de $\bar{x}$ estuviera disponible para esta población, el valor exacto se obtendría fácilmente. Esto se infiere del hecho de que las distribuciones muestrales son distribuciones de frecuencias relativas, lo cual a la vez representa probabilidades.

Puesto que las distribuciones muestrales rara vez están disponibles (si acaso lo están), usted puede utilizar la misma técnica empleada para encontrar las probabilidades asociadas con observaciones simples. Esto es, puede encontrar el área apropiada bajo una curva normal. En este caso la probabilidad aproximada será el área por arriba de 111. De tal forma, usted sólo encontrará las puntuaciones Z para 111 y localizará el área asociada con la columna 3 de la tabla de la curva normal. Observe, sin embargo, que la puntuación Z tomará la forma

$$\boxed{Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}} \tag{4.3}$$

y no la forma de la ecuación 2.24. Como el denominador de esta expresión es la desviación estándar de la distribución muestral, Z es el número de desviaciones estándar entre $\bar{x}$ y μ.

Para el problema en cuestión,[2]

$$Z = \frac{111.0 - 110.023}{\frac{4.970}{\sqrt{15}}} = .76.$$

El área asociada es .2236. Entonces, la probabilidad de seleccionar una muestra de esa población y encontrar que la media de la muestra es mayor que 111 es, aproximadamente, .22. (Nosotros calculamos que la probabilidad real sería .2157). ¿Qué piensa usted que sucederá con esta probabilidad si el tamaño de la muestra se incrementa, por ejemplo a 2.5?

EJEMPLO 4.2

¿Cuál es la probabilidad de seleccionar aleatoriamente 25 observaciones de la población mostrada en la figura 4.2, en la página 78, y encontrar que su valor de media es menor que 107.5?

Solución La probabilidad puede estimarse para encontrar el área bajo la curva normal apropiada con la media 106.597 y una desviación estándar $5.293 / \sqrt{25}$. El área deseada será esa porción de la curva debajo de 107.5. La puntuación Z y el área asociada para la proporción entre 107.5 y 106.597 son, respectivamente,

$$Z = \frac{107.5 - 106.597}{\frac{5.293}{\sqrt{25}}} = .85$$

y .3023. El área debajo de 106.597 es .50 para que la estimación deseada sea .5000 + .3023 = .8023. (Nosotros calculamos que la probabilidad real sería .8034.) ■

EJEMPLO 4.3

Considere que se seleccionan aleatoriamente 100 observaciones de una población cuya media y desviación estándar son 100 y 20, respectivamente. ¿Cuál es la probabilidad de que la media de estas observaciones esté entre 99 y 103?

Solución El área de una curva normal con media 100 y desviación estándar $20 / \sqrt{100}$ que se encuentra entre 99 y 103 es la suma de las áreas entre 99 y 10, y 100 y 103. La puntuación Z y el área entre 99 y 100 son, respectivamente,

$$Z = \frac{99.0 - 100.0}{\frac{20.0}{\sqrt{100}}} = -.50$$

y .1915. Los mismos valores para el área entre 103 y 100 son

$$Z = \frac{103.0 - 100.0}{\frac{20.0}{\sqrt{100}}} = 1.50$$

y .4332. La probabilidad estimada es, entonces, .1915 + .4332 = .6247. ■

[2] Advierta que no empleamos límites reales superior ni inferior debido a los puntos individuales desiguales; las medias no se restringen a valores discretos.

4.2.4 Distribución muestral de $\hat{p}$

Ahora, considere una población formada por alguna característica dicotómica, como vivo-muerto, remisión-no remisión de un tumor, dolor-no dolor, etcétera. Por tradición, cuando se habla en un sentido general, uno de estos dos resultados dicotómicos se llama "éxito" y el otro "fracaso". Ahora suponga que elige aleatoriamente cinco observaciones de esta población y registra la proporción de éxitos en la muestra. Nosotros designaremos la proporción de éxito en una muestra como $\hat{p}$ y la proporción en la población como π. Si usted repitiera este procedimiento muchas veces y formara las proporciones resultantes en una distribución de frecuencias relativas, de esta forma habrá generado la distribución muestral de $\hat{p}$.

Si los miembros de una población con la característica "éxito" se designan con el número 1 y aquéllos con la característica "fracaso" con cero, entonces cada muestra estará formada por unos y ceros, y $\hat{p}$ simplemente será la media de esos valores. Por ejemplo, si una muestra está formada por dos unos y tres ceros, ambas, la media y la proporción serán 2/5 = .4. Asimismo, π será la media de los unos y los ceros que formarán la población. Puesto que en este contexto $\hat{p} = \bar{x}$ y $\pi = \mu$, no es sorprendente que $E[\hat{p}] = \pi$.

La varianza de la población se expresa como $\pi(1-\pi)$. Esto puede verse notando cuál fue la consecuencia de utilizar la fórmula computacional para la varianza de la población (ecuación 2.11, en la página 36) con datos cuyos valores se restringen a unos y ceros.

$$\begin{aligned}\frac{\sum x^2 - \frac{(\sum x)^2}{N}}{N} &= \frac{N\pi - \frac{N^2\pi^2}{N}}{N}\\ &= \frac{N\pi - N\pi^2}{N}\\ &= \pi - \pi^2\\ &= \pi\ (1-\pi).\end{aligned}$$

Elevar al cuadrado los unos y los ceros no cambia sus valores, así que $\sum x^2 = \sum x$. Dado que $\pi = \frac{\sum x}{N}$ se obtiene $\sum x^2 = N\pi$. El resto del resultado es álgebra simple.

Por analogía con la ecuación 4.1 el **error estándar de $\hat{p}$** está dado por

$$\boxed{\sigma\hat{p} = \sqrt{\frac{\pi(1-\pi)}{n}}} \tag{4.4}$$

EJEMPLO 4.4

El panel A de la figura 4.3 muestra la distribución de frecuencias relativas de una población dicotómica con $\pi = .10$, mientras que los paneles B y C muestran las distribuciones muestrales de $\hat{p}$ derivadas de esta población con los tamaños de muestra de 5 y 50, respectivamente. ¿Cuál es el error estándar de $\hat{p}$ para las dos distribuciones muestrales?

Solución Por la ecuación 4.4 el error estándar de la distribución en el panel B es

$$\sigma\hat{p} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{(.10)(.90)}{5}} = .134$$

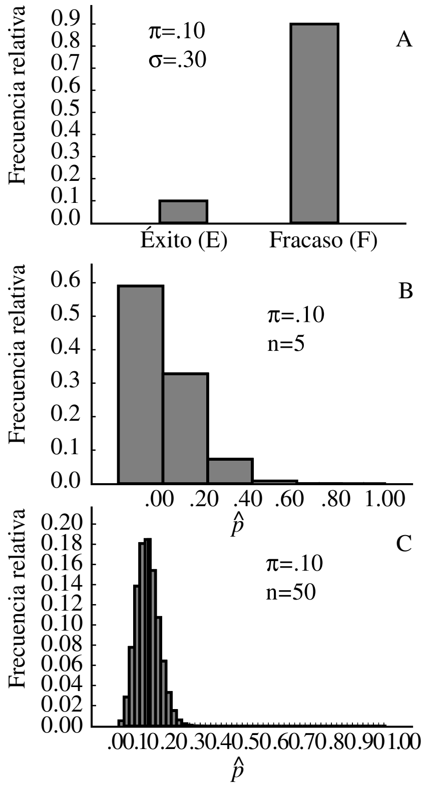


FIGURA 4.3: Población dicotómica con distribuciones muestrales de $\hat{p}$

y para la distribución del panel C

$$\sigma \hat{p} = \sqrt{\frac{(.10)(.90)}{50}} = .0.42$$

■

4.2.5 Uso de la distribución binomial para la aproximación de probabilidades asociada con $\hat{p}$

Si la población es grande y se cuenta con ciertas condiciones (que se analizarán después), la distribución binomial podrá utilizarse para modelar la distribución muestral de $\hat{p}$. La **distribución binomial** es generada por la ecuación

$$\boxed{P(y) = \frac{n!}{y!(n-y)!}\pi^y(1-\pi)^{n-y}} \tag{4.5}$$

donde n es el tamaño de la muestra, y es el número de éxitos, π es la proporción de éxitos en la población y $P(y)$ es la probabilidad de y éxitos en una muestra de tamaño n tomada de una población donde la proporción de éxitos es π. La notación $n!$ se lee como "n factorial" y se calcula como $n\,(n-1)\,(n-2)\cdots((n-(n+1))$. Por esto, 5! se calcula como $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$, y 3! será $3 \cdot 2 \cdot 1 = 6$. Por definición, $0! = 1$. Un ejemplo aclarará el uso de esta ecuación.

Suponga que usted desea encontrar la distribución muestral de $\hat{p}$ para muestras de tamaño igual a 5, extraídas de una población en la cual la proporción de éxitos es .10. Observe que sólo seis valores de $\hat{p}$ son posibles para muestras de tamaño 5. Esto es, pueden ser 0, 1, 2, 3, 4 o 5 éxitos cuyos resultados en valores de $\hat{p}$ de 0/5 = .00, 1/5 = .20, 2/5 = .40, 3/5 = .60, 4/5 = .80 y 5/5 = 1.00. Para $\hat{p} = .00$ la ecuación binomial produce

$$\begin{aligned} P(0) &= \frac{5!}{0!(5-0)!}.10^0(1-.10)^{5-0} \\ &= \frac{\cancel{5!}}{1!\cancel{5!}}.10^0.90^5 \\ &= .90^5 \\ &= .59049. \end{aligned}$$

Observe que en la línea dos los 5! se cancelan, y que tanto 0! como $.10^0$ son iguales a uno, así que el resultado se obtiene en la línea tres. Para $\hat{p} = .20$ la ecuación produce

$$\begin{aligned} P(1) &= \frac{5!}{1!(5-1)!}.10^1(1-.10)^{5-1} \\ &= \frac{5\cdot\cancel{4!}}{1!\cancel{5!}}.10^1.90^4 \\ &= (5)(.10)(.6561) \\ &= .32805. \end{aligned}$$

Usted se puede ahorrar un esfuerzo considerable notando que $5! = 5 \cdot 4!$, así que 4! puede factorizarse del numerador y el denominador de la expresión. En el cálculo de $\hat{p} = .40$ nosotros nos ahorramos un esfuerzo computacional al observar que $5! = 5 \cdot 4 \cdot 3!$, lo cual permite la cancelación de 3! del numerador y el denominador de la expresión.

$$\begin{aligned} P(2) &= \frac{5!}{2!(5-2)!}.10^2(1-.10)^{5-2} \\ &= \frac{5\cdot 4\cdot\cancel{3!}}{2!\cancel{3!}}.10^2.90^3 \\ &= (10)(.01)(.729) \\ &= .0729 \end{aligned}$$

Para $\hat{p} = .60$ y .80

$$\begin{aligned} P(3) &= \frac{5!}{3!(5-3)!}.10^3(1-.10)^{5-3} \\ &= \frac{5\cdot 4\cdot 3\cdot\cancel{2!}}{3!\cancel{2!}}.10^3.90^2 \end{aligned}$$

$$\begin{aligned}
&= (10)(.001)(.81)\\
&= .0081\\
P(4) &= \frac{5!}{4!(5-4)!}.10^4(1-.10)^{5-4}\\
&= \frac{5\cdot\cancel{4}!}{\cancel{4}!1!}.10^4.90^1\\
&= (5)(.0001)(.90)\\
&= .00045.
\end{aligned}$$

Y, finalmente, para $\hat{p} = 1.0$

$$\begin{aligned}
P(5) &= \frac{5!}{5!(5-5)!}.10^5(1-.10)^{5-5}\\
&= \frac{\cancel{5}!}{\cancel{5}!0!}.10^5.90^0\\
&= 10^5\\
&= .00001.
\end{aligned}$$

Los cálculos de arriba se resumen en la tabla 4.1 y se presentan gráficamente en el panel B de la figura 4.3.

TABLA 4.1: Distribuciones muestrales de $\hat{p}$ para $n = 5$ y $\pi = .10$.

Proporción $\hat{p}$	Número de éxitos y	Probabilidad $P(y)$
.00	0	.59049
.20	1	.32805
.40	2	.07290
.60	3	.00810
.80	4	.00045
1.00	5	.00001

EJEMPLO 4.5

Si se determina que el 10% de los residentes de Estados Unidos resultarán positivos a un cierto anticuerpo, ¿cuál es la probabilidad de seleccionar aleatoriamente cinco residentes de Estados Unidos y encontrar que las cinco pruebas resultaron positivas? ¿Cuál es la probabilidad de que *al menos* cuatro (esto es, cuatro o más) resulten positivos? ¿Cuál es la probabilidad de que al menos uno sea positivo?

Solución De la tabla 4.1 se obtiene que la probabilidad de que los cinco residentes resulten positivos es $P(5) = .00001$. La probabilidad de que al menos cuatro resulten positivos es $P(4) + P(5) = .00045 + .00001 = .00046$ y la probabilidad de que al menos uno resulte positivo es $P(1) + P(2) + P(3) + P(4) + P(5) + 1 - P(0) = 1 - .59049 = .40951$. ■

EJEMPLO 4.6

El 38% de todos los donadores de sangre en Estados Unidos tienen sangre tipo "O" positivo. Suponga que, como resultado de una teoría genética en la cual basó su argumentación, un investigador considera que esta proporción es más grande en Islandia. En un estudio preliminar, el investigador seleccionó aleatoriamente 10 donadores de sangre en Islandia y anotó su tipo sanguíneo. ¿Cuál es la probabilidad de que en esta muestra de donadores islandeses, 9 o 10 tengan sangre tipo "O" positivo , si la teoría del investigador es errónea (esto es, la proporción en Islandia *es* .38)? Si el número de sujetos con sangre tipo "O" positivo de hecho es 9 o 10, ¿qué implicaciones tendrá esto para la teoría del investigador?

Solución Dada una proporción de la población de .38, la probabilidad de que la muestra contenga 9 o 10 donadores con sangre tipo "O" positivo es $P(9) + P(10)$.

$$
\begin{aligned}
P(9) &= \frac{10!}{9!(10-9)!}.38^9(1-.38)10^{5-9} \\
&= \frac{10 \cdot \cancel{9}!}{\cancel{9}!1!}.38^9.62^1 \\
&= (10)(.00017)(.62) \\
&= .00105 \\
P(10) &= .3810 = .00006.
\end{aligned}
$$

La probabilidad de que 9 o 10 donadores en la muestra tengan sangre tipo "O" positivo es, entonces, .00105 + .00006 = .00111. Si el número de donadores en la muestra con sangre tipo "O" positivo es de 9 o 10, entonces, la teoría del investigador tiene sustento, pues la probabilidad de alcanzar dicho resultado en una población donde la proporción es .38, es muy pequeña. Es probable, aunque no seguro, que la proporción de personas con sangre tipo "O" positivo en los donadores de sangre islandeses sea mayor a .38. ■

4.2.6 Uso de la curva normal para aproximación de probabilidades asociada con $\hat{p}$

Las distribuciones muestrales descritas en los paneles B y C de la figura 4.3 en la página 81 fueron generados con una población común (panel A), pero difieren en que el tamaño de la muestra para la distribución en el panel B es de cinco, mientras que para el panel C es de 50. Como se observa, los diferentes tamaños de la muestra arrojan formas muy diferentes de distribuciones muestrales. De particular importancia es el hecho de que la distribución en el panel B es a todas luces anormal, mientras que en C aparece casi normal en su forma. Esto se debe al teorema del límite central, el cual (como se recordará) garantiza que las distribuciones muestrales de ciertos estadísticos se acercarán a la normalidad, conforme el tamaño de la muestra se incrementa.

Como sugieren los dos paneles, si se utiliza una aproximación de la curva normal para la distribución basada en muestras con tamaño cinco es probable que no se obtengan resultados satisfactorios, pero la misma aproximación para la distribución basada en muestras de tamaño 50 parece prometedora. El teorema del límite central garantiza que el modelo de la curva normal será apropiado para distribuciones basadas en *algún* tamaño de la muestra, pero, ¿qué tamaño debe ser ése? No hay respuesta segura, pero hay una regla puntual utilizada con frecuencia que afirma que el modelo de la curva normal será satisfactorio siempre y cuando $n\pi$ y $n(1 - \pi)$ sean mayores o iguales a cinco.[3] Para la distribución en el panel B este cálculo es (5)(.1) = .5 y (5)(.9)

---

[3] Algunos autores sostienen que estos valores deberán ser mayores que o iguales a 10

= 4.5, así que no se cuenta con el criterio mencionado, pero para el panel C tenemos (50)(.1) = 5 y (50)(.9) = 45, de manera que el criterio apenas se cumple.

El modelo de la curva normal se aplica en la distribución de $\hat{p}$ de esta forma, con la cual ahora usted está familiarizado. Esto es, la puntuación Z se calcula con el área apropiada y luego se localiza en la tabla de la curva normal. Al sustituir los resultados presentados arriba, en la ecuación 4.3, se obtiene

$$\boxed{Z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}}} \tag{4.6}$$

Para ejemplificar, suponga que se toma una muestra aleatoria de 50 observaciones de la población mostrada en el panel A de la figura 4.3. ¿Cuál es la probabilidad de que la proporción de éxitos en esta muestra sea mayor que .12?

La probabilidad estimada será el área bajo una curva normal con media .10 y desviación estándar

$$\sigma_{\hat{p}} = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{(.10)(.90)}{50}}$$

que cae por encima de .13. Puesto que la proporción de éxitos sólo puede tomar valores de .00, .02, .04,..., .12, .14,..., 1.00, se utiliza el límite real superior del intervalo .12, esto es .13. Se emplea el límite superior porque el problema es encontrar la probabilidad de que la proporción de éxitos *sea mayor* a .12. Tendrá que utilizarse el límite inferior si el problema requiere la probabilidad de obtener una proporción de .12 o *mayor*. Los límites reales de proporciones binomiales superior e inferior pueden calcularse directamente sumando y restando $.5/n$. Para el presente caso, el límite real superior es .12 + .5/50 = .13. A esta forma de utilizar los límites superior e inferior, usando una distribución continua para aproximar probabilidades asociadas con una variable discreta se le conoce como una **corrección de continuidad**. La puntuación Z es entonces

$$Z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{.13 - .10}{\sqrt{\frac{(.10)(.90)}{50}}} = \frac{.03}{.0424} = .71.$$

Si nos remitimos a la tabla de la curva normal en el Apéndice A, obtenemos una área asociada de .2389. El valor calculado de acuerdo con la ecuación 4.5 es $P(7) + P(8) + \ldots + P(50) =$ .2298.

¿Cuál es la probabilidad estimada de que la proporción sea .12? La estimación será el área entre el límite real inferior de .11 y el límite real superior de .13. Como se calculó arriba, la puntuación Z para .13 es .71, mientras que para .11 es

$$Z = \frac{.11 - .10}{\sqrt{\frac{(.10)(.90)}{50}}} = \frac{.01}{.0424} = .24.$$

Al usar estos valores en la tabla de la curva normal se observa que las áreas entre .13 y .10, y .11 y .10 son .2611 y .0948, respectivamente. El área entre .11 y .13 es, entonces, .2611 + .0948 = .1663. La probabilidad calculada de acuerdo con la ecuación 4.5 es $P(6) = .1541$.

EJEMPLO 4.7

Se dice que aproximadamente el 16% de los hombres en Estados Unidos con edades entre 60 y 64 años que mostraron un perfil de riesgo en particular sufrirán un ataque al corazón en los siguientes 10 años [14]. Si se observa una muestra aleatoria de 300 de esos hombres durante los próximos 10 años, ¿cuál es la probabilidad de que menos del 5% sufra un ataque al corazón?

Solución Como el problema especifica que *menos* del 5% experimentará un ataque al corazón, se utilizará el límite real inferior del intervalo del 5%. Este límite es .05 − .5/300 = .048. La puntuación Z es entonces

$$Z = \frac{.048 - .16}{\sqrt{\frac{(.16)(.84)}{300}}} = \frac{-.112}{.021} = -5.33.$$

La tabla de la curva normal no contiene valores Z de esta magnitud, pero puede concluirse, con toda seguridad, que la probabilidad es menor que .0002. (Ésta es la cola del área asociada con $Z = 3.50$, que es el punto más extremo en la tabla). Al utilizar una corrección continua cuando las muestras son grandes generalmente se tiene una menor influencia en el resultado. En este caso, si se hubiera utilizado .05 en lugar de .048, el valor Z habría sido −5.24 lo cual hará una pequeña diferencia en el resultado. ■

EJEMPLO 4.8

Suponga que una comunidad grande se divide equitativamente de acuerdo con su opinión sobre la cantidad máxima que se puede recuperar legalmente a través de una demanda judicial por negligencia médica. Si esta suposición es correcta, ¿cuál es la probabilidad de que un sondeo aleatorio entre 200 miembros de la comunidad producirá 55% o más de respuestas favorables? Calcule la probabilidad con y sin corrección de continuidad.

Solución La corrección de continuidad es .5/200 = .0025. Como la tarea es encontrar la probabilidad de que el 55% *o más* será favorable, se utilizará el límite real inferior de la categoría del 55% o .55 − .0025 = .5475. Puesto que la comunidad en cuestión se divide equitativamente, la proporción favorable en la población se considera como .50. La puntuación Z es entonces

$$Z = \frac{.5475 - .50}{\sqrt{\frac{(.50)(.50)}{200}}} = \frac{.0475}{.035} = 1.36.$$

El área por arriba de 1.36 es .0869. Sin la corrección de continuidad la puntuación Z es .05/.035 = 1.43, que tiene una cola en el área superior de .0764. La probabilidad calculada de acuerdo con la ecuación 4.5 es .0895. ■

4.3 PRUEBA DE HIPÓTESIS

4.3.1 Introducción

En el capítulo 1 indicamos que la disciplina de la estadística/bioestadística puede dividirse conceptualmente en dos áreas relacionadas de estudio: estadística descriptiva y estadística inferencial. Con escasas excepciones, usted ha completado su estudio de la estadística descriptiva y domina de antemano el conocimiento y las habilidades necesarios para el estudio de la inferencia. En esta sección comenzará el estudio de la inferencia a través de la prueba de hipótesis. Más tarde estudiará los intervalos de confianza, otro componente de la inferencia.

Las pruebas de hipótesis que usted conocerá y aprenderá en este capítulo quizá no sean muy importantes desde un punto de vista, pero son esenciales desde otra perspectiva. No tienen importancia porque no se emplean comúnmente en la práctica de la investigación. La razón es que prueban hipótesis simples que no reflejan el tipo de preguntas que interesan a la mayoría de los investigadores de las ciencias de la salud. Sin embargo, esta simplicidad es su fortaleza. Como son relativamente simples, es posible dominar su lógica y su aplicación fundamental sin mayor dificultad. Los mismos métodos de aplicación y lógica que soportan estas pruebas fundamentan las pruebas más complejas que encontrará en capítulos posteriores. De tal forma que con el dominio amplio de las pruebas simples presentadas aquí, usted dominará métodos más complejos, pero sin dificultad. Las pruebas que aprenderá a utilizar en este capítulo son la prueba Z de una media, la prueba t de una media, pruebas de una proporción y prueba de equivalencia de una media.

4.3.2 Método y justificación

La prueba de hipótesis es un método esencial para la toma de decisiones. La decisión relaciona la elección entre dos enunciados competitivos y mutuamente excluyentes, respecto de uno o más parámetros de la población.[4] Los enunciados competitivos, de hecho, se conocen como hipótesis nula y alternativa, respectivamente. La **hipótesis nula**, simbolizada como H_0, es un enunciado preciso[5] concerniente al parámetro (o parámetros) de interés, mientras que la **hipótesis alternativa**, simbolizada como H_A, es un enunciado competitivo menos preciso. Por ejemplo, quizá la hipótesis nula establezca que la media de una población (μ) es igual a 100,[6] mientras que la hipótesis alternativa menos concisa sostiene que la media es mayor que 100. Se emplea el término *nula* porque la hipótesis nula es una declaración de no diferencia. En el ejemplo de arriba, la hipótesis nula enuncia que no hay diferencia entre la media de la población y el valor 100. Por otro lado, la hipótesis alternativa afirma que *hay* una diferencia entre la media de la población y el valor 100 y, en este caso, sigue especificando que la media es mayor que 100.

Decidir si se acepta la afirmación de la hipótesis nula o si la afirmación alternativa se establece como la condición verdadera se hace utilizando el modelo de la distribución muestral del estadístico implicado para establecer el criterio de decisión. Si el criterio se satisface, la hipótesis nula se rechaza en favor de la alternativa. Si no se satisface el criterio, se mantiene el enunciado nulo. Un ejemplo aclarará esto.

Suponga que usted desea probar la hipótesis nula de una media de población igual a 100. Este enunciado se representa como $H_0 : \mu = 100$. La hipótesis alternativa sostiene que la media es mayor que 100 y se representa como $H_A : \mu > 100$. La prueba se lleva a cabo mediante la selección aleatoria de una muestra de un tamaño específico de la población y se calcula la media de la muestra, $\bar{x}$. Puesto que la distribución muestral de $\bar{x}$ tiende a ser aproximadamente normal, se elige la curva normal para probar el modelo, mostrado en la figura 4.4.

---

[4] Hay pruebas de hipótesis que no relacionan los parámetros, pero nosotros no nos ocuparemos de ellas aquí.

[5] Algunas veces la hipótesis nula se enuncia de una manera imprecisa, pero la prueba debe hacerse con un valor específico.

[6] Algunos preferirán enunciar que la hipótesis nula indica que la media es menor que o igual a 100, pero la prueba será realizará con el valor preciso de 100.

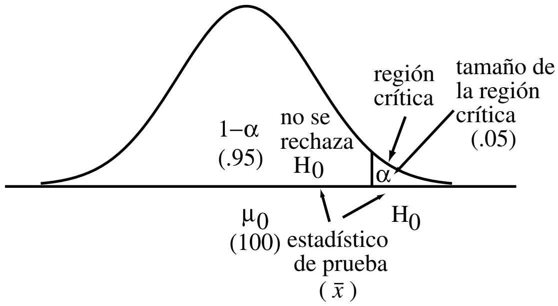


FIGURA 4.4: Ejemplo del modelo para una prueba de hipótesis.

Observe primero que la prueba se realiza de acuerdo con la suposición de que la hipótesis nula es verdadera. De esta manera, el modelo utilizado tiene el valor de la media especificado por la hipótesis nula, simbolizada como μ_0 y que es 100 en el presente ejemplo. (Debe hacerse una distinción clara entre μ, la media de la población, y μ_0, la media de la población *hipotética*). Se especifica una pequeña región en la cola de la curva, llamada la **región crítica**. El *tamaño* de esta región se simboliza con α (**alfa**). Alfa se ubica tradicionalmente en .05 o .01, pero puede encontrarse en otros niveles si así se desea. La ubicación y el tamaño de la región crítica no son arbitrarios. La región está diseñada para cumplir con los dos criterios.

Primero, la probabilidad de que la media de una muestra tome un valor dentro de la región crítica *si la hipótesis nula es verdadera* es pequeña. Esto puede asegurarse escogiendo un valor suficientemente pequeño para α. Si α es .05, entonces, la probabilidad de que el estadístico de prueba tome un valor dentro de la región crítica es únicamente de .05. De tal forma, es improbable que el estadístico de prueba manifieste un valor en la región crítica *si la hipótesis nula es verdadera*.

Segundo, la probabilidad de que el estadístico de prueba tome un valor dentro de la región crítica debe incrementarse cuando la hipótesis nula sea falsa. En específico, mientras la diferencia entre μ_0 y μ se incremente, la probabilidad de que el estadístico de prueba se encuentre dentro de la región crítica también puede incrementar. Usted verá cómo se satisface este criterio al ubicarlo dentro de la región crítica de la cola de la distribución si considera lo siguiente. Suponga que la hipótesis nula es falsa porque la media de la población es 105. Esto obedece a que las medias de la muestra de esta población tenderán a ser más grandes (105 en promedio) de lo que se esperaría si la hipótesis nula fuera verdadera y, por lo tanto, tendrían una mayor probabilidad de ser lo suficientemente grandes como para caer dentro de la región crítica. Si la media de la población fuera aún más grande, digamos 200, las medias de la muestra tenderán a ser un poco grandes y la probabilidad de que se encuentren dentro de la región crítica se incrementará. Regresaremos a este punto más adelante.

Decidir si la hipótesis alternativa será elegida como reflejo de la condición verdadera de la media de la población depende de si el estadístico de prueba toma o no un valor dentro de la región crítica. Si esto sucede, la hipótesis nula se rechaza en favor de la alternativa. ¿Por qué? Porque es improbable que el estadístico de prueba esté dentro de la región crítica si la hipótesis nula es verdadera. De hecho, la probabilidad es únicamente α. Por otra parte, si la hipótesis nula es falsa, la probabilidad se incrementa. Por lo tanto, si el estadístico de prueba está dentro de la región crítica, la explicación lógica para este suceso es que la hipótesis nula es falsa.

Pero, ¿qué sucede si el estadístico de prueba no está dentro de la región crítica? Quizás usted piense que la hipótesis alternativa se rechaza en favor de la nula, pero ése no es el caso. La lógica de la prueba de hipótesis establece la hipótesis nula como la condición por desmentir. La prueba requerida tiene la forma de una probabilidad lo suficientemente pequeña asociada con el valor observado en el estadístico de prueba, bajo una hipótesis nula verdadera. Por lo tanto, se llega a la conclusión de que la hipótesis nula no es desmentida cuando el estadístico de prueba no está en la región crítica. Para usar el lenguaje formal de la prueba de hipótesis: usted "rechaza la hipótesis nula" o "no rechaza la hipótesis nula". Otra forma de reportar el rechazo de la hipótesis nula es declarando que la prueba fue **estadísticamente significativa**, mientras que la decisión de no rechazarla se reporta como **estadísticamente no significativa.**

Quizá en este momento usted se sienta más confundido respecto de la prueba de hipótesis. Además, es probable que usted se pregunte: "¿Para qué sirve esto?" La lógica de la prueba de hipótesis resultará clara cuando comencemos a hacer pruebas y a interpretar los resultados. Ahora, estamos listos para comenzar este proceso. Antes de hacerlo, sin embargo, le sugerimos que vuelva a leer este apartado y estudie con cuidado la figura 4.4. Una breve mirada a la figura 1.1 también podrá ayudarle. En cuanto a la pregunta: "¿Para qué sirve esto?", comenzaremos a tratar ese asunto en el capítulo 5. Por el momento, tenga valor, todo resultará más claro conforme continuemos.

4.3.3 Prueba Z de una media

La prueba Z de una media se utiliza para probar hipótesis nulas de la forma $H_0 : \mu = \mu_0$, como se explicó antes. La hipótesis alternativa toma una de las dos formas referidas como pruebas de una o dos colas (también llamadas pruebas de uno o dos lados). La hipótesis alternativa para las pruebas de una cola es *ya sea* $H_A : \mu > \mu_0$ o $H_A : \mu < \mu_0$ (pero no ambas). La prueba de dos colas tiene la hipótesis alternativa $H_A : \mu \neq \mu_0$. Se ilustrará y se discutirá cada una de ellas.

Prueba de una cola: $\boldsymbol{H_A : \mu > \mu_0}$. Considere que en una comunidad afroestadounidense localizada en un gran asentamiento metropolitano se realiza un censo. El propósito es evaluar el grado de satisfacción/descontento entre los miembros de la comunidad de acuerdo con su acceso a la atención médica. La escala utilizada en el censo produce puntuaciones en un rango que va de cero a 100, en donde las puntuaciones más bajas indican menor satisfacción. Los entrevistadores del censo encuentran que el punto promedio en la comunidad es 47.4, que se interpreta como un nivel subrayado de descontento. El análisis también muestra que la desviación estándar de esos puntos es 24.7.

Por otra parte, como resultado de este censo, se pone a la disposición un número de donaciones federales y no federales para mejorar el acceso a la atención médica en la comunidad. Unos 10 años después del conocimiento de la existencia de este dinero, los políticos desean determinar si la opinión de la comunidad con respecto al acceso a la atención médica ha mejorado.

Es obvio que esta pregunta puede contestarse haciendo un nuevo censo en la comunidad para ver si el punto medio es ahora más elevado de lo que era hace 10 años. Pero no es posible financiar un esfuerzo tan grande. En esta situación, los evaluadores seleccionarán aleatoriamente a 100 miembros de la comunidad para aplicarles una encuesta. La media de esta muestra, entonces, será usada para llevar a cabo la prueba Z de una media de la siguiente forma.

La hipótesis nula sometida a prueba es que el valor medio en la comunidad (población) es 47.4, mientras que la hipótesis alternativa indica que el valor medio es mayor que 47.4. Enunciado más formalmente,

$$H_0 : \mu = 47.4$$
$$H_A : \mu > 47.4$$

Observe que la hipótesis nula afirma que la media observada previamente aún es la condición verdadera en la población, mientras que la hipótesis alternativa sostiene que la media en la comunidad es ahora mayor que la previamente observada. La lógica de la prueba requiere que la condición nula sea refutada o se sostenga.

Debido a que la distribución muestral de $\bar{x}$ tiende a ser normal en su forma, se usará la curva normal como modelo para la distribución muestral. Alfa se ubica en .05 y la región crítica se establece como en la figura 4.4. Los investigadores saben que la probabilidad de que la media de los 100 miembros de la comunidad muestreados sea lo suficientemente grande como para caer dentro de la región crítica es únicamente de cinco en cien si la hipótesis nula es verdadera, pero también saben que tenderá a ser más grande si la hipótesis alternativa es verdadera. Por lo tanto, puede realizarse una prueba de H_0 para hacer notar, o no, que $\bar{x}$ está dentro de la región crítica. Si está en la región crítica, H_0 se rechazará, de otra forma, se mantendrá. Pero, ¿cómo podemos saber si $\bar{x}$ está dentro de la región crítica? Esto se hace a través de dos métodos diferentes, a los cuales nos referiremos como los métodos del valor-*p* versus alfa, y *Z* obtenida versus *Z* crítica. Utilizaremos cada uno de éstos para realizar la prueba *Z* de una media para el problema citado arriba.

Método del valor-*p* versus alfa

Considere que la media de la muestra $\bar{x}$ es 52.8. ¿El valor está dentro de la región crítica? La respuesta cae en el panel A de la figura 4.5. Fíjese en que la media del modelo utilizado para realizar la prueba es $\mu_0 = 47.4$, y que la región crítica está en la cola derecha de la curva. La puntuación *Z* para la media de la muestra se calcula por

$$\boxed{Z = \frac{\hat{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}} \tag{4.7}$$

la cual es simplemente la ecuación 4.3 con μ_0 sustituida por μ, para indicar que se utilizó el valor de la media hipotética en lugar de μ. Por esta ecuación

$$Z = \frac{.52.8 - 47.4}{\frac{24.7}{\sqrt{100}}} = 2.19$$

El valor-*p* (o simplemente *p*) para una prueba de una cola con hipótesis alternativa de la forma $H_A : \mu > \mu_0$ es la probabilidad de obtener un estadístico de prueba tan grande como, o mayor que el valor del estadístico de prueba observado en realidad. En este caso, el valor-*p* es la probabilidad de mantener la media de la muestra, cuyo valor es 52.8 o más. El área en la columna 3 del Apéndice A para la puntuación *Z* de 2.19 es .0143, lo cual es el valor-*p* deseado (mostrado con el sombreado gris en la región crítica de la curva en el panel A de la figura 4.5). El hecho de que el valor-*p* sea menor que alfa (.05) implica que $\bar{x}$ está dentro de la región crítica, lo cual, a la vez, significa que H_0 se rechaza. En general, la hipótesis nula se rechaza cuando $p \leq \alpha$, dado que esto implica que el estadístico de prueba está dentro de la región crítica. De aquí se infiere que H_0 no se rechaza cuando $p > \alpha$.

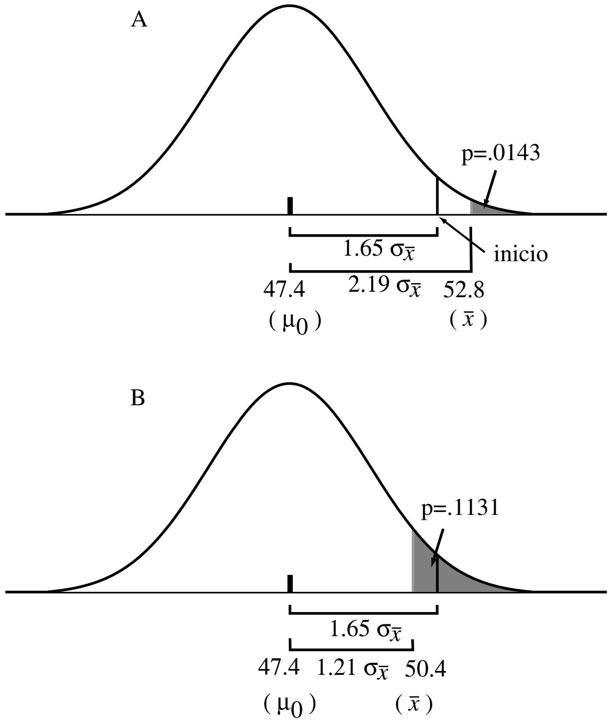


FIGURA 4.5: Pruebas Z de una media de $H_0: \mu = 47.4$ contra la hipótesis alternativa $H_A: \mu > 47.4$ con $\alpha = .05$.

La lógica para usar el área de la curva arriba de $\bar{x}$ (el valor-*p*) para comparar con α y determinar si $\bar{x}$ está en la región crítica puede verse con más claridad considerando lo siguiente. Suponga que $\bar{x}$ estuviera localizado en el punto de inicio de la región crítica (etiquetado "inicio" en el panel A de la figura 4.5). En este caso el área por arriba de $\bar{x}$ sería exactamente .05. Cualquier valor más grande que $\bar{x}$ estará dentro de la región crítica y tendrá una área menor por arriba de ella. En contraste, un valor más pequeño de $\bar{x}$ caerá afuera de la región crítica y tendrá una área mayor por arriba de ella.

A el riesgo de prolongar aún más esta explicación, considere que el resultado de la prueba pudo haber tenido la media de la muestra en 50.4. La puntuación Z en este caso será

$$Z = \frac{50.4 - 47.4}{\frac{24.7}{\sqrt{100}}} = 1.21.$$

La columna 3 del Apéndice A muestra que el área asociada es .1131, que se destaca con la sombra gris en la curva del panel B de la figura 4.5. Puesto que .1131 es mayor que .05, puede afirmarse que $\bar{x}$ no está dentro de la región crítica y que la hipótesis nula no se rechaza. ¿Este resultado indica que el nivel de satisfacción/descontento es aún 47.4? No, esto significa que somos incapaces de demostrar que el nivel de satisfacción/descontento es mayor que 47.4.

Método de *Z* obtenida versus *Z* crítica

El método del valor-*p* versus alfa para determinar si un estadístico de prueba está dentro de la región crítica se realiza calculando la puntuación *Z* del estadístico de prueba, que se utiliza entonces para obtener el área de la curva que cae por arriba del estadístico. Debido que alfa también es una área de la curva, la decisión respecto de la hipótesis nula se hace con base en la comparación de dos áreas. El método de *Z* obtenida versus *Z* crítica comienza de la misma forma. Esto es, se calcula la puntuación *Z* del estadístico de prueba. Este valor se determina como ***Z* obtenida**. Usted recordará que la puntuación *Z* indica el número de desviaciones estándar (o errores estándar en este caso) que un punto o estadístico cae dentro de la media de la distribución. En el ejemplo mostrado en el panel A de la figura 4.5, la puntuación *Z* calculada indica que la media de la muestra de 52.8 se localiza 2.19 desviaciones estándar por arriba de la media de la distribución. La ***Z* crítica** es el número de desviaciones estándar entre la media de la distribución y el punto de inicio de la región crítica. En otras palabras, es la puntuación *Z* el que cae en el límite delantero de la región crítica.

La *Z* crítica puede obtenerse del Apéndice A, buscando el área apropiada para encontrar el valor asociado de *Z*. Como deseamos encontrar el valor de *Z* que tenga .05 de la curva por arriba de él, vemos en la columna 3 de la tabla para .05. Por desgracia, .05 no se encuentra. En su lugar, encontramos dos áreas que son igualmente cercanas a .05, en particular .0495 y .0505. Por lo general, elegimos el área más cercana a la deseada, pero en este caso las dos áreas son igualmente cercanas a .05. La puntuación *Z* de 1.64 corta .0505 en la cola, mientras que una *Z* de 1.65 corta .0495 también en la cola. Podríamos interpolar y usar una *Z* de 1.645, pero, para simplificar, cuando se encuentran dos áreas igualmente cercanas por lo común se utiliza el área más pequeña. Esto significa que podemos tomar el valor de *Z* asociado con .0495, que es 1.65.

Así que la distancia de la media de la distribución (47.4) al límite delantero de la región crítica es (aproximadamente) 1.65 desviaciones estándar. La *Z* obtenida indica que $\bar{x}$ es 2.19 desviaciones estándar de la media de la distribución, mientras que la *Z* crítica indica que el límite delantero de la región crítica se localiza 1.65 desviaciones estándar por arriba de la media de la distribución. Es claro que $\bar{x}$ está más allá de la media de la distribución, que es el límite delantero de la región crítica, así que el estadístico de prueba se localiza necesariamente en la región crítica. (Véase el panel A de la figura 4.5.) Así, mientras el método del valor-*p* versus alfa compara dos áreas, el método de *Z* obtenida versus *Z* crítica compara dos puntuaciones *Z*.

Para el ejemplo mostrado en el panel B, el estadístico de prueba se localiza a 1.21 desviaciones estándar de la media de la distribución, así que no está tan lejos del límite delantero de la región crítica y por ello no se puede localizar dentro. En general, para la prueba de una cola con la hipótesis alternativa $H_A : \mu > \mu_0$ la región crítica se localiza en la cola derecha de la curva, así que la hipótesis nula se rechaza cuando la *Z* obtenida es mayor que o igual a la *Z* crítica.

EJEMPLO 4.9

Utilice la siguiente información para realizar la prueba *Z* de una media. Justifique su decisión respecto de la hipótesis nula con base en ambos métodos, el valor-*p* versus alfa y *Z* obtenida versus *Z* crítica.

$$H_0 : \mu = 120 \qquad \sigma = 40 \qquad \bar{x} = 128.2$$
$$H_A : \mu > 120 \qquad n = 150 \qquad \alpha = .01$$

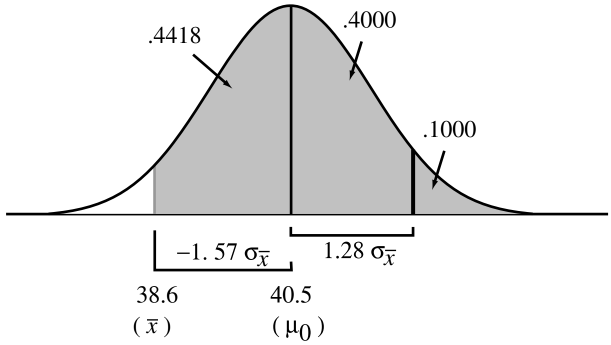


FIGURA 4.6: Prueba Z de una media de $H_0: \mu = 40.5$ contra la hipótesis alternativa $H_A: \mu > 40.5$, con $\alpha = .10$.

Solución El valor Z asociado con 128.2 es

$$Z = \frac{\hat{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{128.2 - 120.0}{\frac{40.0}{\sqrt{150}}} = 2.51.$$

El valor-*p* es el área por arriba de 128.2. La columna 3 del Apéndice A muestra que el área por arriba de Z = 2.51 es .0060. Como este valor es menor que α = .01, la hipótesis nula se rechaza. La columna 3 de la tabla de curva normal también muestra que el valor más cercano a .01 es .0099, que tiene un valor Z asociado de 2.33. Por consiguiente, el valor Z en el límite delantero de la región crítica es 2.33. Dado que la Z obtenida de 2.51 es más grande que la Z crítica de 2.33, la hipótesis nula se rechaza. ■

EJEMPLO 4.10

Utilice la siguiente información para realizar la prueba Z de una media. Justifique su decisión respecto de la hipótesis nula con base en ambos métodos, el valor-*p* versus alfa y Z obtenida versus Z crítica.

$$H_0: \mu = 40.5 \qquad \alpha = 11.5 \qquad \bar{x} = 38.6$$
$$H_A: \mu > 40.5 \qquad n = 90 \qquad \alpha = .10$$

Solución El valor Z asociado con 38.6 es

$$Z = \frac{38.6 - 40.5}{\frac{11.5}{\sqrt{90}}} = -1.57.$$

El valor-*p* es el área por arriba de 38.6, el cual a su vez está compuesto por el área entre 38.6, la media de la distribución de 40.5, y el área por arriba de 40.5. Utilizando $Z = -1.57$, la columna 2 del Apéndice A indica que el área entre 38.6 y 40.5 es .4418. Dado que el área por arriba de 40.5 es .5000, el valor-*p* es .4418 + .5000 = .9418, el cual es sin duda mayor que α = .10, así que la hipótesis nula no se rechaza. (Véase la figura 4.6.)

El valor más cercano a .10 en la columna 3 del Apéndice A es .1003, el cual tiene un valor Z asociado de 1.28. Por lo tanto, el límite delantero de la región crítica se localiza en un punto que está 1.28 errores estándar por *arriba* de la media de la distribución. En contraste, la Z obtenida de

-1.57 indica que $\bar{x}$ se localiza 1.57 errores estándar por *debajo* de la media de la distribución. Entonces, es evidente que $\bar{x}$ no está dentro de la región crítica. (Véase la figura 4.6.) ■

Prueba de una cola: $H_A : \mu < \mu_0$. Una segunda forma de la prueba de una cola emplea la hipótesis alternativa $H_A : \mu < \mu_0$, lo cual implica una media de la población menor que el valor especificado por la hipótesis nula. Con el fin de incrementar la probabilidad de lograr un resultado significativo de la prueba cuando H_0 es falsa y H_A es verdadera, la región crítica debe estar localizada en la cola derecha de la distribución más que en la izquierda. Considere el siguiente ejemplo.

El Inventario de Burnout de Maslach [35] es un instrumento que mide tres aspectos del agotamiento o *burnout* manifestado por enfermeras profesionales. La tabla de la norma para el inventario reporta la media y la desviación estándar para enfermeras de Estados Unidos en la subescala de agotamiento emocional, como 22.19 y 9.53, respectivamente. (Puntos más altos indican un mayor grado de agotamiento emocional). Suponga que un investigador que trabaja con datos sobre las enfermeras cree que, después de un largo periodo de dificultades, las condiciones laborales para las enfermeras en Estados Unidos han mejorado gradualmente desde que se implantaron las leyes. El investigador va más allá en su hipótesis y sugiere que como resultado de estas mejoras laborales, el agotamiento emocional ha disminuido. Con el fin de probar esta hipótesis, el investigador administra el inventario a 121 enfermeras de un hospital local.[7] La puntuación media de las 121 enfermeras en la subescala de agotamiento emocional es 20.58. El investigador lleva a cabo la prueba Z de una cola para una media con $\alpha = .05$ y la siguiente hipótesis

$$H_0 : \mu = 22.19$$
$$H_A : \mu < 22.19$$

Z se calcula como

$$Z = \frac{\hat{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{20.58 - 22.19}{\frac{9.53}{\sqrt{121}}} = -1.86.$$

Método del valor-*p* versus alfa

El valor-*p* para una prueba de una cola con hipótesis alternativa de la forma $H_A : \mu < \mu_0$ es el área por *debajo* del estadístico de prueba. La columna 3 del Apéndice A revela que este valor es .0314. Como $p = .0314 < \alpha = .05$, la hipótesis nula se rechaza. (Véase el panel A de la figura 4.7.)

Método de *Z* obtenida versus *Z* crítica

Puesto que la región crítica para las pruebas con esta alternativa está en el lado izquierdo de la curva, la Z crítica es negativa. Para el problema en cuestión la Z crítica es -1.65, lo cual indica que el límite delantero de la región crítica cae en un punto que se localiza 1.65 errores estándar por debajo de la media de la distribución. Por lo tanto, la hipótesis nula se rechaza cuando la Z obtenida es menor que o igual a la Z crítica. Como $-1.86 < -1.65$, la hipótesis nula se rechaza. El panel A de la figura 4.7 muestra los elementos de este análisis.

A partir de este estudio, el investigador concluye que el agotamiento emocional entre las enfermeras de Estados Unidos ha disminuido. Pero esta conclusión debe ser tentativa, puesto que 121 enfermeras incluidas en el estudio no constituyen una muestra aleatoria de la población total de enfermeras en Estados Unidos. Esta falta de selección aleatoria de sujetos para el estudio es el estado habitual de los sucesos, así que las conclusiones obtenidas con este tipo de estudios rara vez son definitivas.

[7] Ésta no es una prueba aleatoria, pero refleja las limitaciones prácticas de muchas investigaciones.

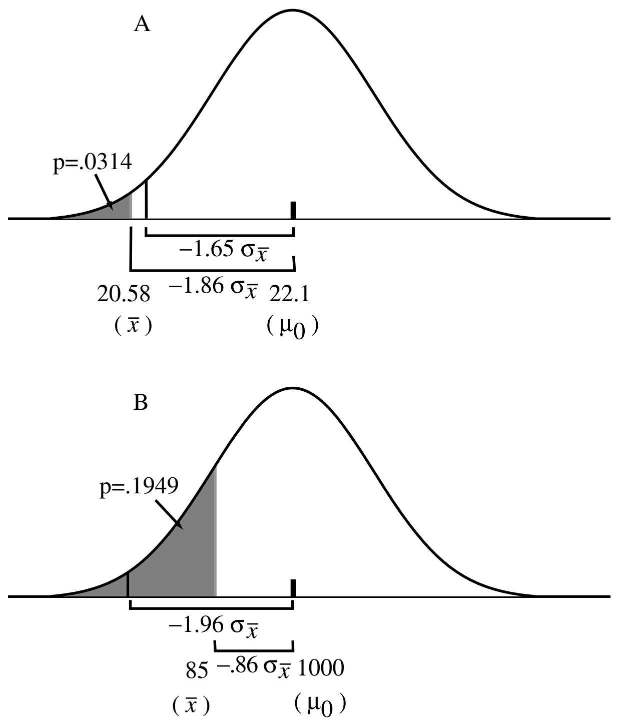


FIGURA 4.7: Pruebas Z de una media con hipótesis alternativa $H_A : \mu < \mu_0$.

EJEMPLO 4.11

Utilice la siguiente información para realizar la prueba Z de una media. Justifique su decisión respecto de la hipótesis nula con base en ambos métodos, el valor-p versus alfa y Z obtenida versus Z crítica.

$$H_0 : \mu = 1000 \qquad \sigma = 235 \qquad \bar{x} = 985$$
$$H_A : \mu < 1000 \qquad n = 180 \qquad \alpha = .025$$

Solución El valor de Z asociado con 985 es

$$Z = \frac{\hat{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{985 - 1000}{\frac{235}{\sqrt{180}}} = -.86.$$

El valor-p es el área por debajo de 985. Utilizando $Z = -.86$ la columna 3 del Apéndice A muestra que esta área será .1949. Dado que .1949 es mayor que .025, la hipótesis nula no se rechaza.

Como la puntuación Z de 1.96 corta en .025 a la cola de la distribución y la región crítica está por debajo de la media de la distribución, Z crítica es −1.96. Dado que −.86 no es menor que −1.96, la hipótesis nula no se rechaza. La lógica de este análisis se muestra en el panel B de la figura 4.7. ■

Prueba de dos colas: $H_A : \mu \neq \mu_0$. En muchas situaciones los investigadores son renuentes o incapaces de especificar si una hipótesis nula falsa implicará un valor de μ mayor que o menor que μ_0. En tales situaciones se realiza una prueba de dos colas. Con el fin de detectar cualquier eventualidad, se coloca una región crítica en *ambas* colas de la distribución. La hipótesis nula se rechaza si el estadístico de prueba toma un valor en cualquiera de las dos regiones. Para poder mantener la probabilidad de que un estadístico de prueba caiga dentro de una región crítica en α cuando H_0 es verdadera, el tamaño de cada región se establece en $\alpha/2$.[8]

Suponga que en el ejemplo de las enfermeras de la página 94, las condiciones de trabajo para ellas han cambiado durante algún tiempo: algunas condiciones han mejorado, mientras que otras parecen haberse deteriorado. El investigador desea determinar si el nivel de agotamiento emocional en una porción de la población de enfermeras de Estados Unidos se ha visto afectado por estos cambios. Como algunos cambios en el lugar de trabajo han sido positivos y otros negativos, es de interés ver si el efecto global ha aumentado o disminuido el nivel promedio de agotamiento. El investigador realiza una prueba Z de dos colas para una media con $\alpha = .05$ y la siguiente hipótesis

$$H_0 : \mu = 22.19$$
$$H_A : \mu \neq 22.19$$

Como se calculó previamente, la puntuación Z para la media de la muestra de 20.58 es −1.86.

Método del valor-*p* versus alfa

El valor-*p* para la prueba de dos colas se obtiene duplicando el área de la cola definida por el estadístico de prueba. El área de la cola es el área por arriba del estadístico de prueba si éste se encuentra por arriba de la media de la distribución, o es el área debajo del estadístico de prueba si éste se encuentra por debajo de la media de la distribución. En el ejemplo real, la media de la muestra de 20.58 está por debajo de la media de la distribución de 22.19, así que el área de la cola es el área debajo de 20.58. Al utilizar $Z = -1.86$, el Apéndice A muestra que esta área será .0314. El valor-*p* es entonces $2 \times .0314$. Dado que $.0628 > \alpha = .05$, la hipótesis nula no se rechaza. Este valor-*p* se muestra como la combinación de las áreas sombreadas en el panel A de la figura 4.8. Observe que el área representada por la combinación de las áreas sombreadas (valor-*p*) excede las áreas combinadas de las dos regiones críticas (α).

Método de *Z* obtenida versus *Z* crítica

Observe primero que hay dos valores Z críticos para la prueba de dos colas, uno de los cuales es positivo y el otro negativo. Esto resulta del hecho de que hay una región crítica por arriba y por debajo de la media de la distribución. Debido a que cada una de estas regiones abarca $\alpha/2$ de la curva, es este valor el que se utiliza en el Apéndice A para encontrar la Z crítica. Dado que $\alpha =$ .05 en el ejemplo en cuestión, necesitamos encontrar el valor de Z que corta .025 en la cola de la curva. La columna 3 del Apéndice A muestra que este valor será 1.96. Por consiguiente, los *valores* críticos para esta prueba son ±1.96.

La hipótesis nula se rechaza si *a*) la Z obtenida es menor que o igual al valor negativo de Z crítica, *o b*) la Z obtenida es mayor que o igual al valor positivo de Z crítica. Esto significa que H_0 se rechaza si el estadístico de prueba está en cualquier región crítica. El panel A de la figura 4.8 muestra que el límite delantero de la región crítica en la cola izquierda de la curva está 1.96 errores estándar por debajo de la media de distribución, mientras que $\bar{x}$ está 1.86 errores estándar por debajo de la media de la distribución. Es claro que $\bar{x}$ no está lo suficientemente lejos por debajo de la media de la distribución como para alcanzar el límite delantero de la región crítica.

[8] La división equitativa de α entre las dos colas no es obligatoria pero casi siempre se hace.

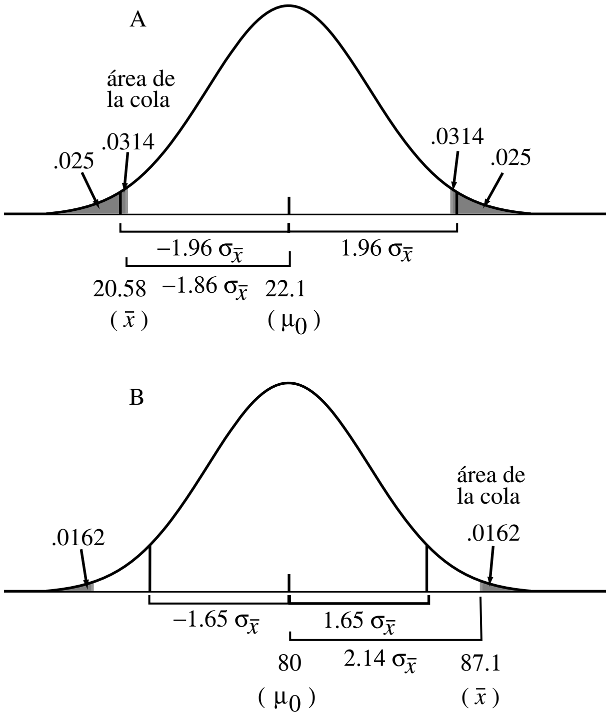


FIGURA 4.8: Pruebas *Z* de una media con hipótesis alternativa $H_A : \mu \neq \mu_0$.

EJEMPLO 4.12

Utilice la siguiente información para realizar la prueba *Z* de una media. Justifique su decisión respecto de la hipótesis nula con base en ambos métodos, el valor-*p* versus alfa y *Z* obtenida versus *Z* crítica.

$$H_0 : \mu = 80 \qquad \sigma = 21 \qquad \bar{x} = 87.1$$
$$H_A : \mu \neq 80 \qquad n = 40 \qquad \alpha = .10$$

Solución El valor de *Z* asociado con 87.1 es

$$Z = \frac{\hat{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{87.1 - 80.0}{\frac{21}{\sqrt{40}}} = 2.14.$$

Como 87.1 está por arriba de la media de la distribución de 80.0, el área de la cola por arriba de 87.1 cuya referencia al Apéndice A muestra que es .0162. El valor-*p* es entonces $p = 2 \times .0162 = .0324$. Debido a que $p = .0324 < \alpha = .10$, la hipótesis nula se rechaza. El investigador puede confiar (aunque no totalmente) en que la media de la población es mayor que 80. Los detalles de esta prueba se describen en el panel B de la figura 4.8.

Los valores críticos de *Z* para una prueba de dos colas en $\alpha = .10$ son los valores que cortan .05 en cada cola de la distribución. Como se observó arriba, estos valores son ±1.65. Dado que el estadístico de prueba está 2.14 errores estándar por arriba de la media de la distribución y la cola derecha de la región crítica comienza en el punto que está 1.65 errores estándar por arriba de la media de la distribución, el estadístico de prueba se encuentra claramente en la cola derecha de la región crítica. Los detalles de esta lógica se indican en el panel B de la figura 4.8. ■

EJEMPLO 4.13

Utilice la siguiente información para realizar la prueba Z de una media. Justifique su decisión respecto de la hipótesis nula con base en ambos métodos, el valor-p versus alfa y Z obtenida versus Z crítica.

$$\begin{array}{lll} H_0 : \mu = 150 & \sigma = 50.4 & \bar{x} = 142.4 \\ H_A : \mu \neq 150 & n = 105 & \alpha = .01 \end{array}$$

Solución El valor de Z asociado con 142.4 es

$$Z = \frac{142.4 - 150.0}{\frac{50.4}{\sqrt{105}}} = -1.55.$$

La región de la cola definida por el estadístico de prueba es el área por debajo de 142.4, que según el Apéndice A es .0606. El valor-p es entonces $p = 2 \times .0606 = .1212$. Como $.1212 > .01$, la hipótesis nula no se rechaza.

El Apéndice A muestra las dos áreas de la cola que están igualmente cerca de .005. En particular, .0049 y .0051. Como se observó con anterioridad, cuando las dos áreas de la cola están igualmente cerca del punto que se está buscando, se usará la más pequeña de las dos áreas. En este caso, el resultado es un valor de Z de 2.58. La Z crítica es entonces ± 2.58. Dado que el estadístico de prueba está 1.55 errores por debajo de la media de la distribución, no alcanza el límite delantero de la cola izquierda de la región crítica. Por consiguiente, la hipótesis nula no se rechaza. ■

Suposiciones que fundamentan la prueba *Z* de una media. Como usted habrá notado, el fundamento para la prueba de hipótesis requiere que una o más regiones críticas estén establecidas de tal manera que para obtener la probabilidad de un estadístico de prueba se tome un valor pequeño en la región (o regiones) cuando la hipótesis nula sea verdadera. De esta forma, si el estadístico de prueba toma un valor en la región (o regiones), la lógica sugiere que la hipótesis nula probablemente no es verdadera. Pero las regiones críticas están construidas en conjunción con un *modelo* de la distribución muestral del estadístico más que con la distribución muestral real. Es así como surge la pregunta referente a las consecuencias si el modelo usado para la prueba de hipótesis es inapropiado; entonces, ¿las probabilidades estimadas obtenidas del modelo serán erróneas? En tal circunstancia, la probabilidad de que un estadístico tome un valor en la región (o regiones) crítica(s) puede ser muy diferente de la probabilidad estimada que ofrece el modelo. Por ejemplo, suponga que el investigador establece α en .05. El investigador cree entonces que la probabilidad de que el estadístico de prueba tome un valor en la región (o regiones) crítica(s) es .05. Pero si el modelo no representa exactamente la distribución muestral del estadístico, la probabilidad real podría ser un tanto diferente, digamos .25. Es obvio que en tal circunstancia el investigador no puede confiar en los resultados de una prueba de hipótesis.

¿En qué circunstancias el investigador puede estar seguro de que el modelo utilizado para la prueba de hipótesis dará probabilidades exactas en relación con la región (o regiones) crítica(s)? Estas circunstancias se conocen como las **suposiciones fundamentales** de una prueba. La prueba Z de una media tiene dos suposiciones fundamentales, a saber, las suposiciones **1.** de normalidad, y **2.** de independencia de las observaciones. A continuación comentaremos cada una de ellas.

Normalidad

La suposición de normalidad requiere que los datos de muestreo para la prueba se obtengan como resultado de la población normalmente distribuida. Pero, como usted sabe, la distribución normal es un modelo matemático con atributos muy específicos, uno de los cuales se extiende hacia el infinito en cualquier dirección. En realidad, las muestras nunca son tomadas de poblaciones con una distribución *perfectamente* normal, así que, en un sentido, esta suposición siempre se viola. Es más práctico, aunque un poco menos exacto, decir que la suposición de normalidad requiere que los datos se obtengan de una población *aproximadamente* normal. Más adelante abundaremos sobre este tema.

Independencia

La suposición de independencia requiere que cada observación en la muestra no esté relacionada con alguna otra observación en la muestra. Así, por ejemplo, la presión sanguínea de un miembro de la muestra no tiene influencia sobre la presión sanguínea de otro miembro de la muestra, como tampoco se ve influida por ésta. Asimismo, el hecho de que una presión sanguínea tenga un valor específico, digamos 120, no cambia la probabilidad de que la siguiente presión sanguínea tenga algún valor específico, digamos 95. Ejemplificar cómo puede violarse esta suposición ayudará a aclarar el concepto.

Suponga que se muestrean 10 pacientes oftalmológicos de una población. En los 10 pacientes se mide la agudeza visual para ambos ojos. De esta manera, la muestra consiste en 20 puntuaciones de agudeza visual; pero estas puntuaciones no son del todo independientes. Si se sabe que la agudeza visual en el ojo izquierdo del primer paciente es 20/200, esto puede (o no) ser indicativo de que la medición del ojo derecho también refleje agudeza reducida. Esto radica en el hecho de que una persona con agudeza visual reducida en un ojo a menudo tiene agudeza visual reducida en el otro, como ocurre con la retinopatía diabética o las cataratas.

En un segundo ejemplo suponga que un investigador selecciona aleatoriamente a grupos familiares para participar en un estudio sobre hipertensión juvenil. Una vez que un grupo familiar es seleccionado, se registran las presiones sanguíneas de todos los niños menores de 18 años en el grupo familiar. Es factible que estos datos violen la suposición de independencia porque, debido a semejanzas genéticas, es altamente posible que los niños en un grupo familiar dado presenten tendencias similares con respecto a las presiones sanguíneas registradas. La suposición de independencia puede ser violada en formas sutiles, así que los investigadores casi siempre tienen esta suposición en mente cuando utilizan métodos estadísticos que requieren independencia con el fin de asegurar exactitud en las probabilidades estimadas.

Consecuencias de las violaciones de la suposición. No es posible especificar las consecuencias que tendrán las violaciones a las suposiciones que fundamentan la prueba Z de una media, porque es imposible enumerar las formas que tomarán. Por ejemplo, las consecuencias de violar la suposición de normalidad dependen, entre otras cosas, de la forma de la población no normal. Pero hay un número infinito de tales formas. Sin embargo, una vez aclarado esto, es posible extraer algunas conclusiones generales mientras se tiene en mente que tales generalidades no se cumplen para circunstancias específicas.

Violación de la suposición de normalidad

Quizá la mejor forma de entender las consecuencias que tiene violar la suposición de normalidad es considerar cómo se relacionan con la población mostrada en la figura 4.2 de la página 78. Sin embargo, antes de hacer esto sería útil distinguir entre dos tipos de α.

Alfa nominal (α_N) es la probabilidad *pretendida* de que un estadístico de prueba tomará un valor en la región (o regiones) crítica(s) cuando la hipótesis nula sea verdadera. Es el tamaño de

TABLA 4.2: α_E para la prueba Z de dos colas para una media cuando $\alpha_N = .05$ y la muestra es de una población no normal.

n	Cola izquierda	Cola derecha	Combinadas (α_E)
5	.020	.026	.046
10	.024	.028	.052
20	.023	.026	.049
30	.024	.027	.051
50	.024	.026	.050
100	.024	.026	.050

la región (o regiones) crítica(s) en el modelo de prueba. Nosotros nos hemos referido previamente a este valor como α, pero usaremos α_N durante la discusión presente. **Alfa empírica** (α_E) es la probabilidad *real* de que un estadístico de prueba tome un valor en la región (o regiones) crítica(s) cuando la hipótesis nula es verdadera. Cuando se cumplen ambas suposiciones de la prueba Z de una media, $\alpha_E = \alpha_N$.

La tabla 4.2 expone los valores de α_E para la prueba Z de dos colas para una media cuando la muestra pertenece a la población de la figura 4.2. Si no hay violaciones, el valor que tome la probabilidad del estadístico de prueba en cada una de las dos regiones será .025. Por consiguiente, la diferencia entre la probabilidad observada y .025 en cada una de las colas se atribuye a la no normalidad de la población. Asimismo, en ausencia de violaciones la probabilidad combinada será .05.

Como se puede ver en esta tabla, para muestras con tamaño cinco

$$\alpha_E - \alpha_N = .046 - .050 = -.004$$

Cuando $\alpha_E < \alpha_N$ en una prueba se dice que es **conservativa** en la fase de alguna violación. Cuando $\alpha_E > \alpha_N$ se dice que la prueba es **liberal** o **anticonservativa**. Así, la prueba Z de una media es conservativa cuando $n = 5$, pero es liberal cuando $n = 10$. El punto importante por distinguir es que la diferencia entre α_E y α_N nunca es grande, especialmente cuando $n \geq 10$. Cuando la diferencia entre α_E y α_N es pequeña, se dice que la prueba es **robusta** ante la violación particular. ¿Qué tan pequeña debe ser esa diferencia para caracterizar a la prueba como robusta? Aunque existen muchas reglas[9] para hacer esta determinación, en gran medida es un asunto de percepción individual.

Observe también que las probabilidades en las dos colas no son iguales. Esto es típico cuando la muestra es de poblaciones asimétricas. En tales situaciones, a menudo ocurre que el nivel en una cola es conservativo, mientras que en la otra es liberal; así que la probabilidad combinada está en el nivel nominal (o cerca) de la prueba. Esto implica que la prueba Z de una media es más robusta para pruebas de dos colas que para las de una.

En resumen, la robustez de la prueba Z de una media es razonable en condiciones de normalidad y se incrementa conforme aumenta el tamaño de la muestra. Para poblaciones asimétricas, la prueba de dos colas es a menudo más robusta que la prueba de una cola. Las probabilidades empíricas no pueden ser satisfactorias cuando las muestras son pequeñas y/o las poblaciones son muy asimétricas.

[9] Una regla semejante declara que una prueba es robusta si $.5\alpha_N \leq \alpha_E \leq 1.5\alpha_N$.

TABLA 4.3: α_E para la prueba Z de dos colas para una media cuando $\alpha_N = .05$, los datos no son del todo independientes y la muestra es de una población normal.

	Pequeña			Moderada		
n	Cola izquierda	Cola derecha	Combi-nadas	Cola izquierda	Cola derecha	Combi-nadas
10	.037	.037	.074	.055	.055	.110
50	.037	.037	.074	.055	.055	.110
100	.037	.037	.074	.055	.055	.110

Violación de la suposición de independencia

A diferencia de la suposición de normalidad, generalmente, la prueba Z de una media no es robusta en condiciones de suposición de independencia. Como ejemplo, suponga que un investigador ingenuo realiza un estudio sobre agudeza visual. El investigador prueba la agudeza visual de ambos ojos en cinco pacientes y entonces lleva a cabo una prueba Z de dos colas para una media usando las 10 agudezas visuales. La tabla 4.3 muestra la probabilidad que tomará el valor en las dos regiones críticas cuando hay dependencias pequeñas y moderadas entre las agudezas visuales de los dos ojos.[10]

Como se observa en esta tabla, cuando se hace la prueba de las agudezas visuales de los 10 ojos y la dependencia es pequeña, la probabilidad en cada región crítica es de .037, lo que produce una probabilidad global de .074. Las probabilidades se incrementan a .055 en cada cola y .110 global cuando el grado de dependencia es moderado. Advierta que incrementar el número de ojos de 50 a 100 no tiene efectos en la mejora de los resultados de la prueba liberal. La falta de robustez de esta prueba se refleja en que la dependencia moderada duplica la probabilidad global pretendida de .05.

Las violaciones de la suposición de independencia también pueden causar resultados conservativos, aunque es más común obtener resultados liberales. Los puntos importantes por notar son: **1.** la prueba Z de una media generalmente no es robusta contra las violaciones de la suposición de independencia, y **2.** incrementar el tamaño de la muestra no reduce el efecto de la violación.

Resumen. La prueba Z de una media se utiliza para probar hipótesis nulas de la forma $H_0 : \mu = \mu_0$ contra las alternativas de una y dos colas. Para la alternativa de una cola

$$H_A : \mu < \mu_0$$

el valor-p se define como el área por debajo de $\bar{x}$, mientras que la Z crítica es el valor de Z que corta a α en la cola inferior. Para la alternativa

$$H_A : \mu > \mu_0$$

el valor-p es el área por debajo de $\bar{x}$ y la Z crítica corta a α en la cola superior. La alternativa de dos colas es

$$H_A : \mu \neq \mu_0$$

[10] Pequeña y moderada se definen operacionalmente como correlaciones de .2 y .5, respectivamente. Usted estudiará los coeficientes de correlación en el capítulo 8.

El valor-p para la alternativa de dos colas es el doble del área de la cola definida por $\bar{x}$. Hay dos valores críticos de Z, uno de los cuales corta a $\alpha/2$ en la cola superior, mientras que el otro corta la misma área en la cola inferior.

La prueba generalmente es robusta ante violaciones de la suposición de normalidad y esta característica va en aumento conforme el tamaño de la muestra se incrementa. Ninguna de estas cualidades es verdadera a tal grado que la suposición de independencia se vea afectada.

4.3.4 Prueba *t* de una media

Existe una dificultad asociada con el uso de la prueba Z de una media que tiene una base lógica y una práctica. La dificultad lógica reside en el hecho de que la prueba se realiza con el fin de determinar el parámetro desconocido μ. Pero para realizar la prueba, usted debe conocer α (véase la ecuación 4.7 en la página 90), que también es un parámetro. Es probable que si α es conocida, μ también lo sea, por esta razón se utiliza una prueba inferencial innecesaria. A pesar de que hay situaciones donde α es conocida y μ desconocida, éstas no son comunes. La dificultad práctica, entonces, es cómo llevar a cabo la prueba de hipótesis sin el conocimiento de α.

La clave fue presentada por William S. Gossett [45] (bajo el seudónimo "Student"), en 1908, con mejoras/modificaciones aportadas después por R. A. Fisher [17].

La solución implica el reemplazo del parámetro (generalmente) desconocido α por el estadístico conocido s en la expresión para el error estándar de $\bar{x}$. (Usted recordará que s es la desviación estándar de la muestra.) Esto da por resultado el estadístico

$$\boxed{t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}} \tag{4.8}$$

El modelo más apropiado para la distribución muestral de este estadístico no es la curva normal, sino más bien un modelo referido como la **distribución *t***. Para ser claros, la distribución t modela la distribución que se obtendrá si se extraen muestras repetidas de una población; t, como se representa en la ecuación 4.8, fue calculada para cada muestra, y los estadísticos resultantes se ordenaron entonces en una distribución de frecuencias relativas. La distribución tendrá media cero cuando la hipótesis nula sea verdadera y media $\mu - \mu_0$ cuando la hipótesis nula no sea verdadera. Aunque matemáticamente son distintas, en apariencia las curvas t son similares a la curva normal, como se aprecia en la figura 4.9.

En realidad, hay un número infinito de distintas distribuciones t, cada una de las cuales es identificada por sus *grados de libertad* (*degrees of freedom*, df). Una explicación profunda de los grados de libertad va más allá del alcance de este libro, pero en términos generales, podemos decir que los **grados de libertad** se refieren a la cantidad de información disponible en s para estimar α. Las curvas t con 3 y 20 grados de libertad se ilustran en la figura 4.9. Observe que la curva con 20 grados de libertad es más similar a la curva normal que la curva con 3 grados de libertad. En general, conforme los grados de libertad aumentan, la curva t asociada se hace muy similar a la curva normal. En teoría, con grados de libertad infinitos la curva t asociada es idéntica a la curva normal.

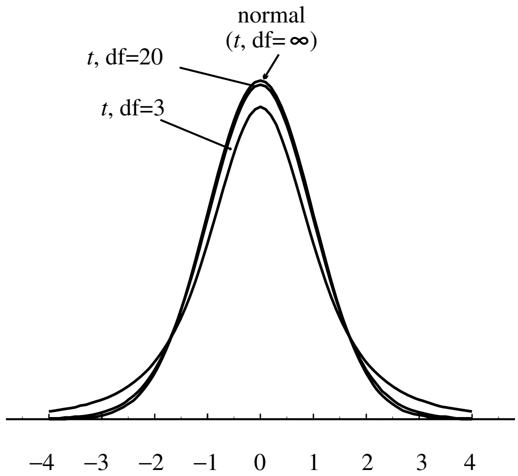


FIGURA 4.9: Curva normal con distribuciones *t*.

Con el fin de encontrar áreas por debajo de cualquier curva t dada, usted necesitará una tabla de áreas (similar a la tabla de la curva normal en el Apéndice A) *para esa curva particular*. Esto es impráctico porque estaremos trabajando con muchas curvas t diferentes. Por eficiencia, entonces, las tablas t no proporcionan áreas como lo hace la tabla de la curva normal, pero dan valores críticos comúnmente empleados para cada curva.

Considere la tabla t en el Apéndice B. Las primeras dos filas de esta tabla se refieren a los intervalos de confianza, que estudiaremos en la última parte de este capítulo. La tercera y cuarta filas dan valores de α para pruebas de hipótesis de una y dos colas, respectivamente. Los números en la parte principal de la tabla son los valores críticos apropiados para cada valor de α. La tabla se utiliza de la siguiente manera.

Suponga que usted desea encontrar el valor crítico para una prueba de una cola con hipótesis alternativa $H_A : \mu > \mu_0$ en $\alpha = .05$ para una distribución t con tres grados de libertad. Al localizar .05 en la tercera fila y leyendo hacia abajo en la fila de tres grados de libertad se observa el valor crítico apropiado como 2.353. Esto significa que un valor t de 2.353 corta a .05 en la cola superior de una curva t con tres grados de libertad. Una versión de la cola inferior de la prueba requerirá un valor crítico de –2.353.

¿Cuál será el valor crítico para una prueba t de dos colas en $\alpha = .05$ para una distribución t con 10 grados de libertad? Al localizar .05 en la cuarta fila y leyendo hacia abajo en la fila de 10 grados de libertad se puede ver el valor 2.228. Observe que 2.228 corta a .025 en la cola de la curva. Los valores críticos para la prueba de dos colas serán entonces ±2.228.

Los grados de libertad se calculan de modo distinto para pruebas diferentes. Para la prueba t de una media, los grados de libertad son $n - 1$, donde n es el número de observaciones en la muestra.

Excepto para la forma en la cual la t obtenida es calculada, los mecanismos de conducción de una prueba t de una media son los mismos que para la prueba Z de una media. La t obtenida es calculada y comparada con la t crítica. Si la t obtenida está en la región crítica, la hipótesis nula se rechaza. De lo contrario, la hipótesis nula se sostiene. Puesto que no tenemos tablas de las áreas de varias distribuciones t, no podremos usar el método del valor-p versus alfa y nos veremos restringidos a probar mediante el método de la t obtenida versus la t crítica. Cuando se realizan pruebas mediante un paquete computacional, el valor-p será suministrado por el software, así que el método del valor-p versus alfa es el método empleado generalmente.

EJEMPLO 4.14

Utilice los datos de la muestra presentados aquí para llevar a cabo la prueba t de una media indicada.

$H_0 : \mu = 8.0$ Muestra: 6.0 8.0 5.5 4.5 8.5 4.0 3.5
$H_A : \mu < 8.0$
$\alpha = .01$

Solución Es posible obtener la desviación estándar de la muestra mediante la aplicación de la ecuación 2.16 a las puntuaciones sumadas y la suma de las puntuaciones al cuadrado como se observa abajo.

X	X^2
6.0	36.00
8.0	64.00
5.5	30.25
4.5	20.25
8.5	72.25
4.0	16.00
3.5	12.25
40.0	251.00

$$s = \sqrt{\frac{\sum x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{251 - \frac{(40)^2}{7}}{7-1}} = \sqrt{\frac{22.429}{6}} = 1.933$$

y

$$\bar{x} = \frac{40}{7} = 5.714.$$

Por la ecuación 4.8 la t obtenida es entonces

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{5.714 - 8.00}{\frac{1.933}{\sqrt{7}}} = \frac{-2.286}{.731} = -3.127$$

Como se puede ver en la tabla del Apéndice B, la prueba de una cola con grados de libertad 7 − 1 = 6 llevada a cabo en $\alpha = .01$ tiene un valor asociado de 3.143. Como se indica en la hipótesis alternativa, la región crítica está en la cola inferior, así que el valor crítico es −3.143. Como el valor obtenido de −3.127 es mayor que el valor crítico, la hipótesis nula no se rechaza. El vínculo del valor obtenido con el valor crítico se muestra en el panel A de la figura 4.10. ■

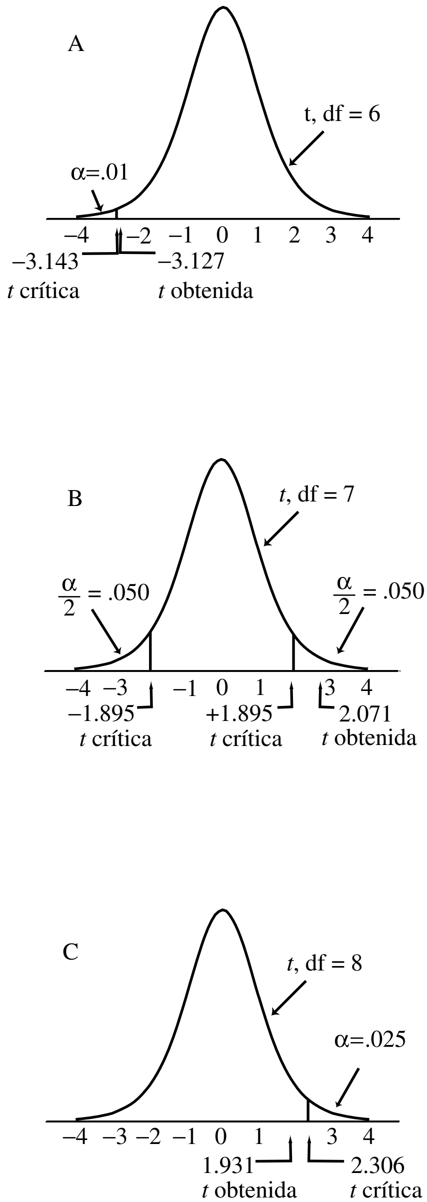


FIGURA 4.10: Ubicaciones de los valores obtenido y crítico para tres pruebas *t* de una media.

EJEMPLO 4.15

Utilice los datos de la muestra presentados aquí para llevar a cabo la prueba t de una media indicada.

$H_0: \mu = 0.0$ Muestra: –0.5 1.0 2.5 –1.0 3.5 –0.5 3.0 2.5
$H_A: \mu \neq 0.0$
$\alpha = .10$

Solución Las sumas necesarias para los cálculos de s y $\bar{x}$ son

X	X^2
–0.5	0.25
1.0	1.00
2.5	6.25
–1.0	1.00
3.5	12.25
–0.5	0.25
3.0	9.00
2.5	6.25
10.5	36.25

Así que

$$s = \sqrt{\frac{\sum x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{36.25 - \frac{(10.50)^2}{8}}{8-1}} = \sqrt{\frac{22.469}{7}} = 1.792$$

y

$$\bar{x} = \frac{10.5}{8} = 1.313.$$

La t obtenida es entonces

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{1.313}{\frac{1.792}{\sqrt{8}}} = \frac{1.313}{.634} = 2.071.$$

La t crítica para una prueba de dos colas en $\alpha = .10$ y siete grados de libertad es ±1.895. Como 2.071 es mayor que 1.895, la hipótesis nula se rechaza. Las posiciones relativas de los valores crítico y obtenido de t en el modelo de la prueba se muestran en el panel B de la figura 4.10. ■

EJEMPLO 4.16

Utilice los datos de la muestra presentado aquí para realizar la prueba t de una media indicada.

$H_0: \mu = 20$ Muestra: 22 19 17 26 21 20 29 27 22
$H_A: \mu > 20$
$\alpha = .025$

Solución Las sumas necesarias para el cálculo de s son

X	X^2
22	484
19	361
17	289
26	676
21	441
20	400
29	841
27	729
22	484
203	4705

así que

$$s = \sqrt{\frac{4705 - \frac{(203)^2}{9}}{9-1}} = \sqrt{\frac{126.222}{8}} = 3.972$$

y

$$\bar{x} = \frac{203}{9} = 22.556.$$

La t obtenida es entonces

$$t = \frac{22.556 - 20.000}{\frac{3.972}{\sqrt{9}}} = \frac{2.556}{1.324} = 1.931.$$

La t crítica para una prueba de una cola en $\alpha = .025$ y ocho grados de libertad es 2.306. Como la t obtenida de 1.931 es menor que la t crítica de 2.306, la hipótesis nula no se rechaza. Las posiciones relativas de los valores crítico y obtenido de t en el modelo de la prueba se muestran en el panel C de la figura 4.10. ■

Suposiciones que fundamentan la prueba *t* de una media. Las suposiciones que fundamentan la prueba t de una media son las mismas que aquellas para la prueba Z de una media: normalidad de la población e independencia de las observaciones.

Consecuencias de las violaciones de las suposiciones. Las violaciones de la suposición de independencia para la prueba t de una media producen las mismas consecuencias que para la prueba Z de una media y no serán reafirmadas aquí. La tabla 4.4 muestra α_E para la prueba t de una media realizada con las mismas muestras que se utilizaron en la construcción de la tabla 4.2.

Una comparación de la tabla 4.4 con la tabla 4.2 muestra los mismos patrones generales para la prueba t que se vieron para la prueba Z. Esto es, **1.** falta de simetría para los resultados en dos colas debido al muestreo de una población asimétrica, **2.** resultados más cercanos a α_N para los dos lados combinados que para las colas individuales, y **3.** el aumento de robustez con el incremento del tamaño de la muestra. También es notable el hecho de que la prueba t por lo general no es tan robusta como la prueba Z.

Resumen. La prueba t de una media difiere de una prueba Z de una media en que la desviación estándar de la muestra (s), más que la desviación estándar de la población (σ), se utiliza en la ecuación para el valor obtenido. Como resultado de esta sustitución, el modelo apropiado para

TABLA 4.4: α_E para la prueba *t* de dos colas para una media cuando $\alpha_N = 0.05$ y la muestra es de una población no normal.

n	Cola izquierda	Cola derecha	Combinadas (α_E)
5	.020	.026	.046
10	.024	.028	.052
20	.023	.026	.049
30	.024	.027	.051
50	.024	.026	.050
100	.024	.026	.050

la distribución muestral del estadístico t es la familia de las distribuciones t, las cuales fueron catalogadas por sus grados de libertad. La prueba t comparte la falta de robustez de la prueba Z a las violaciones de la suposición de independencia y es un tanto menos robusta que la prueba Z a las violaciones de la suposición de normalidad.

4.3.5 Pruebas de una muestra para una proporción

En este apartado discutiremos dos pruebas para la proporción de una población. La primera, a la cual nos referiremos como la prueba *exacta*, está basada en la distribución binomial (véase apartados 4.2.4 y 4.2.5 en las páginas 80 y 81, respectivamente), mientras que la segunda, a la cual nos referiremos como la prueba *aproximada*, está basada en la curva normal (véase apartado 4.2.6 en la página 84). La razón para esta terminología proviene de que, si se satisfacen ciertas suposiciones, la distribución binomial da probabilidades exactas; mientras que el método de la curva normal necesariamente produce un resultado aproximado.

Ambos métodos pueden utilizarse para probar la hipótesis nula

$$H_0 : \pi = \pi_0$$

donde π es la proporción de las observaciones en una población que encuentra algunos criterios especificados y π_0 es la proporción hipotética para la misma población. Las alternativas de una cola toman la forma

$$H_A : \pi < \pi_0$$

o

$$H_A : \pi > \pi_0$$

mientras la alternativa de dos colas es expresada como

$$H_A : \pi \neq \pi_0$$

La prueba exacta.

Prueba de una cola

Usted recordará de las secciones 4.2.4 y 4.2.5 que la distribución binomial puede utilizarse para modelar la distribución muestral de $\hat{p}$. De manera natural, de aquí se infiere que el mismo modelo puede emplearse como base para pruebas de hipótesis.

TABLA 4.5: Distribución muestral de $\hat{p}$ para $n = 10$ y $\pi = .38$.

Proporción $\hat{p}$	Número de éxitos	Probabilidad $P(y)$
.00	0	.00839
.10	1	.05144
.20	2	.14188
.30	3	.23189
.40	4	.24872
.50	5	.18293
.60	6	.09343
.70	7	.03272
.80	8	.00752
.90	9	.00102
1.00	10	.00006

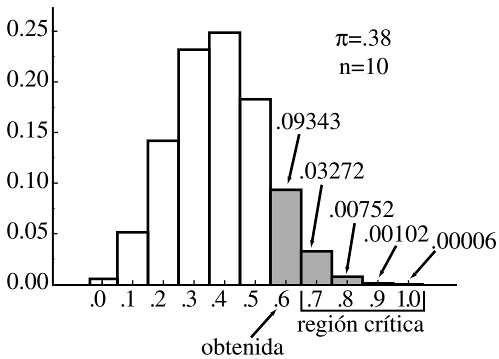


FIGURA 4.11: Prueba binomial con alternativa H_A : $\pi > .38$.

Por ejemplo, suponga que usted muestrea aleatoriamente 10 observaciones de una población dicotómica y desea llevar a cabo la siguiente prueba.

$$H_0 : \pi = .38 \qquad \hat{p} = .60 \qquad n = 10$$
$$H_A : \pi > .38 \qquad \alpha = .05$$

Usted reconocerá este escenario como el estudio de los donadores de sangre islandeses descrito en el ejemplo 4.6 de la página 84. Se utilizó la ecuación 4.5 para construir la distribución muestral de $\hat{p}$ señalada en la tabla 4.5. Al igual que sucede con todas las pruebas de hipótesis, esta distribución se construyó de acuerdo con la suposición de una hipótesis nula verdadera (esto es, que $\pi =$ .38). Esta distribución se presenta gráficamente en la figura 4.11.

Distribuciones discretas como la binomial presentan ciertas dificultades que deben ser reconocidas. Para la presente prueba de una cola, hemos indicado la región crítica en la figura 4.11. Observe que la probabilidad de $\hat{p}$ tomando un valor en la región crítica cuando la hipótesis nula es verdadera no es .05 como se pretendía, sino .03272 + .00752 + .00102 + .00006 = .04132. Si la región crítica se expandiera para incluir $\hat{p} = .60$, α sería .04132 + .09343 = .013475, la cual es mucho mayor que la pretendida .05. Cuando trabajamos con distribuciones muestrales discretas, la región crítica se construye así para abarcar la probabilidad más grande posible que no exceda la α pretendida. En este caso, ese valor es .04132. Note que para una prueba con hipótesis alternativa HA : $\pi > .38$, la región crítica contendrá únicamente $\hat{p} = .0$, que tiene una probabilidad asociada de .00839 (véase la tabla 4.5). Al expandir la región e incluir $\hat{p} = .1$ la probabilidad es .00839 + .05144 = .05983, que es mayor que la α pretendida.

El valor-*p* para la prueba de una cola con alternativa de la forma $H_A : \pi > \pi_0$ es la probabilidad de obtener un valor de $\hat{p}$ más grande o igual que el valor realmente observado. En este caso la $\hat{p}$ obtenida es de .60, así que el valor-*p* es .09343 + .03272 + .00752 + .00102 + .00006 = .13475. Como este valor es mayor que α, la hipótesis nula no se rechaza. El valor-*p* se muestra en la región sombreada en la figura 4.11.

Una vez que se ha definido la región crítica, puede utilizarse también para la prueba la metodología del valor obtenido versus el valor crítico. El valor crítico es simplemente el valor de $\hat{p}$ que define el área más grande posible de la cola que no exceda α. Como se observa en la figura 4.11, el valor crítico $\hat{p}$ es .70 en el presente ejemplo. Puesto que la $\hat{p}$ obtenida de .60 es menor que .70, la hipótesis nula no se rechaza.

Considere ahora la prueba

$$\begin{array}{lll} H_0 : \pi = .45 & \hat{p} = .00000 & n = 7 \\ H_A : \pi > .45 & \alpha = .20 & \end{array}$$

El valor-*p* para una prueba con alternativa de la forma $H_A : \pi < \pi_0$ es la probabilidad de obtener un valor de $\hat{p}$ menor que o igual al valor observado. En este caso el valor-*p* será $P(0)$, el cual mediante la ecuación 4.5 es .01522. Debido a que el valor-*p* de .01522 es menor que $\alpha = .20$ la hipótesis nula se rechaza.

Se puede obtener el valor crítico de $\hat{p}$ notando que $P(0) + P(1) = .01522 + .08719 = .10241 < \alpha$ y $P(0) + P(1) + P(2) = .01522 + .08719 + .21402 = .31643 > \alpha$. Por consiguiente, $\hat{p}$ crítica es la proporción asociada con un éxito, que es 1/7 = .14286. Como la proporción obtenida de .00000 es menor que la proporción crítica de .14286, la hipótesis nula se rechaza. La relación entre el valor-*p*, α, $\hat{p}$ obtenida y crítica se muestra en la figura 4.12, donde el valor-*p* es sombreado.

EJEMPLO 4.17

¿Cuál es el menor nivel posible de α que puede utilizarse para la siguiente prueba de hipótesis?

$$H_0 : \pi = .65 \qquad H_A : \pi * > .65 \qquad n = 9$$

¿Cuál sería el menor nivel posible de α si la alternativa fuera $H_A : \pi < .65$?

Suponga que un investigador decide probar la hipótesis nula de arriba contra la alternativa $H_A : \pi > .65$ utilizando $\alpha = .15$. ¿Cuál será el nivel *real* de α? ¿Cuál será este valor para la hipótesis alternativa $H_A : \pi < .65$?

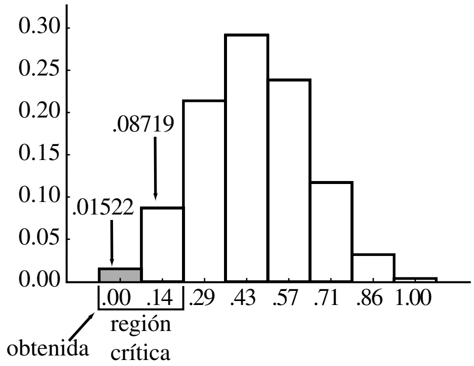


FIGURA 4.12: Prueba binomial con hipótesis alternativa $H_A : \pi < .45$.

Solución El menor nivel posible para la alternativa $H_A : \pi > .65$ es, mediante la ecuación 4.5,

$$P(9) = \frac{9!}{9!(0)!}.65^9(1-.65)^0 = .65^9 = .02071$$

mientras que para la alternativa de la cola inferior es

$$P(0) = \frac{9!}{0!(9)!}.65^0(1-.65)^9 = .35^9 = .00008.$$

El nivel real de la prueba de la cola superior está definido por la mayor área posible de la cola superior que no excede la α pretendida. Este valor puede obtenerse haciendo notar que P (7) + P (8) + P (9) = .21619 + .10037 + .02071 = .33727, el cual es mayor que la α pretendida de .15. En contraste, P (8) + P (9) = .10037 + .02071 = .12108, que es menor que la α pretendida. Por lo tanto, el nivel real de la prueba de la cola superior será .12108.

El nivel real de la prueba de la cola inferior será .05359 puesto que P (0) + P (1) + P (2) + P (3) + P (4) = .00008 + .00132 + .00979 + .04241 + .11813 = .17173, el cual es mayor que α = .15, mientras que P (0) + P (1) + P (2) + P (3) = .00008 + .00132 + .00979 + .04241 = .05359, que es menor que .15. ■

EJEMPLO 4.18

Lleve a cabo la siguiente prueba de hipótesis. Reporte los resultados para ambos métodos del valor-*p* versus alfa y $\hat{p}$ obtenida versus $\hat{p}$ crítica. ¿Cuál será el resultado si α se estableciera en .01?

$$H_0 : \pi = .05 \qquad \hat{p} = 1.0 \qquad n = 5$$
$$H_A : \pi > .50 \qquad \alpha = .05$$

Solución Como la hipótesis alternativa especifica una prueba de cola superior, el valor-*p* es la probabilidad de obtener un valor de $\hat{p}$ mayor que o igual al valor realmente observado. Debido a que $\hat{p}$ = 10, el valor-*p* es P (5) = .03125, el cual es menor que α = .05, así que la hipótesis nula se rechaza.

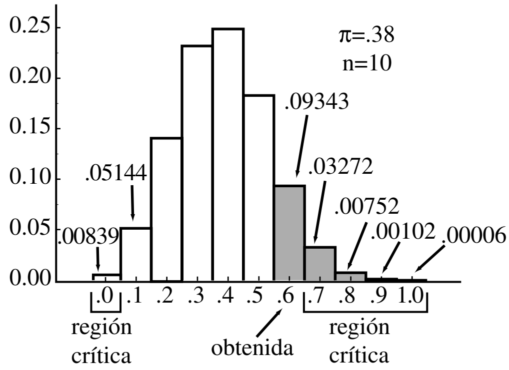


FIGURA 4.13: Prueba binomial con alternativa $H_A: \pi \neq 0.38$, $\alpha = .10$.

Puesto que $P(4) + P(5) = .15625 + .03125 = .18750 > \alpha = .05$ y $P(5) = .03125 < \alpha = .05$, $\hat{p} = 5/5$ crítica $= 5/5 = 1.0$. Dado que $\hat{p}$ obtenida es igual a la $\hat{p}$ crítica, la hipótesis nula se rechaza. ■

Prueba de dos colas

Así como con las pruebas t y Z, las pruebas de hipótesis de dos colas en la distribución binomial emplean regiones críticas en ambas colas de la distribución. Sin embargo, a diferencia de aquellas pruebas no hay un método generalmente aceptado para la construcción de dichas regiones o el cálculo de los valores-p de dos colas. El método que presentamos aquí se encuentra entre los más utilizados.[11]

Puesto que la distribución binomial es discreta, por lo general no es posible construir regiones críticas con probabilidades asociadas exactamente igual a $\alpha/2$, o incluso tener probabilidades iguales en las dos regiones.[12] La estrategia entonces es construir regiones cuyas probabilidades asociadas estén tan cerca de $\alpha/2$ como sea posible sin exceder este valor. Por ejemplo, considere la prueba

$$\begin{array}{lll} H_0: \pi = .38 & \hat{p} = .60 & n = 10 \\ H_A: \pi \neq .38 & \alpha = .10 & \end{array}$$

La distribución muestral para esta prueba se presenta en la tabla 4.5. Una descripción de las regiones críticas y la mitad del valor-p se muestran en la figura 4.13.

Observe primero que la parte izquierda de la región crítica tiene una probabilidad asociada de .00839. Si expandiéramos esta región para incluir $\hat{p}$ de .1, la probabilidad de que $\hat{p}$ caiga en esta región bajo una hipótesis nula verdadera sería $.00839 + .05144 = .05983$, que es mayor que $\alpha/2 = .05$. La región de la parte derecha tiene una probabilidad de $.00006 + .00102 + .00752 + .03272 = .04132$. De nuevo, no podemos expandir esta región sin exceder $\alpha/2$. Esto significa que la probabilidad de un error tipo I no es .10 como se pretendía, sino únicamente es $.00839 + .04132 = .04971$. También se infiere que la $\hat{p}$ crítica es .00 y .7.

---

[11] En el capítulo 10 presentaremos un método diferente en conjunción con la prueba exacta de Fisher. El método demostrado ahí también es aplicable a este problema.

[12] Las dos regiones serán equilibradas cuando π sea igual a .5.

TABLA 4.6: Distribuciones muestrales de $\hat{p}$ para n = 8 y π = .35, .50 y .55.

Proporción $\hat{p}$	Número de éxitos	$\pi = .35$ $P(y)$	$\pi = .50$ $P(y)$	$\pi = .55$ $P(y)$
.000	0	.03186	.00391	.00168
.125	1	.13726	.03125	.01644
.250	3	.25869	.10937	.07033
.375	4	.27859	.21875	.17192
.500	5	.18751	.27344	.26266
.625	6	.08077	.21875	.25683
.750	7	.02175	.10937	.15695
.875	8	.00335	.03125	.05481
1.000	9	.00023	.00391	.00837

El valor-*p* bajo este esquema se obtiene duplicando el valor-*p* de una cola con el valor-*p* de una cola definido como la más pequeña de las probabilidades de dos colas.[13] En este caso el valor-*p* será

$$2(.00006 + .00102 + .00752 + .3272 + .09343) = .26950.$$

Observe que cualquier valor de $\hat{p}$ que cae en la región crítica producirá un valor-*p* menor que α, mientras que cualquier $\hat{p}$ que caiga afuera de una región crítica producirá un valor-*p* mayor que α.

La prueba de hipótesis se lleva a cabo a través del método del valor-*p* versus α, poniendo énfasis en que el valor-*p* de .26950 es mayor que $\alpha = .10$, así que la hipótesis nula no se rechaza. Se llega a la misma conclusión por el método de la $\hat{p}$ obtenida versus la $\hat{p}$ crítica, subrayando que la $\hat{p}$ obtenida de .6 está entre los valores críticos de .0 y .7. De esto se infiere que .6 no está en ninguna región crítica.

EJEMPLO 4.19

Utilice las probabilidades binomiales en la tabla 4.6 para realizar la siguiente prueba de hipótesis. Reporte los resultados para ambos métodos, el valor-*p* versus alfa y la $\hat{p}$ obtenida versus la $\hat{p}$ crítica.

$$\begin{array}{lll} H_0 : \pi = .35 & \hat{p} = .25 & n = 8 \\ H_A : \pi \neq .35 & \alpha = .05 & \end{array}$$

Solución Como se observa en el panel A de la figura 4.14, no hay región crítica en la cola inferior de la distribución. Esto sucede porque el valor más pequeño posible de $\hat{p}$ (esto es, 0) tiene probabilidad de .03186, el cual es mayor que $\alpha/2 = .05/2 = .025$. La región de la cola superior consiste en valores de $\hat{p}$ de .875 y 1.000. Una $\hat{p}$ de .750 no puede incluirse en la región crítica porque las probabilidades sumadas .00023 + .00335 + .02175 = .02533 excederán $\alpha/2 = .025$.

[13] Esto es, la más pequeña de las dos probabilidades obtenidas cuando las probabilidades sumadas de los valores $\hat{p}$ menores que o iguales al valor obtenido de $\hat{p}$ se comparan con aquellos valores obtenidos de la suma de probabilidades de $\hat{p}$ que son más grandes o iguales que la $\hat{p}$ obtenida.

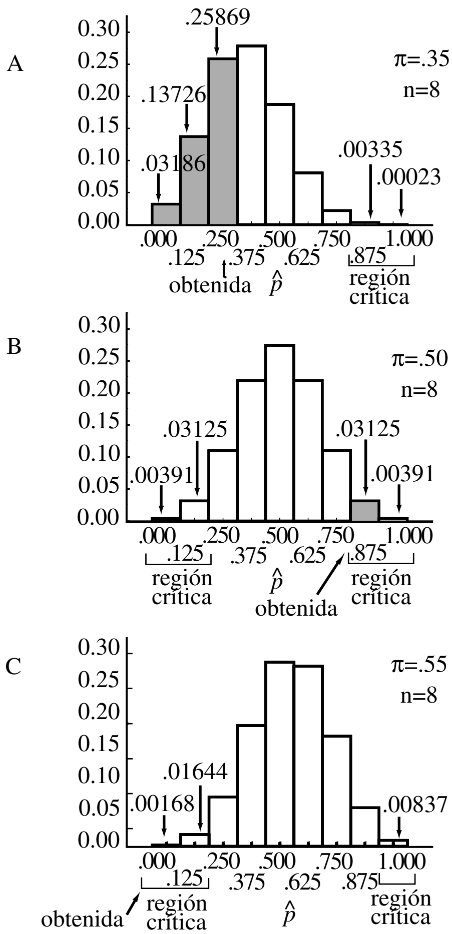


FIGURA 4.14: Pruebas binomiales de dos colas.

Hay un solo valor crítico de $\hat{p}$ que es .875. La $\hat{p}$ obtenida de .250 no está en la región crítica, así que la hipótesis nula no se rechaza. Calculamos el valor-*p* como dos veces el valor-*p* de una cola o 2(.03186 + .13726 + .25869) = .85562, el cual excede $\alpha = .05$, así que la hipótesis nula no se rechaza. ■

EJEMPLO 4.20

Utilice las probabilidades binomiales en la tabla 4.6 para llevar a cabo la siguiente prueba de hipótesis. Reporte los resultados para ambos métodos, el valor-*p* versus alfa y la $\hat{p}$ obtenida versus la crítica.

$$H_0 : \pi = .50 \qquad \hat{p} = .875 \qquad n = 8$$
$$H_A : \pi \neq .50 \qquad \alpha = .10$$

Solución Como se puede ver en el panel B de la figura 4.14, la región crítica de la cola inferior consiste en los valores .000 y .125, mientras que la región de la cola superior consiste en los valores .875 y 1.000. Observe que ninguna de estas regiones puede incrementarse sin las probabilidades asociadas que exceden de $\alpha/2 = .10/2 = .05$. Los valores críticos de $\hat{p}$ son entonces .125 y .875. Debido a que la $\hat{p}$ obtenida de .875 está en la región crítica superior, la hipótesis nula se rechaza.

El valor-*p* de dos colas es el doble del valor-*p* de una cola o

$$2\,(.00391 + .03125) = .07032.$$

Como esto es menor que $\alpha = .10$, la hipótesis nula se rechaza. El valor-*p* de una cola se representa como la porción sombreada del panel B en la figura 4.14. ■

EJEMPLO 4.21

Utilice las probabilidades binomiales en la tabla 4.6 para llevar a cabo la siguiente prueba de hipótesis. Reporte los resultados para ambos métodos, el valor-*p* versus alfa y la obtenida versus la $\hat{p}$ crítica.

$$\begin{array}{lll} H_0 : \pi = .55 & \hat{p} = .000 & n = 8 \\ H_A : \pi \neq .55 & \alpha = .05 & \end{array}$$

Solución Como se observa en el panel C de la figura 4.14, la región crítica de la cola inferior consiste en los valores .000 y .125, mientras que la región de la cola superior consiste únicamente en 1.000. Observe que ninguna de estas regiones puede incrementarse sin las probabilidades asociadas que exceden de $\alpha/2 = .05/2 = .025$. Los valores críticos de $\hat{p}$ son, entonces, .125 y 1.000. Debido a que la $\hat{p}$ obtenida de .000 está en la región crítica inferior, la hipótesis nula se rechaza.

El valor-*p* de dos colas es el doble del valor-*p* de una cola o 2(.00168) = .00336. Como esto es menor que $\alpha = .05$, la hipótesis nula se rechaza. El valor-*p* de una cola se representa como la porción sombreada del panel C en la figura 4.14. ■

Prueba de aproximación. Como usted aprendió en el apartado 4.2.6, la curva normal puede utilizarse para aproximar la distribución muestral de $\hat{p}$ siempre y cuando el tamaño de la muestra sea lo suficientemente grande. De esto se infiere que en tal circunstancia la curva normal puede utilizarse como la base de las pruebas de hipótesis. Estas pruebas son realizan de una manera similar a la utilizada para la prueba *Z* de una media, con la diferencia primordial de que la *Z* obtenida se calcula mediante

$$\boxed{Z = \frac{\overline{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}} \tag{4.9}$$

que es simplemente la ecuación 4.6 con π_0 sustituida por π para indicar que se está utilizando un valor hipotético de π en lugar de π. (Antes de continuar, usted deseará repasar el apartado 4.3.3 en la página 89). La prueba se demuestra abajo.

EJEMPLO 4.22

Utilice la siguiente información para llevar a cabo una prueba aproximada de la hipótesis nula enunciada. Justifique su decisión respecto de la hipótesis nula con base en ambos métodos, el valor-*p* obtenido versus alfa y *Z* obtenida versus *Z* crítica.

$$\begin{array}{lll} H_0 : \pi = .20 & \hat{p} = .217 & n = 350 \\ H_A : \pi > .20 & \alpha = .01 & \end{array}$$

Solución Por la ecuación 4.9 la *Z* obtenida es

$$Z = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{.217 - .200}{\sqrt{\frac{(.20)(.80)}{350}}} = .80.$$

La columna 3 del Apéndice A muestra que el área arriba de $Z = .80$ es .2119. Como este valor es mayor que $\alpha = .01$, la hipótesis nula no se rechaza. La columna 3 de la tabla de la curva normal también muestra que el valor más cercano a 0.01 es .0099, el cual tiene un valor de *Z* asociado de 2.33. Por lo tanto, el valor *Z* en el límite delantero de la región crítica es 2.33. Puesto que la *Z* obtenida de .80 es menor que la *Z* crítica de 2.33, la hipótesis nula no se rechaza.

EJEMPLO 4.23

Utilice la siguiente información para realizar una prueba aproximada de la hipótesis nula enunciada. Justifique su decisión respecto de la hipótesis nula con base en ambos métodos, el valor-*p* obtenido versus alfa y *Z* obtenida versus *Z* crítica.

$$H_0: \pi = .58 \qquad \hat{p} = .53 \qquad n = 400$$
$$H_A: \pi \neq .58 \qquad \alpha = .05$$

Solución Por la ecuación 4.9 la *Z* obtenida es

$$Z = \frac{\bar{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{.53 - .58}{\sqrt{\frac{(.58)(.42)}{400}}} = -2.03.$$

La columna 3 del Apéndice A muestra que el área debajo de $Z = -2.03$ es .0212. Multiplicando esta área por dos se obtiene un valor-*p* de .0424. Dado que este valor es menor que $\alpha = .05$, la hipótesis nula se rechaza. La columna 3 de la tabla de la curva normal también muestra que el área .025 de la cola ($\alpha/2$) tiene una puntuación *Z* asociada de 1.96. Los valores de *Z* crítica para la prueba de dos colas son entonces ±1.96. Puesto que la *Z* obtenida de −2.03 es menor que la *Z* crítica de −1.96, la hipótesis nula se rechaza. ■

Suposiciones y consecuencias de las violaciones para la prueba de una muestra de una proporción. La distribución muestral de $\hat{p}$ es modelada aproximadamente por la distribución binomial cuando las observaciones de éxito/fallo extraídas de una población dicotómica son independientes. La prueba exacta de $\hat{p}$, por lo general, no es robusta contra las violaciones de la suposición de independencia, así que tales violaciones pueden producir resultados incorrectos. Las violaciones de la suposición de independencia pueden ocurrir en circunstancias similares a aquellas discutidas en la página 98 en relación con la prueba *Z* de una media.

La prueba aproximada es válida de acuerdo con las suposiciones de que las observaciones son independientes y que el muestreo es de una población normalmente distribuida. Obviamente, la suposición de normalidad siempre se viola por esta prueba, dado que la muestra de una población es dicotómica. (Véase el panel A de la figura 4.3 en la página 81.) Por consiguiente, la prueba aproximada se utiliza de forma adecuada cuando el tamaño de la muestra es lo suficientemente grande como para asegurar la normalidad aproximada de la distribución muestral de $\hat{p}$ mediante el teorema del límite central. Como discutimos en el apartado 4.2.6, una regla comúnmente invocada afirma que el modelo de la curva normal será satisfactorio mientras $n\pi$ y $n(1 - \pi)$ sean mayores que o iguales a cinco. Una regla más conservadora afirma que el criterio debe ser 10, en vez de cinco. Puede utilizarse una corrección de continuidad (véase el apartado 4.2.6 en la página 84) en conjunción con la prueba aproximada en un intento por mejorar la aproximación, pero estudios hechos por Ramsey y Ramsey [39] desalientan esta práctica.

La prueba aproximada generalmente no es robusta contra las violaciones de la suposición de independencia y es probable que genere resultados erróneos en tales casos.

4.3.6 Pruebas de equivalencia

La prueba de equivalencia es un método para probar más que un procedimiento estadístico específico. De esta forma, las pruebas que usted ha estudiado hasta ahora, así como muchas otras, pueden utilizarse como pruebas de equivalencia. De aquí se infiere que la prueba de equivalencia puede tratar con medias de la población, proporciones y otros parámetros. Por eficiencia, nosotros restringimos esta explicación a las medias y nos ocuparemos de las proporciones después. Consideraremos otros parámetros en capítulos posteriores.

El fundamento para la prueba de equivalencia proviene del hecho de que las pruebas estadísticas estándar están diseñadas para establecer lo que *no* es verdadero en vez de lo que *sí* lo es. Por eso, nosotros entendemos que cuando una hipótesis nula se rechaza usted puede estar seguro de que la hipótesis nula no es verdadera. En contraste, cuando la hipótesis nula no se rechaza usted no puede estar seguro de que la hipótesis nula *es* verdadera. Después aprenderá más acerca de por qué esto es verdad.

También hablaremos acerca de la utilidad de la prueba de equivalencia en el capítulo 5, pero en este momento consideraremos un ejemplo hipotético. Suponga que un fármaco particular, al cual llamaremos fármaco A, se utiliza para tratar la ansiedad. El objetivo del tratamiento es reducir la ansiedad al nivel normal, pero no tanto como para que el paciente se vuelva insensible a las amenazas de su ambiente. El fármaco A es un tratamiento efectivo que muestra numerosos estudios basados en la escala de ansiedad XYZ. Estos estudios muestran que pacientes que tomaron el fármaco A tienen una puntuación media de 80.4 en la escala de ansiedad XYZ, lo cual se considera muy satisfactorio. Un promedio muy por arriba de 80.4 indicará que el fármaco deja demasiada ansiedad en los pacientes, mientras que el promedio muy por debajo de 80.4 mostrará demasiada insensibilidad.

El fármaco A es muy efectivo en el tratamiento de la ansiedad, pero tiene muchos efectos secundarios y por esta razón ha sido reemplazado por el fármaco B, el cual no produce los efectos indeseables. Se realizará un estudio para saber si el fármaco B controla la ansiedad como lo hace el fármaco A. Específicamente el estudio está diseñado para determinar si la población de pacientes que consumen el fármaco B manifiesta una puntuación media en la escala de ansiedad XYZ de 80.4.

Observe que este estudio intenta mostrar que el valor de la media de la población *es* 80.4, en lugar de mostrar que *no es* 80.4. Nosotros designaremos este valor como μ_0. Si el propósito del estudio fuera mostrar que μ_0 no es la media de la población, se emplearía una prueba estándar de significancia como las pruebas Z o t de una media. Como la meta es mostrar que μ_0 es la media, se utilizará una prueba de equivalencia.

El primer paso en la prueba de equivalencia es definir un **intervalo de equivalencia** (*equivalence interval*, *EI*). El *EI* es un conjunto de valores alrededor de μ_0 que está lo bastante cerca de μ_0 como para producir en esencia el mismo resultado que se obtendría si la media fuera μ_0. Por ejemplo, los expertos pueden decidir que un valor de la media poblacional que está a menos de tres puntos de 80.4 no será en términos médicos significativo para 80.4. Por consiguiente, se decide que si la media poblacional es cualquier valor entre 77.4 y 83.4, el fármaco B será considerado, en términos de efectividad, *equivalente* al fármaco A. Esta forma de equivalencia no deberá confundirse con la **bioequivalencia**, la cual se refiere a la equivalencia en el índice y la magnitud de absorción de los fármacos.

La estrategia utilizada para demostrar que el fármaco B produce resultados equivalentes a los del fármaco A pretende mostrar que el valor de la media de la población tratada con el fármaco B es menor que 83.4 y mayor que 77.4. En otras palabras, pretende mostrar que la media poblacional está dentro del *EI*. Esto puede hacerse con las dos pruebas de hipótesis de la siguiente forma.

Prueba uno

$$H_0 : \mu = 83.4$$
$$H_A : \mu < 83.4$$

Prueba dos

$$H_0 : \mu = 77.4$$
$$H_A : \mu > 77.4$$

Observe que si la primera prueba es significativa la interpretación es que la media poblacional es menor que 83.4. El rechazo de la segunda prueba conduce a la conclusión de que la media es mayor que 77.4. En consecuencia, si *ambas* pruebas son significativas puede concluirse que la media poblacional está en el *EI* y que, desde un punto de vista práctico, los dos fármacos son equivalentes para el tratamiento de la ansiedad.

Prueba de dos colas. Si dejamos que EI_U represente el extremo superior (*upper*) del *EI* e EI_L el extremo inferior (*lower*), la hipótesis nula y la alternativa para la prueba de equivalencia (de una media) de dos colas son las siguientes:

$$H_{0E} : \mu \leq EI_L \quad \text{o} \quad \mu \geq EI_U$$
$$H_{AE} : EI_L < \mu < EI_U$$

Observe que la hipótesis nula afirma que la media poblacional no está en el *EI*, mientras que la hipótesis alternativa afirma que la media poblacional está en el *EI*. La hipótesis nula, entonces, es una aseveración de no equivalencia, mientras que la alternativa asevera equivalencia. Para rechazar H_{0E} es necesario mostrar que $\mu > EI_L$ *y* que $\mu < EI_U$. Esto se hace mediante las pruebas de una cola con las siguientes hipótesis.

Prueba uno

$$H_{01} : \mu = EI_U$$
$$H_{A1} : \mu < EI_U$$

Prueba dos

$$H_{02} : \mu = EI_L$$
$$H_{A2} : \mu > EI_L$$

Nosotros utilizamos la notación H_{0E} y H_{AE} para indicar las hipótesis nula y alternativa para la prueba de equivalencia, y H_{01}, H_{A1}, H_{02} y H_{A2} para las hipótesis nula y alternativa para los componentes de las dos pruebas. Utilizando esta notación, H_{0E} se rechaza en favor de H_{AE} única-

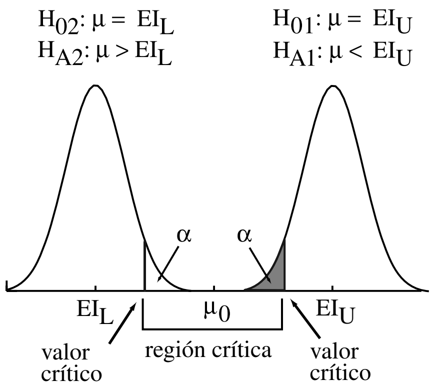


FIGURA 4.15: Prueba de equivalencia de dos colas para una media poblacional.

mente si *tanto* H_{01} *como* H_{02} se rechazan en favor de sus respectivas alternativas. La figura 4.15 representa el procedimiento de la prueba de equivalencia de dos colas. Observe que la región crítica para la prueba de equivalencia es la que abarca las regiones críticas de ambos componentes de las pruebas.

Se deberán reconocer muchas diferencias entre las pruebas de equivalencia de dos colas y de la hipótesis estándar de dos colas. Primero, la prueba de equivalencia se hace conduciendo dos pruebas de *una cola*. De esta manera, una prueba de equivalencia de dos colas en $\alpha = .05$ se realizará mediante dos pruebas de una cola, cada una conducida con $\alpha = .05$ y *no* con $\alpha/2$, como se haría con la prueba estándar. Esto surge del hecho de que cuando H_{0E} es verdadera, la media poblacional deberá ser *ya sea* EI_U *o* EI_L. Si la media poblacional es EI_U, la probabilidad de obtener un resultado significativo es α, que se muestra como el área sombreada de la parte derecha de la curva en la figura 4.15. Si la media es EI_L la probabilidad de un hallazgo significativo es la región crítica del lado izquierdo de la curva. Esto significa que la probabilidad de concluir una equivalencia falsa es *ya sea* el área en la región crítica en el lado derecho de la curva *o* el área de la región crítica en el lado izquierdo de la curva. En cualquier caso, la probabilidad de un encuentro significativo frente a la no equivalencia será α.[14]

Una segunda diferencia, derivada de la primera, es la forma en la cual se calcula el valor-*p*. El valor-*p* para la prueba de equivalencia de dos colas se obtiene calculando el valor-*p* para cada componente de las pruebas de una cola y eligiendo el valor más grande. Por ejemplo, si los valores-*p* de las dos pruebas de dos colas fueran .07 y .001, el valor-*p* para la prueba de equivalencia de dos colas será .07. Este método para obtener el valor-*p* surge del hecho de que ambos componentes de las pruebas deben ser significativos para rechazar la hipótesis nula de no equivalencia a favor del hallazgo de equivalencia. Si el valor-*p* más grande es menor que α, el valor más pequeño deberá ser también menor que α, lo cual significa que ambos componentes de las pruebas son significativos. Por otro lado, si el valor-*p* más grande es mayor que α, entonces al menos uno de los dos componentes de las pruebas no será significativo, lo cual representa que la hipótesis de no equivalencia no se rechaza.

[14] En realidad, quizá sea menor que α, pero nosotros no introduciremos esta complicación aquí.

Finalmente, cuando se lleva a cabo una prueba por el método del valor obtenido versus el valor crítico, usted debe calcular dos valores obtenidos para compararlos con los dos valores críticos. H_{0E} se rechaza cuando ambos valores obtenidos caen en sus respectivas regiones críticas.

EJEMPLO 4.24

Utilice las pruebas Z de una media y la siguiente información para realizar una prueba de equivalencia de dos colas de la hipótesis nula que establece que μ no está en el *EI* de 77.4 a 83.4. Reporte los resultados para ambos métodos, el valor-p versus α y Z obtenida versus Z crítica.

$$\bar{x} = 82.2 \qquad n = 30$$
$$\sigma = 8 \qquad \alpha = .05$$

Solución La prueba de equivalencia se hace conduciendo las dos pruebas siguientes.

Prueba uno

$$H_{01} : \mu = 83.4$$
$$H_{A1} : \mu < 83.4$$

Prueba dos

$$H_{02} : \mu = 77.4$$
$$H_{A2} : \mu > 77.4$$

Para la primera prueba

$$Z = \frac{82.2 - 83.4}{\frac{8}{\sqrt{30}}} = -.82$$

y para la segunda

$$Z = \frac{82.2 - 77.4}{\frac{8}{\sqrt{30}}} = 3.29.$$

La columna 3 del Apéndice A nos muestra los respectivos valores-p, que son .2061 y .0005. Puesto que el más grande de éstos es mayor que $\alpha = .05$, la hipótesis nula no se rechaza.

A pesar de que Z_2 es mayor que su Z crítica asociada de 1.65 y, por lo tanto, es significativa, Z_1 es más grande que su valor crítico asociado de –1.65, por lo que no es significativo. Como las pruebas deben ser significativas para rechazar H_{0E}, la hipótesis nula no se rechaza. La interpretación de este resultado es que la equivalencia no puede ser demostrada. ■

EJEMPLO 4.25

Utilice las pruebas t de una media y la siguiente información para llevar a cabo la prueba de equivalencia de dos colas de la hipótesis nula que establece que μ no está en el *EI* de 2 a –2. Utilice $\alpha = .05$.

Muestra: 3 –1 0 –4 –2 2 1 –3 –1 0

Solución La prueba de equivalencia se hace conduciendo las dos pruebas siguientes.

Prueba uno
$H_{01} : \mu = 2$
$H_{A1} : \mu < 2$

Prueba dos
$H_{02} : \mu = -2$
$H_{A2} : \mu > -2$

Los valores de $\bar{x}$ y s pueden obtenerse de las siguientes sumas

X	X^2
3	9
–1	1
0	0
–4	16
–2	4
2	4
1	1
–3	9
–1	1
0	0
–5	45

Por las ecuaciones 2.1 y 2.16, $\bar{x}$ y s son

$$\bar{x} = \frac{\sum x}{n} = \frac{-5}{10} = -.5$$

y

$$s = \sqrt{\frac{\sum x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{45 - \frac{-5^2}{10}}{9}} = 2.173.$$

Por la ecuación 4.8, los dos estadísticos de prueba se calculan como

$$t_1 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{-.5 - 2}{\frac{2.173}{\sqrt{10}}} = -3.638$$

y

$$t_2 = \frac{-.5 - (-2)}{\frac{2.173}{\sqrt{10}}} = 2.183.$$

El Apéndice B indica que un valor de 1.833 corta .05 en la cola de una distribución t con 9 grados de libertad. Los valores críticos para las dos pruebas de una cola conducidas en $\alpha = .05$ son, por consiguiente, –1.833 y 1.833. Como $t_1 = -3.638 < -1.833$, y $t_2 = 2.183 > 1.833$, ambos componentes de las hipótesis se rechazan. Dado que ambos componentes de las pruebas son significativos, la hipótesis nula de no equivalencia se rechaza en favor de la equivalencia. ■

TABLA 4.7: Distribuciones muestrales de $\hat{p}$ para $n = 8$ y $\pi = 0.3$, y 0.7.

Proporción $\hat{p}$	Número de éxitos	$\pi = .30$ $P(y)$	$\pi = .70$ $P(y)$
.000	0	.05765	.00007
.125	1	.19765	.00122
.250	2	.29648	.01000
.375	3	.25412	.04668
.500	4	.13614	.13614
.625	5	.04668	.25412
.750	6	.01000	.29648
.875	7	.00122	.19765
1.000	8	.00007	.05765

EJEMPLO 4.26

Utilice las pruebas exactas de una muestra para una proporción para realizar una prueba de equivalencia de dos colas de la hipótesis nula que establece que π no está en el *EI* de .3 a .7. Utilice $n = 8$, $\hat{p}$ obtenida de .5 y $\alpha = .20$. Reporte los resultados para ambos métodos, el valor-*p* versus α y $\hat{p}$ obtenida versus la $\hat{p}$ crítica.

Solución La prueba de equivalencia se hace conduciendo las dos pruebas siguientes.

Prueba uno
$H_{01} : \pi = .7$
$H_{A1} : \pi < .7$

Prueba dos
$H_{02} : \pi = .3$
$H_{A2} : \pi > .3$

La tabla 4.7 muestra las probabilidades binomiales para $n = 8$ y $\pi = .3$ y .7. Utilizando las probabilidades para $\pi = .70$, el valor-*p* para la prueba uno es $P(0) + P(1) + P(2) + P(3) + P(4) = .00007 + .00122 + .01000 + .04668 + .13614 = .19411$. Asimismo, utilizando las probabilidades para $\pi = .30$, el valor-*p* para la prueba dos es $P(4) + P(5) + P(6) + P(7) + P(8) = .13614 + .04668 + .01000 + .00122 + .00007 = .19411$. Puesto que ambos valores son menores que $\alpha = .20$, ambas pruebas son significativas, lo cual conduce al rechazo de la hipótesis nula de no equivalencia.

Debido a que $P(0) + P(1) + P(2) + P(3) + P(4) = .19411$, que es menor que $\alpha = .20$, y $P(0) + P(1) + P(2) + P(3) + P(4) + P(5) = .19411 + .25412 = .44823$, que es mayor que $\alpha = .20$, el valor crítico para la prueba uno es $\hat{p} = .50$. Asimismo, debido a que $P(8) + P(7) + P(6) + P(5) + P(4) = .19411$, que es menor que $\alpha = .20$, y $P(8) + P(7) + P(6) + P(5) + P(4) + P(3) = .19411 + .25412 = .44823$, que es mayor que $\alpha = .20$, el valor crítico para la prueba dos también es $\hat{p} = .50$. Dado que $\hat{p} = .50$ obtenida es menor que o igual a $\hat{p}$ crítica de .50, la hipótesis nula para la prueba uno se rechaza. La prueba dos también es significativa porque la $\hat{p}$ obtenida es mayor que

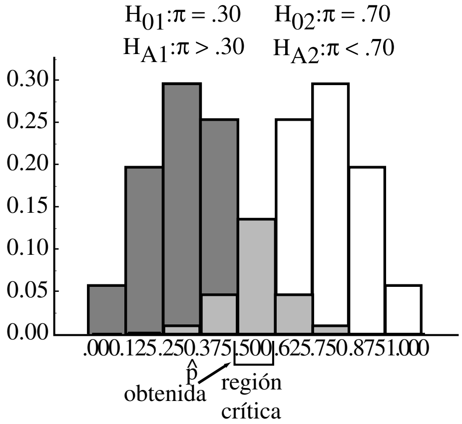


FIGURA 4.16: Prueba de equivalencia de dos colas para una proporción de la población.

o igual a la $\hat{p}$ crítica. Nuevamente, la hipótesis nula de no equivalencia se rechaza. Las dos distribuciones muestrales binomiales y la coyuntura de la región crítica se ilustran en la figura 4.16.

Prueba de una cola. Como veremos en los capítulos 5 y 6, las pruebas de equivalencia de una cola se emplean con mayor frecuencia en la investigación que las pruebas de dos colas. La prueba de una cola se hace conduciendo *cualquiera* de las dos pruebas, la uno *o* la dos, pero no ambas. La elección de la prueba depende de la hipótesis nula de interés. Por ejemplo, suponga que se conoce un procedimiento quirúrgico para restaurar la agudeza visual cerca del nivel normal de 20/20, pero este procedimiento también es conocido por tener un alto riesgo de infección, lo cual puede conducir a una pérdida total de la visión. Hay un nuevo método que tiene una tasa mucho menor de infección, pero la pregunta es si es tan efectivo como el método anterior en términos de restauración visual. Un grupo de expertos en la vista decide que si el nuevo procedimiento produce agudezas promedio de 20/30 o menos, el nuevo método será considerado, funcionalmente, equivalente al tratamiento anterior. Observe que se puede esperar cualquier método para incrementar la visión a un nivel mejor que el normal (esto es, 20/20).

En esta situación, la prueba uno será llevada con el EI_U establecido en el equivalente numérico de 20/30. (El equivalente numérico se llama puntuación LogMar y es .20 para una agudeza de 20/30. El equivalente LogMar de 20/20 es .0.) La hipótesis nula equivalente sostendrá que el promedio de agudeza producido por el nuevo método quirúrgico es mayor que o igual a 20/30 (.20 LogMar), mientras la hipótesis alternativa sostendrá que el promedio es menor que 20/30. La prueba de equivalencia de una cola se hace conduciendo la siguiente prueba en el nivel apropiado de α.[15]

Prueba uno

$$H_{01} : \mu = .20$$
$$H_{A1} : \mu < .20$$

[15] Los puntajes de agudeza visual son típicamente asimétricos de manera que la validez de un resultado depende de la robustez de una media Z o de otra prueba que suponga normalidad.

Si la hipótesis nula se rechaza, los métodos se considerarán equivalentes. En general, las hipótesis nula y alternativa para la prueba de equivalencia de una cola son

$$H_{0E} : \mu \geq EI_U$$
$$H_{AE} : \mu < EI_U$$

o

$$H_{0E} : \mu \leq EI_L$$
$$H_{AE} : \mu > EI_L$$

■

EJEMPLO 4.27

Utilice una prueba Z de una media con la siguiente información para llevar a cabo una prueba de equivalencia de una cola de la hipótesis nula que establece que μ es mayor que o igual al $EI_U =$ .20. Reporte los resultados para ambos métodos, el valor-p versus α y Z obtenida versus Z crítica.

$$\bar{x} = .11 \qquad n = 30$$
$$\sigma = .16 \qquad \alpha = .05$$

Solución La prueba de equivalencia se hace conduciendo la prueba uno como sigue.

Prueba uno

$$H_{01} : \mu = .20$$
$$H_{A1} : \mu < .20$$

Para esta prueba

$$Z_1 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{.11 - .20}{\frac{.16}{\sqrt{30}}} = -3.08.$$

La columna 3 del Apéndice A muestra que el valor-p asociado será .0010. Como este valor es menor que $\alpha = .05$, la hipótesis nula de no equivalencia se rechaza en favor de la alternativa de equivalencia. Puesto que la Z obtenida de -3.08 es menor que la Z crítica de -1.65, la hipótesis nula de no equivalencia se rechaza. ■

EJEMPLO 4.28

Algunas personas con anemia por deficiencia de hierro tienen niveles de hemoglobina en el rango de 8-10 g/dl, mientras que los niveles normales se encuentran alrededor de 15 g/dl. El tratamiento con suplementos de hierro generalmente incrementa los niveles a los valores normales, pero no provoca que esos niveles aumenten por arriba de los normales. Suponga que se ha descubierto un suplemento dietético que puede producirse de forma fácil y económica en países desarrollados. El suplemento será declarado efectivo si puede demostrarse que es equivalente a tratamientos más costosos. En definitiva, si en los países desarrollados el nuevo suplemento produce un promedio de valores de hemoglobina mayor que 13 g/dl en las personas con deficiencia de hierro, el nuevo suplemento será declarado equivalente a los anteriores.

Utilice una prueba Z de una media con la siguiente información para probar la hipótesis nula de equivalencia de que las personas que consumen el nuevo suplemento alcanzan niveles promedio de hemoglobina menores que o iguales a 13 g/dl.[16] Reporte los resultados para ambos métodos, el valor-*p* versus α y Z obtenida versus Z crítica.

$$\bar{x} = 13.8 \qquad n = 150$$
$$\sigma = 1.16 \qquad \alpha = .05$$

Solución La prueba de equivalencia se hace conduciendo la prueba dos como sigue

Prueba dos
$H_0 : \mu = 13$
$H_A : \mu > 13$

Para esta prueba

$$Z_2 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{13.8 - 13.0}{\frac{1.16}{\sqrt{150}}} = 8.45.$$

En la columna 3 del Apéndice A se muestra que el valor-*p* asociado con $Z = 3.5$ es .0002. De esto se infiere que *p* para $Z = 8.45$ será menor que este valor. Debido a que *p* es menor que $\alpha = .05$, la hipótesis nula de no equivalencia se rechaza. Debido que la Z obtenida de 8.45 es mayor que la Z crítica de 1.65, la hipótesis nula de no equivalencia se rechaza. ■

4.3.7 Errores y decisiones correctas en una prueba de hipótesis

Como se habrá dado cuenta, es posible rechazar una hipótesis nula aun cuando ésta sea verdadera. Usted también ha aprendido que la probabilidad de tal ocurrencia se designa como α. También es posible errar al no rechazar una hipótesis nula falsa. En contraste con estos dos tipos de errores, una decisión correcta tendrá lugar cuando la hipótesis nula verdadera no sea rechazada o la hipótesis nula falsa sea rechazada. En este apartado usted aprenderá acerca de cada una de estas eventualidades, así como la probabilidad de su realización. La siguiente explicación se dividirá en dos componentes distintos: **1.** sucesos que ocurren cuando la hipótesis nula es verdadera, y **2.** sucesos que ocurren cuando la hipótesis nula es falsa. Es importante que usted tenga en mente a cuál de estas eventualidades se hace referencia en la discusión siguiente. Los comentarios posteriores se aplican a la prueba de hipótesis en general, pero por simplicidad serán presentados en relación con la prueba Z de una media. La tabla 4.8 presenta un resumen del siguiente discurso y deberá consultarse conforme usted lea.

Sucesos que ocurren cuando la hipótesis nula es verdadera. Un **error tipo I** ocurre cuando se rechaza una hipótesis nula *verdadera*. La *probabilidad* de cometer un error tipo I se llama **nivel de significancia** de la prueba o α. Con el fin de evitar parcialidades, deberá establecerse α antes de efectuar el análisis. Observe también que la probabilidad de un error tipo I está bajo el control del investigador, quien establece el nivel de significancia para la prueba.

El hecho de no rechazar una hipótesis nula *verdadera* da por resultado una **decisión correcta**. La probabilidad de una decisión correcta cuando la hipótesis nula es verdadera es $1 - \alpha$. Las probabilidades de un error tipo I y una decisión correcta para una prueba Z de una cola para una media se muestran en la figura 4.17.

[16] Una vez más, debemos confiar en la robustez de una media Z para las desviaciones de la normalidad de la población.

TABLA 4.8: Resultados asociados con pruebas de hipótesis.

	Hipótesis nula: Verdadera	Hipótesis nula: Falsa
Rechazada	Error tipo I (α)	Decisión correcta (potencia)
Fallo de rechazo	Decisión correcta ($1-\alpha$)	Error tipo II (β)

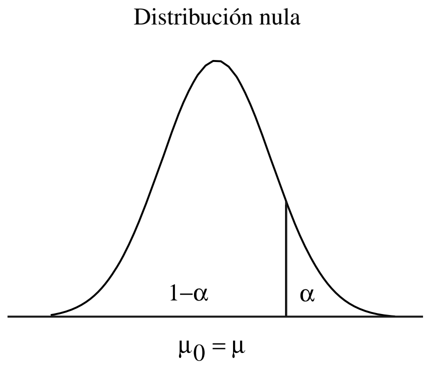


FIGURA 4.17: Probabilidad de un error tipo I y una correcta decisión para una prueba *Z* de una cola para una media.

Sucesos que ocurren cuando la hipótesis nula es falsa. Un **error tipo II** ocurre cuando una hipótesis nula *falsa* no es rechazada. La *probabilidad* de cometer un error tipo II se conoce como **beta** (β).

El rechazo de una hipótesis nula *falsa* da por resultado una **decisión correcta**. La probabilidad de una decisión correcta cuando la hipótesis nula es falsa se llama **potencia.**

Note que cuando la hipótesis nula es verdadera, el valor hipotético de la media poblacional (μ_0) es la media poblacional (μ) (véase la figura 4.17). En esta circunstancia la distribución muestral utilizada para ejecutar la prueba de hipótesis es, de hecho, la distribución muestral del estadístico de prueba. En contraste, cuando la hipótesis nula es falsa la media de la población (μ) es diferente de la media hipotética (μ_0). Esto implica que la distribución muestral del estadístico de prueba se centra alrededor de μ, más que de μ_0, así que el modelo usado para ejecutar la prueba de hipótesis no es una descripción exacta de la distribución muestral del estadístico. Esta situación se ilustra en la figura 4.18.

Es importante comprender que la distribución nula se utiliza para ejecutar la prueba de hipótesis, pero es la distribución **alternativa** la que representa la distribución muestral del estadístico. Esto surge porque la media de la población es μ y no μ_0. Suponga, por ejemplo, que un investigador desea probar la hipótesis nula $\mu = \mu_0 = 100$ y utiliza la curva nula para este propósito. Para el investigador es *desconocido* el hecho de que $\mu = 120$, así que la distribución

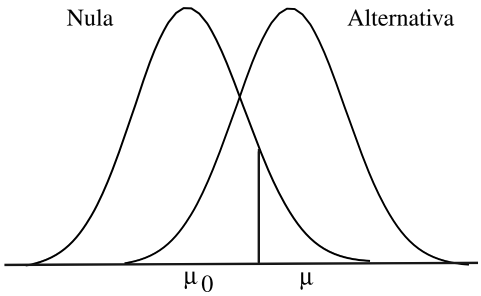


FIGURA 4.18: Distribuciones nula y alternativa.

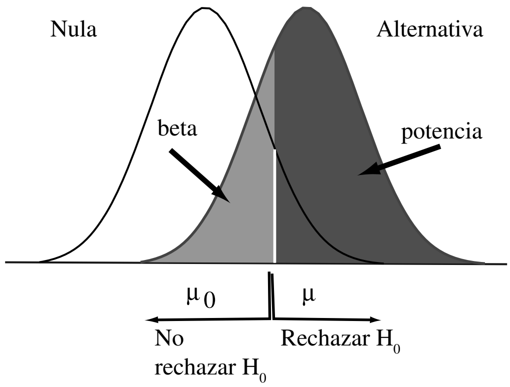


FIGURA 4.19: Potencia y beta para la prueba *Z* de una media con una cola.

muestral se centra alrededor de este valor y no en 100. De esto se infiere que la probabilidad de obtener un estadístico de prueba en cualquier rango especificado se representa por una área bajo la alternativa y no en la curva nula.

Como la potencia es la probabilidad de rechazar una hipótesis nula falsa, puede ilustrarse como la porción de la curva alternativa que representa valores del estadístico de prueba que son lo suficientemente grandes (o pequeños) para causar el rechazo de la hipótesis nula. Considere ahora la figura 4.19. En esta figura la curva alternativa está dividida en dos porciones: la porción que representa la potencia (sombreado oscuro) y la porción que representa beta (sombreado claro). (El límite delantero de la región crítica aparece marcado con una línea blanca). Observe que la hipótesis nula se rechaza cuando el valor del estadístico de prueba es mayor que o igual al valor en el límite delantero de la región crítica. La probabilidad de rechazar la hipótesis nula es, entonces, la porción de la curva alternativa que representa valores del estadístico de prueba iguales a o que exceden el límite delantero de la región crítica. La hipótesis nula no se rechazará cuando el estadístico de prueba está por debajo del límite delantero de la región crítica. Esta probabilidad (β) es representada por la porción de la curva alternativa que queda debajo de la región crítica. Como se observa en la figura, la potencia = $1 - \beta$ y a menudo se representa como tal en estadística y la literatura de investigación.

EJEMPLO 4.29

Conteste las siguientes preguntas de acuerdo con el panel A de la figura 4.20. *a*) ¿Cuál es la hipótesis alternativa para la prueba de significancia? *b*) ¿Qué área(s) representa(n) a la potencia? *c*) ¿Qué área(s) representa(n) a beta? *d*) ¿Qué representa el área "e"?

Solución *a*) La alternativa enuncia $H_A : \mu < \mu_0$. Esto proviene del hecho de que la región crítica está en la cola del lado izquierdo de la distribución nula. *b*) La potencia está representada por las áreas a y c. Esto proviene del hecho de que las áreas a y c representan la probabilidad de que un estadístico de prueba será lo suficientemente pequeño como para cumplir con el criterio de rechazo de H_0, como se establece mediante la región crítica de la curva nula. *c*) Beta es mostrada por las áreas b y d. Dichas áreas constituyen la porción de la curva alternativa que cae fuera de la región crítica y, por lo tanto, representan la probabilidad de que la prueba dé por resultado el no rechazo de la hipótesis nula. *d*) Puesto que la hipótesis nula es falsa, el área e no representa la probabilidad de ningún resultado. Todas las probabilidades se muestran como áreas de la curva alternativa. ■

EJEMPLO 4.30

Conteste las siguientes preguntas según su relación con el panel B de la figura 4.20. *a*) ¿Cuál es la hipótesis alternativa para la prueba de significancia? *b*) ¿Qué área(s) representa(n) a la potencia? *c*)¿Qué área(s) representa(n) a beta? *d*) ¿Qué representa el área "a"? *e*) ¿Qué representa el área d?

Solución *a*) La alternativa enuncia $H_A : \mu > \mu_0$. Esto proviene del hecho de que la región crítica está en la cola del lado derecho de la distribución nula. *b*) La potencia está representada por las áreas b y c. Esto se deriva del hecho de que las áreas b y c representan la probabilidad de que un estadístico de prueba será lo suficientemente grande para cumplir con el criterio de rechazo de H_0, como establece la región crítica de la curva nula. *c*) Beta es mostrada por el área e. Esta área se encuentra en la porción de la curva alternativa que cae fuera de la región crítica y, por lo tanto, representa la probabilidad de que la hipótesis nula no sea rechazada. *d*) Puesto que la hipótesis nula es falsa, el área a no representa la probabilidad de ningún resultado. Todas las probabilidades son mostradas como áreas de la curva alternativa. *e*) Al igual que el área a, el área d no representa ningún resultado. ■

Factores que determinan la potencia y beta. La potencia y beta están determinadas por *a*) el nivel de significancia de la prueba (α), *b*) el tamaño de la muestra (*n*), y *c*) la forma de la distribución alternativa. Ahora profundizaremos en ello.

