Universidad Continental

Estadística II

Guía de Trabajo

Visión

Ser una de las 10 mejores universidades privadas del Perú al año 2020, reconocidos por nuestra excelencia académica y vocación de servicio, líderes en formación integral, con perspectiva global; promoviendo la competitividad del país.

Misión

Somos una universidad privada, innovadora y comprometida con el desarrollo del Perú, que se dedica a formar personas competentes, íntegras y emprendedoras, con visión internacional; para que se conviertan en ciudadanos responsables e impulsen el desarrollo de sus comunidades, impartiendo experiencias de aprendizaje vivificantes e inspiradoras; y generando una alta valoración mutua entre todos los grupos de interés.

Universidad Continental
Material publicado con fines de
estudio Código: AAUC 00176

PRESENTACIÓN

. La asignatura de Estadística II está diseñada para proporcionar al estudiante, los conocimientos y habilidades necesarias en la recolección, procesamiento, análisis e interpretación de datos numéricos así como también en el desarrollo de la teoría de la Inferencia Estadística, la cual ha aumentado el alcance de las aplicaciones de la estadística, para una toma de decisiones.

En general, contiene un compendio de guías prácticas para ser desarrolladas de manera (secuencial), está estructurada en cuatro unidades: en la primera unidad se desarrollara Diseño de Experimentos y técnicas de muestreo probabilísticos como introducción a la Inferencia Estadística; en la segunda unidad, Prueba de Hipótesis y Análisis de Varianza; en la tercera unidad, Estadística no Paramétrica ANOVA en la cuarta unidad, Correlación y Regresión.

La elaboración de la presente guía es fruto de la investigación que ha sido enriquecido a partir de la revisión de manuales y libros de la asignatura, utilizando organizadores y demás cuadros para que sea de fácil entendimiento para el estudiante.

Es recomendable que el estudiante antes de desarrollar la guía el material de trabajo desarrolle una permanente lectura de estudio para entender el procedimiento y trabaje con seriedad, teniendo como herramienta el uso de la computadora aplicando SPSS, Excel,y otros software estadísticos.

Al estudiante proporcionamos el presente material con la intensión de guiarlos en su aprendizaje en la asignatura pero así mismo debe ser complementada con la biografía propuesta en el silabo del curso.

El equipo de docentes

ÍNDICE

PRESENTACIÓN 3
ÍNDICE 4
PRIMERA UNIDAD 5
SEGUNDA UNIDAD 51
TERCERA UNIDAD 109
CUARTA UNIDAD 181
RECURSOS DIGITALES 249

PRIMERA UNIDAD
INTRODUCCION A LA INFERENCIA ESTADISTICA

- GUÍA DE PRÁCTICA N° 1: Diseño de experimentos y técnicas de muestreo
- GUÍA DE PRÁCTICA Nº 2: Estimaciones y tamaño de muestra
- GUÍA DE PRÁCTICA Nº 3: Estimaciones y tamaño de muestra Media Poblacional
- GUÍA DE PRÁCTICA N° 4: Intervalo de confianza de varianza y desviación standar

TEXTO N° 1

Diseño de experimentos y técnicas de muestreo probabilístico.

Compilado y adaptado de Mario Triola Estadística PEARSON; 2009

Págs. De 21 al 30

1-4 Diseño de experimentos

Concepto clave Esta sección contiene mucha información y muchas definiciones. De todas las definiciones, el concepto de una *muestra aleatoria simple* es especialmente importante, por el papel que tiene en el resto de este libro y en la estadística en general, de modo que asegúrese de entender muy bien su definición. Además, comprenda con claridad que el método que se utilice para reunir datos es crítico y absolutamente importante. Reconozca que, para diseñar experimentos, se requiere de una gran reflexión y cuidado para asegurarse de obtener resultados válidos.

Si los datos muestrales no se reúnen de manera adecuada, éstos podrían resultar inútiles por completo, de tal forma que ninguna cantidad de tortura estadística los salvaría.

Los métodos estadísticos se rigen por los datos. Por lo regular obtenemos datos de dos fuentes distintas: los *estudios observacionales* y los *experimentos*.

Definiciones

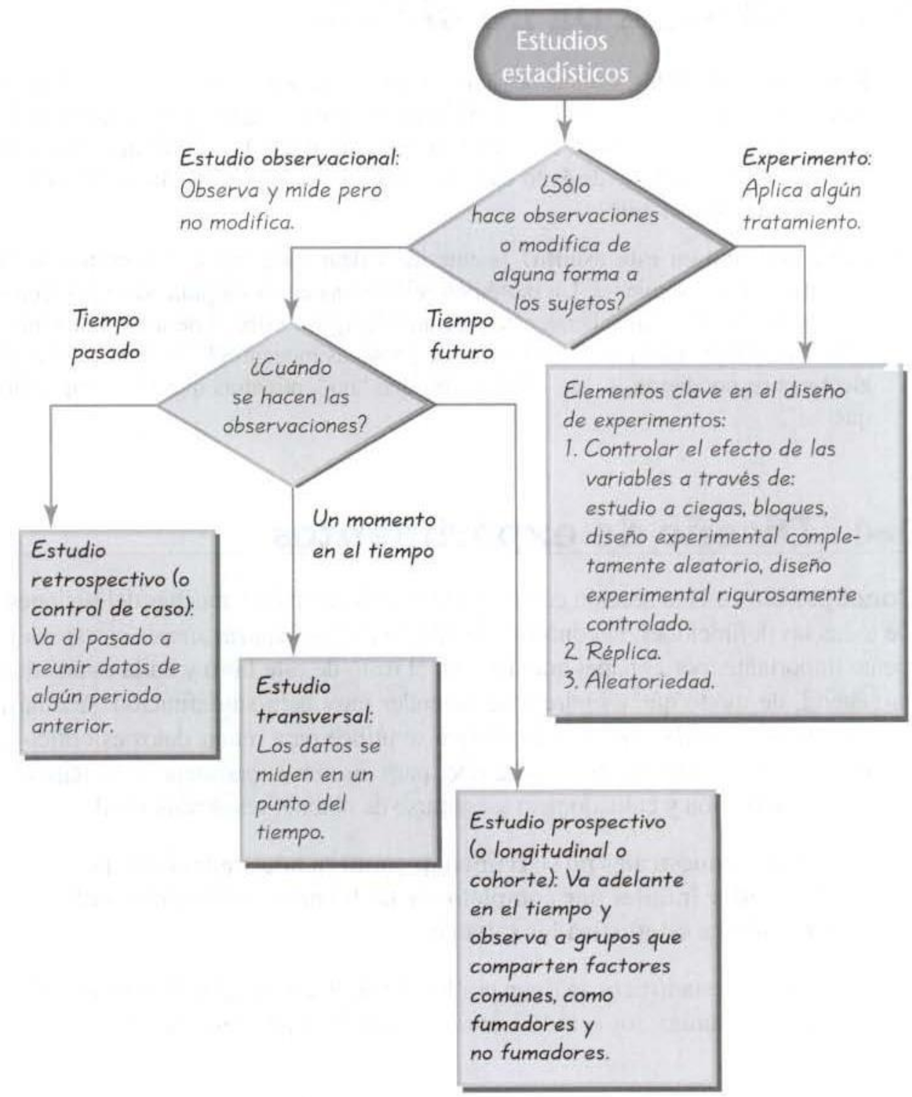
En un **estudio observacional**, vemos y medimos características específicas, pero no intentamos *modificar* a los sujetos que estamos estudiando.

En un **experimento** aplicamos algunos *tratamientos* y luego procedemos a observar sus efectos sobre los sujetos (en los experimentos, a los sujetos se les denomina **unidades experimentales**).

Tipos de estudios observacionales Una encuesta Gallup es un buen ejemplo de un estudio observacional, en tanto que el ensayo clínico del fármaco Lipitor es un buen ejemplo de un experimento. La encuesta Gallup es observacional en el sentido de que simplemente se observan personas (a menudo haciendo entrevistas) sin modificarlas de ninguna forma. Sin embargo, el ensayo clínico de Lipitor implica el tratamiento de algunas personas con el fármaco, de manera que se modifica a los individuos tratados. Existen distintos tipos de estudios observacionales, como se ilustra en la figura 1-3. Estos términos, que suelen utilizarse en muchas revistas científicas diferentes, se definen a continuación.

Figura 1-3

Estudios estadísticos



Definiciones

En un **estudio transversal**, los datos se observan, miden y reúnen en un solo momento.

En un **estudio retrospectivo** (o **de control de caso**), los datos se toman del pasado (mediante el examen de registros, entrevistas y otros).

En un **estudio prospectivo** (o **longitudinal** o **de cohorte**), los datos se reunirán en el futuro y se toman de grupos (llamados *cohortes*) que comparten factores comunes.

Hay una diferencia importante entre el muestreo realizado en estudios retrospectivos y estudios prospectivos. En los estudios retrospectivos regresamos en el tiempo a reunir datos acerca de características resultantes de interés, como un grupo de conductores que murieron en accidentes automovilísticos y otro grupo de conductores que no murieron en este tipo de accidentes. En los estudios prospectivos recurrimos al futuro siguiendo grupos con un factor potencialmente causal y grupos que no lo tienen, como un grupo de conductores que utilizan teléfonos celulares y un grupo de conductores que no los utilizan.

Las tres definiciones anteriores se aplican a los estudios observacionales. No obstante, por ahora nos enfocaremos en los experimentos. Los resultados de los experimentos en ocasiones se estropean debido a la *confusión*.

Definición

La **confusión** ocurre en un experimento cuando uno no es capaz de distinguir entre los efectos de diferentes factores.

Trate de planear el experimento para que no se presente confusión.

Por ejemplo, suponga que un profesor de Vermont experimenta con una nueva política de asistencia ("su calificación promedio en el curso bajará un punto por cada clase que falte"); sin embargo, llega un invierno excepcionalmente benigno, sin nieve y sin temperaturas muy frías, que son factores que habían limitado la asistencia en años anteriores. Si la asistencia mejora, no podemos determinar si ello se debe a la nueva política de asistencia o al invierno benigno. Los efectos de la política de asistencia y del clima se han confundido. Generalmente es muy importante controlar los efectos de las variables. Además de la confusión, los experimentos también se pueden arruinar por otros factores, como el hecho de no lograr reunir una muestra que sea representativa de la población. En general, la organización de los experimentos requiere de un gran cuidado y una extensa planeación. En la figura 1-3 se observa que los siguientes tres aspectos son muy importantes para el diseño de experimentos:

1. Control de los efectos de las variables.
2. Uso de la réplica.
3. Empleo de la aleatoriedad.

Ahora nos enfocaremos en el papel que juegan estos tres factores en el diseño de experimentos.

Control de los efectos de las variables

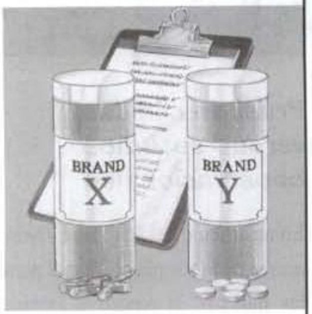
En la figura 1-3 se muestra que uno de los elementos clave en el diseño de experimentos es el control de los efectos de las variables. Se adquiere control al utilizar dispositivos como un estudio a ciegas, un diseño de bloques aleatorio, un diseño experimental completamente aleatorio o un diseño experimental rigurosamente controlado.

Estudio a ciegas En 1954 se diseñó un experimento masivo para probar la efectividad de la vacuna de Salk para prevenir la poliomielitis que había matado o paralizado a miles de niños. En ese experimento, a un grupo de tratamiento se le administró la vacuna real de Salk; mientras que a un segundo grupo se le dio un placebo que no contenía ningún fármaco. En experimentos que implican el uso de

LA ESTADÍSTICA EN LAS NOTICIAS

Pruebas clínicas versus estudios observacionales

En un artículo del *New York Times* acerca de la terapia hormonal para las mujeres, la reportera Denise Grady escribió acerca de un reporte de tratamientos probados en ensayos clínicos controlados aleatorios. Ella declaró que "pruebas como ésta, donde los pacientes se designan al azar para un tratamiento o un placebo, se consideran el estándar por excelencia en la investigación médica. En cambio, los estudios observacionales, donde los pacientes deciden por sí mismos si toman un fármaco, se consideran menos confiables [...] Los investigadores dicen que los estudios observacionales tal vez han dado una falsa imagen color de rosa del reemplazo hormonal, ya que las mujeres que optan por recibir los tratamientos son más saludables y tienen mejores hábitos, que las mujeres que no lo hacen".



Los efectos Hawtorne y del experimentador

El conocido efecto placebo ocurre cuando un sujeto no tratado cree incorrectamente que está recibiendo un tratamiento real, y reporta una mejoría en sus síntomas. El efecto Hawthorne ocurre cuando, por alguna razón, los sujetos tratados responden de manera diferente por el simple hecho de formar parte del experimento. (Este fenómeno se denominó "efecto Hawthorne" porque se observó por primera vez en un estudio realizado con obreros en la planta Hawthorne, de Western Electric). Ocurre un efecto del experimentador (a veces llamado efecto Rosenthall) cuando el investigador o experimentador influye involuntariamente en los sujetos mediante factores como la expresión facial, el tono de voz o la actitud.

placebos, a menudo existe un **efecto placebo** que ocurre cuando un sujeto no tratado reporta una mejoría en los síntomas. (La mejoría reportada en el grupo placebo puede ser real o imaginaria). El efecto placebo se puede disminuir o evitar usando un **estudio a ciegas**, una técnica en que el sujeto no sabe si está recibiendo un tratamiento o un placebo. El estudio a ciegas nos permite determinar si el efecto del tratamiento es significativamente diferente del efecto placebo. El experimento de la poliomielitis fue un **estudio a ciegas doble**, lo que quiere decir que el estudio se realizó en dos niveles: **1.** los niños inyectados no sabían si estaban recibiendo la vacuna de Salk o un placebo, y **2.** los médicos que aplicaron las inyecciones y evaluaron los resultados tampoco lo sabían.

Bloques Cuando se diseña un experimento, con frecuencia sabemos de antemano que existen algunos factores que pueden tener un efecto importante sobre las variables que se están considerando. Por ejemplo, al diseñar un experimento para probar la eficacia de un nuevo fertilizante sobre el crecimiento de los árboles, sabemos que el contenido de humedad del suelo (seco o húmedo) puede afectar el crecimiento de los árboles. En este caso, debemos utilizar bloques. Un **bloque** es un conjunto de sujetos que son similares, pero los bloques son diferentes en aspectos que pueden afectar el resultado del experimento. La prueba de los efectos de un fertilizante podría incluir un bloque de árboles en tierra seca y un bloque de árboles en tierra húmeda. Al diseñar un experimento para probar la eficacia de un nuevo fármaco para enfermedades cardiacas, podríamos crear un bloque de hombres y un bloque de mujeres, ya que se sabe que el corazón de los hombres y el de las mujeres se comporta de manera diferente. Después de identificar los bloques, se procede a asignar *aleatoriamente* los tratamientos a los sujetos de cada bloque.

Diseño de bloque aleatorio: **Utilice este diseño experimental si está realizando un experimento para probar uno o más tratamientos diferentes, si existen distintos grupos de sujetos similares y si los grupos difieren en aspectos que podrían afectar las respuestas a los tratamientos:**

1. **Forme bloques (o grupos) de sujetos con características similares.**
2. **Asigne los tratamientos de manera aleatoria a los sujetos dentro de cada bloque.**

Por ejemplo, suponga que desea probar la eficacia de un fertilizante sobre el peso de 12 álamos, y que tiene dos terrenos: uno de los terrenos es seco y el otro es húmedo. Un diseño experimental inadecuado utilizaría el fertilizante en 6 árboles plantados en suelo húmedo; mientras que los 6 árboles sin tratamiento se plantarían en terreno seco. De esta forma no sabríamos si las diferencias se deben al tratamiento con el fertilizante o al tipo de tierra o a ambos. Véase la figura 1-4*a*). Sería mucho mejor plantar 6 árboles en el terreno seco y 6 árboles en el terreno húmedo. Luego, se usaría la aleatoriedad para identificar 3 árboles que se tratarían con el fertilizante y 3 que no recibirían el tratamiento, como se observa en la figura 1-4*b*). Los resultados nos permitirían ver la eficacia del fertilizante. (La elección de una muestra con un tamaño de 12 árboles es totalmente arbitraria. Un mayor número de árboles produciría mejores resultados. En capítulos posteriores abordamos el tema de la determinación del tamaño de la muestra. Asimismo, el experimento debe diseñarse de forma cuidadosa, para que el fertilizante que se aplique a uno de los árboles no se esparza a los árboles adyacentes. Otros factores importantes como el riego, la temperatura y la luz del sol se deben controlar cuidadosamente. El diseño experimental

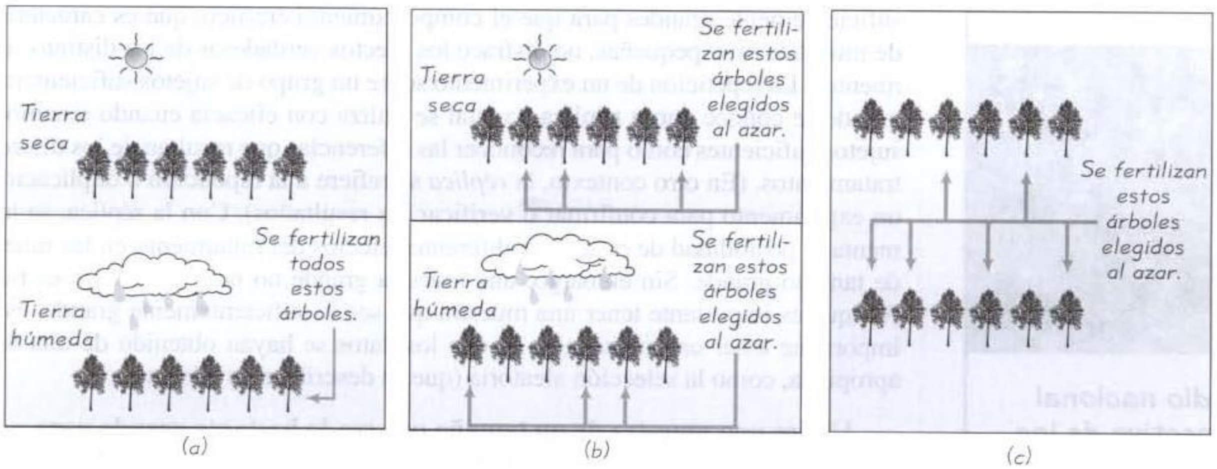


Figura 1-4 Diseño de un experimento con diferentes tratamientos

(a) Diseño experimental inadecuado: Utilizar el fertilizante para tratar a todos los árboles de la tierra húmeda y no utilizar el fertilizante para los árboles de la tierra seca. No habrá forma de saber si los resultados se deben al tratamiento o al tipo de tierra.

(b) Diseño de bloques aleatorio: Formar un bloque de árboles de tierra húmeda y otro bloque de árboles de tierra seca. Dentro de cada bloque se utiliza la aleatoriedad para determinar cuáles se tratarán con el fertilizante y cuáles no.

(c) Diseño experimental completamente aleatorio: De manera aleatoria se determina cuáles árboles se tratarán con el fertilizante y cuáles no.

general requiere de mucha mayor reflexión y cuidado de lo que aquí se ha descrito. Tomar un curso sobre el diseño de experimentos sería una forma muy adecuada de aprender mucho más de lo que podemos describir aquí).

Diseño experimental completamente aleatorio En un **diseño experimental completamente aleatorio**, los sujetos se asignan a distintos grupos de tratamiento mediante un proceso de *selección aleatoria*. Observe la figura 1-4c), donde se muestra que, de los 12 árboles, se seleccionan 6 al azar para el tratamiento con el fertilizante. Otro ejemplo de un diseño experimental completamente aleatorio es el experimento de la poliomielitis: Los niños se asignaron al grupo de tratamiento o al grupo placebo, mediante un proceso de selección aleatoria (equivalente a lanzar una moneda).

Diseño rigurosamente controlado Otro método para signar sujetos a los tratamientos consiste en utilizar el **diseño rigurosamente controlado**, donde los sujetos se *eligen cuidadosamente*, de manera que quienes reciban cada tratamiento sean similares en los aspectos que son importantes para el experimento. En un experimento para probar la efectividad de un fármaco diseñado para disminuir la presión sanguínea, si el grupo placebo incluye a un hombre de 30 años de edad con sobrepeso, fumador, que consume grandes cantidades de bebidas alcohólicas, sal y grasas, el grupo de tratamiento también debería incluir a una persona con características similares (lo cual, en este caso, sería fácil de conseguir).

Réplica y tamaño de la muestra

Además de controlar los efectos de las variables, otro elemento fundamental del diseño experimental es el tamaño de las muestras. Las muestras deben ser lo

Estudio nacional prospectivo de los niños

Un buen ejemplo de un estudio prospectivo es el National Children's Study que comenzó en 2005, en el cual se realiza un seguimiento de 100,000 niños, desde el nacimiento hasta los 21 años de edad. Los niños son de 96 regiones geográficas diferentes. Su objetivo consiste en mejorar la salud de los niños al identificar los efectos de factores ambientales como la dieta, la exposición a sustancias químicas, la vacunación, las películas y la televisión. Se planea que alrededor de 2008 haya algunos resultados preliminares. El estudio abordara preguntas como éstas: ¿De qué manera interactúan los genes y el ambiente para fomentar o prevenir la conducta violenta en los adolescentes? ¿La falta de ejercicio y una dieta inadecuada son las únicas razones del sobrepeso de muchos niños? ¿Las infecciones afectan el progreso del desarrollo, el asma, la obesidad y las enfermedades cardiacas? ¿De qué manera la planeación y construcción de las ciudades y de los vecindarios fomentan o disminuyen las lesiones?

suficientemente grandes para que el comportamiento errático, que es característico de muestras muy pequeñas, no disfrace los efectos verdaderos de los distintos tratamientos. La repetición de un experimento sobre un grupo de sujetos suficientemente grande se conoce como **réplica**, la cual se utiliza con eficacia cuando tenemos los sujetos suficientes como para reconocer las diferencias que resultan de los diferentes tratamientos. (En otro contexto, la *réplica* se refiere a la repetición o duplicación de un experimento para confirmar o verificar los resultados). Con la réplica, se incrementa la posibilidad de reconocer diferentes efectos del tratamiento en las muestras de tamaño grande. Sin embargo, una muestra grande no necesariamente es buena. Aunque es importante tener una muestra que sea lo suficientemente grande, es más importante tener una muestra en la que los datos se hayan obtenido de una forma apropiada, como la selección aleatoria (que se describe más adelante).

Utilice una muestra de un tamaño que sea lo bastante grande para distinguir la verdadera naturaleza de cualquiera de los efectos, y obtenga la muestra utilizando un método apropiado, como uno basado en la *aleatoriedad*.

En el experimento diseñado para probar la vacuna de Salk, a 200,000 niños se les administró la vacuna real de Salk y a otros 200,000 niños se les dio un placebo. Fue posible observar la eficacia de la vacuna porque en realidad en el experimento real se utilizaron muestras de un tamaño lo suficientemente grande. No obstante, aun cuando los grupos placebo y de tratamiento fueron muy grandes, el experimento habría sido un fracaso si los sujetos no se hubieran asignado a los dos grupos para que fueran similares en los aspectos importantes para el experimento.

Aleatoriedad y otras estrategias de muestreo

En la estadística, como en la vida, uno de los peores errores consiste en reunir datos de una forma inapropiada. Insistiremos en este punto tan importante:

Si los datos muestrales no se reúnen de forma adecuada, resultarían tan inútiles que ninguna cantidad de tortura estadística podrá salvarlos.

En la sección 1-3 vimos que una muestra de respuestas voluntarias es aquella donde los sujetos deciden por sí mismo si responden o no. Este tipo de muestras son muy comunes, aunque sus resultados suelen ser inútiles para hacer inferencias válidas acerca de poblaciones más grandes.

Ahora definiremos algunos de los procedimientos de muestreo más comunes.

Definiciones

En una **muestra aleatoria** los miembros de la población se seleccionan de forma que cada *miembro individual* tenga la misma posibilidad de ser elegido.

Una **muestra aleatoria simple** de n sujetos se selecciona de manera que cada posible *muestra del mismo tamaño* n tenga la misma posibilidad de ser elegida.

Una **muestra probabilística** implica seleccionar miembros de una población de forma que cada miembro tenga una posibilidad conocida (aunque no necesariamente la misma) de ser elegido.

EJEMPLO Muestra aleatoria, muestra aleatoria simple, muestra probabilística Imagine un salón de clases con 60 estudiantes acomodados en seis filas de 10 estudiantes cada una. Suponga que el profesor selecciona una muestra de 10 estudiantes lanzando un dado y seleccionando la fila correspondiente al resultado. ¿El resultado es una muestra aleatoria? ¿Una muestra aleatoria simple? ¿Una muestra probabilística?

SOLUCIÓN La muestra es aleatoria porque cada estudiante individual tiene la misma posibilidad (una posibilidad en seis) de resultar seleccionado. La muestra *no* es aleatoria simple porque no todas las muestras de tamaño 10 tienen la misma posibilidad de ser escogidas. Por ejemplo, este diseño muestral de usar un dado para seleccionar una fila hace imposible la selección de 10 estudiantes que estén en filas diferentes (aunque hay una posibilidad en seis de seleccionar la muestra que consiste en los 10 estudiantes de la primera fila). Se trata de una muestra probabilística porque cada estudiante tiene una posibilidad conocida (una posibilidad en seis) de ser elegido.

Importante: A lo largo de este libro, utilizaremos una variedad de procedimientos estadísticos diferentes, y muchas veces tendremos como requisito reunir una *muestra aleatoria simple*, como se definió anteriormente. Con el muestreo aleatorio esperamos que todos los componentes de la población estén representados (aproximadamente) de manera proporcional. Las muestras aleatorias se seleccionan usando muchos procedimientos diferentes, incluyendo el uso de computadoras para generar números aleatorios. (Antes de la llegada de las computadoras, generalmente se usaban las tablas de números aleatorios. Si quiere leer algo verdaderamente interesante, consulte el libro *A Million Random Digits*, publicado por Free Press, que contiene un millón de dígitos generados aleatoriamente). A diferencia del muestreo realizado con descuido o por casualidad, el muestreo aleatorio exige una planeación y ejecución muy cuidadosas.

Además del muestreo aleatorio, existen otras técnicas de muestreo en uso, y aquí se describen las más comunes. Observe la figura 1-5, que describe los diferentes métodos de muestreo. No olvide que en el resto del libro sólo se utilizarán el muestreo aleatorio y el muestreo aleatorio simple.

Definiciones

En el **muestreo sistemático**, elegimos algún punto de partida y luego seleccionamos cada *k*-ésimo (por ejemplo, cada quincuagésimo) elemento en la población.

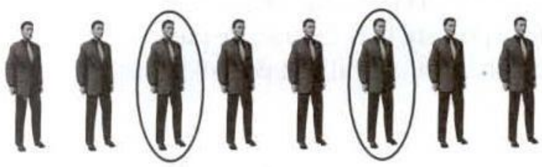
En el **muestreo de conveniencia**, simplemente se utilizan resultados que sean muy fáciles de obtener.

En el **muestreo estratificado** subdividimos a la población en al menos dos subgrupos (o estratos) diferentes, de manera que los sujetos que pertenecen al mismo subgrupo compartan las mismas características (como el género o la categoría de edad), y luego obtenemos una muestra de cada subgrupo (o estrato).

En el **muestreo por conglomerados** primero dividimos el área de la población en secciones (o conglomerados), y luego elegimos al azar algunos de estos conglomerados, y después elegimos *a todos* los miembros de los conglomerados seleccionados.

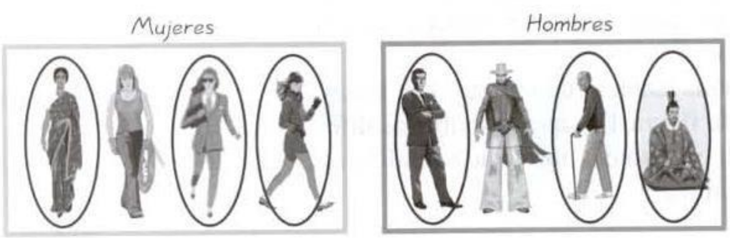
Muestreo aleatorio:
Cada miembro de la población tiene la misma probabilidad de resultar seleccionado. A menudo se usan computadoras para generar números telefónicos aleatorios.

Muestreo aleatorio simple:
Se selecciona una muestra de n sujetos, de manera que cada posible muestra del mismo tamaño n tenga la misma posibilidad de ser elegida.



Muestreo sistemático:
Se selecciona un punto de partida, después se elige cada k-ésimo (por ejemplo, cada quincuagésimo) elemento de la población.

Muestreo de conveniencia:
Se utilizan resultados que son fáciles de obtener.



Muestreo estratificado:
Se subdivide a la población en al menos dos subgrupos (o estratos diferentes), de manera que los sujetos del mismo subgrupo compartan las mismas características (como el género o la categoría de edad), y después se obtiene una muestra de cada subgrupo.

Muestreo por conglomerados:
Se divide el área de la población en secciones (o conglomerados), luego se eligen al azar algunos de estos conglomerados, y después se elige a todos los miembros de los conglomerados seleccionados.

Es fácil confundir el muestreo estratificado y el muestreo por conglomerados, ya que ambos implican la formación de subgrupos. Sin embargo, el muestreo por conglomerados usa a *todos* los miembros de una *muestra* de conglomerados; en tanto que el muestreo estratificado emplea una *muestra* de los miembros de *todos* los estratos. Un ejemplo de muestreo por conglomerados es una encuesta previa a las elecciones, donde se seleccionan aleatoriamente 30 distritos electorales de un número mayor de distritos, y luego se encuesta a todas las personas de cada uno de esos distritos elegidos. Este método es mucho más rápido y mucho menos costoso que elegir a una persona de cada uno de los muchos distritos del área de la población. Los resultados del muestreo estratificado o por conglomerados se ajustan o ponderan para corregir cualquier representación desproporcionada de los grupos.

Para una muestra de tamaño fijo, si usted selecciona al azar sujetos de diferentes estratos, es probable que obtenga resultados más consistentes (y menos variables), que si simplemente seleccionó una muestra al azar de la población general. Por tal razón, el muestreo estratificado se utiliza con frecuencia para reducir la variación en los resultados. Muchos de los procedimientos que se analizarán después en este libro tienen como requisito que los datos muestrales formen una *muestra aleatoria simple*, y ni el muestreo estratificado ni el muestreo por conglomerados satisfacen tal requisito.

Muestreo de etapas múltiples La figura 1-5 ilustra cinco procedimientos de muestreo comunes; sin embargo, los encuestadores profesionales y los investigadores gubernamentales a menudo recolectan datos utilizando cierta combinación de los cinco métodos. Un **diseño de muestreo de etapas múltiples** implica la selección de una muestra en diferentes pasos, los cuales suelen incluir distintos procedimientos de muestreo. Consideremos las estadísticas gubernamentales sobre el desempleo que se obtienen de hogares encuestados. No es práctico hacer visitas personales a cada miembro de una muestra aleatoria simple, ya que los hogares individuales están distribuidos por todo el país. Sería demasiado costoso en tiempo y en dinero encuestar a cada miembro de una muestra aleatoria simple. En cambio, el U.S. Census Bureau y el Bureau of Labor Statistics se combinan para realizar una encuesta llamada Current Population Survey, que se utiliza para obtener datos que describen factores tales como la tasa de desempleo, la matrícula de universidades y los salarios semanales. La encuesta incorpora un diseño de etapas múltiples, que se describe de manera general a continuación:

1. Todo el país (Estados Unidos) se divide en 2007 regiones diferentes llamadas *unidades muestrales primarias* (UMP). Las unidades muestrales primarias son áreas metropolitanas, grandes condados o grupos de condados más pequeños.
2. En cada uno de los 50 estados, se selecciona una muestra de unidades muestrales primarias. Para la Current Population Survey se utilizan 792 de las unidades muestrales primarias. (Se usan las 432 unidades muestrales primarias de las poblaciones más grandes, y se seleccionan al azar 360 unidades muestrales primarias de las 1575 restantes).
3. Cada una de las 792 unidades muestrales primarias elegidas se divide en bloques, y se utiliza un muestreo estratificado para seleccionar una muestra de bloques.
4. En cada bloque seleccionado, se identifican conglomerados de hogares cercanos entre sí. Los conglomerados se eligen al azar, y se entrevista a todos los hogares de los conglomerados elegidos.

Observe que este diseño de muestreo de etapas múltiples incluye los muestreos aleatorio, estratificado y por conglomerados en diferentes etapas. El resultado final

¿Debe usted creer en un estudio estadístico?

En el libro *Statistical Reasoning for Everyday Life*, segunda edición, los autores Jeff Bennett, William Briggs y Mario Triola proporcionan los siguientes ocho lineamientos para evaluar de forma crítica un estudio estadístico. **1.** Identificar el objetivo del estudio, la población considerada y el tipo de estudio. **2.** Considerar la fuente, especialmente con respecto a la posibilidad de un sesgo. **3.** Analizar el procedimiento de muestreo. **4.** Buscar problemas en la definición o medida de las variables de interés. **5.** Tener cuidado con variables confusas que podrían invalidar las conclusiones, **6.** Considerar el contexto y la redacción de cualquier encuesta. **7.** Verificar que las gráficas representen los datos de forma adecuada y que las conclusiones estén justificadas. **8.** Considerar si las conclusiones logran los objetivos del estudio, si tienen sentido y si poseen un significado práctico.

es un diseño de muestreo complejo; no obstante, es mucho más práctico y menos costoso que utilizar un diseño más sencillo, como el muestreo aleatorio simple.

Errores de muestreo

No importa lo bien que usted planee y ejecute el proceso de recolección de muestras, es probable que ocurra algún error en los resultados. Por ejemplo, seleccione 1000 adultos al azar, pregúnteles si se graduaron de bachillerato y registre el porcentaje de respuestas afirmativas en la muestra. Si usted elige otra muestra de 1000 adultos al azar, es probable que obtenga un porcentaje *diferente* en esa muestra.

Definiciones

Un **error de muestreo** es la diferencia entre el resultado de una muestra y el verdadero resultado de la población; este error es consecuencia de las fluctuaciones por el azar.

Un **error que no es de muestreo** sucede cuando los datos muestrales se obtienen, registran o analizan de forma incorrecta (como cuando se selecciona una muestra sesgada, cuando se usa un instrumento de medición defectuoso o cuando se copian los datos de forma incorrecta).

Si recolectamos con cuidado una muestra, de manera que sea representativa de la población, entonces podemos utilizar los procedimientos que se describen en este libro para analizar el error muestral; sin embargo, debemos ser muy cuidadosos de no minimizar el error que no es de muestreo.

En el "Problema del capítulo", nos preguntamos si los datos experimentales originales que derivaron el concepto de los "seis grados de separación" eran datos *buenos*. Señalamos que Stanley Milgram envió 60 cartas a individuos de Wichita, Kansas, y que les pidió que reenviaran a las cartas a una mujer específica en Cambridge, Massachusetts. Participaron 50 de los 60 sujetos, pero sólo tres cartas llegaron al objetivo. Dos experimentos posteriores también mostraron bajas tasas de terminación, aunque finalmente Milgram logró una tasa de terminación del 35%, y calculó que el número promedio de intermediarios era de alrededor de 6. Los experimentos de Milgram tuvieron grandes fallas en la recolección de datos muestrales, ya que éstos no eran representativos de la población de residentes de Estados Unidos. El experimento original incluyó a pobladores de Wichita y sus tasas de respuesta y terminación fueron muy bajas. Incluso con una tasa de terminación del 35%, bien sería posible que sólo hayan participado los individuos interesados en el experimento, los cuales tal vez conocían a más personas, dando como resultado un número menor de intermediarios. Sus anuncios fueron diseñados para reclutar personas que estuvieran "orgullosas de sus habilidades sociales y que tuvieran confianza en su capacidad de encontrar a alguien atravesando las barreras de clase". Los sujetos reclutados eran de una región geográfica específica, eran más sociables y más adinerados que la gente común, y la tasa de respuesta fue muy baja. En consecuencia, el resultado de 6 como el número de grados de separación no estaba justificado por los datos (esto no necesariamente significa que el valor de 6 sea incorrecto). El diseño experimental tenía graves errores y cualquier conclusión basada en los resultados sería muy sospechosa. Por desgracia, no se ha realizado un experimento bien diseñado para verificar si el número 6 es correcto o incorrecto. (En 2001, Duncan Watts, de Columbia University, inició un experimento por correo electrónico a gran escala; sin embargo, las personas que utilizan el correo electrónico no necesariamente son representativas de la población mundial. No obstante, se aprende mucho de sus resultados).

TEXTO N° 2

Estimación de la proporción de una población.

Determinación del tamaño muestral

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 320 a 326, 328

7-1 Panorama general

En este capítulo comenzaremos a trabajar con el verdadero núcleo de la estadística inferencial al utilizar datos muestrales para hacer inferencias acerca de poblaciones. En específico, usaremos datos muestrales para hacer estimaciones de parámetros de población. Por ejemplo, el problema del capítulo se refiere a terapeutas de contacto que identificaron correctamente el campo de energía humano únicamente en el 44% de 280 ensayos. Con base en el estadístico muestral de 44%, estimaremos el porcentaje de aciertos para toda la población de los terapeutas de contacto.

Las dos aplicaciones principales de la estadística inferencial implican el uso de datos muestrales para **1.** estimar el valor de un parámetro de una población y **2.** probar alguna aseveración (o hipótesis) acerca de una población. En este capítulo presentamos métodos para estimar valores de estos importantes parámetros de población: proporciones, medias y varianzas. También presentamos métodos para determinar los tamaños de muestra necesarios para estimar estos parámetros. En el capítulo 8 estudiaremos los métodos básicos para probar aseveraciones (o hipótesis) que se hacen acerca de un parámetro de una población.

7-2 Estimación de la proporción de una población

Concepto clave En esta sección presentamos métodos importantes donde se utiliza una proporción muestral para estimar el valor de una proporción poblacional con un *intervalo de confianza*. También presentamos métodos para calcular el tamaño muestral necesario para estimar una proporción poblacional. En esta sección se explican conceptos generales que se utilizan en las siguientes secciones y en los capítulos posteriores, por lo que es muy importante su comprensión.

Una estrategia de estudio: El tiempo dedicado a esta sección estará bien invertido, ya que presentamos el concepto de intervalo de confianza, un concepto general que se aplicará en las siguientes secciones del capítulo. Sugerimos que primero lea esta sección con el único objetivo de tratar de comprender qué son los intervalos de confianza, para qué sirven y por qué son indispensables. Segundo, trate de desarrollar la habilidad de construir estimaciones del intervalo de confianza de las proporciones de una población. Tercero, aprenda a interpretar correctamente un intervalo de confianza. Cuarto, lea la sección una vez más y trate de comprender la teoría subyacente. Usted tendrá mucho más éxito si comprende lo que está haciendo, en vez de aplicar a ciegas y de manera mecánica los pasos para obtener una respuesta que podría o no tener sentido.

En esta sección sólo consideraremos casos en los que es posible utilizar la distribución normal para aproximar la distribución muestral de proporciones de muestra. Los siguientes requisitos se aplican a los métodos de esta sección.

Requisitos

1. La muestra es aleatoria simple.
2. Las condiciones para la distribución binomial se satisfacen. Esto es, hay un número fijo de ensayos, los ensayos son independientes, hay dos categorías de resultados y las probabilidades permanecen constantes para cada ensayo. (Véase la sección 5-3).
3. Existen al menos 5 éxitos y al menos 5 fracasos. (Cuando p y q se desconocen, estimamos sus valores utilizando la proporción muestral, de manera que este

requisito es una forma de verificar que $np \geq 5$ y $nq \geq 5$ se cumplan para que la distribución normal sea una aproximación adecuada para la distribución binomial. Además, existen procedimientos para tratar situaciones en las que la distribución normal no es una aproximación adecuada. Véase el ejercicio 51).

Notación para proporciones

p = proporción de la *población*

$\hat{p} = \dfrac{x}{n}$ = proporción muestral de x *éxitos* en una muestra de tamaño n

$\hat{q} = 1 - \hat{p}$ = proporción muestral de *fracasos* en una muestra de tamaño n

Proporción, probabilidad y porcentaje Esta sección se enfoca en la proporción poblacional p, aunque también podemos trabajar con probabilidades o porcentajes. Cuando trabaje con un porcentaje, expréselo en forma decimal. (Por ejemplo, exprese el 44% como 0.44, de manera que $\hat{p} = 0.44$). Si desea estimar una proporción poblacional con un solo valor, el mejor estimado es $\hat{p}$. Puesto que $\hat{p}$ consiste en un solo valor, se denomina estimado puntual.

Definición

Un **estimado puntual** es un valor individual (o punto) que se usa para aproximar un parámetro de población.

La proporción muestral $\hat{p}$ es el mejor estimado puntual en la proporción poblacional p.

Usamos $\hat{p}$ como el estimado puntual de p, ya que no está sesgado y es el más consistente de los estimadores que podrían usarse. No está sesgado en el sentido de que la distribución de las proporciones muestrales tiende a concentrarse alrededor del valor de p; esto es, las proporciones muestrales $\hat{p}$ no tienden sistemáticamente a subestimar ni a sobreestimar p. (Véase la sección 6-4). La proporción muestral $\hat{p}$ es el estimador más consistente en el sentido de que la desviación estándar de las proporciones muestrales tiende a ser menor que la desviación estándar de cualquier otro estimador sin sesgo.

PROBLEMA 7 DEL CAPÍTULO

EJEMPLO Tasa de éxito de la terapia de contacto En el problema del capítulo señalamos que, en 280 ensayos con terapeutas de contacto, la mano correcta fue elegida en 123 ensayos, de manera que la tasa de éxito es $\hat{p} = 123/280 = 0.44$. Utilice estos resultados de prueba para calcular el mejor estimado puntual de la proporción de todas las selecciones correctas que resultarían si se pusiera a prueba a todos los terapeutas de contacto.

SOLUCIÓN Puesto que la proporción muestral es el mejor estimado puntual de la proporción poblacional, concluimos que el mejor estimado puntual de p es 0.44. (Utilizando el sentido común, podríamos examinar el diseño del experimento y concluir que la verdadera proporción poblacional es 0.5, pero si sólo utilizamos los resultados muestrales encontramos que 0.44 es el mejor estimado de la proporción poblacional p).

¿Por qué necesitamos intervalos de confianza?

En el ejemplo anterior vimos que 0.44 era el *mejor* estimado puntual de la proporción poblacional p, pero no tenemos indicación precisa de qué tan *bueno* es nuestro mejor estimado. Como el estimado puntual tiene el grave defecto de no revelar nada acerca de qué tan bueno es, los especialistas en estadística han diseñado ingeniosamente otro tipo de estimado: el *intervalo de confianza* o *estimado del intervalo*, que consiste en un rango (o un intervalo) de valores en vez de un solo valor.

Definición

Un intervalo de confianza (o estimado del intervalo) es un rango (o un intervalo) de valores que se usa para estimar el valor real de un parámetro de población. El intervalo de confianza suele abreviarse como IC.

Un intervalo de confianza se asocia con un nivel de confianza, como 0.95 (o 95%). El nivel de confianza nos da la tasa de éxitos del procedimiento que se utiliza para construir el intervalo de confianza. El nivel de confianza suele expresarse como la probabilidad o área $1 - \alpha$ (alfa griega minúscula). El valor de α es el complemento del *nivel de confianza*. Para un nivel de confianza de 0.95 (o 95%), $\alpha = 0.05$. Para un nivel de confianza de 0.99 (o 99%), $\alpha = 0.01$.

Definición

El **nivel de confianza** es la probabilidad $1 - \alpha$ (a menudo expresada como el valor de porcentaje equivalente), que es la proporción de veces que el intervalo de confianza realmente contiene el parámetro de población, suponiendo que el proceso de estimación se repite un gran número de veces. El nivel de confianza también se llama **grado de confianza** o **coeficiente de confianza**.

Las opciones más comunes para el nivel de confianza son 90% (con $\alpha = 0.10$), 95% (con $\alpha = 0.05$) y 99% (con $\alpha = 0.01$). La opción del 95% es la más común puesto que provee un buen equilibrio entre precisión (reflejada en el ancho del intervalo de confianza) y confiabilidad (expresada por el nivel de confianza).

A continuación se presenta un ejemplo de un intervalo de confianza basado en los datos muestrales de 280 ensayos de terapeutas de contacto, donde en el 44% de los ensayos se identifica correctamente la mano elegida:

El intervalo de confianza estimado de 0.95 (o 95%) de la proporción poblacional p es $0.381 < p < 0.497$.

Interpretación de un intervalo de confianza

Debemos ser cuidadosos para interpretar los intervalos de confianza correctamente. Existe una interpretación correcta y muchas diferentes y creativas interpretaciones erróneas del intervalo de confianza $0.381 < p < 0.497$.

Muestra pequeña

El Children's Defense Fund se constituyó con el fin de promover el bienestar de los niños. El grupo publicó *Children Out of School in America*, donde se informó que, en una zona, el 37.5% de los jóvenes de 16 y 17 años ya no asistían a la escuela. Esta estadística recibió mucha cobertura por parte de los medios de comunicación masiva, pero se basaba en una muestra de sólo 16 jóvenes. Otra estadística se basó en una muestra de sólo tres estudiantes. (Véase "Firsthand Report: How Flawed Statistics Can Make an Ugly Picture Look Even Worse", *American School Board Journal*, vol. 162).

Correcta: "Tenemos una confianza del 95% de que el intervalo de 0.381 a 0.497 realmente contiene el valor verdadero de p". Esto significa que si seleccionamos muchas muestras diferentes de tamaño 280 y construimos los intervalos de confianza correspondientes, el 95% de ellos incluirían realmente el valor de la proporción poblacional p. (Note que en esta interpretación correcta, el nivel del 95% se refiere a la tasa de éxitos del *proceso*, utilizada para estimar la proporción, y no a la proporción de la población en sí).

Errónea: "Existe un 95% de probabilidades de que el valor real de p esté entre 0.381 y 0.497".

Para cualquier punto específico en el tiempo, una población tiene un valor fijo y constante de p, un intervalo de confianza construido a partir de una muestra que incluye o no a p. De manera similar, si un bebé acaba de nacer y el médico está por anunciar su género, es incorrecto decir que existe una probabilidad de 0.5 de que sea una niña; el bebé es o no una niña, y no hay una probabilidad implicada. Una proporción poblacional p es como el bebé que acaba de nacer: el valor de p es fijo, de manera que los límites del intervalo de confianza contienen o no a p. Por eso es incorrecto decir que existe un 95% de probabilidades de que p se localice entre valores tales como 0.381 y 0.497.

Un nivel de confianza del 95% nos dice que el *proceso* que estamos usando, a la larga, dará por resultado límites del intervalo de confianza que contienen la proporción real de la población el 95% del tiempo. Suponga que la proporción real de todas las identificaciones correctas de la mano por parte de los terapeutas de contacto es $p = 0.5$. Entonces, el intervalo de confianza obtenido de los datos muestrales no incluiría la proporción poblacional, ya que la proporción poblacional real de 0.5 no se encuentra entre 0.381 y 0.497. Esto se ilustra en la figura 7-1, la cual señala los intervalos de confianza típicos que resultan de 20 muestras diferentes. Con un 95% de confianza, esperamos que 19 de las 20 muestras den por resultado intervalos de confianza que contienen el valor real de p; la figura 7-1 ilustra esto con 19 de los intervalos de confianza que contienen p, mientras un intervalo de confianza no contiene p.

Uso de intervalos de confianza para hacer comparaciones *Advertencia: Los intervalos de confianza pueden usarse de manera informal para comparar conjuntos de datos diferentes, pero el traslape de intervalos de confianza no debe usarse para elaborar conclusiones formales y finales acerca de la igualdad de las proporciones.* (Véase "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals" de Schenker y Gentleman, *American Statistician*, vol. 55, núm. 3).

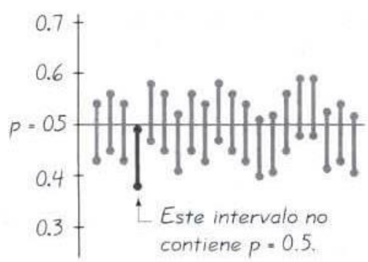


Figura 7-1

Intervalos de confianza de 20 muestras diferentes

Valores críticos

Los métodos de esta sección y muchos de los demás métodos estadísticos que se encuentran en los capítulos siguientes incluyen el uso de una puntuación z estándar, que permite distinguir entre estadísticos muestrales que tienen probabilidades de ocurrir y aquellos que son improbables. Una puntuación z de este tipo se llama *valor crítico* (definido abajo). Los valores críticos se basan en las siguientes observaciones.

1. Sabemos, desde la sección 6-6, que en ciertas condiciones, la distribución muestral de las proporciones muestrales puede aproximarse por una distribución normal, como en la figura 7-2.
2. Las proporciones muestrales tienen una probabilidad relativamente pequeña (denotada por α) de caer en una de las colas sombreadas de gris oscuro de la figura 7-2.
3. Denotando el área de cada cola sombreada como $\alpha/2$, vemos que existe una probabilidad total de α de que una proporción muestral caiga en cualquiera de las dos colas sombreadas de gris oscuro.
4. Por la regla de los complementos (capítulo 4), existe una probabilidad de $1 - \alpha$ de que una proporción muestral caiga dentro de la región interior sombreada de gris claro de la figura 7-2.
5. La puntuación z que separa la región de la cola derecha generalmente se denota por $z_{\alpha/2}$ y se conoce como *valor crítico*, puesto que está en la frontera que separa las proporciones muestrales que tienen probabilidad de ocurrir de aquellas que no tienen probabilidad de ocurrir.

Estas observaciones se formalizan con la notación y definición siguientes.

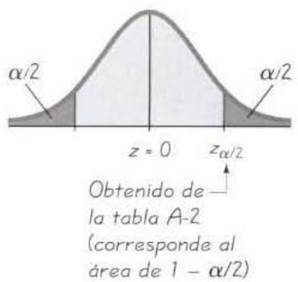


Figura 7-2

Valor crítico $z_{\alpha/2}$ en la distribución normal estándar

Notación para el valor crítico

El valor crítico $z_{\alpha/2}$ es el valor z positivo que está en la frontera vertical que separa una área de $\alpha/2$ en la cola derecha de la distribución normal estándar. (El valor de $-z_{\alpha/2}$ está en la frontera vertical para el área de $\alpha/2$ en la cola izquierda). El subíndice $\alpha/2$ es simplemente un recordatorio de que la puntuación z separa una área de $\alpha/2$ en la cola derecha de la distribución normal estándar.

Definición

Un **valor crítico** es el número en la línea limítrofe que separa estadísticos muestrales que tienen mayor probabilidad de ocurrir de aquellos que no tienen probabilidad de ocurrir. El número $z_{\alpha/2}$ es un valor crítico, una puntuación z con la propiedad de que separa una área de $\alpha/2$ en la cola derecha de la distribución normal estándar (véase la figura 7-2).

EJEMPLO Cálculo de un valor crítico Calcule el valor crítico $z_{\alpha/2}$ que corresponde a un nivel de confianza del 95%.

SOLUCIÓN *Advertencia*: Para calcular el valor crítico z de un nivel de confianza del 95% *no* busque 0.95 en el cuerpo de la tabla A-2. Un nivel de confianza del 95% corresponde a $\alpha = 0.05$. Observe la figura 7-3, donde mostramos que el área en cada una de las colas sombreadas de gris oscuro es $\alpha/2 = 0.025$. Calculamos $z_{\alpha/2} = 1.96$ señalando que toda el área a su izquierda debe ser $1 - 0.025$, o 0.975. Podemos remitirnos a la tabla A-2 y

Figura 7-3 Cálculo de $z_\alpha/2$ para un nivel de confianza del 95%

encontrar que el área de 0.9750 (que se encuentra en el *cuerpo* de la tabla) corresponde exactamente a una puntuación z de 1.96. Para un nivel de confianza del 95%, el valor crítico es, por consiguiente, $z_{\alpha/2} = 1.96$. Para calcular la puntuación z crítica para un nivel de confianza del 95%, busque 0.9750 en el cuerpo la tabla A-2, no 0.95.

El ejemplo anterior mostró que un nivel de confianza del 95% da por resultado un valor crítico de $z_{\alpha/2} = 1.96$. Éste es el valor crítico más común y se lista junto con otros dos valores comunes en la siguiente tabla.

Nivel de confianza	α	valor crítico, $z_{\alpha/2}$
90%	0.10	1.645
95%	0.05	1.96
99%	0.01	2.575

Margen de error

Cuando reunimos un conjunto de datos muestrales, como los datos sobre la terapia de contacto de Emily Rosa en el problema del capítulo (donde el 44% de los 280 ensayos correspondieron a identificaciones correctas), podemos calcular la proporción muestral $\hat{p}$ y esta proporción muestral suele ser diferente de la proporción poblacional p. La diferencia entre la proporción muestral y la proporción de la población se considera un error. Ahora definimos el *margen de error E* como sigue.

Definición

Cuando se utilizan los datos de una muestra aleatoria simple para estimar una proporción poblacional p, el **margen de error**, denotado por $\boldsymbol{E}$, es la diferencia máxima probable (con probabilidad $1 - \alpha$) entre la proporción muestral $\hat{p}$ observada y el valor real de la proporción poblacional p. El margen de error E también se llama *error máximo del estimado* y se calcula multiplicando el valor crítico por la desviación estándar de las proporciones muestrales, como se indica en la fórmula 7-1.

Fórmula 7-1 $$E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$ margen de error para proporciones

Falsificación de datos

El glosario del censo del 2000 define la *falsificación de datos (curbstoning)* como "la práctica por medio de la cual un censor fabrica un cuestionario para una vivienda, sin visitarla". La falsificación de datos ocurre cuando un censor se sienta en la acera (o en cualquier otro lado) y llena las formas inventando las respuestas. Puesto que estos datos no son reales, afectan la validez del censo. En varios estudios se ha investigado la magnitud de la falsificación; uno de ellos reveló que aproximadamente el 4% de los censores realizan esta práctica al menos en una ocasión.

Los métodos de la sección 7-2 suponen que los datos muestrales se reunieron de una forma apropiada, así que si gran parte de los datos se obtuvieron a través de falsificaciones, entonces los estimados de los intervalos de confianza resultantes podrían tener muchos errores.

Dada la forma en que se define el margen de error E, existe una probabilidad de α de que una proporción muestral sea errónea por más de E.

Intervalo de confianza (o estimado de intervalo) para la proporción poblacional p

$$\hat{p} - E < p < \hat{p} + E \qquad \text{donde} \qquad E = z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

El intervalo de confianza suele expresarse en los siguientes formatos equivalentes:

$$\hat{p} \pm E$$

o

$$(\hat{p} - E, \hat{p} + E)$$

En el capítulo 4, cuando las probabilidades se expresaban en forma decimal, redondeábamos a tres dígitos significativos. Aquí utilizamos esa misma regla de redondeo.

Regla de redondeo para estimados de intervalo de confianza de p

Redondee los límites del intervalo de confianza para p a tres dígitos significativos.