Nivel de significancia

La potencia se incrementa conforme aumenta el nivel de significancia (por ejemplo, de .01 a .05). Debido a que beta = 1 − potencia, beta disminuye con incrementos en el nivel de significancia. Esta relación se ilustra en la figura 4.21. Como se aprecia, cuando $\alpha = .01$, la potencia está representada por las áreas d y a, mientras que beta está representada por las áreas b, c, f y e. Cuando α se incrementa por .04 a .05, la potencia se incrementa por las áreas b y c con beta disminuida por esta misma cantidad. Naturalmente, lo inverso también es cierto. Decrementos en el nivel de significancia se asocian con decrementos en la potencia e incrementos en beta. De esta manera, el precio que se paga por el decremento de la probabilidad de un error tipo I es un incremento en la probabilidad de un error tipo II.

Tamaño de la muestra (*n*)

Desde un punto de vista práctico, tal vez el determinante más importante de la potencia y beta es el tamaño de la muestra. Mientras el tamaño de la muestra incrementa, la potencia se incrementa.

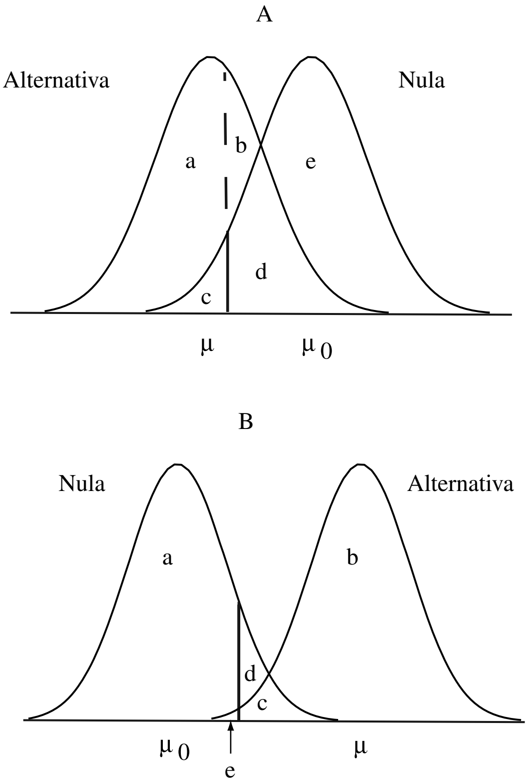


FIGURA 4.20: Dos ejemplos que muestran la potencia y beta para pruebas *Z* de dos colas para una media.

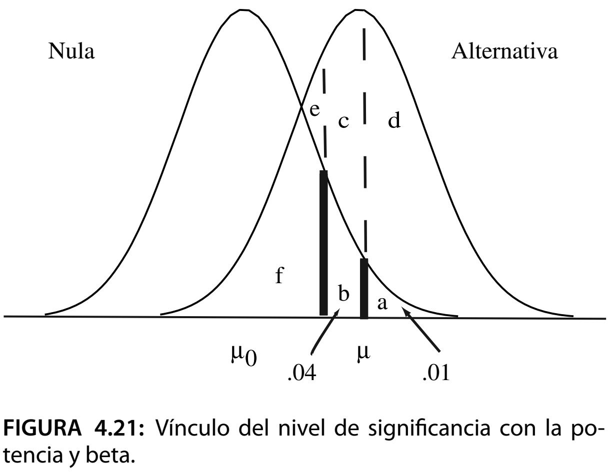


FIGURA 4.21: Vínculo del nivel de significancia con la potencia y beta.

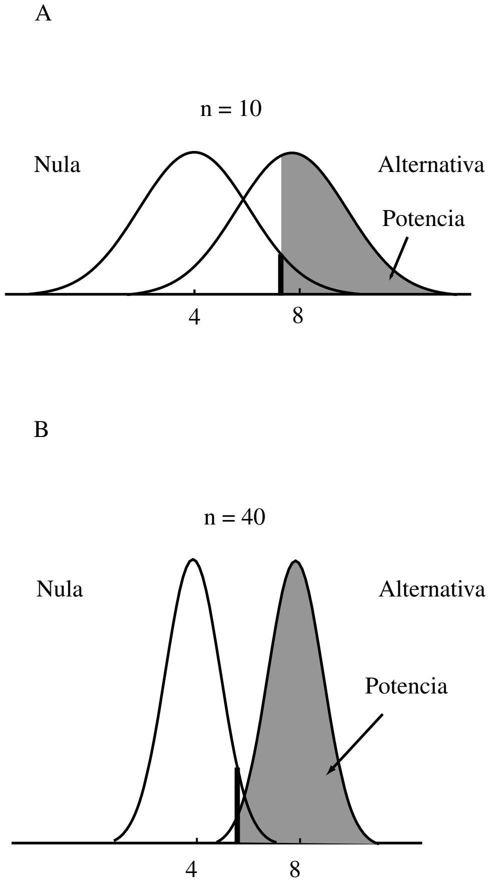


FIGURA 4.22: Vínculo del tamaño de la muestra con la potencia.

De esta forma, un investigador que desea establecer α en algún nivel tradicional (por ejemplo, .05) puede aumentar la potencia incrementando el tamaño de la muestra. Esto se demuestra en la figura 4.22 para la prueba Z de una media.

El panel A muestra las distribuciones nula y alternativa cuando $n = 10$. La porción sombreada de la distribución alternativa muestra la potencia y parece representar ligeramente más que .5 de la curva. El panel B muestra distribuciones con las mismas medias y nivel de significancia, pero con $n = 40$. Como se puede ver, las curvas en el panel B tienen errores estándar más pequeños que la media, lo cual en efecto causa que las dos distribuciones se alejen una de la otra. Observe, sin embargo, que las medias de distribución aún son 4 y 8 como es verdad en las curvas del panel A. El resultado es que una porción mucho más grande de la distribución de la alternativa encuentra el criterio de rechazo especificado por la región crítica de la curva nula.

Puesto que el nivel de significancia a menudo se establece en algún nivel tradicional y la forma de la distribución alternativa está más allá del control del investigador, generalmente el tamaño de la muestra es manipulado para dirigir la potencia en la investigación. Pronto hablaremos más acerca de este tema.

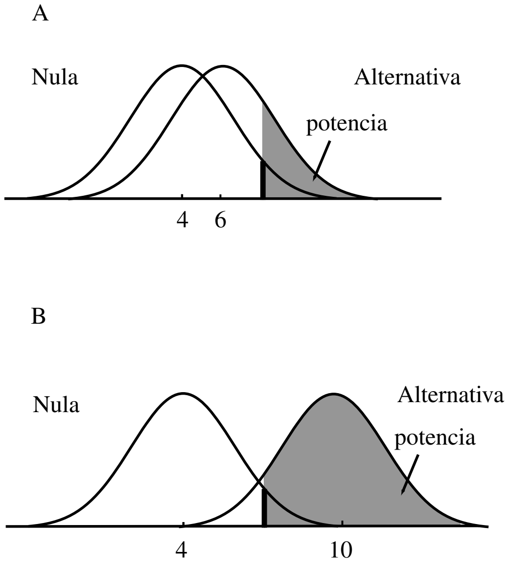


FIGURA 4.23: Potencia de la prueba *Z* de una media como una función de la diferencia entre las medias hipotética y la poblacional actual.

Forma de la distribución alternativa

La potencia de la prueba Z de una media es fuertemente dependiente de la magnitud de la diferencia entre μ_0 y μ. Conforme esta diferencia se incrementa, la potencia aumenta. Conforme la diferencia disminuye, la potencia disminuye hasta que $\mu_0 = \mu$, en cuyo punto la hipótesis nula es verdadera y la probabilidad de rechazo es α. Esta característica de la prueba tiene un fundamento intuitivo. Parece razonable que pequeñas diferencias del valor nulo sean más difíciles de detectar que las grandes. El panel A de la figura 4.23 muestra una situación donde hay una diferencia de dos unidades entre la media poblacional actual ($\mu = 6$) y el valor hipotético ($\mu_0 = 4$). El panel B muestra un incremento en la potencia que corresponde a la media poblacional de 10, que está a seis unidades del valor hipotético.

En general, conforme se incremente el grado en el cual la hipótesis nula es falsa, también se incrementará la potencia de la prueba. Para la prueba binomial esto se expresa como la diferencia entre π_0 y π, pero puede expresarse de muchas formas diferentes para otros tipos de pruebas, de las cuales se hablará en los siguientes capítulos.

Cálculo de la potencia y beta. Por lo general, no es posible calcular la potencia y beta en situaciones de investigación aplicada. La razón es que para calcular estas cantidades, usted debe poseer cierta información acerca de la distribución alternativa que comúnmente no está disponible. Por ejemplo, para la prueba *Z* de una media debe conocerse μ para poder hacer estos cálculos. Pero si se conociera μ, no habría necesidad de una prueba de hipótesis. No obstante es útil, según el grado de interés que se tenga sobre el entendimiento de la potencia y beta, para ejecutar unos cuantos cálculos hipotéticos.

Por ejemplo, ¿cuál será la potencia de una prueba Z de una media en las siguientes condiciones?

$$H_0 : \mu = 4 \qquad \sigma_{\bar{x}} = 2 \qquad \mu = 10$$
$$H_A : \mu > 4 \qquad \alpha = .05$$

Referirnos a los paneles A y B de la figura 4.24 ayudará a clarificar la siguiente explicación. Con el fin de hallar beta, es necesario encontrar la porción de la curva alternativa que cae por debajo del límite delantero de la región crítica en la curva nula. Con el fin de localizar esta área primero calculamos

$$\frac{\mu_0 - \mu}{\sigma_{\bar{x}}} = \frac{4 - 10}{2} = -3.00.$$

Esto indica que $\mu_0 = 4$ está tres errores estándar por debajo de $\mu = 10$. Debido a que el valor de la Z crítica para una prueba Z de una cola en $\alpha = .05$ es 1.65, se infiere que si el valor es añadido a −3.00 obtendremos el número de errores estándar que quedan entre el límite delantero de la región crítica y μ. Para propósitos de calcular la potencia designaremos la Z crítica como Z_α y la puntuación Z indicando la distancia entre μ y el límite delantero de la región crítica como Z_β. (Véase el panel B de la figura 4.24.) En este ejemplo, $Z_\beta = -3.00 + 1.65 = -1.35$. La porción de la curva alternativa que queda por debajo de este valor representa a beta, mientras que la porción por arriba de éste es la potencia. Como se puede ver en la columna 3 de la tabla de la curva normal, la puntuación Z de −1.35 corta .0885 en la cola inferior de la curva alternativa. La potencia es 1 − .0885 = .9115. La probabilidad de que la prueba dará por resultado un error tipo II es entonces .0885, mientras que la probabilidad de una decisión correcta es .9115.

La lógica que fundamenta esta solución fue encontrar el número de errores estándar entre el límite delantero de la región crítica y la media de la curva alternativa (esto es, Z_β). Esto se hizo encontrando el número de errores estándar entre μ_0 y μ, y sumando el número de errores estándar entre μ y el límite delantero de la región crítica (esto es, Z_α). Una vez hecho esto, fue sencillo encontrar el área apropiada en la tabla de la curva normal. Una expresión para Z_β está dada por

$$\boxed{Z_\beta = \frac{\mu_0 - \mu}{\sigma_{\bar{x}}} + Z_\alpha} \tag{4.10}$$

EJEMPLO 4.31

Encuentre la potencia y beta para una media Z en las siguientes condiciones.

$$H_0 : \mu = 4 \qquad \sigma_{\bar{x}} = 2 \qquad \mu = -2$$
$$H_A : \mu < 4 \qquad \alpha = .05$$

Solución Sustituyendo en la ecuación 4.10 se obtiene

$$Z_\beta = \frac{4 - (-2)}{2} + (-1.65) = 1.35.$$

La porción de la curva alternativa que cae por arriba de este valor representa beta, mientras que la porción por debajo de este valor es la potencia. La columna 3 de la tabla de la curva normal muestra que $Z = 1.35$ corta .0885 en la cola superior de la curva normal. La potencia es entonces 1 − .0885 = .9115. ¿Por qué se obtuvieron las mismas respuestas aquí y en el ejemplo de arriba? ■

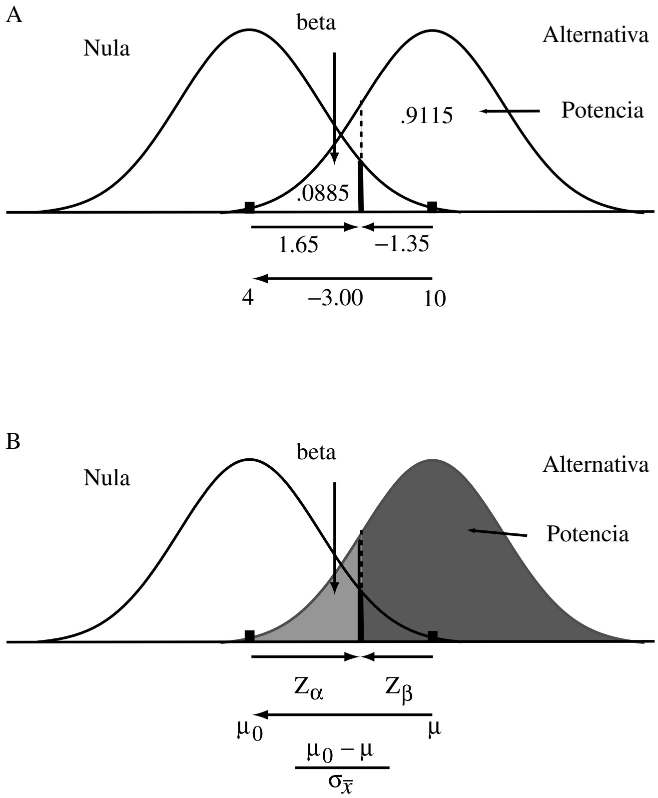


FIGURA 4.24: Cálculo de la potencia y beta para una prueba *Z* de una cola para una media. Las distancias están expresadas en términos de los errores estándar.

EJEMPLO 4.32

Encuentre la potencia y beta para una prueba *Z* de una media en las siguientes condiciones.

$$
\begin{array}{lll}
H_0: \mu = 90 & \sigma = 20 & \alpha = .05 \\
H_A: \mu \neq 90 & n = 25 & \mu = 88
\end{array}
$$

Solución Usando

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = 4$$

en la ecuación 4.10 se obtiene

$$Z_\beta = \frac{90-88}{4} + (-1.96) = -1.46.$$

Se utilizó el valor para $Z_\alpha = -1.96$ porque la prueba bajo consideración es de dos colas con $\alpha =$ .05 y $\mu < \mu_0$. Esto se muestra en la figura 4.25. Observe que la distancia de μ_0 a μ es .5 errores

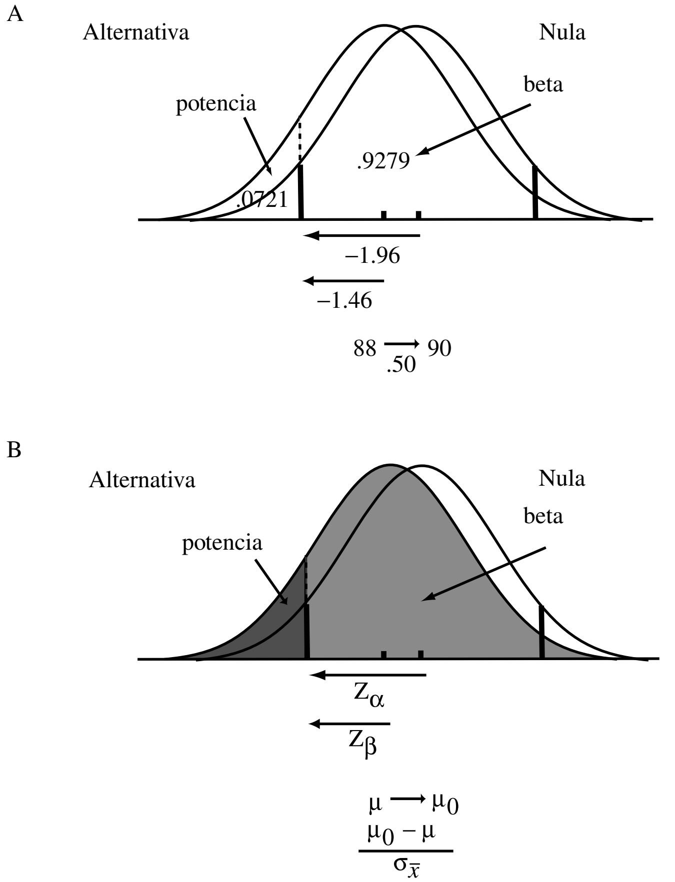


FIGURA 4.25: Cálculo de la potencia y beta para una prueba *Z* de una cola para una media. Las distancias están expresadas en términos de los errores estándar.

estándar. Como el límite delantero de la región crítica está 1.96 errores estándar por debajo de μ_0, la distancia desde μ a la región crítica es −1.96 + .50 = −1.46 errores estándar. La columna 3 de la tabla de la curva normal indica que la potencia es .0721. Beta es 1 – .0721 = .9279. ■

EJEMPLO 4.33

Encuentre la potencia y beta para una prueba *Z* de una media en las siguientes condiciones.

$$H_0: \mu = 90 \qquad \sigma = 20 \qquad \alpha = .05$$
$$H_A: \mu \neq 90 \qquad n = 25 \quad \mu = 92$$

Solución Al usar

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{20}{\sqrt{25}} = 4$$

en la ecuación 4.10 se obtiene

$$Z_\beta = \frac{90-92}{4} + (1.96) = 1.46$$

Se utilizó el valor para $Z_\alpha = 1.96$ porque la prueba bajo consideración es dos colas con $\alpha = .05$ y $\mu > \mu_0$. La potencia es la porción de la curva normal que cae por arriba de $Z = 1.46$. La columna 3 de la tabla de la curva normal muestra que la potencia es .0721. Beta es 1 – .0721 = .9279. ¿Por qué se obtuvieron los mismos valores que en el ejemplo de arriba?

EJEMPLO 4.34

En otro ejemplo encontraremos la potencia y beta para la siguiente prueba exacta para una proporción.

$$\begin{array}{lll} H_0 : \pi = .35 & \alpha = .05 & \pi = .50 \\ H_A : \pi > .35 & n = 8 & \end{array}$$

Solución Las probabilidades binomiales para las distribuciones nula ($\pi = 0.35$) y alternativa ($\pi = .50$) se presentan en la tabla 4.6 de la página 113. $\hat{p}$ crítica puede determinarse por P (6) + P (7) + P (8) = .02175 + .00335 + .00023 = .02533 < $\alpha = .05$, mientras P (5) + P (6) + P (7) + P (8) = .08077 + .02533 = .10610 > $\alpha = .05$. $\hat{p}$ crítica es entonces .750. Por lo tanto, la hipótesis nula se rechaza para cualquier valor de $\hat{p}$, que es más grande o igual que .750. Pero ¿cuál es la probabilidad de que $\hat{p}$ igualará o excederá .750? Dado que $\pi = .50$, esta probabilidad es P (6) + P (7) + P (8) = .10938 + .03125 + .00391 = .14454. Esto, entonces, es la potencia de la prueba. Por definición, beta es 1 – .14454 = .85546. Por consiguiente, parece improbable que se llegue a una decisión correcta para esta prueba. ■

EJEMPLO 4.35

¿Cuál será la potencia y beta si π fuera .55 en vez de .50 en el ejemplo 4.34?

Solución De la tabla 4.6, se sabe que la potencia es

$$.15695 + .05481 + .00837 = .22013.$$

Este valor es más grande que el calculado para $\pi = .50$ esperado, dado que .55 está más allá del valor nulo de .35, que es .50. Beta es 1 – .22013 = .77987. Según usted, ¿qué se podría hacer para incrementar la potencia de estos resultados? ■

Cálculo del tamaño de la muestra. Una pregunta comúnmente planteada por los investigadores es: "¿Cuántos sujetos debo incluir en mi estudio?" Esta pregunta a menudo se resuelve de acuerdo con las limitaciones financieras u otras consideraciones no relacionadas con la estadística. En este apartado examinaremos los aspectos estadísticos de esta pregunta y cómo se relacionan con las pruebas de hipótesis. En general, la solución estadística para esta pregunta requerirá de un programa computacional o de tablas especialmente diseñadas. Simplificaremos este proceso considerando los cálculos del tamaño de la muestra para la prueba Z de una media, que son muy directos. La meta de esta sección es usar la prueba Z de una media para presentar la lógica que fundamenta la solución. La misma lógica general se aplica a pruebas más complejas.

Suponga que un investigador planea usar una prueba Z de una media como herramienta de análisis primaria en su próximo estudio. La hipótesis nula que se someterá a prueba es $H_0 : \mu = 100$, mientras que la hipótesis alternativa es de la forma $H_A : \mu > 100$. Después de algunas reflexiones, el investigador decide que si la media poblacional es 104 o mayor, él querrá tener una probabilidad alta (por ejemplo, 90) de rechazo de la hipótesis nula. Esto implica que el investigador está menos preocupado por detectar medias que difieren por menos de cuatro unidades del valor nulo. Esta decisión quizá esté basada en la opinión de que diferencias de menos de cuatro unidades tienen menos importancia clínica, mientras que diferencias de cuatro o más unidades son más importantes de detectar.

Para hacer más específico el problema, el investigador quiere tener el suficiente tamaño de muestra (n) así como una potencia de .90 para rechazar la hipótesis nula si la media poblacional es 104. Observe que la potencia será mayor que .90 si la media poblacional es mayor que 104. La prueba de una cola será conducida en $\alpha = .05$.

Despejando n en la ecuación 4.10 en la página 132 se obtiene

$$\boxed{n = \frac{\sigma^2 (Z_\beta - Z_\alpha)^2}{(\mu_0 - \mu)^2}} \tag{4.11}$$

Se puede sustituir 100 y 104 para μ_0 y μ, respectivamente, y 1.65 para Z_α. Puesto que .90 de la curva normal cae por arriba de $Z = -1.28$ (véase el panel A de la figura 4.26), este valor se puede sustituir por Z_β. Para propósitos de este ejercicio supondremos que $\sigma^2 = 400$. Esto da entonces

$$n = \frac{400(-1.28 - 1.65)^2}{(100 - 104)^2} = 214.6$$

que se redondea a 215. De esta manera, con 215 sujetos en el estudio, el investigador rechazará la hipótesis nula con probabilidad de .90 si la media verdadera es 104, y obtendrá una mayor probabilidad de rechazo si la media es mayor que 104.

Un problema que surge en la generación de estimaciones del tamaño de la muestra es encontrar un valor de σ^2 para la expresión. Quizá usted argumente que sólo podemos usar s^2 como se hizo con la prueba t, pero recuerde que este cálculo por lo general se realiza *antes* de efectuar el estudio, así que no habrá una muestra de la cual obtener la varianza estimada. Por lo general, la varianza estimada se obtiene de la bibliografía del investigador, de un estudio piloto diseñado para generar la estimación o de algún otro estudio que emplee datos similares. En cualquier caso, las estimaciones del tamaño de la muestra son únicamente eso, estimaciones.

EJEMPLO 4.36

Calcule el tamaño de la muestra requerido para lograr una potencia de .8 con el fin de detectar una media poblacional de 8 para una prueba Z de dos colas de la hipótesis nula $H_0 : \mu = 10$ con $\alpha = .01$. Suponga que $\sigma = 4$.

Solución El 20% de la curva normal cae por arriba y el 80% debajo de un valor Z de .84. (Véase el panel B de la figura 4.26.) Al sustituir este valor para Z_β, y -2.58 para Z_α, 10 para μ_0, 8 para μ, y 16 para σ^2 se obtiene

$$n = \frac{16(.84 - (-2.58))^2}{(10 - 8)^2} = 46.8$$

■

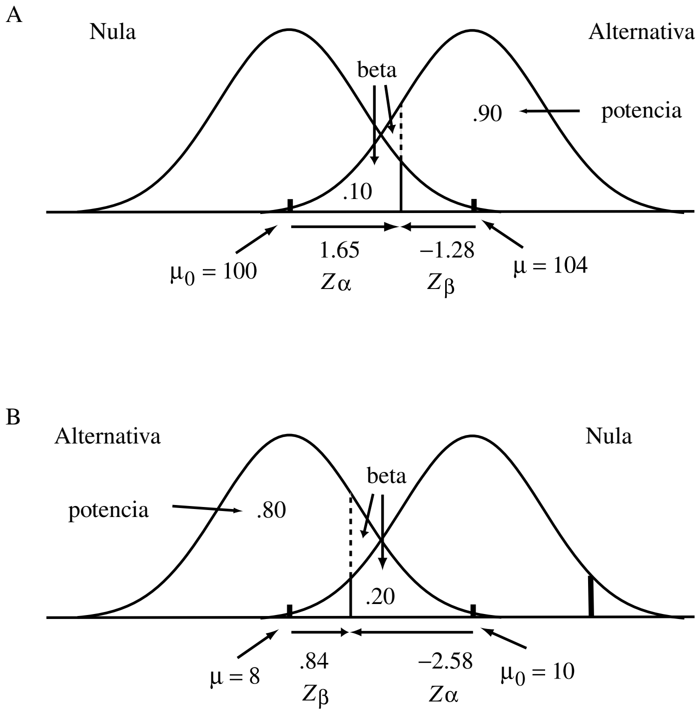


FIGURA 4.26: Cálculos del tamaño de la muestra para pruebas *Z* con una y dos colas para una media. Las distancias están expresadas en términos de errores estándar.

4.4 INTERVALOS DE CONFIANZA

4.4.1 Introducción

Con anterioridad indicamos que hay dos formas básicas de inferencia: la prueba de hipótesis y los intervalos de confianza. Ahora que usted ha completado una introducción a las pruebas de hipótesis regresaremos a la otra forma de inferencia. Quizá la diferencia más importante entre los dos métodos inferenciales radica en las preguntas que implican. Por ejemplo, la prueba *Z* de una media busca información sobre si la media poblacional difiere del valor especificado por la hipótesis nula, mientras que un intervalo de confianza comparable hace la pregunta simple: “¿Cuál es la media poblacional?” Advierta que la última pregunta no implica ninguna forma de hipótesis acerca de la media poblacional, sino simplemente plantea la pregunta de su valor. Como muestra la figura 4.27, ambos métodos emplean estadísticos para dirigir las preguntas formuladas sobre parámetros.

Después de considerar el fundamento de los intervalos de confianza, usted aprenderá a construir tales intervalos para la media poblacional cuando σ es conocida y cuando es desconocida. Usted aprenderá métodos aproximados y exactos para estimar la proporción de la población. En la siguiente sección compararemos los dos métodos y mostraremos que generalmente se prefieren los intervalos de confianza para pruebas de hipótesis cuando ambos métodos son aplicables a un problema en particular.

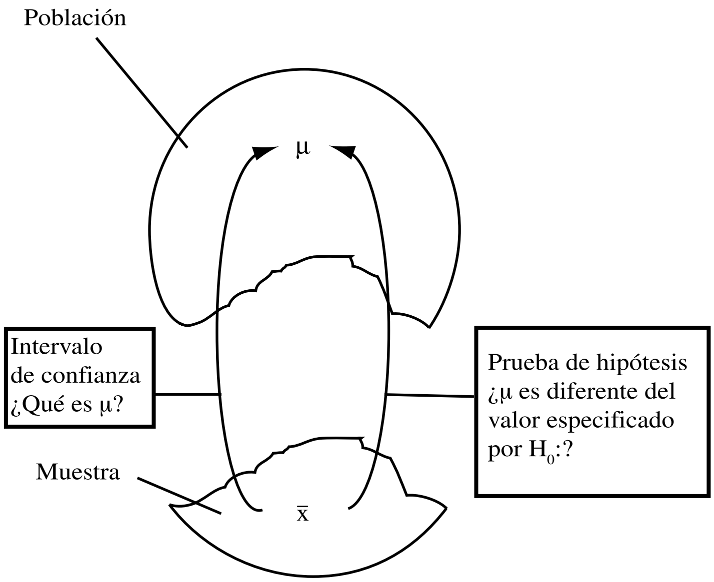


FIGURA 4.27: Preguntas respecto de una media poblacional planteadas por una prueba de hipótesis y un intervalo de confianza.

4.4.2 Razonamiento y método

Intervalos de confianza bilaterales. Considere lo siguiente. Suponga que un investigador selecciona al azar un tamaño de muestra de $n = 100$ de una población normalmente distribuida con desviación estándar (σ) de 50. Se calcula la media de la muestra $\bar{x}$. ¿Cuál es la probabilidad de que $\bar{x}$ tome un valor entre μ y un punto que se localiza un error estándar por arriba de μ? La columna 2 del Apéndice A da esta probabilidad como .3413. Ésta también es la probabilidad de que $\bar{x}$ tome un valor entre μ y un punto que se localiza un error estándar por debajo de μ. De esta forma, la probabilidad de que $\bar{x}$ tome un valor que está dentro de un error estándar de μ es .3413 + .3413 = .6826. Hemos marcado esta región en una curva que representa la distribución muestral de $\bar{x}$ en el panel A de la figura 4.28.

Suponga ahora que un error estándar, que en este ejemplo es

$$\frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = 5,$$

que se suma y se resta de $\bar{x}$. Designaremos $\bar{x} + \frac{\sigma}{\sqrt{n}}$ como U y $\bar{x} - \frac{\sigma}{\sqrt{n}}$ como L. Representamos $\bar{x}$, L y U en el panel A de la figura 4.28. En el ejemplo señalado, la muestra seleccionada aleatoriamente produce un valor de $\bar{x}$, que se encuentra exactamente un error estándar por arriba de μ.

Debe ponerse énfasis en muchos puntos acerca del intervalo L, U (*lower, upper*). Primero, cuando $\bar{x}$ toma un valor que está un error estándar por arriba de μ, el intervalo L, U contiene μ. Esto es, $L \leq \mu \leq U$. Segundo, si $\bar{x}$ ha tomado un valor de un error estándar por debajo de μ (no mostrado en la figura), el intervalo nuevamente contendrá μ, porque μ será igual a U. Tercero, cualquier valor de $\bar{x}$ que esté dentro de un error estándar de μ producirá un intervalo tal que μ será

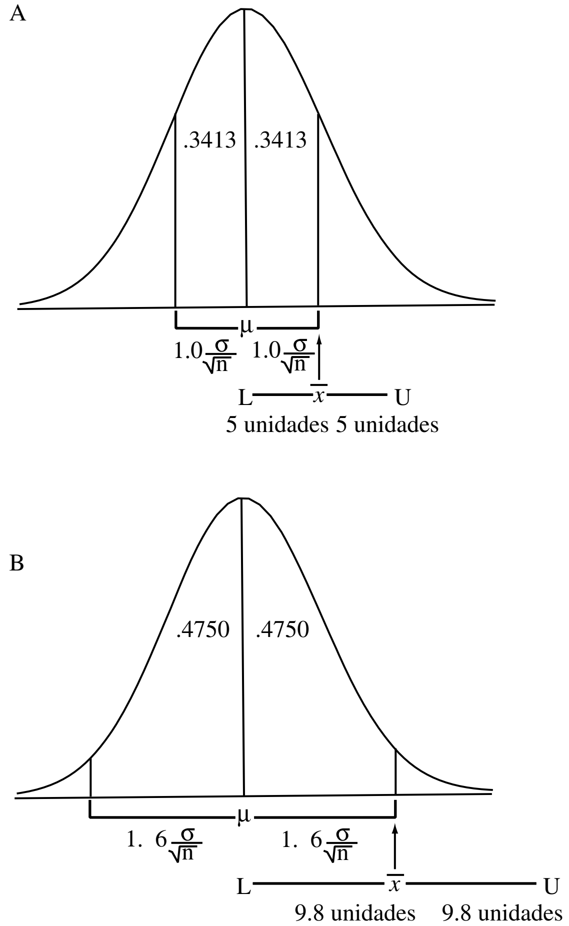


FIGURA 4.28: Razonamiento que fundamenta los intervalos de confianza bilaterales para una media poblacional.

un valor entre L y U, mientras que cualquier valor de $\bar{x}$ que no esté dentro de un error estándar de μ producirá valores de L y U que no contengan μ. Cuarto, la probabilidad[17] de que $\bar{x}$ caiga dentro de un error estándar de μ y, por lo tanto, produzca un intervalo L, U que incluya al valor de μ es aproximadamente .68.

El intervalo L, U se llama un **intervalo de confianza**. L es el **límite inferior** (*lower*) del intervalo de confianza mientras que U es el **límite superior** (*upper*). El **nivel de confianza** o **alcance** del intervalo de confianza es .68. El estadístico alrededor del cual se forma el intervalo, $\bar{x}$ en este caso, se llama **estimación puntual**.

Las implicaciones de lo anterior son las siguientes. Un investigador que desea estimar la media de una población puede seleccionar aleatoriamente una muestra de tamaño n de la población, calcular la media de la muestra y generar el intervalo de confianza como se acaba de describir. El investigador quizá afirme, con un 68% de confianza, que la media poblacional es algún

[17] En este contexto, comentaremos el uso de la palabra probabilidad.

valor entre L y U. Pero el 68% no parece un nivel muy alto de confianza de que μ está en el intervalo. Suponga que el investigador quiere alcanzar un nivel de confianza mayor, digamos, un nivel de 95%.

El nivel de confianza se determina por el *número* de errores estándar sumados y restados a la estimación puntual. ¿Cuántos errores estándar serán sumados y restados a $\bar{x}$ con el fin de formar un intervalo de confianza del 95%? El intervalo tendrá que ser de suficiente longitud para incluir μ en el evento de que $\bar{x}$ tome cualquiera de los valores que forman el .95 central de la curva normal. Esta área es mostrada en el panel B de la figura 4.28. Como se observa en esta figura, .95 de la curva normal cae dentro de 1.96 errores estándar de la media. De esto se infiere que sumar y restar 1.96 errores estándar garantizará que el valor de μ caerá entre L y U cuando $\bar{x}$ esté dentro de 1.96 errores estándar de μ. Por supuesto, si $\bar{x}$ no está dentro de 1.96 errores estándar de μ, el intervalo resultante no contendrá μ. De esta manera, el investigador puede afirmar con un 95% de confianza que μ es algún valor entre L y U. Más adelante se darán detalles de cómo se forma un intervalo de confianza específico.

Intervalos de confianza unilaterales. En ciertas circunstancias los investigadores pueden estar más interesados en un extremo de un intervalo de confianza que en otro. En tales casos, el investigador puede elegir si forma un intervalo de confianza unilateral. Un intervalo de confianza **unilateral** es un enunciado de confianza que consiste *ya sea* en L *o* U, pero no en ambos. Por ejemplo, un ingeniero ambiental puede muestrear rutinariamente la descarga de agua tratada con el fin de estimar concentraciones de amoniaco (medidas como nitrógeno) contenidas allí. Si la concentración es muy alta, los ríos y estanques cercanos pueden resultar afectados negativamente. Bajas concentraciones de amoniaco no serán motivo de preocupación. En este caso el intervalo de confianza unilateral consistirá únicamente en U. Con un nivel de confianza específico, el investigador podrá entonces afirmar que la concentración de amoniaco en la descarga es *como máximo U*. Si U es menor que algunos niveles aceptables, no se emprenderá ninguna acción. De otra forma, se deberán hacer algunos esfuerzos para reducir la concentración sin importar cuál es el valor de L. La ventaja de los intervalos de confianza unilaterales comparada con los intervalos bilaterales será clara con las siguientes explicaciones. Antes de considerar formas específicas de intervalos de confianza unilateral, daremos un fundamento para su construcción.

En el panel A de la figura 4.29 se marcó el punto en el que la distribución muestral de $\bar{x}$ tal que aproximadamente .95 (.5000 + .4505 = .9505) de la curva está por debajo del punto designado. Como se observa, este punto está 1.65 errores estándar por arriba de la media de distribución. Observe que si L se forma sustrayendo 1.65 errores estándar de $\bar{x}$, L siempre será menor que o igual a μ mientras $\bar{x}$ proceda de la región .95 designada. Si $\bar{x}$ es de punto más que 1.65 errores estándar por arriba de μ, L no será menor que o igual a μ. Esto se observa a partir de la figura donde $\bar{x}$ se posicionó 1.65 errores estándar por arriba de μ. Cualquier valor de $\bar{x}$ menor que el valor mostrado producirá una L menor que μ, mientras que cualquier $\bar{x}$ mayor que el valor mostrado producirá una L que no es menor que o igual a μ. Como hay una probabilidad de .95 de que $\bar{x}$ vendrá de un punto que está 1.65 errores estándar por arriba de la media de la distribución, o un punto menor, el investigador puede declarar con .95 de confianza que la media de una población es mayor que o igual a U. Como un ejemplo, si el error estándar es 2.0, L será igual a $\bar{x} - (1.65)\,(2.0)$ o $\bar{x} - 3.30$.

También es posible que un investigador esté interesado en estimar el valor más grande que μ podría tomar, sin preocuparse por un límite inferior. Esto es, el investigador puede desear hacer un enunciado de la forma "La media de la población no es mayor que U". En tal circunstancia, se puede formar un intervalo de confianza unilateral que consiste únicamente en U. Como se observa

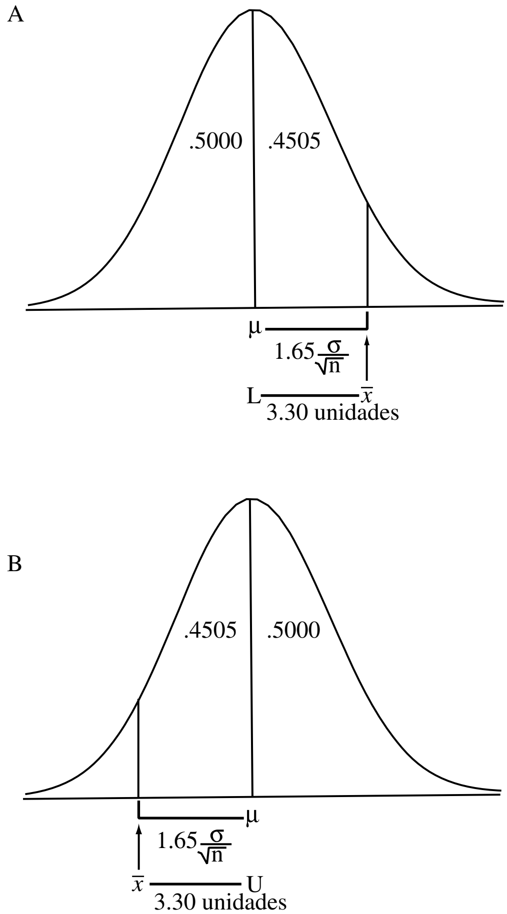


FIGURA 4.29: Fundamento lógico de los intervalos de confianza unilaterales para una media poblacional.

en el panel B de la figura 4.29, cualquier valor de $\bar{x}$ que está 1.65 errores estándar por debajo de μ o más arriba en la distribución producirá un valor de U que es mayor que o igual a μ. De esta manera, el investigador puede tomar una muestra aleatoria de la población, calcular U y luego afirmar con .95 de confianza que μ es menor que U.

4.4.3 Una nota de advertencia

Cuando se pide que enuncie el significado de un intervalo de confianza bilateral del 95%, un estudiante probablemente responderá algo como: "La *probabilidad* de que μ esté entre L y U es .95." Ante una afirmación de este tipo sólo obtendrá el ceño fruncido de su profesor, y si éste está más comprometido con su papel, el estudiante conseguirá una calificación en rojo. El problema puede explicarse mejor por analogía.

Si una canica es seleccionada al azar de una *caja que contiene siete canicas blancas y tres negras*, la probabilidad de que la canica elegida será negra es, por la ecuación 3.1 en la página 52, simplemente la porción de canicas negras en la caja o $\frac{3}{10} = .3$. Ahora suponga que usted ha seleccionado la canica pero no ha abierto la mano. ¿Cuál es la probabilidad de que la canica no observada en *su mano* sea negra? Nuevamente, es la proporción de canicas negras en su mano. Pero, sólo hay una canica en su mano. Entonces, la probabilidad es ya sea uno o cero, dependiendo de si la canica en su mano es negra o blanca.

La misma lógica se aplica a los intervalos de confianza. *Antes* de seleccionar una muestra de la población puede afirmarse que la probabilidad de que el intervalo de confianza construido contenga a μ es .95 (o algún otro nivel específico). La razón para esto es que la proporción de valores posibles de $\bar{x}$ que producirá valores de L y U que contienen a μ es .95. Pero una vez que la muestra ha sido seleccionada y el intervalo de confianza construido, la pregunta "¿Cuál es la probabilidad de que *este* intervalo de confianza contenga a μ?" debe ser contestada, como con la canica en su mano, uno o cero. Al igual que en el caso de la canica, ahora usted ha limitado el problema a un intervalo de confianza, así que la proporción que contiene a μ es ya sea uno o cero.

Debido a ciertas consideraciones filosóficas subyacentes, algunos profesores de estadística requieren enunciados muy precisos respecto del significado de los intervalos de confianza, mientras que otros parecen no estar tan preocupados por ello. Un enunciado aceptable para muchos especialistas en estadística es "el 99% de todos los intervalos de confianza construidos de esta manera contendrán a μ".

4.4.4 Intervalos de confianza para μ cuando σ es conocida

En ésta y las siguientes dos secciones daremos las especificaciones para construir intervalos de confianza para μ y π. Comenzaremos con intervalos de confianza para μ cuando σ es conocida.

Cuando σ es conocida, la curva normal puede utilizarse como un modelo de la distribución muestral de $\bar{x}$, de manera que L y U se obtienen mediante

$$\boxed{L = \bar{x} - Z\frac{\sigma}{\sqrt{n}}} \tag{4.12}$$

y

$$\boxed{U = \bar{x} + Z\frac{\sigma}{\sqrt{n}}} \tag{4.13}$$

Intervalos bilaterales. El nivel de confianza asociado con un intervalo está determinado por Z. Por ejemplo, suponga que se construirá un intervalo de confianza bilateral de .95. El número de errores estándar que habrán de restarse y sumarse a $\bar{x}$ debe ser suficiente para que L y U contengan a μ cuando $\bar{x}$ tome cualquiera de los valores del 95% central alrededor de μ. Esto se demuestra en el panel B de la figura 4.28 en la página 139. Esta figura indica que .95 del área de la curva queda dentro de 1.96 errores estándar de μ. El valor apropiado de Z puede obtenerse notando que $\frac{.95}{2} = .475$ de la curva cae entre μ y el punto deseado. Con el fin de encontrar la puntuación Z apropiada ahora debemos leer la tabla de la curva normal en orden inverso al que hemos utilizado con anterioridad. Esto es, debemos buscar .4750 en la columna 2 de la tabla y luego encontrar el valor Z asociado. Como usted puede ver en la tabla de la curva normal, ese punto es 1.96. De aquí se infiere que sumando y sustrayendo 1.96 errores estándar a $\bar{x}$, podemos asegurar que μ caerá entre L y U cuando $\bar{x}$ tome cualquier valor que esté dentro de 1.96 errores estándar de μ. La probabilidad de que $\bar{x}$ tome un valor tal es .95. Algunos ejemplos ayudarán a aclarar este método de la construcción de intervalos de confianza.

EJEMPLO 4.37

Dada la siguiente información, utilice los intervalos de confianza bilaterales de 90, 95 y 99% para estimar la media de CI de una población de niños cuyas madres recibieron atención prenatal inadecuada. ¿Qué ocurre al intervalo de confianza mientras el nivel de confianza se incrementa?

$$\sigma = 16 \qquad n = 60 \qquad \bar{x} = 90.1$$

Solución El error estándar de la media ($\frac{\sigma}{\sqrt{n}}$) es $\frac{16}{\sqrt{60}} = 2.066$. ¿Cuántos errores estándar deberán restarse y sumarse a $\bar{x}$ con el fin de formar un intervalo de confianza del 90%? Esto puede determinarse notando que .90/2 = .45 de la curva cae entre μ y la puntuación Z deseada. La columna 2 del Apéndice A muestra que no hay un valor dado para .4500 y que hay dos valores .4495 y .4505 que están igualmente cerca del área buscada. Seguiremos la convención de elegir el área en la columna 2 que es la *más cercana* al área deseada cuando el área deseada no está en la tabla; o elegiremos el área más grande en la columna 2 cuando dos áreas están igualmente cerca del área buscada. En este caso hay dos áreas igualmente cerca de .4500, así que escogeremos la más grande, que es .4505. La puntuación Z asociada es 1.65. Observe que aproximadamente .90 de la curva cae 1.65 errores estándar dentro de μ. De esta forma, si se restan o se suman 1.65 errores estándar a $\bar{x}$, podemos estar 90% seguros de que μ quedará incluida. Así, mediante las ecuaciones 4.12 y 4.13, tenemos

$$L = \bar{x} - Z\frac{\sigma}{\sqrt{n}} = 90.1 - (1.65)(2.066) = 86.69$$

y

$$U = \bar{x} + Z\frac{\sigma}{\sqrt{n}} = 90.1 + (1.65)(2.066) = 93.51$$

Como se ha mostrado, un intervalo de confianza del 95% se forma restando y sumando 1.96 errores estándar de manera que

$$L = 90.1 - (1.96)\ (2.066) = 86.05$$

y

$$U = 90.1 + (1.96)(2.066) = 94.15$$

El valor Z que se empleará en la construcción de un intervalo de confianza bilateral de .99 puede encontrarse observando que .99/2 = .4950. Puesto que hay dos valores igualmente cercanos a .4950 en la columna 2 del Apéndice A (esto es, .4949 y .4951), seguiremos la convención de usar el área más grande, en este caso .4951, la cual tiene un valor Z asociado de 2.58. El intervalo será entonces

$$L = 90.1 - (2.58)\ (2.066) = 84.77$$

y

$$U = 90.1 + (2.58)\ (2.066) = 95.43$$

Como muestran los resultados, se gana confianza a expensas de intervalos grandes. ■

EJEMPLO 4.38

Utilice la siguiente información para formar un intervalo de confianza bilateral del 80%.

$$\sigma = 22 \qquad n = 100 \qquad \bar{x} = 220.5$$

Solución El error estándar es $\frac{22}{\sqrt{100}} = 2.20$. El valor Z a utilizarse para el cálculo debe ser de suficiente tamaño para producir valores de L y U que contengan μ cuando $\bar{x}$ tome cualquiera de los valores que formen el .80 central de la curva. Para encontrar este valor de Z, buscamos el área .80/2 = .40 en la columna 2 de la tabla de la curva normal. El área más cercana a .4000 es .3997, la cual tiene un valor Z asociado de 1.28. Al sustituir en las ecuaciones 4.12 y 4.13 se obtiene

$$L = \bar{x} - Z\frac{\sigma}{\sqrt{n}} = 220.5 - (1.28)(2.2) = 217.68$$

y

$$U = \bar{x} + Z\frac{\sigma}{\sqrt{n}} = 220.5 + (1.28)(2.2) = 223.32.$$

Por lo tanto, podemos estar 80% seguros de que la media poblacional es algún valor en el rango de 217.68 y 223.32. ■

Intervalos unilaterales. Suponga que un investigador desea construir un intervalo de confianza del 95% con el fin de obtener un límite inferior de μ. Esto es, el investigador desea encontrar un valor de L que, con un 95% de confianza, pueda ser declarado menor que o igual a μ. Como se observa en el panel A de la figura 4.29, el valor Z aproximado que habrá de utilizarse para este propósito es el valor que tenga .95 – .50 = .45 de la curva entre ese valor y μ. Al utilizar la regla previamente establecida respecto de las áreas que están igualmente cerca de lo deseado, identificamos 1.65 como el valor Z aproximado.

Si $\bar{x}$ es 10.2 y el error estándar es 2.0, entonces, por la ecuación 4.12

$$L = \bar{x} - Z\frac{\sigma}{\sqrt{n}} = 10.2 - (1.65)(2.0) = 6.9.$$

El investigador puede afirmar con un 95% de confianza que la media poblacional es mayor que o igual a 6.9.

Si el investigador estuviera interesado en un límite superior estimado de μ, entonces, por la ecuación 4.13

$$U = \bar{x} + Z\frac{\sigma}{\sqrt{n}} = 10.2 + (1.65)(2.0) = 13.5.$$

Observe que, como con las pruebas de hipótesis de una y dos colas, los intervalos de confianza de uno y dos lados no emplean el mismo valor de Z cuando se construyen intervalos con iguales niveles de confianza. Por ejemplo, se utiliza 1.65 para construir un intervalo unilateral del 95%, pero se emplea 1.96 para un intervalo bilateral del 95%.

EJEMPLO 4.39

Utilice la siguiente información para construir el límite inferior estimado de μ en los niveles de confianza del 90 y 99%.

$$\sigma = 44 \qquad n = 150 \qquad \bar{x} = 105.8$$

Solución Se construirán intervalos de confianza unilaterales porque sólo se requieren los límites inferiores estimados de μ. El error estándar de la media es $\frac{44}{\sqrt{150}} = 3.593$. Para el intervalo del 90%, el valor de Z aproximado tendrá .90 – .50 = .40 de la curva entre sí mismo y la media de la curva. La columna 2 del Apéndice A muestra que este valor será (aproximadamente) 1.28. El límite inferior estimado es entonces

$$L = \bar{x} - Z\frac{\sigma}{\sqrt{n}} = 105.8 - (1.28)(3.593) = 101.2.$$

Para el intervalo .99, el valor Z tendrá .99 – .50 = .49 de la curva entre sí mismo y la media de distribución. La columna 2 del Apéndice A indica que este valor será (aproximadamente) 2.33. El límite inferior estimado es, entonces,

$$L = 105.8 - (2.33)\ (3.593) = 97.4.$$

■

EJEMPLO 4.40

Utilice la siguiente información para obtener los intervalos de confianza de 86 y 94% para el límite superior del valor de μ.

$$\sigma = 4.5 \qquad n = 80 \qquad \bar{x} = 12.0$$

Solución Los valores Z de los dos intervalos son respectivamente 1.08 y 1.56. Utilizando estos valores con un error estándar de $\frac{4.5}{\sqrt{80}} = .503$ se obtiene

$$U = \bar{x} + Z\frac{\sigma}{\sqrt{n}} = 12.0 + (1.08)(.503) = 12.5$$

y

$$U = 12.0 + (1.56)(0.503) = 12.8.$$

Suposiciones. Las suposiciones que fundamentan el intervalo de confianza para μ cuando σ es conocida son las mismas que aquellas que fundamentan la prueba Z de una media (véase la página 98). Las violaciones de una o más de estas suposiciones pueden dar por resultado que el nivel de confianza que en realidad se utilizó difiera del pretendido. La tabla 4.2 en la página 100 resultará útil para conocer el alcance de una de tales violaciones. Esto puede hacerse restando α_E de uno. Por ejemplo, cuando muestras de tamaño cinco son extraídas de una distribución no normal considerada en esa tabla, el nivel real de confianza es 1 – .046 = .954 en vez del pretendido de .950. Asimismo, como se ve en la tabla 4.3, con una violación moderada de la suposición de independencia, el intervalo de confianza del 95% para μ tiene un alcance real de 1 – .11 = .89. El intervalo de confianza para μ cuando σ es conocida es robusto o no robusto de acuerdo con las mismas condiciones que la prueba Z para una media es robusta o no robusta.

4.4.5 Intervalo de confianza para μ cuando σ no es conocida

Usted recordará que en la situación comúnmente encontrada, donde σ no es conocida, se utiliza la prueba t de una media en lugar de la prueba Z de una media. Lo mismo es cierto para intervalos de confianza. Cuando σ no es conocida, se utiliza s como un estimado para el parámetro σ

desconocido, lo cual significa que la distribución pertinente es t en lugar de Z. Las ecuaciones para L y U son entonces

$$\boxed{L = \bar{x} - t\frac{s}{\sqrt{n}}} \tag{4.14}$$

y

$$\boxed{U = \bar{x} + t\frac{s}{\sqrt{n}}} \tag{4,15}$$

donde s es la desviación estándar de la muestra y t es el valor apropiado de la tabla t con $n - 1$ grados de libertad. Por ejemplo, suponga que deseamos construir un intervalo de confianza bilateral del 95% para μ utilizando los datos del apartado 4.3.4, los cuales hemos reproducido aquí por conveniencia.

Muestra: 6.0 8.0 5.5 4.5 8.5 4.0 3.5

La desviación estándar de la muestra puede obtenerse aplicando la ecuación 2.16 de la página 37 a las puntuaciones sumadas y la suma de las puntuaciones al cuadrado, como se muestra abajo.

X	X^2
6.0	36.00
8.0	64.00
5.5	30.25
4.5	20.25
8.5	72.25
4.0	16.00
3.5	12.25
40.0	251.00

$$s = \sqrt{\frac{\sum x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{251 - \frac{(40)^2}{7}}{7-1}} = \sqrt{\frac{22.429}{6}} = 1.933$$

y

$$\bar{x} = \frac{\sum x}{n} = \frac{40}{7} = 5.714.$$

El valor apropiado para t puede encontrarse en la tabla en el Apéndice B con $7 - 1 = 6$ grados de libertad. Como deseamos construir un intervalo de confianza bilateral del 95% utilizamos la fila y la columna así designadas. Esto da un valor para t de 2.447. Por las ecuaciones 4.14 y 4.15

$$L = 5.714 - 2.447\frac{1.933}{\sqrt{7}} = 3.926$$

y

$$U = 5.714 + 2.447\frac{1.933}{\sqrt{7}} = 7.502.$$

Así, el investigador puede tener una confianza del 95% de que esta muestra fue extraída de una población cuya media está entre 3.926 y 7.502. Un estimado del límite inferior unilateral del 95% de μ sería

$$L = 5.714 - 1.943\frac{1.933}{\sqrt{7}} = 4.294. \qquad \blacksquare$$

EJEMPLO 4.41

Los siguientes datos representan una muestra (ficticia) de niveles de glucosa en la sangre, tomados de 10 niños en edades entre 14 y 16 años quienes reportaron que rutinariamente consumen comida rápida tres o más veces a la semana. Utilice este dato para formar un intervalo de confianza bilateral del 99%.

Muestra: 100 99 97 104 124 120 89 122 118 101

Solución La suma de las observaciones y la suma de sus cuadrados son las siguientes.

X	X^2
100	10000
99	9801
97	9409
104	10816
124	15376
120	14400
89	7921
122	14884
118	13924
101	10201
1074	116732

Aplicando las ecuaciones 2.1 y 2.16 se obtiene

$$x = \frac{\sum x}{n} = \frac{1074}{10} = 107.4$$

y

$$s\sqrt{\frac{\sum x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{116732 - \frac{(1074)^2}{10}}{10-1}} = \sqrt{\frac{1384.4}{9}} = 12.403.$$

El Apéndice B muestra que el valor t apropiado para un intervalo de confianza bilateral del 99% basado en 10 – 1 = 9 grados de libertad es 3.250. Luego, por las ecuaciones 4.14 y 4.15,

$$L = \bar{x} - t\frac{s}{\sqrt{n}} = 107.4 - 3.250\frac{12.403}{\sqrt{10}} = 94.65$$

y

$$U = \bar{x} + t\frac{s}{\sqrt{n}} = 107.4 + 3.250\frac{12.403}{\sqrt{10}} = 120.15. \qquad \blacksquare$$

Suposiciones. Las suposiciones que fundamentan el intervalo de confianza para μ cuando σ no es conocida son las mismas que aquellas que fundamentan la prueba *t* de una media, las cuales, a la vez, son las mismas que fundamentan la prueba *Z* de una media (véase la página 98). Las violaciones de una o más de estas suposiciones pueden dar por resultado que el nivel de confianza que en realidad se utilizó difiera del pretendido. La tabla 4.4 resultará útil para conocer el alcance de una de tales violaciones. Esto puede hacerse restando α_E de uno. Por ejemplo, cuando muestras de tamaño cinco son extraídas de una distribución no normal considerada en esa tabla, el nivel real de confianza es 1 – .069 = .931, en vez del pretendido de .950. El intervalo de confianza para μ cuando σ no es conocida es robusto o no robusto de acuerdo con las mismas condiciones que la prueba *t* para una media es robusta o no robusta.

4.4.6 Intervalo de confianza para π

Hay un número aproximado de métodos para construir límites de confianza para una porción de la población (π). Sin embargo, éstos a menudo no son bastante exactos para muchas aplicaciones, especialmente cuando el tamaño de la muestra (*n*) no es grande o cuando $\hat{p}$ está cerca de cero o uno. También hay un método exacto que supera esta dificultad.

En tanto que los investigadores lo utilizan comúnmente, nosotros describiremos y daremos un ejemplo breve de cálculo para un método aproximado basado en la curva normal. Sin embargo, para sus propias aplicaciones, usted deberá considerar utilizar el método exacto, que se describe más adelante.

Un método aproximado. Una aproximación para *L* y *U* está dada por

$$\boxed{L = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}}} \tag{4.16}$$

y

$$\boxed{U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}}} \tag{4.17}$$

donde $\hat{p}$ es la proporción de éxitos en la muestra de tamaño *n* y $\hat{q}$ es la proporción de fracasos, de manera que $\hat{q} = 1 - \hat{p}$. *Z* determina el nivel de confianza y se encuentra mediante la forma descrita en la sección 4.4.4.

Como un ejemplo de aplicación, suponga que un investigador de política de salud desea estimar la proporción de adultos que viven en un condado sureño rural y que tienen alguna forma de seguro de salud. Para cumplir este objetivo, se entrevista a una muestra de 350 adultos que viven en ese condado. De las 350 personas entrevistadas, 112 o 112/350 = .32 reportan que actualmente tienen algún tipo de seguro de salud. Un intervalo de confianza bilateral del 95% puede construirse como

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .32 - 1.96\sqrt{\frac{(.32)(.68)}{350}} = .27$$

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .32 + 1.96\sqrt{\frac{(.32)(.68)}{350}} = .37.$$

Entonces, el investigador puede estar un 95% confiado de que la proporción de adultos que viven en el condado y tienen seguro de salud está entre .27 y .37. Es importante notar el resultado que se hubiera alcanzado si $\hat{p}$ estuviera más cerca de cero (por ejemplo, .02) y n fuera muy reducida (por ejemplo, $n = 25$). En este caso

$$L = .02 - 1.96\sqrt{\frac{(.02)(.98)}{25}} = -.03$$

y

$$U = .02 + 1.96\sqrt{\frac{(.02)(.98)}{25}} = -.07.$$

Pero, por supuesto, las proporciones no pueden ser negativas. En lugar de declarar que la proporción de la población está entre −.03 y .07, el investigador utilizará un intervalo .00 a .07. El punto importante, sin embargo, es que este método para la construcción de intervalos de confianza no es exacto cuando el tamaño de la muestra no es lo bastante grande y $\hat{p}$ está muy cerca de cero o uno.

Suposiciones. Vuelva a la página 116. Además de los comentarios expresados aquí, usted deberá tener en mente que los límites de confianza de las proporciones menores a cero o mayores que uno deberán tomar estos valores como su límite legítimo inferior o superior.

Método exacto. Así como la distribución binomial puede usarse para formar la base de una prueba exacta de hipótesis respecto de π (véase la página 108), esta misma distribución puede formar la base de un intervalo de confianza exacto para estimar este parámetro. Es posible comprender mejor el fundamento de este intervalo exacto después de leer el siguiente apartado. Por el momento, simplemente diremos que esto da confianza en la relación especial que existe entre las pruebas de hipótesis y los intervalos de confianza.

Sin embargo, una vez aclarado esto, debemos añadir que el *método* por el cual estos intervalos se forman no es sencillo y depende de conceptos de estadística matemática que están lejos del alcance de este libro. Por esta razón, le mostraremos *cómo* construir estos intervalos, pero no intentaremos decirle *por qué* este método es apropiado. Estos límites se calculan mediante

$$\boxed{L = \frac{S}{S + (b - S + 1)F_L}} \tag{4.18}$$

y

$$\boxed{U = \frac{(S + 1)F_U}{n - S + (S + 1)F_U}} \tag{4.19}$$

donde S es el número de éxitos (*successes*) en la muestra, n es el número de observaciones en la muestra, y F_L y F_U son los valores apropiados de una distribución F. F_L y F_U pueden obtenerse del Apéndice C. Observe que, a diferencia de la tabla t en el Apéndice B, la tabla F puede consultarse con *dos* diferentes grados de libertad. El primero de ellos, al que llamaremos los grados de libertad del numerador, está listado a través de la parte superior de la tabla; mientras que el segundo, al cual llamaremos grados de libertad del denominador, aparece a lo largo de la orilla de la tabla. Por ejemplo, el valor de F apropiado para utilizarse en el cálculo de un intervalo de confianza bilateral del 95%, suponiendo que los grados de libertad del numerador son iguales a 4 y los grados de libertad del denominador son 20, será 3.51. Con el fin de facilitar el cálculo, utilizaremos la notación df_{LN}, df_{LD}, df_{UN}, y df_{UD} para representar respectivamente los grados de libertad (*degrees of freedom*) del numerador para calcular L, los grados de libertad del denominador para calcular L, los grados de libertad del numerador para calcular U y los grados de libertad del denominador para calcular U. Los grados de libertad utilizados para calcular los límites inferior y superior se obtienen mediante

$$\boxed{df_{LN} = 2(n - S + 1)} \tag{4.20}$$

$$\boxed{df_{LD} = 2S} \tag{4.21}$$

$$\boxed{df_{UN} = 2(S + 1)} \tag{4.22}$$

$$\boxed{df_{UD} = 2(n - S)} \tag{4.23}$$

EJEMPLO 4.42

Una muestra aleatoria de 10 niños con niveles normales de glucosa en la sangre, quienes tienen uno o más hermanos con diabetes, es examinada para buscar anticuerpos asociados con esa enfermedad. El 40% de estos niños resultaron positivos para el anticuerpo. Utilice el método exacto para construir un intervalo de confianza bilateral del 95% y estimar la proporción de niños de este tipo en la población que resultaron positivos al anticuerpo. Compare este intervalo con el construido por el método aproximado explicado en la página 148.

Utilice los mismos datos para construir un intervalo de confianza unilateral exacto del 95% para el límite inferior de la proporción de la población. Compare este intervalo con el construido por el método aproximado explicado en la página 148.

Solución. Empezamos por calcular los grados de libertad necesarios para encontrar F_L. Por las ecuaciones 4.20 y 4.21 y el hecho de que el número de éxitos (S) es

$$df_{LN} = 2(n - S + 1) = 2(10 - 4 + 1) = 14$$

y

$$df_{LD} = 2S = (2)(4) = 8.$$

El Apéndice C muestra que con los grados de libertad del numerador de 14 y los grados de libertad del denominador de 8, el valor apropiado de F para la construcción de un intervalo de confianza bilateral del 95% es 4.13. Utilizando este valor en la ecuación 4.18 se obtiene

$$L = \frac{S}{S + (n - S + 1)F_L} = \frac{4}{4 + (10 - 4 + 1)4.13} = .122.$$

Calculando los grados de libertad para F_U por medio de las ecuaciones 4.22 y 4.23 tenemos que

$$df_{UN} = 2\ (S + 1) = 2\ (4 + 1) = 10$$

y

$$df_{UD} = 2\ (n - S) = 2(10 - 4) = 12.$$

El valor F con los grados de libertad del numerador y denominador de 10 y 12, respectivamente, que se utilizará en la construcción de un intervalo de confianza bilateral del 95% es, por el Apéndice C, 3.37. Entonces, por la ecuación 4.19

$$U = \frac{(S + 1)F_U}{n - S + (S + 1)F_U} = \frac{(4 + 1)3.37}{10 - 4 + (4 + 1)3.37} = .737.$$

El intervalo exacto de confianza bilateral del 95% es entonces .122 a .737. La gran distancia de este intervalo se debe al tamaño pequeño de la muestra.

El intervalo aproximado construido por las ecuaciones 4.16 y 4.17 es

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .4 - 1.96\sqrt{\frac{(.4)(.6)}{10}} = .096$$

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .4 + 1.96\sqrt{\frac{(.4)(.6)}{10}} = .704.$$

El intervalo aproximado de .096 a .704 es diferente del intervalo exacto calculado arriba. Esto es de esperarse porque, por lo general, no podemos confiar en el método aproximado cuando el tamaño de la muestra es pequeño.

El intervalo exacto unilateral para el límite inferior estimado de π puede obtenerse sustituyendo el valor aproximado del Apéndice C en la ecuación 4.18. En el Apéndice C con los grados de libertad previamente calculados por 4.20 y 4.21 (esto es, 14 y 8), el valor F para un intervalo unilateral del 95% es 3.24. Si se sustituye este valor en la ecuación 4.18 se obtiene

$$L = \frac{S}{S + (n - S + 1)F_L} = \frac{4}{4 + (10 - 4 + 1)3.24} = .150.$$

El método aproximado da

$$U = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .4 - 1.65\sqrt{\frac{(.4)(.6)}{10}} = .144.$$ ■

EJEMPLO 4.43

Dados 24 éxitos en una muestra de tamaño 28, construya un intervalo de confianza bilateral exacto del 99% para π. Construya un intervalo de confianza unilateral exacto del 99% para el límite superior de π.

Solución. Los grados de libertad del numerador y del denominador para F_L son, por las ecuaciones 4.20 y 4.21

$$df_{LN} = 2(28 - 24 + 1) = 10$$

y

$$df_{LD} = (2)(24) = 48$$

Del Apéndice C, F_L es 3.01. Entonces, el límite inferior mediante la ecuación 4.18 es

$$L = \frac{24}{24 + (28 - 24 + 1)3.01} = .615.$$

Los grados de libertad del numerador y del denominador para F_U son, mediante las ecuaciones 4.22 y 4.23,

$$df_{UN} = 2(24 + 1) = 50$$

y

$$df_{UD} = 2\ (28 - 24) = 8.$$

Con F_U de 6.22, U es entonces

$$U = \frac{(24+1)6.22}{28-24+(24+1)6.22} = .975.$$

El intervalo de confianza bilateral del 99% es, entonces, .615 y .975.

El límite superior unilateral es

$$U = \frac{(24+1)5.07}{28-24+(24+1)5.07} = .969. \quad \blacksquare$$

Suposiciones. Las suposiciones que fundamentan el método exacto de formulación de intervalos de confianza para la estimación de π son los mismos que aquellos para la prueba exacta de hipótesis, discutida en la página 116. Una nota explicativa debe agregarse para este método de construcción exacta de intervalos de confianza. La exactitud del cálculo de L y U depende del número de lugares decimales utilizados en F_L y F_U. Como la tabla F en este libro considera dos lugares decimales, usted no deberá confiar en más de dos dígitos de exactitud en su cálculo de L y U. Esto puede ser particularmente importante cuando los límites están cercanos a cero o uno. Por ejemplo, si un límite inferior actual es .0003, usted no puede contar con obtener este resultado de la tabla F incluida aquí. En tales casos, podrá obtener valores F más precisos de otras tablas o de un programa computacional comúnmente disponible.

4.5 COMPARACIÓN DE PRUEBAS DE HIPÓTESIS E INTERVALOS DE CONFIANZA

Las pruebas de hipótesis y los intervalos de confianza están más relacionados de lo que inicialmente pudiera parecer. De hecho, por lo general, se prefiere los intervalos de confianza para probar hipótesis. Por esta razón y otras que se expondrán a continuación, los intervalos de confianza, en general, son preferibles a las pruebas de hipótesis en situaciones donde ambos puedan emplearse. En esta sección le mostraremos por qué y cómo los intervalos de confianza pueden utilizarse para realizar pruebas de hipótesis. Terminaremos comparando la información que ofrece cada uno.

4.5.1 Pruebas de hipótesis de dos colas e intervalos de confianza bilaterales

Suponga que se lleva a cabo una prueba Z de dos colas para una media en $\alpha = .05$. La distribución nula utilizada para esta prueba se muestra en los paneles A y B de la figura 4.30. Puesto que la prueba es de dos colas y α es .05, la distancia del valor de la media hipotética (μ_0) al límite delantero de las regiones críticas es 1.96 errores estándar (véase el Apéndice A). En el panel A de esta figura el estadístico de prueba ($\bar{x}$ en este caso) está en una región crítica que se adelanta al rechazo de la hipótesis nula.

Ahora, suponga que un intervalo de confianza del 95% está construido alrededor de $\bar{x}$ y se muestra en el panel A. Observe que un intervalo de confianza bilateral del 95% se forma sumando y restando 1.96 errores estándar a $\bar{x}$. Debido a que ésta también es la distancia de μ_0 al límite de la región crítica, se infiere que cuando $\bar{x}$ está en la región crítica, el valor nulo (μ_0) no estará en el intervalo formado por L y U.

Por comparación, el panel B describe la situación donde el estadístico de prueba no está en la región crítica, lo que conduce a no rechazar la hipótesis nula. En esta situación, μ_0 estará contenida en el intervalo formado por L y U. De los resultados mostrados en los paneles A y B,

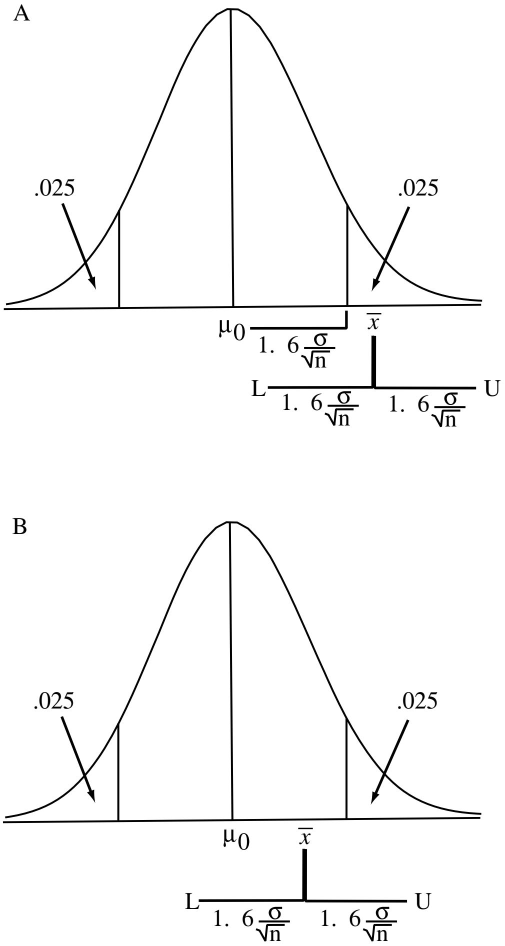


FIGURA 4.30: Vínculo entre la prueba de hipótesis de dos colas y el intervalo de confianza bilateral.

se desprende que un investigador, que desea utilizar una prueba *Z* de dos colas para una media con $\alpha = .05$ y probar la hipótesis nula de que μ es igual a μ_0, podrá realizar la prueba construyendo un intervalo de confianza bilateral del 95% y determinando si μ_0 está en el intervalo resultante. Si μ_0 no está en el intervalo, la hipótesis nula será rechazada. Si μ_0 está en el intervalo la hipótesis nula no será rechazada.

Si el investigador quisiera realizar la prueba de hipótesis con $\alpha = .01$, se requerirá un intervalo de confianza del 99%. En general, si α y el nivel de confianza se expresan como decimales, entonces el nivel de confianza utilizado para llevar a cabo una prueba en el nivel α es $1 - \alpha$. De esta manera, una prueba de hipótesis con $\alpha = .10$ requerirá un intervalo con nivel de confianza de $1 - .10 = .90$.

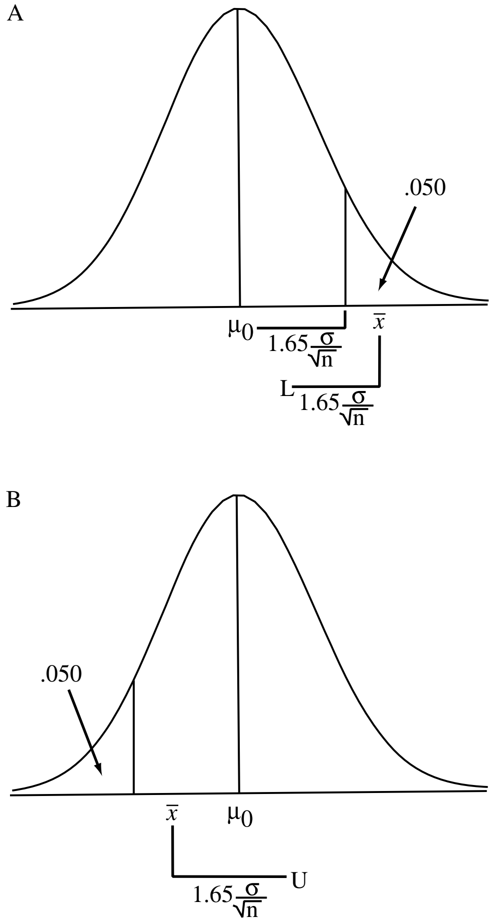


FIGURA 4.31: Relación entre la prueba de hipótesis de dos colas y el intervalo de confianza bilateral.

4.5.2 Pruebas de hipótesis de una cola e intervalos de confianza unilaterales

Así como los intervalos de confianza bilaterales se utilizan para llevar a cabo pruebas de hipótesis de dos colas, los intervalos unilaterales pueden usarse para realizar pruebas de una cola. Por ejemplo, una prueba de una hipótesis nula en contra de una alternativa de la forma

$$H_A : \mu > \mu_0$$

puede realizarse con un intervalo de confianza unilateral para el límite inferior del parámetro estimado. Si $\mu_0 \leq L$ la hipótesis nula se rechaza, de otra forma, no se rechaza. El panel A de la figura 4.31 muestra que cuando el estadístico de prueba está en la región crítica, μ_0 será menor que L.

Se realiza una prueba con hipótesis alternativa de la forma

$$H_A : \mu < \mu_0$$

por medio de un intervalo unilateral para el límite superior del parámetro. Esta prueba se ilustra en el panel B de la figura 4.31. En este ejemplo el estadístico de prueba no está en la región crítica, lo cual implica que $\mu_0 < U$. La hipótesis nula se rechazará si $\mu_0 \geq U$.

En un sentido, entonces, es posible efectuar numerosas pruebas de hipótesis si se da un vitazo a un intervalo de confianza. El investigador sabe que cualquier valor hipotético que caiga entre L y U de un intervalo bilateral no será rechazado, mientras cualquier valor fuera del intervalo será rechazado. Asimismo, el investigador sabe que cualquier valor menor que o igual a L de un intervalo unilateral conducirá a un rechazo, mientras que cualquier valor mayor que L no será rechazado. También es obvio que cualquier valor de μ_0 mayor que o igual a U será rechazado, en tanto que cualquier valor menor que U no será rechazado.

4.5.3 Algunos comentarios adicionales

Además de lo que se mencionó antes, hay dos características de los intervalos de confianza que, en general, los hacen preferibles a las pruebas de hipótesis. Primero, los intervalos de confianza por lo general contestan preguntas más interesantes que las pruebas de hipótesis. Por ejemplo, la pregunta "¿Qué es μ? suele ser más interesante que "¿μ es diferente de 12?". Segundo, si una muestra es muy pequeña para permitir una potencia adecuada, el hecho no es señalado al investigador por una prueba de hipótesis. En contraste, una muestra inadecuada producirá un intervalo de confianza indebidamente largo, lo que alertaría al investigador sobre el hecho de que la muestra es muy pequeña. En general, cuanto más grande es la muestra, más corto es el intervalo de confianza.

Naturalmente, la pregunta que surge es por qué deben utilizarse siempre las pruebas de hipótesis. Y la pregunta derivada de la anterior es por qué usted tiene que aprender tanto acerca de esas pruebas en este capítulo. La razón primordial es que muchas preguntas de interés de los investigadores no pueden responderse mediante los intervalos de confianza. Las únicas respuestas a tales preguntas son dadas por las pruebas de hipótesis. Usted encontrará ejemplos de ello en los siguientes capítulos. También es necesario comprender ciertas pruebas porque comúnmente se citan en la literatura de investigación (aunque en ocasiones de manera inapropiada). Si usted desea comprender plenamente la bibliografía de investigación, deberá conocer las pruebas de hipótesis usadas allí.

4.6 UNA REORIENTACIÓN

Ahora, usted ha terminado de estudiar una introducción elaborada a la inferencia. Mucho de lo que hay que recordar de este libro tiene que ver con los métodos inferenciales comúnmente empleados en la investigación de ciencias de la salud. Si usted comprende los conceptos de este capítulo, estará en muy buena condición para enfrentarse al material restante.

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

$1 - \alpha$ 125
$1 - \beta$ 128
1 – potencia 127
alfa (α) 88

EJERCICIOS

4.1 Dados $\sigma = 50$ y $n = 100$, encuentre la varianza de la distribución muestral de $\bar{x}$.

4.2 ¿Cómo se llama la desviación estándar de la distribución muestral de $\bar{x}$?

4.3 ¿Qué le ocurre a la desviación estándar de la distribución muestral de $\bar{x}$, mientras el tamaño de la muestra se incrementa?

4.4 ¿El teorema del límite central garantiza que la distribución muestral de $\bar{x}$ será normalmente distribuida cuando el tamaño de la muestra es suficientemente grande? Explique su respuesta.

4.5 Utilice la curva normal para determinar la probabilidad aproximada de seleccionar aleatoriamente 25 observaciones de una población, cuya media es 50 y cuya varianza es 100, y encontrar que la media de la muestra es menor que 45.

4.6 Suponga que se seleccionan aleatoriamente 10 observaciones de una población dicotómica en la cual la proporción de éxitos es 0.40. Utilice la distribución binomial para encontrar la probabilidad de que la proporción de éxitos en la muestra será menor que o igual a .30.

4.7 Utilice la curva normal para aproximar la probabilidad descrita en el ejercicio 4.6.

4.8 Dados $\bar{x} = 135$, $\sigma = 40$ y $n = 90$, pruebe la hipótesis nula $H_0 : \mu = 125$ contra la hipótesis alternativa $H_A : \mu \neq 125$ con $\alpha = .10$. Reporte los resultados para ambos métodos, es decir, valor-p versus alfa y valor crítico versus valor obtenido.

4.9 Utilice los siguientes datos para probar la hipótesis nula $H_0 : \mu = 5$ contra la hipótesis alternativa $H_A : \mu < 5$ con el nivel de significancia de .05.

$$\text{Muestra}: 3 \quad 3 \quad 2 \quad 1 \quad 0 \quad 6 \quad 5 \quad 4$$

4.10 Utilice un *intervalo de confianza* construido con los datos en el ejercicio 4.9 para llevar a cabo una prueba de hipótesis de dos colas $H_0 : \mu = 0$ con un nivel .01 de significancia. Explique por qué usted rechaza o no la hipótesis nula.

4.11 Dados $\sigma = 24$, $n = 144$ y $\bar{x} = .52$, lleve a cabo la siguiente prueba con un nivel de significancia de .05.

$$H_{0E} : \mu \le -1.0 \text{ o } \mu \ge 1.0$$
$$H_{AE} : -1.0 < \mu < 1.0$$

¿Cuáles son los términos utilizados para las pruebas de este tipo? ¿Es ésta una prueba de una o de dos colas? ¿Por qué usted no utilizó las pruebas t para realizar la prueba?

4.12 Utilice los siguientes datos para llevar a cabo la siguiente prueba con un nivel de significancia de .05.

$H_{0E} : \mu \le 10$
$H_{AE} : \mu > 10$
$\text{Muestra}: 14 \quad 10 \quad 12 \quad 16 \quad 11 \quad 19 \quad 15 \quad 14$

¿Por qué usted no utilizó una prueba Z?

4.13 Dados $n = 10$ y $\hat{p} = .1$, utilice un método exacto para llevar a cabo la siguiente prueba con un nivel de significancia de 0.05. Utilice ambos métodos, el valor-p versus alfa y el valor obtenido versus el valor crítico.

$H_0 : \pi = .07$
$H_A : \pi \ne .7$

4.14 Utilice un método aproximado para llevar a cabo la prueba descrita en el ejercicio 4.13. ¿Cómo se comparan los dos valores-p?

4.15 Dados $n = 9$ y $\hat{p} = .11$, utilice un método exacto para llevar a cabo la siguiente prueba con un nivel de significancia de .05. Utilice ambos métodos, el valor-p versus alfa y el valor obtenido versus el valor crítico.

$$H_0 : \pi = .5$$
$$H_A : \pi < .5$$

4.16 Defina cada uno de los siguientes términos.

a) Error tipo I
b) Error tipo II
c) Potencia
d) Beta
e) α
f) $1 - \alpha$

4.17 ¿Cuáles son los factores que determinan la potencia de una prueba inferencial?

4.18 Encuentre la potencia y beta para una prueba Z de una media en las siguientes condiciones.

$$H_0 : \mu = 40 \quad \sigma = 20 \quad \alpha = .05$$
$$H_A : \mu > 40 \quad n = 25 \quad \mu = 42$$

4.19 Recalcule la potencia y beta para la prueba descrita en el ejercicio 4.18, suponiendo que el tamaño de la muestra se ha incrementado a 100.

4.20 Encuentre la potencia y beta para la siguiente prueba de significancia.

$$H_0 : \pi = .40$$
$$H_A : \pi < .40$$

Donde

$$\alpha = .05 \quad n = 8 \quad \pi = .10$$

4.21 Calcule el tamaño de la muestra requerido para alcanzar la potencia de .9 y detectar una media poblacional de 6, para una prueba Z de dos colas de la hipótesis nula $H_0 : \mu = 10$ con $\alpha = .05$. Suponga que $\sigma = 2$.

4.22 Utilice los siguientes datos para construir un intervalo de confianza bilateral del 95% para la estimación de μ.

$$\text{Muestra}: 9 \quad 7 \quad 5 \quad 4 \quad 8 \quad 9 \quad 5 \quad 8 \quad 7$$

4.23 Suponga que en una muestra aleatoria de 200 personas mayores de 65 años, se encontró que 150 de los sujetos muestreados fueron vacunados contra viruela en cierto momento de sus vidas. Utilice un método aproximado para construir un intervalo de confianza bilateral del 95% y estimar la proporción de personas en la población de interés que recibieron tales vacunas.

4.24 Recalcule el intervalo de confianza descrito en el ejercicio 4.23 con la condición de que 300 de 400 personas muestreadas recibieron la vacuna. ¿Qué efecto tuvo el incremento en el tamaño de la muestra en el intervalo de confianza?

4.25 Dados seis éxitos en una muestra de tamaño 10, construya un intervalo de confianza exacto del 95% para un límite inferior estimado de π.

4.26 Dados cinco éxitos en una muestra de tamaño 10, utilice un método exacto para construir un intervalo de confianza bilateral para la estimación de π.

A. Las siguientes preguntas se refieren al estudio de caso A (página 469).

4.27 La edad promedio de los participantes en este estudio es reportada como 31.5 años. Este valor puede ser tomado como una estimación de la edad promedio de la población. ¿Qué término se utiliza para las estimaciones de este tipo? Utilice este valor junto con otra información del estudio para construir un intervalo de confianza del 95% (bilateral). Interprete este intervalo.

4.28 Resumiendo sus resultados, los autores del estudio reportan: "Al comenzar a usar... 33 sujetos prefirieron el confort de los lentes acondicionados frente a 26 que prefirieron los de control; sin embargo, esto no fue estadísticamente significativo".

a) Suponiendo que los autores desean saber si los sujetos prefieren un tipo de lentes (tratados) sobre los otros (no tratados), ¿qué tipo de prueba de hipótesis piensa usted que llevaron a cabo? Enuncie la hipótesis nula y la alternativa, luego indique dos métodos que pudieron haber utilizado para llevar a cabo esta prueba.[18]

b) ¿Puede usted proponer otro método de análisis para evaluar esta pregunta?

c) ¿El resultado reportado por los autores puede interpretarse como que los sujetos no tienen preferencia por algún tipo de lentes?

d) ¿Será buena idea combinar los datos de am y pm para propósitos de análisis? Después de todo, esto duplicaría el tamaño de la muestra y, en consecuencia, se incrementaría la potencia estadística.

e) Analice los datos am para el grupo "no" solo y reporte sus resultados. Los autores reportaron este resultado como significativo ($P < .05$). ¿Está usted de acuerdo? ¿Qué podría concluir como resultado de este análisis? ¿Cree que una prueba de hipótesis o un intervalo de confianza sería más apropiado para este análisis? (Nota: F_U con 22 y 4 grados de libertad para un intervalo de confianza del 95% es 8.53).

f) Analice los datos pm para el grupo "no" solo y reporte sus resultados. Los autores reportaron este resultado como significativo ($P < .05$). ¿Está usted de acuerdo? ¿Qué podría concluir como resultado de este análisis?

B. Las siguientes preguntas se refieren al estudio de caso B (página 470).

4.29 Utilice los datos en la tabla J.3 para construir intervalos de confianza bilaterales del 95% para cada uno de los tres grupos y estimar la proporción de sujetos de cada grupo que identificaron correctamente el tipo de brazalete que llevaban puesto. Observe que una identificación correcta para el grupo débil es "imitación". ¿Puede usted especular como fueron ellos capaces de identificar correctamente el brazalete?

D. Las siguientes preguntas se refieren al estudio de caso D (página 471).

4.30 Utilice un intervalo de confianza para estimar la media del conteo CD4 para los sujetos positivos de VIH.

4.31 Suponga que le pidieron utilizar los datos del conteo CD4 para probar la hipótesis nula $H_0 : \mu = 500$ contra una alternativa de dos colas por medio de una prueba Z de una media. ¿Esta prueba parece ofrecer un método razonable para este problema? ¿Por qué?

F. Las siguientes preguntas se refieren al estudio de caso F (página 473).

4.32 Construya intervalos de confianza exactos y aproximados para la estimación de la proporción de casos recurrentes de tuberculosis en la población que son atribuidos a una infección externa. ¿Qué tanto usted podría decir que el intervalo aproximado se acerca al intervalo exacto? (Nota: F_U con 26 y 8 grados de libertad para un intervalo de confianza del 95 es 3.93).

O. Las siguientes preguntas se refieren al estudio de caso O (página 477).

4.33 ¿Puede usted argumentar contra la afirmación hecha por el autor en el punto *e*)?

[18] No se preocupe demasiado si usted no ve esto con suficiente claridad por ahora. Se requiere cierta experiencia para aprender a responder a preguntas de esta clase. Ahora, ha tenido su primera experiencia al respecto.

CAPÍTULO 5

Métodos de muestras apareadas

5.1 INTRODUCCIÓN

Los **datos apareados** se presentan con cierta frecuencia en una variedad de contextos de investigación. Por ejemplo, los investigadores pueden evaluar la efectividad relativa de dos técnicas de cirugía láser para el tratamiento de la retinopatía diabética si aplican una técnica en uno de los ojos de los pacientes que padecen esta enfermedad y aplican el segundo método en el otro ojo. Si después de algún tiempo se mide la agudeza visual en cada ojo para determinar cuál técnica de cirugía dio por resultado una mejor visión, las dos agudezas obtenidas en cada paciente constituirían un par de datos.

Como segundo ejemplo, los investigadores que deseen determinar si un remedio para el resfriado que se vende sin prescripción médica tiene como efecto colateral indeseable la elevación de la presión arterial sistólica, podrían tomar medidas de línea base de presión arterial para cada paciente, administrar el remedio para el resfriado y, después de un periodo de tiempo específico, medir la presión arterial por segunda vez. Después, podrían compararse las medias de las dos presiones arteriales tomadas en dos diferentes momentos para determinar si ocurrió un incremento. De nuevo, las dos mediciones tomadas en cada persona constituirían un par de datos.

Como ejemplo final, podría exponerse a un grupo de mellizos a dos formas de vacuna contra alguna enfermedad infantil. Al azar se asigna a un mellizo de cada par para que reciba la primera vacuna, y el otro mellizo recibe la segunda vacuna. La efectividad relativa de ambas vacunas podría evaluarse comparando las proporciones de mellizos que toman cada tipo de vacuna y que desarrollan la enfermedad. Los indicadores de si cada mellizo del par desarrolla o no la enfermedad serían un par de datos.

Este capítulo estudia pruebas de hipótesis e intervalos de confianza diseñados para comparar promedios o proporciones de datos apareados. Como veremos, muchos de estos métodos son sólo casos especiales de métodos con los que usted ya está familiarizado.

TABLA 5.1: Mediciones de presión arterial sistólica antes y después del tratamiento.

	Pre-tratamiento	Post-tratamiento	diferencia d	d^2
	95	99	4	16
	111	120	9	81
	97	97	0	0
	132	130	–2	4
	144	148	4	16
	100	122	22	484
	120	131	11	121
	110	109	–1	1
	131	140	9	81
	154	153	–1	1
	105	131	26	676
	119	120	1	1
	107	114	7	49
	101	110	9	81
	118	116	–2	4
$\sum$	1744	1840	96	1616

5.2 MÉTODOS RELACIONADOS CON LA DIFERENCIA DE MEDIAS

5.2.1 La prueba *t* (de diferencia) de muestras apareadas

Razonamiento. Al igual que en el ejemplo referido con anterioridad, suponga que un investigador está interesado en determinar si un medicamento para el resfriado que se vende sin prescripción médica tiene como efecto colateral indeseable la elevación de la presión arterial sistólica. Para este fin, el investigador toma la presión arterial de 15 personas del tipo en el que está interesado. Luego, cada persona recibe la dosis recomendada del medicamento. Treinta minutos después de administrar el remedio, se toman de nuevo las presiones arteriales de cada persona. La pregunta que nos interesa es: "¿Son más altas las presiones arteriales después de tomar el medicamento?" La tabla 5.1 muestra presiones arteriales (ficticias) de las 15 personas antes (pre-tratamiento) y después (post-tratamiento) de administrar el fármaco.

Es posible determinar el cambio en la presión arterial de cada persona si restamos su valor pre-tratamiento del valor post-tratamiento. Estas **puntuaciones de diferencia** están representadas en la tabla por *d*. La media de estas puntuaciones de diferencia, que designaremos como *d*, es 96/15 = 6.40. Observe que éste es el mismo valor que se obtendría si la media pre-tratamiento fuera restada de la media post-tratamiento o 122.67 – 116.27 = 6.40.[1]

Así, el cambio promedio en la presión arterial medida antes y después de tomar el remedio para el resfriado es de 6.40 unidades. Pero existen al menos dos explicaciones para esta diferencia

[1] Puede utilizar las reglas de la sumatoria explicadas en el apartado 2.3.2 para verificar estos datos como un resultado general.

de medias. Primero, notamos que cada vez que se toman dos medidas de presión arterial a la misma persona, los valores resultantes rara vez son los mismos aun cuando no haya intervención alguna entre las mediciones. Por lo tanto, pudiera ser que el medicamento no tenga ningún efecto sobre las presiones arteriales. Sin embargo, por azar, las mediciones posteriores fueron mayores en promedio que las mediciones pre-tratamiento. Esto implica que, si repitiéramos el experimento, quizá obtendríamos una diferencia de medias de −6.40 o algún otro valor muy cercano a cero. Una segunda explicación es que el medicamento sí tiene un efecto colateral indeseable. Como resultado, las presiones post-tratamiento individuales tienden a ser más elevadas de lo que hubieran sido sin el medicamento, produciendo así una diferencia de medias de 6.40 unidades. ¿Cuál de estas explicaciones debe creerse? Una prueba de significancia podría ayudarnos a decidir.

Suponga que consideramos los 15 valores *d* como una muestra aleatoria tomada de una población de esos valores. También consideramos la población de valores *d* como aquellos que resultarían si el medicamento para el resfriado no afectara la presión arterial. Es decir, algunos valores serán positivos y algunos negativos, y así la media de la población será cero. Designaremos la media de esta población de diferencias de puntuaciones como μ_d. De esta forma, si el medicamento no afecta, podemos considerar la muestra de *d*s como aleatoria de una población de diferencias de puntuaciones cuya media es cero.

Por otra parte, si el medicamento sí tiene un efecto colateral indeseable, la muestra debe ser considerada como si proviniera de una población cuya media es mayor que cero. La siguiente explicación le ayudará a conceptualizar las dos poblaciones.

Piense en una población grande de personas cuyas presiones arteriales se registran en dos diferentes momentos sin la intervención de ningún tratamiento. La población estadística se compone de la diferencia entre las dos mediciones. Ya que no se administró ningún tratamiento entre las mediciones, las *d*s constan de valores positivos y negativos cuya media es cero.

Por otra parte, suponga que se toman las presiones arteriales de una población grande de personas en dos diferentes momentos y que entre las mediciones se administra un tratamiento que tiende a elevar la presión arterial. En este caso habría más *d*s positivas que negativas en la población y μ_d sería mayor que cero.

Para determinar si el medicamento elevó las presiones arteriales o no, sería útil poder determinar cuál de las dos poblaciones hipotéticas es más probable que represente la población de la cual se obtuvo la muestra de *d*s.

La prueba. Para continuar con el tema anterior, si conducimos la siguiente prueba de hipótesis podremos obtener información referente a cuál de las dos poblaciones hipotéticas representa mejor la población de la cual tomamos los datos de nuestra muestra.

$$H_0 : \mu_d = 0$$
$$H_A : \mu_d > 0$$

En esencia, la hipótesis nula afirma que no ocurrió ningún cambio en la media de la presión arterial, en tanto que la alternativa sostiene que ocurrió un cambio que dio como resultado una media post-tratamiento más alta. Pero, ¿cómo debe conducirse esta prueba? De acuerdo con lo estudiado en el capítulo 4, usted sabe que se puede utilizar una prueba *t* de una media con las diferencias de puntuaciones para conducir la prueba de significancia mencionada con anterioridad. En efecto, una prueba *t* de **muestras apareadas** o de **diferencia apareada** no es más que una prueba *t* de una media realizada con las diferencias de puntuaciones (d). Entonces, la t obtenida es

$$\boxed{t = \frac{\bar{d} - \mu_{d0}}{\frac{s_d}{\sqrt{n}}}} \tag{5.1}$$

donde $\bar{d}$ es la media de la muestra de diferencias de puntuaciones, μ_{d0} es la media hipotética de la población de diferencias de puntuaciones, y s_d es la desviación estándar de la muestra de las diferencias de puntuaciones. Usted reconocerá la ecuación 5.1 como la ecuación para el estadístico t de una media con $n - 1$ grados de libertad. Se han agregado subíndices para recordarle que la prueba se realiza con las diferencias de las puntuaciones. En la práctica, μ_{d0} casi siempre es cero, aunque no necesariamente. Ocurre una excepción con las pruebas de equivalencia, en donde μ_{d0} por lo general no es cero.

EJEMPLO 5.1

Utilice una prueba t de muestras apareadas con $\alpha = .05$ para llevar a cabo la prueba de hipótesis indicada anteriormente con los datos de la tabla 5.1.

Solución Al usar las sumatorias de la tabla 5.1 tenemos que,

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{1616 - \frac{(96)^2}{15}}{15-1}} = \sqrt{\frac{1001.6}{14}} = 8.458$$

y

$$\bar{d} = \frac{\sum d}{n} = \frac{96}{15} = 6.40$$

Por la ecuación 5.1, la t obtenida es, entonces,

$$t = \frac{6.40}{\frac{8.458}{\sqrt{15}}} = \frac{6.40}{2.184} = 2.930.$$

El Apéndice B muestra que el valor crítico para una prueba de una cola con 14 grados de libertad, realizada con $\alpha = .05$, es 1.761, de tal suerte que la hipótesis nula se rechaza. Por lo tanto, podemos concluir que sí hubo un cambio y que lo más probable es que la diferencia observada entre las medias no se deba al azar. ■

¿Este hallazgo significa que el medicamento *causó* la elevación de la presión arterial? Es probable, pero se deben considerar dos factores. Primero, siempre existe la posibilidad de un error tipo I. Segundo, y tal vez más importante, debemos tener en mente que la prueba t nos dice que ocurrió un cambio en las presiones arteriales pero no nos dice *por qué*. Suponga, por ejemplo, que las personas se pusieron ansiosas con el experimento y sufrieron una ligera elevación de la presión arterial como resultado de tal ansiedad. La prueba t no puede diferenciar entre el fármaco, la ansiedad o cualquier otra fuente que sea un agente causante; simplemente confirma que el cambio no fue producto del azar. La discusión sobre causa y efecto tendrá más importancia conforme usted avance en el libro.

TABLA 5.2: Puntuaciones LogMar asociadas con dos tratamientos láser.

Tratamiento uno	Tratamiento dos
.0	.2
.8	1.1
.4	.9
1.0	.5
.5	.2
.4	.7
.5	.5

EJEMPLO 5.2

Los investigadores están interesados en comparar la efectividad de dos tratamientos de cirugía láser para la retinopatía diabética. A los pacientes que manifiestan la enfermedad en ambos ojos se les trata uno de ellos con el primer método quirúrgico y el otro con el segundo método. Después de cierto tiempo se mide la agudeza visual de cada ojo. La pregunta que nos ocupa es: "¿Uno de los métodos de tratamiento produce mejor visión (medida por medio de la agudeza) que el otro?"

La determinación de cuál ojo será tratado con qué método se hace al azar. Es importante la asignación aleatoria de los ojos a los tratamientos para evitar sesgos. Por ejemplo, si esta asignación se le dejara al cirujano, podría preferir usar el primer tratamiento para ojos con afectaciones más severas. Una comparación posterior de las agudezas visuales mostraría que el primer método de tratamiento produjo una visión más deficiente que el segundo. Pero ésta no sería una comparación justa, ya que el método se habría usado en ojos con afectaciones más graves. Las puntuaciones de agudeza visual, expresadas en unidades LogMar,[2] obtenidas después de los tratamientos se presentan en la tabla 5.2.

Utilice los datos provistos para llevar a cabo una prueba t de dos colas de muestras apareadas con $\alpha = .05$ para ayudar a determinar si uno de los tratamientos fue más efectivo que el otro.

Solución En la tabla 5.3 se presentan las puntuaciones LogMar, sus diferencias, los cuadrados de sus diferencias y la suma de las diferencias de las puntuaciones y sus cuadrados.

Utilizando estas sumatorias tenemos que,

$$s_d = \sqrt{\frac{.81 - \frac{(-.5)^2}{7}}{7-1}} = \sqrt{\frac{.774}{6}} = \sqrt{.129} = .359$$

y

$$\bar{d} = \frac{-.5}{7} = -.071.$$

Al aplicar la ecuación 5.1, la t obtenida es

$$t = \frac{\bar{d} - \mu_{d0}}{\frac{s_d}{\sqrt{n}}} = \frac{-.071}{\frac{.359}{\sqrt{7}}} = \frac{-.071}{.136} = -.522.$$

[2] Para una explicación de las puntuaciones LogMar consulte el tema de las pruebas de equivalencia de una cola en la página 123.

TABLA 5.3: Las puntuaciones LogMar, sus diferencias, los cuadrados de sus diferencias y la suma de las diferencias de las puntuaciones y sus cuadrados.

	Tratamiento uno	Tratamiento dos	d	d^2
	.0	.2	−.2	.04
	.8	1.1	−.3	.09
	.4	.9	−.5	.25
	1.0	.5	.5	.25
	.5	.2	.3	.09
	.4	.7	−.3	.09
	.5	.5	.0	.00
Σ	3.6	4.1	−.5	.81

Como los valores críticos de una prueba t de dos colas con seis grados de libertad realizada al nivel .05 son +2.447 y −2.447, la hipótesis nula no se rechaza. Pero, ¿esto qué implica en lo que se refiere a la pregunta planteada por el investigador? ¿Significa que no hay ninguna diferencia en la efectividad de los dos tratamientos? Si usted respondió "sí" a esta última pregunta, haría bien en repasar el tema sobre el error tipo II del capítulo 4. El punto es que usted no conoce beta (la probabilidad de un error tipo II), de manera que no conoce la probabilidad de equivocarse al afirmar que la hipótesis nula es verdadera. Por lo tanto, no debe utilizar la imposibilidad de rechazar la hipótesis nula como evidencia de que ésta (la cual sostiene que no hay diferencias en las agudezas visuales promedio) es verdadera. Una antigua máxima dice que "la ausencia de evidencia no es evidencia de ausencia". Una interpretación más realista de este resultado es que "no hubo evidencia suficiente para concluir que los tratamientos difirieron en su eficacia". ■

Una pregunta más mundana que atormenta a los estudiantes es cuál de los dos grupos de puntuaciones debe ser restado del otro para obtener los valores d. En el ejemplo del remedio contra el resfriado restamos las puntuaciones pre-tratamiento de las puntuaciones post-tratamiento. El resultado fue $d = 6.40$ con una t de 2.930. El hecho de que d fuera positiva indica que la presión arterial post-tratamiento promedio fue más alta que la presión arterial pre-tratamiento promedio. Como la prueba fue significativa, rechazamos la hipótesis nula en favor de la alternativa que sostenía que la media de la presión arterial post-tratamiento es más alta que la media de la presión arterial pre-tratamiento.

Pero, ¿cuál hubiera sido el resultado si hubiéramos restado las puntuaciones post-tratamiento de las puntuaciones pre-tratamiento? En este caso, d hubiera sido −6.40 y la t obtenida sería de −2.930, lo que nos llevaría a concluir que la media de la presión arterial pre-tratamiento fue bastante más baja que la media de la presión arterial post-tratamiento. En realidad no importa si decimos que la posterior es considerablemente más alta que la previa, o que la previa es mucho más baja que la posterior. La conclusión es la misma. No obstante, observe que de haber seguido el curso de la segunda resta, se hubiera tenido que colocar la región crítica en el extremo izquierdo de la distribución, lo que implica un valor crítico negativo. En resumidas cuentas, siempre y cuando usted establezca e interprete el resultado correctamente, no importa en qué dirección se lleve a cabo la resta.

5.2.2 Establecimiento de la equivalencia a través de pruebas *t* de muestras apareadas

Razonamiento. Como se ha señalado en varias partes de este libro, el rechazo de una hipótesis nula provee buena evidencia (aunque no concluyente) de que la hipótesis nula es falsa. En contraste, no rechazar una hipótesis nula por lo general no provee buena evidencia de que la hipótesis nula es verdadera. En situaciones en las que se desea establecer la validez de la hipótesis nula debe emplearse una prueba de equivalencia para demostrar que la hipótesis nula es (aproximadamente) verdadera. (Véase el apartado 4.3.6 de la página 117.)

A veces sucede que los investigadores quisieran establecer que no hay *ninguna* diferencia entre tratamientos o entre las medias antes y después del tratamiento, en vez de demostrar que *sí* hay una diferencia. En tales casos, es posible utilizar la prueba *t* de muestras apareadas para establecer la equivalencia.

Como ejemplo, suponga que un fabricante busca obtener la aprobación de la FDA para una modificación que se ha hecho en un aparato automatizado de monitoreo de presión arterial. La modificación no mejorará el desempeño del aparato, pero sí permite una manufactura más fácil y, por lo tanto, más barata. Antes de aprobar la modificación, la FDA podría exigir que se lleven a cabo una serie de estudios para establecer que el aparato modificado trabaja tan bien como uno no modificado. Uno de estos estudios podría servir para demostrar que ambos aparatos, cuando se usan con el mismo grupo de pacientes, producen valores medios equivalentes de presión arterial. Después de las consultas apropiadas podría decidirse que las medias de las presiones arteriales registradas por ambos aparatos se considerarán equivalentes si las medias difieren por no más de cuatro milímetros de mercurio. La tabla 5.4 muestra presiones arteriales (ficticias) de 16 personas según las mediciones hechas por aparatos modificados y no modificados.

La pregunta pertinente es si puede considerarse que esta muestra de diferencias de puntuaciones proviene de una población de diferencias de puntuaciones cuya media está entre 4 y –4. Si éste fuera el caso, los dos aparatos pueden declararse equivalentes ya que la diferencia en las medias de las presiones arteriales no excede los cuatro milímetros de mercurio. Si la diferencia de la media de la población no está en este rango, los aparatos no pueden ser declarados equivalentes.

La prueba. Como la prueba *t* de muestras apareadas es una aplicación de la prueba *t* de una media de diferencias de puntuaciones, se puede llevar a cabo una prueba de equivalencia para las medias de datos apareados a través de los métodos relacionados con la prueba *t* de una media explicados en la sección 4.3.6 de la página 117.[3] Utilizando la notación de ese apartado, las hipótesis alternativa y nula para una prueba de equivalencia de dos colas basada en la prueba *t* de muestras apareadas se pueden enunciar como

$$H_{0E} : \mu_d \leq EI_L \text{ o } \mu_d \geq EI_U$$
$$H_{AE} : EI_L < \mu_d < EI_U$$

donde EI_L y EI_U son los extremos superior e inferior (*lower, upper*) del intervalo de equivalencia. En esencia, la hipótesis nula plantea que la diferencia entre medias no reside en el intervalo de equivalencia, mientras que la hipótesis alternativa afirma que la diferencia está en el intervalo. Usted recordará que la hipótesis nula se prueba al realizar dos pruebas de una cola al nivel α. Para rechazar la hipótesis nula de equivalencia usted debe demostrar que $\mu_d > EI_L$ *y* que $\mu_d < EI_U$ lo

[3] Quizá quiera repasar esta sección antes de continuar.

TABLA 5.4: Medidas de presión arterial obtenidas con aparatos de monitoreo modificados y no modificados.

	Modificados	No modificados	diferencia d	d^2
	98	99	1	1
	111	109	–2	4
	97	100	3	9
	132	133	1	1
	144	148	4	16
	100	100	0	0
	120	119	–1	1
	110	109	–1	1
	131	136	5	25
	154	153	–1	1
	105	107	2	4
	119	120	1	1
	107	107	0	0
	100	101	1	1
	118	116	–2	4
	122	127	5	25
$\sum$	1868	1884	16	94

que quiere decir que *las dos* pruebas siguientes deben ser significativas.

Prueba uno	Prueba dos
$H_{01} : \mu_d = EI_U$	$H_{02} : \mu_d = EI_L$
$H_{A1} : \mu_d < EI_U$	$H_{A2} : \mu_d > EI_L$

La hipótesis nula y la alternativa para la prueba de equivalencia de una cola son *una* de las siguientes

$$H_{0E} : \mu_d \geq EI_U$$
$$H_{A1} : \mu_d < EI_U$$

o

$$H_{0E} : \mu_d \leq EI_L$$
$$H_{AE} : \mu_d > EI_L$$

La primera hipótesis de equivalencia de una cola presentada arriba se pone a prueba por medio de la prueba uno y la segunda por medio de la prueba dos.

EJEMPLO 5.3

Haga una prueba de equivalencia de dos colas con los datos de la tabla 5.4, como se explicó. Determine las hipótesis nula de equivalencia y alternativa de equivalencia antes de realizar la prueba.

Solución Las hipótesis de equivalencia son

$$H_{0E}: \mu_d \leq -4 \text{ o } \mu_d \geq 4$$
$$H_{AE}: -4 < \mu_d < 4$$

Para probar la hipótesis nula de equivalencia se realizarán las siguientes pruebas t de muestras apareadas con $\alpha = .05$.

Prueba uno	Prueba dos
$H_{01}: \mu_d = 4$	$H_{02}: \mu_d = -4$
$H_{A1}: \mu_d < 4$	$H_{A2}: \mu_d > -4$

Utilizando las sumas de la tabla 5.4

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{94 - \frac{(16)^2}{16}}{16-1}} = \sqrt{\frac{78}{15}} = 2.280$$

y

$$\bar{d} = \frac{\sum d}{n} = \frac{16}{16} = 1.000.$$

Entonces, los valores de la t obtenida para las pruebas uno y dos son

$$t_1 = \frac{1.0 - 4.0}{\frac{2.28}{\sqrt{16}}} = -5.263 \text{ y } t_2 = \frac{1.000 - (-4.0)}{\frac{2.28}{\sqrt{16}}} = 8.772.$$

Los valores t críticos basados en 16 – 1 = 15 grados de libertad para las dos pruebas de una cola son −1.753 y 1.753. Se deduce que ambos resultados son significativos y llevan al rechazo de la hipótesis nula de que los dos aparatos no son equivalentes en lo que se refiere a la media de las presiones arteriales, lo que favorece a la alternativa que sostiene la equivalencia.

El panel A de la figura 5.1 describe la lógica que subyace en esta prueba. En sentido estricto, la equivalencia implica una diferencia media de cero, pero se ha decidido, con base en consideraciones prácticas, que una diferencia media (de población) entre cuatro y menos cuatro está lo suficientemente cercana a cero para ser declarada equivalente. La prueba uno muestra que la diferencia media es menor que cuatro. La prueba dos muestra que la diferencia es más grande que menos cuatro, con lo que se demuestra que la diferencia media se encuentra dentro del rango definido como equivalente.

¿Qué concluiría usted si la hipótesis nula de equivalencia no se hubiera rechazado? ¿Concluiría que los dos aparatos no producen medias equivalentes? ¡No! Usted sólo podría concluir que no pudo demostrar que son equivalentes. ¿Por qué? Porque usted no conoce beta. ■

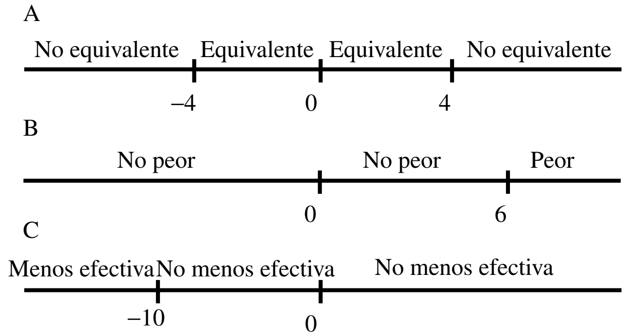


FIGURA 5.1: Lógica que subyace en las pruebas *t* de equivalencia de muestras apareadas.

EJEMPLO 5.4

Suponga que se realiza un estudio para determinar si se puede reducir la dosis de un fármaco para disminuir el colesterol y así evitar el daño hepático que a veces se observa con la dosis que se aplica actualmente. La pregunta que nos interesa aquí es: "¿Una dosis menor producirá niveles de colesterol menores que los obtenidos con la dosis más alta?" Para responder esta pregunta, 18 personas que han estado tomando la dosis más alta cambian a la dosis más baja. Se mide el colesterol mientras los pacientes están tomando la dosis más alta, y se vuelve a medir después de un tiempo especificado en el que han estado tomando la dosis más baja. En la tabla 5.5 se muestran las evaluaciones (ficticias) del colesterol de las 18 personas.

Los investigadores deciden que si el colesterol promedio se eleva menos de seis puntos, se puede declarar que la dosis reducida produce resultados que "no son peores" a los producidos por la dosis más alta. Utilice una prueba de equivalencia de una cola con $\alpha = .05$ para resolverlo.

Solución Como los investigadores quieren determinar si el aumento en el colesterol es menor a seis puntos, se puede utilizar una prueba de equivalencia de una cola con la forma

$$H_{0E} : \mu_d \geq 6$$
$$H_{AE} : \mu_d < 6$$

La prueba se lleva a cabo a través de una prueba *t* de muestras apareadas con las hipótesis

$$H_{01} : \mu_d = 6$$
$$H_{A1} : \mu_d < 6$$

Utilizando las sumas de la tabla anterior tenemos que,

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{730 - \frac{(40)^2}{18}}{18-1}} = \sqrt{\frac{641.111}{17}} = 6.141$$

y

$$\bar{d} = \frac{\sum d}{n} = \frac{40}{18} = 2.222.$$

TABLA 5.5: Evaluaciones de colesterol de 18 pacientes tratados con dosis completas y reducidas de un medicamento para disminuir el colesterol.

	Dosis alta	Dosis baja	diferencia d	d^2
	198	202	4	16
	180	178	–2	4
	165	175	10	100
	152	158	6	36
	211	214	3	9
	261	264	3	9
	140	140	0	0
	200	206	6	36
	188	180	–8	64
	154	152	–2	4
	204	206	2	4
	144	148	4	16
	155	160	5	25
	199	194	–5	25
	190	197	7	49
	160	164	4	16
	189	178	–11	121
	190	204	14	196
Σ	3280	3320	40	730

Por lo tanto,

$$t_1 = \frac{\bar{d} - \mu_{d0}}{\frac{s_d}{\sqrt{n}}} = \frac{2.222 - 6.0}{\frac{6.141}{\sqrt{18}}} = -2.610.$$

La t crítica para una prueba de una cola a un nivel $\alpha = .05$ con 17 grados de libertad es -1.740. La hipótesis nula se rechaza, lo que nos lleva a concluir que la dosis reducida no es peor que la dosis más alta en lo que respecta al control del colesterol.

El panel B de la figura 5.1 describe la lógica de esta prueba. Los investigadores han decidido que un incremento promedio de menos de seis unidades es lo bastante pequeño como para ser considerado "no peor". No hay ninguna posibilidad de obtener mejores resultados, por lo que no se necesita una prueba de dos colas. Una vez que la hipótesis nula se rechaza en favor de la alternativa que sostiene que la diferencia de medias es menor que seis, se sustenta la proposición de "no peor". No le preguntaremos cuál sería la interpretación correcta si la hipótesis nula no se hubiera rechazado, pues estamos seguros de que usted ya la conoce. (La conoce, ¿cierto?) ■

EJEMPLO 5.5

Un fabricante de la versión genérica de un medicamento utilizado para el tratamiento del SIDA desea demostrar que la forma genérica no es menos eficaz para incrementar el conteo de CD4 que la medicina de marca. Con tal fin se trata a 14 pares de sujetos apareados con las dos formas del fármaco. Cada par se iguala con base en la edad (con una diferencia máxima de dos años) y un conteo inicial de CD4 (en un rango de 10/mm^3). Al azar, se selecciona a un miembro de cada par y se le administra la forma genérica del medicamento, mientras que al otro miembro se le administra la fórmula patentada. Después de un tiempo de tratamiento se evalúa el conteo de CD4 de cada sujeto. Los conteos de CD4 (ficticios) de los 14 pares de sujetos se muestran en la tabla 5.6.

Después de consultar con un grupo de expertos, los investigadores deciden que es posible declarar al medicamento genérico equivalente al de marca si el conteo promedio de CD4 de los pacientes que toman la forma genérica no está más de 10/mm^3 por debajo del nivel promedio de los sujetos que toman el medicamento de marca. Utilice una prueba de equivalencia de una cola con $\alpha = .05$ para determinar si los dos fármacos pueden declararse equivalentes.

Solución Como los investigadores desean determinar si la disminución promedio de los niveles de CD4 atribuible al uso de la forma genérica es menor a 10/mm^3, se puede utilizar una prueba de equivalencia de una cola de la forma

$$H_{0E} : \mu_d \le -10$$
$$H_{AE} : \mu_d > -10$$

La prueba se lleva a cabo mediante una prueba t de muestras apareadas con las hipótesis

$$H_{02} : \mu_d = -10$$
$$H_{A2} : \mu_d > -10$$

Utilizando las sumas de la tabla anterior tenemos que,

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{2673 - \frac{(-85)^2}{14}}{14-1}} = \sqrt{\frac{2156.929}{13}} = 12.881$$

y

$$\bar{d} = \frac{\sum d}{n} = \frac{-85}{14} = -6.071.$$

Por lo tanto,

$$t_2 = \frac{\bar{d} - \mu_{d0}}{\frac{s_d}{\sqrt{n}}} = \frac{-6.071 - (-10.0)}{\frac{12.881}{\sqrt{14}}} = 1.141.$$

La t crítica de una prueba de una cola a un nivel $\alpha = .05$ con 13 grados de libertad es 1.771. Por lo tanto, la hipótesis nula no se rechaza. Esto significa que no se puede afirmar que el medicamento genérico *no* es menos eficaz.

El panel C de la figura 5.1 describe la lógica de esta prueba. Los investigadores han decidido que una disminución en la media de menos de 10 unidades es lo suficientemente pequeña para ser considerada “no menos eficaz”. Sin embargo, esto no fue demostrado por la prueba.

Como ejercicio final, considere las implicaciones para las anteriores pruebas de equivalencia si d se hubiera generado restando las dos variables en la dirección opuesta a la que realmente se empleó. ¿Hubiera cambiado alguna de las conclusiones? Trate de hacer su propio bosquejo de la figura 5.1 como aparecería si la resta se hubiera hecho en la dirección opuesta. ■

TABLA 5.6: Conteos de CD4 de pacientes tratados con el medicamento de marca y el genérico.

	De marca	Genérico	Diferencia d	d^2
	302	306	4	16
	400	363	–37	1369
	398	405	7	49
	225	225	0	0
	221	211	–10	100
	261	245	–16	256
	154	138	–16	256
	300	280	–20	400
	188	192	4	16
	154	145	–9	81
	204	206	2	4
	190	200	10	100
	210	211	1	1
	199	194	–5	25
Σ	3280	3321	–85	2673

5.2.3 Intervalo de confianza para la diferencia de medias de muestras apareadas

Razonamiento. La prueba *t* de muestras apareadas busca si existe una diferencia entre las medias de dos variables apareadas. Una pregunta relacionada y generalmente más informativa es: "¿Qué tan grande es la diferencia entre las medias de las variables apareadas?" Esta diferencia se puede calcular con un intervalo de confianza.

En el apartado 4.4.5 aprendió cómo calcular la media de una población cuando no se conoce σ. Este método se puede aplicar de forma directa al problema de calcular la media de una población de diferencias de puntuaciones.

Con anterioridad se explicó que en la situación en la que no existe ninguna diferencia entre las medias de variables apareadas, el modelo apropiado para la población de diferencias de puntuaciones tendría una media de cero. En la situación en la que existe un efecto del tratamiento, un cambio u otro mecanismo que cause una diferencia en las medias de las dos variables apareadas, el mejor modelo para la población de diferencias de puntuaciones es una población cuya media no es cero. Además, la magnitud de esta diferencia entre medias se puede calcular mejor formando el intervalo de confianza correcto sobre las diferencias de puntuaciones de la muestra. Si se crean estos intervalos, los investigadores pueden responder preguntas tales como "¿de qué magnitud fue el cambio promedio que ocurrió?", o "¿qué tan grande es la diferencia entre los resultados de la media de los dos grupos de tratamiento?". Así, en vez de preguntar: "¿Un medicamento crea una mayor respuesta promedio que otro?", como se haría con una prueba *t* de muestras apareadas, el investigador puede responder a la pregunta: "¿Qué tanto más grande es la respuesta promedio producida por un medicamento que la producida por el otro?", mediante un intervalo de confianza.

El intervalo de confianza. El método para construir intervalos de confianza para la media de una población de diferencias de puntuaciones es el mismo que el expresado en las ecuaciones 4.14 y 4.15 de la página 416. Aquí repetimos esas ecuaciones con subíndices para indicar que se están usando puntuaciones de diferencias en el cálculo.

Entonces, las ecuaciones para L y U son

$$\boxed{L = \bar{d} - t\frac{s_d}{\sqrt{n}}} \tag{5.2}$$

y

$$\boxed{U = \bar{d} + t\frac{s_d}{\sqrt{n}}} \tag{5.3}$$

donde s_d es la desviación estándar de la muestra de las diferencias de puntuaciones y t es el valor apropiado de la tabla t con $n - 1$ grados de libertad.

EJEMPLO 5.6

Utilice las diferencias de puntuaciones de la tabla 5.1 de la página 160 para construir un intervalo de confianza unilateral del 95% para el límite inferior de μ_d. Interprete el resultado.

Solución Como se calculó previamente, s_d y $\bar{d}$ son

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{1616 - \frac{(96)^2}{15}}{15-1}} = \sqrt{\frac{1001.60}{14}} = 8.458$$

y

$$\bar{d} = \frac{\sum d}{n} = \frac{96}{15} = 6.40$$

Al observar en el Apéndice B que el valor correcto de t (con 14 grados de libertad) para construir un intervalo de confianza unilateral del 95% es 1.761, la ecuación 5.2 produce

$$S = 6.40 - 1.761\frac{8.458}{\sqrt{15}} = 2.554$$

Una interpretación estadística de este intervalo sostendría que es posible tener una confianza del 95% de que μ_d es mayor que o igual a 2.554. Desde el punto de vista del investigador, con un 95% de confianza se puede afirmar que el incremento promedio de la presión arterial medida antes y después de tomar el remedio contra el resfriado fue de *por lo menos* 2.554 milímetros de mercurio. ■

EJEMPLO 5.7

Los datos de la tabla 5.1 se utilizaron previamente para realizar una prueba t de una cola de muestras apareadas de la forma

$$H_0 : \mu_d = 0$$
$$H_A : \mu_d > 0$$

con $\alpha = .05$. El resultado fue estadísticamente significativo. ¿El intervalo de confianza que usted calculó sustenta este resultado? ¿Por qué?

Solución Del intervalo de confianza es posible ver que la hipótesis nula fue rechazada. Esto se deduce del hecho de que L es mayor que cero.[4] ■

EJEMPLO 5.8

Utilice las diferencias de puntuaciones de la tabla 5.3 de la página 164 para construir un intervalo de confianza bilateral del 95%. ¿Qué significa este intervalo?

Solución Como se calculó previamente,

$$s_d = \sqrt{\frac{.81 - \frac{(-.5)^2}{7}}{7-1}} = \sqrt{\frac{.774}{6}} = .359$$

y

$$\bar{d} = \frac{-.5}{7} = -.071.$$

Al observar en el Apéndice B que el valor correcto de t (con seis grados de libertad) para un intervalo de confianza bilateral del 95% es 2.447, las ecuaciones 5.2 y 5.3 producen

$$S = \bar{d} - t\frac{s_d}{\sqrt{n}} = -.071 - 2.447\frac{.359}{\sqrt{7}} = -.403$$

y

$$I = \bar{d} + t\frac{s_d}{\sqrt{n}} = -.071 + 2.447\frac{.359}{\sqrt{7}} = .261.$$

Una interpretación estadística de este intervalo sostendría que se puede tener una confianza del 95% en que μ_d está entre $-.403$ y $.261$. Desde el punto de vista del investigador es posible aseverar con una confianza del 95% que la diferencia promedio en la agudeza visual entre los pacientes que se sometieron al tratamiento láser uno y los que recibieron el tratamiento láser dos estuvo entre $-.403$ y $.261$ unidades LogMar. Con base en este intervalo, ¿puede usted descartar una diferencia de cero? ■

EJEMPLO 5.9

Los datos de la tabla 5.2 se usaron con anterioridad para realizar una prueba t de dos colas de muestras apareadas de la forma

$$H_0 : \mu_d = 0$$
$$H_A : \mu_d \neq 0$$

con $\alpha = .05$. El resultado no fue estadísticamente significativo. ¿Sustenta este resultado el intervalo de confianza que usted acaba de calcular? ¿Por qué?

Solución Del intervalo de confianza se puede ver que la hipótesis nula no fue rechazada. Esto se deduce del hecho de que cero se localiza en el intervalo $-.403$ a $.261$.[5] ■

---

[4] Si esta explicación no ha quedado clara, debe repasar el apartado 4.5 de la página 152.

[5] De nuevo, debe repasar el apartado 4.5 de la página 152 si la explicación no ha quedado clara.

5.2.4 Suposiciones

Las suposiciones que subyacen en la prueba t de muestras apareadas, en las pruebas de equivalencia basadas en la prueba t de muestras apareadas, y en el intervalo de confianza para la diferencia de medias son las mismas que las de la prueba t de una media y, por consiguiente, de la prueba Z de una media (véase el apartado 4.3.3). Éstas son las suposiciones de normalidad e independencia. Debe tomar en cuenta que estas suposiciones se aplican a las diferencias de puntuaciones en lugar de las distribuciones de dos componentes (por ejemplo, distribuciones pre y post-tratamiento) usadas para generar las diferencias de puntuaciones.

Aun cuando las distribuciones de dos componentes son radicalmente no normales, a menudo las diferencias de puntuaciones son simétricas y con una forma hasta cierto punto normal. Como resultado, los métodos estadísticos mencionados tienden a ser bastante sólidos ante datos no normales. Como sucede con la prueba t de una media, no se puede confiar en que estos métodos sean sólidos si se viola la suposición de independencia.

5.3 MÉTODOS RELACIONADOS CON LAS PROPORCIONES

5.3.1 Prueba de McNemar de una proporción de muestras apareadas

Razonamiento. Volvamos a un ejemplo mencionado al principio de este capítulo. Se probarán dos vacunas en un intento por determinar si una es superior a la otra para prevenir cierta enfermedad infantil. Como se sospecha que puede haber una predisposición genética a esta enfermedad, el diseño del estudio requiere que se utilicen mellizos en el experimento. Se selecciona al azar a un mellizo de cada par para que reciba la primera vacuna; el otro mellizo es inoculado con la segunda vacuna. Al usar mellizos se garantiza que las predisposiciones genéticas estén distribuidas equitativamente entre las dos vacunas, de ese modo se previenen los sesgos que pueden resultar si tales predisposiciones no se hubieran representado por igual en los dos grupos. Después de un periodo determinado se toma nota de cuáles niños contrajeron la enfermedad. El interés reside en determinar si una vacuna fue más efectiva que la otra para prevenir la enfermedad.

La tabla 5.7 muestra los resultados de este estudio (ficticio). Se designa a cada miembro de los 16 pares de mellizos como D o $\overline{D}$ para indicar, respectivamente, que el mellizo contrajo la enfermedad o no la contrajo. La última columna de la tabla muestra cuál vacuna fue aprobada (desaprobada) por los resultados obtenidos de cada par de mellizos. Se otorga un uno cuando el mellizo que recibió la vacuna uno contrajo la enfermedad y el otro no. Un cero indica que el mellizo que recibió la vacuna dos contrajo la enfermedad y el otro no. Así, si la vacuna uno es más efectiva que la vacuna dos se esperaría ver muchos ceros y pocos unos, lo que mostraría más casos de enfermedad en el grupo de la vacuna dos. Muchos unos y pocos ceros indicarían más casos de enfermedad en el grupo de la vacuna uno. Observe que cuando se obtiene el mismo resultado en ambos mellizos, no se registra ningún valor, ya que no se ha obtenido de ese par, ninguna información de superioridad de una vacuna sobre la otra. Los datos que no proveen ninguna información en cuanto a la pregunta que nos ocupa (por ejemplo, cuál vacuna es superior) se denominan **no informativos**.

Si las dos vacunas tienen la misma capacidad para prevenir la enfermedad, se esperaría ver aproximadamente el mismo número de resultados que favorecieran a una vacuna y a la otra. Es decir, se esperaría ver aproximadamente el mismo número de unos y de ceros en la columna de resultados. Esta observación provee la base para una prueba de significancia.

TABLA 5.7: Presencia o ausencia de una enfermedad infantil en mellizos inoculados con dos vacunas diferentes.

Par de mellizos	Vacuna uno	Vacuna dos	Resultado
1	$\overline{D}$	D	1
2	D	D	–
3	$\overline{D}$	$\overline{D}$	–
4	D	$\overline{D}$	1
5	D	$\overline{D}$	1
6	$\overline{D}$	$\overline{D}$	–
7	D	$\overline{D}$	1
8	D	D	–
9	$\overline{D}$	$\overline{D}$	–
10	$\overline{D}$	D	0
11	D	$\overline{D}$	1
12	$\overline{D}$	$\overline{D}$	–
13	D	$\overline{D}$	1
14	$\overline{D}$	$\overline{D}$	–
15	$\overline{D}$	$\overline{D}$	–
16	D	$\overline{D}$	1
			$\sum = 6$

La prueba. Si se omiten los datos no informativos, los ocho resultados de la tabla 5.7 podrían considerarse como una muestra de una población dicotómica como a la que nos referimos en el apartado 4.2.4. Además, si no hay diferencia en la efectividad de las dos vacunas, la población estaría formada por un número igual de unos y ceros, lo que significaría que la proporción de éxitos en la población (π) sería .5. En contraste, si una vacuna fuera más efectiva que la otra, el modelo de población correcto favorecería a una vacuna sobre la otra, lo que significa que tendría más unos que ceros o más ceros que unos. Esto implica que π sería mayor que .5 o menor que .5 dependiendo de cuál vacuna fue favorecida. Se deduce que una prueba de la hipótesis nula

$$H_0 : \pi = .5$$

contra una alternativa apropiada es una prueba de la aseveración de que las vacunas tienen el mismo efecto. El rechazo de esta hipótesis lleva a la conclusión de que las vacunas difieren en su efectividad.

En el apartado 4.3.5 usted aprendió a poner a prueba hipótesis relacionadas con la proporción π de población. La prueba de McNemar es sólo una aplicación a un resultado dicotómico de los métodos que usted aprendió en el apartado 4.3.5, como se ilustra en la tabla 5.7. La hipótesis nula se pone a prueba como señalamos anteriormente, es decir, $\pi = .5$. Como usted sabe, ambos métodos para probar una proporción de población, el aproximado y el exacto, son posibles. Ilustraremos ambos en su relación con la prueba de McNemar.

TABLA 5.8: Tabla dos por dos para datos dicotómicos apareados.

		Primera variable	
		+	−
Segunda variable	+	*a*	*b*
	−	*c*	*d*

Método aproximado

Por medio de la ecuación 4.9 se puede llevar a cabo una prueba basada en la aproximación de la curva normal para la distribución muestral de $\hat{p}$. Como la prueba de McNemar especifica el valor nulo como .5, se puede modificar la ecuación 4.9 para producir lo siguiente

$$\boxed{Z = \frac{\hat{p} - .5}{\frac{.5}{\sqrt{n}}}} \tag{5.4}$$

Aunque las observaciones en la tabla 5.7 son muy pocas[6] para permitir un análisis apropiado a través de la ecuación 5.4, de todas formas utilizaremos estos datos con propósitos ilustrativos. Si tenemos en cuenta que $\hat{p} = \frac{6}{8} = .75$ y $n = 8$, la ecuación 5.4 produce

$$Z = \frac{.75 - .50}{\frac{.50}{\sqrt{8}}} = 1.414$$

Si realizamos una prueba de dos colas con $\alpha = .05$, el Apéndice A muestra que los valores críticos son +1.96 y −1.96, por lo que la hipótesis nula no se rechaza.

Los epidemiólogos y otras personas que generalmente utilizan tablas de contingencia prefieren una forma de cálculo diferente que se da por

$$\boxed{\chi^2 = \frac{(b-c)^2}{b+c}} \tag{5.5}$$

donde χ^2 representa el estadístico chi-cuadrada (que usted estudiará más adelante), y b y c son recuentos de las celdas de una tabla de frecuencias, como se muestra en la tabla 5.8.

Esta tabla contiene las frecuencias de dos variables dicotómicas. Los símbolos + y – se usan para indicar las dos condiciones de las variables dicotómicas. Para el presente ejemplo las frecuencias de la tabla 5.7 serían

[6] Véase el apartado 4.2.6 de la página 84 sobre los requisitos del tamaño de la muestra para este estadístico.

		Vacuna uno	
		D	$\overline{D}$
Vacuna dos	D	2	2
	$\overline{D}$	6	6

Aplicando la ecuación 5.5 obtenemos

$$\chi^2 = \frac{(2-6)^2}{8} = 2.00$$

El valor crítico para esta prueba se encuentra al revisar el Apéndice D con un grado de libertad. Este valor es 3.841 para $\alpha = .05$. Como el valor χ^2 obtenido debe ser mayor que o igual al χ^2 crítico para poder rechazar la hipótesis nula, el resultado no es significativo.

De manera interesante, las ecuaciones 5.4 y 5.5 representan dos formas de la misma prueba, por lo que se obtiene el mismo resultado sin importar cuál ecuación se aplique. Para estas pruebas, $Z^2 = \chi^2$, por lo que para este ejemplo $1.414^2 = 2.00$ (redondeado) para los dos valores obtenidos y $1.96^2 = 3.841$ (de nuevo, redondeado) para los valores críticos. Como se mencionó antes, los investigadores que generalmente usan tablas de contingencia prefieren la ecuación 5.5, pero la ecuación 5.4 puede ser más intuitiva.

EJEMPLO 5.10

Se deben comparar dos formas de filtro solar, con diferentes ingredientes activos, con respecto del nivel de protección que ofrecen contra el daño causado por el sol en la piel. Para este fin, se aplica una forma del filtro solar en uno de los brazos, elegido al azar, de cada uno de 15 sujetos voluntarios, y el otro filtro solar en el otro brazo. Después de un periodo especificado de exposición al sol se examina al microscopio la piel de cada brazo y se califica de acuerdo a si el nivel de protección es satisfactorio, S, o no satisfactorio, $\overline{S}$. Los resultados (ficticios) de cada sujeto se presentan en la tabla 5.9.

Utilice los dos métodos de cálculo anteriores para realizar una prueba de McNemar de dos colas con $\alpha = .05$ para los datos de los filtros solares. ¿Cuál es su conclusión en cuanto a la eficacia relativa de los dos filtros solares?

Solución Se construye un resultado dicotómico, como se indica en la tabla 5.10. Si se ignoran los resultados no informativos, se observa que la proporción de resultados que favorecen al producto uno es $\frac{9}{10} = .90$ y $n = 10$. Luego, con la ecuación 5.4 obtenemos

$$Z = \frac{.90 - .50}{\frac{.50}{\sqrt{10}}} = 2.530$$

El Apéndice A muestra que los valores críticos para una prueba de dos colas realizada con $\alpha = .05$ son $+1.96$ y -1.96, lo que produce el rechazo de la hipótesis nula. Desde un punto de vista estadístico, la hipótesis nula y la alternativa son

$$H_0 : \pi = .5 \text{ y } H_A : \pi \neq .5$$

Desde el contexto de este estudio, la hipótesis nula asevera que no existe ninguna diferencia en la eficacia de los dos productos, mientras que la hipótesis alternativa sostiene que tal diferen-

TABLA 5.9: Niveles satisfactorios o insatisfactorios de protección contra el daño en la piel causado por el sol, de dos productos de filtro solar.

Sujeto	Producto uno	Producto dos
1	$\bar{S}$	S
2	S	S
3	S	$\bar{S}$
4	S	$\bar{S}$
5	S	$\bar{S}$
6	S	$\bar{S}$
7	$\bar{S}$	$\bar{S}$
8	$\bar{S}$	$\bar{S}$
9	S	$\bar{S}$
10	S	$\bar{S}$
11	S	S
12	$\bar{S}$	$\bar{S}$
13	S	$\bar{S}$
14	S	$\bar{S}$
15	S	$\bar{S}$

TABLA 5.10: Niveles satisfactorios o insatisfactorios de protección contra el daño en la piel causado por el sol con variable resultante.

Sujeto	Producto uno	Producto dos	Resultado
1	$\bar{S}$	S	0
2	S	S	–
3	S	$\bar{S}$	1
4	S	$\bar{S}$	1
5	S	$\bar{S}$	1
6	S	$\bar{S}$	1
7	$\bar{S}$	$\bar{S}$	–
8	$\bar{S}$	S	–
9	S	$\bar{S}$	1
10	S	$\bar{S}$	1
11	S	S	–
12	$\bar{S}$	$\bar{S}$	–
13	S	$\bar{S}$	1
14	S	$\bar{S}$	1
15	S	$\bar{S}$	1
			$\sum = 9$

cia sí existe. En este caso, se rechaza la afirmación de que no hay diferencia, en favor de la afirmación de que los dos productos sí difieren en eficacia. Además, es posible concluir que el producto uno es superior al producto dos en términos de la protección que ofrece.

Los datos pueden ordenarse de la siguiente manera para su análisis a través del segundo método de cálculo.

		Producto uno S	Producto uno $\bar{S}$
Producto dos	S	2	1
	$\bar{S}$	9	3

Al aplicar la ecuación 5.5, se obtiene

$$\chi^2 = \frac{(1-9)^2}{10} = 6.40$$

Como 6.40 excede el valor crítico de 3.841 obtenido del Apéndice D, la hipótesis nula se rechaza. La inspección de la tabla de resumen indica que hay nueve casos en los que el producto uno previene el daño, en tanto que el producto dos no, y un caso en el que el producto dos brinda protección y el producto uno no. La conclusión es que el producto uno ofrece una protección superior a la que brinda el producto dos. ■

Método exacto

Antes (en la página 108) usted aprendió a utilizar la distribución binomial para llevar a cabo pruebas exactas de hipótesis con la forma

$$H_0 : \pi = \pi_0$$

donde π es la proporción de observaciones en una población que cumple con algún criterio especificado, y π_0 es la proporción planteada como hipótesis para la misma población. Como la prueba de McNemar puede reducirse a una prueba de la hipótesis

$$H_0 : \pi = .5$$

se sigue que el método exacto que usted aprendió previamente (véase la página 108) puede utilizarse para llevar a cabo una versión exacta de la prueba de McNemar. Como en ese apartado se expuso el método exacto a fondo, no lo revisaremos aquí.[7] Mejor aplicaremos el método exacto a los dos ejemplos previamente analizados por medio del método aproximado.

EJEMPLO 5.11

Utilice los datos de la tabla 5.7 para llevar a cabo una prueba de McNemar exacta de dos colas con $\alpha = .05$. ¿En qué difiere este resultado con el que se obtuvo de la prueba aproximada?

Solución La distribución muestral de $\hat{p}$ en la condición $\pi = .5$ y $n = 8$ se puede construir mediante la ecuación 4.5 y se presenta en la tabla 5.11. Tomando en cuenta que $p = \frac{6}{8} = .75$ y empleando el método ya aprendido (página 108) para calcular el valor-p para una prueba exacta de dos colas, este valor se calcula como 2 (.10937 + .03125 + .00391) = .28906. Como este valor

[7] Quizá quiera repasar la explicación que inicia en la página 108 antes de continuar.

es mayor que $\alpha = .05$, no se rechaza la hipótesis nula. Ésta es la misma decisión a la que se llega con la prueba aproximada. Sin embargo, debe observar el hecho de que el valor Z obtenido de 1.41, calculado en conexión con la prueba aproximada, tiene un valor-p asociado de $p = 2 \times .0793 = .1586$, que es considerablemente diferente al valor obtenido con la prueba exacta.

EJEMPLO 5.12

Utilice los datos de la tabla 5.9 para llevar a cabo una prueba McNemar exacta con hipótesis alternativa

$$H_A : \pi > .5$$

Solución La distribución muestral de $\hat{p}$ en la condición $\pi = .5$ y $n = 10$ se puede construir mediante la ecuación 4.5 y se presenta en la tabla 5.12. Al observar que $\hat{p} = \frac{9}{10} = .90$ y emplear el método explicado en la página 108 para calcular el valor-p para una prueba exacta de cola con hipótesis alternativa de la forma $H_A : \pi > \pi_0$, este valor se calcula como $.00977 + .00098 = .01075$. Como este valor es menor que $\alpha = .05$, la hipótesis nula se rechaza. ■

5.3.2 Establecimiento de la equivalencia para una proporción de muestras apareadas

Razonamiento. Como vimos en el apartado 5.2.2, a veces se realizan estudios con el propósito de demostrar que dos tratamientos o métodos producen resultados equivalentes (similares). Cuando los datos son apareados y el resultado es dicotómico, se puede demostrar la equivalencia mostrando que la proporción de resultados que favorecen un tratamiento se localiza en algún intervalo de equivalencia establecido alrededor de .5. Por ejemplo, suponga que, desde siempre, los pacientes que padecen cierto trastorno mental han sido hospitalizados para su tratamiento. Ahora se cree que este trastorno podría tratarse con igual éxito en la consulta externa y con un costo bastante más bajo. Para poner a prueba esta hipótesis se forman pares de pacientes con base en la edad, el género y la gravedad del trastorno. Al azar, se selecciona a un miembro de cada par para internarlo y recibir tratamiento, y el otro miembro es tratado en la consulta externa. Nuestro interés es demostrar que los dos métodos de tratamiento producen resultados equivalentes (similares) en vez de demostrar que un tratamiento es superior al otro. Después de un análisis de costo-beneficio, y otras consideraciones, se decide que las dos opciones de tratamiento pueden considerarse funcionalmente equivalentes si la proporción de resultados apareados que favorecen el proceso ambulatorio es mayor que .40. (Tome en cuenta que una proporción de .5 indicaría una equivalencia estricta). La tabla 5.13 muestra los resultados (ficticios) de 18 pacientes tratados con los dos métodos. Los símbolos S y $\overline{S}$ se usan para indicar los tratamientos, satisfactorio e insatisfactorio, respectivamente.

La pregunta pertinente es si π, la proporción de los resultados apareados que favorecen al tratamiento en la consulta externa, es mayor que .40. Si éste es el caso, los tratamientos serán declarados equivalentes. Si no, los tratamientos no se considerarán equivalentes. Note que los investigadores descartan la posibilidad de que el tratamiento en la consulta externa sea superior al tratamiento de hospital y, por tanto, están interesados en la equivalencia unilateral.