Con base en los resultados anteriores, podemos resumir el procedimiento para construir un estimado del intervalo de confianza de una proporción poblacional p como sigue.

Procedimiento para construir un intervalo de confianza para p

1. Verifique que los supuestos requeridos se cumplen (la muestra es aleatoria simple, las condiciones para la distribución binomial se satisfacen y existen al menos 5 éxitos y al menos 5 fracasos).
2. Remítase a la tabla A-2 y encuentre el valor crítico $z_{\alpha/2}$ que corresponde al nivel de confianza deseado. (Por ejemplo, si el nivel de confianza es del 95%, el valor crítico es $z_{\alpha/2} = 1.96$).
3. Evalúe el margen de error $E = z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.
4. Utilizando el valor del margen de error E calculado y el valor de la proporción muestral $\hat{p}$, calcule los valores de $\hat{p} - E$ y $\hat{p} + E$. Sustituya esos valores en el formato general para el intervalo de confianza:

$$\hat{p} - E < p < \hat{p} + E$$

o

$$\hat{p} \pm E$$

o

$$(\hat{p} - E, \hat{p} + E)$$

5. Redondee los límites del intervalo de confianza resultantes a tres dígitos significativos.

Fundamentos del margen de error Puesto que la distribución de proporciones muestrales es aproximadamente normal (ya que ambas condiciones $np \geq 5$ y $nq \geq 5$ se satisfacen), podemos utilizar los resultados de la sección 6-6 para concluir que μ y σ están dadas por $\mu = np$ y $\sigma = \sqrt{npq}$. Estos dos parámetros pertenecen a n ensayos, pero los convertimos a una base por ensayo dividiendo entre n como sigue:

$$\text{Medida de proporciones muestrales: } \mu = \frac{np}{n} = p$$

$$\text{Desviación estándar de proporciones muestrales: } \sigma = \frac{\sqrt{npq}}{n} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}}$$

El primer resultado tal vez parezca trivial, puesto que ya estipulamos que la proporción real de la población es p. El segundo resultado no es trivial y resulta útil para describir el margen de error E, pero reemplazamos el producto pq por $\hat{p}\hat{q}$ porque no conocemos todavía el valor de p (ya que es el valor que tratamos de estimar). La fórmula 7-1 para el margen de error refleja el hecho de que $\hat{p}$ tiene una probabilidad de $1 - \alpha$ de estar dentro de $z_{\alpha/2}\sqrt{pq/n}$ de p. El intervalo de confianza para p, como se dio previamente, refleja el hecho de que existe una probabilidad de $1 - \alpha$ de que $\hat{p}$ difiera de p menos que el margen de error $E = z_{\alpha/2}\sqrt{pq/n}$.

Determinación del tamaño muestral

Suponga que queremos reunir datos muestrales con el objetivo de estimar alguna proporción de la población. ¿Cómo sabemos *cuántos* elementos muestrales deben obtenerse? Si tomamos la expresión para el margen de error E (fórmula 7-1), y luego despejamos n, obtenemos la fórmula 7-2, la cual requiere que $\hat{p}$ sea un estimado de la proporción poblacional p; pero si no se conoce un estimado como éste (como suele ser el caso), reemplazamos $\hat{p}$ por 0.5 y reemplazamos $\hat{q}$ por 0.5, con el resultado que se da en la fórmula 7-3.

Tamaño muestral para la estimación de la proporción *p*

Cuando se conoce un estimado $\hat{p}$: **Fórmula 7-2** $n = \dfrac{[z_{\alpha/2}]^2\hat{p}\hat{q}}{E^2}$

Cuando se desconoce el estimado $\hat{p}$: **Fórmula 7-3** $n = \dfrac{[z_{\alpha/2}]^2 \cdot 0.25}{E^2}$

TEXTO N° 3

Estimación de una media poblacional: o

conocida

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 338 a 343

7-3 Estimación de una media de población: σ conocida

Concepto clave En la sección 7-2 presentamos el estimado puntual y el intervalo de confianza como herramientas para el uso de una proporción muestral con el fin de estimar una proporción poblacional; esta sección presenta métodos para utilizar datos muestrales y calcular un estimado puntual y un estimado del intervalo de confianza de una media poblacional. Un requisito fundamental en esta sección es que, además de tener datos muestrales, también conozcamos σ, la desviación estándar de la población. También presentamos un método para calcular el tamaño muestral que se necesitaría para estimar una media poblacional.

Requisitos

1. La muestra es aleatoria simple. (Todas las muestras del mismo tamaño tienen la misma probabilidad de ser seleccionadas).
2. El valor de la desviación estándar poblacional σ es conocido.
3. Cualquiera o ambas de estas condiciones se satisfacen: la población está normalmente distribuida o $n > 30$.

Conocimiento de σ Los requisitos anteriores incluyen el conocimiento de la desviación estándar poblacional σ, pero la siguiente sección presenta métodos para estimar una media poblacional sin conocer el valor de σ.

Requisito de normalidad Los requisitos incluyen la propiedad de que la población se distribuya normalmente o que $n > 30$. Si $n \leq 30$, la población no necesita tener una distribución exactamente normal, sino aproximadamente normal. Podemos considerar que el requisito de normalidad se satisface si no hay valores extremos y si un histograma de los datos muestrales no se aleja mucho de la forma de campana. (Se dice que los métodos de esta sección son *robustos*, es decir, no se ven muy afectados si los datos se alejan de la normalidad, siempre y cuando no se alejen demasiado).

Requisitos del tamaño muestral Esta sección utiliza la distribución normal como la distribución de medias muestrales. Si la población original no está distribuida normalmente, entonces decimos que las medias de muestras con tamaño $n > 30$ tienen una distribución que puede aproximarse a una distribución normal. La condición de que el tamaño muestral sea $n > 30$ se usa por lo regular como directriz, pero no es posible identificar un tamaño muestral mínimo específico que sea suficiente para todos los casos. El tamaño muestral mínimo realmente depende de cuánto se desvía la distribución de la población de una distribución normal. Tamaños muestrales de 15 a 30 son adecuados si la población parece tener una distribución que no es lejana a la normal, pero algunas otras poblaciones tienen distribuciones que son extremadamente diferentes de la normal y pueden necesitarse tamaños muestrales de 50, 100 o más. Usaremos el criterio simplificado de $n > 30$ como justificación para el tratamiento de la distribución de medias muestrales como una distribución normal.

En la sección 7-2 vimos que la proporción muestral $\hat{p}$ es el mejor estimado puntual de la proporción poblacional p. Por razones similares, la media muestral $\bar{x}$ es el mejor estimado puntual de la media poblacional μ.

La media muestral $\bar{x}$ es el mejor estimado puntual de la media de la población.

Por lo regular la media muestral $\bar{x}$ brinda el mejor estimado, por las siguientes dos razones:

1. Para todas las poblaciones, la media muestral $\bar{x}$ es un **estimador sin sesgo** de la media poblacional μ, lo que significa que la distribución de medias muestrales tiende a concentrarse alrededor del valor de la media poblacional μ. [Es decir, las medias muestrales no tienden sistemáticamente a sobreestimar el valor de μ, ni tienden sistemáticamente a subestimar el valor de μ, sino que tienden a coincidir con este valor (como se ilustra en la sección 6-4).
2. Para muchas poblaciones, la distribución de las medias muestrales $\bar{x}$ tiende a ser más consistente (con *menos variación*) que la distribución de otros estadísticos muestrales.

EJEMPLO Pulso cardiaco de mujeres El pulso cardiaco de las personas es sumamente importante. Sin él, ¿dónde estaríamos? El conjunto de datos 1 del apéndice B incluye pulsos cardiacos (en latidos por minuto) de mujeres seleccionadas al azar; los estadísticos son los siguientes: $n = 40$, $\bar{x} = 76.3$ y $s = 12.5$. Utilice esta muestra para calcular el mejor estimado puntual de la media poblacional μ de los pulsos cardiacos de todas las mujeres.

SOLUCIÓN Para los datos muestrales, $\bar{x} = 76.3$. Como la media muestral $\bar{x}$ es el mejor estimado puntual de la media poblacional μ, concluimos que el mejor estimado puntual de los pulsos cardiacos de todas las mujeres es 76.3.

Intervalos de confianza

Aunque un estimado puntual es el *mejor* valor individual para estimar un parámetro poblacional, no nos da ninguna indicación precisa de qué tan *bueno* es este mejor estimado. Sin embargo, un intervalo de confianza nos ofrece información que nos permite comprender mejor la exactitud del estimado. El intervalo de confianza se asocia con un nivel de confianza, como 0.95 (o 95%). El nivel de confianza nos da la tasa de éxitos del procedimiento que se utiliza para construir el intervalo de confianza. Como se describió en la sección 7-2, α es el complemento del nivel de confianza. Para un nivel de confianza de 0.95 (o 95%), $\alpha = 0.05$. Para un nivel de confianza de 0.99 (o 99%), $\alpha = 0.01$.

Margen de error Cuando reunimos un conjunto de datos muestrales, como los datos de los 40 pulsos de mujeres que se incluyen en el conjunto de datos 1 del apéndice B, podemos calcular la media muestral $\bar{x}$ y esa media muestral por lo regular es diferente de la media poblacional μ. La diferencia entre la media muestral y la media poblacional es un error. En la sección 6-5 vimos que $\sigma/\sqrt{n}$ es la desviación estándar de las medias muestrales. Utilizando $\sigma/\sqrt{n}$ y la notación $z_{\alpha/2}$ que se presentó en la sección 7-2, ahora podemos usar el *margen de error E* que se expresa como sigue:

Fórmula 7-4 $E = z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$ margen de error para la media (con base en σ conocida)

La fórmula 7-4 refleja el hecho de que la distribución del muestreo de la media muestral $\bar{x}$ es *exactamente* una distribución normal con media μ y desviación

Estimación del tamaño de las poblaciones silvestres

La Ley Nacional de Administración de los Bosques de Estados Unidos protege a las especies en peligro de extinción, entre las que destaca el búho moteado del norte; la ley impidió que la industria silvícola talara vastas regiones de árboles en el noroeste del Pacífico. Se pidió a biólogos y a especialistas en estadística que analizaran el problema, y ellos concluyeron que estaban disminuyendo las tasas de supervivencia y los tamaños de las poblaciones de los búhos hembra, quienes desempeñan un papel importante en la supervivencia de la especie. Los biólogos y especialistas en estadística también estudiaron el salmón en los ríos Snake y Columbia del estado de Washington, así como los pingüinos en Nueva Zelanda. En el artículo "Sampling Wildlife Populations" (*Chance*, vol. 9, núm. 2), los autores Brian Manly y Lyman McDonald comentan que, en estudios de esta clase, "los biólogos ganan habilidades de modelamiento, que son una característica distintiva de la buena estadística. Por su parte, los especialistas en estadística aprenden a compenetrarse en la realidad de los problemas, ya que los biólogos los introducen en asuntos cruciales".

Los números de serie de tanques capturados revelan el tamaño de la población

Durante la Segunda Guerra Mundial, especialistas en espionaje del bando de los aliados querían determinar el número de tanques que Alemania estaba produciendo. Las técnicas de espionaje tradicionales produjeron resultados poco confiables, pero los especialistas en estadística obtuvieron estimaciones exactas al analizar los números de serie de los tanques capturados. Por ejemplo, los registros muestran que Alemania realmente produjo 271 tanques en junio de 1941. La estimación basada en los números de serie fue de 244, en tanto que los métodos de espionaje tradicionales dieron como resultado una estimación extrema de 1550. (Véase "An Empirical Approach to Economic Intelligence in World War II", de Ruggles y Brodie, *Journal of the American Statistical Association*, vol. 42).

estándar $\sigma/\sqrt{n}$, siempre y cuando la población tenga una distribución normal con media μ y desviación estándar σ. Si la población no está distribuida normalmente, las muestras grandes producen medias muestrales con una distribución que *se aproxima* a la normal. (La fórmula 7-4 requiere del conocimiento de la desviación estándar poblacional σ, pero la sección 7-4 presentará un método para calcular el margen de error E cuando se desconoce σ).

Utilizando el margen de error E, ahora podemos identificar el intervalo de confianza para la media poblacional μ (si se satisfacen los requisitos de esta sección). Los tres formatos que suelen usarse para expresar el intervalo de confianza se presentan en el siguiente cuadro.

Estimación del intervalo de confianza de la media poblacional μ (con σ conocida)

$$\bar{x} - E < \mu < \bar{x} + E \qquad \text{donde} \qquad E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

o

$$\bar{x} \pm E$$

o

$$(\bar{x} - E, \bar{x} + E)$$

Definición

Los dos valores $\bar{x} - E$ y $\bar{x} + E$ se llaman **límites del intervalo de confianza**.

Procedimiento para construir un intervalo de confianza para μ (con σ conocida)

1. Verifique que los supuestos requeridos se satisfagan. (Tenemos una muestra aleatoria simple, σ es conocida, y la población parece estar distribuida normalmente o $n > 30$).
2. Remítase a la tabla A-2 y calcule el valor crítico $z_{\alpha/2}$ correspondiente al nivel de confianza deseado. (Por ejemplo, si el nivel de confianza es del 95%, el valor crítico es $z_{\alpha/2} = 1.96$).
3. Evalúe el margen de error $E = z_{\alpha/2} \cdot \sigma/\sqrt{n}$.
4. Utilizando el valor calculado del margen de error E y el valor de la media muestral $\bar{x}$, calcule los valores de $\bar{x} - E$ and $\bar{x} + E$. Sustituya esos valores en el formato general para el intervalo de confianza:

$$\bar{x} - E < \mu < \bar{x} + E$$

o

$$\bar{x} \pm E$$

o

$$(\bar{x} - E, \bar{x} + E)$$

5. Redondee los valores resultantes usando la siguiente regla de redondeo.

Regla de redondeo para intervalos de confianza utilizados para estimar μ

1. Cuando utilice el *conjunto de datos original* para construir un intervalo de confianza, redondee los límites del intervalo de confianza a un decimal más del que se usa para el conjunto de datos original.
2. Cuando el conjunto de datos original se desconoce y sólo se utiliza el *resumen de estadísticos* (n, $\bar{x}$, s), redondee los límites del intervalo de confianza al mismo número de espacios decimales utilizados para la media muestral.

Interpretación de un intervalo de confianza Al igual que en la sección 7-2, debemos ser cuidadosos para interpretar correctamente los intervalos de confianza. Después de obtener un estimado del intervalo de confianza de la media poblacional μ, como un intervalo de confianza del 95% de $72.4 < \mu < 80.2$, existe una interpretación correcta y muchas interpretaciones erróneas.

Correcta: "Tenemos una confianza del 95% de que el intervalo de 72.4 a 80.2 realmente contiene el valor verdadero de μ". Esto significa que si seleccionamos muchas muestras diferentes del mismo tamaño y construimos los intervalos de confianza correspondientes, a la larga, el 95% de éstos contendrían realmente el valor de μ. (Como en la sección 7-2, esta interpretación correcta se refiere a la tasa de éxitos del *proceso* que se usa para estimar la media poblacional).

Errónea: Puesto que μ es una constante fija, sería incorrecto decir que "existe un 95% de probabilidades de que μ se localice entre 72.4 y 80.2". También sería incorrecto decir que "el 95% de todos los valores de los datos están entre 72.4 y 80.2" y que "el 95% de las medias muestrales caen entre 72.4 y 80.2".

EJEMPLO Pulsos cardiacos de mujeres Para la muestra de pulsos cardiacos de mujeres en el conjunto de datos 1 del apéndice B, tenemos $n = 40$ y $\bar{x} = 76.3$, y la muestra es aleatoria simple. Suponga que sabemos que σ es 12.5. Utilice un nivel de confianza del 95% y calcule lo siguiente:

a. El margen de error E.
b. El intervalo de confianza para μ.

SOLUCIÓN

REQUISITO ✔ Primero debemos verificar que se cumplan los requisitos. La muestra es aleatoria simple. Se supone que conocemos el valor de σ (12.5). Con $n > 30$, se satisface el requisito de que "la población se distribuye normalmente o $n > 30$". Por lo tanto, los requisitos se cumplen y podemos continuar con los métodos de esta sección. ✔

a. El nivel de confianza del 0.95 implica que $\alpha = 0.05$, entonces $z_{\alpha/2} = 1.96$ (como se mostró en un ejemplo de la sección 7-2). El margen de error E se calcula usando la fórmula 7-4 como sigue. Los lugares decimales adicionales se usan para minimizar los errores de redondeo en el intervalo de confianza calculado en el inciso *b*).

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 1.96 \cdot \frac{12.5}{\sqrt{40}} = 3.8737901$$

continúa

b. Con $\bar{x} = 76.3$ y $E = 3.8737901$, construimos el intervalo de confianza como sigue:

$$\bar{x} - E < \mu < \bar{x} + E$$
$$76.3 - 3.8737901 < \mu < 76.3 + 3.8737901$$
$$72.4 < \mu < 80.2 \quad \text{(redondeado a un decimal como en } \bar{x}\text{)}$$

INTERPRETACIÓN Este resultado también podría expresarse como 76.3 ± 3.9 o como (72.4, 80.2). Con base en la muestra con $n = 40$, $\bar{x} = 76.3$, y suponiendo que σ es 12.5, el intervalo de confianza para la media de la población μ es $72.4 < \mu < 80.2$ y este intervalo tiene un nivel de confianza de 0.95. Esto significa que si seleccionamos muchas muestras diferentes de 40 mujeres y construimos los intervalos de confianza como lo hicimos aquí, el 95% de ellos incluirían realmente el valor de la media poblacional μ.

Fundamentos del intervalo de confianza La idea básica que subyace en la construcción de intervalos de confianza se relaciona con el teorema del límite central, que indica que si reunimos muestras aleatorias simples de una población distribuida normalmente, las medias muestrales se distribuyen de manera normal, con media μ y desviación estándar $\sigma/\sqrt{n}$. Si reunimos muestras aleatorias simples de tamaño $n > 30$ de cualquier población, la distribución de medias muestrales es aproximadamente normal, con media μ y desviación estándar $\sigma/\sqrt{n}$. El formato del intervalo de confianza es realmente una variación de la ecuación que ya se usó con el teorema del límite central. En la expresión $z = (\bar{x} - \mu_{\bar{x}})/\sigma_{\bar{x}}$, sustituya $\sigma_{\bar{x}}$ por $\sigma/\sqrt{n}$, sustituya $\mu_{\bar{x}}$ por μ, y entonces despeje μ para obtener

$$\mu = \bar{x} - z\frac{\sigma}{\sqrt{n}}$$

Se usan los valores positivo y negativo para los resultados de z en los límites del intervalo de confianza que estamos empleando.

Consideremos el caso específico de un nivel de confianza del 95%, de manera que $\alpha = 0.05$ y $z_{\alpha/2} = 1.96$. Para este caso, hay una probabilidad de 0.05 de que una media muestral esté a más de 1.96 desviaciones estándar (o $z_{\alpha/2} \cdot \sigma/\sqrt{n}$, lo que denotamos como E) de la media poblacional μ. Por otra parte, existe una probabilidad del 0.95 de que una media muestral esté dentro de 1.96 desviaciones estándar (o $z_{\alpha/2} \cdot \sigma/\sqrt{n}$) a partir de μ (véase la figura 7-4). Si la media muestral $\bar{x}$ está dentro de $z_{\alpha/2} \cdot \sigma/\sqrt{n}$ de la media poblacional μ, entonces μ debe estar entre $\bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n}$ y $\bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n}$; esto se expresa en el formato general de nuestro intervalo de confianza (con $z_{\alpha/2} \cdot \sigma/\sqrt{n}$, denotado como E): $\bar{x} - E < \mu < \bar{x} + E$.

Determinación del tamaño muestral requerido para estimar μ

Ahora examinaremos esta importante pregunta: cuando planeamos reunir una muestra aleatoria simple de datos que se usarán para estimar una media poblacional μ, *¿cuántos* valores muestrales deben obtenerse? Por ejemplo, suponga que queremos estimar el peso medio de pasajeros de líneas aéreas (un valor importante por razones de seguridad). ¿Cuántos pasajeros deben seleccionarse al azar y pesarse? La determinación del tamaño de una muestra aleatoria simple es un aspecto muy importante,

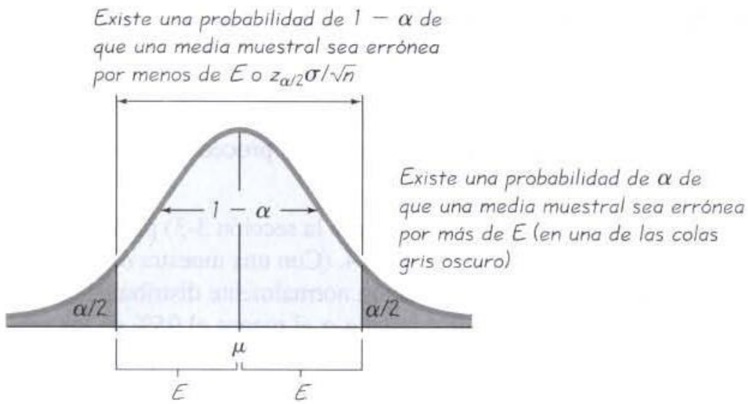


Figura 7-4

Distribución de medias muestrales con σ conocida

puesto que muestras que son innecesariamente grandes desperdician tiempo y dinero, en tanto que muestras muy pequeñas conducen a resultados deficientes.

Si empezamos con la expresión para el margen de error E (fórmula 7-4) y despejamos el tamaño muestral n, obtenemos lo siguiente.

Tamaño muestral para estimar la media μ

Fórmula 7-5
$$n = \left[\frac{z_{\alpha/2}\sigma}{E}\right]^2$$

donde $z_{\alpha/2}$ = puntuación z crítica basada en el nivel de confianza deseado
E = margen de error deseado
σ = desviación estándar poblacional

La fórmula 7-5 es relevante puesto que indica que el tamaño muestral no depende del tamaño de la población (N); el tamaño muestral depende del nivel de confianza deseado, del margen de error deseado y del valor de la desviación estándar σ. (Véase el ejercicio 40 para casos en los que se selecciona una muestra relativamente grande sin reemplazo a partir de una población finita).

El tamaño muestral debe ser un número entero, ya que representa el número de valores muestrales que deben encontrarse. Sin embargo, la fórmula 7-5 suele dar un resultado que no es un número entero, de manera que utilizamos la siguiente regla de redondeo. (Esta regla se basa en el principio de que cuando es necesario redondear, el tamaño de muestra requerido debe redondearse *hacia arriba* para que sea al menos adecuadamente grande en oposición a un tamaño ligeramente más pequeño).

Regla de redondeo para el tamaño muestral n

Cuando se calcula el tamaño muestral n, si el uso de la fórmula 7-5 no produce un número entero, siempre *incremente* el valor de n al siguiente número entero *mayor*.

TEXTO N° 4

Estimación de una media poblacional: o

desconocida

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 349 a 354

7-4 Estimación de la media poblacional: σ desconocida

Concepto clave En esta sección se presentan métodos para construir un estimado del intervalo de confianza de una media poblacional cuando *no se conoce* la desviación estándar. (En la sección 7-3 se presentaron métodos para estimar μ cuando se conoce σ). Cuando se desconoce σ, se utiliza la *distribución t de Student* (en vez de la distribución normal), suponiendo que ciertos requisitos (los cuales se señalan abajo) se satisfacen. Como generalmente se desconoce σ en circunstancias reales, los métodos de esta sección son muy realistas y prácticos, y se utilizan con frecuencia.

Requisitos

1. La muestra es aleatoria simple.
2. La muestra proviene de una población distribuida normalmente o $n > 30$.

Como en la sección 7-3, el requisito de una población distribuida normalmente no es estricto. Por lo regular, podemos considerar que la población está distribuida normalmente después de usar los datos muestrales para confirmar que no existen valores extremos y que el histograma tiene una forma que no es muy lejana a la de una distribución normal. Además, al igual que en la sección 7-3, el requisito de que el tamaño muestral sea $n > 30$ suele usarse como directriz, pero el tamaño muestral mínimo realmente depende de cuánto se aleja la distribución de la población de la distribución normal. [Si se sabe que una población se distribuye normalmente, la distribución de medias muestrales $\bar{x}$ es *exactamente* una distribución normal con media μ y desviación estándar $\sigma/\sqrt{n}$; si la población no está distribuida normalmente, muestras grandes ($n > 30$) producen medias muestrales con una distribución que es *aproximadamente* normal, con media μ y desviación estándar $\sigma/\sqrt{n}$].

Al igual que en la sección 7-3, la media muestral $\bar{x}$ es el mejor estimado puntual (o estimado de un solo valor) de la media poblacional μ.

La media muestral $\bar{x}$ es el mejor estimado puntual de la media poblacional μ.

He aquí el aspecto clave de esta sección: si σ no se conoce, pero los requisitos anteriores se satisfacen, utilizamos la *distribución t de Student* (en vez de la distribución normal), que desarrolló William Gosset (1876-1937). Gosset fue un empleado de la cervecería Guinness Brewery que necesitaba una distribución que pudiera utilizarse con muestras pequeñas. La cervecería irlandesa donde trabajaba no permitía la publicación de resultados de investigaciones, entonces Gosset publicó bajo el seudónimo de *Student*. (En aras de la investigación y para servir a sus lectores, el autor visitó la cervecería Guinness Brewery y probó una muestra del producto. ¡Qué comprometido!)

Puesto que no conocemos el valor de σ, lo estimamos con el valor de la desviación estándar muestral s, pero esto introduce otra fuente de falta de confiabilidad, en especial con las muestras pequeñas. Para mantener un intervalo de confianza en algún nivel deseado, como el 95%, compensamos esta falta de confiabilidad adicional haciendo más ancho el intervalo de confianza: utilizamos valores críticos $t_{\alpha/2}$ (de una distribución t de Student), los cuales son más grandes que los valores críticos de $z_{\alpha/2}$ de la distribución normal.

Distribución *t* de Student

Si una población tiene una distribución normal, entonces la distribución de

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

es una **distribución *t* de Student** para todas las muestras de tamaño n. La distribución t de Student, conocida a menudo como **distribución *t***, se utiliza para calcular valores críticos denotados por $t_{\alpha/2}$.

Pronto analizaremos algunas de las propiedades importantes de la distribución t, pero antes presentamos los componentes necesarios para la construcción de intervalos de confianza. Comencemos con el valor crítico denotado por $t_{\alpha/2}$. Un valor de $t_{\alpha/2}$ se puede encontrar en la tabla A-3 localizando el número apropiado de *grados de libertad* en la columna izquierda y avanzando por el renglón correspondiente hasta encontrar el número que aparece directamente abajo del área adecuada en la parte superior.

Definición

El número de **grados de libertad** para un conjunto de datos muestrales recolectados es el número de valores muestrales que pueden variar después de haber impuesto ciertas restricciones a todos los valores de los datos.

Por ejemplo, si 10 estudiantes tienen puntuaciones de examen con una media de 80, podemos asignar con libertad valores a las primeras 9 puntuaciones, pero la décima puntuación se calcula. La suma de las 10 puntuaciones debe ser 800, entonces la décima puntuación debe ser igual a 800 menos la suma de las primeras 9 puntuaciones. Puesto que esas primeras 9 puntuaciones pueden seleccionarse con libertad para adoptar cualquier valor, decimos que existen 9 grados de libertad disponibles. Para las aplicaciones de esta sección, el número de grados de libertad es simplemente el tamaño muestral menos 1.

$$\textbf{grados de libertad} = n - 1$$

EJEMPLO Cálculo de un valor crítico Una muestra de tamaño $n = 23$ es una muestra aleatoria simple seleccionada de una población distribuida normalmente. Calcule el valor crítico $t_{\alpha/2}$ correspondiente a un nivel de confianza del 95%.

SOLUCIÓN Puesto que $n = 23$, el número de grados de libertad está dado por $n - 1 = 22$. Utilizando la tabla A-3, localizamos el renglón 22 con respecto a la columna de la extrema izquierda. Al igual que en la sección 7-2, un nivel de confianza del 95% corresponde a $\alpha = 0.05$, de manera que encontramos los valores listados en la columna para una *área de 0.05 en dos colas*. El valor correspondiente al renglón para 22 grados de libertad y la columna para una área de 0.05 en dos colas es 2.074; entonces $t_{\alpha/2} = 2.074$.

Ahora que sabemos cómo encontrar valores críticos denotados por $t_{\alpha/2}$, podemos describir el margen de error E de este intervalo de confianza.

Margen de error E para la estimación de μ (con σ desconocida)

Fórmula 7-6 $$E = t_{\alpha/2}\frac{s}{\sqrt{n}}$$

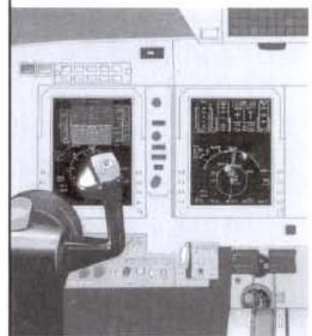
donde $t_{\alpha/2}$ tiene $n - 1$ grados de libertad. La tabla A-3 lista valores de $t_{\alpha/2}$.

Intervalo de confianza para la estimación de μ (con σ desconocida)

$$\bar{x} - E < \mu < \bar{x} + E$$

donde

$$E = t_{\alpha/2}\frac{s}{\sqrt{n}}$$



Extractos de una circular del Departamento del Transporte

Los siguientes extractos de una circular del departamento de transporte de Estados Unidos atañen algunos de los requisitos de exactitud para el equipo de navegación empleado en aviones. Observe el uso del intervalo de confianza. "El total de las contribuciones de error del equipo a bordo, combinado con los errores técnicos de vuelo correspondientes incluidos en la lista, no debe exceder lo siguiente, con un nivel de confianza del 95% (2-sigma), durante un periodo igual al ciclo de actualización". "El sistema de vías y rutas aéreas de Estados Unidos tiene anchuras de protección de ruta que se utilizan en un sistema VOR con una exactitud de ±4.5 grados con base en una probabilidad del 95%".

El siguiente procedimiento utiliza el margen de error anterior en la construcción de estimados del intervalo de confianza de μ.

Procedimiento para construir un intervalo de confianza para μ (con σ desconocida)

1. Verifique que los requisitos se satisfacen. (Tenemos una muestra aleatoria simple y la población parece estar distribuida normalmente o $n > 30$).
2. Utilizando $n - 1$ grados de libertad, remítase a la tabla A-3 y encuentre el valor crítico $t_{\alpha/2}$ que corresponde al nivel de confianza deseado. (Para el nivel de confianza, remítase al "área en dos colas").
3. Evalúe el margen de error $E = t_{\alpha/2} \cdot s/\sqrt{n}$.
4. Utilizando el valor del margen de error E calculado y el valor de la media muestral $\bar{x}$, calcule los valores de $\bar{x} - E$ y $\bar{x} + E$. Sustituya estos valores en el formato general para el intervalo de confianza:

$$\bar{x} - E < \mu < \bar{x} + E$$

o

$$\bar{x} \pm E$$

o

$$(\bar{x} - E, \bar{x} + E)$$

5. Redondee los límites del intervalo de confianza resultantes. Si utiliza el conjunto original de datos, redondee a un decimal más del que se usa para el conjunto original de datos. Si utiliza un resumen de estadísticos (n, $\bar{x}$, s), redondee los límites del intervalo de confianza al mismo número de lugares decimales utilizados para la media muestral.

Gráfica de tallo y hojas de las edades

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

EJEMPLO Construcción de un intervalo de confianza En el diagrama de tallo y hojas que aparece al margen, se incluyen las edades de solicitantes que no lograron un ascenso (según datos de "Debating the Use of Statistical Evidence in Allegations of Age Discrimination", de Barry y Boland, *American Statistician*, vol. 58, núm. 2). Existe el tema más importante de si ciertos solicitantes fueron víctimas de discriminación por edad, pero por ahora nos enfocaremos en el simple aspecto de utilizar esos valores como una muestra con el propósito de estimar la media de una población más grande. Suponga que la muestra es aleatoria simple y utilice los datos muestrales con un nivel de confianza del 95% para calcular lo siguiente:

a. El margen de error E

b. El intervalo de confianza para μ

SOLUCIÓN

REQUISITO ✔ Primero debemos verificar que los dos requisitos para esta sección se satisfacen. Estamos suponiendo que la muestra es aleatoria simple. Ahora revisamos el requisito de que "la población se distribuya normalmente o $n > 30$". Puesto que $n = 23$, debemos verificar que la distribución sea aproximadamente normal. La forma de la gráfica de tallo y hojas sugiere una distribución normal. Además, una gráfica cuantilar normal confirma que los datos muestrales provienen de una población con una distribución aproximadamente normal. Por consiguiente, los requisitos se satisfacen y procedemos con los métodos de esta sección. ✔

a. El nivel de confianza de 0.95 implica que $\alpha = 0.05$, de manera que $t_{\alpha/2} = 2.074$ (utilice la tabla A-3 con gl $= n - 1 = 22$, como se mostró en el ejemplo anterior). Después de encontrar que los estadísticos muestrales son $n = 23$, $\bar{x} = 47.0$ y $s = 7.2$, el margen de error E se calcula utilizando la fórmula 7-6 como sigue. Se utilizan decimales adicionales para minimizar los errores de redondeo en el intervalo de confianza calculado en el inciso *b*).

$$E = t_{\alpha/2}\frac{s}{\sqrt{n}} = 2.074 \cdot \frac{7.2}{\sqrt{23}} = 3.11370404$$

b. Con $\bar{x} = 47.0$ y $E = 3.11370404$, construimos el intervalo de confianza de la siguiente manera:

$$\bar{x} - E < \mu < \bar{x} + E$$
$$47.0 - 3.11370404 < \mu < 47.0 + 3.11370404$$
$$43.9 < \mu < 50.1$$ (redondeado a un decimal más que los datos originales)

INTERPRETACIÓN Este resultado también podría expresarse en la forma de 47.0 ± 3.1 o $(43.9, 50.1)$. Con base en los resultados muestrales dados, tenemos una confianza del 95% de que los límites de 43.9 años y 50.1 años realmente contienen el valor de la media poblacional μ.

Ahora listamos las propiedades importantes de la distribución *t* que utilizamos en esta sección.

Propiedades importantes de la distribución *t* de Student

1. La distribución *t* de Student es diferente para distintos tamaños de muestra. (Véase la figura 7-5 para los casos $n = 3$ y $n = 12$).
2. La distribución *t* de Student tiene la misma forma de campana simétrica que la distribución normal estándar, pero refleja una mayor variabilidad (con distribuciones más amplias) de lo que se espera con muestras pequeñas.

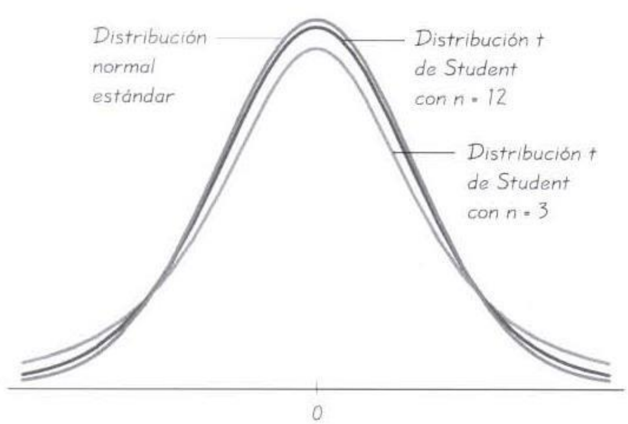


Figura 7-5

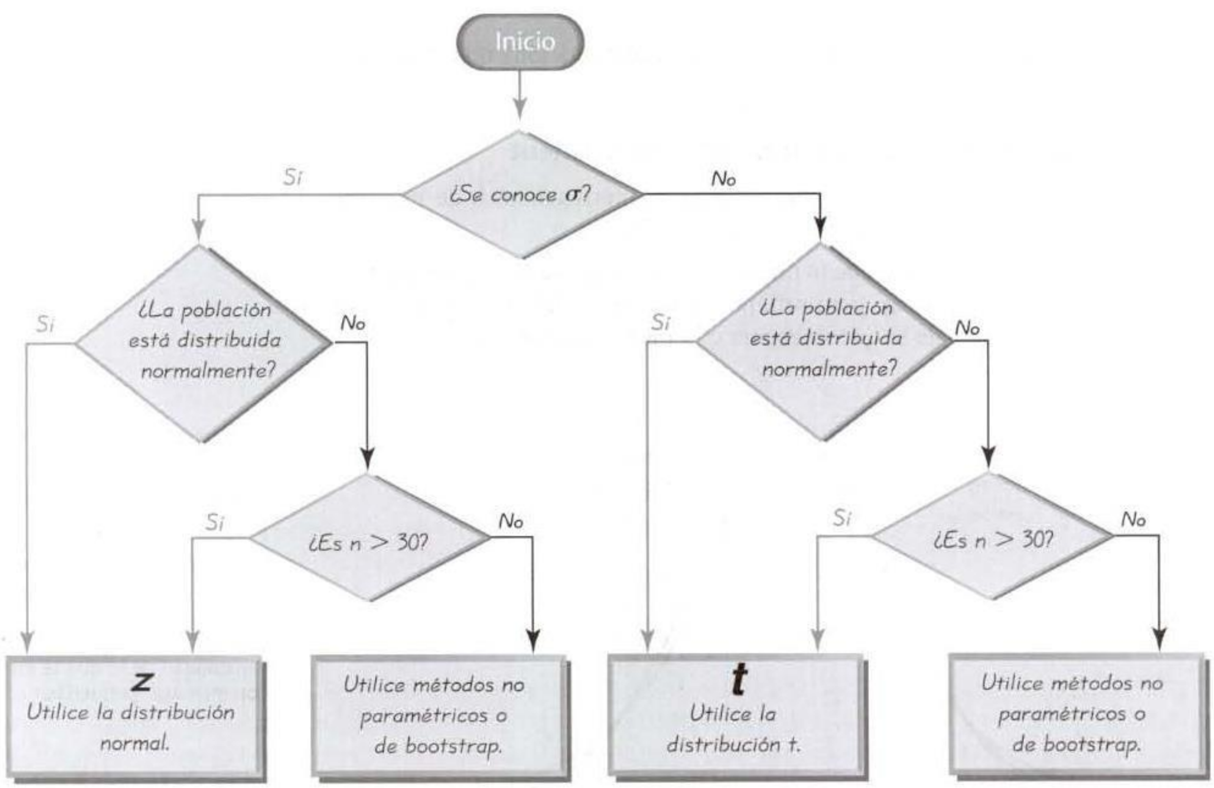
Distribuciones *t* de Student para $n = 3$ y $n = 12$

La distribución *t* de Student tiene la misma forma y simetría general de la distribución normal estándar, pero refleja una mayor variabilidad de lo que se espera con muestras pequeñas.

3. La distribución t de Student tiene una media de $t = 0$ (así como la distribución normal estándar tiene una media de $z = 0$).
4. La desviación estándar de la distribución t de Student varía con el tamaño muestral, pero es mayor que 1 (a diferencia de la distribución normal estándar, que tiene $\sigma = 1$).
5. Conforme el tamaño muestral n se hace más grande, la distribución t de Student se acerca más a la distribución normal estándar.

Elección de la distribución apropiada

En ocasiones es difícil decidir entre utilizar la distribución normal estándar z o la distribución t de Student. El diagrama de flujo de la figura 7-6 y la tabla 7-1 resumen los aspectos clave a considerarse cuando se construyen intervalos de confianza para estimar μ, la media poblacional. En la figura 7-6 o en la tabla 7-1, note que si tenemos una muestra pequeña ($n \leq 30$) obtenida de una distribución que difiere drásticamente de una distribución normal, no podemos usar los métodos descritos en este capítulo. Una alternativa es utilizar métodos no paramétricos (véase el capítulo 13); otra alternativa es usar el método de *bootstrap* por computadora. En ambos enfoques no se hacen supuestos acerca de la población original. El método *bootstrap* se describe en el proyecto tecnológico al final del capítulo.



TEXTO N° 5

Estimación de la varianza poblacional.

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 363 a 369

7-5 Estimación de la varianza poblacional

Concepto clave En esta sección presentamos métodos para **1.** calcular un intervalo de confianza de una desviación estándar o una varianza poblacional y **2.** determinar el tamaño muestral requerido para estimar una desviación estándar o una varianza poblacional. En esta sección se presenta la distribución chi cuadrada, la cual se utiliza para calcular un estimado de un intervalo de confianza de σ o de σ^2.

Requisitos

1. La muestra es aleatoria simple.
2. La población debe tener valores distribuidos normalmente (aun si la muestra es grande).

El supuesto de una población distribuida normalmente se mencionó en secciones anteriores, pero este requisito es mucho más importante aquí. Para los métodos de esta sección, los alejamientos de una distribución normal pueden generar errores muy graves. En consecuencia, el requisito de tener una distribución normal es mucho más estricto, y debemos revisar la distribución de los datos construyendo histogramas y gráficas cuantilares normales, como se describe en la sección 6-7.

Cuando consideramos estimados de proporciones y medias, utilizamos las distribuciones normal y t de Student. Cuando desarrollamos estimados de varianzas o desviaciones estándar utilizamos otra distribución, conocida como la distribución chi cuadrada. Examinaremos características importantes de esta distribución antes de proceder con el desarrollo de intervalos de confianza.

Distribución chi cuadrada

En una población distribuida normalmente con varianza σ^2, suponga que seleccionamos al azar muestras independientes de tamaño n y, para cada muestra, calculamos la varianza muestral s^2 (que es el cuadrado de la desviación estándar muestral s). El estadístico muestral $\chi^2 = (n-1)s^2/\sigma^2$ tiene una distribución llamada **distribución chi cuadrada**.

Distribución chi cuadrada

Fórmula 7-7 $$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

donde n = tamaño muestral
s^2 = varianza muestral
σ^2 = varianza poblacional

Denotamos chi cuadrada por χ^2, que se pronuncia "ji cuadrada". Para calcular valores críticos de la distribución chi cuadrada, remítase a la tabla A-4. La distribución chi cuadrada se determina por el número de grados de libertad y en este capítulo usamos $n-1$ grados de libertad.

$$\textbf{grados de libertad} = n - 1$$

En capítulos posteriores encontraremos situaciones en las que los grados de libertad no son $n-1$, por lo que no debemos hacer la generalización incorrecta de que el número de grados de libertad es siempre $n-1$.

Propiedades de la distribución del estadístico chi cuadrada

1. La distribución chi cuadrada no es simétrica, a diferencia de las distribuciones normal y t de Student (véase la figura 7-8). (Conforme el número de grados de

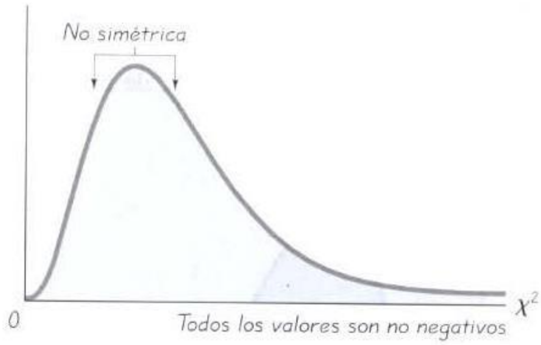


Figura 7-8

Distribución chi cuadrada

libertad se incrementa, la distribución se vuelve más simétrica, como ilustra la figura 7-9).

2. Los valores de chi cuadrada pueden ser cero o positivos, pero no pueden ser negativos (véase la figura 7-8).
3. La distribución chi cuadrada es diferente para cada número de grados de libertad (véase la figura 7-9), y en esta sección el número de grados de libertad está dado por gl $= n - 1$. Conforme el número de grados de libertad se incrementa, la distribución chi cuadrada se aproxima a una distribución normal.

Puesto que la distribución chi cuadrada es sesgada y no simétrica, el intervalo de confianza no se ajusta al formato de $s^2 \pm E$ y debemos hacer cálculos separados para los límites de confianza superior e inferior. Si se utiliza la tabla A-4 para calcular valores críticos, observe su siguiente característica:

En la tabla A-4 cada valor crítico de χ^2 corresponde a una área que se encuentra en el renglón superior de la tabla, y esa área representa la *región acumulativa localizada a la derecha* del valor crítico.

La tabla A-2 para la distribución normal estándar proporciona áreas acumulativas de la *izquierda*, pero la tabla A-4 para la distribución chi cuadrada provee áreas acumulativas de la *derecha*.

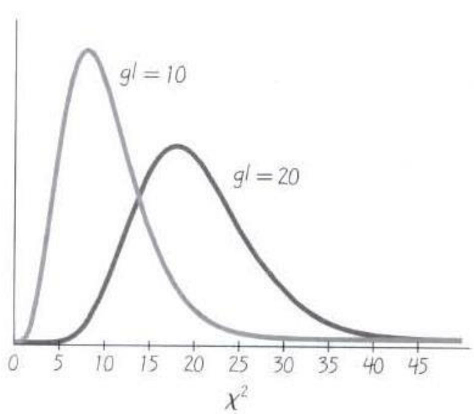


Figura 7-9

Distribución chi cuadrada para gl = 10 y gl = 20

EJEMPLO Valores críticos Calcule los valores críticos de χ^2 que determinan las regiones críticas que contienen una área de 0.025 en cada cola. Suponga que el tamaño muestral relevante es 10, de manera que el número de grados de libertad es $10 - 1$, o 9.

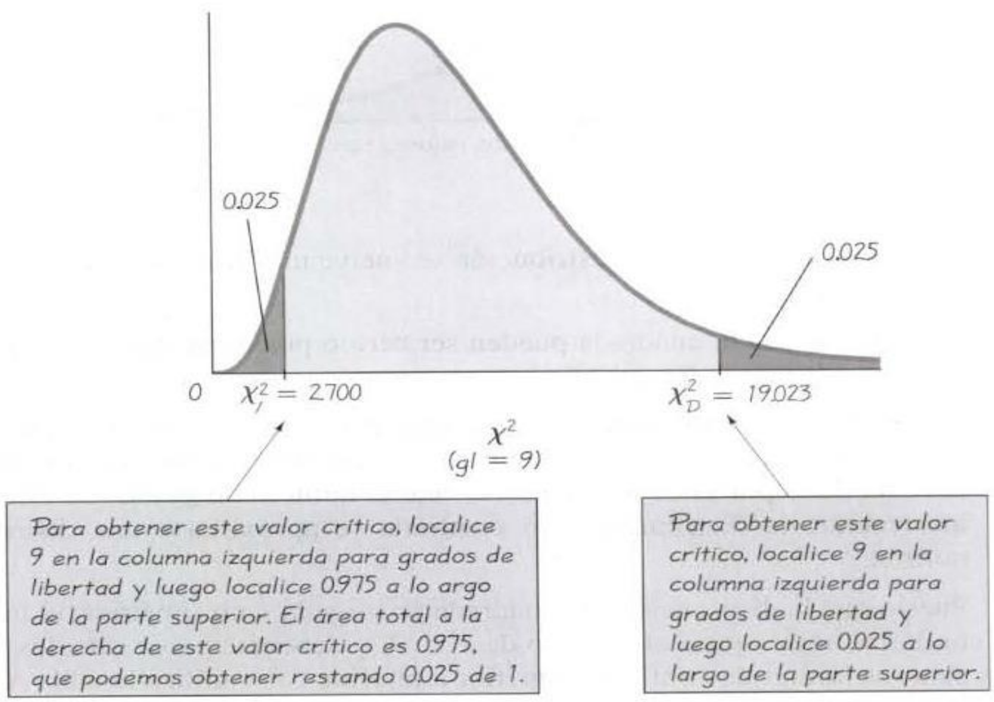


Figura 7-10 Valores críticos de la distribución chi cuadrada

SOLUCIÓN Vea la figura 7-10 y remítase a la tabla A-4. El valor crítico a la derecha ($\chi^2 = 19.023$) se obtiene de manera directa localizando 9 en la columna de grados de libertad a la izquierda y 0.025 a lo largo de la parte superior. El valor crítico de $\chi^2 = 2.700$ a la izquierda otra vez corresponde a 9 en la columna de grados de libertad, pero debemos localizar 0.975 (que se calcula al restar 0.025 de 1) a lo largo de la parte superior, puesto que los valores en el renglón superior son siempre *áreas a la derecha* del valor crítico. Remítase a la figura 7-10 y vea que el área total a la derecha de $\chi^2 = 2.700$ es 0.975. La figura 7-10 nos indica que, para una muestra de 10 valores tomados de una población distribuida normalmente, el estadístico chi cuadrada $(n - 1)s^2/\sigma^2$ tiene una probabilidad de 0.95 de caer dentro de los valores críticos de chi cuadrada de 2.700 y 19.023.

Observe que, cuando se obtienen valores críticos de χ^2 de la tabla A-4, los números de grados de libertad son enteros consecutivos del 1 al 30, seguidos por 40, 50, 60, 70, 80, 90 y 100. Cuando un número de grados de libertad (por ejemplo 52) no se encuentra en la tabla, generalmente se utiliza el valor crítico más cercano. Por ejemplo, si el número de grados de libertad es 52, remítase a la tabla A-4 y utilice 50 grados de libertad. (Si el número de grados de libertad está exactamente a la mitad de dos valores de la tabla, como por ejemplo 55, simplemente calcule la media de los dos valores χ^2). Para números de grados de libertad mayores de 100, use la ecuación que se incluye en el ejercicio 27, una tabla más detallada o un programa de cómputo de estadística.

Estimadores de σ^2

En la sección 6-4 señalamos que las varianzas muestrales s^2 tienden a coincidir con (o centrarse en) el valor de la varianza poblacional σ^2, por lo que decimos que s^2 es un *estimador sin sesgo* de σ^2. Es decir, las varianzas muestrales s^2 no tienden sistemáticamente a sobreestimar el valor de σ^2, ni tampoco tienden sistemáticamente a subestimar σ^2. En vez de ello, tienden a coincidir con el valor de la propia σ^2. Además, los valores de s^2 tienden a producir errores más pequeños por estar más cercanos a σ^2 que otras medidas de variación sin sesgo. Por estas razones, generalmente se utiliza s^2 para estimar σ^2. [Sin embargo, existen otros estimadores de σ^2 que podrían considerarse mejores que s^2. Por ejemplo, aun cuando $(n - 1)s^2/(n + 1)$ es un estimador sesgado de σ^2, tiene la propiedad muy deseable de minimizar la media de los cuadrados de los errores y, por lo tanto, tiene una mayor probabilidad de acercarse a σ^2. Véase el ejercicio 28].

La varianza muestral s^2 es el mejor estimado puntual de la varianza poblacional σ^2.

Puesto que s^2 es un estimador sin sesgo de σ^2, esperaríamos que s fuera un estimador sin sesgo de σ, pero no es así. (Véase la sección 5.4). Sin embargo, si el tamaño muestral es grande, el sesgo es tan pequeño que podemos utilizar s como un estimado de σ razonablemente bueno. Aunque s es un estimado sesgado, se usa con frecuencia como un estimado puntual de σ.

La desviación estándar muestral s suele utilizarse como un estimado puntual de σ (aunque es un estimado sesgado).

Si bien s^2 es el mejor estimado puntual de σ^2, no existe una indicación de qué tan bueno es en realidad. Para compensar esta deficiencia, desarrollamos un estimado de intervalo (o intervalo de confianza) que es más informativo.

Intervalo de confianza (o estimado de intervalo) para la varianza poblacional σ^2

$$\frac{(n-1)s^2}{\chi_D^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_I^2}$$

Esta expresión se utiliza para calcular un intervalo de confianza para la varianza σ^2, pero un intervalo de confianza (o un estimado de intervalo) para la desviación estándar σ se calcula tomando la raíz cuadrada de cada componente, como se indica abajo.

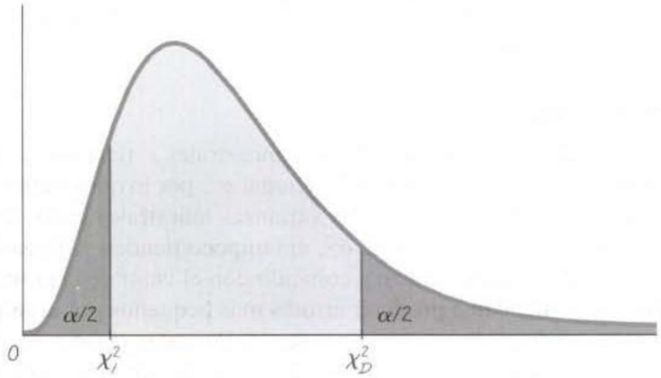
$$\sqrt{\frac{(n-1)s^2}{\chi_D^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_I^2}}$$

Las notaciones χ_D^2 y χ_I^2 en las expresiones anteriores se describen como sigue. (Observe que algunos otros libros de texto utilizan $\chi_{\alpha/2}^2$ en vez de χ_D^2 y utilizan $\chi_{1-\alpha/2}^2$ en vez de χ_I^2).

Figura 7-11

Distribución chi cuadrada con valores críticos χ_I^2 y χ_D^2

Los valores críticos χ_I^2 y χ_D^2 separan las áreas extremas correspondientes a varianzas muestrales que son improbables (con probabilidad α).



Notación

Con una área total de α dividida por igual entre las dos colas de una distribución chi cuadrada, χ_I^2 denota el valor crítico de la cola izquierda y χ_D^2 denota el valor crítico de la cola derecha (como se ilustra en la figura 7-11).

Con base en los resultados anteriores, podemos resumir el procedimiento para construir un estimado del intervalo de confianza de σ o σ^2 como sigue.

Procedimiento para construir un intervalo de confianza para σ o σ^2

1. Verifique que los requisitos se satisfagan. (La muestra es aleatoria simple y un histograma o gráfica cuantilar normal sugiere que la población tiene una distribución que es muy cercana a la distribución normal).
2. Utilizando $n - 1$ grados de libertad, remítase a la tabla A-4 y encuentre los valores críticos χ_D^2 y χ_I^2 correspondientes al nivel de confianza deseado.
3. Evalúe los límites del intervalo de confianza superior e inferior utilizando el siguiente formato para el intervalo de confianza:

$$\frac{(n-1)s^2}{\chi_D^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_I^2}$$

4. Si se desea un estimado del intervalo de confianza de σ, calcule la raíz cuadrada de los límites del intervalo de confianza superior e inferior y cambie σ^2 por σ.
5. Redondee los límites del intervalo de confianza resultantes. Si se utiliza el conjunto original de datos, redondee a un decimal más del que se usa para el conjunto original de datos. Si se utiliza la desviación estándar o varianza muestrales, redondee los límites del intervalo de confianza al mismo número de espacios decimales.

Advertencia: Los intervalos de confianza pueden usarse de manera informal para comparar conjuntos diferentes de datos, pero el traslape de intervalos de confianza no debe usarse para obtener conclusiones formales ni finales acerca de la igualdad de las varianzas o de las desviaciones estándar. En capítulos posteriores se incluirán procedimientos para decidir si dos poblaciones tienen varianzas o desviaciones estándar iguales, y esos métodos no tendrán las deficiencias asociadas con comparaciones basadas en el traslape de los intervalos de confianza.

Práctica de Estadística II - N° 01

TEMA N° 1: Diseño de experimentos y muestreo

Sección : …………………………..………………..	Apellidos : ………………………..……………….
Docente : Escribir el nombre del docente	Nombres : …………………………………………….
	Fecha : ……/……/2017 Duración: 45 minutos
	Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

I. Conocimientos estadísticos y pensamiento crítico

1.1 ¿Cuál es la diferencia entre una muestra aleatoria y una muestra aleatoria simple?
1.2 ¿Cuál es la diferencia entre un estudio observacional y un experimento?
1.3 Cuando se realiza un experimento para probar la eficacia de una nueva vacuna, ¿qué es un estudio a ciegas y porque es importante?
1.4 Al realizar un experimento para probar la eficacia de una nueva vacuna, un investigador decidió utilizar los bloques, con un bloque de hombres y un bloque de mujeres. ¿Cómo ayudaría el uso de los bloques al experimento?

II. Determine si la descripción dada corresponde a un estudio observacional o a un experimento

- **Terapia de contacto**. Emily Rosa, de 9 años de edad, se convirtió en la autora de un artículo en el Journal of the American Medical Association, después de poner a prueba a terapeutas de contacto profesionales. Usando una mampara de cartón, ella colocaba la mano encima de la mano del terapeuta, quien debía de identificar la mano que Emily elegió.
- **Tratamiento contra la sífilis**. Ha surgido una gran controversia en torno del estudio de pacientes con sífilis que no recibieron un tratamiento que los habría curado. Su salud fue vigilada por años después de que se descubrió que padecían esa enfermedad.
- **Control de calidad**. La Food and Drug Administration de Estados Unidos elige al azar una muestra de grageas de aspirina Bayer, y mide la exactitud de la cantidad de aspirina en cada gragea.
- **Brazaletes magnéticos**. A los pasajeros de un barco de crucero se les dan brazaletes magnéticos, que aceptan usar en un intento por disminuir o eliminar los efectos del mareo.

III. Identifique el tipo de estudio observacional (transversal, retrospectivo o prospectivo)

- **Psicología del trauma.** Un investigador del hospital Monte Sinaí de la ciudad de Nueva York, planea obtener datos al hacer un seguimiento (hasta el año 2015) a los hermanos de las víctimas que perecieron en el ataque terrorista al Word Trade Center el 11 de septiembre de 2001.
- **Investigación de los conductores en estado de ebriedad.** Un investigador de la Universidad Johns Hopkins obtiene datos sobre los efectos del alcohol al conducir, examinando informes de accidentes automovilísticos de los últimos cinco años.

- **Audiencias televisivas.** Nielsen Media Research Company encuesta a 5000 hogares para determinar la proporción de éstos que sintonizan el programa *Saturday Night Live*.
- **Estadísticas del éxito.** Un economista reúne datos de ingresos al seleccionar y entrevistar actualmente a un grupo de sujetos; después se remonta al pasado para ver si tuvieron la sabiduría de tomar un curso de estadística entre 1980 y 2005.

IV. Identifique el tipo de muestreo que se utilizó: aleatorio, sistemático, de conveniencia, estratificado o por conglomerados

- **Puesto de revisión de sobriedad.** El autor fue un observador en un puesto de revisión de sobriedad de la policía, donde se detenía y entrevistaba a cada quinto conductor. (El autor fue testigo del arresto de un ex alumno).
- **Encuestas de salida.** En épocas de elecciones presidenciales, los medios noticiosos organizan una encuesta de salida, en la que se eligen estaciones de sondeo al azar y se encuesta a todos los votantes conforme abandonan el lugar.
- **Educación y deportes.** Un investigador de la empresa de equipo deportivo Spaulding estudia la relación entre el nivel académico y la participación en cualquier deporte. El investigador hace una encuesta a 40 golfistas, 40 tenistas y 40 nadadores, todos elegidos al azar.

V. Muestras aleatorias y muestras aleatorias simples.

- **Muestra de conveniencia.** Un profesor de estadística obtiene una muestra de estudiantes, al seleccionar a los primeros 10 que entran a su salón de clases. ¿Este plan de muestreo da como resultado una muestra aleatoria? ¿Una muestra aleatoria simple? Explique.
- **Muestra sistemática.** Un ingeniero de control de calidad selecciona cada diezmilésimo dulce M&M que se produce. ¿Este plan de muestreo da como resultado una muestra aleatoria? ¿Una muestra aleatoria simple? Explique.
- **Muestreo de estudiantes.** Un salón de clases tiene 36 estudiantes sentados en seis filas diferentes, con seis estudiantes en cada fila. El profesor tira un dado para determinar una fila, y luego lo tira nuevamente para elegir a un estudiante específico de la fila. Este proceso se repite hasta completar una muestra de 6 estudiantes. ¿Este plan de muestreo da como resultado una muestra aleatoria? ¿Una muestra aleatoria simple? Explique.
- **Muestreo de píldoras de vitaminas.** Un inspector de la Food and Drug Administration de Estados Unidos obtiene píldoras de vitaminas producidas en una hora en la empresa Health Supply Company. Luego las mezcla exhaustivamente y extrae una muestra de 10 píldoras para probar la cantidad exacta del contenido vitamínico. ¿Este plan de muestreo da como resultado una muestra aleatoria? ¿Una muestra aleatoria simple? Explique.

Referencias bibliográficas consultadas y/o enlaces recomendados

- Triola, M. (2010). Estadística (10a.ed.). México: Pearson.

Práctica de Estadística II - N° 02

TEMA N° 1: Estimación de una proporción poblacional

Sección : ………………………..……………….. Docente : Escribir el nombre del docente	Apellidos : ………………………..………………. Nombres : ……………………………………………. Fecha : ……/……/2017 Duración: 45 minutos Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

1. Si n = 400 y X = 25, construya una estimación del intervalo de confianza con lectura en la tabla:
   a. 90%   b. 99%
2. Seis de cada diez familias de cierta ciudad del Valle del Mantaro poseen una computadora personal. Halle el intervalo de confianza para la proporción de familias con una computadora personal, en muestras de 35 familias de esa ciudad, correspondiente al 90% de confianza.
3. El editor de un periódico desea estimar la proporción de periódicos impresos con algún defecto, tal como borraduras en exceso, disposición errónea de las hojas, páginas faltantes o duplicadas. Se selecciona una muestra aleatoria de 200 periódicos, 35 de ellos contienen algún tipo de defecto. Realice e interprete un intervalo de confianza del 90% para la proporción de periódicos impresos que tienen defectos con lectura en la tabla.

4. Una empresa telefónica desea estimar la proporción de hogares en los que se contrataría una línea telefónica adicional. Se seleccionó una muestra aleatoria de 500 hogares. Los resultados indican que a un costo reducido, 135 de los hogares contratarían una línea telefónica adicional. Construya e interprete una estimación del intervalo de confianza del 99% de la proporción poblacional de hogares que contratarían una línea telefónica adicional con lectura en la tabla.
5. Miles de peruanos organizan sus planes de viaje al exterior por Internet. En una encuesta reciente, se reportó que el 25% compra boletos de avión en Internet. Suponga que la encuesta se basó en 180 peruanos que respondieron.
   a. Construya una estimación del intervalo de confianza del 95% para la proporción poblacional de peruanos que compran boletos de avión en Internet con lectura en la tabla.
   b. Construya una estimación del intervalo de confianza del 90% para la proporción poblacional de peruanos que compran boletos de avión en Internet con lectura en la tabla.
   c. ¿Cuál intervalo es más amplio? Explique por qué esto es cierto.
6. Se hace una encuesta entre mujeres trabajadoras en Ecuador. De 1000 mujeres encuestadas, el 55% piensa que las empresas deben reservar los puestos de trabajo durante seis meses o menos para aquellas con permiso de maternidad, y el 45% considera que deberían reservar sus puestos más de seis meses.
   a. Con lectura en la tabla construya un intervalo de confianza del 95% para la proporción de las mujeres trabajadoras de Ecuador quienes creen que las empresas deberían reservar los puestos de trabajo durante seis meses o menos para aquellas con permiso de maternidad.
   b. Con lectura en la tabla construya un intervalo de confianza del 95% para la proporción de las mujeres trabajadoras de Ecuador quienes creen que las empresas deberían reservar los puestos de trabajo durante más de seis meses para aquellas con permiso de maternidad.

Referencias bibliográficas consultadas y/o enlaces recomendados

- Triola, M. (2010). Estadística (10a.ed.). México: Pearson.

Práctica de Estadística II - N° 03

TEMA N° 3: Estimación de una media poblacional

Sección : ………………………..……………….. Docente : Escribir el nombre del docente	Apellidos : ………………………..………………. Nombres : ……………………………………………. Fecha : ……/……/2017 Duración: 45 minutos Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

7.

7. Una asociación de empresarios desea conocer el ingreso promedio anual de los gerentes de Huancayo. En una muestra aleatoria de 50 administradores, la media es de 34 330 nuevos soles y la desviación estándar es 1560 soles.

a) ¿Cuál es la media poblacional? (Estimación puntual).

b) ¿Cuál es el intervalo razonable de valores para la media poblacional? Utilice 1-a=95%. Interprete los resultados obtenidos. Interprete.

8. Estime el cociente de inteligencia promedio de los estudiantes de ingeniería de cierta Universidad, para un nivel de significación del 1%, a partir de una muestra aleatoria de 25 estudiantes donde se obtuvo un promedio de 127 puntos y una desviación estándar poblacional de 5,4 puntos. Interprete su respuesta.
9. Un fabricante de llantas quiere investigar cuál es el tiempo de vida del recubrimiento de sus llantas. En una muestra de 10 llantas (muestra pequeña) que se recorrieron 50 000 kilómetros, se encontró que el espesor medio de recubrimiento restante era de 0,32 pulgadas con una desviación estándar de la muestra es de 0,09 pulgadas.

a) Determine un intervalo de confianza de 95% para la media poblacional. Interprete su respuesta. Interprete.

b) ¿Sería razonable que el fabricante concluyera que después de 50 000 kilómetros la media poblacional del espesor de recubrimiento restante es de 0,30 pulgadas?

10. El gerente de un establecimiento comercial grande quiere determinar la cantidad promedio que gastan los clientes cada vez que visitan el establecimiento. En una muestra de 20 clientes las cantidades gastadas fueron las siguientes: 48; 42; 46; 51; 24; 42; 55; 38; 53; 49; 51; 47; 62; 62; 49; 61; 51; 53; 59 y 44.

a) ¿Cuál es la mejor estimación de la media poblacional?

b) Determine un intervalo de confianza del 99% e interprete el resultado.

c) ¿Sería razonable concluir que la media poblacional es 50?, ¿y que se dice para 60?

11. El dueño de una gasolinera quiere estimar la cantidad promedio de galones de gasolina que vende a sus clientes. De su registro de ventas toma un muestra aleatoria de 60 ventas, y encuentra que la cantidad media de galones vendidos es de 8,60 galones, y la desviación estándar es de 2,30 galones. ¿Cuál es la estimación puntual de la media poblacional? Determina el intervalo de confianza al 99% para la media poblacional de galones vendidos.
12. Un investigador quiere determinar el ingreso medio mensual de los trabajadores de las empresas constructoras. El error al estimar la media es de 100 soles con un nivel de confianza de 90%. Además se conoce mediante un informe anterior que la desviación estándar poblacional es de 1000 soles. ¿De qué tamaño deberá ser la muestra?
13. Se estima que la desviación estándar de una población es 10. Se quiere estimar la media poblacional con un error máximo de 2, y un nivel de confianza del 95% ¿De qué tamaño deberá ser la muestra?
14. Se quiere realizar una encuesta para determinar el número medio de horas que un ingeniero se comunica vía celular con el personal de la empresa. Un estudio piloto indica que la media semanal es 12 horas, con 3 horas de desviación estándar. Se desea que el error máximo al estimar la cantidad media de horas sea un cuarto de hora, para un nivel de confianza del 95%, ¿a cuántos ingenieros habrá de entrevistar?

15. Se realizó un estudio para estimar los costos hospitalarios para víctimas de accidentes que usaban cinturones de seguridad. Veinte casos que se seleccionaron aleatoriamente presentan una distribución que parece tener forma de campana, con una media de 9004 dólares y una desviación estándar de 5629 dólares.

a. Construya el intervalo de confianza del 99% para la media de todos los costos de este tipo.

b. Si usted fuera director de una compañía de seguros que ofrece tarifas más bajas para conductores que usan cinturones de seguridad, y desea un estimado conservador para la peor situación posible, ¿qué cantidad debe aplicar como posible costo hospitalario para una víctima de accidente que utiliza cinturón de seguridad?

16. En una muestra de siete automóviles, se verificó las emisiones de óxido nitroso (en gramos por milla); de los cuales obtuvo siguientes resultados:

0,06; 0,11; 0,16; 0,15; 0,14; 0,08; 0,15

Suponiendo que esta muestra es representativa de los automóviles en circulación, construya un estimado del intervalo de confianza del 98% de la cantidad media de emisiones de óxido nitroso para todos los automóviles. Si la agencia de protección ambiental requiere que las emisiones de óxido nitroso sean menores que 0,165 gramos/milla, ¿sería posible concluir con seguridad que se está cumpliendo tal requisito?

Referencias bibliográficas consultadas y/o enlaces recomendados

- Triola, M. (2010). Estadística (10a.ed.). México: Pearson.

Práctica de Estadística II - N° 04

TEMA N° 3: intervalo de confianza de la varianza y desviación estándar

Sección : ……………………….……………...	Apellidos : ……………………..……………….
	Nombres : …………………………………………….
Docente : Escribir el nombre del docente	Fecha : ……/……/2017 Duración: 45 minutos
	Tipo de Práctica: Individual ( ) Grupal (X)

17.

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

18. Un ingeniero industrial hace una análisis de la producción de la panadería Mantaro Valley que hace donas que se empacan en cajas con etiquetas que dicen contener 12 donas y pesan un total de 42 onzas. Si la variación entre las donas es muy grande, algunas cajas contendrán menos peso (estafando a los consumidores) y otras más (disminuyendo las ganancias).El supervisor de control de calidad encontró que es posible resolver el problema si las donas tienen una media de 3.50 onzas y una desviación estándar de 0.06 onzas o menor. Se seleccionan al azar 18 donas de la línea de producción y se pesan, con los resultados que se dan aquí (en onzas). Construya un intervalo de confianza para la desviación estándar y luego determine si el supervisor de control de calidad está en problemas.

3.43	3.37	3.58	3.50	3.68	3.61
3.61	3,30	3.37	3.58	3.50	3,32
3.42	3.52	3.66	3.50	3.36	3.42

19. Los valores que se listan son tiempos de espera (en minutos) de proveedores de una planta de metal A, con la Técnica A: donde los proveedores forman una sola fila de espera para tres ventanillas del almacén; y por otro lado la Técnica B: donde los proveedores pueden formarse en cualquiera de tres filas diferentes que se alinean a tres ventanillas del almacén.

Técnica A						Técnica B					
6.4	6.6	6.3	7.8	6.2	7.7	7.3	7.7	7.3	7.7	8.0	7.8
6.4	6.6	6.3	7.8	6.2	7.7	7.3	7.7	7.3	7.7	8.0	7.8
6.1	6.2	6.3	6.2	6.3	7.8	6.5	7.9	7.8	7.9	7.7	6.1

20. Una de las maneras de medir el grado de satisfacción de los empleados de una misma categoría en cuanto a la política salarial, es a través de las desviaciones estándar de sus salarios que se portan de manera normal. La fábrica A afirma ser más homogénea en la política salarial que la fábrica B. Para verificar esa afirmación, se escoge una muestra aleatoria de 30 empleados no especializados de A, y 63

de B, obteniendo las dispersiones de 50 y 30 de salario mínimo respectivamente, ¿cuál sería su conclusión si utiliza un intervalo del 95% para el cociente de varianzas?

Referencias bibliográficas consultadas y/o enlaces recomendados

- Triola, M. (2010). Estadística (10a.ed.). México: Pearson.

SEGUNDA UNIDAD

PRUEBA DE HIPOTESIS Y ANALISIS DE VARIANZA

- GUÍA DE PRÁCTICA N° 5: Fundamentos De Prueba de Hipótesis y Varianza
- GUÍA DE PRÁCTICA Nº 6: Prueba de Hipótesis respecto a la desviación estandar
- GUÍA DE PRÁCTICA Nº 7: Prueba de Hipótesis ,datos apareados
- GUÍA DE PRÁCTICA N° 8: Prueba de Hipótesis ,Inferencias a partir de dos muestras

TEXTO N° 6

Fundamentos de la prueba de hipótesis.

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 386 a 395, de 398 a 399

8-1 Panorama general

Las dos actividades principales de la estadística inferencial son el uso de datos para **1.** *estimar* un parámetro poblacional (como se hizo en el capítulo 7), y **2.** probar una hipótesis o afirmación con respecto a un parámetro poblacional (como se hará en este capítulo).

Definición

En estadística, una **hipótesis** es una aseveración o afirmación acerca de una propiedad de una población.

Una **prueba de hipótesis** (o **prueba de significancia**) es un procedimiento estándar para probar una aseveración acerca de una propiedad de una población.

Los siguientes son ejemplos de hipótesis que pueden someterse a prueba por medio de los procedimientos estudiados en este capítulo.

- **Negocios** El encabezado de una nota periodística afirma que la mayoría de los empleados consiguen trabajo por medio de redes de contactos.
- **Medicina** Investigadores médicos aseveran que la temperatura corporal media de adultos sanos no es igual a 98.6°F.
- **Seguridad de aeronaves** La Federal Aviation Administration afirma que el peso promedio de un pasajero de aeronave (con equipaje de mano) es mayor que las 185 libras de hace 20 años.
- **Control de calidad** Cuando se usa equipo nuevo para fabricar altímetros de aviones, los nuevos altímetros son mejores porque la variación en los errores es reducida y, por lo tanto, las lecturas son más consistentes. (En muchas industrias, la calidad de los bienes y servicios a menudo se puede mejorar al reducir la variación).

Los métodos que se presentan en este capítulo se basan en la regla del suceso infrecuente (sección 4-1) para la estadística inferencial, de manera que la repasaremos antes de continuar.

Regla del suceso infrecuente para la estadística inferencial

Si, bajo un supuesto dado, la probabilidad de un suceso observado particular es excepcionalmente pequeña, concluimos que el supuesto probablemente es incorrecto.

Siguiendo esta regla, probamos una aseveración analizando datos muestrales en un intento por distinguir entre resultados que pueden *ocurrir fácilmente por azar* y resultados cuya ocurrencia es *extremadamente improbable debido al azar*. Podemos explicar la ocurrencia de resultados extremadamente improbables al decir que en realidad ha ocurrido un suceso infrecuente o que el supuesto subyacente no es verdadero. Apliquemos este razonamiento en el siguiente ejemplo.

EJEMPLO Selección del género ProCare Industries, Ltd., alguna vez ofreció un producto llamado "Gender Choice", el cual, según aseveraciones publicitarias, permitía a las parejas "incrementar hasta en un 80% sus posibilidades de tener una niña". Suponga que realizamos un experimento con 100 parejas que desean tener niñas, y que las 100 parejas siguen el "sistema casero fácil de usar" de Gender Choice, descrito en el paquete rosa diseñado para concebir niñas. Suponiendo que Gender Choice no tiene efecto alguno, y basados en el sentido común, sin un método estadístico formal, ¿qué debemos concluir acerca del supuesto de que Gender Choice no tiene efecto alguno, si 100 parejas lo utilizaron y tuvieron 100 bebés, de los cuales

a. 52 fueron niñas?

b. 97 fueron niñas?

SOLUCIÓN

a. Generalmente esperamos que nazcan alrededor de 50 niñas por cada 100 nacimientos. El resultado de 52 niñas es cercano a 50, por lo que no debemos concluir que el producto Gender Choice es eficaz. El resultado de 52 niñas podría ocurrir fácilmente por azar, de manera que no existe evidencia suficiente para afirmar que Gender Choice sea eficaz.

b. Es extremadamente improbable que el resultado de 97 niñas en 100 nacimientos suceda por azar. Nosotros podríamos explicar el nacimiento de 97 niñas de dos maneras: o se trata de un evento *extremadamente* infrecuente que ha ocurrido por azar, o Gender Choice es eficaz. La probabilidad extremadamente baja de que resulten 97 niñas sugiere que Gender Choice es eficaz.

El aspecto central del ejemplo anterior es que debemos concluir que el producto es eficaz sólo si obtenemos *significativamente* más niñas de las que esperaríamos normalmente. Aun cuando los resultados de 52 niñas y 97 niñas están "por arriba del promedio", el resultado de 52 niñas no es significativo, mientras que el de 97 niñas es un resultado significativo.

Este breve ejemplo ilustra el método básico utilizado en la prueba de hipótesis. El método formal incluye una variedad de términos y condiciones convencionales incorporados en un procedimiento organizado. Le sugerimos que inicie el estudio de este capítulo con la lectura de las secciones 8-2 y 8-3, de manera informal, para tener una idea general de estos conceptos, y que después lea nuevamente la sección 8-2, ahora con mayor detenimiento, para familiarizarse con la terminología.

8-2 Fundamentos de la prueba de hipótesis

Concepto clave En esta sección se presentan los componentes individuales de una prueba de hipótesis, y las siguientes secciones utilizan esos componentes en procedimientos detallados. Es necesario comprender el papel de los siguientes componentes: hipótesis nula, hipótesis alternativa, estadístico de prueba, región crítica, nivel de significancia, valor crítico, valor P, error tipo I y error tipo II. En la parte 1 de esta sección se presentan los conceptos básicos; es importante comprenderlos antes de considerar el concepto de potencia de una prueba, que se analiza en la parte 2.

Parte 1: Conceptos básicos de la prueba de hipótesis

A continuación se describen los objetivos de esta sección, los cuales debemos alcanzar antes de considerar el análisis de la potencia en la parte 2.

Objetivos de esta sección

- Dada una aseveración, identificar la hipótesis nula y la hipótesis alternativa, y expresar ambas de forma simbólica.

Detectores de mentiras

¿Por qué no se exige que todos los sospechosos de un crimen sean sometidos a la prueba del detector de mentiras para así prescindir de los juicios? El Council of Scientific Affairs de la American Medical Association afirma que "está establecido que la clasificación de los sujetos con base en la culpabilidad puede realizarse con una precisión del 75 al 97%, pero la tasa de falsos positivos suele ser lo suficientemente alta como para excluir el uso de esta prueba (del polígrafo) como único criterio para determinar la culpabilidad o inocencia". Un "falso positivo" es una indicación de culpabilidad cuando el sujeto en realidad es inocente. Incluso con una precisión tan alta como del 97%, el porcentaje de resultados falsos positivos puede ser del 50%, de manera que la mitad de los sujetos inocentes aparecerían incorrectamente como culpables.

- Dados una aseveración y datos muestrales, calcular el valor del estadístico de prueba.
- Dado un nivel de significancia, identificar el valor (o los valores) crítico(s).
- Dado un valor del estadístico de prueba, identificar el valor P.
- Plantear la conclusión de una prueba de hipótesis en términos sencillos y sin tecnicismos.

Es recomendable estudiar el siguiente ejemplo hasta comprenderlo exhaustivamente. Una vez que lo logre, ya habrá captado el principal concepto de la estadística.

EJEMPLO Selección del género y probabilidad Vamos a referirnos de nuevo a los empaques color rosa de Gender Choice. ProCare Industries afirmaba que las parejas que usaban los productos de Gender Choice con empaque color rosa tendrían niñas en una proporción mayor al 50% o 0.5. Consideremos de nuevo un experimento en el que 100 parejas usan Gender Choice en un intento por concebir una niña y supongamos que los 100 bebés incluyen exactamente 52 niñas. Procederemos a formalizar parte del análisis, pero hay dos aspectos que pueden ser confusos:

1. **Suponga que $p = 0.5$:** Al tratar de determinar si 52 niñas en 100 nacimientos representan una evidencia de la eficacia de Gender Choice, suponemos que $p = 0.5$, de manera que podemos determinar si el resultado de 52 niñas puede ocurrir fácilmente por azar (sin efecto del tratamiento) o si es improbable que este resultado ocurra por azar (de manera que el tratamiento sea efectivo).
2. **Use P (52 o más niñas):** Al determinar si el resultado de "52 niñas" puede ocurrir al azar, utilice la probabilidad de *52 o más niñas*. [Repase la sección "Uso de las probabilidades para determinar resultados infrecuentes" de la sección 5-2, donde señalamos que "x éxitos en n ensayos es un número *inusualmente alto* de éxitos si $P(x \text{ o más}) \leq 0.05$"].

En circunstancias normales, la proporción de niñas es $p = 0.5$, de manera que la aseveración de que Gender Choice es eficaz puede expresarse como $p > 0.5$. Respaldamos la aseveración de que $p > 0.5$ sólo si un resultado como el de 52 niñas es improbable (con una escasa probabilidad, como menor que o igual a 0.05). Si se utiliza una distribución normal como aproximación de la distribución binomial (véase la sección 6-6), encontramos que P(52 o más niñas en 100 nacimientos) $= 0.3821$. La figura 8-1 muestra que, con una probabilidad de 0.5, el resultado de 52 niñas en 100 nacimientos no es infrecuente, de manera que *no* rechazamos el azar como una explicación razonable. Concluimos que la proporción de niñas nacidas de parejas que usan Gender Choice *no* es significativamente mayor que el número que esperaríamos por el azar. He aquí los aspectos clave de este ejemplo:

- Aseveración: En las parejas que utilizan Gender Choice, la proporción de niñas es $p > 0.5$.
- Supuesto de trabajo: La proporción de niñas es $p = 0.5$ (sin efecto del Gender Choice).
- La muestra dio por resultado 52 niñas de entre 100 nacimientos, por lo tanto, la proporción muestral es $\hat{p} = 52/100 = 0.52$.
- Suponiendo que $p = 0.5$, empleamos una distribución normal como aproximación de la distribución binomial para calcular que P(al menos 52 niñas en 100 nacimientos) $= 0.3821$.

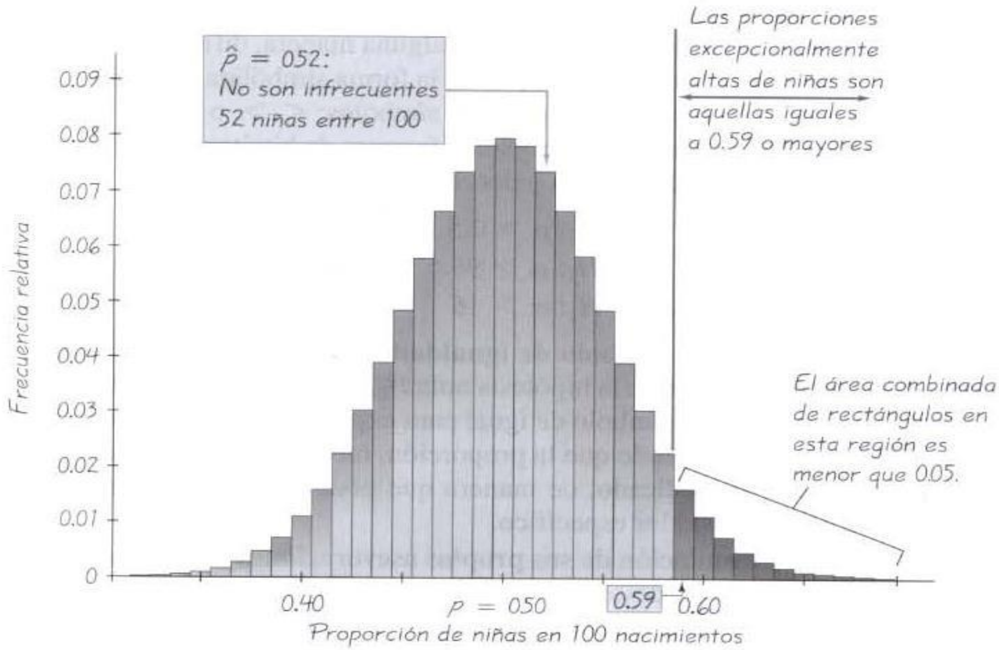


Figura 8-1 Distribución muestral de proporciones de niñas en 100 nacimientos

- Existen dos explicaciones posibles para el resultado de 52 niñas en 100 nacimientos: o bien ocurrió un suceso aleatorio (con una probabilidad de 0.3821), o la proporción de niñas nacidas de parejas que usan Gender Choice es mayor que 0.5. Como la probabilidad de obtener al menos 52 niñas por azar es tan alta (0.3821), consideramos que el azar es una explicación razonable. No existe evidencia suficiente para sustentar la aseveración de que Gender Choice es eficaz para concebir más niñas que lo esperado por el azar. (En realidad fue este tipo de análisis el que condujo a que Gender Choice fuera retirado del mercado).

En la sección 8-3 describiremos los pasos específicos que se utilizan en la prueba de hipótesis; antes de ello, describamos los componentes de una **prueba de hipótesis** formal o **prueba de significancia**. Estos términos suelen emplearse en una gran variedad de disciplinas cuando se requieren métodos estadísticos.

Componentes de una prueba de hipótesis formal

Hipótesis nula y alternativa

- La **hipótesis nula** (denotada por H_0) es la afirmación de que el valor de un parámetro de población (como una proporción, media o desviación estándar) es *igual a* un valor aseverado. Las siguientes son hipótesis nulas típicas del tipo considerado en este capítulo:

$$H_0: p = 0.5 \qquad H_0: \mu = 98.6 \qquad H_0: \sigma = 15$$

La hipótesis nula se prueba en forma directa, en el sentido de que suponemos que es verdadera, y llegamos a una conclusión para rechazar H_0 o no rechazar H_0.

- La **hipótesis alternativa** (denotada por H_1 o H_a o H_A) es la afirmación de que el parámetro tiene un valor que, de alguna manera, difiere de la hipótesis nula. Para los métodos de este capítulo, la forma simbólica de la hipótesis alternativa debe emplear alguno de estos símbolos: $<$, $>$, o bien, $\neq$. A continuación se presentan nueve ejemplos diferentes de hipótesis alternativas que incluyen proporciones, medias y desviaciones estándar:

Proporciones:	$H_1: p > 0.5$	$H_1: p < 0.5$	$H_1: p \neq 0.5$
Medias:	$H_1: \mu > 98.6$	$H_1: \mu < 98.6$	$H_1: \mu \neq 98.6$
Desviaciones estándar:	$H_1: \sigma > 15$	$H_1: \sigma < 15$	$H_1: \sigma \neq 15$

Nota sobre el uso del símbolo de igualdad en H_0: Algunos libros de texto utilizan los símbolos $\leq$ y $\geq$ en la hipótesis nula H_0, pero la mayoría de las revistas científicas emplean sólo el símbolo de igual para expresar igualdad. Realizamos la prueba de hipótesis suponiendo que la proporción, media o desviación estándar es *igual a* algún valor especificado, de manera que podemos trabajar con una sola distribución teniendo un valor específico.

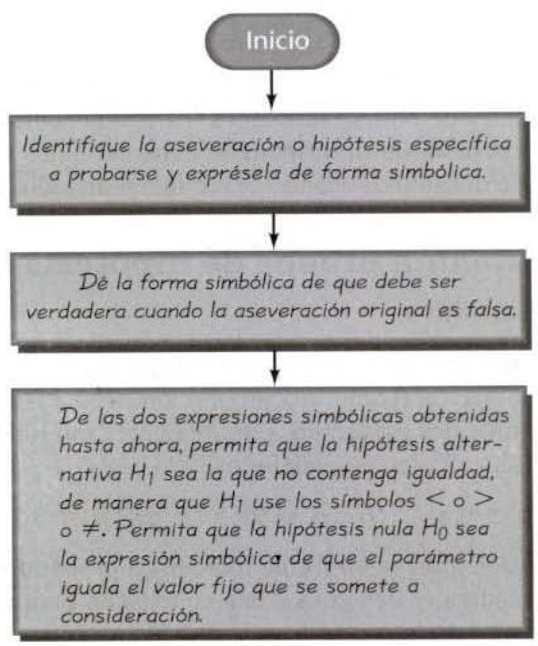
Nota sobre la formulación de sus propias aseveraciones (hipótesis): Si usted está realizando un estudio y desea emplear una prueba de hipótesis para *sustentar* su aseveración, ésta debe redactarse de tal manera que se convierta en la hipótesis alternativa. Esto quiere decir que su aseveración debe expresarse utilizando sólo estos símbolos: $<$, $>$, o bien, $\neq$). No puede utilizar una prueba de hipótesis para sustentar la aseveración de que algún parámetro es *igual a* algún valor especificado.

Por ejemplo, si usted ha creado un método de selección del género, que aumenta la probabilidad de concebir una niña, redacte su aseveración como $p > 0.5$, para que ésta pueda ser sustentada. (En el contexto de tratar de sustentar la meta de la investigación, la hipótesis alternativa en ocasiones se conoce como la *hipótesis de investigación*). Para el propósito de la prueba, usted supondrá que $p = 0.5$, pero usted esperará que $p = 0.5$ sea rechazada para que $p > 0.5$ se sustente.

Nota sobre la identificación de H_0 y H_1: La figura 8-2 resume los procedimientos para identificar las hipótesis nula y alternativa. Observe que la afirmación

Figura 8-2

Identificación de H_0 and H_1



original puede convertirse en la hipótesis nula, en la hipótesis alternativa o tal vez no corresponda con exactitud a ninguna de las dos.

Por ejemplo, en ocasiones probamos la validez de la aseveración de alguien más, como la afirmación de la Coca-Cola Bottling Company de que "la cantidad media de Coca-Cola en las latas es de al menos 12 onzas". Esta afirmación puede expresarse en símbolos tales como $\mu \geq 12$. En la figura 8-2 vemos que si la aseveración original es falsa, entonces $\mu < 12$. La hipótesis alternativa se vuelve $\mu < 12$, pero la hipótesis nula es $\mu = 12$. Podremos considerar la aseveración original después de determinar si existe suficiente evidencia para rechazar la hipótesis nula de $\mu = 12$.

EJEMPLO Identificación de las hipótesis nula y alternativa Remítase a la figura 8-2 y utilice las aseveraciones para expresar las hipótesis nula y alternativa de forma simbólica.

a. La proporción de empleados que consiguen trabajo por medio de una red de contactos es mayor que 0.5.

b. El peso medio de los pasajeros de avión, con su equipaje de mano, es a lo sumo de 195 libras (la cifra que la Federal Aviation Administration difunde actualmente).

c. La desviación estándar de las puntuaciones de CI de actores es igual a 15.

SOLUCIÓN Véase la figura 8-2, que muestra el procedimiento de los tres pasos.

a. En el paso 1 de la figura 8-2, expresamos la aseveración dada como $p > 0.5$. En el paso 2 observamos que si $p > 0.5$ es falso, entonces $p \leq 0.5$ debe ser verdadero. En el paso 3 vemos que la expresión $p > 0.5$ no contiene igualdad, por lo que permitimos que la hipótesis alternativa H_1 sea $p > 0.5$, y dejamos que H_0 sea $p = 0.5$.

b. En el paso 1 de la figura 8-2, expresamos "una media de a lo sumo 195 libras" en símbolos como $\mu \leq 195$. En el paso 2 observamos que si $\mu \leq 195$ es falso, entonces $\mu > 195$ debe ser verdadero. En el paso 3 vemos que la expresión $\mu > 195$ no contiene igualdad, por lo que permitimos que la hipótesis alternativa H_1 sea $\mu > 195$ y que H_0 sea $\mu = 195$.

c. En el paso 1 de la figura 8-2 expresamos la aseveración dada como $\sigma = 15$. En el paso 2 observamos que si $\sigma = 15$ es falso, entonces $\sigma \neq 15$ debe ser verdadero. En el paso 3, permitimos que la hipótesis alternativa sea $\sigma \neq 15$, y que H_0 sea $\sigma = 15$.

Estadístico de prueba

- El **estadístico de prueba** es un valor que se utiliza para tomar la decisión sobre la hipótesis nula, y se calcula convirtiendo al estadístico muestral (como la proporción muestral $\hat{p}$, la media muestral $\bar{x}$, o la desviación estándar muestral s) en una puntuación (como z, t o χ^2), bajo el supuesto de que la hipótesis nula es verdadera. En este capítulo empleamos los siguientes estadísticos de prueba:

Estadístico de prueba para proporciones $$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Un tamaño muestral grande no es lo suficientemente bueno

Los datos muestrales sesgados no deben emplearse para hacer inferencias, sin importar cuán grande sea la muestra. Por ejemplo, en *Women and Love: A Cultural Revolution in Progress*, Shere Hite basa sus conclusiones en 4500 respuestas que recibió después de enviar por correo 100,000 cuestionarios a diversos grupos de mujeres. Por lo general, una muestra *aleatoria* de 4500 sujetos da buenos resultados, pero la muestra de Hite está sesgada y ha sido criticada por considerarse que en ella tienen excesiva representación las mujeres con fuertes sentimientos acerca de los temas planteados. Como la muestra de Hite está sesgada, sus inferencias no son válidas, aun cuando el tamaño de muestra de 4500 pueda parecer lo suficientemente grande.

Estadístico de prueba para medias $$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{o} \quad t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Estadístico de prueba para desviaciones estándar $$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

El estadístico de prueba para una media usa la distribución normal o la distribución t de Student, dependiendo de los requisitos que se satisfagan. En este capítulo se utilizarán los mismos criterios descritos en la sección 7-4. (Véase la figura 7-6 y la tabla 7-1).

EJEMPLO Cálculo del estadístico de prueba Una encuesta de $n = 703$ empleados seleccionados al azar, reveló que el 61% (o $\hat{p} = 0.61$) de ellos consiguió trabajo por medio de una red de contactos. Calcule el valor del estadístico de prueba para la aseveración de que la mayoría de los empleados (más del 50%) consiguen trabajo por medio de una red de contactos. (En la sección 8-3 veremos que existen supuestos que deben verificarse. Para este ejemplo, suponga que se satisfacen los supuestos requeridos y concéntrese en el cálculo del estadístico de prueba indicado).

SOLUCIÓN El ejemplo anterior demostró que la aseveración da por resultado las siguientes hipótesis nula y alternativa: H_0: $p = 0.5$ y H_1: $p > 0.5$. Como trabajamos bajo el supuesto de que la hipótesis nula es verdadera, con $p = 0.5$, obtenemos el siguiente estadístico de prueba:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.61 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{703}}} = 5.83$$

INTERPRETACIÓN De capítulos previos sabemos que la puntuación z de 5.83 es "poco común" (porque es mayor que 2). Parece que, además de ser mayor que el 50%, el resultado muestral de 61% es *significativamente* mayor que el 50%. Observe la figura 8-3, donde demostramos que la proporción muestral de 0.61 (del 61%) cae dentro del rango de valores considerados significativos, ya que están tan arriba de 0.5 que no es probable que ocurran por azar (suponiendo que la proporción de la población es $p = 0.5$).

Región crítica, nivel de significancia, valor crítico y valor *P*

- La **región crítica** (o **región de rechazo**) es el conjunto de todos los valores del estadístico de prueba que pueden provocar que rechacemos la hipótesis nula. Por ejemplo, observe la región sombreada más oscura en la figura 8-3.
- El **nivel de significancia** (denotado por α) es la probabilidad de que el estadístico de prueba caiga en la región crítica, cuando la hipótesis nula es verdadera. Si el estadístico de prueba cae en la región crítica, rechazamos la hipótesis nula, de manera que α es la probabilidad de cometer el error de rechazar la hipótesis nula cuando es verdadera. Se trata de la misma α presentada en

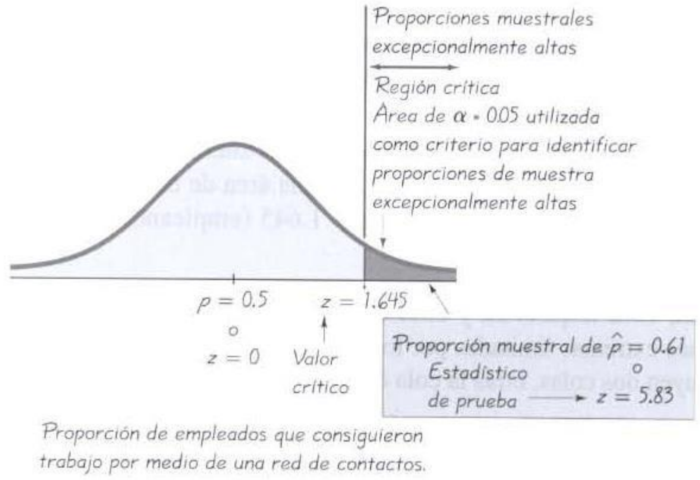


Figura 8-3 Región crítica, valor crítico y estadístico de prueba

la sección 7-2, donde definimos el nivel de confianza para un intervalo de confianza como la probabilidad $1 - \alpha$. Las opciones comunes para α son 0.05, 0.01 y 0.10, aunque la más común es 0.05.

- Un **valor crítico** es cualquier valor que separa la región crítica (donde rechazamos la hipótesis nula) de los valores del estadístico de prueba que no conducen al rechazo de la hipótesis nula. Los valores críticos dependen de la naturaleza de la hipótesis nula, de la distribución muestral que se aplique y del nivel de significancia α. Observe la figura 8-3, donde el valor crítico de $z = 1.645$ corresponde a un nivel de significancia de $\alpha = 0.05$. (Los valores críticos se estudiaron antes, en el capítulo 7).

EJEMPLO Cálculo de valores críticos Con un nivel de significancia de $\alpha = 0.05$, calcule los valores z críticos para cada una de las siguientes hipótesis alternativas (suponiendo que la distribución normal puede emplearse como aproximación de la distribución binomial):

a. $p \neq 0.5$ (de manera que la región crítica está en *ambas* colas de la distribución normal)

b. $p < 0.5$ (de manera que la región crítica está en la cola *izquierda* de la distribución normal)

c. $p > 0.5$ (de manera que la región crítica está en la cola *derecha* de la distribución normal)

SOLUCIÓN

a. Observe la figura 8-4a). Las colas sombreadas contienen una área total de $\alpha = 0.05$, por lo que cada cola contiene una área de 0.025. Empleando los métodos de la sección 6-2, los valores de $z = 1.96$ y $z = -1.96$ separan las regiones de la cola izquierda y la cola derecha. Por lo tanto, los valores críticos son $z = 1.96$ y $z = -1.96$.

Gane $1,000,000 por sus poderes sobrenaturales

El mago James Randi instituyó una fundación educativa que ofrece un premio de $1 millón a quien pueda demostrar poderes paranormales, sobrenaturales u ocultos. Cualquiera que posea un poder como el de adivinar el futuro, percepción extrasensorial (PES) o la habilidad para comunicarse con los muertos, puede ganar el premio si pasa ciertos procedimientos de prueba. Primero se realiza una prueba preliminar y después una formal, pero hasta ahora nadie ha aprobado la prueba preliminar. La prueba formal se diseñaría con métodos estadísticos sólidos, y probablemente incluiría un análisis con una prueba de hipótesis formal. Según la fundación, se consultan "especialistas competentes en estadística cuando se necesita evaluar los resultados o diseñar experimentos". En la página de Internet de la fundación, randi.org, se puede encontrar información sobre la solicitud.

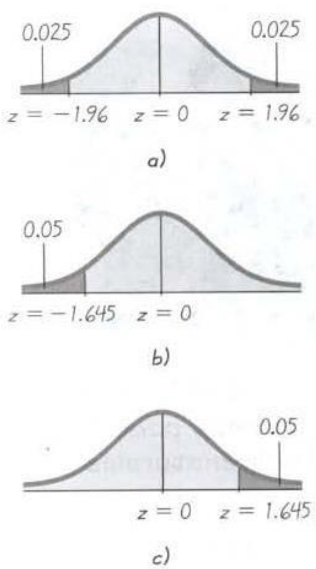


Figura 8-4
Cálculo de valores críticos

b. Observe la figura 8-4b). Con una hipótesis alternativa de $p < 0.5$, la región crítica se encuentra en la cola izquierda. Con una área de cola izquierda de 0.05, se obtiene que el valor crítico es $z = -1.645$ (empleando los métodos de la sección 6-2).

c. Observe la figura 8-4c). Con una hipótesis alternativa de $p > 0.5$, la región crítica está en la cola derecha. Con una área de cola derecha de 0.05, se obtiene que el valor crítico es $z = 1.645$ (empleando los métodos de la sección 6-2).

Dos colas, cola izquierda y cola derecha Las *colas* en una distribución son las regiones extremas limitadas por los valores críticos. Algunas pruebas de hipótesis incluyen dos colas, otras la cola derecha y otras la cola izquierda.

- **Prueba de dos colas:** La región crítica se encuentra en las dos regiones extremas (colas) bajo la curva [como en la figura 8-4a)].
- **Prueba de cola izquierda:** La región crítica se encuentra en la región extrema izquierda (cola) bajo la curva [como en la figura 8-4b)].
- **Prueba de cola derecha:** La región crítica se encuentra en la región extrema derecha (cola) bajo la curva [como en la figura 8-4c)].

En la prueba de dos colas, el nivel de significancia α está dividido equitativamente entre las dos colas que constituyen la región crítica. Por ejemplo, en una prueba de dos colas con un nivel de significancia de $\alpha = 0.05$, existe una área de 0.025 en cada una de las dos colas. En las pruebas de cola derecha o cola izquierda, el área de la región crítica en una cola es α (véase la figura 8-4).

Al examinar la hipótesis alternativa, podemos determinar si la prueba es de cola derecha, de cola izquierda o de dos colas. La cola corresponderá a la región crítica que contiene los valores que entrarán en conflicto, de manera significativa, con la hipótesis nula. En las figuras al margen se resume información útil (véase la figura 8-5), la cual indica que el signo de desigualdad de H_1 señala en la dirección de la región crítica. El símbolo $\neq$ suele expresarse en lenguaje de programación como $< >$, y esto nos recuerda que una hipótesis alternativa, como $p \neq 0.5$ corresponde a una prueba de dos colas.

- El **valor *P*** (o **valor *p*** o **valor de probabilidad**) es la probabilidad de obtener un valor del estadístico de prueba que sea *al menos tan extremo* como el que representa a los datos muestrales, suponiendo que la hipótesis nula es verdadera. La hipótesis nula se rechaza si el valor P es muy pequeño, tanto como 0.05 o menos. Los valores P se calculan con el procedimiento resumido en la figura 8-6.

Decisiones y conclusiones

El procedimiento convencional de prueba de hipótesis requiere que siempre probemos la hipótesis nula, de manera que nuestra conclusión inicial siempre será una de las siguientes:

1. Rechazo de la hipótesis nula.
2. No rechazo de la hipótesis nula.

Criterio de decisión La decisión de rechazar o no rechazar la hipótesis nula suele realizarse por medio del método tradicional (o método clásico) de prueba de hipótesis, el método del valor P, o bien, la decisión puede basarse en intervalos de confianza. En años recientes, el uso del método del valor P ha aumentado, junto con la inclusión de valores P en los resultados de programas de cómputo.

Método tradicional: *Rechace* H_0 si el estadístico de prueba cae dentro de la región crítica.
No rechace H_0 si el estadístico de prueba no cae dentro de la región crítica.

Método del valor P: *Rechace* H_0 si el valor de $P \leq \alpha$ (donde α es el nivel de significancia, tal como 0.05).
No rechace H_0 si el valor $P > \alpha$.

Otra opción: En vez de usar un nivel de significancia como $\alpha = 0.05$, simplemente identifique el valor P y deje la decisión al lector.

Intervalos de confianza: Como un estimado del intervalo de confianza de un parámetro de población contiene los valores posibles de tal parámetro, rechace la aseveración de que el parámetro de población tiene un valor que no está incluido en el intervalo de confianza.

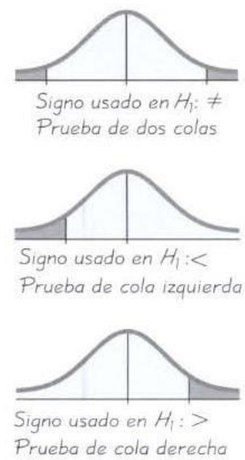


Figura 8-5
Pruebas de dos colas, cola izquierda y cola derecha

EJEMPLO Cálculo de valores *P* Primero determine si las condiciones planteadas dan por resultado una prueba de cola derecha, de cola izquierda o de dos colas; después utilice la figura 8-6 para calcular el valor P, luego saque una conclusión acerca de la hipótesis nula.

a. Se utiliza un nivel de significancia de $\alpha = 0.05$ para probar la aseveración de que $p > 0.25$, y los datos muestrales dan por resultado un estadístico de prueba de $z = 1.18$.

b. Se utiliza un nivel de significancia de $\alpha = 0.05$ para probar la aseveración de que $p \neq 0.25$, y los datos muestrales dan por resultado un estadístico de prueba de $z = 2.34$.

SOLUCIÓN

a. Con la aseveración de que $p > 0.25$, se trata de una prueba de cola derecha. Al utilizar la figura 8-6 para una prueba de cola derecha, vemos que el valor P es el área a la derecha del estadístico de prueba $z = 1.18$. Nos remitimos a la tabla A-2 y encontramos que el área a la *derecha* de $z = 1.18$ es 0.1190. El valor P de 0.1190 es mayor que el nivel de significancia $\alpha = 0.05$, por lo que no rechazamos la hipótesis nula. El valor P de 0.1190 es relativamente grande, lo que indica que los resultados muestrales podrían suceder fácilmente por azar.

b. Con la aseveración de $p \neq 0.25$, se trata de una prueba de dos colas. Al utilizar la figura 8-6 para una prueba de dos colas, observamos que el valor P es *dos veces* el área a la derecha de $z = 2.34$. Nos remitimos a la tabla A-2 y encontramos que el área a la derecha de $z = 2.34$ es 0.0096, de manera que el valor $P = 2 \times 0.0096 = 0.0192$. El valor P de 0.0192 es menor o igual que el nivel de significancia, por lo que rechazamos la hipótesis nula. El pequeño valor P de 0.0192 indica que los resultados muestrales no podrían suceder por azar.

Tabla 8-1	Errores tipo I y tipo II		
		Verdadero estado de las cosas	
		La hipótesis nula es verdadera	La hipótesis nula es falsa
Decisión	Decidimos rechazar la hipótesis nula	**Error tipo I** (rechazo de una hipótesis nula verdadera) α	Decisión correcta
	Decidimos no rechazar la hipótesis nula	Decisión correcta	**Error tipo II** (no rechazo de una hipótesis nula falsa) β

EJEMPLO Redacción de la conclusión final Suponga que un reportero asevera que "más de la mitad" (más del 50%) de los empleados consiguen trabajo por medio de redes de contactos. Esta aseveración de $p > 0.5$ se convierte en la hipótesis alternativa, mientras que $p = 0.5$ se convierte en la hipótesis nula. Además, suponga que la evidencia muestral hace que rechacemos la hipótesis nula de $p = 0.5$. Enuncie la conclusión en términos sencillos y sin tecnicismos.

SOLUCIÓN Remítase a la figura 8-7. La aseveración original no incluye la condición de igualdad, y rechazamos la hipótesis nula. Por lo tanto, la redacción de la conclusión final debe ser la siguiente: "Los datos muestrales sustentan la aseveración de que la mayoría de los empleados consiguen trabajo por medio de redes de contactos".

Errores tipo I y tipo II Cuando probamos una hipótesis nula, llegamos a la conclusión de rechazarla o no rechazarla. Tales conclusiones pueden ser correctas o incorrectas (incluso cuando hacemos todo correctamente). La tabla 8-1 resume los dos distintos tipos de errores que pueden cometerse, junto con los dos tipos de decisiones correctas. Distinguimos entre los dos tipos de errores denominándolos errores tipo I y tipo II.

- **Error tipo I:** El error de rechazar la hipótesis nula cuando en realidad es verdadera. Se utiliza el símbolo α (alfa) para representar la probabilidad de un error tipo I.
- **Error tipo II:** El error de no rechazar la hipótesis nula cuando en realidad es falsa. Se utiliza el símbolo β (beta) para representar la probabilidad de un error tipo II.

Debido a que los estudiantes suelen considerar difícil recordar cuál error es el tipo I y cuál es el error tipo II, recomendamos una herramienta mnemotécnica, como podría ser "revisión no refinada" (**R**e**V**isió**N** **N**o **R**e**F**i**N**ada". Si utilizamos algunas de las consonantes de estas palabras podemos recordar que el error tipo I es RVN: rechazar verdadera nula (hipótesis), mientras que el error tipo II es NRFN: no rechazar falsa nula (hipótesis).

Notación

α (alfa) = probabilidad de un error tipo I (la probabilidad de rechazar la hipótesis nula cuando es verdadera)

β (beta) = probabilidad de un error tipo II (la probabilidad de no rechazar una hipótesis nula cuando es falsa)

EJEMPLO Identificación de errores tipo I y tipo II Suponga que estamos realizando una prueba de hipótesis de la aseveración de que $p < 0.5$. He aquí las hipótesis nula y alternativa

$$H_0: p = 0.5$$

$$H_1: p < 0.5$$

Escriba afirmaciones que identifiquen

a. Un error tipo I.

b. Un error tipo II.

SOLUCIÓN

a. Un error tipo I se comete cuando se rechaza una hipótesis nula verdadera. El siguiente es un error tipo I: concluir que existe evidencia suficiente para sustentar $p < 0.5$, cuando en realidad $p = 0.5$.

b. Un error tipo II se comete al no rechazar la hipótesis nula cuando es falsa. El siguiente es un error tipo II: no rechazar $p = 0.5$ (y, por lo tanto, no sustentar $p < 0.5$) cuando en realidad $p < 0.5$.

TEXTO N° 7

Prueba de una aseveración respecto de una proporción.

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 407 a 411

8-3 Prueba de una aseveración respecto de una proporción

Concepto clave En esta sección se presentan procedimientos completos para probar una hipótesis (o aseveración) respecto de una proporción poblacional. Aquí se utilizan los componentes presentados en la sección 8-2 para el método del valor *P*, el método tradicional y el uso de intervalos de confianza. Además de poner a prueba aseveraciones respecto de proporciones poblacionales, podemos emplear los mismos procedimientos para poner a prueba aseveraciones sobre probabilidades o equivalentes decimales de porcentajes. Los métodos de esta sección utilizan la distribución normal como aproximación de la distribución de probabilidad binomial.

Los siguientes son ejemplos de los tipos de aseveraciones que podremos someter a prueba:

- Más del 50% de los empleados consiguen trabajo por medio de redes de contactos.
- Los sujetos que toman el fármaco Lipitor, que reduce el colesterol, experimentan dolores de cabeza en una proporción mayor que el 7% registrado entre quienes no toman Lipitor.
- El porcentaje de televidentes nocturnos que ven *The Late Show with David Letterman* es igual al 20%.

Las aseveraciones sobre una proporción poblacional suelen probarse utilizando una distribución normal como aproximación de la distribución binomial (véase la sección 6-6). En vez de utilizar exactamente los mismos métodos de la sección 6-6, empleamos una forma diferente, pero equivalente, del estadístico de prueba mostrado a continuación, y no incluimos la corrección por continuidad (debido a que su efecto tiende a ser muy pequeño en muestras grandes). Si no se satisfacen todos los requisitos, debemos utilizar otros métodos que no se describen en esta sección. En esta sección todos los ejemplos y ejercicios incluyen casos en que los requisitos se cumplen, de manera que la distribución muestral de proporciones muestrales puede aproximarse por medio de la distribución normal.