La prueba. Como los intervalos de equivalencia considerados son para proporciones de población, los métodos explicados en el apartado 4.3.6 de la página 117 para la equivalencia de proporción son aplicables aquí.[8] Utilizando la notación del apartado 4.3.6, la hipótesis nula y

[8] Quizá quiera repasar esta sección antes de continuar.

TABLA 5.11: Distribución muestral de $\hat{p}$ para $n = 8$ y $\pi = .50$.

Proporción $\hat{p}$	Número de éxitos y	Probabilidad $P(y)$
.000	0	.00391
.125	1	.03125
.250	2	.10937
.375	3	.21875
.500	4	.27344
.625	5	.21875
.750	6	.10937
.875	7	.03125
1.00	8	.00391

TABLA 5.12: Distribución muestral de $\hat{p}$ para $n = 10$ y $\pi = .50$.

Proporción $\hat{p}$	Número de éxitos y	Probabilidad $P(y)$
.000	0	.00098
.100	1	.00977
.200	2	.04395
.300	3	.11719
.400	4	.20508
.500	5	.24609
.600	6	.20508
.700	7	.11719
.800	8	.04395
.900	9	.00977
1.00	10	.00098

TABLA 5.13: Resultados satisfactorios o insatisfactorios para pares de pacientes que padecen cierto trastorno mental cuando se les da tratamiento hospitalario o en consulta externa.

Par	Paciente hospitalizado	Paciente en consulta externa
1	$\bar{S}$	S
2	$\bar{S}$	$\bar{S}$
3	$\bar{S}$	S
4	$\bar{S}$	S
5	S	$\bar{S}$
6	S	$\bar{S}$
7	S	S
8	S	$\bar{S}$
9	$\bar{S}$	S
10	S	$\bar{S}$
11	S	S
12	S	S
13	$\bar{S}$	S
14	S	$\bar{S}$
15	S	$\bar{S}$
16	S	$\bar{S}$
17	S	$\bar{S}$
18	$\bar{S}$	S

la alternativa para una prueba de equivalencia de dos colas para proporciones apareadas puede enunciarse como

$$H_{0E}: \pi \leq EI_L \quad \text{o} \quad \pi \geq EI_U$$
$$H_{AE}: EI_L < \pi < EI_U$$

donde EI_L y EI_U son, respectivamente, los extremos inferior y superior del intervalo de equivalencia. En esencia, la hipótesis nula plantea que la proporción de resultados apareados que favorece una condición sobre la otra no reside en el intervalo de equivalencia, en tanto que la hipótesis alternativa asevera que esta proporción se localiza en el intervalo. Recordará que la hipótesis nula se pone a prueba realizando dos pruebas de una cola al nivel α. Para rechazar la hipótesis nula usted debe demostrar que $\pi > EI_L$ y que $\pi < EI_U$, lo que significa que las dos pruebas siguientes deben ser significativas.

Prueba uno	Prueba dos
$H_{01}: \pi = EI_U$	$H_{02}: \pi = EI_L$
$H_{A1}: \pi < EI_U$	$H_{A2}: \pi > EI_L$

La hipótesis nula y la alternativa para la prueba de equivalencia de una cola son *una* de las siguientes

$$H_{0E}: \pi \geq EI_U$$
$$H_{AE}: \pi < EI_U$$

o

$$H_{0E}: \pi \leq EI_L$$
$$H_{AE}: \pi > E_{IL}$$

La primera hipótesis de equivalencia de una cola planteada se pone a prueba con la prueba uno y la segunda mediante la prueba dos.

Como usted sabe, por lo estudiado en el apartado 4.3.5, se pueden realizar pruebas de la forma representada por las pruebas uno y dos, ya sea por medios aproximados o exactos. Ambos métodos se ilustrarán en los siguientes ejemplos.

EJEMPLO 5.13

Utilice los métodos aproximado y exacto para llevar a cabo una prueba de equivalencia de una cola con $\alpha = .05$ con respecto a los datos de la tabla 5.13. Plantee las hipótesis nula y alternativa de equivalencia antes de realizar las pruebas.

Solución Una variable resultante que utiliza un uno para indicar un par de pacientes cuyo resultado favorece el tratamiento de consulta externa y un cero para un resultado que favorece el tratamiento hospitalario se construye como se muestra en la tabla 5.14. Si se ignoran los resultados no informativos, de esta tabla se puede ver que la proporción de resultados que favorecen el tratamiento de consulta externa es $\frac{6}{14} = .429$.

Las hipótesis de equivalencia son

$H_{0E}: \pi \leq .40$

$$H_{AE}: \pi > .40$$

La hipótesis nula de equivalencia se puede poner a prueba por medio de la prueba dos

$$H_{02}: \pi = .40$$
$$H_{A2}: \pi > .40$$

La Z obtenida por el método aproximado, usando la ecuación 4.9, es

$$Z_2 = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{.429 - .40}{\sqrt{\frac{(.429)(.571)}{14}}} = .22$$

Como este valor es menor que la Z crítica de 1.65, la hipótesis nula no se rechaza. Se deduce que la equivalencia no está demostrada.

La prueba exacta se lleva a cabo generando la distribución muestral de $\hat{p}$ en la condición $\pi = .40$ y $n = 14$. Esto se logra aplicando la ecuación 4.5 y el resultado se muestra en la tabla 5.15.[9]

Por medio del método descrito en la página 108 para calcular el valor-p para una prueba exacta de una cola, $p = .20660 + .15741 + .09182 + .04081 + .01360 + .00330 + .00055 + .00006 + .00000 = .51415$. Dado que este valor es mayor que $\alpha = .05$, la hipótesis nula no se rechaza, que es el mismo resultado que se obtuvo con la prueba aproximada. La columna 2 del Apéndice A muestra que el área de la cola asociada con $Z = .22$ es .4129, por lo que el valor-p para la prueba aproximada es definitivamente más pequeño que para la prueba exacta. ■

[9] ¿Es necesario generar todos los valores en esta tabla para realizar la prueba?

TABLA 5.14: Resultados de los pares de pacientes que padecen cierto trastorno mental.

Par	Paciente hospitalizado	Paciente en consulta externa	Resultado
1	$\bar{S}$	S	1
2	$\bar{S}$	$\bar{S}$	–
3	$\bar{S}$	S	1
4	$\bar{S}$	S	1
5	S	$\bar{S}$	0
6	S	$\bar{S}$	0
7	S	S	–
8	S	$\bar{S}$	0
9	$\bar{S}$	S	1
10	S	$\bar{S}$	0
11	S	S	–
12	S	S	–
13	$\bar{S}$	S	1
14	S	$\bar{S}$	0
15	S	$\bar{S}$	0
16	S	$\bar{S}$	0
17	S	$\bar{S}$	0
18	$\bar{S}$	S	1
			$\sum = 6$

TABLA 5.15: Distribución muestral de $\hat{p}$ para $n = 14$ y $\pi = .40$.

Proporción $\hat{p}$	Número de éxitos y	Probabilidad $P(y)$
.000	0	.00078
.071	1	.00731
.143	2	.03169
.214	3	.08452
.286	4	.15495
.357	5	.20660
.429	6	.20660
.500	7	.15741
.571	8	.09182
.643	9	.04081
.714	10	.01360
.786	11	.00330
.857	12	.00055
.929	13	.00006
1.00	14	.00000

TABLA 5.16: Resultados positivos y negativos obtenidos bajo dos condiciones.

Par	Condición uno	Condición dos
1	+	−
2	+	−
3	−	+
4	−	−
5	+	−
6	−	+
7	+	+
8	+	−
9	−	+
10	−	+
11	+	−
12	+	−
13	−	+
14	−	+
15	+	−
16	−	+
17	−	+
18	+	−

EJEMPLO 5.14

La tabla 5.16 muestra los resultados positivos (+) y negativos (–) obtenidos en dos condiciones. Utilice los métodos aproximado y exacto para llevar a cabo una prueba de equivalencia de dos colas de la hipótesis nula que plantea que la proporción de resultados que favorece la condición uno no se localiza en el intervalo .3 a .7. Utilice un nivel $\alpha = .05$.

Solución La tabla 5.17 muestra el resultado codificado de cada par. Como el interés reside en la proporción de resultados que favorecen la condición uno, esta condición se codifica con un uno; en tanto que los resultados que favorecen la condición dos se codifican con un cero. De esta tabla se puede ver que $\hat{p} = \frac{8}{16} = .50$.

Como la prueba es de dos colas, se deben realizar tanto la prueba uno como la prueba dos. La hipótesis nula de equivalencia será rechazada sólo si ambas pruebas son significativas. Las hipótesis para estas pruebas son las siguientes.

Prueba uno	Prueba dos
$H_{01} : \pi = .7$	$H_{02} : \pi = .3$
$H_{A1} : \pi < .7$	$H_{A2} : \pi > .3$

La prueba aproximada se lleva a cabo mediante la aplicación de la ecuación 4.9, lo que nos da los valores Z obtenidos de

$$Z_1 = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{.50 - .70}{\sqrt{\frac{(.70)(.30)}{16}}} = -1.75$$

y para la segunda

$$Z_2 = \frac{.50 - .30}{\sqrt{\frac{(.30)(.70)}{16}}} = 1.75$$

Como se puede ver en el Apéndice A, los valores críticos de Z para estas dos pruebas de una cola son −1.65 y 1.65, respectivamente. Se deduce que ambas hipótesis nulas se rechazan, lo que nos lleva al rechazo de la hipótesis nula de equivalencia.

La prueba de equivalencia exacta se realiza usando la ecuación 4.5 para construir distribuciones exactas de $\hat{p}$ en las condiciones $\pi = .7$ y $\pi = .3$ para $n = 16$. Estas distribuciones se presentan en la tabla 5.18. De acuerdo con el método explicado en la página 108 para calcular el valor-*p* para una prueba exacta de una cola, el valor-*p* para la prueba uno es .04868 + .01854 + .00556 + .00130 + .00023 + .00003 + .00000[10] + .00000 + .00000 = .07434, y se obtienen los mismos valores para la prueba dos. Como resultado, ninguna de las hipótesis se rechaza, lo que significa que la hipótesis nula de equivalencia no se rechaza. En este caso, la prueba aproximada es significativa y la prueba exacta no lo es. ■

5.3.3 Intervalo de confianza para una proporción de muestras apareadas

Razonamiento. Mientras que la prueba de McNemar enfrenta la pregunta "¿La proporción de resultados que favorecen una condición y no la otra (π) es diferente a .5?", el intervalo de confianza correspondiente plantea la pregunta: "¿Qué proporción de los resultados (π) favorece una condición sobre la otra?" Como se vio en el apartado 4.5, generalmente se prefieren los intervalos de confianza a las pruebas de hipótesis en situaciones en las que ambos métodos son aplicables. En la mayoría de las situaciones en las que es aplicable la prueba de McNemar, también son aplicables los intervalos de confianza del tipo de los estudiados en esta sección. Cuando ése es el caso, por lo regular se prefiere el intervalo de confianza.

El intervalo de confianza. En el apartado 4.4.6 aprendimos a construir intervalos de confianza para una proporción de la población (π). Los métodos aproximado y exacto que usted aprendió son directamente aplicables al problema que se presenta.[11] Por esa razón no se revisará el proceso para construir intervalos de confianza para π, ya que esto sería redundante, pero sí se verán algunas aplicaciones en muestras.

EJEMPLO 5.15

Suponga que se hacen 14 observaciones apareadas relacionadas con un estudio en particular. De los 14 pares, cuatro resultados favorecen el tratamiento uno, seis favorecen el tratamiento dos y cuatro son no informativos. Utilice intervalos de confianza del 95% aproximados y exactos, como se describe en el apartado 4.4.6, para construir intervalos de confianza bilaterales y calcular la proporción de resultados que favorecen el tratamiento uno.

Solución Las ecuaciones 4.16 y 4.17 producen el siguiente intervalo aproximado

$$S = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .4 - 1.96\sqrt{\frac{(.4)(.6)}{10}} = .096$$

[10] Estos valores no son realmente cero, pero se reportan así por el número de decimales usados en la aproximación.

[11] Quizá quiera repasar el apartado 4.6.6 de la página 148 antes de continuar.

TABLA 5.17: Resultados positivos y negativos de dos condiciones.

Par	Condición uno	Condición dos	Resultado
1	+	−	1
2	+	−	1
3	−	+	0
4	−	−	−
5	+	−	1
6	−	+	0
7	+	+	−
8	+	−	1
9	−	+	0
10	−	+	0
11	+	−	1
12	+	−	1
13	−	+	0
14	−	+	0
15	+	−	1
16	−	+	0
17	−	+	0
18	+	−	1
18			$\sum = 8$

TABLA 5.18: Distribuciones muestrales de $\hat{p}$ para $n = 16$ y $\pi = .3$ y $.7$.

Proporción $\hat{p}$	Número de éxitos y	$\pi = .30$ $P(y)$	$\pi = .70$ $P(y)$
.0000	0	.00332	.00000
.0625	1	.02279	.00000
.1250	2	.07325	.00000
.1875	3	.14650	.00003
.2500	4	.20405	.00023
.3125	5	.20988	.00130
.3750	6	.16490	.00556
.4375	7	.10096	.01854
.5000	8	.04868	.04868
.5625	9	.01854	.10096
.6250	10	.00556	.16490
.6875	11	.00130	.20988
.7500	12	.00023	.20405
.8125	13	.00003	.14650
.8750	14	.00000	.07325
.9375	15	.00000	.02279
1.0000	16	.00000	.00332

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .4 + 1.96\sqrt{\frac{(.4)(.6)}{10}} = .704$$

Utilizando $S = 4$ en las ecuaciones 4.20 y 4.21, los grados de libertad necesarios para calcular el valor F requerido para calcular el valor exacto de L son

$$df_{LN} = 2(n - S + 1) = 2(10 - 4 + 1) = 14$$

y

$$df_{LD} = 2S = (2)(4) = 8.$$

El Apéndice C muestra que con un numerador de 14 grados de libertad y un denominador de 8 grados de libertad, el valor apropiado de F que se debe usar para la construcción del límite inferior de un intervalo de confianza del 95% es 4.13. Si utilizamos este valor en la ecuación 4.18, obtenemos

$$L = \frac{S}{S + (n - S + 1)F_L} = \frac{4}{4 + (10 - 4 + 1)4.13} = .122.$$

Los grados de libertad necesarios para obtener el valor F requerido para calcular el valor exacto de U son, mediante las ecuaciones 4.22 y 4.23,

$$df_{UN} = 2(S + 1) = 2(4 + 1) = 10$$

y

$$df_{UD} = 2(n - S) = 2(10 - 4) = 12.$$

El valor F con un numerador y un denominador con 10 y 12 grados de libertad, respectivamente, que debe usarse en la construcción del límite superior de un intervalo de confianza del 95% es, según el Apéndice C, 3.37. Luego, por la ecuación 4.19

$$U = \frac{(S + 1)F_U}{n - S + (S + 1)F_U} = \frac{(4 + 1)3.37}{10 - 4 + (4 + 1)3.37} = .737.$$

Entonces, el intervalo de confianza del 95% bilateral exacto es de .122 a .737. Esto se compara con el intervalo aproximado de .096 a .704. La discrepancia entre estos intervalos es de esperarse dado el pequeño tamaño de la muestra. La considerable longitud de ambos intervalos también se debe al tamaño de la muestra. ■

EJEMPLO 5.16

Se realiza un estudio (ficticio) para determinar si la serie de síntomas reportados por los veteranos de la Guerra del Golfo es más frecuente en este grupo que en un grupo comparable de veteranos que no participaron en esta acción. Con este fin, se entrevistan 14,791 pares semejantes de veteranos en los que un miembro de cada par ha participado en esa guerra y el otro no. El apareamiento se hace con base en la edad, el género, la rama del servicio militar, las fechas de servicio y la ocupación militar.

En 628 de estos pares, el participante de la Guerra del Golfo reportó haber experimentado uno o más de los síntomas que constituyen el presunto síndrome, en tanto que el otro miembro del par no. En 174 pares, el individuo que no participó en la guerra reportó uno o más síntomas, en tanto que el participante de la Guerra del Golfo no. Trece mil novecientos ochenta y nueve pares fueron no informativos. Utilice estos datos para formar intervalos de confianza bilaterales del 95% para la proporción de resultados en los que el participante reportó uno o más síntomas, en tanto que el no participante no. Forme intervalos aproximados y exactos.

Solución Considerando que $\hat{p} = \frac{628}{628+174} = .783$, las ecuaciones 4.16 y 4.17 producen el siguiente intervalo aproximado

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .783 - 1.96\sqrt{\frac{(.783)(.217)}{802}} = .754$$

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .783 + 1.96\sqrt{\frac{(.783)(.217)}{802}} = .812.$$

Usando $S = 628$ en las ecuaciones 4.20 y 4.21, los grados de libertad necesarios para obtener el valor F requerido para calcular el valor exacto de L son

$$df_{LN} = 2(n - S + 1) = 2(802 - 628 + 1) = 350$$

y

$$df_{LD} = 2S = (2)(628) = 1256.$$

El denominador calculado de 1256 grados de libertad no se muestra en el Apéndice C. Sin embargo, este valor es tan grande que permite el uso de ∞ grados de libertad. Así, con un numerador de 350 grados de libertad y un denominador de ∞ grados de libertad, el valor F correcto que se debe usar para la construcción del límite inferior de un intervalo de confianza del 95% es 1.15.[12] Si usamos este valor en la ecuación 4.18 obtenemos,

$$L = \frac{S}{S + (n - S + 1)F_L} = \frac{628}{628 + (802 - 628 + 1)1.15} = .757.$$

Los grados de libertad necesarios para encontrar el valor F requerido para calcular el valor exacto de U mediante las ecuaciones 4.22 y 4.23 son

$$df_{UN} = 2(S + 1) = 2(628 + 1) = 1258$$

y

$$df_{UD} = 2(n - S) = 2(802 - 628) = 348.$$

Como ninguno de estos grados de libertad se encuentra en el Apéndice C, usaremos ∞ para los grados de libertad del numerador y 350 para los grados de libertad del denominador,[13] lo que nos da un valor F de 1.17. Luego, por la ecuación 4.19

$$U = \frac{(S + 1)F_U}{n - S + (S + 1)F_U} = \frac{(628 + 1)1.17}{802 - 628 + (628 + 1)1.17} = .809.$$

---

[12] De hecho, el valor F asociado con 350 y 1256 grados de libertad es 1.18 cuando se reporta con dos decimales.

[13] Si éste fuera un estudio real, y no un ejercicio de un libro de texto, utilizaríamos un programa de cómputo para calcular el valor F con 1258 y 348 grados de libertad.

Observe que el intervalo aproximado de .754 a .812 está muy cerca del intervalo exacto de .757 a .809. Note también que ambos intervalos son realmente estrechos, por lo que dan cálculos de π que serían satisfactorios para muchas aplicaciones de investigación. Ambos resultados son atribuibles al gran tamaño de la muestra.

Con base en estos intervalos, ¿cuál es su conclusión sobre la pregunta de si los participantes en la Guerra del Golfo reportan uno o más síntomas más a menudo que los que no participaron? ¿Su conclusión cambiaría si el intervalo hubiera sido de .470 a .530? ¿Por qué? ■

5.3.4 Suposiciones

Véase la página 116.

5.4 MÉTODOS RELACIONADOS CON LAS TASAS DE RIESGO DE MUESTRAS APAREADAS

5.4.1 Antecedentes

El **riesgo** de algún evento en un grupo en particular es la *probabilidad* de que el evento ocurra a un miembro de ese grupo. Por ejemplo, suponga que un grupo de trabajadores industriales está expuesto (E), por rutina, a un solvente químico que se sospecha está relacionado con el cáncer de vejiga (D). El riesgo de cáncer de vejiga para este grupo es la probabilidad de que un miembro del grupo experimente esta enfermedad. Usando la notación del capítulo 3, el riesgo para el grupo puede caracterizarse como $P\ (D \mid E)$, que se lee como "la probabilidad de enfermedad dada la exposición". De igual manera, podemos caracterizar el riesgo para los trabajadores de la misma fábrica que no están expuestos al solvente como $P\ (D \mid \overline{E})$, que se lee como "la probabilidad de enfermedad dada la no exposición".

Parece razonable que un investigador de higiene industrial quiera comparar la probabilidad de que un trabajador expuesto experimente la enfermedad con la probabilidad de que un trabajador no expuesto la experimente. Si la probabilidad de los trabajadores expuestos es mayor que la de los trabajadores no expuestos, la implicación puede ser una asociación entre la exposición al solvente y el cáncer de vejiga. Existen diversas formas de hacer una comparación entre las dos probabilidades, incluyendo el cálculo de la diferencia por medio de una simple resta. Una manera común de comparar las dos probabilidades consiste en transformarlas en un cociente conocido como **tasa de riesgo** (*risk ratio*, RR) que se define por la ecuación 5.6.

$$\boxed{RR = \frac{P\ (D \mid E)}{P\ (D \mid \overline{E})}} \tag{5.6}$$

Usted reconocerá lo anterior como la ecuación 3.11 de la página 59.[14] Si el riesgo de enfermedad[15] es el mismo en los dos grupos, la tasa de riesgo es igual a uno, lo que implicaría que no existe ninguna relación entre la exposición y la enfermedad. Por otra parte, un tasa de riesgo mayor que uno implicaría que la exposición se asocia con un riesgo de enfermedad mayor que el de la no exposición. Se dice que las tasas menores que uno son **protectores** porque la exposición implica un riesgo menor de enfermedad. Esto podría ocurrir, por ejemplo, cuando las

[14] En este punto quizá quiera repasar el apartado 3.3.6 de la página 59.

[15] Si bien nos referimos a enfermedad y exposición en la ecuación 5.6, los términos son genéricos y representan cualesquiera variables que se estén examinando.

personas reciben una vacuna. Representaremos el parámetro y el estadístico de la tasa de riesgo como RR y $\widehat{RR}$, respectivamente.

La muestra de tasas de riesgo de muestras apareadas se define con la ecuación 5.7, donde a, b y c son como se especifica en la tabla 5.8 de la página 176. Observe que el valor de la tasa de riesgo de la muestra dependerá de la manera en la que usted ordene la tabla. Por lo general, usted querrá evaluar el riesgo de alguna exposición o condición de no exposición o alguna otra condición. En este caso la variable de exposición se colocará al lado izquierdo de la tabla, y la variable de no exposición o segunda condición con la que se hará la comparación, en la parte superior. Lo importante es que debe estar consciente de lo que $\widehat{RR}$ estima. Se trata de la probabilidad de que ocurra la variable del lado izquierdo de la tabla, dividida entre la probabilidad de que ocurra la variable de la parte superior de la tabla. El resultado derivado de un arreglo de la tabla será el recíproco del resultado obtenido con el segundo arreglo. Esto es, si usted calcula que $\widehat{RR}$ es 1.2 con un arreglo de los datos, el valor obtenido con el otro arreglo será 1.0/1.2 = .833. Tome nota de la manera en que las tablas están ordenadas en los siguientes ejemplos.

$$\boxed{\widehat{RR} = \frac{a+b}{a+c}} \tag{5.7}$$

5.4.2 Prueba de la hipótesis *RR* = 1 para muestras apareadas

Razonamiento. Como en la situación anterior, suponga que se sospecha que la exposición a cierto solvente químico usado por lo común en fábricas incrementa el riesgo de cáncer de vejiga. Para poner a prueba esta sospecha, durante un tiempo podría darse seguimiento a los trabajadores de una industria en particular que están rutinariamente expuestos al solvente para determinar si desarrollan la enfermedad. Para hacer comparaciones, cada uno de estos trabajadores podría formar un par con un trabajador de la misma industria que no esté expuesto al solvente. Después del periodo de seguimiento, se calcula la tasa de riesgo de la muestra dividiendo la proporción de trabajadores expuestos que desarrollen la enfermedad entre la proporción de trabajadores no expuestos que desarrollen la enfermedad. Si los dos riesgos fueran .005 y .001, respectivamente, la tasa de riesgo sería $\frac{.005}{.001} = 5$. Esto indica que el riesgo de cáncer de vejiga para los trabajadores expuestos es cinco veces mayor que el de los trabajadores no expuestos. Pero, ¿éste es sólo resultado de la muestra específica usada en el estudio? Si se repitiera el estudio con otros trabajadores, ¿el resultado sería notablemente diferente? Aún más importante, ¿la tasa de riesgo en la población uno está indicando que no hay relación entre la exposición y la enfermedad (es decir, el riesgo en los dos grupos es el mismo), o difiere de una que indique que sí existe una relación? Una prueba de hipótesis ayudará a responder estas preguntas.

La prueba. Se puede realizar una prueba de la hipótesis nula $H_0 : RR = 1$ si observamos que cada vez que la tasa de riesgo es 1.0, la proporción de resultados que favorecen una condición sobre la otra es de .5. Cuando la tasa de riesgo no es 1.0, la proporción de resultados que favorecen una condición sobre la otra no es de .5. Esto significa que usando la prueba de McNemar de la hipótesis nula $H_0 : \pi = .5$ podemos simultáneamente poner a prueba la hipótesis nula $H_0 : RR = 1$. Así, para esta prueba, están disponibles tanto el método aproximado como el exacto.

EJEMPLO 5.17

Utilice los datos en la tabla 5.7 de la página 175 para calcular la tasa de riesgo muestra para el riesgo de enfermedad en mellizos que reciben la vacuna dos en comparación con la de mellizos que reciben la vacuna uno; después ponga a prueba la hipótesis nula $H_0 : RR = 1$.[16] Utilice las ecuaciones 5.4 y 5.5 de la página 176 para las pruebas de hipótesis.

Solución Tomando los datos de la tabla 5.7 para formar una tabla de dos por dos obtenemos

		Vacuna uno D	Vacuna uno $\overline{D}$
Vacuna dos	D	2	2
	$\overline{D}$	6	6

Por medio de la ecuación 5.7, la tasa de riesgo de la muestra es

$$\widehat{RR} = \frac{(2+2)}{(2+6)} = .50$$

Así, la probabilidad de enfermedad de un mellizo que recibe la vacuna dos es sólo la mitad de la del mellizo que recibe la vacuna uno. Previamente se realizaron las pruebas de hipótesis. (Véase la primera parte de la página 175). Los valores de Z y χ^2 fueron, respectivamente, 1.414 y 2.0, lo que indica un resultado no significativo. Por consiguiente, no podemos demostrar que la tasa de riesgo difiere de uno, por lo que no puede concluirse que exista una relación entre el tipo de vacuna usada y la enfermedad. ■

EJEMPLO 5.18

Utilice los datos de la tabla 5.7 para llevar a cabo una prueba exacta de la hipótesis nula $H_0 : RR = 1$.

Solución Esta prueba se realizó previamente (véase la primera parte de la página 175) y el valor-p resultante fue de .28906. Como este valor es mayor que $\alpha = .05$, la hipótesis nula $H_0 : RR = 1$ no se rechaza. ■

EJEMPLO 5.19

Evans y Frick [13] (citado por Greenland [20]) presentan los resultados del sistema de reporte de accidentes mortales (*Fatal Accident Reporting System*, FARS), que se relacionan con accidentes mortales de motocicletas con dos pasajeros en los que tanto el conductor como el pasajero eran varones y ninguno usaba casco. Reportan que en 226 casos ambos murieron, en 546 casos sólo el conductor murió y en 378 casos murió sólo el pasajero. Utilice estos datos para calcular la tasa de riesgo comparando el riesgo de muerte del conductor con el del pasajero. Utilice las ecuaciones 5.4 y 5.5 para poner a prueba la hipótesis nula $RR = 1$.

Solución Es conveniente ordenar los datos en una tabla de dos por dos, como en la tabla 5.8 de la página 176.

		Pasajero murió	Pasajero sobrevivió
Conductor	murió	226	546
	sobrevivió	378	

[16] De ahora en adelante las pruebas de hipótesis serán de dos colas con un nivel $\alpha = .05$, a menos de que se especifique algo diferente.

Observe que no se tiene información sobre el número de casos en los que ni el conductor ni el pasajero murieron. Esto no es relevante para el problema que nos ocupa, ya que esta frecuencia no entra en el cálculo de $\widehat{RR}$, ni en las pruebas de hipótesis.

Por medio de la ecuación 5.7, la tasa de riesgo de la muestra es

$$\widehat{RR} = \frac{a+b}{a+c} = \frac{226+546}{226+378} = 1.278$$

Luego, por estos datos, el conductor tiene 1.278 veces el riesgo de muerte del pasajero. Para emplear la ecuación 5.4 consideramos que la celda a es no informativa en lo que se refiere a la prueba de significancia. Además, la proporción de casos en los que el conductor muere pero el pasajero no,[17] $\hat{p}$ es

$$\hat{p}\,\frac{b}{b+c} = \frac{546}{546+378} = .591$$

Entonces, con la ecuación 5.4,

$$Z = \frac{\hat{p}-.5}{\frac{.5}{\sqrt{n}}} = \frac{.591-.50}{\frac{.50}{\sqrt{924}}} = 5.53$$

Como este valor es mayor que 1.96, la hipótesis nula que asevera que $RR = 1$ se rechaza. Por consiguiente, concluimos que los conductores tienen un riesgo de muerte mayor que los pasajeros.

La misma conclusión se obtiene si se aplica la ecuación 5.5:

$$\chi^2 = \frac{(b-c)^2}{b+c} = \frac{(546-378)^2}{546+378} = 30.55$$

Este valor es mayor que el valor crítico de χ^2 de 3.841 del Apéndice D, por lo que la hipótesis nula se rechaza. ■

5.4.3 Establecimiento de equivalencia mediante la tasa de riesgo de muestras apareadas

Razonamiento. Como sucede con las diferencias de medias y las proporciones apareadas, a veces es deseable demostrar que un riesgo asociado con alguna condición es equivalente a un riesgo asociado con alguna otra condición. Esto implicaría el uso de una prueba de equivalencia de dos colas basada en tasas de riesgo. Quizá es más común que se desee demostrar que un riesgo asociado con alguna condición *no es mayor que* un riesgo asociado con alguna otra condición. Esto implica una prueba de una cola.

Por ejemplo, suponga que se pondrá en marcha un programa diseñado para controlar la mosca de la fruta mediante el riego aéreo del químico malatión en grandes áreas geográficas. Como algunos activistas ambientales afirman que tal práctica es dañina para los seres humanos, se podría diseñar un estudio para demostrar que el riesgo después de regar malatión no es mayor que el riesgo antes de tal acción. Con este fin, ciertos sujetos que viven en las áreas que serán regadas se someterían a un examen médico antes y después de que se lleve a cabo el riego. Cada condición examinada se califica como presente o no presente en cada sujeto.

---

[17] En el formato de tabla usado en conjunción con la prueba de McNemar, las entradas en la celda b se codificaron como 1, mientras que en la celda c se codificaron como 0.

La inspección de los datos recabados muestra que para una condición específica (por ejemplo, frecuencia cardiaca elevada), 126 personas manifiestan la condición antes y después del riego; 414 después del riego pero no antes; 390 antes pero no después, y 999 no manifiestan la condición en ningún momento. Los investigadores pondrán a prueba la hipótesis nula de que el riesgo después de rociar el químico es mayor que el riesgo antes de esta acción, contra la hipótesis alternativa de que el riesgo después de rociar el químico no es mayor que el riesgo antes de tal acción.

¿Qué diferencia hay entre el análisis explicado anteriormente y uno que pone a prueba la hipótesis nula de no diferencia en el riesgo contra una hipótesis alternativa de que el riesgo se incrementa después de rociar malatión? ¿Hacia dónde se inclinan las evidencias en los dos análisis? Dejaremos que usted reflexione sobre estas preguntas.

La prueba. Los intervalos de equivalencia para las tasas de riesgo tienen por lo regular, pero no necesariamente, una simetría de aproximadamente 1.0 en el sentido que $EI_U = \frac{1}{EI_L}$ y viceversa. Así, si EI_L es .8, EI_U por lo general sería $\frac{1.0}{.8} = 1.25$.

Utilizando la notación del apartado 4.3.6, la hipótesis nula y la alternativa para una prueba de equivalencia de dos colas para tasas de riesgo apareadas se pueden plantear como

$$H_{0E} : RR \leq EI_L \quad \text{o} \quad RR \geq EI_U$$
$$H_{AE} : EI_L < RR < EI_U$$

donde EI_L e EI_U son los extremos inferior y superior del intervalo de equivalencia. En esencia, la hipótesis nula plantea que la tasa de riesgo no reside en el intervalo de equivalencia, mientras que la alternativa afirma que la tasa de riesgo está en el intervalo. Usted recordará que la hipótesis nula se pone a prueba realizando dos pruebas de una cola a nivel α. Para poder rechazar la hipótesis nula de equivalencia usted debe demostrar que $RR > EI_L$ *y* que $RR < EI_U$, lo que significa que *las dos* pruebas siguientes deben ser significativas.

Prueba uno	Prueba dos
$H_{01} : RR = EI_U$	$H_{02} : RR = EI_L$
$H_{A1} : RR < EI_U$	$H_{A2} : RR > EI_L$

Las hipótesis nula y alternativa para la prueba de equivalencia de una cola es *una* de las siguientes

$$H_{0E} : RR \geq EI_U$$
$$H_{AE} : RR < EI_U$$

o

$$H_{0E} : RR \leq EI_L$$
$$H_{AE} : RR > EI_L$$

La primera hipótesis de equivalencia de una cola planteada arriba se pone a prueba por medio de la prueba uno y la segunda mediante la prueba dos.

Como la prueba de McNemar está limitada a una prueba de la hipótesis $H_0 : RR = 1$, no es útil cuando se trata de poner a prueba la equivalencia. Sin embargo, hay una serie de pruebas aproximadas para este fin. Una de ellas se realiza mediante

$$\boxed{Z = \frac{\ln \widehat{RR} - \ln (RR_0)}{\sqrt{\frac{b+c}{(a+b)(a+c)}}}} \tag{5.8}$$

Los símbolos a, b, y c se muestran en la tabla 5.8 de la página 176, $\widehat{RR}$, es la tasa de riesgo de la muestra como se define en la ecuación 5.7 y RR_0 es la tasa de riesgo hipotético de la población. El símbolo ln () indica que se debe tomar el logaritmo natural del valor que está entre paréntesis.

EJEMPLO 5.20

Calcule la tasa de riesgo para los datos del malatión que se analizaron con anterioridad. Suponiendo que un incremento en el riesgo de menos de 1.1 se considere aceptable, realice una prueba de equivalencia de una cola para demostrar que el incremento del riesgo es menor que este valor. Plantee las hipótesis nula y alternativa de equivalencia e interprete los resultados de las pruebas.

Solución Es conveniente construir una tabla de dos por dos con los datos, como se muestra a continuación.

		Antes del riego	
		elevada	no elevada
Después del riego	elevada	126	414
	no elevada	390	999

Por medio de la ecuación 5.7, la tasa de riesgo de la muestra es

$$\widehat{RR} = \frac{a+b}{a+c} = \frac{126+414}{126+390} = 1.047$$

Como los investigadores están tratando de determinar si la tasa de riesgo es menor que 1.1, la prueba será de una cola, con la hipótesis nula de equivalencia

$$H_{0E} : RR \geq 1.1$$

y la hipótesis alternativa

$$H_{AE} : RR < 1.1$$

Si se realiza la prueba uno mediante la ecuación 5.8, tenemos que

$$Z_1 = \frac{\ln (\widehat{RR}) - \ln (RR_0)}{\sqrt{\frac{b+c}{(a+b)(a+c)}}} = \frac{\ln (1.047) - \ln (1.1)}{\sqrt{\frac{414+390}{(126+414)(126+390)}}} = -.919.$$

Como este valor es mayor que el valor crítico de −1.65, la prueba no es significativa. Por lo tanto, los investigadores no pudieron demostrar que el incremento en el riesgo estaba en el rango aceptable (es decir, menos de 1.1). ■

EJEMPLO 5.21

Utilice los datos de la siguiente tabla para calcular $\widehat{RR}$. Después realice una prueba de equivalencia de dos colas usando un *EI* de .833 a 1.2. Interprete sus resultados.

		Variable dos +	Variable dos −
Variable uno	+	301	771
	−	780	151

Solución Por medio de la ecuación 5.7, la tasa de riesgo de la muestra es

$$\widehat{RR} = \frac{301+771}{301+780} = .992.$$

Esto sugeriría un leve efecto protector para la variable uno. Si realizamos la prueba uno mediante la ecuación 5.8, tenemos que

$$Z_1 = \frac{\ln(\widehat{RR}) - \ln(RR_0)}{\sqrt{\frac{b+c}{(a+b)(a+c)}}} = \frac{\ln(.992) - \ln(1.2)}{\sqrt{\frac{771+780}{(301+771)(301+780)}}} = -5.203.$$

Como este valor es menor que el valor crítico de −1.65, la prueba uno es significativa.

Si realizamos la prueba dos mediante la ecuación 5.8, tenemos que

$$Z_2 = \frac{\ln(.992) - \ln(.833)}{\sqrt{\frac{771+780}{(301+771)(301+780)}}} = 4.775.$$

Como este valor es mayor que el valor crítico de 1.65, la prueba dos es significativa. Ya que ambas pruebas son significativas, la hipótesis nula de equivalencia se rechaza. Por consiguiente, concluimos que la variable uno y la variable dos producen resultados equivalentes, definidos por el intervalo de equivalencia. ■

5.4.4 Intervalo de confianza para la tasa de riesgo de muestras apareadas

Razonamiento. Previamente (sección 4.5) indicamos que por lo general los intervalos de confianza son preferibles a las pruebas de hipótesis. Así, en vez de preguntar: "¿Es la tasa de riesgo de la población diferente de 1?", una pregunta más informativa sería: "¿Cuál es la tasa de riesgo de la población?". En tanto que los métodos exactos para construir intervalos de confianza para *RR* son problemáticos, existen métodos aproximados bastante más sencillos.

El intervalo de confianza. Los límites inferior y superior de la tasa de riesgo de muestras apareadas se pueden obtener a través de las siguientes ecuaciones.

$$\boxed{L = \exp\left[\ln(\widehat{RR}) - Z\sqrt{\frac{b+c}{(a+b)(a+c)}}\right]} \tag{5.9}$$

y

$$\boxed{U = \exp\left[\ln(\widehat{RR}) + Z\sqrt{\frac{b+c}{(a+b)(a+c)}}\right]} \tag{5.10}$$

Los símbolos *exp* y *ln* indican, respectivamente, que se deben tomar la exponencial natural y el logaritmo natural de la expresión contenida entre corchetes, mientras que a, b, y c son las frecuencias de una tabla de dos por dos, como se explicó con anterioridad.

EJEMPLO 5.22

Utilice los datos de Evans y Frick [13] (reproducidos a continuación) para construir un intervalo bilateral de confianza del 95%. ¿Qué significa este intervalo? Utilice *el intervalo* para llevar a cabo una prueba de dos colas de la hipótesis nula $RR = 1$. Explique cómo llegó a la conclusión de esta prueba.

		Pasajero	
		murió	sobrevivió
Conductor	murió	226	546
	sobrevivió	378	

Solución Por medio de la ecuación 5.7

$$\widehat{RR}\frac{a+b}{a+c} = \frac{226+546}{226+378} = 1.278.$$

Luego, por medio de la ecuación 5.9,

$$\begin{aligned} L &= \exp\left[\ln(\widehat{RR}) - Z\sqrt{\frac{b+c}{(a+b)(a+c)}}\right] \\ &= \exp\left[\ln(1.278) - 1.96\sqrt{\frac{546+378}{(226+546)(226+378)}}\right] \\ &= \exp\left[.245 - 1.96\sqrt{.002}\right] \\ &= \exp\left[.157\right] \\ &= 1.170. \end{aligned}$$

Y por medio de la ecuación 5.10.

$$\begin{aligned} U &= \exp\left[\ln(\widehat{RR}) + Z\sqrt{\frac{b+c}{(a+b)(a+c)}}\right] \\ &= \exp\left[.245 + 1.96\sqrt{.002}\right] \\ &= \exp\left[.333\right] \\ &= 1.395. \end{aligned}$$

De este intervalo podemos concluir que los conductores de motocicletas (en las circunstancias antes mencionadas) tienen entre 1.170 y 1.395 veces más riesgo de morir, comparados con los pasajeros. También podemos concluir que si se realizara una prueba de dos colas de la hipótesis nula $RR = 1$ a un nivel $\alpha = .05$, el resultado sería significativo. Se llega a esta conclusión considerando que el valor nulo de 1.0 no está en el intervalo.[18] ■

EJEMPLO 5.23

Utilice los datos del malatión expuestos antes (y reproducidos a continuación) para construir un intervalo de confianza unilateral del 95% y calcular el límite inferior del incremento de la frecuencia cardiaca. ¿Qué significa este intervalo? Utilice *el intervalo* para llevar a cabo una prueba de una cola de la hipótesis nula $RR = 1$ contra la alternativa $RR > 1$. Explique cómo llegó a su conclusión en esta prueba.

		Antes del riego	
		elevada	no elevada
Después del riego	elevada	126	414
	no elevada	390	999

Solución Por medio de la ecuación 5.7

$$\widehat{RR} = \frac{a+b}{a+c} = \frac{126+414}{126+390} = 1.047.$$

Luego, por medio de la ecuación 5.9

$$\begin{aligned} L &= \exp\left[\ln(\widehat{RR}) - Z\sqrt{\frac{b+c}{(a+b)(a+c)}}\right] \\ &= \exp\left[\ln(1.047) - 1.65\sqrt{\frac{414+390}{(126+414)(126+390)}}\right] \\ &= \exp\left[.046 - 1.65\sqrt{.003}\right] \\ &= \exp\left[-.044\right] \\ &= .957. \end{aligned}$$

De este intervalo podemos concluir que el riesgo de un incremento de la frecuencia cardiaca después de rociar malatión, comparado con el riesgo antes de tal práctica es mayor que o igual a .957. Tome en cuenta que un valor de .957 sería protector, un valor de 1.0 indicaría que no hay ninguna diferencia en el riesgo, y un valor mayor que 1.0 implicaría un riesgo elevado. Ya que todos estos valores son posibles, nos queda poca información sobre el riesgo asociado con la exposición a este químico. Como es lógico, podemos concluir que si se realizara una prueba de una cola de la hipótesis nula $RR = 1$ y de la alternativa $RR > 1$ con $\alpha = .05$, el resultado no sería significativo. Se llega a esta conclusión considerando que el valor nulo de 1.0 está en el intervalo.[19] ■

[18] Véase el apartado 4.5 de la página 152 si esta explicación no está del todo clara.
[19] Véase el apartado 4.5 de la página 152 si esta explicación no está del todo clara.

5.4.5 Suposiciones

Como el método para probar $H_0 : RR = 1$ se basa en la prueba de McNemar, las suposiciones que subyacen en la prueba son las planteadas en la página 116.

Los métodos para realizar pruebas de equivalencia y para construir intervalos de confianza que se presentan aquí dependen de una aproximación de la curva normal. Los tamaños de las muestras deben ser bastante grandes para asegurar que la aproximación dé resultados lo suficientemente precisos. Greenland [20] sugiere una regla general para el tamaño de la muestra, la cual requiere que tanto el número de personas expuestas como el de las no expuestas sean mayores que o iguales a cinco. Además, se supone que los pares de observaciones son independientes entre sí. Esto es, los resultados obtenidos de un par no influyen ni están influidos por los resultados obtenidos de cualquier otro par.

5.5 MÉTODOS RELACIONADOS CON LAS RAZONES DE PROBABILIDAD DE MUESTRAS APAREADAS

5.5.1 Antecedentes

En ciertas circunstancias, que se analizarán en el capítulo 6, la tasa de riesgo no es un indicador apropiado del riesgo relativo. En tales casos, es común usar la razón de probabilidad como medida.

Las **posibilidades** u oportunidades asociadas con algún evento de un grupo en particular se refieren a la *probabilidad* de que el evento ocurra a un miembro de ese grupo, dividida entre la probabilidad de que el evento no ocurra a un miembro de ese grupo. Por ejemplo, suponga que se ha identificado a un grupo de trabajadores industriales que ha sido diagnosticado con cáncer de vejiga. Las posibilidades de que un miembro de este grupo haya estado expuesto a cierto solvente industrial que se sospecha causa cáncer de vejiga es, usando la notación del capítulo 3, $\frac{P(E|D)}{P(\bar{E}|D)}$, que se lee como “la probabilidad de exposición dada la enfermedad dividida entre la probabilidad de no exposición dada la enfermedad”. Si, por ejemplo, este valor fuera 2.0, diríamos que la probabilidad de exposición al solvente de los trabajadores que tienen cáncer de vejiga es el doble de la probabilidad de no haber estado expuestos al solvente.

De igual forma, podemos describir las posibilidades de exposición al solvente para los trabajadores de la misma fábrica que no sufren de cáncer de vejiga como $\frac{P(E|\bar{D})}{P(\bar{E}|\bar{D})}$, que se lee como “la probabilidad de exposición sin enfermedad dividida entre la probabilidad de no exposición sin enfermedad”. Si, por ejemplo, este valor fuera .80, diríamos que la probabilidad de exposición al solvente para los trabajadores que no tienen cáncer de vejiga es .8 de la probabilidad de que no estuvieron expuestos al solvente.

Parece razonable que un investigador de higiene industrial quiera comparar las posibilidades de que un trabajador con cáncer de vejiga haya estado expuesto al solvente con las posibilidades de que un trabajador sin cáncer de vejiga haya estado expuesto. Si las posibilidades de exposición para los trabajadores con cáncer de vejiga son mayores que las posibilidades de los trabajadores sin cáncer de vejiga, esto puede implicar una asociación entre la exposición al solvente y el cáncer de vejiga. Existen diferentes formas de hacer una comparación entre las dos posibilidades, incluyendo el cálculo de la diferencia mediante una simple resta. Una manera común de comparar las dos posibilidades es ordenarlas en un cociente conocido como **razón de probabilidad, oportunidad relativa** o **razón de oportunidades** *(odds ratio*, OR), que se define por la ecuación 5.11. Observe que

$$OR = \frac{\frac{P(E|D)}{P(\bar{E}|D)}}{\frac{P(E|\bar{D})}{P(\bar{E}|\bar{D})}}$$

que se simplifica en

$$\boxed{OR = \frac{P(E|D)P(\bar{E}|\bar{D})}{P(\bar{E}|D)P(E|\bar{D})}}$$

La ecuación 3.12 de la página 60[20] expresa las posibilidades de enfermedad para ambos grupos, expuesto y no expuesto, mientras que la ecuación 5.11 expresa las posibilidades de exposición para los grupos con y sin enfermedad. Ambas formas son usadas por los investigadores.

Será de utilidad distinguir entre dos formas generales de diseños de investigación[21] con las cuales se emplea comúnmente la razón de probabilidad de muestras apareadas. En el apartado 5.4.1 aparece un ejemplo en el que se hizo un seguimiento a trabajadores que estaban rutinariamente expuestos a un solvente junto con trabajadores que no estaban expuestos, para así determinar quiénes desarrollaban cáncer de vejiga. La razón de probabilidad calculada para este diseño expresaba las posibilidades de enfermedad (cáncer de vejiga) para los grupos de trabajadores expuestos y no expuestos al solvente. Este tipo de diseños, en el que se sigue a sujetos expuestos y no expuestos para determinar si presentan la enfermedad se conoce como diseños de **cohorte prospectiva**. Los datos de este tipo de diseño normalmente se ordenan para su análisis como se muestra en la tabla 5.19 de la siguiente página.

En contraste, en esta sección se dio un ejemplo en el que trabajadores que ya habían desarrollado la enfermedad (cáncer de vejiga) fueron comparados con trabajadores que no la habían desarrollado. La razón de probabilidad calculada para este diseño expresaba las posibilidades de exposición (al solvente) para los trabajadores con y sin la enfermedad (cáncer de vejiga). Los diseños de este tipo, en los que se comparan sujetos que ya tienen la enfermedad con sujetos que no la tienen, con el fin de determinar cuáles tienen mayor exposición, se denominan diseños de **control de casos**. Los datos de este tipo de diseño normalmente se ordenan para su análisis como se muestra en la tabla 5.20.

Para la siguiente explicación, supondremos que tenemos a la mano un estudio de control de casos; aunque con las debidas sustituciones de exposición por enfermedad y viceversa, los comentarios se aplican de igual forma a los estudios prospectivos.

Si las posibilidades de exposición[22] en los dos grupos son las mismas, la razón de probabilidad es uno, lo que implicaría que no hay relación entre la enfermedad y la exposición. Por otra parte, una razón de probabilidad mayor que uno implicaría que la enfermedad está asociada con una exposición mayor que la ausencia de enfermedad. Se dice que los cocientes que son menores a uno son **protectores** porque la enfermedad implica menos exposición que la ausencia de enfermedad. Esto podría ocurrir, por ejemplo, cuando las personas con la enfermedad están menos expuestas a un suplemento vitamínico diario que las personas sin la enfermedad. Representaremos el parámetro y el estadístico de la razón de probabilidad como OR y $\widehat{OR}$, respectivamente. La muestra de razón de probabilidad de muestras apareadas se define por la ecuación 5.12, donde b y c son como se especifica en la tabla 5.8 de la página 176.

$$\boxed{\widehat{OR} = \frac{b}{c}}$$

[20] En este punto quizá quiera repasar el material de la página 60.

[21] Éstos son solamente dos de los muchos diseños que hay.

[22] Si bien nos referimos a enfermedad y exposición en la ecuación 5.11, los términos son genéricos y representan cualesquiera variables que se estén examinando.

TABLA 5.19: Arreglo típico de datos para el estudio prospectivo de muestras apareadas.

		No expuestos	
		Enfermedad	No enfermedad
Expuestos	Enfermedad	a	b
	No Enfermedad	c	d

TABLA 5.20: Arreglo típico de datos para el estudio de control de casos de muestras apareadas.

		No enfermedad	
		Expuestos	No expuestos
Enfermedad	Expuestos	a	b
	No expuestos	c	d

5.5.2 Prueba de la hipótesis $OR = 1$ para muestras apareadas

Razonamiento. Como en la situación antes mencionada, suponga que se sospecha que la exposición a un solvente químico en particular, usado con frecuencia en fábricas, incrementa el riesgo de cáncer de vejiga. Para poner a prueba esta sospecha se forman pares de trabajadores que han sido diagnosticados con cáncer de vejiga con trabajadores de la misma industria que no tienen cáncer de vejiga. Los antecedentes laborales de estos trabajadores se examinan para determinar quiénes han estado expuestos al solvente y quiénes no. La razón de probabilidad de la muestra se calcula por medio de la ecuación 5.12. Si este cociente fuera 5.0, por ejemplo, la indicación sería que las posibilidades de exposición al solvente para los trabajadores con cáncer de vejiga son cinco veces mayores a las de los trabajadores sin cáncer. Pero, ¿éste es simplemente el resultado de la muestra específica usada en el estudio? Si se repitiera el estudio con otros trabajadores, ¿el resultado sería considerablemente diferente? Aún más importante, ¿la razón de probabilidad en la población es uno, lo que indica que no hay ninguna relación entre enfermedad y exposición (es decir, las posibilidades en los dos grupos son las mismas), o es diferente de uno, lo que indica que sí existe tal relación? Una prueba de hipótesis ayudará a responder estas preguntas.

La prueba. Se puede realizar una prueba de la hipótesis nula $H_0 : OR = 1$ observando que cuando la razón de probabilidad es 1.0, la proporción de resultados que favorece a una condición sobre la otra es .5. Cuando la razón de probabilidad no es 1.0, la proporción de resultados que favorece a una condición sobre la otra no es .5. Esto significa que si se usa la prueba de McNemar de la hipótesis nula $H_0 : \pi = .5$, podemos poner a prueba simultáneamente la hipótesis nula $H_0 : OR = 1$. Usted recordará que sucedía lo mismo con la tasa de riesgo. Se deduce que la prueba de McNemar pone a prueba simultáneamente las hipótesis $\pi = .5$, $RR = 1$ y $OR = 1$. Así, para esta prueba, están disponibles tanto el método aproximado como el exacto.

EJEMPLO 5.24

Suponga que en el estudio del cáncer de vejiga analizado antes, se encontraron 13 pares de trabajadores en los que ambos miembros del par habían estado expuestos al solvente, 25 pares en los que el trabajador con cáncer estuvo expuesto y el miembro sin cáncer no estuvo expuesto, 5 pares en los que el trabajador con cáncer no estuvo expuesto y el otro miembro, también con cáncer, sí estuvo expuesto, y 55 pares en los que ninguno de los miembros estuvo expuesto.

Utilice los datos anteriores para calcular la razón de probabilidad de la muestra. ¿Qué significa este cociente? Utilice una prueba de McNemar aproximada para probar la hipótesis nula $OR = 1$. Interprete sus resultados.

Solución Por comodidad, los datos se ordenaron en una tabla de dos por dos como se muestra en la tabla 5.21.

De la ecuación 5.12,

$$\widehat{OR} = \frac{b}{c} = \frac{25}{5} = 5.0$$

Así, las posibilidades de exposición para las personas con cáncer de vejiga son cinco veces mayores que las de las personas sin la enfermedad.

Por la ecuación 5.5,

$$\chi^2 = \frac{(b-c)^2}{b+c} = \frac{(25-5)^2}{25+5} = \frac{400}{30} = 13.33$$

El valor crítico de esta prueba se encuentra utilizando el Apéndice D con un grado de libertad. Este valor es 3.841 para $\alpha = .05$. Debido a que el valor obtenido de 13.33 es mayor que el valor crítico de 3.84, la hipótesis nula $H_0 : \pi = .5$ se rechaza, lo que significa que $H_0 : \widehat{OR} = 1$ también se rechaza. Ahora se puede afirmar que el cáncer de vejiga está relacionado con la exposición al solvente en cuestión. ¿Cree usted que este resultado demuestra que el solvente *causa* cáncer de vejiga? ■

EJEMPLO 5.25

Suponga que los datos de la tabla 5.21 se modificaron como se muestra en la tabla 5.22. Utilice estos datos para calcular $\widehat{OR}$. Después, lleve a cabo una prueba exacta de dos colas de la hipótesis nula $H_0 : OR = 1$.

Solución Por la ecuación 5.12,

$$\widehat{OR} = \frac{b}{c} = \frac{6}{4} = 1.5$$

Se puede hacer una prueba exacta de la hipótesis nula $H_0 : OR = 1$ al poner a prueba la hipótesis nula $H_0 : \pi = .5$ con la prueba de McNemar. Con este fin, observamos que las frecuencias de las celdas a y d son no informativas[23] y que la proporción de pares en los que el trabajador con cáncer estuvo expuesto al solvente y que el trabajador sin cáncer no estuvo expuesto $(\hat{p})$ es $\frac{b}{b+c} = \frac{6}{6+4} = .6$.

Se puede construir la distribución muestral de $\hat{p}$ en la condición $\pi = .5$ y $n = 10$ mediante la ecuación 4.5, como se presenta en la tabla 5.23. Al observar que $\hat{p} = .60$ y empleando el método explicado en el apartado que aparece al principio de la página 108 para encontrar el valor-p para una prueba exacta de dos colas, este valor se calcula como .20508 + .11719 + .04395 + .00977 + .00098 = .37697 para la cola derecha, de modo que $p = (2)(.37697) = .75394$. Como este valor es mayor que $\alpha = .05$, no se rechaza la hipótesis nula. Por lo tanto, no hemos podido demostrar que OR es diferente de uno. ■

---

[23] Véase la página 174.

TABLA 5.21: Datos de un estudio de control de caso que relaciona el cáncer de vejiga con la exposición a cierto solvente.

		Sin cáncer	
		Solvente	Sin solvente
Cáncer	Solvente	13	25
	Sin solvente	5	55

TABLA 5.22: Datos de un estudio de control de caso que relaciona el cáncer de vejiga con la exposición a cierto solvente.

		Sin cáncer	
		Solvente	Sin solvente
Cáncer	Solvente	9	6
	Sin solvente	4	29

TABLA 5.23: Distribución muestral de $\hat{p}$ para $n = 10$ y $\pi = .50$.

Proporción $\hat{p}$	Número de éxitos y	Probabilidad $P(y)$
.000	0	.00098
.100	1	.00977
.200	2	.04395
.300	3	.11719
.400	4	.20508
.500	5	.24609
.600	6	.20508
.700	7	.11719
.800	8	.04395
.900	9	.00977
1.00	10	.00098

5.5.3 Establecimiento de equivalencia mediante la razón de probabilidades de muestras apareadas

Razonamiento. Como sucede con las tasas de riesgo, a veces es aconsejable demostrar que las posibilidades asociadas con alguna condición son equivalentes a las posibilidades asociadas con alguna otra condición. Esto implicaría el uso de una prueba de equivalencia de dos colas basada en razones de probabilidad. Quizá es más común la conveniencia de demostrar que las posibilidades asociadas con una condición *no son mayores que* las posibilidades asociadas con alguna otra condición. Esto implica una prueba de una cola.

Por ejemplo, reconsidere el programa analizado anteriormente que se puso en marcha para controlar la mosca de la fruta mediante el riego aéreo del químico malatión en grandes áreas geográficas. Como algunos activistas ambientales argumentan que tal práctica es dañina para los seres humanos, se podría diseñar un estudio para demostrar que las probabilidades de algún resultado indeseable después del riego no son mayores que las probabilidades previas al riego. Con este fin, ciertas personas que viven en las áreas que se regarán podrían someterse a un examen médico antes y después de que se lleve a cabo el riego. Cada condición examinada se califica como presente o no presente en cada sujeto.

La inspección de los datos recabados muestra que para una condición específica (por ejemplo, frecuencia cardiaca elevada), 126 personas manifiestan la condición antes y después del riego; 414 después del riego pero no antes; 390 antes pero no después y 999 no manifiestan la condición en ningún momento. Los investigadores pondrán a prueba la hipótesis nula de que las probabilidades después de rociar malatión son mayores que las probabilidades previas a tal práctica, contra la hipótesis alternativa de que las probabilidades después de la aspersión no son mayores que las probabilidades antes de la misma.

La prueba. Como sucede con las tasas de riesgo, los intervalos de equivalencia para las razones de posibilidades tienen por lo regular, aunque no necesariamente, una simetría de aproximadamente 1.0 en el sentido que $EI_U = \frac{1}{EI_L}$ y viceversa. Así, si EI_L es .8, EI_U generalmente sería $\frac{1.0}{.8} = 1.25$.

Usando la notación del apartado 4.3.6, la hipótesis nula y la alternativa para una prueba de equivalencia de dos colas para razones de probabilidad apareadas se pueden plantear como

$$H_{0E} : OR \leq EI_L, \text{ o bien, } OR \geq EI_U$$
$$H_{AE} : EI_L < OR < EI_U$$

donde EI_L y EI_U son, respectivamente, los extremos inferior y superior del intervalo de equivalencia. En esencia, la hipótesis nula plantea que la razón de probabilidad no reside en el intervalo de equivalencia, en tanto que la hipótesis alternativa afirma que la razón de probabilidad está en el intervalo. Usted recordará que la hipótesis nula se pone a prueba realizando dos pruebas de una cola al nivel α. Para rechazar la hipótesis nula de equivalencia, usted debe mostrar que $OR > EI_L$ *y* que $OR < EI_U$, lo que significa que *las dos* pruebas siguientes deben ser significativas.

Prueba uno	Prueba dos
$H_{01} : OR = EI_U$	$H_{02} : OR = EI_L$
$H_{A1} : OR < EI_U$	$H_{A2} : OR > EI_L$

Las hipótesis nula y alternativa para la prueba de equivalencia de una cola son *una* de las siguientes.

$$H_{0E} : OR \geq EI_U$$
$$H_{AE} : OR < EI_U$$

o

$$H_{0E} : OR \leq EI_L$$
$$H_{AE} : OR > EI_L$$

La primera hipótesis de equivalencia de una cola planteada arriba se somete a la prueba 1 y la segunda a través de la prueba 2.

Como la prueba de McNemar está limitada a una prueba de la hipótesis $H_0 : OR = 1$, no es útil cuando se está haciendo una prueba de equivalencia. Sin embargo, para muestras apareadas, existe una relación entre π y OR por una parte, y entre $\hat{p}$ y $\widehat{OR}$ por otra, lo que permite que las pruebas relacionadas con OR se conviertan en pruebas de π; por consiguiente, se pueden hacer pruebas con métodos con los que usted ya está familiarizado. Esto también quiere decir que se pueden realizar tanto pruebas aproximadas como exactas.

La relación entre π y OR está dada por

$$\boxed{\pi = \frac{OR}{1+OR}} \tag{5.13}$$

y entre $\hat{p}$ y $\widehat{OR}$ por

$$\boxed{\hat{p} = \frac{\widehat{OR}}{1+\widehat{OR}}} \tag{5.14}$$

EJEMPLO 5.26

Calcule la razón de probabilidad para los datos del malatión contenidos en la siguiente tabla. Suponiendo que un incremento en la razón de probabilidad menor que 1.1 se considera aceptable, realice una prueba aproximada de equivalencia de una cola para demostrar que el incremento en la razón de probabilidad es menor que este valor. Plantee las hipótesis nula y alternativa de equivalencia, e interprete los resultados de la prueba.

		Antes del riego	
		elevada	no elevada
Después del riego	elevada	126	414
	no elevada	390	999

Solución Por medio de la ecuación 5.12, la razón de posibilidades de la muestra es

$$\widehat{OR} = \frac{b}{c} = \frac{414}{390} = 1.062.$$

Como los investigadores están intentando determinar si la razón de probabilidad es menor que 1.1, la prueba será de una cola con la hipótesis nula de equivalencia

$$H_{0E} : OR \geq 1.1$$

y la hipótesis alternativa

$$H_{AE} : OR < 1.1$$

Con el fin de realizar la prueba uno, primero usamos la ecuación 5.13 para convertir el valor *OR* nulo de 1.1 a π. Esto está dado por

$$\pi_0 = \frac{OR}{1+OR} = \frac{1.1}{1.0+1.1} = .524.$$

De igual forma,

$$\hat{p} = \frac{\widehat{OR}}{1+\widehat{OR}} = \frac{1.062}{1.0+1.062} = .515.$$

Con estas conversiones, ahora estamos listos para poner a prueba la hipótesis nula $H_0 : OR = 1.1$ contra la alternativa $H_A : OR < 1.1$ contrastando $H_0 : \pi = .524$ con la alternativa $H_A : \pi < .524$. Como usted bien sabe, esta prueba puede llevarse a cabo mediante una prueba Z para proporciones, como la que permite la ecuación 4.9.

Si se realiza la prueba uno mediante la ecuación 4.9 de la página 115 tenemos que

$$Z_1 = \frac{\hat{p}-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{.515-.524}{\sqrt{\frac{.524(1-.524)}{804}}} = -.51.$$

El valor $n = 804$ se obtiene con $b + c = 414 + 390 = 804$. Como el valor Z obtenido de $-.51$ es mayor que el valor Z crítico de -1.65, la prueba no es significativa. Por consiguiente, los investigadores no pudieron demostrar que el incremento del riesgo estaba en un rango aceptable (es decir, menor que 1.1). ■

EJEMPLO 5.27

Utilice los datos en la siguiente tabla para calcular $\widehat{OR}$. Después, realice una prueba exacta de equivalencia de dos colas usando el *EI* de .833 a 1.2. Interprete sus resultados.

		Variable dos	
		+	−
Variable uno	+	13	8
	−	7	19

Solución Por medio de la ecuación 5.12,

$$\widehat{OR} = \frac{b}{c} = \frac{8}{7} = 1.143.$$

Ya que las razones de probabilidad de muestras apareadas pueden transformarse en proporciones, los métodos para llevar a cabo pruebas exactas de equivalencia explicados en las secciones 4.3.6 y 5.3.2 son útiles para realizar pruebas exactas de equivalencia para la razón de probabilidad de muestras apareadas.

Primero transformamos el intervalo de equivalencia expresado como razones de probabilidad en un intervalo de equivalencia expresado como una proporción. Así, por medio de la ecuación 5.13, los límites de equivalencia superior e inferior se convierten en

$$I_U = \frac{\widehat{OR}}{1+\widehat{OR}} = \frac{1.2}{1.0+1.2} = .545$$

TABLA 5.24: Distribuciones muestrales de $\hat{p}$ para $n = 15$ y $\pi = .545$, y .454.

Proporción $\hat{p}$	Número de éxitos y	$\pi = .545$ $P(y)$	$\pi = .454$ $P(y)$
.000	0	.00001	.00011
.067	1	.00013	.00143
.133	2	.00112	.00829
.200	3	.00580	.02989
.267	4	.02084	.07455
.333	5	.05491	.13638
.400	6	.10962	.18900
.467	7	.16882	.20206
.533	8	.20221	.16801
.600	9	.18838	.10866
.667	10	.13539	.05421
.733	11	.07371	.02049
.800	12	.02943	.00568
.867	13	.00814	.00109
.933	14	.00139	.00013
1.000	15	.00011	.00001

y

$$I_L = \frac{\widehat{OR}}{1+\widehat{OR}} = \frac{.833}{1.0+.833} = .454.$$

Por medio de la ecuación 5.14,

$$\hat{p} = \frac{1.143}{1.0+1.143} = .533.$$

Note que este resultado puede obtenerse de manera más simple como

$$\hat{p} = \frac{b}{b+c} = \frac{8}{8+7} = .533.$$

Con el fin de llevar a cabo las versiones exactas de las pruebas uno y dos, será necesario generar las distribuciones exactas de $\hat{p}$ bajo las condiciones $\pi = .545$ y $\pi = .454$ donde $n = 15$. Esto se realiza mediante la ecuación 4.5 de la página 81. Estas distribuciones se presentan en la tabla 5.24.

Observando que $\hat{p} = .533$, el valor-p para la prueba uno se obtiene calculando la probabilidad de que $\hat{p}$ tome un valor de .533 o menor, con la condición $\pi = .545$ y $n = 15$. Este valor se calcula como .20221 + .16882 + .10962 + .05491 + .02084 + .00580 + .00112 + .00013 + .00001 = .56346. El valor-p para la prueba dos se obtiene calculando la probabilidad de que $\hat{p}$ tome un valor de .533 o mayor, con la condición $\pi = .454$ y $n = 15$. Este valor se calcula como .16801 + .10866 + .05421 + .02049 + .00568 + .00109 + .00013 + .00001 = .35828. Para rechazar la hipótesis nula de equivalencia, estas dos pruebas deben ser significativas. Ya que ninguna de las pruebas cumple con el criterio de .05, la hipótesis nula no se rechaza y la equivalencia no se demuestra. ■

5.5.4 Intervalo de confianza para una razón de posibilidades de muestras apareadas

Razonamiento. Anteriormente indicamos (sección 4.5) que los intervalos de confianza generalmente son preferibles a las pruebas de hipótesis. Así, en lugar de preguntar: "¿La razón de probabilidad es diferente de uno?", una pregunta generalmente más informativa es: "¿Cuál es la razón de probabilidad de la población?". Existen los métodos aproximado y exacto para construir intervalos de confianza para *OR*. Presentaremos cada uno en el siguiente apartado.

El intervalo de confianza.

Método aproximado

Existe una serie de métodos para formar intervalos de confianza aproximados para la razón de probabilidad de muestras apareadas. Aquí describiremos brevemente dos de ellos.

El primero usa las ecuaciones 4.16 y 4.17 (repetidas aquí con algunas modificaciones como las ecuaciones 5.15 y 5.16) para formar un intervalo de confianza para π. La relación entre la razón de probabilidad de muestras apareadas y la proporción se usa entonces para convertir los extremos de este intervalo (es decir, L y U) en un intervalo para el cálculo de la razón de probabilidad de la población. Esta conversión se lleva a cabo mediante la ecuación 5.17.

$$\boxed{L = \hat{p} - Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \tag{5.15}$$

$$\boxed{U = \hat{p} + Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \tag{5.16}$$

$$\boxed{\widehat{OR} = \frac{\hat{p}}{1-\hat{p}}} \tag{5.17}$$

donde $n = b + c$.

EJEMPLO 5.28

Utilice los datos de la tabla 5.25 de la página siguiente para calcular $\widehat{OR}$. Forme un intervalo de confianza bilateral de 95% para estimar *OR*.

Solución Por medio de la ecuación 5.12,

$$\widehat{OR} = \frac{b}{c} = \frac{13}{11} = 1.182.$$

Por medio de la ecuación 5.14,

$$\hat{p}\frac{\widehat{OR}}{1+\widehat{OR}} = \frac{1.182}{1+1.182} = .542.$$

TABLA 5.25: Datos de un estudio de control de caso que relaciona el cáncer de vejiga con la exposición a cierto solvente.

		Controles	
		Expuestos	No expuestos
Casos	Expuestos	19	13
	No expuestos	11	19

Entonces, por medio de la ecuación 5.16,

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .542 + 1.96\sqrt{\frac{.542(1-.542)}{24}} = .741$$

y por la ecuación 5.15

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .542 - 1.96\sqrt{\frac{.542(1-.542)}{24}} = .343.$$

Pero .343 y .741 representan un intervalo de confianza para la estimación de π. Se puede obtener un intervalo de confianza para el cálculo de *OR* convirtiendo estos extremos en expresiones de razones de probabilidad. Esto se logra mediante la ecuación 5.17:

$$U = \frac{\hat{p}}{1-\hat{p}} = \frac{.741}{1-.741} = 2.861$$

y

$$L = \frac{\hat{p}}{1-\hat{p}} = \frac{.343}{1-.343} = .522.$$

Así, el intervalo de confianza bilateral del 95% para el cálculo de *OR* va de .522 a 2.861.

Un segundo método usado comúnmente emplea las ecuaciones 5.18 y 5.19.

$$\boxed{L = \exp\left(\ln(\widehat{OR}) - Z\sqrt{\frac{1}{b}+\frac{1}{c}}\right)} \tag{5.18}$$

$$\boxed{U = \exp\left(\ln(\widehat{OR}) + Z\sqrt{\frac{1}{b}+\frac{1}{c}}\right)} \tag{5.19}$$

Los términos *b* y *c* son las frecuencias de las celdas, como se describe en la tabla 5.19 o 5.20. ■

EJEMPLO 5.29

Utilice las ecuaciones 5.18 y 5.19 con los datos de la tabla 5.25 para formar un intervalo de confianza bilateral del 95% para la estimación de *OR*.

Solución Como se calculó previamente, $\widehat{OR} = 1.182$. De la tabla 5.25, b y c son, respectivamente, 13 y 11. Entonces, por medio de las ecuaciones 5.19 y 5.18

$$U = \exp\left(\ln(\widehat{OR}) + Z\sqrt{\frac{1}{b}+\frac{1}{c}}\right) = \exp\left(\ln(1.182) + 1.96\sqrt{\frac{1}{13}+\frac{1}{11}}\right) = 2.638$$

y

$$L = \exp\left(\ln(\widehat{OR}) - Z\sqrt{\frac{1}{b}+\frac{1}{c}}\right) = \exp\left(\ln(1.182) - 1.96\sqrt{\frac{1}{13}+\frac{1}{11}}\right) = .530.$$

Como se puede ver, éste es un intervalo bastante amplio. ¿Qué habría hecho más corto este intervalo? ■

Método exacto

Con anterioridad usted aprendió a construir un intervalo de confianza aproximado para el cálculo de la *RR*, primero empleando un método conocido para el cálculo de π y, después, convirtiendo los extremos del intervalo resultante en expresiones que dan lugar para el cálculo del *RR*. El mismo método se puede emplear para formar un intervalo de confianza exacto para el cálculo de la *OR*.

En la página 149[24] usted empleó las ecuaciones 4.18, 4.19, 4.20, 4.21, 4.22 y 4.23 para formar intervalos de confianza exactos para el cálculo de π. Aquí presentamos estas ecuaciones con un ligero cambio en la notación para reflejar el estilo de notación de la tabla de contingencia que estamos usando en esta sección. Los extremos del intervalo de confianza obtenidos mediante el uso de estas ecuaciones se pueden convertir, a través de la ecuación 5.17 de la página 208, para lograr un cálculo aproximado de *OR*.

$$\boxed{L = \frac{b}{b+(c+1)F_L}} \tag{5.20}$$

$$\boxed{U = \frac{(b+1)F_U}{c+(b+1)F_U}} \tag{5.21}$$

En estas ecuaciones, los términos b y c son las frecuencias de las celdas, como se ilustra en las tablas 5.19 o 5.20 de la página 201, y F_L y F_U son los valores apropiados de una distribución F, los cuales se pueden obtener del Apéndice C. Observe que, a diferencia de la tabla t en el Apéndice B, la tabla F se debe consultar con *dos* diferentes grados de libertad. El primero de ellos, que denominaremos grados de libertad del numerador, se lista en la parte superior de la tabla, en tanto que el segundo, al cual denominaremos grados de libertad del denominador, se presenta de arriba abajo, en la orilla izquierda de la tabla. Así, por ejemplo, el valor F apropiado que se debe usar en el cálculo de un intervalo de confianza bilateral del 95%, suponiendo que los grados de libertad del numerador son cuatro y los grados de libertad del denominador son 20, sería 3.51. Para facilitar los cálculos, utilizaremos la notación df_{LN}, df_{LD}, df_{UN} y df_{UD} para representar, respectivamente, los grados de libertad (*degrees of freedom*) del numerador para calcular L, los grados de

[24] Quizá quiera repasar este apartado antes de continuar.

libertad del denominador para calcular L, los grados de libertad del numerador para calcular U, y los grados de libertad del denominador para calcular U.

$$\boxed{df_{LN} = 2(c+1)} \tag{5.22}$$

$$\boxed{df_{LD} = 2b} \tag{5.23}$$

$$\boxed{df_{UN} = 2(b+1)} \tag{5.24}$$

$$\boxed{df_{UN} = 2c} \tag{5.25}$$

EJEMPLO 5.30

Utilice los datos de la tabla 5.25 para formar un intervalo de confianza bilateral del 95% para el cálculo de *OR*.

Solución Para obtener F_L primero debemos calcular df_{LN} y df_{LD}, que son, respectivamente,

$$df_{LN} = 2(c+1) = 2(11+1) = 24$$

y

$$df_{LD} = 2b = (2)(13) = 26.$$

Aunque no se muestra en al Apéndice C, el valor F apropiado para un intervalo de confianza bilateral del 95% con 24 grados de libertad de numerador y 26 de denominador, es 2.22. De la ecuación 5.20 se deduce que

$$L = \frac{b}{b+(c+1)F_L} = \frac{13}{13+(11+1)2.22} = .328$$

Para obtener F_U primero debemos calcular df_{UN} y df_{UD}, que son respectivamente,

$$df_{UN} = 2(b+1) = 2(13+1) = 28$$

y

$$df_{UD} = 2c = (2)(11) = 22.$$

Aunque no se muestra en el Apéndice C, el valor F apropiado para un intervalo de confianza bilateral del 95% con 28 grados de libertad de numerador y 22 de denominador, es 2.29. De la ecuación 5.21 se deduce que

$$U = \frac{(b+1)F_U}{c+(b+1)F_U} = \frac{(13+1)2.29}{11+(13+1)2.29} = .745$$

El intervalo de confianza calculado hasta ahora, $L = .328$ y $U = .745$, estima π. Para construir un intervalo para el cálculo de *OR* usamos la ecuación 5.17 para convertir los extremos de este intervalo en expresiones de la razón de probabilidad. Así, obtenemos

$$L = \frac{\hat{p}}{1-\hat{p}} = \frac{.328}{1-.328} = .488$$

y

$$U = \frac{.745}{1-.745} = 2.922.$$

Los intervalos de confianza obtenidos mediante los dos métodos aproximados fueron .522 a 2.861, y .530 a 2.638, que son bastante cercanos. El método exacto produjo un intervalo de .488 a 2.922. Mientras que en la literatura científica se emplean mucho los métodos aproximados, el método exacto es preferible. ■

5.5.5 Suposiciones

Las suposiciones que fundamentan los métodos presentados aquí para las razones de probabilidad son en esencia las mismas que las que se expusieron en la página 116.

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

cohorte prospectiva 200
datos apareados 159
datos no informativos 174
equivalencia mediante pruebas t de muestras apareadas 165
equivalencia para la razón de probabilidad apareada 204
equivalencia para la tasa de riesgo apareada 193
equivalencia para una proporción de muestras apareadas 180
intervalo de confianza para la diferencia de medias de muestras apareadas 171
intervalo de confianza para la proporción de muestras apareadas 186
intervalo de confianza para la razón de posibilidades de muestras apareadas 208
intervalo de confianza para la tasa de riesgo apareada 196
población de diferencias de puntuaciones 165
posibilidades 199
prueba de McNemar 174
prueba t de diferencias apareadas 161
prueba t de muestras apareadas 161
puntuaciones de diferencia 160
razón de probabilidad 199
razón de probabilidad de muestras apareadas 199
riesgo 190
tasa de riesgo 190
tasa de riesgo de muestras apareadas 190

EJERCICIOS

5.1 Se evaluará la efectividad de una dieta diseñada para reducir el colesterol sérico, para lo cual se mide el nivel de colesterol de ocho pacientes identificados con riesgo de enfermedad cardiovascular, antes y después de estar sometidos a la dieta por 16 semanas. Utilice los siguientes datos para

a) realizar una prueba t de dos colas de muestras apareadas con $\alpha = .05$, y

b) construir un intervalo de confianza bilateral del 95% para el cálculo de μ_d.

c) ¿A qué conclusión llega sobre la efectividad de la dieta?

Paciente	Antes de la dieta	Después de la dieta
1	213	199
2	252	241
3	195	197
4	222	220
5	267	248
6	216	224
7	209	209
8	255	237

5.2 Se realiza un estudio en el que un método invasivo estándar (por ejemplo, extracción de sangre) para supervisar la glucosa en pacientes diabéticos se compara con un método no invasivo (por ejemplo, sin extracción de sangre) recién desarrollado, con el fin de determinar si los dos sistemas producen resultados equivalentes. Para ello, se emplean ambos aparatos de forma simultánea para evaluar los niveles de glucosa en 10 pacientes diabéticos. Los dos aparatos se considerarán equivalentes si la diferencia de medias de los niveles de glucosa producidos por los dos métodos se encuentra entre ±4 puntos. Las dos mediciones de cada paciente aparecen a continuación.

Invasivo	No invasivo
140	144
84	82
200	200
249	247
71	64
131	138
140	132
122	123
139	146
119	117

a) ¿Cuál prueba de equivalencia sería apropiada para este estudio, la de una cola o la de dos colas? ¿Por qué?

b) Plantee las hipótesis nula y alternativa de equivalencia para el estudio.

c) Realice la prueba de equivalencia a un nivel $\alpha = .05$ y reporte el resultado.

d) ¿A qué conclusión llega sobre la equivalencia de los dos dispositivos?

5.3 Se solicita a pacientes con quemaduras faciales que califiquen su apariencia general como "satisfactoria" (S) o "insatisfactoria" (I) antes y después de someterse a un régimen de tratamiento de reducción de cicatrices.

a) Suponga que cinco pacientes califican su apariencia como I antes del tratamiento y S después (I-S), dos califican su apariencia como S antes e I después del tratamiento (S-I), mientras que uno califica su apariencia como S antes y S después del tratamiento (S-S), y otro más califica I antes e I después del tratamiento. (I-I).

i. Utilice una versión exacta de dos colas de la prueba de McNemar con $\alpha = .05$ para determinar si el tratamiento es efectivo para mejorar la percepción que tiene el paciente de su apariencia.

ii. Utilice un intervalo de confianza bilateral exacto del 95% para calcular la proporción de pacientes que perciben que su apariencia ha mejorado.

b) Dados I-S = 71, I-I = 22, S-S = 11, y S-I = 20,

i. Utilice una versión aproximada de dos colas de la prueba de McNemar con $\alpha = .05$ para determinar si el tratamiento es efectivo para mejorar la percepción que tiene el paciente de su apariencia.

ii. Utilice un intervalo de confianza bilateral aproximado del 95% para calcular la proporción de pacientes que perciben que su apariencia ha mejorado.

5.4 A los oficiales de patrulla de una fuerza de policía metropolitana de gran tamaño se les asignan motocicletas (control del tránsito) o autos patrulla (no control del tránsito). Suponga que los oficiales de cada categoría forman pares con respecto a su tiempo de servicio, edad y género. Luego se examina a cada par para ver si presentan cáncer de piel en la cara, el cuello o las manos. La tabla provista a continuación muestra los resultados de los pares de oficiales. La categoría "cáncer" indica la presencia de uno o más tipos de cáncer de piel en las áreas examinadas.

		Auto patrulla	
		cáncer	sin cáncer
Motocicleta	cáncer	10	34
	sin cáncer	18	444

a) Calcule la tasa de riesgo para comparar el riesgo de cáncer de piel de los oficiales en motocicleta con el de los oficiales en autos patrulla. ¿Qué significado tiene este cociente?

b) Construya un intervalo de confianza del 95% para el cálculo de *RR*.

5.5 Suponga que el estudio de cáncer de piel del ejercicio 5.4 hubiera recabado datos de manera retrospectiva. Esto es, suponga que los oficiales con cáncer de piel hubieran formado pares con oficiales sin cáncer de piel con respecto a su tiempo de servicio, edad y género. Luego, los pares de oficiales se clasifican dependiendo de si fueron asignados a motocicletas o autos patrulla. Estos datos de control de caso se muestran a continuación.

		Sin cáncer de piel	
		motocicleta	auto patrulla
Cáncer de piel	motocicleta	10	24
	auto patrulla	9	444

a) Calcule la razón de probabilidad para comparar las posibilidades de ser un oficial de motocicleta con cáncer de piel y sin cáncer de piel. ¿Qué significado tiene esta razón de probabilidad?

b) Construya intervalos de confianza del 95% para el cálculo de *OR* mediante dos métodos aproximados diferentes. ¿Qué tanto coinciden los intervalos?

c) Utilice un método exacto para construir un intervalo de confianza del 95% para el cálculo de *OR*. ¿Cómo se compara este resultado con los dos resultados aproximados?

5.6 Utilice los datos de la tabla provista a continuación para calcular $\widehat{OR}$; después, realice la siguiente prueba. Interprete el resultado.

H_{0E} : $OR \leq .80$, o bien, $OR \geq 1.25$
H_{AE} : $.8 < OR < 1.25$

		No enfermedad	
		Expuestos	No expuesto
Enfermedad	Expuesto	10	24
	No expuesto	9	444

A. Las siguientes preguntas se refieren al estudio de caso A (página 469).

5.7 Los autores reportan que "Un total de 16 sujetos cambiaron su preferencia entre el inicio y el fin del uso, de los cuales ocho que cambiaron su preferencia de los lentes de control a los lentes acondicionados, formaron pares de forma exacta con los otros ocho que cambiaron exactamente de manera opuesta...". Después reportan que el resultado de una prueba estadística específica no fue significativo.

¿Qué prueba cree usted que realizaron? Haga la misma prueba. ¿Coincide usted con el resultado que ellos obtuvieron? ¿Cuál es su interpretación de este resultado?

B. Las siguientes preguntas se refieren al estudio de caso B (página 470).

5.8 Suponga que los investigadores quisieran calcular el cambio promedio en las puntuaciones de la escala A de WOMAC desde la línea base hasta la semana 12. ¿Qué técnica se podría usar para este cálculo? Especifique las ecuaciones que se deben usar, si es que las hay. ¿Existe suficiente información en el estudio de caso para llevar a cabo la técnica que usted recomienda? Explique.

K. Las siguientes preguntas se refieren al estudio de caso K (página 474).

5.9 Los autores realizaron una prueba *t* de muestras apareadas con los datos reportados en la tabla J.6. Surge la pregunta de por qué usaron las medias de las escuelas para este propósito en lugar de las puntuaciones apareadas de niños individuales. Después de todo, usar puntuaciones de estudiantes individuales daría muchos más grados de libertad. Hubo dos razones. Primero, dificultades logísticas y confidenciales hicieron que el emparejamiento de datos de los estudiantes fuera complicado. La segunda razón se basó en una consideración estrictamente estadística. ¿Cuál cree usted que fue la base de esta consideración? (Pista: véase la página 99 para encontrar información relacionada con las presiones arteriales juveniles).

5.10 Realice una prueba *t* de muestras apareadas de dos colas a un nivel $\alpha = .01$ de las medias de las escuelas. Interprete el resultado.

5.11 ¿Puede usted imaginar una forma más informativa para analizar estos datos? Realice el análisis que usted propone. Explique por qué este resultado es más informativo.

M. Las siguientes preguntas se refieren al estudio de caso M (página 475).

5.12 Calcule el cambio promedio en el nivel de oxígeno que ocurrió después de tomar el medicamento para dormir.

5.13 Calcule el cambio promedio en el nivel de dióxido de carbono que ocurrió después de tomar el medicamento para dormir.

CAPÍTULO 6

Métodos para dos muestras independientes

6.1 INTRODUCCIÓN

No siempre resulta práctico recabar datos con las configuraciones de muestras apareadas que se vieron en el capítulo 5. Por ejemplo, es posible que un investigador desee agrupar los datos de varios individuos por peso, edad y género, pero que se encuentre con que hay algunos sujetos en los dos grupos existentes que son lo suficientemente similares respecto de estos atributos como para que esta estrategia sea viable. El mismo problema surgiría en estudios en los que se agrupan sujetos al azar mediante una designación de condiciones de tratamiento alternas. Como resultado, la mayoría de los datos recabados en contextos de investigación es dispar.

Consideremos un estudio en el cual se debe comparar la eficacia de dos medicamentos diseñados para el tratamiento de la hipertensión. Los investigadores podrían descubrir que no es posible formar pares de sujetos lo bastante similares como para ser tratados como muestras apareadas. En circunstancias como ésta, una estrategia simple sería asignar a los sujetos a los dos tratamientos al azar sin considerar las características inherentes a ellos (es decir, peso, condiciones médicas preexistentes, etcétera). El proceso de asignación aleatoria garantiza que ningún sesgo sistemático interfiera en la formación de los dos grupos.

Los métodos que estudiamos en el capítulo 5 no son apropiados para el análisis de datos dispares. Es por ello que este capítulo maneja pruebas de hipótesis e intervalos de confianza diseñados para comparar medias o proporciones de datos dispares.

6.2 MÉTODOS RELACIONADOS CON LAS DIFERENCIAS ENTRE MEDIAS

6.2.1 La prueba *t* de muestras independientes

Razonamiento. Como se mencionó con anterioridad, suponga que se debe hacer un estudio para determinar si un medicamento es más eficaz que otro para el tratamiento de la hipertensión. Con este fin, los investigadores asignan al azar a 30 sujetos hipertensos a uno de los dos grupos de tratamiento. Los sujetos asignados al grupo 1 reciben un régimen de tratamiento basado en el primer fármaco y los asignados al grupo 2 reciben un tratamiento que se basa en el

TABLA 6.1: Las presiones arteriales sistólicas de los pacientes hipertensos después del tratamiento con uno de los dos regímenes de medicamentos.

Grupo 1	Grupo 2
129	138
111	120
140	137
139	154
144	148
120	122
131	131
129	128
131	140
154	145
119	131
138	120
142	144
110	129
140	141
$\bar{x}_1 = 131.80$	$\bar{x}_2 = 135.20$

segundo. La pregunta que nos atañe es: "¿Los dos regímenes de tratamiento producen niveles de presión arterial diferentes en los pacientes hipertensos?" La tabla 6.1 muestra las presiones arteriales sistólicas (ficticias) de 30 sujetos, tomadas después de recibir tratamiento mediante los dos regímenes de medicamentos.

Una estrategia razonable para contestar la pregunta sobre si los dos tratamientos producen niveles diferentes de presión arterial sería calcular las medias de las presiones para cada uno de los dos grupos, y luego compararlas para determinar si difieren. En este caso, los sujetos que tomaron la sustancia 1 tuvieron una presión promedio de 131.80, en tanto que los sujetos que tomaron la sustancia 2 tuvieron una presión promedio de 135.20. Así que, la presión arterial promedio del primer grupo fue de 3.4 unidades menos que la del segundo grupo.

Pero existen por lo menos dos explicaciones para esta diferencia. Primero, notamos que cada vez que asignamos sujetos a los grupos al azar, las medias de los dos grupos con seguridad diferirán en algún grado en casi todas las medidas imaginables. De hecho, nos sorprendería encontrar que la estatura promedio de estos dos grupos fue de 68.4 pulgadas para ambos, o que los dos pesos promedio fueron exactamente de 151.38 libras.

Entonces, aun si los dos tratamientos tuvieran igual efecto sobre la variable de interés, esperaríamos ver cierta diferencia en los promedios de los dos grupos. La primera explicación para la diferencia de 3.4 unidades que se observa en nuestro estudio ficticio es, entonces, que surgió no por alguna diferencia en el efecto de los medicamentos, sino por la casualidad relacionada con la manera en la que se formaron los grupos.

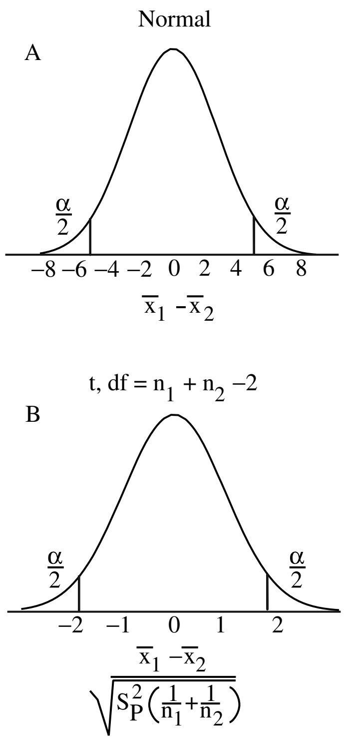


FIGURA 6.1: Distribución de $\bar{x}_1 - \bar{x}_2$ (panel A) y un estadístico *t* con $n_1 + n_2 - 2$ grados de libertad (panel B).

Una segunda explicación es que el medicamento administrado al grupo 1 es más efectivo al reducir las presiones de los pacientes hipertensos que la sustancia administrada al segundo grupo. Esto explica por qué las presiones en el grupo 1 tienden a ser más bajas que en el grupo 2. Pero, ¿cuál de estas explicaciones debe creerse? Una prueba de significancia podría ser útil para decidir el asunto.

Piense en una distribución muestral formada por la repetida selección de una muestra aleatoria de cierta población, dividiendo al azar cada muestra en dos componentes muestra y calculando la cantidad $\bar{x}_1 - \bar{x}_2$ donde $\bar{x}_1$ y $\bar{x}_2$ son las medias de las dos muestras componentes.[1] Si este proceso se repitiera muchas veces, los valores resultantes de $\bar{x}_1 - \bar{x}_2$ conformarían una distribución muestral, como se presenta en el panel A de la figura 6.1. Es importante comprender que esta distribución representa la distribución muestral de $\bar{x}_1 - \bar{x}_2$ *cuando ambas muestras provienen de poblaciones con la misma media*. Es claro que si ambas muestras provienen de la misma población, entonces, vienen de poblaciones con la misma media. Designaremos la media de la población de la cual se tomó la primera muestra como μ_1 y la media de la población de

[1] De forma equivalente, podemos pensar en las dos muestras componentes sólo como dos muestras aleatorias tomadas de una población común.

la cual se tomó la segunda muestra como μ_2. Como ambas muestras provienen de la misma población, es claro que $\mu_1 = \mu_2$.

Observe que la media de esta distribución muestral es cero y que cuanto más lejanos estén los valores de $\bar{x}_1 - \bar{x}_2$ de cero, menos probable es su ocurrencia. Pero, ¿cómo podría usarse esta distribución para ayudar a contestar la pregunta formulada con anterioridad? Es decir, ¿la diferencia de 3.4 puntos entre las medias de los dos grupos tratados con los dos diferentes medicamentos para la hipertensión se debió a la casualidad aleatoria conforme se formaron los grupos, o la sustancia administrada al grupo 1 fue más efectiva que la que se administró al grupo 2?

Las dos muestras se tomaron originalmente de una población común, por lo que $\mu_1 = \mu_2$. Si las dos sustancias produjeran exactamente el mismo resultado en las dos muestras, entonces sería cierto que $\mu_1 = \mu_2$. En tales circunstancias no esperaríamos que $\bar{x}_1 - \bar{x}_2$ estuviera en una región crítica. ¿Por qué? Porque la probabilidad de que $\bar{x}_1 - \bar{x}_2$ esté en una región crítica cuando $\mu_1 = \mu_2$ es solamente α.

Pero, ¿qué pasaría si la sustancia 1 fuera más efectiva que la sustancia 2? En este caso, las presiones en el grupo 1 se reducirían al tomar el fármaco, por lo que el valor absoluto de $\bar{x}_1 - \bar{x}_2$ tendería a ser mayor que en el caso donde ambas sustancias tuvieran el mismo efecto. Por ejemplo, en nuestro estudio hipotético de hipertensión, $\bar{x}_1 - \bar{x}_2 = 131.8 - 135.2 = -3.4$. Pero, suponga que la sustancia 1 hubiera sido más efectiva como para que $\bar{x}_1 = 120.4$. Entonces, $\bar{x}_1 - \bar{x}_2 = 120.4 - 135.2 = -14.8$. Cuanto más efectiva sea la sustancia 1, en comparación con la sustancia 2, más abajo estará $\bar{x}_1 - \bar{x}_2$ de la media de la distribución, cuyo valor es cero, y la oportunidad de que $\bar{x}_1 - \bar{x}_2$ se localice en la región crítica inferior será mayor. ¿Qué pasaría si la sustancia 2 fuera la más efectiva para la reducción de la presión arterial?

En resumen, no esperamos que $\bar{x}_1 - \bar{x}_2$ caiga en la región crítica cuando los tratamientos producen el mismo resultado. De hecho, la probabilidad de que esto ocurra es solamente α. Sin embargo, cuando los tratamientos producen diferentes resultados, esperamos que el valor de $\bar{x}_1 - \bar{x}_2$ se aleje del cero y se acerque a una región crítica. Aún más, cuando difieren los efectos del tratamiento, no podemos sostener que $\mu_1 = \mu_2$. Más bien, se debe decir que $\mu_1 \neq \mu_2$.

De lo anterior, cuando $\bar{x}_1 - \bar{x}_2$ cae en la región crítica, podemos decir que la diferencia observada entre estos valores se debió a efectos diferenciales del tratamiento y no a la casualidad de cómo se formaron los grupos. Ahora vemos cómo se realiza esta prueba.

La prueba. La hipótesis nula que debe analizarse mediante la prueba *t* de muestras independientes es a menudo (pero no siempre)

$$H_0 : \mu_1 = \mu_2$$

o de manera equivalente,

$$H_0 : \mu_1 - \mu_2 = 0$$

La alternativa de dos colas es a menudo (pero no siempre)

$$H_A : \mu_1 \neq \mu_2$$

o de manera equivalente,

$$H_A : \mu_1 - \mu_2 \neq 0$$

Las alternativas de una cola son a menudo (pero no siempre) de la forma

$$H_A : \mu_1 < \mu_2$$

o de manera equivalente,

$$H_A : \mu_1 - \mu_2 < 0$$

o

$$H_A : \mu_1 > \mu_2$$

o de manera equivalente,

$$H_A : \mu_1 - \mu_2 > 0$$

En esencia, la hipótesis nula afirma que no hubo diferencia en los efectos de los tratamientos administrados en los dos grupos, en tanto que la alternativa de dos colas sostiene que sí existe una diferencia. Las alternativas de una cola especifican la forma de esa diferencia.

Para poder llevar a cabo la prueba de hipótesis, $\bar{x}_1 - \bar{x}_2$ no puede referirse directamente a la distribución muestral que aparece en el panel A de la figura 6.1 de la página 217, sino que debe ser dividida entre un estimado (o entre una estimación) del error estándar (véase el denominador de la ecuación 6.1), lo que significa que el estadístico de prueba se debe referir a una distribución *t* como en el panel B de la figura 6.1. A diferencia de la prueba *t* de una media que usted ya conoce, los grados de libertad para la prueba *t* de muestras independientes son $n_1 + n_2 - 2$. El estadístico de prueba se calcula mediante la ecuación 6.1.

$$\boxed{t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}} \tag{6.1}$$

En la ecuación 6.1, $\bar{x}_1$ y $\bar{x}_2$ representan, respectivamente, las medias de la muestra 1 y de la muestra 2, en tanto que, n_1 y n_2 representan el número de observaciones en cada una de las dos muestras. El símbolo δ_0 representa la diferencia que se plantea como hipótesis entre μ_1 y μ_2, que normalmente es cero. No obstante, como usted verá en el análisis de pruebas de equivalencia basadas en la prueba *t* de muestras independientes, éste no siempre es el caso.

El término s_P^2 requiere explicación. Usted recordará que la prueba *t* de una media utilizó un estimado (o una estimación) de la desviación estándar de la población de la cual se tomó la muestra. Usamos la desviación estándar muestral (s) para este fin. En este caso, necesitamos estimar la varianza de la población de la que se tomaron las muestras. Podríamos usar la varianza muestra (s^2) de una de las dos muestras para este fin, pero sería inútil, ya que emplea una estimación basada en una sola de las dos muestras. Una mejor estrategia sería basar la estimación de la varianza en *ambas* muestras. Esto se puede hacer mediante una forma particular de promediar que los estadísticos denominan **agrupamiento** (de ahí el subíndice *P*, por *pooling*). Así, s_P^2 es una estimación de la varianza de la población basada en un promedio o agrupamiento de la información en las dos muestras. A continuación podemos ver que se utilizan las dos varianzas muestrales.

$$s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

s_1^2 y s_2^2 representan las varianzas de la primera y segunda muestra, respectivamente.

El numerador de esta expresión es simplemente la suma de las sumas de los cuadrados de las dos muestras. Por lo tanto, la ecuación 6.2 nos da un método más eficiente para calcular s_P^2.

Hemos colocado paréntesis en el numerador para ayudarle a identificar las expresiones de las dos sumas de cuadrados. Los subíndices 1 y 2 se usan para identificar las dos muestras.

$$\boxed{s_P^2 = \frac{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}\right) + \left(\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}\right)}{n_1 + n_2 - 2}} \tag{6.2}$$

EJEMPLO 6.1

Utilice los datos de la tabla 6.1 de la página 216 para realizar una prueba t de dos colas de muestras independientes con $\alpha = .05$. Interprete el resultado referente a las dos terapias de medicamentos.

Solución Para fines de cálculo será conveniente acomodar los datos como aparecen en la tabla 6.2 de la siguiente página.

Utilizando las cantidades de la tabla 6.2,

$$\bar{x}_1 = \frac{1977}{15} = 131.8$$

$$\bar{x}_2 = \frac{2028}{15} = 135.2$$

y

$$\begin{aligned} s_P^2 &= \frac{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}\right) + \left(\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}\right)}{n_1 + n_2 - 2} \\ &= \frac{\left(262767 - \frac{(1977)^2}{15}\right) + \left(275706 - \frac{(2028)^2}{15}\right)}{15 + 15 - 2} \\ &= \frac{(2198.4) + (1520.4)}{28} \\ &= 132.814 \end{aligned}$$

Así, nuestra mejor estimación de la varianza de la población de la que se tomaron las dos muestras es 132.814. Por la ecuación 6.1 de la página anterior, t obtenida es

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\ &= \frac{131.8 - 135.2}{\sqrt{132.814\left(\frac{1}{15} + \frac{1}{15}\right)}} \\ &= \frac{-3.4}{4.208} \\ &= -.808 \end{aligned}$$

TABLA 6.2: Presiones arteriales ordenadas para análisis mediante la prueba *t* de muestras independientes.

	Grupo 1		Grupo 2	
	X_1	X_1^2	X_2	X_2^2
	129	16641	138	19044
	111	12321	120	14400
	140	19600	137	18769
	139	19321	154	23716
	144	20736	148	21904
	120	14400	122	14884
	131	17161	131	17161
	129	16641	128	16384
	131	17161	140	19600
	154	23716	145	21025
	119	14161	131	17161
	138	19044	120	14400
	142	20164	144	20736
	110	12100	129	16641
	140	19600	141	19881
Σ	1977	262767	2028	275706

El Apéndice B muestra que los valores críticos para una prueba *t* de dos colas con

$$n_1 + n_2 - 2 = 15 + 15 - 2 = 28$$

grados de libertad y realizada con $\alpha = .05$ son -2.048 y 2.048, por lo que la hipótesis nula no se rechaza. Entonces, ¿podemos concluir que las dos sustancias no difirieron en su efecto sobre la presión arterial? ¡No! Una vez más, no es posible utilizar el no rechazo de la hipótesis nula como evidencia de que la hipótesis nula es cierta. ¿Por qué no? Porque usted no conoce la probabilidad de un error tipo II (beta). En esta situación, debemos decir que no encontramos suficiente evidencia de que haya un efecto diferencial del medicamento para llegar a una conclusión así. ■

EJEMPLO 6.2

Un grupo de trabajadores expuestos a un moho tóxico en el sistema de ventilación del edificio en el que trabajaron se someten a una prueba para evaluar los síntomas experimentados antes del descubrimiento del moho. La escala clasifica la experiencia de síntomas de cero (sin síntomas) a 40 (varios síntomas, algunos de los cuales fueron severos). Los investigadores creen que los hombres tienden a minimizar tales experiencias y, por lo tanto, sus puntuaciones serán significativamente más bajas que las de las mujeres.

Utilice los datos de la tabla 6.3 de la página 223 para realizar una prueba *t* de una cola de muestras independientes con $\alpha = .05$ para evaluar la teoría del investigador. Comience enunciando claramente las hipótesis nula y alternativa.

Solución Si los hombres son denominados como el grupo 1 y las mujeres como el grupo 2, entonces, las hipótesis nula y alternativa pueden enunciarse de la siguiente manera:

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 < \mu_2$$

¿Cómo se verían afectadas estas hipótesis si los hombres fueran denominados como el grupo 2 y las mujeres como el grupo 1?

Para fines de cálculo será conveniente ordenar los datos como se muestra en la tabla 6.4. Si se usan las cantidades de la tabla 6.4,

$$\bar{x}_1 = \frac{\sum x}{n} = \frac{147}{11} = 13.364$$
$$\bar{x}_2 = \frac{\sum x}{n} = \frac{372}{15} = 24.800$$

y

$$\begin{aligned} s_P^2 &= \frac{\left(\sum x_1^2 - \frac{(\Sigma x_1)^2}{n_1}\right) + \left(\sum x_2^2 - \frac{(\Sigma x_2)^2}{n_2}\right)}{n_1 + n_2 - 2} \\ &= \frac{\left(2287 - \frac{(147)^2}{11}\right) + \left(10222 - \frac{(372)^2}{15}\right)}{11 + 15 - 2} \\ &= \frac{(322.545) + (996.400)}{24} \\ &= 54.956 \end{aligned}$$

Así, nuestra mejor estimación de la varianza de la población de la que se tomaron las dos muestras es 54.956. Mediante la ecuación 6.1 de la página 219, la t obtenida es

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\ &= \frac{13.364 - 24.800}{\sqrt{54.956\left(\frac{1}{11} + \frac{1}{15}\right)}} \\ &= \frac{-11.346}{2.943} \\ &= -3.886 \end{aligned}$$

El Apéndice B nos muestra que el valor crítico para una prueba t de una cola con 11 + 15 – 2 = 24 grados de libertad para $\alpha = .05$ es −1.711, por lo que se rechaza la hipótesis nula. Por lo tanto, podemos concluir que los hombres que experimentaron el moho tóxico obtuvieron puntuaciones significativamente más bajas en la escala que las mujeres. No obstante, observe que el diseño de

TABLA 6.3: Puntuaciones en la escala de síntomas de hombres y mujeres expuestos al moho tóxico.

Hombres	Mujeres
14	28
9	20
16	22
7	31
10	13
20	10
13	32
23	29
5	30
11	9
19	38
	27
	27
	30
	26
$\bar{x}_1 = 13.364$	$\bar{x}_2 = 24.800$

TABLA 6.4: Puntuaciones en la escala de síntomas ordenadas para análisis mediante la prueba *t* de muestras independientes.

	Grupo 1		Grupo 2	
	X_1	X_1^2	X_2	X_2^2
	14	196	28	784
	9	81	20	400
	16	256	22	484
	7	49	31	961
	10	100	13	169
	20	400	10	100
	13	169	32	1024
	23	529	29	841
	5	25	30	900
	11	121	9	81
	19	361	38	1444
			27	729
			27	729
			30	900
			26	676
$\sum$	147	2287	372	10222

este estudio no permitió la asignación aleatoria de los sujetos a los grupos. De hecho, no podemos asignar sujetos a un género de forma aleatoria. Debemos ser particularmente cautelosos con la forma en que se interpretan los resultados en ausencia de una asignación al azar. En este caso, no podemos asegurar que la diferencia significativa encontrada entre los dos grupos se deba al género. Tal vez, por ejemplo, los hombres tendieron a que les asignaran trabajos en áreas del edificio donde había menos moho.

En este caso, la hipótesis alternativa fue de la forma $H_A : \mu_1 < \mu_2$, lo cual implica una región crítica en el extremo izquierdo de la distribución. Una alternativa de la forma $H_A : \mu_1 > \mu_2$ requeriría una región crítica en el extremo derecho. ■

6.2.2 Cómo establecer la equivalencia por medio de las pruebas *t* de muestras independientes

Razonamiento. Como hemos apuntado en otras ocasiones, el rechazo de la hipótesis nula representa buena evidencia (aunque no lo suficientemente positiva) de la falsedad de ésta. En contraste, el fracaso al rechazar la hipótesis nula, por lo general, no aporta buena evidencia de que la hipótesis es verdadera. En situaciones donde se desea establecer la validez de la hipótesis nula, debe emplearse una prueba de equivalencia para mostrar que la hipótesis nula es verdadera (aproximadamente). (Véase la sección 4.3.6 en la página 117).

En ocasiones, los investigadores quieren demostrar que *no* hay diferencia entre los tratamientos, en vez de demostrar que *hay* una diferencia. En semejantes casos, las pruebas *t* de muestras independientes pueden usarse para establecer la equivalencia.

Por ejemplo, una compañía farmacéutica quizá quiera demostrar que un nuevo medicamento diseñado para el tratamiento de la hipertensión es más efectivo que uno anterior. Quizás este nuevo fármaco no induce algunos efectos indeseables asociados comúnmente con el antiguo medicamento, o tal vez su manufactura es más barata. También puede ser que, según las propiedades farmacéuticas del nuevo medicamento, no sea posible que esta sustancia sea más efectiva que el tratamiento anterior.

La prueba. Usando la notación de la sección 4.3.6 de la página 117, las hipótesis nula y alternativa para una prueba de equivalencia de dos colas basada en la prueba *t* de muestras independientes se enuncian como

$$H_{0E} : \mu_1 - \mu_2 \leq EI_L \text{ o } \mu_1 - \mu_2 \geq EI_U$$
$$H_{AE} : EI_L < \mu_1 - \mu_2 < EI_U$$

donde EI_L y EI_U son los extremos inferior y superior del intervalo de equivalencia, respectivamente. En esencia, la hipótesis nula plantea que la diferencia entre las medias no reside en el intervalo de equivalencia, mientras que la alternativa afirma que la diferencia yace en el intervalo. Usted recordará que la hipótesis nula se comprueba realizando dos pruebas de una cola al nivel de α. Para poder rechazar la hipótesis nula de equivalencia, debe demostrar que $\mu_1 - \mu_2 > EI_L$ *y* que $\mu_1 - \mu_2 < EI_U$, lo que significa que *las dos* pruebas siguientes deben ser significativas.

Prueba 1	Prueba 2
$H_{01} : \mu_1 - \mu_2 = EI_U$	$H_{02} : \mu_1 - \mu_2 = EI_L$
$H_{A1} : \mu_1 - \mu_2 < EI_U$	$H_{A2} : \mu_1 - \mu_2 > EI_L$

Las hipótesis nula y alternativa para la prueba de equivalencia de una cola es *una* de las siguientes

$$H_{0E} : \mu_1 - \mu_2 \geq EI_U$$
$$H_{AE} : \mu_1 - \mu_2 < EI_U$$

o

$$H_{0E} : \mu_1 - \mu_2 \leq EI_L$$
$$H_{AE} : \mu_1 - \mu_2 > EI_L$$

La primera hipótesis de equivalencia de una cola presentada arriba se somete a la prueba 1 y la segunda a la prueba 2.

EJEMPLO 6.3

Suponga que los datos de la tabla 6.1 de la página 216 se recabaron en relación con un estudio de equivalencia diseñado para demostrar que la sustancia administrada al grupo 1 no es más efectiva que la sustancia administrada al grupo 2, en lo que se refiere al control de la hipertensión. Se decide que la sustancia 1 se declarará *no más efectiva* que la sustancia 2, si la media en el nivel de la presión arterial alcanzada por la sustancia 1 está menos de cinco unidades por debajo de la alcanzada por la sustancia 2.

Utilice la prueba t de muestras independientes para realizar una prueba de equivalencia de una cola con estos datos para $\alpha = .05$. Enuncie las hipótesis nula y alternativa antes de empezar la prueba.

Solución La hipótesis nula de equivalencia toma la forma

$$H_{0E} : \mu_1 - \mu_2 \leq -5$$

mientras que la alternativa sostiene que

$$H_{AE} : \mu_1 - \mu_2 > -5$$

La prueba se lleva a cabo mediante la prueba 2 con las hipótesis nula y alternativa

$$H_{02} : \mu_1 - \mu_2 = -5$$

y

$$H_{A2} : \mu_1 - \mu_2 > -5$$

Antes encontramos que los valores de $\bar{x}_1, \bar{x}_2$ y s_P^2 eran 131.8, 135.2 y 133.779, respectivamente. Ya que la hipótesis nula especifica una diferencia que no es cero entre μ_1 y μ_2, el término δ_0 de la ecuación 6.1 es igualmente distinto a cero y debe incluirse en el cálculo. Entonces, la t obtenida es

$$t_2 = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{131.8 - 135.2 - (-5)}{\sqrt{132.814\left(\frac{1}{15} + \frac{1}{15}\right)}} = \frac{1.6}{4.208} = .380$$

El Apéndice B muestra que el valor crítico para una prueba t de una cola con 15 + 15 – 2 = 28 grados de libertad conducida para $\alpha = .05$ es 1.701.[2] Como la t obtenida de .380 es menor que la t crítica de 1.701, la hipótesis nula no se rechaza. Por lo tanto, no se ha demostrado la equivalencia. ■

[2] Note que la t crítica es positiva porque la alternativa especifica que $\mu_1 - \mu_2$ es mayor que el valor determinado por la hipótesis nula.

EJEMPLO 6.4

Utilice los datos de la tabla 6.5 de la siguiente página para realizar una prueba de equivalencia de dos colas con $\alpha = .05$. Utilice el intervalo de equivalencia $EI_L = -4$ y $EI_U = 4$. Enuncie las hipótesis nula y alternativa.

Solución Las hipótesis nula y alternativa plantean que

$$H_{0E} : \mu_1 - \mu_2 \leq -4 \text{ o } \mu_1 - \mu_2 \geq 4$$

y

$$H_{AE} : -4 < \mu_1 - \mu_2 < 4.$$

Será conveniente para fines de cálculo ordenar los datos como se muestra en la tabla 6.6 de la siguiente página.

Usando las cantidades de la tabla 6.6,

$$\bar{x}_1 = \frac{\sum x}{n} = \frac{108}{16} = 6.750$$

y

$$\bar{x}_2 = \frac{\sum x}{n} = \frac{102}{16} = 6.375$$

$$\begin{aligned} s_P^2 &= \frac{\left(\sum x_1^2 - \frac{(\Sigma x_1)^2}{n_1}\right) + \left(\sum x_2^2 - \frac{(\Sigma x_2)^2}{n_2}\right)}{n_1 + n_2 - 2} \\ &= \frac{\left(800 - \frac{(108)^2}{16}\right) + \left(736 - \frac{(102)^2}{16}\right)}{16 + 16 - 2} \\ &= \frac{(71.00) + (85.75)}{30} \\ &= 5.225 \end{aligned}$$

Así, nuestra mejor estimación de la varianza de la población de la cual se tomaron las dos muestras es 5.225. Mediante la prueba 1 y la ecuación 6.1 de la página 219, t_1 es

$$t_1 = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{6.750 - 6.375 - 4.000}{\sqrt{5.225\left(\frac{1}{16} + \frac{1}{16}\right)}} = \frac{-3.625}{.808} = -4.486$$

El Apéndice B demuestra que el valor crítico para una prueba t de una cola con $16 + 16 - 2 = 30$ grados de libertad realizada con $\alpha = .05$ es -1.697. Se deduce que la hipótesis nula $H_0 : \mu_1 - \mu_2 = 4$ se rechaza en favor de la alternativa $H_A : \mu_1 - \mu_2 < 4$.

Mediante la prueba 2 y la ecuación 6.1, la t_2 obtenida es

$$t_2 = \frac{6.750 - 6.375 - (-4.0)}{\sqrt{5.225\left(\frac{1}{16} + \frac{1}{16}\right)}} = \frac{4.375}{.808} = 5.415$$

TABLA 6.5: Datos de práctica para la prueba de equivalencia de dos colas basada en la prueba *t* de muestras independientes.

Grupo 1	Grupo 2
8	8
9	3
5	4
7	10
4	8
4	3
10	5
8	3
3	7
10	9
9	5
7	6
6	6
5	8
6	10
7	7
$\bar{x}_1 = 6.750$	$\bar{x}_2 = 6.375$

TABLA 6.6: Datos de práctica ordenados para análisis mediante la prueba de equivalencia de dos colas basada en la prueba *t* de muestras independientes.

	Grupo 1		Grupo 2	
	X_1	X_1^2	X_2	X_2^2
	8	64	8	64
	9	81	3	9
	5	25	4	16
	7	49	10	100
	4	16	8	64
	4	16	3	9
	10	100	5	25
	8	64	3	9
	3	9	7	49
	10	100	9	81
	9	81	5	25
	7	49	6	36
	6	36	6	36
	5	25	8	64
	6	36	10	100
	7	49	7	49
Σ	108	800	102	736

Ya que la t_2 obtenida es mayor que la t crítica de 1.697, se rechaza la hipótesis nula $H_0 : \mu_1 - \mu_2 = -4$ en favor de la alternativa $H_A : \mu_1 - \mu_2 > -4$.

Como *ambas*, la prueba 1 y la prueba 2, son significativas, la hipótesis nula de equivalencia se rechaza en favor de la alternativa. De ese modo se establece la equivalencia como se define por el intervalo de equivalencia. ■

6.2.3 Intervalo de confianza para la diferencia entre las medias de dos muestras independientes

Razonamiento. La prueba t de muestras independientes intenta determinar si existe una diferencia entre las medias de dos poblaciones (o si la diferencia es de algún valor determinado). Una pregunta relacionada y más informativa es: "¿Qué tan grande es la diferencia entre las medias de población?" Esta diferencia se puede calcular con un intervalo de confianza.

Mediante la aplicación, si se asigna al azar a los sujetos a un tratamiento con una de las dos diferentes sustancias, la prueba t de muestras independientes intenta determinar si existe una diferencia en el efecto producido por las dos sustancias. En contraste, un intervalo de confianza para la diferencia entre las medias basado en muestras independientes se aboca a la pregunta: "¿Qué tan grande es la diferencia entre los efectos de los dos medicamentos?"

El intervalo de confianza. El intervalo de confianza para la diferencia entre las medias basado en muestras independientes tiene las siguientes formas para L y U.

$$\boxed{L = (\bar{x}_1 - \bar{x}_2) - t\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{6.3}$$

y

$$\boxed{U = (\bar{x}_1 - \bar{x}_2) + t\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{6.4}$$

donde $\bar{x}_1, \bar{x}_2$ y s_P^2 son como se definieron antes para la prueba t de muestras independientes, y t es el valor t apropiado con $n_1 + n_2 - 2$ grados de libertad.

EJEMPLO 6.5

Utilice los datos de la tabla 6.1 de la página 216 para formar un intervalo de confianza de dos colas del 95% para la estimación de $\mu_1 - \mu_2$. Interprete el resultado. Utilice el intervalo obtenido para realizar una prueba t de dos colas de muestras independientes.

Solución Como se calculó previamente en la página 220, $\bar{x}_1 = 131.8$, $\bar{x}_2 = 135.2$ y $s_P^2 = 132.814$. Entonces, mediante las ecuaciones 6.3 y 6.4

$$\begin{aligned} L &= (\bar{x}_1 - \bar{x}_2) - t\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \\ &= (131.8 - 135.2) - 2.048\sqrt{132.814\left(\frac{1}{15} + \frac{1}{15}\right)} \\ &= 12.018 \end{aligned}$$

y

$$
\begin{aligned}
U &= (\bar{x}_1 - \bar{x}_2) + t\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \\
&= (131.8 - 135.2) + 2.048\sqrt{132.814\left(\frac{1}{15} + \frac{1}{15}\right)} \\
&= 5.128
\end{aligned}
$$

■

Una interpretación estrictamente estadística de este intervalo sostendría que uno puede afirmar con una confianza del 95% que la diferencia $\mu_1 - \mu_2$ está entre −12.018 y 5.218. Desde el punto de vista de una investigación, este intervalo sostiene con un 95% de confianza, que el nivel promedio de presión arterial realizado por los pacientes tratados con la sustancia 1 menos el nivel promedio de presión arterial de los pacientes tratados con la sustancia 2 está entre −12.018 y 5.218. Observe que éste es un resultado insatisfactorio para un investigador que está tratando de evaluar la relativa efectividad de las dos sustancias, ya que la diferencia puede ser negativa (lo que indica una ventaja para la sustancia 1), positiva (lo que indica una ventaja para la sustancia 2) o cero (indicando que no hay diferencia).

Este intervalo puede usarse para realizar una prueba de dos colas de la hipótesis $H_0 : \mu_1 - \mu_2 = 0$ mediante el método descrito en el apartado 4.5 de la página 152. Por este método, simplemente hay que ver si el valor determinado por la hipótesis nula está contenido en el intervalo. Si el valor nulo está entre L y U, la hipótesis nula no se rechaza. De otra forma, se rechaza la hipótesis nula. En este caso, cero está entre −12.018 y 5.218 por lo que no se rechaza la hipótesis nula. Éste es el mismo resultado que se obtuvo cuando se realizó una prueba t de muestras independientes con estos mismos datos, en la página 220.

Como comparación, suponga por un momento que el intervalo calculado fuera −10.00 a −5.00. La conclusión en este caso sería que la sustancia 1 mantuvo la ventaja sobre la sustancia 2, en tanto que la presión arterial promedio de los pacientes tratados con este medicamento estaría entre cinco y 10 puntos por debajo del promedio alcanzado por los pacientes tratados con el medicamento 2. En este caso, una prueba de significancia rechazaría la hipótesis nula porque el valor determinado por la hipótesis nula (cero) no está en el intervalo −10.00 a −5.00. Es importante destacar que, mientras la prueba t de muestras independientes sólo afirmaría la superioridad de la sustancia 1, el intervalo de confianza provee un cálculo de la magnitud de esta ventaja.

EJEMPLO 6.6

Utilice los datos en la tabla 6.3 de la página 223, para formar un intervalo de confianza del 95% de una cola que proporcione un límite superior estimado de la diferencia entre la media de la escala de puntajes de los síntomas de hombres y mujeres. Interprete el intervalo así obtenido. Utilice el intervalo para desarrollar una prueba t de muestras independientes de una cola con hipótesis alternativa y nula $H_0 : \mu_1 - \mu_2 = 0$ y $H_A : \mu_1 - \mu_2 < 0$.

Solución Como se calculó previamente en la página 222, $\bar{x}_1 = 13.364$, $\bar{x}_2 = 24.800$ y $s_P^2 = 54.956$. Entonces, por la ecuación 6.4 de la siguiente página

$$U = (13.364 - 24.800) + 1.711\sqrt{54.956\left(\frac{1}{11} + \frac{1}{15}\right)} = -6.401$$

Observe que los grados de libertad asociados con $t = 1.711$ son $11 + 15 - 2 = 24$.

Este intervalo indica, con un 95% de confianza, que la media de la puntuación de los síntomas de los hombres menos la media de la puntuación de los síntomas de las mujeres es *cuando mucho* –6.401. Como U es menor que cero, la hipótesis nula $H_0 : \mu_1 - \mu_2 = 0$ se rechaza en favor de la alternativa $H_A : \mu_1 - \mu_2 < 0$.[3] Como recordará, éste fue el resultado obtenido en la página 222. ■

6.2.4 Suposiciones

Las suposiciones subyacentes a la prueba t de muestras independientes, las pruebas de equivalencia basadas en la prueba t de muestras independientes y los intervalos de confianza para la diferencia entre las medias de muestras independientes son las siguientes. (1) La suposición de **normalidad** especifica que las muestras provienen de poblaciones normalmente distribuidas. Estos procedimientos generalmente (mas no siempre) son sólidos contra los incumplimientos de esta suposición y lo son especialmente cuando los tamaños de la muestra son mayores que o iguales a 30 y son iguales o aproximados. La suposición de **homogeneidad de varianza** especifica que las dos muestras provienen de poblaciones que tienen la misma varianza. Siempre que la heterogeneidad de varianza no sea muy extrema (es decir, 10 a uno), estos procedimientos son generalmente (mas no siempre) sólidos en las mismas condiciones indicadas con anterioridad para la suposición de normalidad. La suposición de **independencia** requiere que cada caso en las dos muestras no esté relacionado con todos los demás casos en las dos muestras.[4] Los procedimientos tratados aquí no pueden considerarse sólidos cuando no se cumple la suposición de independencia.

6.3 MÉTODOS RELACIONADOS CON PROPORCIONES

6.3.1 Una prueba de muestras independientes para la diferencia entre proporciones

Razonamiento. Suponga que se desea evaluar un programa diseñado para orientar a las adolescentes embarazadas sobre una adecuada nutrición durante la gestación. Como parte de la evaluación, se asigna al azar a las adolescentes para recibir el programa educacional o no recibirlo. El resultado de interés es el bajo peso al nacer. La pregunta que nos ocupa es: "¿Las adolescentes embarazadas que reciben orientación nutricional registran una proporción de bebés con bajo peso al nacer diferente de las adolescentes embarazadas que no reciben la orientación?"

Una estrategia razonable para contestar la pregunta sobre si el programa de orientación produjo proporciones de bebés con bajo peso al nacer diferentes, sería calcular la proporción de bebés con bajo peso obtenido en cada uno de los dos grupos y después comparar las dos proporciones para determinar si difieren. En el presente caso, suponga que 314 madres recibieron la orientación nutricional, de las cuales 23 tuvieron bebés con bajo peso. En contraste, 39 de las 316 madres en el grupo sin orientación tuvieron bebés con bajo peso. Así, la proporción de bebés con bajo peso al nacer en el programa nutricional fue $\hat{p}_1 = \frac{23}{314} = .073$, en tanto que en el grupo sin instrucción fue $\hat{p}_2 = \frac{39}{316} = .123$. La diferencia entre estas dos proporciones es, entonces, $.073 - .123 = -.050$, lo cual indica que el grupo con orientación tuvo una proporción menor de bebés con bajo peso al nacer.

Existen por lo menos dos explicaciones para esta diferencia. Primero, notamos que cada vez que asignamos sujetos a los grupos al azar, las proporciones de algunos resultados en los dos

---

[3] Debe revisar el apartado 4.5 de la página 152 si este resultado no queda claro.

[4] Véase la sección 4.3.3 de la página 99 para ejemplos y más detalles.

grupos muy seguramente diferirán en algún grado en casi todas las medidas imaginables. De hecho, nos sorprendería encontrar que la proporción de personas asignadas al azar con un historial de enfermedades cardiacas fuera .131 en los dos grupos o que las proporciones de fumadores fuera exactamente la misma.

Así, aun cuando el programa de orientación no fuera mejor que el de no orientación, esperaríamos ver alguna diferencia en las proporciones de los dos grupos. Entonces, la primera explicación para la diferencia de –.050 que se observó en nuestro estudio (ficticio) es que esta diferencia surgió, no por alguna diferencia en el efecto del tratamiento administrado, sino por la casualidad relacionada a la manera en la que se formaron los grupos.

Una segunda explicación es que la orientación nutricional que se dio al grupo 1 fue efectiva al reducir la incidencia de bebés con bajo peso al nacer. Esto explicaría por qué la proporción en el grupo 1 fue más baja que en el grupo 2. Pero, ¿cuál de estas explicaciones debe creerse? Una prueba de significancia podría ayudar a decidir el asunto.

Piense en una distribución muestral formada al seleccionar repetidamente una muestra al azar de cierta población dicotómica, dividiendo aleatoriamente cada muestra en dos muestras componentes y después calculando la cantidad $\hat{p}_1 - \hat{p}_2$, donde $\hat{p}_1$ y $\hat{p}_2$ son las proporciones de cierto resultado obtenido de las dos muestras componentes.[5] Si este proceso se repitiera muchas veces, los valores resultantes de $\hat{p}_1 - \hat{p}_2$ podrían conformar una distribución muestral. Es importante comprender que esta distribución representa la distribución muestral de $\hat{p}_1 - \hat{p}_2$ *cuando ambas muestras provienen de poblaciones que tienen la misma proporción*. Es claro que si ambas muestras provienen de la misma población, entonces, vienen de poblaciones con la misma proporción. Designaremos la proporción de la población de la cual se tomó la primera muestra como π_1 y la de la población de la cual se tomó la segunda muestra como π_2. Como ambas muestras provienen de la misma población, es claro que $\pi_1 = \pi_2$.

Observe que la media de esta distribución muestral sería cero y que cuanto más lejanos estén los valores de $\hat{p}_1 - \hat{p}_2$ de cero, menos probable es su ocurrencia. Pero, ¿cómo podría usarse esta distribución para ayudar a contestar la pregunta sobre si la diferencia de –.050 entre las proporciones de bebés con bajo peso al nacer en los dos grupos, se debió al hecho de que el programa de orientación que se dio al grupo 1 fue efectivo al reducir la incidencia de bebés con bajo peso?

Las dos muestras originalmente se tomaron de una población común, por lo que $\pi_1 = \pi_2$. Si el programa de orientación produjera exactamente el mismo resultado que la no orientación, entonces sería cierto que $\pi_1 = \pi_2$. En tales circunstancias no esperaríamos que $\hat{p}_1 - \hat{p}_2$ estuviera en una región crítica. ¿Por qué? Porque la probabilidad de que $\hat{p}_1 - \hat{p}_2$ esté en una región crítica cuando $\pi_1 = \pi_2$ es solamente α.

Pero, ¿qué pasaría si la orientación nutricional fuera más efectiva que la no orientación? En este caso la incidencia de pesos bajos al nacer en el grupo 1 se reduciría por haber recibido el programa, por lo que el valor absoluto de $\hat{p}_1 - \hat{p}_2$ tendería a ser mayor (más alejado de cero) que en el caso en el que ambas estrategias tuvieran un igual efecto. Por ejemplo, en nuestro estudio hipotético, $\hat{p}_1 - \hat{p}_2 = .073 - .123 = -.050$. Pero, suponga que la orientación hubiera sido más efectiva, de tal manera que $\hat{p}_1 = .011$. Entonces, $\hat{p}_1 - \hat{p}_2 = .011 - .123 = -.112$. Cuanto más efectiva sea la orientación en comparación con la no orientación, $\hat{p}_1 - \hat{p}_2$ estará más abajo de cero y la probabilidad de que $\hat{p}_1 - \hat{p}_2$ se coloque en la región crítica inferior será mayor. ¿Qué pasaría si la no orientación fuera superior a la orientación?

---

[5] Igualmente, podemos pensar en las dos muestras componentes simplemente como dos muestras aleatorias tomadas de una población común.

En resumen, no esperamos que $\hat{p}_1 - \hat{p}_2$ esté en una región crítica cuando los tratamientos producen el mismo resultado. De hecho, la probabilidad de que esto suceda es solamente α. Sin embargo, cuando los tratamientos producen resultados diferentes, esperamos que el valor de $\hat{p}_1 - \hat{p}_2$ se aleje del cero y vaya hacia una región crítica. Es más, cuando los efectos de los tratamientos difieren no podemos afirmar que $\pi_1 = \pi_2$. Más bien, se debe decir que $\pi_1 \neq \pi_2$.

De lo anterior, cuando $\hat{p}_1 - \hat{p}_2$ cae en una región crítica, podemos decir que la diferencia observada entre estos valores se debió a efectos diferenciales del tratamiento y no a la casualidad de cómo se formaron los grupos. Ahora veremos cómo se realiza esta prueba.

La prueba. La hipótesis nula que debe analizarse mediante la prueba Z de muestras independientes es a menudo (pero no siempre)

$$H_0 : \pi_1 = \pi_2$$

o de manera equivalente,

$$H_0 : \pi_1 - \pi_2 = 0$$

La alternativa de dos colas es a menudo (pero no siempre)

$$H_A : \pi_1 \neq \pi_2$$

o de manera equivalente,

$$H_A : \pi_1 - \pi_2 \neq 0$$

Las alternativas de una cola son a menudo (pero no siempre) de la forma

$$H_A : \pi_1 < \pi_2$$

o de manera equivalente,

$$H_A : \pi_1 - \pi_2 < 0$$

o

$$H_A : \pi_1 > \pi_2$$

o de manera equivalente,

$$H_A : \pi_1 - \pi_2 > 0$$

En esencia, la hipótesis nula afirma que no hubo diferencia en los efectos de los tratamientos administrados en los dos grupos, en tanto que la alternativa de dos colas sostiene que sí existe una diferencia. Las alternativas de una cola especifican la forma de esa diferencia.

Para poder llevar a cabo la prueba de hipótesis, $\hat{p}_1 - \hat{p}_2$ no puede referirse directamente a una distribución muestral como se describió antes, sino que debe dividirse entre el error estándar (véase el denominador de la ecuación 6.5), lo que significa que el estadístico de prueba se puede referir a una distribución normal. El estadístico de prueba se calcula mediante la ecuación 6.5.

$$\boxed{Z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}}} \tag{6.5}$$

En esta ecuación $\hat{p}_1$ y $\hat{p}_2$ representan la proporción de éxitos en la primera y segunda muestra, respectivamente; $\hat{q}_1$ y $\hat{q}_2$ representan la proporción de fracasos en las dos muestras (esto es, $\hat{q}_1 = 1 - \hat{p}_1$ y $\hat{q}_2 = 1 - \hat{p}_2$ y n_1 y n_2 representan los tamaños de las dos muestras. El símbolo δ_0 representa la diferencia planteada como hipótesis entre π_1 y π_2, que normalmente es cero. No obstante, como verá en la explicación de pruebas de equivalencia basadas en la prueba Z de muestras independientes para la diferencia entre proporciones, éste no siempre es el caso. Cabe hacer notar que la ecuación 6.5 nos da una prueba *aproximada*. Hay una prueba exacta disponible, pero no la explicaremos aquí.

EJEMPLO 6.7

Utilice la información que se dio para el estudio de orientación nutricional para conducir una prueba Z de una cola de la hipótesis nula $H_0 : \pi_1 = \pi_2$ contra la alternativa $H_A : \pi_1 < \pi_2$. ¿A qué conclusión llegó respecto de la efectividad del programa de orientación?

Solución Como mencionamos antes, 314 madres recibieron orientación nutricional, de las cuales 23 tuvieron bebés de bajo peso. En contraste, 39 de las 316 madres en el grupo sin orientación tuvieron bebés de bajo peso. Así, $\hat{p}_1 = \frac{23}{314} = .073$, $\hat{p}_2 = \frac{39}{316} = .123$, $\hat{q}_1 = 1 - .073 = .927$, $\hat{q}_2 = 1 - .123 = .877$, $n_1 = 314$ y $n_2 = 316$. Entonces, por la ecuación 6.5 de la página anterior

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{.073 - .123}{\sqrt{\frac{(.073)(.927)}{314} + \frac{(.123)(.877)}{316}}} = \frac{-.050}{.0236} = -2.12$$

El Apéndice A muestra que el valor crítico de una prueba Z de una cola realizada con $\alpha = .05$ es -1.65, que nos lleva al rechazo de la hipótesis nula. Podemos concluir, entonces, que el programa de orientación fue efectivo al reducir la incidencia de partos de bajo peso. ■

EJEMPLO 6.8

Un estudio muestra que 61 de 414 adultos que crecieron en un hogar monoparental reportan haber sufrido, por lo menos, un incidente de abuso sexual durante la infancia. En contraste, 74 de 501 adultos que crecieron en hogares con ambos padres reportan el mismo abuso. Utilice una prueba Z de muestras independientes para la diferencia entre proporciones con el fin de probar la hipótesis nula $H_0 : \pi_1 = \pi_2$ contra la alternativa de dos colas. ¿Cuál es su conclusión sobre si existe una diferencia en proporciones de abuso entre los hogares monoparentales y los hogares con ambos padres?

Solución De lo anterior se sigue que $\hat{p}_1 = \frac{61}{414} = .147$, $\hat{p}_2 = \frac{74}{501} = .148$, $\hat{q}_1 = 1 - .147 = .853$, $\hat{q}_2 = 1 - .148 = .852$, $n_1 = 414$, y $n_2 = 501$. Entonces, por la ecuación 6.5 de la página anterior

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{.147 - .148}{\sqrt{\frac{(.147)(.853)}{414} + \frac{(.148)(.852)}{501}}} = \frac{-.001}{.0235} = -.04$$

El Apéndice A muestra que los valores críticos para una prueba Z de dos colas con $\alpha = .05$ son -1.96 y 1.96, lo que conduce al no rechazo de la hipótesis nula. Se deduce que no pudimos demostrar una diferencia entre las proporciones de abusos en los hogares de uno y dos padres. ■

6.3.2 Cómo establecer la equivalencia por medio de una prueba *Z* de muestras independientes para la diferencia entre proporciones

Razonamiento. Como usted sabe, rechazar una hipótesis nula ofrece buena evidencia de que la hipótesis es falsa. En contraste, no rechazar una hipótesis nula, por lo general, no provee buena evidencia de que la hipótesis nula es verdadera. En situaciones donde se desea establecer la validez de la hipótesis nula debe emplearse una prueba de equivalencia para demostrar que la hipótesis nula es (aproximadamente) verdadera. (Véase el apartado 4.3.6, en la página 117).

A veces ocurre que los investigadores quieren establecer que no existe *ninguna* diferencia entre los tratamientos, en vez de demostrar que *sí* existe una diferencia. En tales casos se puede utilizar una prueba *Z* de muestras independientes para la diferencia entre proporciones con el fin de establecer la equivalencia.

Por ejemplo, quizás una compañía farmacéutica desee demostrar que una nueva sustancia diseñada para el tratamiento de la depresión es similar a una antigua versión de la misma en lo concerniente a la incidencia de un efecto secundario en particular.

La prueba. Usando la notación de la sección 4.3.6 de la página 117, las hipótesis nula y alternativa para una prueba de equivalencia de dos colas basada en una prueba *Z* de muestras independientes para la diferencia entre proporciones se puede enunciar como

$$H_{0E} : \pi_1 - \pi_2 \leq EI_L \text{ o } \pi_1 - \pi_2 \geq EI_U$$
$$H_{AE} : EI_L < \pi_1 - \pi_2 < EI_U$$

donde EI_L y EI_U son los extremos inferior y superior del intervalo de equivalencia. En esencia, la hipótesis nula plantea que la diferencia entre proporciones no reside en el intervalo de equivalencia, mientras que la alternativa afirma que la diferencia está en el intervalo. Recordará que la hipótesis nula se comprueba realizando dos pruebas de una cola al nivel α. Para poder rechazar la hipótesis nula de equivalencia, debe demostrarse que $\pi_1 - \pi_2 > EI_L$ y que $\pi_1 - \pi_2 < EI_U$, lo que significa que las dos pruebas siguientes deben ser significativas.

Prueba 1	Prueba 2
$H_{01} : \pi_1 - \pi_2 = EI_U$	$H_{02} : \pi_1 - \pi_2 = EI_L$
$H_{A1} : \pi_1 - \pi_2 < EI_U$	$H_{A2} : \pi_1 - \pi_2 > EI_L$

Las hipótesis nula y alternativa para la prueba de equivalencia de una cola son *una* de las siguientes

$$H_{0E} : \pi_1 - \pi_2 \geq EI_U$$
$$H_{AE} : \pi_1 - \pi_2 < EI_U$$

o

$$H_{0E} : \pi_1 - \pi_2 \leq EI_L$$
$$H_{AE} : \pi_1 - \pi_2 > EI_L$$

La primera hipótesis de equivalencia de una cola mencionada anteriormente se somete a la prueba 1 y, la segunda a la prueba 2.

EJEMPLO 6.9

Se conduce un ensayo clínico para determinar si la incidencia de efectos secundarios asociados con una nueva sustancia elaborada para el tratamiento de la depresión es similar a la incidencia asociada con una antigua versión de la misma. Se determina que las dos sustancias se considerarán equivalentes en lo que concierne a la incidencia del efecto secundario indicado, si la diferencia entre las proporciones de los pacientes que están tomando la nueva sustancia y la anterior, que experimentan el efecto secundario, está en el intervalo de equivalencia −.04 a .04. La proporción de pacientes que toman la nueva sustancia y que experimentan el efecto secundario es .07, mientras que para los pacientes que toman la antigua sustancia es .06. Cada grupo se compuso de 100 pacientes. Utilice una prueba de equivalencia de dos colas con $\alpha = .05$ para determinar esta resolución. Comience por enunciar las hipótesis nula y alternativa.

Solución Las hipótesis nula y alternativa de equivalencia son como sigue.

$$H_{0E}: \pi_1 - \pi_2 \le -.04 \text{ o } \pi_1 - \pi_2 \ge .04$$
$$H_{AE}: -.04 < \pi_1 - \pi_2 < .04$$

Por la ecuación 6.5, la Z obtenida para la prueba 1 es

$$Z_1 = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{(.07 - .06) - .040}{\sqrt{\frac{(.07)(.93)}{100} + \frac{(.06)(.94)}{100}}} = \frac{-.03}{.03486} = -.86$$

Entonces, la Z obtenida para la prueba dos es

$$Z_2 = \frac{(.07 - .06) - (-.040)}{\sqrt{\frac{(.07)(.93)}{100} + \frac{(.06)(.94)}{100}}} = \frac{.05}{.03486} = 1.43$$

Del Apéndice A se puede determinar que el valor crítico para la prueba 1 es −1.65 y que para la prueba 2 es 1.65. Para rechazar la hipótesis nula de equivalencia, *ambas* pruebas deben ser significativas. En este caso, ninguna de las pruebas es significativa, lo que nos lleva a no rechazar la hipótesis nula de equivalencia. Así que no se pudo establecer la equivalencia. ■

EJEMPLO 6.10

Se realiza un estudio para determinar si una fórmula para bebés mejorada brinda protección contra una enfermedad común de la infancia equivalente a la que obtienen los bebés que toman leche materna. Se determina que si la proporción de bebés que toman fórmula y que contraen la enfermedad es mayor, por menos de .03, que la proporción de bebés que toman leche materna y que contraen la enfermedad, se declarará la equivalencia. Cincuenta y dos de los 390 bebés alimentados con fórmula contrajeron la enfermedad, en tanto que 94 de los 750 bebés que toman leche materna contrajeron la enfermedad. Utilice una prueba de una cola con $\alpha = .05$ para probar la equivalencia. Declare las hipótesis de equivalencia nula y alternativa.

Solución Si designamos a los bebés alimentados con fórmula como el grupo 1, las hipótesis de equivalencia nula y alternativa se plantean como

$$H_{0E}: \pi_1 - \pi_2 \ge .03$$
$$H_{AE}: \pi_1 - \pi_2 < .03$$

De la información proporcionada anteriormente, $\hat{p}_1 = \frac{52}{390} = .133$, $\hat{p}_2 = \frac{94}{750} = .125$, $\hat{q}_1 =$, $1 - .133 = .867$, $\hat{q}_2 = 1 - .125 = .875$, $n_1 = 390$, y $n_2 = 750$. Por la ecuación 6.5 de la página 232, la Z obtenida para la prueba 1 es

$$Z_1 = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{(.133 - .125) - .030}{\sqrt{\frac{(.133)(.867)}{390} + \frac{(.125)(.875)}{750}}} = \frac{-.022}{.0210} = -.1.05$$

Del Apéndice A se puede determinar que el valor crítico para la prueba 1 es −1.65, lo que da por resultado el no rechazo de la hipótesis nula. Por consiguiente, no se estableció la equivalencia entre la fórmula para el bebé y la leche materna, en lo que se refiere a la prevención de la enfermedad infantil especificada. ■

6.3.3 Intervalo de confianza para una diferencia entre proporciones basada en dos muestras independientes

Razonamiento. La prueba Z de muestras independientes para una diferencia entre proporciones intenta determinar si existe una diferencia entre dos proporciones de población. Una pregunta relacionada, y por lo común más informativa, es: "¿Qué tan grande es la diferencia entre las proporciones de población?" Esta diferencia se calcula con un intervalo de confianza.