Prueba de aseveraciones sobre una proporción poblacional *p*

Requisitos

1. Las observaciones muestrales son una muestra aleatoria simple.
2. Se satisfacen las condiciones para una *distribución binomial*. (Existe un número fijo de ensayos independientes con probabilidades constantes, y cada ensayo tiene dos categorías de resultados de "éxito" y "fracaso").
3. Se satisfacen las condiciones $np \geq 5$ y $nq \geq 5$, por lo tanto, **la distribución binomial de proporciones muestrales puede aproximarse con una distribución normal, con $\mu = np$ y $\sigma = \sqrt{npq}$** (como se describió en la sección 6-6). Observe que p es la proporción *supuesta* que se utiliza en la aseveración y no la proporción muestral.

continúa

Proceso de aprobación de un fármaco

Lograr la aprobación de la FDA para un fármaco nuevo es costoso y requiere de mucho tiempo. El proceso inicia con un **estudio de fase I**, en el que se prueba la seguridad del fármaco con un grupo pequeño de voluntarios (de 20 a 100). En la **fase II**, se prueba la eficacia del fármaco en ensayos aleatorios con un grupo más grande de sujetos (entre 100 y 300). Esta fase a menudo incluye sujetos asignados al azar a un grupo de tratamiento o a un grupo placebo. En la **fase III**, la meta consiste en comprender mejor la eficacia del fármaco, así como sus efectos adversos. En la fase III generalmente participan de 1000 a 3000 sujetos, y suele requerir varios años de pruebas. Lisa Gibbs escribió en la revista *Money* que "la industria (farmacéutica) afirma que por cada 5000 tratamientos que se someten a prueba, sólo cinco llegan a los ensayos clínicos y sólo uno termina en las farmacias". Los estimados del costo total varían desde $40 millones hasta $1,500 millones.

Notación

n = tamaño de muestra o número de ensayos

$\hat{p} = \dfrac{x}{n}$ (proporción *muestral*)

p = proporción de la población (utilizada en la hipótesis nula)

$q = 1 - p$

Estadístico de prueba para probar una aseveración sobre una proporción

$$z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$$

Valores *P*: Utilice la distribución normal estándar (tabla A-2) y remítase a la figura 8-6.
Valores críticos: Utilice la distribución normal estándar (tabla A-2).

PROBLEMA 8 DEL CAPÍTULO

EJEMPLO Obtención de empleo por medio de redes de contactos El problema del capítulo incluyó los siguientes resultados de encuesta: de 703 empleados elegidos al azar, el 61% obtuvo trabajo por medio de redes de contactos. Utilice los datos muestrales, con un nivel de significancia de 0.05, para probar la aseveración de que la mayoría de los empleados (más del 50%) consiguen su trabajo por medio de redes de contactos. El siguiente es un resumen de la aseveración y de los datos muestrales:

Aseveración: La mayoría de los empleados consigue trabajo por medio de redes de contactos. Es decir, $p > 0.5$.

Datos muestrales: $n = 703$ y $\hat{p} = 0.61$

Daremos soluciones utilizando el método del valor *P*, el método tradicional y los intervalos de confianza. Sin embargo, antes de continuar debemos verificar que se satisfagan los requisitos necesarios.

REQUISITO ✔ Lea con atención los tres siguientes requisitos.

1. Al parecer, la encuesta se realizó utilizando métodos de muestreo adecuados, de manera que podemos suponer que la muestra es aleatoria simple.
2. Hay un número fijo (703) de ensayos independientes con dos categorías (el empleado consiguió trabajo por medio de redes de contactos o no). (Técnicamente, los ensayos no son independientes, pero pueden tratarse como tales utilizando el siguiente lineamiento presentado en la sección 5-3: "Cuando se hace un muestreo sin reemplazo, los sucesos pueden manejarse como si fueran independientes si el tamaño muestral no es mayor que el 5% del tamaño poblacional. Es decir, $n \le 0.05N$".)
3. Los requisitos $np \ge 5$ y $nq \ge 5$ se satisfacen con $n = 703$, $p = 0.5$ y $q = 0.5$. [Obtenemos $np = (703)(0.5) = 351.5 \ge 5$ y $nq = (703)(0.5) = 351.5 \ge 5$].

Con los tres requisitos satisfechos, ahora procedemos a realizar una prueba de hipótesis formal. A continuación se ejemplifican el método del valor *P*, el método tradicional y el uso de intervalos de confianza. ✔

El método del valor *P*

Cuando ponemos a prueba la aseveración $p > 0.5$ del ejemplo anterior, los siguientes pasos corresponden al método del valor *P* de la prueba de hipótesis, que se resumen en la figura 8-8.

Paso 1: La aseveración original en forma simbólica es $p > 0.5$.

Paso 2: El opuesto de la aseveración original es $p \leq 0.5$.

Paso 3: De las dos expresiones simbólicas anteriores, la expresión $p > 0.5$ no contiene igualdad, por lo que se convierte en la hipótesis alternativa. La hipótesis nula es la afirmación de que p iguala el valor fijo de 0.5. Por consiguiente, podemos expresar H_0 y H_1 de la siguiente manera:

$$H_0: p = 0.5$$
$$H_1: p > 0.5$$

Paso 4: En ausencia de cualquier circunstancia especial, seleccionamos $\alpha = 0.05$ para el nivel de significancia.

Paso 5: Para esta prueba se usa la distribución normal y, en virtud de que estamos probando una aseveración acerca de una proporción poblacional p, el estadístico de prueba $\hat{p}$ es relevante y la distribución muestral de las proporciones muestrales $\hat{p}$ se aproxima por medio de una distribución normal.

Paso 6: El estadístico de prueba es $z = 5.83$, que se calcula de la siguiente manera:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.61 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{703}}} = 5.83$$

Ahora obtenemos el valor *P* utilizando el siguiente procedimiento, que se muestra en la figura 8-6:

Prueba de cola derecha:	Valor P = área a la derecha del estadístico de prueba z
Prueba de cola izquierda:	Valor P = área a la izquierda del estadístico de prueba z
Prueba de dos colas:	Valor P = *dos veces* el área de la región extrema limitada por el estadístico de prueba z

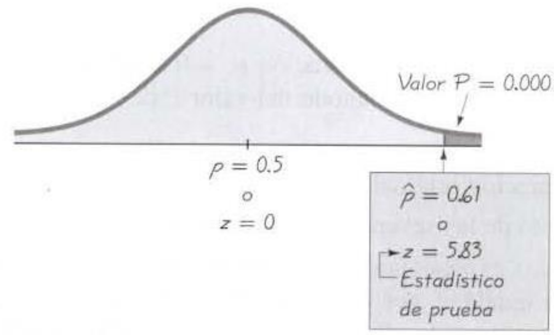
Puesto que la prueba de hipótesis que estamos realizando es de cola derecha con un estadístico de prueba de $z = 5.83$, el valor *P* es el área a la derecha de $z = 5.83$. Si nos remitimos a la tabla A-2, observamos que para los valores de $z = 3.50$ y mayores utilizamos 0.0001 para el área acumulativa a la *derecha* del estadístico de prueba. Por lo tanto, el valor *P* es 0.0001. La figura 8-10 muestra el estadístico de prueba y el valor *P* para este ejemplo.

Paso 7: Puesto que el valor *P* de 0.0001 es menor o igual que el nivel de significancia $\alpha = 0.05$, rechazamos la hipótesis nula.

Paso 8: Concluimos que existe suficiente evidencia muestral para sustentar la aseveración de que la mayoría de los empleados consiguen trabajo por medio de redes de contactos. (Consulte la figura 8-7 para ayudarse a redactar esta conclusión final).

Figura 8-10

Método del valor P



El método tradicional

El método tradicional de prueba de hipótesis se resume en la figura 8-9, y los pasos 1 al 5 son iguales a los pasos 1 al 5 del método del valor P, como se puede ver arriba. Así que continuamos con el paso 6 del método tradicional.

Paso 6: El estadístico de prueba es $z = 5.83$, como vimos en el método del valor P. Ahora calculamos el valor crítico (en vez del valor P). Se trata de una prueba de cola derecha, de manera que el área de la región crítica es una área de $\alpha = 0.05$ en la cola derecha. Si nos remitimos a la tabla A-2 y aplicamos los métodos de la sección 6-2, obtenemos que el valor crítico de $z = 1.645$ se encuentra en el límite de la región crítica. Observe la figura 8-3 de la página 395, que muestra la región crítica, el valor crítico y el estadístico de prueba.

Paso 7: Como el estadístico de prueba se localiza dentro de la región crítica, rechazamos la hipótesis nula.

Paso 8: Concluimos que existe suficiente evidencia muestral para sustentar la aseveración de que la mayoría de los empleados consiguen trabajo por medio de redes de contactos. (Consulte la figura 8-7 para ayudarse a redactar esta conclusión final).

Método del intervalo de confianza

La aseveración de $p > 0.5$ puede probarse con un nivel de significancia de 0.05, construyendo un intervalo de confianza del 90%. (Véase la tabla 8-2. En general, para las pruebas de hipótesis de dos colas, construya un intervalo de confianza con un nivel de confianza de $1 - \alpha$; pero las pruebas de hipótesis de una cola, con nivel de significancia α, requieren un intervalo de confianza con un nivel de confianza de $1 - 2\alpha$).

Ahora utilicemos el método del intervalo de confianza para probar la aseveración de $p > 0.5$, con datos muestrales que consisten en $n = 703$ y $\hat{p} = 0.61$ (como en el problema del capítulo). Si deseamos un nivel de significancia de $\alpha = 0.05$ en una prueba de cola derecha, empleamos un nivel de confianza del 90% con los métodos de la sección 7-2 para obtener este resultado: $0.580 < p < 0.640$. Como tenemos una confianza del 90% de que el valor verdadero de p está contenido dentro de los límites de 0.580 y 0.640, tenemos evidencia suficiente para sustentar la aseveración de que $p > 0.5$.

Advertencia: Cuando se prueban aseveraciones sobre una proporción poblacional, el método tradicional y el método del valor P son equivalentes en el sentido de que siempre dan los mismos resultados, pero el método del intervalo de confianza es un poco diferente. Tanto el método tradicional como el método del valor P utilizan la misma desviación estándar basada en la *proporción aseverada* p, pero el intervalo de confianza emplea una desviación estándar estimada con base en la *proporción muestral* $\hat{p}$. Como consecuencia, es posible que en algunos casos los métodos tradicional y del valor P de prueba de una aseveración sobre una proporción produzcan una conclusión diferente a la del método del intervalo de confianza (véase el ejercicio 29). Una buena estrategia consiste en usar un intervalo de confianza para estimar una proporción poblacional, pero usar el método del valor P o el método tradicional para probar una hipótesis.

Cuando pruebe una aseveración sobre una proporción poblacional p, tenga cuidado en identificar correctamente la proporción muestral $\hat{p}$. En ocasiones la proporción muestral $\hat{p}$ está dada directamente, pero en otros casos debe calcularse. Observe los siguientes ejemplos.

Afirmación dada	Cálculo de $\hat{p}$
El 10% de los automóviles deportivos observados son rojos.	$\hat{p}$ está dada directamente: $\hat{p} = 0.10$
96 hogares encuestados tienen servicio de televisión por cable y 54 no lo tienen.	$\hat{p}$ debe calcularse utilizando $\hat{p} = x/n$. $\hat{p} = \frac{x}{n} = \frac{96}{96 + 54} = 0.64$

TEXTO N° 8

Prueba de una aseveración respecto de una media: o conocida y desconocida

Compilado y adaptado de Mario Triola Estadística PEARSON; 2009

Págs. De 418 a 419 y de 426 a 428

8-4 Prueba de una aseveración respecto de una media: σ conocida

Concepto clave En esta sección se presentan métodos para poner a prueba una aseveración respecto de una media poblacional, cuando se conoce el valor de la desviación estándar poblacional. La siguiente sección presenta métodos para probar una aseveración respecto de una media cuando no se conoce el valor de σ. Aquí se usa la distribución normal con los mismos componentes de las pruebas de hipótesis que se presentaron en la sección 8-2.

Los requisitos, el estadístico de prueba, los valores críticos y el valor P se resumen de la siguiente manera:

Prueba de aseveraciones acerca de una media poblacional (σ desconocida)

Requisitos

1. La muestra es aleatoria simple.
2. Se conoce el valor de la desviación estándar poblacional σ.
3. Se satisface una o ambas de las siguientes condiciones: la población se distribuye normalmente o $n > 30$.

Estadístico de prueba para probar una aseveración sobre una media (σ conocida)

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

Valores P: Utilice la distribución normal estándar (tabla A-2) y remítase a la figura 8-6.

Valores críticos: Utilice la distribución normal estándar (tabla A-2).

Antes de iniciar cualquier procedimiento de prueba de hipótesis, debemos *explorar* primero el conjunto de datos y verificar que los requisitos de la prueba específica se cumplan. Con los métodos descritos en los capítulos 2 y 3, investigue las medidas de tendencia central, la variación y la distribución dibujando una gráfica; calcule la media, la desviación estándar y el resumen de los cinco números; también identifique cualquier valor extremo.

Comerciales

Las cadenas de televisión tienen sus propios departamentos para revisar los comerciales y verificar sus aseveraciones. La National Advertising Division, una rama del Council of Better Business Bureaus, investiga las afirmaciones publicitarias. La Federal Trade Commission y los fiscales locales de distrito también realizan este proceso. Hace algún tiempo, Firestone tuvo que eliminar la afirmación de que sus neumáticos frenaban un 25% más rápido, y Warner Lambert tuvo que gastar $10 millones para informar a sus clientes que Listerine no previene ni cura el resfriado. Muchos anuncios engañosos son retirados de manera voluntaria, y muchos otros escapan al escrutinio simplemente porque los mecanismos regulatorios no pueden revisar una cantidad tan grande de comerciales.

8-5 Prueba de una aseveración respecto de una media: σ desconocida

Concepto clave En la sección anterior se presentaron métodos para probar una aseveración sobre una media poblacional, cuando se conoce el valor de la desviación estándar poblacional σ. En esta sección se presentan métodos para probar una aseveración respecto de una media poblacional cuando se desconoce el valor de σ. Los métodos de esta sección son muy prácticos y realistas, porque generalmente se desconoce σ. Aquí se utiliza la distribución t de Student, que se presentó en la sección 7-4. Los requisitos, el estadístico de prueba, el valor P y los valores críticos se resumen a continuación.

Prueba de aseveraciones acerca de una media poblacional (σ desconocida)

Requisitos

1. La muestra es aleatoria simple.
2. Se *desconoce* el valor de la desviación estándar poblacional σ.
3. Se satisfacen una o ambas de las siguientes condiciones: la población se distribuye de manera normal o $n > 30$.

Estadístico de prueba para probar una aseveración acerca de una media (σ desconocida)

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

Valores P y valores críticos: Utilice la tabla A-3 y utilice gl $= n - 1$ para el número de grados de libertad. (Véase la figura 8-6 para los procedimientos del cálculo del valor P).

Requisito de normalidad El requisito de una población distribuida normalmente no es estricto, y generalmente podemos tratar a la población como si tuviera una distribución normal después de utilizar los datos muestrales para confirmar que no existen valores extremos y que el histograma tenga una forma que no se aleje mucho de la normalidad. Se dice que esta prueba t es *robusta* con respecto a su alejamiento de la normalidad, lo que significa que la prueba funciona bastante bien si no se aleja demasiado de la forma normal.

Tamaño muestral Usamos el criterio simplificado de $n > 30$ como justificación para tratar la distribución de medias muestrales como una distribución normal, pero el tamaño muestral mínimo realmente depende de qué tanto la distribución poblacional se aleja de una distribución normal. Como no conocemos el valor de σ, lo estimamos con el valor de la desviación estándar muestral s, aunque esto introduce otra fuente de falta de confiabilidad, especialmente en el caso de muestras pequeñas. Para compensar esta falta de confiabilidad adicional, calculamos los valores P y los valores críticos utilizando la distribución t en vez de la distribución normal que se empleó en la sección 8-4, donde se conocía σ. Las siguientes son las propiedades importantes de la distribución t de Student:

Propiedades importantes de la distribución t de Student

1. La distribución t de Student difiere para tamaños de muestra distintos (véase la figura 7-5 en la sección 7-4).
2. La distribución t de Student tiene la misma forma general de campana que la distribución normal estándar; su forma más ancha refleja una mayor variabilidad, lo que se espera cuando se utiliza s para estimar σ.
3. La distribución t de Student tiene una media de $t = 0$ (del mismo modo que la distribución normal estándar tiene una media de $z = 0$).
4. La desviación estándar de la distribución t de Student varía de acuerdo con el tamaño muestral y es mayor que 1 (a diferencia de la distribución normal estándar, que tiene $\sigma = 1$).
5. Conforme aumenta el tamaño muestral n, la distribución t de Student se acerca más a la distribución normal estándar.

Elección de la distribución apropiada

Cuando se prueban aseveraciones acerca de medias poblacionales, en ocasiones se aplica la distribución normal, en otras la distribución t de Student y en algunas no se aplica ninguna de las dos, por lo que debemos utilizar métodos no paramétricos o técnicas *bootstrap* de muestreo. (Los métodos no paramétricos, que no requieren una distribución en particular, se estudian en el capítulo 13; la técnica *bootstrap* de muestreo se describe en el Proyecto tecnológico al final del capítulo 7). Consulte las páginas 354 y 355, donde la figura 7-6 y la tabla 7-1 resumen las decisiones que deben tomarse al elegir entre la distribución normal y la t de Student. En ellas se observa que cuando se prueban aseveraciones acerca de medias poblacionales, la distribución t de Student se aplica en estas condiciones:

Utilice la distribución t de Student cuando se desconozca σ y cuando cualquiera o ambas de las siguientes condiciones se satisfagan:

La población se distribuye normalmente o $n > 30$.

EJEMPLO Control de calidad de los dulces M&M El conjunto de datos 13 del apéndice B incluye los pesos de 13 dulces M&M rojos, elegidos al azar de una bolsa que contiene 465 M&M. A continuación se presentan los pesos (en gramos), los cuales tienen una media de $\bar{x} = 0.8635$ y una desviación estándar de $s = 0.0576$ g. En el empaque se afirma que el peso neto del contenido es 396.9 g, de manera que los M&M deben tener un peso medio de al menos $396.9/465 = 0.8535$ g para dar la cantidad anunciada. Utilice los datos muestrales con un nivel de significancia de 0.05, para probar la aseveración que hizo un gerente de producción de que los M&M tienen en realidad una media mayor que 0.8535 g, de manera que los consumidores están recibiendo más que la cantidad indicada en la etiqueta. Utilice el método tradicional, siguiendo el procedimiento descrito en la figura 8-9.

0.751	0.841	0.856	0.799	0.966	0.859	0.857
0.942	0.873	0.809	0.890	0.878	0.905	

SOLUCIÓN

REQUISITOS ✔ Vea los tres requisitos listados arriba. Se trata de una muestra aleatoria simple; no estamos empleando un valor conocido de σ. El tamaño

La ética en los experimentos

Con frecuencia se pueden obtener datos muestrales encuestando o simplemente observando a miembros seleccionados de la población. Muchas otras situaciones requieren que, de alguna manera, manipulemos circunstancias para obtener datos muestrales. En ambos casos surgen cuestiones éticas. Investigadores en Tuskegee, Alabama, negaron el tratamiento de penicilina eficaz a víctimas de sífilis para poder estudiar la enfermedad. ¡Este experimento continuó por un periodo de 27 años!

¿Los zurdos mueren antes?

Un estudio realizado por los psicólogos Diane Halpern y Stanley Coren recibió una gran atención de los medios de comunicación masiva y generó un gran interés cuando concluyó que las personas zurdas no viven tanto tiempo como las personas diestras. Con base en el estudio, parecía que las personas zurdas viven un promedio de nueve años menos que las diestras. El estudio de Halpern y Coren ha sido criticado por utilizar datos defectuosos, ya que se emplearon datos de segunda mano al encuestar a los parientes de individuos que habían muerto recientemente. El mito de que los zurdos mueren más jóvenes se convirtió en una idea generalizada que ha perdurado durante muchos años. Sin embargo, estudios más recientes indican que los zurdos *no* viven menos que los diestros.

muestral es $n = 13$, que no es mayor que 30, pero no se encuentran valores extremos, y una gráfica cuantilar normal sugiere que los pesos se distribuyen normalmente porque los puntos se acercan a la línea recta y no muestran un patrón sistemático. (En la sección anterior se presenta la gráfica cuantilar normal). Un histograma también sugeriría que los pesos se distribuyen de forma normal. Los requisitos se satisfacen y podemos efectuar la prueba de hipótesis. ✔

Seguiremos el procedimiento tradicional que se resume en la figura 8-9.

Paso 1: La aseveración de que la media es mayor que 0.8535 g se expresa simbólicamente como $\mu > 0.8535$.

Paso 2: La alternativa (en forma simbólica) a la aseveración original es $\mu \leq 0.8535$.

Paso 3: Puesto que la afirmación $\mu > 0.8535$ no contiene la condición de igualdad, se convierte en la hipótesis alternativa. La hipótesis nula es la aseveración de que $\mu = 0.8535$.

$$H_0: \mu = 0.8535$$
$$H_1: \mu > 0.8535 \qquad \textbf{(aseveración original)}$$

Paso 4: Como se especifica en el planteamiento del problema, el nivel de significancia es $\alpha = 0.05$.

Paso 5: Puesto que la aseveración se refiere a la *media poblacional* μ, y como los requisitos para utilizar el estadístico de prueba t se satisfacen, empleamos la distribución t. (Remítase a la figura 7-6 o a la tabla 7-1 para consultar los criterios de elección entre las distribuciones normal y t).

Paso 6: El estadístico de prueba se calcula de la siguiente manera:

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{0.8635 - 0.8535}{\frac{0.0576}{\sqrt{13}}} = 0.626$$

El valor crítico de $t = 1.782$ se obtiene de la tabla A-3. Primero localice el número correcto de grados de libertad en la columna de la izquierda. En este ejemplo utilizamos gl $= n - 1 = 12$. Como se trata de una prueba de cola derecha, con $\alpha = 0.05$, nos remitimos a la columna que indica una área de 0.05 en una cola. El estadístico de prueba y el valor crítico se muestran en la pantalla de resultados de STATDISK que se presenta en la siguiente página.

Paso 7: Como el estadístico de prueba de $t = 0.626$ no se localiza dentro de la región crítica, no rechazamos la hipótesis nula.

INTERPRETACIÓN (Remítase a la figura 8-7 para ayudarse a redactar la conclusión). No se rechaza la hipótesis nula. Con base en los datos muestrales disponibles, no tenemos evidencia suficiente para sustentar la aseveración de que el peso medio de los dulces M&M es mayor que 0.8535 g. Aunque la media muestral de 0.8635 g excede a 0.8535 g, no lo rebasa por una cantidad *significativa*.

TEXTO N° 9

Prueba de una aseveración respecto de una desviación estándar o varianza

Compilado y adaptado de Mario Triola Estadística PEARSON; 2009

Págs. De 436 a 437

8-6 Prueba de una aseveración respecto de una desviación estándar o de una varianza

Concepto clave En esta sección se presentan métodos para probar una aseveración respecto de una desviación estándar poblacional σ o varianza poblacional σ^2. Los métodos de esta sección utilizan la distribución chi cuadrada, que se explicó en la sección 7-5. A continuación se resumen los supuestos, el estadístico de prueba, el valor P y los valores críticos.

Prueba de aseveraciones acerca de σ o σ^2

Requisitos

1. La muestra es aleatoria simple.
2. La población tiene una distribución normal. (Éste es un requisito mucho más estricto que el de una distribución normal cuando se prueban aseveraciones acerca de medias, como en las secciones 8-4 y 8-5).

Estadístico de prueba para probar una aseveración acerca de σ o σ^2

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Valores P y valores críticos: Utilice la tabla A-4, con gl $= n - 1$ para el número de grados de libertad. (La tabla A-4 está basada en *áreas acumulativas de la derecha*).

No utilice los métodos de esta sección con una población que tiene una distribución que se aleja mucho de la normalidad. En las secciones 8-4 y 8-5 vimos que los métodos de prueba de aseveraciones acerca de medias requieren de una población distribuida de forma normal, y que esos métodos funcionan razonablemente

Figura 8-13 Propiedades de la distribución chi cuadrada

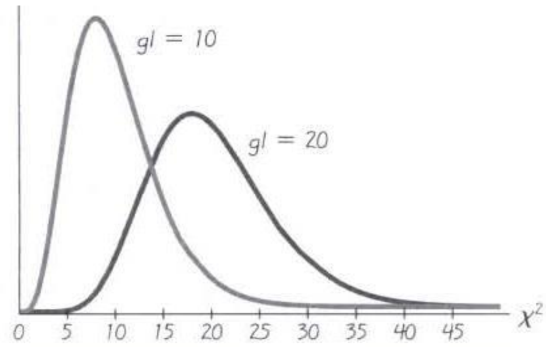


Figura 8-14 Distribución chi cuadrada para gl = 10 y gl = 20

bien siempre y cuando la distribución poblacional no se aleje mucho de la normalidad. Sin embargo, las pruebas de aseveraciones acerca de desviaciones estándar o varianzas no son tan *robustas* con respecto a formas alejadas de la normalidad, lo que quiere decir que no funcionan muy bien con distribuciones que no tienen una distribución normal. Por consiguiente, la condición de una población distribuida normalmente es un requisito mucho más estricto en esta sección.

La distribución chi cuadrada se explicó en la sección 7-5, donde señalamos las siguientes propiedades importantes.

Propiedades de la distribución chi cuadrada

1. Todos los valores de χ^2 son no negativos y la distribución no es simétrica (véase la figura 8-13).
2. Existe una distribución χ^2 diferente para cada número de grados de libertad (véase la figura 8-14).
3. Todos los valores críticos se encuentran en la tabla A-4, utilizando

$$\textbf{grados de libertad} = n - 1$$

La tabla A-4 está basada en áreas acumulativas de la zona *derecha* (a diferencia de los datos de la Tabla A-2 que representan áreas acumulativas de la zona izquierda). Para obtener los valores críticos en la tabla A-4, primero se localiza el renglón correspondiente al número apropiado de grados de libertad (donde gl $= n - 1$). Luego, se utiliza el nivel de significancia α para determinar la columna correcta. Los siguientes ejemplos se basan en un nivel de significancia de $\alpha = 0.05$, pero se puede emplear cualquier otro nivel de significancia de manera similar.

Prueba de cola derecha:	Puesto que el área a la *derecha* del valor crítico es 0.05, localice 0.05 en la parte superior de la tabla A-4.
Prueba de cola izquierda:	Con una área de cola izquierda de 0.05, el área a la *derecha* del valor crítico es 0.95, así que localice 0.95 en la parte superior de la tabla A-4.
Prueba de dos colas:	Divida el nivel de significancia de 0.05 entre la cola derecha y la cola izquierda, de manera que las áreas a la *derecha* de los dos valores críticos sean 0.975 y 0.025, respectivamente. Localice 0.975 y 0.025 en la parte superior de la tabla A-4 (véase la figura 7-10 y el ejemplo en la página 366).

TEXTO N° 10

Inferencias acerca de dos proporciones

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 456 a 462

9-1 Panorama general

En los capítulos 7 y 8 se estudiaron conceptos importantes y fundamentales de la estadística inferencial: métodos para *estimar* valores de parámetros poblacionales y métodos para *probar hipótesis* (o aseveraciones) acerca de parámetros poblacionales. Los capítulos 7 y 8 incluyen métodos que se aplican a una sola muestra, utilizada para hacer una inferencia acerca de un solo parámetro poblacional. Sin embargo, en realidad existen muchas situaciones importantes y significativas en las que es necesario comparar *dos* conjuntos de datos muestrales. Los siguientes son ejemplos típicos incluidos en este capítulo, el cual presenta métodos para utilizar datos muestrales de dos poblaciones de manera que se puedan hacer inferencias acerca de tales poblaciones.

- Poner a prueba la aseveración de que, cuando se trata el síndrome del túnel carpiano, la cirugía es más exitosa que la aplicación de un entablillado.
- Cuando se prueba la eficacia de la vacuna de Salk en la prevención de la poliomielitis paralítica, determinar si el grupo de tratamiento tiene una menor incidencia de poliomielitis que el grupo al que se administró un placebo.
- Cuando se prueba la eficacia del Lipitor, determinar si los sujetos tienen niveles más bajos de colesterol después de tomar el fármaco.
- Dados dos grupos similares de sujetos con depresión bipolar, determinar si el grupo tratado con paroxetina obtiene puntuaciones más bajas en la escala de depresión Hamilton que el grupo que recibió un placebo.

Los capítulos 7 y 8 incluyeron métodos que se aplicaron a proporciones, medias y medidas de variación (desviación estándar y varianza), y en el presente capítulo nos referiremos a esos mismos parámetros. Este capítulo aplica los mismos métodos utilizados en los capítulos 7 y 8 a situaciones que implican comparaciones de dos muestras en vez de una sola muestra.

top 20 9-2 Inferencias acerca de dos proporciones

Concepto clave En esta sección se presentan métodos en los que se utilizan dos proporciones muestrales para construir un estimado de un intervalo de confianza de la diferencia entre las proporciones poblacionales correspondientes o para probar una aseveración acerca de dos proporciones poblacionales. Si bien esta sección se basa en proporciones, podemos utilizar los mismos métodos para tratar con probabilidades o podemos tratar con porcentajes utilizando los equivalentes decimales correspondientes. Por ejemplo, tal vez queramos determinar si existe una diferencia entre el porcentaje de reacciones adversas en un grupo placebo y el porcentaje de reacciones adversas en un grupo de tratamiento con un fármaco. Podemos convertir los porcentajes a sus valores decimales correspondientes y proceder a utilizar los métodos de esta sección.

Cuando se prueba una hipótesis acerca de dos proporciones poblacionales o cuando se construye un estimado del intervalo de confianza de la diferencia entre dos proporciones poblacionales, los requisitos y la notación son los siguientes. Observe que cuando se prueba la hipótesis nula de $p_1 = p_2$, no hay necesidad de estimar los parámetros individuales p_1 y p_2, sino que estimamos su valor común con la proporción muestral agrupada que se describe en el cuadro de resumen.

Requisitos

1. Tenemos proporciones de dos muestras aleatorias simples que son *independientes*. (Las muestras son independientes si los valores muestrales seleccionados de una población no están relacionados ni apareados de alguna forma con los valores muestrales seleccionados de la otra población).
2. Para ambas muestras, el número de éxitos es de al menos 5 y el número de fracasos es de al menos 5.

Notación para dos proporciones

Para la población 1 permitimos que

p_1 = proporción *poblacional*

n_1 = tamaño muestral

x_1 = número de éxitos en la muestra

$\hat{p} = \dfrac{x_1}{n_1}$ (la proporción muestral)

$\hat{q}_1 = 1 - \hat{p}_1$

Se adjuntan los significados correspondientes a p_2, n_2, x_2, $\hat{p}_2$ y $\hat{q}_2$, que provienen de la población 2.

Proporción muestral agrupada

La proporción de la muestra agrupada se denota por $\overline{p}$ y está dada por:

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Denotamos el complemento de $\overline{p}$ como $\overline{q}$, de manera que $\overline{q} = 1 - \overline{p}$.

Estadístico de prueba para dos proporciones (con H_0: $p_1 = p_2$)

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\overline{p}\,\overline{q}}{n_1} + \dfrac{\overline{p}\,\overline{q}}{n_2}}}$$

donde $p_1 - p_2 = 0$ **(supuesto en la hipótesis nula)**

$$\hat{p}_1 = \frac{x_1}{n_1} \quad \text{y} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

$$\overline{q} = 1 - \overline{p}$$

Valor *P*: Utilice la tabla A-2. (Use el valor calculado del estadístico de prueba z y obtenga el valor P siguiendo el procedimiento que se resume en la figura 8-6).

Valores críticos: Utilice la tabla A-2. (Con base en el nivel de significancia α, obtenga valores críticos utilizando los procedimientos de la sección 8-2).

Estimado del intervalo de confianza de $p_1 - p_2$

El estimado del intervalo de confianza de $p_1 - p_2$ es:

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

donde el margen de error E está dado por $E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$

Cálculo del número de éxitos x_1 y x_2 Los cálculos para pruebas de hipótesis e intervalos de confianza requieren que tengamos valores específicos de x_1, n_1, x_2 y n_2. Algunas veces los datos muestrales disponibles incluyen estos números específicos, pero algunas veces es necesario calcular los valores de x_1 y x_2. Por ejemplo, considere la afirmación de que "cuando se hizo una encuesta a 1125 personas, el 47% de ellas dijeron que nunca viajan por avión o que lo hacen pocas veces". A partir de esta afirmación podemos ver que $n_1 = 1125$ y $\hat{p}_1 = 0.47$, pero no tenemos el número real de éxitos x_1. Sin embargo, a partir de $\hat{p}_1 = x_1/n_1$, sabemos que

$$x_1 = n_1 \cdot \hat{p}_1$$

de manera que $x_1 = 1125 \cdot 0.47 = 528.75$. Pero usted no puede tener 528.75 personas que nunca viajan por avión o que lo hacen pocas veces, de manera que el número de éxitos x_1 debe ser el número entero 529. Ahora podemos utilizar $x_1 = 529$ en los cálculos que requieren de este valor. En realidad es bastante sencillo: el 47% de 1125 es $0.47 \cdot 1125$, que da por resultado 528.75, cifra que se redondea a 529.

Pruebas de hipótesis

Ahora consideraremos pruebas de hipótesis acerca de dos proporciones poblacionales, pero sólo nos ocuparemos de pruebas que tienen la hipótesis nula $p_1 = p_2$. (Para aseveraciones de que la diferencia entre p_1 y p_2 es igual a una constante que no sea cero, véase el ejercicio 35 en esta sección). El siguiente ejemplo ayudará a aclarar los papeles que desempeñan x_1, n_1, $\hat{p}_1$, $\bar{p}$, etcétera. En específico, usted debe reconocer que bajo el supuesto de igualdad de proporciones, el mejor estimado de la proporción común se obtiene al combinar ambas muestras en una muestra grande, de manera que $\bar{p}$ es el estimador de la proporción poblacional común.

PROBLEMA 9 DEL CAPÍTULO

EJEMPLO ¿La cirugía es mejor que el entablillado? El problema del capítulo incluye los resultados de una prueba clínica en la que se dio tratamiento a pacientes con síndrome de túnel carpiano, y los resultados se resumen en la tabla 9-1. Utilice los datos muestrales de la tabla 9-1, con un nivel de significancia de 0.05 para probar la aseveración de que la tasa de éxito de la cirugía es mejor que la tasa de éxito del entablillado.

SOLUCIÓN

REQUISITOS ✔ Primero debemos verificar que se satisfagan los requisitos necesarios. Dado el diseño del experimento, es razonable suponer que se trata de una muestra aleatoria simple. Además, el grupo de tratamiento con cirugía es independiente del grupo de tratamiento con entablillado. Para el segundo requisito, observe que el grupo de tratamiento con cirugía tiene 67 éxitos en 73 pacientes, de manera que existen 6 fracasos. Por lo tanto, el grupo de tratamiento

El margen de error líder

Los autores Stephen Ansolabehere y Thomas Belin escribieron lo siguiente en su artículo "Poll Faulting" (revista *Chance*): "Nuestra mayor crítica al reporte de los resultados de la encuesta es para el margen de error de una proporción individual (±3%), tomando en cuenta que la atención de los medios de comunicación está claramente dirigida al *liderazgo* o delantera de un candidato". Señalan que el liderazgo es realmente la *diferencia* entre dos proporciones $(p_1 - p_2)$ y explican cómo desarrollaron la siguiente regla práctica: el liderazgo es aproximadamente $\sqrt{3}$ veces más grande que el margen de error para cualquier proporción individual. Para una encuesta típica de preelección, un margen de error reportado de ±3% se convierte en alrededor de ±5% para el liderazgo de un candidato sobre el otro. Los autores señalan que debe reportarse el margen de error para el liderazgo.

con cirugía tiene al menos 5 éxitos y al menos 5 fracasos. Por otra parte, el grupo de tratamiento con entablillado tiene 60 éxitos en 83 pacientes, de manera que el número de fracasos es 23. Por lo tanto, el grupo de tratamiento con entablillado tiene al menos 5 éxitos y al menos 5 fracasos. Para cada una de las dos muestras revisamos que tengan al menos 5 éxitos y al menos 5 fracasos. La verificación de los requisitos se completó con éxito y procedemos con la prueba de hipótesis. ✔

Ahora utilizaremos el método del valor P para la prueba de hipótesis, tal como se resume en la figura 8-8. En los siguientes pasos estipulamos que los pacientes sometidos a cirugía constituyen la muestra 1, y que los pacientes tratados con entablillado constituyen la muestra 2.

Paso 1: La aseveración de una mayor proporción de éxitos en el grupo de tratamiento con cirugía se expresa como $p_1 > p_2$.

Paso 2: Si $p_1 > p_2$ es falso, entonces $p_1 \leq p_2$.

Paso 3: Puesto que nuestra aseveración de $p_1 > p_2$ no contiene igualdad, se convierte en la hipótesis alternativa. La hipótesis nula es la afirmación de igualdad, entonces tenemos

$$H_0: p_1 = p_2 \qquad H_1: p_1 > p_2 \qquad \textbf{(aseveración original)}$$

Paso 4: El nivel de significancia es $\alpha = 0.05$.

Paso 5: Utilizaremos la distribución normal (con el estadístico de prueba dado con anterioridad) como una aproximación de la distribución binomial. Estimamos el valor común de p_1 y p_2 con el estimado de la muestra agrupada $\overline{p}$, calculado como se indica a continuación, con espacios decimales adicionales para minimizar los errores de redondeo en cálculos posteriores.

$$\overline{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{67 + 60}{73 + 83} = 0.81410256$$

Con $\overline{p} = 0.81410256$, se deduce que $\overline{q} = 1 - 0.81410256 = 0.18589744$.

Paso 6: Ahora podemos calcular el valor del estadístico de prueba.

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\overline{p}\,\overline{q}}{n_1} + \frac{\overline{p}\,\overline{q}}{n_2}}}$$

$$= \frac{\left(\frac{67}{73} - \frac{60}{83}\right) - 0}{\sqrt{\frac{(0.81410256)(0.18589744)}{73} + \frac{(0.81410256)(0.18589744)}{83}}}$$

$$= 3.12$$

El valor P de 0.0009 se calcula como sigue: se trata de una prueba de cola derecha, por lo que el valor P es el área ubicada a la derecha del estadístico de prueba $z = 3.12$ (como se indica en la figura 8-6). Remítase a la tabla A-2 y encuentre que el área a la izquierda del estadístico de prueba $z = 3.12$ es 0.9991; por lo tanto, el valor P es

El autor como testigo

El autor fue requerido para testificar en la Suprema Corte del estado de Nueva York en el caso de un ex alumno que impugnaba una reelección perdida ante la oficina del ayuntamiento del condado Dutchess. El autor testificó utilizando la estadística para demostrar que el comportamiento de votación en un distrito impugnado fue significativamente diferente del comportamiento en todos los demás distritos. Cuando el abogado de la oposición se refirió a los resultados de un intervalo de confianza, preguntó si el 5% de error (de un intervalo de confianza del 95%) podría añadirse a los tres puntos porcentuales del margen de error para obtener un error total de 8%, indicando de esa forma que no entendía el concepto básico de un intervalo de confianza. El juez citó el testimonio del autor, apoyó la demanda del ex alumno y ordenó una nueva elección en el distrito impugnado. Este juicio después fue anulado por la corte de apelación con base en que las irregularidades de la votación debían haberse impugnado antes de la elección, no después.

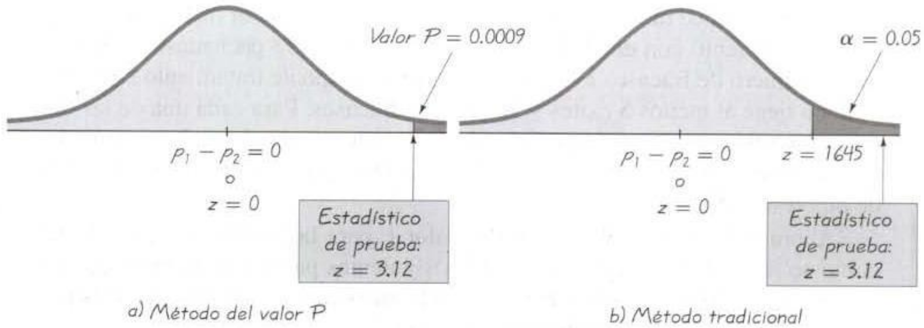


Figura 9-1 Prueba de la aseveración de que la cirugía es mejor que el entablillado

$1 - 0.9991 = 0.0009$. El estadístico de prueba y el valor P se presentan en la figura 9-1*a*).

Paso 7: Puesto que el valor P de 0.0009 es menor que el nivel de significancia de $\alpha = 0.05$, rechazamos la hipótesis nula de $p_1 = p_2$.

INTERPRETACIÓN Debemos retomar la afirmación original de que el porcentaje de éxito con la cirugía es mayor que el porcentaje de éxito con el entablillado. Puesto que rechazamos la hipótesis nula, concluimos que existe evidencia suficiente para sustentar la aseveración de que la proporción de éxitos con la cirugía es mayor que con el entablillado. (Véase la figura 8-7 para la redacción de la conclusión final). Los investigadores tenían una justificación al escribir en un artículo científico que "el tratamiento con cirugía abierta de liberación del túnel carpiano produjo mejores resultados que el tratamiento de entablillado de la muñeca en los pacientes con STC (síndrome del túnel carpiano)". Con base en el ensayo clínico y el análisis estadístico posterior de los resultados, los médicos tienen mayores fundamentos cuando recomiendan un tratamiento para el síndrome del túnel carpiano, una afección que a menudo es muy dolorosa.

Método tradicional de prueba de hipótesis

El ejemplo anterior ilustra el método del valor P para prueba de hipótesis, pero sería bastante fácil utilizar en su lugar el método tradicional. En el paso 6, en vez de calcular el valor P, podríamos calcular el valor crítico. Con un nivel de significancia de $\alpha = 0.05$ en una prueba de cola derecha, basada en una distribución normal, remítase a la tabla A-2 para encontrar que una área de $\alpha = 0.05$ en la cola derecha corresponde al valor crítico de $z = 1.645$. Consulte la figura 9-1*b*), donde podemos observar que el estadístico de prueba se localiza en la región crítica limitada por el valor crítico de $z = 1.645$. Una vez más, rechazamos la hipótesis nula y concluimos que hay suficiente evidencia para sustentar la aseveración de que la cirugía es más exitosa que el entablillado.

Intervalos de confianza

Podemos construir un estimado del intervalo de confianza de la diferencia entre las proporciones poblacionales ($p_1 - p_2$) utilizando el formato anterior. Si un estimado del intervalo de confianza de $p_1 - p_2$ no incluye a 0, tenemos evidencia que sugiere que p_1 y p_2 tienen valores diferentes. Sin embargo, la desviación estándar utilizada

¿Ayuda la aspirina a prevenir ataques cardiacos?

En un estudio reciente realizado a 22,000 médicos, la mitad recibió dosis normales de aspirina mientras que a la otra mitad se le dieron placebos. El estudio duró seis años, con un costo de \$4.4 millones. Entre quienes tomaron la aspirina, 104 sufrieron ataques cardiacos. De los que tomaron los placebos, 189 sufrieron ataques cardiacos. (Estas cifras están basadas en datos de *Time* y el *New England Journal of Medicine*, vol. 318, núm. 4). Éste es un experimento clásico con un grupo de tratamiento (quienes tomaron la aspirina) y un grupo placebo (los que tomaron tabletas que tenían la apariencia y el sabor de tabletas de aspirina, pero que no contenían aspirina). Podemos utilizar los métodos que se presentan en este capítulo para señalar el hecho de que los resultados muestran una tasa menor estadísticamente significativa de ataques cardiacos en el grupo muestral que tomó aspirina.

para un intervalo de confianza difiere de la que se utiliza para una prueba de hipótesis. Cuando se prueban las aseveraciones acerca de diferencias entre dos proporciones poblacionales, el método tradicional y el método del valor P son equivalentes en el sentido de que siempre producen los mismos resultados, pero el estimado del intervalo de confianza de la diferencia podría sugerir otra conclusión (véase el ejercicio 34). Si se obtienen diferentes conclusiones, se debe a que los métodos tradicional y del valor P utilizan una desviación estándar *exacta* con base en la suposición de que no existe diferencia entre las proporciones poblacionales (como se estableció en la hipótesis nula). Sin embargo, el intervalo de confianza se construye utilizando una desviación estándar que se basa en valores *estimados* de las dos proporciones poblacionales. Como consecuencia, si usted desea estimar la diferencia entre dos proporciones poblacionales, hágalo construyendo un intervalo de confianza, pero si usted desea probar alguna aseveración acerca de dos proporciones poblacionales, utilice el método del valor P o el método tradicional.

Además, *no pruebe la igualdad de dos proporciones poblacionales determinando si existe un traslape entre dos estimados individuales del intervalo de confianza de las dos proporciones poblacionales individuales*. Cuando se compara con el estimado del intervalo de confianza de $p_1 - p_2$, el análisis del traslape de dos intervalos de confianza individuales es más conservador (rechazando la igualdad con menos frecuencia) y tiene menos potencia (porque es menos probable rechazar $p_1 = p_2$ cuando en realidad $p_1 \neq p_2$). (Véase "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals", de Schenker y Gentleman, *American Statistician*, vol. 55, núm. 3). Consulte el ejercicio 33.

PROBLEMA 9 DEL CAPÍTULO

EJEMPLO ¿La cirugía es mejor que el entablillado? Utilice los datos muestrales que se presentan en la tabla 9-1 para construir un estimado del intervalo de confianza del 90% de la diferencia entre las dos proporciones poblacionales. (El nivel de confianza del 90% es comparable al nivel de significancia de $\alpha = 0.05$ que se utilizó en la anterior prueba de hipótesis de cola derecha. Véase la tabla 8-2 en la sección 8-2).

SOLUCIÓN

REQUISITO ✔ La solución del ejemplo anterior comienza con la misma verificación de requisitos que se necesita aquí. Se cumple la verificación de requisitos, así que podemos continuar con la construcción del intervalo de confianza. ✔

Con un nivel de confianza del 90%, $z_{\alpha/2} = 1.645$ (de la tabla A-2). Primero calculamos el valor del margen de error E como se indica.

$$E = z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} = 1.645\sqrt{\frac{\left(\frac{67}{73}\right)\left(\frac{6}{73}\right)}{73} + \frac{\left(\frac{60}{83}\right)\left(\frac{23}{83}\right)}{83}} = 0.0966$$

Con $\hat{p}_1 = 67/73 = 0.9178$, $\hat{p}_2 = 60/83 = 0.7299$ y $E = 0.0966$, el intervalo de confianza se evalúa como sigue.

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

$$(0.9178 - 0.7229) - 0.0966 < (p_1 - p_2) < (0.9178 - 0.7229) + 0.0966$$

$$0.0983 < (p_1 - p_2) < 0.292$$

Experimento en relación con la poliomielitis

En 1954 se realizó un experimento para probar la efectividad de la vacuna de Salk como protección contra los devastadores efectos de la poliomielitis. Aproximadamente 200,000 niños recibieron una inyección de solución salina inocua, y otros 200,000 una inyección de la vacuna. El experimento fue "doble ciego" porque los niños inyectados no sabían si estaban recibiendo la vacuna real o el placebo, en tanto que los doctores que aplicaban las inyecciones y evaluaban los resultados tampoco lo sabían. Sólo 33 de los 200,000 niños vacunados padecieron posteriormente poliomielitis paralítica, mientras que 115 de los 200,000 inyectados con la solución salina padecieron posteriormente poliomielitis paralítica. Un análisis estadístico de éstos y otros resultados llevó a la conclusión de que la vacuna de Salk realmente era efectiva contra la poliomielitis paralítica.

INTERPRETACIÓN Los límites del intervalo de confianza no contienen a 0, lo que sugiere que existe una diferencia significativa entre las dos proporciones. Además, con base en esos resultados, tenemos una confianza del 90% de que el porcentaje de éxitos con la cirugía es mayor que el porcentaje de éxitos con el entablillado por una cantidad que está entre el 9.83 y el 29.2%.

Fundamentos: ¿Por qué funcionan los procedimientos de esta sección? El estadístico de prueba dado para la prueba de hipótesis se justifica por lo siguiente:

1. Con $n_1p_1 \geq 5$ y $n_1p_1 \geq 5$, la distribución de $\hat{p}_1$ puede aproximarse con una distribución normal con media p_1, una desviación estándar $\sqrt{p_1q_1/n_1}$ y una varianza p_1q_1/n_1. Estas conclusiones se basan en las secciones 6-6 y 7-2, y también se aplican a la segunda muestra.
2. Puesto que $\hat{p}_1$ y $\hat{p}_2$ se aproximan por medio de una distribución normal, $\hat{p}_1 - \hat{p}_2$ también se aproximará por medio de una distribución normal con media $p_1 - p_2$ y varianza

$$\sigma^2_{(\hat{p}_1-\hat{p}_2)} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}$$

(El resultado anterior se basa en esta propiedad: la varianza de las *diferencias* entre dos variables aleatorias independientes es la *suma* de sus varianzas individuales).

3. Puesto que generalmente no se conocen los valores de p_1, q_1, p_2 y q_2, y a partir de la hipótesis nula, suponemos que $p_1 = p_2$ y podemos agrupar (o combinar) los datos muestrales. El estimado combinado del valor común de p_1 y p_2 es $\bar{p} = (x_1 + x_2)/(n_1 + n_2)$. Si reemplazamos p_1 y p_2 por $\bar{p}$ y reemplazamos q_1 y q_2 por $\bar{q} = 1 - \bar{p}$, la varianza del paso 2 nos lleva a la siguiente desviación estándar:

$$\sigma_{(\hat{p}_1-\hat{p}_2)} = \sqrt{\frac{\bar{p}\,\bar{q}}{n_1} + \frac{\bar{p}\,\bar{q}}{n_2}}$$

4. Sabemos ahora que la distribución de $p_1 - p_2$ es aproximadamente normal, con media $p_1 - p_2$ y desviación estándar como se indica en el paso 3, por lo tanto, el estadístico de prueba z tiene la forma dada antes.

La forma del intervalo de confianza requiere una expresión para la varianza diferente de la que se dio en el paso 3. En el paso 3 suponemos que $p_1 = p_2$, pero si no hacemos esta suposición (como en la construcción del intervalo de confianza), estimamos la varianza de $p_1 - p_2$ como

$$\sigma^2_{(\hat{p}_1-\hat{p}_2)} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2} = \frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}$$

y la desviación estándar se convierte en

$$\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

La pena de muerte como correctivo

Un argumento utilizado comúnmente para apoyar la pena de muerte es que ésta desanima a otros individuos de cometer asesinatos. Jeffrey Grogger, de la Universidad de California, analizó los datos sobre los homicidios diarios en ese estado durante cuatro años, en una época en que las ejecuciones eran frecuentes. Entre sus conclusiones, publicadas en el *Journal of the American Statistical Association* (vol. 85, núm. 410), está la siguiente: "El análisis realizado de forma consistente indica que estos datos no sustentan la hipótesis de que la ejecución desanima el asesinato en el corto plazo". Éste es uno de los temas más importantes de política social y los esfuerzos de personas como el profesor Grogger ayudan a disipar las ideas erróneas, de manera que nos permiten tener información precisa para analizar temas como éste.

TEXTO N° 11

Inferencias acerca de dos medias: muestras independientes

Compilado y adaptado de Mario Triola Estadística PEARSON; 2009

Págs. De 469 a 470, 475

9-3 Inferencias acerca de dos medias: Muestras independientes

Concepto clave En esta sección se presentan métodos para utilizar datos muestrales de dos muestras independientes para someter a prueba hipótesis acerca de dos medias poblacionales o para construir estimados de intervalos de confianza de la diferencia entre dos medias poblacionales. En la parte 1 de esta sección se incluyen situaciones en las que se desconocen las desviaciones estándar de las dos poblaciones, entre las cuales no se supone igualdad. La parte 2 incluye otras dos situaciones: **1.** se conocen las dos desviaciones estándar poblacionales; **2.** se desconocen las dos desviaciones estándar poblacionales, pero se supone igualdad entre ellas. Por razones que se explicarán más adelante, es necesario poner especial atención a los métodos que se describen en la parte 1, ya que son los más realistas y tienen un mejor desempeño general.

Parte 1: Muestras independientes con σ_1 y σ_1 desconocidas y sin suposición de igualdad

Comencemos con definiciones formales que distinguen entre muestras *independientes* y *dependientes*.

Definiciones

Dos muestras son **independientes** si los valores muestrales seleccionados de una población no están relacionados, apareados o asociados de alguna manera con los valores muestrales seleccionados de la otra población.

Dos muestras son **dependientes** (o consisten en **datos apareados**) si los miembros de una muestra se pueden usar para determinar los miembros de la otra muestra. [Las muestras que consisten en datos apareados (por ejemplo, datos de esposo/esposa) son dependientes. Además de los datos muestrales apareados, la dependencia también puede ocurrir con muestras relacionadas a través de asociaciones tales como miembros de la familia]. (En este libro usaremos el término *datos apareados*, ya que describe mejor la naturaleza de los datos).

EJEMPLO Prueba de fármaco

Muestras independientes: Se trata a un grupo de sujetos con el fármaco reductor del colesterol Lipitor, mientras que a un segundo grupo separado de sujetos se les da un placebo. Estos dos grupos muestrales son independientes, puesto que los individuos en el grupo de tratamiento no están en ninguna forma apareados o asociados con miembros correspondientes en el grupo placebo.

Datos apareados (o muestras dependientes): Se prueba la eficacia de una dieta utilizando los pesos de los sujetos medidos antes y después de someterse a la dieta. Cada valor "antes" se aparea con el valor "después", puesto que cada par de mediciones antes/después proviene de la misma persona.

Esta sección considera dos muestras independientes, y la siguiente sección se enfoca en datos apareados. Cuando se utilizan dos muestras independientes para probar una aseveración acerca de la diferencia $\mu_1 - \mu_2$, o para construir un estimado del intervalo de confianza de $\mu_1 - \mu_2$, utilice el siguiente cuadro.

Requisitos

1. σ_1 y σ_2 se desconocen y no se hace una suposición sobre la igualdad de σ_1 y σ_2.
2. Las dos muestras son *independientes*.
3. Ambas muestras son *aleatorias simples*.
4. Cualquiera o ambas de estas condiciones se satisfacen: los dos tamaños muestrales son grandes (con $n_1 > 30$ y $n_2 > 30$) o ambas muestras provienen de poblaciones que tienen distribuciones normales. (En muestras pequeñas, el requisito de normalidad es menos estricto, en el sentido de que los procedimientos se comportan bien siempre y cuando no existan valores extremos ni grandes sesgos).

Estadístico de prueba de hipótesis para dos medias: muestras independientes

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Grados de libertad: Cuando calcule valores críticos o valores P, utilice lo siguiente para determinar el número de grados de libertad, denotados por gl. (Si bien estos dos métodos generalmente dan por resultado números diferentes de grados de libertad, la conclusión de una prueba de hipótesis rara vez se ve afectada por la elección del método).

1. En este libro utilizamos el estimado sencillo y conservador: gl = el más pequeño de $n_1 - 1$ y $n_2 - 1$.
2. Los programas de cómputo de estadística por lo regular utilizan el estimado más exacto, pero más difícil, dado en la fórmula 9-1. (Nosotros no utilizaremos la fórmula 9-1 en los ejemplos y ejercicios de este libro).

Fórmula 9-1

$$\text{gl} = \frac{(A + B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}}$$

donde $$A = \frac{s_1^2}{n_1} \quad \text{y} \quad B = \frac{s_2^2}{n_2}$$

Valores *P*: Remítase a la tabla A-3. Utilice el procedimiento resumido en la figura 8-6.
Valores críticos: Remítase a la tabla A-3.

Estimado del intervalo de confianza de $\mu_1 - \mu_2$: muestras independientes

El estimado del intervalo de confianza de la diferencia $\mu_1 - \mu_2$ es

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

donde $$E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

y el número de grados de libertad gl se obtiene como se describió antes para las pruebas de hipótesis. (En este libro utilizamos gl = el menor de $n_1 - 1$ y $n_2 - 1$).

aleatorias independientes es igual a la varianza de la primera variable aleatoria *más* la varianza de la segunda variable aleatoria. Es decir, la varianza de los valores muestrales $\bar{x}_1 - \bar{x}_2$ tiende a igualar $s_1^2/n_1 + s_2^2/n_2$, siempre y cuando $\bar{x}_1$ y $\bar{x}_2$ sean independientes. (Véase el ejercicio 35).

Parte 2: Métodos alternativos

La parte 1 de esta sección se refirió a situaciones en las que no se conocen las dos desviaciones estándar poblacionales, entre las cuales no se supone igualdad. En la parte 2, consideramos otras dos situaciones: **1.** se conocen las dos desviaciones estándar poblacionales; **2.** se desconocen las dos desviaciones estándar poblacionales, pero se supone igualdad entre ellas. Ahora describiremos los procedimientos para estos casos alternativos.

Método alternativo: σ_1 y σ_2 conocidas

En realidad, las desviaciones estándar poblacionales σ_1 y σ_2 casi nunca se conocen, pero si son conocidas, el estadístico de prueba y el intervalo de confianza están basados en una distribución normal y no en una distribución *t*. Veamos el siguiente resumen.

Requisitos

1. Se conocen las dos desviaciones estándar poblacionales.
2. Las dos muestras son *independientes*.
3. Ambas muestras son *aleatorias simples*.
4. Cualquiera de estas condiciones (o ambas) se satisfacen: los dos tamaños muestrales son *grandes* (con $n_1 > 30$ y $n_2 > 30$) o las dos muestras provienen de poblaciones que tienen distribuciones normales. (Para muestras pequeñas, el requisito de normalidad es menos estricto en el sentido de que los procedimientos funcionan bien siempre y cuando no existan valores extremos ni desviaciones de la normalidad demasiado pronunciadas).

Prueba de hipótesis para dos medias: muestras independientes y $\sigma_1 = \sigma_2$ conocidas

Estadístico de prueba: $$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

Valor *P* y valores críticos: Remítase a la tabla A-2.

Estimado del intervalo de confianza de $\mu_1 - \mu_2$: muestras independientes y $\sigma_1 = \sigma_2$ conocidas

Intervalo de confianza: $(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$

donde $$E = z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

TEXTO N° 12

Inferencias a partir de datos apareados.
Comparación de la varianza de dos muestras.

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 484 a 485 y de 495 a 498

9-4 Inferencias a partir de datos apareados

Concepto clave Dos muestras son **dependientes** (o consisten en **datos apareados**) si los miembros de una muestra pueden utilizarse para determinar los miembros de la otra muestra. En esta sección estudiamos métodos para poner a prueba aseveraciones acerca de la diferencia media de datos apareados. Para cada par de datos de valores muestrales, calculamos la diferencia entre los dos valores, y luego utilizamos esas diferencias muestrales para probar aseveraciones acerca de la diferencia poblacional o para construir estimados de intervalos de confianza para la diferencia poblacional. En el capítulo 10 también hablaremos sobre datos apareados, sólo que entonces nos interesaremos en la *asociación* entre dos variables y no en la *media* de las diferencias.

En los datos apareados, existe alguna relación para que cada valor en una muestra se aparee con un valor correspondiente en la otra muestra. A continuación se presentan algunos ejemplos típicos de datos apareados:

- Los datos muestrales son datos apareados de mediciones de colesterol de lipoproteínas de baja densidad (LDL), tomadas antes y después del tratamiento con Lipitor. Ejemplo: LDL antes de Lipitor = 182; LDL después de Lipitor = 155.
- Los datos muestrales son datos apareados de los índices de masa corporal (IMC) del esposo y de la esposa. Ejemplo: IMC del esposo = 25.1; IMC de la esposa = 19.7.
- Los datos muestrales son las estaturas de candidatos ganadores de la presidencia apareados con las estaturas de los candidatos que recibieron el segundo número más alto de votos. Ejemplo: estatura de Truman = 69 pulgadas; estatura de Dewey = 68 pulgadas.

Para tratar con inferencias acerca de medias y datos apareados, abajo se incluyen resúmenes de los requisitos, la notación, el estadístico de prueba de hipótesis y el intervalo de confianza. Puesto que la prueba de hipótesis y el intervalo de confianza utilizan la misma distribución y el mismo error estándar, son equivalentes en el sentido de que arrojan las mismas conclusiones. En consecuencia, la hipótesis nula de que la diferencia de la media es igual a 0 puede probarse determinando si el intervalo de confianza incluye a 0. Una prueba de hipótesis de dos colas con un nivel de significancia de 0.05 puede probarse con un 95% de confianza, mientras que una prueba de hipótesis de una cola con nivel de significancia de 0.05 puede probarse con un intervalo de confianza del 90%. [Véase la tabla 7.2 para casos comunes].

Requisitos

1. Los datos muestrales consisten en datos apareados.
2. Las muestras son aleatorias simples.
3. Cualquiera o ambas de estas condiciones se satisfacen: el número de datos apareados de datos muestrales es grande ($n > 30$) o los pares de valores tienen diferencias que se toman de una población con una distribución aproximadamente normal. (Si existe una desviación radical de la distribución normal, no debemos utilizar los métodos que se estudian en esta sección, pero quizá podamos utilizar los métodos no paramétricos que se analizan en el capítulo 13).

Notación para datos apareados

d = diferencia individual entre los dos valores en un solo dato apareado

μ_d = valor medio de las diferencias d para la *población* de todos los datos apareados

$\bar{d}$ = valor medio de las diferencias d para los datos *muestrales* apareados (igual a la media de los valores $x - y$)

s_d = desviación estándar de las diferencias d para la *muestra* de datos apareados

n = número de *pares* de datos

Estadístico de prueba de hipótesis para datos apareados

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

donde los grados de libertad = $n - 1$.

Valores P y valores críticos: tabla A-3 (distribución t)

Intervalos de confianza para datos apareados

$$\bar{d} - E < \mu_d < \bar{d} + E$$

donde $$E = t_{\alpha/2}\frac{s_d}{\sqrt{n}}$$

Valores críticos de $t_{\alpha/2}$: Utilice la tabla A-3 con $n - 1$ grados de libertad.

9-5 Comparación de la variación en dos muestras

Concepto clave Puesto que la característica de variación entre los datos es extremadamente importante, esta sección presenta la prueba F para comparar dos varianzas (o desviaciones estándar) poblacionales utilizando dos muestras. En esta sección estudiaremos la distribución F, que se utiliza para la prueba F. Es sumamente importante estar conscientes de una grave desventaja de este procedimiento: la prueba F para comparar dos varianzas (o desviaciones estándar) poblacionales es *muy* sensible a las desviaciones que se alejan de la distribución normal. Al final de esta sección analizaremos brevemente alternativas a la prueba F que no son tan sensibles a esta característica.

Prueba *F* para comparar varianzas

En esta sección utilizamos varianzas muestrales (o desviaciones estándar) para comparar dos varianzas (o desviaciones estándar) poblacionales. Debemos saber que la varianza es el cuadrado de la desviación estándar y también debemos conocer la siguiente notación.

Medidas de variación

s = desviación estándar *muestral* $\quad$ s^2 = varianza *muestral* (desviación estándar muestral al cuadrado)

σ = desviación estándar *poblacional* $\quad$ σ^2 = varianza *poblacional* (desviación estándar poblacional al cuadrado)

Los cálculos de esta sección se simplificarán considerablemente si designamos las dos muestras de manera que s_1^2 represente a *la más grande* de las dos varianzas muestrales. Matemáticamente no importa cuál muestra se designe como la muestra 1, así que la vida será más fácil si permitimos que s_1^2 represente a la mayor de las dos varianzas muestrales, como en el estadístico de prueba incluido en el cuadro de resumen.

Requisitos

1. Las dos poblaciones son *independientes* una de la otra. (En la sección 9-2 aprendimos que dos muestras son independientes si la muestra seleccionada de una población no está relacionada con la muestra seleccionada de la otra población. Las muestras no están apareadas o asociadas).
2. Las dos poblaciones están *distribuidas normalmente.* (Este supuesto es importante, ya que los métodos esta sección son sumamente sensibles a las desviaciones de la normalidad).

Notación para pruebas de hipótesis con dos varianzas o desviaciones estándar

s_1^2 = *la más grande* de dos varianzas muestrales

n_1 = tamaño de la muestra que tiene la varianza *más grande*

σ_1^2 = varianza de la población de donde se obtiene la muestra con la varianza *más grande*

Los símbolos s_2^2, n_2 y σ_2^2 se utilizan para la otra muestra y la otra población.

Estadístico de prueba para pruebas de hipótesis con dos varianzas

$$F = \frac{s_1^2}{s_2^2} \qquad \text{(donde } s_1^2 \text{ es } \textit{la más grande} \text{ de las dos varianzas muestrales)}$$

Valores críticos: Utilice la tabla A-5 para obtener valores críticos F que se determinan por lo siguiente:

1. El nivel de significancia α (la tabla A-5 tiene cuatro páginas de valores críticos para $\alpha = 0.025$ y 0.05).
2. **Grados de libertad del numerador = $n_1 - 1$**
3. **Grados de libertad del denominador = $n_2 - 1$**

Para dos poblaciones distribuidas normalmente con varianzas iguales (es decir, $\sigma_1^2 = \sigma_2^2$), la distribución muestral del estadístico de prueba $F = s_1^2/s_2^2$ es la **distribución F** que se muestra en la figura 9-5 con los valores críticos que se listan en la tabla de A-5. Si usted continúa repitiendo el experimento de seleccionar muestras al azar a partir de dos poblaciones distribuidas normalmente con varianzas iguales, la distribución de la proporción s_1^2/s_2^2 de las varianzas muestrales es la distribución F.

En la figura 9-5, observe las siguientes propiedades de la distribución F:

- La distribución F no es simétrica.
- Los valores de la distribución F no pueden ser negativos.
- La forma exacta de la distribución F depende de dos diferentes grados de libertad.

Valores críticos: Para calcular un valor crítico, primero remítase a la parte de la tabla A-5 correspondiente a α (para una prueba de una cola) o $\alpha/2$ (para una prueba de dos colas), luego intercepte la columna que representa los grados de libertad para s_1^2 con el renglón que representa los grados de libertad para s_2^2. Puesto que estamos estipulando que la varianza muestral más grande es s_1^2, todas

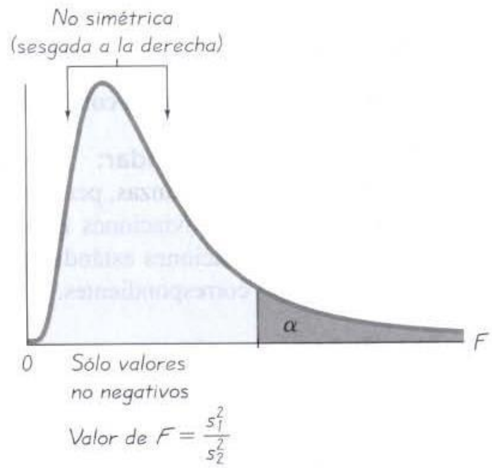


Figura 9-5 Distribución *F*

Existe una distribución *F* diferente para cada par diferente de grados de libertad para el numerador y el denominador.

las pruebas de una cola serán de cola derecha y todas las pruebas de dos colas requerirán que encontremos sólo el valor crítico localizado a la derecha. Buenas noticias: No tenemos necesidad de calcular un valor crítico separando una región crítica de cola izquierda. (Puesto que la distribución F no es simétrica y sólo tiene valores no negativos, un valor crítico de cola izquierda no puede encontrarse utilizando el negativo del valor crítico de cola derecha; en vez de ello, el valor crítico de cola izquierda se calcula utilizando el recíproco del valor de cola derecha con los números de grados de libertad invertidos. Véase el ejercicio 23).

Con frecuencia tenemos números de grados de libertad que no están incluidos en la tabla a A-5. Podríamos utilizar interpolación lineal para aproximar los valores que no aparecen, pero en la mayoría de los casos esto no es necesario puesto que el estadístico de prueba F es menor que el valor crítico más bajo posible o mayor que el valor crítico más alto posible. Por ejemplo, la tabla de A-5 indica que para $\alpha = 0.025$ en la cola derecha, 20 grados de libertad para el numerador y 34 grados de libertad para el denominador, el valor crítico F está entre 2.0677 y 2.1952. Cualquier estadístico de prueba F menor que 2.0677 determina que no se rechace la hipótesis nula, y cualquier estadístico de prueba F mayor que 2.1952 dará por resultado el rechazo de la hipótesis nula; la interpolación sólo es necesaria si el estadístico de prueba F cae entre 2.0677 y 2.1952. El uso de la tecnología adecuada elimina por completo este problema al dar valores P o valores críticos para cualquier número de grados de libertad.

Interpretación del estadístico de prueba *F* Si en realidad las dos poblaciones tienen varianzas iguales, entonces la proporción s_1^2/s_2^2 tiende a 1, puesto que los valores de s_1^2 y s_2^2 tienden a acercarse. Pero si las dos poblaciones tienen varianzas radicalmente diferentes, s_1^2 y s_2^2 tienden a ser números muy distintos. Si denotamos la más grande de las varianzas muestrales como s_1^2, vemos que la proporción s_1^2/s_2^2 será un número grande siempre que s_1^2 y s_2^2 tengan valores lejanos entre sí. En consecuencia, un valor de F cercano a 1 será evidencia a favor de la

Menor variación, mayor calidad

Ford y Mazda estaban produciendo transmisiones similares que, al parecer, se fabricaban con las mismas especificaciones. Pero las transmisiones hechas en Estados Unidos requerían más reparaciones por garantía que las transmisiones hechas en Japón. Cuando los investigadores inspeccionaron muestras de las cajas de engranajes de las transmisiones japonesas, al principio pensaron que sus instrumentos de medición estaban defectuosos porque no detectaban ninguna variabilidad entre las cajas de engranajes de las transmisiones Mazda. Se dieron cuenta de que, aunque las transmisiones estadounidenses cumplían con las especificaciones, las transmisiones Mazda no sólo estaban dentro de las especificaciones, sino también uniformemente cercanas al valor deseado. Al reducir la variabilidad entre las cajas de engranajes de las transmisiones, Mazda redujo los costos de inspección, de desecho, de corrección y de las reparaciones por garantía.

conclusión de que $\sigma_1^2 = \sigma_2^2$, y un valor grande de F será evidencia en contra de la conclusión de igualdad de las varianzas poblacionales.

Los valores de F grandes son evidencia en contra de $\sigma_1^2 = \sigma_2^2$.

Aseveraciones acerca de desviaciones estándar: El estadístico de prueba F se aplica a una aseveración acerca de dos varianzas, pero también podemos utilizarlo para aseveraciones acerca de dos desviaciones estándar poblacionales. Cualquier aseveración acerca de dos desviaciones estándar poblacionales puede replantearse en términos de las varianzas correspondientes.

Práctica de Estadística II - N° 05

TEMA N° 3: PRUEBA DE HIPOTESIS DE UNA PROPORCION Y MEDIA CON DESVIACIÓN CONOCIDA Y DESCONOCIDA

Sección : ………………………..……………….. Docente : Escribir el nombre del docente	Apellidos : ………………………..………………. Nombres : ……………………………………………. Fecha : ……/……/2017 Duración: 45 minutos Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

21. Un fabricante afirma que el 30% de todos los consumidores prefiere el servicio que brinda su empresa. Con el fin de evaluar esta afirmación se tomó una muestra aleatoria de 400 consumidores y se encontró que 100 de ellos prefieren dicho servicio. ¿Es ésta, suficiente evidencia para inferir que el porcentaje de preferencia del servicio no es del 30%? Utilice el nivel de significación de α=1%.
22.
23. En un establecimiento de comida rápida de Plaza VEA se asegura que el 90% de sus órdenes se entregan en menos de diez minutos. En una muestra de 100 órdenes, 82 se entregaron dentro de ese lapso, ¿Puede concluirse, en el nivel de significación α=0,01 que menos del 90% de las órdenes e entregan en menos de diez minutos?

24. Una cadena de tiendas de descuento expide su propia tarjeta de crédito. El gerente del departamento de tarjetas de créditos desea averiguar si el saldo insoluto medio mensual es mayor que S/. 400. El nivel de significación se fija en 0,05. En una revisión aleatoria de 172 saldos insolutos se encontró que la media muestral es de S/. 407 y la desviación estándar es S/. 38. ¿Debería concluir el funcionario de crédito que la media poblacional es mayor que S/. 400 o es razonable suponer que la diferencia de S/. 7 (obtenido de S/. 407-S/.400 = S/. 7) se debe al azar?

25. El departamento de quejas de una compañía de seguros encuentra que el costo medio de atender una queja es de S/. 60. Una compañía mostró que esta cantidad era mayor que en otras compañías de seguros, por lo que se tomaron medidas para disminuir los costos. Para evaluar el efecto de estas medidas se tomó una muestra aleatoria de 26 reclamaciones recientes. El costo por reclamación fue de S/. 57 y la desviación S/. 10 ¿Pueden concluir que las medidas tomadas para reducir los costos fueron efectivas? Use el nivel de significación 1%.

26. Una encuesta reveló que de 785 sujetos seleccionados aleatoriamente y que completaron cuatro años de estudios universitarios, 144 fuman y 641 no fuman. Utilice un nivel de significancia de 0.01 para probar la aseveración de que el porcentaje de fumadores que tienen cuatro años de estudios universitarios es menor que el porcentaje del 27% de la población general. ¿Por qué los graduados universitarios que fuman tienen una tasa menor del resto?

Referencias bibliográficas consultadas y/o enlaces recomendados

- Triola, M. (2010). Estadística (10a.ed.). México: Pearson.

Práctica de Estadística II - N° 06

Inferencia para varianzas y desviaciones estándar e inferencia de dos poblaciones

Sección : …………………………..………………...	Apellidos : ……………………….…………………
	Nombres : ……………………………………………
Docente : Escribir el nombre del docente	Fecha : ……/……/2017 Duración: 45 minutos
	Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

01. Pesos al nacer. Se realizó un estudio de los hijos de madres que consumieron cocaína durante el embarazo y se obtuvieron los siguientes datos muestrales de pesos al nacer: $\bar{x}$ = 190, = 2700 g, y s = 645 g (según datos de "Cognitive Outcomes of Preschool Children with Prenatal Cocaine Exposure", de Singer *et al., Journal of the American Medical Association,* vol. 291, núm. 20). Utilice un nivel de significancia de 0.05 para probar la aseveración de que la desviación estándar de los pesos al nacer de hijos de consumidoras de cocaína difiere de la desviación estándar de 696 g de los pesos al nacer de hijos de mujeres que no consumieron cocaína durante el embarazo. (Como la, tabla A-4 tiene un máximo de 100 grados de libertad, mientras que aquí se requieren 189 grados de libertad, utilice los siguientes valores críticos obtenidos por medio de STATDISK: X_L^2 = 152.8222 *y* X_R^2 = 228.9638.) Con base en el resultado, ¿parece que la cocaína consumida por las madres afecta la variación de los pesos de sus bebés?

02. Utilice un nivel de significancia de 0.05 para probar la aseveración de que las longitudes de ciertas varillas cromadas varían menos que las longitudes de las varillas en general. La desviación estándar de las longitudes de la población de varillas es de 2.5 pulgadas. A continuación se listan las longitudes (en pulgadas) de varillas cromadas seleccionadas al azar.

71.2	71.7	70.3	69.9	69.5	70.5	71.4	72.0	70.2
70.5	69.9	69.5	69.2	70.0	71.0	66.5	70.2	71.8

03. Un supervisor de control de calidad en una enlatadora sabe que la cantidad exacta contenida en cada lata varía, pues hay ciertos factores imposibles de controlar que afectan la cantidad de llenado. El llenado medio por lata es importante pero igualmente importante es la variación σ^2 de la cantidad de llenado. Si σ^2 es grande, algunas latas contendrán muy poco y otras, demasiado. Las agencias reguladoras especifican que la desviación estándar de la cantidad de llenado debe ser menor que 0.1 onzas. El supervisor de control de calidad muestreó n=10 latas y midió la cantidad de llenado en cada una. Los datos se reproducen a continuación.

7.96 7.90 7.98 8.01 7.97 7.96 8.03 8.02 8.04 8.02

04. Una compañía asegura que el mercado para su producto X tiene una aceptación de iguales proporciones en la ciudad A que en la ciudad B. Un especialista en mercado pone en duda dicha afirmación y para tal fin tomó una muestra aleatoria de 500 amas de casa en la ciudad A y encontró que el 59.6% de las mismas prefería el artículo X. Por otra parte tomó una muestra aleatoria de 300 amas de casa en la ciudad B y encontró que el 50% de las mismas preferían el artículo X. ¿Existe una diferencia real entre las dos ciudades? Nivel de significación 5%

05. Una encuesta indaga por la preferencia de películas de terror en los adolescentes de la ciudad. De un total de 400 entrevistados 160 eran mujeres de las que 17 preferían este género de películas. Del grupo de varones 15 declararon que también las preferían a otros género de películas. Al nivel del 10% de significancia. ¿Existe un número más elevado de varones adolescentes que prefieren el género de terror en las películas?

06. En los últimos años se han registrado un gran número de lavadoras de ropa, tanto electrónicas como mecánicas, con serias fallas en su funcionamiento. Se desea analizar y comparar el costo de reparación de cada tipo de artefacto. Para ello se ha seleccionado una muestra aleatoria de cada tipo de lavadora y se han registrado en la tabla los costos de reparación en soles. Se sabe que el número de fallas tiene distribución normal.

N°	1	2	3	4	5	6	7	8	9	10	11
ELECTRONICO	178	161	194	204	185	179	173	172	108	181	185
MECANICO	128	89	150	191	188	209	53	131	184	97	112

A un nivel de significación del 5%, ¿los costos de reparación de ambos tipos de lavadoras son homogéneos?

07. Un investigador desea verificar el nivel de consumo de menestras en los habitantes de la región, dado que las menestras son una fuente de fibra, proteínas y otros compuestos vitamínicos muy importantes. Con una muestra de 23 personas de la zona rural estima una media de 0,76 Kgr por año con una desviación estándar de 0.23Kg, mientras que en una muestra de 35 personas en la zona urbana brinda una media de 0.79Kg por año con una desviación estándar de 0.32Kg. Por los resultados obtenidos, ¿Puede afirmarse al nivel del 0.10 que en las zonas urbanas se consume más menestras que en las zonas rurales?

08. Una compañía de transportes requiere comprar un gran lote de buses para el transporte urbano con el fin de reemplazar su parque automotor y para tal fin desea comprobar la afirmación hecha por el proveedor de la marca B, en el sentido de que la marca A es menos ahorradora de combustible. Para tal fin la empresa toma una muestra aleatoria de 35 vehículos marca A y encuentra que la misma tiene un promedio en el rendimiento de 18 km/g con una desviación estándar de 8 km/g, mientras que una muestra de 32 vehículos marca B presenta un promedio de 22 km/g con desviación estándar de 3 km/g. ¿Qué decisión debe tomar el gerente de la compañía con un nivel de significación del 5%?

09. El jefe de personal de una gran empresa afirma que la diferencia de los promedios de antigüedad entre los obreras y obreros de la compañía es de 3.5 años. El presidente de la compañía considera que ésta diferencia es superior. Para comprobar dicha situación, se toma una muestra aleatoria de 40 obreras cuyo promedio de antigüedad es de 12.4 años con desviación estándar de 1.5 años y de un grupo de 45 obreros cuyo promedio de antigüedad es de 8.3 años con desviación estándar de 1.7 años. Comprobar la hipótesis con un nivel de significación del 5%.

10. En Southern Copper Corporation realizaron un concurso de promoción. A continuación se muestran las edades de los ingenieros solicitantes. Algunos de los solicitantes que no tuvieron éxito para obtener la promoción se quejaron de que hubo discriminación por edad en la competencia, utilice un α=0.01 para poner a prueba la aseveración de que los solicitantes sin éxito provienen de una población con una edad media mayor que la de los solicitantes exitosos., ¿parece haber discriminación por la edad?

Edades de solicitantes sin éxito										Edades de solicitantes con éxito									
34	37	37	38	41	42	43	44	44	45	27	33	36	37	38	38	39	42	42	43
45	45	46	48	49	53	53	54	54	55	43	44	44	44	45	45	45	45	46	46
56	57	60								47	47								

11. Un inspector de tránsito registró la antigüedad de automóviles particulares y taxis elegidos al azar en la ciudad de Huancayo. Utilice un nivel de significancia de 0.02 para probar la aseveración de que hay una diferencia entre la antigüedad media de un automóvil y la antigüedad media de un taxi. Esperaríamos que los taxis fueran más nuevos, pero, ¿que sugieren los resultados?
A continuación se listan las antigüedades (en años).

Automóviles										Taxis									
4	0	8	11	14	3	4	4	3	5	8	8	0	3	8	4	3	3	6	11
8	3	3	7	4	6	6	1	8	2	7	7	6	9	5	10	8	4	3	4
15	11	4	1	6	1	8													

12. Los conteos de glóbulos blancos en la sangre sirven para diagnosticar problemas de la médula ósea. A continuación se presentan los conteos de glóbulos blancos de muestras aleatorias simples de hombres y mujeres. Utilice un nivel de significancia de 0.05 para someter a prueba la afirmación de que los hombres y las mujeres tienen conteos medios diferentes de glóbulos blancos. suponga que las dos muestras son aleatorias simples independientes, seleccionadas de poblaciones distribuidas normalmente. También suponga que la desviación estándar poblacional es 0,98 y 1,02 para mujeres y varones respectivamente.

Mujeres:	8.90	6.50	9.45	7.65	6.40	5.15	16.60
	5.90	9.30	8.55	10.80	4.85	4.90	8.75
	4.05	9.05	5.05	6.40	4.05	7.60	4.95
Hombres:	5.25	5.95	10.05	5.45	5.30	5.55	6.85
	6.40	7.85	7.70	5.30	6.50	4.55	7.10
	4.40	4.90	10.75	11.00	9.60		

13. Utilice los datos muéstrales que se listan a continuación y con un nivel de significancia de 0.05 ponga a prueba la aseveración de que la cantidad media de alquitrán en cigarrillos largos con filtro es menor que la cantidad media de alquitrán en cigarrillos largos sin filtro. Todas las mediciones son en miligramos, suponga que las dos muestras son aleatorias simples independientes, seleccionadas de poblaciones distribuidas normalmente. También suponga que las desviaciones estándar poblacionales son iguales.

Con filtro	16	15	16	14	16	1	16	18	10	14
	11	14	13	13	13	16	16	8	16	11
Sin filtro	23	23	24	26	25	26	21	24		

14. Muchos estudiantes han tenido la experiencia poco placentera de sentir pánico en los exámenes porque la primera pregunta era excepcionalmente difícil. Se estudió el efecto que tiene el orden de las preguntas de exámenes sobre la ansiedad. Las siguientes puntuaciones son mediciones de la "ansiedad debilitante por exámenes" ¿Existe evidencia suficiente para sustentar la aseveración de que el orden de las preguntas de examen tiene un efecto en la calificación? suponga que las dos muestras son aleatorias simples independientes, seleccionadas de poblaciones distribuidas normalmente.

Preguntas ordenadas de fácil a difícil					Preguntas ordenadas de difícil a fácil			
24.64	39.29	16.32	32.83	28.02	33.62	34.02	26.63	30.26
33.31	20.60	21.13	26.69	28.90	35.91	26.68	29.49	35.32
26.43	24.23	7.10	32.86	21.06	27.24	32.34	29.34	33.53

15. Se compran fusibles etiquetados con 100 Ω a dos distribuidores diferentes. La especificación para este tipo de fusibles es que su resistencia verdadera esté dentro del 5% de su resistencia etiquetada. En una muestra de 180 fusibles del distribuidor A, 150 de éstos satisfacían la especificación. En otra muestra de 270 fusibles comprados al distribuidor B, 233 cumplían la especificación. El distribuidor A es el proveedor actual, pero si los datos demuestran convincentemente que una proporción mayor de los resistores del distribuidor B satisface la especificación, se hará el cambio. ¿Se debe hacer el cambio?

Práctica de Estadística II - N° 07

Tema 07: Inferencia para la media de datos apareados

Sección : …………………………..………………..	Apellidos : …………………………..……………….
Docente : Escribir el nombre del docente	Nombres : …………………………………………….
	Fecha : ……/……/2017 Duración: 45 minutos
	Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

01. Se desea comparar el monto (en cientos de soles) del impuesto predial que calculan dos empresas contratadas por el Municipio a una muestra de residencias. Para ello se seleccionó una muestra aleatoria de diez propiedades residenciales para calcular el impuesto predial correspondiente. Las empresas presentaron los montos que se muestran en la tabla, para un nivel de significancia de 0,05 ¿Puede concluirse que hay diferencia en los montos calculados del impuesto predial medio de las residencias por las empresas?

Empresa 1	135	110	131	142	105	130	131	110	125	149
Empresa 2	128	105	119	140	98	123	127	115	122	145

02. Se realiza un estudio para medir el efecto del cambio ambiental en estudiantes extranjeros. Uno de los aspectos del estudio es una comparación del peso corporal (en libras) de los estudiantes al ingresar al país, y su peso un año más tarde. Se sospecha que los alimentos peruanos, que son más nutritivos, provocan el aumento de peso. Se utiliza el nivel $\alpha=0,01$. Se selecciona una muestra aleatoria de 11 estudiantes extranjeros para el estudio, ¿Cuál es su conclusión?

Antes	124	157	98	190	103	135	149	176	200	180	256
Después	142	157	96	212	116	134	150	184	209	180	269

03. El gerente de un balneario de aguas curativas anuncia un programa de reducción de peso y afirma que el participante promedio pierde más de 6 kilogramos. En la siguiente tabla se muestra el resultado en 8 personas, ¿cuál sería su decisión para un nivel de significación del 1%? Además determine e interprete el intervalo de confianza para el valor medio de las diferencias para la población de datos apareados

Antes	91,8	100	94,1	88,2	80,4	87,7	91,8	94,5
Después	86,4	96,8	87,3	81,8	73,2	79,0	85,0	84,5

04. El gerente de una empresa diseñó un plan de incentivos para los vendedores. A fin de evaluar este plan innovador, se seleccionaron aleatoriamente 12 vendedores y se registró su ingreso promedio semanal antes y después de aplicar el plan. ¿Hubo un incremento significativo en el ingreso promedio semanal de los vendedores debido al plan innovador de incentivos, para $\alpha=0,05$?

Antes	320	290	421	510	210	402	625	560	360	431	506	505
Después	340	285	475	510	210	500	631	560	365	431	525	619

05. Del ejercicio anterior determine el intervalo de confianza para el valor medio de las diferencias para la población de datos apareados.

Práctica de Estadística II - N° 08

TEMA N° 8: Estimación de dos varianzas poblacionales

Sección : ……………………..………………...

Docente : Escribir el nombre del docente

Apellidos : ………………………..…………….
Nombres : ……………………………………………
Fecha : ……/……/2017 Duración: 45 minutos
Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

01. El contenido de azúcar, en mg/ml, de un caldo utilizado para fabricar un producto farmacéutico fue medido varias veces en cada uno de tres días sucesivos.

Día 1:	5.0	4.8	5.1	5.1	4.8	5.1	4.8
	4.8	5.0	5.2	4.9	4.9	5.0	
Día 2:	5.8	4.7	4.7	4.9	5.1	4.9	5.4
	5.3	5.3	4.8	5.7	5.1	5.7	
Día 3:	6.3	4.7	5.1	5.9	5.1	5.9	4.7
	6.0	5.3	4.9	5.7	5.3	5.6	

a) ¿Puede concluir que la variabilidad del proceso es mayor el segundo día que el primero?
b) ¿Puede concluir que la variabilidad del proceso es mayor el tercer día que el segundo?

02. La estabilidad de mediciones en un producto manufacturado es importante para mantener la calidad del producto. De hecho, a veces es mejor tener una pequeña variación en el valor medido de alguna característica importante de un producto, así como tener la media del proceso ligeramente fuera del objetivo, que sufrir una amplia variación con valor medio que perfectamente se ajuste a los requisitos. Esta última situación puede producir un porcentaje más alto de productos defectuosos que la primera. Un fabricante de focos eléctricos sospechaba que una de sus líneas de producción estaba produciendo focos con una amplia variación en duración de vida útil. Para probar su teoría, comparó las duraciones de vida útil de n = 50 focos muestreados al azar de la línea sospechosa y n = 50 de una línea que parecía estar "en control". Las medias muéstrales y varianzas para las dos muestras fueron como sigue:

"Líneas sospechosa"	Línea "en control"
$\bar{x}_1 = 1520$	$\bar{x}_2 = 1476$
$s_1^2 = 92000$	$s_2^2 = 37000$

¿Los datos dan suficiente evidencia para indicar que los focos producidos por la "línea sospechosa" tienen una varianza más grande en duración que los producidos por la línea que se supone están en control? Pruebe usando α= 0.05?

03. Unos científicos realizaron un experimento que comprendía 10 corredores sanos y 10 ciclistas sanos, para determinar si hay diferencias significativas en mediciones de presión dentro del compartimiento del músculo anterior para corredores y ciclistas.7 Los datos, es decir, presión del compartimiento, en milímetros de mercurio (Hg), son como sigue:

	Corredores		Ciclistas	
Afección	Media	Desviación estándar	Media	Desviación estándar
En reposo	14.5	3.92	11.1	3.98
80% máximo consumo de O_2	12.2	3.49	11.5	4.95
Máximo consumo de O_2	19.1	16.9	12.2	4.47

Para cada una de las tres variables medidas en este experimento, pruebe para ver si hay una diferencia significativa en las varianzas para corredores contra ciclistas. Encuentre los valores p aproximados para cada una de estas pruebas. ¿Será apropiada una prueba t de dos muestras con una estimación agrupada de y2 para estas tres variables? Explique.

04. Los datos adjuntos se obtuvieron en un estudio para evaluar el potencial de licuefacción en una planta de energía nuclear propuesta .Antes de probar la resistencia cíclica, se recopilaron muestras de suelo mediante un método de jarro y un método de bloque y se obtuvieron los siguientes valores observados de densidad en seco (lb/pie3).

Muestreo con jarro	101.1	111.1	107.6	98.1
	99.5	98.7	103.3	108.9
	109.1	104.1	110.0	98.4
	105.1	104.5	105.7	103.3
	100.3	102.6	101.7	105.4
	99.6	103.3	102.1	104.3
Muestreo con bloque	107.1	105.0	98.0	97.9
	103.3	104.6	100.1	98.2
	97.9	103.2	96.9	

¿Podríamos afirmar que las varianzas de las medidas de la densidad en seco (lb/pie3) con ambos métodos de muestreo son homogéneas?

TERCERA UNIDAD
RANALISIS DE VARIANZA ANOVA

- GUÍA DE PRÁCTICA Nº 9:Anova de un factor y de dos factores
- GUÍA DE PRÁCTICA Nº 10: Estadística no paramétrica :Prueba de signo
- GUÍA DE PRÁCTICA Nº 11: Estadística no paramétrica ,tablas de contingencia
- GUÍA DE PRÁCTICA Nº 12: Estadística no paramétrica, correlación de Rangos Prueba de Rachas

TEXTO N° 13

Análisis de varianza

Compilado y adaptado de Mario Triola
Estadística PEARSON; 2009

Págs. De 636 a 649 y de 655 a 660

12-1 Panorama general

En la sección 12-2 explicamos un método importante para probar la igualdad de tres o más medias poblacionales. En la sección 9-3 estudiamos procedimientos para probar la hipótesis de que *dos* medias poblacionales son iguales, pero los métodos de esa sección no pueden aplicarse cuando se incluyen tres o más medias. En vez de referirnos al objetivo principal de probar medias iguales, el término *análisis de varianza* se refiere al *método* que empleamos, el cual está basado en un análisis de varianzas muestrales.

Definición

El **análisis de varianza (ANOVA)** es un método de prueba de igualdad de tres o más medias poblacionales, por medio del análisis de las varianzas muestrales.

¿Por qué no probar sencillamente dos muestras al mismo tiempo? ¿Por qué necesitamos un nuevo procedimiento, cuando podemos probar la igualdad de dos medias utilizando los métodos presentados en el capítulo 9? Por ejemplo, si deseamos utilizar los datos muestrales de la tabla 12-1 para probar la aseveración de que las tres poblaciones tienen la misma media, ¿por qué no simplemente tomamos dos a la vez y probamos H_0: $\mu_1 = \mu_2$, luego H_0: $\mu_2 = \mu_3$, y así sucesivamente? Para los datos de la tabla 12-1, el método de probar la igualdad de dos medias a la vez requiere de seis pruebas diferentes de hipótesis, de manera que el grado de confianza podría ser tan bajo como 0.95^6 (o 0.735). En general, conforme incrementamos el número de pruebas de significancia individuales, aumentamos el riesgo de obtener una diferencia únicamente por el azar (en vez de una diferencia real en las medias). El riesgo de un error tipo I (encontrar una diferencia en uno de los pares cuando en realidad no existe tal diferencia) es demasiado alto. El método del análisis de varianza nos ayuda a evitar este problema en particular (rechazar una hipótesis nula verdadera) utilizando una prueba de igualdad de varias medias.

Distribución *F*

Los métodos del ANOVA de este capítulo requieren de la distribución F que se presentó por primera vez en la sección 9-5, en la cual señalamos que la distribución F tiene las siguientes propiedades importantes (véase la figura 12-1):

1. La distribución F no es simétrica; está sesgada hacia la derecha.
2. Los valores de F pueden ser 0 o positivos, pero no pueden ser negativos.
3. Existe una distribución F diferente para cada par de grados de libertad para el numerador y el denominador.

Los valores críticos de F se localizan en la tabla A-5.

El análisis de varianza (ANOVA) está basado en una comparación de dos estimados diferentes de la varianza común de las distintas poblaciones. Estos estimados (la *varianza entre muestras* y la *varianza dentro de las muestras*) se describirán en la sección 12-2. El término *un factor* se utiliza porque los datos muestrales están separados en grupos según una característica o factor. En la sección 12-3 estudiaremos el análisis de varianza de dos factores, el cual nos permite comparar poblaciones separadas en categorías por medio de dos características (o factores). Por ejemplo, podríamos separar la estatura de las personas utilizando los siguientes dos factores: **1.** género (hombre o mujer) y **2.** mano dominante derecha o izquierda.

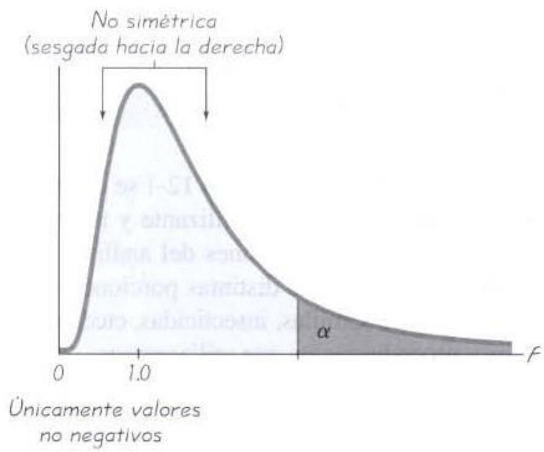


Figura 12-1

Distribución *F*

Existe una distribución *F* distinta para cada par de grados de libertad diferente para el numerador y el denominador.

Estrategia de estudio sugerida: Puesto que los procedimientos empleados en este capítulo requieren de cálculos complicados, enfatizaremos el uso y la interpretación de programas de cómputo, tales como STATDISK, Minitab y Excel, o de una calculadora TI-83/84 Plus. Sugerimos que inicie la sección 12-2 enfocándose en el siguiente concepto clave: estamos utilizando un procedimiento para probar la aseveración de que tres o más medias son iguales. A pesar de que los detalles de los cálculos son complicados, nuestro procedimiento será fácil debido a que está basado en un valor *P*. Si el valor *P* es pequeño, como 0.05 o menor, se rechaza la igualdad de las medias. De otra manera no se rechaza la igualdad de las medias. Después de comprender este procedimiento básico y sencillo, proceda a la comprensión de los fundamentos subyacentes.

12-2 ANOVA de un factor

Concepto clave En esta sección se presenta el método del *análisis de varianza de un factor*, que se utiliza para probar las hipótesis de que tres o más medias poblacionales son iguales, como en H_0: $\mu_1 = \mu_2 = \mu_3$. Como los cálculos son muy complicados, recomendamos interpretar los resultados obtenidos por medio de un programa de cómputo o de una calculadora TI-83/84 Plus. Sugerimos la siguiente estrategia de estudio:

1. Comprenda que un valor *P* pequeño (como 0.05 o menos) conduce al rechazo de la hipótesis nula de igualdad de medias. Con un valor *P* grande (como uno mayor que 0.05), no rechace la hipótesis nula de igualdad de medias.
2. Trate de comprender el fundamento subyacente al estudiar los ejemplos de esta sección.
3. Familiarícese con la naturaleza de los valores de la SC (suma de cuadrados) y de los CM (cuadrados medios), así como con el papel que desempeñan en la determinación del estadístico de prueba *F*, pero utilice programas estadísticos de cómputo o una calculadora para obtener esos valores.

El método que empleamos se denomina **análisis de varianza de un factor** (o **análisis de varianza de una entrada**) porque empleamos una sola propiedad o característica para categorizar las poblaciones. En ocasiones a esta característica se le llama *tratamiento* o *factor*.

Definición

Un **tratamiento** (o **factor**) es una propiedad o característica que nos permite distinguir entre sí a las distintas poblaciones.

Por ejemplo, los pesos de los álamos en la tabla 12-1 se distinguen de acuerdo con el tratamiento (ninguno, fertilizante, riego, fertilizante y riego). Se utiliza el término *tratamiento* porque las primeras aplicaciones del análisis de varianza implicaron experimentos de agricultura en los que distintas porciones de tierra se trataban con diferentes fertilizantes, tipos de semillas, insecticidas, etcétera. El siguiente recuadro incluye los requisitos y procedimientos que utilizaremos.

Requisitos

1. Las poblaciones tienen distribuciones que son aproximadamente normales. (Este requisito no es demasiado estricto, ya que el método funciona bien, a menos que la población tenga una distribución muy diferente de la normal. Si una población tiene una distribución muy diferente a la normal, utilice la prueba de Kruskal-Wallis, descrita en la sección 13-5).
2. Las poblaciones tienen la misma varianza σ^2 (o desviación estándar σ). (Este requisito no es demasiado estricto, ya que el método funciona bien a menos que las varianzas poblacionales difieran en grandes cantidades. El especialista en estadística de la Universidad de Wisconsin, George E. P. Box demostró que, siempre y cuando los tamaños muestrales sean iguales (o casi iguales), las varianzas pueden diferir de tal forma que la más grande sea hasta nueve veces el tamaño de la más pequeña, y los resultados del ANOVA continúan siendo esencialmente confiables).
3. Las muestras son aleatorias simples (es decir, muestras del mismo tamaño que tienen la misma probabilidad de ser elegidas).
4. Las muestras son independientes entre sí (es decir, no están aparejadas o asociadas de ninguna forma).
5. Las diferentes muestras provienen de poblaciones que están categorizadas de una sola forma. (Ésta es la base del nombre del método: análisis de varianza de *un factor*).

Procedimiento de prueba de H_0: $\mu_1 = \mu_2 = \mu_3 = \ldots$

1. **Utilice STATDISK, Minitab, Excel o una calculadora TI-83/84 Plus para obtener los resultados.**
2. **Identifique el valor P en los resultados.**
3. **Plantee una conclusión con base en estos criterios:**
   - **ISi el valor $P \leq \alpha$, rechace la hipótesis nula de medias iguales y concluya que al menos una de las medias poblacionales es diferente de las otras.**
   - **Si el valor $P > \alpha$, no rechace la hipótesis nula de medias iguales.**

Tenga cuidado al interpretar los resultados: Cuando concluimos que existe suficiente evidencia para rechazar la aseveración de medias poblacionales iguales, no podemos concluir a partir del ANOVA que cualquier media en particular es distinta de las demás. (Existen otras pruebas que pueden utilizarse para identificar las medias específicas que son diferentes, las cuales se conocen como *procedimientos de comparación múltiple* y que estudiaremos más adelante en esta sección.

EJEMPLO Pesos de álamos Dados los pesos de álamos listados en la tabla 12-1, y un nivel de significancia de $\alpha = 0.05$, utilice STATDISK, Minitab, Excel o una calculadora TI-83/84 Plus para probar la aseveración de que las cuatro muestras provienen de poblaciones con medias diferentes. Vea a las siguientes pantallas de resultados.

STATDISK

Source:	DF:	SS:	MS:	Test Stat, F:	Critical F:	P-Value:
Treatment:	3	4.682415	1.560805	5.731353	3.238868	0.0073483
Error:	16	4.35724	0.2723275			
Total:	19	9.039655	0.4757713			

Reject the Null Hypothesis
Reject equality of means

Minitab

One-way ANOVA: None, Fert, Irrig, Fert&Irrig

```
Source  DF     SS     MS     F      P
Factor   3  4.682  1.561  5.73  0.007
Error   16  4.357  0.272
Total   19  9.040

S = 0.5219   R-Sq = 51.80%   R-Sq(adj) = 42.76%
```

Excel

ANOVA						
Source of Variation	*SS*	*df*	*MS*	***F***	***P-value***	***F crit***
Between Groups	4.682415	3	1.560805	**5.731352875**	**0.007348294**	**3.238871522**
Within Groups	4.35724	16	0.2723275			
Total	9.039655	19				

TI-83/84 Plus

```
One-way ANOVA
 F=5.731352875
 p=.0073482944
 Factor
  df=3
  SS=4.682415
↓ MS=1.560805
```

```
One-way ANOVA
↑ MS=1.560805
 Error
  df=16
  SS=4.35724
  MS=.2723275
 Sxp=.521850074
```

SOLUCIÓN

REQUISITO ✔ Cuando se investiga el requisito de normalidad de las cuatro poblaciones diferentes, la única muestra cuestionable es la segunda del grupo de tratamiento con fertilizante. Sólo para esa muestra, la gráfica cuantilar normal y el histograma sugieren que tal vez no se trate de una distribución normal, y el problema es el valor de 1.34 kg, que al parecer es un valor extremo; por lo tanto, debemos ser cuidadosos. Lo más adecuado sería realizar el análisis con y sin la inclusión de ese valor. Consulte el ejercicio 5, en donde vemos que en este caso el valor extremo de 1.34 kg no tiene un efecto drástico en los resultados.

continúa

Las varianzas son muy diferentes, pero la más grande no es más de nueve veces el tamaño de la más pequeña, de manera que se satisface el requisito de varianzas iguales. Suponemos que tenemos muestras aleatorias simples. Sabemos que las muestras son independientes, y cada valor pertenece exactamente a un grupo. Por lo tanto, los requisitos se satisfacen y podemos proceder con la prueba de hipótesis. ✔

La hipótesis nula es H_0: $\mu_1 = \mu_2 = \mu_3$ y la hipótesis alternativa es la aseveración de que al menos una de las medias es diferente de las otras.

Paso 1: Utilice algún recurso tecnológico para obtener los resultados del ANOVA, como alguno de los que se muestran en este ejemplo.

Paso 2: Todas las pantallas de resultados indican que el valor P es 0.007, redondeado.

Paso 3: Puesto que el valor P es menor que el nivel de significancia de α = 0.05, rechazamos la hipótesis nula de igualdad de medias.

INTERPRETACIÓN Existe evidencia suficiente para sustentar la aseveración de que las cuatro medias poblacionales no son iguales. Con base en la muestra de pesos de la tabla 12-1. Con base en la muestra de pesos de la tabla 12-1, concluimos que esos pesos provienen de poblaciones con medias diferentes. Con base en esta prueba de ANOVA, no podemos concluir que cualquier media en particular sea diferente de las demás.

Fundamentos

Suponiendo que las poblaciones tienen la misma varianza σ^2, el estadístico de prueba F es la razón o el cociente de los siguientes dos estimados de σ^2: **1.** la variación *entre* muestras (con base en la variación entre medias muestrales), y **2.** variación *dentro* de muestras (con base en las varianzas muestrales). Un estadístico de prueba F significativamente *grande* (ubicado a la extrema derecha de la gráfica de distribución F) constituye evidencia en contra de medias poblacionales iguales, de manera que la prueba es de cola derecha. La figura 12-2 indica la relación entre el estadístico de prueba F y el valor P.

Estadístico de prueba del ANOVA de un factor

$$F = \frac{\text{varianza entre las muestras}}{\text{varianza dentro de las muestras}}$$

El numerador del estadístico de prueba F mide la variación entre medias muestrales. El estimado de la varianza en el denominador depende únicamente de las varianzas muestrales y no se ve afectado por las diferencias entre las medias muestrales. Como consecuencia, las medias muestrales que tienen valores cercanos dan como resultado un estadístico de prueba F pequeño y concluimos que no existen diferencias significativas entre las medias muestrales. Pero si el valor de F es excesivamente *grande*, entonces rechazamos la aseveración de igualdad de medias. (Los ambiguos términos "pequeño" y "excesivamente grande" se vuelven objetivos por medio del valor P correspondiente, el cual nos indica si el estadístico de prueba F está o no en la región crítica). Puesto que valores excesivamente grandes de F reflejan medias desiguales, la prueba es de cola derecha.

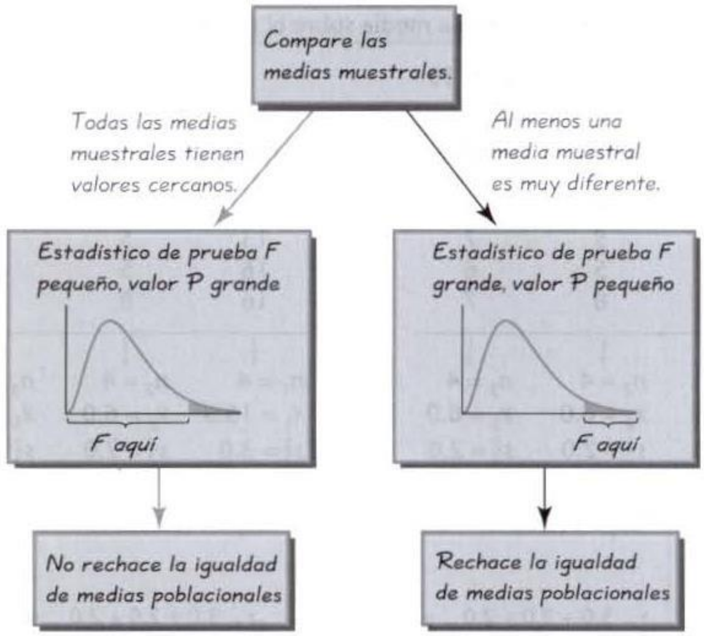


Figura 12-2

Relación entre el estadístico de prueba F y el valor P

Cálculos con tamaños muestrales *n* iguales

Para comprender realmente los efectos del análisis de varianza de un factor, vea la tabla 12-2. Compare el conjunto de datos A con el conjunto de datos B y observe que la única diferencia es que añadimos 10 a cada valor de la muestra 1 en el conjunto de datos A para obtener los valores de la muestra 1 en el conjunto de datos B. Si todos los conjuntos de datos tienen el mismo tamaño muestral (como en $n = 4$ para la tabla 12-2), los cálculos requeridos no son demasiado difíciles. Primero, calcule la varianza *entre* muestras al evaluar $ns_{\bar{x}}^2$, donde $s_{\bar{x}}^2$ es la varianza de las medias muestrales y n es el tamaño de cada una de las muestras. Es decir, considere las medias muestrales como un conjunto ordinario de valores y calcule la varianza. (Del teorema del límite central, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ se puede despejar σ para obtener $\sigma = \sqrt{n} \cdot \sigma_{\bar{x}}$, de manera que podemos estimar σ^2 con $ns_{\bar{x}}^2$). Por ejemplo, las medias muestrales del conjunto de datos A en la tabla 12-2 son 5.5, 6.0 y 6.0. Estos tres valores tienen una varianza de $s_{\bar{x}}^2 = 0.0833$, de manera que la

$$\text{varianza entre las muestras} = ns_{\bar{x}}^2 = 4(0.0833) = 0.3332$$

A continuación, estime la varianza *dentro* de las muestras, calculando s_p^2, que es la varianza agrupada que se obtiene al calcular la media de las varianzas muestrales. Las varianzas muestrales en la tabla 12-2 son 3.0, 2.0 y 2.0, de forma que la

$$\text{varianza dentro de muestras} = s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$$

Finalmente, evalúe el estadístico de prueba F de la siguiente manera:

$$F = \frac{\text{varianza entre muestras}}{\text{varianza dentro de muestras}} = \frac{ns_{\bar{x}}^2}{s_p^2} = \frac{0.3332}{2.3333} = 0.1428$$

Resistencia a las encuestas

Las encuestas basadas en muestras relativamente pequeñas pueden ser bastante precisas, siempre y cuando la muestra sea aleatoria o representativa de la población. Sin embargo, el incremento en las tasas de rechazo a las encuestas está haciendo que sea más difícil obtener muestras aleatorias. El Council of American Survey Research Organizations reportó que, en un año reciente, el 38% de los consumidores se rehusaron a responder encuestas. El director de una compañía de investigación de mercados dijo que "las personas tienen temor de ser seleccionadas y les preocupa que las generalizaciones se realicen con base únicamente en aquellos que cooperan". Los resultados de la industria de investigación de mercados, multimillonaria en dólares, afectan los productos que compramos, los programas de televisión que vemos y muchas otras facetas de nuestras vidas.