Mediante la aplicación, si se asigna al azar a los sujetos a un tratamiento con una de las dos diferentes sustancias, la prueba Z de muestras independientes para una diferencia entre proporciones intentará determinar si existe una diferencia en el efecto producido por las dos sustancias, como lo expresa cierto resultado dicotómico. En contraste, un intervalo de confianza para la diferencia entre las proporciones basado en dos muestras independientes se aboca a la pregunta: "¿Qué tan grande es la diferencia entre el efecto de las dos sustancias?"

El intervalo de confianza. El intervalo de confianza para la diferencia entre las proporciones basado en muestras independientes tiene las siguientes formas para L y U.

$$\boxed{L = (\hat{p}_1 - \hat{p}_2) - \left(Z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1 - 1} + \frac{\hat{p}_2\hat{q}_2}{n_2 - 1}} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)} \tag{6.6}$$

y

$$\boxed{U = (\hat{p}_1 - \hat{p}_2) + \left(Z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1 - 1} + \frac{\hat{p}_2\hat{q}_2}{n_2 - 1}} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)} \tag{6.7}$$

donde $\hat{p}_1$ y $\hat{p}_2$ son las proporciones de éxitos en las dos muestras, $\hat{q}_1$ y $\hat{q}_2$ son las proporciones defracasos en las dos muestras (es decir, $\hat{q}_1 = 1 - \hat{p}_1$, $\hat{q}_2 = 1 - \hat{p}_2$), n_1 y n_2 son los respectivos tamaños de las muestras y Z es el valor Z apropiado para el intervalo especificado.

EJEMPLO 6.11

Una encuesta (ficticia) aplicada a adolescentes de entre 12 y 16 años de edad reporta que 106 de 299 varones y 66 de 313 mujeres encuestados indican que por rutina fuman tres o más cigarrillos al día. Construya un intervalo de confianza de dos colas del 95% para calcular la diferencia entre las proporciones de varones y mujeres en la población que fuma dos o más cigarrillos al día. Utilice este intervalo de confianza para probar la hipótesis nula $H_0 : \pi_1 = \pi_2$ contra la alternativa $\pi_1 \neq \pi_2$ con un nivel de significancia de .05. Plantee la razón para su decisión considerando la hipótesis nula.

Solución Por las ecuaciones 6.6 y 6.7 de la página anterior

$$
\begin{aligned}
L &= (\hat{p}_1 - \hat{p}_2) - \left(Z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1 - 1} + \frac{\hat{p}_2\hat{q}_2}{n_2 - 1}} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) \\
&= (.355 - .211) - \left(1.96\sqrt{\frac{(.355)(.645)}{298} + \frac{(.211)(.789)}{312}} + \frac{1}{2}\left(\frac{1}{299} + \frac{1}{313}\right)\right) \\
&= .070
\end{aligned}
$$

y

$$
\begin{aligned}
U &= (\hat{p}_1 - \hat{p}_2) + \left(Z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1 - 1} + \frac{\hat{p}_2\hat{q}_2}{n_2 - 1}} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right) \\
&= (.355 - .211) + \left(1.96\sqrt{\frac{(.355)(.645)}{298} + \frac{(.211)(.789)}{312}} + \frac{1}{2}\left(\frac{1}{299} + \frac{1}{313}\right)\right) \\
&= .218
\end{aligned}
$$

Así, el intervalo de confianza del 95% es .070 a .218.

Este intervalo puede usarse para realizar una prueba de dos colas de la hipótesis nula $H_0 : \pi_1 - \pi_2 = 0$ por el método descrito en el apartado 4.5 de la página 152. Mediante este método, simplemente hay que observar el intervalo para saber si el valor estipulado por la hipótesis nula está contenido ahí. Si el valor nulo está entre L y U, la hipótesis nula no se rechaza. De otra manera, la hipótesis se rechaza. En este caso, cero no está entre .070 y .218, por lo que se rechaza la hipótesis nula. Realizar pruebas de hipótesis de esta forma mediante el uso de las ecuaciones 6.6 y 6.7 puede dar resultados un tanto superiores a los obtenidos mediante la ecuación 6.5 de la página 232, debido al hecho de que las ecuaciones anteriores pueden aproximarse más al modelo de la curva normal [18]. ■

EJEMPLO 6.12

En el ejemplo 6.7 de la página 233 se llevó a cabo una prueba Z de una cola de muestras independientes para la diferencia entre proporciones relacionadas con la evaluación de un programa (ficticio) de orientación nutricional. Se sometió a prueba la hipótesis nula $H_0 : \pi_1 = \pi_2$ contra la alternativa $H_A : \pi_1 < \pi_2$. En el transcurso de esta prueba descubrió que $\hat{p}_1 = .073$, $\hat{p}_2 = .123$, $\hat{q}_1 = .927$, $\hat{q}_2 = .877$, $n_1 = 314$ y $n_2 = 316$. Utilice esta información para construir un intervalo de confianza de una cola para el límite superior de $\pi_1 - \pi_2$. Interprete este intervalo y úselo para realizar una prueba de significancia similar a la que efectuó antes. Explique cómo realizó la prueba.

Solución Por la ecuación 6.7 de la página 236,

$$U = (.073 - .123) + \left(1.65\sqrt{\frac{(0.73)(.927)}{313} + \frac{(.123)(.877)}{315}} + \frac{1}{2}\left(\frac{1}{314} + \frac{1}{316}\right)\right)$$
$$= -.008$$

Desde el punto de vista de una investigación, este intervalo sostiene, con una confianza del 95%, que la proporción de bebés con bajo peso al nacer en el grupo de mamás que recibieron la orientación estuvo *por lo menos* .008 por debajo de la del grupo sin orientación. La hipótesis nula antes planteada se rechaza porque el extremo superior del intervalo está por debajo de cero.[6] ■

6.3.4 Suposiciones

Las suposiciones subyacentes a la prueba Z de muestras independientes para la diferencia entre proporciones, la prueba de equivalencia basada en la prueba Z de muestras independientes para la diferencia entre proporciones y el intervalo de confianza para una diferencia entre proporciones basada en dos muestras independientes son esencialmente las mismas que en la prueba aproximada para una proporción, que se expusieron en la página 116.

6.4 MÉTODOS RELACIONADOS CON LAS TASAS DE RIESGO DE MUESTRAS INDEPENDIENTES

6.4.1 Antecedentes

Como se explicó en el apartado 5.4.1 de la página 190,[7] el **riesgo** de algún evento en un grupo en particular es la *probabilidad* de que el evento le ocurra a un miembro de ese grupo. Usando la notación del capítulo 3, el riesgo de un grupo que ha sido expuesto a un agente potencialmente dañino (o benéfico) puede ser calificado como $P\ (D \mid E)$, que se lee: "La probabilidad de enfermedad dada la exposición". De igual forma, podemos calificar el riesgo de los sujetos en un grupo no expuesto como $P\ (D \mid \bar{E})$, que se lee: "La probabilidad de enfermedad dada la no exposición". Estas dos probabilidades se pueden comparar formándolas en una tasa de riesgo como sigue.

$$\boxed{RR = \frac{P\ (D \mid E)}{P\ (D \mid \bar{E})}} \tag{6.8}$$

Usted la reconocerá como la ecuación 3.11 de la página 59.[8] Si el riesgo de enfermedad[9] en los dos grupos es el mismo, la tasa de riesgo es 1, lo que implicaría que no hay relación entre la exposición y la enfermedad. Por otra parte, una tasa de riesgo mayor que 1 implicaría que la exposición está asociada con un mayor riesgo de enfermedad que la no exposición. Se dice que los riesgos menores que 1 son **protectores** porque la exposición implica un riesgo reducido de enfermedad. Esto podría ocurrir, por ejemplo, cuando las personas son expuestas a una vacuna. Como mencionamos previamente, representaremos las formas del parámetro y el estadístico de la tasa de riesgo como RR y $\widehat{RR}$, respectivamente.

---

[6] Véase el apartado 4.5 de la página 152 si esta explicación no queda del todo clara.

[7] Quizá desee revisar ese apartado antes de continuar.

[8] Quizá desee revisar la explicación al inicio de la página 59 en este punto.

[9] Si bien nos referimos a enfermedad y exposición en la ecuación 6.8, los términos son genéricos y representan cualesquiera que sean las variables examinadas.

TABLA 6.7: Tabla de dos por dos para datos dicotómicos dispares.

		Enfermedad sí	Enfermedad no
Expuestos	sí	a	b
	no	c	d

Será conveniente referirse a la tabla 6.7 para las siguientes explicaciones. Si bien utilizaremos los términos genéricos "expuestos" y "enfermedad" para esta tabla, es factible emplear otras variables como "género" y "acceso a servicios médicos". Un estudio típico en el que se pueden emplear tasas de riesgo daría seguimiento a dos grupos de individuos, uno para quienes hubieran experimentado alguna forma de exposición y otro para quienes no hubieran estado expuestos, para así determinar entonces si el riesgo de cierto resultado es diferente en los dos grupos. Los estudios de este tipo se denominan estudios de **cohorte potencial**.

La tasa de riesgo muestral de muestras independientes se define mediante la ecuación 6.9, donde a, b, c y d son las frecuencias de celdas, como se especificó en la tabla 6.7. Observe que el riesgo para las personas en la primera fila se está comparando con el de las personas en la segunda fila. Así, por ejemplo, una tasa de riesgo calculada de 2.0 indicaría que las personas en la primera fila corren el doble de riesgo del resultado especificado que las personas en la segunda fila. Debe tener esto en cuenta cuando construya tablas para análisis.

$$\boxed{\widehat{RR} = \frac{a/(a+b)}{c/(c+d)}} \tag{6.9}$$

6.4.2 Prueba de la hipótesis RR = 1 para muestras independientes

Razonamiento. Volviendo a un ejemplo presentado en el capítulo 5, suponga que se sospecha que la exposición a un solvente químico en particular usado en fábricas incrementa el riesgo de cáncer de vejiga. Para probar esta sospecha, por un tiempo determinado se podría dar seguimiento a los trabajadores de cierta industria, que por rutina están expuestos al solvente, para determinar si desarrollan la enfermedad. Con el propósito de lograr comparaciones, se podría dar seguimiento a un grupo de trabajadores de la misma industria que no estén expuestos al solvente. Después del periodo de seguimiento, la tasa de riesgo muestral se calcula dividiendo la proporción de trabajadores que desarrollan la enfermedad entre la proporción de trabajadores no expuestos que desarrollan la enfermedad. Si los dos riesgos fueran .005 y .001, respectivamente, la tasa de riesgo sería $\frac{.005}{.001} = 5$. Esto indica que el riesgo de cáncer de vejiga para los trabajadores expuestos es cinco veces mayor que el de los trabajadores no expuestos. Pero, ¿es éste simplemente un resultado de la muestra usada en el estudio? Si se repitiera el estudio con otros trabajadores, ¿el resultado podría ser notablemente diferente? Lo que es más importante, ¿la tasa de riesgo en la población 1 está indicando que no hay relación entre la exposición y la enfermedad (es decir, el riesgo en los dos grupos es el mismo), o es diferente al que indica que sí existe esa relación? Una prueba de hipótesis ayuda a responder esta pregunta.

La prueba. Se puede realizar una prueba de hipótesis nula $RR = 1$ contra alternativas de una cola o de dos colas mediante la ecuación 6.10 de la siguiente página. Las frecuencias a, b, c y d son como se muestran en la tabla 6.7. La hipotética tasa de riesgo de la población está representada por RR_0. Puesto que el logaritmo de 1 es cero, la expresión (RR_0) puede omitirse cuando se esté probando $H_0 : RR = 1$.

$$Z = \frac{\ln(\widehat{RR}) - \ln(RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} \tag{6.10}$$

EJEMPLO 6.13

Suponga que los resultados del ejemplo del solvente industrial mencionado con anterioridad hubieran sido los siguientes: 42 de los trabajadores expuestos al solvente desarrollaron cáncer de vejiga y 2,981 no. De los trabajadores no expuestos, 21 desarrollaron la enfermedad y 4,088 no. Utilice estos datos para calcular e interpretar la tasa de riesgo muestral. Después compruebe la hipótesis nula $H_0 : RR = 1$ contra la alternativa de dos colas.

Solución Si se forman los datos en una tabla de dos por dos, tenemos que

		Cáncer de vejiga	
		sí	no
Expuestos al solvente	sí	42	2,981
	no	21	4,088

Por la ecuación 6.9 de la página anterior, la tasa de riesgo muestral es

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)} = \frac{42/(42+2981)}{21/(21+4088)} = 2.718$$

Esto significa que el riesgo de cáncer de vejiga entre los trabajadores expuestos al solvente es 2.718 veces mayor que el de los trabajadores que no estuvieron expuestos.

Por la ecuación 6.10,

$$Z = \frac{\ln(\widehat{RR}) - \ln(RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} = \frac{\ln(2.718)}{\sqrt{\frac{2981/42}{42+2981} + \frac{4088/21}{21+4088}}} = \frac{1.00}{\sqrt{.023 + .047}} = 3.78$$

El Apéndice A muestra que los valores críticos para una prueba Z de dos colas con $\alpha = .05$ son -1.96 y 1.96. Como la Z obtenida de 3.78 excede la Z crítica de 1.96, se rechaza la hipótesis nula. Por lo tanto, se concluye que el riesgo para el grupo expuesto es mayor que el del grupo no expuesto. ■

EJEMPLO 6.14

Se realiza un estudio a gran escala para determinar si la ingestión habitual de ácido fólico (vitamina B_9) entre las mujeres en edad reproductiva y con riesgo de tener bebés que presenten espina bífida reducirá el riesgo de esta enfermedad. Con tal fin, se alienta a un grupo grande de mujeres que han sido valoradas como de alto riesgo a que tomen un suplemento diario de ácido fólico. Un grupo similar de mujeres no recibe esta recomendación. De los bebés vivos registrados para el grupo con la recomendación, se diagnostican 10 con espina bífida y 12,344 que no tienen la enfermedad. En contraste, 17 de los bebés nacidos en el grupo sin la recomendación presentaron la enfermedad y 11,202 no.

Calcule e interprete la tasa de riesgo para los dos grupos. Compruebe la hipótesis nula H_0 : $RR = 1$ contra la alternativa de una cola H_A : $RR < 1$ en el nivel .05.

Solución Si formamos los datos en una tabla de dos por dos, tenemos que

		Espina bífida	
		sí	no
Recomendación	sí	10	12,344
	no	17	11.202

Por la ecuación 6.9 de la página 239 la tasa de riesgo muestral es

$$\widehat{RR} = \frac{a\,/\,(a+b)}{c\,/\,(c+d)} = \frac{10\,/\,(10+12344)}{17\,/\,(17+11202)} = .534$$

Esto significa que el riesgo de que un bebé nazca con espina bífida en el grupo "con la recomendación" es solamente de .534 del riesgo del grupo "sin la recomendación". Por la ecuación 6.10 de la página anterior

$$Z = \frac{\ln\,(\widehat{RR}) - \ln\,(RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} = \frac{\ln\,(.534)}{\sqrt{\frac{12344/10}{10+12344} + \frac{11202/17}{17+11202}}} = \frac{-.627}{\sqrt{.100+.059}} = -1.57$$

El Apéndice A muestra que el valor crítico de una prueba Z de una cola con la alternativa estipulada conducida con $\alpha = .05$ es -1.65. Como la Z obtenida de -1.57 no está en la región crítica definida por una Z de -1.65, no se rechaza la hipótesis nula. Como resultado, los investigadores en este estudio ficticio no pudieron demostrar un efecto protector para las madres con la recomendación de incluir ácido fólico en su dieta. ■

6.4.3 Cómo establecer la equivalencia por medio de la tasa de riesgo de muestras independientes

Razonamiento. Como sucede con las diferencias entre medias y proporciones, a veces se desea demostrar que el riesgo asociado con alguna condición es equivalente al riesgo asociado con alguna otra condición. Esto implicaría el uso de una prueba de equivalencia de dos colas basada en tasas de riesgos. Quizá es más común el caso en el cual se desea demostrar que el riesgo asociado con una condición *no es mayor que* el riesgo asociado con otra condición. Esto implica una prueba de una cola.

Recordando un ejemplo del capítulo 5, suponga que se pondrá en marcha un programa diseñado para controlar la mosca de la fruta mediante la aspersión aérea del químico malatión en grandes áreas geográficas. Como algunos activistas ambientales han afirmado que esta aspersión es dañina para los humanos, se podría diseñar un estudio para demostrar que el riesgo asociado con la aspersión no es mayor que el riesgo para las personas no expuestas al aerosol. Con este fin, los sujetos que viven en una zona recientemente fumigada quizá se sometan a un examen médico al igual que las personas que viven en una región que no haya sido fumigada. Cada condición examinada se evalúa como presente o no presente en cada sujeto.

El análisis de los datos recabados muestra que para una condición específica (es decir, ritmo cardiaco elevado), 126 personas en el área fumigada manifiestan la condición y 819 no. En el área no fumigada, 119 residentes manifiestan la condición y 811 no. Los investigadores someterán a prueba la hipótesis nula de que el riesgo en la zona fumigada es mayor que el riesgo en la zona no fumigada contra la alternativa de que el riesgo en el área fumigada no es mayor que el de la zona no fumigada.

¿Qué diferencia hay entre el análisis descrito aquí y uno que someta a prueba la hipótesis nula de no diferencia en el riesgo contra una alternativa de que el riesgo se eleva en las áreas fumigadas? ¿En dónde reside el peso de la prueba en los dos análisis? Dejaremos que reflexione sobre esta pregunta.[10]

La prueba. Los intervalos de equivalencia para las tasas de riesgo, por lo regular, (aunque no necesariamente) son simétricos respecto a 1.0 en el sentido de que $EI_U = \frac{1}{EI_L}$ y viceversa. Así, si EI_L es .8, EI_U sería normalmente $\frac{1.0}{.8} = 1.25$.

Si se usan los datos de la sección 4.3.6 de la página 117, las hipótesis nula y alternativa para una prueba de equivalencia de dos colas para tasas de riesgo independientes se pueden enunciar como

$$H_{0E}: RR \le EI_L \text{ o } RR \ge EI_U$$
$$H_{AE}: EI_L < RR < EI_U$$

donde EI_L y EI_U son los extremos inferior y superior del intervalo de equivalencia, respectivamente. En esencia, la hipótesis nula plantea que la tasa de riesgo no reside en el intervalo de equivalencia, mientras que la alternativa afirma que la tasa de riesgo está en el intervalo. Recordará que la hipótesis nula se comprueba realizando dos pruebas de una cola en el nivel α. Para rechazar la hipótesis nula de equivalencia, se debe probar que $RR > EI_L$ *y* que RR $< EI_U$, lo que significa que *las dos* pruebas siguientes deben ser significativas.

Prueba 1	Prueba 2
$H_{01}: RR = EI_U$	$H_{02}: RR = EI_L$
$H_{A1}: RR < EI_U$	$H_{A2}: RR > EI_L$

Las hipótesis nula y alternativa para la prueba de equivalencia de una cola son *una* de las siguientes:

$$H_{0E}: RR \ge EI_U$$
$$H_{AE}: RR < EI_U$$

o

$$H_{0E}: RR \le EI_L$$
$$H_{AE}: RR > EI_L$$

[10] No se desaliente si no puede comprender todo. Las pruebas de equivalencia a menudo son un poco difíciles de comprender. Sólo siga trabajando en ellas.

La primera hipótesis de equivalencia de una cola enunciada antes se somete a la prueba 1 y la segunda a la prueba 2.

Las pruebas 1 y 2 se pueden efectuar usando la ecuación 6.10 de la página 240. Un ejemplo demostrará cómo se logra.

EJEMPLO 6.15

Calcule la tasa de riesgo para los datos del malatión que se utilizaron antes. Suponiendo que un incremento en el riesgo de menos de 1.1 se considera aceptable, realice una prueba de equivalencia de una cola para demostrar que el incremento en el riesgo es menor que este valor. Enuncie las hipótesis nula y alternativa de equivalencia e interprete los resultados de la prueba.

Solución Es conveniente formar los datos en una tabla de dos por dos como se muestra a continuación.

		Ritmo cardiaco elevado: sí	Ritmo cardiaco elevado: no
Expuestos al malatión	sí	126	819
	no	119	811

Por la ecuación 6.9 de la página 239, la tasa de riesgo muestral es

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)} = \frac{126/(126+819)}{119/(119+811)} = 1.042$$

Puesto que los investigadores intentan determinar si la tasa de riesgo es menor que 1.1, la prueba será de una cola con la hipótesis nula de equivalencia

$$H_{0E} : RR \geq 1.1$$

y la alternativa

$$H_{AE} : RR < 1.1$$

Si se realiza la prueba 1 por medio de la ecuación 6.10, tenemos que

$$Z_1 = \frac{\ln(\widehat{RR}) - \ln(RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} = \frac{\ln(1.042) - \ln(1.100)}{\sqrt{\frac{819/126}{126+819} + \frac{811/119}{119+811}}} = \frac{.041 - .095}{\sqrt{.007 + .007}} = -.456$$

Como este valor es mayor que el valor crítico de −1.65, la prueba no es significativa. Por lo tanto, los investigadores no pudieron demostrar que el incremento en el riesgo estuvo dentro del límite aceptable (es decir, menos de 1.1). ■

EJEMPLO 6.16

Utilice los datos de la siguiente tabla para calcular $\widehat{RR}$. Después, lleve a cabo una prueba de equivalencia de dos colas usando un *EI* de .833 a 1.200. Interprete sus resultados.

		Resultado positivo: sí	Resultado positivo: no
Expuestos	sí	301	771
	no	345	820

Solución Por la ecuación 6.9 de la página 239, la tasa de riesgo muestral es

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)} = \frac{301/(301+771)}{345/(345+820)} = .948$$

Esto sugeriría que la exposición da un pequeño efecto protector. Si se conduce la prueba 1 por medio de la ecuación 6.10 de la página 240, tenemos que

$$Z_1 = \frac{\ln(\widehat{RR}) - \ln(RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} = \frac{\ln(2.718)}{\sqrt{\frac{771/301}{301+771} + \frac{820/345}{345+820}}} = \frac{-.053 - .182}{\sqrt{.002 + .002}} = -3.72$$

Como este valor es menor que el valor crítico de −1.65, la prueba 1 es significativa.

Si se realiza la prueba 2 mediante la ecuación 6.10, tenemos que

$$Z_2 = \frac{\ln(.948) - \ln(.833)}{\sqrt{\frac{771/301}{301+771} + \frac{820/345}{345+820}}} = \frac{-.053 - (-.183)}{\sqrt{.002 + .002}} = 2.06$$

Como este valor es mayor que el valor crítico de 1.65, la prueba 2 es significativa. Como ambas pruebas son significativas, se rechaza la hipótesis nula de equivalencia. Por lo tanto, concluimos que la exposición y la no exposición producen resultados equivalentes, como lo define el intervalo de equivalencia. ■

6.4.4 Intervalo de confianza para la tasa de riesgo de muestras independientes

Razonamiento. Antes indicamos (sección 4.5 en la página 152) que se prefieren los intervalos de confianza a las pruebas de hipótesis. Así, en vez de preguntar: "¿Es diferente a 1 la tasa de riesgo de la población?", planteamos una pregunta más informativa: "¿Cuál es la tasa de riesgo de la población?" Si bien los métodos exactos para construir intervalos de confianza para *RR* son problemáticos, se dispone de métodos aproximados bastante simples.

El intervalo de confianza. Los límites inferiores y superiores para la tasa de riesgo de muestras independientes se pueden obtener por las siguientes ecuaciones.

$$\boxed{L = \exp\left[\ln(\widehat{RR}) - Z\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}\right]} \qquad (6.11)$$

y

$$\boxed{U = \exp\left[\ln(\widehat{RR}) + Z\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}\right]} \qquad (6.12)$$

Los símbolos exp y ln indican, respectivamente, que se deben tomar la exponencial y el logaritmo naturales de la expresión adjunta, en tanto que, a, b y c son frecuencias en una tabla de dos por dos, como se discutió previamente en relación con la tabla 6.7 de la página 239. $\widehat{RR}$ es la tasa de riesgo muestral, como lo define la ecuación 6.9 de la página 239.

EJEMPLO 6.17

En el ejemplo 6.13 de la página 240 se realizó una prueba de la hipótesis nula $H_0 : RR = 1$ usando los datos de la tabla dada. Utilice estos datos para construir un intervalo de confianza del 95%. ¿Qué significa este intervalo? ¿Implica este intervalo que la hipótesis nula $H_0 : RR = 1$ se rechazó? ¿Qué razones tiene para su respuesta?

		Cáncer de vejiga: sí	Cáncer de vejiga: no
Expuestos al solvente	sí	42	2,981
	no	21	4,088

Solución Por la ecuación 6.9, la tasa de riesgo muestral es

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)} = \frac{42/(42+2981)}{21/(21+4088)} = 2.718$$

Entonces, por la ecuación 6.11 de la página anterior

$$\begin{aligned} L &= \exp\left[\ln(\widehat{RR}) - Z\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}\right] \\ &= \exp\left[\ln(2.718) - 1.96\sqrt{\frac{2981/42}{42+2981} + \frac{4088/21}{21+4088}}\right] \\ &= \exp\left[1.000 - 1.96\sqrt{.023+.047}\right] \\ &= \exp[.4811] \\ &= 1.618 \end{aligned}$$

Y, por la ecuación 6.12 de la página anterior,

$$\begin{aligned} U &= \exp\left[\ln(\widehat{RR}) + Z\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}\right] \\ &= \exp\left[\ln(2.718) + 1.96\sqrt{\frac{2981/42}{42+2981} + \frac{4088/21}{21+4088}}\right] \\ &= \exp\left[1.000 + 1.96\sqrt{.023+.047}\right] \\ &= \exp[1.519] \\ &= 4.568 \end{aligned}$$

De este intervalo podemos concluir que las personas expuestas al solvente están entre 1.618 y 4.568 veces más en riesgo que las personas no expuestas. También podemos concluir que una prueba de dos colas de la hipótesis nula $RR = 1$ efectuada con $\alpha = .05$ sería significativa. Llegamos a esta conclusión observando que el valor nulo de 1.0 no está en el intervalo.[11] ■

EJEMPLO 6.18

En el ejemplo 6.14 de la página 240 se usaron datos ficticios de un estudio que evaluaba la función del ácido fólico en la prevención de la espina bífida, para llevar a cabo una prueba de la hipótesis nula $H_0 : RR = 1$ contra la alternativa $H_A : RR < 1$. Estos datos se dan en la tabla a continuación. Utilícelos para construir un intervalo de confianza de una cola del 95% para el límite superior de la tasa de riesgo de la población. Use este intervalo para efectuar la prueba antes mencionada. ¿Cómo justifica este intervalo su conclusión?

		Espina bífida	
		sí	no
Recomendación	sí	10	12,344
	no	17	11,202

Solución Por la ecuación 6.9 de la página 239, la tasa de riesgo muestral es

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)} = \frac{10/(10+12344)}{17/(17+11202)} = .534$$

Y por la ecuación 6.12 de la página 244,

$$\begin{aligned} U &= \exp\left[\ln(\widehat{RR}) + Z\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}\right] \\ &= \exp\left[\ln(.534) + 1.65\sqrt{\frac{12344/10}{10+12344} + \frac{11202/17}{17+11202}}\right] \\ &= \exp\left[-.627 + 1.65\sqrt{.100+.059}\right] \\ &= \exp[.031] \\ &= 1.031 \end{aligned}$$

Como el límite superior de 1.031 es mayor que 1, no se rechaza la hipótesis nula. Éste es el mismo resultado obtenido previamente con la prueba de hipótesis. ■

6.4.5 Suposiciones

Los métodos que se presentan aquí para realizar las pruebas de hipótesis, las pruebas de equivalencia y para construir los intervalos de confianza dependen de una aproximación a la curva normal. Los tamaños de las muestras deben ser lo bastante grandes para asegurar que la aproximación provea resultados suficientemente precisos. Estos procedimientos también requieren la suposición de independencia y generalmente no son sólidos ante las violaciones de ésta.

---

[11] Véase el apartado 4.5 de la página 152, si esta explicación no es del todo clara.

6.5 MÉTODOS RELACIONADOS CON RAZONES DE PROBABILIDAD DE MUESTRAS INDEPENDIENTES

6.5.1 Antecedentes

Otro estadístico, quizás empleado más comúnmente, usado para relacionar enfermedad y exposición es la razón de probabilidad. La **posibilidad** asociada con algún evento en un grupo en particular es la *probabilidad* de que el evento le ocurra a un miembro de ese grupo dividida entre la probabilidad de que el evento no le ocurra a un miembro de ese grupo. Por ejemplo, suponga que a un grupo de trabajadores industriales se les ha diagnosticado cáncer de vejiga. La posibilidad de que un miembro de ese grupo haya estado expuesto a un solvente industrial específico que se sospecha causa cáncer de vejiga es, usando la notación del capítulo 3, $\frac{P\ (E|D)}{P\ (\bar{E}|D)}$, que se lee: "La probabilidad de exposición dada la enfermedad dividida entre la probabilidad de no exposición dada la enfermedad." Si, por ejemplo, este valor fuera 2.0, diríamos que la probabilidad de exposición al solvente para los trabajadores con cáncer de vejiga es del doble de la probabilidad de que no hayan estado expuestos.

De igual forma, podemos calificar la posibilidad de exposición al solvente para los trabajadores en la misma fábrica, que no tienen cáncer de vejiga como $\frac{P\ (E|\bar{D})}{P\ (\bar{E}|\bar{D})}$, que se lee: "La probabilidad de exposición dada la no enfermedad dividida entre la probabilidad de no exposición dada la no enfermedad." Si, por ejemplo, este valor fuera .80, diríamos que la probabilidad de exposición al solvente para los trabajadores que no tienen cáncer de vejiga es .8 veces la probabilidad de que no hayan estado expuestos al solvente.

Parece razonable que un investigador en higiene industrial quiera comparar las posibilidades de que un trabajador con cáncer de vejiga haya estado expuesto al solvente contra las posibilidades de que un trabajador que no tiene cáncer de vejiga haya estado expuesto al solvente. Si la posibilidad de exposición para los trabajadores con cáncer de vejiga es mayor que la posibilidad para los trabajadores que no tienen cáncer de vejiga, la implicación puede ser una asociación entre exposición al solvente y cáncer de vejiga. Hay diferentes formas de hacer una comparación entre las dos posibilidades, incluyendo el cálculo de la diferencia por simple sustracción. Una manera utilizada con frecuencia para comparar las dos posibilidades es formarlas en un cociente denominado **razón de probabilidad** (*odds ratio*, que también se traduce como razón de oportunidades, razón de momios, razón de posibilidades y oportunidad relativa), que se define por la ecuación 6.13. Tome en cuenta que

$$OR = \frac{\dfrac{P\ (E|D)}{P\ (\bar{E}|D)}}{\dfrac{P\ (E|\bar{D})}{P\ (\bar{E}|\bar{D})}}$$

lo que se simplifica en

$$\boxed{OR = \frac{P\ (E|D)\ P\ (\bar{E}|\bar{D})}{P\ (\bar{E}|D)\ P\ (E|\bar{D})}} \qquad (6.13)$$

La ecuación 3.12 de la página 60[12] expresa la posibilidad de enfermedad para los grupos expuesto y no expuesto, en tanto que la ecuación 6.13 expresa la posibilidad de exposición para los grupos con y sin enfermedad. Ambas formas son utilizadas por los investigadores.

[12] En este punto quizá desee revisar la información de la página 60.

TABLA 6.8: Orden típico de los datos para el estudio cohorte potencial de muestras independientes.

		Enfermedad	
		sí	no
Expuestos	sí	a	b
	no	c	d

TABLA 6.9: Orden típico de los datos para el estudio de control de caso de muestras independientes.

		Expuestos	
		sí	no
Enfermedad	sí	a	b
	no	c	d

Será útil distinguir entre dos formas generales de diseños de investigación[13] con los que normalmente se emplea la razón de probabilidad. En el apartado 6.4.2 vimos un ejemplo en el que durante cierto tiempo se dio seguimiento a un grupo de trabajadores que por rutina estaban expuestos a un solvente, así como a trabajadores que no estaban expuestos, con el fin de determinar cuáles desarrollaban cáncer de vejiga. La razón de probabilidad calculada para este diseño expresaría la posibilidad de enfermedad (cáncer de vejiga) para los grupos de trabajadores expuestos (al solvente) y no expuestos. Los diseños de este tipo en los que sujetos expuestos y no expuestos son estudiados por algún tiempo para determinar quién desarrolla la enfermedad se denominan diseños de **cohorte potencial**. Los datos de este tipo de diseño normalmente se ordenarían para análisis como aparecen en la tabla 6.8.

En contraste, en esta sección vimos un ejemplo en el que se comparaba a trabajadores que ya habían desarrollado la enfermedad (cáncer de vejiga) con trabajadores que no habían desarrollado la enfermedad. La razón de probabilidad calculada para este diseño expresaría la posibilidad de exposición (al solvente) para los trabajadores con la enfermedad (cáncer de vejiga) y sin ella. Los diseños de este tipo, en los que se hacen comparaciones entre sujetos que ya tienen la enfermedad y sujetos que no la tienen para así determinar quién tiene una mayor exposición, se denominan diseños de **control de caso**. Los datos de este tipo de diseño normalmente se ordenarían para análisis como se presentan en la tabla 6.9.

Para la siguiente explicación, supondremos que se está discutiendo un estudio de control de caso; aunque con las debidas sustituciones de exposición por enfermedad y viceversa, los comentarios se aplican igualmente en los estudios de cohorte potencial.

Si la posibilidad de exposición[14] en los dos grupos es la misma, la razón de probabilidad es 1, lo que implicaría que no existe relación alguna entre enfermedad y exposición. Por otra parte, una razón de probabilidad mayor que 1 implicaría que la enfermedad está asociada con una mayor exposición que la ausencia de enfermedad. Los cocientes menores de 1 se denominan **protectores** porque enfermedad implica menos exposición que la ausencia de enfermedad. Esto podría ocurrir, por ejemplo, cuando las personas con la enfermedad tienen menos exposición a un suplemento vitamínico diario que las personas sin la enfermedad. Representaremos las formas de parámetro y estadística de la razón de probabilidad como OR y $\widehat{OR}$, respectivamente. La razón de

[13] Éstos sólo son dos de entre muchos diseños.

[14] Si bien nos referimos a exposición y enfermedad en la ecuación 6.13, los términos son genéricos y representan cualesquiera variables que se estén representando.

probabilidad muestral de muestras independientes se define por la ecuación 6.14. Observe que

$$\widehat{OR} = \frac{\frac{a}{b}}{\frac{c}{d}}$$

lo que se simplifica

$$\boxed{\widehat{OR} = \frac{ad}{bc}} \tag{6.14}$$

donde a, b, c y d son como lo muestran las tablas 6.8 y 6.9 de la página anterior.

6.5.2 Prueba de la hipótesis OR = 1 para muestras independientes

Razonamiento. Como en la situación anterior, suponga que se sospecha que la exposición a cierto solvente químico que se usa comúnmente en la manufactura incrementa el riesgo de cáncer de vejiga. Para probar esta sospecha, se interroga a los trabajadores de una industria en particular a los que se les ha diagnosticado cáncer de vejiga sobre si su trabajo implicaba la exposición al solvente. De igual manera, también se interroga a un grupo de trabajadores en la misma industria que no tienen la enfermedad en relación con la exposición a esa sustancia.

Con este dato, la razón de probabilidad puede calcularse mediante la ecuación 6.14. Si este cociente fuera 5.0, por ejemplo, la indicación sería que la posibilidad de exposición al solvente para los trabajadores con cáncer de vesícula es cinco veces la de los trabajadores libres de cáncer. Pero, ¿éste es simplemente un resultado de la muestra específica usada en el estudio? Si el estudio se repitiera con otros trabajadores, ¿el resultado sería notablemente diferente? Lo que es más importante, ¿la razón de probabilidad en la población 1 está indicando que no existe relación alguna entre enfermedad y exposición (es decir, la posibilidad en los dos grupos es la misma), o es diferente a 1, indicando que sí existe esa relación? Una prueba de hipótesis ayudará a responder esta pregunta.

La prueba. Por medio de la ecuación 6.15 se puede realizar una prueba de la hipótesis nula $OR = 1$ contra las alternativas de una cola o de dos colas. Las frecuencias a, b, c y d en esta ecuación son como se muestra en las tablas 6.8 y 6.9 de la página anterior. La razón de probabilidad planteada como hipótesis para la población se representa por $O\,R_0$. Como el logaritmo de 1 es cero, se puede omitir la expresión $\ln(O\,R_0)$ cuando se esté probando $H_0 : O\,R = 1$.

$$\boxed{Z = \frac{\ln(\widehat{OR}) - \ln(O\,R_0)}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}}} \tag{6.15}$$

EJEMPLO 6.19

Suponga que en el estudio sobre el cáncer de vejiga examinado con anterioridad, 13 de los trabajadores que tenían cáncer de vejiga estuvieron expuestos al solvente y 56 no. De los trabajadores sin cáncer de vejiga, cuatro habían estado expuestos al solvente y 65 no. Use este dato para calcular la razón de probabilidad muestral. ¿Qué significa este cociente? Utilice la ecuación 6.15 para probar la hipótesis nula $OR = 1$. Interprete sus resultados.

Solución Por conveniencia, los datos se ordenaron en una tabla de dos por dos, como se muestra en la tabla 6.10.

TABLA 6.10: Datos de un estudio de control de caso que relacionan el cáncer de vejiga con la exposición al solvente.

		Expuestos el solvente sí	Expuestos el solvente no
Cáncer de vejiga	sí	13	56
	no	4	65

De la ecuación 6.14 de la página anterior

$$\widehat{OR} = \frac{ad}{bc} = \frac{(13)(65)}{(56)(4)} = 3.772$$

Esto significa que la posibilidad de exposición al solvente para los trabajadores con cáncer de vejiga fue 3.772 veces la de los trabajadores que no padecen la enfermedad.

Por la ecuación 6.15 de la página anterior,

$$Z = \frac{\ln(\widehat{OR}) - \ln(OR_0)}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}} = \frac{\ln(3.772)}{\sqrt{\frac{1}{13} + \frac{1}{56} + \frac{1}{4} + \frac{1}{65}}} = \frac{1.328}{.600} = 2.21$$

Los valores críticos para una prueba Z de dos colas conducida con $\alpha = .05$, como se reportó en el Apéndice A, son 1.96 y −1.96. Como el valor Z obtenido de 2.21 excede 1.96, se rechaza la hipótesis nula. Por consiguiente, se concluye que la razón de probabilidad de la población es mayor que 1, lo que indica que los trabajadores con cáncer de vejiga tienen una posibilidad mayor de haber estado expuestos al solvente que los trabajadores que no tienen la enfermedad. ■

EJEMPLO 6.20

En el ejemplo 6.14 de la página 240 usamos los datos de la siguiente tabla para evaluar la relación entre la ingestión de ácido fólico y la espina bífida. Encontramos que la razón de probabilidad muestral fue .534. Una prueba posterior de la hipótesis nula $H_0 : RR = 1$ no fue significativa. Utilice este dato para calcular e interpretar la razón de probabilidad muestral. Realice una prueba de la hipótesis nula $H_0 : OR = 1$ contra la alternativa $H_A : OR < 1$ en el nivel .05. Interprete los resultados de esta prueba.

		Espina bífida sí	Espina bífida no
Recomendación	sí	10	12,344
	no	17	11,202

Solución Por la ecuación 6.14 de la página anterior

$$\widehat{OR} = \frac{ad}{bc} = \frac{(10)(11202)}{(12344)(17)} = .534$$

Observe que se acerca mucho al valor obtenido para $\widehat{RR}$. Es lo que normalmente ocurre cuando se trata de alguna enfermedad inusual. Entonces, por la ecuación 6.15 de la página 249,

$$Z_1 = \frac{\ln(\widehat{OR}) - \ln(OR_0)}{\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}} = \frac{\ln(.534)}{\sqrt{\frac{1}{10}+\frac{1}{12344}+\frac{1}{17}+\frac{1}{11202}}} = \frac{-.627}{.399} = -1.57$$

El valor crítico para una prueba Z de una cola realizada con $\alpha = .05$, como se reportó en el Apéndice A, es -1.65. Ya que el valor Z obtenido de -1.57 es mayor que este valor crítico, no se rechaza la hipótesis nula. Por consiguiente, no pudimos demostrar un efecto protector para las madres a las que se les alentó a incluir ácido fólico en su dieta. ■

6.5.3 Cómo establecer la equivalencia por medio de la razón de probabilidad de muestras independientes

Razonamiento. Como en el caso de las diferencias entre medias, las diferencias entre proporciones y las tasas de riesgo, a veces se desea demostrar que la posibilidad asociada con alguna condición es equivalente a la posibilidad asociada con alguna otra condición. Esto implicaría el uso de una prueba de equivalencia de dos colas basada en razones de probabilidad. Quizá es más común el caso en el cual se desea demostrar que la posibilidad asociada con una condición *no es mayor que* la posibilidad asociada con otra condición. Esto implica una prueba de una cola.

Suponga, por ejemplo, que epidemiólogos militares realizarán un estudio para evaluar la relación entre cierta enfermedad cutánea que aflige a algunos miembros del personal de una base militar y la exposición a una munición en particular. Después de la debida consideración, se determina que el estudio debe intentar demostrar que la posibilidad de exposición a la munición para el personal con la enfermedad cutánea no es mayor que la posibilidad para otros miembros del personal de la base que no padecen le enfermedad. También se determina que una razón de probabilidad menor a 1.1 bastará para demostrar que la posibilidad de exposición para las personas con la enfermedad no fue mayor que la posibilidad de las personas sin la enfermedad.

Note que en un estudio estándar (es decir, de no equivalencia) el peso de la prueba sería demostrar que la exposición *fue* mayor, mientras que en este estudio la intención es demostrar que la exposición *no fue* mayor. De cierta manera, el estudio estándar supondría que la munición es segura hasta probar que es insegura, mientras que el estudio de equivalencia supone que la munición es insegura hasta que se demuestre que es segura. La diferencia entre los estudios estándar y de equivalencia no siempre es fácil de comprender, pero la distinción es lo suficientemente importante para justificar el esfuerzo que se necesita para hacer que sea clara.

La prueba. Como ocurre con las tasas de riesgo, los intervalos de equivalencia para las razones de probabilidad, por lo regular, pero no necesariamente, son simétricos respecto a 1.0 en el sentido de que $EI_U = \frac{1}{EI_L}$ y viceversa. Así, si EI_L es .8, EI_U normalmente sería $\frac{1.0}{.8} = 1.25$.

Usando la notación de la sección 4.3.6 de la página 117, las hipótesis nula y alternativa para una prueba de equivalencia de dos colas para las razones de probabilidad de muestras independientes se enuncian como

$$H_{0E}: OR \le EI_L \quad \text{o} \quad OR \ge EI_U$$
$$H_{AE}: EI_L < OR < EI_U$$

donde EI_L y EI_U son los límites inferior y superior del intervalo de equivalencia. En esencia, la hipótesis nula plantea que la razón de probabilidad no reside en el intervalo de equivalencia y la hipótesis alternativa afirma que la razón de probabilidad está en el intervalo. Como recordará, la hipótesis nula se comprueba efectuando dos pruebas de una cola en el nivel α. Para poder rechazar la hipótesis nula de equivalencia, debe demostrarse que $OR > EI_L$ *y* que $OR < EI_U$, lo que significa que *ambas* pruebas siguientes deben ser significativas.

Prueba 1	Prueba 2
$H_{01} : OR = EI_U$	$H_{02} : OR = EI_L$
$H_{A1} : OR < EI_U$	$H_{A2} : OR > EI_L$

Las hipótesis nula y alternativa para la prueba de equivalencia de una cola son *una* de las siguientes:

$$H_{0E} : OR \geq EI_U$$
$$H_{AE} : OR < EI_U$$

o

$$H_{0E} : OR \leq EI_L$$
$$H_{AE} : OR > EI_L$$

La primera hipótesis de equivalencia de una cola enunciada antes se somete a la prueba 1 y la segunda a la prueba 2.

EJEMPLO 6.21

Calcule la razón de probabilidad para los datos sobre la enfermedad cutánea contenidos en la siguiente tabla. Suponiendo que una razón de probabilidad de menos de 1.1 sería aceptada para demostrar que la posibilidad de exposición del personal con la enfermedad no fue mayor que la posibilidad de exposición del personal sin la enfermedad, realice una prueba de equivalencia de una cola para demostrar que la razón de probabilidad es menor que este valor. Enuncie las hipótesis nula y alternativa de equivalencia e interprete los resultados de la prueba.

		Expuestos a la munición	
		sí	no
Enfermedad cutánea	sí	126	819
	no	119	811

Solución Por la ecuación 6.14 de la página 249, la razón de probabilidad muestral es

$$\widehat{OR} = \frac{ad}{bc} = \frac{(126)(811)}{(819)(119)} = 1.048$$

Las hipótesis nula y alternativa para la prueba de equivalencia de una cola son:

$$H_{0E} : OR \geq 1.1$$
$$H_{AE} : OR < 1.1$$

Entonces, al efectuar la prueba 1 mediante la ecuación 6.15 de la página 249,

$$Z_1 = \frac{\ln(\widehat{OR}) - \ln(OR_0)}{\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}} = \frac{\ln(1.048) - \ln(1.100)}{\sqrt{\frac{1}{126}+\frac{1}{819}+\frac{1}{119}+\frac{1}{811}}} = \frac{-.048}{.137} = -.350$$

Como ese valor es mayor que el valor crítico de −1.65, no se rechaza la hipótesis nula. Por lo tanto, los investigadores no pudieron demostrar que la posibilidad de exposición para el personal con la enfermedad cutánea no es mayor que la posibilidad para el personal sin la enfermedad. ■

EJEMPLO 6.22

Utilice los datos de la siguiente tabla para calcular $\widehat{OR}$. ¿Qué significa ese cociente? Después, lleve a cabo una prueba de equivalencia de dos colas usando un *EI* que va de .833 a 1.200. Interprete sus resultados.

		Variable 2	
		sí	no
Variable 1	sí	48	12
	no	24	6

Solución Por la ecuación 6.14 de la página 249, la razón de probabilidad muestral es

$$\widehat{OR} = \frac{ad}{bc} = \frac{(48)(6)}{(12)(24)} = 1.000$$

Este cociente indica que la posibilidad de un “sí” para la variable 2 es la misma para los que tienen un “sí” para la variable 1, así como para los que tienen un “no” para la variable 1. Podemos mover el ejemplo de lo abstracto a una forma más concreta si sustituimos “enfermedad” por la variable 1 y “exposición” por la variable 2. Ahora podemos interpretar que el cociente indica que las posibilidades de exposición son las mismas para aquellos con enfermedad y sin enfermedad.

La prueba de equivalencia de dos colas se realiza mediante la prueba 1 *y* la prueba 2, como se efectuó en la ecuación 6.15

$$Z_1 = \frac{\ln(\widehat{OR}) - \ln(OR_0)}{\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}} = \frac{\ln(1.000) - \ln(1.200)}{\sqrt{\frac{1}{48}+\frac{1}{12}+\frac{1}{24}+\frac{1}{6}}} = \frac{-.183}{.559} = -.327$$

y

$$Z_2 = \frac{\ln(1.000) - \ln(.833)}{\sqrt{\frac{1}{48}+\frac{1}{12}+\frac{1}{24}+\frac{1}{6}}} = \frac{.182}{.559} = -.326$$

Del Apéndice A se puede ver que los valores críticos para la prueba 1 y la prueba 2 son −1.65 y 1.65, respectivamente. Como ninguna de estas pruebas es significativa, no se rechaza la hipótesis nula de equivalencia. Recordará que *ambas* pruebas deben ser significativas para que se rechace la hipótesis nula. Por lo tanto, no pudimos establecer la equivalencia en este caso. ■

6.5.4 Intervalo de confianza para la razón de probabilidad de muestras independientes

Razonamiento. Antes indicamos (apartado 4.5 en la página 152) que, generalmente, se prefieren los intervalos de confianza a las pruebas de hipótesis. Así, en vez de preguntar: "¿Es la razón de probabilidad de la población diferente de 1?", una pregunta más informativa es: "¿Cuál es la razón de probabilidad de la población?" Mientras que los métodos exactos para construir intervalos de confianza para *OR* son problemáticos, existen métodos aproximados suficientemente simples.

El intervalo de confianza. Los límites inferior y superior para la razón de probabilidad de muestras independientes se pueden obtener mediante las siguientes ecuaciones.

$$\boxed{L = \exp\left[\ln(\widehat{OR}) - Z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right]} \tag{6.16}$$

y

$$\boxed{U = \exp\left[\ln(\widehat{OR}) + Z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right]} \tag{6.17}$$

Los símbolos exp y ln indican, respectivamente, que se deben tomar la exponencial y el logaritmo naturales de la expresión adjunta, en tanto que a, b, c y d son frecuencias en una tabla de dos por dos, como se explicó previamente en relación con las tablas 6.8 y 6.9 de la página 248. $\widehat{OR}$ es la razón de probabilidad muestral, como lo define la ecuación 6.14 (página 249).

EJEMPLO 6.23

En el ejemplo 6.19 de la página 249 usamos los datos de la tabla 6.11 de la página siguiente para llevar a cabo una prueba de dos colas de la hipótesis nula $H_0 : OR = 1$. Utilice este dato para construir un intervalo de confianza de dos colas del 95% para el cálculo de *OR*. ¿Qué significa este intervalo? ¿Este intervalo implica que la hipótesis nula $H_0 : OR = 1$ se rechazó? ¿Cuál fue la razón para su respuesta?

Solución Por la ecuación 6.14 de la página 249, la razón de probabilidad muestral es

$$\widehat{OR} = \frac{ad}{bc} = \frac{(13)(65)}{(56)(4)} = 3.772$$

Entonces, por las ecuaciones 6.16 y 6.17

$$\begin{aligned}
L &= \exp\left[\ln(\widehat{OR}) - Z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right]w \\
&= \exp\left[\ln(3.772) - 1.96\sqrt{\frac{1}{13}+\frac{1}{56}+\frac{1}{4}+\frac{1}{65}}\right] \\
&= \exp[1.328 - 1.96\ (.600)] \\
&= \exp[.152] \\
&= 1.164
\end{aligned}$$

TABLA 6.11: Datos de un estudio de control de caso que relacionan el cáncer de vejiga con la exposición al solvente.

		Expuestos el solvente	
		sí	no
Cáncer de vejiga	sí	13	56
	no	4	65

y

$$
\begin{aligned}
U &= \exp\left[\ln(\widehat{OR}) + Z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right] \\
&= \exp\left[\ln(3.772) + 1.96\sqrt{\frac{1}{13}+\frac{1}{56}+\frac{1}{4}+\frac{1}{65}}\right] \\
&= \exp\,[1.328 + 1.96\ (.600)] \\
&= \exp\,[2.504] \\
&= 12.231
\end{aligned}
$$

Así, podemos tener un 95% de confianza de que las posibilidades de enfermedad para las personas expuestas al solvente son entre 1.164 y 12.231 veces las de personas no expuestas.

A partir de este intervalo podemos afirmar que una prueba de dos colas de la hipótesis nula $H_0: OR = 1$ sería rechazada. Esto se deduce del hecho de que el valor estipulado por la hipótesis nula, 1 en este caso, no está en el intervalo. ■

EJEMPLO 6.24

En el ejemplo 6.20 de la página 250 usamos datos ficticios de un estudio que evaluaba la función del ácido fólico en la prevención de espina bífida para desempeñar una prueba de la hipótesis nula H_0: $OR = 1$ contra la alternativa $H_A: OR < 1$. Estos datos se presentan en la tabla a continuación. Utilice estos datos para construir un intervalo de confianza de una cola para el límite superior de la razón de probabilidad de la población. Utilice este intervalo para realizar la prueba que aquí se menciona. ¿Cómo justifica este intervalo su conclusión?

		Espina bífida	
		sí	no
Recomendación	sí	10	12,344
	no	17	11,202

Solución Por la ecuación 6.14 de la página 249, la razón de probabilidad muestral es

$$\widehat{OR} = \frac{ad}{bc} = \frac{(10)\ (11202)}{(12344)\ (17)} .534$$

Entonces, por la ecuación 6.17 de la página 254,

$$
\begin{aligned}
U &= \exp\left[\ln(\widehat{OR}) + Z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right] \\
&= \exp\left[\ln(.534) + 1.65\sqrt{\frac{1}{10}+\frac{1}{12344}+\frac{1}{17}+\frac{1}{11202}}\right] \\
&= \exp\left[-.627 + 1.65\,(.399)\right] \\
&= \exp\left[.031\right] \\
&= 1.031
\end{aligned}
$$

Este límite superior indica que la posibilidad de tener un bebé afectado por espina bífida para las madres a quienes se les recomendó que incluyeran ácido fólico en su dieta es 1.031 *o menos* veces la de las madres que no recibieron esta recomendación. Como el límite superior de 1.031 es mayor que 1, no se rechaza la hipótesis nula. Éste es el mismo resultado obtenido previamente con la prueba de hipótesis. ■

6.5.5 Suposiciones

Los métodos que se presentan aquí para realizar las pruebas de hipótesis, las pruebas de equivalencia y para construir los intervalos de confianza dependen de una aproximación de la curva normal. Los tamaños de las muestras deben ser lo bastante grandes para asegurar que la aproximación provea resultados suficientemente precisos. Estos procedimientos también requieren la suposición de independencia, y generalmente no son sólidos ante las violaciones de ésta.

6.5.6 Cómo estimar el riesgo de enfermedad a partir de datos de un control de caso

En las páginas 239 y 248 hablamos brevemente de dos diseños de estudio comunes, los estudios de cohorte y de control de caso. Posteriormente, usamos estos diseños para demostrar el cálculo y la interpretación de la tasa de riesgo y la razón de probabilidad. En esta sección hacemos más comentarios respecto del uso de estos estadísticos en relación con datos recabados en el contexto de los estudios de control de caso.

Como recordará, en los estudios de control de caso se compara a los sujetos con enfermedad con los sujetos sin enfermedad en cuanto a su exposición a algún factor de riesgo. Estos tipos de estudios normalmente se llevan a cabo más fácilmente que los estudios de cohorte, los cuales requieren dar seguimiento por un tiempo a los sujetos con y sin exposición a algún factor de riesgo para determinar cuáles manifiestan alguna enfermedad. No obstante, una deficiencia de los estudios de control de caso es que, aunque permiten evaluar el riesgo o la posibilidad de exposición, no se puede determinar el riesgo de enfermedad. El riesgo de enfermedad, a menudo, es de interés primordial para los investigadores de las ciencias de la salud. En esta sección demostraremos por qué el riesgo de enfermedad no puede determinarse directamente de los estudios de control de caso y más adelante se mostrará que, en ciertas condiciones, es posible obtener un *estimado* de la tasa de riesgo para la enfermedad en los estudios de control de caso mediante el uso de la razón de probabilidad.

Para facilitar la explicación, supongamos que las celdas de la tabla 6.12 de la siguiente página contienen frecuencias de *población* para exposición y enfermedad. Para distinguir estas celdas de frecuencias de población de las frecuencias de la muestra, *a*, *b*, *c* y *d*, usaremos las denominaciones

TABLA 6.12: Tabla de contingencia de la población de la relación entre exposición y enfermedad.

		Enfermedad	
		sí	no
Expuestos	sí	A	B
	no	C	D

correspondientes A, B, C y D. Usando esta notación, el riesgo de enfermedad para los sujetos expuestos es la proporción de los sujetos expuestos en la población que tienen la enfermedad o

$$\frac{A}{A+B}$$

De igual forma, el riesgo de enfermedad en el grupo no expuesto es

$$\frac{C}{C+D}$$

De ello, se sigue que

$$RR = \frac{\frac{A}{A+B}}{\frac{C}{C+D}}$$

Es claro que para determinar RR, debemos ser capaces de calcular la proporción de sujetos en los grupos expuesto y no expuesto que tienen la enfermedad. Esto implica que los datos de la muestra deben ser representativos de los grupos expuesto y no expuesto en la población. Pero, en un estudio de control de caso, el investigador *selecciona* el número de personas que tienen y no tienen la enfermedad. Esto significa que el investigador fija la proporción de personas con o sin enfermedad en los grupos expuesto y no expuesto. Esto, desde luego, es inaceptable para calcular la posibilidad y el riesgo de enfermedad. El método más efectivo para calcular la posibilidad y el riesgo de enfermedad es obtener muestras aleatorias de miembros expuestos y no expuestos de la población.

Entonces, de manera ideal, desearíamos muestrear aleatoriamente a sujetos expuestos y no expuestos de la población para calcular RR para enfermedad. Pero, ¿qué pasaría en un estudio de control de caso? Como recordará, en este ejemplo se selecciona a los sujetos no con base en la exposición/no exposición, sino con base en el estatus de enfermedad/no enfermedad. Mientras que una muestra aleatoria de sujetos con y sin enfermedad permitiría cálculos del riesgo de exposición en estos dos grupos, las muestras resultantes no proveerían información alguna respecto del riesgo de enfermedad en los grupos expuesto y no expuesto. Como estos datos permiten el cálculo del riesgo en los grupos enfermedad/no enfermedad, podemos estimar el riesgo de exposición, pero no el de enfermedad.

Sin embargo, observe que cuando la enfermedad es inusual, A y C son pequeñas en comparación con B y D. Esto implica que

$$\frac{A}{A+B} \approx \frac{A}{B}$$

y

$$\frac{C}{C+D} \approx \frac{C}{D}$$

(El símbolo ≈ se lee "es aproximadamente igual a".) Así, *cuando la enfermedad es inusual*,

$$\text{RR}= \frac{\frac{A}{A+B}}{\frac{c}{c+D}} \approx \frac{\frac{A}{B}}{\frac{C}{D}} = \frac{AD}{BC}$$

La expresión

$$\frac{AD}{BC}$$

es la razón de probabilidad de la población, que se estima por la razón de probabilidad muestral. Todo esto implica que, cuando la enfermedad es inusual, la razón de probabilidad muestral para la enfermedad puede usarse para calcular la tasa de riesgo de enfermedad de la población. Por ejemplo, si $A = 100$, $B = 5{,}000$, $C = 75$ y $D = 4{,}900$,

$$\text{RR} = \frac{\frac{100}{100+5000}}{\frac{75}{75+4900}} \; 1.30 \approx \frac{(100)\ (4900)}{(5000)\ (75)} = 1.31$$

Muchas de las situaciones relacionadas con el uso de tasas de riesgo y razones de probabilidad para el análisis de datos recabados en diversos contextos de investigación van más allá del alcance de este libro. Para más detalles, es recomendable consultar textos de investigación epidemiológica como Breslow y Day [5] o Lilienfeld y Lilienfeld [31].

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

equivalencia de μ_1 y μ_2 224
equivalencia de posibilidades 251
equivalencia de riesgo 241
equivalencia para π_1 y π_2 234
estudio de cohorte potencial 239, 248
estudio de control de caso 248
exposición protectora 238, 248
homogeneidad de varianza 230
IC para $\mu_1 - \mu_2$ 228
IC para $\pi_1 - \pi_2$ 236
IC para *OR* 254
IC para *RR* 244
$\mu_1 - \mu_2$ 215
$\hat{p}_1 - \hat{p}_2$ 231
posibilidad 247
prueba para $\pi_1 - \pi_2$ 230
prueba para $OR = 1$ 249
prueba para $RR = 1$ 239
razón de probabilidad 247
riesgo 238
RR y *OR* cuando la enfermedad es inusual 258
s_P^2 219
suposición de normalidad 230
tasa de riesgo 238
$\bar{x}_1 - \bar{x}_2$ 217

EJERCICIOS

6.1 Se realiza un estudio para determinar si los niños que viven en un radio de una milla cerca de un restaurante de comida rápida tienen un índice promedio de masa corporal más alto (IMC) que los niños que no viven tan cerca. Como regla, las categorías IMC representan lo siguiente: 18.5 o menos = bajo peso, 18.5 a 24.9 = normal, 25.0 a 29.9 = sobrepeso y 30.0 o más significa obesidad. Los datos se presentan a continuación.

Realice una prueba de la hipótesis $\mu_1 = \mu_2$ contra la alternativa $\mu_1 > \mu_2$ en el nivel de .05 de significancia. Denomine el grupo "en un radio de una milla" como el grupo 1. ¿A qué conclusión llega respecto de esta pregunta?

Grupo uno	Grupo dos
26.2	25.9
24.5	20.1
20.0	22.2
30.2	29.7
28.4	28.0
18.6	29.4
21.5	20.2
21.7	20.7
29.9	26.3
18.3	18.2

6.2 Suponga que se descubre que una escuela primaria ha sido construida cerca del lugar de una fábrica de sustancias químicas desaparecida hace tiempo. Ahora, se sospecha que las sustancias producidas por la fábrica causan problemas respiratorios complejos. El consejo de la escuela debe probar que el terreno es seguro en cuanto a la salud del aparato respiratorio se refiere.

Como parte de su estudio, el consejo envía una lista de 30 síntomas respiratorios a todos los alumnos que asistieron a la escuela primaria durante cuatro años o más y que ahora tienen 30 años o más. La misma lista se envía a un grupo de antiguos estudiantes que cumplen con las mismas características en lo que se refiere a asistencia a una escuela primaria, pero sin que esta última estuviera cercana al lugar de la planta química. Se determina que si se puede demostrar que la media de los síntomas marcados por el grupo que asistió a la escuela en cuestión es mayor por menos de dos que la media marcada por los alumnos de la escuela control, la escuela sospechosa será considerada segura en lo concerniente a síntomas respiratorios.

A continuación se dan las respuestas de los asistentes de ambas escuelas. Denomine a los asistentes de la escuela sospechosa como el grupo 1.

Grupo uno	Grupo dos
6	12
4	1
20	22
17	19
9	3
2	0
8	5
14	9
12	11
0	15
24	5
8	17

a) Realice una prueba de equivalencia para cumplir con la responsabilidad dada al consejo de la escuela. ¿Ésta sería más probablemente una prueba de una cola o de dos colas? ¿Por qué?

b) Enuncie las hipótesis nula y alternativa de equivalencia.

c) ¿Cómo podría el análisis ser diferente si el consejo de la escuela fuera responsable de demostrar que el terreno de la escuela es inseguro?

6.3 Utilice los datos del ejercicio 6.1 para construir un intervalo de confianza de dos colas del 95%. ¿Qué calcula este intervalo?

6.4 Utilice los datos del ejercicio 6.1 para construir un intervalo de confianza de una cola del 95% para el límite inferior de $\mu_1 - \mu_2$. ¿Este intervalo sustenta el resultado de su prueba de hipótesis para estos datos? ¿Cómo?

6.5 Suponga que una encuesta realizada por un grupo educativo en contra del tabaquismo demuestra que .312 de los varones del último año de bachillerato y .288 de las mujeres reportan que, por rutina, fuman uno o más cigarrillos al día. Suponiendo que la encuesta se aplicó a 1,777 varones y 1,821 mujeres, lleve a cabo una prueba de dos colas de la hipótesis nula $\pi_1 = \pi_2$ con un nivel de significancia de .05. Interprete el resultado.

6.6 Los investigadores sospechan que cierto tratamiento médico, muy costoso y usado con frecuencia, no representa más beneficio para los pacientes que un placebo. Para investigar esta impresión, se lleva a cabo un estudio a gran escala en el que se asigna a 1,400 pacientes al azar para recibir el tratamiento de cuidado estándar o un placebo. De los 700 pacientes que reciben el tratamiento activo, 313 reportan un efecto benéfico. En contraste, 317 de los pacientes tratados con el placebo reportan un beneficio. Los investigadores determinan que es razonable declarar el tratamiento activo "no mejor que el placebo" si la proporción de pacientes que se benefician del tratamiento es mayor, por menos de .04, que la proporción de los que se benefician del placebo.
Use una prueba de equivalencia con $\alpha = .05$ para analizar el resultado del estudio. Enuncie las hipótesis nula y alternativa antes de efectuar la prueba.

6.7 Utilice los datos del ejercicio 6.5 para construir un intervalo de confianza de dos colas del 95%. Si este intervalo se usara para conducir una prueba de dos colas de $H_0 : \pi_1 = \pi_2$ con un nivel de significancia de .05, ¿el resultado sería significativo? ¿Por qué? ¿Qué calcula el intervalo?

6.8 Se cree que los pacientes que padecen de neuritis óptica (caracterizada por la inflamación del nervio óptico) y que tienen tres o más lesiones en el tallo cerebral (demostrado por una resonancia magnética) corren un riesgo mayor de desarrollar esclerosis múltiple (EM) en los próximos 10 años que los pacientes con neuritis óptica que no manifiestan tales lesiones. Para examinar esta opinión, se da seguimiento a 719 pacientes de neuritis óptica, de los cuales 291 tienen tres o más lesiones y 428 no. De aquellos pacientes con lesiones, 196 desarrollaron EM y 191 de aquellos sin lesiones desarrollaron la enfermedad.
a) Calcule la tasa de riesgo comparando el riesgo de EM para los pacientes con y sin lesiones en el tallo cerebral. Interprete este cociente.
b) Pruebe la hipótesis $H_0 : RR = 1$ contra la alternativa $H_A : RR \neq 1$ con un nivel de significancia de .05.

6.9 Por mucho tiempo se ha afirmado que las personas que residen cerca de líneas de alto voltaje tienen un riesgo mayor de desarrollar ciertas formas de cáncer que las personas que no viven cerca de estas líneas. Muchos estudios han fallado al demostrar una relación entre la proximidad a estas líneas y el riesgo de cáncer. Una compañía eléctrica que busca el permiso para construir una red de suministro cerca de un desarrollo habitacional se ha enfrentado a la resistencia del Consejo de Planeación del Condado (CPC). El consultor del CPC afirma que mientras los estudios anteriores no han podido demostrar la relación entre la cercanía de residencia a estas líneas y los tipos de cáncer en cuestión, no ha habido estudios para demostrar la seguridad de la proximidad a tales líneas.
Un estudio comisionado por la compañía eléctrica muestra que de las 9,848 personas que residen en un radio de 500 yardas cerca de las líneas de alto voltaje, 590 han desarrollado uno de los tipos de cáncer en cuestión. De los 13,112 residentes que viven a más de 500 yardas de esas líneas, 577 contrajeron uno de los tipos de cáncer.
a) Usando tasas de riesgo, lleve a cabo una prueba de equivalencia utilizando el intervalo de confianza .91 a 1.1 y $\alpha = .05$.
b) Enuncie las hipótesis nula y alternativa para la prueba de equivalencia.
c) Enuncie la hipótesis nula que usted cree fue utilizada para los estudios anteriores que fallaron al demostrar la relación entre las variables en cuestión.
d) Analice la diferencia entre las pruebas anteriores (comunes) y la prueba de equivalencia usada aquí en lo que se refiere a objetivos de análisis.

6.10 Utilice los datos del ejercicio 6.8 para construir un intervalo de confianza de dos colas del 95% para el cálculo de RR. Interprete este intervalo. ¿Este intervalo sustenta el resultado de su prueba de hipótesis? ¿Por qué?

6.11 Vuelva a analizar el problema descrito en el ejercicio 6.8 usando razones de probabilidad en lugar de tasas de riesgos.

6.12 Responda las preguntas del ejercicio 6.9 usando nuevamente razones de probabilidad en vez de tasas de riesgo.

6.13 Utilice los datos del ejercicio 6.8 para construir un intervalo de confianza de dos colas del 95% para el cálculo de OR. Interprete este intervalo. ¿Este intervalo sustenta el resultado de su prueba de hipótesis? ¿Por qué?

A. Las siguientes preguntas se refieren al estudio de caso A (página 469).

6.14 Utilice los datos a.m. para construir un intervalo de confianza de dos colas del 95% para el cálculo de $\pi_n - \pi_y$, donde π_n representa la proporción de sujetos que contestaron no a la pregunta sobre duración que escogieron el lente tratado y π_y representa la misma proporción para el grupo que contestó sí a la pregunta sobre duración. Interprete este intervalo.

6.15 Use el IC construido en el ejercicio 6.14 para llevar a cabo una prueba de dos colas de la hipótesis nula $H_0 : \pi_n = \pi_y$ con un nivel de .05. Explique por qué rechazó o no rechazó la hipótesis nula.

6.16 Construya el intervalo de confianza descrito en el ejercicio 6.14 usando los datos p.m. Lleve a cabo la prueba de hipótesis relacionada. ¿Qué puede decir sobre los resultados de las dos pruebas en lo que se refiere al estudio?

B. Las siguientes preguntas se refieren al estudio de caso B (página 470).

6.17 ¿Usted esperaría que una prueba *t* de dos colas de muestras independientes que usa como línea de base los datos WOMAC A para comparar las medias de los grupos estándar y placebo, realizada con α = .05, produciría un resultado significativo? Explique su respuesta. Suponga que el resultado *fuera* significativo, ¿cómo explicaría este hecho?

6.18 Utilice los datos de la línea de base WOMAC A para construir un intervalo de confianza de dos colas del 95% para el cálculo de la diferencia entre las medias de los grupos estándar y placebo. (*Sugerencia:* Piense en cómo se pueden usar la desviación estándar de la muestra y el tamaño de la muestra para obtener la suma de cuadrados de un grupo en particular.) ¿Este intervalo sustenta su opinión en 6.17?

6.19 Utilice los datos de 12 semanas de WOMAC A para construir un intervalo de confianza de dos colas del 95% para el cálculo de la diferencia entre las medias de los grupos estándar y placebo. Desde el punto de vista del investigador, ¿qué se calcula con este intervalo?

6.20 Utilice el intervalo construido en el ejercicio 6.19 para llevar a cabo una prueba de dos colas de la hipótesis $H_0 : \mu_s = \mu_p$, donde μ_s y μ_p representan las medias de los grupos estándar y placebo, respectivamente. Explique cómo llegó a su conclusión sobre si había que rechazar la hipótesis nula.

6.21 Utilice los datos de la tabla J.3 para probar la siguiente hipótesis nula contra las alternativas de dos colas con un nivel de significancia de .05. Observe que π_s, π_w y π_p representan las proporciones en cada grupo que identificaron correctamente el tipo de brazalete utilizado. Una identificación correcta para el grupo débil es "de imitación".

a) $H_0 : \pi_s = \pi_w$
b) $H_0 : \pi_s = \pi_p$
c) $H_0 : \pi_w = \pi_p$

6.22 Usando μ_w y μ_p para representar la puntuación media WOMAC A para los grupos débil y placebo a las 12 semanas, respectivamente, lleve a cabo la siguiente prueba con un nivel de significancia de .05. ¿El resultado justifica tratar a los grupos débil y placebo como equivalentes?

$$H_{0E} : \mu_w - \mu_p \leq -1.0 \quad \text{o} \quad \mu_w - \mu_p \geq 1.0$$
$$H_{AE} : -1.0 < \mu_w - \mu_p < 1.0$$

6.23 Usando π_w y π_p para representar la proporción de sujetos en los grupos débil y placebo respectivamente, que indican que están usando un brazalete de imitación, lleve a cabo la siguiente prueba con un nivel de significancia de .05. ¿El resultado justifica tratar a los grupos débil y placebo como equivalentes?

$$H_{0E} : \pi_w - \pi_p \leq -.02 \quad \text{o} \quad \pi_w - \pi_p \geq .02$$
$$H_{AE} : -.02 < \pi_w - \pi_p < .02$$

C. Las siguientes preguntas se refieren al estudio de caso C (página 471).

6.24 ¿Sobre qué estadísticos (por ejemplo, diferencia entre medias, etc.) podría basarse una evaluación de los dos tratamientos?

6.25 ¿Sugeriría usted que se utilice una prueba de hipótesis o un intervalo de confianza para este propósito?

6.26 Calcule una tasa de riesgo comparando el riesgo de muerte para el grupo invasivo con el del grupo no invasivo para cada uno de los periodos. Interprete cada uno de estos cocientes.

6.27 Construya intervalos de confianza del 95% para el cálculo de *RR* para cada uno de los periodos.

6.28 ¿Cuál es su conclusión sobre la efectividad de los dos tratamientos?

D. Las siguientes preguntas se refieren al estudio de caso D (página 471).

6.29 Los investigadores reportan que "los pacientes infectados con VIH tuvieron puntuaciones NPZ-8 significativamente más bajas (t[18] = 2.26, $P < .05$) y PBV más bajos (t[18] = 1.79, $P < .01$) que los participantes del control de salud". La notación t[18] = 2.26, $P < .05$ significa que se realizó una prueba con 18 grados de libertad, que produjo un *t* obtenida de 2.26 con un valor-*p* asociado de menos de .05. Una interpretación similar se aplica a la prueba de las puntuaciones PBV.

a) ¿Qué tipo de prueba *t* piensa que usaron los investigadores? ¿Qué le hace creer que éste es el caso?
b) ¿Ve usted algo peculiar sobre los valores-*p* reportados?
c) Realice las mismas dos pruebas. ¿Obtiene el mismo valor *t* obtenido? Suponiendo pruebas de dos colas, indique si cada prueba es significativa al nivel .05 o .01, o si no es significativa a ningún nivel.

d) ¿Observó usted algo respecto de estas pruebas que pudiera llevarle a cuestionar la validez de cualquiera de las mismas?

6.30 Construya un intervalo de confianza del 95% para calcular la diferencia entre la media de las puntuaciones NPZ-8 de los sujetos infectados con VIH con evaluaciones positivas de la etapa ADC y sujetos infectados con VIH con evaluaciones negativas de la etapa ADC.

E. Las siguientes preguntas se refieren al estudio de caso E (página 472).

6.31 Forme una razón de probabilidad para expresar las posibilidades de una persona que da positivo teniendo la enfermedad contra las posibilidades de una persona que da negativo teniendo la enfermedad. Construya un intervalo de confianza del 95% para este cálculo de *OR*.

G. Las siguientes preguntas se refieren al estudio de caso G (página 473).

6.32 Calcule la magnitud de la diferencia entre la proporción de mellizos dicigóticos y hermanos no mellizos con EM.

H. Las siguientes preguntas se refieren al estudio de caso H (página 473).

6.33 ¿Está usted de acuerdo con la conclusión a la que llegaron los autores de este estudio? ¿Qué razones tiene para su respuesta?

6.34 ¿Cree usted que los autores de este estudio tuvieron el privilegio de estudiar bioestadística con este texto?

6.35 Cuando se discutió el tratamiento de la cistinosis (una enfermedad que puede afectar el funcionamiento de los riñones), Markello *et al.* [34] enuncian lo siguiente (se omiten las citas).

> Para 1987, se había probado la eficacia de la cisteamina oral en preservar el funcionamiento renal y mejorar el crecimiento. Desde entonces, *se ha demostrado la bioequivalencia entre la cisteamina y la fosfocisteamina*, un fotoéster de cisteamina más aceptable... [cursivas agregadas]

Estos autores citan a Smolin *et al.* [43] como la autoridad para su enunciado de bioequivalencia.[15] ¿Cree usted que se estableció la bioequivalencia en el estudio de Smolin *et al.*? Explique su respuesta.

6.36 ¿Qué modificaciones le haría al estudio de Smolin *et al.* para establecer la bioequivalencia de los dos tratamientos?

I. Las siguientes preguntas se refieren al estudio de caso I (página 473).

6.37 ¿Cree usted que el análisis debe efectuarse mediante una prueba de equivalencia de una cola o de dos colas?

6.38 Use la notación de las hipótesis de equivalencia para enunciar las hipótesis nula y alternativa para el estudio.

6.39 Realice la prueba con un nivel de .05. ¿Pudieron los investigadores declarar la equivalencia?

6.40 Los investigadores plantearon que realizarían la prueba de equivalencia construyendo un intervalo de confianza de una cola del 95% y anotando si el límite superior del intervalo estuvo por encima o por debajo de .004. ¿Es ésta una forma razonable de realizar la prueba? Explique. Cite una página o páginas en el texto que sustenten su opinión.

6.41 Los investigadores apuntan que otros ensayos clínicos habían usado .01 como el criterio para la equivalencia. Realice nuevamente la prueba usando el valor menos riguroso de .01. ¿Se habría declarado la equivalencia, si se hubiera usado este criterio?

J. Las siguientes preguntas se refieren al estudio de caso J (página 474).

6.42 Enuncie las hipótesis nula y alternativa de equivalencia para este estudio.

6.43 Realice la prueba de equivalencia con $\alpha = .05$. ¿A qué conclusión llegó en cuanto a las terapias oral e intravenosa?

N. Las siguientes preguntas se refieren al estudio de caso N (página 476).

6.44 Evalúe cada uno de los factores potenciales de riesgo listados en la tabla 1 calculando la razón de probabilidad que compara la posibilidad de ser un abusador para quienes tienen el factor de riesgo potencial y para quienes no lo tienen. Construya un intervalo de confianza del 95% para *OR*. ¿Cuáles son sus conclusiones?

O. Las siguientes preguntas se refieren al estudio de caso O (página 477).

6.45 Comente la importancia del punto (i).

[15] Véase la página 117.

CAPÍTULO 7

Métodos de muestras múltiples

7.1 INTRODUCCIÓN

En el capítulo 6 usted estudió métodos estadísticos diseñados para el análisis de datos recopilados en diseños de investigación en los que se emplean dos grupos independientes. Como ejemplo, se pueden asignar individuos al azar a uno de dos grupos; en un grupo se administra un tratamiento a los sujetos y en el otro se administra un placebo. Si el resultado es continuo, se pueden emplear en el análisis métodos relacionados con la prueba t de muestras independientes o el intervalo de confianza correspondiente. Si el resultado fuera dicotómico, podrían ser útiles diversos métodos basados en proporciones.

Pero, suponga que un investigador desea comparar tres condiciones de tratamiento, digamos, tres diferentes terapias farmacológicas. O quizás el investigador quiera estudiar dos tratamientos activos y un placebo. En este capítulo usted encontrará métodos que se emplean para el análisis de datos recabados de dos *o más* grupos. El primer método que usted estudiará es el análisis de varianza (ANOVA) de un factor, que se considera una extensión de la prueba t de muestras independientes. El segundo es una forma de chi-cuadrada que puede verse como una extensión de la prueba Z para diferencias entre proporciones. Después de que usted haya dominado estos dos estadísticos, abordaremos el tema de los procedimientos de comparación múltiple (MCP).

Conforme comencemos el estudio de los estadísticos mencionados anteriormente, usted notará que los métodos de análisis de varianza y chi-cuadrada se presentan como pruebas de hipótesis y que no se proveen intervalos de confianza ni pruebas de equivalencia. Esto resulta del hecho de que la lógica en estos métodos no los dota de tales formas estadísticas. Aunque los intervalos de confianza se calculan comúnmente en conexión con los MCP, por lo general son problemáticos si no se usa una computadora, razón por la cual no se cubrirán en este texto. Tampoco se representarán las pruebas de equivalencia para los MCP ya que por lo general no se usan en la práctica de la investigación actual. De hecho, usted ha completado sus estudios sobre pruebas equivalentes en lo que a este texto se refiere.

7.2 PRUEBA F DEL ANÁLISIS DE VARIANZA (ANOVA) DE UN FACTOR

7.2.1 Hipótesis

Usted recordará que la hipótesis nula que se somete a prueba por medio de la prueba *t* de muestras independientes es de la forma $H_0 : \mu_1 = \mu_2$, la cual esencialmente plantea que los tratamientos provistos en ambos grupos producen resultados equivalentes en lo referente a las medias. La hipótesis nula de un análisis de varianza de un factor amplía este concepto a grupos múltiples y se plantea como

$$\boxed{H_0 : \mu_1 = \mu_2 = \cdots = \mu_k} \tag{7.1}$$

lo que afirma que todas las medias de la población son iguales. La notación indica que la igualdad se extiende a cualquier cantidad de grupos y que al último de ellos se denomina *k*. Note que en el caso de dos grupos, la ecuación 7.1 se reduciría a $H_0 : \mu_1 = \mu_2$, que es la hipótesis que se somete a prueba por medio de la prueba *t* de muestras independiente s.

La hipótesis alternativa es cualquier condición que haga que la hipótesis nula sea falsa. Así, dados tres grupos, cualquiera de las condiciones alternativas siguientes, salvo un error tipo II, causaría el rechazo de la hipótesis nula.

1. $\mu_1 = \mu_2 \neq \mu_3$
2. $\mu_1 \neq \mu_2 = \mu_3$
3. $\mu_1 = \mu_3 \neq \mu_2$
4. $\mu_1 \neq \mu_2 = \mu_3$

Es importante comprender que, para $k = 3$, cuando 7.1 se rechaza, no hay manera de saber cuál de las cuatro condiciones listadas arriba causó el rechazo. Como verá, es este hecho el que da origen a los MCP que se presentarán más adelante en este capítulo. Para ponerlo en perspectiva, suponga que se realiza un estudio para comparar tres métodos quirúrgicos. La variable resultante es la cantidad de pérdida de sangre de los pacientes durante el procedimiento. La hipótesis nula plantea que la pérdida promedio de sangre para los tres métodos quirúrgicos es la misma. Si se rechaza la hipótesis nula, el investigador sabrá que esta hipótesis es falsa; es decir, que la media de la pérdida de sangre para las tres técnicas no es la misma. Pero, ¿por qué es falsa la hipótesis nula? ¿Es porque el método 1 produce el mismo resultado que el método 2, pero estos métodos no producen el mismo resultado que el método 3, como se plantea en la alternativa 1? ¿O es porque los tres métodos producen diferentes medias de pérdida de sangre como se afirma en la alternativa 4? El ANOVA de un factor no puede dar respuesta a esta pregunta. Sólo puede declarar como falsa la hipótesis nula o no declararla como tal.

7.2.2 *F* obtenida

De la misma manera que usted usó la *Z* obtenida, la *Z* crítica y el estadístico *t* para realizar ciertas pruebas de significancia, también el ANOVA de un factor utiliza los estadísticos *F* obtenida y *F* crítica para este fin. La *F* obtenida se da por la siguiente ecuación

$$\boxed{F = \frac{M\,S_b}{M\,S_w}} \tag{7.2}$$

Como se observa, la F obtenida es un cociente de dos cantidades —el cuadrado medio entre grupos, simbolizado por MS_b[1] y el cuadrado medio dentro de grupos, simbolizado por MS_w.[2] Será útil examinar cada una de estas cantidades por separado.

El **cuadrado medio dentro (*mean square within*, MS_w)**. El cuadrado medio dentro es también un cociente y se define como sigue

$$\boxed{MS_w = \frac{SS_w}{N - k}} \tag{7.3}$$

donde SS_w es la suma de cuadrados dentro *(sum of squares within)*, N es el número total de observaciones, y k es el número de grupos. La cantidad $N - k$ indica los **grados de libertad del denominador**. Por ejemplo, si hay tres grupos con cinco sujetos en cada uno, $N = 15$, $k = 3$ y los grados de libertad del denominador son $15 - 3 = 12$.

La **suma de cuadrados dentro** se define como

$$\boxed{SS_w = SS_1 + SS_2 + \cdots + SS_k} \tag{7.4}$$

donde SS_1 es la suma de cuadrados[3] del primer grupo, SS_2 es la suma de cuadrados del segundo grupo, y SS_k es la suma de cuadrados del último grupo. Quizá recuerde del capítulo 2 que la suma de cuadrados de un conjunto de datos se calcula por medio de

$$SS = \sum (x - \bar{x})^2$$

o de manera equivalente

$$SS = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

Así, la ecuación 7.4 puede escribirse

$$\boxed{SS_w = \left[\sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n_1}\right] + \left[\sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n_2}\right] + \cdots + \left[\sum x_k^2 - \frac{\left(\sum x_k\right)^2}{n_k}\right]} \tag{7.5}$$

Los subíndices 1, 2, . . . , k se refieren al grupo representado por los datos. Así, x_1 representa las observaciones del primer grupo, x_2 las del segundo, y así sucesivamente.

EJEMPLO 7.1

Los datos (ficticios) de la tabla 7.1 de la siguiente página representan los pesos (en libras) de sujetos que se han sometido a tres dietas diferentes. Utilice los datos para calcular MS_w.

Solución Para su análisis, será conveniente ordenar los datos como se muestra en la tabla 7.2 de la siguiente página. En esta tabla cada observación se muestra junto con su cuadrado. Las

[1] El cuadrado medio entre también se conoce como el cuadrado medio de los tratamientos.
[2] El cuadrado medio dentro también se conoce como el cuadrado medio del error.
[3] Usted estudió la suma de cuadrados para un conjunto de datos en la página 36.

TABLA 7.1: Pesos (en libras) de sujetos sometidos a tres diferentes dietas.

Dieta 1	Dieta 2	Dieta 3
198	214	174
211	200	176
240	259	213
189	194	201
178	188	158

TABLA 7.2: Pesos (en libras) de sujetos sometidos a tres diferentes dietas ordenados para su análisis.

	Dieta 1		Dieta 2		Dieta 3	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
	198	39204	214	45796	174	30276
	211	44521	200	40000	176	30976
	240	57600	259	67081	213	45369
	189	35721	194	37636	201	40401
	178	31684	188	35344	158	24964
Σ	1016	208730	1055	225857	922	171986

sumas de las observaciones y sus cuadrados se muestran en la parte inferior de la tabla. Utilizando estos resultados, las sumas de los cuadrados para los grupos individuales se calculan como sigue.

$$S S_1 = \sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n} = 208730 - \frac{(1016)^2}{5} = 2278.8$$

$$S S_2 = \sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n_2} = 225857 - \frac{(1055)^2}{5} = 3252.0$$

$$S S_3 = \sum x_3^2 - \frac{\left(\sum x_3\right)^2}{n_3} = 171986 - \frac{(922)^2}{5} = 1969.2$$

Note que las sumas de cuadrados calculadas de esta forma se obtienen *dentro* de los grupos. Esto contrasta con la suma de los cuadrados entre grupos, como veremos dentro de poco.

Por medio de la ecuación 7.4

$$S S_w = S S_1 + S S_2 + S S_3 = 2278.8 + 3252.0 + 1969.2 = 7500.0$$

Luego, por medio de la ecuación 7.3

$$M S_w = \frac{SS_w}{N - K} = \frac{7500.0}{15 - 3} = 625$$

Los grados de libertad del denominador son 15 − 3 = 12. ■

El cuadrado medio entre (*mean square between*, $\boldsymbol{MS_b}$). Al igual que el cuadrado medio dentro, el cuadrado medio entre es un cociente de una suma de cuadrados entre los grados de libertad. Más puntualmente,

$$\boxed{MS_b = \frac{SS_b}{k-1}} \tag{7.6}$$

donde SS_b es la suma de cuadrados entre y k es el número de grupos. La cantidad $k - 1$ se denomina **grados de libertad del numerador**. Por ejemplo, si hay tres grupos, los grados de libertad del numerador son 3 − 1 = 2. La **suma de cuadrados entre** se define como

$$\boxed{SS_w = n\left[\sum_{j=1}^{k} \bar{x}_j^2 - \frac{\left(\sum_{j=1}^{k} \bar{x}_j\right)^2}{k}\right]} \tag{7.7}$$

donde n es el número de observaciones en *cada* grupo, y $\bar{x}_j$ son las medias grupales. Esta ecuación implica que con cada grupo se asocia un número igual de observaciones. Dentro de poco presentaremos una ecuación de suma de cuadrados que relaja esta restricción. En la ecuación 7.7 la expresión contenida entre corchetes [] es de particular interés. Para este cálculo se obtiene la media de cada grupo con la suma de cuadrados *entre* las medias grupales que se estén calculando. La expresión entre corchetes es sólo la conocida fórmula de suma de cuadrados en la que sustituimos x por $\bar{x}$. Como hay k medias, se remplaza la n en la ecuación con k. Observe también que utilizamos el subíndice j en lugar del i, que es más conocido, porque con anterioridad hemos utilizado la letra i para designar las observaciones dentro de muestras.

EJEMPLO 7.2

Utilice la información de la tabla 7.2 para calcular MS_b.

Solución Usando las sumas de la tabla 7.2, las medias de los tres grupos son, respectivamente,

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{1016}{5} = 203.2$$

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{1055}{5} = 211.0$$

$$\bar{x}_3 = \frac{\sum x_3}{n_3} = \frac{922}{5} = 184.4$$

La suma de las tres medias, así como la suma de los tres cuadrados medios son las siguientes.

	$\bar{x}$	$\bar{x}^2$
	203.2	41290.24
	211.0	44521.00
	184.4	34003.36
$\sum$	598.6	119814.60

Por medio de la ecuación 7.7

$$
\begin{aligned}
SS_b &= n\left[\sum_{j=1}^{k}\bar{x}_j^2 - \frac{\left(\sum_{j=1}^{k}\bar{x}_j\right)^2}{k}\right] \\
&= 5\left[119814.60 - \frac{(598.6)^2}{3}\right] \\
&= (5)\,(373.95) \\
&= 1869.75
\end{aligned}
$$

Luego, por medio de la ecuación 7.6

$$
MS_b = \frac{SS_b}{k-1} = \frac{1869.75}{3-1} = 934.88
$$

Se escogió la forma de la ecuación 7.7 para enfatizar que esta suma de cuadrados se obtiene de la suma de cuadrados calculada con las medias de la muestra, de manera que esta suma de cuadrados se obtiene *entre* los grupos. La razón de poner este énfasis será más clara con posterioridad. Si bien es pedagógicamente atractiva, esta forma de la ecuación tiene la desventaja de ser aplicable sólo a grupos con el mismo número de observaciones. Una forma más útil, aunque menos intuitiva, de la ecuación de la suma de cuadrados entre está dada por

$$
\boxed{SS_b = \frac{\left(\sum_{i=1}^{n_1} x_{i1}\right)^2}{n_1} + \frac{\left(\sum_{i=1}^{n_2} x_{i2}\right)^2}{n_2} + \cdots + \frac{\left(\sum_{i=1}^{n_k} x_{ik}\right)^2}{n_k} - \frac{\left(\sum_{Todas} x_{..}\right)^2}{N}} \tag{7.8}
$$

Los términos que aparecen antes del signo menos en la ecuación 7.8 indican que las observaciones en cada grupo deben sumarse, y luego esta suma se debe elevar al cuadrado y el resultado debe dividirse entre el número de observaciones en el grupo. Este cálculo se lleva a cabo para cada grupo y luego se suman los resultados. El término que aparece después del signo menos indica que *todas* las observaciones deben sumarse y que el resultado debe elevarse al cuadrado. Este término se divide entre N, que representa el número total de observaciones —es decir, $n_1 + n_2 + ... + n_k$. Utilizando las sumas de la tabla 7.2 de la página 266, con la ecuación 7.8 tenemos que,

$$
\begin{aligned}
SS_b &= \frac{\left(\sum_{i=1}^{n_1} x_{i1}\right)^2}{n_1} + \frac{\left(\sum_{i=1}^{n_2} x_{i2}\right)^2}{n_2} + \cdots + \frac{\left(\sum_{i=1}^{n_k} x_{i3}\right)^2}{n_k} - \frac{\left(\sum_{Todas} x_{..}\right)^2}{N} \\
&= \frac{(1016)^2}{5} + \frac{(1055)^2}{5} + \frac{(922)^2}{5} - \frac{(2993)^2}{15} \\
&= 5999073 - 597203.267 \\
&= 1869.73
\end{aligned}
$$

que, redondeando, es el mismo resultado obtenido mediante la ecuación 7.7. ■

TABLA 7.3: Tabla de ANOVA de un factor.

Fuente de la variación	Suma de cuadrados	df	Cuadrados medios	F	Valor crítico de F	Valor-p
Entre	SS_b	$k-1$	$SS_b/k-1$	MS_b/MS_w	(tabla)	(computadora)
Dentro	SS_b	$N-k$	$SS_b/N-k$			
Total	SS_t	$N-1$				

7.2.3 La prueba de significancia

Por medio de la ecuación 7.2, la F obtenida es

$$F = \frac{MS_b}{MS_w} = \frac{934.88}{625.00} = 1.50$$

El valor crítico de F se obtiene, primero, observando que los grados de libertad del numerador para el análisis son $k - 1 = 3 - 1 = 2$ y los grados de libertad del denominador son $N - k = 15 - 3 = 12$. Para utilizar el Apéndice C, los grados de libertad del numerador se ubican en la parte superior de la tabla y los grados de libertad del denominador a un costado de la tabla. Para $\alpha = .05$ con 2 y 12 grados de libertad, el Apéndice C muestra que el valor crítico de F es 3.89. La hipótesis nula se rechaza cuando la F obtenida es mayor o igual que el valor crítico de F. En este caso 1.50 no es mayor ni igual que 3.89, por lo que no se rechaza la hipótesis nula.

¿Cómo se debe interpretar el hecho de que no se rechace H_0? ¿Significa que no existen diferencias entre los sistemas de pérdida de peso en lo que respecta a la pérdida de peso? ¡NO! Simplemente significa que no pudimos demostrar que *sí hay* diferencias. (¿Ha escuchado esto antes?)

7.2.4 La tabla de ANOVA

Los resultados de un análisis de varianza tradicionalmente se resumen en una **tabla de ANOVA**. La tabla 7.3 muestra cómo se construye una de estas tablas. La suma total de cuadrados (SS_t) es la suma de los cuadrados que se obtendrían si la suma de cuadrados de todos los datos se calculara sin importar a qué grupo pertenecen. También se obtiene como $SS_b + SS_w$. Esto significa que la suma total de cuadrados para la variable resultante puede dividirse en componentes que están relacionados con el efecto del tratamiento (SS_b) y con un componente de error aleatorio (SS_w). El valor-p asociado con el estadístico F requiere de una computadora para su cálculo. La tabla de ANOVA para el análisis de las dietas se muestra en la tabla 7.4 de la siguiente página.

EJEMPLO 7.3

En el ejemplo 6.1 de la página 220 usted utilizó los datos de la tabla 6.2 de la página 221 para realizar una prueba t de dos colas de muestras independientes con $\alpha = .05$. Utilice estos mismos datos para llevar a cabo una prueba F de un ANOVA de un factor.

Solución De la tabla 6.2, $\sum x_1 = 1977$, $\sum x_1^2 = 262767$, $\sum x_2 = 2028$, $\sum x_2^2 = 275706$

TABLA 7.4: Tabla de ANOVA de un factor para el análisis de las dietas.

Fuente de variación	Suma de cuadrados	df	Cuadrados medios	F	Valor crítico de F	Valor-p
Entre	1869.75	2	934.88	1.50	3.89	(computadora)
Dentro	7500.00	12	625.00			
Total	9369.75	14				

y $n_1 = n_2 = 15$. Utilizando estos valores

$$S\,S_1 = \sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n_1} = 262767 - \frac{(1977)^2}{15} = 2198.4$$

y

$$S\,S_2 = \sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n_2} = 275706 - \frac{(2028)^2}{15} = 1520.4.$$

Entonces, por medio de la ecuación 7.4

$$S\,S_w = S\,S_1 + S\,S_2 = 2198.4 + 1520.4 = 3718.8.$$

Luego, por medio de la ecuación 7.3

$$M\,S_w = \frac{S\,S_w}{N-k} = \frac{3718.8}{30-2} = 132.814.$$

Usando la ecuación 7.8

$$S\,S_b = \frac{\left(\sum_{i=1}^{n_1} x_{i1}\right)^2}{n_1} + \frac{\left(\sum_{i=1}^{n_2} x_{i2}\right)^2}{n_2} - \frac{\left(\sum_{Todas} x_{..}\right)^2}{N} = \frac{(1977)^2}{15} + \frac{(2028)^2}{15} - \frac{(4005)^2}{30} = 86.7.$$

Al dividir esta cantidad entre los grados de libertad del numerador $k - 1 = 2 - 1 = 1$, obtenemos $MS_b = 86.7$. Entonces, la F obtenida es

$$F = \frac{M\,S_b}{M\,S_w} = \frac{86.7}{132.814} = .653.$$

La tabla C indica que el valor crítico de F con 1 y 28 grados de libertad es 4.20. Como la F obtenida es menor que este valor, no se rechaza la hipótesis nula.

Como lo demuestra este ejemplo, tanto la prueba t de muestras independientes como el ANOVA de un factor pueden utilizarse para el análisis cuando se hacen pruebas de diferencia entre dos medias. Note que cuando sólo dos grupos están implicados, la hipótesis nula para la prueba ANOVA se convierte en $H_0 : \mu_1 = \mu_2$, que es la misma hipótesis puesta a prueba por medio de la prueba t. De hecho, cuando sólo hay dos grupos implicados, la F obtenida es igual al cuadrado de la t obtenida, y la misma relación es cierta para los valores críticos (de dos colas).[4] Debido a esta relación, las dos pruebas siempre producirán el mismo resultado, sin importar si se rechaza o no la hipótesis nula cuando dos grupos estén implicados. La prueba ANOVA tiene la ventaja de ser aplicable a dos grupos así como a más de dos grupos. ■

[4] En este contexto, las dos colas se refieren a la prueba t. Salvo raras excepciones, las pruebas F siempre son de una cola.

TABLA 7.5: Calificaciones de la calidad de la atención de urgencias en cuatro hospitales metropolitanos.

Hospital 1	Hospital 2	Hospital 3	Hospital 4
10	18	14	9
12	15	16	8
16	14	16	10
9	18	14	11
	12	17	6
	12	18	
	13	20	
	14		

EJEMPLO 7.4

Como parte de un estudio de control de calidad, se pide a enfermeras del departamento de urgencias de cuatro hospitales en una zona metropolitana determinada, que evalúen de forma anónima la calidad de la atención que se da en sus instalaciones en una serie de dimensiones. Cada dimensión se evalúa con una escala que va de cero (pésimo) a 20 (excelente). En la tabla 7.5 se muestran las calificaciones de una de estas dimensiones. Utilice estos datos para realizar una prueba F de ANOVA de un factor con un nivel $\alpha = .05$. Haga un resumen de sus resultados en una tabla de ANOVA.

Solución Es conveniente ordenar los datos del análisis como se muestra en la tabla 7.6 de la siguiente página. En esta tabla aparece cada observación junto con su cuadrado. Las sumas de las observaciones y sus cuadrados se presentan al final de la tabla. Con base en estos resultados, las sumas de cuadrados de los grupos individuales se calculan de la siguiente forma.

$$SS_1 = \sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n_1} = 581 - \frac{(47)^2}{4} = 28.75$$

$$SS_2 = \sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n_2} = 1722 - \frac{(116)^2}{8} = 40.00$$

$$SS_3 = \sum x_3^2 - \frac{\left(\sum x_3\right)^2}{n_3} = 1917 - \frac{(115)^2}{7} = 27.71$$

$$SS_4 = \sum x_4^2 - \frac{\left(\sum x_4\right)^2}{n_4} = 402 - \frac{(44)^2}{5} = 14.8$$

Por medio de la ecuación 7.4,

$$SS_w = SS_1 + SS_2 + SS_3 + SS_4 = 28.75 + 40.00 + 27.71 + 14.80 = 111.26.$$

Luego, por medio de la ecuación 7.3,

$$MS_w = \frac{SS_w}{N-k} = \frac{111.26}{24-4} = 5.56.$$

TABLA 7.6: Calificaciones de la calidad de la atención de urgencia en cuatro hospitales metropolitanos, ordenados para su análisis.

	Hospital 1		Hospital 2		Hospital 3		Hospital 4	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2	X_4	X_4^2
	10	100	18	324	14	196	9	81
	12	144	15	225	16	256	8	64
	16	256	14	196	16	256	10	100
	9	81	18	324	14	196	11	121
			12	144	17	289	6	36
			12	144	18	324		
			13	169	20	400		
			14	196				
Σ	47	581	116	1722	115	1917	44	402

Por medio de la ecuación 7.8

$$
\begin{aligned}
SS_b &= \frac{\left(\sum_{i=1}^{n_1} x_{i1}\right)^2}{n_1} + \frac{\left(\sum_{i=1}^{n_2} x_{i2}\right)^2}{n_2} + \frac{\left(\sum_{i=1}^{n_3} x_{i3}\right)^2}{n_3} + \frac{\left(\sum_{i=1}^{n_4} x_{i4}\right)^2}{n_4} - \frac{\left(\sum_{Todas} x_{..}\right)^2}{N} \\
&= \frac{(47)^2}{4} + \frac{(116)^2}{8} + \frac{(115)^2}{7} + \frac{(44)^2}{5} - \frac{(322)^2}{24} \\
&= 190.57.
\end{aligned}
$$

Si dividimos SS_b entre los grados de libertad del numerador tenemos que

$$MS_b = \frac{SS_b}{k-1} = \frac{190.57}{4-1} = 63.52.$$

Entonces, la F obtenida es

$$F = \frac{MS_b}{MS_w} = \frac{63.52}{5.56} = 11.42.$$

Al revisar la tabla C con 3 y 20 grados de libertad, encontramos un valor crítico de F de 3.10. Puesto que la F obtenida es mayor que el valor crítico de F, se rechaza la hipótesis nula. Podemos concluir, por ende, que los datos de los hospitales no provienen de la media de una población común. Estos resultados se resumen en la tabla 7.7 de la siguiente página. ■

7.2.5 Dos características importantes de MS_b y MS_w

Como usted sabe, se rechaza la hipótesis nula cuando la F obtenida es igual o excede al valor crítico de F. Esto implica que el valor de la F obtenida aumenta de cierta forma cuando la hipótesis nula es falsa. Entender cómo sucede esto ampliará su comprensión del ANOVA y es el tema de este apartado.

TABLA 7.4: Tabla de ANOVA de un factor de las calificaciones de la calidad de la atención de los departamentos de urgencias de cuatro hospitales.

Fuente de la variación	Suma de cuadrados	df	Cuadrados medios	F	Valor crítico de F	Valor-p
Entre	190.57	3	63.52	11.42	3.10	(computadora)
Dentro	111.26	20	5.56			
Total	301.83	23				

Cuando la hipótesis nula es verdadera, MS_w y MS_b calculan la varianza de la población (o poblaciones) de la que se tomaron las muestras. Debido a que MS_w y MS_b simplemente son dos métodos diferentes de calcular σ^2 cuando la hipótesis nula es verdadera, esperaríamos que la F obtenida tome un valor cercano a 1 en esta circunstancia. Pero, ¿qué sucede con la F obtenida cuando la hipótesis nula es falsa? Una demostración aclarará este asunto.

Reconsideremos los datos de la tabla 7.2 de la página 266. Si suponemos que los pesos de estas personas se obtuvieron bajo una hipótesis nula verdadera, podemos crear un facsímil de una hipótesis nula falsa mediante una simple modificación. Por ejemplo, suponga que la dieta 3 fue más efectiva que las otras dos, por lo que las personas que siguieron esta dieta perdieron 20 libras más de lo que habrían perdido en cualquier otro caso. A continuación se muestran los pesos y los pesos cuadrados de las personas del grupo 3. Estos pesos se crearon sustrayendo 20 libras del peso de cada miembro del grupo 3, tal como se había reportado en la tabla 7.2.

	X_3	X_3^2
	154	23716
	156	24336
	193	37249
	181	32761
	138	19044
Σ	822	137106

¿Qué repercusión tendrá este notable efecto adelgazador sobre la F obtenida? Primero veamos lo que sucede con MS_w. Como se calculó previamente,

$$SS_1 = 2278.8$$
$$SS_2 = 3252.0$$
$$SS_3 = 1969.2$$

Como no modificamos los pesos de los sujetos de los grupos 1 y 2, podemos usar las sumas de los cuadrados previamente calculados para estas personas. Usando las sumas de la tabla anterior, recalculamos SS_3 como

$$SS_3 = \sum x_3^2 - \frac{\left(\sum x_3\right)^2}{n_3} = 137106 - \frac{(822)^2}{5} = 1969.2$$

Pero ésta es la misma suma de cuadrados obtenida sin la drástica pérdida de peso. Deducimos que MS_w no cambia por el efecto experimentado por el grupo 3. Ahora examinamos la repercusión (si existe) sobre SS_b.

Como se calculó antes (página 267)

$$\bar{x}_1 = 203.2$$
$$\bar{x}_2 = 211.0$$
$$\bar{x}_3 = 184.4$$

Como no alteramos las observaciones para las dietas 1 y 2, las medias de estos grupos permanecerán como se calcularon previamente. La nueva media para el grupo tres será

$$\bar{x}_3 = \frac{\sum x_3}{n_3} = \frac{822}{5} = 164.4.$$

La suma de las dos medias originales y la nueva media para la dieta 3, así como la suma de sus valores cuadrados son de la siguiente manera.

	$\bar{x}$	$\bar{x}$
	203.2	41290.24
	211.0	44521.00
	164.4	27027.36
$\sum$	578.6	112838.60

Por la ecuación 7.7

$$
\begin{aligned}
SS_b &= n\left[\sum_{j=1}^{k} \bar{x}_j^2 - \frac{\left(\sum_{j=1}^{k} \bar{x}_j\right)^2}{k}\right] \\
&= 5\left[112838.60 - \frac{(578.6)^2}{3}\right] \\
&= (5)(1245.95) \\
&= 6229.75
\end{aligned}
$$

que es considerablemente más grande que el valor de 1869.73 obtenido bajo la hipótesis nula verdadera (página 268).

Por medio de la ecuación 7.6,

$$MS_b = \frac{SS_b}{k-1} = \frac{6229.75}{3-1} = 3114.875.$$

La F obtenida es ahora

$$F = \frac{MS_b}{MS_w} = \frac{3114.875}{625.000} = 4.98.$$

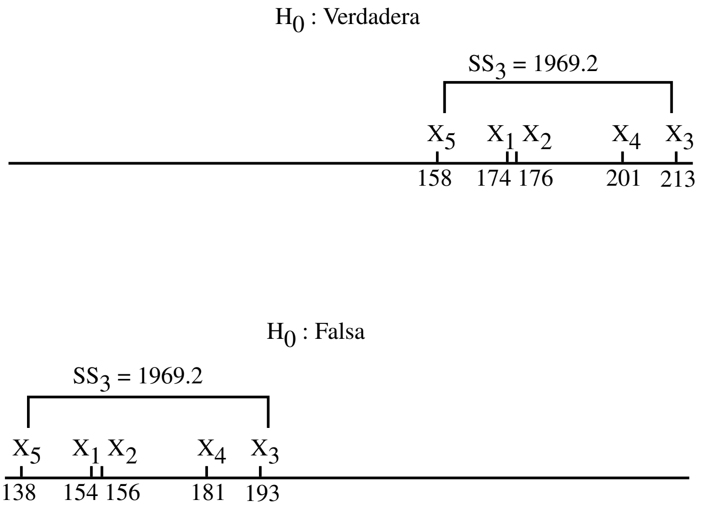


FIGURA 7.1: Representación de SS_3 bajo las hipótesis nulas verdadera y falsa.

que es mayor que el valor crítico de F de 3.89, de manera que se rechaza la hipótesis nula del mismo efecto de las dietas.

Como se mencionó antes, cuando la hipótesis nula es verdadera, MS_w y MS_b calculan la varianza de la población de tal manera que la F obtenida normalmente es cercana a 1. Pero, como se observa en la demostración anterior, cuando la hipótesis nula es falsa, MS_w no cambia por esta condición y continúa siendo un estimado de la varianza de la población. En contraste, MS_b se infla bajo una hipótesis nula falsa, incrementando la magnitud de la F obtenida.

En las figuras 7.1 y 7.2 de la siguiente página se muestra la relación entre las dos sumas de cuadrados y la condición (es decir, verdadera o falsa) de las hipótesis nulas. En la figura 7.1 se observa que la pérdida de peso que experimentó el tercer grupo simplemente recorrió sus pesos en la línea numérica pero no alteró las posiciones relativas de los pesos en el grupo. Como consecuencia, las sumas de los cuadrados calculadas a partir estos valores no cambiaron.

En contraste, como se aprecia en la figura 7.2, la pérdida de peso en el grupo 3 provocó que la media de ese grupo se desviara de las medias de los grupos 1 y 2. El efecto es la dispersión de las medias y, por consiguiente, el incremento de la suma de cuadrados obtenida de estos valores. ¿Qué cree usted que sucedería a la razón F si, además de la pérdida de 20 libras que experimentó el grupo 3, cada miembro del grupo 2 hubiera aumentado 20 libras? ¿Es usted lo suficientemente curioso para hacer los cálculos y verificar su respuesta?[5]

En este punto se debe hacer una advertencia. Modelamos el efecto del tratamiento sustrayendo una constante (es decir, 20 libras) del peso de cada miembro de un grupo seleccionado. Cuando todos los miembros de un grupo responden al tratamiento de la misma manera *aditiva*,[6] que es diferente de la respuesta en otros grupos, como en este caso, el resultado se denomina **alternativa de cambio**. Contraste esta forma de respuesta con una en la que algunos miembros de un grupo

[5] Cuidado, porque de ser así, usted está en peligro de convertirse en un bioestadístico.
[6] Esto es, al sumar o restar la misma constante.

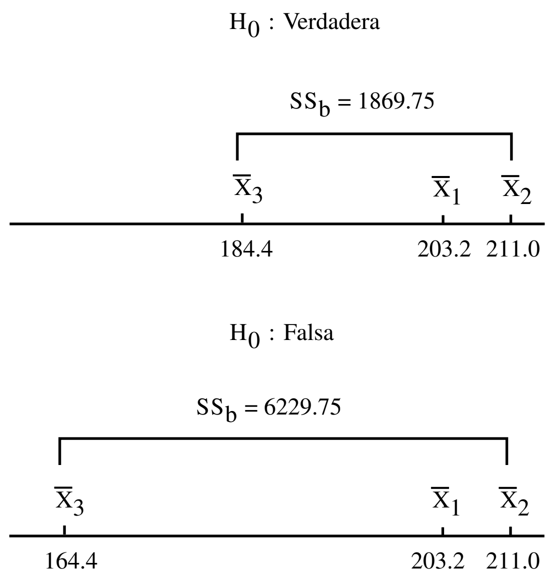


FIGURA 7.2: Representación de SS_b bajo las hipótesis nulas verdadera y falsa.

no responden en absoluto, en tanto que otros se benefician enormemente con el tratamiento. Esta forma puede causar un incremento en la suma de cuadrados del grupo, así como un cambio en la media. Esto podría tener un efecto nocivo en la potencia de la prueba F. Aunque la prueba F está diseñada principalmente para detectar alternativas de cambio, ha funcionado bien en una gran variedad de disciplinas donde a veces no se encuentran alternativas de cambio.

7.2.6 Suposiciones

Las suposiciones de la prueba F del ANOVA son las mismas que para la prueba t de muestras independientes que analizamos en el apartado 6.2.4 de la página 230.

7.3 LA PRUEBA CHI-CUADRADA DE 2 POR *K*

7.3.1 Hipótesis

En el capítulo 6 de la página 230 usted aprendió a poner a prueba la hipótesis nula $H_0 : \pi_1 = \pi_2$ mediante una prueba Z de muestras independientes. La prueba chi-cuadrada de 2 por k extiende este concepto para poner a prueba la igualdad de cualquier número de proporciones. Esta hipótesis nula se plantea como

$$\boxed{H_0 : \pi_1 = \pi_2 = \cdots = \pi_k} \tag{7.9}$$

que afirma que todas las proporciones de la población son iguales. La notación indica que la igualdad se extiende a cualquier número de grupos y que al último de éstos se le denomina grupo k. Observe que, en el caso de dos grupos, la ecuación 7.9 se reduciría a $H_0 : \pi_1 = \pi_2$, que es la hipótesis puesta a prueba por medio de la prueba Z de muestras independientes.

La hipótesis alternativa es cualquier condición que haga que la hipótesis nula sea falsa. Así, dados tres grupos, cualquiera de las siguientes condiciones, salvo un error tipo II, causaría el rechazo de la hipótesis nula.

1. $\pi_1 = \pi_2 \neq \pi_3$
2. $\pi_1 \neq \pi_2 = \pi_3$
3. $\pi_1 = \pi_3 \neq \pi_2$
4. $\pi_1 \neq \pi_2 \neq \pi_3$

Es importante comprender que, cuando se rechaza la ecuación 7.9, no hay forma de saber cuál de las cuatro condiciones causó el rechazo. Como usted verá, es este hecho lo que origina los MCP que se presentarán más adelante en este capítulo. Para poner esto en perspectiva, suponga que se lleva a cabo un estudio para comparar tres tratamientos para una enfermedad terminal. La variable resultante es un indicador de la supervivencia del paciente después de cinco años de tratamiento. La hipótesis nula plantea que las proporciones de supervivencia después de cinco años son iguales en los tres tratamientos. Si la hipótesis nula se rechaza, el investigador sabrá que la hipótesis nula es falsa; es decir, que los tres tratamientos no producen las mismas proporciones de sobrevivientes después de cinco años. Pero, ¿por qué es falsa la hipótesis nula? ¿Es porque el tratamiento 1 produce el mismo resultado que el tratamiento 2, pero estos tratamientos no producen el mismo resultado que el tratamiento 3, como se plantea en la alternativa 1? ¿O es porque los tres tratamientos producen proporciones diferentes de sobrevivientes, como se plantea en la alternativa 4? La prueba chi-cuadrada no puede responder esta pregunta; sólo puede declarar que la hipótesis nula es o no falsa.

7.3.2 χ^2 obtenida

Como ocurre con otros estadísticos con los que usted ahora está familiarizado, la prueba de hipótesis se realiza calculando un valor obtenido y su comparación subsiguiente con un valor crítico. Para la prueba chi-cuadrada, el valor obtenido se calcula por medio de

$$\boxed{\chi^2 = \sum_{\textit{todas las celdas}} \left[\frac{(f_o - f_e)^2}{f_e} \right]} \tag{7.10}$$

donde f_o y f_e son las frecuencias observada y esperada, respectivamente. La **frecuencia observada** es simplemente el número de resultados que ocurren en la celda en cuestión, como se muestra en la tabla 7.8 de la siguiente página. En esta tabla hemos usado subíndices dobles para indicar el renglón y la columna de los datos en cada celda. Por ejemplo, si el resultado uno indica “muerto después de cinco años” y el resultado dos representa “vivo después de cinco años”, entonces f_{o11} sería el número de personas del grupo uno que murieron en la marca de cinco años y f_{o21} sería el número de personas del grupo 1 que están vivas hasta este momento. Las entradas f_{o12} y f_{o22} representarían las mismas cantidades para el grupo 2. Esta tabla se puede extender para representar datos de cualquier cantidad de grupos.

La **frecuencia esperada** representa el número de personas que se espera se encuentren en cada celda *si la hipótesis nula es verdadera*. Este concepto requerirá cierta explicación.[7] Como

[7] No se preocupe demasiado si no entiende todo lo que sigue. Puede llevar una vida *bastante* completa y productiva sin el dominio total de este concepto.

TABLA 7.8: Descripción de una tabla de chi-cuadrada de 2 por 3.

	Grupo 1	Grupo 2	Grupo 3
Resultado 1	f_{o11} f_{e11}	f_{o12} f_{e12}	f_{o13} f_{e13}
Resultado 2	f_{o21} f_{e21}	f_{o22} f_{e22}	f_{o23} f_{e23}

se expresa en la ecuación 3.5 de la página 56, dos sucesos son independientes si el producto de sus probabilidades es igual a su probabilidad conjunta. Expresado en términos del problema en cuestión podemos decir que si el tratamiento recibido es independiente al hecho de si el paciente vive o no, entonces

$$P\,(\mathrm{G}_1\ \mathrm{D}) = P\,(\mathrm{G}_1)\,P\,(\mathrm{D})$$

donde G_1 representa la pertenencia al grupo 1 y D indica estar muerto. Así, la expresión de arriba dice que si la pertenencia al grupo y el resultado son independientes, la probabilidad de estar muerto y ser un miembro del grupo 1 es simplemente igual al producto de las dos probabilidades individuales. La misma forma de planteamiento se puede aplicar a cada celda de la tabla. El concepto clave aquí es reconocer que una afirmación de independencia entre la pertenencia al grupo y el resultado es equivalente a una afirmación de que la hipótesis nula (véase la ecuación 7.9 de la página 276) es verdadera. Después de todo, si el tratamiento recibido en particular no tiene nada que ver con el hecho de que el paciente muera o no, entonces la proporción de individuos vivos (o muertos) en cada grupo es la misma.

Para continuar, si la hipótesis nula es verdadera, la probabilidad de estar en el grupo 1 *y* estar muerto es igual a la probabilidad de estar en el grupo 1 multiplicada por la probabilidad de estar muerto. La probabilidad de estar en el grupo 1 es justo la proporción de personas del grupo 1 o

$$\frac{N_{G_1}}{N}$$

donde N_{G1} es el número total de pacientes del grupo 1 y N es el total de pacientes de la tabla. De igual forma, la probabilidad de estar muerto es simplemente la proporción de personas que están muertas o

$$\frac{N_M}{N}$$

donde N_D es el número total de personas que están muertas. Así, si la hipótesis nula es verdadera, la probabilidad de estar muerto y estar en el grupo uno simplemente es

$$\left(\frac{N_{G_1}}{N}\right)\left(\frac{N_M}{N}\right)$$

Como esta expresión es una probabilidad, podemos considerarla como una proporción: la proporción del número total de pacientes que se encontrarán en esta celda *si la hipótesis nula es*

TABLA 7.9: Datos del estudio del tratamiento, ordenados para el análisis de chi-cuadrada.

	Grupo 1	Grupo 2	Grupo 3	
Muertos	[17] (20.81)	[29] (25.03)	[11] (11.16)	57
Vivos	[52] (48.19)	[54] (57.97)	[26] (25.84)	132
	69	83	37	$N = 189$

verdadera. El *número* de personas que esperaríamos encontrar en esta celda sería entonces la proporción que buscamos multiplicada por el número total de pacientes o

$$\left(\frac{N_{G_1}}{N}\right)\left(\frac{N_M}{N}\right)N = \frac{\left(N_{G_1}\right)\left(N_M\right)}{N}$$

Para que esto sea aplicable a cualquier celda de la tabla escribiremos la frecuencia esperada (f_e) para cualquier celda como

$$\boxed{f_e = \frac{\left(N_R\right)\left(N_C\right)}{N}} \tag{7.11}$$

donde N_R es el total de renglón de la celda cuya frecuencia esperada se está calculando y N_C es el total de columna de la misma celda.

Una vez que se obtienen f_o y f_e para cada celda, la cantidad

$$\frac{\left(f_o - f_e\right)^2}{N}$$

se calcula para cada celda y luego se suman los resultados para generar el valor obtenido de chi-cuadrada. Un ejemplo le será de ayuda.

EJEMPLO 7.5

Suponga que se obtuvieron los siguientes resultados en el ejemplo anterior del tratamiento de una enfermedad terminal. De los pacientes que recibieron el tratamiento 1, 17 están muertos después de cinco años, en tanto que 52 todavía viven. Para el tratamiento 2, 29 están muertos mientras que 54 siguen vivos, y para el grupo 3, 11 están muertos y 26 aún viven. Utilice estos datos para calcular el valor obtenido de chi-cuadrada.

Solución Las frecuencias, observada (corchetes) y esperada (paréntesis), así como los totales de los renglones y las columnas se ordenan como se muestra en la tabla 7.9. Las frecuencias esperadas se calcularon mediante la ecuación 7.11, como se muestra a continuación.

$$f_{e11} = \frac{(N_M)(N_{G1})}{N} = \frac{(57)(69)}{189} = 20.81$$

$$f_{e12} = \frac{(N_M)(N_{G2})}{N} = \frac{(57)(83)}{189} = 25.03$$

$$f_{e13} = \frac{(N_M)(N_{G3})}{N} = \frac{(57)(37)}{189} = 11.16$$

$$f_{e21} = \frac{(N_A)(N_{G1})}{N} = \frac{(132)(69)}{189} = 48.19$$
$$f_{e22} = \frac{(N_A)(N_{G2})}{N} = \frac{(132)(83)}{189} = 57.97$$
$$f_{e23} = \frac{(N_A)(N_{G3})}{N} = \frac{(132)(37)}{189} = 25.84$$

Luego, por medio de la ecuación 7.10, la chi-cuadrada obtenida es

$$\begin{aligned}\chi^3 &= \sum_{\textit{todas las celdas}} \left[\frac{(f_o - f_e)^2}{f_e}\right] \\ &= \frac{(17-20.81)^2}{20.81} + \frac{(29-25.03)^2}{25.03} + \frac{(11-11.16)^2}{11.16} + \frac{(52-48.19)^2}{48.19} + \frac{(54-57.97)^2}{57.97} + \frac{(26-25.84)^2}{25.84} \\ &= .70 + .63 + .00 + .30 + .27 + .00 \\ &= 1.9\end{aligned}$$

El valor crítico se obtiene al revisar el Apéndice D con $k - 1$ grados de libertad, donde k es el número de grupos. En el presente caso, el Apéndice D muestra que a un nivel $\alpha = .05$ y con $3 - 1 = 2$ grados de libertad, el valor crítico de χ^2 es 5.991. La hipótesis nula se rechaza cuando la chi-cuadrada obtenida es mayor que o igual al valor crítico de chi-cuadrada. Como 1.9 es menor que 5.991, no se rechaza la hipótesis nula. Concluimos, por consiguiente, que no se puede demostrar una diferencia entre las proporciones de la población. En términos de investigación, concluimos que no pudimos demostrar una diferencia en la efectividad de los tres tratamientos. ■

EJEMPLO 7.6

En el ejemplo 6.7 de la página 233 se empleó el estadístico Z para probar $H_0 : \pi_1 = \pi_2$ utilizando datos de un estudio (ficticio) referente a la instrucción nutricional y bebés con bajo peso al nacer. En ese estudio, 314 madres recibieron capacitación en nutrición y 23 de ellas tuvieron bebés con bajo peso al nacer. En contraste, 39 de las 316 madres en el grupo sin capacitación en nutrición tuvieron bebés con bajo peso al nacer. Utilice la chi-cuadrada para poner a prueba la hipótesis anterior. Esto es, ponga a prueba la hipótesis de que la proporción de bebés con bajo peso al nacer, de madres que recibieron capacitación en nutrición, es la misma que la proporción de bebés de madres que no recibieron instrucción. ¿Ve usted una relación entre la chi-cuadrada obtenida y la Z obtenida de -2.12 que calculó previamente?

Solución Las frecuencias observada (en corchetes) y esperada (en paréntesis) se muestran en la tabla 7.10 de la siguiente página. Las frecuencias observadas se obtuvieron considerando que 23 de las 314 madres instruidas tuvieron bebés con bajo peso al nacer, por lo que $314 - 23 = 291$ tuvieron bebés que no presentaron bajo peso al nacer. De igual manera, como 39 de las 316 madres sin instrucción tuvieron bebés con bajo peso, $316 - 39 = 277$ de estas madres no tuvieron bebés con bajo peso al nacer. Las frecuencias esperadas se calcularon mediante la ecuación 7.11, como se muestra a continuación.

$$f_{e11} = \frac{(N_L)(N_I)}{N} = \frac{(62)(314)}{630} = 30.90$$
$$f_{e12} = \frac{(N_L)(N_{NI})}{N} = \frac{(62)(316)}{630} = 31.10$$

TABLA 7.10: Datos del estudio de bajo peso al nacer, ordenados para el análisis de chi-cuadrada.

	Con instrucción	Sin instrucción	
Con bajo peso al nacer	[23] (30.90)	[39] (31.10)	62
Sin bajo peso al nacer	[291] (283.10)	[277] (284.90)	568
	314	316	N = 630

$$f_{e21} = \frac{(N_{NL})(N_I)}{N} = \frac{(568)(314)}{630} = 283.10$$
$$f_{e22} = \frac{(N_{NL})(N_{NI})}{N} = \frac{(568)(316)}{630} = 284.90$$

La notación N_L, N_{NL}, N_I y N_{NI} representa, respectivamente, el número de bebés con bajo peso al nacer, el número de bebés que no presentaron bajo peso al nacer, el número de madres que recibieron instrucción nutricional y el número de madres que no recibieron esa instrucción.

Luego, por medio de la ecuación 7.10, la chi-cuadrada obtenida es

$$\chi^3 = \sum_{todas\ las\ celdas} \left[\frac{(f_o - f_e)^2}{f_e} \right]$$

$$= \frac{(23-30.90)^2}{30.90} + \frac{(39-31.10)^2}{31.10} + \frac{(291-283.10)^2}{283.10} + \frac{(277-284.90)^2}{284.90}$$
$$= 2.02 + 2.01 + .22 + .22$$
$$= 4.47$$

El valor crítico se obtiene al consultar el Apéndice D con $k - 1 = 2 - 1 = 1$ grados de libertad. Con un nivel $\alpha = .05$, el valor crítico de χ^2 es 3.841. La hipótesis nula se rechaza cuando la chi-cuadrada obtenida es mayor que o igual al valor crítico de chi-cuadrada. Como 4.47 es mayor que 3.841, se rechaza la hipótesis nula. Concluimos, por lo tanto, que sí existe una diferencia entre las proporciones de la población. En términos de la investigación, concluimos que hay una diferencia en las proporciones de bebés con bajo peso al nacer, producida por las madres que recibieron información nutricional y las que no recibieron esta información.

Observe que la chi-cuadrada obtenida de 4.47 es el cuadrado (redondeando) del valor *Z* obtenido de −2.12 que se calculó en la página 233. Como la prueba *Z* de muestras independientes para una diferencia entre proporciones pone a prueba la misma hipótesis nula que la prueba chi-cuadrada con 1 grado de libertad, estas pruebas producen el mismo resultado en términos de la decisión de rechazo o no rechazo. Así como la prueba *F* de un ANOVA de un factor puede considerarse como una generalización de la prueba *t* de muestras independientes, también chi-cuadrada de 2 por *k* puede considerarse una generalización de la prueba *Z* de muestras independientes para una diferencia entre proporciones.

Como comentario final, recuerde que mientras f_e es el número de observaciones que se espera en una celda en particular si la hipótesis nula es verdadera, f_o es el número que en realidad se encuentra allí. Es lógico que cuanto más grande sea la diferencia entre estos dos valores, mayor será la evidencia en contra de la hipótesis nula. El numerador del estadístico chi-cuadrada (ecuación 7.10) provoca que la chi-cuadrada obtenida aumente de tamaño como resultado de esta diferencia. Cuando esta diferencia es grande, la chi-cuadrada obtenida es grande. Si f_e y f_o tuvieran siempre el mismo valor, la chi-cuadrada obtenida sería cero. Así, el rechazo ocurre cuando la chi-cuadrada obtenida es igual o mayor que el valor crítico. ■

7.3.3 Suposiciones

La prueba chi-cuadrada de 2 por k que aquí se presenta es un aproximado más que una prueba exacta.[8] Para que la aproximación sea suficientemente precisa, una regla general plantea que por lo menos el 80% de la f_e en la tabla debe ser mayor que o igual a 5 y ninguna celda debe tener una f_e menor que 1.

También se supone que las observaciones son independientes. Por ejemplo, un resultado no puede influir en otro resultado ni ser influido por éste. En el último ejemplo, esto significaría que el hecho de que una madre tenga un bebé de bajo peso al nacer no debería estar relacionado con que otra madre tenga un bebé de bajo peso. Podría darse un error si unos mellizos nacidos de la misma madre fueran incluidos en el análisis. La prueba de chi-cuadrada no se considera sólida si se viola el supuesto de independencia.

7.4 PROCEDIMIENTOS DE COMPARACIÓN MÚLTIPLE

7.4.1 Introducción

Usted conoció los errores tipo I en la página 125 y aprendió que esos errores ocurren cuando una hipótesis nula verdadera se rechaza, y que la probabilidad de un evento así se denomina α. En este apartado aprenderá una segunda forma del error tipo I. Para diferenciar entre uno y otro, nos referiremos a la forma que ya conocemos como **error por comparación** o **PCE** (*per comparison error*) y designaremos la probabilidad de un PCE como α_{PCE}. Nos referiremos a la nueva forma del error tipo I como **error familiar** o **FWE** (*familywise error*) y usaremos el símbolo α_{FWE} para representar la probabilidad de este suceso.

Si usted tuviera que realizar una sola prueba al nivel α_{PCE}, la probabilidad de rechazar una hipótesis nula verdadera sería α_{PCE}. Pero, suponga que realizara 100 de estas pruebas al nivel α_{PCE}. Si todas estas 100 hipótesis nulas fueran verdaderas, ¿cuál sería la probabilidad de que rechazara una o más de estas hipótesis nulas verdaderas? Es lógico que si usted realizara cada prueba, por ejemplo, a un nivel $\alpha_{PCE} = .05$, habría muchas posibilidades de que rechazara *por lo menos* una hipótesis nula verdadera. Para hacer el ejemplo más extremo, ¿qué probabilidad cree que habría de rechazar una o más hipótesis nulas verdaderas, si hiciera 1 millón de estas pruebas? Por supuesto, esta probabilidad estaría muy cerca de 1.

Cuando por lo menos una[9] hipótesis nula verdadera se rechaza en una serie (o "familia") de pruebas, se dice que ha ocurrido un **error familiar**. Un error familiar se convierte en una preocupación en dos contextos al menos. Llamaremos al primero de éstos "análisis de comparación múltiple" y al segundo, "análisis de extremos múltiples".[10]

---

[8] Una prueba exacta es posible pero requiere de un programa especial de computación y, por lo tanto, no lo tratamos aquí.

[9] "Por lo menos una" significa lo mismo que "una o más".

[10] Muchos autores no hacen ninguna distinción entre las dos formas y simplemente las agrupan como "comparaciones múltiples".

El **análisis de comparación múltiple** se refiere a la situación en la que se comparan múltiples grupos sobre una sola variable resultante. Por ejemplo, la tabla 7.1 de la página 266 muestra los pesos de sujetos que participaron en tres diferentes estrategias de dieta. Como usted sabe ahora, una prueba F de un ANOVA de un factor solamente se ocupa de saber si todas las medias de la población son iguales. Pero un investigador muy probablemente tendría preguntas más específicas. Por ejemplo, ¿hay alguna diferencia en la efectividad de las dietas 1 y 2? ¿Hay alguna diferencia en la efectividad de las dietas 1 y 3? En general, hay $\frac{k(k-1)}{2}$ pares de comparaciones que pueden hacerse.[11] En el caso de tres grupos, esto significaría que se pueden hacer $\frac{(3)(3-1)}{2} = 3$ pares de comparaciones. Las hipótesis nulas para estas comparaciones son

1. $\mu_1 = \mu_2$
2. $\mu_1 = \mu_3$
3. $\mu_2 = \mu_3$

Es evidente que podríamos ocuparnos de las preguntas implícitas en estas hipótesis realizando tres pruebas t de muestras independientes. Si hubiera cuatro grupos quizá querríamos probar todos los $\frac{(4)(3)}{2} = 6$ pares de hipótesis. Pero, ¿cuáles son las implicaciones para un error familiar cuando hacemos una serie de pruebas de este tipo? A medida que hacemos más pruebas, aumenta la probabilidad de obtener *algunos* resultados significativos. Si realizamos seis de estas pruebas y encontramos que una es significativa, no sabemos si este resultado significativo se originó porque la hipótesis nula específica tratada con esta prueba es falsa o si nosotros simplemente hicimos tantas pruebas que tarde o temprano obtuvimos un resultado significativo. Ésta es otra forma de decir que el resultado se puede atribuir a un incremento en α_{FWE}.

El **análisis de extremos múltiples** se refiere a la situación en la que se comparan dos grupos en múltiples medidas resultantes. Por ejemplo, si quisiéramos comparar cinco químicas sanguíneas de dos grupos de pacientes, posiblemente haríamos cinco pruebas t, una para cada química sanguínea. Como usted está realizando múltiples pruebas, nuevamente se enfrentará al problema del error familiar.

La relación entre α_{PCE} y α_{FWE} es importante y se ilustra en la tabla 7.11 de la siguiente página. Las entradas de esta tabla se generaron usando métodos de simulación computarizados en los que, mediante pruebas t de muestras independientes, se analizaron muestras aleatorias de poblaciones distribuidas normalmente. En la primera columna se observa el nivel de significancia que se usó en cada prueba, en tanto que en la segunda se presenta el número de muestras (es decir, grupos) implicadas en los análisis. La tercera columna muestra el número de pruebas realizadas, la cuarta la tasa de errores familiares que generaron las múltiples pruebas, y la última muestra la **tasa de error por familia,** que explicaremos a continuación.

Para aclararlo, la primera columna del primer renglón muestra que la primera serie de pruebas se realizó a un nivel preestablecido de $\alpha_{PCE} = .05$. La segunda columna muestra que estuvieron implicados tres grupos, los cuales, como indica la tercera columna, producen $\frac{(3)(2)}{2} = 3$ pares de pruebas. La cuarta columna indica que las tres pruebas t tuvieron una tasa de error familiar de 0.122. Esto significa que la probabilidad de rechazar por lo menos una de las tres hipótesis

---

[11] A menudo, los investigadores se interesan solamente por un subconjunto de estas comparaciones, por lo que no todas se llevan a cabo. Existen ventajas importantes al omitir las comparaciones que no nos interesan.

TABLA 7.11: Tasas de error por comparación y familiar de los números de comparaciones establecidos.

α_{PCE}	Número de grupos	Número de comparaciones	α_{FWE}	PFE
.05	3	3	.122	0.150
	5	10	.286	0.499
	10	45	.630	2.249
	20	190	.920	9.508
.01	3	3	.027	0.030
	5	10	.075	0.100
	10	45	.231	0.451
	20	190	.528	1.898

nulas fue de 0.122. La última columna muestra la **tasa de error por familia (PFE)** definido como el promedio de rechazos que se dan en un conjunto de comparaciones. Note que el PFE no es una probabilidad. Esto es evidente si se observa que el promedio de rechazos erróneos es, algunas veces, mayor que 1.

Se deben destacar tres puntos importantes de esta tabla. Primero, α_{FWE} aumenta conforme el número de pruebas se incrementa. De hecho, cuando se realizaron 10 pruebas con $\alpha_{PCE} = .05$, la probabilidad de obtener un resultado significativo fue mayor que uno en cuatro, comparado con .122, cuando se hicieron tres pruebas. El segundo punto que hay que destacar es que α_{FWE} puede disminuir drásticamente al reducir α_{PCE}. Para el caso en el que se condujeron 10 pruebas, el índice de error familiar de .286, obtenido cuando el nivel de comparación fue .05, disminuyó a .075 cuando α_{PCE} se redujo a .01. En tercer lugar, α_{FWE} siempre es más grande que α_{PCE}.

7.4.2 Control de los errores familiares

Cuando usted rechaza una sola hipótesis nula, la interpretación es clara. Usted tiene una probabilidad de α_{PCE} de hacerlo incorrectamente. Ya que esta probabilidad es pequeña, puede confiar en que la hipótesis nula es falsa. Cuando usted realiza una *serie* de pruebas y rechaza una o más hipótesis nulas, la interpretación no es tan clara. ¿Rechazó usted estas hipótesis porque son falsas o porque la tasa de errores familiares tipo I es tan alta que los rechazos eran altamente probables incluso frente a hipótesis nulas verdaderas? Usted confió en su resultado de la prueba única porque pudo controlar la probabilidad de un rechazo falso con α_{PCE}. Podría tener la misma confianza en sus resultados para pruebas múltiples si pudiera controlar α_{FWE} a un nivel establecido, por ejemplo .05.

Se han desarrollado muchos métodos para el control de los errores familiares. Algunos de éstos son útiles en tanto que otros parecen tener imperfecciones, y por consiguiente, no son tan útiles. Algunas de estas pruebas están diseñadas para emplearse en contextos de investigación específicos, mientras que otras están diseñadas para un uso más general. No intentaremos discutir o siquiera listar estas técnicas aquí, sino que mejor explicaremos el uso de estos tres métodos. No recomendamos el uso del primero de ellos, conocido como el método de Bonferroni, por razones que explicaremos a continuación. Lo incluimos aquí por su simplicidad y porque ayuda a

TABLA 7.12: Efectos del ajuste de Bonferroni en las tasas de error por familia para números de comparaciones establecidos.

α_{PCE}	Número de grupos	Número de comparaciones	α_{FWE}	PFE
.0167	3	3	.044	0.050
.0050	5	10	.040	0.050
.0011	10	45	.037	0.050
.0003	20	190	.036	0.050

comprender el problema de las pruebas múltiples. El método reductor de Bonferroni es una versión mejorada del método antiguo y es generalmente superior a esa forma de la prueba. Si bien la técnica de Bonferroni y la técnica reductora de Bonferroni son aplicables a una variedad de situaciones de pruebas de extremos múltiples y comparaciones múltiples, la prueba HSD de Tukey está restringida a análisis de comparación múltiple, pero resulta muy útil en ese contexto.

Método de Bonferroni para controlar errores familiares. Como se muestra en la tabla 7.11, α_{FWE} puede disminuir al reducir α_{PCE}. Pero suponga que desea establecer α_{FWE} en algún valor establecido, por ejemplo, .05. ¿Qué tan abajo debe fijar α_{PCE} para hacer que α_{FWE} sea 0.05? Uno de los métodos más antiguos, más simples y más utilizados para encontrar este nivel se conoce como el **ajuste de Bonferroni**.[12] Por medio de este ajuste

$$\boxed{\alpha_{PCE} = \frac{\alpha_{FWE}}{NT}} \tag{7.12}$$

donde N_T representa el número de pruebas a realizar. Así, por ejemplo, si deseamos controlar α_{FWE} a .05 mientras realizamos tres pruebas, cada una se llevaría a cabo con un nivel de significancia de $\frac{.05}{3} = .17$. Para 10 pruebas el nivel apropiado sería $\frac{.05}{10} = .005$. Como se aprecia, a medida que el número de pruebas aumenta debemos reducir α_{PCE} para mantener el nivel deseado del error familiar.

La tabla 7.12 muestra los resultados de la aplicación de un ajuste de Bonferroni a α_{FWE}. Como lo demuestra esta tabla, la técnica de Bonferroni no establece α_{FWE} al nivel especificado, pero garantiza que no se eleve por encima de ese nivel. Así, los datos de la tabla 7.12 para α_{FWE} siempre se encuentran por debajo del nivel establecido de .05 pero sin ser iguales a .05. Ahora presentaremos una técnica que, aunque no es tan común como el ajuste de Bonferroni, generalmente es superior.

Método de reducción de Bonferroni para controlar errores familiares. En 1979, Holm [24] propuso una modificación al procedimiento de Bonferroni, que por lo general es más (y nunca menos) poderoso y mantiene el error familiar al mismo nivel que el procedimiento clásico. Este Bonferroni modificado, o mejor dicho, procedimiento de reducción de Bonferroni, se ilustra en la figura 7.3 de la siguiente página y se lleva a cabo como sigue.

1. Se calculan los estadísticos de la prueba múltiple.
2. Se obtiene el valor-p de cada estadístico calculado en 1.

[12] Llamado así en honor de Carlo Emilio Bonferroni, 1892-1960.

	Paso 1	Paso 2	Paso 3	...	Paso NT
Valor *p*	P (1)	P (2)	P (3)	...	P (NT)
Reducción	$\frac{FWE}{NT}$	$\frac{FWE}{NT-1}$	$\frac{FWE}{NT-2}$	...	$\frac{FWE}{1}$
Clásico	$\frac{FWE}{NT}$	$\frac{FWE}{NT}$	$\frac{FWE}{NT}$	...	$\frac{FWE}{NT}$

FIGURA 7.3: Ejemplo del procedimiento de reducción de Bonferroni de comparación múltiple.

3. Los valores p se ordenan del más pequeño al más grande y el más pequeño se denomina $p_{(1)}$, el segundo más pequeño $p_{(2)}$, y así sucesivamente, hasta que el más grande se denomina $p_{(NT)}$ y donde NT es el número de pruebas.
4. En el primer paso, $p_{(1)}$ se compara con $\frac{\alpha_{FWE}}{NT}$. Si $p_{(1)} \leq \frac{\alpha_{FWE}}{NT}$, la prueba se declara significativa y se sigue con el segundo paso. Si $p_{(1)} > \frac{\alpha_{FWE}}{NT}$, la prueba se declara no significativa

   y se suspenden las pruebas, y las comparaciones restantes se declaran no significativas.
5. Si el primer paso es significativo, el paso 2 se lleva a cabo comparando $p_{(2)}$ con $\frac{\alpha_{FWE}}{NT-1}$. Si $p_{(2)} \leq \frac{\alpha_{FWE}}{NT-1}$, el resultado se declara significativo y las pruebas continúan al siguiente paso. De otra forma, la prueba se declara no significativa y se suspenden las pruebas, declarando todas las comparaciones restantes como no significativas.
6. Los pasos se siguen tal como se muestra en la figura 7.3 hasta que se obtiene un resultado no significativo o hasta que se completa el último paso.