Tabla 12-2 Efecto de una media sobre el estadístico de prueba F

A → añadir 10 → B

	A: Muestra 1	A: Muestra 2	A: Muestra 3	B: Muestra 1	B: Muestra 2	B: Muestra 3
	7	6	4	17	6	4
	3	5	7	13	5	7
	6	5	6	16	5	6
	6	8	7	16	8	7
	$n_1 = 4$	$n_2 = 4$	$n_3 = 4$	$n_1 = 4$	$n_2 = 4$	$n_3 = 4$
	$\bar{x}_1 = 5.5$	$\bar{x}_2 = 6.0$	$\bar{x}_3 = 6.0$	$\bar{x}_1 = 15.5$	$\bar{x}_2 = 6.0$	$\bar{x}_3 = 6.0$
	$s_1^2 = 3.0$	$s_2^2 = 2.0$	$s_3^2 = 2.0$	$s_1^2 = 3.0$	$s_2^2 = 2.0$	$s_3^2 = 2.0$

	A	B
Varianza entre muestras	$ns_{\bar{x}}^2 = 4\ (0.0833) = 0.3332$	$ns_{\bar{x}}^2 = 4\ (30.0833) = 120.3332$
Varianza dentro de muestras	$s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$	$s_p^2 = \frac{3.0 + 2.0 + 2.0}{3} = 2.3333$
Estadístico de prueba F	$F = \frac{ns_{\bar{x}}^2}{s_p^2} = \frac{0.3332}{2.3333} = 0.1428$	$F = \frac{ns_{\bar{x}}^2}{s_p^2} = \frac{120.3332}{2.3333} = 51.5721$
Valor P (obtenido con Excel)	*Valor P* = 0.8688	*Valor P* = 0.0000118

El valor crítico de F se calcula suponiendo una prueba de cola derecha, ya que los valores grandes de F corresponden a diferencias significativas entre medias. Con k muestras, cada una con n valores, el número de grados de libertad se obtiene como sigue.

Grados de libertad:
(k = número de muestras y n = tamaño muestral)

grados de libertad del numerador $= k - 1$

grados de libertad del denominador $= k\,(n - 1)$

Para el conjunto de datos A de la tabla 12-2, $k = 3$ y $n = 4$, de manera que los grados de libertad son 2 para el numerador y $3(4 - 1) = 9$ para el denominador. Con $\alpha = 0.05$, 2 grados de libertad para el numerador y 9 grados de libertad para el denominador, el valor crítico F de la tabla A-5 es 4.2565. Si utilizáramos el método tradicional de prueba de hipótesis con el conjunto de datos A de la tabla 12-2, veríamos que esta prueba de cola derecha tiene un estadístico de prueba $F = 0.1428$ y un valor crítico de $F = 4.2565$, de manera que el estadístico de prueba no se encuentra en la región crítica y, por lo tanto, no rechazamos la hipótesis nula de igualdad de medias.

Para ver realmente cómo funciona el estadístico de prueba F, considere ambos conjuntos de datos muestrales en la tabla 12-2. Observe que las tres muestras de la parte A son idénticas a las tres muestras de la parte B, excepto que en la parte B añadimos 10 a cada valor de la muestra 1 de la parte A. Las tres medias muestrales de la parte A son muy cercanas, pero existen diferencias sustanciales en la parte B. Las tres varianzas muestrales de la parte A son idénticas a las de la parte B.

La suma de 10 a cada dato de la primera muestra de la tabla 12-2 tiene un efecto drástico en el estadístico de prueba, ya que F cambia de 0.1428 a 51.5721. La suma de 10 a cada dato de la primera muestra también tiene un efecto drástico en el valor P, el cual cambia de 0.8688 (no significativo) a 0.0000118 (significativo). Observe que la varianza entre muestras en la parte A es 0.3332, pero en la parte B es 120.3332 (lo que indica que las medias muestrales en la parte B están más separadas). Observe también que las varianzas dentro de las muestras son de 2.3333 en ambas partes, ya que la varianza dentro de una muestra no se ve afectada cuando sumamos una constante a cada valor muestral. *El cambio en el estadístico F y el valor P es atribuible únicamente a los cambios en* $\bar{x}_1$. Esto ilustra que el estadístico de prueba F es muy sensible a las *medias* muestrales, aun cuando se obtiene a través de dos estimados distintos de la *varianza* poblacional común.

He aquí el punto clave de la tabla 12-2: los conjuntos de datos A y B son idénticos, excepto que en el conjunto de datos B se añadió 10 a cada valor de la primera muestra. La suma de 10 a cada valor de la primera muestra causa que las tres medias muestrales se aparten más, con el resultado de que el estadístico de prueba F se incrementa y el valor P disminuye.

La ética en los reportes de encuestas

La American Association for Public Opinion Research creó un código de ética para aplicarse en los reportes de noticias de resultados de encuesta. Este código requiere que se incluya lo siguiente: **1.** identificación del patrocinador, **2.** fecha de la realización de la encuesta, **3.** tamaño de la muestra, **4.** naturaleza de la población muestreada, **5.** tipo de encuesta utilizada y **6.** redacción exacta de las preguntas de la encuesta. Las encuestas financiadas por el gobierno de Estados Unidos se someten a una verificación que evalúa el riesgo para los sujetos encuestados, el mérito científico de la encuesta y la garantía del consentimiento de los sujetos para participar.

Cálculos con tamaños muestrales desiguales

Mientras que los cálculos requeridos para los casos con tamaños muestrales iguales son razonables, las cosas se complican bastante cuando los tamaños muestrales no son iguales. Se aplica el mismo razonamiento básico, porque calculamos un estadístico de prueba F que es el cociente de dos estimados diferentes de la varianza poblacional común σ^2, pero esos estimados implican medidas *ponderadas* que toman en cuenta los tamaños muestrales, tal como se indica a continuación.

$$F = \frac{\text{varianza entre muestras}}{\text{varianza dentro de muestras}} = \frac{\left[\dfrac{\Sigma n_i(\bar{x}_i - \bar{\bar{x}})^2}{k-1}\right]}{\left[\dfrac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}\right]}$$

donde
$\bar{\bar{x}}$ = media de todos los valores muestrales combinados
k = número de medias poblacionales que se están comparando
n_i = número de valores en la i-ésima muestra
$\bar{x}_i$ = media de los valores en la i-ésima muestra
s_i^2 = varianza de los valores en la i-ésima muestra

El factor de n_i está incluido de manera que las muestras más grandes llevan más peso. El denominador del estadístico de prueba es sencillamente la media de las varianzas muestrales, pero se trata de una media ponderada cuyos pesos se basan en los tamaños muestrales.

Como el cálculo de este estadístico de prueba puede conducir a grandes errores de redondeo, los diferentes programas estadísticos de cómputo suelen emplear una expresión distinta (pero equivalente) que implica la notación de la SC (suma de cuadrados) y los CM (cuadrados medios). A pesar de que la siguiente notación y sus componentes son complicados y tediosos, la idea básica es la misma: el estadístico de prueba F es una razón con un numerador que refleja la variación *entre* las medias de las muestras y un denominador que refleja la variación *dentro* de las muestras. Si las poblaciones tienen medias iguales, el cociente F tiende a ser pequeño, pero si las medias poblacionales no son iguales, el cociente F tiende a ser significativamente grande. A continuación se describen los componentes más importantes del método ANOVA.

La **SC (total)** o suma total de cuadrados es una medida de la variación total $\overline{\overline{x}}$ en todos los datos muestrales combinados.

Fórmula 12-1 $\quad \mathrm{SC(total)} = \Sigma(x - \overline{\overline{x}})^2$

La SC(total) se puede separar en los componentes de la SC(del tratamiento) y la SC(del error), descritas como sigue.

La **SC(del tratamiento)**, también llamada SC(del factor), SC(entre grupos) o SC(entre muestras), es una medida de la variación entre las medias muestrales.

Fórmula 12-2

$$\begin{aligned}\mathrm{SC(del\ tratamiento)} &= n_1(\overline{x}_1 - \overline{\overline{x}})^2 + n_2(\overline{x}_2 - \overline{\overline{x}})^2 + \cdots + n_k(\overline{x}_k - \overline{\overline{x}})^2 \\ &= \Sigma n_i(\overline{x}_i - \overline{\overline{x}})^2\end{aligned}$$

Si las medias poblacionales ($\mu_1, \mu_2, \ldots, \mu_k$) son iguales, entonces las medias muestrales $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k$ tenderán a acercarse entre sí y también a acercarse a $\overline{\overline{x}}$. El resultado será un valor de SC(del tratamiento) relativamente pequeño. Sin embargo, si las medias poblacionales no son todas iguales, entonces al menos una de $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_k$ tenderá a estar lejos de las demás y también de $\overline{\overline{x}}$. El resultado será un valor relativamente grande de SC(del tratamiento).

La **SC(del error)**, también conocida como SC(dentro de grupos) o SC(dentro de muestras), es una suma de cuadrados que representa la variación que se supone común a todas las poblaciones consideradas.

Fórmula 12-3

$$\begin{aligned}\mathrm{SC(del\ error)} &= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2 \\ &= \Sigma(n_i - 1)s_i^2\end{aligned}$$

Dadas las expresiones anteriores para SC(total), SC(del tratamiento) y SC(del error), siempre deben mantenerse las siguientes relaciones.

Fórmula 12-4 $\quad \mathrm{SC(total)} = \mathrm{SC(del\ tratamiento)} + \mathrm{SC(del\ error)}$

SC(del tratamiento) y SC(del error) son ambas sumas de cuadrados, y si dividimos cada una de ellas entre su número correspondiente de grados de libertad, obtenemos los cuadrados *medios*. Algunas de las siguientes expresiones para los cuadrados medios incluyen la notación N:

N = número total de valores en todas las muestras combinadas

CM(del tratamiento) es un cuadrado medio de tratamiento, que se obtiene como sigue:

Fórmula 12-5 $$\text{CM(del tratamiento)} = \frac{\text{SC(del tratamiento)}}{k-1}$$

CM(del error) es un cuadrado medio del error, que se obtiene como sigue:

Fórmula 12-6 $$\text{CM(del error)} = \frac{\text{SC(del error)}}{N-k}$$

CM(total) es un cuadrado medio de la variación total, que se obtiene como sigue:

Fórmula 12-7 $$\text{CM(total)} = \frac{\text{SC(total)}}{N-1}$$

Estadístico de prueba para ANOVA con tamaños muestrales desiguales

Al contrastar la hipótesis nula H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ con la hipótesis alternativa de que no todas estas medias son iguales, el estadístico de prueba

Fórmula 12-8 $$F = \frac{\text{CM(del tratamiento)}}{\text{CM(del error)}}$$

tiene una distribución F (cuando la hipótesis nula H_0 es verdadera) con grados de libertad dados por

$$\text{grados de libertad del numerador} = k - 1$$
$$\text{grados de libertad del denominador} = N - k$$

Este estadístico de prueba es esencialmente el mismo que se utilizó antes, y su interpretación también es igual a la descrita con anterioridad. El denominador depende únicamente de las varianzas muestrales que miden la variación dentro de los tratamientos y no se ve afectado por las diferencias entre las medias muestrales. En contraste, el numerador se ve afectado por las diferencias entre las medias muestrales. Si las diferencias entre las medias muestrales son extremas, causarán que el numerador sea excesivamente grande, por lo que F también será excesivamente grande. Como consecuencia, los valores muy grandes de F sugieren medias desiguales y, por lo tanto, la prueba ANOVA es de cola derecha.

Las tablas constituyen un formato adecuado para resumir los resultados más importantes en los cálculos del ANOVA, y la tabla 12-3 tiene un formato que suele utilizarse en el despliegue de resultados de las computadoras. (Véase las pantallas anteriores de STATDISK, Minitab y Excel). Las cifras de la tabla 12-3 resultan de los datos de la tabla 12-1.

Diseño del experimento En el análisis de varianza de un factor (o de una entrada) utilizamos un factor como base para separar los datos en diferentes categorías.

Tabla 12-3 Tabla de ANOVA para los datos de la tabla 12-1

Fuente de variación	Suma de cuadrados (SC)	Grados de libertad	Cuadrado medio (CM)	Estadístico de prueba F	Valor P
Tratamientos	4.6824	3	1.5608	5.7314	0.0073
Error	4.3572	16	0.2723		
Total	9.0397	19			

Si concluimos que las diferencias entre las medias son significativas, no podemos estar completamente seguros de que las diferencias se puedan explicar por el factor utilizado. Es posible que la variación de algún otro factor desconocido sea el responsable. Una forma de reducir el efecto de factores extraños es planear el experimento de manera que tenga un **diseño completamente aleatorizado**, en el que se da a cada elemento la misma posibilidad de pertenecer a las diferentes categorías o tratamientos. Por ejemplo, usted podría asignar sujetos a dos diferentes grupos de tratamiento y a un grupo placebo por medio de un proceso de selección aleatoria equivalente a sacar papeles de un tazón. Otra forma de reducir el efecto de factores extraños es el uso de un **diseño rigurosamente controlado**, en el cual los elementos se eligen cuidadosamente de manera que el resto de los factores no tengan variabilidad. En general, los buenos resultados requieren que el experimento se diseñe y ejecute de forma cuidadosa.

Identificación de medias diferentes

Después de hacer una prueba con el análisis de varianza, podemos concluir que existe evidencia suficiente para rechazar una aseveración de igualdad de medias poblacionales, pero no podemos concluir a partir de un ANOVA que alguna media *en particular* sea diferente de las demás. Existen varios procedimientos formales e informales para identificar las medias específicas que son diferentes. Un método informal consiste en usar la misma escala para construir gráficas de cuadro con los conjuntos de datos, para ver si uno o más de ellos es muy diferente de los otros. Otro método consiste en construir estimados de intervalos de confianza de las medias a partir de los conjuntos de datos, y luego comparar esos intervalos de confianza para ver si uno o más de ellos no se traslapan con los demás.

Anteriormente señalamos que no es adecuado aparear las muestras y realizar pruebas de hipótesis individuales con los procedimientos descritos en la sección 9-3. Con cuatro poblaciones, este método (hacer dos a la vez) requiere de seis pruebas de hipótesis diferentes, de manera que si cada prueba se realiza con un nivel de significancia de 0.05, el nivel general de confianza para las seis pruebas sería demasiado bajo, como 0.95^6 (o 0.735), de manera que el nivel de significancia sería tan alto como $1 - 0.735 = 0.265$. Este nivel de significancia indica que el riesgo de cometer un error tipo I (encontrar una diferencia en uno de los pares cuando en realidad no existe tal diferencia), es demasiado alto.

Existen varios procedimientos para identificar cuáles medias difieren de las demás. Algunas de las pruebas, llamadas **pruebas de rango**, nos permiten localizar subconjuntos de medias que no son significativamente diferentes de las demás.

Otras pruebas, denominadas **pruebas de comparación múltiple**, utilizan pares de medias, pero hacen ajustes para superar el problema de tener un nivel de significancia que aumenta conforme se incrementa el número de pruebas individuales. No existe consenso sobre cuál es la mejor prueba, pero algunas de las más comunes son la prueba de Duncan, la prueba de Student-Newman-Keuls (o prueba SNK), la prueba de Tukey (o prueba de diferencia significativa honesta de Tukey), la prueba de Scheffé, la prueba de Dunnett, la prueba de la diferencia mínima significativa y la prueba de Bonferroni. Ahora aplicaremos esta última con el fin de ver un ejemplo de uno de los métodos para identificar cuáles medias difieren de las demás. El procedimiento es el siguiente:

Prueba de comparación múltiple de Bonferroni

Paso 1. Realice una prueba t separada para cada par de muestras, pero haga los ajustes que se describen en los siguientes pasos.

Paso 2. Para obtener un estimado de la varianza σ^2 que es común a todas las poblaciones implicadas, utilice el valor del CM(del error), que utiliza todos los datos muestrales disponibles. El valor del CM(del error) generalmente se obtiene al realizar la prueba del análisis de varianza. Utilice el valor del CM(del error) para calcular el valor del estadístico de prueba t, como se indica abajo. Este estadístico de prueba en particular se basa en la opción de la muestra 1 y la muestra 4; cambie los subíndices y utilice otro par de muestras hasta haber probado todos los pares de muestras posibles.

$$t = \frac{\bar{x}_1 - \bar{x}_4}{\sqrt{\text{CM(error)} \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_4}\right)}}$$

Paso 3. Después de calcular el valor del estadístico de prueba t para un par específico de muestras, calcule el valor t crítico o el valor P, pero haga el siguiente ajuste para que el nivel de significancia general no se salga de control.

Valor *P*: Utilice el estadístico de prueba t con gl $= N - k$, donde N es el número total de valores muestrales y k es el número de muestras, y calcule el valor P de la manera acostumbrada, pero ajústelo multiplicándolo por el número de pares de muestras diferentes posibles. (Por ejemplo, con cuatro muestras, existen seis pares posibles diferentes, de manera que el valor P se ajusta multiplicándolo por 6).

Valor crítico: Cuando calcule el valor crítico, ajuste el nivel de significancia α dividiéndolo entre el número de pares de muestras posibles. (Por ejemplo, con cuatro muestras, existen seis pares posibles diferentes, de manera que el nivel de significancia se ajusta dividiéndolo entre 6).

Observe que en el paso 3 del procedimiento de Bonferroni se realiza una prueba individual con un nivel de significancia mucho más bajo, o bien, el valor P aumenta de manera importante. Por ejemplo, si tenemos cuatro muestras, existen seis pares diferentes de muestras posibles. Si utilizamos un nivel de significancia de 0.05, cada una de las seis comparaciones apareadas utiliza un nivel de significancia de $0.05/6 = 0.00833$. Por lo tanto, el rechazo de la igualdad de medias requiere de diferencias que estén muy separadas. El grado de confianza para las seis pruebas puede ser tan bajo como $(1 - 0.00833)^6 = 0.95$, de manera que la posibilidad de cometer un error tipo I en una de las seis pruebas es de alrededor de 0.05. Este ajuste en el paso 3 compensa el hecho de que estamos realizando varias pruebas en vez de una sola.

EJEMPLO Uso de la prueba de Bonferroni En un ejemplo previo de esta sección se utilizó el análisis de varianza con los datos muestrales de la tabla 12-1. Concluimos que existe evidencia suficiente para justificar el rechazo de la aseveración de igualdad de medias. Utilice la prueba de Bonferroni, con un nivel de significancia de 0.05, para identificar cuál de las medias difiere de las demás.

SOLUCIÓN La prueba de Bonferroni requiere de una prueba t separada para cada diferente par de muestras posible. Las hipótesis nulas que deben probarse son las siguientes:

$$H_0\colon \mu_1 = \mu_2 \qquad H_0\colon \mu_1 = \mu_3 \qquad H_0\colon \mu_1 = \mu_4$$
$$H_0\colon \mu_2 = \mu_3 \qquad H_0\colon \mu_2 = \mu_4 \qquad H_0\colon \mu_3 = \mu_4$$

Comenzamos con $H_0\colon \mu_1 = \mu_4$. En la tabla 12-1 observamos que $\bar{x}_1 = 0.184$, $n_1 = 5$, $\bar{x}_4 = 1.334$ y $n_4 = 5$. A partir de los resultados anteriores del análisis de varianza, también sabemos que para los datos de la tabla 12-3, CM(del error) $= 0.2723275$. Ahora podemos evaluar el estadístico de prueba:

$$t = \frac{\bar{x}_1 - \bar{x}_4}{\sqrt{\text{CM(del error)} \cdot \left(\dfrac{1}{n_1} + \dfrac{1}{n_4}\right)}} = \frac{0.184 - 1.334}{\sqrt{0.2723275 \cdot \left(\dfrac{1}{5} + \dfrac{1}{5}\right)}} = -3.484$$

El número de grados de libertad es gl $= N - k = 20 - 4 = 16$. Con un estadístico de prueba $t = -3.484$ y con gl $= 16$, el valor P de dos colas es 0.003065, pero ajustamos este valor P al multiplicarlo por 6 (el número de diferentes pares de muestras posibles) para obtener un valor P final de 0.01839. Como este valor P es pequeño (menor que 0.05), rechazamos la hipótesis nula. Parece que las muestras 1 y 4 tienen medias significativamente diferentes. [Si queremos obtener el valor crítico, usamos gl $= 16$ y un nivel de significancia ajustado de $0.05/6 = 0.00833$. (Este nivel de significancia ajustado no se aleja mucho del área de 0.01, que se incluye en la tabla A-3. Por lo tanto, los valores t críticos para la prueba de dos colas están cerca de -2.921 y 2.921). Los valores críticos reales son -3.008 y 3.008. El estadístico de prueba de -3.484 se encuentra en la región crítica, de manera que nuevamente rechazamos la hipótesis nula de que $\mu_1 = \mu_4$)].

En vez de continuar haciendo las pruebas de hipótesis separadas para los cinco pares restantes, observe la pantalla de SPSS, que incluye todos los resultados de la prueba de Bonferroni. (El tercer renglón de los resultados numéricos corresponde a los resultados obtenidos aquí). La pantalla indica que la media de la muestra 4 difiere significativamente de las otras tres medias muestrales. Con base en la prueba de Bonferroni, parece que los pesos de los álamos tratados con fertilizante y riego tienen una media diferente de las tres categorías de tratamiento restantes.

SPSS Bonferroni Results

Multiple Comparisons

Dependent Variable: WEIGHT
Bonferroni

(I) SAMPLE	(J) SAMPLE	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-.1480	.33005	1.000	-1.1409	.8449
	3.00	.0200	.33005	1.000	-.9729	1.0129
	4.00	-1.1500*	.33005	.018	-2.1429	-.1571
2.00	1.00	.1480	.33005	1.000	-.8449	1.1409
	3.00	.1680	.33005	1.000	-.8249	1.1609
	4.00	-1.0020*	.33005	.047	-1.9949	-.0091
3.00	1.00	-.0200	.33005	1.000	-1.0129	.9729
	2.00	-.1680	.33005	1.000	-1.1609	.8249
	4.00	-1.1700*	.33005	.016	-2.1629	-.1771
4.00	1.00	1.1500*	.33005	.018	.1571	2.1429
	2.00	1.0020*	.33005	.047	.0091	1.9949
	3.00	1.1700*	.33005	.016	.1771	2.1629

*. The mean difference is significant at the .05 level.

12-3 ANOVA de dos factores

Concepto clave El procedimiento de análisis de varianza que estudiamos en la sección 12-2 se conoce como análisis de varianza de *un* factor (o análisis de varianza de una entrada), porque los datos están categorizados en grupos de acuerdo con un *solo* factor (o tratamiento). En esta sección estudiaremos el método del *análisis de varianza de dos factores*, que se utiliza con datos separados en categorías formadas de acuerdo con *dos* factores. El método de esta sección requiere que primero hagamos una prueba de *interacción* entre los dos factores. Después, hacemos una prueba para determinar si el factor de renglón tiene algún efecto, y también para determinar si el factor de columna tiene algún efecto.

La tabla 12-4 es un ejemplo de datos clasificados de acuerdo con dos factores. Un factor es la variable de renglón del terreno (terreno 1 y terreno 2), en tanto que el segundo factor es la variable de columna de tratamiento (ninguno, fertilizante, riego, fertilizante y riego). A las subcategorías de la tabla 12-4 se les conoce como *celdas*, de manera que la tabla 12-4 tiene ocho celdas, con cinco valores cada una.

Cuando analizamos los datos muestrales de la tabla 12-4 estudiamos el análisis de varianza de un solo factor, por lo que parecería razonable hacer simplemente un ANOVA de una entrada para el factor del terreno y otro ANOVA de una entrada para el factor de tratamiento. Por desgracia, si realizamos dos ANOVA de un factor por separado, perdemos información e ignoramos totalmente una característica muy importante: el efecto de la interacción entre los dos factores.

Definición

Existe una **interacción** entre dos factores si el efecto de uno de los factores cambia en las diferentes categorías del otro factor.

Como ejemplo de una *interacción* entre dos factores, considere el apareamiento de alimentos. La crema de maní y la mermelada interactúan bien, pero la salsa de tomate y el helado interactúan de tal forma que el resultado es un sabor desagradable. Los médicos deben tener cuidado de evitar la prescripción de fármacos que tienen interacciones que producen efectos adversos. Se descubrió que el fármaco

Tabla 12-4 Pesos de álamos (en kg)

	Sin tratamiento	Fertilizante	Riego	Fertilizante y riego
Terreno 1	0.15	1.34	0.23	2.03
(fértil, húmedo)	0.02	0.14	0.04	0.27
	0.16	0.02	0.34	0.92
	0.37	0.08	0.16	1.07
	0.22	0.08	0.05	2.38
Terreno 2	0.60	1.16	0.65	0.22
(arenoso, seco)	1.11	0.93	0.08	2.13
	0.07	0.30	0.62	2.33
	0.07	0.59	0.01	1.74
	0.44	0.17	0.03	0.12

antimicótico Nizoral (ketoconazole) interactuaba con el fármaco antihistamínico Seldane (terfenadine) de tal manera que el Seldane no se metabolizaba de manera adecuada, provocando arritmias cardiacas en algunos pacientes. Posteriormente el Seldane fue retirado del mercado.

Exploración de los datos Exploremos los datos de la tabla 12-4 calculando la media de cada celda y construyendo una gráfica. En la tabla 12-5 se listan las medias de las celdas individuales. Esas medias van de 0.164 kg hasta 1.334 kg, de manera que parecen variar considerablemente. La figura 12-3 ilustra gráficas de esas medias e indica que las medias del terreno 2 son mayores que las medias del terreno 1 para tres de las cuatro categorías de tratamiento. Puesto que los segmentos lineales del terreno 2 parecen ser aproximadamente *paralelos* a los segmentos lineales correspondientes del terreno 1, aparentemente los pesos por terreno se comportan de la misma forma para las distintas categorías del tratamiento, de manera que no hay un efecto de interacción. En general, si una gráfica como la de la figura 12-3 da por resultado segmentos lineales que son aproximadamente *paralelos*, entonces tenemos evidencia de que *no hay una interacción* entre las variables de renglón y de columna. Si los segmentos lineales del terreno 2 no fueran paralelos a los segmentos lineales del terreno 1, tendríamos evidencia de una interacción entre el terreno y el tratamiento. Estas observaciones basadas en la tabla 12-5 y en la figura 12-3 son muy subjetivas, por lo que procederemos con el método más objetivo del análisis de varianza de dos factores.

Tabla 12-5 Medias (en kg) de las celdas de la tabla 12-4

	Tratamiento			
	Ninguno	Fertilizante	Riego	Fertilizante y riego
Terreno 1 (fértil, húmedo)	0.184	0.332	0.164	1.334
Terreno 2 (arenoso, seco)	0.458	0.630	0.278	1.308

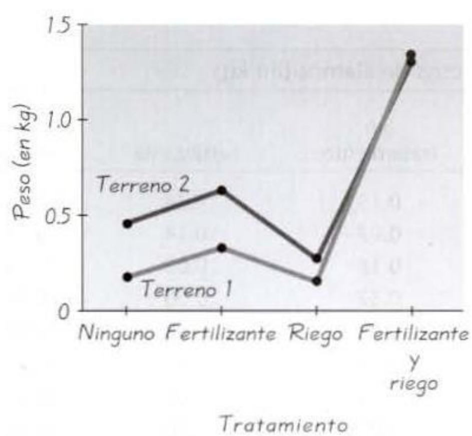


Figura 12-3 Gráfica de interacción de medias (en kg) de las celdas de la tabla 12-4

Encuestas y psicólogos

Los resultados de las encuestas se pueden ver muy afectados por la redacción de las preguntas. Las personas interpretan de forma diferente una frase como "durante los últimos años". Durante los últimos años (en realidad desde 1980), los investigadores de encuestas y los psicólogos han trabajado en conjunto para mejorar las encuestas disminuyendo el sesgo e incrementando la exactitud. En un caso, los psicólogos estudiaron el hallazgo de que del 10 al 15 por ciento de los encuestados afirmaron haber votado en la última elección, cuando en realidad no lo hicieron. Experimentaron con teorías de problemas de memoria, el deseo de ser considerado responsable y la tendencia que manifiestan quienes generalmente votan de decir que votaron en la elección más reciente, aun cuando no lo hayan hecho. Se encontró que sólo esta última explicación formaba parte del problema.

Cuando aplicamos el ANOVA de dos factores para los datos de la tabla 12-4, consideramos tres posibles efectos sobre el peso de los álamos: **1.** los efectos de una interacción entre el terreno y el tratamiento; **2.** los efectos del terreno; **3.** los efectos del tratamiento. Los cálculos son bastante complicados, *de manera que supondremos que se utilizó un programa de cómputo o una calculadora TI-83/84 Plus*. (Al final de esta sección se describen los procedimientos para el uso de herramientas tecnológicas). A continuación se presentan los resultados de Minitab para los datos de la tabla 12-4.

Minitab

Source	DF	SS	MS	F	P
Site	1	0.2722	0.27225	**0.81**	**0.374**
Treatment	3	7.5470	2.51567	**7.50**	**0.001**
Interaction	3	0.1716	0.05721	**0.17**	**0.915**
Error	32	10.7267	0.33521		
Total	39	18.7176			

Ahora presentaremos los requisitos y el procedimiento básico para el análisis de varianza (ANOVA) de dos factores. El procedimiento también se resume en la figura 12-3.

Requisitos

1. Para cada celda, los valores muestrales provienen de una población con una distribución que es aproximadamente normal.
2. Las poblaciones tienen la misma varianza σ^2 (o desviación estándar σ).
3. Las muestras son aleatorias simples. (Es decir, las muestras del mismo tamaño tienen la misma probabilidad de ser seleccionadas).
4. Las muestras son independientes entre sí. (Las muestras no están apareadas o asociadas de ninguna manera).
5. Los valores muestrales se categorizan en dos factores. (Ésta es la base del nombre del método: análisis de varianza de *dos factores*).
6. Todas las celdas tienen el mismo número de valores muestrales. (Este diseño se conoce como diseño *balanceado*).

Procedimiento del ANOVA de dos factores (véase la figura 12-4)

Paso 1: *Efecto de interacción:* En el análisis de varianza de dos factores, inicie probando la hipótesis nula de que no existe interacción entre los dos factores. Si utilizamos Minitab para los datos de la tabla 12-4, obtenemos el siguiente estadístico de prueba:

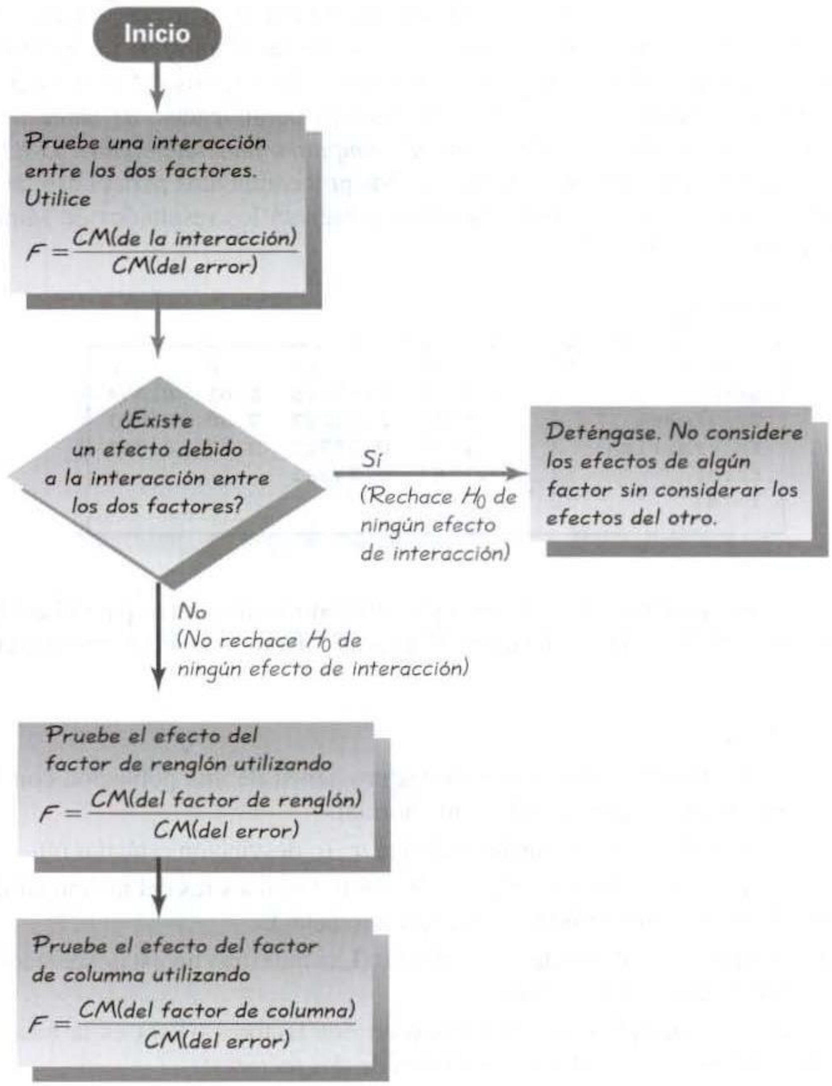
$$F = \frac{\text{CM(de la interacción)}}{\text{CM(del error)}} = \frac{0.05721}{0.33521} = 0.17$$

Interpretación: El valor P correspondiente aparece en los resultados de Minitab como 0.915, por lo que no rechazamos la hipótesis nula de ninguna interacción entre los dos factores. No parece que los pesos de los álamos estén afectados por una interacción entre el terreno y el tratamiento.

Paso 2: *Efectos renglón/columna*: Si rechazamos la hipótesis nula de ninguna interacción entre factores, entonces tenemos que detenernos aquí; no debemos proceder con las dos pruebas adicionales. (Si existe una interacción

Figura 12-4

Procedimiento del ANOVA de dos factores



entre los factores, no debemos considerar los efectos de alguno de los factores sin considerar los del otro).

Si no rechazamos la hipótesis nula de ninguna interacción entre los factores, entonces debemos proceder a probar las siguientes dos hipótesis:

H_0: No existen efectos del factor de renglón (es decir, las medias de renglón son iguales).

H_0: No existen efectos del factor de columna (es decir, las medias de columna son iguales).

En el paso 1 no rechazamos la hipótesis nula de ninguna interacción entre los factores, por lo que procedemos con las siguientes dos pruebas de hipótesis identificadas en el paso 2.

Para el factor de renglón del terreno obtenemos

$$F = \frac{\text{CM(del terreno)}}{\text{CM(del error)}} = \frac{0.27225}{0.33521} = 0.81$$

Interpretación: Este valor no es significativo, ya que el valor P correspondiente aparece en los resultados de Minitab como 0.374. No rechazamos la hipótesis nula de que no existen efectos por el terreno. Es decir, el terreno no parece tener un efecto sobre el peso de los álamos.

Para el factor de columna del tratamiento obtenemos

$$F = \frac{\text{CM(del tratamiento)}}{\text{CM(del error)}} = \frac{2.51567}{0.33521} = 7.50$$

Interpretación: Este valor es significativo, ya que el valor P correspondiente se indica como 0.001. Por lo tanto, rechazamos la hipótesis nula de ningún efecto del tratamiento. Parece que el tratamiento tiene un efecto sobre el peso de los álamos. Con base en los datos muestrales de la tabla 12-4, concluimos que los diferentes tratamientos parecen tener pesos medios diferentes, aunque al parecer los pesos tienen medias iguales en ambos terrenos. (Vea la figura 12-3 y observe que los segmentos lineales cambian drásticamente en las diferentes categorías de tratamiento, pero los segmentos lineales del terreno 1 y del terreno 2 no son muy diferentes entre sí).

Caso especial: Una observación por celda y ninguna interacción La tabla 12-4 contiene 5 observaciones por celda. Si nuestros datos muestrales consisten únicamente en una observación por celda, perdemos CM(de la interacción), SC(de la interacción) y gl(de la interacción), ya que estos valores están basados en varianzas muestrales calculadas para cada celda individual. Si existe sólo una observación por celda, no hay variación dentro de las celdas individuales y esas varianzas muestrales no pueden ser calculadas. Cuando tenemos una observación por celda procedemos de la siguiente manera: *si parece razonable suponer (con base en el conocimiento de las circunstancias) que no existe interacción entre los dos factores, haga dicha suposición y después proceda como antes a probar las siguientes dos hipótesis por separado:*

H_0: No existen efectos del factor de renglón.

H_0: No existen efectos del factor de columna.

Como ejemplo, suponga que tenemos únicamente el primer valor de cada celda de la tabla 12-4, lo que produce los datos de la tabla 12-6. En la tabla 12-6, las dos medias por renglón son 0.9375 y 0.6575. ¿Es esta diferencia significativa, lo que sugiere que existe un efecto debido al terreno? En la tabla 12-6, las cuatro medias de columna son 0.3750, 1.2500, 0.4400 y 1.1250. ¿Son significativas estas diferencias, lo que sugiere que existe un efecto debido al tratamiento? Suponiendo que los pesos no se ven afectados por alguna interacción entre el terreno y el tratamiento,

Tabla 12-6 Pesos (en kg) de álamos en el año 1

	Tratamiento			
	Ninguno	Fertilizante	Riego	Fertilizante y riego
Terreno 1 (fértil, húmedo)	0.15	1.34	0.23	2.03
Terreno 2 (arenoso, seco)	0.60	1.16	0.65	0.22

a continuación se presentan los resultados de Minitab. (Si creemos que existe una interacción, el método descrito aquí no se aplica).

Minitab

Source	DF	SS	MS	F	P
Site	1	0.15680	0.156800	0.28	0.634
Treatment	3	1.23665	0.412217	0.73	0.598
Error	3	1.68690	0.562300		
Total	7	3.08035			

Factor de renglón: Primero empleamos los resultados de la pantalla de Minitab para probar la hipótesis nula de ningún efecto del factor de renglón del terreno.

$$F = \frac{\text{CM(del terreno)}}{\text{CM(del error)}} = \frac{0.156800}{0.562300} = 0.28$$

Este estadístico de prueba no es significativo, debido a que el valor P correspondiente en la pantalla de Minitab es 0.634. No rechazamos la hipótesis nula; parece que los pesos de los álamos no se ven afectados por el terreno.

Factor de columna: Ahora utilizamos la pantalla de Minitab para probar la hipótesis nula de ningún efecto del factor de columna de la categoría de tratamiento. El estadístico de prueba es

$$F = \frac{\text{CM(de la edad)}}{\text{CM(del error)}} = \frac{0.412217}{0.562300} = 0.73$$

Este estadístico de prueba no es significativo, ya que el valor P correspondiente en la pantalla de Minitab es 0.598. No rechazamos la hipótesis nula, de manera que parece que el peso del álamo no se ve afectado por el tratamiento. Utilizando los datos de la tabla 12-6, concluimos que los pesos de los álamos no parecen verse afectados por el terreno ni por el tratamiento, pero cuando utilizamos 5 valores en cada celda (tabla 12-4), concluimos que los pesos parecen verse afectados por el tratamiento. Éste es el poder de las muestras grandes.

En esta sección explicamos brevemente una rama importante de la estadística. Pusimos énfasis en la interpretación de resultados de computadora y omitimos los cálculos y fórmulas manuales, que son bastante engorrosos.

12-3 DESTREZAS Y CONCEPTOS BÁSICOS

Conocimientos estadísticos y pensamiento crítico

1. **ANOVA de dos factores.** ¿Qué es el *análisis de varianza de dos factores* y para qué se utiliza?
2. **Interacción.** ¿Qué es una interacción y qué papel desempeña en el análisis de varianza de dos factores?
3. **Interacción.** Cuando se realiza un análisis de varianza de dos factores, si hay una interacción entre los dos factores, ¿por qué no debemos continuar con pruebas separadas para los efectos de los factores de renglón y de columna?

TEXTO N° 14

Prueba del signo.

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 676 a 685

13-1 Panorama general

Los métodos de estadística inferencial presentados en los capítulos 7, 8, 9, 10 y 12 se llaman *métodos paramétricos* porque se basan en el muestreo de una población con parámetros específicos, como la media μ, la desviación estándar σ o la proporción p. Por lo regular, estos métodos paramétricos deben cumplir con algunas condiciones bastante estrictas, como el requisito de que los datos muestrales provengan de una población distribuida normalmente. Este capítulo describe métodos no paramétricos, los cuales no tienen esos estrictos requisitos.

Definiciones

Las **pruebas paramétricas** tienen requisitos acerca de la naturaleza o forma de las poblaciones implicadas; las **pruebas no paramétricas** no requieren que las muestras provengan de poblaciones con distribuciones normales o con cualquier otro tipo particular de distribución. En consecuencia, las pruebas de hipótesis no paramétricas suelen llamarse **pruebas de distribución libre**.

Aunque el término *no paramétrica* sugiere que la prueba no está basada en un parámetro, existen algunas pruebas no paramétricas que sí dependen de un parámetro como la mediana. Sin embargo, las pruebas no paramétricas no requieren de una distribución particular, por lo que algunas veces se les conoce como pruebas de *distribución libre*. Aunque *distribución libre* es una descripción más precisa, por lo regular se utiliza el término *no paramétrica*. A continuación se enuncian las principales ventajas y desventajas de los métodos no paramétricos.

Ventajas de los métodos no paramétricos

1. Los métodos no paramétricos pueden aplicarse a una amplia variedad de situaciones puesto que no tienen los requisitos más estrictos de los métodos paramétricos correspondientes. En particular, los métodos no paramétricos no requieren de poblaciones distribuidas normalmente.
2. A diferencia de los métodos paramétricos, los métodos no paramétricos a menudo pueden aplicarse a datos categóricos, como el género de quienes responden una encuesta.
3. Los métodos no paramétricos, por lo regular, implican cálculos más sencillos que los métodos paramétricos correspondientes y, por lo tanto, son más fáciles de comprender y aplicar. (Sin embargo, como la tecnología ha simplificado los cálculos, es probable que la facilidad de los cálculos no sea un factor tan importante).

Desventajas de los métodos no paramétricos

1. Los métodos no paramétricos tienden a desperdiciar información porque los datos numéricos exactos suelen reducirse a una forma cualitativa. Por ejemplo, en la prueba del signo no paramétrica (descrita en la sección 13-2), las pérdidas de peso de las personas sometidas a una dieta se registran

simplemente como signos negativos; las magnitudes reales de las pérdidas de peso se ignoran.

2. Las pruebas no paramétricas no son tan eficientes como las pruebas paramétricas, de manera que con una prueba no paramétrica generalmente necesitamos evidencia más fuerte (como una muestra más grande o diferencias mayores) para rechazar una hipótesis nula.

Cuando se satisfacen los requisitos de distribuciones poblacionales, las pruebas no paramétricas generalmente son menos eficaces que sus contrapartes paramétricas, pero la reducción en la eficiencia puede compensarse con un tamaño muestral más grande. Por ejemplo, en la sección 13-6 se presentará un concepto llamado *correlación de rangos*, que tiene una tasa de eficiencia de 0.91 cuando se compara con la correlación lineal presentada en el capítulo 10. Esto significa que, si todo permanece igual, la correlación de rangos no paramétrica requiere 100 observaciones muestrales para obtener los mismos resultados que 91 observaciones muestrales analizadas por medio de la correlación lineal paramétrica, suponiendo que se satisfacen los requisitos más estrictos para la aplicación del método paramétrico. La tabla 13-2 lista los métodos no paramétricos cubiertos en este capítulo, junto con el método paramétrico correspondiente y la tasa de **eficiencia**. La tabla 13-2 indica que varias pruebas no paramétricas tienen tasas de eficiencia por encima de 0.90, por lo que la eficiencia más baja tal vez no sea un factor esencial para elegir entre los métodos paramétricos y no paramétricos. Sin embargo, puesto que las pruebas paramétricas tienen tasas de eficiencia más altas que sus contrapartes no paramétricas, generalmente es mejor utilizar las pruebas paramétricas cuando sus supuestos requeridos se satisfacen.

Rangos

Las secciones 13-3 a 13-6 utilizan métodos basados en rangos, que describiremos a continuación.

Definición

Los datos están *ordenados* cuando se acomodan de acuerdo con algún criterio, por ejemplo, del más pequeño al más grande o del mejor al peor. Un **rango** es un número asignado a un elemento muestral individual de acuerdo con su lugar en la lista ordenada. Al primer elemento se le asigna un rango de 1, al segundo elemento se le asigna un rango de 2 y así sucesivamente.

Tabla 13-2 Eficiencia: Comparación de pruebas paramétricas y no paramétricas

Aplicación	Prueba paramétrica	Prueba no paramétrica	Tasa de eficiencia de una prueba no paramétrica con población normal
Datos muestrales apareados	Prueba *t* o prueba *z*	Prueba del signo	0.63
		Prueba de rangos con signo de Wilcoxon	0.95
Dos muestras independientes	Prueba *t* o prueba *z*	Prueba de la suma de rangos de Wilcoxon	0.95
Varias muestras independientes	Análisis de varianza (prueba *F*)	Prueba de Kruskal-Wallis	0.95
Correlación	Correlación lineal	Prueba de correlación de rangos ordenados	0.91
Aleatoriedad	Prueba no paramétrica	Prueba de rachas	Sin bases para comparar

Manejo de rangos empatados: Si ocurre un empate en los rangos, el procedimiento habitual es calcular la media de los rangos implicados y luego asignar este rango medio a cada uno de los elementos empatados, como en el siguiente ejemplo.

EJEMPLO Los números 4, 5, 5, 5, 10, 11, 12 y 12 tienen rangos dados de 1, 3, 3, 3, 5, 6, 7.5 y 7.5, respectivamente. Vea la siguiente tabla y observe el procedimiento para manejar empates.

Datos ordenados	Rango preliminar		Rango
4	1		**1**
5	2		**3**
5	3	La medida es 3.	**3**
5	4		**3**
10	5		**5**
11	6		**6**
12	7	La medida es 7.5.	**7.5**
12	8		**7.5**

13-2 Prueba del signo

Concepto clave El objetivo principal de esta sección es entender el procedimiento de la *prueba del signo*, el cual implica convertir valores de datos en signos positivos y negativos, y luego hacer una prueba para ver si hay una cantidad desproporcionadamente mayor de uno u otro signo.

Definición

La **prueba del signo** es una prueba no paramétrica (de distribución libre) que utiliza signos positivos y negativos para probar diferentes aseveraciones, incluyendo:

1. Aseveraciones que implican datos muestrales apareados
2. Aseveraciones que implican datos nominales
3. Aseveraciones acerca de la mediana de una sola población

Concepto básico de la prueba del signo

La idea básica que subyace en la prueba del signo es el análisis de las frecuencias de los signos positivos y negativos para determinar si son significativamente diferentes. Por ejemplo, suponga que probamos un tratamiento diseñado para incrementar la probabilidad de que un bebé sea niña. Si se trata a 100 mujeres y 51 de ellas tienen niñas, el sentido común sugiere que no existe evidencia suficiente para afirmar que el tratamiento es efectivo, puesto que 51 niñas entre 100 bebés no son significativas. Pero ¿qué sucede con 52 niñas y 48 niños? ¿O con 90 niñas y 10 niños? La prueba del signo nos permite determinar cuándo este tipo de resultados son significativos.

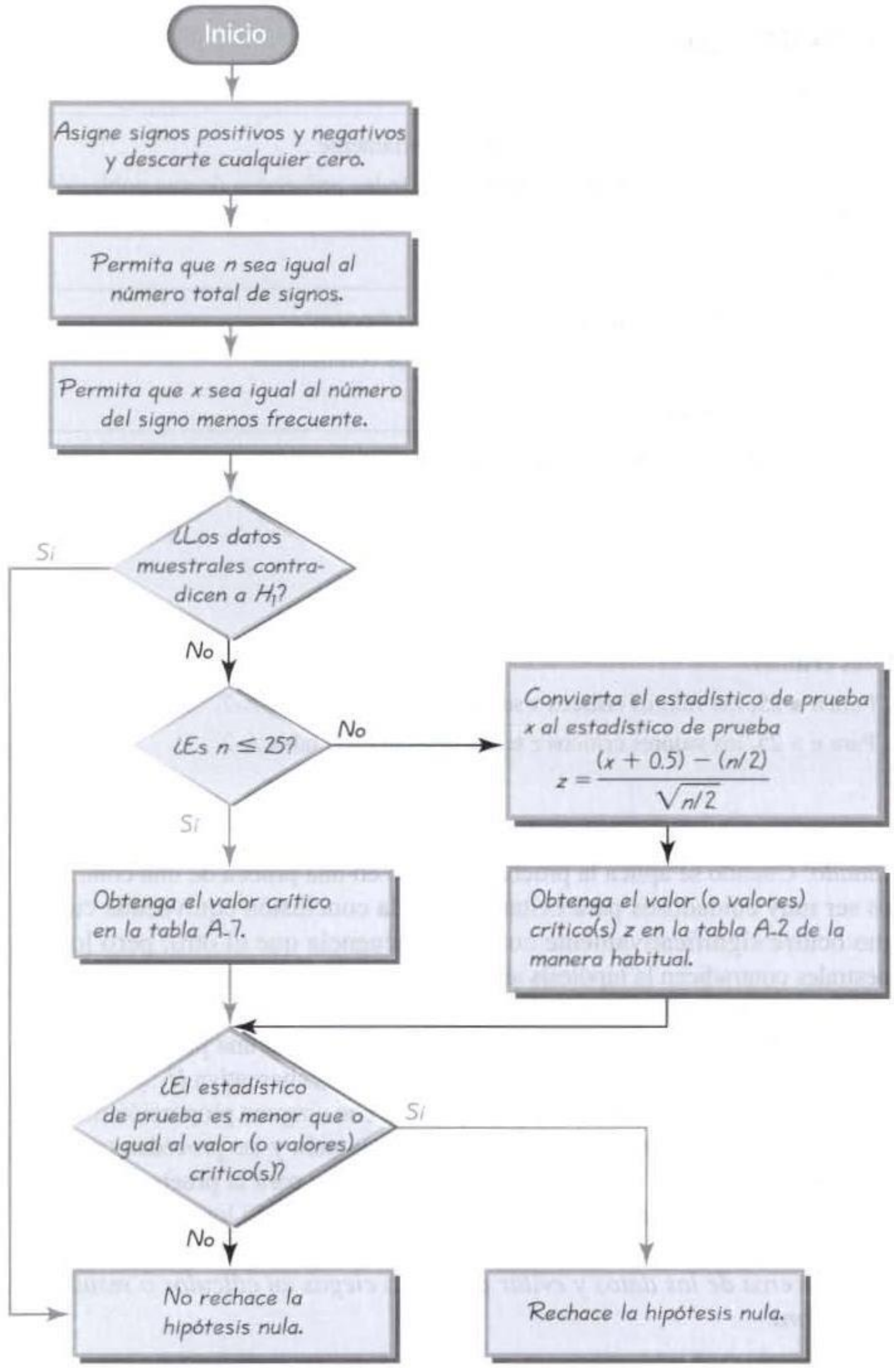


Figura 13-1

Procedimiento de la prueba del signo

Por razones de consistencia y simplicidad, utilizaremos un estadístico de prueba con base en el número de veces que ocurre el signo *menos frecuente*. En el cuadro adjunto se resumen los supuestos relevantes, la notación, el estadístico de prueba y los valores críticos. La figura 13-1 resume el procedimiento de la prueba del signo, que se ilustrará con los ejemplos que siguen.

Asistencia a clases y calificaciones

En un estudio de 424 estudiantes de licenciatura de la Universidad de Michigan, se encontró que los estudiantes con los peores registros de asistencia tendían a obtener las calificaciones más bajas. (¿Quién se sorprende?). Aquellos que estuvieron ausentes menos del 10% del tiempo tendieron a recibir calificaciones de B o superiores. El estudio también reveló que los estudiantes que se sientan al frente en el salón de clases tienden a obtener calificaciones significativamente mejores.

Prueba del signo

Requisitos

1. Los datos muestrales se seleccionaron aleatoriamente.
2. *No* existe el requisito de que los datos muestrales provengan de una población con una distribución particular, como una distribución normal.

Notación

x = el número de veces que ocurre el signo *menos frecuente*

n = el número total de signos positivos y negativos combinados

Estadístico de prueba

Para $n \leq 25$: x (el número de veces que ocurre el signo menos frecuente)

Para $n > 25$: $$z = \frac{(x + 0.5) - \left(\frac{n}{2}\right)}{\frac{\sqrt{n}}{2}}$$

Valores críticos

1. Para $n \leq 25$, los valores críticos x se encuentran en la tabla A-7.
2. Para $n > 25$, los valores críticos z se encuentran en la tabla A-2.

Cuidado: Cuando se aplica la prueba del signo en una prueba de una cola, necesitamos ser muy cuidadosos para evitar obtener la conclusión equivocada cuando un signo ocurre significativamente con más frecuencia que el otro, pero los datos muestrales contradicen la hipótesis alternativa. Por ejemplo, suponga que estamos probando la aseveración de que una técnica de selección del género favorece a los niños, pero obtenemos una muestra de 10 niños y 90 niñas. Con una proporción muestral de niños igual a 0.10, los datos contradicen la hipótesis alternativa H_1: $p > 0.5$. No hay forma de sustentar la aseveración de que $p > 0.5$ con ninguna proporción muestral menor que 0.5, por lo que no rechazamos la hipótesis nula y no procedemos con la prueba del signo. La figura 13-1 resume el procedimiento para la prueba del signo e incluye esta revisión: ¿Contradicen los datos muestrales a H_1? Si los datos muestrales van en el sentido opuesto de H_1, no rechace la hipótesis nula. *Siempre es importante reflexionar acerca de los datos y evitar confiar a ciegas en cálculos o resultados de computadora.*

Aseveraciones que implican datos apareados

Cuando se utiliza la prueba del signo con datos que están ordenados en pares, convertimos los datos en bruto a datos con signos positivos y negativos como sigue:

1. Restamos cada valor de la segunda variable del valor correspondiente de la primera variable.
2. Registramos sólo el *signo* de la diferencia encontrada en el paso 1. *Excluimos los empates:* es decir, excluimos todos los datos apareados en los que ambos valores son iguales.

Éste es el concepto clave que subyace en la aplicación de la prueba del signo:

Si dos conjuntos de datos tienen medianas iguales, el número de signos positivos debe ser aproximadamente igual al número de signos negativos.

EJEMPLO ¿El tipo de semilla afecta el crecimiento del maíz? En 1908 William Gosset publicó el artículo "The Probable Error of a Mean" bajo el seudónimo de "Student" (*Biometrika*, vol. 6, núm. 1). Él incluyó los datos que se listan en la tabla 13-3 para dos tipos diferentes de semillas de maíz (normales y secadas en horno), que se utilizaron en parcelas de tierra *adyacentes*. Los valores corresponden a las cosechas de cabezas de maíz (o mazorcas) en libras por acre. Utilice la prueba del signo con un nivel de significancia de 0.05 para probar la aseveración de que no hay diferencia entre las cosechas de las semillas normales y las de las semillas secadas en horno.

SOLUCIÓN

REQUISITO ✔ El único requisito es que los datos muestrales se obtengan al azar. No hay un requisito sobre la distribución de la población, como el hecho de que los datos muestrales provengan de una población distribuida normalmente. Con base en el diseño de este experimento, suponemos que los datos muestrales son aleatorios. ✔

La siguiente es la idea básica: si no existe diferencia entre las cosechas de las semillas normales y las cosechas de las semillas secadas en horno, el número de signos positivos y negativos debe ser aproximadamente igual. En la tabla 13-3 tenemos 7 signos negativos y 4 signos positivos. ¿Son aproximadamente iguales los números de signos positivos y negativos, o son significativamente diferentes? Seguimos los mismos pasos básicos de prueba de hipótesis, tal como se describieron en la figura 8-9, y aplicamos el procedimiento de la prueba del signo que se resume la figura 13-1.

Pasos 1, 2 y 3: La hipótesis nula es la aseveración de que no hay diferencia entre las cosechas de las semillas normales y las cosechas de las semillas secadas en horno, y la hipótesis alternativa es la aseveración de que existe una diferencia.