Es importante que se suspendan las pruebas cuando se obtiene el primer resultado no significativo. Si no se sigue esta "regla de suspensión", α_{FWE} no se limitará al nivel deseado.

Note que en el primer paso, la prueba es idéntica al ajuste clásico de Bonferroni, como se definió en la ecuación 7.12. Los pasos subsiguientes, si se llevan a cabo, son menos rigurosos que el procedimiento clásico de prueba. Así que, por ejemplo, si se deben realizar cinco pruebas con α_{FWE} limitado a .05, entonces α_{PCE} para el primer paso es $\frac{.05}{5}$, para el segundo paso es $\frac{.05}{4}$ y para los pasos posteriores son $\frac{.05}{3} = .0167$, $\frac{.05}{2} = .025$ y para el último paso es $\frac{.05}{1} = .05$. En contraste, el método clásico emplea el nivel más riguroso de .01 para todas las pruebas.

EJEMPLO 7.7

Suponga que se realiza un ensayo clínico para comparar los resultados de cirugías de trasplante de córnea en los que las córneas donadas se obtienen de donadores mayores (≥ 60 años de edad) y de donadores de menor edad (< 60 años de edad). Con este fin, se asigna aleatoriamente a los pacientes a uno de los dos grupos, y el primer grupo recibe córneas de los donadores mayores y el segundo de los donadores de menor edad. Cuatro de los resultados son interesantes para los investigadores: (A) si el paciente experimenta rechazo de la córnea dentro de los cinco años posteriores a la cirugía, (B) si la córnea se vuelve opaca dentro de los cinco años posteriores a la cirugía, (C) si se requiere más intervención médica dentro del primer año posterior a la cirugía, y (D) si la visión es de 20/30 o mejor dentro de los 30 días posteriores a la cirugía.

Los resultados son los siguientes. De los 60 pacientes de cada uno de los dos grupos, .12 de los pacientes del grupo 1 y .10 de los pacientes del grupo 2 experimentaron rechazo; .21 de los pacientes del grupo 1 y .03 de los pacientes del grupo 2 desarrollaron córneas opacas; .06 de los pacientes del grupo 1 y .07 del grupo 2 requirieron más tratamiento médico, y .68 de los pacientes del grupo 1 y .89 de los pacientes del grupo 2 alcanzaron una visión de 20/30 o mejor.

Realice análisis de extremos múltiples de estos resultados pero asegúrese de que el error familiar no exceda .05. ¿Qué forma de prueba estadística escogió usted para los análisis? ¿Por qué?

Solución Ya que los resultados son dicotómicos y se están comparando dos grupos, emplearemos la prueba Z de muestras independientes para una diferencia entre proporciones, como se describió en el apartado 6.3.1 de la página 230. Las Z obtenidas (con subíndices apropiados para identificar el resultado sobre el que se está conduciendo la prueba) se calculan a través de la ecuación 6.5 de la página 232, como sigue.

$$Z_A = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{.12 - .10}{\sqrt{\frac{(.12)(.88)}{60} + \frac{(.10)(.90)}{60}}} = .35$$

$$Z_B = \frac{.21 - .03}{\sqrt{\frac{(.21)(.79)}{60} + \frac{(.03)(.97)}{60}}} = 3.16$$

$$Z_C = \frac{.06 - .07}{\sqrt{\frac{(.06)(.94)}{60} + \frac{(.07)(.93)}{60}}} = -.22$$

$$Z_D = \frac{.68 - .89}{\sqrt{\frac{(.68)(.32)}{60} + \frac{(.89)(.11)}{60}}} = -2.90$$

Por medio del método descrito en la página 96 para calcular el valor-p de dos colas de una prueba Z, los valores-p asociados con los estadísticos anteriores son los siguientes.

$$p_A = .7264$$
$$p_B = .0016$$
$$p_C = .8258$$
$$p_D = .0038$$

Los cuatro valores-p, junto con una denominación de la prueba de la que cada uno se deriva, se presentan en orden ascendente a continuación. También se muestran los valores de reducción (S-D) de α_{PCE} para cada prueba de significancia. Como podemos ver, la prueba de B es significativa (S) porque el valor-p de .0016 es menor que el valor α_{PCE} de .0125 que se calculó al dividir .05 entre 4. De igual forma, la prueba de D es significativa porque el valor-p de .0038 es menor que el valor α_{PCE} de .0167 que se calculó al dividir .05 entre 3. La prueba de A no es significativa (NS) porque .7264 es mayor que .0250. Es importante comprender que, automáticamente, C se declara no significativa en este punto debido a la regla de suspensión. El investigador que conduce estas pruebas puede estar seguro de que α_{FWE} no es mayor que .05.

Prueba	B	D	A	C
Valor-p	.0016	.0038	.7264	.8258
S-D α_{PCE}	.0125	.0167	.0250	.0500
	S	S	NS	NS

■

EJEMPLO 7.8

Un investigador que trabaja en un estudio que emplea múltiples grupos de sujetos desea analizar una serie de hipótesis nulas mediante pruebas t de muestras independientes. A continuación se presentan las hipótesis nulas y sus respectivos valores-p, asociados con cada prueba. Utilice estos resultados para realizar un procedimiento de reducción de Bonferroni en el que α_{FWE} no sea mayor que .05. ¿Cómo se comparan estos resultados con los que se obtendrían de las pruebas de Bonferroni clásicas?

H_0:	Valor-p
$\mu_1 = \mu_3$	.0111
$\mu_2 = \mu_4$	.0419
$\mu_2 = \mu_5$	.0090
$\mu_3 = \mu_4$	.0200
$\mu_4 = \mu_5$	.0181

Solución A continuación se listan en orden ascendente los cinco valores p, junto con la prueba de hipótesis de la cual cada uno se deriva. También se muestran los valores de reducción de α_{PCE} (S-D) y los valores del procedimiento clásico de Bonferroni de α_{PCE} (CB) para cada prueba de significancia.

Como se observa aquí, la prueba de $\mu_2 - \mu_5$ es significativa porque el valor p de .0090 es menor que el valor α_{PCE} de .0100, que se calculó al dividir .05 entre 5. De igual manera, la prueba de $\mu_1 - \mu_3$ es significativa porque el valor p de .0111 es menor que el valor α_{PCE} de .0125 que se calculó al dividir .05 entre 4. La prueba de $\mu_4 - \mu_5$ no es significativa porque .0181 es mayor que .0167. Es importante comprender que, automáticamente, las pruebas de $\mu_3 - \mu_4$ y $\mu_2 - \mu_4$ se declaran no significativas en este punto debido a la regla de suspensión. Estas últimas dos pruebas no son significativas a pesar de que sus valores-p son menores que su valor asociado de α_{FWE}. Quizá se sienta tentado a declarar estas pruebas significativas, pero debe tener en mente que la violación de la regla de suspensión invalida el procedimiento. El investigador que lleva a cabo estas pruebas puede estar seguro de que α_{FWE} no es mayor que .05.

Note que si el investigador hubiera empleado el método clásico de Bonferroni, que por desgracia todavía es de uso corriente, sólo $\mu_2 - \mu_5$ habría sido significativa.

Prueba	$\mu_2 = \mu_5$	$\mu_1 = \mu_3$	$\mu_4 = \mu_5$	$\mu_3 = \mu_4$	$\mu_2 = \mu_4$
Valor p	.0090	.0111	.0181	.0200	.0419
S-D α_{PCE}	.0100	.0125	.0167	.0250	.0500
CB α_{PCE}	.0100	.0100	.0100	.0100	.0100
	S	S	NS	NS	NS

■

Método HSD de Tukey para controlar errores familiares. Como se dijo previamente, los métodos de Bonferroni y de reducción de Bonferroni para controlar el error familiar tienen muchas aplicaciones, incluyendo el análisis de comparación múltiple y el análisis de extremos múltiples. En contraste, la prueba HSD (diferencia honestamente significativa) de Tukey [47] está diseñada para usarse en situaciones de comparaciones múltiples en donde se deben llevar a cabo todas las comparaciones de pares de medias de grupos. Estas pruebas se realizan calculando el estadístico de prueba, comúnmente representado por q, para cada una de las $\frac{k(k-1)}{2}$ comparaciones y el estadístico q resultante se consulta en la tabla apropiada de valores críticos. El estadístico de prueba se define de la siguiente manera.

$$\boxed{q_{ij} = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{MS_w}{n_h}}}} \tag{7.13}$$

Los subíndices i y j denotan los dos grupos que se están comparando, por lo que $\bar{x}_i$ y $\bar{x}_j$ son las medias de los grupos i y j, respectivamente. MS_w es el cuadrado medio, tal como se calculó para un ANOVA de un factor mediante las ecuaciones 7.3 y 7.4 de la página 265. El símbolo n_h representa la media armónica de los tamaños de las dos muestras y se calcula

$$n_h = \frac{2}{\frac{1}{n_i} + \frac{1}{n_j}}$$

Cuando $n_i = n_j$, $n_h = n$, que es el tamaño de la muestra de cualquiera de los grupos.

EJEMPLO 7.9

Utilice los datos del estudio sobre las dietas de la tabla 7.1 en la página 266 para realizar una prueba HSD de Tukey. Comience por plantear las hipótesis nulas que han de contrastarse, después lleve a cabo las pruebas y, finalmente, exponga sus conclusiones. Mantenga α_{FWE} a un nivel de .05.

Solución Debido a que hay tres grupos y deseamos hacer todas las comparaciones de pares, tendremos $\frac{3(2)}{2} = 3$ hipótesis para poner a prueba. Éstas son

$$H_0 : \mu_1 = \mu_2$$
$$H_0 : \mu_1 = \mu_3$$
$$H_0 : \mu_2 = \mu_3$$

Los cálculos previos (véase la página 267) obtenidos al llevar a cabo un ANOVA de un factor de estos datos producen lo siguiente.

$$\bar{x}_1 = 203.2$$
$$\bar{x}_2 = 211.0$$
$$\bar{x}_3 = 184.4$$

De la página 266 obtenemos

$$MS_w = 625$$

Como el tamaño de las muestras es el mismo para todos los grupos, n_h será

$$n_h = \frac{2}{\frac{1}{n_i} + \frac{1}{n_j}} = \frac{2}{\frac{1}{5} + \frac{1}{5}} = 5$$

para todas las comparaciones. Los estadísticos de prueba para las tres comparaciones son, por medio de la ecuación 7.13,

$$q_{12} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{MS_w}{n_h}}} = \frac{203.2 - 211.0}{\sqrt{\frac{625}{5}}} = -.698$$

$$q_{13} = \frac{\bar{x}_1 - \bar{x}_3}{\sqrt{\frac{MS_w}{n_h}}} = \frac{203.2 - 184.4}{\sqrt{\frac{625}{5}}} = 1.682$$

$$q_{23} = \frac{\bar{x}_2 - \bar{x}_3}{\sqrt{\frac{MS_w}{n_h}}} = \frac{211.0 - 184.4}{\sqrt{\frac{625}{5}}} = 2.379$$

Los valores críticos de q[13] se obtienen del Apéndice E. En la tabla se revisa el número de medias en el análisis y los grados de libertad apropiados. Los grados de libertad para la prueba de Tukey son los mismos que los grados de libertad del denominador para el ANOVA de un factor, concretamente, $N - k$. Como hay un total de 15 sujetos y tres grupos, los grados de libertad para estas pruebas son 15 − 3 = 12. Remitiéndonos al Apéndice E para 3 medias y 12 grados de libertad, encontramos un valor crítico de 3.773. La decisión de rechazar o no rechazar la hipótesis nula se toma de la misma manera que para una prueba *t*. Es decir, para una prueba de dos colas, se rechazará la hipótesis nula si la q obtenida es mayor que o igual al valor de la tabla o si es menor que o igual al negativo del valor de la tabla.

Como se observa, ninguna de las hipótesis se rechaza, por lo que no se puede demostrar ninguna diferencia entre las medias de grupo. Este resultado no nos sorprende, ya que el ANOVA de un factor que se realizó con estos datos no fue significativo. ■

EJEMPLO 7.10

Utilice los indicadores de la calidad de la atención de emergencia de la tabla 7.5 de la página 271 para realizar una prueba HSD de Tukey. Mantenga el error familiar a un nivel de .05. ¿Cuántas comparaciones tendrán que hacerse? ¿Cuáles son sus conclusiones?

Solución Utilizando las sumatorias de la tabla 7.6 de la página 272, calculamos

$$\bar{x}_1 = \frac{47}{4} = 11.75$$

$$\bar{x}_2 = \frac{116}{8} = 14.50$$

$$\bar{x}_3 = \frac{115}{7} = 16.43$$

$$\bar{x}_4 = \frac{44}{5} = 8.80$$

[13] La distribución de q se conoce como la **distribución del rango estudentizado**.

De la página 271 obtenemos

$$MS_w = 5.56$$

Como el tamaño de las muestras no es igual,[14] n_h será diferente en las seis comparaciones. Para q_{12}

$$n_h = \frac{2}{\frac{1}{n_i} + \frac{1}{n_j}} = \frac{2}{\frac{1}{4} + \frac{1}{8}} = 5.33$$

Dejaremos que usted se encargue de los demás cálculos de n_h.

Las $\frac{4(3)}{2} = 6$ comparaciones son, entonces,

$$q_{12} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{MS_w}{n_h}}} = \frac{11.75 - 14.50}{\sqrt{\frac{5.56}{5.33}}} = -2.693$$

$$q_{13} = \frac{\bar{x}_1 - \bar{x}_3}{\sqrt{\frac{MS_w}{n_h}}} = \frac{11.75 - 16.43}{\sqrt{\frac{5.56}{5.09}}} = -4.478$$

$$q_{14} = \frac{\bar{x}_1 - \bar{x}_4}{\sqrt{\frac{MS_w}{n_h}}} = \frac{11.75 - 8.80}{\sqrt{\frac{5.56}{4.44}}} = 2.636$$

$$q_{23} = \frac{\bar{x}_2 - \bar{x}_3}{\sqrt{\frac{MS_w}{n_h}}} = \frac{14.50 - 16.43}{\sqrt{\frac{5.56}{7.47}}} = -2.237$$

$$q_{24} = \frac{\bar{x}_2 - \bar{x}_4}{\sqrt{\frac{MS_w}{n_h}}} = \frac{14.50 - 8.80}{\sqrt{\frac{5.56}{6.15}}} = 5.995$$

$$q_{34} = \frac{\bar{x}_3 - \bar{x}_4}{\sqrt{\frac{MS_w}{n_h}}} = \frac{14.50 - 8.80}{\sqrt{\frac{5.56}{5.83}}} = 7.813$$

Al revisar el Apéndice E con cuatro medias y $N - k = 24 - 4 = 20$ grados de libertad, encontramos una q crítica de 3.958. Como $q_{13} = -4.478$ es menor que -3.958 y $q_{24} = 5.995$ y $q_{34} = 7.813$ son mayores que 3.958, declaramos estas comparaciones significativas. Concluimos, por lo tanto, que las calificaciones de la calidad del servicio en el hospital 1 son significativamente más bajas que los del hospital 3, y que las calificaciones de los hospitales 2 y 3 son mayores que las del hospital 4. Podemos confiar en la validez de estas conclusiones porque el error familiar se ha mantenido en .05. ■

7.4.3 Comentarios adicionales con respecto de los procedimientos de comparación múltiple

A menudo, para tener alguna ventaja en la estadística usted debe pagar un precio. Esto es especialmente cierto para los MCP. Para tener control sobre α_{FWE}, usted debe reducir el nivel de significancia al que se realiza cada prueba. Como usted sabe, reducir α reduce también la

---

[14] La prueba HSD de Tukey se creó originalmente para muestras del mismo tamaño, por lo que se usaba n en lugar de n_h en la ecuación 7.13. Cuando el tamaño de las muestras no es igual, es decir, $n_h \neq n$, la prueba en ocasiones recibe el nombre de prueba de Tukey-Kramer.

potencia. Como resultado, a veces sucede que una prueba ANOVA de un factor o una prueba chi-cuadrada de 2 por k demuestra que no todos los valores paramétricos son iguales, sino que todas las comparaciones múltiples subsiguientes son no significativas. Entre otras razones, esto puede ser el resultado de una falta de potencia al nivel de la comparación individual.

También debe saber que el uso de MCP es un tema controversial entre los especialistas en estadística. Por razones filosóficas, algunos de ellos creen que estos métodos no deberían utilizarse. Al tomar su propia decisión al respecto en un contexto de investigación en particular, le sugerimos que considere la siguiente pregunta. ¿Qué me preocupa más: *a*) rechazar una o más hipótesis nulas verdaderas en este grupo de comparaciones, o *b*) no rechazar una o más hipótesis nulas falsas en este conjunto de pruebas? Si *a*) es de interés primordial, entonces, probablemente debería hacer los ajustes necesarios para controlar α_{FWE}. Si *b*) es más importante, quizá no deba hacer estos ajustes. También puede escoger una postura comprometida si decide controlar α_{FWE} al .10 en vez del más tradicional nivel de .05. Mientras que esto eleva un poco la probabilidad de un error familiar, también permite hacer ajustes menos rigurosos a α_{PCE}, y así no reduce el poder como en el caso de usar un nivel de .05.

Por último, como se dijo antes, existen muchos métodos para controlar el error familiar. Hemos demostrado solamente tres de ellos aquí y hemos descartado el uso de uno de ellos. Es recomendable que lea más sobre estos métodos al realizar sus trabajos de investigación. Un enfoque particularmente lúcido de este tema lo tiene Kirk [27], en tanto que puede encontrar una discusión más técnica en Hochberg y Tamhane [22].

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

ajuste de Bonferroni 285
alternativa de cambio 275
análisis de comparación múltiple 283
análisis de extremos múltiples 283
análisis de varianza (ANOVA) de un factor 264
cuadrado medio dentro (MS_w) 267
cuadrado medio entre (MS_b) 265
chi-cuadrada de 2 por k 276
efecto aditivo del tratamiento 275
error por comparación 282
error familiar 282
frecuencia esperada (f_e) 277
frecuencia observada (f_o) 277
grados de libertad del denominador 265
grados de libertad del numerador 267
$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ 264
$H_0 : \pi_1 = \pi_2 = \cdots = \pi_k$ 276
método de reducción de Bonferroni 285
método HSD de Tukey 289
procedimientos de comparación múltiple (MCP) 282
suma de cuadrados dentro (SS_w) 265
suma de cuadrados entre (SS_b) 267

EJERCICIOS

7.1 Investigadores interesados en el estrés y su relación con el trabajo en entornos industriales realizan un estudio en el que se compara el número de pulsaciones de tres grupos de trabajadores. El primer grupo consiste en personal de oficina que realiza funciones rutinarias, el segundo grupo trabaja con materiales peligrosos y el tercero realiza trabajos similares a los del segundo grupo pero sin entrar en contacto con materiales peligrosos. Las pulsaciones se registran a la mitad de la hora de la comida para los tres grupos. Las pulsaciones obtenidas son las siguientes.

Trabajadores de oficina	Materiales peligrosos	Materiales no peligrosos
58	88	65
64	59	70
71	74	79
66	80	66
79	81	74
74	69	79
70	90	60

a) Calcule SS_w y MS_w

b) Calcule SS_b y MS_b

c) Ponga a prueba la hipótesis $\mu_1 = \mu_2 = \mu_3$ al nivel .05 de significancia. ¿Cuál es su conclusión para la pregunta en cuestión?

7.2 Por rutina, en los centros de readaptación juvenil se alienta a los internos a que se sometan voluntariamente a pruebas de VIH. Para incrementar el índice de participación en el programa de pruebas, las autoridades han desarrollado tres diferentes programas de capacitación que enfatizan, respectivamente, los conocimientos acerca del virus, el compromiso con parejas futuras y la importancia del tratamiento temprano. Para determinar si los tres enfoques tienen una eficacia diferente al motivar la participación en el programa de pruebas, cada uno se pone en marcha en un centro de readaptación diferente. En un cuarto centro no se da ninguna capacitación, lo que sirve como control.

Los resultados son los siguientes. Centro de readaptación 1: 46 participan, 51 no. Centro de readaptación 2: 52 participan, 44 no. Centro de readaptación 3: 36 participan, 37 no. Centro de readaptación 4 (control): 24 participan, 64 no. Utilice estos datos para poner a prueba la hipótesis $H_0 : \pi_1 = \pi_2 = \pi_3 = \pi_4$ a un nivel de significancia de .05. ¿Este análisis realmente responde la pregunta planteada por los investigadores? ¿Qué pregunta de investigación es la que enfrenta este análisis?

7.3 Realice un análisis de los datos del ejercicio 7.2 que enfrenta la pregunta que interesa a los investigadores.

7.4 Utilice los datos del ejercicio 7.2 para poner a prueba las siguientes hipótesis, pero no permita que el error familiar sea mayor de .05. Use pruebas de dos colas.

(A) $H_0 : \pi_1 = \pi_4$

(B) $H_0 : \pi_2 = \pi_4$

(C) $H_0 : \pi_3 = \pi_4$

7.5 La siguiente cita referente al error familiar fue tomada de un conocido texto de bioestadística. ¿Está usted de acuerdo con esta cita? Dé las razones para su respuesta.

> Para realizar esta **prueba de comparación múltiple**, seleccionamos un nivel de significancia general, que denota *la probabilidad de que una o más de las hipótesis nulas sea falsa*.

7.6 Utilice los datos del ejercicio 7.1 para realizar la prueba HSD de Tukey. Mantenga el error familiar a un nivel de .05. ¿Cuáles son sus conclusiones? Dado el resultado del ANOVA obtenido en el ejercicio 7.1, ¿le sorprende este resultado? Explique su respuesta.

B. Las siguientes preguntas se refieren al estudio de caso B (página 470).

7.7 ¿Cuál cree usted que sea la probabilidad de que resulte significativa la F obtenida de un ANOVA de un factor realizado con los datos de línea base de la escala WOMAC A para los tres grupos? Explique su respuesta.

7.8 Suponga que un ANOVA de un factor realizado con los datos de línea base de la escala WOMAC A resultara significativo. ¿Qué concluiría usted de este resultado?

7.9 Utilice los datos de la semana 12 de la escala WOMAC A para poner a prueba la hipótesis $H_0 : \mu_s = \mu_w = \mu_p$ al nivel .05, donde los subíndices representan los grupos estándar, débil y placebo, respectivamente. Interprete el resultado.

7.10 Realice una prueba HSD de Tukey con los datos de la semana 12 de la escala WOMAC A. Mantenga el error familiar en .05.

7.11 Utilice los datos de la tabla J.3 para poner a prueba la hipótesis $H_0 : \pi_s = \pi_w = \pi_p$ al nivel .05 de significancia, donde π representa la proporción de sujetos que identifican correctamente su brazalete como real o falso. La respuesta correcta para el grupo débil es falso. Interprete el resultado.

7.12 Ponga a prueba las hipótesis

$H_0 : \pi_s = \pi_w$

$H_0 : \pi_s = \pi_p$

$H_0 : \pi_w = \pi_p$

sin permitir que el error familiar se eleve por encima de .05. Interprete los resultados.

C. Las siguientes preguntas se refieren al estudio de caso C (página 471).

7.13 Ponga a prueba la hipótesis $H_0 : RR = 1$ para cada uno de los periodos. No permita que el error familiar se eleve por encima de .05 para estas pruebas.

D. Las siguientes preguntas se refieren al estudio de caso D (página 471).

7.14 Ponga a prueba la hipótesis nula $H_0 : \mu_1 = \mu_2 = \mu_3$, donde μ_1 es la puntuación media de NPZ-8 de las personas infectadas con VIH que tuvieron evaluaciones positivas de la etapa de ADC, μ_2 es la media de los sujetos infectados con VIH que tuvieron evaluaciones negativas de la etapa de ADC y μ_3 es la media de las personas que no tienen el VIH. Realice esta prueba a un nivel $\alpha = .05$. *Si* esta prueba es significativa, lleve a cabo una prueba HSD de Tukey y establezca el error familiar en .05.

G. Las siguientes preguntas se refieren al estudio de caso G (página 473).

7.15 Ponga a prueba la hipótesis nula de que las proporciones de mellizos monocigóticos, dicigóticos y hermanos no mellizos con esclerosis múltiple son las mismas. Utilice $\alpha = .05$. Interprete el resultado.

7.16 Pruebe la diferencia en las proporciones de casos de esclerosis múltiple para todos los pares de grupos. No permita que el error familiar se eleve por encima de .05. Interprete sus resultados.

CAPÍTULO 8

Estimación de relaciones

8.1 ANTECEDENTES

Muchos de los procedimientos estadísticos con los cuales usted ya está familiarizado han sido diseñados para estimar relaciones. Por ejemplo, la ANOVA en un sentido y las pruebas chi-cuadrada de 2 por *k* pueden utilizarse para estimar si existe una relación entre los temas tratados en un estudio y un resultado determinado.

En este capítulo consideraremos el coeficiente de correlación de Pearson del producto-momento (P-M), el cual está diseñado para estimar una forma específica de relación entre dos variables continuas; en particular, el grado en el cual están relacionadas linealmente. También explicaremos brevemente la prueba chi-cuadrada para la independencia, cuya función es determinar si dos categorías de variables son independientes. Debido a que la prueba chi- cuadrada para la independencia es una extensión directa de la prueba chi-cuadrada de 2 por *k*, con la cual usted ya está familiarizado, en este apartado sólo daremos un breve vistazo a esta prueba.

8.2 COEFICIENTE DE CORRELACIÓN DE PEARSON DEL PRODUCTO-MOMENTO (P-M)

El coeficiente de correlación de Pearson del producto-momento (P-M) toma valores entre −1 y 1, inclusive. Para entender el P-M, hay que comprender este valor. El P-M ofrece dos piezas de información, a las que calificaremos como la naturaleza y la fortaleza de la relación entre dos variables continuas. En su momento, se explicará cada una de ellas. Sin embargo, antes de llegar a ese punto será útil aprender cómo se calcula el P-M.

8.2.1 Cálculo del coeficiente de correlación del producto-momento

Las ecuaciones para el cálculo de P-M son muchas y muy variadas, pero algebraicamente son idénticas. En esta sección presentaremos formas[1] conceptuales y computacionales de estas ecuaciones, así como una forma basada en las puntuaciones z.[2] La forma conceptual y la pun-

[1] Véase la página 36.
[2] Véase la página 42.

tuación z se presentan para facilitar su comprensión del P-M, mientras que la forma computacional será utilizada para los cálculos. La forma conceptual está dada por

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\left[\sum (x-\bar{x})^2\right]\left[\sum (y-\bar{y})^2\right]}} \tag{8.1}$$

donde *r* es el coeficiente de correlación de la muestra; *x* y *y* son dos variables que serán correlacionadas, y *n* es el número de observaciones apareadas. Observe que si las dos expresiones en el denominador (es decir, las sumas de los cuadrados de *x* y *y*) se dividieran entre $n - 1$, los resultados serán las varianzas muestrales de *x* y *y*. Asimismo, si el numerador del coeficiente de correlación estuviera dividido entre $n - 1$, el resultado es lo que llamamos **covarianza** de *x* y *y*. La covarianza es una medida de relación lineal entre dos variables, pero no es posible interpretarla debido a que depende de la escala sobre la cual las dos variables son medidas y, por lo tanto, no está definida por 1 y –1 como el P-M. Es posible escribir *r* con los divisores $n - 1$ referidos aquí, pero estos términos se cancelan en la expresión, por lo tanto, no son necesarios.

La forma computacional de *r* está dada por

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} \tag{8.2}$$

La ecuación 8.2 es algebraicamente equivalente a la ecuación 8.1, pero se rediseñó para facilitar el cálculo. Usted reconocerá las sumas de las formas cuadradas en el denominador. También notará que el único término en esta ecuación que no se ha utilizado con anterioridad en un cálculo es $\sum xy$. Éste se obtiene sumando los productos obtenidos de la multiplicación de cada valor de *x* por cada valor de *y*.

Una forma poco usual, pero conceptualmente relevante, está dada por

$$\boxed{r = \frac{\sum z_x z_y}{n-1}} \tag{8.3}$$

donde z_x y z_y son las variables *x* y *y* expresadas en la forma de la puntuación *z*. Regresaremos a la ecuación 8.3 en una sección posterior.

EJEMPLO 8.1

Utilice la ecuación 8.2 para calcular P-M para los datos de la tabla 8.1 de la siguiente página. Por el momento usted no necesita preocuparse por el significado de los datos de esta tabla.

Solución Para propósitos de cálculo sería conveniente ordenar los datos de la tabla 8.1 como se muestra en la tabla 8.2 de la siguiente página. Designaremos la puntuación por acceso como la variable *x* y la puntuación por el grado de salud como la variable *y*.

TABLA 8.1: Puntos del grado de salud y acceso al servicio médico para 15 sujetos.

Número de sujeto	Puntuación por acceso	Puntuación por salud
1	3	2
2	6	6
3	13	9
4	1	1
5	7	5
6	8	7
7	13	10
8	10	8
9	2	2
10	4	3
11	5	4
12	11	9
13	4	5
14	3	4
15	9	8

TABLA 8.2: Puntuaciones por el grado de salud y acceso al servicio médico ordenados para el cálculo de P-M.

	Número de sujeto	x	y	xy	x^2	y^2
	1	3	2	6	9	4
	2	6	6	36	36	36
	3	13	9	117	169	81
	4	1	1	1	1	1
	5	7	5	35	49	25
	6	8	7	56	64	49
	7	13	10	130	169	100
	8	10	8	80	100	64
	9	2	2	4	4	4
	10	4	3	12	16	9
	11	5	4	20	25	16
	12	11	9	99	121	81
	13	4	5	20	16	25
	14	3	4	12	9	16
	15	9	8	72	81	64
Σ		99	83	700	869	575

Utilizando las sumas en la tabla 8.2 y mediante la ecuación 8.2 se obtiene

$$
\begin{aligned}
r &= \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} \\
&= \frac{700 - \frac{(99)(83)}{15}}{\sqrt{\left[869 - \frac{(99)^2}{15}\right]\left[575 - \frac{(83)^2}{15}\right]}} \\
&= \frac{152.20}{\sqrt{[215.60][115.73]}} \\
&= .964
\end{aligned}
$$

■

EJEMPLO 8.2

Utilice la ecuación 8.2 para calcular el P-M con los datos en la tabla 8.3 de la siguiente página. Por ahora, usted no necesita preocuparse por el significado de los datos de esta tabla.

Solución Para propósitos de cálculo sería conveniente ordenar los datos de la tabla 8.3 como se muestra en la tabla 8.4 de la siguiente página. Designaremos el porcentaje de almuerzos gratuitos como la variable x y el porcentaje del uso de casco como la variable y.

Utilizando las sumas en la tabla 8.4 y mediante la ecuación 8.2 se obtiene

$$
\begin{aligned}
r &= \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}} \\
&= \frac{2703 - \frac{(194)\ (161)}{9}}{\sqrt{\left[5356 - \frac{(194)^2}{9}\right]\left[3409 - \frac{(161)^2}{9}\right]}} \\
&= \frac{-767.44}{\sqrt{[1174.22][528.89]}} \\
&= -.974
\end{aligned}
$$

Ahora que usted puede calcular P-M regresaremos al asunto del significado del resultado obtenido. En específico, examinaremos la naturaleza y la fortaleza de la relación entre x y y. ■

8.2.2 Naturaleza de la relación

Considere los datos en la tabla 8.1 en la página anterior. Estos datos son ficticios y representan dos mediciones tomadas a 15 sujetos. La primera es una puntuación obtenida con base en un cuestionario que representa el grado en el cual el sujeto tiene acceso al servicio médico. El rango de esta escala va de 1 a 13, donde 1 representa el no acceso y 13 indica que se tiene acceso completo a todas las facetas del servicio médico (es decir, atención médica, dental, etcétera). La segunda

TABLA 8.3: Porcentaje de estudiantes que tienen almuerzo gratuito o a un costo menor, y porcentaje de estudiantes que rutinariamente usan casco cuando montan su bicicleta en 9 escuelas.

Número de escuela	Porcentaje de almuerzo gratuito o a costo menor	Porcentaje del uso de casco
1	17	18
2	15	25
3	25	15
4	4	29
5	36	8
6	40	8
7	20	19
8	8	28
9	29	11

TABLA 8.4: Porcentaje de estudiantes que tienen almuerzo gratuito o a un costo menor, y porcentaje de estudiantes que rutinariamente usan casco cuando montan su bicicleta en 9 escuelas, ordenados para el cálculo de P-M.

	Número de escuelas	x	y	xy	x^2	y^2
	1	17	18	306	289	324
	2	15	25	375	225	625
	3	25	15	375	625	225
	4	4	29	116	16	841
	5	36	8	288	1296	64
	6	40	8	320	1600	64
	7	20	19	380	400	361
	8	8	28	224	64	784
	9	29	11	319	841	121
Σ		199	161	2703	5356	3409

medición es una puntuación global del índice de salud con un rango de 1 a 10. Uno representa un estado global muy deficiente de salud, mientras que 10 indica una salud global muy buena. Observe el patrón en estos datos. Las personas con alta puntuación de acceso también tienden a mostrar una puntuación alta de salud y viceversa. En este momento, las personas con puntuaciones bajas de acceso tienden a mostrar una puntuación baja de salud.

Este patrón puede observarse con más claridad en la gráfica bivariable de la figura 8.1 en la página 301. Una **gráfica bivariable** utiliza puntos para representar pares de puntuaciones para cada sujeto. Por ejemplo, si se lee hacia abajo sobre el eje X (acceso) y sobre el eje Y (salud), usted verá que un sujeto tuvo puntuaciones de acceso y salud de 11 y 9, respectivamente. Un segundo sujeto tuvo puntuaciones de 2 y 2. Es posible ver un patrón distinto en esta figura. Mientras las puntuaciones de acceso se incrementan, las puntuaciones de salud también tienden a aumentar. Cuando existe este tipo de relaciones, esto es, cuando los valores altos de una variable tienden a estar asociados con los valores altos de otra, y los valores bajos de una variable tienden a estar asociados con los valores bajos de otra, se dice que los datos están **positivamente correlacionados.** De acuerdo con esta circunstancia, el P-M será un número positivo. Éste es el caso para los datos de la tabla 8.1, cuya correlación calculada en la página 298 fue de .964.

Ahora considere los datos de la tabla 8.3 en la página anterior y la gráfica bivariable de la figura 8.2 de la siguiente página. Una medición rigurosa, pero comúnmente utilizada, de los estatus socioeconómicos globales de una escuela es el porcentaje de estudiantes que califican para el almuerzo gratuito o de costo menor. Los porcentajes de niños que califican para el almuerzo gratuito o de costo menor y los porcentajes reportados para el uso rutinario del casco cuando montan su bicicleta se muestran en la tabla 8.3 y en la figura 8.2; los porcentajes corresponden a nueve escuelas. En esta tabla y, más claramente en la figura, se observa que las escuelas con bajos porcentajes de niños que reciben el almuerzo gratuito o de costo menor tienden a mostrar porcentajes más altos reportados para la rutina de usar casco cuando montan en bicicleta, mientras que las escuelas con altos porcentajes de estudiantes que reciben el almuerzo gratuito o de menor costo típicamente utilizan con menos frecuencia el casco. Cuando existen relaciones de este tipo, esto es, cuando los altos valores de una variable tienden a estar asociados con los bajos valores de otra variable, y los bajos valores de una variable tienden a estar asociados con los altos valores de otra, se dice que los datos están **negativamente correlacionados**. En esta circunstancia, el P-M será un número negativo. En tal caso, para los datos de la tabla 8.3 la correlación calculada en la página 298 fue de –.974.

8.2.3 Fortaleza de la relación

En la descripción de la naturaleza de la relación entre dos variables se dijo que dos variables están positivamente correlacionadas cuando los valores altos en una de ellas *tienden* a estar asociados con los valores altos en la segunda variable, y los valores bajos en una variable *tienden* a estar asociados con valores bajos en la segunda variable. La palabra *tienden* también se utilizó en la descripción de la correlación negativa. Pero, ¿qué tan fuerte es esta tendencia? ¿Cuándo los datos están positivamente correlacionados, cada sujeto con un puntaje alto en una variable tiene también un puntaje alto en la otra, o algunos sujetos rompen el patrón? ¿Un sujeto con puntaje alto en una variable tiene una puntuación igualmente alta en la otra? Estas preguntas se refieren a la fortaleza de la relación entre dos variables.

Como recordará, el P-M puede tomar valores entre –1 y 1. El P-M tiene la máxima fortaleza en cualquiera de estos valores y pierde fortaleza mientras estos valores disminuyen hacia cero. En cero,[3] el coeficiente de correlación indica la fortaleza mínima. Comencemos por ver los ejemplos en los cuales $r = 1.0$ y $r = -1.0$.

Considere los datos de la tabla 8.5 en la página 302. Parece que las variables x y y en esta tabla están positivamente correlacionadas, pero, ¿qué tan fuerte es esa correlación? La aplicación de la ecuación 8.2 en la página 296[4] mostraría que P-M = 1.0 para estos datos. Pero, ¿qué tan exacto es $r = 1.0$?

Un coeficiente de correlación de 1.0 significa que cada sujeto tuvo exactamente la misma puntuación en las dos variables, cuando la escala de diferencias es eliminada expresando las dos variables en términos de las puntuaciones z.[5] Los datos en la tabla 8.6 muestran las variables en la tabla 8.5, expresadas como puntuaciones z. Le mostraremos cómo los dos primeros valores de

---

[3] Trataremos el caso de correlación cero en una sección aparte.
[4] Le dejamos a usted la tarea de realizar este cálculo.
[5] Antes de continuar, sería conveniente repasar la página 42.

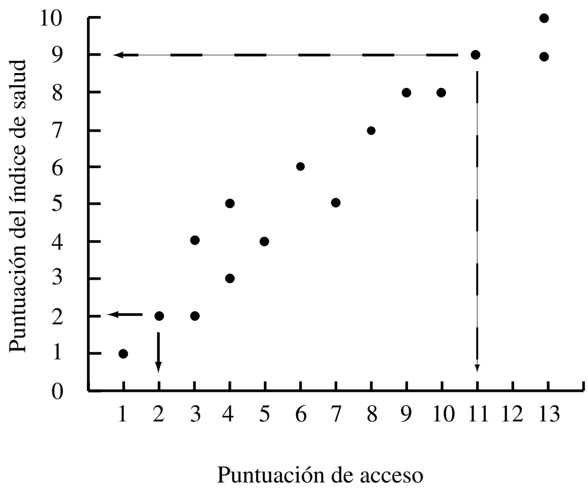


FIGURA 8.1: Gráfica bivariable de puntuaciones de cuidado de la salud e índice de salud positivamente correlacionados.

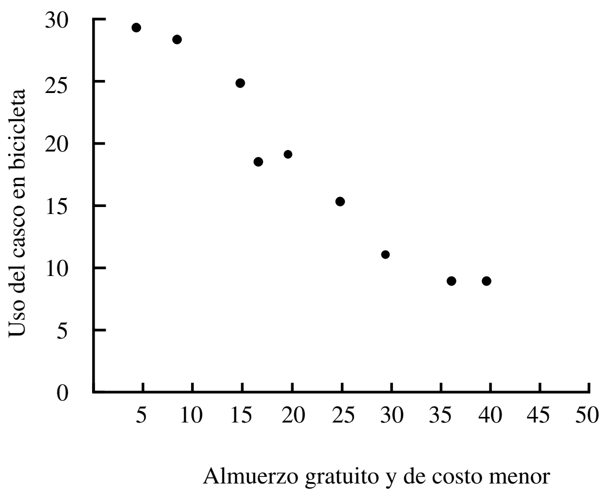


FIGURA 8.2: Gráfica bivariable de porcentajes de estudiantes que tienen almuerzo gratuito o de costo menor y porcentajes del uso de casco al montar en bicicleta correlacionados negativamente en nueve escuelas.

TABLA 8.5: Conjunto de datos donde $r = 1.0$.

Número de sujeto	x	y	x^2	y^2
1	4	11	16	121
2	6	15	36	225
3	11	25	121	625
4	4	11	16	121
5	9	21	81	441
6	5	13	25	169
7	8	19	64	361
8	13	29	169	841
Σ	60	144	528	2904

x y y fueron convertidos en puntuaciones z, pero le dejamos la tarea de confirmar las otras conversiones.

Los datos de muestreo se convierten en puntuaciones z a través de la aplicación de la ecuación 2.23, en la página 42 (que repetimos aquí para mayor comodidad).

$$z = \frac{x - \bar{x}}{s}$$

Aquí, x es la puntuación a ser convertida y $\bar{x}$ y s son la media y la desviación estándar de la muestra, respectivamente. Utilizando las sumas de la tabla 8.5 y calculando la media y la desviación estándar de x mediante la ecuación 2.1 (en la página 25) y la ecuación 2.16 (en la página 37) se obtiene

$$\bar{x} = \frac{\sum x}{n} = \frac{60}{8} = 7.5$$

y

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{528 - \frac{(60)^2}{8}}{7}} = 3.338$$

Entonces, para $x = 4$ y $x = 6$,

$$z_4 = \frac{x - \bar{x}}{s} = \frac{4.0 - 7.5}{3.338} = -1.049$$

y

$$z_6 = \frac{x - \bar{x}}{s} = \frac{6.0 - 7.5}{3.338} = -.449$$

Utilizando las ecuaciones 2.1 y 2.16 para y resulta

$$\bar{y} = \frac{\sum y}{n} = \frac{144}{8} = 18.0$$

TABLA 8.6: Valores de x y y de la tabla 8.5 expresados en puntuaciones z.

Número de sujeto	z_x	z_y	$z_x z_y$
1	−1.049	−1.049	1.100
2	−.449	−.449	.202
3	1.049	1.049	1.100
4	−1.049	−1.049	1.100
5	.449	.449	.202
6	−.749	−.749	.561
7	.150	.150	.023
8	1.648	1.648	2.716
Σ	00.000[a]	00.000	7.000[b]

[a] La suma de z_x y z_y es cero cuando los cálculos se efectúan con un número suficiente de lugares decimales.
[b] La suma de $z_x z_y$ es 7.000 cuando los cálculos se efectúan con un número suficiente de lugares decimales.

y

$$s = \sqrt{\frac{\sum y^2 - \frac{(\sum y)^2}{n}}{n-1}} = \sqrt{\frac{2904 - \frac{(144)^2}{8}}{7}} = 6.676.$$

Entonces, para $y = 11$ y $y = 15$,

$$z_{11} = \frac{y - \bar{y}}{s} = \frac{11.0 - 18.0}{6.676} = -1.049$$

y

$$z_{15} = \frac{y - \bar{y}}{s} = \frac{15.0 - 18.0}{6.676} = -.449.$$

Note que los sujetos 1 y 2 tienen exactamente los mismos puntos en x y y cuando estas variables son convertidas a la escala común de puntuaciones z. Esto también es verdadero para los sujetos restantes, como se puede ver en la tabla 8.6. Confirmamos que $r = 1$ por la aplicación de la ecuación 8.3, con lo cual se obtiene

$$r = \frac{\sum z_x z_y}{n-1} = \frac{7}{7} = 1.000.$$

Puesto que P-M evalúa el grado al cual x y y están linealmente relacionadas, cuando $r = 1.0$ la gráfica bivariable de x y y muestra que los puntos individuales caen en una línea. Esto puede observarse en la figura 8.3 de la siguiente página. En contraste, la gráfica de la figura 8.1 de los datos para los cuales $r = .964$ sugiere una relación lineal pero no constituye una línea, lo cual muestra que no hay una relación lineal perfecta.

Compare las puntuaciones z de la tabla 8.6, donde $r = 1.0$ con los datos de la tabla 8.7, los cuales fueron generados a partir de la tabla 8.1, donde $r = .964$. Aunque cada sujeto tuvo puntuaciones z similares en las variables de acceso y salud, sus puntuaciones no fueron idénticas dando

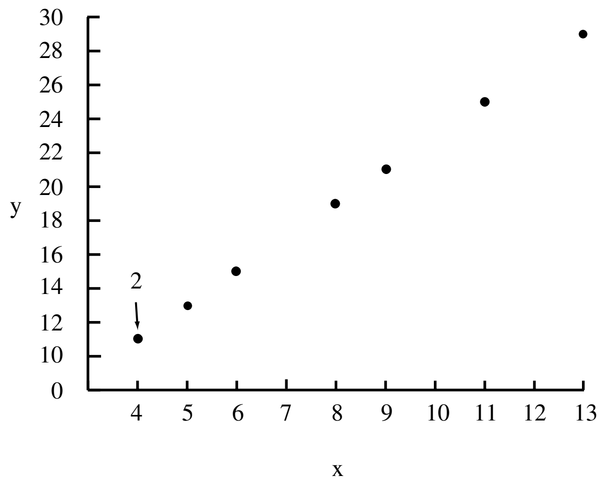


FIGURA 8.3: Gráfica bivariable de datos de la tabla 8.5 para los cuales $r = 1.0$.

como resultado un P-M menor que 1.00. No obstante, la similitud de los dos grupos de puntuaciones z fue suficiente para producir una correlación *cercana* perfecta de .964. El valor del coeficiente de correlación calculado con anterioridad puede confirmarse mediante la aplicación de la ecuación 8.3, la cual, usando los resultados de la tabla 8.7, da por resultado

$$\begin{aligned} r &= \frac{\sum z_x z_y}{n-1} \\ &= \frac{13.490}{4} \\ &= .964. \end{aligned}$$

Ahora, volveremos al caso en el cual $r = -1.0$. La aplicación de la ecuación 8.2 a los datos en la tabla 8.8 en la página siguiente producirá $r = -1.0$. Estos datos se convirtieron a puntuaciones z en la tabla 8.9.

Observe que cuando $r = -1.0$, la magnitud de la puntuación del sujeto para las dos variables es la misma, pero siempre con signo opuesto. De tal forma, si un sujeto está 1.5 desviaciones estándar por arriba de la media de la variable x, se encontrará 1.5 desviaciones estándar por debajo de la media de la variable y. Al igual que cuando $r = 1.0$, la gráfica bivariable de x y y cae en una línea cuando $r = -1.0$, pero, como se aprecia en la figura 8.4, la línea tiene una pendiente negativa. Observe que en los datos mostrados en la figura 8.2 de la página 301, $r = -.974$ tiende hacia una línea inclinada negativa, pero no constituye una línea. Esto es porque los datos están negativamente correlacionados, pero no a la perfección.

Así, en general, para datos positivamente correlacionados, cuanto más cercano esté r de 1, más similares serán las puntuaciones de las variables x y y cuando ambas variables estén expresadas en la escala común de puntuaciones z. Cuando $r = 1.0$ las puntuaciones z para las dos variables son idénticas. Para datos negativamente correlacionados, cuanto más cerca esté r de –1, más fuerte es la tendencia de las dos variables a tener igual magnitud pero con signos opuestos. Ahora examinaremos qué pasa cuando r es cero.

TABLA 8.7: Puntuaciones de acceso al cuidado médico y de salud de la tabla 8.1 expresadas en puntuaciones z.

	Número de sujeto	Puntuación por acceso	Puntuación por salud	Acceso X salud
	1	−.917	−1.229	1.127
	2	−.153	.162	−.025
	3	1.631	1.206	1.967
	4	−1.427	−1.577	2.250
	5	.102	−.185	−.019
	6	.357	.510	.182
	7	1.631	1.554	2.535
	8	.866	.858	.743
	9	−1.172	−1.229	1.440
	10	−.663	−.881	.584
	11	−.408	−.533	.217
	12	1.121	1.206	1.352
	13	−.663	−.185	.123
	14	−.917	−.533	.489
	15	.612	.858	.525
Σ		00.000	00.000[a]	13.490

[a] La suma de las puntuaciones z de salud es 0.000 cuando los cálculos se efectúan con suficientes lugares decimales.

TABLA 8.8: Conjunto de datos donde $r = -1.0$.

	Número de sujeto	x	y
	1	4	29
	2	6	25
	3	11	15
	4	4	29
	5	9	19
	6	5	27
	7	8	21
	8	13	11
Σ		60	176

TABLA 8.9: Valores de *x* y *y* de la tabla 8.8 expresados como puntuaciones *z*.

Número de sujeto	z_x	z_y	$z_x z_y$
1	−1.049	1.049	−1.100
2	−.449	−.449	−.202
3	1.049	− 1.049	− 1.100
4	−1.049	1.049	− 1.100
5	.449	−.449	−.202
6	−.749	.749	−.561
7	.150	−.150	−.023
8	1.648	− 1.648	− 2.716
Σ	00.000	00.000	− 7.000[a]

[a] La suma de $z_x z_y$ es −7.000 cuando los cálculos se efectúan con suficientes lugares decimales.

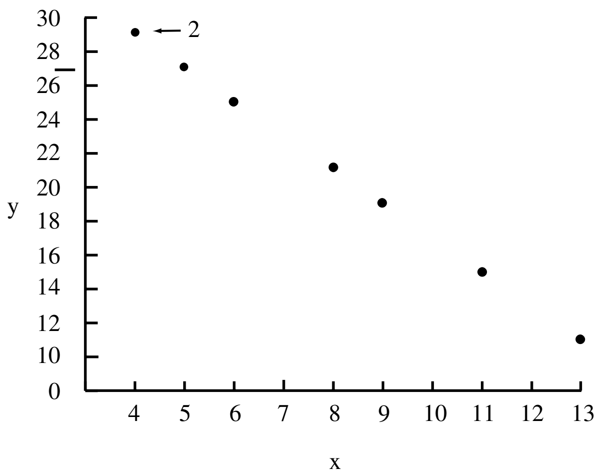


FIGURA 8.4: Gráfica bivariable de los datos de la tabla 8.8, para los cuales $r = -1.0$.

8.2.4 Correlación cero

Los investigadores ingenuos algunas veces interpretan la correlación cero entre x y y como si indicara que no hay relación entre las dos variables. Éste no es necesariamente el caso. Una correlación cero existirá cuando no hay relación entre x y y, pero también puede darse cuando ciertas relaciones *no lineales* están expresadas en los datos. Considere la gráfica bivariable de la figura 8.5.

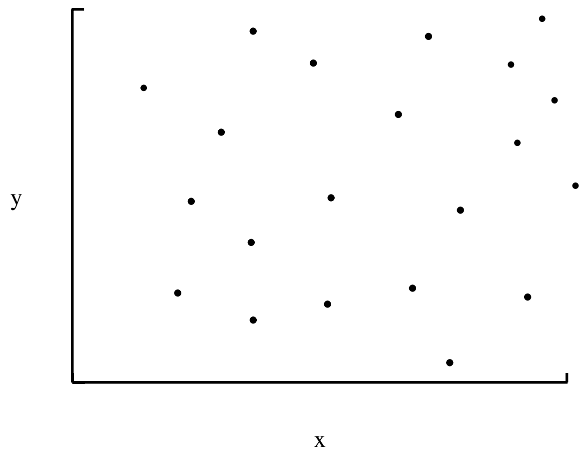


FIGURA 8.5: Gráfica bivariable en la cual no hay relación entre x y y.

En esta gráfica se representa una situación en la cual no hay relación entre las dos variables. De hecho, ningún patrón es observable en estos datos; tal conjunto generará una correlación cero porque x y y no están relacionadas. Ahora considere los datos de la tabla 8.10.

La aplicación de la ecuación 8.2 a estos datos mostrará que $r = 0.0$; pero, ¿esto significa que no hay relación entre x y y? La respuesta es claramente negativa, como se aprecia en la gráfica de la figura 8.6 de la siguiente página.

TABLA 8.10: Conjunto de datos para los cuales $r = 0.0$.

Número de sujeto	x	y
1	1	1
2	2	2
3	3	3
4	4	4
5	5	4
6	6	3
7	7	2
8	8	1

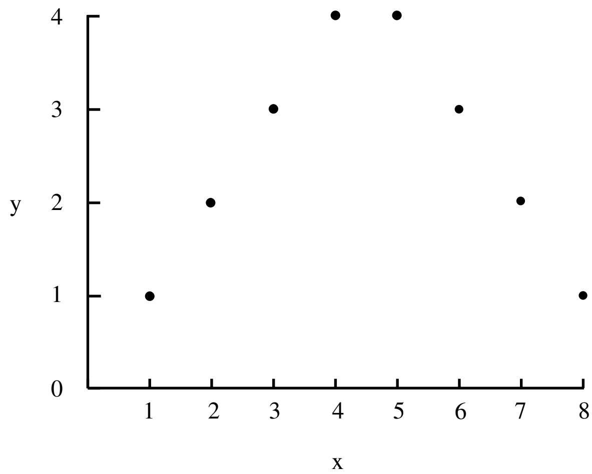


FIGURA 8.6: Gráfica bivariable de datos para los cuales $r = 0.0$, pero es evidente una relación entre x y y.

Es obvio que para valores de x entre 1 y 4, y se incrementa, mientras x aumenta; pero para valores de x entre 5 y 8, y disminuye con incrementos en x. P-M simplemente no está diseñado para detectar relaciones de este tipo. P-M está diseñado para detectar relaciones *lineales*. Esto es, relaciones que tienden a estar caracterizadas por una línea recta. Muchas otras formas no lineales no son detectables por P-M.

Para resumir, la correlación cero no necesariamente indica que no hay relación entre x y y.[6] Puede ser que no haya una relación, pero también es posible que haya una relación no evidente para P-M.

8.2.5 Relación causa-efecto

En ocasiones alguien podría sentirse tentado a interpretar fuertes correlaciones entre variables como un indicador de que una variable causa la otra. Ésta es la advertencia común que hacen los autores de textos introductorios a la estadística para prevenir a los estudiantes contra tales conclusiones. Dos variables pueden estar correlacionadas porque una causa la otra, pero también pueden estar correlacionadas cuando no existe una relación causal. Por esa razón, debemos ser extremadamente cautelosos al afirmar la existencia de una relación causal entre dos variables con base en la correlación observada entre ellas. Desde luego, sería imprudente sólo decir "no lo haga". Considere los siguientes ejemplos.

Investigadores pioneros en el ramo notaron una fuerte correlación positiva entre el número estimado de cigarrillos fumados a lo largo de la vida de una persona y la cantidad de daño en los pulmones observado en la autopsia. ¿Fumar daña los pulmones humanos? Numerosos estudios han demostrado que así es. Sin embargo, no se puede obtener una conclusión con base en la alta correlación positiva entre el número de cigarrillos fumados y la patología observada. Un ejemplo trivial pero revelador mostrará por qué.

---

[6] Hay excepciones teóricas como en el caso de la normalidad bivariable; pero éstas son de interés más teórico que práctico.

Suponga que por cada uno de los 365 días en un año dado nosotros registramos el dato de **1.** la cantidad de helado consumido ese día en Florida y **2.** el número de personas que se ahogaron en albercas o en la playa ese día en Florida. ¿Usted cree que esas dos variables están correlacionadas? Es obvio que la respuesta es afirmativa. En los días en que se consumieron grandes cantidades de helado hubo una tendencia a registrarse más ahogamientos. Cuando se consume poco helado, hay pocos ahogamientos.¿Comer helado provoca que la gente se ahogue? Más ridículamente, ¿los ahogamientos causan que la gente coma helado?[7]

La razón para la correlación entre estas dos variables es obvia, ambas están relacionadas con la temperatura. En días calurosos se consume más helado y más gente va a las albercas o a la playa. En días fríos, se come menos helado y hay menos gente nadando. Por lo tanto, pero estas variables (consumo de helado y ahogamientos) se encuentran sustancialmente relacionadas debido a su relación con una tercera variable, y no a una relación directa entre ellas. Es posible encontrar este tipo de situaciones en casi cualquier contexto de investigación. Hemos cumplido nuestra labor al prevenir al lector al respecto.

8.2.6 Prueba de hipótesis e intervalo de confianza

En este punto hemos caracterizado P-M como un estadístico descriptivo, es decir, un estadístico que describe la relación entre dos variables. Pero así como usamos la media muestral ($\bar{x}$) para hacer inferencias acerca de la media poblacional (μ), podemos usar el coeficiente de correlación de la muestra (r) para hacer inferencias acerca del coeficiente de correlación de la población (ρ). Haremos esto mediante los mecanismos de la prueba de hipótesis y los intervalos de confianza.

Prueba de $H_0 : \rho = 0$. En muchos casos el investigador estará interesado en determinar si existe una correlación entre x y y en una población. En tal circunstancia la hipótesis nula natural que se somete a prueba es

$$H_0 : \rho = 0$$

donde ρ es la correlación de la población. Esta prueba puede realizarse por medio de una prueba t cuya forma es la siguiente

$$\boxed{t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}} \tag{8.4}$$

Los grados de libertad para esta prueba son $n - 2$, donde n es el número de pares de puntos. La hipótesis alternativa de dos colas es

$$H_A : \rho \neq 0$$

con alternativas de una cola tomando una de las siguientes formas.

$$H_A : \rho < 0$$

o

$$H_A : \rho > 0$$

[7] Como es evidente, no existe un registro donde haya testigos de un ahogamiento que afirmen que la víctima fue escuchada pidiendo un cono doble de chocolate mientras se sumergía en el agua por tercera vez.

EJEMPLO 8.3

En la página 298 se indicó que la correlación entre los 15 pares de puntuaciones de salud y acceso al servicio médico fue .964. Realice una prueba de significancia para este coeficiente con las siguientes hipótesis nula y alternativa.

$$H_0 : \rho = 0$$
$$H_A : \rho > 0$$

Solución Por medio de la ecuación 8.4,

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{.964}{\sqrt{\frac{1-.964^2}{15-2}}} = 13.072$$

La referencia en el Apéndice B muestra que la t crítica para la prueba de una cola realizada con $\alpha = .05$ y 13 grados de libertad es 1.771. Puesto que la t obtenida excede este valor, la hipótesis nula se rechaza. Como resultado, el investigador puede asegurar que la correlación poblacional es mayor que cero. ■

Una prueba de $H_0 : \rho = \rho_0$. La mayoría de las veces el investigador querrá saber si la correlación poblacional difiere de cero. Ocasionalmente, sin embargo, el investigador querrá saber si la correlación poblacional difiere de algún otro valor que no sea cero. Debido a que la distribución muestral de r es simétrica cuando $\rho = 0$, la prueba t es robusta (o adecuada) para pruebas de $H_0 : \rho = 0$. Cuando el valor nulo específico es diferente de cero, la distribución muestral de r es asimétrica bajo la hipótesis nula, así que la prueba t quizá produzca resultados engañosos debido al fracaso de la suposición de normalidad. Como resultado, la ecuación 8.4 generalmente no será apropiada para la prueba. Fisher [15, 16] propuso un método de pruebas que supera en gran medida este problema. El siguiente estadístico de prueba está basado en el método de Fisher.

$$\boxed{Z = \frac{.5 \ln\left(\frac{1+r}{1-r}\right) - .5 \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{\sqrt{n-3}}}}} \tag{8.5}$$

Aquí, ln es el logaritmo natural, ρ_0 es el valor hipotético del coeficiente de correlación de la población, y n es el número de pares. Este estadístico tiene una distribución aproximadamente normal, así que la prueba puede realizarse mediante referencia a la curva normal.

EJEMPLO 8.4

A partir de la literatura, un investigador sabe que la correlación entre una forma abreviada de la Escala de valoración de actitud hacia el comportamiento sexual riesgoso y una forma más elaborada de la escala es .57. En un intento por mejorar la correlación entre las formas corta y larga de la escala, el investigador remplaza algunos reactivos en la forma corta. Luego, ambos instrumentos son administrados a un grupo de 18 estudiantes universitarios. El investigador encuentra una correlación de .71 entre la nueva versión de la forma corta y la forma larga. Utilice esta información para llevar a cabo la prueba de dos colas de la hipótesis nula.

$$H_0 : \rho = .57$$

Solución Por la ecuación 8.5

$$Z = \frac{.5\ \ln\left(\frac{1+r}{1-r}\right) - .5\ \ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{\sqrt{n-3}}}} = \frac{.5\ \ln\left(\frac{1+.71}{1-.71}\right) - .5\ \ln\left(\frac{1+.57}{1-.57}\right)}{\sqrt{\frac{1}{18-3}}} = \frac{.887-.648}{.258} = .926$$

Del Apéndice A se puede ver que los valores críticos de Z para una prueba de dos colas llevada a cabo con $\alpha = .05$ son -1.96 y 1.96. Como la Z obtenida está entre estos dos valores, la hipótesis nula no se rechaza. Esto significa que el investigador no fue capaz de demostrar un cambio en la correlación después de modificar la forma corta de la escala. ■

Un intervalo de confianza para la estimación de ρ. Un intervalo de confianza basado en los fundamentos de la ecuación 8.5 es posible y, probablemente, empleado con más frecuencia que el método presentado aquí. Sin embargo, tal método es por lo general superior y, por lo tanto, recomendable.

Las estimaciones para el límite inferior y superior de ρ pueden obtenerse a partir de:

$$\boxed{L = \frac{(1+F)r + (1-F)}{(1+F) + (1-F)r}} \tag{8.6}$$

$$\boxed{U = \frac{(1+F)r - (1-F)}{(1+F) - (1-F)r}} \tag{8.7}$$

En estas ecuaciones r es el coeficiente de correlación de la muestra y F es el valor apropiado del Apéndice C. Los grados de libertad para F son $n-2$ para ambos grados de libertad en el numerador y el denominador. Intervalos de uno y dos lados son posibles como se indica en la porción izquierda superior de la tabla de F.

EJEMPLO 8.5

Utilice los datos de la Escala de valoración de la conducta sexual riesgosa referida anteriormente para construir un intervalo de confianza de dos lados del 95% para la estimación de ρ. La correlación observada en ese estudio fue .71, que se obtuvo de los datos recolectados entre 18 estudiantes. Utilice el intervalo resultante para llevar a cabo la prueba de dos colas de $H_0 : \rho = .57$ con $\alpha =$ .05. ¿Cómo obtuvo su resultado?

Solución Por la ecuación 8.6,

$$L = \frac{(1+2.76).71 + (1-2.76)}{(1+2.76) + (1-2.76).71} = \frac{.910}{2.510} = .363$$

y

$$U = \frac{(1+2.76).71 - (1-2.76)}{(1+2.76) - (1-2.76).71} = \frac{4.430}{5.010} = .884$$

$F = 2.76$ se obtuvo consultando el Apéndice C para el intervalo de confianza de dos lados con grados de libertad del numerador y el denominador de $18 - 2 = 16$. El investigador puede, por lo tanto, estar 95 por ciento seguro de que el coeficiente de correlación de la población cae entre .363 y .884. La prueba de hipótesis de dos colas no será rechazada porque el valor hipotético de .57 cae entre los dos límites. ■

EJEMPLO 8.6

Dados $r = .82$ y $n = 52$, encuentre la estimación del límite inferior para ρ en la cual se puede tener un 99% de confianza.

Solución Por la ecuación 8.6

$$
\begin{aligned}
L &= \frac{(1+F)r+(1-F)}{(1+F)+(1-F)r} \\
&= \frac{(1+1.95).82+(1-1.95)}{(1+1.95)+(1-1.95).82} \\
&= \frac{1.469}{2.171} \\
&= .68
\end{aligned}
$$

$F = 1.95$ se obtuvo consultando el Apéndice C para el intervalo de confianza de un lado del 99% con 50 grados de libertad del numerador y el denominador. ■

8.2.7 Hipótesis

Contrariamente a la creencia de algunos investigadores, no hay suposiciones estadísticas que fundamenten el cálculo de r y su posterior uso como un estadístico descriptivo. Sin embargo, ciertas suposiciones se aplican a las pruebas de hipótesis y los intervalos de confianza asociados con este estadístico. Existen dos suposiciones primarias que preocupan. En primer lugar, se supone que los pares de datos son muestreados a partir de una distribución normal bivariable. Es posible pensar en esta distribución como una versión de tres dimensiones de la curva normal. Las pruebas de hipótesis y los intervalos de confianza pueden, o no, ser robustos ante las violaciones de esta suposición dependiendo de la forma de la distribución de la población y el valor de ρ. En general, los investigadores deben ser cautelosos cuando utilizan los métodos inferenciales presentados aquí y, en particular, deberán serlo cuando cualquier x y/o y parezca desviarse de la normalidad en un grado significativo.[8]

Se supone también que cada par de datos es independiente de todos los otros pares de datos. Una violación de esta suposición puede darse, por ejemplo, si el mismo sujeto fue parte de la prueba para x y y en dos momentos diferentes. No se pueden hacer inferencias respecto de ρ, basadas en estos métodos, cuando se viola esta suposición.

8.3 PRUEBA CHI-CUADRADA PARA LA INDEPENDENCIA

La prueba chi-cuadrada para la independencia se utiliza para probar la hipótesis nula en la que dos variables discretas son independientes contra la alternativa de que no sean independientes. Usted ya ha tratado con esta prueba, aunque en forma restringida, en el apartado 7.3 de la página 276. Esto se debe a que la prueba chi-cuadrada para la independencia es una forma más general que la prueba chi-cuadrada de por 2 por k con la que usted trató en esa sección.[9]

---

[8] La distribución normal de x y y no garantiza una normalidad bivariable, pero teniendo una x y/o y no normal se garantiza que la normalidad bivariable no se ha alcanzado.

[9] Antes de continuar, le recomendamos ampliamente releer el apartado 7.3 de la página 276.

TABLA 8.11: Descripción de una tabla chi-cuadrada de *j* por *k*.

		Variable 1			
		Categoría 1	Categoría 2	Categoría …	Categoría k
Variable 2	Categoría 1	f_{o11} f_{e11}	f_{o12} f_{e12}	… …	f_{o1k} f_{e1k}
	Categoría 2	f_{o21} f_{e21}	f_{o22} f_{e22}	… …	f_{o2k} f_{e2k}
	Categoría ⋮	⋮ ⋮	⋮ ⋮	⋮… ⋮…	⋮ ⋮
	Categoría j	f_{oj1} f_{ej1}	f_{oj2} f_{ej2}	… …	f_{ojk} f_{ejk}

La chi-cuadrada obtenida de esta forma se calcula mediante la ecuación 7.10 de la página 277, la cual repetimos aquí para mayor comodidad.

$$\chi^2 = \sum_{\text{todas las celdas}} \left[\frac{(f_o - f_e)^2}{f_e} \right]$$

Aquí f_o y f_e son las frecuencias observada y esperada. La **frecuencia observada** es simplemente el número de resultados que tienen lugar en una celda dada como se muestra en la tabla 8.11. En esta tabla utilizamos dos subíndices para indicar la fila y la columna de cada entrada de la celda. Por ejemplo, si la variable 1 representa el estatus de vacunación en contra de la hepatitis B con categoría "no vacunado", la categoría 2 representa "vacunado", y la categoría 3 representa "no se sabe"; la variable 2 indica el lugar de residencia, entonces f_{o11} será el número de personas en el lugar 1 que han sido vacunadas y f_{o21} será el número de personas en el lugar 2 quienes sí han sido vacunadas. Las entradas f_{o13} y f_{o23} representarán los números en esos lugares cuyo estatus es desconocido.

Los f_e son los números **esperados** en cada celda *si la hipótesis nula es verdadera* y se calculan mediante la ecuación 7.11 de la página 279, la cual se repite aquí por comodidad.

$$f_e = \frac{(N_R)\,(N_C)}{N}$$

Aquí N_R es la fila total de la celda cuya frecuencia esperada se está calculando, y N_C es la columna total para la misma celda.[10]

Esta notación "...", utilizada en la tabla 8.11, indica que el número de columnas continúa hasta la última columna designada como columna *k*. Asimismo, la notación "⋮" indica que el número de filas continúa hasta la última fila, la cual se designa como fila *j*. Todo esto significa que la tabla puede tener cualquier número de filas y columnas. Los grados de libertad para el estadístico

[10] Véase el apartado 7.3 en la página 276 para más detalles.

de prueba se calculan por medio de

$$\boxed{\chi^2_{df} = (j-1)(k-1)} \tag{8.8}$$

donde j y k son el número de filas y columnas en la tabla, respectivamente.

EJEMPLO 8.7

Suponga que en tres condados rurales se lleva a cabo una encuesta para determinar el estatus de vacunación en contra de la hepatitis B. Se encuentra que en el condado 1, 41 personas han sido vacunadas, 126 no han sido vacunadas y se desconoce el estatus de 452. En el condado 2, hay 202 personas vacunadas, 210 no han sido vacunadas y se desconoce el estatus de 440. En el último condado 330 personas han sido vacunadas, 614 no han sido vacunadas y se desconoce el estatus de 680.

Utilice estos datos para llevar a cabo el análisis chi-cuadrada. Interprete sus resultados.

Solución Las frecuencias observadas (corchetes) y esperadas (paréntesis) así como los totales de filas y columnas se ordenan como se muestra en la figura 8.12 de la siguiente página. Las frecuencias esperadas se calculan mediante la ecuación 7.11 como sigue.

$$f_{e11} = \frac{(N_{\text{Uno}})\,(N_{\text{V}})}{N} = \frac{(619)\,(573)}{3095} = 114.60$$

$$f_{e12} = \frac{(N_{\text{Uno}})\,(N_{\text{NV}})}{N} = \frac{(619)\,(950)}{3095} = 190.00$$

$$f_{e13} = \frac{(N_{\text{Uno}})\,(N_{\text{U}})}{N} = \frac{(619)\,(1572)}{3095} = 314.40$$

$$f_{e21} = \frac{(N_{\text{Dos}})\,(N_{\text{V}})}{N} = \frac{(852)\,(573)}{3095} = 157.74$$

$$f_{e22} = \frac{(N_{\text{Dos}})\,(N_{\text{NV}})}{N} = \frac{(852)\,(950)}{3095} = 261.52$$

$$f_{e23} = \frac{(N_{\text{Dos}})\,(N_{\text{U}})}{N} = \frac{(852)\,(1572)}{3095} = 432.74$$

$$f_{e31} = \frac{(N_{\text{Tres}})\,(N_{\text{V}})}{N} = \frac{(1624)\,(573)}{3095} = 300.66$$

$$f_{e32} = \frac{(N_{\text{Tres}})\,(N_{\text{NV}})}{N} = \frac{(1624)\,(950)}{3095} = 498.48$$

$$f_{e33} = \frac{(N_{\text{Tres}})\,(N_{\text{U}})}{N} = \frac{(1624)\,(1572)}{3095} = 824.86$$

TABLA 8.12: Datos de la encuesta ordenados para el análisis chi-cuadrada.

	Vacunado	No vacunado	Desconocido	
Condado 1	[41] (114.60)	[126] (190.00)	[452] (314.40)	619
Condado 2	[202] (157.74)	[210] (261.52)	[440] (432.74)	852
Condado 3	[330] (300.66)	[614] (498.48)	[680] (824.86)	1624
	573	950	1572	$N = 3095$

Entonces, por la ecuación 7.10, la chi-cuadrada obtenida es

$$\begin{aligned}
\chi^2 &= \sum_{\text{todas las celdas}} \left[\frac{(f_o - f_e)^2}{f_e} \right] \\
&= \frac{(41-114.60)^2}{114.60} + \frac{(126-190.00)^2}{190.00} + \frac{(452-314.40)^2}{314.40} + \frac{(202-157.74)^2}{157.74} \\
&+ \frac{(210-261.52)^2}{261.52} + \frac{(440-432.74)^2}{432.74} + \frac{(330-300.66)^2}{300.66} + \frac{(614-498.48)^2}{498.48} \\
&+ \frac{(680-824.86)^2}{824.86} \\
&= 47.27 + 21.56 + 60.22 + 12.42 + 10.15 + .12 + 2.86 + 26.77 + 25.44 \\
&= 206.81
\end{aligned}$$

El valor crítico se obtiene del Apéndice D con

$$(j - 1)(k - 1) = (3 - 1)(3 - 1) = 4 \text{ grados de libertad}$$

donde j es el número de filas y k el número de columnas en la tabla. En el caso presente, el Apéndice D muestra que para $\alpha = .05$ y cuatro grados de libertad, la χ^2 crítica es 9.488. La hipótesis nula se rechaza cuando la chi-cuadrada obtenida, es mayor que, o igual a, la chi-cuadrada crítica . Puesto que 206.81 es mayor que 9.488, la hipótesis nula no se rechaza. Por consiguiente, concluimos que el estatus de vacunación y el condado de residencia no son independientes. Esto significa que el estatus de vacunación depende del condado de residencia. Dicho de otra forma: las proporciones de los estatus de vacunados, no vacunados y estatus desconocido de las personas en los tres condados no son las mismas. ■

8.3.1 Suposiciones

La prueba chi-cuadrada para la independencia presentada aquí es una aproximación más que una prueba exacta.[11] Para que la aproximación sea lo suficientemente exacta, la regla empírica afirma

[11] Es posible realizar una prueba exacta, pero requiere de un programa de cómputo especializado, por lo que no se analizará aquí.

que al menos el 80% de la f_e en la tabla deberá ser mayor que, o igual a, cinco y ninguna celda deberá tener f_e menor que 1.

También se supone que las observaciones son independientes. Por ejemplo, un resultado no puede influir ni estar influido por otro resultado. En el último ejemplo, una violación pudo haber ocurrido si algunos individuos dieron múltiples veces sus respuestas y sus datos se consideraron varias veces en el análisis. La prueba chi-cuadrada no puede contarse entre las resistentes contra la violación de la suposición de independencia.

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

EJERCICIOS

8.1 Dadas las variables x y y, defina o explique los siguientes términos:
a) Correlación positiva.
b) Correlación negativa.
c) Correlación cero.
d) Fortaleza de la relación.
e) Relación lineal.
f) Correlación de 1.0 y −1.0.

8.2 Las puntuaciones APGAR son asignadas a los recién nacidos uno y cinco minutos después del nacimiento e indican el estado general de salud del bebé. Puntuaciones de siete a 10 son típicas e indican que el bebé sólo requiere cuidado posnatal de rutina. Puntuaciones de cuatro a seis indican que el bebé quizá necesite asistencia, mientras que puntuaciones de tres o menos indican que el bebé requiere de asistencia inmediata para mantenerlo con vida.
Suponga que se realiza un estudio para determinar la consistencia con la cual los puntos APGAR son asignados. Para tal fin, dos doctores observaron 12 nacimientos y asignaron independientemente los puntos APGAR a los bebés. Utilice los puntos APGAR en la tabla que se presenta a continuación para llevar a cabo cada una de las siguientes tareas. Designaremos las puntuaciones que otorga el primer médico como x y aquellas asignadas por el segundo médico como y.

Número de bebé	x	y
1	9	7
2	8	9
3	7	8
4	8	8
5	6	7
6	4	4
7	9	8
8	7	7
9	2	3
10	8	7
11	7	8
12	9	9

a) A partir de la inspección visual, ¿cree usted que hay una correlación positiva, negativa o de cero entre las dos variables? ¿Por qué?
b) Calcule el coeficiente de correlación de Pearson del producto-momento para los datos.

c) Lleve a cabo una prueba de dos colas de la hipótesis $H_0 : \rho = 0$, en el nivel de significancia de .05.

d) Construya un IC del 95% de dos lados para la estimación de ρ.

8.3 Como parte de un estudio piloto de seguridad, algunos motociclistas arrestados por conducir bajo la influencia del alcohol también se les realizó una prueba para detectar sustancias controladas. Los resultados de esas pruebas pueden ser positivo (+), negativo (–) o inconcluso (I). Para utilizar los recursos eficientemente, los investigadores desean determinar si los resultados de las pruebas están relacionados con la hora del arresto. Para tal fin, los tiempos de arresto se clasificaron en una de tres categorías: de medianoche a 8:00 AM (categoría 1), de 8:00 AM a 4:00 PM (categoría 2) y de 4:00 PM a medianoche (categoría 3).

Los resultados del estudio piloto son como sigue: categoría 1, 19 (+), 44 (–), 6 (I); categoría 2, 7 (+), 29 (–), 3 (I); categoría 3, 13 (+), 51 (–), 9 (I). Pruebe la hipótesis de que la hora del día y el resultado de la prueba son independientes contra la alternativa de que sí están vinculados. Utilice $\alpha = .05$.

D. Las siguientes preguntas se refieren al estudio de caso D (página 471).

8.4 Los investigadores reportan un –.50 de orden de correlación del rango de Spearman entre PBV y NPZ-8 para los 20 participantes estudiados. Calcule la correlación de Pearson para los mismos datos. ¿Consideraría los resultados como marcadamente diferentes de los generados por el rango de Spearman?

8.5 Dada la naturaleza asimétrica de la variable NPZ-8, ¿cree usted que el coeficiente de correlación que calculó en 8.4 es una expresión válida de la relación lineal entre PBV y NPZ-8? Explique su respuesta.

8.6 ¿Tendría usted alguna reserva acerca de realizar una prueba de hipótesis $H_0 : \rho = 0$ por medio de la ecuación 8.4 en la página 309 para la correlación de Pearson que calculó? Explique.

8.7 ¿La *naturaleza* de la relación expresada por la correlación de Pearson que calculó en el ejercicio 8.4 tiene sentido? Explique.

8.8 Calcule nuevamente la correlación de Pearson usando únicamente los 15 sujetos infectados de VIH. ¿El hecho de no considerar los cinco sujetos sanos hace una diferencia?

8.9 Puesto que la meta que establecieron los investigadores para este estudio fue determinar si la función neuropsicológica de personas infectadas con VIH está correlacionada con la pérdida de volumen cerebral, ¿podría decirse que su meta se cumplió?

8.10 Calcule la correlación de Pearson entre NPZ-8 y CD4 para los 15 sujetos infectados. Construya un IC del 95% para estimar ρ. ¿Tiene usted alguna reserva concerniente a este intervalo?

L. Las siguientes preguntas se refieren al estudio de caso L (página 475).

8.11 Lleve a cabo una prueba de significancia para determinar si la severidad de la lesión está vinculada con la edad. Utilice $\alpha = .05$. ¿Cuál es su conclusión?

M. Las siguientes preguntas se refieren al estudio de caso M (página 475).

8.12 Como resultado de la inspección visual, ¿parece haber una correlación positiva, negativa o de cero entre las mediciones antes y después del oxígeno?

8.13 Calcule P-M para los valores antes y después del oxígeno. Estime ρ con un intervalo de confianza del 95%.

8.14 Como se publicó originalmente, el valor "después" del oxígeno para el sujeto #4 se registró incorrectamente en la tabla como 4.9; el valor correcto es de 8.3. Repita M.8.13 utilizando el valor incorrecto de 4.9 en vez de 8.3. ¿Registrar este valor incorrecto en el análisis hizo una diferencia tan grande que la correlación resulta afectada?

CAPÍTULO 9

Regresión lineal

9.1 ANTECEDENTES

En el capítulo 8 aprendimos que para los pares de observaciones realizadas sobre un cierto número de sujetos u objetos, una correlación positiva entre las dos variables medidas implica que los valores altos de una variable tienden a estar asociados con los valores altos de la otra, y que los valores bajos de una tienden a estar asociados con los valores bajos de la otra. De igual manera, si las dos variables están correlacionadas negativamente, existe una relación inversa entre ellas.

Esta información sería útil si, por alguna razón, usted tuviera que observar una de las dos variables, digamos, x y predecir el valor de la otra variable, y. Es razonable que si un sujeto en particular tuviera un valor x alto y usted tuviera que predecir (o adivinar) su valor y, usted escogería un valor y alto si supiera que los datos están correlacionados positivamente. De igual modo, usted escogería un valor y bajo si supiera que los datos están correlacionados negativamente. Quizá sus cálculos no sean muy exactos, pero probablemente lo sean más que si usted no tuviera ningún conocimiento respecto de la relación entre x y y.

Se puede formalizar el uso de la relación entre dos variables para predecir el valor de una usando el valor de la otra, de tal manera que obtengamos las mejores predicciones posibles.[1] Además, y quizá más importante, esta metodología resultará útil para "explicar" la variación de una variable como consecuencia de su relación con otra u otras variables. Los modelos usados para generar estas predicciones y explicaciones son los temas de este capítulo.

Los modelos que se desarrollarán en el presente capítulo pueden clasificarse como modelos de regresión lineal simple o múltiple. Los modelos de regresión lineal simple (RLS) utilizan sólo una variable para predecir el valor de alguna otra, en tanto que los modelos de regresión lineal múltiple (RLM) usan una o más variables de predicción con este fin. Así, los modelos RLS constituyen la forma más sencilla de RLM. Por esta razón, comenzaremos con el modelo más simple.

[1] Más adelante definiremos la frase "las mejores predicciones posibles".

9.2 REGRESIÓN LINEAL SIMPLE

El modelo de regresión lineal simple se formula de la siguiente manera:

$$\boxed{\hat{y} = a + bx} \tag{9.1}$$

donde $\hat{y}$ (denominada "*y* con sombrero") es el valor *previsto* de *y*, *a* y *b* son constantes por definir a continuación y *x* es la variable sobre la cual se deben hacer las predicciones. Por ejemplo, si $a = 2$ y $b = 4$, entonces el valor previsto de *y* para un sujeto cuyo valor de la variable *x* es 10 sería

$$\hat{y} = 2 + 4(10) = 42$$

Pero, ¿qué tan buena es esta predicción? Una forma simple de calificar la exactitud de esta predicción es encontrar la diferencia entre el valor *y* real del sujeto y su valor previsto de *y*. Esta cantidad, $y - \hat{y}$, se conoce como un **residual**. Una pregunta obvia en este punto es: "¿Cómo se escogieron *a* y *b*?" Antes de contestar esta pregunta, será útil calcular *a* y *b* para algún conjunto de datos y evaluar las predicciones hechas a partir del modelo formulado.

9.2.1 Cálculo de *a* y *b*

Las ecuaciones para el cálculo de *a* y *b* son las siguientes.

$$\boxed{a = \hat{y} - (b)(\overline{x})} \tag{9.2}$$

$$\boxed{b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}} \tag{9.3}$$

Hay dos puntos notables en estas ecuaciones. Primero, *a* es una función de *b* por lo que *b* se debe calcular antes que *a*. Segundo, la ecuación para *b* es similar a la ecuación usada para el cálculo del coeficiente de correlación de Pearson del producto-momento.[2] Esta similitud nos lleva a la segunda ecuación para el cálculo de *b*, que nos será de utilidad cuando se conozca el coeficiente de correlación.

$$\boxed{b = r\sqrt{\frac{SS_y}{SS_x}}} \tag{9.4}$$

Aquí, *r* es el coeficiente de correlación de Pearson del producto–momento, SS_y es la suma *y* de cuadrados (por las siglas de *sum of squares*) y SS_x es la suma *x* de cuadrados. Por la ecuación 9.4 se puede ver que *b* será cero cuando *r* es cero, y *b* será positivo o negativo cuando *r* es positivo o negativo. Esto se deduce porque $\sqrt{\frac{SS_y}{SS_x}}$ es la raíz cuadrada del cociente de dos sumas de cuadrados y, por consiguiente, siempre será positivo.

[2] Véase la ecuación 8.2 de la página 296.

EJEMPLO 9.1

Utilice los datos de la tabla 8.2 de la página 297 para elaborar un modelo de RLS para la predicción de puntuaciones de salud a partir de las puntuaciones de acceso a servicios médicos.

Solución Las puntuaciones de acceso se designan como x y las puntuaciones de salud como y; utilizando las sumas de la tabla 8.2 con la ecuación 9.3, se obtiene

$$b = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} = \frac{700 - \frac{(99)(83)}{15}}{869 - \frac{(99)^2}{15}} = \frac{152.20}{215.60} = .7059$$

Observe que, en general, b representa el cambio en $\hat{y}$ que ocurre por cada unidad de cambio en x. Para este caso, un incremento de una unidad en la puntuación de acceso da como resultado un incremento de .7059 en la puntuación de salud prevista. De igual forma, una disminución de una unidad en la puntuación de acceso da como resultado una disminución de .7059 en la puntuación de salud prevista.

Ya que previamente calculamos que r (en la página 298) sería .9635,[3] obtenemos el mismo resultado mediante la ecuación 9.4 si calculamos

$$SS_x = \sum x^2 - \frac{(\sum x)^2}{n} = 869 - \frac{(99)^2}{15} = 215.6000$$

y

$$SS_y = \sum y^2 - \frac{(\sum y)^2}{n} = 575 - \frac{(83)^2}{15} = 115.7333$$

Entonces

$$b = r\sqrt{\frac{SS_y}{SS_x}} = .9635\sqrt{\frac{115.7333}{215.6000}} = .7059$$

que es el mismo resultado obtenido antes.

Podemos obtener a calculando

$$\overline{y} = \frac{\sum y}{n} = \frac{83}{15} = 5.5333$$

y

$$\overline{x} = \frac{\sum x}{n} = \frac{99}{15} = 6.6000$$

Luego, por la ecuación 9.2

$$a = \overline{y} - b\overline{x} = 5.5333 - (.7059)(6.6000) = .8744$$

Entonces, el modelo de predicción es

$$\hat{y} = a + bx = .8744 + .7059x$$

Las puntuaciones de salud previstas obtenidas con la aplicación de este modelo se muestran en la tabla 9.1.

---

[3] Efectuamos éste y otros cálculos con más cifras decimales para aumentar la exactitud de los resultados.

9.2.2 Las sumas de cuadrados residuales y de regresión y los coeficientes de determinación y no determinación

La tabla 9.1 muestra las puntuaciones de salud (y) y de acceso a servicios médicos (x), tomadas de la tabla 8.2 de la página 297. También se muestran las puntuaciones de salud previstas $\hat{y}$, los cuadrados residuales ($(y-\hat{y})^2$) y la cantidad $(\hat{y}-\overline{y})^2$, que será explicada a continuación. Los valores previstos se obtuvieron mediante la aplicación del modelo construido con anterioridad. Por ejemplo, los valores previstos para los primeros dos sujetos se calcularon de la siguiente manera:

$$\hat{y}_1 = .8744 + .7059x_1 = .8744 + .7059\ (3) = 2.9921$$

y

$$\hat{y}_2 = .8744 + .7059x_2 = .8744 + .7059\ (6) = 5.1098.$$

Los residuales elevados al cuadrado para los primeros dos sujetos son, entonces,

$$(y_1 - \hat{y}_1)^2 = (2 - 2.9921)^2 = .9843$$

y

$$(y_2 - \hat{y}_2)^2 = (6 - 5.1098)^2 = .7925.$$

Las primeras dos observaciones en la columna con el encabezado $(\hat{y}-\overline{y})^2$ se obtuvieron de la siguiente manera. Tomando en cuenta que $\overline{y} = 83/15 = 5.5333$,

$$(\hat{y}_1 - \overline{y}) = (2.9921 - 5.5333)^2 = 6.4577$$

y

$$(\hat{y}_2 - \overline{y}) = (5.1098 - 5.5333)^2 = .1794.$$

La suma y de cuadrados (SS_y) representa la variación de la variable resultante (y). Esto es, representa el hecho de que algunos sujetos tienen puntuaciones de salud muy por encima del valor de la puntuación promedio de salud (5.5333), en tanto que otros sujetos tienen mucho menos. ¿Cómo se explica esta variación? Para responder esta pregunta, primero tenga en mente que

$$\boxed{SS_y = SS_{reg} + SS_{res}} \tag{9.5}$$

donde SS_{reg} es la **suma de los cuadrados de regresión** y se calcula mediante

$$\boxed{SS_{reg} = \sum (\hat{y} - \overline{y})^2} \tag{9.6}$$

y SS_{res} es la **suma de los cuadrados residuales** y se calcula mediante

$$\boxed{SS_{res} = \sum (y - \hat{y})^2} \tag{9.7}$$

Cuando estas dos sumas de la tabla 9.1, en la siguiente página, se suman, el resultado es el

$$SS_y = 107.4322 + 8.2898 = 115.7720$$

TABLA 9.1: Puntuaciones de salud previstas y sumas de cuadrados residuales y de regresión.

Número de sujetos	Salud y	$\hat{y}$	Acceso x	$(y-\hat{y})^2$	$(\hat{y}-\bar{y})^2$
1	2	2.9921	3	.9843	6.4577
2	6	5.1098	6	.7925	.1794
3	9	10.0511	13	1.1048	20.4105
4	1	1.5803	1	.3367	15.6262
5	5	5.8157	7	.6654	.0797
6	7	6.5216	8	.2289	.9767
7	10	10.0511	13	.0026	20.4105
8	8	7.9334	10	.0044	5.7605
9	2	2.2862	2	.0819	10.5437
10	3	3.6980	4	.4872	3.3683
11	4	4.4039	5	.1631	1.2755
12	9	8.6393	11	.1301	9.6472
13	5	3.6980	4	1.6952	3.3683
14	4	2.9921	3	1.0159	6.4577
15	8	7.2275	9	.5968	2.8703
Σ	83	83.000	99	8.2898	107.4322

cual, exceptuando los errores de redondeo, es el mismo resultado obtenido antes cuando se calculó SS_y directamente. ■

Examinemos estos dos componentes de SS_y más de cerca. Primero, note que $y_i - \hat{y}_i$ es una medida de la cantidad por la cual erró el modelo al predecir el valor y del sujeto i. Si deseamos distinguir el error total en la predicción de las puntuaciones de un grupo de n sujetos quizá se sienta tentado a tomar $\sum(y-\hat{y})$ como tal, pero esto no sería una medida satisfactoria porque la suma siempre es cero. En vez de sumar la diferencia entre y y $\hat{y}$, se suman las diferencias elevadas al cuadrado, como se indica en la ecuación 9.7.

Con anterioridad, planteamos la pregunta sobre cómo se escogieron los valores de a y b. La respuesta es que son los valores que minimizan a SS_{res}. Es en este sentido que a y b producen el mejor modelo posible.[4] Por esta razón, los modelos de este tipo se denominan **modelos de mínimos cuadrados**.

El **coeficiente de no determinación** se simboliza por $1-\hat{R}^2$ y es la proporción de SS_{res} a SS_y, o más formalmente,

$$\boxed{1-\hat{R}^2 = \frac{SS_{res}}{SS_y}} \tag{9.8}$$

[4] También hay otros "mejores modelos posibles". Por ejemplo, se puede construir un modelo que minimice a $\sum|y-\hat{y}|$.

El **coeficiente de no determinación** es, entonces, la proporción de la suma y de cuadrados que no está explicada por el modelo.

Pero, ¿qué pasa con el otro componente de SS_y? Si $SS_y = SS_{reg} + SS_{res}$ y definimos SS_{res}/SS_y como la proporción de SS_y no explicada por x, entonces, es razonable definir la proporción que *sí es* explicada por x como la proporción de SS_{res} a SS_y. El **coeficiente de determinación** se simboliza por $\hat{R}^2$ y se define formalmente como

$$\boxed{\hat{R}^2 = \frac{SS_{reg}}{SS_y}} \tag{9.9}$$

El **coeficiente de determinación** es, entonces, la proporción de la suma y de cuadrados explicada por el modelo.

EJEMPLO 9.2

Encuentre e interprete los coeficientes de no determinación y determinación para los datos de salud y acceso de la tabla 9.1 de la página anterior.

Solución Usando $SS_y = 115.722$, $SS_{reg} = 107.4322$ y $SS_{res} = 8.2898$, como se calculó anteriormente,

$$1 - \hat{R}^2 = \frac{SS_{res}}{SS_y} = \frac{8.2898}{115.722} = .072$$

y

$$\hat{R}^2 = \frac{SS_{reg}}{SS_y} = \frac{107.4322}{115.722} = .928$$

Esto significa que aproximadamente el 93% de la variación de las puntuaciones de salud estuvo relacionado con las puntuaciones de acceso a servicios médicos del individuo.[5] Esto también significa que aproximadamente el 7% de esta variación no se explica por el acceso a servicios médicos y, por consiguiente, debe atribuirse a otros factores desconocidos. ■

9.2.3 Notas sobre el cálculo de SS_{res} y SS_{reg}

En los cálculos realizados con anterioridad calculamos todos los valores de $\hat{y}$ para encontrar SS_{res} y SS_{reg}. Utilizamos este método para ayudarle a comprender los coeficientes de determinación y no determinación. Sin embargo, no es necesario calcular la $\hat{y}$ para obtener estas cantidades. Ahora demostraremos esos métodos alternativos, pues serán útiles más adelante, cuando estudiemos la regresión lineal múltiple.

Si sustituimos $a + bx$ por $\hat{y}$ en la ecuación 9.7 de la página 322, usamos algo de álgebra básica y las reglas sumatorias descritas en el apartado 2.3.2 de la página 13, se obtiene el siguiente resultado.

$$\boxed{SS_{res} = \sum y^2 - a\sum y - b\sum xy} \tag{9.10}$$

[5] Ésta es una proporción muy poco realista para ser atribuida al acceso a servicios médicos, pero queríamos impresionarlo con un modelo fuerte ya que es su primer contacto con la regresión lineal simple.

Si sustituimos los valores calculados antes en la ecuación 9.10 obtenemos

$$SS_{res} = 575 - (.8744)(83) - (.7059)(700) = 8.2948$$

que, sin el redondeo,[6] es el mismo resultado obtenido previamente.

Podemos calcular SS_{reg} directamente de la siguiente manera.

$$\boxed{SS_{reg} = b^2 SS_x} \tag{9.11}$$

Si sustituimos los valores calculados antes, tenemos que

$$SS_{reg} = \left(.7059^2\right)(25.6) = 107.4324$$

que una vez más, sin el redondeo, es el resultado calculado previamente.

9.2.4 Más comentarios sobre los coeficientes de determinación y no determinación

Cuando la correlación entre x y y sea uno o uno negativo, el modelo de regresión nos dará la predicción perfecta de y. Esto significa que para cada sujeto $\hat{y} = y$, de manera que $SS_{res} = \sum \left(y - \hat{y}\right)^2$ se convierte en $\sum (y - y)^2 = 0$, lo que significa que $1 - \hat{R}^2 = \frac{SS_{res}}{SS_y}$ es cero. Esto también implica que $SS_{reg} = \sum \left(\hat{y} - \overline{y}\right)$ se convierte en $\sum (y - \overline{y})^2 = SS_y$, de manera que $\hat{R}^2 = \frac{SS_{reg}}{SS_y} = \frac{SS_y}{SS_y}$ es 1. Entonces, cuando $r = 1$ o -1, toda la variación en y se debe a x, de manera que no hay variación en y que quede sin explicar.

En contraste, cuando la correlación entre y y x es cero, b es, por la ecuación 9.4, también cero para que $a = \overline{y} - b\overline{x}$ se convierta simplemente en $\overline{y}$. El modelo de regresión es, entonces, $\hat{y} = \overline{y}$. Así, cuando x no tiene relación lineal con y el modelo predecirá $\overline{y}$ para cada sujeto sin importar su puntuación x. En este caso, $SS_{res} = \sum \left(y - \hat{y}\right)^2$ se convierte en $\sum (y - \overline{y})^2 = SS_y$, por lo que $1 - \hat{R}^2$ se convierte en $\frac{SS_y}{SS_y} = 1.0$. Al mismo tiempo, SS_{reg} será $\sum (\overline{y} - \overline{y})^2$, que es cero. Entonces, cuando no existe relación lineal entre x y y , x no explicará ninguna de las variaciones en y, por lo que no hay explicación alguna para esta variación.

Los símbolos $\hat{R}^2$ y $1 - \hat{R}^2$ para los coeficientes de determinación y no determinación, respectivamente, no son arbitrarios. Si permitimos que r_{yx} represente la correlación de Pearson del producto-momento entre y y x y $r_{y\hat{y}}$ represente la correlación entre y y $\hat{y}$, entonces en general, el símbolo $\hat{R}$ representa $r_{y\hat{y}}$, por lo que se deduce que $\hat{R}^2 = r^2_{y\hat{y}}$ y $1 - \hat{R}^2 = 1 - r^2_{y\hat{y}}$. En el presente caso esto es fácil de verificar, ya que para la regresión lineal simple (no así para la regresión múltiple que discutiremos posteriormente), $r_{y\hat{y}} = r_{yx}$, que antes calculamos como .9635. Observamos que para los datos de la tabla 8.2, $\hat{R}^2 = r^2_{y\hat{y}} = r^2_{yx} = .9635^2 = .928$, que es el mismo resultado obtenido con la ecuación 9.9. El coeficiente de no determinación para estos datos sería $1 - .9635^2 = .072$, que es el resultado obtenido mediante la ecuación 9.8.

[6] Antes calculamos este valor como 8.2898.

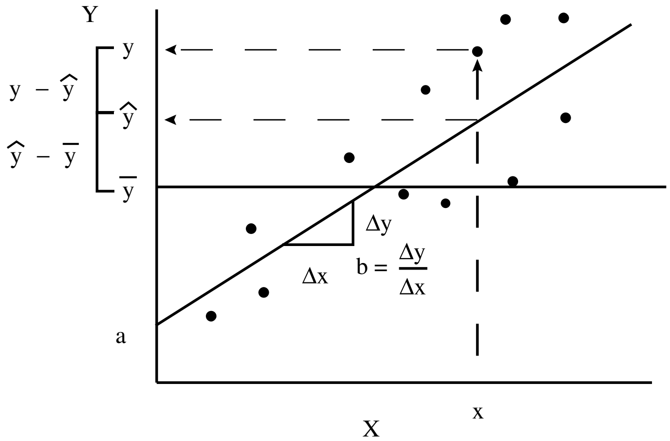


FIGURA 9.1: Descripción de la relación entre un modelo de regresión lineal simple, los valores previstos de y y $y - \hat{y}$ y $\hat{y} - \bar{y}$.

Además de las expresiones en las ecuaciones 9.8 y 9.9, podemos caracterizar los coeficientes de no determinación y determinación como

$$\boxed{1 - \hat{R}^2 = 1 - r_{y\hat{y}}^2} \qquad (9.12)$$

y

$$\boxed{\hat{R}^2 = r_{y\hat{y}}^2} \qquad (9.13)$$

La relación entre el modelo de RLS, los valores previstos y las cantidades $y - \hat{y}$ y $\hat{y} - \bar{y}$ se muestran en la figura 9.1. Usted reconocerá $\hat{y} = a + bx$ como la ecuación para una línea. Esta **línea de regresión de mínimos cuadrados** se describe en la figura junto con la gráfica bivariable de algún conjunto de datos arbitrarios. Como se muestra en la figura, los términos a y b en el modelo representan la ordenada y la pendiente de y, respectivamente. Si hablamos en términos geométricos, esta línea relaciona cada valor de x con un valor previsto de y, como se muestra para uno de los valores de x. La línea se construye para minimizar las distancias verticales al cuadrado de las observaciones desde la línea. Es decir, $\sum (y - \hat{y})^2$ es menor de lo que se lograría con cualquier otra línea.

9.2.5 Inferencia respecto de *b* y $\hat{R}^2$

Tal como se puede probar la significancia de la media muestral y otros estadísticos que usted ya conoce, se puede hacer con a, b y $\hat{R}^2$. En general, una prueba para a es de poco interés, por lo que no ahondaremos en ella. Las pruebas para b y $\hat{R}^2$ producen el mismo resultado en RLS, pero no así en RLM. Describiremos de ambas pruebas para que usted esté familiarizado con ellas cuando comience su estudio de los modelos de regresión lineal múltiple.

Mediante la ecuación 9.14 se puede hacer una prueba de la hipótesis nula

$$H_0 : \beta = 0$$

donde β es la contraparte del parámetro de b.[7]

$$\boxed{t = \frac{b}{\sqrt{\frac{MSres}{SSx}}}} \tag{9.14}$$

MS_{res} es el **residual cuadrático medio** que se define como SS_{res} dividido entre sus grados de libertad correspondientes, que son $n - 2$. SS_x es la suma x de cuadrados, como se vio en la página 320. El estadístico t resultante se refiere a una distribución t con $n - 2$ grados de libertad.

Se pueden obtener los límites inferior y superior de un intervalo de confianza para estimar β mediante

$$\boxed{L = b - t\sqrt{\frac{MSres}{SSx}}} \tag{9.16}$$

$$\boxed{U = b + t\sqrt{\frac{MSres}{SSx}}} \tag{9.15}$$

donde t tiene $n - 2$ grados de libertad.

Mediante la ecuación 9.17 se puede efectuar una prueba de la hipótesis nula

$$H_0 : R^2 = 0$$

donde R^2 es la contraparte del parámetro $\hat{R}^2$.

$$\boxed{F = \frac{\hat{R}^2}{\frac{1-\hat{R}^2}{n-2}}} \tag{9.17}$$

El estadístico F resultante se refiere a una distribución F con grados de libertad de numerador y denominador de 1 y $n - 2$, respectivamente.

En el caso de la RLS, estas dos pruebas siempre producirán el mismo resultado en términos de la decisión de rechazar o no rechazar. Esto es razonable, ya que si no existe relación lineal entre x y y, b será cero, al igual que la proporción de varianza explicada por x. En los modelos de RLM estas pruebas t y F generalmente no producirán el mismo resultado porque se enfocan en diferentes cuestiones y, por lo mismo, prueban diferentes hipótesis.

EJEMPLO 9.3

Utilice los resultados obtenidos con la información contenida en la tabla 9.1 para probar las hipótesis $H_0 : b = 0$ y $\hat{R}^2 = 0$. Construya e interprete un intervalo de confianza bilateral del 95% para el cálculo de β.

Solución Previamente encontramos que SS_{res} es 8.2898. Como el tamaño (n) de la muestra fue 15, los grados de libertad correspondientes son $15 - 2 = 13$, por lo tanto,

$$MS_{res} = \frac{SS_{res}}{n-2} = \frac{8.2898}{13} = .6377$$

---

[7] No debe confundirse β con el mismo símbolo cuando se usa para denotar la probabilidad de un error tipo II.

Previamente se calculó SS_x en 215.60. Usando estos resultados en la ecuación 9.14 junto con $b = .7059$, tenemos que

$$t = \frac{b}{\sqrt{\frac{MS_{res}}{SS_x}}} = \frac{.7059}{\sqrt{\frac{.6377}{215.60}}} = 12.980$$

El Apéndice B muestra que, para una prueba de dos colas realizada con $\alpha = .05$, t crítica es ± 2.160. Como el valor t obtenido de 12.980 es mayor que el valor crítico +2.160, la hipótesis nula se rechaza.

Por las ecuaciones 9.15 y 9.16, los límites inferior y superior para un intervalo de confianza del 95% son

$$L = b - t\sqrt{\frac{MSres}{SSx}} = .7059 - 2.160\sqrt{\frac{.6377}{215.60}} = .588$$

y

$$U = b + t\sqrt{\frac{MSres}{SSx}} = .7059 + 2.160\sqrt{\frac{.6377}{215.60}} = .823$$

Esto significa que podemos tener una confianza del 95% de que en la población, por cada cambio de una unidad en la puntuación de acceso, el cambio en la puntuación de salud prevista está entre .588 y .823. Observe que el resultado de la prueba de significancia está confirmado, ya que cero no se encuentra en este intervalo.

El coeficiente de determinación se puede probar mediante la ecuación 9.17, que nos da

$$F = \frac{\hat{R}^2}{\frac{1-\hat{R}^2}{n-2}} = \frac{.928}{\frac{1-.928}{13}} = 167.56$$

Sin el redondeo, éste es el cuadrado del valor t obtenido que se calculó antes. El Apéndice C muestra que, para 1 y 13 grados de libertad, F crítica para $\alpha = .05$ es 4.67. Como F obtenida de 167.56 excede este valor, se rechaza la hipótesis nula. Como resultado, podemos confiar en que el coeficiente de determinación de la población es mayor que cero. ■

9.2.6 Una inconsistencia lógica

Al principio de este capítulo indicamos que los modelos de regresión resultan útiles para predecir el valor de una variable (y) a partir de otra variable (x). También dijimos que estos modelos se emplean para evaluar la proporción de variación en una variable (y) explicada por alguna otra variable (x). Demostramos cómo se lleva a cabo esta segunda función mediante los coeficientes de determinación y no determinación. Pero, en este punto, existe una dificultad lógica relacionada con la función de predicción.

Debe tenerse disponibles a x y y para construir el modelo que se empleará en las predicciones. Pero si usted conoce y, ¿por qué no habría de "verlo" solamente, en vez de predecirlo? La respuesta es que los estudios diseñados para generar predicciones se dan en dos fases. En la primera, se recaban tanto x como y para cada sujeto y después se utilizan para calcular el modelo (es decir, a y b). Entonces, el modelo se utiliza para predicciones en situaciones en las que no hay ningún valor y disponible.

Por ejemplo, suponga que el administrador de un hospital está preocupado porque un porcentaje importante de las nuevas enfermeras dejan el empleo después de poco tiempo. El administrador quisiera realizar un estudio para determinar si es factible predecir la permanencia en el empleo al momento de la solicitud.

Con este fin, a todas las solicitantes para los puestos de enfermera se les aplica la prueba de Maslach (*Maslach Burnout Inventory* [35]) —descrita en la página 94—, ya que el administrador sospecha que el desgaste es, por lo menos, un factor del problema. Los expedientes se conservan de acuerdo con el tiempo que las enfermeras permanecen en el personal del hospital. Cuando se han recabado suficientes datos, se construye el modelo de la permanencia en el empleo (y) que se predice a partir de la puntuación de la escala de desgaste (x). Si el modelo parece un pronosticador adecuado, se puede aplicar la escala de desgaste a las futuras solicitantes para los puestos de enfermera y así predecir cuánto tiempo durarán en el empleo.[8]

9.3 REGRESIÓN LINEAL MÚLTIPLE

El modelo de regresión lineal múltiple es una de las herramientas estadísticas más poderosas y flexibles a disposición de los investigadores. Por esta razón, libros y cursos enteros están dedicados por completo a su estudio. La cobertura aquí tendrá que restringirse a algunos conceptos básicos que se relacionan con este modelo. De igual forma, los cálculos se restringirán al caso de dos predictores, ya que los modelos con más predictores son difíciles de calcular sin la ayuda de una computadora.

9.3.1 El modelo

El modelo de regresión lineal múltiple se puede expresar como

$$\boxed{\hat{y} = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p} \tag{9.18}$$

donde $\hat{y}$ es el valor previsto de y, a, b_1, $b_2 \ldots b_p$ son constantes y x_1, $x_2 \ldots x_p$ son las variables a partir de las cuales se harán las predicciones. El subíndice p se usa para indicar el número de predictores, el cual está limitado sólo por la cantidad de datos disponibles para la construcción del modelo.

En este apartado y los siguientes, le mostraremos cómo construir el modelo en el caso de dos predictores y cómo realizar *pruebas F parciales* y *de modelo completo* para modelos de cualquier tamaño. También comentaremos la utilidad del modelo de RLM.

9.3.2 Cálculo del modelo

Usaremos los datos de la tabla 9.2 de la página siguiente para demostrar los cálculos usados para elaborar un modelo de dos predictores.[9]

[8] Rara vez una sola variable serviría como un predictor adecuado para un fenómeno tan multifacético como el tiempo de permanencia en un empleo. Esto implica que para un estudio de este tipo podrían usarse los modelos de RLM.

[9] En general, los cálculos para los modelos de RLM se llevan a cabo a través de manipulaciones de matrices y no de los métodos aritméticos demostrados aquí. Sin embargo, tales cálculos son muy complejos y van mucho más allá del alcance de este capítulo.

TABLA 9.2: Datos usados para la construcción de un modelo de RLM de dos predictores.

Número de sujetos	y	x_1	x_2	y^2	x_1^2	x_2^2	yx_1	yx_2	$x_1 x_2$
1	9	2	5	81	4	25	18	45	10
2	15	8	9	225	64	81	120	135	72
3	4	1	3	16	1	9	4	12	3
4	8	5	5	64	25	25	40	40	25
5	11	1	7	121	1	49	11	77	7
6	7	7	5	49	49	25	49	35	35
7	23	10	10	529	100	100	230	230	100
8	4	9	10	16	81	100	36	40	90
9	5	2	1	25	4	1	10	5	2
10	13	7	5	169	49	25	91	65	35
11	15	9	7	225	81	49	135	105	63
12	22	9	10	484	81	100	198	220	90
13	11	5	1	121	25	1	55	11	5
14	4	8	3	16	64	9	32	12	24
15	17	8	10	289	64	100	136	170	80
Σ	168	91	91	2430	693	699	1165	1202	641

Comenzaremos por calcular algunos valores intermedios que serán usados en las ecuaciones que siguen.[10]

$$SS_y = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 2430 - \frac{(168)^2}{15} = 548.4000$$

$$SS_{x1} = \sum x_1^2 - \frac{\left(\sum x_1\right)^2}{n} = 693 - \frac{(91)^2}{15} = 140.9333$$

$$SS_{x2} = \sum x_2^2 - \frac{\left(\sum x_2\right)^2}{n} = 699 - \frac{(91)^2}{15} = 146.9333$$

$$SS_{yx1} = \sum yx_1 - \frac{\left(\sum y\right)\left(\sum x_1\right)}{n} = 1165 - \frac{(168)(91)}{15} = 145.8000$$

$$SS_{yx2} = \sum yx_2 - \frac{\left(\sum y\right)\left(\sum x_2\right)}{n} = 1202 - \frac{(168)(91)}{15} = 182.8000$$

$$SS_{x1x2} = \sum x_1x_2 - \frac{\left(\sum y\right)\left(\sum x_1\right)}{n} = 1165 - \frac{(168)(91)}{15} = 145.8000$$

$$SS_{x1x2} = \sum x_1x_2 - \frac{\left(\sum x_1\right)\left(\sum x_2\right)}{n} = 641 - \frac{(91)(91)}{15} = 88.9333$$

[10] Mejor póngase cómodo, esto podría ser tedioso.

Entonces, las ecuaciones para a, b_1 y b_2 son

$$\boxed{a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2} \tag{9.19}$$

$$\boxed{b_1 = \frac{(SS_{x2})(SS_{yx1}) - (SS_{x1x2})(SS_{yx2})}{(SS_{x1})(SS_{x2}) - (SS_{x1x2})^2}} \tag{9.20}$$

$$\boxed{b_2 = \frac{(SS_{x1})(SS_{yx2}) - (SS_{x1x2})(SS_{yx1})}{(SS_{x1})(SS_{x2}) - (SS_{x1x2})^2}} \tag{9.21}$$

Si sustituimos los resultados obtenidos anteriormente en las ecuaciones 9.20 y 9.21, tenemos que

$$b_1 = \frac{(146.9333)(145.8000) - (145.8000)(182.8000)}{(140.9888)(146.9333) - (88.9333)^2} = .4036$$

y

$$b_2 = \frac{(140.9333)(182.8000) - (88.9333)(145.8000)}{(140.9888)(146.9333) - (88.9333)^2} .9998$$

Si observamos que

$$\bar{y} = \frac{\sum y}{n} = \frac{168}{15} = 11.2000$$

$$\bar{x}_1 = \frac{\sum x_1}{n} = \frac{91}{15} = 6.0667$$

y

$$\bar{x}_2 = \frac{\sum x_2}{n} = \frac{91}{15} = 6.0667$$

por lo que el cálculo para a es, por la ecuación 9.19,

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 = 11.2000 - (.4036)(6.0667) - (.9998)(6.0667) = 2.6860$$

Así, el modelo de dos predictores es

$$\bar{y} = a + b_1\bar{x}_1 + b_2\bar{x}_2 = 2.6860 + .4036x_1 + .9998x_2$$

El valor de .4036 para b_1 significa que para un incremento de una unidad en x_1 sin cambios en x_2, $\hat{y}$ prevista se incrementará por .4036. El valor de .9998 para b_2 se interpreta de manera similar con x_1 como constante.

Podríamos usar este modelo para calcular todos los valores de $\hat{y}$, por lo que las ecuaciones 9.6 y 9.7 se pueden usar en conjunto con SS_y para encontrar $\hat{R}^2$ y $1 - \hat{R}^2$, pero SS_{reg} se puede calcular más directamente como

$$\boxed{SC_{reg} = b_1SC_{yx1} + b_2SC_{yx2} + \ldots + b_pSC_{yxp}} \tag{9.22}$$

donde p representa el número de predictores en el modelo. En este caso de dos predictores, SS_{reg} se calcula como

$$SS_{reg} = b_1SS_{yx1} + b_2SS_{yx2} = (.4036)(145.8000) + (.9998)(182.8000) = 241.6083$$

De la ecuación 9.9 de la página 324,

$$\hat{R}^2 = \frac{SS_{reg}}{SS_y} = \frac{241.6083}{548.4000} = .4406$$

Esto significa que aproximadamente el 44% de la variación en y se explica por la combinación de x_1 y x_2. Esto también implica que aproximadamente $1-\hat{R}^2 = 1 - .4406 = .5594$ de la variación en y no está explicada.

9.3.3 Pruebas de significancia para $\hat{R}^2$ y para las *b*

Se puede realizar una prueba de la hipótesis nula

$$H_0 : R^2_{y,1,\ldots,p} = 0$$

o de manera equivalente

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_p = 0$$

para modelos con cualquier cantidad de predictores, mediante la siguiente prueba F.

$$\boxed{F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}}} \tag{9.23}$$

En esta ecuación, p es el número de variables predictoras en el modelo y N es el número de observaciones, por ejemplo, los sujetos. El estadístico F obtenido tiene grados de libertad de numerador y denominador de p y $N - p - 1$, respectivamente.

Una prueba del modelo construido a partir de los datos de la tabla 9.2, de la página 330, se llevaría a cabo de la siguiente manera:

$$F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}} = \frac{\frac{.4406}{2}}{\frac{1-4406}{12}} = 4.73$$

En el Apéndice C se puede ver que para los grados de libertad del numerador y del denominador de 2 y 12, respectivamente, y $\alpha = .05$, F crítica es 3.89. Como la F obtenida de 4.73 es mayor que este valor, se rechaza la hipótesis nula. Esto significa que el investigador puede estar seguro de que el modelo sí explica algo de la variación en la variable y.

EJEMPLO 9.4

Un investigador está interesado en tratar de explicar la variación en las presiones arteriales que ocurren entre los hombres adultos. Con este fin, el investigador reúne datos sobre **1.** el ejercicio diario promedio, medido en horas (x_1), **2.** la edad (x_2) y **3.** el peso (x_3) para 74 sujetos. La variable resultante (y) es la presión arterial sistólica. El modelo (ficticio) producido por estos datos tiene un $\hat{R}^2$ de .2106 y es como sigue

$$\hat{y} = 33.5522 + .1710x_1 + .1033x_2 + .4471x_3$$

a) ¿Cuál sería la presión arterial sistólica prevista para un hombre de 21 años que se ejercita en promedio una hora diaria y pesa 165 libras? *b*) ¿Cuál sería el valor previsto para un hombre de la misma edad que también se ejercita en promedio una hora diaria pero pesa 185 libras? *c*) ¿Qué significa el valor de b_3 = .4471? *d*) Pruebe la significancia del modelo $\hat{R}^2$. *e*) Interprete los resultados de esta prueba.

Solución *a*) La presión sistólica prevista para un hombre de 21 años de edad que se ejercita en promedio una hora por día y pesa 165 libras sería

$$\hat{y} = 33.5522 + (.1710)(1) + (.1033)(21) + (.4471)(165) = 109.6640$$

b) El valor previsto para un hombre de la misma edad que también se ejercita en promedio una hora por día, pero pesa 20 libras más, sería

$$109.6640 + (20)(.4471) = 118.6060$$

c) El valor de b_3 = .4471 significa que, cuando otras variables permanecen sin cambios, por cada libra de incremento (o disminución) en el peso, la presión arterial sistólica prevista se incrementará (o disminuirá) en .4471. Ésta es la base del cálculo en *b*).
d) Se puede realizar una prueba de la hipótesis nula

$$H_0 : R^2 = 0$$

o de manera equivalente

$$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$$

mediante la ecuación 9.23

$$F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}} = \frac{\frac{.2106}{3}}{\frac{1-2106}{74-3-1}} = 6.22$$

El Apéndice C indica que para una prueba en el nivel α= .05 con grados de libertad de numerador y denominador de 3 y 70, respectivamente, la *F* crítica es 2.74. Como la *F* obtenida de 6.22 es mayor que este valor, se rechaza la hipótesis nula.
e) La prueba de significancia asegura al investigador que una parte de la variación en las presiones arteriales sistólicas se explica por la combinación de ejercicio, edad y peso. Un estimado de esta proporción está dado por $\hat{R}^2$ = .2106. ■

9.3.4 La prueba F parcial

Hasta este momento hemos manejado los modelos de RLM como estadísticos. Es decir, hemos hablado de uno, dos y hasta tres modelos predictores y hemos indicado que los modelos pueden tener cualquier número de predictores siempre y cuando el investigador tenga suficientes datos para sostener su construcción.[11] Pero, ¿cómo se toma la decisión de cuántos predictores debe haber en un modelo dado? La respuesta para esta pregunta puede ser muy compleja y depende, entre otras cosas, del propósito con el que se está usando el modelo y las hipótesis en las que se interesa el investigador.

Sin importar los detalles subyacentes en el asunto del número de predictores que deben incluirse en el modelo, la respuesta a menudo tiene que ver con la cuestión de si el hecho de agregar

[11] Es matemáticamente necesario que $p < N$. Desde un punto de vista práctico, N debe ser muchas veces mayor que p.

más variables a un modelo existente ampliará R^2. Por ejemplo, un investigador que está interesado en identificar variables que expliquen la variación en la presión arterial sistólica quizá quiera determinar si, al añadir el peso del sujeto al modelo que ya contiene su promedio de ejercicio diario y su edad, aumentará el modelo $\hat{R}^2$ considerablemente. Dicho de otra manera, ¿el peso de un paciente puede explicar una porción significativa de la variación en la presión arterial sistólica más allá de lo que la explican el ejercicio y la edad? Antes de examinar los métodos que se usan para responder tales preguntas, nos será útil presentar la siguiente notación.

Tenemos que $x_1, x_2, ..., x_p$ constituyen un conjunto de variables disponibles para un modelo de RLM. $\hat{R}^2_{y,1}$ representa el valor de $\hat{R}^2$ que se obtiene cuando sólo se usa x_1 como predictor. De igual forma, $\hat{R}^2_{y,2}$ representaría el valor de $\hat{R}^2$ que se obtiene cuando sólo se usa x_2 como predictor. Usamos la notación $\hat{R}^2_{y,1,2}$ para representar el valor de $\hat{R}^2$ que se obtiene cuando se usan x_1 y x_2 como predictores. La misma notación se usa para cualquier subconjunto de predictores. Por ejemplo, $\hat{R}^2_{y,1,3,4}$ representaría el valor de $\hat{R}^2$ obtenido cuando x_1, x_3 y x_4 son las variables utilizadas como predictores. También usamos la notación $\hat{R}^2_{y,G}$ y $\hat{R}^2_{y,p}$ para representar las $\hat{R}^2_p$ de dos modelos cuyas variables son indeterminadas pero en las cuales las variables que constituyen $\hat{R}^2_{y,p}$ son un subconjunto menor que las que constituyen $\hat{R}^2_{y.L}$.[12] Por ejemplo, $\hat{R}^2_{y.L}$ podría ser $\hat{R}^2_{y,1,3,4}$, mientras que $\hat{R}^2_{y,p}$ podría ser $\hat{R}^2_{y.1,4}$. Note que las variables en $\hat{R}^2_{y,p}$ son un subconjunto menor que las de $\hat{R}^2_{y,G}$.

EJEMPLO 9.5

Dado $\hat{R}^2_{y.G} = \hat{R}^2_{y.1,3,5,7,9}$, indique cuál de los coeficientes de determinación podría constituir $\hat{R}^2_{y,p}$. Especifique por qué cada una de las alternativas se incluye o se excluye como posibilidad.

a) $\hat{R}^2_{y.1,3,5,7,9}$

b) $\hat{R}^2_{y.1,7,9}$

c) $\hat{R}^2_{y.1,4,7,9}$

d) $\hat{R}^2_{y.1,3,5,7}$

Solución

a) se excluye porque no es un subconjunto *menor* de $\hat{R}^2_{y,G}$, es decir, tiene el mismo número de variables que $\hat{R}^2_{y,G}$.

b) se incluye porque tiene menos variables que $\hat{R}^2_{y,G}$ y todas sus variables están en $\hat{R}^2_{y,G}$.

c) se excluye porque tiene una variable (x_4) que no está en $\hat{R}^2_{y,G}$.

---

[12] L representa el modelo más grande (por *larger*) de los dos y S el más pequeño (por *smaller*).

d) se incluye porque tiene menos variables que $\hat{R}^2_{y.L}$ y todas sus variables están en $\hat{R}^2_{y,G}$. ■

Surge una pregunta importante sobre si (o en qué condiciones),

$$\hat{R}^2_{y.1} + \hat{R}^2_{y.2} + \ldots + \hat{R}^2_{y.p} = \hat{R}^2_{y.1,2,\ldots p}$$

La respuesta es que esta igualdad será cierta sólo cuando todas las variables x no estén correlacionadas. Este resultado tiene muchas implicaciones para la interpretación de los resultados obtenidos de los modelos de RLM.

Se puede realizar una prueba de la hipótesis nula

$$H_0 : R^2_{y.L} - R^2_{y.S} = 0$$

mediante la siguiente prueba F:

$$\boxed{F = \frac{\dfrac{\hat{R}^2_{y.L} - \hat{R}^2_{y.S}}{pL - ps}}{\dfrac{1 - \hat{R}^2_{y.L}}{N - pL - 1}}} \tag{9.24}$$

En esta ecuación p_L y p_S se refieren al número de variables en los modelos, el más grande y el más pequeño, respectivamente, y N es el número total de observaciones (por ejemplo, los sujetos) en el análisis. El estadístico F resultante tiene grados de libertad de numerador y denominador de $p_L - p_S$ y $N - p_L - 1$, respectivamente. Las pruebas F que comparan modelos de esta forma se denominan **pruebas *F* parciales**.

EJEMPLO 9.6

En la página 332 usamos los datos de la tabla 9.2 para calcular $\hat{R}^2_{y.12} = .4406$. Además, hemos calculado[13] $\hat{R}^2_{y.1} = .2750$ y $\hat{R}^2_{y.2} = .4147$. Utilice estos resultados para responder las siguientes preguntas.

a) ¿x_1 por sí sola explica una proporción significativa de la variación en y?

b) ¿x_2 por sí sola explica una proporción significativa de la variación en y?

c) ¿x_1 y x_2 explican una proporción significativa de la variación en y cuando se usan juntas en un modelo?

d) ¿Un modelo que usa x_1 y x_2 como predictores explica una proporción significativamente mayor de la variación en y que un modelo que sólo usa una de las dos variables?

---

[13] El cálculo de estos resultados no se muestra aquí, pero siéntase con la libertad de usar los métodos del apartado 9.2.3 para verificar su validez.

Solución Las respuestas a las preguntas *a*), *b*) y *c*) se pueden obtener mediante la aplicación de la ecuación 9.23 de la página 332. Para la pregunta *a*) queremos probar la hipótesis

$$H_0 : \hat{R}^2_{y.1} = 0$$

La aplicación de la ecuación 9.23 da por resultado

$$F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}} = \frac{\frac{.2750}{1}}{\frac{1-.2750}{15-1-1}} = 4.93$$

La referencia al Apéndice C muestra que para $\alpha = .05$ y los grados de libertad del numerador y del denominador de $p = 1$ y $N - p - 1 = 15 - 1 - 1 = 13$, respectivamente, F crítica es 4.67. Como la F obtenida de 4.93 es mayor que este valor, la hipótesis nula se rechaza. Concluimos, por consiguiente, que x_1 por sí sola explica una proporción significativa de la variación en y.

Para la pregunta *b*) queremos probar la hipótesis

$$H_0 : R^2_{y.2} = 0$$

Al aplicar la ecuación 9.23 se obtiene

$$F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}} = \frac{\frac{.4147}{1}}{\frac{1-.4147}{15-1-1}} = 9.21$$

Nuevamente, la referencia al Apéndice C muestra que para $\alpha = .05$ y los grados de libertad del numerador y del denominador de $p = 1$ y $N - p - 1 = 15 - 1 - 1 = 13$, respectivamente, F crítica es 4.67. Como la F obtenida de 9.21 es mayor que este valor, la hipótesis nula se rechaza. Concluimos, por consiguiente, que x_2 por sí sola explica una porción significativa de la variación en y.

Para la pregunta *c*) queremos probar la hipótesis

$$H_0 : R^2_{y.12} = 0$$

Al aplicar la ecuación 9.23 se obtiene

$$F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}} = \frac{\frac{.4406}{1}}{\frac{1-.4406}{15-2-1}} = 4.73$$

La referencia al Apéndice C muestra que para $\alpha = .05$ y los grados de libertad del numerador y del denominador de $p = 2$ y $N - p - 1 = 15 - 2 - 1 = 12$, respectivamente, la F crítica es 3.89. Como la F obtenida de 4.73 es mayor que este valor, la hipótesis nula se rechaza. Concluimos, por consiguiente, que x_1 y x_2 usadas juntas en el mismo modelo explican una proporción significativa de la variación en y.

Para la pregunta *d*) usaremos la ecuación 9.24 de la página anterior para probar las hipótesis

$$H_0 : R^2_{y.12} - R^2_{y.1} = 0$$

y

$$H_0 : R^2_{y.12} - R^2_{y.2} = 0$$

Note que la primera hipótesis trata la pregunta: "¿Un modelo que usa x_1 y x_2 explica una proporción significativamente mayor de la variación en *y* que un modelo que solamente usa x_1 ? La segunda hipótesis plantea la cuestión: "¿Un modelo que usa x_1 y x_2 explica una proporción significativamente mayor de la variación en *y* que un modelo que solamente usa x_2?

Para la prueba de la primera hipótesis calculamos

$$F = \frac{\dfrac{\hat{R}^2_{y.L} - \hat{R}^2_{y.S}}{p_L - p_S}}{\dfrac{1-\hat{R}^2_{y.L}}{N - p_L - 1}} = \frac{\dfrac{.4406 - 2750}{2-1}}{\dfrac{1-.4406}{15-2-1}} = 3.55$$

Usando los grados de libertad de numerador y denominador de 1 y 12, respectivamente, la *F* crítica para una prueba en el nivel .05, de acuerdo con el Apéndice C, es 4.75. Como la *F* obtenida de 3.55 es menor que este valor, no rechazamos la hipótesis nula. Por consiguiente, no podemos demostrar que el modelo de dos predictores explica una mayor parte de la variación en *y* que el modelo que sólo contiene x_1.

Para la prueba de la segunda hipótesis calculamos

$$F = \frac{\dfrac{\hat{R}^2_{y.L} - \hat{R}^2_{y.S}}{p_L - p_S}}{\dfrac{1-\hat{R}^2_{y.L}}{N - p_L - 1}} = \frac{\dfrac{.4406 - 4147}{2-1}}{\dfrac{1-.4406}{15-2-1}} = .56$$

Usando los grados de libertad de numerador y denominador de 1 y 12, respectivamente, la *F* crítica para una prueba en el nivel .05, de acuerdo con el Apéndice C, es 4.75. Como la *F* obtenida de .56 es menor que este valor, no rechazamos la hipótesis nula. Por consiguiente, no podemos demostrar que el modelo de dos predictores explica una mayor parte de la variación en *y* que el modelo que sólo contiene x_2.

Mientras que no pudimos demostrar que el modelo de dos predictores es un mejor predictor de *y* que cualquiera de los modelos de un predictor, probablemente escogeríamos x_2 si decidiéramos utilizar sólo un predictor. De hecho, la diferencia entre $\hat{R}^2_{y.12}$ y $\hat{R}^2_{y.2}$ parece pequeña. No obstante, debe considerarse que la selección de las variables para el modelo no está solamente en función de pruebas estadísticas. Se deben considerar otros factores, como el costo, la dificultad para adquirir datos y otras cuestiones prácticas. ■

EJEMPLO 9.7

En el ejemplo 9.4 de la página 332 presentamos un caso en el que un investigador está interesado en tratar de explicar la variación en las presiones arteriales entre los hombres adultos. Con este fin, el investigador reúne datos sobre **1.** el ejercicio diario promedio, medido en horas (x_1), **2.** la edad (x_2) y **3.** el peso (x_3) de 74 sujetos. La variable resultante (y) es la presión arterial sistólica. El modelo (ficticio) generado por estos datos tiene un $\hat{R}^2$ asociado de .2106. ¿El ejercicio (x_1) aumenta $\hat{R}^2$ considerablemente cuando esta variable se agrega a un modelo que ya contiene la edad y el peso, si $\hat{R}^2_{y.23} = .1905$?

Solución Deseamos probar la hipótesis

$$H_0 : R^2_{y.123} - R^2_{y.23} = 0$$

Observe que esto compara un modelo que considera la edad, el peso y el ejercicio con un modelo que incluye solamente la edad y el peso. La prueba F parcial para probar esta hipótesis es

$$F = \frac{\dfrac{\hat{R}^2_{y.L} - \hat{R}^2_{y.S}}{p_L - p_S}}{\dfrac{1-\hat{R}^2_{y.L}}{N - p_L - 1}} = \frac{\dfrac{.2106-1905}{3-2}}{\dfrac{1-.2106}{74-3-1}} = 1.78$$

El valor crítico para un estadístico F con $3 - 2 = 1$ y $74 - 3 - 1 = 70$ grados de libertad es 3.98, por lo que la hipótesis nula no se rechaza. Esto significa que no podemos demostrar que si agregamos la variable del ejercicio a un modelo que considera la edad y el peso, aumentará la proporción de variación en la presión arterial que explica el modelo. ■

9.4 SUPOSICIONES

Las suposiciones subyacentes a la regresión lineal simple y la regresión lineal múltiple se pueden resumir como **1.** normalidad, **2.** homogeneidad de varianza y **3.** independencia. Consideraremos cada una de éstas en el contexto de la RLS y luego las estudiaremos para la RLM.

En la página 321 utilizamos datos de la tabla 8.2 de la página 297 para construir un modelo de RLS para la predicción de puntuaciones de salud a partir de las puntuaciones de acceso a servicios médicos. Después llevamos a cabo pruebas estadísticas relacionadas con este modelo. Para comprender las suposiciones respecto de estas pruebas, considere una serie de poblaciones que consisten en puntuaciones de salud. La primera población consta de puntuaciones de salud de personas con la misma puntuación de acceso, digamos de 1. La segunda población consta de puntuaciones de salud de personas con puntuaciones de acceso de 2, la tercera consta de puntuaciones de salud de personas con puntuaciones de acceso de 3, y así sucesivamente. La normalidad y homogeneidad de las suposiciones de varianza se relacionan con estas poblaciones. Más específicamente, se presupone que todas estas poblaciones están distribuidas normalmente y tienen la misma varianza.

Si consideramos un modelo de RLM de dos predictores, entonces consideramos una serie de poblaciones en las que cada sujeto tiene la misma puntuación x_1 y la misma puntuación x_2. Esto no significa que x_1 y x_2 tengan el mismo valor, más bien que todos los sujetos en una población en particular tienen el mismo valor de x_1 (por ejemplo, 5) y el mismo valor de x_2 (por ejemplo, 20). Este concepto es el mismo para cualquier número de predictores. Como antes, se supone que todas las poblaciones están distribuidas normalmente con la misma varianza. Observe que los valores x, por sí solos, no necesitan cumplir con estas suposiciones.

La suposición de independencia significa que la respuesta (y) de cada sujeto no está relacionada con la respuesta de cualquier otro sujeto. Aunque no necesariamente es cierto en un conjunto específico de condiciones, los modelos de RLS y RLM tienden a ser los más robustos ante las violaciones de la suposición de normalidad y los menos robustos ante las violaciones de la suposición de independencia.

9.5 ALGUNOS COMENTARIOS ADICIONALES RESPECTO DE LA UTILIDAD DE LA RLM

Antes en este capítulo planteamos que el "...modelo de regresión lineal múltiple es una de las herramientas estadísticas más poderosas y flexibles a disposición de los investigadores". Usted ya lo ha comprobado al ver su habilidad para predecir y explicar la variación. Pero esto difícilmente transmite el panorama completo.

El modelo de RLM es una expresión de lo que se conoce como el modelo lineal general. Esto significa que con una acertada selección de variables codificadas[14] independientes y dependientes, este modelo puede usarse para realizar todas las pruebas estadísticas siguientes.

1. Prueba *t* de muestras independientes.
2. ANOVA de una vía.
3. ANOVA factorial.
4. Análisis de covarianza.
5. Análisis discriminante (de dos grupos).

Además, diversas pruebas para las interacciones pueden llevarse a cabo mediante este modelo. También se pueden detectar pruebas para relaciones curvilíneas como la descrita en la figura 8.6 de la página 308 y otras formas.[15]

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

$\hat{R}^2$, 323-324
coeficiente de determinación, 324
coeficiente de no determinación, 323
línea de regresión de mínimos cuadrados, 326
modelo de los mínimos cuadrados, 323
prueba *F* de modelo completo, 329
prueba *F* parcial, 333
regresión lineal múltiple, 329
regresión lineal simple, 320
residual, 320
residual cuadrático medio, 327
suma de los cuadrados de regresión, 322
suma de los cuadrados residuales, 322

[14] Un esquema simple de identificación podría asignar un 1 para los sujetos femeninos y un 0 para los sujetos masculinos.

[15] No es necesario decir que debe tomar un curso de RLM tan pronto como termine éste.

EJERCICIOS

9.1 Responda lo siguiente.

a) Escriba dos expresiones diferentes para el coeficiente de determinación (R^2).

b) Escriba dos expresiones diferentes para el coeficiente de no determinación ($1–R^2$).

c) Para el modelo de RLS, ¿qué valores toman SS_{res} y SS_{reg} cuando $r_{yx} = 1.0$?

d) Para el modelo de RLS, ¿qué valores toman SS_{res} y SS_{reg} cuando $r_{yx} = 0.0$?

9.2 Indique cuál de los siguientes enunciados son ciertos y cuáles falsos.

a) En general, $\hat{R}^2_{y.1,2,3} = \hat{R}^2_{y.1} + \hat{R}^2_{y.2} + \hat{R}^2_{y.3}$.

b) Una prueba de hipótesis
$H_0 : R^2_{y.1,2,3}$
equivale a una prueba de hipótesis
$H_0 : \beta_1 = \beta_2 = \beta_3 = 1$.

c) Un buen modelo de predicción produce residuales grandes (en términos relativos).

d) Los valores grandes (en términos relativos) de $\hat{y} - \bar{y}$ por lo general se asocian con los buenos modelos de predicción.

9.3 ¿Cuál de las siguientes hipótesis se debe probar para poder realizar una prueba que determine si agregar las variables x_1 y x_3 a un modelo que contiene las variables x_2 y x_4 incrementa R^2?

a) $R^2_{y.1,2,3,4} - R^2_{y.2,4} = 0$

b) $R^2_{y.1,2,3,4} - R^2_{y.1,3} = 0$

c) $R^2_{y.1,3} = 0$

d) $R^2_{y.2,4} = 0$

9.4 Un investigador interesado en la relación entre el índice de masa corporal (IMC) y el colesterol sérico desea ajustar un modelo de RLS en el que el colesterol sérico total pueda predecirse a partir del IMC con los siguientes datos. Utilice estos datos para responder cada uno de los siguientes planteamientos.

Colesterol total	IMC
165	25.9
155	20.1
141	22.2
228	30.7
190	28.0
155	29.4
132	20.2
170	20.7
188	26.3
150	18.2

a) Construya un modelo de RLS.

b) Encuentre los residuales para los primeros tres sujetos.

c) Encuentre los coeficientes de determinación y no determinación.

d) Pruebe la hipótesis $H_0 : R^2 = 0$.

e) Pruebe la hipótesis $H_0 : \beta = 0$.

f) Construya un intervalo de confianza bilateral del 95% para el cálculo de β. ¿Este intervalo de confianza concuerda con el resultado de sus pruebas de hipótesis? Explique.

9.5 Conteste las siguientes preguntas realizando la prueba de significancia indicada con base en la siguiente información. Considere $N = 40$ en todos los casos.

1. $\hat{R}^2_{y.1,2,3,4} = .68$
2. $\hat{R}^2_{y.2,3,4} = .43$
3. $\hat{R}^2_{y.4} = .03$
4. $\hat{R}^2_{y.1,3} = .27$
5. $\hat{R}^2_{y.2,4} = .14$
6. $\hat{R}^2_{y.1} = .10$

a) ¿Las cuatro variables usadas en conjunto explican cualquier variación de *Y*?

b) ¿Agregar x_1 al modelo que ya contiene x_2, x_3 y x_4 explica cualquier variación adicional en *Y*?

c) ¿Cuando se usan x_2 y x_4 juntas, éstas explican cualquier variación en *Y*?

d) ¿Agregar las variables x_2 y x_4 a un modelo que ya contiene las variables x_1 y x_3 incrementa la proporción de variación explicada en *Y*?

D. Las siguientes preguntas se refieren al estudio de caso D (página 471).

9.6 A partir de su estudio, los investigadores concluyen lo siguiente: "Estas correlaciones sugieren que la cuantificación de PBV puede ofrecer un predictor objetivo y sustituto fácilmente adquirido de discapacidad neuropsicológica y disfunción cognoscitiva/motriz clínicamente evidente entre las personas infectadas con VIH".

a) Construya un modelo de RLS para predecir puntuaciones NPZ-8 a partir de PBV. Utilice sólo los datos de los 15 sujetos infectados para el cálculo.

b) Interprete el término *b* en el modelo en términos de NPZ-8 y PBV.

c) Encuentre los coeficientes de determinación y no determinación asociados con este modelo.

d) Utilice el modelo para predecir las puntuaciones NPZ-8 de los primeros cinco sujetos.

e) Considerando su observación del coeficiente de determinación y los valores previstos para los primeros cinco sujetos, ¿estaría de acuerdo con la conclusión de los autores?

9.7 Construya un modelo de RLM para predecir las puntuaciones NPZ-8 a partir de los valores PBV y CD4. Utilice sólo los datos de los 15 sujetos infectados.

a) Encuentre el coeficiente de determinación para el modelo de dos predictores.

b) ¿Agregar CD4 al modelo incrementa R^2 al nivel .05 considerablemente?

c) Utilice el modelo de dos predictores para predecir las puntuaciones NPZ-8 a partir de los valores PBV y CD4 para los primeros cinco sujetos. ¿El modelo de dos predictores parece predecir mejor que el modelo de un predictor?

d) ¿PBV y CD4 juntos explican una parte significativamente importante de la variación en las puntuaciones NPZ-8?

M. Las siguientes preguntas se refieren al estudio de caso M (página 475).

9.8 Construya un modelo de RLS para predecir las puntuaciones "después" del dióxido de carbono a partir de los valores "después" del oxígeno.

a) Interprete *b*.

b) Estime β con un intervalo de confianza del 95%. Interprete este intervalo.

c) ¿Cree usted que una prueba de dos colas de la hipótesis nula $R^2 = 0$ realizada con $\alpha = .05$ sería significativa? Dé las razones para su respuesta.

d) Calcule R^2 para el modelo.

e) ¿Cree usted que la correlación entre las dos variables es positiva o negativa? ¿Por qué?

CAPÍTULO 10

Métodos basados en el principio de permutación

10.1 INTRODUCCIÓN

La mayoría de los textos de introducción a la estadística cuentan con un capítulo titulado “Métodos no paramétricos”[1] o “Pruebas de distribución libre”. Por lo general, este capítulo presenta de manera breve algunas reconocidas pruebas no paramétricas basadas en rangos[2] cuyas distribuciones muestrales se obtienen a través del método de permutación. Estas pruebas son casos especiales de una clase más extensa de procedimientos basados en la permutación que emplean observaciones originales en vez de rangos. Con muy pocas excepciones no se discute esta clase más amplia y sólo se revisan los casos especiales. Además, no se explica el principio de permutación subyacente en ambos tipos de pruebas.

Creemos que este enfoque deja al alumno una visión demasiado limitada del tema. Como resultado, en este texto se abandona el enfoque más tradicionalista (como lo hemos hecho en ocasiones anteriores) y nos enfocamos en este tipo más amplio de pruebas con los métodos más conocidos basados en rangos presentados como casos especiales. Hemos escogido este camino por tres razones. Primero, con la llegada de los algoritmos rápidos y las poderosas computadoras, se está utilizando más el amplio tipo de pruebas en la investigación aplicada. Mientras que estos métodos se han comprendido bien por muchos años, la falta de plataformas computacionales poderosas ha dificultado su aplicación. Segundo, los estudiantes que no cuentan con antecedentes matemáticos sólidos pueden comprender mejor estas pruebas de lo que lograrían de otras formas; y, tercero, la flexibilidad de estas pruebas añade una poderosa dimensión a las numerosas herramientas estadísticas del investigador.

La comprensión de los métodos basados en permutaciones depende de la comprensión de los conceptos matemáticos de permutaciones y combinaciones. Por esta razón, comenzaremos con estos conceptos. Luego, se discutirá una serie de métodos basados en permutaciones y se demostrará cómo se obtienen los casos especiales equivalentes basados en rangos. En una sección final, discutiremos algunas de las características de estos métodos y las compararemos con las pruebas con las que usted ya está familiarizado.