H_0: No existe diferencia (la mediana de las diferencias es igual a 0).

H_1: Existe una diferencia (la mediana de las diferencias no es igual a 0).

Paso 4: El nivel de significancia es $\alpha = 0.05$.

Paso 5: Utilizamos la prueba no paramétrica del signo.

continúa

Tabla 13-3 Cosechas de maíz de diferentes semillas

Normales	1903	1935	1910	2496	2108	1961	2060	1444	1612	1316	1511
Secadas en horno	2009	1915	2011	2463	2180	1925	2122	1482	1542	1443	1535
Signo de la diferencia	−	+	−	+	−	+	−	−	+	−	−

Paso 6: El estadístico de prueba x es el número de veces que se presenta el signo menos frecuente. La tabla 13-3 incluye diferencias con 7 signos negativos y 4 signos positivos. (Si hubiera cualquier diferencia igual a 0, la descartaríamos). Permitimos que x sea igual al menor de 7 y 4, así que $x = 4$. Además, $n = 11$ (el número total de signos positivos y negativos combinados). Nuestra prueba es de dos colas con $\alpha = 0.05$. Nos remitimos a la tabla A-7, donde se encuentra el valor crítico de 1 para $n = 11$ y $\alpha = 0.05$ en dos colas. (Véase la figura 13-1).

Paso 7: Con un estadístico de prueba de $x = 4$ y un valor crítico de 1, no rechazamos la hipótesis nula de no diferencia. [Véase la nota 2 incluida en la tabla A-7: "La hipótesis nula se rechaza si el número del signo menos frecuente (x) es menor que o igual al valor en la tabla". Puesto que $x = 4$ *no* es menor que o igual al valor crítico de 1, no rechazamos la hipótesis nula].

Paso 8: No hay suficiente evidencia para sustentar el rechazo de la aseveración de que la mediana de las diferencias es igual a 0; esto es, no existe suficiente evidencia para justificar el rechazo de la aseveración de que no existe una diferencia entre las cosechas de las semillas normales y las cosechas de las semillas secadas en horno. Ésta es la misma conclusión a la que se llegaría utilizando la prueba paramétrica t con datos apareados de la sección 9-4, pero los resultados de la prueba del signo no siempre coinciden con los resultados de la prueba paramétrica.

Aseveraciones que implican datos nominales

En el capítulo 1 definimos los datos nominales como aquellos que consisten sólo en nombres, etiquetas o categorías. La naturaleza de los datos nominales limita los cálculos posibles, pero podemos identificar la *proporción* de datos muestrales que pertenecen a una categoría en particular y podemos probar aseveraciones acerca de la proporción poblacional p correspondiente. El siguiente ejemplo utiliza datos nominales que consisten en el género (niñas/niños). La prueba del signo se utiliza representando a las niñas con signos positivos $(+)$ y a los niños con signos negativos $(-)$. (Créanme, esto signos se eligieron arbitrariamente). También observe el procedimiento para manejar casos en los que $n > 25$.

EJEMPLO Selección del género El Genetics and IVF Institute realizó un ensayo clínico de sus métodos de selección del género. Para cuando se escribía este libro, los resultados incluían a 325 bebés nacidos de padres que utilizaron el método XSORT para aumentar la probabilidad de concebir una niña, y 295 de esos bebés fueron niñas. Utilice la prueba del signo con un nivel de significancia de 0.05 y pruebe la aseveración de que este método de selección del género no tiene ningún efecto.

SOLUCIÓN

REQUISITO ✔ El único requisito es que los datos muestrales se seleccionen al azar. Con base en el diseño de este experimento, podemos suponer que los datos muestrales son aleatorios. Ahora podemos proceder con la prueba del signo. ✔

Permita que p denote la proporción poblacional de niñas. La aseveración de ningún efecto implica que las proporciones de niñas y niños sean iguales a

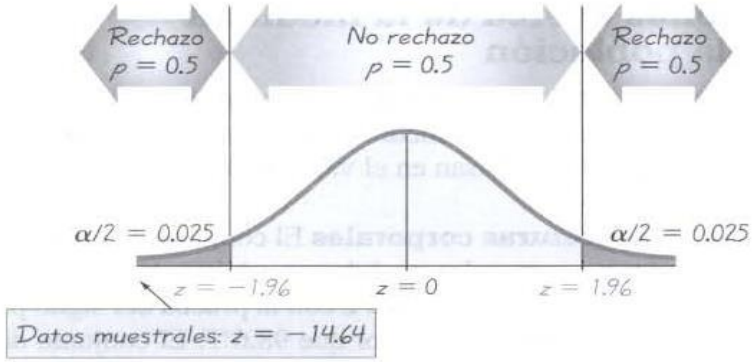


Figura 13-2 Prueba del efecto de un método de selección del género

0.5, de manera que $p = 0.5$. Por lo tanto, las hipótesis nula y alternativa pueden establecerse de la siguiente manera:

H_0: $p = 0.5$ **(la proporción de niñas es igual a 0.5)**
H_1: $p \neq 0.5$

Si denotamos a las niñas con signo positivo (+) y a los niños con signo negativo (−), tenemos 295 signos positivos y 30 signos negativos. Ahora consulte el procedimiento de prueba del signo que se resume en la figura 13-1. El estadístico de prueba x es el menor de 295 y 30, así que $x = 30$. Esta prueba implica dos colas puesto que un número desproporcionadamente alto o bajo de niñas nos llevará a rechazar la aseveración de $p = 0.5$. Los datos muestrales no contradicen la hipótesis alternativa puesto que 295 y 30 no son precisamente iguales. (Esto es, los datos muestrales son consistentes con la hipótesis alternativa de una diferencia). Continuando con el procedimiento de la figura 13-1, notamos que el valor de $n = 325$ es superior a 25, por lo que el estadístico de prueba x se convierte (utilizando una corrección por continuidad) al estadístico de prueba z como sigue:

$$z = \frac{(x + 0.5) - \left(\frac{n}{2}\right)}{\frac{\sqrt{n}}{2}}$$

$$= \frac{(30 + 0.5) - \left(\frac{325}{2}\right)}{\frac{\sqrt{325}}{2}} = -14.64$$

Con $\alpha = 0.05$ en una prueba de dos colas, los valores críticos son $z = \pm 1.96$. El estadístico de prueba $z = -14.64$ es menor que -1.96 (véase la figura 13-2), por lo que rechazamos la hipótesis nula de que la proporción de niñas es igual a 0.5. Tenemos evidencia muestral suficiente para justificar el rechazo de la aseveración de que el método de selección del género no tiene efecto alguno (ya que las proporciones de niñas y niños son iguales a 0.5). Parece que este método afecta el género de los bebés.

Aseveraciones acerca de la mediana de una sola población

El siguiente ejemplo ilustra el procedimiento para utilizar la prueba del signo al probar una aseveración acerca de la mediana de una sola población. Vea cómo los signos positivos y negativos se basan en el valor aseverado de la mediana.

EJEMPLO Temperaturas corporales El conjunto de datos 2 del apéndice B incluye temperaturas corporales medidas en adultos. Utilice las 106 temperaturas listadas para las 12:00 a.m. del día 2 con la prueba del signo para probar la aseveración de que la mediana es menor que 98.6°F. El conjunto de datos tiene 106 sujetos, 68 sujetos con temperaturas por debajo de 98.6°F, 23 sujetos con temperaturas por arriba de 98.6°F y 15 sujetos con temperaturas iguales a 98.6°F.

SOLUCIÓN

REQUISITO ✔ El único requisito es que los datos muestrales se seleccionen al azar y, con base en el diseño de este experimento, suponemos que los datos muestrales son aleatorios. Ahora podemos proceder con la prueba del signo. ✔

La aseveración de que la mediana es menor que 98.6°F es la hipótesis alternativa, mientras que la hipótesis nula es la aseveración de que la mediana es igual a 98.6°F.

H_0: La mediana es igual a 98.6°F. **(mediana = 98.6°F)**
H_1: La mediana es menor que 98.6°F. **(mediana < 98.6°F)**

Siguiendo el procedimiento descrito en la figura 13-1, descartamos los 15 ceros, utilizamos el signo negativo $(-)$ para denotar cada temperatura que está por debajo de 98.6°F y utilizamos el signo positivo $(+)$ para denotar cada temperatura que está por encima de 98.6°F. Por lo tanto, tenemos 68 signos negativos y 23 signos positivos, entonces $n = 91$ y $x = 23$ (el número del signo menos frecuente). Los datos muestrales no contradicen la hipótesis alternativa, puesto que la mayoría de las 91 temperaturas están por debajo de 98.6°F. (Si los datos muestrales indicaran un conflicto con la hipótesis alternativa, podríamos terminar inmediatamente la prueba concluyendo que no rechazamos la hipótesis nula). El valor de n excede a 25, por lo que convertimos el estadístico de prueba x al estadístico de prueba z:

$$z = \frac{(x + 0.5) - \left(\frac{n}{2}\right)}{\frac{\sqrt{n}}{2}}$$

$$= \frac{(23 + 0.5) - \left(\frac{91}{2}\right)}{\frac{\sqrt{91}}{2}} = -4.61$$

En esta prueba de una cola con $\alpha = 0.05$, utilizamos la tabla A-2 para obtener el valor crítico z de -1.645. En la figura 13-3 podemos ver que el estadístico de prueba $z = -4.61$ cae dentro de la región crítica; por lo tanto, rechazamos la hipótesis nula. Con base en la evidencia muestral disponible, sustentamos la aseveración de que la mediana de la temperatura corporal de adultos saludables es menor que 98.6°F.

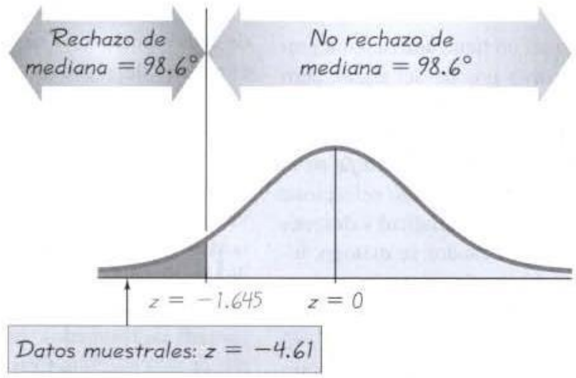


Figura 13-3

Prueba de la aseveración de que la mediana es menor que 98.6°F

En esta prueba del signo para la aseveración de que la mediana está por debajo de 98.6°F, obtenemos un estadístico de prueba de $z = -4.61$ con un valor P de 0.00000202, pero una prueba paramétrica de la aseveración de que $\mu < 98.6$°F produce un estadístico de prueba $t = -6.611$, con un valor P de 0.000000000813. Puesto que el valor P de la prueba del signo no es tan bajo como el valor P de la prueba paramétrica, vemos que la prueba del signo no es tan sensible como la prueba paramétrica. Ambas pruebas nos llevan al rechazo de la hipótesis nula, pero la prueba del signo no considera que los datos muestrales sean tan extremos, en parte porque la prueba del signo utiliza sólo información acerca de la *dirección* de los datos, ignorando las *magnitudes* de los valores de los datos. En la siguiente sección se estudiará la prueba de rangos con signo de Wilcoxon, que supera con creces esta desventaja.

Fundamentos para el estadístico de prueba que se utiliza cuando $n > 25$: Cuando se calculan valores críticos para la prueba del signo, utilizamos la tabla A-7 sólo para n hasta 25. Cuando $n > 25$, el estadístico de prueba z se basa en una aproximación normal a la distribución de probabilidad binomial con $p = q = 1/2$. Recuerde que en la sección 6-6 vimos que la aproximación normal a la distribución binomial es aceptable cuando $np \geq 5$ y $nq \geq 5$. Recuerde también que en la sección 5-4 vimos que $\mu = np$ y $\sigma = \sqrt{npq}$ para distribuciones de probabilidad binomial. Puesto que esta prueba del signo supone que $p = q = 1/2$, satisfacemos los prerrequisitos de que $np \geq 5$ y $nq \geq 5$ siempre y cuando $n \geq 10$. Además, con el supuesto de que $p = q = 1/2$, obtenemos $\mu = np = n/2$ y $\sqrt{npq} = \sqrt{n/4} = \sqrt{n}/2$, por lo tanto

$$z = \frac{x - \mu}{\sigma}$$

se convierte en

$$z = \frac{x - \left(\frac{n}{2}\right)}{\frac{\sqrt{n}}{2}}$$

Finalmente, reemplazamos x con $x + 0.5$ como una corrección por continuidad. Esto es, los valores de x son discretos, pero puesto que estamos utilizando una distribución de probabilidad continua, un valor discreto como 10 se representa realmente con el intervalo de 9.5 a 10.5. Como x representa el signo menos frecuente, actuamos conservadoramente interesándonos sólo por $x + 0.5$; así obtenemos el estadístico de prueba z, como en la ecuación y en la figura 13-1.

TEXTO N° 15

Prueba de rangos con signos de wilcoxon para datos apareados.

Prueba de la suma de rangos de wilcoxon para dos muestras independientes

Compilado y adaptado de Mario Triola Estadística PEARSON; 2009

Págs. De 689 a 693, y 695 a 697

13-3 Prueba de rangos con signo de Wilcoxon para datos apareados

Concepto clave Esta sección se ocupa de la *prueba de rangos con signo de Wilcoxon*, que se usa con datos muestrales apareados. Esta prueba se utiliza con la hipótesis nula de que la población de diferencias de los datos apareados tiene una mediana igual a cero.

La prueba del signo (sección 13-2) también se puede usar con datos apareados, pero la prueba del signo sólo utiliza los signos de las diferencias. Al utilizar los rangos, la prueba de rangos con signo (o de rangos signados) de Wilcoxon toma en cuenta las magnitudes de las diferencias. Puesto que la prueba de rangos con signo de Wilcoxon incorpora y utiliza más información que la prueba del signo, tiende a arrojar conclusiones que reflejan mejor la verdadera naturaleza de los datos.

Definición

La **prueba de rangos con signo de Wilcoxon** es una prueba no paramétrica que utiliza rangos ordenados de datos muestrales que consisten en datos apareados. Se usa para probar la hipótesis nula de que la población de diferencias tiene una mediana de cero, de manera que las hipótesis nula y alternativa son las siguientes:

H_0: Los datos apareados tienen diferencias que provienen de una población con una mediana igual a cero.

H_1: Los datos apareados tienen diferencias que provienen de una población con una mediana diferente de cero.

(La prueba de rangos con signo de Wilcoxon también puede usarse para probar la aseveración de que una muestra proviene de una población con una mediana específica. Véase el ejercicio 13 para esta aplicación).

Brecha de género en las pruebas de fármacos

Un estudio de la relación entre los ataques cardiacos y las dosis administradas de aspirina incluyó a 22,000 médicos varones. Este estudio, como muchos otros, excluyó a las mujeres. La General Accounting Office criticó hace poco a los institutos nacionales de salud por no incluir a ambos sexos en muchos estudios, ya que los resultados de pruebas médicas en hombres no necesariamente se aplican a las mujeres. Por ejemplo, los corazones de las mujeres son diferentes de los de los hombres en muchos aspectos importantes. Cuando saquemos conclusiones con base en resultados muestrales, debemos ser cuidadosos al generalizar las inferencias a una población más grande que aquella de la cual se obtuvo la muestra.

Prueba de rangos con signo de Wilcoxon

Requisitos

1. Los datos consisten en datos apareados que se seleccionaron aleatoriamente.
2. La población de las diferencias (calculadas a partir de los pares de datos) tiene una distribución que es aproximadamente *simétrica*, lo que quiere decir que la mitad izquierda de su histograma es aproximadamente una imagen de espejo de la mitad derecha. (*No* existe el requisito de que los datos tengan una distribución normal).

Notación

El procedimiento para calcular la suma de rangos T se incluye después de este recuadro.
T = la más pequeña de las siguientes dos sumas:

1. La suma de los valores absolutos de los rangos negativos de las diferencias d que no sean cero.
2. La suma de los rangos positivos de las diferencias d que no sean cero.

Estadístico de prueba

Si $n \leq 30$, el estadístico de prueba es T.

Si $n > 30$, el estadístico de prueba es $z = \dfrac{T - \dfrac{n(n+1)}{4}}{\sqrt{\dfrac{n(n+1)(2n+1)}{24}}}$

Valores críticos

1. Si $n \leq 30$, el valor crítico T se encuentra en la tabla a A-8.
2. Si $n > 30$, los valores críticos z se encuentran en la tabla A-2.

Procedimiento de la prueba de rangos con signo de Wilcoxon

Paso 1: Para cada par de datos, calcule la diferencia d restando el segundo valor del primero. Mantenga los signos, pero descarte cualquier par para el que $d = 0$.

Paso 2: *Ignore los signos de las diferencias*, luego acomode las diferencias de la menor a la mayor y reemplácelas por el valor del rango correspondiente (como se describe en la sección 13-1). Cuando las diferencias tengan el mismo valor numérico, asígneles la media de los rangos implicados en el empate.

Paso 3: Agregue a cada rango el signo de la diferencia de la que provino. Esto es, inserte aquellos signos que se ignoraron en el paso 2.

Paso 4: Calcule la suma de los valores absolutos de los rangos negativos. También calcule la suma de los rangos positivos.

Paso 5: Permita que T sea *la más pequeña* de las dos sumas calculadas en el paso 4. Podría utilizarse cualquier suma, pero para simplificar el procedimiento seleccionamos arbitrariamente la más pequeña de las dos sumas. (Véase la notación para T en el recuadro anterior).

Paso 6: Permita que n sea el número de pares de datos para los que la diferencia d no es 0.

Paso 7: Determine el estadístico de prueba y los valores críticos con base en el tamaño muestral, como se indica en el recuadro anterior.

Paso 8: Cuando plantee la conclusión, rechace la hipótesis nula si los datos muestrales le llevan a un estadístico de prueba que se ubica en la región crítica, esto es, cuando el estadístico de prueba sea menor o igual que el valor (o los valores) crítico(s). De otra forma, no rechace la hipótesis nula.

EJEMPLO ¿El tipo de semilla afecta el crecimiento del maíz? En 1908 William Gosset publicó el artículo "The Probable Error of a Mean" bajo el seudónimo de "Student" (*Biometrika*, vol. 6, núm. 1). Él incluyó los datos que se listan en la tabla 13-4 para dos tipos diferentes de semillas de maíz (normales y secadas en horno), que se utilizaron en parcelas de tierra *adyacentes*. Los valores corresponden a las cosechas de cabezas de maíz (o mazorcas) en libras por acre. Utilice la prueba de rangos con signos de Wilcoxon, con un nivel de significancia de 0.05, para probar la aseveración de que no hay diferencia entre las cosechas de las semillas normales y de las semillas secadas en horno.

SOLUCIÓN

REQUISITOS ✔ Debemos tener datos apareados seleccionados al azar. Los datos están apareados y, dado el diseño de este experimento, es razonable suponer que los datos apareados fueron elegidos al azar. Además, el siguiente histograma generado por Minitab indica que la distribución de las diferencias es aproximadamente simétrica, tal como se requiere. (Es decir, el lado izquierdo de la gráfica es aproximadamente una imagen de espejo del lado derecho. A simple vista no parecen ser simétricos, pero con sólo 11 valores, la diferencia entre el lado izquierdo y el lado derecho no es demasiado pronunciada). ✔

continúa

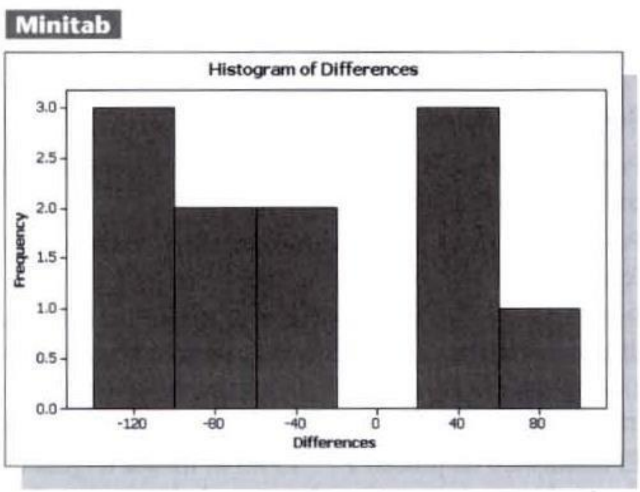


Tabla 13-4 Cosechas de maíz de diferentes semillas

Normales	1903	1935	1910	2496	2108	1961	2060	1444	1612	1316	1511
Secadas en horno	2009	1915	2011	2463	2180	1925	2122	1482	1542	1443	1535
Diferencias d	−106	20	−101	33	−72	36	−62	−38	70	−127	−24
Rangos de |diferencias|	10	1	9	3	8	4	6	5	7	11	2
Rangos con signo	−10	1	−9	3	−8	4	−6	−5	7	−11	−2

Las hipótesis nula y alternativa son como sigue:

H_0: Las cosechas de las semillas normales y de las semillas secadas en horno son tales que la mediana de la población de las diferencias es igual a cero.

H_1: La mediana de la población de diferencias no es igual a cero.

El nivel de significancia es $\alpha = 0.05$. Estamos utilizando el procedimiento de la prueba de rangos con signo de Wilcoxon, por lo que el estadístico de prueba se calcula aplicando el procedimiento de ocho pasos presentado con anterioridad en esta sección.

Paso 1: En la tabla 13-4 el renglón de diferencias se obtiene calculando esta diferencia para cada par de datos:

d = cosecha de las semillas normales − cosecha de las semillas secadas en horno

Paso 2: Ignorando sus signos, ordenamos los rangos de las diferencias absolutas de la menor a la mayor. (Si existiera algún empate en los rangos tendríamos que manejarlos asignando la media de los rangos implicados a cada uno de los valores empatados. Además, tendríamos que descartar cualquier diferencia de 0).

Paso 3: El renglón inferior de la tabla 13-4 se crea insertando a cada rango el signo de la diferencia correspondiente. Si en realidad no existe diferencia entre las cosechas de los dos tipos de semillas (como en la hipótesis nula), esperamos que la suma de los rangos positivos sea aproximadamente igual a la suma de los valores absolutos de los rangos negativos.

Paso 4: Ahora calculamos la suma de los valores absolutos de los rangos negativos y también calculamos la suma de los rangos positivos.

Suma de los valores absolutos de los rangos negativos: 51 (de $10 + 9 + 8 + 6 + 5 + 11 + 2$)

Suma de los rangos positivos: 15 (de $1 + 3 + 4 + 7$)

Paso 5: Permitiendo que T sea la menor de las dos sumas calculadas en el paso 4, encontramos que $T = 15$.

Paso 6: Permitiendo que n sea el número de pares de datos para los que la diferencia d no es 0, tenemos $n = 11$.

Paso 7: Puesto que $n = 11$, tenemos que $n \leq 30$, por lo cual utilizamos un estadístico de prueba de $T = 15$ (y no calculamos un estadístico de prueba z). Además, puesto que $n \leq 30$, utilizamos la tabla de A-8 para encontrar el valor crítico de 11 (utilizando $n = 11$ y $\alpha = 0.05$ en dos colas).

Paso 8: El estadístico de prueba $T = 15$ no es menor o igual que el valor crítico de 11, por lo que no rechazamos la hipótesis nula. Parece que no hay una diferencia entre las cosechas de las semillas normales y de las semillas secadas en horno.

Si utilizamos la prueba del signo con el ejemplo anterior, llegaremos a la misma conclusión. Aunque la prueba de signo y la prueba de rangos con signo de Wilcoxon coinciden en este caso en particular, existen otros casos en los que no concuerdan.

Fundamentos: En este ejemplo los rangos sin signo de 1 hasta 11 tienen un total de 66, de manera que si no existen diferencias significativas, cada uno de los dos totales de rangos con signo debe ser de alrededor de $66 \div 2$, o 33. Esto es, los rangos negativos y los rangos positivos deberían repartirse como 33-33 o algo cercano, tal como 31-35. La tabla de valores críticos indica que a un nivel de significancia de 0.05, con 11 pares de datos, un reparto de 11-55 representa una desviación significativa de la hipótesis nula, y cualquier reparto que esté más separado (como 10-56 o 2-64) también representará una desviación significativa de la hipótesis nula. Por el contrario, repartos como 12-52 no representan desviaciones significativas de un reparto de 33-33, y no justificarían el rechazo de la hipótesis nula. La prueba de rangos con signo de Wilcoxon está basada en el total del rango más bajo, por lo que en vez de analizar los dos números que constituyen el reparto, consideramos sólo el número más bajo.

La suma $1 + 2 + 3 + \cdots + n$ de todos los rangos es igual a $n(n + 1)/2$, y si ésta es una suma de rangos a dividirse por igual entre dos categorías (positivo y negativo), cada uno de los dos totales debería estar cerca de $n(n + 1)/4$, que es la mitad de $n(n + 1)/2$. El reconocimiento de este principio nos ayuda a entender el estadístico de prueba que se usa cuando $n > 30$. El denominador en esa expresión representa una desviación estándar de T y se basa en el principio de que

$$1^2 + 2^2 + 3^3 + \cdots + n^2 = \frac{n(n + 1)(2n + 1)}{6}$$

La prueba de rangos con signo de Wilcoxon puede utilizarse sólo para datos apareados. La siguiente sección describirá una prueba de suma de rangos que puede aplicarse a dos conjuntos de datos independientes que no están asociados en pares.

13-4 Prueba de la suma de rangos de Wilcoxon para dos muestras independientes

Concepto clave Esta sección describe l*a prueba de la suma de rangos de Wilcoxon*, que utiliza los rangos de los valores de dos conjuntos independientes de datos muestrales para probar la hipótesis nula de que las dos poblaciones tienen medianas iguales.

La prueba de rangos con signo de Wilcoxon (sección 13-3) implica datos apareados, pero la prueba de suma de rangos de Wilcoxon de esta sección implica dos muestras independientes que no están relacionadas ni asociadas o apareadas. (Para evitar confusiones entre la prueba de suma de rangos de Wilcoxon para muestras independientes y la prueba de rangos con signo de Wilcoxon para datos apareados, considere el uso de las siglas ISR del Impuesto Sobre la Renta como recurso mnemotécnico para recordarnos "independiente: suma de rangos").

Definición

La **prueba de la suma de rangos de Wilcoxon** es una prueba no paramétrica que utiliza rangos de datos muestrales de dos poblaciones independientes. Se utiliza para probar la hipótesis nula de que las dos muestras independientes provienen de poblaciones con medianas iguales. La hipótesis alternativa es la aseveración de que las dos poblaciones tienen medianas diferentes.

H_0: Las dos muestras provienen de poblaciones con medianas iguales.

H_1: Las dos muestras provienen de poblaciones con medianas diferentes.

Concepto básico: La prueba de la suma de rangos de Wilcoxon es equivalente a la **prueba *U* de Mann-Whitney** (véase el ejercicio 13), que se incluye en algunos otros libros de texto y programas de cómputo (como el Minitab). La idea fundamental que subyace en la prueba de la suma de rangos de Wilcoxon es la siguiente: si dos muestras se obtienen de poblaciones idénticas y los valores individuales se acomodan en rangos como un conjunto combinado de valores, entonces el rango alto y el bajo deberían caer de manera uniforme entre las dos muestras. Si los rangos bajos se encuentran predominantemente en una muestra y los rangos altos se encuentran predominantemente en la otra muestra, sospechamos que las dos poblaciones tienen medianas diferentes.

Prueba de la suma de rangos de Wilcoxon

Requisitos

1. Hay dos muestras independientes de datos seleccionados al azar.
2. Cada una de las dos muestras tiene más de 10 valores. (Para muestras con 10 valores o menos, en libros de referencia están disponibles tablas especiales, como las *CRC Standard Probability and Statistics Tables and Formulae*, publicadas por CRC Press).
3. *No* existe el requisito de que las dos poblaciones tengan una distribución normal o cualquier otra distribución particular.

Notación

n_1 = tamaño de la muestra 1

n_2 = tamaño de la muestra 2

R_1 = suma de rangos de la muestra 1, que se calcula utilizando el procedimiento que se describe a continuación

R_2 = suma de rangos de la muestra 2, que se calcula utilizando el procedimiento que se describe a continuación

R = lo mismo que R_1 (suma de rangos de la muestra 1)

μ_R = media de los valores muestrales R que se espera cuando las dos poblaciones tienen medianas iguales

σ_R = desviación estándar de los valores muestrales R que se espera cuando las dos poblaciones tienen medianas iguales

Estadístico de prueba

$$z = \frac{R - \mu_R}{\sigma_R}$$

donde

$$\mu_R = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma_R = \sqrt{\frac{n_1 n_2(n_1 + n_2 + 1)}{12}}$$

n_1 = tamaño de la muestra a partir de la cual se calcula la suma de rangos R

n_2 = tamaño de la otra muestra

R = suma de rangos de la muestra con tamaño n_1

Valores críticos: Los valores críticos pueden encontrarse en la tabla A-2 (puesto que el estadístico de prueba está basado en la distribución normal).

Procedimiento para calcular el valor del estadístico de prueba

1. Combine temporalmente las dos muestras en una muestra grande, entonces reemplace cada valor muestral por su rango. (El valor más bajo toma un rango de 1, el siguiente valor más bajo toma un rango de 2, etcétera. Si los valores están empatados, asígneles la media de los rangos implicados en el empate. Consulte la sección 13-1 para ver una descripción de los rangos y el procedimiento para manejar empates).
2. Calcule la suma de los rangos de las dos muestras.

3. Calcule el valor del estadístico de prueba z como se indicó en el recuadro anterior, donde cualquier muestra puede utilizarse como la "muestra 1". (Si ambos tamaños muestrales son mayores que 10, entonces la distribución muestral de R es aproximadamente normal, con media μ_R y desviación estándar σ_R, y el estadístico de prueba es como se mostró en el recuadro anterior).

Note que, a diferencia de las pruebas de hipótesis correspondientes en la sección 9-3, la prueba de la suma de rangos de Wilcoxon *no* requiere poblaciones distribuidas normalmente. Además, la prueba de la suma de rangos de Wilcoxon puede utilizarse con datos en el nivel de medición ordinal, como los datos que consisten en rangos. En contraste, los métodos paramétricos de la sección 9-3 no se pueden utilizar con datos en el nivel de medición ordinal. En la tabla 13-2 notamos que la prueba de la suma de rangos de Wilcoxon tiene una tasa de eficiencia de 0.95 en comparación con la prueba paramétrica t o con la prueba z. Puesto que esta prueba tiene una alta tasa de eficiencia y requiere cálculos más fáciles, suele preferirse sobre las pruebas paramétricas descritas en la sección 9-3, aun cuando se satisfaga el requisito de normalidad.

TEXTO N° 16

Experimentos multinomiales: Bondad de ajuste

Tablas de contingencia: Independencia y homogeneidad

Compilado y adaptado de Mario Triola Estadística PEARSON; 2009

Págs. De 591 a 596, y 606 a 615

11-2 Experimentos multinomiales: Bondad de ajuste

Concepto clave Con datos separados en diferentes categorías, someteremos a prueba la hipótesis de que la distribución de los datos coincide con alguna distribución aseverada (es decir, "se ajusta" a ella). La prueba de hipótesis empleará la distribución chi cuadrada con los conteos de frecuencias observados y los conteos de frecuencias que se esperarían con la distribución aseverada. El estadístico de prueba chi cuadrada es una medida de la discrepancia entre las frecuencias observadas y las esperadas.

Comenzamos con la definición de un *experimento multinomial*, que es muy similar a la definición de un experimento binomial que se estudió en la sección 5-3, excepto que un experimento multinomial tiene más de dos categorías (a diferencia de un experimento binomial, que tiene exactamente dos categorías).

Definición

Un **experimento multinomial** es un experimento que satisface las siguientes condiciones:

1. El número de ensayos es fijo.
2. Los ensayos son independientes.
3. Todos los resultados de cada ensayo deben clasificarse exactamente en una de varias categorías diferentes.
4. Las probabilidades para las diferentes categorías permanecen constantes en cada ensayo.

EJEMPLO **Últimos dígitos de pesos** Miles de sujetos se estudian de manera rutinaria como parte del National Health Examination Survey. Los procedimientos de examen son muy exactos. Por ejemplo, cuando se obtienen pesos de sujetos, es sumamente importante pesarlos de verdad en vez de pedirles que reporten su peso. Se sabe que cuando la gente reporta su peso, generalmente da un peso más bajo que el real. Entonces, ¿cómo pueden verificar los investigadores que los pesos se obtuvieron por medio de mediciones reales y no por el reporte de los sujetos? Un método consiste en analizar los *últimos dígitos* de los pesos. Cuando la gente reporta su peso, tiende a redondear la cifra, a menudo *hacia* el entero inferior. Los últimos dígitos de los pesos reportados suelen tener un número desproporcionado de ceros y cincos, en comparación con los últimos dígitos de los pesos obtenidos a través de un proceso de medición. En contraste, cuando realmente se pesa a las personas, los últimos dígitos tienden a distribuirse de manera uniforme, de modo que 0, 1, 2,..., 9 se presentan aproximadamente con la misma frecuencia. El autor obtuvo los pesos de 80 estudiantes elegidos al azar, cuyos últimos dígitos se resumen en la tabla 11-2. Más adelante analizaremos los datos; por ahora, simplemente verificaremos que se satisfagan las cuatro condiciones de un experimento multinomial.

continúa

Tabla 11-2
Últimos dígitos de pesos

Último dígito	Frecuencia
0	35
1	0
2	2
3	1
4	4
5	24
6	1
7	4
8	7
9	2

SOLUCIÓN He aquí la verificación de que se satisfacen las cuatro condiciones del experimento multinomial:

1. El número de ensayos (últimos dígitos) es el número fijo 80.
2. Los ensayos son independientes, puesto que el último dígito de cualquier peso individual no afecta al último dígito de cualquier otro peso.
3. Cada resultado (último dígito) se clasifica exactamente en una de 10 categorías diferentes. Las categorías se identifican como 0, 1, 2,..., 9.
4. Al poner a prueba la aseveración de que los 10 dígitos son igualmente probables, cada dígito posible tiene la probabilidad de 1/10, y se supone que esa probabilidad permanece constante para cada sujeto.

En esta sección presentamos un método para probar la aseveración de que en un experimento multinomial, las frecuencias observadas en las diferentes categorías se ajustan a una distribución en particular. Puesto que hacemos una prueba de qué tan bien se ajusta una frecuencia de distribución observada a alguna distribución teórica especificada, este método suele llamarse *prueba de bondad de ajuste*.

Definición

La **prueba de bondad de ajuste** se utiliza para probar la hipótesis de que una distribución de frecuencias se ajusta a (o coincide con) alguna distribución aseverada.

Por ejemplo, utilizando los datos de la tabla 11-2, podemos probar la hipótesis de que los datos se ajustan a una distribución uniforme, con igual probabilidad para todos los dígitos. Nuestras pruebas de bondad de ajuste incorporarán la siguiente notación.

Notación

O representa la *frecuencia observada* de un resultado.
E representa la *frecuencia esperada* de un resultado.
k representa el número de *categorías diferentes* o resultados.
n representa el *número total de ensayos*.

Cálculo de frecuencias esperadas

En la tabla 11-2 las frecuencias observadas O se denotan por 35, 0, 2, 1, 4, 24, 1, 4, 7 y 2. La suma de las frecuencias observadas es 80, de manera que $n = 80$. Si suponemos que los 80 dígitos se obtuvieron de una población en la que todos los dígitos son igualmente probables, entonces *esperamos* que cada dígito se presente en 1/10 de los 80 ensayos, de manera que cada una de las 10 frecuencias esperadas está dada por $E = 8$. Si generalizamos este resultado, obtenemos un procedimiento sencillo para calcular las frecuencias esperadas, siempre y cuando supongamos que todas las frecuencias esperadas son iguales: simplemente divida el número total de observaciones entre el número de categorías diferentes ($E = n/k$). En otros casos en los que no todas las frecuencias esperadas son iguales, a menudo podemos calcular las frecuencias esperadas para cada categoría multiplicando la suma de todas las frecuencias observadas por la probabilidad p de la categoría, de manera que $E = np$. Aquí resumimos estos dos procedimientos.

- **Si todas las frecuencias esperadas son iguales, entonces cada frecuencia esperada es la suma de todas las frecuencias observadas dividida entre el número de categorías, de manera que $E = n/k$.**
- **Si las frecuencias esperadas no son todas iguales, entonces cada frecuencia esperada se calcula multiplicando la suma de todas las frecuencias observadas por la probabilidad para la categoría, de manera que $E = np$ para cada categoría.**

Aun cuando estas dos fórmulas para E pueden ser muy buenas, sería mejor utilizar un método informal basado en la comprensión de las circunstancias. Sólo pregúntese: "¿Cómo se pueden repartir las frecuencias observadas entre las diferentes categorías, de manera que exista una coincidencia perfecta con la distribución aseverada?" Además, reconozca que todas las frecuencias *observadas* deben ser números enteros, puesto que representan conteos reales, en tanto que las frecuencias *esperadas* no requieren ser números enteros. Por ejemplo, cuando se tira un dado 33 veces, la frecuencia esperada para cada posible resultado es $33/6 = 5.5$. Se espera que el número 3 se presente con una frecuencia de 5.5, aunque es imposible obtener el resultado de que el 3 se presente exactamente 5.5 veces.

Sabemos que las frecuencias muestrales por lo regular se desvían un poco de los valores que esperamos teóricamente, y ahora planteamos la pregunta clave: ¿Son estadísticamente significativas las diferencias entre los valores *observados* O reales y los valores teóricos *esperados* E? Necesitamos una medida de la discrepancia entre los valores O y E, así que utilizamos el estadístico de prueba dado con los supuestos y los valores críticos. (Más adelante explicaremos cómo se desarrolló este estadístico de prueba, pero usted puede ver que incluye diferencias de $O - E$ como componente clave).

Requisitos

1. Los datos se seleccionaron al azar.
2. Los datos muestrales consisten en conteos de frecuencias para cada una de las diferentes categorías.
3. Para cada categoría, la frecuencia *esperada* es al menos de 5. (La frecuencia esperada para una categoría es la frecuencia que ocurriría si los datos realmente tuvieran la distribución que se asevera. No existe ningún requisito de que la frecuencia *observada* para cada categoría deba ser al menos de 5).

Estadístico de prueba para pruebas de bondad de ajuste en experimentos multinomiales

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Valores críticos

1. Los valores críticos se encuentran en la tabla A-4, utilizando $k - 1$ grados de libertad, donde k = número de categorías.
2. Las pruebas de hipótesis por bondad de ajuste siempre son de *cola derecha*.

LA ESTADÍSTICA EN LAS NOTICIAS

Los asientos de avión más seguros

Muchos de nosotros creemos que, en un accidente aéreo, los asientos de la parte de atrás son los más seguros. Los expertos en seguridad no están de acuerdo en que alguna parte específica de un avión sea más segura que las otras. Algunos aviones chocan primero con la nariz cuando caen, pero otros chocan con la cola al despegar. Matt McCormick, un experto en supervivencia del National Transportation Safety Board, dijo a la revista *Travel* que "no existe ningún lugar seguro para sentarse". Se pueden utilizar pruebas de bondad de ajuste con la hipótesis nula de que todas las secciones de un avión son igualmente seguras. Los aviones accidentados podrían dividirse en las secciones frontal, media y posterior. Entonces, las frecuencias observadas de decesos podrían compararse con las frecuencias que se esperarían con una distribución de decesos uniforme. El estadístico de prueba χ^2 refleja el tamaño de las discrepancias entre las frecuencias observadas y las esperadas, y revelaría si algunas secciones son más seguras que las otras.

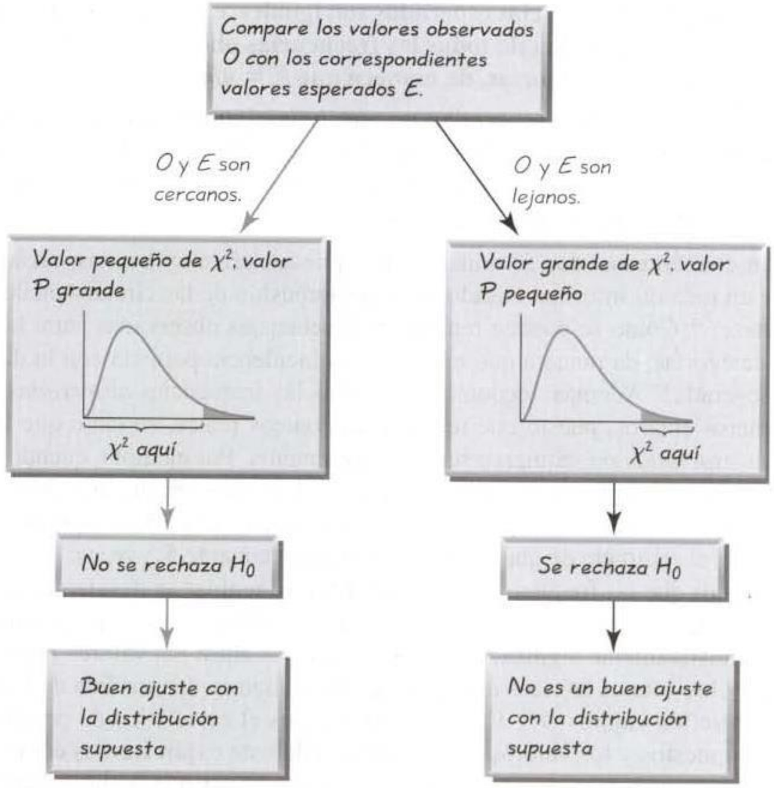


Figura 11-3 Relaciones entre el estadístico de prueba χ^2, el valor P y la bondad de ajuste

El estadístico de prueba χ^2 se basa en las diferencias entre valores observados y esperados, de manera que una *concordancia cercana* entre los valores observados y esperados conducirá a un valor de χ^2 *pequeño* y un valor *P grande*. Una discrepancia grande entre los valores observados y esperados conducirá a un valor de χ^2 *grande* y un valor *P pequeño*. De esta forma, las pruebas de hipótesis de esta sección siempre son de cola derecha, puesto que el valor crítico y la región crítica se localizan en el extremo derecho de la distribución. Estas relaciones se resumen e ilustran en la figura 11-3.

Una vez que sabemos calcular el valor del estadístico de prueba y el valor crítico, podemos probar hipótesis utilizando el mismo procedimiento general que se describió en el capítulo 8.

EJEMPLO Análisis de los últimos dígitos de pesos: frecuencias iguales esperadas Consulte los últimos dígitos de los 80 pesos en la tabla 11-2. Ponga a prueba la aseveración de que los dígitos *no* se presentan con la misma frecuencia. Con base en los resultados, ¿qué concluye acerca del procedimiento usado para obtener los pesos?

SOLUCIÓN

REQUISITO ✔ Es necesario que los datos muestrales se seleccionen al azar, que consistan en conteos de frecuencias, que provengan de un experimento multinomial y que cada frecuencia esperada sea al menos de 5. Antes señalamos que los datos provienen de estudiantes elegidos al azar. Los datos consisten en conteos de frecuencias. En el ejemplo anterior se estableció que las condiciones para un experimento multinomial se satisfacen. El análisis previo de los valores esperados incluyó el resultado de que cada frecuencia esperada es 8, de manera que cada frecuencia esperada sí satisface el requisito de ser al menos de 5. Todos los requisitos se satisfacen y podemos proceder con la prueba de hipótesis. ✔

La aseveración de que los dígitos no ocurren con la misma frecuencia es equivalente a la aseveración de que las frecuencias relativas o probabilidades de las 10 celdas ($p_0, p_1, \ldots, p_9$) no son todas iguales. Aplicaremos el método tradicional para prueba de hipótesis (véase la figura 8-9).

Paso 1: La aseveración original es que los dígitos no se presentan con la misma frecuencia. Es decir, al menos una de las probabilidades $p_0, p_1, \ldots, p_9$ es diferente de las otras.

Paso 2: Si la aseveración original es falsa, entonces todas las probabilidades son las mismas. Es decir, $p_0 = p_1 = \cdots = p_9$.

Paso 3: La hipótesis nula debe contener la condición de igualdad, así que tenemos

$$H_0: \quad p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$$

H_1: Al menos una de las probabilidades es diferente de las otras.

Paso 4: No se especificó un nivel de significancia, así que seleccionamos $\alpha = 0.05$, una elección muy común.

Paso 5: Puesto que probamos la aseveración de que la distribución de los últimos dígitos es una distribución uniforme, utilizamos la prueba de bondad de ajuste descrita en esta sección. Se utiliza la distribución χ^2 con el estadístico de prueba que se dio al principio.

Paso 6: Las frecuencias observadas O se listan en la tabla 11-2. Cada frecuencia esperada E correspondiente es igual a 8 (porque los 80 dígitos estarían distribuidos de manera uniforme a lo largo de las 10 categorías). La tabla 11-3 muestra el cálculo del estadístico de prueba χ^2. El estadístico de prueba es $\chi^2 = 156.500$. El valor crítico es $\chi^2 = 16.919$ (obtenido de la tabla A−4 con $\alpha = 0.05$ en la cola derecha y con grados de libertad iguales a $k - 1 = 9$). El estadístico de prueba y el valor crítico se muestran en la figura 11-4.

Paso 7: Puesto que el estadístico de prueba cae dentro de la región crítica, existe evidencia suficiente para rechazar la hipótesis nula.

Paso 8: Existe suficiente evidencia para sustentar la aseveración de que los últimos dígitos no se presentan con la misma frecuencia relativa. Ahora tenemos evidencia muy fuerte que sugiere que los pesos realmente no se midieron. Es razonable especular que se trata de valores reportados y no de mediciones reales.

El ejemplo anterior incluyó la hipótesis nula de que las probabilidades para las diferentes categorías son todas iguales. Los métodos de esta sección también pueden utilizarse cuando las probabilidades (o frecuencias) hipotéticas son diferentes, como en el ejemplo siguiente.

Tabla 11-3 Cálculo del estadístico de prueba χ^2 para los últimos dígitos de pesos

Último dígito	Frecuencia observada O	Frecuencia esperada E	$O - E$	$(O - E)^2$	$\frac{(O-E)^2}{E}$
0	35	8	27	729	91.1250
1	0	8	−8	64	8.0000
2	2	8	−6	36	4.500
3	1	8	−7	49	6.125
4	4	8	−4	16	2.000
5	24	8	16	256	32.000
6	1	8	−7	49	6.125
7	4	8	−4	16	2.000
8	7	8	−1	1	0.125
9	2	8	−6	36	4.500
	80	80			$\chi^2 = \sum \frac{(O-E)^2}{E} = 156.500$

(Con excepción de los errores por redondeo, estos dos totales deben coincidir)

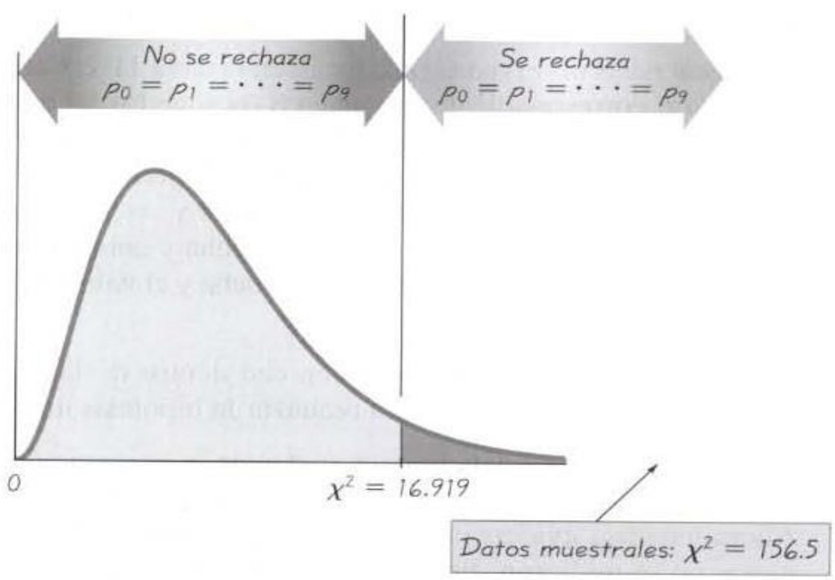


Figura 11-4 **Prueba de** $p_0 = p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = p_7 = p_8 = p_9$

11-3 Tablas de contingencia: Independencia y homogeneidad

Concepto clave En esta sección estudiaremos *tablas de contingencia* (o *tablas de frecuencias de dos factores*), que incluyen conteos de frecuencia para datos categóricos ordenados en una tabla, con al menos dos renglones y al menos dos columnas. Presentamos un método para poner a prueba la aseveración de que las variables de renglón y de columnas son independientes unas de otras. Utilizaremos el mismo método para una prueba de homogeneidad, en la que ponemos a prueba la aseveración de que distintas poblaciones tienen la misma proporción de algunas características.

Comenzamos con la definición de una tabla de contingencia.

Definición

Una **tabla de contingencia** (o **tabla de frecuencias de dos factores**) es una tabla en la que las frecuencias corresponden a dos variables. (Una variable se utiliza para categorizar renglones, y una segunda variable se utiliza para categorizar columnas).

La tabla 11-4 es un ejemplo de una tabla de contingencia con dos renglones y tres columnas, y los datos en las celdas son conteos de frecuencias. Los datos de la tabla 11-4 corresponden a un estudio retrospectivo (o de casos y controles). La variable de renglón tiene dos categorías: controles y casos. Los sujetos del grupo de control eran motociclistas elegidos al azar en ciertos lugares de la carretera. Los sujetos del grupo de casos eran motociclistas gravemente heridos o fallecidos. La variable de columna se utiliza para el color del casco que usaban. La pregunta importante es la siguiente: ¿El color del casco del motociclista se relaciona de alguna forma con el riesgo de lesiones relacionadas con choques? (Los datos se basan en "Motorcycle Rider Conspicuity and Crash Related Injury: Case-Control Study", de Wells *et al.*, *BMJ USA*, vol. 4).

Esta sección presenta dos tipos de prueba de hipótesis basadas en tablas de contingencia. Primero consideramos las pruebas de independencia, las cuales se usan

Tabla 11-4 Estudio de casos y controles de conductores de motocicleta

	Color del caso		
	Negro	Blanco	Amarillo/anaranjado
Controles (sin lesiones)	491	377	31
Casos (lesionados o fallecidos)	213	112	8

para determinar si una variable de renglón de una tabla de contingencia es independiente de su variable de columna. Luego, consideramos las pruebas de homogeneidad, que se utilizan para determinar si situaciones diferentes tienen las mismas proporciones de alguna característica. Ambos tipos de pruebas de hipótesis utilizan los *mismos* métodos básicos. Comenzamos con las pruebas de independencia.

Prueba de independencia

Una de las dos pruebas incluidas en esta sección es la *prueba de independencia* entre la variable de renglón y la variable de columna.

Definición

Una **prueba de independencia** pone a prueba la hipótesis nula de que no existe asociación entre la variable de renglón y la variable de columna en una tabla de contingencia. (Para la hipótesis nula, utilizaremos la afirmación de que "las variables de renglón y de columna son independientes").

Es muy importante reconocer que, en este contexto, la palabra *contingencia* se refiere a dependencia, pero sólo se trata de una dependencia estadística, y no puede utilizarse para establecer un vínculo directo de causa-efecto entre las dos variables en cuestión. Cuando se prueba la hipótesis nula de independencia entre las variables de renglón y de columna en una tabla de contingencia, los requisitos, el estadístico de prueba y los valores críticos son como se describe en el siguiente cuadro.

Requisitos

1. Los datos muestrales son seleccionados al azar y se representan como conteos de frecuencias en una tabla de dos factores.
2. La hipótesis nula H_0 es la afirmación de que las variables de renglón y de columna son *independientes*; la hipótesis alternativa H_1 es la afirmación de que las variables de renglón y de columna son dependientes.
3. Para cada celda de la tabla de contingencia, la frecuencia *esperada E* es al menos de 5. (No existe el requisito de que cada frecuencia *observada* deba ser al menos de 5. Además, no existe el requisito de que la población deba tener una distribución normal o cualquier otra distribución específica).

Estadístico de prueba para una prueba de independencia

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Valores críticos

1. Los valores críticos se encuentran en la tabla A-4, utilizando

   $$\textbf{grados de libertad} = (r - 1)(c - 1)$$

   donde r es el número de renglones y c es el número de columnas.
2. En una prueba de independencia de una tabla de contingencia, la región crítica se localiza *sólo en la cola derecha*.

El estadístico de prueba nos permite medir el grado de discordancia entre las frecuencias observadas en la realidad y aquellas que se esperarían teóricamente cuando las dos variables son independientes. Los valores grandes del estadístico de prueba χ^2 están en la región de la extrema derecha de la distribución chi cuadrada y reflejan diferencias significativas entre las frecuencias observadas y esperadas. En muestreos grandes repetidos, la distribución del estadístico de prueba χ^2 puede aproximarse por medio de la distribución chi cuadrada, siempre y cuando todas las frecuencias esperadas sean al menos de 5. El número de grados de libertad $(r - 1)(c - 1)$ refleja el hecho de que, puesto que conocemos el total de las frecuencias en una tabla de contingencia, podemos asignar con libertad frecuencias a sólo $r - 1$ renglones y $c - 1$ columnas antes de que se determine la frecuencia para cada celda. [Sin embargo, no podemos tener frecuencias negativas o frecuencias tan grandes que la suma de cualquier renglón (o columna) exceda al total de las frecuencias observadas para ese renglón (o columna)].

La frecuencia esperada E para cada celda puede calcularse multiplicando el total de las frecuencias de renglón por el total de las frecuencias de columna, y dividiendo el producto entre el gran total de todas las frecuencias, como se ilustra a continuación.

Frecuencia esperada para una tabla de contingencia

$$\text{frecuencia esperada} = \frac{(\text{total de renglón})(\text{total de columna})}{(\text{gran total})}$$

EJEMPLO Cálculo de la frecuencia esperada Remítase a la tabla 11-4 y calcule la frecuencia esperada para la primera celda, cuando la frecuencia es 491.

SOLUCIÓN La primera celda se ubica en el primer renglón (con un total de 899) y la primera columna (con un total de 704), y la suma de todas las frecuencias en la tabla es 1232.

$$E = \frac{(\text{total por renglón})(\text{total por columna})}{(\text{gran total})} = \frac{(899)(704)}{1232} = 513.714$$

INTERPRETACIÓN Al interpretar este resultado para la primera celda, podemos decir que, aunque 491 motociclistas del grupo de control usaron cascos negros, habríamos esperado que 513.714 de ellos usaran cascos negros si el grupo (de controles o de casos) es independiente del color del casco usado. Existe una discrepancia entre $O = 491$ y $E = 513.714$, y este tipo de discrepancias son componentes fundamentales del estadístico de prueba.