---

[1] Algunos autores prefieren unir el término no paramétrico con un guión, otros no.
[2] Éste y otros términos se definirán más adelante.

10.2 ALGUNOS PRELIMINARES

Las distribuciones muestrales de muchas pruebas de permutación se basan en permutaciones y combinaciones de enteros y otros números. Quizá usted ha estudiado estos métodos en el pasado, pero los revisaremos aquí. Los utilizaremos en las aplicaciones siguientes.

10.2.1 Permutaciones

Considere un orden arbitrario de izquierda a derecha de los enteros 1, 2 y 3. Un orden como éste sería 3, 2, 1, mientras que otro sería 1, 3, 2. Cada una de estas disposiciones de los números se denomina **permutación** de los enteros 1, 2 y 3. Nos interesa determinar cuántas de estas disposiciones se pueden obtener de estos números. Con tan pocos números, podríamos listar cada orden y sólo contar el número que se obtenga de esta forma. A continuación se lleva a cabo para las siguientes permutaciones.

1	1	2	3
2	1	3	2
3	2	1	3
4	2	3	1
5	3	1	2
6	3	2	1

Como no hay otras disposiciones de los datos sin repetir alguno de ellos, podemos estar seguros de que hemos listado todas las posibles permutaciones de los enteros 1, 2 y 3. Incluso, haremos notar que existen seis arreglos posibles.

Una inspección más detallada de las seis disposiciones nos revela que teníamos tres opciones para llenar la primera posición de cada permutación. Esto es, podríamos escoger los números 1, 2 o 3 para la primera posición. Una vez que llenamos la primera posición nos quedaban solamente dos opciones para la segunda posición. Así, en las permutaciones 1 y 2 donde llenamos la primera posición con el entero 1, podíamos escoger los enteros 2 o 3 para la segunda posición. Una vez que se llenaron las dos primeras posiciones, quedaba sólo un entero para la tercera posición. Entonces, había tres formas de llenar la primera posición y, para cada una de ellas, dos formas de llenar la segunda posición; luego, para cada una de éstas, solamente había una forma de llenar la tercera posición; el total de permutaciones fue $3 \cdot 2 \cdot 1 = 6$.

Como segundo ejemplo, consideremos el número de permutaciones que se pueden realizar con los números 1, 2, 3 y 4. Listaremos todas las disposiciones en la siguiente página. Lo primero que notamos es que la adición de un dígito ha aumentado considerablemente el número de permutaciones, de seis a 24. También vemos que ahora tenemos cuatro opciones para la primera posición. Luego de llenar la primera posición, nos quedan tres opciones para la segunda, dos para la tercera y un solo valor para la cuarta. Esto significa que tenemos $4 \cdot 3 \cdot 2 \cdot 1 = 24$ permutaciones de los cuatro dígitos.

TABLA 10.1: Todas las posibles permutaciones de los enteros 1, 2, 3 y 4.

1	1	2	3	4	13	3	1	2	4
2	1	2	4	3	14	3	1	4	2
3	1	3	2	4	15	3	2	1	4
4	1	3	4	2	16	3	2	4	1
5	1	4	2	3	17	3	4	1	2
6	1	4	3	2	18	3	4	2	1
7	2	1	3	4	19	4	1	2	3
8	2	1	4	3	20	4	1	3	2
9	2	3	1	4	21	4	2	1	3
10	2	3	4	1	22	4	2	3	1
11	2	4	1	3	23	4	3	1	2
12	2	4	3	1	24	4	3	2	1

Esto nos lleva a una conclusión general. Si tenemos n objetos (números, perros, amigos, etc.) podemos ordenarlos en $n!$ diferentes permutaciones. La notación $n!$ se lee "factorial n" y significa $n(n-1)(n-2)\cdots(1)$. Si P_n representa el número de formas en que se pueden permutar n objetos, entonces podemos escribir lo siguiente.

$$\boxed{P_n = n!} \tag{10.1}$$

EJEMPLO 10.1

¿Cuántas permutaciones distintas de los números 3.1, 0.0, 13.0, 99.7 y .6 son posibles?

Solución Como hay cinco números, la ecuación 10.1 indica que el total de permutaciones distintas es $n! = 5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. ■

10.2.2 Combinaciones

Ahora considere el problema de determinar el número de formas en las que podemos dividir n objetos, números, etcétera, en dos grupos. Para empezar, supongamos que deseamos dividir los números del 1 al 4 en dos grupos de dos números cada uno. Se debe ignorar el orden en los dos grupos. Una posibilidad es colocar 1 y 2 en el primer grupo, y 3 y 4 en el segundo. Esta separación se llama **combinación**. Observe que al colocar los números 2 y 1 en el grupo uno y 4 y 3 en el segundo, no se forma una nueva combinación ya que los dos grupos contienen los mismos números que antes, simplemente los hemos listado en un orden diferente. Una segunda combinación se podría lograr colocando 1 y 3 en el primer grupo, y 2 y 4 en el segundo. Pero, ¿de cuántas formas se puede lograr esta tarea? Como hicimos antes con las permutaciones, podríamos escribir todas las combinaciones posibles y simplemente contar el número obtenido. Como se puede ver en este listado, hay seis posibles combinaciones.

	Grupo 1		Grupo 2	
1	1	2	3	4
2	1	3	2	4
3	1	4	2	3
4	2	3	1	4
5	2	4	1	3
6	3	4	1	2

Como hicimos con las permutaciones, sería útil si pudiéramos encontrar una expresión matemática que nos diera este número directamente, en vez de tener que listar todas las posibilidades. Antes de enfrentar directamente este problema, convendría reexaminar las permutaciones de la página anterior.

EJEMPLO 10.2

Ya que las 4! *permutaciones* de la página anterior ordenan los números en todas las configuraciones posibles, debe ser cierto que todas las posibles *combinaciones* están representadas ahí. ¿Cuántas *combinaciones* puede usted encontrar en las 4! = 24 *permutaciones* listadas en la página anterior? Suponga que las primeras dos posiciones para cada permutación representan el grupo 1 y las dos últimas el grupo 2. Liste las permutaciones que representan cada combinación.

Solución A continuación se presenta cada combinación con sus respectivas permutaciones.

Combinación	Permutaciones
uno	1, 2, 7, 8
dos	3, 4, 13, 14
tres	5, 6, 19, 20
cuatro	9, 10, 15, 16
cinco	11, 12, 21, 22
seis	17, 18, 23, 24

Podemos ver en esta lista que hay cuatro permutaciones en cada combinación. No es de sorprender, porque para cada combinación hay dos números en el primer grupo que se pueden permutar de dos formas, y para cada una de estas permutaciones los dos números en el segundo grupo se pueden permutar de dos formas. Por ejemplo, en la primera combinación, la primera permutación de los números del grupo uno es 1 y 2. Para esta permutación del primer grupo hay dos permutaciones en el segundo grupo: 3, 4 y 4, 3. Para la segunda permutación en el primer grupo, 2, 1, también hay dos permutaciones en el segundo grupo. Como hay dos formas de permutar los números del primer grupo y dos formas de permutar los números del segundo grupo, el número total de permutaciones para cada combinación es $2 \cdot 2 = 4$.

En general, supongamos que n_1 representa el número de objetos del primer grupo y n_2 el número de objetos en el segundo, entonces, el número de permutaciones en el primer grupo para una combinación dada será $n_1!$ y el número en el segundo grupo será $n_2!$, por lo que cada combinación contendrá en sí misma $n_1!n_2!$ permutaciones.

EJEMPLO 10.3

Si se deben dividir siete objetos en dos grupos con cuatro en el primero y tres en el segundo, ¿cuántas permutaciones habrá en cada una de las combinaciones posibles?

Solución Como $n_1 = 4$ y $n_2 = 3$, habrá $4!3! = 24 \cdot 6 = 144$ permutaciones en cada combinación.

Si C_{n2}^{n1} representa el número de combinaciones que se pueden lograr colocando n_1 objetos en el primer grupo y n_2 en el segundo, entonces podemos hacer el siguiente razonamiento. Para cualquier conjunto de $n!$ permutaciones habrá C_{n2}^{n1} combinaciones, cada una de las cuales contendrá $n_1!n_2!$ permutaciones, por lo que $n! = C_{n2}^{n1}$ donde $n = n_1 + n_2$. Si resolvemos C_{n2}^{n1} tendremos el número de formas en las que podemos dividir n objetos en dos grupos.

$$\boxed{C_{n2}^{n1} = \frac{n!}{n_1!n_2!}} \tag{10.2}$$

■

EJEMPLO 10.4

¿En cuántas formas se pueden dividir siete personas en dos grupos, si se deben colocar cuatro personas en el primer grupo y tres en el segundo?

Solución La ecuación 10.2 da

$$C_3^4 = \frac{n!}{n_1!n_2!} = \frac{7!}{4!3!}$$

Nos podemos ahorrar algo del trabajo de cálculo si observamos que $7! = 7 \cdot 6 \cdot 5 \cdot 4!$, por lo que

$$C_3^4 = \frac{7 \cdot 6 \cdot 5 \cdot \cancel{4!}}{\cancel{4!}3!} = \frac{210}{6} = 35$$

■

EJEMPLO 10.5

¿En cuántas formas se pueden dividir los enteros del 1 al 8 en dos grupos, si debe haber cuatro enteros en cada uno?

Solución Mediante la ecuación 10.2

$$C_4^4 = \frac{n!}{n_1!n_2!} = \frac{8!}{4!4!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot \cancel{4!}}{4!\cancel{4!}} = \frac{1680}{24} = 70$$ ■

10.3 APLICACIONES

Ahora consideraremos varias pruebas cuyas distribuciones muestrales se basan en el principio de permutación.[3] Estas pruebas por lo general se caracterizan por ser "de libre distribución" o "no paramétricas". *No paramétrico* se refiere al hecho de que las distribuciones muestrales de estas pruebas no dependen de la especificación de parámetros de población.[4] Por ejemplo, la distribución muestral de la prueba Z de una media requiere que se especifique σ (así como otros parámetros). De igual manera, la distribución muestral de la prueba *t* de muestras independientes requiere que las poblaciones estén distribuidas normalmente, lo que a su vez implica valores de ciertos parámetros como el sesgo y la curtosis.[5] Como resultado de este requerimiento, estas pruebas se caracterizan por ser **paramétricas**. En contraste, la prueba de Wilcoxon para muestras independientes, que veremos pronto, a menudo se caracteriza como la contraparte no paramétrica de la prueba *t* de muestras independientes. La prueba de Wilcoxon no requiere que se especifique ninguno de los parámetros de población, por lo que el sesgo de población puede ser cero o cualquier otro valor (limitado). Esta prueba se caracteriza, por ende, como **no paramétrica**.

Además, como no es necesario especificar los parámetros para obtener las distribuciones muestrales para estas pruebas, tampoco es necesario especificar la forma (por ejemplo, normal) de la población muestreada. Por esa razón, tales pruebas se conocen como pruebas de **distribución libre**.[6] Entonces, mientras que la normalidad de la población es uno de los supuestos de la prueba *t* de muestras independientes, por lo que no es ni paramétrica ni de distribución libre, no sucede lo mismo con la prueba de la suma de rangos de Wilcoxon, por lo que esta última es tanto no paramétrica como de distribución libre. Si bien existe una diferencia entre los términos no paramétrico y de distribución libre, la mayoría de las pruebas que poseen una de las dos características también posee la otra, por lo que a menudo los términos se usan de forma indistinta.[7]

También deseamos resaltar que mientras este capítulo se limita a las pruebas de hipótesis, los métodos aquí descritos también se pueden utilizar para construir intervalos de confianza. Sin embargo, construir intervalos de confianza mediante el método de permutación puede ser aún más difícil que la prueba de hipótesis; así que, para este tema será mejor ayudarse con una computadora. Asimismo, es el problema de la complejidad de cálculo lo que hace necesario el uso de pequeños conjuntos de datos muestrales usados para ilustrar cada una de las siguientes pruebas. En particular, ésta no es una dificultad, ya que nuestro interés primordial es mostrarle los principios subyacentes en estas pruebas.

---

[3] Como ahora sabe, las combinaciones son subconjuntos de permutaciones, por lo que las pruebas basadas en combinaciones también pueden ser consideradas como pruebas de permutación.

[4] Hay algunas excepciones. Por ejemplo, la prueba de Wilcoxon para muestras independientes supone que el muestreo se hace a partir de una población simétrica, lo que a su vez implica un parámetro sesgado de cero.

[5] El sesgo para la distribución normal es cero y la curtosis es tres.

[6] De hecho, aunque está fuera del alcance de este libro, estas pruebas pueden justificarse sin el concepto de una población.

[7] En ocasiones, los especialistas en estadística protestan por este asunto, pero carece de importancia para nuestros propósitos.

Ahora, enfocaremos nuestra atención en varias pruebas de permutación. En adelante se presentarán dos pruebas para cada concepto estadístico tratado (por ejemplo, la correlación). El primero utilizará observaciones originales, en tanto que el segundo empleará rangos.

10.3.1 Correlación

En la página 309 usted aprendió a utilizar la ecuación 8.4 para comprobar la hipótesis nula $H_0 : \rho = 0$, en donde ρ es el coeficiente de correlación del producto-momento de Pearson. Uno de los supuestos subyacentes de esa prueba es que la muestra provino de una población bivariable normal. En esta sección usted aprenderá a probar hipótesis relacionadas con la correlación de la población sin tener que hacer este supuesto restrictivo. Se demostrarán dos pruebas. La primera, conocida como la prueba de correlación de Pitman, utiliza observaciones originales; y la segunda, conocida como la prueba de Hoteling y Pabst, utiliza rangos en vez de las observaciones originales. Se pone énfasis en la lógica de estas pruebas, así como en la relación entre las dos.

En esta sección, y en las posteriores, se hará hincapié en la comprensión de los principios subyacentes y no en la eficiencia. Para todos los fines prácticos, las pruebas basadas en observaciones originales requieren de computadoras para su realización, que a su vez, utilizan diversos algoritmos para aumentar su eficiencia. No nos ocuparemos de esos métodos aquí.

La prueba de correlación de Pitman.

Razonamiento

Suponga que un investigador obtiene dos mediciones sobre cada uno de cuatro sujetos cuyos datos se muestran a continuación.

Sujeto	x	y
1	4	7
2	0	0
3	3	3
4	1	4

Mediante la ecuación 8.2 de la página 296, la correlación del producto-momento de Pearson entre x y y es .822. El investigador desea probar la hipótesis nula $H_0 : \rho = 0$, pero está casi seguro de que la población de la cual se muestrearon los datos está sesgada y, por lo tanto, no distribuida de forma normal. Como resultado, el investigador siente que no puede depender de la ecuación 8.4 de la página 309 para producir resultados válidos. Se puede realizar una prueba que no requiera la suposición de normalidad para asegurar su validez evocando la siguiente lógica.

Si la hipótesis nula es cierta, es decir, si no hay correlación entre x y y, entonces, los apareamientos de los datos anteriores son arbitrarios. Esto es, cada valor x pudo haber sido igualmente apareado con cualquiera de los otros valores y, o con los que se tienen en el conjunto de datos. Por casualidad, los apareamientos se dieron de esta manera, pero igualmente podría haberse hecho cualquier otro apareamiento bajo una hipótesis nula verdadera.

En contraste, si hay una relación positiva o negativa entre x y y, los apareamientos no serán aleatorios. Por el contrario, para una relación positiva, los valores x altos tenderán a estar asociados con los valores y altos, y valores x bajos con los valores y bajos. Para la asociación negativa existirá una relación inversa.

Mitad inferior de la distribución muestral

−.949 −.822 −.759 −.759 −.569 −.569 −.569 −.316 −.253 −.253 −.190 −.190

Mitad superior de la distribución muestral

.190 .190 .253 .253 .316 .569 .569 .569 .759 .759 .822 .949

r obtenida

FIGURA 10.1: Distribución muestral de permutación de *r* para un conjunto de datos en particular.

Se deduce que *bajo una hipótesis nula verdadera* los valores *y* podrían tomar cualquiera de las disposiciones $n! = 4! = 24$, como se observa en la siguiente página. Esto significa que cualquiera de los 24 coeficientes de correlación obtenidos de las 24 disposiciones de *y* serían igualmente probables *si la hipótesis nula es verdadera*.

La distribución muestral de permutación de *r* se construye si acomodamos los 24 valores de *r* obtenidos de las 24 permutaciones de *y* en orden creciente, como se muestra en la figura 10.1. Si la hipótesis nula es verdadera, esperaríamos que la *r obtenida* (de los datos recabados en el estudio) estuviera cerca de cero, ya que una hipótesis nula verdadera significa que la correlación en la población es cero. Por otra parte, si $\rho > 0$ esperaríamos que *r* tomara valor en la cola derecha de la distribución y que ocurriera lo opuesto cuando $\rho < 0$. En este caso, la *r* obtenida es .822. ¿Qué probabilidad hay de obtener un valor tan extremo de *r* cuando la hipótesis nula es verdadera? Es simplemente la proporción de *r*'s que toma un valor tan extremo (es decir, lejos de cero), o más, como el obtenido de los datos. En este caso, sólo dos de los 24 valores posibles de *r* están tan lejos, o más, del cero como el obtenido de los datos: .822 y .949. Ahora podemos decir que de los 24 valores posibles de *r* que podrían haberse obtenido bajo una hipótesis nula verdadera sólo 2 son tan extremos como el observado en los datos. La probabilidad de alcanzar un valor tan extremo de *r* bajo una hipótesis nula verdadera es, entonces, 2/24 = .083. Éste es, de hecho, el valor-*p* para una prueba con la alternativa $\rho > 0$.

En general, el valor-*p* para una prueba con la alternativa de una cola $\rho > 0$ es la proporción de valores en la distribución muestral de permutación que es mayor que o igual a la observada en los datos. Cuando la alternativa de una cola es de la forma $\rho < 0$, el valor-*p* para ésta es la proporción de valores en la distribución de permutación menor que o igual a la observada en los datos. El valor-*p* de dos colas para la alternativa $\rho \neq 0$ se calcula encontrando la proporción de observaciones en la cola superior (inferior) de la distribución que es mayor (menor) que o igual a la *r* obtenida y multiplicando este valor por 2. La proporción mayor que o igual a la *r* obtenida se multiplica por 2 cuando *r* es mayor que cero, en tanto que la proporción menor que o igual a la *r* obtenida se multiplica por 2 cuando ésta es menor que cero.

TABLA 10.2: Todos los posibles conjuntos de pares de datos y coeficientes de correlación para un conjunto de datos en particular.

1

4	7
0	0
3	3
1	4

$r = .822$

2

4	0
0	7
3	3
1	4

$r = -.949$

3

4	3
0	7
3	0
1	4

$r = -.759$

4

4	7
0	3
3	0
1	4

$r = .253$

5

4	0
0	3
3	7
1	4

$r = -.190$

6

4	3
0	0
3	7
1	4

$r = .569$

7

4	0
0	0
3	7
1	3

$r = .759$

8

4	0
0	4
3	7
1	3

$r = .253$

9

4	7
0	4
3	0
1	3

$r = .190$

10

4	4
0	7
3	0
1	3

$r = .569$

11

4	0
0	7
3	4
1	3

$r = -.822$

12

4	7
0	0
3	4
1	3

$r = .949$

13

4	7
0	3
3	4
1	0

$r = .759$

14

4	3
0	7
3	4
1	0

$r = -.253$

15

4	4
0	7
3	3
1	0

$r = -.190$

16

4	7
0	4
3	3
1	0

$r = .569$

17

4	3
0	4
3	7
1	0

$r = .316$

18

4	4
0	3
3	7
1	0

$r = .569$

19

4	4
0	3
3	0
1	7

$r = -.316$

20

4	0
0	4
3	0
1	7

$r = -.569$

21

4	0
0	4
3	3
1	7

$r = .759$

22

4	4
0	0
3	3
1	7

$r = .253$

23

4	3
0	0
3	4
1	7

$r = .190$

24

4	0
0	3
3	4
1	7

$r = -.569$

Realización de la prueba

Los pasos para realizar la prueba de correlación de Pitman son los siguientes.

1. Calcule r para los datos obtenidos del estudio. Llame a este valor r obtenida.
2. Habiendo fijado una de las dos variables (por ejemplo, x), haga todos los posibles conjuntos de pares de x y y formando todas las permutaciones $n!$ probables de la variable no fija (por ejemplo, y), como se muestra para ejemplificar el conjunto de datos de la página 351.
3. Calcule r para cada conjunto de pares de datos, como se indica en la página 351.
4. Ordene los valores de r obtenidos en el paso 3 del menor al mayor. Ésta es la distribución muestral de permutación de r.
5. Para una prueba de una cola con la alternativa $H_A : \rho > 0$ calcule la proporción de valores en la distribución de permutación que son mayores que o iguales a la r obtenida. Éste es el valor-p para la prueba de una cola.
6. Para una prueba de una cola con la alternativa $H_A : \rho < 0$ calcule la proporción de valores en la distribución de permutación que son menores que o iguales a la r obtenida. Éste es el valor-p para la prueba de una cola.
7. Para la prueba de dos colas con la alternativa $H_A : \rho \neq 0$:

   a) calcule la proporción de valores en la distribución de permutación que son mayores que o iguales a la r obtenida, si ésta es mayor que cero. Multiplique este valor por 2 para encontrar el valor-p de dos colas.

   b) calcule la proporción de valores en la distribución de permutación que son menores que o iguales a la r obtenida, si ésta es menor que cero. Multiplique este valor por 2 para encontrar el valor-p de dos colas.

EJEMPLO 10.6

Utilice la distribución muestral de permutación de r que aparece en la figura 10.1 para encontrar el valor-p de dos colas para la siguiente prueba.

$$H_0 : \rho = 0$$
$$H_A : \rho < 0$$

suponiendo que la r obtenida es $-.949$.

Solución Como la alternativa es de la forma $\rho < 0$, debemos realizar una prueba de una cola en la que el valor-p es la proporción de valores en la distribución muestral menor que o igual a la obtenida de los datos. En este caso, sólo el valor de r de la permutación igual a $-.949$ cumple con el criterio de ser menor que o igual a la r obtenida de $-.949$, por lo que $p = \frac{1}{24} = .042$. Si α fuera .05, concluiríamos que $\rho < 0$. ■

EJEMPLO 10.7

Utilice la distribución muestral de permutación de r mostrada en la figura 10.1 para encontrar el valor-p para la siguiente prueba.

$$H_0 : \rho = 0$$
$$H_A : \rho \neq 0$$

suponiendo que la r obtenida es .759.

Solución Como la r obtenida es mayor que cero, encontramos la proporción de valores de permutación mayor que o igual a la r obtenida. Observamos que dos valores de .759, un valor de .822 y uno de .949 son mayores que o iguales a la r obtenida, por lo que el valor-p deseado es (2)(4/24) = .333. ■

EJEMPLO 10.8

Dados cinco pares de datos xy, ¿cuál es el valor-p más pequeño que podría obtenerse de una prueba de permutación de $H_0 : \rho = 0$? ¿Sería posible efectuar una prueba de dos colas con $\alpha = .01$? ¿Y, una prueba de una cola?

Solución Como hay $n! = 5! = 120$ conjuntos de pares xy posibles, de los cuales cada uno contribuirá con un valor a la distribución de permutación, conoceremos el valor-p más pequeño cuando la r obtenida tome el valor del más pequeño o del más grande de las r's de permutación. Suponiendo que no existen valores ligados en los extremos de la distribución de permutación, el valor-p de una cola más pequeño posible sería $\frac{1}{120} = .0083$ y el más pequeño para una prueba de dos colas sería $(2)\left(\frac{1}{120}\right) = .0167$. Por consiguiente, no se puede realizar una prueba de dos colas con $\alpha = .01$ porque nunca se podría alcanzar una conclusión importante con este tamaño muestral. Sin embargo, sería posible efectuar una prueba de una cola con $\alpha = .01$. ■

Suposiciones

La suposición fundamental en la prueba de correlación de Pitman se relaciona con la independencia. No sólo debe ser cierto que x y y no estén correlacionados bajo una hipótesis nula verdadera, también debe ser cierto que haya independencia entre todos los valores de x así como entre todos los valores de y. Por ejemplo, suponga que se han tomado las agudezas visuales del ojo izquierdo (x_1) y del derecho (y_1) de un paciente. En otro momento, una vez más el paciente se somete a una prueba produciendo las medidas (x_2) y (y_2). Así, en un conjunto de n pares de mediciones que deben estar correlacionadas, x_1 y x_2, así como y_1 y y_2, son de la misma persona. Suponga también que el sujeto tiene vista deficiente. En estas condiciones, los valores de (y_1) y (y_2) son dependientes, por lo que el valor de y_2 dependerá del valor de y_1. En esta circunstancia todas las permutaciones $n!$ de y no son completamente probables, ya que para los valores fijos de x_1 x_2, y_2 debe tomar algún valor cerca de y_1 en vez de cualquiera de los otros valores de y. No se puede depender de que la prueba de Pitman sea sólida ante violaciones de este tipo.

Prueba de Hoteling y Pabst para la correlación de rangos. La prueba de Hoteling y Pabst (H & P) es similar a la prueba de correlación de Pitman, excepto porque los datos originales se convierten en rangos antes de que se calculen la r obtenida y la distribución de permutación. Se puede demostrar que, en ausencia de valores duplicados para x o y, calcular el coeficiente de correlación producto-momento de Pearson sobre los rangos, como se muestra a continuación, produce el mismo resultado que un bien conocido estadístico de correlación no paramétrico llamado **rho de Spearman**. Por esta razón, la prueba descrita en esta sección, que emplea el coeficiente de correlación producto-momento de Pearson, es también una prueba de rho de Spearman.

Razonamiento

Para realizar la prueba de Pitman para ocho pares de observaciones tendríamos que encontrar 8! = 40, 320 conjuntos de pares de datos y calcular el coeficiente de correlación para cada uno. La cantidad de pares de datos aumenta rápidamente si consideramos una n mayor. Valdría la pena generar esta distribución de permutación para una sola prueba, pero no podemos evitar sentirnos desmoralizados por el hecho de que para cada nuevo conjunto de datos se debe repetir la tarea. Pero, considere lo que pasaría si remplazáramos cada observación en el conjunto de datos original por su rango, en donde el valor x más pequeño reciba un rango de 1, el segundo más pequeño un rango de 2 y así sucesivamente; el valor más grande recibe un rango de n. Entonces, clasificaríamos la variable y de la misma manera.

Podríamos, entonces, calcular el coeficiente de correlación sobre los rangos y no sobre las puntuaciones originales. Esto produciría un resultado diferente, pero aún así nos daría una medida de la relación entre x y y. La ventaja de esta estrategia es que las permutaciones tendrían que llevarse a cabo solamente una vez para cualquier tamaño de muestra dado. Cuando se tuviera que probar un nuevo conjunto de datos con la misma n, aparecerían los rangos anteriores del análisis, aunque los datos originales hubieran cambiado. Se deduce que la distribución de permutación sería la misma sin importar los valores de los datos originales, ya que las correlaciones se están conduciendo sobre sus rangos. Sólo queda encontrar la distribución de permutación para cada tamaño de prueba (n).

Ahora, efectuaremos una prueba H & P sobre los mismos datos usados previamente para la prueba de Pitman. Estos datos se repiten aquí.

Sujeto	x	y
1	4	7
2	0	0
3	3	3
4	1	4

La prueba H & P requiere que remplacemos los datos originales por sus respectivos rangos. El resultado se muestra a continuación. Utilizamos la notación R_x y R_y para denotar los rangos de x y y.

Sujeto	R_x	R_y
1	4	4
2	1	1
3	3	2
4	2	3

La r obtenida para los rangos es .800. Los 24 conjuntos de permutaciones para los rangos con sus valores de r correspondientes se muestran en la página siguiente. La distribución muestral de permutación resultante se muestra en la figura 10.2 de la página 356. Como se observa en esta distribución, hay cuatro valores de la distribución muestral de permutación, es decir, tres valores de .80 y uno de 1.00, que son mayores que o iguales al valor r obtenido de .80. Así, el valor-p para una prueba de una cola sería 4/24 = .167 y para una prueba de dos colas sería (2)(4/24) = .333.

TABLA 10.3: Todos los posibles conjuntos de pares de rangos y coeficientes de correlación para $n = 4$ pares de datos.

1

4	4
1	1
3	2
2	3

$r = .80$

2

4	1
1	4
3	2
2	3

$r = -1.00$

3

4	2
1	4
3	1
2	3

$r = -.80$

4

4	4
1	2
3	1
2	3

$r = .40$

5

4	1
1	2
3	4
2	3

$r = -.20$

6

4	2
1	1
3	4
2	3

$r = .40$

7

4	3
1	1
3	4
2	2

$r = .80$

8

4	1
1	3
3	4
2	2

$r = -.40$

9

4	4
1	3
3	1
2	2

$r = .20$

10

4	3
1	4
3	1
2	2

$r = -.40$

11

4	1
1	4
3	3
2	2

$r = -.80$

12

4	4
1	1
3	3
2	2

$r = 1.00$

13

4	4
1	2
3	3
2	1

$r = .80$

14

4	2
1	4
3	3
2	1

$r = -.40$

15

4	3
1	4
3	2
2	1

$r = -.20$

16

4	4
1	3
3	2
2	1

$r = .40$

17

4	2
1	3
3	4
2	1

$r = .00$

18

4	3
1	2
3	4
2	1

$r = .60$

19

4	3
1	2
3	1
2	4

$r = .00$

20

4	2
1	3
3	1
2	4

$r = -.60$

21

4	1
1	3
3	2
2	4

$r = -.80$

22

4	3
1	1
3	2
2	4

$r = .40$

23

4	2
1	1
3	3
2	4

$r = .20$

24

4	1
1	2
3	3
2	4

$r = -.40$

Mitad inferior de la distribución muestral

−1.00 −.80 −.80 −.80 −.60 −.40 −.40 −.40 −.40 −.20 −.20 .00

Mitad superior de la distribución muestral

.00 .20 .20 .40 .40 .40 .40 .60 .80 .80 .80 1.00

r obtenida

FIGURA 10.2: Distribución muestral de permutación de r calculada sobre rangos donde $n = 4$.

A diferencia de la prueba de Pitman, que utiliza observaciones originales, la prueba H & P sobre rangos no prueba $H_0 : \rho = 0$, sino que prueba la hipótesis menos específica de que x y y son independientes. La alternativa de dos colas asegura que las dos variables no son independientes, en tanto que las versiones de una cola afirman una relación positiva o negativa entre las variables. También debe quedar claro que la correlación sobre rangos es, a diferencia de la correlación sobre puntuaciones originales, no una expresión del grado de linealidad entre x y y, sino una evaluación de la relación monotónica entre x y y. Una **relación monotónica** es aquella en la que los aumentos y las reducciones en una variable están acompañados por aumentos o reducciones en la otra variable, pero no necesariamente en línea recta. Un ejemplo sería si x tomara los valores 1, 2 y 3 mientras que y tomara los valores 1, 4 y 11.

Ahora realizaremos la prueba H & P con un nuevo conjunto de datos como se indica aquí.

Sujeto	x	y
1	942	13
2	101	14
3	313	18
4	800	10

La prueba H & P requiere que coloquemos los datos originales con sus respectivos rangos. El resultado se muestra abajo.

Sujeto	R_x	R_y
1	4	2
2	1	3
3	2	4
4	3	1

La r obtenida para estos rangos es −.60. La referencia a la distribución muestral de permutación de la figura 10.2 indica que los cinco valores de permutación son menores que o iguales a −.60, por lo que el valor-p de la cola inferior es 5/24 = .208, mientras que para una prueba de dos colas es (2)(5/24) = .417. Es importante comprender que si hubiéramos preferido realizar la prueba de Pitman, habríamos tenido que generar una nueva distribución muestral de permutación para los datos. Como remplazamos los datos originales por los rangos y ya habíamos generado la distribución de permutación para los rangos del 1 al 4 para cada variable, ya teníamos la distribución apropiada.

Remplazar los datos originales por rangos también nos permite construir tablas de valores críticos para la prueba H & P. Para lograrlo, sólo necesitamos generar la distribución muestral de permutación para los conjuntos de rangos para varios tamaños de muestras y ubicar los valores de permutación de tal forma que α o $\alpha/2$ de los valores de permutación sean mayores que o iguales al valor identificado. Por ejemplo, si examinamos la figura 10.2 de la página anterior, encontramos que una correlación de 1.0 tiene un valor-*p* asociado (de una cola) de 1/24 = .042, en tanto que .80 tiene un valor asociado de 4/24 = .167, por lo que no hay un valor crítico para una prueba de una cola efectuada con α = .005, .010 o .025. De igual manera, no hay ningún valor crítico para una prueba de dos colas al nivel .010, .020 o .050. Sin embargo, el valor de 1.0 se puede emplear para una prueba de una cola al nivel .05 o una prueba de dos colas al nivel .10. El nivel de significancia no sería exactamente .05 o .10, pero no excedería esos valores. El mismo valor crítico (1.0) podría usarse para pruebas de una cola o de dos colas en los niveles .10 y .20, respectivamente.

El punto aquí es que los valores críticos de las pruebas basadas en la permutación no pueden darnos el nivel exacto de significancia deseada. Como resultado, los valores críticos se seleccionan para que el rango del error tipo I sea menor que o igual al nivel declarado. A medida que el tamaño de la muestra aumenta este problema prácticamente desaparece.

El Apéndice F provee valores críticos de la cola superior para $n = 4$ a 90 para el coeficiente de correlación de rangos. Para pruebas de la cola superior, usted sólo necesita comparar el coeficiente de correlación de rangos obtenido contra al valor crítico apropiado. Si la *r* obtenida es mayor que o igual a *r* crítica, se rechaza la hipótesis nula. Para pruebas de la cola inferior, se compara la *r* obtenida con el negativo del valor tabulado. Si la *r* obtenida es menor que o igual a este valor, se rechaza la hipótesis nula. Para pruebas de dos colas se rechaza la hipótesis nula si la *r* obtenida es mayor que o igual al valor en la tabla o si la *r* obtenida es menor que o igual al negativo del valor en la tabla.

Para *n* mayor que 90, la correlación de rangos se puede sustituir en la ecuación 8.4 (página 309) por el resultado que se refiere a una tabla *t* con $n - 2$ grados de libertad. El resultado es una muy buena aproximación al valor de la permutación cuando $n > 90$.

Realización de la prueba

Los pasos para realizar la prueba H & P para la correlación de rangos son los siguientes.

1. Remplace las observaciones originales con sus rangos respectivos y dé al valor *x* más bajo un rango de 1, al segundo valor más bajo un rango de dos, etcétera, hasta que el valor más grande sea remplazado con un rango *n*. La variable *y* se remplaza con rangos de la misma forma.
2. Calcule *r* para los rangos obtenidos en el paso 1. Denomine este valor *r* obtenida.
3. Si $4 \leq n \leq 90$, refiera la *r* obtenida al Apéndice F y
   a) para una prueba de la cola superior rechace la hipótesis nula, si la *r* obtenida es mayor que o igual al valor de la tabla.
   b) para una prueba de la cola inferior rechace la hipótesis nula si la *r* obtenida es menor que o igual al negativo del valor de la tabla.

c) para una prueba de dos colas rechace la hipótesis nula si la r obtenida es mayor que o igual al valor de la tabla *o* si la r obtenida es menor que o igual al negativo del valor de la tabla.

4. Para $n > 90$ utilice la ecuación 8.4 de la página 309 y refiera el resultado a la tabla B con $n - 2$ grados de libertad.

EJEMPLO 10.9

Utilice los datos provistos para calcular el coeficiente de correlación de rangos. Utilice el valor resultante de r para realizar una prueba H & P de dos colas con $\alpha = .05$ de la hipótesis nula de que x y y son independientes.

Sujeto	x	y
1	1314	20
2	880	16
3	414	10
4	1774	18
5	101	11
6	902	15
7	544	12
8	722	13
9	377	9
10	1200	17

Solución Comenzamos por remplazar las observaciones originales con sus rangos, como se indica aquí.

Sujeto	R_x	R_y
1	9	10
2	6	7
3	3	2
4	10	9
5	1	3
6	7	6
7	4	4
8	5	5
9	2	1
10	8	8

Para propósitos de cálculo será conveniente ordenar los rangos como se indica en la tabla 10.4.

Si usamos las cantidades anteriores con la ecuación 8.2 (modificadas apropiadamente para los rangos) nos da

$$r = \frac{\sum R_x R_y - \frac{\left(\sum R_x\right)\left(\sum R_y\right)}{n}}{\sqrt{\left[\sum R_x^2 - \frac{\left(\sum R_x\right)^2}{n}\right]\left[\sum R_y^2 - \frac{\left(\sum R_y\right)^2}{n}\right]}}$$

TABLA 10.4: Pesos (en libras) de sujetos sometidos a tres diferentes dietas ordenados para su análisis.

Sujeto	R_x	R_y	$R_x R_y$	R_x^2	R_y^2
1	9	10	90	81	100
2	6	7	42	36	49
3	3	2	6	9	4
4	10	9	90	100	81
5	1	3	3	1	9
6	7	6	42	49	36
7	4	4	16	16	16
8	5	5	25	25	25
9	2	1	2	4	1
10	8	8	64	64	64
Σ	55	55	380	385	385

$$
\begin{aligned}
&= \frac{380 - \dfrac{(55)(55)}{10}}{\sqrt{\left[385 - \dfrac{(55)^2}{10}\right]\left[385 - \dfrac{(55)^2}{10}\right]}} \\
&= \frac{77.5}{\sqrt{[82.5][82.5]}} \\
&= .939.
\end{aligned}
$$

La referencia al Apéndice F indica que para $n = 10$, r crítica para una prueba de dos colas efectuada con $\alpha = .05$ es $\pm.648$. Como la r obtenida de .939 excede .648, se rechaza la hipótesis nula. Podemos concluir, entonces, que x y y no son independientes y que existe una relación positiva entre las dos variables. ■

EJEMPLO 10.10

Un investigador cree que existe una relación negativa entre una medida de densidad ósea y el consumo diario promedio de cafeína en mujeres de 70 años o más. Para probar esta hipótesis se hacen dos evaluaciones en un grupo de 80 mujeres mayores. Como el investigador sospecha que la distribución del consumo de cafeína está sesgada, selecciona un coeficiente de correlación de rangos para evaluar la relación entre las dos variables. La correlación de rangos calculada es $-.122$. Utilice la prueba H & P para probar la hipótesis nula de que no hay ninguna relación entre la densidad ósea y el consumo de cafeína en mujeres mayores, contra la alternativa de que existe una relación negativa entre las dos variables. Realice la prueba al nivel .05. Interprete los resultados.

Solución La referencia al Apéndice F indica que para $n = 80$, r crítica para una prueba de una cola con $\alpha = .05$, cuya alternativa postula una relación negativa es $-.185$. Como la r obtenida de $-.122$ es mayor que este valor, no se rechaza la hipótesis nula. Como resultado, el investigador no pudo demostrar una relación negativa entre las dos variables. ■

EJEMPLO 10.11

Dado $n = 120$ y una correlación de rangos de .208, realice una prueba de dos colas de la hipótesis nula de que no existe relación alguna entre x y y con $\alpha = .10$.

Solución Mediante la ecuación 8.4 de la página 309

$$r = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{.208}{\sqrt{\frac{1-.208^2}{120-2}}} = 2.310$$

La referencia al Apéndice B indica que los valores t críticos para una prueba de dos colas con $n - 2 = 120 - 2 = 118$ grados de libertad son ± 1.658. Como la t obtenida de 2.310 excede 1.658, se rechaza la hipótesis nula, lo que nos lleva a la conclusión de que existe una relación positiva entre las dos variables. ■

Suposiciones

Hay dos suposiciones subyacentes en la prueba de Hoteling y Pabst para la correlación de rangos. La primera es que todos los valores x son independientes, como lo son los valores y. (Para más detalles, vea las suposiciones subyacentes a la prueba de correlación de Pitman de la página 353). El segundo supuesto requiere que no existan valores duplicados entre las observaciones x ni tampoco entre las observaciones y. Los valores de x que dupliquen los valores de y no violan este supuesto y no implican consecuencias. Considere el siguiente conjunto de datos.

Sujeto	x	y
1	9	12
2	13	16
3	7	10
4	4	18
5	7	13

Observe que hay dos valores x de 7. Ésta es la violación al supuesto. El hecho de que haya una x y una y con valor 13 no marca una diferencia. El problema surge cuando intentamos remplazar las observaciones originales con los rangos. Cuando se sustituyen los rangos para la variable x el 4 sería remplazado con el rango 1, pero ¿qué podemos hacer con las dos puntuaciones de 7? Existen varias estrategias, pero la más común asigna el promedio de los rangos que hubieran sido asignados de no haber existido un vínculo. Por lo general, la segunda puntuación más baja recibiría un rango de 2, mientras que la tercera más baja recibiría un rango de 3. La estrategia del **rango promedio** asigna el promedio del rango 2 y 3 o 2.5 a las dos puntuaciones de 7. Por este método los rangos aparecerían como sigue.

Sujeto	R_x	R_y
1	4	2
2	5	4
3	2.5	1
4	2	5
5	2.5	3

Si hubiera habido tres vínculos en vez de dos, remplazaríamos cada uno con el promedio de los tres rangos que hubieran sido asignados a los tres valores de no existir vínculos.

Pero nuestro problema no está resuelto. Los valores críticos para $n = 5$ en el Apéndice F se obtuvieron para los rangos del 1 al 5, no para los rangos 1, 2.5, 2.5, 4 y 5. Las siguientes son estrategias que pueden usarse cuando se encuentran observaciones ligadas.[8]

1. Conserve las observaciones originales y aplique la prueba de Pitman.[9]
2. Utilice el método del rango promedio descrito antes y genere la distribución muestral de permutación de estos rangos. (Véase la nota al pie número 9).
3. Utilice el método del rango promedio descrito antes y aplique la ecuación 8.4 de la página 309 para la correlación de rangos. El resultado se refiere a una distribución t con $n - 2$ grados de libertad. Este método, por lo general, produce buenos resultados si n es lo suficientemente grande (mayor que 30) y la proporción de observaciones ligadas es pequeña (20% o menos).

10.3.2 Pruebas de muestras pareadas

En el capítulo 5, en la página 162, aprendimos a realizar una prueba t de muestras apareadas mediante la ecuación 5.1 para probar la hipótesis nula $H_0 : \mu_d = 0$, donde μ_d es la media de una población de puntuaciones diferentes. Una de las suposiciones subyacentes a esa prueba es que las puntuaciones diferenciales de la muestra provenían de una población distribuida de forma normal. En esta sección usted aprenderá a realizar pruebas de muestras apareadas que no requieren esta suposición restrictiva. Se demostrarán dos pruebas. La primera utiliza observaciones originales, en tanto que la segunda, conocida como la prueba de los signos de Wilcoxon, utiliza rangos en vez de observaciones originales. Se pone énfasis en la lógica de estas pruebas, así como en la relación entre ambas.

Prueba *t* de permutación para muestras apareadas.

Razonamiento

Suponga que los siguientes resultados son de un estudio en el que se observó a cuatro individuos antes y después de un tratamiento.

Sujeto	Pre-tratamiento	Post-tratamiento	(Diferencia) d
1	95	99	4
2	111	120	9
3	97	102	5
4	132	130	−2

Mediante la ecuación 5.1 de la página 162, el estadístico t de muestras apareadas para este dato es 1.760. Si el investigador desea probar la hipótesis nula $H_0 : \mu_d = 0$, pero quiere evitar la suposición de normalidad de la población de puntuaciones diferentes, se puede realizar una prueba de permutación basada en la siguiente lógica.

[8] Otras estrategias se describen en textos no paramétricos. Véase por ejemplo [7].

[9] Esto se puede llevar a cabo utilizando software comercial.

El primer sujeto obtuvo una puntuación pre-tratamiento de 95 y una puntuación post-tratamiento de 99. Pero, si ni el tratamiento ni cualquier otro factor tuvieran un efecto en las puntuaciones, entonces el orden de estas puntuaciones para el primer sujeto sería una cuestión de probabilidad. Esto es, en ausencia de un tratamiento efectivo, este sujeto sólo recibió una puntuación pre-tratamiento de 99 y una puntuación pos-tratamiento de 95. Lo mismo es verdadero también para otros sujetos. Entonces, el orden de las puntuaciones para cada sujeto *por debajo de la hipótesis nula verdadera es que la efectividad del no tratamiento* sólo sería una cuestión de probabilidad.

De este razonamiento se puede ver que si se revierten las puntuaciones de los sujetos en todas las formas posibles y se calcula un estadístico t de muestras apareadas para cada una de estas configuraciones, podemos generar todos los posibles estadísticos t que se podían haber obtenido de los datos del estudio si la hipótesis nula fuera verdadera. Pero, revertir un par de puntuaciones sólo cambia el signo de la puntuación de diferencia, por lo que las diferencias positivas se convierten en negativas y viceversa. Esto significa que para calcular todos los estadísticos t posibles, sólo necesitamos formar todos los patrones positivos/negativos posibles de las puntuaciones de diferencia y calcular el estadístico en cada uno de estos conjuntos. Note que la primera puntuación de diferencia se puede expresar de dos formas, positiva o negativa. Para cada una de éstas, la segunda diferencia también se puede expresar de dos maneras, por lo que el número de formas de describir las dos primeras puntuaciones de diferencia es $2 \cdot 2 = 4$. Si llevamos esta lógica a todas las puntuaciones de diferencia n vemos que el número de formas de representar puntuaciones de diferencia n es 2^n, que en este problema es $2^4 = 16$. Las 16 configuraciones se muestran en la tabla 10.5.

Habiendo formado los 16 posibles conjuntos de cambio de signo de las puntuaciones de diferencia y calculando un estadístico t de muestras apareadas para cada uno, podemos ordenar los 16 estadísticos t en una distribución muestral de permutación, como se aprecia en la figura 10.3 de la página 363. Si la hipótesis nula es verdadera, esperaríamos que la *t obtenida* (es decir, la que se obtuvo de los datos recabados en el estudio) estuviera cerca de cero, ya que una hipótesis nula significa que la media de la población de puntuaciones de diferencia es cero. Por otra parte, si $\mu_d > 0$, esperaríamos que t tomara cierto valor en la cola del extremo derecho de la distribución y que ocurriera lo opuesto cuando $\mu_d < 0$. En este caso, t obtenida para los datos derivados del estudio es 1.760. ¿Cuál es la probabilidad de obtener un valor tan extremo de t cuando la hipótesis nula es verdadera? Es simplemente la proporción de t's en la distribución muestral que toman un valor tan extremo (es decir, lejos de cero), o más, que el obtenido de los datos. En este caso, sólo dos de los 16 posibles valores de t están tan lejos o más de cero que el obtenido de los datos: 1.760 y 3.397. Ahora podemos decir que de los 16 posibles valores que pudieron haberse obtenido bajo una hipótesis nula cierta, sólo 2 son tan extremos como lo que se observó de los datos. Así, la probabilidad de alcanzar un valor tan extremo de t bajo una hipótesis nula verdadera es $2/16 = 0.125$. Éste es, de hecho, el valor-p para una prueba con la alternativa $\mu_d > 0$. El valor-p de dos colas sería

$$(2)\left(\frac{2}{16}\right) = .250.$$

En general, el valor-p para una prueba con la alternativa de una cola $\mu_d > 0$ es la proporción de valores en la distribución muestral de permutación que son mayores que o iguales al valor observado de los datos. Cuando la alternativa de una cola es de la forma $\mu_d < 0$, el valor-p para esta alternativa de una cola es la proporción de valores en la distribución de permutación que son menores que o iguales al observado de los datos. El valor-p de dos colas para la alternativa $\mu_d \neq 0$ se calcula encontrando la proporción de observaciones en la cola superior (inferior) de la

TABLA 10.5: Todos los posibles conjuntos de cambios de signos algebraicos y estadísticos *t* de muestras apareadas para un conjunto específico de puntuaciones de diferencia.

1	2	3	4	5	6
–4	4	–4	4	–4	4
9	–9	–9	9	9	–9
5	5	5	–5	–5	–5
–2	–2	–2	–2	–2	–2
$t = .661$	$t = -.155$	$t = -.862$	$t = .480$	$t = -.155$	$t = -1.095$

7	8	9	10	11	12
–4	4	–4	4	–4	4
–9	9	9	–9	–9	9
–5	5	5	5	5	–5
–2	2	2	2	2	2
$t = -3.397$	$t = 3.397$	$t = 1.095$	$t = .155$	$t = -.480$	$t = .862$

13	14	15	16
–4	4	–4	4
9	–9	–9	9
–5	–5	–5	5
2	2	2	–2
$t = .155$	$t = -.661$	$t = -1.760$	$t = 1.760$

Mitad inferior de la distribución muestral

–3.397 –1.760 –1.095 –.862 –.661 –.480 –.155 –.155

Mitad superior de la distribución muestral

.155 .155 .480 .661 .862 1.095 1.760 3.397

t obtenida

FIGURA 10.3: Distribución muestral de permutación de los estadísticos *t* de muestras apareadas para un conjunto de datos en particular.

distribución que son mayores (menores) que o iguales a la t obtenida y multiplicando este valor por 2. La proporción mayor que o igual a la t obtenida se multiplica por 2 cuando t es mayor que cero, en tanto que la proporción menor que o igual a la t obtenida se multiplica por 2 cuando la t obtenida es menor que cero.

Realización de la prueba

Los pasos para realizar la prueba t de permutación para muestras apareadas son los siguientes.

1. Calcule t para los datos obtenidos del estudio. Denomine este valor t obtenida.
2. Forme todos los posibles conjuntos de cambios de signos algebraicos de las puntuaciones de diferencia como se muestra para los datos del ejemplo en la página anterior.
3. Calcule el estadístico t de muestras apareadas para cada conjunto de puntuaciones de diferencia como se muestra en la página 363.
4. Ordene los valores de t obtenida del paso 3 del menor al mayor. Ésta es la distribución muestral de permutación de t.
5. Para una prueba de una cola con la alternativa $H_A : \mu_d > 0$, calcule la proporción de valores en la distribución de permutación que son mayores que o iguales a la t obtenida. Éste es el valor-p para la prueba de una cola.
6. Para una prueba de una cola con la alternativa $H_A : \mu_d < 0$, calcule la proporción de valores en la distribución de permutación que son menores que o iguales a la t obtenida. Éste es el valor-p para la prueba de una cola.
7. Para la prueba de dos colas con la alternativa $H_A : \mu_d \neq 0$:
   a) calcule la proporción de valores en la distribución de permutación que son mayores que o iguales a la t obtenida, si ésta es mayor que cero. Multiplique este valor por 2 para encontrar el valor-p de dos colas.
   b) calcule la proporción de valores en la distribución de permutación que son menores que o iguales a la t obtenida si ésta es menor que cero. Multiplique este valor por 2 para encontrar el valor-p de dos colas.

EJEMPLO 10.12

Utilice la distribución muestral de permutación del estadístico t de muestras apareadas que se observa en la figura 10.3 para encontrar el valor-p para la siguiente prueba

$$H_0 : \mu_d = 0$$
$$H_A : \mu_d < 0$$

suponiendo que la t obtenida es −.862.

Solución Como la alternativa es de la forma $\mu_d < 0$, debemos realizar una prueba de una cola con un valor-p que sea la proporción de los valores en la distribución muestral que son menores que o iguales al obtenido de los datos. En este caso los valores de t que satisfacen el criterio de ser menores que o iguales a la t obtenida de −.862 son −3.397, −1.760, −1.095 y −.862, por lo que $p = \frac{4}{16} = .250$. Si α fuera .05, no rechazaríamos la hipótesis nula. ■

EJEMPLO 10.13

Utilice la distribución muestral de permutación de t de la figura 10.3 para encontrar el valor-p para la siguiente prueba.

$$H_0 : \mu_d = 0$$
$$H_A : \mu_d \neq 0$$

suponiendo que la t obtenida es 3.397.

Solución Como la t obtenida es mayor que cero, encontramos la proporción de los valores de permutación que son mayores que o iguales a la t obtenida. Observamos que sólo el valor de permutación de 3.397 cumple con este criterio, por lo que, el valor-p deseado es (2)(1/16) = .125. ■

EJEMPLO 10.14

Dado $n = 10$, ¿cuántos valores de t conformarían la distribución muestral de permutación para la prueba t de muestras apareadas?

Solución

$$2^n = 2^{10} = 1024$$ ■

Suposiciones

La suposición fundamental en la prueba t de permutación para muestras apareadas es que todas las puntuaciones de diferencia son independientes. Puede darse una violación de este supuesto, por ejemplo, si en un estudio pre y post-tratamiento, el mismo sujeto fuera tratado en dos diferentes momentos y se analizaran dos veces sus puntuaciones pre y post-tratamiento. Así, este sujeto tendría dos puntuaciones de diferencia en el conjunto de datos. Si las dos puntuaciones de diferencia se relacionaran de alguna manera, ocurriría una violación.

Como un segundo ejemplo, podría ocurrir un incumplimiento si se incluyeran en un análisis las puntuaciones de diferencia para hermanos. Si el estudio fuera sobre pérdida de peso y hubiera un componente genético relacionado con la pérdida de peso, entonces la cantidad de peso perdido en los hermanos podría relacionarse. No se puede contar con que esta prueba sea sólida ante una violación de la suposición de independencia.

Prueba de los signos de Wilcoxon (o prueba de rangos signados de Wilcoxon). Esta prueba es similar a la prueba t de permutación para muestras apareadas, excepto que las puntuaciones de diferencia se convierten en rangos antes de que se calcule la t obtenida y se genere una distribución muestral de permutación.

Razonamiento

Para realizar la prueba t de permutación para muestras apareadas para 10 pares de datos (por ejemplo, puntuaciones de diferencia), tendríamos que encontrar $2^{10} = 1024$ conjuntos de puntuaciones de diferencia y calcular un estadístico t para cada conjunto. Para 20 pares de datos, el número de conjuntos de puntuaciones de diferencia y estadísticos t sería $2^{20} = 1,048,576$. Esta tarea descomunal, y otras consideraciones, animaron a Frank Wilcoxon [49] a idear una prueba de permutación de muestras apareadas basada en rangos. Una ventaja de una estrategia así es que una vez que la distribución de permutación para el estadístico de prueba es generada para un tamaño muestral dado (n), se puede usar la distribución para cualquier muestra del mismo tamaño, sin importar los valores de los datos originales. Esto se deduce del hecho de que el dato original se remplaza con rangos para los que ya se ha construido la distribución.

La conversión de los datos originales en rangos se logra de la siguiente manera. Primero, se suprimen del análisis todas las puntuaciones de diferencia de cero y se reduce el tamaño de la muestra (n) como es debido. Después, se remplazan las puntuaciones de diferencia con los rangos de sus valores absolutos, sustituyendo el valor absoluto más pequeño por 1, el segundo más pequeño por 2, y así sucesivamente hasta que la puntuación de diferencia más grande (ignorando el signo algebraico) sea remplazada con n. El signo algebraico (más o menos) de cada observación original es asignado al rango para esa observación. Por ejemplo, suponga que el siguiente conjunto de puntuaciones de diferencia debe convertirse en rangos.

Diferencia
3
–9
1
–4
–2
5
8
–7
0

La conversión sería como se muestra en la siguiente tabla.

(1)	(2)	(3)	(4)
3	3	3	3
–9	9	8	–8
1	1	1	1
–4	4	4	–4
–2	2	2	–2
5	5	5	5
8	8	7	7
–7	7	6	–6

La columna (1) de la tabla anterior muestra puntuaciones de diferencia originales sin la diferencia cero, (2) muestra los valores absolutos de las puntuaciones de diferencia, (3) los rangos de las puntuaciones de diferencia de valor absoluto, y (4) los rangos con los signos de las puntuaciones originales. Observe que la puntuación de diferencia de cero fue removida antes de que se diera el proceso de clasificación, por lo que, n ahora es 8.

Utilizando los datos empleados para la prueba t de permutación para muestras apareadas (repetida aquí para su comodidad), la clasificación sería de la siguiente forma.

Sujeto	Pre-tratamiento	Post-tratamiento	(Diferencia) d
1	95	99	4
2	111	120	9
3	97	102	5
4	132	130	–2

Si ignoramos el signo algebraico, el rango de 1 sería asignado a 2, 2 a 4, 3 a 5 y 4 a 9. Adjuntar el signo algebraico produce el siguiente conjunto de rangos con signo.

TABLA 10.6: Todos los posibles conjuntos de cambios de signos algebraicos y estadísticos t de muestras apareadas para puntuaciones de diferencia ordenados cuando $n = 4$.

1	2	3	4	5	6
−2	2	−2	2	−2	2
4	−4	−4	4	4	−4
3	3	3	−3	−3	−3
−1	−1	−1	−1	−1	−1
$t = .679$	$t = .000$	$t = -.679$	$t = .322$	$t = -.322$	$t = -1.134$

7	8	9	10	11	12
−2	2	−2	2	−2	2
−4	4	4	−4	−4	4
−3	3	3	3	3	−3
−1	1	1	1	1	1
$t = -3.873$	$t = 3.873$	$t = 1.134$	$t = .322$	$t = -.322$	$t = .679$

13	14	15	16
−2	2	−2	2
4	−4	−4	4
−3	−3	−3	3
1	1	1	−1
$t = .000$	$t = -.679$	$t = -1.852$	$t = 1.852$

Sujeto	(Diferencia) d	Rangos con signo
1	4	2
2	9	4
3	5	3
4	−2	−1

Si aplicamos a los rangos la ecuación 5.1 de la página 162, tendremos un estadístico t de muestras apareadas de 1.852.

Para generar la distribución muestral de permutación para esta prueba de rangos debemos generar los enteros del 1 al 4 y darles valores positivos y negativos en todos los patrones posibles. Esto se lleva a cabo en la tabla 10.6. Con los valores t de cada conjunto se forma una distribución muestral de permutación, como se observa en la figura 10.4 de la siguiente página.

Desde el punto de vista técnico, la hipótesis probada mediante este estadístico sostiene que la población de puntuaciones de diferencia es simétrica respecto de cero, mientras que la alterna-

Mitad inferior de la distribución muestral

−3.873 −1.852 −1.134 −.679 −.679 −.322 −.322 .000

Mitad superior de la distribución muestral

.000 .322 .322 .679 .679 1.134 1.852 3.873

$\uparrow$

t obtenida

FIGURA 10.4: Distribución muestral de permutación de estadísticos t de muestras apareadas con base en rangos para $n = 4$.

tiva sostiene que éste no es el caso. Si suponemos que la población de puntuaciones de diferencia es simétrica,[10] entonces la hipótesis nula probada es la misma que la que se sometió a la prueba t de permutación para muestras apareadas, esto es, $H_0 : \mu_d = 0$. Así, para probar esta hipótesis nula contra la alternativa $H_A : \mu_d > 0$, vemos en la figura 10.4 que existen dos valores de distribución de permutación que son mayores que o iguales a la t obtenida de 1.852. Estos valores son 1.852 y 3.873, por lo que el valor-p para la prueba de una cola es $\frac{2}{16} = .125$.

Remplazar datos originales con rangos nos permite construir tablas de valores críticos para esta prueba. Para lograrlo, sólo necesitamos generar la distribución muestral de permutación para los conjuntos de rangos con signos para diversos tamaños de muestras y localizar los valores de permutación de tal manera que α o $\alpha/2$ de los valores de permutación sean mayores que o iguales al valor identificado. Por ejemplo, en la figura 10.4 se observa que un valor t de 3.873 tiene una valor-p (de una cola) correspondiente de 1/16 = .0625, mientras que un valor de 1.852 tiene un valor asociado de 2/16 = .125, por lo que no existe ningún valor crítico para una prueba de una cola efectuada con α = .005, .010, .025 o .050. De igual forma, no existe valor crítico para una prueba de dos colas al nivel .010, .020, .050 o .100. Sin embargo, el valor crítico de 3.873 puede usarse para una prueba de una cola al nivel .100 o una prueba de dos colas en .200. El nivel de significancia no sería exactamente .100 o .200, pero no excedería esos valores. Este problema se reduce a medida que el tamaño de la muestra aumenta.

El Apéndice G provee valores críticos para n = 4 a 90 para el estadístico t de muestras apareadas de rangos o la prueba de los rangos con signos de Wilcoxon. Para las pruebas de colas superiores, usted sólo necesita comparar el estadístico t del rango obtenido contra su valor crítico apropiado. Si la t obtenida es mayor que o igual a la t crítica se rechaza la hipótesis nula. Para las pruebas de colas inferiores, la t obtenida se compara con el negativo del valor tabulado. Si la t obtenida es menor que o igual a este valor, se rechaza la hipótesis nula. La hipótesis nula para una prueba de dos colas se rechaza, si la t obtenida es mayor que o igual al valor de la tabla o es menor que o igual al negativo del valor de la tabla. Observe que los valores críticos en esta tabla se refieren al tamaño de la muestra (n) y no a los grados de libertad como se hace con las tablas t convencionales.

[10] Un tema para discusiones posteriores.

Para n mayor que 90, estadístico t del rango obtenido se puede consultar en una tabla t con $n - 1$ grados de libertad. El resultado es una muy buena aproximación al valor de la permutación cuando $n > 90$.

Realización de la prueba

Los pasos para realizar la prueba de rangos con signos de Wilcoxon son los siguientes.

1. Calcule las puntuaciones de diferencia restando una observación en cada par de datos de la otra; la observación seleccionada para la resta debe ser constante en todos los pares de datos. Elimine cualquier diferencia de puntuación de cero.
2. Remplace las puntuaciones de diferencia obtenidas en el paso 1 con los rangos de sus valores absolutos; el más pequeño debe recibir el rango 1, el segundo más pequeño el rango 2, y así sucesivamente hasta que el más grande reciba el rango n. Adjunte el signo de las puntuaciones de diferencia a los rangos obtenidos de esta forma.
3. Utilice la ecuación 5.1 de la página 162 para calcular el estadístico t de muestras apareadas para los rangos obtenidos en el paso 2. Denomine a este valor t obtenida.
4. Para $4 \leq n \leq 90$, refiera la t obtenida al Apéndice G y
   a) para una prueba de la cola superior rechace la hipótesis nula si la t obtenida es mayor que o igual al valor de la tabla.

   b) para una prueba de la cola inferior rechace la hipótesis nula si la t obtenida es menor que o igual al negativo del valor de la tabla.

   c) para una prueba de dos colas rechace la hipótesis nula si la t obtenida es mayor que o igual al valor de la tabla, *o* si la t obtenida es menor que o igual al negativo del valor de la tabla.
5. Para $n > 90$, refiera la t obtenida al Apéndice B con $n - 1$ grados de libertad.

EJEMPLO 10.15

Realice una prueba de rangos con signos de Wilcoxon de dos colas con $\alpha = .05$ sobre los datos (ficticios) de presión arterial de la tabla 10.7 de la siguiente página.

Solución Los datos están ordenados para su análisis como sigue.

Diferencia	Diferencia absoluta	Rango	Rango con signo
–4	4	4	–4
–14	14	8	–8
–1	1	1	–1
2	2	2	2
3	3	3	3
–9	9	7	–7
6	6	6	6
–17	17	9	–9
–30	30	10	–10
5	5	5	5

TABLA 10.7: Mediciones pre y post-tratamiento de la presión arterial sistólica.

Pre-tratamiento	Post-tratamiento
151	147
174	160
150	150
171	170
144	146
139	142
159	150
140	146
137	120
179	149
146	151

La primera columna indica las puntuaciones de diferencia, por ejemplo, la puntuación post-tratamiento de cada sujeto menos la puntuación pre-tratamiento. Observe que la puntuación de diferencia de cero ha sido eliminada. La segunda columna da los valores absolutos de las puntuaciones de diferencia, mientras que la tercera indica los rangos de los valores absolutos de las puntuaciones de diferencia. La última columna proporciona el rango con signo de cada sujeto, obtenido al agregar el signo de la puntuación de diferencia al rango mostrado en la tercera columna.

La t obtenida es el estadístico t de muestras apareadas calculado sobre los rangos con signo. Para facilitar este cálculo tomamos en cuenta que la suma de los rangos con signo y la suma de los cuadrados de los rangos con signo son, respectivamente

$$\sum R_d = (-4) + (-8) + (-1) + (2) + (3) + (-7) + (6) + (-9) + (-10) + (5) = -23$$

y

$$\sum R_d = (-4)^2 + (-8)^2 + (-1)^2 + (2)^2 + (3)^2 + (-7)^2 + (6)^2 + (-9)^2 + (-10)^2 + (5)^2 = 385.$$

El estadístico t de muestras apareadas (ecuación 5.1 de la página 162), en donde se sustituye R_d por d para indicar que el análisis se lleva a cabo sobre los rangos con signo de las puntuaciones de diferencia y no en las puntuaciones de diferencia, es como sigue.

$$t = \frac{\bar{R}_d}{\frac{s_{Rd}}{\sqrt{n}}}$$

La desviación estándar indica los rangos con signo (usando la ecuación 2.16 de la página 37, en donde se sustituye R_d por x) es

$$s_{Rd} = \sqrt{\frac{\sum R_d^2 - \frac{\left(\sum R_d\right)^2}{n}}{n-1}} = \sqrt{\frac{385 - \frac{(-23)^2}{10}}{10-1}} = 6.075$$

TABLA 10.8: Tiempos de reacción ante emergencias de conducción simuladas relacionadas con el uso del teléfono celular.

Sin teléfono celular	Con teléfono celular
.318	.322
.301	.341
.384	.391
.290	.289
.411	.401
.371	.399
.371	.400
.333	.338

El rango con signo de la media es, entonces,

$$\bar{R}_d = \frac{\sum R_d}{n} = \frac{-23}{10} = -2.3$$

Si sustituimos estos valores en la ecuación anterior para t tenemos que,

$$t = \frac{-2.3}{\frac{6.075}{\sqrt{10}}} = -1.197$$

El Apéndice G indica que para $n = 10$, los valores críticos para una prueba de rangos con signo de Wilcoxon de dos colas realizada con $\alpha = .05$ son 2.424 y –2.424. Como la t obtenida de –1.197 está entre estos dos valores, no se rechaza la hipótesis nula. ■

EJEMPLO 10.16

Los investigadores están interesados en determinar el efecto del uso del teléfono celular en los tiempos de reacción ante emergencias de conducción simuladas. Con este fin, ocho sujetos en simuladores de manejo "conducen" durante de una hora. En un punto elegido al azar a esa hora, el sujeto experimenta una emergencia simulada que demanda un frenado inmediato para evitar una colisión. Cada sujeto se somete al simulacro de manejo dos veces, una mientras habla por teléfono y otra sin teléfono. El orden en el que se dan las pruebas es azaroso. El tiempo de reacción se define como tiempo transcurrido, medido en fracciones de segundo, entre la introducción de la emergencia y la aplicación del freno. Estos tiempos se dan en la tabla 10.8.

Utilice estos tiempos de reacción para realizar una prueba de una cola de rangos con signos de Wilcoxon en la que la alternativa indique un incremento en el tiempo de reacción cuando se utiliza el teléfono celular. Realice esta prueba con $\alpha = .05$.

Solución A continuación se muestran las puntuaciones de diferencia, es decir, los tiempos de reacción a la hora de usar el teléfono celular menos los tiempos de reacción sin teléfonos celulares, los valores absolutos de las puntuaciones de diferencia, los rangos de los valores absolutos de las puntuaciones de diferencia y los rangos con signo de los valores absolutos de las puntuaciones de diferencia.

Diferencia	Diferencia absoluta	Rango	Rango con signo
.004	.004	2	2
.040	.040	8	8
.007	.007	4	4
−.001	.001	1	−1
−.010	.010	5	−5
.028	.028	6	6
.029	.029	7	7
.005	.005	3	3

La t obtenida es el estadístico t de muestras apareadas calculado sobre los rangos con signo. Para facilitar este cálculo tenemos que la suma de los rangos con signo y la suma de los cuadrados de los rangos con signo son, respectivamente

$$\sum R_d = (2)+(8)+(4)+(-1)+(-5)+(6)+(7)+(3) = 24$$

y

$$\sum R_d^2 = (2)^2+(8)^2+(4)^2+(-1)^2+(-5)^2+(6)^2+(7)^2+(3) = 204$$

Utilizando estos resultados,

$$s_{Rd} = \sqrt{\frac{\sum R_d^2 - \frac{\left(\sum R_d\right)^2}{n}}{n-1}} = \sqrt{\frac{204 - \frac{(24)^2}{8}}{8-1}} = 4.342$$

y

$$\bar{R}_d = \frac{\sum R_d}{n} = \frac{24}{8} = 3.0$$

La t obtenida es, entonces,

$$t = \frac{\bar{R}_d}{\frac{s_{Rd}}{\sqrt{n}}} = \frac{3.0}{\frac{4.342}{\sqrt{8}}} = 1.954$$

El Apéndice G indica que para $n = 8$, el valor crítico para una prueba de una cola de rangos con signo de Wilcoxon realizada con $\alpha = .05$ es 2.225. Como la t obtenida de 1.954 es menor que este valor, no se rechaza la hipótesis nula. No pudimos, por consiguiente, demostrar que el uso del teléfono celular incrementa el tiempo de reacción en un accidente automovilístico. ■

EJEMPLO 10.17

Dado $n = 120$ y t obtenida de 2.626, realice una prueba de dos colas de rangos con signo de Wilcoxon de la hipótesis nula $H_0 : \mu_d = 0$ con $\alpha = .10$.

Solución Como $n > 90$, la t obtenida se refiere a una tabla t con $n - 1$ grados de libertad. La referencia al Apéndice B con $120 - 1 = 119$ grados de libertad para $\alpha = .10$ nos da valores críticos de dos colas de −1.658 y 1.658. Como la t obtenida de 2.626 es mayor que la t crítica de 1.658, se rechaza la hipótesis nula.

Suposiciones

Existen dos suposiciones fundamentales en la prueba de rangos con signo de Wilcoxon. La primera es que todos los valores de d y, de ahí, todos los rangos con signo son independientes. (Consulte las suposiciones subyacentes en la prueba de permutación de muestras apareadas de la página 365 para más detalles). La segunda suposición requiere que no haya ningún valor duplicado entre los valores absolutos de d. Considere el siguiente conjunto de datos.

Diferencia	Diferencia absoluta	Rango	Rango con signo
2	2	2.5	2.5
4	4	6.0	6.0
7	7	8.0	8.0
1	1	1.0	1.0
4	4	6.0	6.0
–4	4	6.0	–6.0
–2	2	2.5	–2.5
–8	8	9.0	–9.0
3	3	4.0	4.0
9	9	10.0	10.0

Note que para las diferencias absolutas existen dos valores de 2 y tres de 4. Ésta es la violación del supuesto. El problema surge cuando intentamos remplazar estos valores absolutos con rangos. Cuando se sustituyen rangos para $|d|$, el 1 sería remplazado con el rango 1, pero, ¿qué podemos hacer con los dos valores de 2? Existen diversas estrategias, pero la más común asigna el promedio de los rangos que hubieran sido asignados si no hubieran existido vínculos. Por lo general, la segunda puntuación más pequeña recibiría un rango de 2, en tanto que la tercera más pequeña recibiría el rango 3. La estrategia del **rango promedio** asigna el promedio del rango 2 y 3 o 2.5 a las dos puntuaciones de 2. El cuarto valor absoluto más pequeño es 3, que recibe el rango 4. De nuevo, nos enfrentamos a valores ligados de cuatro, ya que existen tres de estos valores. Si no hubiera habido vínculos, el quinto, sexto y séptimo valores más pequeños recibirían los rangos 5, 6 y 7, por lo que las observaciones ligadas recibirían el promedio de los rangos 5, 6 y 7 o 6.0. Los valores absolutos de siete, ocho y nueve no están ligados y, por lo tanto, reciben los rangos 8.0, 9.0 y 10.0.

Pero nuestro problema no está resuelto. Los valores críticos para $n = 10$ en el Apéndice G se obtuvieron para los rangos del 1 al 10, no para los rangos 1, 2.5, 2.5, 4, 6, 6, 6, 7, 8, 9 y 10. Las siguientes son estrategias que puede usar cuando se encuentre con observaciones ligadas.[11]

1. Conserve las observaciones originales y aplique la prueba de permutación, tal y como se describió previamente (Véase la página 361).[12]
2. Utilice el método del rango promedio descrito antes y genere la distribución muestral de permutación del estadístico t calculado sobre estos rangos. (Véase la nota al pie de la página 12).
3. Utilice el método del rango promedio descrito antes y refiera la t obtenida a una tabla t usando $n - 1$ grados de libertad. Este método generalmente produce buenos resultados si

[11] En los textos no paramétricos se describen otras estrategias. Vea por ejemplo [7].
[12] Esto se puede llevar a cabo utilizando software comercial.

n es lo suficientemente grande (mayor que 30) y la proporción de observaciones ligadas es pequeña (20% o menos).

Antes dijimos que, desde el punto de vista técnico, la hipótesis probada mediante el estadístico de rangos con signo sostiene que la población de puntuaciones de diferencia es simétrica respecto de cero, mientras que la alternativa afirma que éste no es el caso. Si suponemos que la población de puntuaciones de diferencia es simétrica, entonces, la hipótesis nula probada es la misma que la probada mediante la prueba *t* de permutación para muestras apareadas, es decir, $H_0 : \mu_d = 0$. Pero, ¿qué tan realista es la suposición de que la población de puntuaciones de diferencia es simétrica?

Cuando las poblaciones "pre" y "post" son idénticas, como sería el caso en la ausencia de un efecto del tratamiento, la población con puntuaciones de diferencia es simétrica *sin importar la forma de las poblaciones pre y post.* Así, las poblaciones pre y post podrían ser sesgadas de manera radical, pero mientras tengan la misma forma, la población de puntuaciones de diferencia será simétrica. Se deduce que la hipótesis nula probada será $H_0 : \mu_d = 0$. En situaciones en las que existen razones para creer que las poblaciones pre y post no tienen la misma forma, probablemente esto no sea cierto. ■

10.3.3 Dos muestras independientes

En el capítulo 6, en la página 219, usted aprendió a realiza una prueba *t* de muestras independientes mediante la ecuación 6.1 para probar la hipótesis nula $H_0 : \mu_1 = \mu_2$, donde μ_1 y μ_2 son las medias de las poblaciones de las que se extrajeron las dos muestras. Uno de las suposiciones en esta prueba es que las muestras provinieron de poblaciones con distribución normal. Ahora usted aprenderá a realiza pruebas de muestras independientes que no requieren este supuesto restrictivo. Se demostrarán dos pruebas. La primera utiliza observaciones originales; la segunda, conocida como prueba de Wilcoxon para muestras independientes,[13] utiliza rangos en vez de las observaciones originales. Se hace hincapié en la lógica de estas pruebas, así como en la relación entre las dos.

La prueba *t* de permutación para muestras independientes

Razonamiento

Suponga que los siguientes resultados se obtienen de un estudio en el que se ha asignado al azar a seis individuos a dos grupos.

Grupo 1	Grupo 2
9	0
12	4
13	6

Mediante la ecuación 6.1 de la página 219, el estadístico *t* de muestras independientes para este dato es 3.748. Si el investigador desea probar la hipótesis nula $H_0 : \mu_1 = \mu_2$, pero quiere evitar el supuesto de normalidad de la población se puede realizar una prueba de permutación basada en la siguiente lógica.

Puesto que los seis sujetos fueron asignados al azar a los dos grupos, la puntuación de cualquier sujeto en particular tiene exactamente las mismas probabilidades de aparecer en

[13] También conocida como la prueba U de Mann-Whitney.

Mitad inferior de la distribución muestral

−3.748 −1.765 −1.200 −.983 −.791 −.615 −.452 −.452 −.147 .000

Mitad superior de la distribución muestral

.000 .147 .452 .452 .615 .791 .983 1.200 1.765 3.748

t obtenida

FIGURA 10.5: Distribución muestral de permutación del estadístico *t* de muestras independientes para un conjunto de datos en particular.

cualquiera de los dos grupos *si la hipótesis nula de no efectos por tratamiento es verdadera*. Por otra parte, en presencia de un efecto por tratamiento que tiende a aumentar o disminuir las puntuaciones de los miembros de un grupo, las puntuaciones de ese grupo tenderán a ser más elevadas o más bajas que las del otro grupo.

Así, cuando la hipótesis nula de no efectos por tratamiento es verdadera, las puntuaciones de los sujetos se distribuyen al azar entre los dos grupos con el resultado de que ningún grupo tiende a manifestar puntuaciones significativamente más elevadas o más bajas que el otro. Se deduce que un estadístico *t* calculado sobre estos datos generalmente tomaría un valor cercano a cero. Cuando un efecto por tratamiento está presente y, por lo tanto, las puntuaciones de un grupo tienden a ser más elevadas que las del otro, el estadístico *t* tiende a alejarse del cero.

Como se enunció antes, si no existen efectos por tratamiento, la puntuación de cualquier sujeto dado tiene con exactitud las mismas probabilidades de aparecer en un grupo o en otro. Así, si la hipótesis nula es verdadera, el orden de los datos observados sólo es una de las posibles disposiciones implicadas por el proceso de asignación aleatoria. Pero, ¿de cuántas formas podrían haberse asignado los dos grupos?

Del apartado 10.2.2 y la ecuación 10.2 de la página 347 vemos que las seis observaciones podrían asignarse a los dos grupos en

$$C_{n2}^{n1} = \frac{n!}{n_1!n_2!} = \frac{6!}{3!3!} = 20$$

formas diferentes. Las 20 combinaciones se presentan en la tabla 10.9 de la siguiente página.

Habiendo formado las 20 posibles disposiciones de los datos en dos grupos y calculando un estadístico *t* de muestras independientes para cada una, podemos ordenar los 20 estadísticos *t* en una distribución muestral de permutación, como se ilustra en la figura 10.5.

Si la hipótesis nula es verdadera, esperaríamos que la *t obtenida* (a partir de los datos del estudio) estuviera cerca de cero, ya que una hipótesis nula verdadera implica que la diferencia entre las medias de las dos poblaciones de las que se extrajeron las muestras es cero. Una segunda interpretación es que cualquier diferencia entre las medias de las muestras se debe al proceso de asignación aleatoria y no a cualquier efecto por tratamiento. Por otra parte, si $\mu_1 > \mu_2$, esperaríamos

TABLA 10.9: Todas las combinaciones posibles de seis observaciones y los estadísticos t asociados para un conjunto de puntuaciones en particular.

grupo 1	grupo 2
0	9
4	12
6	13

$t = -3.748$

grupo 1	grupo 2
0	6
4	12
9	13

$t = -1.765$

grupo 1	grupo 2
0	6
4	9
12	13

$t = -.983$

grupo 1	grupo 2
0	6
4	9
13	12

$t = -.791$

grupo 1	grupo 2
0	4
6	12
9	13

$t = -1.200$

grupo 1	grupo 2
0	4
6	9
12	13

$t = -.615$

grupo 1	grupo 2
0	4
6	9
13	12

$t = -.452$

grupo 1	grupo 2
0	4
9	6
12	13

$t = -.147$

grupo 1	grupo 2
0	4
9	6
13	12

$t = .000$

grupo 1	grupo 2
0	4
12	6
13	9

$t = -.452$

grupo 1	grupo 2
4	0
6	12
9	13

$t = -.452$

grupo 1	grupo 2
4	0
6	9
12	13

$t = .000$

grupo 1	grupo 2
4	0
6	9
13	12

$t = .147$

grupo 1	grupo 2
4	0
9	6
12	13

$t = .452$

grupo 1	grupo 2
4	0
9	6
13	12

$t = .615$

grupo 1	grupo 2
4	0
12	6
13	9

$t = 1.200$

grupo 1	grupo 2
6	0
9	4
12	13

$t = .791$

grupo 1	grupo 2
6	0
9	4
13	12

$t = .983$

grupo 1	grupo 2
6	0
12	4
13	9

$t = 1.765$

grupo 1	grupo 2
9	0
12	4
13	6

$t = -3.748$

que t tomara cierto valor en el extremo derecho de la distribución, ya que las puntuaciones en la muestra 1 tenderían a ser más grandes que las de la segunda muestra. Lo opuesto sería cierto cuando $\mu_1 < \mu_2$. En este caso, la t obtenida para los datos derivados del estudio es 3.748. ¿Qué probabilidad hay de obtener un valor tan extremo de t cuando la hipótesis nula es cierta? Es simplemente la proporción de t's en la distribución muestral que toman un valor así de extremo (lejos de cero) o más, que el obtenido de los datos. En este caso, sólo uno de los 20 posibles valores de t está tan lejos, o más del cero, que el obtenido de los datos: 3.748. Ahora podemos decir que de los 20 valores posibles de t que podrían haberse obtenido bajo una hipótesis nula verdadera, sólo uno es tan extremo como el observado de los datos. La probabilidad de alcanzar un valor tan extremo de t bajo una hipótesis nula verdadera es, entonces, 1/20 = .050. Éste es, de hecho, el valor-p para una prueba con la alternativa $\mu_1 > \mu_2$. El valor-p de dos colas sería $(2)\left(\frac{1}{20}\right) = .100$.

En general, el valor-p para una prueba con la alternativa de una cola $\mu_1 > \mu_2$ es la proporción de valores en la distribución muestral de permutación que son mayores que o iguales al observado de los datos. Cuando la alternativa de una cola es de la forma $\mu_1 < \mu_2$, el valor-p para esta alternativa de una cola es la proporción de valores en la distribución de permutación que son menores que o iguales al observado de los datos. El valor-p de dos colas para la alternativa $\mu_1 \neq \mu_2$ se calcula encontrando la proporción de observaciones en la cola superior (inferior) de la distribución que son mayores (menores) que o iguales a t obtenido y multiplicado este valor por 2. La proporción mayor que o igual a la t obtenida se multiplica por 2 cuando t es mayor que cero, mientras que la proporción menor que o igual a la t obtenida se multiplica por 2 cuando la t obtenida es menor que cero.

Realización de la prueba

Los pasos para realiza la prueba t de permutación para muestras independientes son los siguientes.

1. Calcule t para los datos obtenidos del estudio. Denomine este valor t.
2. Forme todas las posibles combinaciones de datos de dos grupos, como se ilustra para un conjunto de datos muestra en la página anterior.
3. Calcule el estadístico t de muestras independientes para cada combinación, como se ilustra en la página 376.
4. Ordene los valores de t obtenidos del paso 3, del menor al mayor. Ésta es la distribución muestral de permutación de t.
5. Para una prueba de una cola con la alternativa $H_A : \mu_1 > \mu_2$, calcule la proporción de valores en la distribución de permutación que son mayores que o iguales a la t obtenida. Éste es el valor-p para la prueba de una cola.
6. Para una prueba de una cola con la alternativa $H_A : \mu_1 < \mu_2$, calcule la proporción de valores en la distribución de permutación que son menores que o iguales a la t obtenida. Éste es el valor-p para una prueba de una cola.
7. Para la prueba de dos colas con la alternativa $H_A : \mu_1 \neq \mu_2$:

a) calcule la proporción de valores en la distribución de permutación que son mayores que o iguales a la t obtenida, si ésta es mayor que cero. Multiplique este valor por 2 para encontrar el valor-p de dos colas.

b) calcule la proporción de valores en la distribución de permutación que son menores que o iguales a la t obtenida, si ésta es menor que cero. Multiplique este valor por 2 para encontrar el valor-p de dos colas.

EJEMPLO 10.18

Utilice la distribución muestral de permutación del estadístico t de muestras independientes de la figura 10.5 para encontrar el valor-p para la siguiente prueba.

$$H_0 : \mu_1 = \mu_2$$
$$H_A : \mu_1 < \mu_2$$

suponiendo que la t obtenida es -1.200.

Solución Como la alternativa es de la forma $\mu_1 < \mu_2$, debemos realizar una prueba de una cola, siendo el valor-p la proporción de los valores en la distribución muestral que son menores que o iguales al obtenido de los datos. En este caso los valores de t que satisfacen el criterio de ser menores que o iguales a la t obtenida de -1.200 son -3.748, -1.765 y -1.200, por lo que $p = \frac{3}{20} = .150$. Si α fuera .05, no rechazaríamos la hipótesis nula. ■

EJEMPLO 10.19

Dado $n_1 = n_2 = 5$, ¿cuántos valores de t conformarían la distribución muestral de permutación para la prueba t de muestras independientes?

Solución

$$C_5^5 = \frac{n!}{n_1!n_2!} = \frac{10!}{5!5!} = 252$$ ■

Suposiciones

La suposición fundamental que subyace en la prueba t de permutación para muestras independientes es que todas las puntuaciones son independientes. Podría ocurrir una violación a este supuesto, por ejemplo, si en un ensayo clínico, el mismo sujeto fuera evaluado en dos puntos en el tiempo e introduciendo las puntuaciones resultantes del análisis. Así, este sujeto tendría dos puntuaciones en el conjunto de datos. Si las dos puntuaciones estuvieran relacionadas de alguna manera, podría ocurrir una violación.

Como segundo ejemplo, podría ocurrir un incumplimiento si las puntuaciones para los hermanos se incluyeran en un análisis. Si el estudio tratara sobre pérdida de peso y existiera un componente genético relacionado con la pérdida de peso, entonces la cantidad de peso perdido por los hermanos podría relacionarse. No podemos contar con que esta prueba sea robusta ante un incumplimiento de la suposición de independencia.

Prueba de Wilcoxon para muestras independientes (prueba de suma de rangos). Esta prueba[14] es similar a la prueba t de permutación para muestras independientes, excepto que las puntuaciones se convierten en rangos antes de calcular la t obtenida y generar la distribución muestral de permutación.

[14] También conocida como la prueba U de Mann-Whitney.

Razonamiento

Para realizar la prueba t de permutación para muestras independientes para ocho observaciones en cada uno de dos grupos, tendríamos que encontrar $C_8^8 = 12{,}870$ combinaciones de datos y calcular un estadístico t de muestras independientes para cada una. El número de cálculos aumenta con rapidez a medida que aumenta el tamaño de la muestra.

Esta descomunal tarea, y otras consideraciones, animaron a Frank Wilcoxon [49] a idear una prueba de permutación de muestras independientes basada en rangos. Una ventaja de una estrategia así es que una vez que la distribución de permutación para el estadístico de prueba es generada para tamaños de muestra dados (n_1 y n_2), se puede usar la distribución para cualquier muestra del mismo tamaño, sin importar los valores de los datos originales. Esto se deduce del hecho de que los datos originales se remplazan con rangos para los que ya se ha construido la distribución.

La conversión de los datos originales a rangos se logra de la siguiente manera. La puntuación más pequeña en el conjunto de datos recibe el rango 1, la segunda más pequeña el rango 2 y así sucesivamente; la observación más grande recibe el rango $n_1 + n_2$. Esto es, si hay tres observaciones en cada muestra, el valor más pequeño recibe el rango 1 y el valor más grande recibe el rango 6. Note que los rangos se asignan sin importar a qué grupo pertenecen. Por ejemplo, suponga que deseamos convertir los datos utilizados previamente junto con la prueba t de permutación para muestras independientes en rangos. Estos datos se repiten aquí para su comodidad.

Tratamiento	Control
9	0
12	4
13	6

Después de remplazar las observaciones originales con los rangos, el resultado es como sigue.

Tratamiento	Control
4	1
5	2
6	3

Si aplicamos la ecuación 6.1 de la página 219 a los rangos, tenemos un estadístico t de muestras independientes de 3.674.

Para generar la distribución muestral de permutación para esta prueba de rangos, debemos generar los enteros del 1 al 6 y formar todas las combinaciones posibles, en las cuales 3 rangos se asignan a cada uno de los dos grupos. Se debe calcular un estadístico t para cada una de estas combinaciones. Esto se hace en la tabla 10.10 de la siguiente página. Los valores t de cada combinación conforman una distribución muestral de permutación, como se ilustra en la figura 10.6 de la página 381.

En sentido estricto, la hipótesis que se prueba mediante este estadístico sostiene que las dos muestras se sustrajeron de poblaciones *idénticas*, mientras que la alternativa afirma que éste no es el caso. Así, un tratamiento que aumenta (disminuye) la varianza, la media, la mediana, el sesgo o cualquier otro aspecto del grupo tratado causaría que la hipótesis nula fuera falsa. Sin embargo, esta prueba no es igualmente potente ante todas estas condiciones. En particular es

TABLA 10.10: Todas las combinaciones posibles de los rangos 1 al 6 y los estadísticos *t* asociados.

grupo 1	grupo 2
1	4
2	5
3	6

$t = -3.674$

grupo 1	grupo 2
1	3
2	5
4	6

$t = -1.871$

grupo 1	grupo 2
1	3
2	4
5	6

$t = -1.118$

grupo 1	grupo 2
1	3
2	4
6	5

$t = -.612$

grupo 1	grupo 2
1	2
3	5
4	6

$t = -1.118$

grupo 1	grupo 2
1	2
3	4
5	6

$t = -.612$

grupo 1	grupo 2
1	2
3	4
6	5

$t = -.196$

grupo 1	grupo 2
1	2
4	3
5	6

$t = -.196$

grupo 1	grupo 2
1	2
4	3
6	5

$t = .196$

grupo 1	grupo 2
1	2
5	3
6	4

$t = .612$

grupo 1	grupo 2
2	1
3	5
4	6

$t = -.612$

grupo 1	grupo 2
2	1
3	4
5	6

$t = -.196$

grupo 1	grupo 2
2	1
3	4
6	5

$t = .196$

grupo 1	grupo 2
2	1
4	3
5	6

$t = .196$

grupo 1	grupo 2
2	1
4	3
6	5

$t = .612$

grupo 1	grupo 2
2	1
5	3
6	4

$t = 1.118$

grupo 1	grupo 2
3	1
4	2
5	6

$t = .612$

grupo 1	grupo 2
3	1
4	2
6	5

$t = 1.118$

grupo 1	grupo 2
3	1
5	2
6	4

$t = 1.871$

grupo 1	grupo 2
4	1
5	2
6	3

$t = 3.674$

Mitad inferior de la distribución muestral

−3.674 −1.871 −1.118 −1.118 −.612 −.612 −.612 −.196 −.196 −.196

Mitad superior de la distribución muestral

.196 .196 .196 .612 .612 .612 1.118 1.118 1.871 3.674

t obtenida

FIGURA 10.6: Distribución muestral de permutación de estadísticos r para muestras independientes basadas en rangos para $n_1 = n_2 = 3$.

potente para alternativas desplazadas, como se describe en la página 275. James V. Bradley [4] enuncia lo siguiente a este respecto.

> Aquellos que no tienen experiencia en la experimentación práctica pueden desmoralizarse por la imprecisión de la hipótesis alternativa, y preferirán un enunciado definitivo de medias desiguales a un enunciado de no identidad que con probabilidad incluye desigualdad de medias. No obstante, si bien en teoría es posible que las poblaciones de tratamiento no sean idénticas aunque tengan medias exactamente iguales, y si bien es absurdo recrear estas poblaciones en nuestra cabeza, también es increíblemente difícil crearlas en un laboratorio. Eso es, en la mayoría de las áreas de investigación, si se cambia la variable manipulada (tratamiento o condición), es casi imposible inducir un cambio en *cualquiera* de los aspectos (media, mediana, moda, varianza, amplitud intercuartil, décimo percentil, forma, etcétera) o combinaciones de aspectos de la distribución de la población de la variable medida sin producir *algún* cambio en *todos* los aspectos. La situación buscada tampoco es fácil de encontrar. Así, por resultar práctico en la mayoría de las áreas de investigación, un veredicto de la prueba estadística de poblaciones no idénticas equivale a un veredicto de medias desiguales, varianzas desiguales, etcétera.

Siguiendo el tono de este argumento, una prueba t de rangos significativa (o prueba de Wilcoxon para muestras independientes) se puede interpretar como indicador de una diferencia entre las medias de población. Sin embargo, debemos tener presente que la hipótesis nula engloba un enunciado más amplio que uno enfocado sólo en una diferencia entre medias.

Si deseamos probar la hipótesis nula de poblaciones idénticas (o el efecto de la ausencia de tratamiento de cualquier clase) para los datos muestrales presentados antes en los que la t obtenida fue 3.674, observamos que sólo un valor en el muestreo de permutación es mayor que o igual a t obtenida (3.674), por lo que el valor-p para una prueba de una cola es $\frac{1}{20} = .05$. El valor-p para una alternativa de dos colas sería $(2)\left(\frac{1}{20}\right) = .100$. Si suponemos que este resultado es significativo, por el argumento presentado antes, podemos interpretar que este resultado indica un efecto por tratamiento que (entre otras cosas) da como resultado una respuesta media más elevada para un grupo que para el otro.

EJEMPLO 10.20

Utilice los datos presentados aquí para realizar una prueba de una cola de Wilcoxon para muestras independientes con $\alpha = .20$ para demostrar que la respuesta media para el grupo de tratamiento es menor que la del grupo de control. Interprete el resultado.

Tratamiento	Control
.9	.0
1.2	2.4
.3	6.0

Solución Remplazar estas observaciones con rangos produce

Tratamiento	Control
3	1
4	5
2	6

Si aplicamos la ecuación 6.1 de la página 219 a los rangos, tenemos un estadístico t de muestras independientes de $-.612$. La referencia a la figura 10.6 de la página anterior indica que siete valores de la distribución muestral de permutación son menores que o iguales a la t obtenida, por lo que el valor-p de una cola es $\frac{7}{20} = .35$ y no se rechaza la hipótesis nula. En términos prácticos, esto significa que no pudimos demostrar el efecto del tratamiento. ■

Remplazar los datos originales con rangos nos permite construir tablas de valores críticos para esta prueba. Para lograrlo, sólo necesitamos generar la distribución muestral de permutación para los conjuntos de rangos para los diversos tamaños de muestras y localizar los valores de permutación para que α o $\alpha/2$ de los valores de permutación sean mayores que o iguales al valor identificado. Por ejemplo, si examinamos la figura 10.6 de la página anterior, indica que un valor t de 3.674 tiene un valor-p (de una cola) asociado con $1/20 = .050$ mientras que un valor de 1.871 tiene una valor asociado de $2/20 = .100$, por lo que no hay valor crítico para una prueba de una cola efectuada con $\alpha = .005$, .010 o .025. De igual manera no existe valor crítico para una prueba de dos colas al nivel .010, .020 o .050. Sin embargo, el valor crítico de 3.674 se puede usar para una prueba de dos colas al nivel .050 o una prueba de dos colas al .100. Conforme van aumentando los tamaños de las muestras y la distribución muestral de permutación se va conformando de más valores distintos de t, las pruebas se pueden realizar a más niveles de α. Como la distribución muestral de permutación es simétrica, los valores críticos de la cola inferior se obtienen como negativos de los valores de la tabla.

El Apéndice H da los valores críticos para $n_1 = n_2 = 3$ a 60 para el estadístico t de muestras independientes de rangos o la prueba de Wilcoxon para muestras independientes. Esta tabla provee los valores críticos sólo para las situaciones en las que los tamaños de la muestras para los dos grupos son iguales.[15] Para las pruebas de la cola superior usted sólo necesita comparar el

[15] En la actualidad se están construyendo tablas más amplias para los tamaños desiguales de las muestras.

estadístico t del rango obtenido contra el valor crítico apropiado. Si la t obtenida es mayor que o igual a la t crítica, se rechaza la hipótesis nula. Para las pruebas de la cola inferior, la t obtenida se compara con el negativo del valor tabulado. Si la t obtenida es menor que o igual a este valor, se rechaza la hipótesis nula. La hipótesis nula para una prueba de dos colas se rechaza si la t obtenida es mayor que o igual al valor de la tabla, o es menor que o igual al negativo del valor de la tabla. Observe que los valores críticos en esta tabla se refieren al tamaño de la muestra (n_1 o n_2) en vez de los grados de libertad, como se hace con las tablas t convencionales.

Para $n_1 = n_2$ mayor que 60, el estadístico t del rango obtenido se puede referir a una tabla t con $n_1 + n_2 - 2$ grados de libertad. El resultado es una muy buena aproximación al valor de la permutación cuando $n_1 = n_2 > 60$. Esta aproximación, por lo general, también es aceptable para tamaños desiguales de muestras, siempre y cuando los tamaños de las muestras sean lo suficientemente grandes, por ejemplo, n_1 y n_2 son mayores que 15. La aproximación es muy buena cuando ambos tamaños de la muestra son mayores que 60.

Realización de la prueba

Los pasos para realizar la prueba de Wilcoxon para muestras independientes son los siguientes.

1. Remplace las puntuaciones originales con sus rangos respectivos y dé a la puntuación más baja un rango de 1, a la segunda más baja un rango de 2, y así sucesivamente hasta que la puntuación más grande sea remplazada con un rango $n_1 + n_2$. La clasificación se lleva a cabo sin importar el grupo.

2. Use la ecuación 6.1 de la página 219 para calcular el estadístico t de muestras independientes para los rangos obtenidos en el paso 1. Denomine este valor t obtenida.

3. Para tamaños iguales de muestras donde $3 \leq n_1 = n_2 \leq 60$, refiera t obtenida al Apéndice H y

   a) para una prueba de la cola superior rechace la hipótesis nula si la t obtenida es mayor que o igual al valor de la tabla.

   b) para una prueba de la cola inferior rechace la hipótesis nula si la t obtenida es menor que o igual al negativo del valor de la tabla.

   c) para una prueba de dos colas rechace la hipótesis nula si la t obtenida es mayor que o igual al valor de la tabla *o* si la t obtenida es menor que o igual al negativo del valor de la tabla.

4. Para $n_1 = n_2 > 60$, refiera t obtenida al Apéndice B con $n_1 + n_2 - 2$ grados de libertad.

5. Para tamaños de muestras desiguales, la t obtenida se puede referir a las tablas construidas para tamaños desiguales de muestras (creadas por los autores) o, si n_1 y n_2 son mayores que 15, la referencia puede ser a una tabla t con $n_1 + n_2 - 2$ grados de libertad.

EJEMPLO 10.21

Realice una prueba de Wilcoxon de dos colas para muestras independientes con $\alpha = .05$ sobre los datos (ficticios) de presión arterial de la tabla 10.11 de la siguiente página.

Solución Comenzamos por remplazar las medidas de presión arterial con sus respectivos rangos y formando el resultado para análisis como se ilustra en la tabla 10.12.

TABLA 10.11: Presiones arteriales de pacientes hipertensos después del tratamiento con medicamentos y placebos.

Medicamento	Placebo
129	146
131	130
128	127
117	160
144	165
119	133
138	148
142	150
110	171
140	141

TABLA 10.12: Rangos de las presiones arteriales ordenados para análisis mediante la prueba *t* de muestras independientes.

	Medicamento		Placebo	
	R_1	R_1^2	R_2	R_2^2
	6	36	15	225
	8	64	7	49
	5	25	4	16
	2	4	18	324
	14	196	19	361
	3	9	9	81
	10	100	16	256
	13	169	17	289
	1	1	20	400
	11	121	12	144
Σ	73	725	137	2145

Utilizando las cantidades de la tabla 10.12 calculamos el estadístico t de muestras independientes como se demostró en la página 219.[16]

El rango medio para el primer grupo es

$$\bar{R}_1 = \frac{\sum R_1}{n_1} = \frac{73}{10} = 7.3$$

[16] Utilizamos el símbolo R en vez de X en los cálculos como un recordatorio de que estamos analizando rangos y no las observaciones originales.

mientras que para el segundo es

$$\bar{R}_2 = \frac{\sum R_1}{n_1} = \frac{73}{10} = 7.3$$

La estimación combinada de la varianza (mediante la ecuación 6.2 de la página 220) es

$$\begin{aligned} s^2_{P_R} &= \frac{\left(\sum R_1^2 - \frac{\left(\sum R_1\right)^2}{n_1}\right) + \left(\sum R_2^2 - \frac{\left(\sum R_1\right)^2}{n_1}\right)}{n_1 + n_2 - 2} \\ &= \frac{\left(725 - \frac{(73)^2}{10}\right) + \left(2145 - \frac{(137)^2}{10}\right)}{10 + 10 - 2} \\ &= 25.567 \end{aligned}$$

Mediante la ecuación 6.1 de la página 219, la t obtenida, con las sustituciones apropiadas de R por x, es entonces

$$\begin{aligned} t &= \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{s^2_{P_R}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \\ &= \frac{7.3 - 13.7}{\sqrt{25.567\left(\frac{1}{10} + \frac{1}{10}\right)}} \\ &= -2.830 \end{aligned}$$

La referencia al Apéndice H indica que para $n_1 = n_2 = 10$, la t crítica para una prueba de Wilcoxon de dos colas para muestras independientes realizada con $\alpha = .05$ es -2.248 y 2.248. Como la t obtenida de -2.830 es menor que la t crítica de -2.248, se rechaza la hipótesis nula. Técnicamente, esto significa que las muestras no provinieron de la misma población. Por argumentos anteriores, podemos concluir que los pacientes tratados con el medicamento registraron presiones arteriales más bajas que los pacientes tratados con el placebo. ■

EJEMPLO 10.22

Los investigadores están interesados en determinar el efecto del uso del teléfono celular sobre los tiempos de reacción ante emergencias de conducción simuladas. Con este fin, 16 sujetos son asignados al azar a uno de dos grupos. Ambos grupos se colocan ante simuladores de manejo donde "conducen" por un periodo de una hora. En un punto al azar durante esa hora, cada sujeto experimenta una emergencia simulada que demanda un frenado inmediato para evitar un choque. Durante los simulacros, los sujetos del primer grupo deben mantener una conversación por teléfono celular con uno de los investigadores, en tanto que los sujetos del segundo grupo simplemente "conducen" sin la interferencia del teléfono celular. El tiempo de reacción se define como tiempo transcurrido, medido en fracciones de segundo, entre la introducción de la emergencia simulada y la aplicación del freno. Estos tiempos para los dos grupos se dan en la tabla 10.13.

Utilice estos tiempos de reacción para realizar una prueba de Wilcoxon de una cola para muestras independientes en que la alternativa indique un aumento en el tiempo de reacción cuando se usa el teléfono celular. Realice la prueba con $\alpha = .05$.

TABLA 10.13: Tiempos de reacción de usuarios y no usuarios de teléfono celular ante una emergencia de conducción simulada.

Teléfono	Sin teléfono
.322	.297
.295	.298
.399	.211
.215	.199
.444	.370
.377	.299
.300	.236
.331	.290

Solución Comenzamos por remplazar las medidas de tiempo de reacción con sus respectivos rangos y formando el resultado para análisis como se ilustra en la tabla 10.14.

Utilizando las cantidades de la tabla 10.14 calculamos el estadístico t de muestras independientes como se demostró en la página 219.[17]

El rango medio para el grupo uso de teléfono celular es

$$\bar{R}_1 = \frac{\sum R_1}{n_1} = \frac{87}{8} = 10.875$$

mientras que para el grupo sin teléfono celular es

$$\bar{R}_2 = \frac{\sum R_2}{n_2} = \frac{49}{8} = 6.125$$

La estimación combinada de la varianza (mediante la ecuación 6.2 de la página 220) es

$$s_{P_R}^2 = \frac{\left(\sum R_1^2 - \frac{\left(\sum R_1\right)^2}{n_1}\right) + \left(\sum R_2^2 - \frac{\left(\sum R_2\right)^2}{n_2}\right)}{n_1 + n_2 - 2}$$

$$= \frac{\left(1087 - \frac{(87)^2}{8}\right) + \left(409 - \frac{(49)^2}{8}\right)}{8 + 8 - 2}$$

$$= 17.839$$

[17] Utilizamos el símbolo R en vez de X en los cálculos como un recordatorio de que estamos analizando rangos y no las observaciones originales.

TABLA 10.14: Rangos de tiempos de reacción ordenados para análisis mediante la prueba *t* de muestras independientes.

	Teléfono		Sin teléfono	
	R_1	R_1^2	R_2	R_2^2
	11	121	7	49
	6	36	8	64
	15	225	2	4
	3	9	1	1
	16	256	13	169
	14	196	9	81
	10	100	4	16
	12	144	5	25
Σ	87	1087	49	409

Mediante la ecuación 6.1 de la página 219, la t obtenida, con las sustituciones apropiadas de R por x, es

$$\begin{aligned} t &= \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{s_{P_R}^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}} \\ &= \frac{10.875 - 6.125}{\sqrt{17.839\left(\dfrac{1}{8} + \dfrac{1}{8}\right)}} \\ &= 2.249 \end{aligned}$$

La referencia al Apéndice H indica que para $n_1 = n_2 = 8$, la t crítica para una prueba de Wilcoxon de una cola para muestras independientes realizada con $\alpha = .05$ es 1.944. Como la t obtenida de 2.249 es mayor que la t crítica de 1.944, se rechaza la hipótesis nula. Por los argumentos anteriores, podemos concluir que los sujetos que usaron teléfonos celulares mostraron tiempos de reacción mayores que los sujetos que no usaron teléfonos celulares. ■

EJEMPLO 10.23

Dados $n_1 = n_2 = 63$ y una t obtenida de -2.626, realice una prueba de Wilcoxon de dos colas para muestras independientes con $\alpha = .05$.

Solución Como $n_1 = n_2 > 60$, la t obtenida se refiere a una tabla t con $n_1 + n_2 - 2$ grados de libertad. La referencia al Apéndice B con $63 + 63 - 2 = 124$ grados de libertad con $\alpha = .05$ da valores críticos de dos colas de -1.979 y 1.979. Como la t obtenida de -2.626 es menor que la t crítica de -1.979, se rechaza la hipótesis nula. ■

TABLA 10.15: Ejemplo de observaciones vinculadas a un diseño de dos grupos.

Grupo 1	Grupo 2
14	76
3	44
17	73
22	76
3	90
0	11
76	49
23	53

EJEMPLO 10.24

Dados $n_1 = 18$ y $n_2 = 24$ y una t obtenida de 1.334, realice una prueba de Wilcoxon de una cola para muestras independientes con $\alpha = .05$ para demostrar que los valores del grupo 1 son mayores que los valores del grupo 2.

Solución Como $n_1 \neq n_2$ y $n_1 > 15$ y $n_2 > 15$, podemos realizar una prueba aproximada refiriendo la t obtenida a una distribución t con $n_1 + n_2 - 2 = 18 + 24 - 2 = 40$ grados de libertad. El Apéndice B indica que la t crítica para una prueba de una cola realizada con 40 grados de libertad y $\alpha = .05$ es 1.684. Como la t obtenida de 1.334 es menor que este valor, no se rechaza la hipótesis nula. ■

Suposiciones

Existen dos suposiciones subyacentes en la prueba de Wilcoxon para muestras independientes. La primera es que todos los valores en las dos muestras y, de ahí, todos los rangos, son independientes. (Véase suposiciones subyacentes en la prueba de permutación de muestras independientes de la página 378 para más detalles). La segunda suposición requiere que no existan valores duplicados entre los datos de las dos muestras. Considere los datos de la tabla 10.15.

Observe que hay dos valores de 3 y tres de 76. Ésta es la violación al supuesto. El problema se da cuando intentamos remplazar estos valores de los datos originales con rangos. Cuando se sustituyen rangos, el valor cero sería remplazado con el rango 1, pero, ¿qué podemos hacer con los dos valores de 3? Existen diversas estrategias, pero la más común asigna el promedio de los rangos que hubieran sido asignados de no haber existido un vínculo. Por lo general, la segunda puntuación más baja recibiría un rango de 2, mientras que la tercera más baja recibiría un rango de 3. La estrategia "rango promedio" asigna el promedio de los rangos 2 y 3 o 2.5 a las dos puntuaciones de 3. El cuarto valor más pequeño es 11, que recibe el rango 4. Continuamos con la clasificación dando el rango 5 al 14, el rango 6 al 17, el rango 7 al 22, el rango 8 al 23, el rango 9 al 44, el rango 10 al 49, el rango 11 al 53 y el rango 12 al 73. El siguiente valor más alto es 76, y hay tres de éste. Como los siguientes tres valores recibirían los rangos 13, 14 y 15 si no hubiera ningún vínculo, asignamos el promedio de estos rangos, es decir, 14, a los tres valores de 76. La puntuación 90 recibe el rango 16. El conjunto completo de rangos se muestra en la tabla 10.16.

TABLA 10.16: Ejemplo de rangos asignados a observaciones ligadas en un diseño de dos grupos.

Grupo 1	Grupo 2
5	14
2.5	9
6	12
7	14
2.5	16
1	4
14	10
8	11

Ahora, la pregunta que surge es cómo este dato ligado podría formar la base de una prueba de hipótesis. Después de todo, los valores críticos para $n_1 = n_2 = 8$ en el Apéndice H se obtuvieron para los rangos 1 al 16, no para los rangos 1, 2.5, 2.5, 4, y así sucesivamente. Las siguientes son estrategias que pueden usarse cuando se encuentren observaciones ligadas.[18]

1. Conserve las observaciones originales y aplique la prueba de permutación descrita previamente (Véase la página 374).[19]

2. Utilice el método del rango promedio descrito antes y genere la distribución muestral de permutación de estos rangos. (Véase la nota al pie de página 19).

3. Utilice el método del rango promedio descrito antes y refiera la t obtenida a una tabla t usando $n_1 + n_2 - 2$ grados de libertad. Este método por lo común produce buenos resultados si $n_1 + n_2$ es lo bastante grande (mayor que 30) y la proporción de observaciones ligadas es pequeña (20% o menos).

10.3.4 Muestras independientes múltiples

En el capítulo 7 usted aprendió a realizar una prueba F de un ANOVA de un factor para probar la hipótesis nula $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$, donde $\mu_1, \mu_2 \cdots \mu_k$ son las medias de las poblaciones de las cuales se extrajeron las pruebas. Uno de los supuestos subyacentes en esta prueba es que las muestras provinieron de poblaciones normalmente distribuidas. En esta sección usted aprenderá a realizar pruebas ANOVA de muestras múltiples que no requieren este supuesto restrictivo. Se demostrarán dos pruebas. La primera usa observaciones originales y la segunda, conocida como la prueba Kruskal-Wallis, usa rangos en vez de observaciones. Se enfatiza la lógica de estas pruebas, así como la relación entre las dos.

[18] En los textos no paramétricos se describen otras estrategias. Véase por ejemplo [7].
[19] Esto se puede llevar a cabo utilizando software comercial.

La prueba F de permutación para un ANOVA de un factor.

Razonamiento

Suponga que los siguientes resultados se obtienen de un estudio en el que seis sujetos tienen que ser asignados al azar a tres grupos de tratamiento.

Grupo 1	Grupo 2	Grupo 3
4	−1	7
9	0	12

Mediante la ecuación 7.2 de la página 264 el estadístico F para un ANOVA de un factor para este dato es 6.196. Si el investigador desea probar la hipótesis nula $H_0 : \mu_1 = \mu_2 = \mu_3$, pero quiere evitar la suposición de normalidad de la población, se puede llevar a cabo una prueba de permutación basada en la siguiente lógica.

Como se asignó a los seis sujetos al azar a los tres grupos, la puntuación de cualquiera de los sujetos tiene la misma probabilidad de aparecer en uno de los tres grupos que en cualquiera de los otros *si la hipótesis nula de efecto no diferencial del tratamiento es verdadera*. Por otra parte, ante la presencia de un efecto del tratamiento que tiende a incrementar o reducir las puntuaciones de los miembros de ciertos grupos, las puntuaciones de esos grupos tienden a ser más elevadas o más bajas que las de otros grupos.

Así, cuando la hipótesis nula de no efecto del tratamiento es verdadera, las puntuaciones de los sujetos se distribuyen al azar entre los tres grupos y el resultado es que ningún grupo tiende a mostrar puntuaciones significativamente más elevadas o más bajas que el otro. Se deduce que un estadístico F calculado sobre estos datos por lo general tomaría un valor cercano a 1. Cuando está presente un efecto del tratamiento, el estadístico F tiende a ser más grande que 1.[20]

Como se afirmó antes, si no hay un efecto por el tratamiento, la puntuación de cualquier sujeto dado tiene la misma probabilidad de aparecer en un grupo o en uno de los otros dos. Así, si la hipótesis nula es verdadera, la disposición de datos observada es sólo una de las disposiciones posibles implicadas por el proceso de asignación al azar. Pero, ¿de cuántas maneras podrían haberse asignado las puntuaciones a los tres grupos?

Siguiendo la lógica del apartado 10.2.2 y la ecuación 10.2 de la página 347 podemos deducir que las seis observaciones podrían asignarse a los tres grupos en

$$\frac{n!}{n_1!n_2!n_3!} = \frac{6!}{2!2!2!} = 90$$

maneras.

En general, n sujetos se pueden asignar a k grupos en

$$\boxed{\frac{n!}{n_1!n_2!\cdots n_k!}} \tag{10.3}$$

maneras. Como se observa, la ecuación 10.2 de la página 347 es sólo un caso especial de la forma más general provista por la expresión 10.3. No intentaremos mostrar las 90 disposiciones de datos para los datos considerados aquí, pero sí proveeremos seis ejemplos de tales disposiciones en la tabla 10.17.

[20] Véase la sección 7.2.5 de la página 272.

TABLA 10.17: Seis de las 90 disposiciones posibles de seis observaciones a tres grupos con los estadísticos *F* asociados para un conjunto de puntuaciones en particular.

grupo 1	grupo 2	grupo 3
4	–1	7
9	0	12

$F = 6.196$

grupo 1	grupo 2	grupo 3
–1	4	9
0	7	12

$F = 19.158$

grupo 1	grupo 2	grupo 3
–1	0	9
4	7	12

$F = 3.229$

grupo 1	grupo 2	grupo 3
0	–1	9
4	7	12

$F = 2.910$

grupo 1	grupo 2	grupo 3
0	–1	9
7	4	12

$F = 3.229$

grupo 1	grupo 2	grupo 3
–1	0	9
7	4	12

$F = 2.910$

12 observaciones superiores en la distribución muestral

6.196 6.196 6.196 6.196 6.196 6.196 19.158 19.158 19.158 19.158 19.158 19.158

↑

F obtenida

FIGURA 10.7: 12 observaciones superiores en la distribución muestral de permutación de los estadísticos *F* para un conjunto de datos en particular.

Habiendo formado todas las 90 posibles disposiciones de los datos en tres grupos y calculando un estadístico *F* para cada uno, podemos ordenar los 90 estadísticos *F* en una distribución muestral de permutación como se ilustra en la figura 10.7. Observe que, por razones de espacio, no presentamos los 90 estadísticos *F* sino solamente los 12 estadísticos superiores de la distribución.[21]

Si la hipótesis nula es verdadera, esperaríamos que la *F obtenida* (a partir de los datos recabados en el estudio) esté cerca de cero, ya que una hipótesis verdadera implica que el numerador y denominador del estadístico *F* estiman la varianza de las poblaciones de las que se extrajeron las muestras.[22] Una segunda interpretación es que cualquier variación entre las medias de las muestras se debe al proceso de asignación aleatoria y no a algún efecto del tratamiento. Por

[21] Usted recordará del capítulo 7 que la prueba F de un ANOVA de un factor es intrínsecamente unilateral.

[22] Véase el capítulo 7.

otra parte, si H_0 no es verdadera, esperaríamos que F tomara algún valor en la cola del extremo derecho de la distribución ya que la variación muestral entre (es decir, MS_b) sería mayor de lo que se esperaría si esta variación se debiera al proceso de asignación aleatoria solamente. En este caso, la F obtenida para los datos derivados del estudio es 6.196. ¿Qué probabilidad hay de obtener un valor tan extremo de F cuando la hipótesis nula es verdadera? Es sólo la proporción de F's en la distribución muestral que toman un valor tan grande como, o más grande que, el obtenido de los datos. En este caso, 12 de los 90 valores posibles de F son iguales o exceden el obtenido de los datos, es decir, 6.196. Ahora podemos decir que de los 90 valores posibles de F que hubieran podido obtenerse bajo una hipótesis nula verdadera, sólo 12 son tan extremos como el observado en la tabla. La probabilidad de alcanzar un valor tan extremo de F bajo una hipótesis nula verdadera es, entonces, 12/90 = .133. Éste es el valor-p para una prueba de la hipótesis nula $\mu_1 = \mu_2 = \mu_3$.

Realización de la prueba

Los pasos para realizar la prueba F de un ANOVA de un factor son los siguientes.

1. Calcule F para los datos obtenidos del estudio. Denomine a este valor F obtenida.
2. Forme todas las disposiciones posibles de los datos en los k grupos.
3. Calcule el estadístico F de un ANOVA de un factor para cada una de estas disposiciones.
4. Ordene los valores de F obtenidos del paso 3 del menor al mayor. Ésta es la distribución muestral de permutación de F.
5. Calcule la proporción de los valores en la distribución de permutación que son mayores que o iguales a la F obtenida. Éste es el valor-p para la prueba.

EJEMPLO 10.25

Utilice la distribución muestral de permutación de un estadístico F de un ANOVA de un factor mostrada en la figura 10.7 para encontrar el valor-p para la siguiente prueba.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

suponiendo que la F obtenida es 19.158 y $n_1 = n_2 = n_3 = 2$.

Solución El valor-p para la prueba es la proporción de estadísticos F en la distribución muestral de permutación que son iguales o exceden la F obtenida de 19.158. En este caso, los valores de F que satisfacen el criterio de ser mayores que o iguales a la F obtenida son los seis valores de 19.158, por lo que $p = \frac{6}{90} = .067$. Si α fuera .100 rechazaríamos la hipótesis nula. *Dados los datos en este estudio*, ¿podríamos rechazar H_0 con $\alpha = .050$? (*Pista:* ¡NO! Porque el valor-p más pequeño posible para estos datos es 0.067). ■

EJEMPLO 10.26

¿De cuántas maneras pueden asignarse 12 sujetos a cuatro grupos si se debe asignar a tres sujetos en cada uno? Suponga que se deben asignar dos sujetos al primer grupo, cuatro al segundo, cinco al tercero y uno al cuarto grupo?

Solución Para el caso en el que se deben asignar tres sujetos a cada uno de los cuatro grupos obtenemos, mediante la expresión 10.3

$$\frac{n!}{n_1!n_2!\cdots n_k!} = \frac{12!}{3!3!3!3!} = \frac{479,001,600}{1296} = 369,600$$

Para $n_1 = 2$, $n_2 = 4$, $n_3 = 5$ y $n_4 = 1$,

$$\frac{n!}{n_1!n_2!\cdots n_k!} = \frac{12!}{2!4!5!1!} = \frac{479,001,600}{5760} = 83,160$$ ■

Suposiciones

Las suposiciones subyacentes para la prueba F de un ANOVA de un factor son las mismas que las de la prueba t de muestras independientes de permutación que explicamos en la página 378.

Prueba Kruskal-Wallis. Esta prueba es similar a la prueba F de permutación para un ANOVA de un factor, excepto que las puntuaciones se convierten en rangos antes de calcular F obtenida y generar la distribución muestral de permutación.

Razonamiento

Para realizar la prueba F de permutación para un ANOVA de un factor con tres observaciones en cada uno de cuatro grupos, previamente encontramos que tendríamos que formar 369,600 disposiciones de datos calculando un estadístico F para cada una. El número de cálculos aumenta con rapidez a medida que aumentan el tamaño de la muestra y el número de grupos.

Esta descomunal tarea, y otras consideraciones, animaron a W. H. Kruskal y W. A. Wallis [28] a idear una prueba de permutación de muestras independientes múltiples basada en rangos. Una ventaja de una estrategia así es que una vez que se genera la distribución de permutación para el estadístico de prueba para los tamaños de muestras dados (n_1, n_2, etcétera), se puede usar la distribución para cualquier muestra de los mismos tamaños sin importar los valores de los datos originales. Esto se deduce del hecho de que el dato original se remplaza con rangos para los cuales ya se ha construido la distribución.

La conversión de los datos originales a rangos se logra de la siguiente manera. La puntuación más pequeña en el conjunto de datos recibe el rango 1, la segunda más pequeña el rango 2 y así sucesivamente; la observación más grande recibe el rango $n_1 + n_2 + \cdots + n_k$, donde n_k es el número de observaciones en el grupo k. Esto es, si hay cuatro observaciones en cada una de las tres muestras, el valor más pequeño recibe el rango 1 y el más grande el rango 4 + 4 + 4 = 12. Note que los rangos se asignan sin importar a qué grupo pertenecen. Por ejemplo, suponga que deseamos convertir en rangos los datos utilizados previamente junto con la prueba F de permutación de un ANOVA de un factor. Estos datos se repiten aquí para su comodidad.

Grupo 1	Grupo 2	Grupo 3
4	−1	7
9	0	12

Las observaciones originales se remplazarían con rangos de la siguiente manera.

Grupo 1	Grupo 2	Grupo 3
3	1	4
5	2	6

TABLA 10.18: Seis de las 90 disposiciones posibles de los rangos 1 al 6 de tres grupos con los estadísticos *F* asociados.

grupo 1	grupo 2	grupo 3
3	1	4
5	2	6

$F = 4.333$

grupo 1	grupo 2	grupo 3
1	3	5
2	4	6

$F = 16.000$

grupo 1	grupo 2	grupo 3
1	2	5
3	4	6

$F = 4.333$

grupo 1	grupo 2	grupo 3
2	1	5
3	4	6

$F = 3.273$

grupo 1	grupo 2	grupo 3
2	1	5
4	3	6

$F = 4.333$

grupo 1	grupo 2	grupo 3
1	2	5
4	3	6

$F = 3.273$

Si aplicamos la ecuación 7.2 de la página 264 a los rangos, obtenemos un estadístico *F* para un ANOVA de un factor de 4.333.

Para generar la distribución muestral de permutación para esta prueba de rangos, debemos generar los enteros 1 al 6 y formar todas las disposiciones posibles en las que dos rangos se asignan a cada uno de los tres grupos. Se debe calcular un estadístico *F* para cada una de estas disposiciones. En la página 390 se demostró que son posibles 90 de estas disposiciones de rangos. Con fines de demostración, se muestran seis de las 90 disposiciones posibles en la tabla 10.18.

Habiendo formado las 90 disposiciones posibles de los rangos en tres grupos y calculando un estadístico *F* para cada una, podemos ordenar los 90 estadísticos *F* en una distribución muestral de permutación como se ilustra en la figura 10.8 de la siguiente página. Observe que, por razones de espacio, no presentamos los 90 estadísticos de permutación aquí, sino sólo los 18 estadísticos superiores de la distribución.[23]

Si la hipótesis nula es verdadera, esperaríamos que la *F obtenida* (de los rangos de los datos recabados en el estudio) estuviera cerca de 1.[24] Por otra parte, si H_0 no es verdadera, esperaríamos que *F* tomara algún valor en la cola del extremo derecho de la distribución, ya que la variación muestral entre (es decir, MS_b) sería mayor de lo que se esperaría si esta variación se debiera sólo al proceso de asignación aleatoria. En este caso, la *F* obtenida para los rangos de los datos

[23] Usted recordará del capítulo 7 que la prueba *F* de un ANOVA de un factor es intrínsecamente de una cola.

[24] Véase el capítulo 7.

18 observaciones superiores en la distribución muestral											
4.333	4.333	4.333	4.333	4.333	4.333	4.333	4.333	4.333	4.333	4.333	4.333
↑ F obtenida											
18 observaciones superiores en la distribución muestral (continuación)											
16.000	16.000	16.000	16.000	16.000	16.000						

FIGURA 10.8: 18 observaciones superiores en la distribución muestral de permutación de los estadísticos de rango *F* obtenidos de todas las disposiciones posibles de los rangos 1 al 6 en tres grupos compuestos por dos rangos cada uno.

derivados del estudio es 4.333. ¿Qué probabilidad hay de obtener un valor tan extremo de *F* cuando la hipótesis nula es verdadera? Es simplemente la proporción de estadísticos *F* en la distribución muestral que toman un valor tan grande como, o más grande que, el obtenido de los datos. En este caso, 12 de los 90 valores posibles de *F* son iguales a *F* obtenida de 4.333, mientras que otros seis exceden este valor. Ahora podemos decir que de los 90 valores posibles de *F* que podrían haberse obtenido bajo la hipótesis nula, sólo 12 + 6 = 18 son tan extremos como el observado de los datos. La probabilidad de alcanzar un valor tan extremo de *F* bajo una hipótesis nula verdadera es entonces 18/90 = .20. Éste es el valor-*p* para una prueba de la hipótesis nula de poblaciones idénticas o efectos iguales del tratamiento.[25]

Remplazar los datos originales con rangos también nos permite construir tablas de valores críticos para esta prueba. Para lograrlo, sólo necesitamos generar la distribución muestral de permutación para los conjuntos de rangos para diversos números de grupos y tamaños de muestras y ubicar los valores de permutación de tal forma que α de los valores de permutación sean mayores que o iguales al valor identificado. Por ejemplo, si examinamos la figura 10.8, encontramos que un valor *F* de 16.0 tiene un valor-*p* asociado de 6/90 = .067. Como 16.0 es el valor más extremo en la distribución, es obvio que para tres grupos con dos observaciones en cada uno, se pueden realizar pruebas con $\alpha = .10$ pero no en los niveles .05, .025, .01 o .005. A medida que los tamaños de las muestras aumentan y la distribución muestral de permutación se compone de más valores distintos de *F*, se pueden realizar pruebas en más niveles de α.

El Apéndice I provee valores críticos para la prueba Kruskal-Wallis cuando el número de observaciones por grupo es de tres a 30 y el número de grupos es de dos a 10. Observe que para usar esta tabla el número de observaciones debe ser el mismo para cada grupo.[26]

EJEMPLO 10.27

Utilice los datos provistos aquí para realizar una prueba Kruskal-Wallis con $\alpha = .05$.

Grupo 1	Grupo 2	Grupo 3	Grupo 4
3	4	16	0
1	2	12	7
5	9	8	10

[25] Véase el tema de la hipótesis nula probada mediante el procedimiento de muestras independientes en la página 381.

[26] En la actualidad se están construyendo los valores críticos para tamaños de muestra desiguales.

Solución Remplazar las observaciones originales con rangos produce la siguiente tabla.

Grupo 1	Grupo 2	Grupo 3	Grupo 4
4	5	12	1
2	3	11	7
6	9	8	10

Si R_1 representa los rangos del primer grupo, R_1^2 los rangos cuadrados del primer grupo y los subíndices 2, 3 y 4 los grupos restantes, entonces calculamos

$$\sum R_1 = 4 + 2 + 6 = 12$$
$$\sum R_1^2 = 4^2 + 2^2 + 6^2 = 56$$
$$\sum R_2 = 5 + 3 + 9 = 17$$
$$\sum R_2^2 = 5^2 + 3^2 + 9^2 = 115$$
$$\sum R_3 = 12 + 11 + 8 = 31$$
$$\sum R_3^2 = 12^2 + 11^2 + 8^2 = 329$$
$$\sum R_4 = 1 + 7 + 10 = 18$$
$$\sum R_4^2 = 1^2 + 7^2 + 10^2 = 150$$

Las sumas de los cuadrados para los cuatro grupos son como sigue.

$$SS_1 = \sum R_1^2 - \frac{\left(\sum R_1\right)^2}{n_1} = 56 - \frac{(12)^2}{3} = 8.000$$
$$SS_2 = \sum R_2^2 - \frac{\left(\sum R_2\right)^2}{n_2} = 115 - \frac{(17)^2}{3} = 18.667$$
$$SS_3 = \sum R_3^2 - \frac{\left(\sum R_3\right)^2}{n_2} = 329 - \frac{(31)^2}{3} = 8.667$$
$$SS_4 = \sum R_4^2 - \frac{\left(\sum R_4\right)^2}{n_4} = 150 - \frac{(18)^2}{3} = 42.000$$

Mediante la ecuación 7.4 de la página 265 la suma de los cuadrados es

$$SS_w = SS_1 + SS_2 + SS_3 + SS_4 = 8.000 + 18.667 + 8.667 + 42.000 = 77.334$$

Mediante la ecuación 7.3 de la página 265 el cuadrado medio es

$$MS_w = \frac{SS_w}{N-k} = \frac{77.334}{12-4} = 9.667$$

donde SS_w es la suma de los cuadrados, N es el número total de observaciones y k es el número de grupos.

Mediante la ecuación 7.7 de la página 267 sustituyendo R por x para indicar que el cálculo es para los rangos y no para las observaciones originales, la suma de cuadrados entre es

$$SS_b = n\left[\sum_{j=1}^{k} \bar{R}_j^2 - \frac{\left(\sum_{j=1}^{k} \bar{R}_j\right)^2}{k}\right]$$

donde n es el número de observaciones en *cada* grupo y $\bar{R}_j$ son las medias grupales de rangos.

De los cálculos anteriores obtenemos

$$\bar{R}_1 = \frac{\sum R_1}{n_1} = \frac{12}{3} = 4.000$$
$$\bar{R}_2 = \frac{\sum R_2}{n_2} = \frac{17}{3} = 5.667$$
$$\bar{R}_3 = \frac{\sum R_3}{n_3} = \frac{31}{3} = 10.333$$
$$\bar{R}_4 = \frac{\sum R_4}{n_4} = \frac{18}{3} = 6.000$$

Entonces

$$\sum \bar{R} = 4.000 + 5.667 + 10.333 + 6.000 = 26.000$$

y

$$\sum \bar{R}^2 = 4.000^2 + 5.667^2 + 10.333^2 + 6.000^2 = 190.886$$

Haciendo las sustituciones apropiadas en la ecuación 7.7 tenemos

$$SS_b = n\left[\sum_{j=1}^{k} \bar{R}_j^2 - \frac{\left(\sum_{j=1}^{k} \bar{R}_j\right)^2}{k}\right] = 3\left[190.886 - \frac{(26.000)^2}{4}\right] = 65.658$$

Mediante la ecuación 7.6 de la página 267 el cuadrado medio entre es

$$MS_b = \frac{SS_b}{k-1} = \frac{65.658}{4-1} = 21.886$$

donde SS_b es la suma de cuadrados entre y k es el número de grupos. Finalmente, mediante la ecuación 7.2 de la página 264, la F obtenida es

$$F = \frac{MS_b}{MS_w} = \frac{21.886}{9.667} = 2.264$$

Con $\alpha = .05$, cuatro grupos y tres observaciones por grupo, el Apéndice I da un F crítica de 4.483. Como la F obtenida de 2.264 es menor que este valor, no podemos rechazar la hipótesis nula. Por lo tanto, no pudimos demostrar ningún efecto del tratamiento en este estudio. ■

Realización de la prueba

Los pasos para realizar la prueba de rangos Kruskal-Wallis son los siguientes.

1. Remplace las puntuaciones originales con sus respectivos rangos y dé a la puntuación más baja el rango 1, a la segunda más baja el rango 2, y así sucesivamente hasta que la puntuación más baja reciba el rango $n_1 + n_2 + \cdots + n_k$. La clasificación se realiza sin importar el grupo.
2. Utilice la ecuación 7.2 de la página 264 para calcular el estadístico F de un ANOVA de un factor para los rangos obtenidos en el paso 1. Denomine a este valor obtenido F obtenida.
3. Para tamaños iguales de muestras donde $3 \leq n \leq 30$ y $2 \leq k \leq 10$, refiera la F obtenida al Apéndice I y rechace H_0 si la F obtenida es mayor que o igual a la F crítica.
4. Para tamaños iguales de muestras con $n > 30$, refiera la F obtenida a una tabla F como el Apéndice C con grados de libertad para numerador y denominador de $k - 1$ y $N - k$, respectivamente. Aquí, k representa el número de grupos y N el número total de observaciones en los grupos k combinados.
5. Para tamaños desiguales de muestras, la F obtenida se puede referir a las tablas construidas para tamaños desiguales de muestras (creadas por los autores) o, si todos los grupos tienen tamaños de muestras mayores que 15, la referencia puede ser a una tabla F con $k - 1$ y $N - k$ grados de libertad.

EJEMPLO 10.28

En un estudio de control de calidad, se comparan tres planes de seguros en cuanto al tiempo de hospitalización (medido en días) para los pacientes admitidos bajo los tres planes. Cuando se recaban los datos, tenemos que se admitieron 119 pacientes en el plan I, 142 en el plan II y 92 en el plan III. Como los datos del tiempo de hospitalización normalmente están sesgados, los investigadores deciden que una prueba Kruskal-Wallis sería más apropiada que la prueba F de un ANOVA de un factor que supone normalidad. Para este fin, el tiempo de hospitalización para los 353 pacientes se convierte en rangos y se calcula un estadístico F sobre los rangos resultantes. La F obtenida para los rangos fue 3.22. Pruebe la significancia de este hallazgo con $\alpha = .05$. ¿A qué conclusión llegó usted sobre el tiempo de hospitalización para los tres planes?

Solución Si bien el Apéndice I no provee un valor crítico para esta prueba, los tamaños grandes de muestras hacen de la distribución F una aproximación adecuada. Usando los grados de libertad de $k - 1 = 3 - 1 = 2$ del numerador y los grados de libertad $N - k = 353 - 3 = 350$ del denominador, el Apéndice C provee un valor crítico de 3.02 para $\alpha = .05$. Como la F obtenida de 3.22 excede este valor, podemos rechazar la hipótesis nula de poblaciones idénticas. De acuerdo con las palabras de James Bradley reproducidas en la página 381, también podríamos concluir que el promedio del tiempo de hospitalización difiere para alguno de los tres planes de seguro. ■

Suposiciones

Las suposiciones subyacentes en la prueba Kruskal-Wallis son las mismas que en la prueba de muestras independientes de Wilcoxon, explicadas en la página 388. Consulte también las páginas 388 y 389 sobre las observaciones ligadas. Note que, cuando hay vínculos, si bien la prueba de muestras independientes se puede referir a una distribución t, la prueba Kruskal-Wallis se referiría a una distribución F. En el caso de dos grupos, las pruebas de Wilcoxon y Kruskal-Wallis dan el mismo resultado.

TABLA 10.19: Tabla de contingencia que muestra los resultados positivos y negativos de dos grupos de tratamiento.

	+	−	
Tratamiento 1	7	3	10
Tratamiento 2	1	9	10
	8	12	

10.3.5 Tablas de contingencia

Hasta ahora nos hemos enfocado en resultados continuos como presiones arteriales o tiempos de reacción que son apropiados para análisis mediante correlaciones o pruebas basadas en estadísticos t o F. Pero el principio de permutación es igualmente aplicable al análisis de tablas de contingencia. Si bien el principio de permutación puede aplicarse a tablas de contingencia de cualquier tamaño, restringiremos nuestra atención a la tabla dos por dos, ya que estas tablas se pueden analizar sin usar un software especializado. Comenzaremos por la lógica que subyace en la prueba de tablas dos por dos respecto de la prueba de Fisher; después presentaremos el método de cálculo.

Razonamiento. Suponga que durante un ensayo clínico, se asigna al azar a 20 pacientes a uno de dos tratamientos, 10 pacientes en cada uno. El resultado para cada paciente se evalúa como positivo (+) o negativo (–). Los resultados del ensayo se presentan en la tabla 10.19.

Como se observa, 7 o $\frac{7}{10} = .7$ de los pacientes que recibieron el tratamiento 1 obtuvieron un resultado positivo mientras que 1 o $\frac{1}{10} = .1$ de los pacientes que recibieron el tratamiento 2 dieron un resultado positivo. La pregunta es qué significa esta diferencia en los resultados positivos entre los dos grupos. Una posibilidad es que el tratamiento administrado a los pacientes del grupo 1 fue más efectivo que el que se aplicó a los pacientes del grupo 2, por lo que produjo resultados más positivos. Pero existe una segunda posibilidad. Podría ser que el proceso de asignación al azar colocó a siete de los pacientes con resultados positivos en el grupo 1 mientras que, por casualidad, colocó sólo a un paciente con resultado positivo en el grupo 2. Observe que la primera explicación implica una diferencia en la efectividad de los tratamientos administrados a los dos grupos, mientras que la segunda niega cualquier diferencia. Pero, ¿cómo se podría tomar una decisión sobre cuál de las explicaciones debe creerse?

Un método de decisión basado en la permutación generaría todas las configuraciones de resultados posibles bajo una hipótesis nula de no diferencia en el tratamiento y después usaría esta distribución para tener acceso a la probabilidad del resultado alcanzado en el estudio. En este caso, el proceso de asignación al azar podría haber colocado a ninguno, uno, dos, tres, cuatro, cinco, seis, siete o a ocho de los pacientes + en el grupo 1. La figura 10.20 de la siguiente página incluye todas las disposiciones de esta tabla. Observe que al especificar el número de pacientes + en el grupo 1 automáticamente se especifican los valores en las otras tres celdas de la tabla. Por ejemplo, la tabla 2 de la figura 10.10 indica el orden de los datos cuando se observa un resultado positivo en el grupo 1. En este caso, el número de resultados negativos en el grupo uno *debe* ser nueve porque se asignó un total de 10 pacientes al grupo 1. De igual forma, como había un total de ocho resultados positivos en el estudio, el grupo 2 debe tener siete resultados positivos cuando el grupo 1 tiene uno de estos resultados. La misma lógica demanda que los tres resultados

TABLA 10.20: Todas las tablas posibles de márgenes fijos formadas a partir de un conjunto de datos en particular.

(1)

	+	−	
uno	0	10	10
dos	8	2	10
	8	12	

(2)

	+	−	
uno	1	9	10
dos	7	3	10
	8	12	

(3)

	+	−	
uno	2	8	10
dos	6	4	10
	8	12	

(4)

	+	−	
uno	3	7	10
dos	5	5	10
	8	12	

(5)

	+	−	
uno	4	6	10
dos	4	6	10
	8	12	

(6)

	+	−	
uno	5	5	10
dos	3	7	10
	8	12	

(7)

	+	−	
uno	6	4	10
dos	2	8	10
	8	12	

(8)

	+	−	
uno	7	3	10
dos	1	9	10
	8	12	

(9)

	+	−	
uno	8	2	10
dos	0	10	10
	8	12	

negativos estén en el grupo 2. Así, cuando se genera una distribución de permutación, sólo necesitamos especificar el número de resultados positivos en el grupo 1.[27]

Con base en su conocimiento de los métodos de permutación adquiridos hasta este momento, parecería que un procedimiento válido para probar la permutación podría desarrollarse listando todos los valores del número de resultados positivos obtenidos en el grupo 1 bajo todas las posibles asignaciones aleatorias posibles de los dos grupos. Para ver esto de forma intuitiva, suponga que repetidamente debemos asignar a 10 sujetos a cada grupo. Para cada asignación registramos el número de resultados positivos en el grupo 1. Continuamos este proceso hasta que se hayan obtenido todas las disposiciones de datos posibles en las cuatro celdas. Los números de resultados positivos en el grupo 1 podrían listarse del menor al mayor para formar la distribución muestral de permutación. Entonces, se podría realizar una prueba contando el número de resultados en la distribución de permutación que son mayores que (menores que) o iguales al número de resultados positivos en el grupo 1 obtenidos en el estudio. Éste es, de hecho, el método (asistido por computadora) empleado con más frecuencia para la prueba de permutación de tablas de contingencia.[28]

En el caso de las tablas dos por dos existe un enfoque más directo a la generación de la distribución muestral de permutación. Este enfoque, con la correspondiente prueba de significancia, se conoce como la prueba exacta de Fisher.

[27] Podríamos haber usado cualquiera de las otras tres celdas, si así lo hubiéramos querido.

[28] De hecho, en los últimos años se han desarrollado complejos algoritmos que resulta innecesario listar todos los resultados.

Por el método de Fisher podemos determinar cuántos ceros, unos, dos, etcétera, aparecerán en la distribución muestral de permutación, liberándonos así de la carga de generar todas las aleatorizaciones posibles para encontrar estos valores. Como ejemplo, veamos cuántos 2 aparecerán en la distribución.

Para tomar esta determinación introducimos la notación $+_1, +_2, \ldots, +_8$ para identificar a los ocho pacientes que alcancen un resultado positivo. Se observarán dos resultados positivos en el grupo 1 si $+_1$ y $+_2$ son aleatorizados al grupo 1, o si $+_1$ y $+_3$ son aleatorizados al grupo 1 o si se encuentra cualquiera de las posibles combinaciones con dos resultados positivos asignadas al grupo 1. Pero, ¿de cuántas maneras se pueden dividir ocho sujetos con resultados positivos entre los dos grupos para que dos de estos sujetos estén asignados al grupo 1? De su estudio del apartado 10.2.2 de la página 345 reconocerá éste como un problema de combinaciones. Mediante la ecuación 10.2 de la página 347

$$C_{n_2}^{n_1} = \frac{n!}{n_1!n_2!} = C_6^2 = \frac{8!}{2!6!} = 28$$

Así, hay 28 formas en las que se puede llevar a cabo la asignación aleatoria de los ocho sujetos para que dos de estos pacientes queden asignados al grupo 1. Pero los pacientes con un resultado negativo también deben ser aleatorizados. Para *cada una* de las 28 posibles asignaciones de los pacientes con resultados positivos habrá

$$C_{n_2}^{n_1} = \frac{n!}{n_1!n_2!} = C_4^8 = \frac{12!}{8!4!} = 495$$

asignaciones de pacientes con resultados negativos.

Esto significa que hay

$$\left(C_6^2\right)\left(C_4^8\right) = (28)\,(495) = 13,\,860$$

asignaciones aleatorias que producirán dos resultados positivos en el grupo 1. Este proceso puede realizarse para las nueve tablas de la figura 10.20. Esto daría el número de ceros, unos, dos, etcétera, que podrían resultar del proceso de asignación aleatoria. Los resultados podrían combinarse para formar la distribución muestral de permutación deseada.

Prueba exacta de Fisher. La tabla 10.21 de la siguiente página representa los resultados de una aleatorización de sujetos a uno de dos grupos. En la aleatorización, se asignaron g sujetos al grupo 1, mientras que se asignaron h sujetos al grupo 2. Observe que especificar la frecuencia de cualquier celda determina de manera automática las frecuencias de las otras tres celdas.

También será de utilidad notar que los valores máximo y mínimo de a para los márgenes especificados e, f, g y h serán, respectivamente

$$\boxed{a_{mín} = g - mín\,(f, g)} \tag{10.4}$$

y

$$\boxed{a_{máx} = mín\,(e, g)} \tag{10.5}$$

TABLA 10.21: Tabla de contingencia que presenta los resultados de una asignación aleatoria de sujetos a uno de dos grupos.

	+	−	
Tratamiento 1	a	b	$g = a + b$
Tratamiento 2	c	d	$h = c + d$
	$e = a + c$	$f = b + d$	

Deseamos encontrar la probabilidad de que la aleatorización produzca cada uno de los valores posibles de a. Es decir, los valores de $a_{mín}$ a $a_{máx}$.[29] Note que esta probabilidad sería sólo la proporción de todas las aleatorizaciones posibles que produzcan el valor a determinado.

Para encontrar la proporción de aleatorizaciones que produzcan un valor de a determinado, primero debemos encontrar el número de éstas. Este número se puede entonces dividir entre el número total de aleatorizaciones para producir la probabilidad deseada.

Comenzamos por apuntar que el número fijo de pacientes con resultados positivos es e. Debemos encontrar el número de aleatorizaciones que produzcan exactamente a resultados positivos en el grupo 1. Hay dos componentes en este problema. Primero debemos preguntar: "¿De cuántas maneras podemos dividir e sujetos en dos grupos colocando a a sujetos en un grupo y c en el otro?" Por lo que usted estudió en el apartado 10.2.2 de la página 345, sabe que este valor es C_c^a. Pero también debemos considerar las asignaciones de los f pacientes con resultados negativos. Por el mismo razonamiento, esto será C_d^b. Ahora podemos decir que para cada una de las C_c^a posibles asignaciones de los sujetos con resultados positivos habrá C_d^b asignaciones de pacientes con resultados negativos. Así, el número de aleatorizaciones posibles que produzcan la frecuencia a es $\left(C_c^a\right)\left(C_d^b\right)$. Podemos obtener la *probabilidad* de alcanzar la frecuencia a si dividimos $\left(C_c^a\right)\left(C_d^b\right)$ entre el número total de aleatorizaciones posibles. Este número será C_h^g. Observe que este número refleja *todas* las aleatorizaciones posibles sin fijar a a un valor específico.

Para resumir, podemos decir que la probabilidad asociada con cualquier valor de a es

$$P\,(a) = \frac{\left(C_c^a\right)\left(C_d^b\right)}{C_h^g}$$

Usando la notación de la ecuación 10.2 de la página 347 y si n representa el número total de sujetos (es decir, $e + f$ o $g + h$) podemos escribir esta expresión como

$$P\,(a) = \frac{\left(\dfrac{e!}{a!c!}\right)\left(\dfrac{f!}{b!d!}\right)}{\dfrac{n!}{g!h!}}$$

lo que se simplifica en

$$\boxed{P\,(a) = \frac{e!f!g!h!}{a!b!c!d!n!}} \tag{10.6}$$

[29] Las notaciones *máx* (x, y) y *mín* (x, y) indican, respectivamente, que deben emplearse el máximo y el mínimo de xy.

TABLA 10.22: Probabilidades de *a*.

a	$P(a)$
0	.00036
1	.00953
2	.07502
3	.24006
4	.35008
5	.24006
6	.07502
7	.00953
8	.00036

lo que se conoce como la **función hipergeométrica**.

EJEMPLO 10.29

Dados los valores marginales en la tabla 10.19 de la página 399, encuentre los valores máximo y mínimo para los números de pacientes con resultados positivos asignados al grupo 1 como resultado del proceso de aleatorización. Utilice la ecuación 10.6 de la siguiente página para encontrar las probabilidades asociadas con cada uno de estos valores.

Solución Mediante las ecuaciones 10.4 y 10.5, el número más alto posible y el más bajo posible de resultados positivos en el grupo 1 son, respectivamente,

$$a_{mín} = g - mín\,(f, g) = 10 - mín\,(12, 10) = 10 - 10 = 0$$

y

$$a_{máx} = mín\,(e, g) = mín\,(8, 10) = 8$$

Las frecuencias para los valores de los resultados positivos en el grupo 1 del cero al ocho se muestran en la tabla 10.20 de la página 400. Usando las frecuencias en el cuadro 1 de la tabla 10.20 en la ecuación 10.6 da

$$P\,(0) = \frac{8!12!10!10!}{0!10!8!2!20!} = .00036$$

La probabilidad para la tabla 2 sería

$$P\,(1) = \frac{8!12!10!10!}{1!9!7!3!20!} = .00953$$

Le damos las probabilidades restantes en la tabla 10.22, pero dejaremos que usted realice los cálculos.[30]

Podemos realizar una prueba de la hipótesis nula

$$H_0 : \pi_1 = \pi_2$$

[30] No olvide que usted puede cancelar términos, así como lo hizo en el apartado 4.2.5 de la página 81, para ahorrarse algo del trabajo de cálculo.

donde π_1 y π_2 son las proporciones de las poblaciones de resultados positivos para los grupos 1 y 2 refiriendo el valor obtenido de a (7, en este ejemplo) a la distribución en 10.22. Para la alternativa de una cola

$$H_A : \pi_1 > \pi_2$$

la probabilidad de obtener un valor de a mayor que o igual a 7 es $P(7) + P(8) = .00953 + .00036 = .00989$, que es el valor-$p$ de la prueba. Por lo tanto, podemos rechazar la hipótesis nula con $\alpha = .01$ y así concluir que el tratamiento administrado al grupo 1 fue más efectivo al producir resultados positivos de lo que fue el tratamiento dado al grupo 2. ■

Realización de la prueba

Los pasos para realizar la prueba exacta de Fisher son los siguientes.

1. Utilice las ecuaciones 10.4 y 10.5 para encontrar los valores mínimo y máximo para a.
2. Utilice los resultados del paso 1 para construir todas las tablas de aleatorización posibles que tengan los mismos valores marginales que los obtenidos del conjunto de datos original.
3. Utilice la ecuación 10.6 de la página 402 para encontrar las probabilidades para cada tabla numerada en el paso 2.
4. Ordene los valores de a junto con sus respectivas probabilidades de $a_{mín}$ a $a_{máx}$.
5. Para las pruebas de una cola el valor-p se encuentra

   a) sumando las probabilidades para todos los valores de a que son menores que o iguales a a obtenido si la hipótesis alternativa especifica que $H_A : \pi_1 < \pi_2$, o

   b) sumando las probabilidades para todos los valores de a que son mayores que o iguales a a obtenido si la hipótesis alternativa especifica que $H_A : \pi_1 > \pi_2$.

6. Para las pruebas de dos colas el valor-p se encuentra[31]

   a) apuntando la suma de probabilidades para todos los valores de a que son menores que o iguales a a obtenido, así como la suma de todos los valores de a que son mayores que o iguales a a obtenido. Denomine al menor de estos valores P_1.

   b) Si P_1 se calcula de la cola inferior, comience por sumar las probabilidades en la cola superior comenzando con $a_{máx}$. Continúe sumando en tanto que la suma sea menor que o igual a P_1. Denomine esta suma P_2.

   c) Si P_1 se calcula de la cola superior, comience por sumar las probabilidades en la cola inferior comenzando con $a_{mín}$. Continúe sumando en tanto que la suma sea menor que o igual a P_1. Denomine esta suma P_2.

   d) El valor-p de dos colas es $P_1 + P_2$.

---

[31] Se han propuesto diversos métodos para calcular el valor-p de dos colas para la prueba exacta de Fisher. El que aquí se presenta es de los que se utilizan con mayor frecuencia.

TABLA 10.23: Tabla de contingencia que relaciona el hábito de fumar con la enfermedad.

	D	$\overline{D}$	
S	4	2	6
$\overline{S}$	1	8	9
	5	10	

EJEMPLO 10.30

Suponga que se efectúa un estudio para determinar si la proporción de fumadores que padecen cierta enfermedad difiere de la proporción de no fumadores que manifiestan la misma enfermedad. Utilice los datos de la tabla 10.23 y la prueba exacta de Fisher[32] para probar la hipótesis nula

$$H_0 : \pi_1 = \pi_2$$

contra la alternativa

$$H_A : \pi_1 > \pi_2$$

Después, pruebe la hipótesis nula contra la alternativa

$$H_A : \pi_1 \neq \pi_2$$

Realice ambas pruebas al nivel de significancia de .05.

Solución Mediante las ecuaciones 10.4 y 10.5 de la página 401, los números mínimos y máximos de fumadores con la enfermedad que aparecen en las tablas de aleatorización serán

$$a_{mín} = g - mín\ (f, g) = 6 - mín\ (10, 6) = 6 - 6 = 0$$

y

$$a_{máx} = mín\ (e, g) = mín\ (5, 6) = 5$$

La figura 10.24 de la siguiente página muestra todas las configuraciones de tabla posibles asociadas con la tabla 10.23. Si usamos estos valores de la tabla con la ecuación 10.6 de la página 402, tendremos las probabilidades de la tabla 10.25 de la siguiente página.

Calculamos $P(0)$ y $P(5)$ aquí pero dejaremos que usted resuelva los cálculos restantes.

$$P\ (0) = \frac{e!f!g!h!}{a!b!c!d!n!} = \frac{5!10!6!9!}{0!6!5!4!15!} = .04196$$

$$P\ (0) = \frac{5!10!6!9!}{5!1!0!9!15!} = .00200$$

Para la alternativa de una cola $H_A : \pi_1 > \pi_2$, aplicamos el paso 5b y encontramos

$$P\ (4) + P\ (5) = .04496 + .00200 = .04696$$

[32] ¿Qué otros métodos conoce que puedan usarse para realizar esta prueba?

TABLA 10.24: Todas las tablas posibles de márgenes fijos formadas a partir de un conjunto de datos en particular.

(1)

	D	$\bar{D}$	
S	0	6	6
$\bar{S}$	5	4	9
	5	10	

(2)

	D	$\bar{D}$	
S	1	5	6
$\bar{S}$	4	5	9
	5	10	

(3)

	D	$\bar{D}$	
S	2	4	6
$\bar{S}$	3	6	9
	5	10	

(4)

	D	$\bar{D}$	
S	3	3	6
$\bar{S}$	2	7	9
	5	10	

(5)

	D	$\bar{D}$	
S	4	2	6
$\bar{S}$	1	8	9
	5	10	

(6)

	D	$\bar{D}$	
S	5	1	6
$\bar{S}$	0	9	9
	5	10	

TABLA 10.25: Probabilidades de *a*.

a	$P(a)$
0	.04196
1	.25175
2	.41958
3	.23976
4	.04496
5	.00200

que es el valor-*p* de la prueba. Como este valor es menor que $\alpha = .05$ se rechaza la hipótesis nula.

Para encontrar el valor-*p* de dos colas, aplicamos el paso 6a, primero encontrando la probabilidad de que *a* tome un valor mayor que o igual a *a* obtenido de 4. Este valor es

$$P(4) + P(5) = .04496 + .00200 = .04696$$

La probabilidad de que *a* tome un valor menor que o igual al valor obtenido de 4 se puede dar si encontramos $P(0) + P(1) + P(2) + P(3) + P(4)$ o, de manera más simple, si apuntamos que $P(0) + P(1) + P(2) + P(3) = 1 - .04696 = .95304$, por lo que la probabilidad de que *a* tome un valor menor que o igual a 4 es $.95304 + P(4) = .95304 + .04496 = .99800$.

Como .04696 es menor que .99800, determinamos $P_1 = .04696$.

Como P_1 se obtuvo de la cola superior de la distribución de permutación, aplicamos el paso 6c y comenzamos la suma con $a_{mín}$, que es cero en este caso. Encontramos que $P(0) = .4196$. No podemos añadir otras probabilidades a este valor porque $P(0) + P(1) = .04196 + .25175 = .29371$, que es mayor que $P_1 = .04696$. Por consiguiente, $P_2 = .04196$ y el valor-*p* de dos colas para esta prueba es $P_1 + P_2 = .04696 + .04196 = .08892$, por lo que la prueba de dos colas no es significativa con $\alpha = .05$. ■

TABLA 10.26: Tabla de contingencia que relaciona el abuso infantil con los trastornos de alimentación.

	Trastorno	Sin trastorno	
Víctimas de abuso	7	4	11
Sujetos que no sufren abuso	3	3	6
	10	7	

EJEMPLO 10.31

Un investigador desea determinar si el abuso infantil está relacionado con los trastornos de alimentación desarrollados después. Para este fin se recabaron los datos mostrados en la tabla 10.26. Utilice estos datos para realizar la prueba exacta de Fisher al nivel de significancia de .05. Pruebe la hipótesis

$$H_0 : \pi_1 = \pi_2$$

contra la alternativa

$$H_A : \pi_1 \neq \pi_2$$

¿Usted creería que estos datos se recabaron de una población general o de un grupo en particular?

Solución Los valores mínimo y máximo de a son, mediante las ecuaciones 10.4 y 10.5

$$a_{mín} = g - mín\,(f, g) = 11 - mín\,(7, 11) = 11 - 7 = 4$$

y

$$a_{máx} = mín\,(e, g) = mín\,(10, 11) = 10$$

Usamos estos valores para formar las tablas de aleatorización de la tabla 10.27 de la siguiente página.[33]

La probabilidad de observar cada una de las tablas de la tabla 10.27 bajo el proceso de aleatorización se da en la tabla 10.28 de la siguiente página. Mostramos el cálculo de dos de estas probabilidades, pero dejaremos que usted resuelva las demás. Usando las frecuencias de celda y marginales de las tablas 1 y 4 de la figura 10.27, calculamos, mediante la ecuación 10.6

$$P\,(4) = \frac{e!f!g!h!}{a!b!c!d!n!} = \frac{10!7!11!6!}{4!7!6!0!17!} = .01697$$
$$P\,(7) = \frac{10!7!11!6!}{7!4!3!3!17!} = .33937$$

Para encontrar el valor-p de dos colas, aplicamos el paso 6a de la página 404, primero encontrando la probabilidad de que a tome un valor menor que o igual a a obtenido de 7. Este valor es

$$P(4) + P(5) + P(6) + P(7) = .01697 + .14253 + .35633 + .33937 = .85520$$

[33] ¿Qué pasa si usted usa un valor para a que está fuera de este rango? Pruebe estableciendo $a = 3$.

TABLA 10.27: Todas las tablas posibles de márgenes fijos formadas a partir de un conjunto de datos que relaciona el abuso con los trastornos de la alimentación.

(1)

	Trastorno	Sin trastorno	
Víctimas de abuso	4	7	11
Sujetos que no sufren abuso	6	0	6
	10	7	

(2)

	Trastorno	Sin trastorno	
Víctimas de abuso	5	6	11
Sujetos que no sufren abuso	5	1	6
	10	7	

(3)

	Trastorno	Sin trastorno	
Víctimas de abuso	6	5	11
Sujetos que no sufren abuso	4	2	6
	10	7	

(4)

	Trastorno	Sin trastorno	
Víctimas de abuso	7	4	11
Sujetos que no sufren abuso	3	3	6
	10	7	

(5)

	Trastorno	Sin trastorno	
Víctimas de abuso	8	3	11
Sujetos que no sufren abuso	2	4	6
	10	7	

(6)

	Trastorno	Sin trastorno	
Víctimas de abuso	9	2	11
Sujetos que no sufren abuso	1	5	6
	10	7	

(7)

	Trastorno	Sin trastorno	
Víctimas de abuso	10	1	11
Sujetos que no sufren abuso	0	6	6
	10	7	

TABLA 10.28: Probabilidades de *a*.

a	$P(a)$
4	.01697
5	.14253
6	.35633
7	.33937
8	.12726
9	.01697
10	.00057

La probabilidad de la cola superior es

$$P(7) + P(8) + P(9) + P(10) = .33937 + .12726 + .01697 + .00057 = .48417$$

Como .48417 es menor que .85520, establecemos $P_1 = .48417$.

Como P_1 se obtuvo como una suma de la cola superior, encontramos P_2 comenzando por sumar las probabilidades de la cola inferior, con $a_{mín} = 4$. Esta suma es

$$P(4) + P(5) = .01697 + .14253 = .15950$$

Observe que no podemos añadir la probabilidad para $P(6)$ porque daría como resultado una suma de .51583, que es más grande que P_1. Por consiguiente, el valor-*p* de dos colas es $P_1 + P_2 = .48417 + .15950 = .64367$, por lo que no se rechaza la hipótesis nula. También notamos que los datos eran una probabilidad y no una muestra de la población general porque, aun el grupo de no abuso tuvo una alta incidencia de trastornos de alimentación: 3/6 = .5. ■

Suposiciones

La suposición fundamental que subyace en la prueba exacta de Fisher es que todas las observaciones son independientes. Hemos hablado sobre la suposición de independencia en relación con muchas otras pruebas y no lo repetiremos aquí.[34] Baste con decir que no se puede depender de esta prueba para el control de errores del tipo I cuando se viola este supuesto.

10.4 MÁS SOBRE LOS MÉTODOS BASADOS EN LA PERMUTACIÓN

Hemos dedicado aproximadamente 100 páginas a este tema, pero no hemos dicho todo. Es imposible cerrar este capítulo sin decir un poco más sobre los métodos basados en la permutación y nuestra presentación a este respecto.

- Las formas de cálculo para los estadísticos no paramétricos presentadas en este capítulo no son las comúnmente usadas en los textos no paramétricos. Por ejemplo, para las pruebas de rangos le hemos enseñado simplemente a convertir las puntuaciones originales en rangos y a calcular estadísticos paramétricos estándar con los que usted ya está familiarizado en cuanto a los rangos resultantes. En los tratamientos estándar, se dan formas de cálculo más simples. Pero esto requiere que el alumno domine nuevos estadísticos con los que no está familiarizado. Estas formas a menudo también están relacionadas con tablas de valores críticos no intuitivos[35] y confusos. En otros tiempos, la facilidad de cálculo era de gran importancia, lo que hizo que las formas más simples fueran sumamente aceptadas. Con la llegada de las calculadoras de bolsillo y las computadoras, la facilidad de cálculo ha disminuido en importancia. Así, hemos decidido que usted no tenga que aprender un nuevo conjunto de estadísticos, sino que simplemente aplique aquellos con los que ya está familiarizado. Subrayamos que ambos métodos de cálculo dan el mismo resultado.
- No todas las pruebas no paramétricas basadas en rangos se pueden expresar como sus equivalentes paramétricas realizadas sobre transformaciones en rangos de los datos originales. Sin embargo, como se demostró en este capítulo, esto es cierto de la mayoría de las pruebas basadas en rangos presentadas en textos de introducción. Nos atrevemos a añadir que esto es cierto de las pruebas no paramétricas basadas en rangos más comúnmente empleadas.

---

[34] Véase por ejemplo el tema de independencia relacionado con la prueba chi cuadrada presentada en la página 282.

[35] Por ejemplo, algunas tablas requieren que los valores críticos de la cola inferior se usen para pruebas de la cola superior.

- Un mito antiguo y a menudo reiterado sostiene que las pruebas no paramétricas basadas en rangos son intrínsecamente menos potentes (o eficientes) que sus equivalentes paramétricas. Por ejemplo, Kuzma y Bohnenblust [30] dicen de las pruebas no paramétricas "son menos eficientes (es decir, requieren un tamaño muestral más grande para rechazar una hipótesis falsa) que las pruebas paramétricas equiparables". Estos mismos autores concluyen: "Al utilizar métodos no paramétricos, usted debe tener cuidado de verlos como métodos estadísticos complementarios y no como alternativas atractivas".
  El primer planteamiento es falso. Como resultado, no apoyamos la conclusión dada en el segundo planteamiento. La verdad del asunto es que *las pruebas no paramétricas son necesariamente menos potentes que sus equivalentes paramétricas sólo cuando los supuestos de estas últimas se cumplen a la perfección*.[36] A decir verdad, algunas pruebas no paramétricas, como las pruebas de rangos con signos y de muestras independientes de Wilcoxon, pueden tener ventajas poderosas *muy grandes* sobre sus contrapartes paramétricas, como las pruebas t de muestras apareadas e independientes, en situaciones comunes. [1, 2] En la práctica del análisis de datos, los métodos paramétricos pueden tener ventajas poderosas en algunas situaciones, mientras que los métodos no paramétricos mantienen ciertas ventajas en otras. En verdad, ninguno de los conjuntos de pruebas mantiene un dominio absoluto sobre el otro en este sentido.
- En la época de su desarrollo, la mayoría de los métodos de permutación que utilizaban puntuaciones originales fueron de gran importancia teórica en vez de práctica. Esto se debió a que ante las grandes dificultades que implicaba el cálculo, su uso no sólo fuera impráctico, sino imposible, al menos para muestras grandes. Con la llegada de las computadoras modernas y los nuevos algoritmos de cálculo, estas pruebas han entrado al mundo de los métodos estadísticos prácticos. Esto lo demuestra el hecho de que hay muchos paquetes de software disponibles que implementan estos métodos. No daremos una lista completa, pero mencionaremos que entre los paquetes más conocidos y usados está el StatXact de Cytel Software Corporation [9]. Entre los más completos está SC (Statistical Calculator) de Mole Software [10]. Un lenguaje de programación bastante simple y flexible diseñado específicamente para problemas de permutación y similares se encuentra en el software Resampling Stats de Resampling Stats Inc. [41]. Además, muchos de los paquetes antiguos que implementaron los métodos paramétricos y los no paramétricos basados en rangos han agregado métodos basados en la permutación para puntuaciones originales y tablas de contingencia a su cuadro de procedimientos.
- En el apartado 2.2.2 de la página 10, indicamos que las mediciones del nivel ordinal por tradición han presentado cierta dificultad en lo que se refiere a las pruebas inferenciales. Los métodos de permutación superan muchos de estos problemas.
- **Interacción** es un importante concepto estadístico en el que el efecto de una variable depende de su asociación con otra(s) variable(s). Un defecto importante de los métodos de permutación basados en rangos es su incapacidad para detectar las interacciones de manera significativa. Los métodos de permutación basados en datos originales son un tanto mejores en este aspecto, pero sólo dan resultados aproximados y no exactos.
- Sólo hemos mencionado unos cuantos métodos de permutación en este capítulo. Hay muchos más disponibles. De hecho, los métodos de permutación se pueden usar para desarrollar procedimientos especiales para problemas para los que no hay ningún procedimiento no paramétrico disponible.

[36] Una situación rara vez encontrada en la práctica.

PALABRAS Y FRASES CLAVE

Al terminar de leer este capítulo, usted estará familiarizado con las siguientes palabras y frases:

combinaciones 346
distribución muestral de permutación 350
distribución libre 348
estrategia del rango promedio 360
factorial 345
función hipergeométrica 403
interacción 410
no paramétrico 348
paramétrico 348
permutaciones 344
principio de permutación 348
prueba de correlación de Pitman 349
prueba de los signos de Wilcoxon 365
prueba de Wilcoxon para muestras independientes 378
prueba exacta de Fisher 401
prueba *F* de permutación de ANOVA de un factor 390
prueba Hoteling y Pabst (H & P) 353
prueba Kruskal-Wallis 393
prueba *t* de permutación para muestras independientes 374
prueba *t* de permutación para muestras apareadas 361
pruebas basadas en rangos 353
rangos 354
relación monotónica 356
rho de Spearman 354

EJERCICIOS

10.1 Se debe administrar consecutivamente una serie de cinco pruebas médicas distintas a un paciente. ¿En cuántos órdenes podrían administrarse las cinco pruebas? Suponga que hubiera seis pruebas.

10.2 En un ensayo clínico se debe asignar aleatoriamente a los pacientes a los grupos de tratamiento.

a) Si 10 pacientes son asignados a un grupo de tratamiento y a un grupo de control, y hay cinco pacientes en cada uno, ¿cuántos agrupamientos de pacientes distintos son posibles?

b) Si 15 pacientes son asignados a tres grupos, con cinco pacientes en cada uno, ¿cuántos agrupamientos de pacientes distintos son posibles?

10.3 Dados tres pares de observaciones,

a) ¿cuántos valores de *r* aparecerían en la distribución muestral de la prueba de correlación de Pitman?

b) ¿sería posible realizar una prueba con $\alpha = .05$? ¿Por qué?

10.4 Se realiza un estudio para determinar si hay una relación entre la edad y los resultados de las pruebas de detección de cáncer con un antígeno prostático específico (PSA). Los datos se muestran a continuación. Utilice estos datos para realizar la prueba Hoteling y Pabst para la correlación de rangos para probar la hipótesis nula de no relación contra la alternativa de una relación positiva. Utilice $\alpha = .05$.

PSA	Edad
9.1	78
1.2	24
2.3	60
6.0	77
3.9	48
1.0	33
6.1	75
2.7	51
5.9	88
9.9	79

10.5 Suponga que se obtiene una correlación de rangos de .88 para $n = 124$ pares de datos. Pruebe este coeficiente para determinar su significancia mediante una prueba de dos colas con $\alpha = .10$.

10.6 Dados seis pares de datos, ¿cuántas disposiciones de datos serían necesarias para realizar la distribución muestral para la prueba *t* de permutación para muestras apareadas?

10.7 Los siguientes datos representan los pesos de 10 mujeres, registrados antes de usar un anticonceptivo oral y después de usar un anticonceptivo oral por seis meses. Utilice la prueba de los signos de Wilcoxon para determinar si existe una diferencia significativa entre los pesos registrados bajo las dos condiciones. Utilice $\alpha = .05$.

Sin anticonceptivos	Anticonceptivos
128	130
145	145
103	108
139	135
125	124
160	163
144	153
122	129
134	128
161	169

10.8 Suponga que debe llevarse a cabo una prueba t de permutación para muestras independientes sobre los datos recabados de dos grupos que constan de cinco sujetos cada uno. ¿Cuántos estadísticos t conformarán la distribución muestral? ¿Cuántos estadísticos conformarían la distribución muestral si se usara la prueba de muestras independientes de Wilcoxon?

10.9 Los siguientes datos representan los tiempos de espera (expresados en minutos) de pacientes admitidos en salas de emergencia de instituciones públicas y privadas, para ser atendidos. Utilice la prueba de muestras independientes de Wilcoxon para comparar los tiempos de espera para las dos salas. Lleve a cabo una prueba de dos colas con $\alpha = .05$. ¿Cuál es su conclusión?

Pública	Privada
128	34
56	18
13	49
94	66
155	177
22	4
19	8
3	1
99	15
41	17

10.10 Suponga que se debe usar una prueba F de permutación de un ANOVA de un factor para analizar los datos obtenidos de un ensayo clínico en el que había cinco sujetos en cada uno de tres grupos. ¿Cuántos estadísticos F aparecerán en la distribución muestral?

10.11 Se asigna aleatoriamente a los estudiantes inscritos en una clase de introducción a la salud pública a una de tres secciones de instrucción. En la primera sección se usan métodos tradicionales de clase, en la segunda sección se usa un modo basado en Internet que incluye la participación del instructor, y en la tercera se enseña mediante clases grabadas en video sin la participación del instructor. A continuación se dan los promedios finales del curso de los 15 estudiantes que participaron en el estudio. Utilice la prueba Kruskal-Wallis, realizada con $\alpha = .05$ para determinar si los métodos de instrucción tuvieron diferentes efectos sobre los promedios del curso.

Clase	Internet	Video
87	90	100
77	76	99
56	44	89
92	58	88
80	97	70

10.12 La prueba exacta de Fisher es útil en particular para situaciones en que las frecuencias de celda esperadas son demasiado pequeñas para permitir el uso válido de la prueba chi cuadrada. Utilice la tabla de abajo para

a) calcular los valores mínimo y máximo de a,

b) listar todas las tablas de aleatorización,

c) calcular la probabilidad asociada con cada tabla, y

d) efectuar pruebas de la hipótesis nula

$$H_0 : \pi_1 = \pi_2$$

contra las alternativas

$$H_A : \pi_1 \neq \pi_2$$

y

$$H_A : \pi_1 < \pi_2$$

en el nivel .05 de significancia.

	+	−
Grupo 1	6	4
Grupo 2	3	2

A. Las siguientes preguntas se refieren al estudio de caso A (página 469).

10.13 Los autores señalan que se usaron métodos paramétricos y no paramétricos para analizar datos sobre salud ocular. Durante la discusión de los resultados de estos análisis afirman: "Las diferencias, si se encuentran, entre datos paramétricos y no paramétricos son consistentes, y la potencia de las pruebas no paramétricas es reducida". Comente este planteamiento.

B. Las siguientes preguntas se refieren al estudio de caso B (página 470).

10.14 Suponga que los investigadores que realizan el análisis inexpuesto en la página 293 (7.9) deciden llevar a cabo el análisis utilizando una prueba no paramétrica. ¿Qué pruebas podrían usarse para este fin? ¿Qué hipótesis nulas estarían asociadas con las pruebas que usted sugiere?

D. Las siguientes preguntas se refieren al estudio de caso D (página 471).

10.15 Los investigadores usaron pruebas *t* de muestras independientes para demostrar que los pacientes infectados con VIH tuvieron puntuaciones NPZ-8 y PBV significativamente más bajas que cinco sujetos sanos. Dada la distribución de las puntuaciones NPZ-8 de la población no normal, algunos investigadores podrían preferir utilizar un procedimiento no paramétrico para tomar estas decisiones.

a) ¿Cuáles pruebas no paramétricas podrían usarse para este fin?

b) Utilice una prueba no paramétrica de dos colas para comparar los datos NPZ-8 y PBV en pacientes infectados con VIH y sanos en el nivel .05.

c) ¿Se violó algún supuesto de las pruebas no paramétricas que usted utilizó en 10.15b? ¿Por qué?

d) ¿Usó una prueba exacta o aproximada para los análisis en 10.15b? ¿Por qué?

e) Explique cómo podría haberse realizado una versión exacta de las pruebas efectuadas en 10.15b?

10.16 Los investigadores reportan una correlación de orden de rangos de Spearman de −.50 entre PBV y NPZ-8 para los 20 participantes del estudio.

a) Calcule la correlación de orden de rangos de Spearman entre PBV y NPZ-8 para los 20 participantes del estudio. ¿Se obtiene el resultado reportado por los investigadores?

b) Calcule la correlación de orden de rangos de Spearman entre PBV y NPZ-8 para los 15 participantes del estudio infectados con VIH. ¿Hizo alguna diferencia la exclusión de los 5 participantes sanos del estudio?

c) Pruebe los coeficientes obtenidos en 10.16a y 10.16b para determinar la significancia. Utilice pruebas de dos colas con $\alpha = .05$

d) Dado el objetivo de este estudio, ¿cuál de los coeficientes de correlación antes mencionados cree usted que es el más apropiado?

10.17 Utilice una prueba no paramétrica para probar la hipótesis de que las puntuaciones NPZ-8 de los sujetos infectados con VIH que tienen evaluaciones positivas de la etapa ADC, de los sujetos infectados con VIH que tienen evaluaciones negativas de la etapa ADC y de los sujetos que dieron negativo para VIH vienen de poblaciones idénticas. Interprete el resultado. Compare este resultado con el obtenido para 7.14 de la página 294.

K. Las siguientes preguntas se refieren al estudio de caso K (página 474).

10.18 Utilice las medias de la escuela de la tabla 1 para llevar a cabo una prueba no paramétrica con el fin de determinar si las medias escolares post-prueba difieren significativamente de las medias pre-prueba.

10.19 ¿Es probable que el supuesto de normalidad de la población fuera un asunto de consideración si estos datos se analizaran mediante una prueba *t* de muestras pareadas? Dé las razones para su respuesta.

N. Las siguientes preguntas se refieren al estudio de caso N (página 476).

10.20 Realice una prueba exacta de Fisher para cada uno de los factores potenciales de riesgo. (Se puede ahorrar mucho esfuerzo con algo de cancelación creativa, como se demostró en la página 82). Interprete los resultados.

O. Las siguientes preguntas se refieren al estudio de caso O (página 477).

10.21 ¿Está usted de acuerdo con el segundo punto? Explique.

10.22 ¿Es estrictamente cierto el comentario hecho en el tercer punto sobre las pruebas no paramétricas en relación con el orden de rangos? Explique.

10.23 ¿Es cierta la afirmación hecha en el sexto punto? Explique.

10.24 ¿El ejemplo dado en el séptimo punto constituye una aplicación legítima de la prueba Kruskal-Wallis? Explique.

10.25 ¿El octavo punto es estrictamente cierto? Explique.

APÉNDICE A

Tabla de la curva normal

Áreas debajo de la curva normal estándar

Z (1)	(2)	(3)	Z (1)	(2)	(3)	Z (1)	(2)	(3)
0.00	.0000	.5000	0.26	.1026	.3974	0.52	.1985	.3015
0.01	.0040	.4960	0.27	.1064	.3936	0.53	.2019	.2981
0.02	.0080	.4920	0.28	.1103	.3897	0.54	.2054	.2946
0.03	.0120	.4880	0.29	.1141	.3859	0.55	.2088	.2912
0.04	.0160	.4840	0.30	.1179	.3821	0.56	.2123	.2877
0.05	.0199	.4801	0.31	.1217	.3783	0.57	.2157	.2843
0.06	.0239	.4761	0.32	.1255	.3745	0.58	.2190	.2810
0.07	.0279	.4721	0.33	.1293	.3707	0.59	.2224	.2776
0.08	.0319	.4681	0.34	.1331	.3669	0.60	.2257	.2743
0.09	.0359	.4641	0.35	.1368	.3632	0.61	.2291	.2709
0.10	.0398	.4602	0.36	.1406	.3594	0.62	.2324	.2676
0.11	.0438	.4562	0.37	.1443	.3557	0.63	.2357	.2643
0.12	.0478	.4522	0.38	.1480	.3520	0.64	.2389	.2611
0.13	.0517	.4483	0.39	.1517	.3483	0.65	.2422	.2578
0.14	.0557	.4443	0.40	.1554	.3446	0.66	.2454	.2546
0.15	.0596	.4404	0.41	.1591	.3409	0.67	.2486	.2514
0.16	.0636	.4364	0.42	.1628	.3372	0.68	.2517	.2483
0.17	.0675	.4325	0.43	.1664	.3336	0.69	.2549	.2451
0.18	.0714	.4286	0.44	.1700	.3300	0.70	.2580	.2420
0.19	.0753	.4247	0.45	.1736	.3264	0.71	.2611	.2389
0.20	.0793	.4207	0.46	.1772	.3228	0.72	.2642	.2358
0.21	.0832	.4168	0.47	.1808	.3192	0.73	.2673	.2327
0.22	.0871	.4129	0.48	.1844	.3156	0.74	.2703	.2297
0.23	.0910	.4090	0.49	.1879	.3121	0.75	.2734	.2266
0.24	.0948	.4052	0.50	.1915	.3085	0.76	.2764	.2236
0.25	.0987	.4013	0.51	.1950	.3050	0.77	.2793	.2207

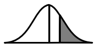
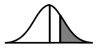
Áreas debajo de la curva normal estándar

Z (1)	(2)	(3)	Z (1)	(2)	(3)	Z (1)	(2)	(3)
0.78	.2823	.2177	1.04	.3508	.1492	1.30	.4032	.0968
0.79	.2852	.2148	1.05	.3531	.1469	1.31	.4049	.0951
0.80	.2881	.2119	1.06	.3554	.1446	1.32	.4066	.0934
0.81	.2910	.2090	1.07	.3577	.1423	1.33	.4082	.0918
0.82	.2939	.2061	1.08	.3599	.1401	1.34	.4099	.0901
0.83	.2967	.2033	1.09	.3621	.1379	1.35	.4115	.0885
0.84	.2995	.2005	1.10	.3643	.1357	1.36	.4131	.0869
0.85	.3023	.1977	1.11	.3665	.1335	1.37	.4147	.0853
0.86	.3051	.1949	1.12	.3686	.1314	1.38	.4162	.0838
0.87	.3078	.1922	1.13	.3708	.1292	1.39	.4177	.0823
0.88	.3106	.1894	1.14	.3729	.1271	1.40	.4192	.0808
0.89	.3133	.1867	1.15	.3749	.1251	1.41	.4207	.0793
0.90	.3159	.1841	1.16	.3770	.1230	1.42	.4222	.0778
0.91	.3186	.1814	1.17	.3790	.1210	1.43	.4236	.0764
0.92	.3212	.1788	1.18	.3810	.1190	1.44	.4251	.0749
0.93	.3238	.1762	1.19	.3830	.1170	1.45	.4265	.0735
0.94	.3264	.1736	1.20	.3849	.1151	1.46	.4279	.0721
0.95	.3289	.1711	1.21	.3869	.1131	1.47	.4292	.0708
0.96	.3315	.1685	1.22	.3888	.1112	1.48	.4306	.0694
0.97	.3340	.1660	1.23	.3907	.1093	1.49	.4319	.0681
0.98	.3365	.1635	1.24	.3925	.1075	1.50	.4332	.0668
0.99	.3389	.1611	1.25	.3944	.1056	1.51	.4345	.0655
1.00	.3413	.1587	1.26	.3962	.1038	1.52	.4357	.0643
1.01	.3438	.1562	1.27	.3980	.1020	1.53	.4370	.0630
1.02	.3461	.1539	1.28	.3997	.1003	1.54	.4382	.0618
1.03	.3485	.1515	1.29	.4015	.0985	1.55	.4394	.0606

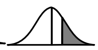
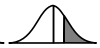
Áreas debajo de la curva normal estándar

Z (1)	(2)	(3)	Z (1)	(2)	(3)	Z (1)	(2)	(3)
1.56	.4406	.0594	1.82	.4656	.0344	2.08	.4812	.0188
1.57	.4418	.0582	1.83	.4664	.0336	2.09	.4817	.0183
1.58	.4429	.0571	1.84	.4671	.0329	2.10	.4821	.0179
1.59	.4441	.0559	1.85	.4678	.0322	2.11	.4826	.0174
1.60	.4452	.0548	1.86	.4686	.0314	2.12	.4830	.0170
1.61	.4463	.0537	1.87	.4693	.0307	2.13	.4834	.0166
1.62	.4474	.0526	1.88	.4699	.0301	2.14	.4838	.0162
1.63	.4484	.0516	1.89	.4706	.0294	2.15	.4842	.0158
1.64	.4495	.0505	1.90	.4713	.0287	2.16	.4846	.0154
1.65	.4505	.0495	1.91	.4719	.0281	2.17	.4850	.0150
1.66	.4515	.0485	1.92	.4726	.0274	2.18	.4854	.0146
1.67	.4525	.0475	1.93	.4732	.0268	2.19	.4857	.0143
1.68	.4535	.0465	1.94	.4738	.0262	2.20	.4861	.0139
1.69	.4545	.0455	1.95	.4744	.0256	2.21	.4864	.0136
1.70	.4554	.0446	1.96	.4750	.0250	2.22	.4868	.0132
1.71	.4564	.0436	1.97	.4756	.0244	2.23	.4871	.0129
1.72	.4573	.0427	1.98	.4761	.0239	2.24	.4875	.0125
1.73	.4582	.0418	1.99	.4767	.0233	2.25	.4878	.0122
1.74	.4591	.0409	2.00	.4772	.0228	2.26	.4881	.0119
1.75	.4599	.0401	2.01	.4778	.0222	2.27	.4884	.0116
1.76	.4608	.0392	2.02	.4783	.0217	2.28	.4887	.0113
1.77	.4616	.0384	2.03	.4788	.0212	2.29	.4890	.0110
1.78	.4625	.0375	2.04	.4793	.0207	2.30	.4893	.0107
1.79	.4633	.0367	2.05	.4798	.0202	2.31	.4896	.0104
1.80	.4641	.0359	2.06	.4803	.0197	2.32	.4898	.0102
1.81	.4649	.0351	2.07	.4808	.0192	2.33	.4901	.0099

Áreas debajo de la curva normal estándar

Z (1)	 (2)	 (3)	Z (1)	 (2)	 (3)	Z (1)	 (2)	 (3)
2.34	.4904	.0096	2.60	.4953	.0047	2.86	.4979	.0021
2.35	.4906	.0094	2.61	.4955	.0045	2.87	.4979	.0021
2.36	.4909	.0091	2.62	.4956	.0044	2.88	.4980	.0020
2.37	.4911	.0089	2.63	.4957	.0043	2.89	.4981	.0019
2.38	.4913	.0087	2.64	.4959	.0041	2.90	.4981	.0019
2.39	.4916	.0084	2.65	.4960	.0040	2.91	.4982	.0018
2.40	.4918	.0082	2.66	.4961	.0039	2.92	.4982	.0018
2.41	.4920	.0080	2.67	.4962	.0038	2.93	.4983	.0017
2.42	.4922	.0078	2.68	.4963	.0037	2.94	.4984	.0016
2.43	.4925	.0075	2.69	.4964	.0036	2.95	.4984	.0016
2.44	.4927	.0073	2.70	.4965	.0035	2.96	.4985	.0015
2.45	.4929	.0071	2.71	.4966	.0034	2.97	.4985	.0015
2.46	.4931	.0069	2.72	.4967	.0033	2.98	.4986	.0014
2.47	.4932	.0068	2.73	.4968	.0032	2.99	.4986	.0014
2.48	.4934	.0066	2.74	.4969	.0031	3.00	.4987	.0013
2.49	.4936	.0064	2.75	.4970	.0030	3.01	.4987	.0013
2.50	.4938	.0062	2.76	.4971	.0029	3.02	.4987	.0013
2.51	.4940	.0060	2.77	.4972	.0028	3.03	.4988	.0012
2.52	.4941	.0059	2.78	.4973	.0027	3.04	.4988	.0012
2.53	.4943	.0057	2.79	.4974	.0026	3.05	.4989	.0011
2.54	.4945	.0055	2.80	.4974	.0026	3.06	.4989	.0011
2.55	.4946	.0054	2.81	.4975	.0025	3.07	.4989	.0011
2.56	.4948	.0052	2.82	.4976	.0024	3.08	.4990	.0010
2.57	.4949	.0051	2.83	.4977	.0023	3.09	.4990	.0010
2.58	.4951	.0049	2.84	.4977	.0023	3.10	.4990	.0010
2.59	.4952	.0048	2.85	.4978	.0022	3.11	.4991	.0009

Áreas debajo de la curva normal estándar

Z (1)	 (2)	 (3)	Z (1)	 (2)	 (3)
3.12	.4991	.0009	3.38	.4996	.0004
3.13	.4991	.0009	3.39	.4997	.0003
3.14	.4992	.0008	3.40	.4997	.0003
3.15	.4992	.0008	3.41	.4997	.0003
3.16	.4992	.0008	3.42	.4997	.0003
3.17	.4992	.0008	3.43	.4997	.0003
3.18	.4993	.0007	3.44	.4997	.0003
3.19	.4993	.0007	3.45	.4997	.0003
3.20	.4993	.0007	3.46	.4997	.0003
3.21	.4993	.0007	3.47	.4997	.0003
3.22	.4994	.0006	3.48	.4997	.0003
3.23	.4994	.0006	3.49	.4998	.0002
3.24	.4994	.0006	3.50	.4998	.0002
3.25	.4994	.0006			
3.26	.4994	.0006			
3.27	.4995	.0005			
3.28	.4995	.0005			
3.29	.4995	.0005			
3.30	.4995	.0005			
3.31	.4995	.0005			
3.32	.4995	.0005			
3.33	.4996	.0004			
3.34	.4996	.0004			
3.35	.4996	.0004			
3.36	.4996	.0004			
3.37	.4996	.0004			

A P É N D I C E B

Valores críticos de la distribución t de Student

Percentiles de la distribución t de Student

IC unilateral	.900	.950	.975	.990	.995
IC bilateral	.800	.900	.950	.980	.990
PH de una cola	.100	.050	.025	.010	.005
PH de dos colas	.200	.100	.050	.020	.010
Grados de libertad					
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787

Percentiles de la distribución t de Student

IC unilateral	.900	.950	.975	.990	.995
IC bilateral	.800	.900	.950	.980	.990
PH de una cola	.100	.050	.025	.010	.005
PH de dos colas	.200	.100	.050	.020	.010
Grados de libertad					
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.691	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
36	1.306	1.688	2.028	2.434	2.719
37	1.305	1.687	2.026	2.431	2.715
38	1.304	1.686	2.024	2.429	2.712
39	1.304	1.685	2.023	2.426	2.708
40	1.303	1.684	2.021	2.423	2.704
41	1.303	1.683	2.020	2.421	2.701
42	1.302	1.682	2.018	2.418	2.698
43	1.302	1.681	2.017	2.416	2.695
44	1.301	1.680	2.015	2.414	2.692
45	1.301	1.679	2.014	2.412	2.690
46	1.300	1.679	2.013	2.410	2.687
47	1.300	1.678	2.012	2.408	2.685
48	1.299	1.677	2.011	2.407	2.682
49	1.299	1.677	2.010	2.405	2.680
50	1.299	1.676	2.009	2.403	2.678

Percentiles de la distribución t de Student

IC unilateral	.900	.950	.975	.990	.995
IC bilateral	.800	.900	.950	.980	.990
PH de una cola	.100	.050	.025	.010	.005
PH de dos colas	.200	.100	.050	.020	.010
Grados de libertad					
51	1.298	1.675	2.008	2.402	2.676
52	1.298	1.675	2.007	2.400	2.674
53	1.298	1.674	2.006	2.399	2.672
54	1.297	1.674	2.005	2.397	2.670
55	1.297	1.673	2.004	2.396	2.668
56	1.297	1.673	2.003	2.395	2.667
57	1.297	1.672	2.002	2.394	2.665
58	1.296	1.672	2.002	2.392	2.663
59	1.296	1.671	2.001	2.391	2.662
60	1.296	1.671	2.000	2.390	2.660
61	1.296	1.670	2.000	2.389	2.659
62	1.295	1.670	1.999	2.388	2.657
63	1.295	1.669	1.998	2.387	2.656
64	1.295	1.669	1.998	2.386	2.655
65	1.295	1.669	1.997	2.385	2.654
66	1.295	1.668	1.997	2.384	2.652
67	1.294	1.668	1.996	2.383	2.651
68	1.294	1.668	1.995	2.382	2.650
69	1.294	1.667	1.995	2.382	2.649
70	1.294	1.667	1.994	2.381	2.648
71	1.294	1.667	1.994	2.380	2.647
72	1.293	1.666	1.993	2.379	2.646
73	1.293	1.666	1.993	2.379	2.645
74	1.293	1.666	1.993	2.378	2.644
75	1.293	1.665	1.992	2.377	2.643

Percentiles de la distribución t de Student

IC unilateral	.900	.950	.975	.990	.995
IC bilateral	.800	.900	.950	.980	.990
PH de una cola	.100	.050	.025	.010	.005
PH de dos colas	.200	.100	.050	.020	.010
Grados de libertad					
76	1.293	1.665	1.992	2.376	2.642
77	1.293	1.665	1.991	2.376	2.641
78	1.292	1.665	1.991	2.375	2.640
79	1.292	1.664	1.990	2.374	2.640
80	1.292	1.664	1.990	2.374	2.639
81	1.292	1.664	1.990	2.373	2.638
82	1.292	1.664	1.989	2.373	2.637
83	1.292	1.663	1.989	2.372	2.636
84	1.292	1.663	1.989	2.372	2.636
85	1.292	1.663	1.988	2.371	2.635
86	1.291	1.663	1.988	2.370	2.634
87	1.291	1.663	1.988	2.370	2.634
88	1.291	1.662	1.987	2.369	2.633
89	1.291	1.662	1.987	2.369	2.632
90	1.291	1.662	1.987	2.368	2.632
91	1.291	1.662	1.986	2.368	2.631
92	1.291	1.662	1.986	2.368	2.630
93	1.291	1.661	1.986	2.367	2.630
94	1.291	1.661	1.986	2.367	2.629
95	1.291	1.661	1.985	2.366	2.629
96	1.290	1.661	1.985	2.366	2.628
97	1.290	1.661	1.985	2.365	2.627
98	1.290	1.661	1.984	2.365	2.627
99	1.290	1.660	1.984	2.365	2.626
100	1.290	1.660	1.984	2.364	2.626

Percentiles de la distribución t de Student

IC unilateral	.900	.950	.975	.990	.995
IC bilateral	.800	.900	.950	.980	.990
PH de una cola	.100	.050	.025	.010	.005
PH de dos colas	.200	.100	.050	.020	.010
Grados de libertad					
101	1.290	1.660	1.984	2.364	2.625
102	1.290	1.660	1.983	2.363	2.625
103	1.290	1.660	1.983	2.363	2.624
104	1.290	1.660	1.983	2.363	2.624
105	1.290	1.659	1.983	2.362	2.623
106	1.290	1.659	1.983	2.362	2.623
107	1.290	1.659	1.982	2.362	2.623
108	1.289	1.659	1.982	2.361	2.622
109	1.289	1.659	1.982	2.361	2.622
110	1.289	1.659	1.982	2.361	2.621
111	1.289	1.659	1.982	2.360	2.621
112	1.289	1.659	1.981	2.360	2.620
113	1.289	1.658	1.981	2.360	2.620
114	1.289	1.658	1.981	2.360	2.620
115	1.289	1.658	1.981	2.359	2.619
116	1.289	1.658	1.981	2.359	2.619
117	1.289	1.658	1.980	2.359	2.619
118	1.289	1.658	1.980	2.358	2.618
119	1.289	1.658	1.980	2.358	2.618
120	1.289	1.658	1.980	2.358	2.617
121	1.289	1.658	1.980	2.358	2.617
122	1.289	1.657	1.980	2.357	2.617
123	1.288	1.657	1.979	2.357	2.616
124	1.288	1.657	1.979	2.357	2.616
∞	1.282	1.645	1.960	2.326	2.576

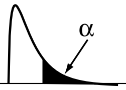
APÉNDICE C

Valores críticos de la distribución F

Valores críticos de la distribución F, $\alpha = .10$

IC unilateral = .90

IC bilateral = .80



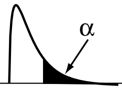
$\alpha = .10$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85

Valores críticos de la distribución F, $\alpha = .10$

IC unilateral = .90

IC bilateral = .80



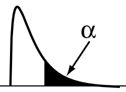
$\alpha = .10$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
31	2.87	2.48	2.27	2.14	2.04	1.97	1.92	1.88	1.84
32	2.87	2.48	2.26	2.13	2.04	1.97	1.91	1.87	1.83
33	2.86	2.47	2.26	2.12	2.03	1.96	1.91	1.86	1.83
34	2.86	2.47	2.25	2.12	2.02	1.96	1.90	1.86	1.82
35	2.85	2.46	2.25	2.11	2.02	1.95	1.90	1.85	1.82
36	2.85	2.46	2.24	2.11	2.01	1.94	1.89	1.85	1.81
37	2.85	2.45	2.24	2.10	2.01	1.94	1.89	1.84	1.81
38	2.84	2.45	2.23	2.10	2.01	1.94	1.88	1.84	1.80
39	2.84	2.44	2.23	2.09	2.00	1.93	1.88	1.83	1.80
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79
41	2.83	2.44	2.22	2.09	1.99	1.92	1.87	1.82	1.79
42	2.83	2.43	2.22	2.08	1.99	1.92	1.86	1.82	1.78
43	2.83	2.43	2.22	2.08	1.99	1.92	1.86	1.82	1.78
44	2.82	2.43	2.21	2.08	1.98	1.91	1.86	1.81	1.78
45	2.82	2.42	2.21	2.07	1.98	1.91	1.85	1.81	1.77
46	2.82	2.42	2.21	2.07	1.98	1.91	1.85	1.81	1.77
47	2.82	2.42	2.20	2.07	1.97	1.90	1.85	1.80	1.77
48	2.81	2.42	2.20	2.07	1.97	1.90	1.85	1.80	1.77
49	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76
55	2.80	2.40	2.19	2.05	1.95	1.88	1.83	1.78	1.75
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
65	2.78	2.39	2.17	2.03	1.94	1.87	1.81	1.77	1.73
70	2.78	2.38	2.16	2.03	1.93	1.86	1.80	1.76	1.72
80	2.77	2.37	2.15	2.02	1.92	1.8	1.79	1.75	1.71
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69
150	2.74	2.34	2.12	1.98	1.89	1.81	1.76	1.71	1.67
350	2.72	2.32	2.10	1.96	1.86	1.79	1.73	1.69	1.65
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63

Valores críticos de la distribución F, $\alpha = .10$

IC unilateral = .90

IC bilateral = .80



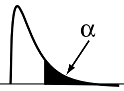
$\alpha = .10$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
2	9.39	9.41	9.42	9.43	9.44	9.47	9.48	9.49	9.49
3	5.23	5.22	5.20	5.20	5.18	5.15	5.14	5.14	5.13
4	3.92	3.90	3.88	3.86	3.84	3.80	3.78	3.77	3.76
5	3.30	3.27	3.25	3.23	3.21	3.15	3.13	3.11	3.10
6	2.94	2.90	2.88	2.86	2.84	2.77	2.75	2.73	2.72
7	2.70	2.67	2.64	2.62	2.59	2.52	2.50	2.48	2.47
8	2.54	2.50	2.48	2.45	2.42	2.35	2.32	2.30	2.29
9	2.42	2.38	2.35	2.33	2.30	2.22	2.19	2.17	2.16
10	2.32	2.28	2.26	2.23	2.20	2.12	2.09	2.06	2.06
11	2.25	2.21	2.18	2.16	2.12	2.04	2.01	1.98	1.97
12	2.19	2.15	2.12	2.09	2.06	1.97	1.94	1.91	1.90
13	2.14	2.10	2.07	2.04	2.01	1.92	1.88	1.86	1.85
14	2.10	2.05	2.02	2.00	1.96	1.87	1.83	1.81	1.80
15	2.06	2.02	1.99	1.96	1.92	1.83	1.79	1.77	1.76
16	2.03	1.99	1.95	1.93	1.89	1.79	1.76	1.73	1.72
17	2.00	1.96	1.93	1.90	1.86	1.76	1.73	1.70	1.69
18	1.98	1.93	1.90	1.87	1.84	1.74	1.70	1.67	1.66
19	1.96	1.91	1.88	1.85	1.81	1.71	1.67	1.64	1.63
20	1.94	1.89	1.86	1.83	1.79	1.69	1.65	1.62	1.61
21	1.92	1.87	1.84	1.81	1.78	1.67	1.63	1.60	1.59
22	1.90	1.86	1.83	1.80	1.76	1.65	1.61	1.58	1.57
23	1.89	1.84	1.81	1.78	1.74	1.64	1.59	1.56	1.55
24	1.88	1.83	1.80	1.77	1.73	1.62	1.58	1.55	1.53
25	1.87	1.82	1.79	1.76	1.72	1.61	1.56	1.53	1.52
26	1.86	1.81	1.77	1.75	1.71	1.59	1.55	1.52	1.50
27	1.85	1.80	1.76	1.74	1.70	1.58	1.54	1.50	1.49
28	1.84	1.79	1.75	1.73	1.69	1.57	1.53	1.49	1.48
29	1.83	1.78	1.75	1.72	1.68	1.56	1.52	1.48	1.47
30	1.82	1.77	1.74	1.71	1.67	1.55	1.51	1.47	1.46

Valores críticos de la distribución F, $\alpha = .10$

IC unilateral = .90

IC bilateral = .80



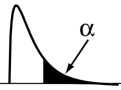
$\alpha = .10$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
31	1.81	1.77	1.73	1.70	1.66	1.54	1.50	1.46	1.45
32	1.81	1.76	1.72	1.69	1.65	1.53	1.49	1.45	1.44
33	1.80	1.75	1.72	1.69	1.64	1.53	1.48	1.44	1.43
34	1.79	1.75	1.71	1.68	1.64	1.52	1.47	1.44	1.42
35	1.79	1.74	1.70	1.67	1.63	1.51	1.47	1.43	1.41
36	1.78	1.73	1.70	1.67	1.63	1.51	1.46	1.42	1.40
37	1.78	1.73	1.69	1.66	1.62	1.50	1.45	1.41	1.40
38	1.77	1.72	1.69	1.66	1.61	1.49	1.45	1.41	1.39
39	1.77	1.72	1.68	1.65	1.61	1.49	1.44	1.40	1.38
40	1.76	1.71	1.68	1.65	1.61	1.48	1.43	1.39	1.38
41	1.76	1.71	1.67	1.64	1.60	1.48	1.43	1.39	1.37
42	1.75	1.71	1.67	1.64	1.60	1.47	1.42	1.38	1.37
43	1.75	1.70	1.67	1.64	1.59	1.47	1.42	1.38	1.36
44	1.75	1.70	1.66	1.63	1.59	1.46	1.41	1.37	1.35
45	1.74	1.70	1.66	1.63	1.58	1.46	1.41	1.37	1.35
46	1.74	1.69	1.65	1.63	1.58	1.46	1.40	1.36	1.34
47	1.74	1.69	1.65	1.62	1.58	1.45	1.40	1.36	1.34
48	1.73	1.69	1.65	1.62	1.57	1.45	1.40	1.35	1.34
49	1.73	1.68	1.65	1.62	1.57	1.44	1.39	1.35	1.33
50	1.73	1.68	1.64	1.61	1.57	1.44	1.39	1.35	1.33
55	1.72	1.67	1.63	1.60	1.55	1.43	1.37	1.33	1.31
60	1.71	1.66	1.62	1.59	1.54	1.41	1.36	1.31	1.29
65	1.70	1.65	1.61	1.58	1.53	1.40	1.35	1.30	1.28
70	1.69	1.64	1.60	1.57	1.53	1.39	1.34	1.29	1.27
80	1.68	1.63	1.59	1.56	1.51	1.38	1.32	1.27	1.24
100	1.66	1.61	1.57	1.54	1.49	1.35	1.29	1.24	1.21
150	1.64	1.59	1.55	1.52	1.47	1.33	1.26	1.20	1.17
350	1.62	1.56	1.52	1.49	1.44	1.29	1.22	1.15	1.11
∞	1.60	1.55	1.50	1.47	1.42	1.26	1.18	1.10	1.00

Valores críticos de la distribución F, $\alpha = .05$

IC unilateral = .95

IC bilateral = .90



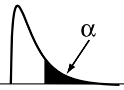
$\alpha = .05$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21

Valores críticos de la distribución F, $\alpha = .05$

IC unilateral = .95

IC bilateral = .90



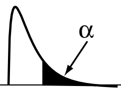
$\alpha = .05$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
31	4.16	3.30	2.91	2.68	2.52	2.41	2.32	2.25	2.20
32	4.15	3.29	2.90	2.67	2.51	2.40	2.31	2.24	2.19
33	4.14	3.28	2.89	2.66	2.50	2.39	2.30	2.23	2.18
34	4.13	3.28	2.88	2.65	2.49	2.38	2.29	2.23	2.17
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16
36	4.11	3.26	2.87	2.63	2.48	2.36	2.28	2.21	2.15
37	4.11	3.25	2.86	2.63	2.47	2.36	2.27	2.20	2.14
38	4.10	3.24	2.85	2.62	2.46	2.35	2.26	2.19	2.14
39	4.09	3.24	2.85	2.61	2.46	2.34	2.26	2.19	2.13
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
41	4.08	3.23	2.83	2.60	2.44	2.33	2.24	2.17	2.12
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11
43	4.07	3.21	2.82	2.59	2.43	2.32	2.23	2.16	2.11
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.15	2.09
47	4.05	3.20	2.80	2.57	2.41	2.30	2.21	2.14	2.09
48	4.04	3.19	2.80	2.57	2.41	2.29	2.21	2.14	2.08
49	4.04	3.19	2.79	2.56	2.40	2.29	2.20	2.13	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07
55	4.02	3.16	2.77	2.54	2.38	2.27	2.18	2.11	2.06
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.03
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97
150	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94
350	3.87	3.02	2.63	2.40	2.24	2.12	2.04	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

Valores críticos de la distribución F, $\alpha = .05$

IC unilateral = .95

IC bilateral = .90



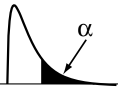
$\alpha = .05$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
2	19.40	19.41	19.42	19.43	19.45	19.48	19.49	19.49	19.50
3	8.79	8.74	8.71	8.69	8.66	8.58	8.55	8.53	8.53
4	5.96	5.91	5.87	5.84	5.80	5.70	5.66	5.64	5.63
5	4.74	4.68	4.64	4.60	4.56	4.44	4.41	4.38	4.36
6	4.06	4.00	3.96	3.92	3.87	3.75	3.71	3.68	3.67
7	3.64	3.57	3.53	3.49	3.44	3.32	3.27	3.24	3.23
8	3.35	3.28	3.24	3.20	3.15	3.02	2.97	2.94	2.93
9	3.14	3.07	3.03	2.99	2.94	2.80	2.76	2.72	2.71
10	2.98	2.91	2.86	2.83	2.77	2.64	2.59	2.55	2.54
11	2.85	2.79	2.74	2.70	2.65	2.51	2.46	2.42	2.40
12	2.75	2.69	2.64	2.60	2.54	2.40	2.35	2.31	2.30
13	2.67	2.60	2.55	2.51	2.46	2.31	2.26	2.22	2.21
14	2.60	2.53	2.48	2.44	2.39	2.24	2.19	2.15	2.13
15	2.54	2.48	2.42	2.38	2.33	2.18	2.12	2.08	2.07
16	2.49	2.42	2.37	2.33	2.28	2.12	2.07	2.03	2.01
17	2.45	2.38	2.33	2.29	2.23	2.08	2.02	1.98	1.96
18	2.41	2.34	2.29	2.25	2.19	2.04	1.98	1.93	1.92
19	2.38	2.31	2.26	2.21	2.16	2.00	1.94	1.90	1.88
20	2.35	2.28	2.22	2.18	2.12	1.97	1.91	1.86	1.84
21	2.32	2.25	2.20	2.16	2.10	1.94	1.88	1.83	1.81
22	2.30	2.23	2.17	2.13	2.07	1.91	1.85	1.80	1.78
23	2.27	2.20	2.15	2.11	2.05	1.88	1.82	1.78	1.76
24	2.25	2.18	2.13	2.09	2.03	1.86	1.80	1.75	1.73
25	2.24	2.16	2.11	2.07	2.01	1.84	1.78	1.73	1.71
26	2.22	2.15	2.09	2.05	1.99	1.82	1.76	1.71	1.69
27	2.20	2.13	2.08	2.04	1.97	1.81	1.74	1.69	1.67
28	2.19	2.12	2.06	2.02	1.96	1.79	1.73	1.68	1.65
29	2.18	2.10	2.05	2.01	1.94	1.77	1.71	1.66	1.64
30	2.16	2.09	2.04	1.99	1.93	1.76	1.70	1.64	1.62

Valores críticos de la distribución F, $\alpha = .05$

IC unilateral = .95

IC bilateral = .90



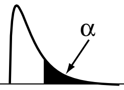
$\alpha = .05$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
31	2.15	2.08	2.03	1.98	1.92	1.75	1.68	1.63	1.61
32	2.14	2.07	2.01	1.97	1.91	1.74	1.67	1.62	1.59
33	2.13	2.06	2.00	1.96	1.90	1.72	1.66	1.60	1.58
34	2.12	2.05	1.99	1.95	1.89	1.71	1.65	1.59	1.57
35	2.11	2.04	1.99	1.94	1.88	1.70	1.63	1.58	1.56
36	2.11	2.03	1.98	1.93	1.87	1.69	1.62	1.57	1.55
37	2.10	2.02	1.97	1.93	1.86	1.68	1.62	1.56	1.54
38	2.09	2.02	1.96	1.92	1.85	1.68	1.61	1.55	1.53
39	2.08	2.01	1.95	1.91	1.85	1.67	1.60	1.54	1.52
40	2.08	2.00	1.95	1.90	1.84	1.66	1.59	1.53	1.51
41	2.07	2.00	1.94	1.90	1.83	1.65	1.58	1.52	1.50
42	2.06	1.99	1.94	1.89	1.83	1.65	1.57	1.52	1.49
43	2.06	1.99	1.93	1.89	1.82	1.64	1.57	1.51	1.48
44	2.05	1.98	1.92	1.88	1.81	1.63	1.56	1.50	1.48
45	2.05	1.97	1.92	1.87	1.81	1.63	1.55	1.50	1.47
46	2.04	1.97	1.91	1.87	1.80	1.62	1.55	1.49	1.46
47	2.04	1.96	1.91	1.86	1.80	1.61	1.54	1.48	1.46
48	2.03	1.96	1.90	1.86	1.79	1.61	1.54	1.48	1.45
49	2.03	1.96	1.90	1.85	1.79	1.60	1.53	1.47	1.44
50	2.03	1.95	1.89	1.85	1.78	1.60	1.52	1.46	1.44
55	2.01	1.93	1.88	1.83	1.76	1.58	1.50	1.44	1.41
60	1.99	1.92	1.86	1.82	1.75	1.56	1.48	1.42	1.39
65	1.98	1.90	1.85	1.80	1.73	1.54	1.46	1.40	1.37
70	1.97	1.89	1.84	1.79	1.72	1.53	1.45	1.38	1.35
80	1.95	1.88	1.82	1.77	1.70	1.51	1.43	1.36	1.32
100	1.93	1.85	1.79	1.75	1.68	1.48	1.39	1.32	1.28
150	1.89	1.82	1.76	1.71	1.64	1.44	1.34	1.26	1.22
350	1.86	1.78	1.72	1.67	1.60	1.39	1.29	1.19	1.14
∞	1.83	1.75	1.69	1.64	1.57	1.35	1.24	1.13	1.00

Valores críticos de la distribución F, $\alpha = .025$

IC unilateral = .975

IC bilateral = .95



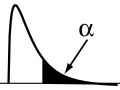
$\alpha = .025$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57

Valores críticos de la distribución F, $\alpha = .025$

IC unilateral = .975

IC bilateral = .95



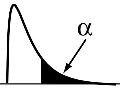
$\alpha = .025$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
31	5.55	4.16	3.57	3.23	3.01	2.85	2.73	2.64	2.56
32	5.53	4.15	3.56	3.22	3.00	2.84	2.71	2.62	2.54
33	5.51	4.13	3.54	3.20	2.98	2.82	2.70	2.61	2.53
34	5.50	4.12	3.53	3.19	2.97	2.81	2.69	2.59	2.52
35	5.48	4.11	3.52	3.18	2.96	2.80	2.68	2.58	2.50
36	5.47	4.09	3.50	3.17	2.94	2.78	2.66	2.57	2.49
37	5.46	4.08	3.49	3.16	2.93	2.77	2.65	2.56	2.48
38	5.45	4.07	3.48	3.15	2.92	2.76	2.64	2.55	2.47
39	5.43	4.06	3.47	3.14	2.91	2.75	2.63	2.54	2.46
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45
41	5.41	4.04	3.45	3.12	2.89	2.74	2.62	2.52	2.44
42	5.40	4.03	3.45	3.11	2.89	2.73	2.61	2.51	2.43
43	5.39	4.02	3.44	3.10	2.88	2.72	2.60	2.50	2.43
44	5.39	4.02	3.43	3.09	2.87	2.71	2.59	2.50	2.42
45	5.38	4.01	3.42	3.09	2.86	2.70	2.58	2.49	2.41
46	5.37	4.00	3.42	3.08	2.86	2.70	2.58	2.48	2.41
47	5.36	3.99	3.41	3.07	2.85	2.69	2.57	2.48	2.40
48	5.35	3.99	3.40	3.07	2.84	2.69	2.56	2.47	2.39
49	5.35	3.98	3.40	3.06	2.84	2.68	2.56	2.46	2.39
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38
55	5.31	3.95	3.36	3.03	2.81	2.65	2.53	2.43	2.36
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
65	5.26	3.91	3.32	2.99	2.77	2.61	2.49	2.39	2.32
70	5.25	3.89	3.31	2.97	2.75	2.59	2.47	2.38	2.30
80	5.22	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.28
100	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24
150	5.13	3.78	3.20	2.87	2.65	2.49	2.37	2.28	2.20
350	5.07	3.73	3.15	2.82	2.60	2.44	2.32	2.23	2.15
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11

Valores críticos de la distribución F, $\alpha = .025$

IC unilateral = .975

IC bilateral = .95



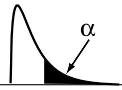
$\alpha = .025$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
2	39.40	39.41	39.43	39.44	39.45	39.48	39.49	39.50	39.50
3	14.42	14.34	14.28	14.23	14.17	14.01	13.96	13.92	13.90
4	8.84	8.75	8.68	8.63	8.56	8.38	8.32	8.28	8.26
5	6.62	6.52	6.46	6.40	6.33	6.14	6.08	6.03	6.02
6	5.46	5.37	5.30	5.24	5.17	4.98	4.92	4.87	4.85
7	4.76	4.67	4.60	4.54	4.47	4.28	4.21	4.16	4.14
8	4.30	4.20	4.13	4.08	4.00	3.81	3.74	3.69	3.67
9	3.96	3.87	3.80	3.74	3.67	3.47	3.40	3.35	3.33
10	3.72	3.62	3.55	3.50	3.42	3.22	3.15	3.10	3.08
11	3.53	3.43	3.36	3.30	3.23	3.03	2.96	2.90	2.88
12	3.37	3.28	3.21	3.15	3.07	2.87	2.80	2.75	2.72
13	3.25	3.15	3.08	3.03	2.95	2.74	2.67	2.62	2.60
14	3.15	3.05	2.98	2.92	2.84	2.64	2.56	2.51	2.49
15	3.06	2.96	2.89	2.84	2.76	2.55	2.47	2.42	2.40
16	2.99	2.89	2.82	2.76	2.68	2.47	2.40	2.34	2.32
17	2.92	2.82	2.75	2.70	2.62	2.41	2.33	2.27	2.25
18	2.87	2.77	2.70	2.64	2.56	2.35	2.27	2.21	2.19
19	2.82	2.72	2.65	2.59	2.51	2.30	2.22	2.16	2.13
20	2.77	2.68	2.60	2.55	2.46	2.25	2.17	2.11	2.09
21	2.73	2.64	2.56	2.51	2.42	2.21	2.13	2.07	2.04
22	2.70	2.60	2.53	2.47	2.39	2.17	2.09	2.03	2.00
23	2.67	2.57	2.50	2.44	2.36	2.14	2.06	1.99	1.97
24	2.64	2.54	2.47	2.41	2.33	2.11	2.02	1.96	1.94
25	2.61	2.51	2.44	2.38	2.30	2.08	2.00	1.93	1.91
26	2.59	2.49	2.42	2.36	2.28	2.05	1.97	1.90	1.88
27	2.57	2.47	2.39	2.34	2.25	2.03	1.94	1.88	1.85
28	2.55	2.45	2.37	2.32	2.23	2.01	1.92	1.86	1.83
29	2.53	2.43	2.36	2.30	2.21	1.99	1.90	1.84	1.81
30	2.51	2.41	2.34	2.28	2.20	1.97	1.88	1.81	1.79

Valores críticos de la distribución F, $\alpha = .025$

IC unilateral = .975

IC bilateral = .95



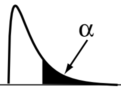
α = .025

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
31	2.50	2.40	2.32	2.26	2.18	1.95	1.86	1.80	1.77
32	2.48	2.38	2.31	2.25	2.16	1.93	1.85	1.78	1.75
33	2.47	2.37	2.29	2.23	2.15	1.92	1.83	1.76	1.73
34	2.45	2.35	2.28	2.22	2.13	1.90	1.82	1.75	1.72
35	2.44	2.34	2.27	2.21	2.12	1.89	1.80	1.73	1.70
36	2.43	2.33	2.25	2.20	2.11	1.88	1.79	1.72	1.69
37	2.42	2.32	2.24	2.18	2.10	1.87	1.77	1.70	1.67
38	2.41	2.31	2.23	2.17	2.09	1.85	1.76	1.69	1.66
39	2.40	2.30	2.22	2.16	2.08	1.84	1.75	1.68	1.65
40	2.39	2.29	2.21	2.15	2.07	1.83	1.74	1.67	1.64
41	2.38	2.28	2.20	2.15	2.06	1.82	1.73	1.66	1.63
42	2.37	2.27	2.20	2.14	2.05	1.81	1.72	1.65	1.62
43	2.36	2.26	2.19	2.13	2.04	1.80	1.71	1.64	1.61
44	2.36	2.26	2.18	2.12	2.03	1.80	1.70	1.63	1.60
45	2.35	2.25	2.17	2.11	2.03	1.79	1.69	1.62	1.59
46	2.34	2.24	2.17	2.11	2.02	1.78	1.69	1.61	1.58
47	2.33	2.23	2.16	2.10	2.01	1.77	1.68	1.60	1.57
48	2.33	2.23	2.15	2.09	2.01	1.77	1.67	1.59	1.56
49	2.32	2.22	2.15	2.09	2.00	1.76	1.66	1.59	1.55
50	2.32	2.22	2.14	2.08	1.99	1.75	1.66	1.58	1.55
55	2.29	2.19	2.11	2.05	1.97	1.72	1.62	1.55	1.51
60	2.27	2.17	2.09	2.03	1.94	1.70	1.60	1.52	1.48
65	2.25	2.15	2.07	2.01	1.93	1.68	1.58	1.49	1.46
70	2.24	2.14	2.06	2.00	1.91	1.66	1.56	1.47	1.44
80	2.21	2.11	2.03	1.97	1.88	1.63	1.53	1.44	1.40
100	2.18	2.08	2.00	1.94	1.85	1.59	1.48	1.39	1.35
150	2.13	2.03	1.95	1.89	1.80	1.54	1.42	1.32	1.27
350	2.09	1.98	1.90	1.84	1.75	1.48	1.35	1.23	1.17
∞	2.05	1.94	1.87	1.80	1.71	1.43	1.30	1.15	1.00

Valores críticos de la distribución F, $\alpha = .01$

IC unilateral = .99

IC bilateral = .98



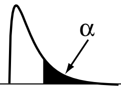
$\alpha = .01$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07

Valores críticos de la distribución F, $\alpha = .01$

IC unilateral = .99

IC bilateral = .98



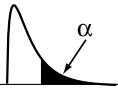
$\alpha = .01$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
31	7.53	5.36	4.48	3.99	3.67	3.45	3.28	3.15	3.04
32	7.50	5.34	4.46	3.97	3.65	3.43	3.26	3.13	3.02
33	7.47	5.31	4.44	3.95	3.63	3.41	3.24	3.11	3.00
34	7.44	5.29	4.42	3.93	3.61	3.39	3.22	3.09	2.98
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96
36	7.40	5.25	4.38	3.89	3.57	3.35	3.18	3.05	2.95
37	7.37	5.23	4.36	3.87	3.56	3.33	3.17	3.04	2.93
38	7.35	5.21	4.34	3.86	3.54	3.32	3.15	3.02	2.92
39	7.33	5.19	4.33	3.84	3.53	3.30	3.14	3.01	2.90
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89
41	7.30	5.16	4.30	3.81	3.50	3.28	3.11	2.98	2.87
42	7.28	5.15	4.29	3.80	3.49	3.27	3.10	2.97	2.86
43	7.26	5.14	4.27	3.79	3.48	3.25	3.09	2.96	2.85
44	7.25	5.12	4.26	3.78	3.47	3.24	3.08	2.95	2.84
45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83
46	7.22	5.10	4.24	3.76	3.44	3.22	3.06	2.93	2.82
47	7.21	5.09	4.23	3.75	3.43	3.21	3.05	2.92	2.81
48	7.19	5.08	4.22	3.74	3.43	3.20	3.04	2.91	2.80
49	7.18	5.07	4.21	3.73	3.42	3.19	3.03	2.90	2.79
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78
55	7.12	5.01	4.16	3.68	3.37	3.15	2.98	2.85	2.75
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
65	7.04	4.95	4.10	3.62	3.31	3.09	2.93	2.80	2.69
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59
150	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53
350	6.71	4.67	3.84	3.37	3.07	2.85	2.69	2.56	2.46
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41

Valores críticos de la distribución F, $\alpha = .01$

IC unilateral = .99

IC bilateral = .98



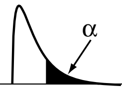
$\alpha = .01$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
2	99.40	99.42	99.43	99.44	99.45	99.48	99.49	99.50	99.50
3	27.23	27.05	26.92	26.83	26.69	26.35	26.24	26.16	26.13
4	14.55	14.37	14.25	14.15	14.02	13.69	13.58	13.50	13.46
5	10.05	9.89	9.77	9.68	9.55	9.24	9.13	9.05	9.02
6	7.87	7.72	7.60	7.52	7.40	7.09	6.99	6.91	6.88
7	6.62	6.47	6.36	6.28	6.16	5.86	5.75	5.68	5.65
8	5.81	5.67	5.56	5.48	5.36	5.07	4.96	4.89	4.86
9	5.26	5.11	5.01	4.92	4.81	4.52	4.41	4.34	4.31
10	4.85	4.71	4.60	4.52	4.41	4.12	4.01	3.94	3.91
11	4.54	4.40	4.29	4.21	4.10	3.81	3.71	3.63	3.60
12	4.30	4.16	4.05	3.97	3.86	3.57	3.47	3.39	3.36
13	4.10	3.96	3.86	3.78	3.66	3.38	3.27	3.20	3.17
14	3.94	3.80	3.70	3.62	3.51	3.22	3.11	3.04	3.00
15	3.80	3.67	3.56	3.49	3.37	3.08	2.98	2.90	2.87
16	3.69	3.55	3.45	3.37	3.26	2.97	2.86	2.78	2.75
17	3.59	3.46	3.35	3.27	3.16	2.87	2.76	2.69	2.65
18	3.51	3.37	3.27	3.19	3.08	2.78	2.68	2.60	2.57
19	3.43	3.30	3.19	3.12	3.00	2.71	2.60	2.52	2.49
20	3.37	3.23	3.13	3.05	2.94	2.64	2.54	2.45	2.42
21	3.31	3.17	3.07	2.99	2.88	2.58	2.48	2.39	2.36
22	3.26	3.12	3.02	2.94	2.83	2.53	2.42	2.34	2.31
23	3.21	3.07	2.97	2.89	2.78	2.48	2.37	2.29	2.26
24	3.17	3.03	2.93	2.85	2.74	2.44	2.33	2.25	2.21
25	3.13	2.99	2.89	2.81	2.70	2.40	2.29	2.20	2.17
26	3.09	2.96	2.86	2.78	2.66	2.36	2.25	2.17	2.13
27	3.06	2.93	2.82	2.75	2.63	2.33	2.22	2.13	2.10
28	3.03	2.90	2.79	2.72	2.60	2.30	2.19	2.10	2.06
29	3.00	2.87	2.77	2.69	2.57	2.27	2.16	2.07	2.03
30	2.98	2.84	2.74	2.66	2.55	2.25	2.13	2.04	2.01

Valores críticos de la distribución F, $\alpha = .01$

IC unilateral = .99

IC bilateral = .98



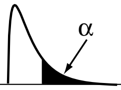
$\alpha = .01$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
31	2.96	2.82	2.72	2.64	2.52	2.22	2.11	2.02	1.98
32	2.93	2.80	2.70	2.62	2.50	2.20	2.08	1.99	1.96
33	2.91	2.78	2.68	2.60	2.48	2.18	2.06	1.97	1.93
34	2.89	2.76	2.66	2.58	2.46	2.16	2.04	1.95	1.91
35	2.88	2.74	2.64	2.56	2.44	2.14	2.02	1.93	1.89
36	2.86	2.72	2.62	2.54	2.43	2.12	2.00	1.91	1.87
37	2.84	2.71	2.61	2.53	2.41	2.10	1.98	1.89	1.85
38	2.83	2.69	2.59	2.51	2.40	2.09	1.97	1.88	1.84
39	2.81	2.68	2.58	2.50	2.38	2.07	1.95	1.86	1.82
40	2.80	2.66	2.56	2.48	2.37	2.06	1.94	1.84	1.80
41	2.79	2.65	2.55	2.47	2.36	2.04	1.92	1.83	1.79
42	2.78	2.64	2.54	2.46	2.34	2.03	1.91	1.82	1.78
43	2.76	2.63	2.53	2.45	2.33	2.02	1.90	1.80	1.76
44	2.75	2.62	2.52	2.44	2.32	2.01	1.89	1.79	1.75
45	2.74	2.61	2.51	2.43	2.31	2.00	1.88	1.78	1.74
46	2.73	2.60	2.50	2.42	2.30	1.99	1.86	1.77	1.73
47	2.72	2.59	2.49	2.41	2.29	1.98	1.85	1.76	1.71
48	2.71	2.58	2.48	2.40	2.28	1.97	1.84	1.75	1.70
49	2.71	2.57	2.47	2.39	2.27	1.96	1.83	1.74	1.69
50	2.70	2.56	2.46	2.38	2.27	1.95	1.82	1.73	1.68
55	2.66	2.53	2.42	2.34	2.23	1.91	1.78	1.68	1.64
60	2.63	2.50	2.39	2.31	2.20	1.88	1.75	1.65	1.60
65	2.61	2.47	2.37	2.29	2.17	1.85	1.72	1.62	1.57
70	2.59	2.45	2.35	2.27	2.15	1.83	1.70	1.59	1.54
80	2.55	2.42	2.31	2.23	2.12	1.79	1.65	1.54	1.49
100	2.50	2.37	2.27	2.19	2.07	1.74	1.60	1.48	1.43
150	2.44	2.31	2.20	2.12	2.00	1.66	1.52	1.39	1.33
350	2.37	2.24	2.13	2.05	1.93	1.58	1.43	1.28	1.20
∞	2.32	2.18	2.08	2.00	1.88	1.52	1.36	1.18	1.00

Valores críticos de la distribución F, $\alpha = .005$

IC unilateral = .995

IC bilateral = .99

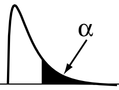


$\alpha = .005$

Den. df	Numerador df 1	2	3	4	5	6	7	8	9
2	198.50	199.00	199.17	199.25	199.30	199.33	199.36	199.37	199.39
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77
6	18.64	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
21	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81
23	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60
27	9.34	6.49	5.36	4.74	4.34	4.06	3.85	3.69	3.56
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52
29	9.23	6.40	5.28	4.66	4.26	3.98	3.77	3.61	3.48
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45

Valores críticos de la distribución F, $\alpha = .005$

IC unilateral = .995
IC bilateral = .99



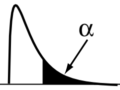
$\alpha = .005$

Den.	Numerador df								
df	1	2	3	4	5	6	7	8	9
31	9.13	6.32	5.20	4.59	4.20	3.92	3.71	3.55	3.42
32	9.09	6.28	5.17	4.56	4.17	3.89	3.68	3.52	3.39
33	9.05	6.25	5.14	4.53	4.14	3.86	3.66	3.49	3.37
34	9.01	6.22	5.11	4.50	4.11	3.84	3.63	3.47	3.34
35	8.98	6.19	5.09	4.48	4.09	3.81	3.61	3.45	3.32
36	8.94	6.16	5.06	4.46	4.06	3.79	3.58	3.42	3.30
37	8.91	6.13	5.04	4.43	4.04	3.77	3.56	3.40	3.28
38	8.88	6.11	5.02	4.41	4.02	3.75	3.54	3.39	3.26
39	8.85	6.09	5.00	4.39	4.00	3.73	3.53	3.37	3.24
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22
41	8.80	6.05	4.96	4.36	3.97	3.70	3.49	3.33	3.21
42	8.78	6.03	4.94	4.34	3.95	3.68	3.48	3.32	3.19
43	8.76	6.01	4.92	4.32	3.94	3.67	3.46	3.30	3.18
44	8.74	5.99	4.91	4.31	3.92	3.65	3.45	3.29	3.16
45	8.71	5.97	4.89	4.29	3.91	3.64	3.43	3.28	3.15
46	8.70	5.96	4.88	4.28	3.90	3.62	3.42	3.26	3.14
47	8.68	5.94	4.86	4.27	3.88	3.61	3.41	3.25	3.12
48	8.66	5.93	4.85	4.25	3.87	3.60	3.40	3.24	3.11
49	8.64	5.91	4.84	4.24	3.86	3.59	3.39	3.23	3.10
50	8.63	5.90	4.83	4.23	3.85	3.58	3.38	3.22	3.09
55	8.55	5.84	4.77	4.18	3.80	3.53	3.33	3.17	3.05
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01
65	8.44	5.75	4.69	4.11	3.73	3.46	3.26	3.10	2.98
70	8.40	5.72	4.66	4.08	3.70	3.43	3.23	3.08	2.95
80	8.33	5.67	4.61	4.03	3.65	3.39	3.19	3.03	2.91
100	8.24	5.59	4.54	3.96	3.59	3.33	3.13	2.97	2.85
150	8.12	5.49	4.45	3.88	3.51	3.25	3.05	2.89	2.77
350	7.98	5.38	4.35	3.78	3.42	3.16	2.96	2.81	2.68
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62

Valores críticos de la distribución F, $\alpha = .005$

IC unilateral = .995

IC bilateral = .99



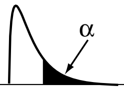
$\alpha = .005$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
2	199.40	199.42	199.43	199.44	199.45	199.48	199.49	199.50	199.50
3	43.69	43.39	43.17	43.01	42.78	42.21	42.02	41.88	41.83
4	20.97	20.70	20.51	20.37	20.17	19.67	19.50	19.37	19.32
5	13.62	13.38	13.21	13.09	12.90	12.45	12.30	12.19	12.14
6	10.25	10.03	9.88	9.76	9.59	9.17	9.03	8.92	8.88
7	8.38	8.18	8.03	7.91	7.75	7.35	7.22	7.12	7.08
8	7.21	7.01	6.87	6.76	6.61	6.22	6.09	5.99	5.95
9	6.42	6.23	6.09	5.98	5.83	5.45	5.32	5.23	5.19
10	5.85	5.66	5.53	5.42	5.27	4.90	4.77	4.68	4.64
11	5.42	5.24	5.10	5.00	4.86	4.49	4.36	4.26	4.23
12	5.09	4.91	4.77	4.67	4.53	4.17	4.04	3.94	3.90
13	4.82	4.64	4.51	4.41	4.27	3.91	3.78	3.69	3.65
14	4.60	4.43	4.30	4.20	4.06	3.70	3.57	3.47	3.44
15	4.42	4.25	4.12	4.02	3.88	3.52	3.39	3.30	3.26
16	4.27	4.10	3.97	3.87	3.73	3.37	3.25	3.15	3.11
17	4.14	3.97	3.84	3.75	3.61	3.25	3.12	3.02	2.98
18	4.03	3.86	3.73	3.64	3.50	3.14	3.01	2.91	2.87
19	3.93	3.76	3.64	3.54	3.40	3.04	2.91	2.82	2.78
20	3.85	3.68	3.55	3.46	3.32	2.96	2.83	2.73	2.69
21	3.77	3.60	3.48	3.38	3.24	2.88	2.75	2.65	2.61
22	3.70	3.54	3.41	3.31	3.18	2.82	2.69	2.59	2.55
23	3.64	3.47	3.35	3.25	3.12	2.76	2.62	2.52	2.48
24	3.59	3.42	3.30	3.20	3.06	2.70	2.57	2.47	2.43
25	3.54	3.37	3.25	3.15	3.01	2.65	2.52	2.42	2.38
26	3.49	3.33	3.20	3.11	2.97	2.61	2.47	2.37	2.33
27	3.45	3.28	3.16	3.07	2.93	2.57	2.43	2.33	2.29
28	3.41	3.25	3.12	3.03	2.89	2.53	2.39	2.29	2.25
29	3.38	3.21	3.09	2.99	2.86	2.49	2.36	2.25	2.21
30	3.34	3.18	3.06	2.96	2.82	2.46	2.32	2.22	2.18

Valores críticos de la distribución F, $\alpha = .005$

IC unilateral = .995

IC bilateral = .99



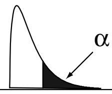
$\alpha = .005$

Den. df	Numerador df 10	12	14	16	20	50	100	350	∞
31	3.31	3.15	3.03	2.93	2.79	2.43	2.29	2.19	2.14
32	3.29	3.12	3.00	2.90	2.77	2.40	2.26	2.16	2.11
33	3.26	3.09	2.97	2.88	2.74	2.37	2.24	2.13	2.09
34	3.24	3.07	2.95	2.85	2.72	2.35	2.21	2.11	2.06
35	3.21	3.05	2.93	2.83	2.69	2.33	2.19	2.08	2.04
36	3.19	3.03	2.90	2.81	2.67	2.30	2.17	2.06	2.01
37	3.17	3.01	2.88	2.79	2.65	2.28	2.14	2.04	1.99
38	3.15	2.99	2.87	2.77	2.63	2.27	2.12	2.02	1.97
39	3.13	2.97	2.85	2.75	2.62	2.25	2.11	2.00	1.95
40	3.12	2.95	2.83	2.74	2.60	2.23	2.09	1.98	1.93
41	3.10	2.94	2.82	2.72	2.58	2.21	2.07	1.96	1.91
42	3.09	2.92	2.80	2.71	2.57	2.20	2.06	1.94	1.90
43	3.07	2.91	2.79	2.69	2.55	2.18	2.04	1.93	1.88
44	3.06	2.89	2.77	2.68	2.54	2.17	2.03	1.91	1.87
45	3.04	2.88	2.76	2.66	2.53	2.16	2.01	1.90	1.85
46	3.03	2.87	2.75	2.65	2.51	2.14	2.00	1.89	1.84
47	3.02	2.86	2.74	2.64	2.50	2.13	1.99	1.87	1.82
48	3.01	2.85	2.72	2.63	2.49	2.12	1.97	1.86	1.81
49	3.00	2.83	2.71	2.62	2.48	2.11	1.96	1.85	1.80
50	2.99	2.82	2.70	2.61	2.47	2.10	1.95	1.84	1.79
55	2.94	2.78	2.66	2.56	2.42	2.05	1.90	1.78	1.73
60	2.90	2.74	2.62	2.53	2.39	2.01	1.86	1.74	1.69
65	2.87	2.71	2.59	2.49	2.36	1.98	1.83	1.70	1.65
70	2.85	2.68	2.56	2.47	2.33	1.95	1.80	1.67	1.62
80	2.80	2.64	2.52	2.43	2.29	1.90	1.75	1.62	1.56
100	2.74	2.58	2.46	2.37	2.23	1.84	1.68	1.55	1.49
150	2.67	2.51	2.38	2.29	2.15	1.76	1.59	1.44	1.37
350	2.58	2.42	2.30	2.20	2.06	1.66	1.48	1.32	1.23
∞	2.52	2.36	2.24	2.14	2.00	1.59	1.40	1.21	1.00

APÉNDICE D

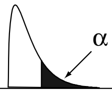
Valores críticos de la distribución chi-cuadrada

Valores críticos de la distribución chi-cuadrada



df	.200	.100	.050	.025	.010	.005	.001
1	1.642	2.706	3.841	5.024	6.635	7.879	10.828
2	3.219	4.605	5.991	7.378	9.210	10.597	13.816
3	4.642	6.251	7.815	9.348	11.345	12.838	16.266
4	5.989	7.779	9.488	11.143	13.277	14.860	18.467
5	7.289	9.236	11.070	12.833	15.086	16.750	20.515
6	8.558	10.645	12.592	14.449	16.812	18.548	22.458
7	9.803	12.017	14.067	16.013	18.475	20.278	24.322
8	11.030	13.362	15.507	17.535	20.090	21.955	26.124
9	12.242	14.684	16.919	19.023	21.666	23.589	27.877
10	13.442	15.987	18.307	20.483	23.209	25.188	29.588
11	14.631	17.275	19.675	21.920	24.725	26.757	31.264
12	15.812	18.549	21.026	23.337	26.217	28.300	32.909
13	16.985	19.812	22.362	24.736	27.688	29.819	34.528
14	18.151	21.064	23.685	26.119	29.141	31.319	36.123
15	19.311	22.307	24.996	27.488	30.578	32.801	37.697
16	20.465	23.542	26.296	28.845	32.000	34.267	39.252
17	21.615	24.769	27.587	30.191	33.409	35.718	40.790
18	22.760	25.989	28.869	31.526	34.805	37.156	42.312
19	23.900	27.204	30.144	32.852	36.191	38.582	43.820
20	25.038	28.412	31.410	34.170	37.566	39.997	45.315
21	26.171	29.615	32.671	35.479	38.932	41.401	46.797
22	27.301	30.813	33.924	36.781	40.289	42.796	48.268
23	28.429	32.007	35.172	38.076	41.638	44.181	49.728
24	29.553	33.196	36.415	39.364	42.980	45.559	51.179
25	30.675	34.382	37.652	40.646	44.314	46.928	52.620
26	31.795	35.563	38.885	41.923	45.642	48.290	54.052
27	32.912	36.741	40.113	43.195	46.963	49.645	55.476
28	34.027	37.916	41.337	44.461	48.278	50.993	56.892
29	35.139	39.087	42.557	45.722	49.588	52.336	58.301
30	36.250	40.256	43.773	46.979	50.892	53.672	59.703

Valores críticos de la distribución chi-cuadrada



df	.200	.100	.050	.025	.010	.005	.001
31	37.359	41.422	44.985	48.232	52.191	55.003	61.098
32	38.466	42.585	46.194	49.480	53.486	56.328	62.487
33	39.572	43.745	47.400	50.725	54.776	57.648	63.870
34	40.676	44.903	48.602	51.966	56.061	58.964	65.247
35	41.778	46.059	49.802	53.203	57.342	60.275	66.619
36	42.879	47.212	50.998	54.437	58.619	61.581	67.985
37	43.978	48.363	52.192	55.668	59.893	62.883	69.346
38	45.076	49.513	53.384	56.896	61.162	64.181	70.703
39	46.173	50.660	54.572	58.120	62.428	65.476	72.055
40	47.269	51.805	55.758	59.342	63.691	66.766	73.402
41	48.363	52.949	56.942	60.561	64.950	68.053	74.745
42	49.456	54.090	58.124	61.777	66.206	69.336	76.084
43	50.548	55.230	59.304	62.990	67.459	70.616	77.419
44	51.639	56.369	60.481	64.201	68.710	71.893	78.750
45	52.729	57.505	61.656	65.410	69.957	73.166	80.077
46	53.818	58.641	62.830	66.617	71.201	74.437	81.400
47	54.906	59.774	64.001	67.821	72.443	75.704	82.720
48	55.993	60.907	65.171	69.023	73.683	76.969	84.037
49	57.079	62.038	66.339	70.222	74.919	78.231	85.351
50	58.164	63.167	67.505	71.420	76.154	79.490	86.661
55	63.577	68.796	73.311	77.380	82.292	85.749	93.168
60	68.972	74.397	79.082	83.298	88.379	91.952	99.607
65	74.351	79.973	84.821	89.177	94.422	98.105	105.988
70	79.715	85.527	90.531	95.023	100.425	104.215	112.317
75	85.066	91.061	96.217	100.839	106.393	110.286	118.599
80	90.405	96.578	101.879	106.629	112.329	116.321	124.839
85	95.734	102.079	107.522	112.393	118.236	122.325	131.041
90	101.054	107.565	113.145	118.136	124.116	128.299	137.208
95	106.364	113.038	118.752	123.858	129.973	134.247	143.344
100	111.667	118.498	124.342	129.561	135.807	140.169	149.449

APÉNDICE E

Valores críticos de q para la prueba DHS de Tukey

Valores críticos de q para la prueba DHS de Tukey, $\alpha = .025$ (1 cola), .05 (2 colas)

	Número de medias								
df	2	3	4	5	6	7	8	9	10
2	6.085	8.331	9.798	10.881	11.734	12.435	13.027	13.539	13.988
3	4.501	5.910	6.825	7.502	8.037	8.478	8.852	9.177	9.462
4	3.926	5.040	5.757	6.287	6.706	7.053	7.347	7.602	7.826
5	3.635	4.602	5.218	5.673	6.033	6.330	6.582	6.801	6.995
6	3.460	4.339	4.896	5.305	5.628	5.895	6.122	6.319	6.493
7	3.344	4.165	4.681	5.060	5.359	5.606	5.815	5.997	6.158
8	3.261	4.041	4.529	4.886	5.167	5.399	5.596	5.767	5.918
9	3.199	3.948	4.415	4.755	5.024	5.244	5.432	5.595	5.738
10	3.151	3.877	4.327	4.654	4.912	5.124	5.304	5.460	5.598
11	3.113	3.820	4.256	4.574	4.823	5.028	5.202	5.353	5.486
12	3.081	3.773	4.199	4.508	4.750	4.950	5.119	5.265	5.395
13	3.055	3.734	4.151	4.453	4.690	4.884	5.049	5.192	5.318
14	3.033	3.701	4.111	4.407	4.639	4.829	4.990	5.130	5.253
15	3.014	3.673	4.076	4.367	4.595	4.782	4.940	5.077	5.198
16	2.998	3.649	4.046	4.333	4.557	4.741	4.896	5.031	5.150
17	2.984	3.628	4.020	4.303	4.524	4.705	4.858	4.991	5.108
18	2.971	3.609	3.997	4.276	4.494	4.673	4.824	4.955	5.071
19	2.960	3.593	3.977	4.253	4.468	4.645	4.794	4.924	5.037
20	2.950	3.578	3.958	4.232	4.445	4.620	4.768	4.895	5.008
21	2.941	3.565	3.942	4.213	4.424	4.597	4.743	4.870	4.981
22	2.933	3.553	3.927	4.196	4.405	4.577	4.722	4.847	4.957
23	2.926	3.542	3.914	4.180	4.388	4.558	4.702	4.826	4.935
24	2.919	3.532	3.901	4.166	4.373	4.541	4.684	4.807	4.915
25	2.913	3.523	3.890	4.153	4.358	4.526	4.667	4.789	4.897
26	2.907	3.514	3.880	4.141	4.345	4.511	4.652	4.773	4.880
27	2.902	3.506	3.870	4.130	4.333	4.498	4.638	4.758	4.864
28	2.897	3.499	3.861	4.120	4.322	4.486	4.625	4.745	4.850
29	2.892	3.493	3.853	4.111	4.311	4.475	4.613	4.732	4.837
30	2.888	3.486	3.845	4.102	4.301	4.464	4.601	4.720	4.824

Valores críticos de q para la prueba DHS de Tukey, $\alpha = .025$ (1 cola), .05 (2 colas)

	Número de medias								
df	2	3	4	5	6	7	8	9	10
31	2.884	3.481	3.838	4.094	4.292	4.454	4.591	4.709	4.812
32	2.881	3.475	3.832	4.086	4.284	4.445	4.581	4.698	4.802
33	2.877	3.470	3.825	4.079	4.276	4.436	4.572	4.689	4.791
34	2.874	3.465	3.820	4.072	4.268	4.428	4.563	4.680	4.782
35	2.871	3.461	3.814	4.066	4.261	4.421	4.555	4.671	4.773
36	2.868	3.457	3.809	4.060	4.255	4.414	4.547	4.663	4.764
37	2.865	3.453	3.804	4.054	4.249	4.407	4.540	4.655	4.756
38	2.863	3.449	3.799	4.049	4.243	4.400	4.533	4.648	4.749
39	2.861	3.445	3.795	4.044	4.237	4.394	4.527	4.641	4.741
40	2.858	3.442	3.791	4.040	4.232	4.389	4.521	4.635	4.735
41	2.857	3.439	3.788	4.035	4.227	4.383	4.515	4.629	4.728
42	2.855	3.436	3.784	4.031	4.222	4.378	4.509	4.623	4.722
43	2.853	3.434	3.780	4.027	4.218	4.373	4.504	4.617	4.716
44	2.851	3.431	3.777	4.023	4.213	4.368	4.499	4.612	4.711
45	2.849	3.428	3.773	4.019	4.209	4.364	4.494	4.607	4.705
46	2.847	3.426	3.770	4.015	4.205	4.360	4.490	4.602	4.700
47	2.846	3.423	3.767	4.012	4.201	4.355	4.485	4.597	4.695
48	2.844	3.421	3.764	4.009	4.198	4.351	4.481	4.593	4.691
49	2.843	3.419	3.762	4.006	4.194	4.348	4.477	4.588	4.686
50	2.841	3.416	3.759	4.003	4.191	4.344	4.473	4.584	4.682
55	2.835	3.407	3.747	3.989	4.176	4.328	4.456	4.566	4.663
60	2.829	3.399	3.738	3.978	4.163	4.314	4.441	4.551	4.647
65	2.825	3.392	3.729	3.969	4.153	4.303	4.429	4.538	4.633
70	2.821	3.387	3.722	3.961	4.144	4.293	4.419	4.527	4.622
75	2.818	3.382	3.716	3.954	4.136	4.285	4.410	4.517	4.612
80	2.814	3.377	3.711	3.948	4.130	4.278	4.402	4.509	4.603
100	2.806	3.365	3.695	3.929	4.110	4.256	4.379	4.484	4.577
120	2.800	3.356	3.685	3.917	4.096	4.241	4.363	4.468	4.560
240	2.786	3.335	3.659	3.887	4.063	4.205	4.324	4.427	4.517
∞	2.772	3.314	3.633	3.858	4.030	4.170	4.286	4.387	4.474

Valores críticos de q para la prueba DHS de Tukey, $\alpha = .005$ (1 cola), .01 (2 colas)

	Número de medias								
df	2	3	4	5	6	7	8	9	10
2	14.036	19.019	22.294	24.717	26.629	28.201	29.530	30.679	31.689
3	8.260	10.619	12.170	13.324	14.241	14.998	15.641	16.199	16.691
4	6.511	8.120	9.173	9.958	10.583	11.101	11.542	11.925	12.264
5	5.702	6.976	7.804	8.421	8.913	9.321	9.669	9.971	10.239
6	5.243	6.331	7.033	7.556	7.972	8.318	8.612	8.869	9.097
7	4.949	5.919	6.542	7.005	7.373	7.678	7.939	8.166	8.367
8	4.745	5.635	6.204	6.625	6.959	7.237	7.474	7.680	7.863
9	4.596	5.428	5.957	6.347	6.657	6.915	7.134	7.325	7.494
10	4.482	5.270	5.769	6.136	6.428	6.669	6.875	7.054	7.213
11	4.392	5.146	5.621	5.970	6.247	6.476	6.671	6.841	6.992
12	4.320	5.046	5.502	5.836	6.101	6.320	6.507	6.670	6.814
13	4.260	4.964	5.404	5.726	5.981	6.192	6.372	6.528	6.666
14	4.210	4.895	5.322	5.634	5.881	6.085	6.258	6.409	6.543
15	4.167	4.836	5.252	5.556	5.796	5.994	6.162	6.309	6.438
16	4.131	4.786	5.192	5.489	5.722	5.915	6.079	6.222	6.348
17	4.099	4.742	5.140	5.430	5.659	5.847	6.007	6.147	6.270
18	4.071	4.703	5.094	5.379	5.603	5.787	5.944	6.081	6.201
19	4.046	4.669	5.054	5.334	5.553	5.735	5.889	6.022	6.141
20	4.024	4.639	5.018	5.293	5.510	5.688	5.839	5.970	6.086
21	4.004	4.612	4.986	5.257	5.470	5.646	5.794	5.924	6.038
22	3.986	4.588	4.957	5.225	5.435	5.608	5.754	5.882	5.994
23	3.970	4.566	4.931	5.195	5.403	5.573	5.718	5.844	5.955
24	3.955	4.546	4.907	5.168	5.373	5.542	5.685	5.809	5.919
25	3.942	4.527	4.885	5.144	5.347	5.513	5.655	5.778	5.886
26	3.930	4.510	4.865	5.121	5.322	5.487	5.627	5.749	5.856
27	3.918	4.495	4.847	5.101	5.300	5.463	5.602	5.722	5.828
28	3.908	4.481	4.830	5.082	5.279	5.441	5.578	5.697	5.802
29	3.898	4.467	4.814	5.064	5.260	5.420	5.556	5.674	5.778
30	3.889	4.455	4.799	5.048	5.242	5.401	5.536	5.653	5.756

Valores críticos de q para la prueba DHS de Tukey, $\alpha = .005$ (1 cola), .01 (2 colas)

	Número de medias								
df	2	3	4	5	6	7	8	9	10
31	3.881	4.443	4.786	5.032	5.225	5.383	5.517	5.633	5.736
32	3.873	4.433	4.773	5.018	5.210	5.367	5.500	5.615	5.716
33	3.865	4.423	4.761	5.005	5.195	5.351	5.483	5.598	5.698
34	3.859	4.413	4.750	4.992	5.181	5.336	5.468	5.581	5.682
35	3.852	4.404	4.739	4.980	5.169	5.323	5.453	5.566	5.666
36	3.846	4.396	4.729	4.969	5.156	5.310	5.439	5.552	5.651
37	3.840	4.388	4.720	4.959	5.145	5.298	5.427	5.538	5.637
38	3.835	4.381	4.711	4.949	5.134	5.286	5.414	5.526	5.623
39	3.830	4.374	4.703	4.940	5.124	5.275	5.403	5.513	5.611
40	3.828	4.370	4.700	4.936	5.119	5.269	5.396	5.505	6.602
41	3.826	4.366	4.692	4.928	5.111	5.259	5.385	5.494	5.591
42	3.822	4.360	4.685	4.920	5.102	5.250	5.375	5.484	5.580
43	3.817	4.354	4.679	4.912	5.094	5.241	5.366	5.474	5.570
44	3.813	4.349	4.672	4.905	5.085	5.232	5.357	5.465	5.560
45	3.809	4.343	4.666	4.898	5.078	5.224	5.349	5.456	5.551
46	3.806	4.338	4.660	4.891	5.071	5.217	5.340	5.447	5.542
47	3.802	4.334	4.654	4.885	5.064	5.209	5.333	5.439	5.533
48	3.799	4.329	4.649	4.879	5.057	5.202	5.325	5.431	5.525
49	3.796	4.325	4.644	4.873	5.051	5.195	5.318	5.424	5.517
50	3.793	4.321	4.639	4.867	5.045	5.189	5.311	5.416	5.510
55	3.779	4.302	4.617	4.843	5.018	5.160	5.281	5.385	5.477
60	3.761	4.287	4.598	4.822	4.996	5.136	5.255	5.358	5.449
65	3.758	4.274	4.583	4.805	4.977	5.116	5.234	5.336	5.427
70	3.750	4.263	4.570	4.790	4.961	5.099	5.216	5.318	5.407
75	3.743	4.253	4.559	4.777	4.947	5.085	5.201	5.301	5.390
80	3.737	4.245	4.549	4.766	4.935	5.073	5.187	5.287	5.375
100	3.718	4.220	4.520	4.733	4.900	5.035	5.147	5.245	5.331
120	3.702	4.200	4.497	4.709	4.872	5.005	5.118	5.214	5.299
240	3.672	4.160	4.450	4.655	4.814	4.943	5.052	5.145	5.227
	3.643	4.120	4.403	4.603	4.757	4.882	4.987	5.078	5.157

APÉNDICE F

Valores críticos del coeficiente de correlación de rangos

Valores críticos para el coeficiente de correlación de rangos

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
n					
4	1.000	1.000	—	—	—
5	.800	.900	1.000	1.000	—
6	.657	.829	.886	.943	1.000
7	.607	.714	.786	.893	.929
8	.524	.643	.738	.833	.881
9	.483	.600	.700	.783	.833
10	.455	.564	.648	.745	.794
11	.427	.536	.618	.709	.755
12	.406	.503	.587	.678	.727
13	.385	.484	.560	.648	.703
14	.367	.464	.538	.622	.675
15	.354	.446	.521	.604	.654
16	.341	.429	.503	.585	.635
17	.328	.414	.488	.566	.615
18	.319	.401	.472	.550	.600
19	.309	.391	.460	.535	.584
20	.299	.380	.448	.522	.570
21	.291	.370	.435	.509	.556
22	.284	.361	.425	.497	.544
23	.278	.353	.416	.486	.532
24	.271	.344	.406	.477	.522
25	.265	.338	.398	.466	.511
26	.259	.330	.389	.456	.501
27	.255	.324	.383	.448	.491
28	.250	.318	.375	.440	.483
29	.245	.312	.369	.433	.475
30	.240	.307	.362	.425	.467

Valores críticos para el coeficiente de correlación de rangos

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
n					
31	.236	.302	.356	.419	.460
32	.232	.296	.351	.412	.453
33	.229	.291	.345	.405	.445
34	.225	.287	.340	.400	.439
35	.222	.283	.335	.394	.433
36	.219	.279	.330	.388	.427
37	.216	.275	.326	.384	.422
38	.213	.271	.321	.378	.416
39	.210	.267	.317	.373	.410
40	.207	.264	.313	.368	.405
41	.204	.261	.309	.363	.400
42	.202	.257	.305	.360	.396
43	.199	.254	.301	.355	.391
44	.197	.251	.298	.351	.386
45	.194	.248	.294	.347	.382
46	.192	.246	.291	.343	.378
47	.190	.243	.288	.340	.374
48	.188	.240	.285	.336	.371
49	.186	.238	.282	.333	.366
50	.184	.235	.279	.330	.363
51	.182	.233	.276	.326	.359
52	.181	.231	.274	.323	.356
53	.179	.229	.271	.320	.353
54	.177	.226	.268	.317	.349
55	.175	.224	.266	.314	.346
56	.174	.222	.264	.311	.343
57	.172	.220	.261	.308	.340
58	.171	.218	.259	.306	.337
59	.169	.216	.257	.303	.335
60	.168	.214	.254	.301	.332

Valores críticos para el coeficiente de correlación de rangos

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
n					
61	.166	.212	.252	.298	.328
62	.165	.211	.250	.296	.326
63	.164	.209	.248	.293	.323
64	.162	.207	.247	.291	.321
65	.161	.206	.244	.289	.318
66	.160	.204	.243	.287	.316
67	.158	.203	.241	.284	.314
68	.157	.201	.239	.282	.311
69	.156	.200	.237	.280	.309
70	.155	.198	.235	.278	.307
71	.154	.197	.234	.276	.305
72	.153	.195	.232	.274	.303
73	.151	.194	.231	.272	.301
74	.151	.193	.229	.271	.298
75	.150	.191	.227	.269	.297
76	.149	.190	.226	.267	.295
77	.147	.189	.224	.265	.292
78	.147	.188	.223	.264	.291
79	.146	.186	.222	.262	.289
80	.145	.185	.220	.260	.288
81	.144	.184	.218	.258	.285
82	.143	.183	.217	.257	.284
83	.142	.182	.216	.255	.282
84	.141	.181	.215	.254	.281
85	.140	.180	.213	.252	.279
86	.139	.179	.212	.251	.277
87	.139	.178	.211	.249	.275
88	.138	.177	.210	.248	.274
89	.137	.175	.209	.247	.273
90	.136	.174	.207	.245	.271

APÉNDICE G

Valores críticos para la prueba de los signos de Wilcoxon o prueba de los rangos signados (como un estadístico t)

Valores críticos para la prueba de los signos de Wilcoxon

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
n					
4	3.873	—	—	—	—
5	1.773	4.243	—	—	—
6	1.872	2.371	4.583	—	—
7	1.721	2.420	2.925	4.899	—
8	1.508	2.225	2.934	4.123	5.196
9	1.605	1.975	2.704	3.414	4.545
10	1.450	2.049	2.424	3.160	3.862
11	1.502	2.009	2.477	3.074	3.592
12	1.481	1.906	2.416	3.071	3.488
13	1.414	1.869	2.296	2.958	3.468
14	1.394	1.875	2.245	2.919	3.339
15	1.407	1.833	2.238	2.815	3.286
16	1.382	1.827	2.261	2.770	3.170
17	1.385	1.785	2.236	2.763	3.114
18	1.359	1.774	2.178	2.709	3.096
19	1.357	1.785	2.152	2.690	3.034
20	1.373	1.767	2.150	2.637	3.006
21	1.365	1.769	2.167	2.612	3.003
22	1.338	1.748	2.156	2.610	2.966
23	1.363	1.746	2.122	2.580	2.953
24	1.336	1.758	2.108	2.570	2.912
25	1.354	1.750	2.110	2.577	2.893
26	1.327	1.726	2.092	2.560	2.891
27	1.341	1.744	2.089	2.559	2.867
28	1.339	1.719	2.098	2.538	2.858
29	1.325	1.733	2.090	2.532	2.831
30	1.322	1.731	2.069	2.510	2.848

Valores críticos para la prueba de los signos de Wilcoxon

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
n					
31	.236	.302	.356	.419	.460
32	.232	.296	.351	.412	.453
33	.229	.291	.345	.405	.445
34	.225	.287	.340	.400	.439
35	.222	.283	.335	.394	.433
36	.219	.279	.330	.388	.427
37	.216	.275	.326	.384	.422
38	.213	.271	.321	.378	.416
39	.210	.267	.317	.373	.410
40	.207	.264	.313	.368	.405
41	.204	.261	.309	.363	.400
42	.202	.257	.305	.360	.396
43	.199	.254	.301	.355	.391
44	.197	.251	.298	.351	.386
45	.194	.248	.294	.347	.382
46	.192	.246	.291	.343	.378
47	.190	.243	.288	.340	.374
48	.188	.240	.285	.336	.371
49	.186	.238	.282	.333	.366
50	.184	.235	.279	.330	.363
51	.182	.233	.276	.326	.359
52	.181	.231	.274	.323	.356
53	.179	.229	.271	.320	.353
54	.177	.226	.268	.317	.349
55	.175	.224	.266	.314	.346
56	.174	.222	.264	.311	.343
57	.172	.220	.261	.308	.340
58	.171	.218	.259	.306	.337
59	.169	.216	.257	.303	.335
60	.168	.214	.254	.301	.332

Valores críticos para la prueba de los signos de Wilcoxon

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
n					
61	1.304	1.680	2.013	2.406	2.688
62	1.301	1.675	2.007	2.405	2.680
63	1.301	1.673	2.012	2.408	2.683
64	1.298	1.675	2.005	2.406	2.682
65	1.297	1.673	2.009	2.407	2.677
66	1.300	1.674	2.009	2.398	2.675
67	1.300	1.678	2.005	2.398	2.676
68	1.296	1.671	2.004	2.402	2.673
69	1.301	1.675	2.007	2.402	2.674
70	1.297	1.674	2.005	2.398	2.670
71	1.295	1.671	2.001	2.397	2.669
72	1.297	1.670	1.999	2.392	2.672
73	1.295	1.672	2.000	2.397	2.670
74	1.296	1.671	2.003	2.392	2.665
75	1.294	1.673	1.998	2.390	2.663
76	1.295	1.671	2.001	2.390	2.664
77	1.293	1.667	1.995	2.388	2.661
78	1.293	1.670	1.997	2.387	2.661
79	1.296	1.670	1.996	2.390	2.657
80	1.296	1.668	1.998	2.389	2.656
81	1.293	1.668	1.997	2.390	2.658
82	1.293	1.666	1.998	2.384	2.656
83	1.295	1.666	1.996	2.385	2.657
84	1.294	1.668	1.997	2.388	2.660
85	1.291	1.667	1.995	2.383	2.650
86	1.295	1.668	1.995	2.381	2.653
87	1.292	1.667	1.993	2.385	2.652
88	1.295	1.664	1.993	2.383	2.649
89	1.296	1.667	1.994	2.382	2.648
90	1.294	1.664	1.994	2.379	2.649

APÉNDICE H

Valores críticos para la prueba de la suma de rangos de Wilcoxon para muestras independientes (como un estadístico t)

Valores críticos para la prueba de la suma de rangos de Wilcoxon

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
$n_1 = n_2$					
3	3.674	1.871	—	—	—
4	1.594	2.898	4.382	—	—
5	1.732	2.077	3.031	3.781	5.000
6	1.526	1.982	2.550	3.321	3.847
7	1.546	1.887	2.497	3.008	3.662
8	1.409	1.944	2.249	2.985	3.450
9	1.408	1.726	2.330	2.757	3.260
10	1.394	1.845	2.248	2.705	3.097
11	1.374	1.836	2.173	2.642	3.070
12	1.352	1.747	2.179	2.662	2.929
13	1.384	1.792	2.171	2.585	2.966
14	1.353	1.712	2.096	2.576	2.906
15	1.368	1.737	2.131	2.558	2.846
16	1.336	1.752	2.105	2.533	2.840
17	1.342	1.681	2.078	2.504	2.824
18	1.344	1.723	2.049	2.513	2.802
19	1.343	1.723	2.055	2.477	2.776
20	1.339	1.719	2.055	2.474	2.783
21	1.333	1.712	2.050	2.467	2.750
22	1.326	1.703	2.043	2.456	2.746
23	1.318	1.693	2.034	2.469	2.738
24	1.309	1.704	2.045	2.451	2.726
25	1.320	1.690	2.031	2.432	2.736
26	1.308	1.696	2.037	2.434	2.718
27	1.315	1.699	2.039	2.432	2.720
28	1.320	1.682	2.021	2.428	2.718
29	1.307	1.683	2.020	2.422	2.714
30	1.309	1.681	2.017	2.414	2.689

Valores críticos para la prueba de la suma de rangos de Wilcoxon

Una cola	.100	.050	.025	.010	.005
Dos colas	.200	.100	.050	.020	.010
$n_1 = n_2$					
31	1.310	1.679	2.012	2.421	2.698
32	1.310	1.689	2.022	2.410	2.688
33	1.309	1.684	2.014	2.412	2.692
34	1.294	1.678	2.020	2.413	2.679
35	1.305	1.672	2.010	2.412	2.679
36	1.301	1.677	2.013	2.409	2.677
37	1.298	1.680	2.002	2.405	2.673
38	1.305	1.682	2.002	2.400	2.669
39	1.300	1.673	1.991	2.394	2.663
40	1.305	1.674	2.000	2.398	2.667
41	1.299	1.674	1.997	2.390	2.659
42	1.303	1.673	2.004	2.391	2.660
43	1.296	1.671	1.999	2.391	2.660
44	1.299	1.669	1.994	2.391	2.658
45	1.300	1.667	1.998	2.389	2.647
46	1.293	1.672	1.992	2.378	2.644
47	1.294	1.668	2.001	2.383	2.649
48	1.287	1.672	1.994	2.379	2.652
49	1.294	1.667	1.995	2.382	2.647
50	1.294	1.669	1.994	2.377	2.648
51	1.286	1.664	1.993	2.379	2.642
52	1.291	1.665	1.991	2.380	2.641
53	1.296	1.666	1.989	2.373	2.641
54	1.294	1.667	1.993	2.372	2.639
55	1.292	1.666	1.990	2.371	2.637
56	1.296	1.660	1.987	2.376	2.640
57	1.293	1.665	1.983	2.374	2.637
58	1.296	1.664	1.991	2.371	2.633
59	1.293	1.662	1.986	2.368	2.629
60	1.289	1.666	1.987	2.370	2.635

APÉNDICE I

Valores críticos para la prueba Kruskal-Wallis (expresados como un estadístico *F*)

Valores críticos para la prueba Kruskal-Wallis, α=.10

n por grupo	Número de grupos								
	2	3	4	5	6	7	8	9	10
3	3.501	4.105	3.170	2.724	2.472	2.289	2.164	2.069	1.989
4	8.400	3.300	2.733	2.438	2.247	2.109	2.008	1.930	1.865
5	4.313	2.898	2.520	2.290	2.131	2.018	1.930	1.860	1.804
6	3.929	2.731	2.413	2.209	2.069	1.966	1.885	1.822	1.769
7	3.562	2.650	2.355	2.162	2.029	1.932	1.856	1.795	1.745
8	3.778	2.621	2.315	2.130	2.003	1.908	1.835	1.778	1.729
9	3.380	2.562	2.280	2.104	1.981	1.891	1.821	1.763	1.716
10	3.406	2.534	2.257	2.085	1.966	1.877	1.809	1.752	1.707
11	3.082	2.506	2.238	2.071	1.954	1.866	1.798	1.745	1.699
12	3.054	2.486	2.223	2.059	1.945	1.858	1.790	1.738	1.692
13	3.213	2.469	2.212	2.048	1.935	1.851	1.784	1.731	1.687
14	3.117	2.452	2.201	2.041	1.928	1.845	1.780	1.726	1.683
15	3.018	2.443	2.192	2.034	1.923	1.839	1.775	1.722	1.679
16	3.068	2.437	2.185	2.028	1.918	1.835	1.765	1.719	1.676
17	2.825	2.421	2.177	2.023	1.913	1.832	1.764	1.716	1.673
18	2.970	2.418	2.172	2.017	1.908	1.827	1.761	1.713	1.671
19	2.857	2.409	2.167	2.014	1.906	1.825	1.758	1.711	1.668
20	2.954	2.404	2.163	2.011	1.902	1.823	1.759	1.709	1.666
21	2.837	2.401	2.159	2.008	1.898	1.820	1.757	1.706	1.665
22	2.900	2.394	2.155	2.003	1.896	1.817	1.756	1.705	1.663
23	2.865	2.389	2.152	2.001	1.894	1.815	1.754	1.704	1.662
24	2.827	2.387	2.148	1.998	1.892	1.813	1.752	1.702	1.661
25	2.856	2.383	2.146	1.995	1.891	1.812	1.750	1.701	1.659
26	2.876	2.377	2.141	1.994	1.888	1.811	1.748	1.699	1.658
27	2.886	2.376	2.141	1.992	1.887	1.808	1.747	1.698	1.657
28	2.830	2.374	2.140	1.990	1.886	1.807	1.747	1.697	1.656
29	2.831	2.370	2.137	1.989	1.885	1.806	1.746	1.697	1.655
30	2.827	2.367	2.134	1.987	1.883	1.805	1.744	1.694	1.654

Valores críticos para la prueba Kruskal-Wallis, $\alpha = .05$

n por grupo	Número de grupos								
	2	3	4	5	6	7	8	9	10
3	—	7.000	4.483	3.640	3.226	2.941	2.718	2.561	2.431
4	19.200	4.826	3.727	3.187	2.854	2.628	2.463	2.339	2.238
5	6.224	4.072	3.377	2.950	2.673	2.482	2.342	2.230	2.142
6	6.500	3.885	3.183	2.812	2.574	2.402	2.271	2.171	2.089
7	5.185	3.646	3.073	2.731	2.508	2.348	2.226	2.130	2.052
8	5.059	3.545	2.998	2.678	2.464	2.309	2.194	2.103	2.028
9	5.428	3.480	2.945	2.634	2.430	2.284	2.171	2.080	2.008
10	5.055	3.415	2.904	2.604	2.404	2.260	2.151	2.065	1.993
11	4.722	3.370	2.870	2.580	2.384	2.243	2.136	2.053	1.981
12	4.428	3.329	2.844	2.560	2.369	2.230	2.123	2.042	1.972
13	4.713	3.295	2.821	2.543	2.354	2.219	2.113	2.032	1.964
14	4.394	3.275	2.804	2.529	2.343	2.210	2.106	2.024	1.958
15	4.327	3.257	2.790	2.517	2.335	2.201	2.099	2.018	1.951
16	4.432	3.231	2.778	2.508	2.326	2.194	2.085	2.013	1.948
17	4.317	3.213	2.764	2.499	2.318	2.188	2.083	2.008	1.943
18	4.354	3.199	2.756	2.491	2.313	2.182	2.078	2.004	1.939
19	4.222	3.189	2.747	2.484	2.308	2.178	2.073	2.000	1.936
20	4.221	3.179	2.739	2.479	2.302	2.174	2.075	1.998	1.933
21	4.204	3.172	2.730	2.473	2.296	2.170	2.072	1.994	1.930
22	4.175	3.162	2.725	2.467	2.293	2.165	2.070	1.991	1.927
23	4.135	3.151	2.722	2.463	2.288	2.162	2.066	1.989	1.926
24	4.089	3.146	2.714	2.458	2.286	2.159	2.063	1.986	1.924
25	4.126	3.141	2.709	2.455	2.283	2.158	2.061	1.985	1.921
26	4.148	3.133	2.703	2.451	2.279	2.156	2.059	1.983	1.920
27	4.158	3.127	2.700	2.450	2.276	2.153	2.056	1.981	1.918
28	4.084	3.125	2.700	2.446	2.275	2.150	2.056	1.980	1.917
29	4.079	3.116	2.695	2.442	2.274	2.149	2.054	1.979	1.916
30	4.067	3.112	2.691	2.440	2.270	2.147	2.052	1.976	1.913

Valores críticos para la prueba Kruskal-Wallis, α=.025

n por grupo	Número de grupos								
	2	3	4	5	6	7	8	9	10
3	—	8.739	6.133	4.792	4.084	3.642	3.315	3.081	2.895
4	—	6.789	4.860	3.995	3.501	3.171	2.936	2.758	2.615
5	9.188	5.570	4.304	3.644	3.236	2.959	2.758	2.602	2.479
6	11.029	5.110	4.016	3.443	3.092	2.843	2.656	2.516	2.404
7	9.046	4.750	3.842	3.321	2.994	2.763	2.592	2.459	2.353
8	7.735	4.589	3.717	3.239	2.930	2.708	2.547	2.421	2.318
9	6.078	4.447	3.636	3.173	2.880	2.671	2.514	2.389	2.290
10	6.680	4.370	3.568	3.127	2.842	2.637	2.487	2.368	2.270
11	6.469	4.292	3.523	3.092	2.813	2.614	2.466	2.349	2.254
12	6.221	4.226	3.480	3.063	2.790	2.595	2.447	2.335	2.241
13	6.315	4.164	3.444	3.037	2.769	2.579	2.434	2.322	2.230
14	6.009	4.126	3.420	3.017	2.752	2.566	2.424	2.311	2.221
15	5.727	4.095	3.397	2.998	2.740	2.555	2.413	2.303	2.212
16	5.926	4.061	3.379	2.986	2.728	2.544	2.393	2.294	2.207
17	5.628	4.034	3.356	2.972	2.716	2.535	2.389	2.289	2.201
18	5.728	4.012	3.344	2.960	2.708	2.527	2.382	2.283	2.195
19	5.609	3.997	3.333	2.950	2.699	2.522	2.376	2.278	2.191
20	5.641	3.980	3.320	2.939	2.692	2.515	2.380	2.274	2.186
21	5.643	3.963	3.306	2.935	2.684	2.508	2.375	2.269	2.183
22	5.624	3.954	3.300	2.927	2.680	2.503	2.371	2.266	2.180
23	5.588	3.935	3.292	2.918	2.673	2.499	2.366	2.262	2.177
24	5.539	3.928	3.281	2.913	2.671	2.494	2.363	2.258	2.174
25	5.482	3.917	3.275	2.908	2.665	2.492	2.359	2.256	2.171
26	5.516	3.902	3.267	2.903	2.661	2.489	2.358	2.253	2.170
27	5.440	3.897	3.261	2.898	2.658	2.486	2.354	2.251	2.166
28	5.448	3.890	3.259	2.893	2.654	2.482	2.352	2.248	2.165
29	5.443	3.881	3.253	2.889	2.652	2.479	2.350	2.248	2.163
30	5.351	3.874	3.247	2.885	2.648	2.477	2.347	2.244	2.160

Valores críticos para la prueba Kruskal-Wallis, $\alpha = .01$

n por grupo	Número de grupos								
	2	3	4	5	6	7	8	9	10
3	—	12.882	8.773	6.630	5.422	4.685	4.189	3.838	3.558
4	—	10.293	6.502	5.188	4.431	3.940	3.593	3.334	3.131
5	25.000	7.953	5.681	4.644	4.019	3.615	3.322	3.103	2.932
6	14.798	7.025	5.213	4.326	3.802	3.437	3.175	2.975	2.819
7	10.969	6.489	4.931	4.133	3.653	3.321	3.081	2.893	2.746
8	11.905	6.115	4.730	4.006	3.556	3.241	3.013	2.836	2.697
9	10.629	5.894	4.601	3.906	3.483	3.186	2.964	2.792	2.657
10	9.593	5.717	4.489	3.837	3.428	3.137	2.927	2.762	2.627
11	9.427	5.584	4.418	3.780	3.383	3.103	2.894	2.734	2.606
12	8.580	5.477	4.349	3.736	3.345	3.072	2.867	2.714	2.586
13	8.799	5.391	4.292	3.698	3.316	3.051	2.851	2.695	2.571
14	8.055	5.322	4.255	3.667	3.290	3.030	2.833	2.679	2.557
15	8.100	5.261	4.217	3.639	3.273	3.015	2.820	2.668	2.547
16	8.065	5.219	4.190	3.621	3.255	3.000	2.789	2.658	2.537
17	7.977	5.178	4.156	3.595	3.237	2.984	2.785	2.648	2.530
18	7.854	5.136	4.134	3.579	3.225	2.973	2.773	2.639	2.521
19	7.709	5.099	4.116	3.566	3.210	2.966	2.765	2.633	2.516
20	7.746	5.075	4.096	3.552	3.203	2.955	2.771	2.627	2.509
21	7.742	5.058	4.077	3.545	3.189	2.948	2.762	2.620	2.506
22	7.542	5.033	4.068	3.532	3.182	2.939	2.759	2.616	2.499
23	7.496	5.005	4.052	3.520	3.174	2.934	2.753	2.610	2.495
24	7.432	4.991	4.037	3.511	3.168	2.928	2.749	2.607	2.492
25	7.484	4.976	4.027	3.500	3.162	2.923	2.743	2.602	2.488
26	7.387	4.954	4.016	3.497	3.159	2.920	2.741	2.600	2.486
27	7.397	4.944	4.008	3.487	3.148	2.913	2.736	2.595	2.481
28	7.388	4.929	4.002	3.481	3.147	2.908	2.731	2.589	2.480
29	7.263	4.916	3.993	3.473	3.145	2.905	2.728	2.589	2.476
30	7.233	4.905	3.983	3.469	3.137	2.903	2.724	2.585	2.471

Valores críticos para la prueba Kruskal-Wallis, $\alpha = .005$

n por grupo	Número de grupos								
	2	3	4	5	6	7	8	9	10
3	—	27.000	11.285	8.106	6.591	5.593	4.936	4.459	4.100
4	—	12.000	7.930	6.205	5.218	4.574	4.120	3.794	3.534
5	99.900	10.092	6.813	5.443	4.650	4.135	3.768	3.494	3.277
6	20.643	8.614	6.204	5.031	4.366	3.902	3.575	3.331	3.132
7	16.608	7.938	5.818	4.782	4.174	3.754	3.458	3.220	3.043
8	13.836	7.382	5.548	4.611	4.045	3.650	3.368	3.152	2.981
9	13.364	7.022	5.379	4.485	3.946	3.579	3.304	3.095	2.932
10	12.598	6.851	5.215	4.388	3.876	3.518	3.259	3.058	2.896
11	10.942	6.636	5.121	4.317	3.818	3.472	3.218	3.022	2.869
12	11.029	6.505	5.034	4.259	3.772	3.436	3.184	2.997	2.843
13	10.339	6.384	4.956	4.208	3.733	3.408	3.163	2.975	2.824
14	10.186	6.282	4.901	4.166	3.701	3.377	3.139	2.954	2.809
15	9.958	6.195	4.854	4.134	3.676	3.360	3.123	2.941	2.795
16	10.049	6.134	4.819	4.109	3.656	3.346	3.087	2.927	2.781
17	9.719	6.082	4.773	4.078	3.630	3.321	3.078	2.915	2.771
18	9.673	6.016	4.744	4.052	3.617	3.310	3.067	2.905	2.764
19	9.324	5.975	4.721	4.037	3.598	3.296	3.055	2.898	2.757
20	9.443	5.939	4.688	4.018	3.588	3.287	3.063	2.889	2.746
21	9.289	5.919	4.669	4.007	3.572	3.276	3.053	2.879	2.743
22	9.120	5.875	4.655	3.988	3.558	3.265	3.047	2.875	2.734
23	9.116	5.835	4.638	3.976	3.548	3.258	3.038	2.867	2.730
24	9.081	5.820	4.618	3.967	3.543	3.251	3.035	2.862	2.726
25	9.021	5.807	4.604	3.954	3.537	3.245	3.027	2.857	2.722
26	8.944	5.770	4.594	3.944	3.534	3.239	3.023	2.856	2.719
27	8.982	5.764	4.580	3.929	3.517	3.232	3.016	2.849	2.710
28	8.750	5.738	4.569	3.925	3.518	3.227	3.012	2.840	2.711
29	8.868	5.718	4.557	3.912	3.511	3.221	3.009	2.842	2.706
30	8.739	5.704	4.543	3.909	3.505	3.218	3.002	2.838	2.701

APÉNDICE J

Estudios de caso[1]

A: Uso de lentes de contacto

Coles *et al.* [6] se enfocaron en la pregunta sobre si los lentes de contacto suaves y nuevos tratados con una solución acondicionadora proporcionarían un mejor grado de comodidad para el paciente que lentes similares sin este tratamiento. Con este fin,

> sesenta y un usuarios de lentes de contacto experimentados sin ninguna característica ocular inusual fueron reclutados de la base de datos de Breman Consultants mediante un anuncio en los periódicos locales. Para resultar electos y poder participar en el estudio, los sujetos debían tener 18 años mínimo y haber estado usando, con éxito, lentes de contacto suaves, al menos por un mes antes del estudio. Se incluyó a personas miopes de hasta −12.00 D con menos de 0.75 D de anisometropía.

Se consideraron apropiados para análisis los datos de 59 de los sujetos, 24 hombres y 35 mujeres. La edad promedio de estos sujetos fue de 31.5 con una desviación estándar de 10.2.

Los pacientes fueron clasificados en dos grupos: aquellos que contestaron "sí" a la pregunta "¿Puede usted usar sus lentes por el tiempo que desee?"; y aquellos que respondieron "no". A los pacientes de ambas categorías se les colocó un lente acondicionado en un ojo seleccionado al azar y se les colocó un lente sin acondicionamiento en el otro ojo. El orden de inserción, es decir, si se colocó primero el lente acondicionado o el lente sin acondicionamiento, también fue al azar. A los pacientes no se les dijo cuál de los lentes estaba acondicionado.

Los lentes fueron insertados durante una visita matutina (a.m.) a la planta de investigación. Después de la inserción se les preguntó a los sujetos con cuál lente se sentían más cómodos. Aproximadamente ocho horas después los sujetos volvieron para una visita vespertina (p.m.) en la que se les pidió nuevamente que seleccionaran el lente más cómodo. Los pacientes no podían indicar "no hay diferencia", por lo que fue necesario hacer elecciones forzadas. Los resultados para esta parte del estudio aparecen a continuación.

En la tabla, no significa "no a la pregunta sobre la duración", sí significa "sí a la pregunta sobre la duración", lente tratado significa "seleccionó el lente tratado con la solución acondicionadora", lente no tratado significa "seleccionó el lente sin la solución acondicionadora", a.m. significa "evaluación durante la visita matutina" y p.m. significa "evaluación durante la visita vespertina".

Además de lo antes mencionado, en cada visita se evaluaron diversas variables de salud ocular para cada ojo. Estas mediciones constituyen una escala continua.

---

[1] Las descripciones de los estudios en esta sección están abreviadas y simplificadas por razones de espacio e irrelevancia para el material de este libro. El lector deberá referirse a la publicación original para más detalles.

TABLA J.1: Conteo de las preferencias de lentes.

	No		Si	
	Lente tratado	Lente no-tratado	Lente tratado	Lente no-tratado
a.m.	10	2	23	24
p.m.	7	5	26	21

TABLA J.2: Medias y desviaciones estándar de las puntuaciones de dolor WOMAC A en la línea base y a las 12 semanas.

	Estándar		Débil		Placebo	
Tiempo	$\bar{x}_1$	s_1	$\bar{x}_2$	s_2	$\bar{x}_3$	s_3
Línea base	10.7	2.1	11.0	2.0	10.9	2.1
12 semanas	7.8	3.9	8.8	3.2	9.3	3.2

B: Brazaletes magnéticos para aliviar el dolor de la osteoartritis

Harlow *et al.* [21] realizaron un ensayo clínico para estudiar la eficacia de brazaletes magnéticos en el alivio del dolor asociado con la osteoartritis de rodilla o cadera. Con este fin se formaron tres grupos al azar. El primer grupo recibió brazaletes magnéticos con potencia estándar (tratamiento estándar); el segundo un brazalete magnético debilitado (tratamiento débil), y el tercero un brazalete desmagnetizado (de imitación o placebo). A los sujetos no se les dijo qué tipo de brazalete recibieron.

El resultado principal fue una puntuación (continua) en la escala de dolor WOMAC A de osteoartritis en miembros inferiores (Western Ontario and McMaster Universities). Las puntuaciones más elevadas en esta escala indican más dolor que las puntuaciones más bajas. La evaluación del dolor se llevó a cabo en la línea base y después de 12 semanas de uso.

Los participantes en el estudio tenían entre 45 y 80 años de edad y fueron seleccionados de entre cinco consultorios rurales de medicina general en Mid Devon, Inglaterra. Todos los sujetos padecían osteoartritis de rodilla o cadera.

Las medias y desviaciones estándar de las edades de los grupos formados al azar fueron las siguientes. Grupo de tratamiento estándar, $\bar{x} = 66.6$, $s = 8.4$; grupo de tratamiento débil, $\bar{x} = 66.8$, $s = 8.3$; grupo placebo o de imitación, $\bar{x} = 66.3$, $s = 9.1$.

Las medias y desviaciones estándar de las puntuaciones de dolor WOMAC A en la línea base y a las 12 semanas aparecen en la tabla J.2. Los tamaños de las muestras de los tres grupos son $n_1 = 65$, $n_2 = 64$ y $n_3 = 64$.

Para evaluar la efectividad de ocultarle a los sujetos el tipo de brazalete que están usando, al final del ensayo cada sujeto tuvo que indicar si creía que estaba usando un brazalete real o uno de imitación. Los resultados se reportan en la tabla J.3.[2]

Una respuesta correcta para los miembros del grupo estándar sería "real" y para el grupo placebo la respuesta correcta sería "de imitación". Como se cree que los imanes colocados en los

[2] No se dio seguimiento a 10 sujetos; por tanto, no aparecen en esta tabla.

TABLA J.3: Indicadores sobre si los sujetos pensaban que estaban usando un brazalete real o de imitación.

Creencia	Estándar	Débil	Placebo
Real	35	24	10
De imitación	5	12	30
No sé	22	23	22

TABLA J.4: Número de muertes en los grupos de tratamiento invasivo y no invasivo para el infarto de miocardio, en varios momentos.

	Invasivo		No invasivo	
Tiempo	Muerto	Vivo	Muerto	Vivo
Alta hospitalaria	21	441	6	452
un mes	23	439	9	449
un año	58	404	36	422
23 meses	80	382	59	399

brazaletes del grupo débil eran demasiado débiles para cualquier fin terapéutico, una respuesta correcta para los miembros de este grupo también sería "de imitación".

C: Tratamiento invasivo versus no invasivo para el infarto de miocardio

Boden *et al.* [3] asignaron al azar a 920 pacientes que habían sufrido un infarto de miocardio recientemente a un tratamiento mediante el método invasivo estándar ($n = 462$) o un método no invasivo ($n = 458$). Uno de los principales resultados fue si el paciente seguía con vida en periodos específicos de tiempo después del tratamiento. La tabla J.4 muestra el número de pacientes muertos y vivos al momento del alta hospitalaria, un mes después del tratamiento, un año después del tratamiento y 23 meses después del tratamiento.

D: Correlación entre el volumen cerebral y el desempeño neurocognitivo en pacientes infectados con VIH

Patel *et al.* [37] realizaron un estudio cuyo fin era determinar si la función neuropsicológica en las personas infectadas con VIH se correlaciona con la pérdida de volumen cerebral. El desempeño neuropsicológico se evaluó mediante la aplicación de una serie de pruebas del funcionamiento neuropsicológico (NPZ-8), en tanto que el volumen cerebral fue medido como el porcentaje del volumen parenquimatoso del cerebro obtenido mediante una resonancia magnética (IRM). Los investigadores plantearon como hipótesis que existe una relación entre estas dos medidas aún antes de ser evidente la disfunción clínica declarada.

La tabla J.5 muestra mediciones del volumen parenquimatoso del cerebro (PBV, por las siglas de *brain parenchymal volume*), del funcionamiento neuropsicológico (NPZ-8), del estado VIH (positivo o negativo), de si el paciente ha manifestado o no la etapa del complejo demencial del SIDA (ADC, por las siglas de *AIDS dementia complex*) y del estado del sistema inmunitario

TABLA J.5: Mediciones de volumen cerebral, funcionamiento neuropsicológico y estado del sistema inmunitario en 15 sujetos VIH positivos y 15 sujetos VIH negativos.

Sujeto	PBV	NPZ-8	Estado VIH	ADC	CD4
1	.791	12	+	+	16
2	.782	5	+	+	324
3	.646	3	+	+	256
4	.740	5	+	+	563
5	.804	2	+	+	321
6	.858	6	+	+	190
7	.729	8	+	+	818
8	.803	0	+	−	355
9	.831	3	+	−	465
10	.826	0	+	−	519
11	.786	0	+	−	87
12	.882	7	+	−	108
13	.889	3	+	−	190
14	.917	0	+	−	573
15	.885	0	+	−	1032
16	.886	0	−		
17	.833	0	−		
18	.851	1	−		
19	.897	0	−		
20	.901	0	−		

(CD4) de los sujetos que participaron en el estudio. Los bajos valores PBV indican menos volumen cerebral, en tanto que puntuaciones NPZ-8 más bajas indican un mejor funcionamiento neuropsicológico. Un valor ADC positivo indica que se ha diagnosticado el complejo demencial del SIDA. Para los pacientes con SIDA, los conteos CD4 más altos se asocian con un mejor funcionamiento del sistema inmunitario.

E: Prueba para la detección de la pancreatitis aguda

La pancreatitis aguda puede ser difícil de diagnosticar de manera oportuna. Para ayudar a esta tarea, Kemppainen *et al.* [25] desarrollaron una rápida prueba para la detección de la pancreatitis basada en una medida inmunocromatográfica de la tripsinogena-2 urinaria.

Para establecer las propiedades del nuevo método, se administró la prueba a 500 pacientes consecutivos que se quejaban de dolor abdominal agudo y que llegaron a las salas de urgencias de dos hospitales. Con posterioridad, se evaluó a cada paciente usando los métodos estándar para determinar si padecían pancreatitis aguda.

De los 53 pacientes a los que se diagnosticó pancreatitis aguda, 50 dieron positivo con la prueba nueva y 3 dieron negativo. De los 447 pacientes a quienes no se diagnosticó pancreatitis aguda, 21 dieron positivo y 426 dieron negativo.

F: La reinfección como una causa de la tuberculosis recurrente

Desde hace mucho se ha pensado que la tuberculosis recurrente es causada normalmente por la reactivación de una infección endógena y no por una nueva infección exógena. Para comprobar la validez de este supuesto, Van Rie *et al*. [48] realizaron un análisis de ADN en pares de muestras de *Mycobacterium tuberculosis* de 16 pacientes que sufrieron una recaída de tuberculosis pulmonar después de un tratamiento curativo para la tuberculosis secundaria. Los pacientes vivían en áreas de Sudáfrica en las cuales la tuberculosis es endémica.

El análisis de ADN mostró que en 12 de los 16 pacientes, la reinfección estaba relacionada con un agente causal diferente al de la recurrencia del agente original. Los investigadores concluyeron que la "reinfección exógena parece ser una causa importante de la tuberculosis secundaria después de un tratamiento previo en una zona con alta incidencia de esta enfermedad".

G: Esclerosis múltiple en mellizos

No ha sido definido con claridad el rol de los factores ambientales y genéticos en la determinación de la susceptibilidad a la esclerosis múltiple. Para ayudar a esta tarea, Ebers *et al.* [11] examinaron los registros de 5,463 pacientes que asistían a 10 clínicas de esclerosis múltiple en Canadá. Se encontró que 70 de estos pacientes tenían un hermano mellizo (27 eran monocigóticos[3] y 43 eran dicigóticos[4]). Además, se estudiaron 4,582 pacientes que no tenían un hermano mellizo, pero sí un hermano. Se encontró que siete de los mellizos monocigóticos tenían un mellizo que también tenía la enfermedad, en tanto que sólo uno de los 43 mellizos dicigóticos tenía un mellizo enfermo. De los 4,582 pacientes no mellizos, 87 tenían un hermano enfermo. (Supondremos que los pacientes no mellizos tenían un solo hermano).

H: Comparación entre dos tratamientos para la cistinosis nefrótica

Smolin *et al.* [43] compararon la efectividad de dos sustancias, la cisteamina (MEA) y la fosfocisteamina (MEAP), para tratar a niños con cistinosis nefrótica. La MEA se usa actualmente para el tratamiento de esta enfermedad, pero la MEAP tiene las ventajas de mejor olor y sabor y puede ser, por tanto, más agradable para los niños. Un resultado fundamental del estudio era determinar si existe diferencia entre la habilidad de las dos sustancias para disminuir el contenido de cistina en los leucocitos.

El estudio se realizó administrando las dos sustancias a seis niños. Los autores reportan que: "El porcentaje de disminución en el contenido de cistina en los leucocitos obtenido con la administración de MEA (61.9%) no fue considerablemente diferente a la disminución observada cuando se administró MEAP (65.3%)". De lo cual concluyen: "MEA y MEAP parecen ser igualmente efectivos en cuanto a sus propiedades para reducir la cistina".

I: Comparación entre dos tratamientos para el infarto agudo de miocardio

Al tratar a pacientes que experimentan un infarto agudo de miocardio es deseable la rápida lisis (disolución) de trombos coronarios (obstrucciones). Investigadores del ensayo COBALT sobre la administración de alteplasa (The Continuous Infusion versus Double-Bolus Administration of

[3] De un solo óvulo fertilizado.

[4] De dos óvulos diferentes.

Alteplase, COBALT) [46] compararon los índices de mortalidad asociados con dos métodos de lisis de los trombos. El primer método, la infusión acelerada (IA) requiere alrededor de 90 minutos, en tanto que el método doble bolus (D-B) requiere aún menos tiempo.

El estudio se realizó asignando al azar a 7,169 pacientes a los dos tratamientos, 3,685 al grupo D-B y 3,684 al grupo IA. El resultado de interés era mortalidad por cualquier causa durante el periodo de 30 días después del tratamiento. El propósito del estudio era determinar si D-B podía ser considerado terapéuticamente equivalente a IA en lo concerniente a la mortalidad después del tratamiento. Siendo más precisos, los investigadores considerarían D-B por lo menos tan bueno como IA si la proporción de muertes asociada con el tratamiento D-B excede la proporción del tratamiento IA por menos de 0.004. La proporción de muertes en el grupo D-B fue de 0.0798 y la proporción del grupo IA fue 0.0753.

J: Tratamientos oral e intravenoso para la fiebre en pacientes con cáncer

La terapia estándar para el tratamiento de la fiebre en pacientes con cáncer y granulocitopenia (conteos bajos de glóbulos blancos) implica agentes antimicróbicos administrados de forma intravenosa. Si se demostrara que la terapia oral puede ser tan efectiva como la terapia intravenosa, se podrían lograr mejoras notables en la calidad de vida y en los costos de la terapia.

Para determinar si es posible establecer esta equivalencia, Kern *et al.* [26] realizaron un estudio en el que, al azar, se asignó a pacientes con cáncer que padecían fiebre y granulocitopenia a los tratamientos oral e intravenoso. Todos los pacientes fueron hospitalizados y se les recetaron dichos tratamientos. Se consideraba que un tratamiento tenía éxito en un paciente en particular si todos los demás se lograban sin cambios en el régimen del tratamiento: "... la temperatura era normal por al menos tres días consecutivos (o dos días en pacientes con fiebre inexplicable y rápida recuperación de la granulocitopenia), los síntomas y las señales de infección en lugares de infección identificables habían desaparecido, el principal agente patógeno había sido erradicado y la principal infección documentada no había reaparecido en la primera semana después del fin del tratamiento".

Los resultados mostraron que 138 de los 161 pacientes que recibieron terapia oral y 127 de los 151 pacientes que recibieron terapia intravenosa se consideraron casos exitosos. Los investigadores decidieron que se declararía la equivalencia si la diferencia absoluta entre las proporciones de éxitos atribuidos a los dos regímenes de tratamiento era .10 o menos.

K: Educación para la seguridad en las granjas

La agricultura es una de las ocupaciones más peligrosas en Estados Unidos. Sólo las volcaduras de tractores representan aproximadamente 132 muertes al año. Además, alrededor de 500 trabajadores sufren lesiones incapacitantes a diario. Por desgracia, en promedio, 104 adolescentes de 19 años o menos mueren en accidentes de granja al año.

Liller *et al*. [32] realizaron un proyecto a gran escala para instruir a los jóvenes sobre los riesgos y los asuntos de seguridad asociados con la agricultura. Como parte del proyecto, se desarrolló un currículo enfocado a la seguridad en las granjas. Con la cooperación del sistema escolar local se seleccionaron cinco escuelas localizadas en áreas del condado primordialmente agrícolas. Por un largo periodo se dieron lecciones de seguridad en las granjas a todos los estudiantes de quinto grado de las cinco escuelas. Los alumnos de quinto grado que participaron hicieron un examen de seguridad especialmente desarrollado en las granjas antes y después de las lecciones.

TABLA J.6: Medias de las puntuaciones pre-prueba y post-prueba de cinco escuelas que participaron en un estudio de seguridad en las granjas.

Escuela	Media pre-prueba	Media post-prueba
1	5.93	8.68
2	5.65	7.83
3	6.84	8.09
4	5.68	7.78
5	6.05	8.03

TABLA J.7: Severidad de las lesiones sufridas pos empleados escolares clasificados por edad.

Edad	Lesión menor	Lesión moderada	Lesión severa
< 30	255	57	14
30 – 39	332	105	44
40 – 49	533	201	46
50 +	561	205	66

Se calificaron todos los exámenes y se registró la puntuación promedio pre-prueba y post-prueba de *cada escuela*. Estos promedios aparecen en la tabla J.6.

L: Evaluación del impacto de las enfermeras de escuela

Perrin, Goad y Williams [38] evaluaron el impacto económico de emplear enfermeras en las secundarias de un gran sistema escolar del sureste. Las enfermeras tenían autorización para tratar al personal así como a los estudiantes. Como parte de su evaluación, examinaron la relación entre la edad de los empleados que sufrían una lesión relacionada con su trabajo y la severidad de dicha lesión. Los datos recabados a este respecto aparecen en la tabla J.7.

Se definió como lesión menor aquella que no requería tratamiento, las lesiones moderadas eran las que requerían tratamiento pero no así hospitalización, y las lesiones severas eran las que requerían hospitalización.

M: Tratamiento de los trastornos del sueño en los alpinistas

Una queja común entre los alpinistas es su inhabilidad para dormir lo suficiente una vez que alcanzan alturas de moderadas a altas. La falta de sueño se puede tratar con ciertas clases de sustancias, pero han surgido dudas sobre si estas sustancias pueden interferir con la respiración, una consideración importante a dichas alturas.

Röggla, Moser y Röggla [40] realizaron un estudio diseñado para evaluar el efecto del temazepam en los niveles de oxígeno y dióxido de carbono de los alpinistas a 171 y 300 metros de altura. Se evaluaron los niveles de oxígeno y dióxido de carbono de siete alpinistas, antes de

TABLA J.8: Niveles de oxígeno y dióxido de carbono de siete alpinistas antes y después de tomar un medicamento para dormir.

	Oxígeno		Dióxido de carbono	
Sujeto	Antes	Después	Antes	Después
1	9.3	8.6	4.3	4.4
2	8.6	8.4	4.4	4.7
3	8.9	8.2	4.4	4.7
4	9.1	8.3	4.0	4.3
5	8.5	8.1	4.1	4.4
6	9.1	8.1	4.4	4.7
7	9.4	8.4	4.0	4.8

TABLA J.9: Factores de riesgo potencial de abuso sexual por adolescentes víctimas de abuso sexual.

Factor de riesgo	**Víctima de abuso solamente (n = 14)**	**Víctimas de abuso y abusador (n = 11)**
Experimentan violencia intrafamiliar	5	10
Son testigos de violencia intrafamiliar	5	9
Rechazo por parte de la familia	8	10
Rechazo por parte de sus iguales	3	5

administrar la sustancia y una hora después. En la tabla J.8 se muestran los resultados obtenidos antes y después de tomar el medicamento a los 300 metros.

N: Factores de riesgo relacionados con el abuso sexual por varones adolescentes víctimas de abuso sexual

Muchos varones adolescentes que sufrieron abuso sexual se convierten en abusadores, mientras que una considerable proporción de ellos no. Skuse *et al.* [42] investigaron los factores de riesgo asociados con el comportamiento abusivo por parte de los adolescentes abusados. En su estudio identificaron a 25 varones adolescentes, de entre 11 y 15 años y 11 meses de edad, víctimas de abuso sexual, de los cuales 11 habían abusado de otros chicos y 14 no lo habían hecho. Los factores de riesgo investigados incluyeron experimentar violencia intrafamiliar, ser testigos de ella, sufrir rechazo por parte de la familia, pobre identificación con una figura paterna y otros.

Los resultados parciales de su estudio se muestran en la tabla J.9.

La primera columna de esta tabla muestra el factor de riesgo potencial, en tanto que la segunda y tercera columnas muestran el número de sujetos en los dos grupos que experimentaron ese factor en particular.

O: Cómo leer un trabajo escolar

En un artículo titulado "Cómo leer un trabajo escolar: Estadística para los no estadísticos. I: Diferentes tipos de datos necesitan diferentes tipos de pruebas estadísticas", Greenhalgh [19] proporciona directrices (o recomendaciones) para los investigadores médicos sobre el uso de los procedimientos estadísticos. Varios de los puntos de este trabajo son de particular interés.[5]

1. El autor da ejemplos de varias escalas de medición y apunta: "A menudo se usan números para etiquetar las propiedades de las cosas". Por ejemplo, se pueden asignar números para representar la estatura y el peso. En contraste, también se pueden usar números para etiquetar la propiedad "ciudad de origen" donde 1 = Londres, 2 = Manchester, 3 = Birmingham, y así sucesivamente. Finalmente, se pueden asignar números para representar "gusto por x" donde 1 = nada, 2 = un poco, y 3 = mucho.

2. El autor enuncia: "En general, las pruebas paramétricas son más potentes que las no paramétricas y por lo tanto, así deben usarse en la medida de lo posible".

3. El autor enuncia: "Las pruebas no paramétricas consideran el orden de rango de los valores (cuál es el más pequeño, cuál viene después, y así sucesivamente) e ignora las diferencias absolutas entre ellos. Como se podrá imaginar, la significancia estadística es más difícil de demostrar con las pruebas no paramétricas...".

4. Después de discutir la curva normal, el autor afirma que algunas variables biológicas, como el peso corporal, muestran una distribución "normal asimétrica".

5. El autor enuncia: "No obstante, usar pruebas basadas en la distribución normal para analizar datos no distribuidos normalmente es, definitivamente, hacer trampa".

6. El autor describe la prueba de Kruskal-Wallis así: "En efecto, una generalización de la prueba *t* pareada o la prueba de Wilcoxon para pares relacionados, donde se hacen tres o más conjuntos de observaciones sobre una sola muestra".

7. Se da un ejemplo de una situación en la que se podría usar la prueba de Kruskal-Wallis para "determinar si el nivel de glucosa en plasma es más elevado una hora, dos horas o tres horas después de una comida".

8. El autor indica que la correlación de rangos de Spearman "evalúa la fuerza de la asociación lineal entre dos variables continuas".

9. Aunque ahonda en el supuesto de normalidad, el autor no trata temas como la homocedasticidad y la independencia.

[5] Las caracterizaciones hechas aquí tienen como fin ayudarle a comprender asuntos estadísticos y no son una evaluación de este artículo. La utilidad de los comentarios hechos en este trabajo deberá ser evaluada sólo después de leer el artículo en su totalidad.

TABLA J.10: Pacientes diabéticos en Tayside, Escocia, clasificados por tipo de diabetes, género y edad.

	Tipo I		Tipo II	
Edad	Varón	Mujer	Varón	Mujer
0-14	29	42	0	0
15-24	70	65	0	0
25-44	252	157	22	16
45-65	90	74	163	152
> 65	10	18	195	242
Σ	451	356	380	410

P: Monitoreo de la glucosa en sangre

Evans *et al*. [12] realizaron un estudio cuyo fin era, en parte, investigar patrones de autocontrol de la concentración de glucosa en la sangre en los pacientes diabéticos tipo I y tipo II. Durante el estudio, los autores construyeron una base de datos que representaba la población de pacientes diabéticos en Tayside, Escocia. Entre otras variables, la base de datos mostraba cantidades de pacientes diabéticos clasificados por el tipo de diabetes (tipo I o tipo II), edad y género. Las categorías de pacientes se muestran en la tabla J.10.

APÉNDICE K

Respuestas a los ejercicios

Capítulo uno

1.1 (a) La organización y la condensación de los datos y (b) la forma de hacer deducciones de poblaciones basándose en muestras.

1.3 El dato se refiere a los registros de las mediciones hechas sobre las características.

A. Las siguientes respuestas se refieren al estudio de caso A.

1.5 La muestra consiste de 24 varones y 35 mujeres que participan en el estudio. Esta muestra se caracteriza por ser de una población popular. Podemos calificar los conteos de las preferencias de lentes en la tabla J.1 como una muestra de una población estadística.

1.7 La población popular consiste en su mayoría de hombres y mujeres que tienen por lo menos 18 años de edad sin ninguna característica ocular inusual, que han usado lentes de contacto suaves sin problemas al menos un mes. También están presentes las personas miopes de hasta −12.00 D con menos de .75 D de anisometropía. Las cantidades de las preferencias de lentes se pueden usar para calcular la preferencia de lentes en la población estadística.

1.9 No. La población estadística no estuvo disponible para el estudio, por lo que no se puede obtener parámetro alguno.

B. Las siguientes respuestas se refieren al estudio de caso B.

1.11 La muestra consiste en 65 + 64 + 64 = 193 sujetos que participaron en el estudio. Como esta muestra se caracteriza por estar compuesta por personas y no por características, se considera como muestra de una población popular. También tenemos tres muestras compuestas de valores de intensidad de dolor. Estas muestras son de una población estadística.

1.13 La población popular consta de personas entre 45 y 80 años de edad que padecen osteoartritis de cadera o rodilla y que ven a médicos generales en el área de Mid Devon, Inglaterra. La población estadística consta del dolor cuya intensidad puede describirse como un valor de la escala de dolor de osteoartritis en miembros inferiores de las universidades Western Ontario y McMaster. La media y la desviación estándar de esta población estadística muy probablemente se encuentran alrededor de 67 y 8 o 9, respectivamente.

1.15 No se proporciona ningún parámetro porque la población estadística no estaba disponible para el estudio.

F. Las siguientes respuestas se refieren al estudio de caso F.

1.17 Probablemente no. Estos sujetos vivieron en áreas de Sudáfrica donde la tuberculosis es endémica y tiene una alta incidencia. No podemos suponer que el mecanismo para la reinfección encontrado en un área de alta incidencia es el mismo que el encontrado en un área de baja incidencia.

Capítulo dos

2.1 (a) Ordinal. Podemos determinar si una enfermera tiene más antigüedad que otra, pero no podemos decir cuánta más (o menos).

(b) La *variable* días de enfermedad es continua ya que se pueden fraccionar los días a cualquier nivel, por ejemplo, 2.313, etc. Sin embargo, los *datos* que se proporcionan aquí podrían considerarse discretos ya que sólo se registraron mediciones de días completos. La clasificación de los datos como continuos o discretos en ocasiones es un poco arbitraria.

(c) i. 1, 8, 5

ii. $2 + 9 + 1 + 0 + 5 + 4 + 6 + 7 + 8 + 8 = 50$

iii. $6 + 3 + 7 + 8 + 9 + 2 + 8 + 9 + 6 + 8 + 5 = 71$

iv. Aquí se pide que sumemos todas las puntuaciones. Pero ya sumamos las primeras 10 en 1 (c) ii y las restantes en 1 (c) iii, por lo que el total es $50 + 71 = 121$.

(d) De las reglas de la suma de la página 13 sabemos que

$$\sum_{i=1}^{n}(x_i + c) = \sum_{i=1}^{n} x_i + nc$$

por lo que, para el problema en cuestión

$$\sum_{i=1}^{n}(x_i + 2) = \sum_{i=1}^{n} x_i + (21)(2) = \sum_{i=1}^{n} x_i + 42$$

(e)

Días de enfermedad	Frecuencia	Frecuencia acumulativa	Frecuencia relativa	Frecuencia relativa acumulativa
9	3	21	.143	1.000
8	5	18	.238	.857
7	2	13	.095	.619
6	3	11	.143	.524
5	2	8	.095	.381
4	1	6	.048	.286
3	1	5	.048	.238
2	2	4	.095	.190
1	1	2	.048	.095
0	1	1	.048	.048

(f) Las tres gráficas se muestran en la figura K.1. Observe que catalogamos el eje x con valores de datos y no con límites superiores e inferiores reales de los intervalos de los valores de datos, como lo hicimos con las figuras 2.2, 2.3 y 2.4. Ésta es una práctica común con datos no agrupados, pero también sería correcto utilizar rótulos de límites superiores e inferiores reales.

(g) i. Por la ecuación 2.1 obtenemos

$$\overline{x} = \frac{\sum x}{n} = \frac{121}{21} = 5.762.$$

Ordenando los días de enfermedad del más pequeño al más grande y aplicando la ecuación 2.3 nos da

$$\text{Mediana } (n \text{ par}) = x_{\frac{n+1}{2}} = x_{\frac{21+1}{2}} = x_{11} = 6.$$

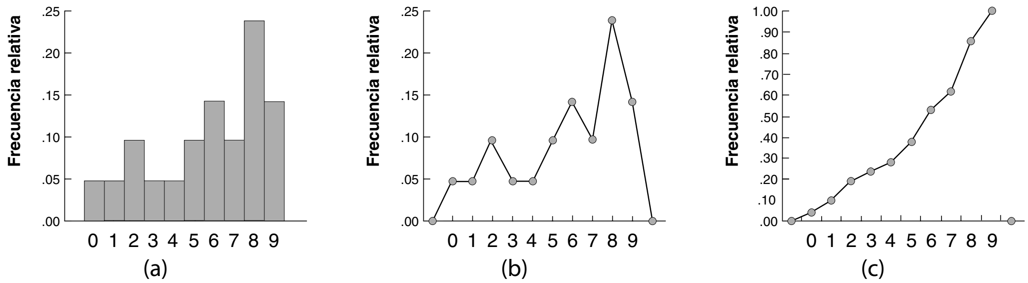


FIGURA K.1: Figuras para el ejercicio 2.1f: (a) Frecuencia relativa (b) Frecuencia relativa (c) Polígono de frecuencia relativa acumulativa.

De lo mencionado en la página 31 y la observación de la distribución de la frecuencia en el ejercicio 1e, notamos que 8 es la observación que ocurre con más frecuencia y es, por lo tanto, la moda.

ii. Por las ecuaciones 2.12 y 2.16

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1} = \frac{853 - \frac{(121)^2}{21}}{21-1} = 7.790$$

y

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{853 - \frac{(121)^2}{21}}{21-1}} = 2.791.$$

iii. Por las ecuaciones 2.6 y 2.7

$$\text{Rango (exclusivo)} = x_L - x_S = 9 - 0 = 9$$

y

$$\text{Rango (inclusivo)} = URL_L - LRL_S = 9.5 - (-.5) = 10$$

iv. Las puntuaciones z a continuación se generaron mediante la ecuación 2.23 usando $\bar{x} = 5.762$ y $s = 2.791$. Por ejemplo, la enfermera 1 tuvo 2 días de enfermedad, por lo que la puntuación z asociada es

$$z = \frac{x - \bar{x}}{s} = \frac{2.0 - 5.762}{2.791} = -1.348.$$

Número de enfermera	z	Número de enfermera	z	Número de enfermera	z
1	–1.348	8	0.444	15	1.160
2	1.160	9	0.802	16	–1.348
3	–1.706	10	0.802	17	0.802
4	–2.064	11	0.085	18	1.160
5	–0.273	12	–0.990	19	0.085
6	–0.631	13	0.444	20	0.802
7	0.085	14	0.802	21	–0.273

v. Por la ecuación 2.17 de la página 38 y las distribuciones de frecuencia y frecuencia acumulativa construidas en el ejercicio 1e, obtenemos lo siguiente.

$$P_{15} = LRL + (w)\left[\frac{(pr)(n) - cf}{f}\right] = 1.5 + (1)\left[\frac{(.15)(21) - 2}{2}\right] = 2.075$$

$$P_{50} = 5.5 + (1)\left[\frac{(.5)(21) - 8}{3}\right] = 6.333$$

$$P_{80} = 7.5 + (1)\left[\frac{(.8)(21) - 13}{5}\right] = 8.26$$

Note que $P_{50} = 6.33$, lo que constituye una definición de la mediana, contrasta con el valor medio de 6 obtenido por el más burdo, pero más comúnmente usado, método empleado en el ejercicio 2.1 (g) i.

vi. De lo mencionado en la página 41 sabemos que el rango percentil dependerá de nuestra definición del rango percentil en relación con las tres puntuaciones. Así, podríamos usar las ecuaciones 2.20, 2.21 o 2.22 para el cálculo. Utilizaremos la ecuación 2.21 pero también sería correcto utilizar cualquiera de las otras dos.

$$PR_2 = 100\left[\frac{(.5)(f) + cf}{n}\right] = 100\left[\frac{(.5)(2) + 2}{21}\right] = 14.286 \text{ o } 14$$

$$PR_5 = 100\left[\frac{(.5)(2) + 6}{21}\right] = 33.333 \text{ o } 33$$

$$PR_8 = 100\left[\frac{(.5)(5) + 13}{21}\right] = 73.810 \text{ o } 74$$

vii. Por las ecuaciones 2.25 y 2.26 de la página 45

$$\text{Sesgo } = \frac{\sum z^3}{n} = \frac{-12.480}{21} = -.594$$

y

$$\text{Curtosis } = \frac{\sum z^4}{n} = \frac{41.931}{21} = 1.997.$$

A. Las siguientes respuestas se refieren al caso de estudio A.

2.3 Sí, a.m./p.m., sí/no, tratados/no tratados. Éstas son variables de nivel nominal.

2.5 Como la edad media es de la muestra de sujetos usada en el estudio, el símbolo correcto sería $\bar{x}$. Usaríamos s para representar la desviación estándar de la muestra (edad).

2.7 Por la ecuación 2.23

$$z = \frac{x - \bar{x}}{s}$$

entonces, las puntuaciones z para los sujetos con edades entre 18 y 75 serían

$$z_{18} = \frac{18.0 - 31.5}{10.2} = -1.32$$

y

$$z_{75} = \frac{75.0 - 31.5}{10.2} = 4.26$$

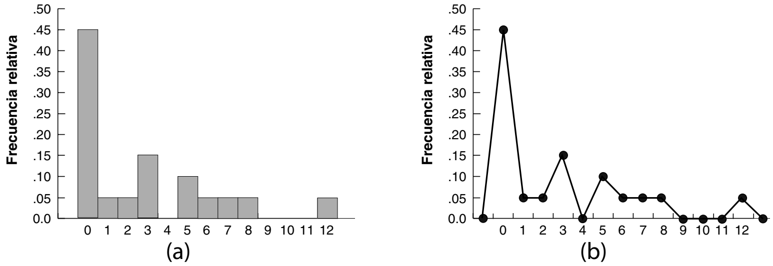


FIGURA K.2: Figuras para el ejercicio 2.9: (a) Histograma de frecuencia relativa NPZ-8 (b) Polígono de frecuencia relativa.

D. Las siguientes respuestas se refieren al estudio de caso D.

2.9 Véase la figura K.2.

2.11 Por la ecuación 2.1.

$$\bar{x} = \frac{\sum x}{n} = \frac{16.537}{20} = .827.$$

Por la ecuación 2.4

$$\text{Mediana } (n \text{ impar}) = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_{10} + x_{11}}{2} = \frac{.831 + .833}{2} = .832$$

o por el método demostrado en la página 30, notamos que cualquier punto entre .8315 y .8325 satisface la definición de la media, por lo que tomamos el punto medio de este intervalo o .832 como la mediana.

Ya que todas las puntuaciones ocurren con frecuencia uno, no existe moda.

2.13 Considerando que el primero, segundo y tercer cuartil son, respectivamente, P_{25}, P_{50} y P_{75}, (ver la página 40) calculamos cada uno de estos valores.

Como (.25) (20) = 5, P_{25} es el punto en la escala debajo del cual caen 5 observaciones. Como ninguna observación cae debajo de –.5 a .5 y nueve caen debajo de .5, el punto debajo del cual caen 5 observaciones debe ser el intervalo –.5 a .5 (es decir, los límites inferior y superior reales del intervalo cero). Si usamos esta información con la ecuación 2.17 que enuncia

$$P_P = LRL + (w)\left[\frac{(pr)(n) - cf}{f}\right]$$

nos da

$$P_{25} = -.5 + (1)\left[\frac{(.25)(20) - 0}{9}\right] = .056.$$

Como (.5)(20) = 10 y 10 observaciones caen debajo del límite superior real de 1.5, P_{50} = 1.5. Considerando que (.75)(20) = 15 y que 14 observaciones caen debajo del límite superior real de 4.5 y 16 caen debajo de 5.5, empleamos la ecuación 2.17 como sigue.

$$P_{75} = 4.5 + (1)\left[\frac{(.75)(20) - 14}{2}\right] = 5$$

2.15 Usando la media y la desviación estándar calculadas en 2.10 y 2.12 y la ecuación 2.23 que enuncia

$$z = \frac{x - \bar{x}}{s}$$

calculamos

$$z_0 = \frac{0 - 2.75}{3.432} = -.801$$

$$z_1 = \frac{1 - 2.75}{3.432} = -.510$$

$$z_2 = \frac{2 - 2.75}{3.432} = -.219$$

$$z_3 = \frac{3 - 2.75}{3.432} = .073$$

$$z_5 = \frac{5 - 2.75}{3.432} = .656$$

$$z_6 = \frac{6 - 2.75}{3.432} = .947$$

$$z_7 = \frac{7 - 2.75}{3.432} = 1.238$$

$$z_8 = \frac{8 - 2.75}{3.432} = 1.530$$

$$z_{12} = \frac{12 - 2.75}{3.432} = 2.695$$

Usando la ecuación 2.25 y tomando en cuenta que la frecuencia asociada con cada valor z produce

$$\begin{aligned}\text{Sesgo} &= \frac{\sum z^3}{n}\\ &= \frac{(9)(-.801^3)+(-.510^3)+(-.219^3)+(3)(.073^3)+(2)(.656^3)+(.947^3)+(1.238^3)+(1.530^3)+(2.695^3)}{20}\\ &= 1.085\end{aligned}$$

Esto muestra un sesgo positivo que fue la conclusión a la que se llegó por inspección en 2.14.

O. Las siguientes respuestas se refieren al estudio de caso O.

2.17 La estatura y el peso se describen mejor como mediciones de la relación de nivel, ciudad de origen como nominal y gusto por x como ordinal.

Capítulo tres

3.1 (a) Por la ecuación 3.3 de la página 56

$$P(\bar{I}\,|F) = \frac{P(\bar{I}F)}{P(F)} = \frac{.36}{.46} = .783.$$

(b) Leyendo de la tabla

$$P(F) = .46$$

(c) Por la ecuación 3.2 de la página 56

$$P(M \cup I) = P(M) + P(I) - P(MI) = .54 + .32 - .22 = .64.$$

(d) Leyendo de la tabla

$$P(FI) = .10.$$

(e) Por la ecuación 3.3 de la página 56

$$P(F|I) = \frac{P(FI)}{P(I)} = \frac{.10}{.32} = .313.$$

(f) Por la ecuación 3.2 de la página 56

$$P(M \cup F) = P(M) + P(F) - P(MF) = .54 + .46 - .00 = 1.0.$$

(Tiene que ser una u otra, ¿cierto?)

3.3 De la información dada sabemos que $P(F) = .50$, $P(T) = .20$ y $P(FT) = .10$ donde F indica femenino y T representa al partidario del incremento del impuesto. Notamos que $P(F)(T) = (.50)(.20) = .10$ que es igual a $P(F)(T)$, por lo que por 3.5, género y apoyo al incremento son independientes.

Podríamos igualmente haber aplicado 3.4 de la página 56 anotando $P(T|F) = \frac{P(TF)}{P(F)} = \frac{.10}{.50} = .20$, que es igual a $P(T) = .20$. ¿Habríamos llegado a la misma conclusión si hubiéramos preferido comparar $P(F \mid T)$ y $P(F)$?

3.5 La tabla de probabilidad es como sigue.

	B	$\overline{B}$	
A	.624	.176	.40
$\overline{A}$	.096	.104	.60
	.70	.30	

Comenzamos por llenar los valores marginales. Tenemos $P(\overline{A}) = .20$, por lo que $P(A) = 1 - P(\overline{A}) = 1 - .20 = .80$. De igual forma, $P(B) = .72$, por lo que $P(\overline{B}) = 1 - .72 = .28$. Ahora nos enfocamos a las probabilidades conjuntas. Como sabemos que $P(\overline{B}|A) = \frac{P(A\overline{B})}{P(A)}$, podemos multiplicar ambos lados por $P(A)$ para obtener

$$P(\overline{B}|A)P(A) = P(A\overline{B})$$

que es la probabilidad conjunta que necesitamos. Entonces calculamos $P(A\overline{B}) = (.22)(.80) = .176$. Ahora podemos utilizar los marginales para llenar las otras probabilidades conjuntas.

$$P(AB) = P(A) - P(A\overline{B}) = .800 - .176 = .624$$
$$P(\overline{A}B) = P(B) - P(AB) = .720 - .624 = .096$$
$$P(\overline{A}\overline{B}) = P(\overline{B}) - P(A\overline{B}) = .280 - .176 = .104$$

3.7 (a) Por la ecuación 3.6 de la página 58

$$\text{Sensibilidad} = P(+|D) = \frac{.065}{.070} = .929$$

(b) Por la ecuación 3.7 de la página 58

$$= P(-|\bar{D}) = \frac{.92}{.93} = .989$$

(c) Por la ecuación 3.8 de la página 58

$$PPV = P(D|+) = \frac{.065}{.075} = .867$$

(d) Por la ecuación 3.9 de la página 58

$$NPV = P(\bar{D}|-) = \frac{.920}{.925} = .995$$

3.9 Pensemos que U representa a una persona de menos de 40 años, $\bar{U}$ a una persona de 40 o más y S un partidario de la vacunación. Entonces, de la información dada tenemos que $P(U) = .40$, $P(S|U) = .72$ y $P(S|\bar{U}) = .52$. Se nos pide encontrar $P(U|S)$. Podríamos proceder como en el ejercicio 3.5, pero un método más directo es aplicar la regla de Bayes como se expresa por medio de la ecuación 3.13 de la página 61, que enuncia

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$$

Para el problema en cuestión se transforma en

$$P(U|S) = \frac{P(S|U)P(U)}{P(S|U)P(U) + P(S|\bar{U})P(\bar{U})} = \frac{(.72)(.40)}{(.72)(.40) + (.52)(.60)} = .48$$

3.11 (a) Debemos encontrar el área entre un punto sobre la curva (98) y la media (80), por lo que solamente tenemos que encontrar la puntuación Z para el punto y leer el área asociada del Apéndice A. Entonces, por la ecuación 2.24 de la página 42

$$Z = \frac{x - \mu}{\sigma} = \frac{98 - 80}{12} = 1.5.$$

A partir de la columna 2 del Apéndice A, encontramos que el área es .4332.

(b) Esta área está en la cola inferior de la curva, por lo que solamente necesitamos encontrar la puntuación Z para 74 y leer el área asociada de la columna tres de la tabla.

$$Z = \frac{74 - 80}{12} = -.5.$$

A partir de la columna 3 del Apéndice A, encontramos que el área es .3085.

(c) Ésta es el área debajo de un punto que está por encima de la media de la distribución. Encontraremos el área entre 80 y 82 y agregaremos el resultado a .5, que es el área debajo de la media de la distribución.

$$Z = \frac{82 - 80}{12} = -.17.$$

De la columna dos obtenemos .0675, por lo que el área debajo de 82 es .5000 + .0675 = .5675.

(d) El área está entre dos puntos, uno encima y otro debajo de la media de la distribución. Encontraremos el área asociada con cada uno y sumaremos las dos para encontrar el área total entre los dos puntos.

$Z = \frac{94-80}{12} = 1.17$ que tiene un área asociada de .3790.

$Z = \frac{72-80}{12} = -.67$ que tiene un área asociada de .2486.

El área total entre los dos puntos es, por consiguiente, .3790 + .2486 = .6276.

(e) Ésta es el área entre dos puntos, ambos por debajo de la media de la distribución. La estrategia es encontrar las áreas asociadas con cada uno de los dos puntos y sustraer el área más pequeña de la más grande para encontrar el área entre los dos puntos.

$Z = \frac{60-80}{12} = -1.67$ que tiene un área asociada de .4525.

$Z = \frac{56-80}{12} = -2.00$ que tiene un área asociada de .4772.

El área entre los dos puntos es, entonces, .4772 – .4525 = .0247.

(f) Ésta es el área encima de un punto que está por encima de la media de la distribución. Así, es un área en la cola del lado derecho de la distribución.

$Z = \frac{104-80}{12} = 2.00$ que tiene un área asociada en la columna tres de .0228.

(g) Ésta es el área debajo de un punto que está por debajo de la media de la distribución. Así, es un área en la cola del lado izquierdo de la distribución.

$Z = \frac{54-80}{12} = -2.17$ que tiene un área asociada en la columna tres de .0150.

(h) Ésta es el área entre dos puntos, ambos por encima de la media de la distribución. La estrategia es encontrar las áreas asociadas con cada uno de los dos puntos y sustraer el área más pequeña de la más grande para encontrar el área entre los dos puntos.

$Z = \frac{82-80}{12} = .17$ que tiene un área asociada de .0675.

$Z = \frac{94-80}{12} = 1.17$ que tiene un área asociada de .3790.

El área entre los dos puntos es, entonces, .3790 – .0675 = .3115.

A. Las siguientes respuestas se refieren al estudio de caso A.

3.13 $P(M) = \frac{24}{59} = .41.$

3.15 (a) $P(T)$ puede leerse directamente del margen de la tabla y es .56.

(b) $P(Y)$ puede leerse directamente del margen de la tabla y es .80.

(c) $P(T|Y) = \frac{23}{47} = .49.$

(d) $P(T|N) = \frac{10}{12} = .83.$[1]

(e) Para 3.15a: La probabilidad de seleccionar al azar a un sujeto que había escogido un lente tratado.

Para 3.15b: La probabilidad de seleccionar al azar a un sujeto que había respondido sí a la pregunta sobre la duración.

Para 3.15c: La probabilidad de seleccionar al azar *de entre los sujetos que respondieron sí a la pregunta sobre la duración*, a un sujeto que escogió el lente tratado.

Para 3.15d: La probabilidad de seleccionar al azar *de entre los sujetos que respondieron no a la pregunta sobre la duración*, a un sujeto que escogió el lente tratado.

---

[1] Note que si hubiéramos usado los valores .17 y .20 de la tabla de probabilidad habríamos obtenido un resultado diferente. Esto es porque estos valores están redondeados, lo que da pie a ciertos errores en el cálculo.

(f) No. Siguiendo 3.5 de la página 56 y considerando que $P(T)P(N) = \left(\frac{33}{59}\right)\left(\frac{12}{59}\right) = .11$ no es igual a $P(TN) = \frac{10}{59} = .17$

3.17 Por la ecuación 3.11

$$RR = \frac{P(U|N)}{P(U|Y)} = \frac{\frac{2}{12}}{\frac{24}{47}} = .33.$$

Esto indica que las personas que no pueden usar lentes todo el tiempo que desean corren menos riesgo de seleccionar el lente no tratado que las personas que pueden usar lentes durante el tiempo que les place. Esto puede indicar que las personas que no pueden usar lentes por el tiempo que desean podrían beneficiarse más con los lentes tratados que las personas que pueden usar su lente por el tiempo que quieren.

3.19 Las posibilidades de seleccionar un lente no tratado para las personas que respondieron sí a la pregunta sobre la duración es, de acuerdo con la página 60.

$$\frac{P(U|Y)}{P(T|Y)} = \frac{\frac{24}{47}}{\frac{23}{47}} = 1.04.$$

Para las personas que respondieron no a la pregunta sobre la duración las posibilidades serían

$$\frac{P(U|N)}{P(T|N)} = \frac{\frac{2}{12}}{\frac{10}{12}} = .20.$$

E. Las siguientes respuestas se refieren al estudio de caso E.

3.21 Creamos una tabla como la siguiente.

	D	$\overline{D}$	
+	.100	.042	.142
−	.006	.852	.858
	.106	.894	

Entonces, por la ecuación 3.6 de la página 58

$$= P(+|D) = \frac{.100}{.106} = .943.$$

Por la ecuación 3.7 de la página 58

$$= P(-|\overline{D}) = \frac{.852}{.894} = .953$$

Por la ecuación 3.8 de la página 58

$$PPV = P(D|+) = \frac{.100}{.142} = .704$$

Por la ecuación 3.9 de la página 58

$$NPV = P(\overline{D}|-) = \frac{.852}{.858} = .993$$

P. Las siguientes respuestas se refieren al estudio de caso de estudio P.

3.22 (a) La probabilidad de seleccionar a un paciente que sea mujer, tenga diabetes tipo I y esté en el rango de edad 25-44.

(b) Esta probabilidad conjunta se puede encontrar al dividir el número de pacientes que tienen las tres características entre el número total de pacientes o $\frac{157}{1597} = .098$.

(c) Esto incluiría a todos los pacientes que sean mujeres *o* que tengan diabetes tipo I *o* que estén en el rango de edad 24-44. En resumen, es decir, todos excepto los varones que tengan diabetes tipo II y que estén en un rango de edad diferente al de 24-44. La revisión de la tabla J.10 muestra que hay 163 + 195 de estos pacientes, por lo que la probabilidad es $1 - \frac{163+195}{1597} = .776$.

(d) Ésta es la probabilidad de seleccionar a un varón de entre los pacientes que estén en el rango de edad 15-24 *o* tengan diabetes tipo II.

(e) Hay un total de 380 + 410 pacientes que tienen diabetes tipo I y otros 70 + 65 pacientes en el rango de edad 15-24 para un total de 925. Como hay 380 + 70 = 450 varones en este grupo, la probabilidad de seleccionar a un varón de entre el grupo designado sería $\frac{450}{925} = .486$.

Capítulo cuatro

4.1 Por la ecuación 4.2 de la página 77

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{2500}{100} = 25$$

4.3 Disminuye.

4.5 Por la ecuación 4.3 de la página 78

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{45 - 50}{\frac{10}{\sqrt{25}}} = -2.5.$$

La referencia a la columna tres del Apéndice A muestra que la probabilidad deseada es .0062.

4.7 Usando el límite superior real de .35 para una proporción de .30 en la ecuación 4.6 de la página 85 da

$$Z = \frac{.35 - .40}{\sqrt{\frac{.40(1-.40)}{10}}} = -.32.$$

La referencia a la columna tres del Apéndice A proporciona un cálculo de probabilidad de .3745; muy cercano al valor exacto, lo que es algo sorprendente dado el pequeño tamaño de la muestra.

4.9 Como no se da σ, será apropiado realizar una prueba *t* de una media para la prueba de hipótesis.

Las cantidades necesarias para el cálculo de s, $\bar{x}$ y t son

	X	X^2
	3	9
	3	9
	2	4
	1	1
	0	0
	6	36
	5	25
	4	16
Σ	24	100

Entonces, por medio de la ecuación 2.16 de la página 37, la ecuación 2.1 de la página 25 y la ecuación 4.8 de la página 102

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{100 - \frac{(24)^2}{8}}{8-1}} = 2.0$$

$$\bar{x} = \frac{\sum x}{n} = \frac{24}{8} = 3.0$$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{3.0 - 5.0}{\frac{2.0}{\sqrt{8.0}}} = -2.828.$$

La referencia al Apéndice B muestra que el valor crítico para esta prueba t de una cola con $8 - 1 = 7$ grados de libertad efectuada con .05 es -1.895. Como t crítico de -2.828 es menor que este valor, se rechaza la hipótesis nula.

4.11 La forma de las hipótesis enunciadas indica que ésta es una prueba de equivalencia de dos colas. Se usan pruebas Z en vez de t porque conocemos σ.

Calculamos Z para la prueba uno

$$Z_1 = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{.52 - 1.0}{\frac{24}{\sqrt{144}}} = -.24$$

y para la prueba dos

$$Z_2 = \frac{.52 - (-1.0)}{\frac{24}{\sqrt{144}}} = .76.$$

Los valores críticos de una cola para las dos pruebas son -1.65 y 1.65, respectivamente, por lo que no se rechaza la hipótesis nula de equivalencia.

4.13 De lo mencionado sobre las pruebas de dos colas exactas para π al principio de la página 112, vemos que el valor crítico de $\hat{p}$ para esta prueba es .30 para la región crítica de la cola inferior y no existe un valor crítico en la región superior. Esto se puede ver si consideramos que $P(0) + P(1) + P(2) + P(3) = .000006 + .000138 + .001447 + .009002 = .010593$, que es menor que $\alpha/2 = 0.025$, en tanto que $P(0) + P(1) + P(2) + P(3) + P(4) = .000006 + .000138 + .001447 + .009002 + .036757 = .047350$, que es mayor que $\alpha/2 = .025$. Para la cola superior, $P(10) = .028248$, que es mayor que $\alpha/2$, por lo que ningún valor de $\hat{p}$ satisface el criterio de tener una probabilidad asociada menor que $\alpha/2$.

Claramente, $\hat{p}$ obtenido de .10 está en la región crítica de la cola inferior, por lo que se rechaza la hipótesis nula. El valor-p de dos colas es $2(.000006 + .000138) =$

.000288, que es menor que α = .05, por lo que también se indica el rechazo por este método. Los cálculos de las probabilidades para los valores individuales de $\hat{p}$ se dan por la ecuación 4.5 de la página 81.

$$P(0) = \tfrac{10!}{0!(10-0)!}.7^0(1-.7)^{10-0} = .000006$$
$$P(1) = \tfrac{10!}{1!(10-1)!}.7^1(1-.7)^{10-1} = .000138$$
$$P(2) = \tfrac{10!}{2!(10-2)!}.7^2(1-.7)^{10-2} = .001447$$
$$P(3) = \tfrac{10!}{3!(10-3)!}.7^3(1-.7)^{10-3} = .009002$$
$$P(4) = \tfrac{10!}{4!(10-4)!}.7^4(1-.7)^{10-4} = .036757$$
$$P(9) = \tfrac{10!}{9!(10-9)!}.7^9(1-.7)^{10-9} = .121061$$
$$P(10) = \tfrac{10!}{10!(10-10)!}.7^{10}(1-.7)^{10-10} = .028248$$

4.15 De lo estudiado sobre las pruebas de una cola exactas para π, al comienzo de la página 108, vemos que el valor-*p* para esta prueba es $P(0) + P(1)$ donde, por medio de la ecuación 4.5 de la página 81,

$$P(0) = \frac{9!}{0!(9-0)!}.5^0(1-.5)^{9-0} = .001953$$
$$P(1) = \frac{9!}{1!(9-1)!}.5^1(1-.5)^{9-1} = .017578$$

por lo que el valor *p* de una cola es .001953 + .017578 = .019531. Como este valor es menor que $\alpha = 0.05$, se rechaza la hipótesis nula.

Ya que $P(0) + P(1) = 0.001953 + 0.017578 = 0.019531$ que es menor que α = .05 y $P(0) + P(1) + P(2)$ = .019531 + .070313 = 0.089844 es mayor que α = .05, $\hat{p}$ crítica es 1/9 = .11. Como $\hat{p}$ obtenida de .11 cae en esta región crítica, se rechaza la hipótesis nula.

4.17 El nivel de significancia de la prueba (α).
El tamaño de la muestra (*n*).
La forma de la distribución alternativa.

4.19 De nuevo, si aplicamos la ecuación 4.10 con el error estándar de la media reflejando *n* = 100 nos da

$$Z_\beta = \frac{40-42}{2} + 1.65 = .65$$

que tiene una potencia asociada de .2578. Beta es, entonces, 1.0 − .2578 = .7422.

4.21 Por la ecuación 4.11 de la página 136 con $\alpha/2$ sustituido por alfa para reflejar la prueba de dos colas, y la discusión respectiva, calculamos

$$n = \frac{\sigma^2(Z_\beta - Z_{\alpha/2})^2}{(\mu_0 - \mu)^2} = \frac{4(1.28-(-1.96))^2}{(10-6)^2} = 2.6244$$

que redondeamos a 3.[2] El pequeño tamaño de la muestra refleja el hecho de que la media de la alternativa está

$$\frac{6-10}{\frac{2}{\sqrt{2.6244}}} = -3.24$$

errores estándar por debajo de la media de la distribución nula.

[2] Es útil bosquejar el problema como lo hemos hecho en la figura 4.26 de la página 137.

4.23 Considerando que $\hat{p}$ y aplicando las ecuaciones 4.16 y 4.17 de la página 148 nos da

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .75 - 1.96\sqrt{\frac{(.75)(.25)}{200}} = .690$$

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .75 + 1.96\sqrt{\frac{(.75)(.25)}{200}} = .810.$$

4.25 De la discusión al principio de la página 149 y las ecuaciones 4.20 y 4.21 de la página 150, calculamos

$$df_{LN} = 2\,(n - S + 1) = 2\,(10 - 6 + 1) = 10$$
$$df_{LD} = 2S = 2 \times 6 = 12$$

El Apéndice C muestra que para los grados de libertad del numerador y el denominador de 10 y 12 respectivamente, el valor F apropiado para un intervalo de confianza unilateral del 95% es 2.75. Si usamos este valor en la ecuación 4.18 de la página 149 nos da

$$L = \frac{S}{S + (n - S + 1)F_L} = \frac{6}{6 + (10 - 6 + 1)2.75} = .304.$$

A. Las siguientes respuestas se refieren al estudio de caso A.

4.27 Como se mencionó en la página 139, cuando se usa un estadístico de muestra como un estimado de un parámetro, éste se define como "estimación puntual".
En el artículo nos dan

$$\bar{x} = 31.5$$
$$s = 10.2$$
$$n = 59$$

Entonces, por medio de las ecuaciones 4.14 y 4.15 de la página 146

$$L = \bar{x} - t\frac{s}{\sqrt{n}} = 31.5 - 2.002\frac{10.2}{\sqrt{59}} = 28.8$$

y

$$U = \bar{x} + t\frac{s}{\sqrt{n}} = 31.5 + 2.002\frac{10.2}{\sqrt{59}} = 34.2.$$

Así, podemos tener una confianza del 95% de que la edad media de la población es de entre 28.8 y 34.2.

B. Las siguientes respuestas se refieren al estudio de caso B.

4.29 En este caso utilizaremos un método aproximado para formar los intervalos de confianza requeridos. Para este fin, dejemos que los subíndices S, W y P indiquen, respectivamente, el grupo estándar, el débil y el placebo. Entonces, por las ecuaciones 4.16 y 4.17 que enuncian

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

y considerando que

$$\hat{p}_S = \frac{35}{62} = .565$$
$$\hat{p}_W = \frac{12}{59} = .203$$
$$\hat{p}_P = \frac{30}{62} = .484$$

por lo que

$$L_S = .565 - 1.96\sqrt{\frac{(.565)(.435)}{62}} = .442$$
$$U_S = .565 - 1 + 96\sqrt{\frac{(.565)(.435)}{62}} = .688$$

y

$$L_W = .203 - 1.96\sqrt{\frac{(.203)(.797)}{59}} = .100$$
$$U_W = .203 + 1.96\sqrt{\frac{(.203)(.797)}{59}} = .306$$

y

$$L_P = .484 - 1.96\sqrt{\frac{(.484)(.516)}{62}} = .360$$
$$U_P = .484 + 1.96\sqrt{\frac{(.484)(.516)}{62}} = .608$$

Es probable que algunos sujetos hayan notado que sus brazaletes atraían objetos metálicos en tanto que otros notaron que no era el caso.

D. Las siguientes respuestas se refieren al estudio de caso D.

4.31 No. Para realizar la prueba *Z* tendríamos que conocer σ y no es así. Podríamos, claro está, probar la hipótesis con una prueba *t* de una media.

O. Las siguientes respuestas se refieren al estudio de caso O.

4.33 No si se sabe que la prueba es robusta ante violaciones de la suposición de normalidad en la circunstancia para la cual se está empleando.

Capítulo cinco

5.1 Las puntuaciones diferenciales, sus cuadrados y las sumas de cada uno son de la siguiente forma.

	d	d^2
	–14	196
	–11	121
	2	4
	–2	4
	–19	361
	8	64
	0	0
	–18	324
Σ	–54	1074

Usando estas sumas, la media y la desviación estándar de las puntuaciones diferenciales son

$$\bar{d} = \frac{\sum d}{n} = \frac{-54}{8} = -6.75$$

y

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{1074 - \frac{(-54)^2}{8}}{8-1}} = 10.068.$$

(a) Por la ecuación 5.1 de la página 162

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} = \frac{-6.75}{\frac{10.068}{\sqrt{8}}} = -1.896.$$

t crítica para una prueba de dos colas con $n - 1 = 8 - 1 = 7$ grados de libertad efectuada con $\alpha = .05$ es, por el Apéndice B, $\pm$ 2.365, por lo que no se rechaza la hipótesis nula.

(b) Por las ecuaciones 5.2 y 5.3 de la página 172

$$L = \bar{d} - t\frac{s_d}{\sqrt{n}} = -6.75 - (2.365)\left(\frac{10.068}{\sqrt{8}}\right) = -15.168$$

y

$$U = \bar{d} + t\frac{s_d}{\sqrt{n}} = -6.75 + (2.365)\left(\frac{10.068}{\sqrt{8}}\right) = 1.668.$$

(c) La prueba t no significativa, o de manera equivalente, el hecho de que el IC contuviera a cero indica que no tenemos suficiente evidencia para afirmar un efecto cambiante en el colesterol para la dieta.

5.3 (a) i. Utilizando lo tratado en la página 179 y descartando (S-S) e (I-I) por ser no informativos, calculamos

$$\hat{p} = \frac{5}{5+2} = .7143.$$

Las probabilidades binomiales para $\pi = .5$ y $n = 7$ son, por la ecuación 4.5 de la página 81, como sigue.

Proporción $\hat{p}$	Número de eventos y	Probabilidad $P(y)$
.0000	0	.00781
.1429	1	.05469
.2857	2	.16406
.4286	3	.27344
.5714	4	.27344
.7143	5	.16406
.8571	6	.05469
1.0000	7	.00781

Si empleamos el método descrito al comienzo de la página 108 para encontrar el valor-p para una prueba de dos colas exacta, obtenemos $p = 2(.00781 + .05469 + .16406) = .45312$. La hipótesis nula no se rechaza, por lo que no podemos afirmar que el tratamiento es efectivo para mejorar la percepción que tiene el paciente de su apariencia.

ii. Usando lo estudiado en la página 149 y considerando que $S = 5$ y $n = 7$, encontramos, por las ecuaciones 4.20 y 4.21 que los grados de libertad del numerador y el denominador para F_L son, respectivamente,

$$df_{LN} = 2(n - S + 1) = 2\ (7 - 5 + 1) = 6$$

y

$$df_{LD} = 2S = 2 \cdot 5 = 10.$$

Entonces, por la ecuación 4.18 de la página 149, el límite inferior del IC es

$$L = \frac{S}{S + (n - S + 1)F_L} = \frac{5}{5 + (7 - 5 + 1)4.07} = .291.$$

Los grados de libertad del numerador y el denominador para U son, por las ecuaciones 4.22 y 4.23

$$Df_{UN} = 2(S + 1) = 2\ (5 + 1) = 12$$

y

$$df_{UD} = 2(n - S) = 2\ (7 - 5) = 4.$$

Luego, por la ecuación 4.19

$$U = \frac{(S + 1)F_U}{n - S + (S + 1)F_U} = \frac{(5 + 1)8.75}{7 - 5 + (5 + 1)8.75} = \frac{52.5}{54.5} = .963.$$

¿Le sorprende descubrir que .5 está en el intervalo? (Con sinceridad esperamos que no).

(b) i. Descartando S-S e I-I por ser no informativos, calculamos

$$\hat{p} = \frac{71}{71 + 20} = .780.$$

Luego, por la ecuación 5.4 de la página 176

$$Z = \frac{\hat{p} - .5}{\frac{.5}{\sqrt{n}}} = \frac{.780 - .5}{\frac{.5}{\sqrt{91}}} = 5.342.$$

La referencia al Apéndice A muestra que Z crítica para una prueba de dos colas con $\alpha = .05$ es ± 1.96. La hipótesis nula $H_0 : \pi = .05$ se rechaza, por lo que podemos asegurar un efecto benéfico para el tratamiento en lo que se refiere a mejorar la percepción que tiene el paciente de su apariencia.

ii. De la discusión de la página 186 y por las ecuaciones 4.16 y 4.17

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .780 - 1.96\sqrt{\frac{(.78)(.22)}{91}} = .695$$

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}\hat{q}}{n}} = .780 + 1.96\sqrt{\frac{(.78)(.22)}{91}} = .865.$$

¿Le sorprende descubrir que .5 no está en el intervalo? (Con sinceridad esperamos que no).

5.5 (a) Por la ecuación 5.12 de la página 200

$$\widehat{OR} = \frac{b}{c} = \frac{24}{9} = 2.667$$

lo que significa que, en esta muestra, las posibilidades de ser un oficial de motocicleta para los oficiales con cáncer de piel son 2.667 veces las posibilidades para aquellos que no tienen cáncer de piel.

(b) Por el primer método, comenzamos por construir un intervalo de confianza para el cálculo de π, luego usamos la relación entre $\widehat{OR}$ y $\hat{p}$ para convertir el intervalo en un cálculo de OR. Para este fin, primero usamos el resultado de 5.5a y la ecuación 5.14 de la página 205 para obtener

$$\hat{p} = \frac{\widehat{OR}}{1 + \widehat{OR}} = \frac{2.667}{1 + 2.667} = .727.$$

Luego, por las ecuaciones 5.15 y 5.16 de la página 208, calculamos

$$L = \hat{p} - Z\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = .727 - 1.96\sqrt{\frac{.727(1 - .727)}{33}} = .575$$

y

$$U = \hat{p} + Z\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = .727 + 1.96\sqrt{\frac{.727(1 - .727)}{33}} = .879.$$

Ahora empleamos la ecuación 5.17 de la página 208 para convertir estos puntos extremos en razones de posibilidades de la siguiente manera.

$$L = \frac{\hat{p}}{1 - \hat{p}} = \frac{.575}{1 - .575} = 1.353$$

y

$$U = \frac{\hat{p}}{1 - \hat{p}} = \frac{.879}{1 - .879} = 7.264.$$

Un segundo método produce el intervalo de confianza directamente por medio de las ecuaciones 5.18 y 5.54 de la página 210 de la siguiente manera.

$$L = \exp\left(1\text{n }(OR) - Z\sqrt{\frac{1}{b} + \frac{1}{c}}\right) = \exp\left(1\text{n }(2.667) - 1.96\sqrt{\frac{1}{24} + \frac{1}{9}}\right) = 1.240$$

$$U = \exp\left(1\text{n }(OR) + Z\sqrt{\frac{1}{b} + \frac{1}{c}}\right) = \exp\left(1\text{n }(2.667) - 1.96\sqrt{\frac{1}{24} + \frac{1}{9}}\right) = 5.738$$

Como se puede ver, las dos aproximaciones producen cálculos muy similares para *L*, pero difieren considerablemente cuando se calcula *U*.

(c) De lo que tratamos en la página 210 y de las ecuaciones 5.22, 5.23, 5.24, 5.25, 5.20 y 5.21 obtenemos

$$df_{LN} = 2(c - 1) = 2\,(9 + 1) = 20$$

y

$$df_{LD} = 2b = 2 \cdot 24 = 48$$

por lo que

$$L = \frac{b}{b + (c+1)F_L} = \frac{24}{24 + (9+1)2.01} = .542.$$

Los grados de libertad para el cálculo de *U* son

$$Df_{UN} = 2(b + 1) = 2\,(24 + 1) = 50$$

y

$$df_{UD} = 2c = 2 \cdot 9 = 18$$

por lo que

$$U = \frac{(b+1)F_U}{c + (b+1)F_U} = \frac{(24+1)2.35}{9 + (24+1)2.35} = .867.$$

Convertir los puntos extremos del intervalo en estimados de *OR* por la ecuación 5.17 nos da

$$L = \frac{\hat{p}}{1 - \hat{p}} = \frac{.542}{1 - .542} = 1.183$$

y

$$U = \frac{\hat{p}}{1 - \hat{p}} = \frac{.867}{1 - .867} = 6.519.$$

Comparar el intervalo exacto con las dos aproximaciones nos lleva a concluir que si bien ambas aproximaciones fueron lo suficientemente precisas, siempre preferiríamos utilizar el intervalo exacto.

A. Las siguientes respuestas se refieren al estudio de caso A.

5.7 Como los datos son apareados y dicotómicos, probablemente hayan empleado la prueba de McNemar.[3]

Si éste no fuera un ejercicio de un libro de texto, usaríamos un método exacto para la prueba, pero, debido a los cálculos implicados, usaremos un método aproximado como el que nos da la ecuación 5.4. Primero, observamos que $\hat{p} = \frac{8}{16} = .50$, por lo que

$$Z = \frac{\hat{p} - .5}{\frac{.5}{\sqrt{n}}} = \frac{.5 - .5}{\frac{.5}{\sqrt{16}}} = 0.$$

Como los valores críticos para una prueba de dos colas efectuada con $\alpha = .05$ son ±1.96, no rechazamos la hipótesis nula, lo que concuerda con el resultado reportado por los autores. Así que, no se pudo demostrar ningún cambio.

[3] Ésta es, de hecho, la prueba que se está usando.

K. Las siguientes respuestas se refieren al estudio de caso K.

5.9 El supuesto de independencia podría haber sido infringido si se hubieran usado mediciones de presión arterial individuales para el análisis. Los alumnos de cierta escuela pueden parecerse entre sí más que los alumnos de otras escuelas debido a factores socioeconómicos u otros, así como los hermanos de una familia en particular pueden parecerse entre sí más que a los niños de una familia diferente.

5.11 Un intervalo de confianza hubiera sido más informativo porque hubiera dado un cálculo de *cuánto* cambió la media en vez de simplemente afirmar que ocurrió dicho cambio.

Al usar los valores para $\bar{d}$ y s_d calculados en 5.10 con las ecuaciones 5.2 y 5.3 nos da

$$L = \bar{d} - t\frac{s_d}{\sqrt{n}} = 2.052 - 4.604\frac{.537}{\sqrt{5}} = .946$$

y

$$U = \bar{d} + t\frac{s_d}{\sqrt{n}} = 2.052 + 4.604\frac{.537}{\sqrt{5}} = 3.158.$$

Así, podemos tener un 99% de confianza de que el cambio en la media de la escuela fue de entre .946 y 3.158 puntos.

M. Las siguientes respuestas se refieren al estudio de caso M.

5.13 Si seguimos los mismos pasos dados en 5.12, obtenemos

$$\bar{d} = \frac{\sum d}{n} = \frac{2.4}{7} = .343$$

y

$$s_d = \sqrt{\frac{\sum d^2 - \frac{(\sum d)^2}{n}}{n-1}} = \sqrt{\frac{1.1 - \frac{(2.4)^2}{7}}{7-1}} = .215$$

Luego, por las ecuaciones 5.2 y 5.3

$$L = \bar{d} - t\frac{s_d}{\sqrt{n}} = .343 - 2.447\frac{.215}{\sqrt{7}} = .144$$

y

$$U = \bar{d} + t\frac{s_d}{\sqrt{n}} = .343 + 2.447\frac{.215}{\sqrt{7}} = .542.$$

Capítulo seis

6.1 Los datos se ordenan para análisis usando una prueba *t* de muestras independientes, como en la tabla K.1. Por la ecuación 6.2 de la página 220

$$\begin{aligned} s_P^2 &= \frac{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}\right) + \left(\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}\right)}{n_1 + n_2 - 2} \\ &= \frac{\left(5913.29 - \frac{(239.3)^2}{10}\right) + \left(5957.57 - \frac{(240.7)^2}{10}\right)}{10 + 10 - 2} \\ &= \frac{186.841 + 163.921}{18} \\ &= 19.487. \end{aligned}$$

TABLA K.1: Datos ordenados para análisis para el ejercicio 1.

	Grupo 1		Grupo 2	
	X_1	X_1^2	X_2	X_2^2
	26.2	686.44	25.9	670.81
	24.5	600.25	20.1	404.01
	20.0	400.00	22.2	492.84
	30.2	912.04	29.7	882.09
	28.4	806.56	28.0	784.00
	18.6	345.96	29.4	864.36
	21.5	462.25	20.2	408.04
	21.7	470.89	20.7	428.49
	29.9	894.01	26.3	691.69
	18.3	334.89	18.2	331.24
$\sum$	239.9	5913.29	240.7	5957.57

Para continuar,

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{239.3}{10} = 23.93$$

y

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{240.7}{10} = 24.07.$$

Luego, por la ecuación 6.1 de la página 219

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{23.93 - 24.07}{\sqrt{19.487\left(\frac{1}{10} + \frac{1}{10}\right)}} = -0.071.$$

El Apéndice B indica que el valor crítico para una prueba t de una cola con 18 grados de libertad efectuada con $\alpha = .05$ es 1.734. La hipótesis nula no se rechaza. Por consiguiente, no encontramos suficiente evidencia para concluir que los niños que viven una milla a la redonda de un restaurante de comida rápida tienen una media de IMC más elevada que los niños que viven más alejados.

6.3 Utilizar algunos de los resultados obtenidos en el ejercicio 6.1 con las ecuaciones 6.3 y 6.4 de la página 228 da

$$L = (\bar{x}_1 - \bar{x}_2) - t\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = (23.93 - 24.07) - 2.101\sqrt{19.487\left(\frac{1}{10} + \frac{1}{10}\right)} = -4.288$$

y

$$U = (\bar{x}_1 - \bar{x}_2) + t\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = (23.93 - 24.07) + 2.101\sqrt{19.487\left(\frac{1}{10} + \frac{1}{10}\right)} = -4.008.$$

Esto nos da un 95% de confianza de que $\mu_1 - \mu_2$ está entre –4.288 y 4.008. Desde el punto de vista de los investigadores, esto indica, con 95% de confianza, que la diferencia entre los pesos promedio de los dos grupos es como se especificó por el *IC*.

Observe que cero está en el intervalo. ¿Qué significa en lo que concierne a una prueba de hipótesis?

6.5 Por la ecuación 6.5 de la página 232

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{.312 - .288}{\sqrt{\frac{(.312)(.688)}{1777} + \frac{(.288)(.712)}{1821}}} = 1.57$$

El Apéndice A muestra que los valores críticos para una prueba Z de dos colas realizada con $\alpha = .05$ son ± 1.96. Entonces, la prueba no es significativa por lo que no pudimos demostrar que existe una diferencia en las proporciones de fumadores masculinos y femeninos.

6.7 Por las ecuaciones 6.6 y 6.7 de la página 236,

$$L = (\hat{p}_1 - \hat{p}_2) - \left(Z\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1 - 1} + \frac{\hat{p}_2\hat{q}2}{n_2 - 1}} + \frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$= (.312 - .288) - \left(1.96\sqrt{\frac{(.312)(.688)}{1777 - 1} + \frac{(.288)(.712)}{1821 - 1}} + \frac{1}{2}\left(\frac{1}{1777} + \frac{1}{1821}\right)\right)$$

$$= -.007$$

y

$$U = (.312 - .288) + \left(1.96\sqrt{\frac{(.312)(.688)}{1777 - 1} + \frac{(.288)(.712)}{1821 - 1}} + \frac{1}{2}\left(\frac{1}{1777} + \frac{1}{1821}\right)\right)$$

$$= .055$$

Una prueba de dos colas de la hipótesis efectuada con $\alpha = .05$ no sería significativa ya que cero se encuentra en el intervalo. El intervalo calcula $\pi_1 - \pi_2$ o, en el contexto del estudio, la diferencia entre las proporciones de fumadores masculinos y femeninos.

6.9 (a) Con fines de análisis, será conveniente ordenar los datos como sigue.

		Cáncer	
		Sí	No
En 500 yardas	Sí	590	9258
	No	577	12535

Por la ecuación 6.9 de la página 239

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)} = \frac{590/(590+9258)}{577/(577+12535)} = 1.361.$$

Realizamos la prueba uno[4] mediante la ecuación 6.10 de la página 240 de la siguiente manera.

$$Z = \frac{\ln(\widehat{RR}) - \ln(RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} = \frac{\ln(1.361) - \ln(1.100)}{\sqrt{\frac{9258/590}{590+9258} + \frac{12535/577}{577+12535}}} = 3.735$$

[4] De hecho, podemos ver que la prueba será no significativa porque el estadístico de prueba no está en el *EI*.

Del Apéndice A encontramos que el valor crítico para una prueba Z de una cola con $\alpha = .05$ es -1.65. Como la Z obtenida de 3.735 es mayor que este valor, no se rechaza la hipótesis nula.

La prueba dos es, entonces,

$$Z = \frac{\ln(RR) - \ln(RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} = \frac{\ln(1.361) - \ln(.910)}{\sqrt{\frac{9258/590}{590+9258} + \frac{12535/577}{577+12535}}} = 7.061.$$

Como 7.061 es mayor que Z crítica de 1.65, la prueba dos es significativa.

Como *ambas* pruebas (uno y dos) deben ser significativas para que la hipótesis nula de equivalencia de dos colas sea rechazada, la prueba de equivalencia no es significativa. Así, no se estableció la seguridad. El punto es que los estudios anteriores no demostraron que los cables de alta tensión eran peligrosos y el presente estudio no pudo demostrar que eran seguros.

(b) $H_{0E}: RR \leq .91$ o $RR \geq 1.10$

$H_{AE}: .91 < RR < 1.1$

(c) $H_0: RR = 1$

¿Cuál sería la conclusión si se probara esta hipótesis en este estudio?

(d) Si $RR = 1$, suponemos que no existe relación alguna entre los cables de alta tensión y el cáncer. Así, una prueba de la hipótesis nula $RR = 1$ supone seguridad e intenta rechazar la hipótesis de seguridad a favor de una alternativa que niega la seguridad.

La hipótesis nula establecida en el ejercicio 6.9b supone peligro ($RR \geq 1.1$ o $RR \leq .91$) e intenta rechazar la hipótesis nula de peligro a favor de una hipótesis alternativa de seguridad.

6.11 (a) Si se usa la tabla construida para la respuesta al ejercicio 6.8a, con la ecuación 6.14 de la página 249, tenemos que

$$\widehat{OR} = \frac{ad}{bc} = \frac{(196)(237)}{(95)(191)} = 2.560$$

(b) Por la ecuación 6.15 de la página 249

$$Z = \frac{\ln(OR) - \ln(OR_0)}{\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}} = \frac{\ln(2.560)}{\sqrt{\frac{1}{196} + \frac{1}{95} + \frac{1}{191} + \frac{1}{237}}} = 5.935.$$

Como la Z crítica es ± 1.96 se rechaza la hipótesis nula, lo que nos lleva a la conclusión de que la probabilidad de que los pacientes con lesiones desarrollen EM es mayor que la probabilidad de los pacientes sin dichas lesiones.

6.13 Por las ecuaciones 6.16 y 6.17 de la página 254

$$\begin{aligned} L &= \exp\left[\ln(OR) - Z\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right] \\ &= \exp\left[\ln(2.560) - 1.96\sqrt{\frac{1}{196} + \frac{1}{95} + \frac{1}{191} + \frac{1}{237}}\right] \\ &= 1.877 \end{aligned}$$

y

$$
\begin{aligned}
U &= \exp\left[1\text{n } (\widehat{OR}) + Z\sqrt{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}+\frac{1}{d}}\right] \\
&= \exp\left[1\text{n } (2.560) + 1.96\sqrt{\frac{1}{196}+\frac{1}{95}+\frac{1}{191}+\frac{1}{237}}\right] \\
&= 3.492.
\end{aligned}
$$

Así, podemos estar un 95% confiados de que la posibilidad de que los pacientes con lesiones desarrollen EM es entre 1.877 y 3.492 veces la posibilidad de los pacientes sin estas lesiones.

Este intervalo indica que una prueba de dos colas de $H_0 : OR = 1$ sería rechazada porque 1 no se localiza en el intervalo que es el resultado obtenido en el ejercicio 6.11b.

A. Las siguientes respuestas se refieren al estudio de caso A.

6.15 La hipótesis nula se rechaza por el hecho de que cero no está en el intervalo de confianza.

B. Las siguientes respuestas se refieren al estudio de caso caso B.

6.17 No, porque no se había administrado tratamiento alguno en ese momento y los sujetos habían sido asignados a los grupos al azar. Suponiendo que la asignación aleatoria se haya realizado conducido apropiadamente, un hallazgo significativo probablemente se debería a un error del tipo I.

6.19 Usando los métodos empleados para la pregunta 6.18 obtenemos

$$SS_S = (65 - 1)(3.9^2) = 973.44$$

y

$$SS_P = (64 - 1)(3.2^2) = 645.12$$

Entonces, obtenemos

$$s_P^2 = \frac{973.44 + 645.12}{65 + 64 - 2} = 12.745.$$

Los límites son, entonces,

$$L = (7.8 - 9.3) - 1.979\sqrt{12.745\left(\frac{1}{65}+\frac{1}{64}\right)} = 2.744$$

y

$$U = (7.8 - 9.3) + 1.979\sqrt{12.745\left(\frac{1}{65}+\frac{1}{64}\right)} = .256$$

Este intervalo calcula la magnitud de la diferencia entre los efectos de los brazaletes placebo magnetizados como lo expresa la diferencia entre la media de las puntuaciones de dolor producida por ambos tratamientos. Como cero no está en el intervalo, podemos afirmar un nivel de dolor reducido como resultado del tratamiento.

6.21 (a) Tomando en cuenta que $\hat{p}_s = \frac{35}{62} = .565$ y $\hat{p}_w = \frac{12}{59} = .203$ y aplicando la ecuación 6.5 de la página 232 tenemos que

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{.565 - .203}{\sqrt{\frac{(.565)(.435)}{62} + \frac{(.203)(.797)}{59}}} = 4.420.$$

Como este valor excede 1.96 rechazamos la hipótesis nula y concluimos que la proporción de sujetos que usan el brazalete estándar que pudieron identificar correctamente el tipo de brazalete que traían puesto es mayor que la proporción de aquellos que usaron el brazalete débil.

(b) $\hat{p}_P = \frac{30}{62} = .484$

$$Z = \frac{.565 - .484}{\sqrt{\frac{(.565)(.435)}{62} + \frac{(.484)(.516)}{62}}} = .906$$

La hipótesis nula no se rechaza, por lo que no pudimos demostrar que existe una diferencia entre las dos proporciones.

(c)

$$Z = \frac{.203 - .484}{\sqrt{\frac{(.203)(.797)}{59} + \frac{(.484)(.516)}{62}}} = -3.415$$

Se rechaza la hipótesis nula.

6.23 Primero notamos que las proporciones de sujetos en los grupos débil y placebo que indican que están usando un brazalete placebo son $\frac{12}{59} = .203$ y $\frac{30}{62} = .484$, respectivamente. Después notamos que la diferencia entre las dos proporciones es $.203 - .484 = -.281$, que es menor que EI_L de $-.02$, por lo que la prueba dos no será significativa. Podemos, por lo tanto, declarar que la prueba dos no es significativa sin recurrir al cálculo del estadístico de prueba, por lo que la hipótesis nula de equivalencia no se rechaza. No podemos declarar que los dos grupos son equivalentes.

C. Las siguientes respuestas se refieren al estudio de caso C.

6.25 Intervalos de confianza, porque son más informativos que las pruebas de hipótesis.

6.27 Los intervalos de confianza para *RR* pueden formarse por medio de las ecuaciones 6.11 y 6.12 como sigue.

$$L = \exp\left[1\text{n } (RR) - Z\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}\right]$$

y

$$U = \exp\left[1\text{n } (RR) + Z\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}\right]$$

Si aplicamos estas ecuaciones a los datos de los cuatro periodos de tiempo produce lo siguiente.

$$L = \exp\left[1\text{n } (3.470) - 1.96\sqrt{\frac{441/21}{21+441} + \frac{452/6}{6+452}}\right] = 1.414$$

alta

$$U = \exp\left[1\text{n } (3.470) + 1.96\sqrt{\frac{441/21}{21+441} + \frac{452/6}{6+452}}\right] = 8.518$$

$$L = \exp\left[\ln\ (2.533) - 1.96\sqrt{\frac{439/23}{23+439} + \frac{449/9}{9+449}} \right] = 1.185$$

un mes

$$U = \exp\left[\ln\ (2.533) + 1.96\sqrt{\frac{439/23}{23+439} + \frac{449/9}{9+449}} \right] = 5.415$$

$$L = \exp\left[\ln\ (1.597) - 1.96\sqrt{\frac{404/58}{58+404} + \frac{422/36}{36+422}} \right] = 1.076$$

un año

$$U = \exp\left[\ln\ (1.597) + 1.96\sqrt{\frac{404/58}{58+404} + \frac{422/36}{36+422}} \right] = 2.371$$

$$L = \exp\left[\ln\ (1.344) - 1.96\sqrt{\frac{382/80}{80+382} + \frac{399/59}{59+399}} \right] = .985$$

23 meses

$$U = \exp\left[\ln\ (1.344) + 1.96\sqrt{\frac{382/80}{80+382} + \frac{399/59}{59+399}} \right] = 1.833$$

D. Las siguientes respuestas se refieren al caso de estudio D.

6.29 (a) La prueba t de muestras independientes. Se usó para comparar dos grupos independientes (sujetos infectados y no infectados con VIH). Además, como había 15 personas en un grupo y 5 en el otro, los grados de libertad para la prueba de muestras independientes sería $15 + 5 - 2 = 18$, que son los grados de libertad reportados para la prueba.

(b) Primero recordemos que un t obtenida de 2.26 debe tener un valor p más pequeño que t de 1.79. Aun así, los autores reportan que el primero es significante en el nivel .05 y el segundo en el nivel .01. Segundo, t crítico para pruebas de una cola y de dos colas con 18 grados de libertad realizadas con $\alpha = .01$ son 2.552 y 2.878 respectivamente, por lo que t obtenida de 1.79 no pudo producir un resultado significante en el nivel .01, como lo reportaron los autores.

(c) Si calculamos s_P^2 para los datos NPZ-8 mediante la ecuación 6.2 tenemos que

$$\begin{aligned} s_P^2 &= \frac{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n_1}\right) + \left(\sum x_2^2 - \frac{(\sum x_2)^2}{n_2}\right)}{n_1 + n_2 - 2} \\ &= \frac{\left(374 - \frac{(54)^2}{15}\right) + \left(1 - \frac{(1)^2}{5}\right)}{15 + 5 - 2} \\ &= 10.022 \end{aligned}$$

entonces

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_P^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{3.6 - .2}{\sqrt{10.022\left(\frac{1}{15} + \frac{1}{5}\right)}} = 2.080$$

Para los datos PBV,

$$s_P^2 = \frac{\left(9.94394 - \frac{(12.169)^2}{15}\right) + \left(3.81950 - \frac{(4.368)^2}{5}\right)}{15 + 5 - 2} = .00418.$$

Entonces

$$t = \frac{.8113 - .8736}{\sqrt{.00418\left(\frac{1}{15} + \frac{1}{5}\right)}} = -1.866$$

t crítica para una prueba de dos colas al nivel .05 es ±2.101 y al nivel .01 es ±2.878. Así, los resultados para ambas pruebas no son significativas. También apuntamos que nuestros valores t obtenidos difieren de aquellos reportados por los autores.

(d) De una inspección visual, parece que la varianza de las puntuaciones NPZ-8 en el grupo VIH es mucho mayor que en el grupo de control. Esto se confirma por los cálculos que demuestran que el primero es aproximadamente 12.829 y el segundo .2. Parece que se viola el supuesto de homogeneidad de varianza. Además, como los tamaños de las pruebas son relativamente pequeños y en definitiva desiguales, no podemos esperar que las propiedades de la prueba t de muestras independientes provean un resultado válido.

E. Las siguientes respuestas se refieren al estudio de caso E.

6.31 Comenzamos por ordenar los datos como sigue

		Enfermedad	
		sí	no
Prueba	+	50	21
	−	3	426

Entonces, por la ecuación 6.14

$$\widehat{OR} = \frac{ad}{bc} = \frac{(50)(426)}{(21)(3)} = 338.095.$$

Por las ecuaciones 6.16 y 6.17

$$\begin{aligned}
L &= \exp\left[1\text{n }(\widehat{OR}) - Z\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right] \\
&= \exp\left[1\text{n }(338.095) - 1.96\sqrt{\frac{1}{50} + \frac{1}{21} + \frac{1}{3} + \frac{1}{426}}\right] \\
&= 97.379 \\
U &= \exp\left[1\text{n }(\widehat{OR}) + Z\sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}\right] \\
&= \exp\left[1\text{n }(338.095) + 1.96\sqrt{\frac{1}{50} + \frac{1}{21} + \frac{1}{3} + \frac{1}{426}}\right] \\
&= 1173.848
\end{aligned}$$

H. Las siguientes respuestas se refieren al estudio de caso H.

6.33 Los autores no tuvieron bases para la conclusión a la que llegaron. Esencialmente, utilizaron el hecho de no rechazar la hipótesis nula como evidencia de que la hipótesis nula es cierta. Su error estuvo compuesto por el hecho de que estaban tratando con una muestra extremadamente pequeña, lo que originó preocupaciones sobre si la prueba tenía suficiente poder para detectar únicamente las diferencias más grandes.

6.35 No, no hay nada en el artículo que lo lleve a uno a creer que se llevó a cabo cualquier tipo de pruebas de equivalencia. De nuevo los autores parecen haber usado simplemente una prueba de significancia convencional y no sabían cómo interpretar un resultado no significativo correctamente.

I. Las siguientes respuestas se refieren al estudio de caso I.

6.37 La siguiente afirmación de los autores implica una prueba de equivalencia de una cola. "Siendo más precisos, los investigadores considerarían D-B por lo menos tan bueno como AI si la proporción de decesos asociados con el tratamiento D-B excediera la proporción del tratamiento AI por menos de .004".

6.39 En referencia a la prueba uno, como .0798 −.0753 = .0045 es mayor que .004, el valor Z será positivo y, por ende, no puede ser menor que o igual al valor crítico de −1.65, por lo que la hipótesis nula de equivalencia no será rechazada.

6.41 No, por la misma razón mencionada en la pregunta 6.39.

J. Las siguientes respuestas se refieren al estudio de caso J.

6.43 Tomando $\hat{p}_1 = \frac{138}{161} = .857$ y $\hat{p}_2 = \frac{127}{151} = .841$, y realizando la prueba uno por medio de la ecuación 6.5 tenemos que

$$Z_1 = \frac{\hat{p}_1 - \hat{p}_2 - \delta_0}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} = \frac{.857 - .841 - .10}{\sqrt{\frac{(.857)(.143)}{161} + \frac{(.841)(.159)}{151}}} = -2.070$$

Entonces, la prueba dos es

$$Z_2 = \frac{.857 - .841 - (-.10)}{\sqrt{\frac{(.857)(.143)}{161} + \frac{(.841)(.159)}{151}}} = 2.859.$$

Los valores críticos para la prueba uno y la prueba dos son −1.65 y 1.65, respectivamente, por lo que se rechazan ambas hipótesis nulas; por lo tanto, la hipótesis nula de equivalencia se rechaza y la equivalencia se establece.

O. Las siguientes respuestas se refieren al estudio de caso O.

6.45 Pruebas como la prueba *t* de muestras independientes son, por lo general menos robustas ante el incumplimiento de estas suposiciones que ante el supuesto de normalidad. Es extraño que se dedicara tanto esfuerzo a un supuesto y se ignoraran otros dos tan importantes.

Capítulo siete

7.1 Para fines de cálculo, será conveniente ordenar los datos de la siguiente manera. Para simplificar la notación, denominaremos las observaciones de los tres grupos 1, 2 y 3, respectivamente.

	Oficina		Peligroso		No peligroso	
	X_1	X_1^2	X_2	X_2^2	X_3	X_3^2
	58	3364	88	7744	65	4225
	64	4096	59	3481	70	4900
	71	5041	74	5476	79	6241
	66	4356	80	6400	66	4356
	79	6241	81	6561	74	5476
	74	5476	69	4761	79	6241
	70	4900	90	8100	60	3600
Σ	482	33474	541	42523	493	35039

(a) Usando las cantidades de la tabla anterior, calculamos las sumas de los cuadrados para cada grupo como sigue.

$$SS_1 = \sum x_1^2 - \frac{(\sum x_1)^2}{n_1} = 33474 - \frac{(482)^2}{7} = 284.857$$

$$SS_2 = \sum x_2^2 - \frac{(\sum x_2)^2}{n_2} = 42523 - \frac{(541)^2}{7} = 711.429$$

$$SS_3 = \sum x_3^2 - \frac{(\sum x_3)^2}{n_3} = 35039 - \frac{(493)^2}{7} = 317.714$$

Por la ecuación 7.4 de la página 265

$$SS_w = SS_1 + SS_2 + ... + SS_k = 284.857 + 711.429 + 317.714 = 1314.000$$

Entonces, por la ecuación 7.3 de la página 265

$$MS_w = \frac{SS_w}{N-k} = \frac{1314.0}{21-3} = 73.0.$$

(b) Por la ecuación 7.8 de la página 268

$$SS_b = \frac{\left(\sum_{i=1}^{n_1} x_{i1}\right)^2}{n_1} + \frac{\left(\sum_{i=1}^{n_2} x_{i2}\right)^2}{n_2} + ... + \frac{\left(\sum_{i=1}^{n_k} x_{ik}\right)^2}{n_k} - \frac{\left(\sum_{All} x..\right)^2}{N}$$

$$= \frac{(482)^2}{7} + \frac{(541)^2}{7} + \frac{(493)^2}{7} + \frac{(1516)^2}{21}$$

$$= 281.238$$

Y por la ecuación 7.6 de la página 267

$$MS_b = \frac{SS_b}{k-1} = \frac{281.238}{3-1} = 140.619.$$

(c) Por la ecuación 7.2 de la página 264, F obtenida es

$$F = \frac{MS_b}{MS_w} = \frac{146.619}{73.0} = 1.93$$

El Apéndice C muestra que, F crítica para una prueba con 2 y 18 grados de libertad realizada con $\alpha = .05$ es 3.55. Como, F obtenida de 1.93 no excede este valor, no se rechaza la hipótesis nula. Por lo tanto, concluimos que fallamos al demostrar cualquier diferencia entre las medias de las pulsaciones de los tres grupos.

7.3 Llevaremos a cabo un análisis de chi-cuadrada de 2 por 3, ya que sólo tres grupos están implicados.

	Instalación uno	Instalación dos	Instalación tres	
Participan	[46] (48.86)	[52] (48.36)	[36] (36.77)	134
No participan	[51] (48.14)	[44] (47.64)	[37] (36.23)	132
	97	96	73	$N = 266$

Los valores esperados son, entonces,

$$f_{e11} = \frac{(134)(97)}{266} = 48.86$$

$$f_{e12} = \frac{(134)(96)}{266} = 48.86$$

$$f_{e13} = \frac{(134)(73)}{266} = 36.77$$

$$f_{e21} = \frac{(132)(97)}{266} = 48.14$$

$$f_{e22} = \frac{(132)(96)}{266} = 47.64$$

$$f_{e23} = \frac{(132)(73)}{266} = 36.23$$

La chi-cuadrada obtenida es, entonces,

$$\begin{aligned}\chi^2 &= \frac{(46.0-48.86)^2}{48.86} + \frac{(52.0-48.36)^2}{48.36} \frac{(36.0-36.77)^2}{36.77} + \frac{(51.0-48.14)^2}{48.14} \\ &\quad + \frac{(44.0-47.64)^2}{47.64} + \frac{(37.0-36.23)^2}{36.23} \\ &= .167 + .274 + .016 + .170 + .278 + .016 \\ &= .921.\end{aligned}$$

Del Apéndice D, vemos que para $k - 1 = 3 - 1 = 2$ grados de libertad, χ^2 crítica para una prueba con $\alpha = .05$ es 5.991. Como χ^2 es menor que χ^2 crítica, no se rechaza la hipótesis nula por lo que no se puede sostener una afirmación sobre un efecto diferencial en los tres métodos de instrucción.

7.5 NO, esta afirmación no es correcta. El error por familia es la probabilidad de que una o más de las hipótesis nula *verdaderas* sean *rechazadas* no la probabilidad de que una o más de las hipótesis nulas sean falsas.

B. Las siguientes respuestas se refieren al estudio de caso B.

7.7 Dada la asignación aleatoria y la falta de cualquier diferencia en el tratamiento adjudicado a los grupos, esperaríamos que un resultado significativo fuera un error tipo I. Así, la probabilidad de este evento sería α.

7.9 Usando las desviaciones estándar reportadas de la tabla 1 y el hecho de que la suma de cuadrados (SS) para cualquier grupo dado puede expresarse como $(n–1)s^2$ calculamos

$$SS_s = (65 - 1)3.9^2 = 973.44$$

$$SS_w = (64 - 1)3.2^2 = 645.12$$

$$SS_p = (64 - 1)\ 3.2^2 = 645.12$$

Entonces, por las ecuaciones 7.4[5] y 7.3

$$SS_{\text{dentro}} = SS_s + SS_w + SS_p = 973.44 + 645.12 + 645.12 = 2263.68$$

$$MS_{\text{dentro}} = \frac{SS_{\text{dentro}}}{N-k} = \frac{2263.68}{193.3} = 11.914.$$

Tomando en cuenta que la cantidad de observaciones para cualquier grupo en particular puede expresarse como $n\bar{x}$, calculamos las cantidades para los grupos estándar, débil y placebo como (65)(7.8) = 507, (64)(8.8) = 563 y (64)(9.3) = 595, respectivamente.[6] La suma de los tres grupos es, entonces, 507 + 563 + 595 = 1665. Luego, por la ecuación 7.8

$$\begin{aligned} SS_b &= \frac{(\sum_{i=1}^{n_1} x_{i1})^2}{n_1} + \frac{(\sum_{i=1}^{n_2} x_{i2})^2}{n_2} + \ldots + \frac{(\sum_{i=1}^{n_k} x_{ik})^2}{n_k} - \frac{(\sum_{All} x..)^2}{N} \\ &= \frac{(507)^2}{65} + \frac{(563)^2}{64} + \frac{(595)^2}{64} - \frac{(1665)^2}{193} \\ &= 75.021. \end{aligned}$$

Entonces, por la ecuación 7.6

$$MS_b = \frac{SS_b}{k-1} = \frac{75.021}{2} = 37.511.$$

Por la ecuación 7.2, la F obtenida es, entonces,

$$F = \frac{MS_b}{MS_{\text{internas}}} = \frac{37.511}{11.914} = 3.148$$

Los grados de libertad para F crítica son 2 y 190. Al no encontrar este valor en la tabla F, conservadoramente usamos 2 y 150 grados de libertad, lo que da una F crítica de 3.06. (¿Por qué calificamos este valor como conservador?) La hipótesis nula se rechaza, lo que nos lleva a la conclusión de que los tres tratamientos no tuvieron el mismo efecto sobre el dolor, como se mide con la escala WOMAC A.

7.11 Los datos se ordenan para análisis como se muestra en los corchetes de las celdas.

	Estándar	Débil	Placebo	
Correcto	[35] (26.09)	[12] (24.83)	[30] (26.09)	77
Incorrecto	[27] (35.91)	[47] (34.17)	[32] (35.91)	106
	62	59	62	$N = 183$

Los valores esperados para las celdas (en paréntesis) se calculan por medio de la ecuación 7.11

$$f_e = \frac{(N_R)(N_C)}{N}.$$

[5] Observe que usamos SS_{dentro} para representar la suma de cuadrados, ya que para este problema usamos SS_w para representar la suma de cuadrados para el grupo con el brazalete débil. Por consistencia, usaremos el mismo subíndice para el cuadrado promedio.

[6] Las medias reportadas en el artículo fueron redondeadas, por lo que es necesario redondear al entero más cercano para compensar la cantidad de observaciones de algunos grupos.

Los valores esperados para las seis celdas son, entonces,

$$f_{e11} = \frac{(77)(62)}{183} = 26.09$$
$$f_{e12} = \frac{(77)(59)}{183} = 24.83$$
$$f_{e13} = \frac{(77)(62)}{183} = 26.09$$
$$f_{e21} = \frac{(106)(62)}{183} = 35.91$$
$$f_{e22} = \frac{(106)(59)}{183} = 34.17$$
$$f_{e23} = \frac{(106)(62)}{183} = 35.91$$

Luego, por la ecuación 7.10

$$\chi^2 = \sum_{\substack{\text{todas las}\\ \text{celdas}}} \left[\frac{(f_0 - f_e)^2}{f_e}\right].$$

Por lo que

$$\begin{aligned}\chi^2 &= \frac{(35-26.09)^2}{26.09} + \frac{(12-24.83)^2}{24.83} + \frac{(30-26.09)^2}{26.09} + \frac{(27-35.91)^2}{35.91}\\ &\quad + \frac{(47-34.17)^2}{34.17} + \frac{(32-35.91)^2}{35.91}\\ &= 3.04 + 6.63 + .59 + 2.21 + 4.82 + .43\\ &= 17.72\end{aligned}$$

Del Apéndice D vemos que χ^2 crítica para una prueba con 2 grados de libertad realizada al nivel .05 es 5.991. La hipótesis nula se rechaza, llevándonos a la conclusión de que los tres grupos no fueron igualmente efectivos en determinar el tipo de brazalete que usaron.

C. Las siguientes respuestas se refieren al estudio de caso C.

7.13 Por cuestiones de análisis, será conveniente ordenar los datos para cada periodo de tiempo como sigue.

	Muertos	Vivos
Invasivo	a	b
No invasivo	c	d

Entonces, por las ecuaciones 6.9 y 6.10

$$\widehat{RR} = \frac{a/(a+b)}{c/(c+d)}$$

y

$$Z = \frac{1\text{n } (RR) - 1\text{n } (RR_0)}{\sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}}.$$

Alta

$$\widehat{RR} = \frac{21/(21+441)}{6/(6+452)} = 3.470$$

$$Z = \frac{\ln (3.470)}{\sqrt{\frac{441/21}{21+441} + \frac{452/6}{6+452}}} = 2.715 \qquad p = 2 \times .0033 = .0066$$

Un mes

$$\widehat{RR} = \frac{23/(23+439)}{9/(9+449)} = 2.533$$

$$Z = \frac{\ln (2.533)}{\sqrt{\frac{439/23}{23+439} + \frac{449/9}{9+449}}} = 2.398 \qquad p = 2 \times .0082 = .0164$$

Un año

$$\widehat{RR} = \frac{58/(58+404)}{36/(36+422)} = 1.597$$

$$Z = \frac{\ln (1.597)}{\sqrt{\frac{404/58}{58+404} + \frac{422/36}{36+422}}} = 2.321 \qquad p = 2 \times .0102 = .0204$$

23 meses

$$\widehat{RR} = \frac{80/(80+382)}{59/(59+399)} = 1.344$$

$$Z = \frac{\ln (1.344)}{\sqrt{\frac{382/80}{80+382} + \frac{399/59}{59+399}}} = 1.866 \qquad p = 2 \times .0307 = .0614$$

Comparando el valor-*p* más pequeño, .0066, con $\frac{.05}{4} = .0125$, da un resultado significativo. El segundo valor-*p* más pequeño, .0164, comparado con $\frac{.05}{3} = .0167$, también es significativo, como en el tercero más pequeño, .0204 cuando se compara con $\frac{.05}{2} = .0250$. El valor-*p* restante, .0614, se compara con $\frac{.05}{1} = .0500$ y no es significativo. Así, podemos declarar diferencias significantes para la fecha de alta, un mes y un año, pero no podemos hacer lo mismo para el dato de 23 meses. Además, podemos hacer estas declaraciones sabiendo que el error por familia no excede a .05.

G. Las siguientes respuestas se refieren al estudio de caso G.

7.15 Los datos se ordenan para análisis como se muestra en los corchetes de las celdas.

	Monocigótico	Dicigótico	No mellizos	
EM	[7] (.55)	[1] (.88)	[87] (93.57)	95
Sin EM	[20] (26.45)	[42] (42.12)	[4495] (4488.43)	4557
	27	43	4582	$N = 4652$

Los valores esperados para las celdas (en paréntesis) se calculan por medio de la ecuación 7.11 así

$$f_{e11} = \frac{(95)(27)}{4652} = .55$$
$$f_{e12} = \frac{(95)(43)}{4652} = .88$$
$$f_{e13} = \frac{(95)(4582)}{4652} = 93.57$$
$$f_{e21} = \frac{(4557)(27)}{4652} = 26.45$$
$$f_{e22} = \frac{(4557)(43)}{4652} = 42.12$$
$$f_{e23} = \frac{(4557)(4582)}{4652} = 4488.43$$

Por la ecuación 7.10

$$\chi^2 = \sum_{\substack{\text{todas las} \\ \text{celdas}}} \left[\frac{(f_0 - f_e)^2}{f_e} \right]$$

Por lo que

$$\begin{aligned}
\chi^2 &= \frac{(7-.55)^2}{.55} + \frac{(1-.88)^2}{.88} + \frac{(87-93.57)^2}{93.57} + \frac{(20-26.45)^2}{26.45} \\
&\quad + \frac{(42-42.12)^2}{42.12} + \frac{(4495-4488.43)^2}{4488.43} \\
&= 75.64 + .02 + .46 + 1.57 + .00 + .00 \\
&= 77.69
\end{aligned}$$

Del Apéndice D podemos ver que χ^2 crítica para una prueba con 2 grados de libertad efectuada al nivel .05 es 5.991. La hipótesis nula es rechazada, lo que nos lleva a la conclusión de que los tres grupos no tuvieron las mismas proporciones de hermanos con EM.

Notamos que dos de las celdas tenían valores esperados de .55 y .88. Unos valores esperados tan pequeños nos llevan a dudar sobre la validez de este análisis.

Capítulo ocho

8.1 Dadas las variables *x* y *y*, defina o explique los siguientes términos.

(a) Indica que los valores altos de *x* tienden a asociarse con los valores altos de *y* y los valores bajos de *x* tienden a asociarse con los valores bajos de *y*.

(b) Indica que los valores altos de *x* tienden a asociarse con los valores bajos de *y* y los valores bajos de *x* tienden a asociarse con los valores altos de *y*.

(c) Indica que no existe una relación *lineal* entre *x* y *y*.

(d) Indica qué tan fuerte es la tendencia lineal para *x* y *y* de asociarse de manera positiva o negativa. La fuerza cobra su valor máximo cuando $r = 1$ o $r = -1$, y está en sus valores mínimos cuando $r = 0$.

(e) La relación entre x y y puede describirse por medio de una línea, como en las figuras 8.1 y 8.2.

(f) La correlación de 1.0 indica que x y y toman el mismo valor cuando se expresan sobre una escala en común (por ejemplo, como puntuaciones z). La correlación de -1.0 indica que, cuando se expresan sobre una escala en común (por ejemplo, como puntuaciones z), x y y tienen el mismo valor absoluto pero con un signo algebraico opuesto.

8.3 Los datos se ordenan para análisis en la siguiente tabla.

	Categoría uno	Categoría dos	Categoría tres	
(+)	[19] (14.87)	[17] (8.40)	[13] (15.73)	39
(−)	[44] (47.27)	[29] (26.72)	[51] (50.01)	124
(I)	[6] (6.86)	[3] (3.88)	[9] (7.26)	18
	69	39	73	$N = 181$

Por la ecuación 7.11

$$f_{e11} = \frac{(39)(69)}{181} = 14.87 \qquad f_{e12} = \frac{(39)(39)}{181} = 8.40 \qquad f_{e13} = \frac{(39)(73)}{181} = 15.73$$

$$f_{e21} = \frac{(124)(69)}{181} = 47.27 \qquad f_{e22} = \frac{(124)(39)}{181} = 26.72 \qquad f_{e23} = \frac{(124)(73)}{181} = 50.01$$

$$f_{e31} = \frac{(18)(69)}{181} = 6.86 \qquad f_{e32} = \frac{(18)(39)}{181} = 3.88 \qquad f_{e33} = \frac{(18)(73)}{181} = 7.26$$

Entonces, por la ecuación 7.10 chi-cuadrada obtenida es

$$\begin{aligned}\chi^2 &= \frac{(19.0-14.87)^2}{14.87} + \frac{(7.0-8.40)^2}{8.40} + \frac{(13.0-15.73)^2}{15.73} + \frac{(44.0-47.27)^2}{47.27} + \frac{(29.0-26.72)^2}{26.72} \\ &\quad \frac{(51.0-50.01)^2}{50.01} + \frac{(6.0-6.86)^2}{6.86} + \frac{(3.0-3.88)^2}{3.88} + \frac{(9.0-7.26)^2}{7.26} \\ &= 1.147 + .233 + .474 + .226 + .195 + .020 + .108 + .200 + .417 \\ &= 3.02.\end{aligned}$$

Del Apéndice D vemos que chi-cuadrada crítica para una prueba con

$$(j-1)(k-1) = (3-1)(3-1) = 4$$

grados de libertad efectuada con $\alpha = .05$ es 9.488. La hipótesis nula no se rechaza, por lo que no podemos afirmar que exista una relación entre el tiempo del arresto y el resultado de la prueba.

D. Las siguientes respuestas se refieren al estudio de caso D.

8.5 Sí, P-M representa la relación lineal entre las dos variables, sin importar la forma de la población bivariante implicada.

8.7 Sí, las evaluaciones PBV más elevadas indican un volumen cerebral mayor, mientras que las puntuaciones NPZ-8 más bajas indican una mejor función neuropsicológica. Entonces, existe una tendencia de que las evaluaciones PBV se asocien con puntuaciones NPZ-8 más bajas.

8.9 No, obtuvieron una correlación de una mezcla de sujetos infectados y no infectados con VIH, lo que es contrario a su meta establecida. Es más, la correlación de sujetos infectados con VIH por sí sola es significativamente más baja que la de la mezcla.

L. Las siguientes respuestas se refieren al estudio de caso L.

8.11

	Lesión menor	Lesión moderada	Lesión severa	
< 30	[255] (226.54)	[57] (76.55)	[14] (22.91)	326
30 – 39	[332] (334.25)	[105] (112.94)	[44] (33.80)	481
40 – 49	[533] (542.03)	[201] (183.15)	[46] (54.82)	780
50+	[561] (578.17)	[205] (195.36)	[66] (58.47)	832
	1681	568	170	$N = 2419$

$$f_{e11} = \frac{(326)(1681)}{2419} = 226.54 \quad f_{e12} = \frac{(326)(568)}{2419} = 76.55 \quad f_{e13} = \frac{(326)(170)}{2419} = 22.91$$

$$f_{e21} = \frac{(481)(1681)}{2419} = 334.25 \quad f_{e22} = \frac{(481)(568)}{2419} = 112.94 \quad f_{e23} = \frac{(481)(170)}{2419} = 33.80$$

$$f_{e31} = \frac{(780)(1681)}{2419} = 542.03 \quad f_{e32} = \frac{(780)(568)}{2419} = 183.15 \quad f_{e33} = \frac{(780)(170)}{2419} = 54.82$$

$$f_{e41} = \frac{(832)(1681)}{2419} = 578.17 \quad f_{e42} = \frac{(832)(568)}{2419} = 195.36 \quad f_{e43} = \frac{(832)(170)}{2419} = 58.47$$

Por la ecuación 7.10, chi-cuadrada obtenida es

$$\begin{aligned}
\chi^2 &= \frac{(255-226.54)^2}{226.54} + \frac{(57-76.55)^2}{76.55} + \frac{(14-22.91)^2}{22.91} \\
&\quad + \frac{(332-334.25)^2}{334.25} + \frac{(105-112.94)^2}{112.94} + \frac{(44-33.80)^2}{33.80} \\
&\quad + \frac{(555-542.03)^2}{542.03} + \frac{(201-183.15)^2}{183.15} + \frac{(46-54.82)^2}{54.82} \\
&\quad + \frac{(561-578.17)^2}{578.17} + \frac{(205-195.36)^2}{195.36} + \frac{(66-58.47)^2}{58.47} \\
&= 3.58 + 4.99 + 3.47 + .02 + .56 + 3.08 + .15 + 1.74 + 1.42 + .51 + .48 + .97 \\
&= 20.97
\end{aligned}$$

Comenzar el Apéndice D con (4–1)(3–1) = 6 grados de libertad produce un valor crítico chi-cuadrado de 12.592 para una prueba efectuada al nivel .05. Ya que la chi-cuadrada obtenida de 20.97 excede este valor, rechazamos la hipótesis nula de independencia y concluimos que la gravedad de las lesiones se relaciona con la edad del sujeto lesionado.

M. Las siguientes respuestas se refieren al estudio de caso M.

8.13 Usando la ecuación 8.2 calculamos

$$r = \frac{522.25 - \frac{(62.9)(58.1)}{7}}{\sqrt{\left[565.89 - \frac{(62.9)^2}{7}\right]\left[482.43 - \frac{(58.1)^2}{7}\right]}}$$
$$= .485$$

Por las ecuaciones 8.6 y 8.7

$$L = \frac{(1+F)r + (1-F)}{(1+F) + (1-F)r}$$
$$= \frac{(1+7.15)(.485) + (1-7.15)}{(1+7.15) + (1-7.15)(.485)}$$
$$= -.425$$
$$U = \frac{(1+F)r - (1-F)}{(1+F) - (1-F)r}$$
$$= \frac{(1+7.15)(.485) - (1-7.15)}{(1+7.15) - (1-7.15)(.485)}$$
$$= .907$$

F = 7.15 se obtuvo consultando en el Apéndice C para un intervalo de confianza con 7 – 2 = 5 grados de libertad del numerador y el denominador.

Capítulo nueve

9.1 (a) Por la ecuación 9.9 de la página 324 y la ecuación 9.13 de la página 326

$$\hat{R}^2 = \frac{SS_{reg}}{SS_v}$$

y

$$\hat{R}^2 = r^2_{y\hat{y}}$$

(b) Por la ecuación 9.8 de la página 323 y la ecuación 9.12 de la página 326

$$1 - \hat{R}^2 = \frac{SS_{res}}{SS_y}$$

y

$$1 - \hat{R}^2 = 1 - r^2_{y\hat{y}}$$

(c) De la sección 9.2.4 de la página 325,

$$SS_{res} = 0 \text{ y } SS_{reg} = SS_y.$$

(d) De la sección 9.2.4 de la página 325,

$$SS_{res} = SS_y \text{ y } SS_{reg} = 0.$$

9.3

$$R^2_{y.1,2,3,4} - R^2_{y.2,4}$$

Se hace una comparación entre un modelo que contiene todas las variables incluyendo x_1 y x_3 y un modelo que contiene todas las variables *excepto* x_1 y x_3. De tal forma que si R^2 es mayor para el modelo más grande, debe ser por la presencia de x_1 y x_3.

9.5 Las respuestas las dan la ecuación 9.23 de la página 332

$$F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}}$$

y la ecuación 9.24 de la página 335

$$F = \frac{\frac{R^2_{y.L} - R^2_{y.S}}{p_L - p_S}}{\frac{1-R^2_{y.L}}{N-p_L-1}}$$

(a) Por la ecuación 9.23

$$F = \frac{\frac{\hat{R}^2_{y.1,2,3,4}}{4}}{\frac{1-\hat{R}^2_{y.1,2,3,4}}{40-4-1}} = \frac{\frac{.68}{4}}{\frac{1-.68}{35}} = 18.59.$$

Para cuatro y 35 grados de libertad, F crítica es 2.64, por lo que rechazamos la hipótesis nula.

Respuesta: **sí**

(b) Por la ecuación 9.24

$$F = \frac{\frac{R^2_{y.1,2,3,4} - R^2_{y.2,3,4}}{4-3}}{\frac{1-R^2_{y.1,2,3,4}}{40-4-1}} = \frac{\frac{.68-.43}{1}}{\frac{1-.68}{35}} = 27.34$$

Para uno y 35 grados de libertad, F crítica es 4.12, por lo que rechazamos la hipótesis nula.

Respuesta: **sí**

(c) Por la ecuación 9.23

$$F = \frac{\frac{\hat{R}^2_{y.2,4}}{2}}{\frac{1-\hat{R}^2_{y.2,4}}{40-2-1}} = \frac{\frac{.14}{2}}{\frac{1-.14}{37}} = 3.01.$$

Para dos y 37 grados de libertad, F crítica es 3.25, por lo que no rechazamos la hipótesis nula.

Respuesta: **evidencia insuficiente para hacer esta afirmación**

(d) Por la ecuación 9.24

$$F = \frac{\frac{R^2_{y.1,2,3,4} - R^2_{y.1,3}}{4-2}}{\frac{1-R^2_{y.1,2,3,4}}{40-4-1}} = \frac{\frac{.68-.27}{2}}{\frac{1-.68}{35}} = 22.42$$

Para dos y 35 grados de libertad, F crítica es 3.27, por lo que rechazamos la hipótesis nula.

Respuesta: **sí**

D. Las siguientes respuestas se refieren al estudio de caso D.

9.7 Si y representa las puntuaciones NPZ-8, x_1 los valores PBV y x_2 los valores CD4, comenzamos por calcular los siguientes valores intermedios (véase los cálculos de la página 329).

$$SS_y = \sum y^2 - \frac{(\Sigma y)^2}{n} = 374 - \frac{(54)^2}{15} = 179.6$$

$$SS_{x_1} = \sum x_1^2 - \frac{(\Sigma x_1)^2}{n} = 9.944 - \frac{(12.169)^2}{15} = .0717$$

$$SS_{x_2} = \sum x_2^2 - \frac{(\Sigma x_2)^2}{n} = 3356299 - \frac{(5817)^2}{15} = 1100466.4$$

$$SS_{yx_1} = \sum yx_1 - \frac{(\Sigma y)(\Sigma x_1)}{n} = 42.962 - \frac{(54)(12.169)}{15} = -.8464$$

$$SS_{yx_2} = \sum yx_2 - \frac{(\Sigma y)(\Sigma x_2)}{n} = 16442 - \frac{(54)(5817)}{15} = -4499.2$$

$$SS_{x_1x_2} = \sum x_1x_2 - \frac{(\Sigma x_1)(\Sigma x_2)}{n} = 4736.929 - \frac{(12.169)(5817)}{15} = 17.7908$$

Por las ecuaciones 9.20 y 9.21

$$\begin{aligned} b_1 &= \frac{(SS_{x_2})(SS_{yx_1}) - (SS_{x_1x_2})(SS_{yx_2})}{(SS_{x_1})(SS_{x_2}) - (SS_{x_1x_2})^2} \\ &= \frac{(1100466.4)(-.8464) - (17.7908)(-4499.2)}{(.0717)(1100466.4) - (17.7908)^2} \\ &= -10.8337 \\ b_2 &= \frac{(SS_{x_1})(SS_{yx_2}) - (SS_{x_1x_2})(SS_{yx1})}{(SS_{x_1})(SS_{x_2}) - (SS_{x_1x_2})^2} \\ &= \frac{(.0717)(-4499.2) - (17.7908)(-.8464)}{(.0717)(1100466.4) - (17.7908)^2} \\ &= -.0039 \end{aligned}$$

Considerando que $\bar{y} = \frac{54}{15} = 3.60$ y $\bar{x} = \frac{12.169}{15} = .8113$ y $\bar{x}_2 = \frac{5817}{15} = 387.8$ y aplicando la ecuación 9.19 tenemos que

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 = 3.6 - (-10.8337)(.8113) - (-.0039)(387.8) = 13.9018.$$

Entonces, el modelo de dos predictores es

$$\hat{y} = a + b_1x_1 + b_2x_2 = 13.9018 - 10.8337x_1 - .0039x_2.$$

(a) Por la ecuación 9.22

$$\begin{aligned} SS_{reg} &= b_1SS_{yx_1} + b_2SS_{yx_2} + \ldots + b_pSS_{yx_p} \\ &= (-10.8337)(-.8464) + (-.0039)(-4499.2) \\ &= 26.7165 \end{aligned}$$

Luego, usando SS_y como se calculó previamente, y la ecuación 9.9

$$\hat{R}^2_{y.12} = \frac{SS_{reg}}{SS_y} = \frac{26.7165}{179.6} = 0.1488.$$

(b) Utilizando el resultado de 9.6c con la ecuación 9.24 obtenemos

$$F = \frac{\frac{\hat{R}^2_{y.L} - \hat{R}^2_{y.S}}{p_L - p_S}}{\frac{1-\hat{R}^2_{y.L}}{N-p_L-1}} = \frac{\frac{.1488-.0556}{2-1}}{\frac{1-.1488}{15-2-1}} = 1.314.$$

La referencia al Apéndice C con 1 y 12 grados de libertad da una F crítica de 4.75 para una prueba al nivel .05. El resultado no es significativo, por lo tanto, no pudimos demostrar un aumento en R^2 como consecuencia de sumar CD4 al modelo.

(c)

$$\hat{y}_1 = 13.9018 - 10.8337\ (.791) - .0039\ (16) = 5.27$$
$$\hat{y}_2 = 13.9018 - 10.8337\ (.782) - .0039\ (324) = 4.17$$
$$\hat{y}_3 = 13.9018 - 10.8337\ (.646) - .0039\ (256) = 5.90$$
$$\hat{y}_4 = 13.9018 - 10.8337\ (.740) - .0039\ (563) = 3.69$$
$$\hat{y}_5 = 13.9018 - 10.8337\ (.804) - .0039\ (321) = 3.94$$

Parece que hay poca mejoría obtenida del uso del modelo de dos predictores.

(d) Usando $\hat{R}^2_{y.12} = .1488$, calculado en el ejercicio 9.7a con la ecuación 9.23 de la página 332, da

$$F = \frac{\frac{\hat{R}^2}{p}}{\frac{1-\hat{R}^2}{N-p-1}} = \frac{\frac{.1488}{2}}{\frac{1-.1488}{15-2-1}} = 1.049$$

La referencia al Apéndice C con 2 y 12 grados de libertad da una F crítica de 3.89 para una prueba al nivel .05. Ya que la F obtenida de 1.049 es menor que este valor, no se rechaza la hipótesis nula. No pudimos, por ende, demostrar que cuando PBV y CD4 se usan juntos justifican cualquier variación en las puntuaciones NPZ-8.

Capítulo diez

10.1 Por la ecuación 10.1 de la página 345, cuando se administran cinco pruebas, $P_n = n! = 5! = 120$. Para seis pruebas, $6! = 720$.

10.3 (a) De la página 350, calculamos $3! = 6$.

(b) No. El valor-p más pequeño posible es $1/6 = .167$.

10.5 Como se mencionó en la página 358, utilizamos la ecuación 8.4 de la página 309 refiriendo el resultado a la tabla B con $n - 2$ grados de libertad, por lo que

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{.88}{\sqrt{\frac{1.88^2}{124-2}}} = 20.464.$$

t crítica para una prueba de dos colas con 122 grados de libertad con $\alpha = .10$ es ± 1.657, por lo que se rechaza la hipótesis nula.

10.7 Los rangos signados (sr), sus cuadrados (sr^2) y la suma de cada uno son como sigue.

	sr	sr^2
	2	4
	5	25
	–4	16
	–1	1
	3	9
	9	81
	7	49
	–6	36
	8	64
Σ	23	285

El estadístico t de muestras apareadas (ecuación 5.1 de la página 162) en donde se sustituye R_d por d para indicar que el análisis se lleva a cabo con rangos signados de puntuaciones diferenciales y no con puntuaciones diferenciales es como sigue.

$$t = \frac{\overline{R}_d}{\frac{s_{R_d}}{\sqrt{n}}} = \frac{2.556}{\frac{5.318}{\sqrt{9}}} = 1.442$$

donde $\overline{R}_d = \frac{23}{9} = 2.556$ y

$$s_{R_d} = \sqrt{\frac{\sum R_d^2 - \frac{(\sum R_d)^2}{n}}{n-1}} = \sqrt{\frac{285 - \frac{(23)^2}{9}}{9-1}} = 5.318.$$

La referencia al Apéndice G muestra que el valor crítico de dos colas para la estadística de los signos para $n = 9$ y $\alpha = .05$ es ±2.704. No se rechaza la hipótesis nula.

10.9 La conversión a rangos da como resultado lo siguiente.

Pública	Privada
18	11
14	8
5	13
16	15
19	20
10	3
9	4
2	1
17	6
12	7

El rango medio para el primer grupo es

$$R_1 = \frac{\sum R_1}{n_1} = \frac{122}{10} = 12.2$$

en tanto que para el segundo es

$$R_2 = \frac{\sum R_2}{n_2} = \frac{88}{10} = 8.8.$$

El cálculo de la varianza combinada (por medio de la ecuación 6.2 de la página 220) es

$$s_{P_R}^2 = \frac{\left(\sum R_1^2 - \frac{(\sum R_1)^2}{n_1}\right) + \left(\sum R_2^2 - \frac{(\sum R_2)}{n_2}\right)}{n_1 + n_2 - 2}$$

$$= \frac{\left(1780 - \frac{(122)^2}{10}\right) + \left(1090 - \frac{(88)^2}{10}\right)}{10 + 10 - 2}$$

$$= 33.733$$

Entonces, por la ecuación 6.1 de la página 219, t obtenida, con las sustituciones correctas de R por x, es

$$t = \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{s_{P_R}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{12.2 - 8.8}{\sqrt{33.733\left(\frac{1}{10} + \frac{1}{10}\right)}}$$

$$= 1.309$$

La referencia al Apéndice H muestra que t crítica para una prueba de dos colas de Wilcoxon realizada con $\alpha = .05$ es ± 2.248. No pudimos rechazar la hipótesis nula y, por ende, no pudimos demostrar una diferencia en los tiempos de espera en las dos instituciones.

10.11 Los rangos de los tres grupos son los siguientes.

Clase	Internet	Filmación
8	11	15
6	5	14
2	1	10
12	3	9
7	13	4

Si R_1 representa los rangos del primer grupo, R_2^1 los rangos cuadrados del primer grupo y los subíndices 2 y 3 los grupos restantes, entonces calculamos

$$\sum R_1 = 8 + 6 + 2 + 12 + 7 = 35$$

$$\sum R_1^2 = 8^2 + 6^2 + 2^2 + 12^2 + 7^2 = 297$$

$$\sum R_2 = 11 + 5 + 1 + 3 + 13 = 33$$

$$\sum R_2^2 = 11^2 + 5^2 + 1^2 + 3^2 + 13^2 = 325$$

$$\sum R_3 = 15 + 14 + 10 + 9 + 4 = 52$$

$$\sum R_3^2 = 15^2 + 14^2 + 10^2 + 9^2 + 4^2 = 618$$

Las sumas de cuadrados para los tres grupos son los siguientes.

$$SS_1 = \sum R_1^2 - + \frac{(\Sigma R_1)^2}{n_1} = 297 - \frac{(35)^2}{5} = 52.0$$

$$SS_2 = \sum R_2^2 - \frac{(\Sigma R_2)^2}{n_2} = 325 - \frac{(33)^2}{5} = 107.2$$

$$SS_3 = \sum R_3^2 - \frac{(\Sigma R_3)^2}{n_3} = 618 - \frac{(52)^2}{5} = 77.2$$

Por la ecuación 7.4 de la página 265 la suma de cuadrados es

$$SS_w = SS_1 + SS_2 + SS_3 = 52.0 + 107.2 + 77.2 = 236.4.$$

Por la ecuación 7.3 de la página 265 el cuadrado medio es

$$MS_w = \frac{SS_w}{N-k} = \frac{236.4}{15-3} = 19.7$$

donde SS_w es la suma de cuadrados dentro, N es el número total de casos y k es el número de grupos.

Por la ecuación 7.7 de la página 267, sustituyendo R por x para indicar que el cálculo es para rangos y no para observaciones originales, la suma de cuadrados entre es

$$SS_b = n\left[\sum_{j=1}^{k} \bar{R}_j^2 - \frac{\left(\sum_{j=1}^{k} \bar{R}_j\right)^2}{k}\right]$$

donde n es el número de observaciones en *cada* grupo, y $\bar{R}_j$ son las medias del grupo de rangos.

De cálculos anteriores obtenemos

$$\bar{R}_1 = \frac{\Sigma R_1}{n_1} = \frac{35}{5} = 7.0$$

$$\bar{R}_2 = \frac{\Sigma R_2}{n_2} = \frac{33}{5} = 6.6$$

$$\bar{R}_3 = \frac{\Sigma R_3}{n_3} = \frac{52}{5} = 10.4$$

Entonces

$$\sum \bar{R} = 7.0 + 6.6 + 10.4 = 24.0$$

y

$$\sum \bar{R}^2 = 7.0^2 + 6.6^2 + 10.4^2 = 200.72.$$

Si hacemos las sustituciones apropiadas en la ecuación 7.7, tenemos que

$$SS_b = 5\left[200.72 - \frac{(24.00)^2}{3}\right] = 43.6$$

Por la ecuación 7.6 de la página 267 el cuadrado medio entre es

$$MS_b = \frac{SS_b}{k-1} = \frac{43.6}{3-1} = 21.8$$

donde SS_b es la suma de cuadrados entre y k es el número de grupos. Finalmente, por la ecuación 7.2 de la página 264 la F obtenida es

$$F = \frac{MS_b}{MS_w} = \frac{21.8}{19.7} = 1.107$$

Con $\alpha = .05$, tres grupos y cinco observaciones por grupo, el Apéndice I da una F crítica de 4.072. Como la F obtenida de 1.107 es menor que este valor, no pudimos rechazar la hipótesis nula y, por lo tanto, no podemos demostrar una diferencia en el efecto de los tres métodos de instrucción.

A. Las siguientes respuestas se refieren al estudio de caso A.

10.13 Este enunciado podría conducir a pensar que las pruebas no paramétricas pueden ser menos o más poderosas que sus equivalentes paramétricos.

D. Las siguientes respuestas se refieren al estudio de caso D.

10.15 (a) La prueba de suma de rangos de Wilcoxon que también se conoce como la prueba Mann-Whitney.

(b) Remplazar las puntuaciones NPZ-8 con rangos produce lo siguiente.

HIV +	HIV −
20.0	5.0
15.5	5.0
13.0	10.0
15.5	5.0
11.0	5.0
17.0	
19.0	
5.0	
13.0	
5.0	
5.0	
18.0	
13.0	
5.0	
5.0	

Usando los subíndices 1 y 2 para representar los grupos VIH positivo y negativo, respectivamente, calculamos s_P^2 sobre los rangos mediante la ecuación 6.2 de la página 220 como sigue.

$$s_{P_R}^2 = \frac{\left(\sum R_1^2 - \frac{(\sum R_1)^2}{n_1}\right) + \left(\sum R_2^2 - \frac{(\sum R_2)^2}{n_2}\right)}{n_1 + n_2 - 2}$$

$$= \frac{\left(2607.5 - \frac{(180)^2}{15}\right) + \left(200 - \frac{(30)^2}{5}\right)}{15 + 5 - 2}$$

$$= 25.972$$

Luego, aplicando la ecuación 6.1 de la página 219 a los rangos, tenemos que

$$t = \frac{\bar{R}_1 - \bar{R}_2}{\sqrt{s_{P_R}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{12.0 - 6.0}{\sqrt{25.972\left(\frac{1}{15} + \frac{1}{5}\right)}} = 2.280.$$

La referencia al Apéndice B da valores críticos para una prueba t de dos colas con 18 grados de libertad efectuada con $\alpha = .05$ como ± 2.101[7], lo que da como resultado el rechazo de la hipótesis nula.

Los rangos para los datos PBV son los siguientes.

HIV +	HIV −
6	16
4	11
1	12
3	18
8	19
13	
2	
7	
10	
9	
5	
14	
17	
20	
15	

s_P^2 es, entonces,

$$s_P^2 = \frac{\left(1664 - \frac{(134)^2}{15}\right) + \left(1206 - \frac{(76)^2}{5}\right)}{15 + 5 - 2} = 28.763$$

[7] Este valor sería visto con cuidado ya que fue tomado de una distribución t y los dos tamaños de muestras no son mayores que 15. El valor crítico exacto es ± 2.263, que también da por resultado el rechazo de la hipótesis nula.

y la t obtenida es

$$t = \frac{8.933 - 15.200}{\sqrt{28.763\left(\frac{1}{15} + \frac{1}{5}\right)}} = -2.263.$$

La referencia al Apéndice B da los valores críticos para una prueba t de dos colas con 18 grados de libertad efectuada con $\alpha = .05$ como ± 2.101 (véase la nota al pie de página 7) lo que da como resultado el rechazo de la hipótesis nula.

(c) El supuesto de las observaciones no ligadas se incumplió en el caso de la prueba de las puntuaciones NPZ-8 pero no en el caso de la prueba de los valores PBV.

(d) Se necesitó una prueba aproximada para el análisis NPZ-8 debido a la presencia de observaciones ligadas. Se podría haber realizado una prueba exacta para el análisis PBV pero debido a que la tabla de valores críticos disponible se construyó para tamaños iguales de muestras, se hizo la referencia a una distribución t para que se realizara una prueba aproximada.

(e) Se podría obtener una versión exacta de la prueba de rangos de las puntuaciones NPZ-8 si se convirtieran en rangos las puntuaciones originales y después se desarrollara la distribución de permutación del estadístico t y refiriendo el estadístico de prueba a esa distribución. Se podría obtener una versión exacta de la prueba de las puntuaciones PBV si se refiriera al estadístico de prueba a una tabla que incluya tamaños de muestras desiguales.

10.17 Los rangos para los tres grupos son los siguientes.

ADC +	ADC −	HIV −
20.0	5.0	5.0
15.5	13.0	5.0
13.0	5.0	10.0
15.5	5.0	5.0
11.0	18.0	5.0
17.0	13.0	
19.0	5.0	
	5.0	

Si designamos los tres grupos con los subíndices 1, 2 y 3, respectivamente, calculamos las sumas de cuadrados para los grupos de la siguiente manera.

$$SS_1 = \sum R_1^2 - \frac{(\sum R_1)^2}{n_1} = 1820.5 - \frac{(111)^2}{7} = 60.357$$

$$SS_2 = \sum R_2^2 - \frac{(\sum R_2)^2}{n_2} = 787 - \frac{(69)^2}{8} = 191.875$$

$$SS_3 = \sum R_3^2 - \frac{(\sum R_3)^2}{n_3} = 200 - \frac{(30)^2}{5} = 20.000$$

Entonces, por 7.4

$$SS_w = SS_1 + SS_2 + ... + SS_k = 60.357 + 191.875 + 20.000 = 272.232$$

Por la ecuación 7.3

$$MS_w = \frac{SS_w}{N-k} = \frac{272.232}{20-3} = 16.014.$$

Por la ecuación 7.8

$$\begin{aligned} SS_b &= \frac{(\sum_{i=1}^{n_1} x_{i1})^2}{n_1} + \frac{(\sum_{i=1}^{n_2} x_{i2})^2}{n_2} + \ldots + \frac{(\sum_{i=1}^{n_k} x_{ik})^2}{n_k} - \frac{(\sum_{Todas} x..)^2}{N} \\ &= \frac{(111)^2}{7} + \frac{(69)^2}{8} + \frac{(30)^2}{5} - \frac{(210)^2}{20} \\ &= 330.268. \end{aligned}$$

Por la ecuación 7.6

$$MS_b = \frac{SS_b}{k-1} = \frac{330.268}{3-1} = 165.134.$$

La F obtenida es, entonces,

$$F = \frac{MS_b}{MS_w} = \frac{165.134}{16.014} = 10.312.$$

Al consultar el Apéndice C con 2 y 17 grados de libertad, encontramos que la F crítica para una prueba con $\alpha = .05$ es 3.59. Como la F obtenida de 10.312 excede este valor, rechazamos la hipótesis de medias de población iguales.

Debido a que esta prueba es aproximada, así como por otras razones, no debemos dar demasiada importancia a una comparación con el ANOVA efectuado sobre las puntuaciones NPZ-8 originales (véase el ejercicio 7.14, página 294) pero es interesante considerar que la F obtenida, calculada sobre los rangos, es más grande que la calculada sobre las puntuaciones originales.

K. Las siguientes respuestas se refieren al estudio de caso K.

10.19 No, por dos razones. (1) Bajo una hipótesis nula verdadera, las puntuaciones diferenciales tienden a ser simétricas sin importar la forma de las distribuciones pre-prueba/post-prueba. La prueba t de muestras apareadas tiende a ser robusta con respecto a la no normalidad de la población mientras que la distribución de la puntuación diferencial sea simétrica. (2) Las puntuaciones pre- y post-prueba son medias y no observaciones individuales y, debido al teorema del límite central, las medias tienden a la normalidad sin importar la forma de la población.

O. Las siguientes respuestas se refieren al estudio de caso O.

10.21 No, esto es sólo un mito. Hay circunstancias en las que las pruebas paramétricas son más potentes y circunstancias en las que las pruebas no paramétricas son más potentes. Los enunciados indiscriminados de este tipo son injustificados.

10.23 Esto es definitivamente falso. Se puede pensar que la prueba Kruskal-Wallis es una generalización de la prueba de Wilcoxon de suma de rangos (o de Mann-Whitney), pero definitivamente no la prueba de signos.

10.25 No, esto se lleva a cabo por medio de la correlación P-M los datos originales, pero la correlación de Spearman utiliza rangos. Como resultado, la evaluación por medio de la prueba basada en rangos no es para una relación en línea recta sino para una relación monótona.

Bibliografía

1. R. Clifford Blair y James J. Higgins, *A comparison of the power of wilcoxon's rank-sum statistic to that of student's t statistic under various non-normal distributions*, Journal of Educational Statistics **5** (1980), 309-335.
2. R. Clifford Blair y James J. Higgins, *Comparison of the power of the paired samples t-test to that of wilcoxon's signed rank test under various population shapes*, Psychological Bulletin **97** (1985), 119-128.
3. William E. Boden, Robert A. O' Rourke, Michael H. Crawford, Alvin S. Blaustein, Prakash C. Deedwania, Robert G. Zoble, Laura F. Wexler, Robert E. Kleiger, Carl J. Pepine, David R. Ferry, Bruce K. Chow y Philip W. Lavori, *Outcomes in patients with acute Non-Q-Wave myocardial infarction randomly assigned to an invasive as compared with a conservative management strategy*, The New England Journal Of Medicine **338** (1998), 1785-1792.
4. James V. Bradley, *Distribution free statistical tests*, 1a ed., Prentice-Hall, Inc., Englewood Cliffs, Nueva Jersey, 1968.
5. Norman E. Breslow y N. E. Day, *Statistical methods in cancer research, volume i-the analysis of case-control studies*, IARC Scientific Publications, núm. 32, Lyon, Francia: International Agency for Research on Cancer, 1980.
6. Chantal Coles, Noel A. Brennan, Vicki Shuley, Jill Woods, Chris Prior, Joseph G. Vehige y A. Simmons, *The influence of lens conditioning on signs and symptoms with new hydrogel contact lenses*, Clinical and Experimental Optometry **87** (2004), 367-371.
7. William J. Conover, *Practical nonparametric statistics*, 2a ed., John Wiley and Sons, Nueva York, Chichester, Brisbane, Toronto, 1980.
8. William J. Conover y Ronald Liman, *Rank transformations as a bridge between parametric and nonparametric statistics*, The American Statistician **35** (1981), 124-129.
9. Cytel Software Corporation, Cambridge, MA, *Statxact: software for exact nonparametric inference*, 2003, versión 6.
10. A. E. Dusoir, *Statistical calculator*, Mole Software, 2004, v2.03.
11. G. C. Ebers, D. E. Bulman, A. D. Sadovnick, D. W. Paty, S. Warren, W. Hader, T. J. Murray, T. P. Seland, P. Duquette, T. Grey *et al.*, *A population-based study of multiple sclerosis in twins*, The New England Journal Of Medicine **315** (1986), 1638-1642.
12. Josie M. M. Evans, Ray W. Newton, Danny A. Ruta, Thomas M. Macdonald, Richard J. Stevenson y Andrew D. Morris, *Frequency of blood glucose monitoring in relation to glycaemic control: observational study with diabetes database*, British Medical Journal **319** (1999), 83-86.
13. L. Evans y M. C. Frick, *Helmet effectiveness of preventing motorcycle driver and passenger fatalities*, Accident Analysis and Prevention **20** (1988), 447-458.
14. Expert Panel On Detection, Evaluation, and Treatment Of High Blood Cholesterol In Adults, *Executive summary of the third report of the national cholesterol education program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III)*, Journal of the American Medical Association **285** (2001), 2486-2497.

15. Ronald A. Fisher, *Frequency distributions of the values of the correlation coefficient in samples from an indefinitely large population*, Biometrika **10** (1915), 507-521.

16. Ronald A. Fisher, *On the "probable error" of a coefficient of correlation deduced from a small sample*, Metron **1**(1921), 3-32.

17. Ronald A. Fisher, *Applications of "Student's" distribution*, Metron **5** (1925), 90-104.

18. Joseph L. Fleiss, *Statistical methods for rates and proportions*, John Wiley and Sons, Nueva York, 1981.

19. Trisha Greenhalgh, *How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests*, British Medical Journal **315** (1997), 364-366.

20. Sander Greenland, *Applications of stratified analysis methods*, Modern Epidemiology (Kenneth Rothmann y Sander Greenland, eds.), Lippincott-Raven, Filadelfia, 2a ed., 1998.

21. Tim Harlow, Colin Greaves, Adrian White, Liz Brown, Anna Hart y Edzard Ernst, *Randomised controlled, trial of magnetic bracelets for relieving pain in osteoarthritis of the hip and knee*, British Medical Journa1 **329** (2004), 18-25.

22. Yosef Hochberg y Ajit C. Tamhane, *Multiple comparison procedures*, John Wiley and Sons, Nueva York, 1987.

23. J. M. Hoenig y D. M. Heisey, *The abuse of power: The pervasive fallacy of power calculations for data analysis*, The American Statistician **55** (2001), 19-24.

24. S. Holm, A *simple sequentially rejective multiple test procedure*, Scandinavian Journal of Statistics **6** (1979), 65-70.

25. Esko A. Kemppainen, Johan I. Hedström, Pauli A. Puolakkainen, Vesa S. Sainio, Reifo K. Haapiainen, Vesa Perhoniemi, Sirpa Osman, Eero O. Kivilaakso y Ulf-Håkan Stenman, *Rapid measurement of urinary trypsinogen-2 as a screening test for acute pancreatitis*, The New England Journal Of Medicine **336** (1997), 1788-1793.

26. Winfried V. Kern, Alain Cometta, Robrecht De Bock, John Langenaeken, Marianne Paesmans, Harold Gaya, Giorgio Zanetti, Thierry Calandra, Michel P. Glauser, Françoise Crokaert, Jean Klastersky, Athanasios Skoutelis, Harry Bassaris, Stephen H. Zinner, Claudio Viscoli, Dan Engelhard y AndrewPadmos for the International Antimicrobial Therapy Cooperative Group of the European Organization for Research and Treatment of Cancer, *Oral versus intravenous empirical antimicrobial therapy for fever in patients with granulocytopenia who are receiving cancer chemotherapy*, The New England Journal Of Medicine **341** (1999), 312-318.

27. Roger E. Kirk, *Experimental design: Procedures for the behavioral sciences*, 3a ed., Wadsworth Publishing Company, Belmont, California, 1994.

28. William H. Kruskal y William A. Wallis, *Use of ranks in one-criterion variance analysis*, Journal of the American Statistical Association **48** (1953), 907-911.

29. Jan W. Kuzma, *Basic statistics for the health sciences*, 3a ed., Mayfield Publishing Company, Mountain View, California, 1998.

30. Jan W. Kuzma y Stephen E. Bohnenblust, *Basic statistics for the health sciences*, 4a ed., Mayfield Publishing Company, Mountain View, California, 2001.

31. A. M. Lilienfeld y D. E. Lilienfeld, *Foundations of epidemiology*, 2a ed., Oxford University Press, Nueva York, 1980.

32. Karen D. Liller, Virginia Noland, Pabitra Rijal, Karen Pesce y Robin Gonzalez, *Development and evaluation of the Kids Count Farm Safety Lesson*, Journal of Agricultural Safety and Health **8** (2002), 411-421.

33. FREDRICK M. LORD, *On the statistical treatment of football numbers*, American Psychologist **8** (1953), 750-751.

34. THOMAS C. MARKELLO, ISA M. BERNARDINI Y WILLIAM A. GAHL, *Improved renal function in children with cystinosis treated with cysteamine*, The New England Journal Of Medicine **328** (1993), 1157-1162.

35. CHRISTINA MASLACH Y SUSAN E. JACKSON, *Maslach burnout inventory: Manual*, 2a ed., Consulting Psychologists Press, Palo Alto, California, 1986.

36. WILLIAM MENDENHALL Y ROBERT J. BEAVER, *Introduction to probability and statistics*, 8a ed., PWS-Kent Publishing Company, Boston, Massachusetts, 1991.

37. SOHIL H. PATEL, DENNIS L. KOLSON, GUILA GLOSSER, ISABEL MATOZZO, YULIN GE, JAMES S. BABB, LOIS J. MANNON y ROBERT I. GROSSMAN, *Correlation between percentage of brain parenchymal volume and neurocognitive performance in hiv-infected patients*, American Journal of Neuroradiology **23** (2002), 543-549.

38. KAREN M. PERRIN, SOMER L. GOAD Y CAROL WILLIAMS, *Can school nurses save money by treating school employees as well as students?*, Journal of School Health **72** (2002), 305-306.

39. PHILIP H. RAMSEY Y PATRICIA P. RAMSEY, *Evaluating the normal approximation to the binomial test*, Journal of Educational Statistics **13** (1988), 173-182.

40. GEORG RÖGGLA, BERTHOLD MOSER Y MARTIN RÖGGLA, *Effect of temazepam on ventilatory response at moderate altitude*, British Medical Journal **320** (2000), 1-5.

41. JULIAN SIMON Y PETER BRUCE, *Resampling stats software*, Resampling Stats, Inc., 1973-2004.

42. DAVID SKUSE, ARMON BENTOVIM, JILL HODGES, JIM STEVENSON, CHRISO ANDREOU, MONICA LANYADO, MICHELLE NEW, BRYM WILLIAMS Y DEAN MCMILLAN, *Risk factors for development of sexually abusive behaviour in sexually victimised adolescent boys: cross sectional study*, British Medical Journal **317** (1998), 175-179.

43. L. A. SMOLIN, K. F. CLARK, J. G. THOENE, W. A. GAHL Y J. A. SCHNEIDER, A *comparison of the effectiveness of cysteamine and phosphocysteamine in elevating plasma cysteamine concentration and decreasing leukocyte free cystine in nephropathic cystinosis*, Pediatric Research **23** (1988), 616-620.

44. STANLEY S. STEVENS, *On the theory of scales of measurement*, Science **161** (1946), 677-680.

45. STUDENT, *The probable error of a mean*, Biometrika **6** (1908), 1-25.

46. THE CONTINUOUS INFUSION VERSUS DOUBLE-BOLUS ADMINISTRATION OF ALTEPLASE (COBALT) INVESTIGATORS, A *comparison of continuous infusion of alteplase with double-bolus administration for acute myocardial infarction*, The New England Journal Of Medicine **337** (1997), 1124-1130.

47. JOHN W. TUKEY, *The problem of multiple comparisons*, Mimeographed monograph, Princeton University, 1953, citado en Yosef Hochberg y Ajit C. Tamhane, *Multiple Comparison Procedures*, John Wiley and Sons, Nueva York, 1987.

48. ANNELIES VAN RIE, ROBIN WARREN, MADELEINE RICHARDSON, THOMAS C. VICTOR, ROBERT P. GIE, DONALD A. ENARSON, NULDA BEYERS y PAUL D. VAN HELDEN, *Exogenous reinfection as a cause of recurrent tuberculosis after curative treatment*, The New England Journal Of Medicine **341** (1999), 1174-1179.

49. FRANK WILCOXON, *Individual comparisons by ranking methods*, Biometrics **1** (1945), 80-83.

Índice analítico

V

W

Z