Para comprender mejor las frecuencias esperadas, suponga que conocemos sólo los totales del renglón y de la columna, como en la tabla 11-5, y que debemos llenar la celda de las frecuencias esperadas suponiendo independencia (o ausencia de relación) entre las variables de renglón y de columna. En el primer renglón, 899 de los 1232 sujetos están en el grupo de control, de manera que $P(\text{grupo de control}) = 899/1232$. En la primera columna, 704 de los 1232 conductores usaron cascos negros, de manera que $P(\text{casco negro}) = 704/1232$. Puesto que estamos suponiendo independencia entre el grupo y el color del casco, la regla de la multiplicación para sucesos independientes $[P(A \text{ y } B) = P(A) \cdot P(B)]$ se expresa como

$$P(\text{grupo de control y casco negro}) = P(\text{grupo de control}) \cdot P(\text{casco negro})$$

$$= \frac{899}{1232} \cdot \frac{704}{1232}$$

Tabla 11-5 Estudio de casos y controles de conductores de motocicleta

	Color del caso			Totales de renglón
	Negro	Blanco	Amarillo/anaranjado	
Controles				899
Casos				333
Totales de columna	**704**	**489**	**39**	**Gran total: 1232**

Conociendo la probabilidad de estar en la celda superior izquierda, ahora podemos calcular el *valor esperado* para esa celda, el cual obtenemos multiplicando la probabilidad para esa celda por el número total de sujetos, como en la siguiente ecuación:

$$E = n \cdot p = 1232\left[\frac{899}{1232} \cdot \frac{704}{1232}\right] = 513.714$$

La forma de este producto sugiere una manera general para obtener la frecuencia esperada de una celda:

$$\text{Frecuencia esperada } E = (\text{gran total}) \cdot \frac{(\text{total de renglón})}{(\text{gran total})} \cdot \frac{(\text{total de columna})}{(\text{gran total})}$$

Esta expresión se simplifica para obtener

$$E = \frac{(\text{total de renglón}) \cdot (\text{total de columna})}{(\text{gran total})}$$

Como ya sabemos calcular los valores esperados, ahora procedemos a utilizar los datos de la tabla de contingencia para probar las hipótesis, como en el siguiente ejemplo.

EJEMPLO Lesiones y color del casco para motocicleta Remítase a los datos de la tabla 11-4. Utilice un nivel de significancia de 0.05 para probar la aseveración de que el grupo (de control o de casos) es independiente del color del casco.

SOLUCIÓN

REQUISITO ✔ Como se requiere: los datos se seleccionaron al azar; los datos consisten en conteos de frecuencias en una tabla de dos factores; se está poniendo a prueba la hipótesis nula de que las variables son independientes y las frecuencias esperadas son al menos de 5. (Las frecuencias esperadas son 513.714, 356.827, 28.459, 190.286, 132.173, 10.541). Puesto que todos los requisitos se satisfacen, procedemos a la prueba de hipótesis. ✔

La hipótesis nula y la hipótesis alternativa son las siguientes:

H_0: El hecho de que el sujeto pertenezca al grupo de control o al grupo de casos es independiente del color del casco. (Esto equivale a decir que las lesiones son independientes del color del casco).

H_1: El grupo y el color del casco son dependientes.

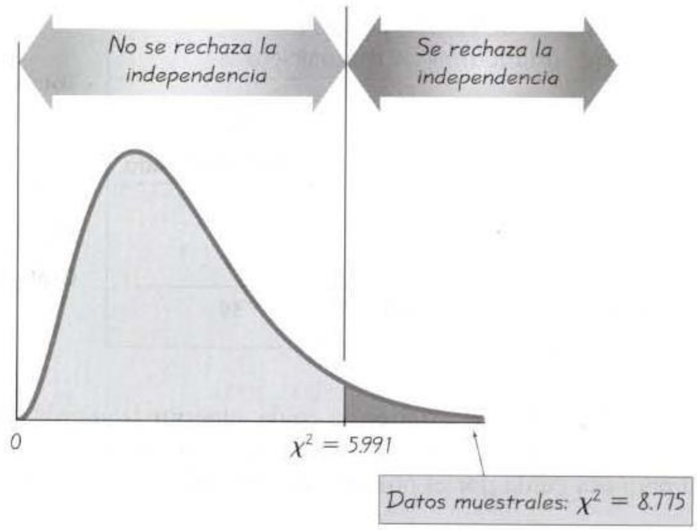
El nivel de significancia es $\alpha = 0.05$.

continúa

Un falso positivo durante ocho años

La Associated Press publicó recientemente una nota informativa acerca de Jim Malone, quien había recibido un resultado de prueba positivo para una infección de VIH. Durante ocho años Jim acudió a reuniones de grupos de apoyo, luchó contra la depresión y perdió peso mientras temía morir de SIDA. Finalmente, se le informó que la prueba original era errónea: él no estaba infectado con VIH. Se realizó una prueba de seguimiento después del primer resultado positivo, y la prueba de confirmación demostró que no tenía una infección de VIH, pero nadie le informó al señor Malone el nuevo resultado. Jim Malone sufrió durante ocho años por el resultado de una prueba que en realidad era un falso positivo.

Figura 11-7
Prueba de independencia para los datos de motociclistas



Puesto que los datos se presentan en una tabla de contingencia, utilizamos la distribución χ^2 con este estadístico de prueba:

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(491 - 513.714)^2}{513.714} + \cdots + \frac{(8 - 10.541)^2}{10.541}$$
$$= 8.775$$

El valor crítico es $\chi^2 = 5.991$ y se encuentra en la tabla A-4, observando que $\alpha = 0.05$ en la cola derecha y que el número de grados de libertad está dado por $(r-1)(c-1) = (2-1)(3-1) = 2$. El estadístico de prueba y el valor crítico se muestran en la figura 11-7. Puesto que el estadístico de prueba está dentro de la región crítica, rechazamos la hipótesis nula de independencia entre el grupo y el color del casco. Al parecer, el color del casco y el grupo (de control o de casos) son dependientes. Puesto que los controles no estaban lesionados y los casos tenían lesiones o habían fallecido, parece que existe una relación entre el color del casco y la seguridad en la motocicleta. Los autores del artículo científico afirmaron que el estudio sustenta la promulgación de leyes que exijan que los motociclistas sean más visibles.

Valores *P*

En el ejemplo anterior se utilizó el método tradicional de prueba de hipótesis, pero podemos utilizar con facilidad el método del valor *P*. STATDISK, Minitab, Excel y la calculadora TI-83/84 Plus proporcionan valores *P* para pruebas de independencia de tablas de contingencia. Si usted no tiene una calculadora o un programa de cómputo adecuados, puede estimar los valores *P* con la tabla A-4, localizando el estadístico de prueba en el renglón correspondiente al número apropiado de grados de libertad. En el ejemplo anterior, consulte el renglón para 2 grados de libertad y observe que el estadístico de prueba de 8.775 se ubica entre 7.378 y 9.210. Por lo tanto, el valor *P* debe estar entre 0.025 y 0.01, de manera que concluimos que $0.01 < P < 0.025$. (El valor *P* real es 0.0124). Sabiendo que el valor *P* es menor que el nivel de significancia de 0.05, rechazamos la hipótesis nula como lo hicimos en el ejemplo anterior.

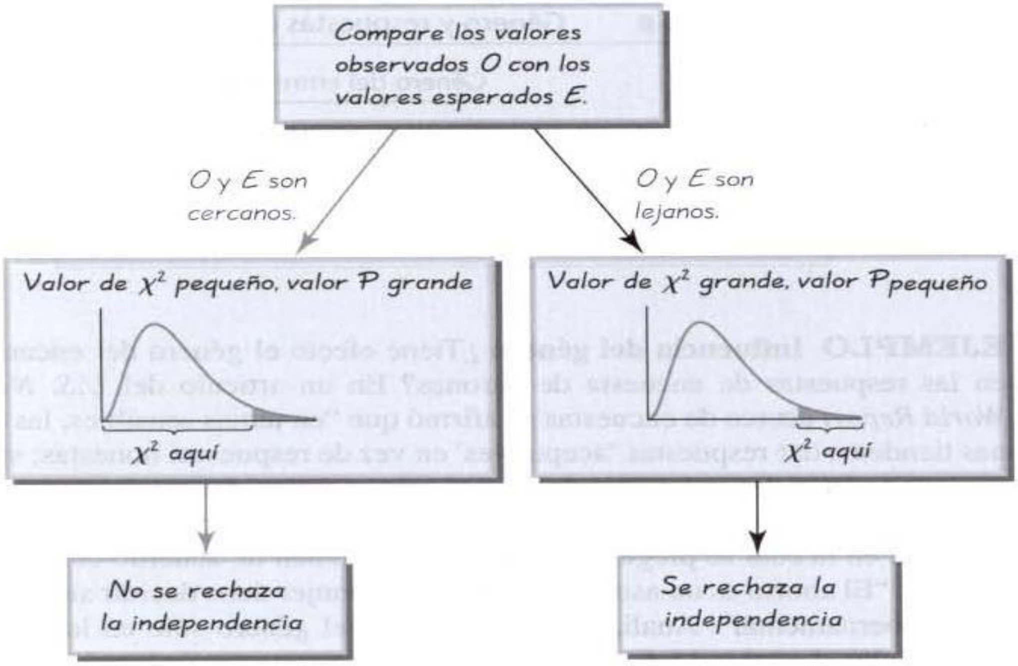


Figura 11-8 Relaciones entre componentes clave en la prueba de independencia

Algual que en la sección 11-2, si las frecuencias observada y esperada son cercanas, el estadístico de prueba χ^2 será pequeño y el valor P será grande. Si las frecuencias observada y esperada se alejan mucho, el estadístico de prueba χ^2 será grande y el valor P será pequeño. Estas relaciones se resumen e ilustran en la figura 11-8.

Prueba de homogeneidad

En el ejemplo anterior, ilustramos una prueba de independencia entre dos variables y utilizamos una población de motociclistas. Sin embargo, algunas otras muestras se obtienen de poblaciones *diferentes*, y tal vez deseemos determinar si esas poblaciones tienen las mismas proporciones de las características en consideración. En estos casos se utiliza la *prueba de homogeneidad*. (La palabra *homogéneo* significa "que tiene la misma calidad" y, en este contexto, estamos haciendo una prueba para determinar si las proporciones son las mismas).

Definición

En una **prueba de homogeneidad** probamos la aseveración de que *poblaciones diferentes* tienen las mismas proporciones de algunas características.

Al hacer una prueba de homogeneidad, podemos utilizar los mismos requisitos, estadístico de prueba, valor crítico y procedimientos descritos en esta sección, con una excepción: en vez de probar la hipótesis nula de independencia entre las variables de renglón y de columna, probamos la hipótesis nula de que las poblaciones diferentes tienen las mismas proporciones de algunas características.

Ventaja del equipo local

En un artículo de la revista *Chance* titulado "Predicting Professional Sports Game Outcomes from Intermediate Game Scores", los autores Harris Cooper, Kristina DeNeve y Frederick Mosteller utilizaron la estadística para analizar dos creencias comunes: **1.** Los equipos tienen una ventaja cuando juegan en casa y **2.** En realidad sólo cuenta el último cuarto de los partidos profesionales de básquetbol. Utilizando una muestra aleatoria de cientos de partidos, encontraron que, en los cuatro deportes más populares, el equipo local gana aproximadamente el 58.6% de los partidos. Además, los equipos de básquetbol que van ganando después de tres cuartos del juego, ganan aproximadamente cuatro de cada cinco ocasiones, aunque los equipos de béisbol que van ganando después de 7 entradas ganan alrededor de 19 de cada 20 ocasiones. Los métodos de análisis estadístico incluyeron la distribución chi cuadrada aplicada a una tabla de contingencia.

Tabla 11-6 Género y respuestas de encuesta

	Género del entrevistador	
	Hombre	Mujer
Hombres que están de acuerdo	560	308
Hombres que están en desacuerdo	240	92

EJEMPLO Influencia del género ¿Tiene efecto el género del encuestador en las respuestas de encuesta de varones? En un artículo del *U.S. News & World Report* acerca de encuestas se afirmó que "en temas sensibles, las personas tienden a dar respuestas 'aceptables' en vez de respuestas honestas; sus respuestas podrían depender del género o el origen étnico del entrevistador". Para sustentar esta aseveración, el Eagleton Institute proporcionó los datos de una encuesta en la cual se preguntó a hombres si estaban de acuerdo con esta afirmación: "El aborto es un asunto privado que la mujer debe decidir sin intervención gubernamental". Analizaremos el efecto del género sólo en los hombres encuestados. La tabla 11-6 está basada en estas respuestas de hombres encuestados. Suponga que la encuesta se diseñó de manera que los entrevistadores varones recibieron instrucciones para obtener 800 respuestas de sujetos varones, y las entrevistadoras mujeres recibieron instrucciones para obtener 400 respuestas de sujetos varones. Utilice un nivel de significancia de 0.05 y pruebe la aseveración de que las proporciones de las respuestas de acuerdo/en desacuerdo son las mismas para los sujetos entrevistados por hombres y los sujetos entrevistados por mujeres.

SOLUCIÓN

REQUISITO ✔ Los datos consisten en conteos de frecuencias independientes; cada observación se puede categorizar de acuerdo con dos variables; y las frecuencias esperadas (que en la tabla de resultados de Minitab aparecen como 578.67, 289.33, 221.33 y 110.67) son al menos de 5. [Las dos variables son: **1.** género del entrevistador, y **2.** si el sujeto estuvo de acuerdo o en desacuerdo]. Puesto que se trata de una prueba de homogeneidad, ponemos a prueba la aseveración de que las proporciones de respuestas de acuerdo/en desacuerdo son iguales para los sujetos entrevistados por hombres y para los sujetos entrevistados por mujeres. Todos los requisitos se satisfacen, así que procedemos con la prueba de hipótesis. ✔

Puesto que tenemos dos poblaciones separadas (sujetos entrevistados por hombres y sujetos entrevistados por mujeres), probamos la homogeneidad con estas hipótesis:

H_0: Las proporciones de las respuestas acuerdo/en desacuerdo son iguales para los sujetos entrevistados por hombres y los sujetos entrevistados por mujeres.

H_1: Las proporciones son diferentes.

El nivel de significancia es $\alpha = 0.05$. Utilizamos el mismo estadístico de prueba χ^2 descrito antes, el cual calculamos por medio del mismo procedimiento. En vez de hacer una lista de los detalles de este cálculo, presentamos la pantalla de Minitab que resulta de los datos de la tabla 11-6.

Minitab

```
Expected counts are printed below observed counts
Chi-Square contributions are printed below expected counts

            C1      C2  Total
   1       560     308    868
        578.67  289.33
         0.602   1.204

   2       240      92    332
        221.33  110.67
         1.574   3.149

Total      800     400   1200

Chi-Sq = 6.529, DF = 1, P-Value = 0.011
```

La pantalla de Minitab indica las frecuencias esperadas de 578.67, 289.33, 221.33 y 110.67. Los resultados también incluyen el estadístico de prueba $\chi^2 = 6.529$ y el valor P de 0.011. Utilizando el método del valor P para probar la hipótesis, rechazamos la hipótesis nula de proporciones iguales (homogéneas) (puesto que el valor P de 0.011 es menor que 0.05). Existe suficiente evidencia para sustentar el rechazo de la aseveración de que las proporciones son iguales. Parece que la respuesta y el género del entrevistador son dependientes. Aunque este análisis estadístico no puede utilizarse para justificar ninguna afirmación acerca de la causalidad, parece que los hombres se ven influidos por el género del entrevistador.

El medio de encuesta puede afectar los resultados

En una encuesta de católicos en Boston, se preguntó a los sujetos si los anticonceptivos debían ponerse a la disposición de mujeres solteras. En entrevistas personales, el 44% de los encuestados dijo que sí. Sin embargo, en un grupo similar encuestado por correo o por teléfono, el 75% de los individuos respondió afirmativamente a la misma pregunta.

EJEMPLO Lanzamiento y giro de monedas de un centavo Cuando se lanza o se hace girar una moneda de un centavo de dólar, ¿existe la misma probabilidad de obtener caras? Utilice los datos de la tabla 11-7, con un nivel de significancia de 0.05, para probar la aseveración de que la proporción de caras es la misma al lanzar y al girar la moneda. (Los datos provienen de resultados experimentales presentados en *Chance News*).

TABLA 11-7 Experimento con monedas

	Caras	Cruces
Lanzamiento	2048	1992
Rotación	953	1047

SOLUCIÓN

REQUISITOS ✔ Como se requiere, los datos son aleatorios y consisten en conteos de frecuencia en una tabla de dos factores. Aquí estamos probando la hipótesis nula de que la proporción de caras con lanzamientos es igual a la proporción de caras con giros. Las frecuencias esperadas son al menos de 5. (Las frecuencias esperadas son 2007.291, 2032.709, 993.709 y 1006.291). Puesto que todos los requisitos se satisfacen, procedemos con la prueba de hipótesis. ✔

Puesto que tenemos dos poblaciones separadas (en un experimento se lanzaron monedas y en un experimento diferente se hicieron girar monedas), deseamos hacer una prueba de homogeneidad con las siguientes hipótesis:

H_0: La proporción de caras es igual con lanzamientos que con giros.

H_1: Las proporciones son diferentes.

El nivel de significancia es $\alpha = 0.05$. Utilizamos el mismo estadístico de prueba χ^2 descrito antes, el cual calculamos por medio del mismo procedimiento.

En vez de hacer una lista de los detalles de este cálculo, presentamos la pantalla de Minitab que resulta de los datos de la tabla 11-7.

Minitab

```
Chi-Square contributions are printed below expected counts

         Heads    Tails          Total
    1     2048     1992           4040
       2007.29  2032.71
         0.826    0.815

    2      953     1047           2000
        993.71  1006.29
         1.668    1.647

Total     3001     3039           6040

Chi-Sq = 4.955, DF = 1, P-Value = 0.026
```

La pantalla de Minitab indica las frecuencias esperadas de 2007.29, 2032.71, 993.71 y 1006.29. Los resultados incluyen también el estadístico de prueba $\chi^2 = 4.955$ y el valor P de 0.026. Utilizando el método del valor P para prueba de hipótesis, rechazamos la hipótesis nula de proporciones iguales (homogéneas) (puesto que el valor P de 0.026 es menor que 0.05). Existe suficiente evidencia para sustentar el rechazo de la aseveración de que las proporciones son iguales. Parece que el lanzamiento de una moneda de un centavo y el giro de una moneda de un centavo produce proporciones diferentes de caras.

Prueba exacta de Fisher

Para el análisis de tablas 2×2, incluimos el requisito de que cada celda debe tener una frecuencia esperada de 5 o más. Este requisito es necesario para que la distribución χ^2 sea una aproximación adecuada para la distribución exacta del estadístico de prueba $\Sigma\frac{(O - E)^2}{E}$. En consecuencia, si una tabla 2×2 tiene una celda con una frecuencia esperada menor que 5, los procedimientos anteriores no deben utilizarse, porque la distribución no es una aproximación adecuada. A menudo se utiliza la *prueba exacta de Fisher* para una tabla 2×2 de este tipo, ya que proporciona un valor P exacto y no requiere de una técnica de aproximación.

Considere los datos de la tabla 11-8, en la cual las frecuencias esperadas aparecen en paréntesis, debajo de las frecuencias observadas. La primera celda tiene una frecuencia esperada menor que 5, por lo que no debemos utilizar los métodos

Tabla 11-8 Cascos y lesiones faciales en accidentes de bicicleta (las frecuencias esperadas están en paréntesis)

	Con casco	Sin casco
Lesiones faciales recibidas	2	13
	(3)	(12)
Todas las lesiones no faciales	6	19
	(5)	(20)

anteriores. Con la prueba exacta de Fisher, calculamos la probabilidad de obtener por azar los resultados observados (suponiendo que el uso de un casco y ser víctima de lesiones faciales son sucesos independientes), y también calculamos la probabilidad de cualquier resultado que sea *más extremo*. (El hecho de referirnos a resultados "más extremos" podría generar confusión, por lo que sería útil repasar el apartado de la sección 5-2 "Uso de las probabilidades para determinar resultados infrecuentes"). Cuando se pone a prueba la hipótesis nula de independencia entre el uso de casco y ser víctima de lesiones faciales, las frecuencias de 2, 13, 6 y 19 pueden reemplazarse por 1, 14, 7 y 18, respectivamente, para obtener resultados *más extremos* con los mismos totales de renglón y de columna. (En ocasiones se critica la prueba exacta de Fisher porque el uso de totales fijos de renglón y de columna a menudo es poco realista). La prueba exacta de Fisher requiere que calculemos las probabilidades de las frecuencias observadas y cada conjunto de frecuencias más extremas. Luego, esas probabilidades se suman para obtener un valor *P* exacto.

Como por lo general los cálculos son muy complejos, es mejor utilizar un programa de cómputo. Para los datos de la tabla 11-8, la prueba exacta de Fisher por medio de STATDISK, SPSS, SAS y Minitab produjo un valor *P* exacto de 0.686. Puesto que este valor *P* exacto no es pequeño (como menor que 0.05), no rechazamos la hipótesis nula de que el uso del casco y ser víctima de lesiones faciales son sucesos independientes.

Datos apareados Además del requisito de que cada celda debe tener una frecuencia esperada al menos de 5, los métodos de esta sección también requieren que las observaciones individuales sean *independientes*. Si una tabla 2×2 consiste en conteos de frecuencias que provienen de *datos apareados*, no contamos con la independencia requerida. En estos casos, podemos utilizar la prueba de McNemar, que se explica en la siguiente sección.

13-5 Prueba de Kruskal-Wallis

Concepto clave En esta sección se describe la *prueba de Kruskal-Wallis*, que utiliza rangos de datos de tres o más muestras independientes para probar la hipótesis nula de que las muestras provienen de poblaciones con medianas iguales.

En la sección 12-2 utilizamos el análisis de varianza de un factor (ANOVA) para probar la hipótesis nula de que tres o más poblaciones tienen la misma media, pero el ANOVA requiere que todas las poblaciones implicadas tengan distribuciones normales. La prueba de Kruskal-Wallis no requiere distribuciones normales.

Definición

La **prueba de Kruskal-Wallis** (también llamada la **prueba** H) es una prueba no paramétrica que utiliza rangos de datos muestrales de tres o más poblaciones independientes. Se utiliza para probar la hipótesis nula de que las muestras independientes provienen de poblaciones con medianas iguales; la hipótesis alternativa es la aseveración de que las poblaciones tienen medianas que no son iguales.

H_0: Las muestras provienen de poblaciones con medianas iguales.

H_1: Las muestras provienen de poblaciones con medianas que no son iguales.

Para aplicar la prueba de Kruskal-Wallis, calculamos el *estadístico de prueba H, el cual tiene una distribución que puede aproximarse por medio la distribución chi cuadrada, siempre y cuando cada muestra tenga al menos cinco observaciones*. Cuando utilizamos la distribución chi cuadrada en este contexto, el número de grados de libertad es $k - 1$, donde k es el número de muestras. (Para una revisión rápida de las características clave de la distribución chi cuadrada, véase la sección 7-5).

Prueba de Kruskal-Wallis

Requisitos

1. Tenemos al menos tres muestras independientes, las cuales se seleccionan al azar.
2. Cada muestra tiene al menos cinco observaciones. (Si las muestras tienen menos de cinco observaciones, remítase a tablas especiales de valores críticos, como las *CRC Standard Probability and Statistics Tables and Formulae*, publicadas por CRC Press).
3. *No* existe el requisito de que las poblaciones tengan una distribución normal o alguna otra distribución particular.

Notación

N = número total de observaciones en todas las muestras combinadas

k = número de muestras

R_1 = suma de los rangos de la muestra 1, que se calcula utilizando el procedimiento que se describe a continuación

n_1 = número de observaciones de la muestra 1

Para la muestra 2, la suma de los rangos es R_2 y el número de observaciones es n_2, y se utiliza una notación similar para las otras muestras.

Estadístico de prueba

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \cdots + \frac{R_k^2}{n_k}\right) - 3(N+1)$$

Valores críticos

1. La prueba es de *cola derecha*.
2. gl = $k - 1$. (Puesto que el estadístico de prueba H puede aproximarse por medio de una distribución chi cuadrada, utilice la tabla A-4 con $k - 1$ grados de libertad, donde k es el número de muestras diferentes).

Procedimiento para calcular el valor del estadístico de prueba H

1. Combine temporalmente todas las muestras en una muestra grande y asigne un rango a cada valor muestral. (Ordene los valores del menor al mayor, y en caso de empates, asigne a cada observación la media de los rangos implicados).
2. En cada muestra, calcule la suma de los rangos y calcule el tamaño muestral.
3. Calcule H utilizando los resultados del paso 2, con la notación y el estadístico de prueba descritos en el recuadro anterior.

El estadístico de prueba H es básicamente una medida de la varianza de las sumas de rangos $R_1, R_2, \ldots, R_k$. Si los rangos están distribuidos de forma equitativa entre los grupos muestrales, entonces H debe ser un número relativamente pequeño. Si las muestras son muy diferentes, entonces los rangos serán excesivamente bajos en algunos grupos y altos en otros, con el efecto neto de que H será grande. En consecuencia, sólo los valores grandes de H nos llevan al rechazo de la hipótesis nula de que las muestras provienen de poblaciones idénticas. *La prueba de Kruskal-Wallis es, por lo tanto, una prueba de cola derecha.*

13-6 Correlación de rangos

Concepto clave En esta sección se describe el método no paramétrico de la correlación de rangos, el cual se utiliza con datos apareados para probar una asociación entre dos variables. En el capítulo 10 utilizamos datos muestrales apareados para calcular valores del coeficiente de correlación lineal r, pero en esta sección utilizaremos *rangos* como base para calcular el coeficiente de correlación de rangos r_s.

Definición

La **prueba de correlación de rangos** (o **prueba de correlación de rangos de Spearman**) es una prueba no paramétrica que utiliza rangos de datos muestrales consistentes en datos apareados. Se utiliza para probar una asociación entre dos variables, por lo que las hipótesis nula y alternativa son las siguientes (donde ρ_s denota el coeficiente de correlación de rangos de la población completa):

H_0: $\rho_s = 0$ (*No* existe correlación entre las dos variables).

H_1: $\rho_s \neq 0$ (Existe una correlación entre las dos variables).

Ventajas: La correlación de rangos tiene las siguientes ventajas sobre los métodos paramétricos analizados en el capítulo 10:

1. El método no paramétrico de correlación de rangos puede utilizarse en una variedad más amplia de circunstancias que el método paramétrico de correlación lineal. Con la correlación de rangos, podemos analizar datos apareados que sean rangos o puedan convertirse en rangos. Por ejemplo, si dos jueces califican a 30 gimnastas, podemos utilizar la correlación de rangos, pero no la correlación lineal. A diferencia de los métodos paramétricos del capítulo 10, el método de correlación de rangos *no* requiere una distribución normal de cualquier población.
2. La correlación de rangos puede utilizarse para detectar algunas relaciones (no todas) que no son lineales. (Más adelante se dará un ejemplo en esta sección).

Desventaja: Una desventaja de la correlación de rangos es su tasa de eficiencia de 0.91, como se describe en la sección 13-1. Esta tasa de eficiencia indica que, con todas las demás circunstancias iguales, el método no paramétrico de correlación de rangos requiere de 100 pares de datos muestrales para tener los mismos resultados que sólo 91 pares de observaciones muestrales analizadas a través del método paramétrico, suponiendo que los requisitos más estrictos del método paramétrico se satisfacen.

Utilizamos la notación r_s para el coeficiente de correlación de rangos para no confundirlo con el coeficiente de correlación lineal r. El subíndice s no tiene nada que ver con la desviación estándar; se usa en honor de Charles Spearman (1863-1945), quien originó el método de correlación de rangos. De hecho, r_s suele llamarse **coeficiente de correlación de rangos de Spearman**. El procedimiento de la correlación de rangos se resume en la figura 13-4.

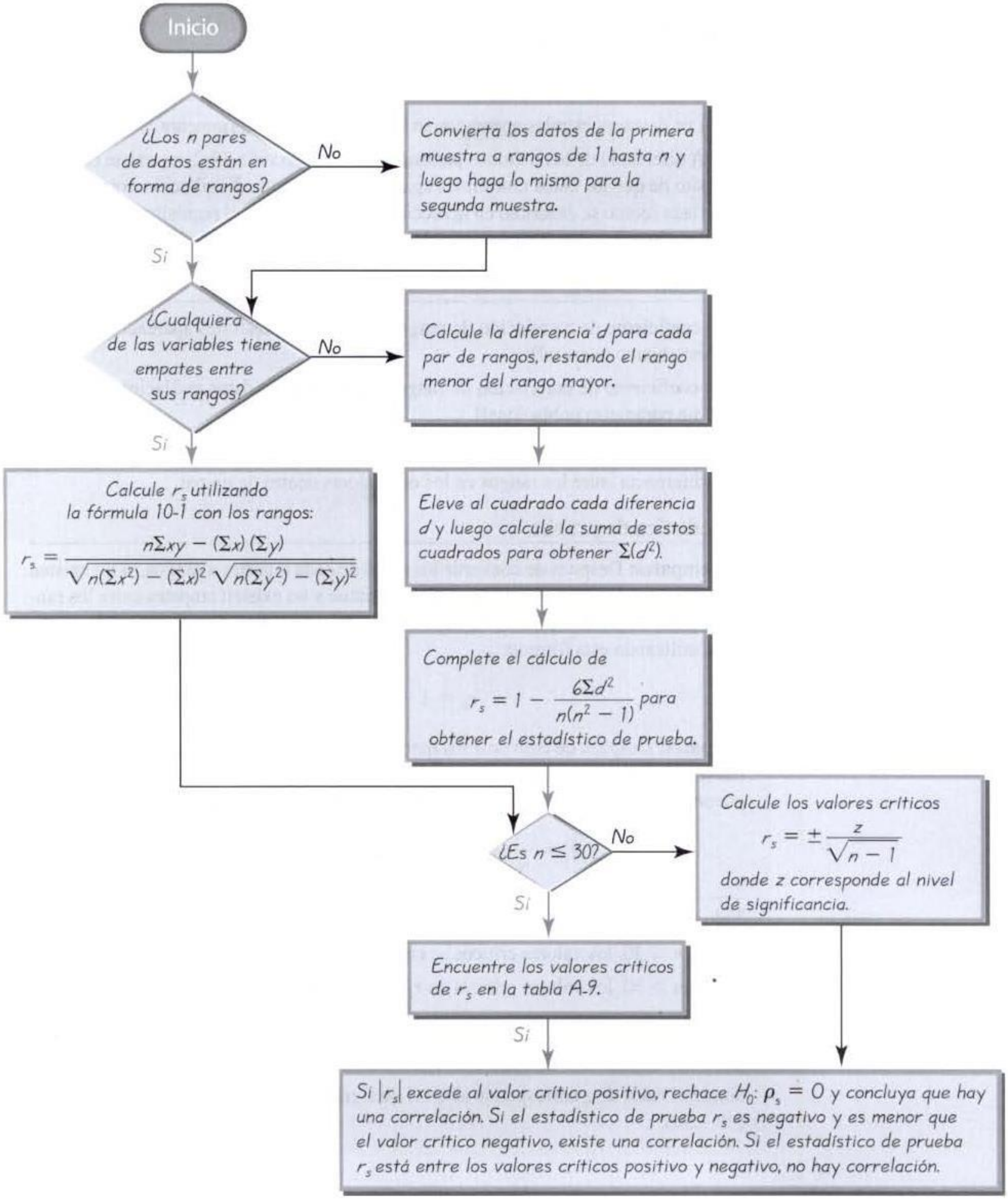


Figura 13-4 **Procedimiento de correlación de rangos para probar H_0: $\rho_s = 0$**

Correlación de rangos

Requisitos

1. Los datos muestrales apareados se seleccionaron aleatoriamente.
2. A diferencia de los métodos paramétricos de la sección 10-2, *no* existe el requisito de que los datos muestrales apareados tengan una distribución normal bivariada (como se describió en la sección 10-2). *No* existe el requisito de una distribución normal para cualquier población.

Notación

r_s = coeficiente de correlación de rangos para datos muestrales apareados (r_s es un estadístico muestral)

ρ_s = coeficiente de correlación de rangos para todos los datos poblacionales (ρ_s es un parámetro poblacional)

n = número de pares de datos muestrales

d = diferencia entre los rangos de los dos valores dentro de un par

Estadístico de prueba

Sin empates: Después de convertir los datos de cada muestra a rangos, si no existen empates entre los rangos para la primera variable y no existen empates entre los rangos para la segunda variable, el valor exacto del estadístico de prueba puede calcularse utilizando esta fórmula:

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

Empates: Después de convertir los datos de cada muestra a rangos, si cualquier variable tiene empates entre sus rangos, el valor exacto del estadístico de prueba r_s puede calcularse utilizando la fórmula 10-1 con los rangos:

$$r_s = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

Valores críticos

1. Si $n \leq 30$, los valores críticos se encuentran en la tabla A-9.
2. Si $n > 30$, los valores críticos de r_s se calculan utilizando la fórmula 13-1.

Fórmula 13-1 $$r_s = \frac{\pm z}{\sqrt{n - 1}}$$ **(valores críticos cuando $n > 30$)**

donde el valor de z corresponde al nivel de significancia. (Por ejemplo, si $\alpha = 0.05$, $z = 1.96$).

13-7 Prueba de rachas para detectar aleatoriedad

Concepto clave En esta sección estudiaremos la prueba de rachas para detectar aleatoriedad, la cual puede utilizarse para determinar si los datos muestrales en una secuencia están en un orden aleatorio. Esta prueba se basa en datos muestrales que tienen dos características y analiza rachas de esas características para determinar si las rachas parecen ser el resultado de algún proceso aleatorio, o si las rachas sugieren que el orden de los datos no es aleatorio.

Definiciones

Una **racha** es una secuencia de datos que tienen la misma característica; la secuencia es precedida y seguida por datos con una característica diferente o por ningún dato en absoluto.

La **prueba de rachas** utiliza el número de rachas en una secuencia de datos muestrales para probar la aleatoriedad del orden de los datos.

Principio fundamental de la prueba de rachas

El principio fundamental de la prueba de rachas puede establecerse brevemente como sigue:

Rechace la aleatoriedad si el número de rachas es muy bajo o muy alto.

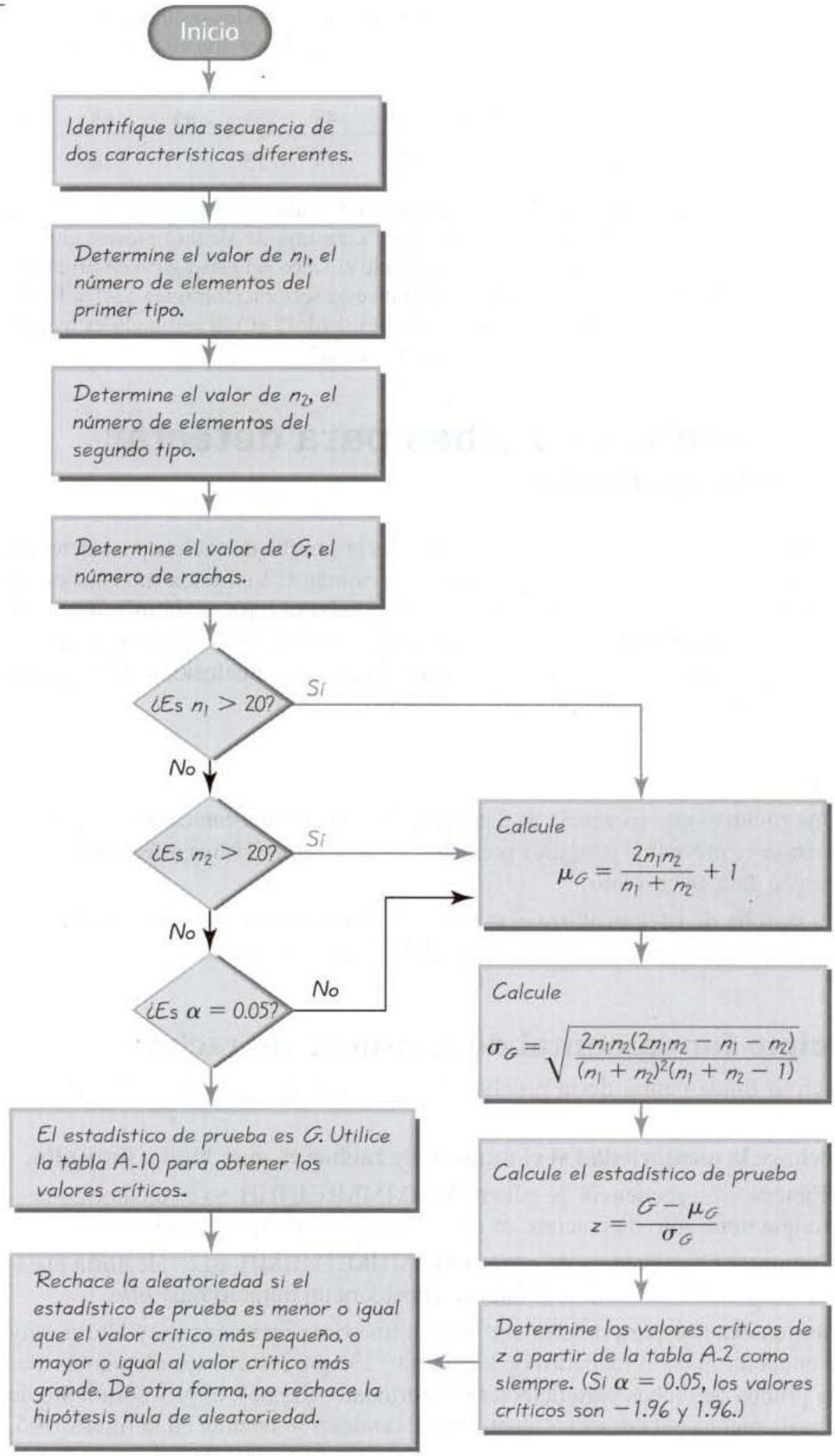
- Ejemplo: La secuencia de género MMMMMHHHHH no es aleatoria puesto que tiene sólo dos rachas, es decir, el número de rachas es muy *bajo*.
- Ejemplo: La secuencia de género MHMHMHMHMH no es aleatoria puesto que existen 10 rachas, lo cual se considera un número muy *alto*.

Los criterios exactos para determinar si un número de rachas es muy alto o muy bajo se encuentran en el recuadro de la página 723, que resume los elementos clave de la prueba de rachas para detectar aleatoriedad. Además, el procedimiento de la prueba de rachas para detectar aleatoriedad también se resume en la figura 13-5.

Es importante señalar que la prueba de rachas para detectar aleatoriedad se basa en el *orden* en el que se presentan los datos; *no* está basada en la *frecuencia* de los datos. Por ejemplo, una secuencia de 3 hombres y 20 mujeres podría parecer aleatoria, pero la prueba de rachas no se ocupa del problema de si 3 hombres y 20 mujeres constituyen una muestra *sesgada* (con un número desproporcionadamente mayor de mujeres).

Figura 13-5

Prueba de rachas para detectar aleatoriedad



Prueba de rachas para detectar aleatoriedad

Requisitos

1. Los datos muestrales están acomodados de acuerdo con algún esquema de orden, por ejemplo, el orden en el que se obtuvieron los valores muestrales.
2. Cada valor de los datos se puede categorizar en una de *dos* categorías separadas (como hombre/mujer).

Notación

n_1 = número de elementos en la secuencia que tienen una característica particular. (La característica elegida para n_1 es arbitraria).

n_2 = número de elementos en la secuencia que tienen la otra característica

G = número de rachas

Estadístico de prueba

Para muestras pequeñas y $\alpha = 0.05$: Si $n_1 \leq 20$ y $n_2 \leq 20$ y el nivel de significancia es $\alpha = 0.05$, el estadístico de prueba es el número de rachas G. Los valores críticos se encuentran en la tabla A-10. He aquí el criterio de decisión:

Rechace la aleatoriedad si el número de rachas G es

- menor o igual al valor crítico más pequeño encontrado en la tabla A-10.
- mayor o igual al valor crítico más grande encontrado en la tabla A-10.

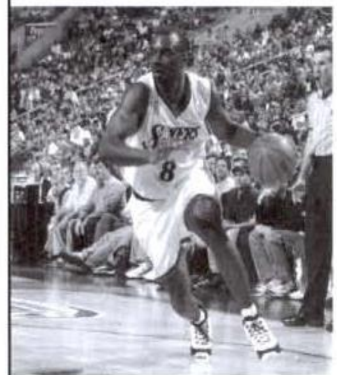
Para muestras grandes o $\alpha \neq 0.05$: If $n_1 > 20$ o $n_2 > 20$ o $\alpha \neq 0.05$, utilice el estadístico de prueba y los valores críticos siguientes.

Estadístico de prueba: $$z = \frac{G - \mu_G}{\sigma_G}$$

donde $$\mu_G = \frac{2n_1n_2}{n_1 + n_2} + 1$$

y $$\sigma_G = \sqrt{\frac{(2n_1n_2)(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

Valores críticos de z: Utilice la tabla A-2.



Rachas de suerte en los deportes

Existe la creencia de que los atletas suelen tener "rachas de suerte", es decir, periodos breves de éxito extraordinario. El psicólogo Amos Tversky, de la Universidad de Stanford, y otros investigadores utilizaron la estadística para analizar los miles de tiros de los 76 de Filadelfia en una temporada completa y la mitad de otra. Encontraron que el número de "rachas de suerte" no difería de lo que se esperaría en pruebas aleatorias, donde el resultado de cada prueba es independiente de cualquier resultado previo. Es decir, la probabilidad de una anotación no depende de las anotaciones o fallas previas.

Práctica de Estadística II - N° 09

TEMA N°9 : ANOVA

Sección : ………………………..………………..

Docente : Escribir el nombre del docente

Apellidos : ………………………..………………
Nombres : ……………………………………………
Fecha : ……/……/2016 Duración: 45 minutos
Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

01. Se está estudiando la resistencia a la compresión de concreto. Se investigan cuatro técnicas diferentes de mezclado. Los datos recabados son los siguientes:

Técnica de mezclado	Resistencia a la compresión (miles de psi)		
1	3,12	3,00	2,86
2	3,20	3,30	2,97
3	2,80	2,90	2,95
4	2,60	2,70	2,60

Pruebe la hipótesis de que las técnicas de mezclado afectan la resistencia del concreto. Emplea α= 0.05

02. Una fábrica de hilados tiene un gran número de telares. Se supone que cada uno de los telares proporciona la misma salida de tela por minuto. Para investigar esta suposición, se eligen tres telares al azar y su salida se mide en diferentes tiempos. Se obtienen los siguientes datos:

Telar	Salida (lb/min)				
1	4.0	4.1	4.2	4.0	4.1
2	3.9	3.8	3.9	4.0	4.0
3	4.1	4.2	4.1	4.0	3.9

¿Son los telares son similares en la salida?

03. Un ingeniero en electrónica está interesado en el efecto de la conductividad de cinco diferentes tipos de recubrimiento para tubos de rayos catódicos utilizados en un dispositivo de despliegue de telecomunicaciones. Se obtuvieron los siguientes datos de conductividad:

Tipo de recubrimiento	Conductividad			
1	143	141		
2	152	149	137	143
3	134	133	132	
4	129	127		
5	147	148	144	

¿Hay alguna diferencia en la conductividad debida al tipo de recubrimiento?

04. Se determinó el tiempo de respuesta en milisegundos para tres tipos diferentes de circuitos utilizados en una calculadora electrónica. Los resultados se presentan a continuación:

Tipo de circuito	Tiempo de respuesta				
1	19	22	20	18	25
2	20	21	33	27	40
3	16	15	18	26	17

¿Tienen los tres circuitos una respuesta homogénea?

05. Se están considerando seis máquinas diferentes para la fabricación de sellos de goma y se están comparando con respecto a la resistencia a la tensión del producto. Se utiliza una muestra aleatoria de cuatro sellos hechos con cada máquina para determinar si la resistencia media a la tensión varía de una máquina a otra. A continuación se presentan las medidas de la resistencia a la tensión en kilogramos por centímetro cuadrado x 10“1:

Máquina					
1	2	3	4	5	6
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5

Realice el análisis de varianza a un nivel de significancia de 0.05 e indique si la resistencia promedio a la tensión de las seis máquinas difiere o no de manera significativa.

06. Un estudio midió la tasa de sorción (ya sea absorción o adsorción) de tres tipos diferentes de solventes químicos orgánicos. Estos solventes se utilizan para limpiar partes industriales metálicas, y son desechos potencialmente riesgosos. Se probaron muestras independientes de solventes de cada tipo y se registraron sus tasas de sorción como un porcentaje molar. ¿Existe una diferencia significativa en la tasa promedio de sorción de los tres solventes

Aromáticos		Cloroalcalinos		Esteres		
1.06	0.95	1.58	1.12	0.29	0.43	0.06
0.79	0.65	1.45		0.06	0.51	

Práctica de Estadística II - N° 10

TEMA N°10 : ESTADISTICA NO PARAMETRICA

Sección : ……………………….……………...

Docente : Escribir el nombre del docente

Apellidos : ……………………….……………….
Nombres : …………………………………………….
Fecha : …../……/2017 Duración: 45 minutos
Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

1..En una universidad se desea saber si los alumnos decían la verdad al solicitarles cierta información, para ello se llevó a cabo un experimento que consistía en preguntarle a cada alumno su respectiva estatura y estas se contrataron con la estatura real, La siguiente tabla contienelasestaturasquesemidieronde11estudiantes
.Utilice un nivel de significancia de 0.05 y pruebe la aseveración de que no hay diferencia entre las estaturas reportadas y las estaturas medidas, es decir que los alumnosnomienten.

Estatura reportada	68	74	82.25	66.5	69	68	71	70	70	67	68
Estatura medida	66.8	73.9	74.3	66.1	67.2	67.9	69.4	69.9	68.6	67.9	67.6

2.-.En una encuesta que se aplicó a 1002 personas, 701 dijeron que votaroen la elección presidencial reciente (según datos DATUM). ¿Hay suficienteevidenciapara sustentar la aseveración de que la mayoría de las personasvotaronenlaelección?

3.-Un médico planea reunir datos muéstrales para probar la aseveración de que la temperatura corporal mediaesmenorque98.6°F. Por restricciones de tiempo, encuentra que sólo alcanzará a reunir datos de 12 personas. Utilice un nivel de significancia de 0.05 y pruebe la aseveración de que dichas temperaturas corporalesprovienendeunapoblaciónconunamediana queesmenorque98.6°F.

97,6	97,5	98,6	98,2	98,2	99,1
98,5	98,1	98,4	97,9	97,9	97,7

4.-El captopril es un fármaco que se diseñó para disminuir la presión sanguínea sistólica. Cuando se probó este fármaco en sujetos, sus lecturas de presión sanguínea sistólica (en milímetros de mercurio) se midieron antes y después de que se tomara el fármaco, con los resultados que se presentan en la tabla adjunta.

¿Hay suficiente evidencia para sustentar la aseveración de que el fármaco no surte efecto? ¿Parece que el captopril disminuye la presión sanguínea sistólica?

Sujeto	A	B	C	D	E	F	G	H	I	J	K
Antes	200	174	198	170	179	182	193	209	185	155	169
Después	191	170	177	167	159	151	176	183	159	145	146

5.-Se estudió el orden de preguntas de examen para ver su efecto en la ansiedad. Los resultados muéstrales se listan abajo. Utilizando un niveldesignificanciade0.05, pruebelaaseveracióndequelasdosmuestrasprovienen depoblacionesconlasmismascalificaciones.

Fácil a difícil

24.64	39.29	16.32	32.83
28.02	33.31	20.60	21.13
26.69	28.90	26.43	24.23
7.10	32.86	21.06	28.89
28.71	31.73	30.02	21.96
25.49	38.81	27.85	30.29
30.72			

Difícil a fácil

33.62	34.02	26.63
35.91	26.68	29.49
27.24	32.34	29.34
27.62	42.91	30.20

6.-A continuación, … el consumo de combustible (en millas/gal) para diferentes automóvil… …ue entró en vigor en 2008 (según datos de USA Today). El nuevo sistema se implementó en respuesta a las quejas de que las antiguas calificaciones eran demasiado elevadas. Utilice un nivel de significancia de 0.01 para someter a prueba la afirmación de que las calificaciones antiguas son más altas que las calificaciones nuevas.

Sujeto	A	B	C	D	E	F	G	H	I	J	K
Antes	200	174	198	170	179	182	193	209	185	155	169
Después	191	170	177	167	159	151	176	183	159	145	146

Práctica de Estadística II - N° 11

TEMA N° 11: ANOVA

Sección : ……………………….………………...	Apellidos : ……………………….………………
Docente : Escribir el nombre del docente	Nombres : ……………………………………………
	Fecha : ……/……/2017 Duración: 45 minutos
	Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

06. Se está estudiando la resistencia a la compresión de concreto. Se investigan cuatro técnicas diferentes de mezclado. Los datos recabados son los siguientes:

Técnica de mezclado	Resistencia a la compresión (miles de psi)		
1	3,12	3,00	2,86
2	3,20	3,30	2,97
3	2,80	2,90	2,95
4	2,60	2,70	2,60

Pruebe la hipótesis de que las técnicas de mezclado afectan la resistencia del concreto. Emplea $\alpha = 0.05$

07. Una fábrica de hilados tiene un gran número de telares. Se supone que cada uno de los telares proporciona la misma salida de tela por minuto. Para investigar esta suposición, se eligen tres telares al azar y su salida se mide en diferentes tiempos. Se obtienen los siguientes datos:

Telar	Salida (lb/min)				
1	4.0	4.1	4.2	4.0	4.1
2	3.9	3.8	3.9	4.0	4.0
3	4.1	4.2	4.1	4.0	3.9

¿Son los telares son similares en la salida?

08. Un ingeniero en electrónica está interesado en el efecto de la conductividad de cinco diferentes tipos de recubrimiento para tubos de rayos catódicos utilizados en un dispositivo de despliegue de telecomunicaciones. Se obtuvieron los siguientes datos de conductividad:

Tipo de recubrimiento	Conductividad			
1	143	141		
2	152	149	137	143
3	134	133	132	
4	129	127		
5	147	148	144	

¿Hay alguna diferencia en la conductividad debida al tipo de recubrimiento?

09. Se determinó el tiempo de respuesta en milisegundos para tres tipos diferentes de circuitos utilizados en una calculadora electrónica. Los resultados se presentan a continuación:

Tipo de circuito	Tiempo de respuesta				
1	19	22	20	18	25
2	20	21	33	27	40
3	16	15	18	26	17

¿Tienen los tres circuitos una respuesta homogénea?

10. Se están considerando seis máquinas diferentes para la fabricación de sellos de goma y se están comparando con respecto a la resistencia a la tensión del producto. Se utiliza una muestra aleatoria de cuatro sellos hechos con cada máquina para determinar si la resistencia media a la tensión varía de una máquina a otra. A continuación se presentan las medidas de la resistencia a la tensión en kilogramos por centímetro cuadrado x 10“1:

Máquina					
1	2	3	4	5	6
17.5	16.4	20.3	14.6	17.5	18.3
16.9	19.2	15.7	16.7	19.2	16.2
15.8	17.7	17.8	20.8	16.5	17.5

11.-Realice el análisis de varianza a un nivel de significancia de 0.05 e indique si la resistencia promedio a la tensión de las seis máquinas difiere o no de manera significativa.

06.Un estudio midió la tasa de sorción (ya sea absorción o adsorción) de tres tipos diferentes de solventes químicos orgánicos. Estos solventes se utilizan para limpiar partes industriales metálicas, y son desechos potencialmente riesgosos. Se probaron muestras independientes de solventes de cada tipo y se registraron sus tasas de sorción como un porcentaje molar. ¿Existe una diferencia significativa en la tasa promedio de sorción de los tres solventes

Aromáticos		**Cloroalcalinos**		**Esteres**		
1.06	0.95	1.58	1.12	0.29	0.43	0.06
0.79	0.65	1.45		0.06	0.51	

CUARTA UNIDAD
CORRELACION Y REGRESION

- GUÍA DE PRÁCTICA Nº 13: Correlación y Regresión, Prueba de Hipótesis
- GUÍA DE PRÁCTICA Nº 14: Correlación y Regresión , Regresión Múltiple
- GUÍA DE PRÁCTICA Nº 15:
- GUÍA DE PRÁCTICA Nº 16:

TEXTO N° 18

Correlación, regresión, variación e intervalos de predicción

Compilado y adaptado de Mario Triola Estadística PEARSON; 2009

Págs. De 517 a 531, de 541 a 552 y de 557 a 561

10-1 Panorama general

Este capítulo presenta métodos importantes para hacer inferencias sobre una *correlación* (o relación) entre dos variables. También se explicará cómo describir este tipo de relación con una ecuación que pueda emplearse para predecir el valor de una variable, dado el valor de la otra variable. Consideramos datos muestrales ordenados en *pares*. La sección 9-4 también utilizó muestras apareadas, pero el objetivo era diferente en esencia. Las inferencias en la sección 9-4 se referían a la *media de las diferencias* entre pares de valores; en cambio, este capítulo tiene el objetivo de hacer inferencias sobre *relaciones* entre dos variables.

En la sección 10-2 se analiza la relación entre dos variables. La sección 10-3 presenta el análisis de regresión, el cual nos permite describir la relación por medio de una ecuación. En la sección 10-4 analizamos las diferencias entre valores predichos y valores reales observados. Las secciones 10-2 a la 10-4 se ocupan de relaciones entre *dos* variables, en tanto que la sección 10-5 presenta la regresión múltiple para relaciones entre tres o más variables. Por último, en la sección 10-6 se describen algunos métodos básicos para desarrollar un modelo matemático que puede emplearse para describir la relación entre dos variables. Aunque la sección 10-3 está limitada a las relaciones lineales, la sección 10-6 incluye algunas relaciones comunes no lineales.

top 20 10-2 Correlación

Concepto clave En esta sección se explica el *coeficiente de correlación lineal* r que es una medida numérica de la fuerza de la relación entre dos variables que representan datos cuantitativos. Utilizando datos muestrales apareados (que en ocasiones se llaman **datos bivariados**), calculamos el valor de r (generalmente con la ayuda de recursos tecnológicos) y luego utilizamos este valor para concluir que existe (o no) una relación entre las dos variables. En esta sección sólo consideramos las relaciones *lineales*, lo que quiere decir que cuando se grafican, los puntos se aproximan al patrón de una línea recta. Puesto que los programas de cómputo o las calculadoras suelen emplearse para calcular el valor de r, es importante enfocarse en los conceptos de esta sección, sin entretenerse demasiado con cálculos aritméticos tediosos.

Parte 1: Conceptos básicos de correlación

Iniciamos con la definición básica de *correlación*, un término que se utiliza comúnmente en el contexto de una relación entre dos variables.

Definición

Una correlación existe entre dos variables cuando una de ellas está relacionada con la otra de alguna manera.

La tabla 10-1, por ejemplo, consiste en datos apareados de las duraciones y los intervalos posteriores a las erupciones del géiser Old Faithful. (El "intervalo de tiempo posterior" a una erupción es el tiempo que transcurre hasta la siguiente erupción). Determinaremos si existe una correlación entre la variable x (duración) y la variable y (intervalo de tiempo posterior).

Exploración de los datos

Antes de trabajar con los métodos más formales de cálculo de esta sección, primero debemos explorar el conjunto de datos para ver qué podemos aprender. Con frecuencia podemos encontrar una relación entre dos variables al construir un diagrama de dispersión. Cuando examinamos un diagrama de dispersión, debemos estudiar el patrón general de los puntos graficados. Si existe un patrón, es necesario observar su dirección. Si los datos van hacia arriba, esto sugiere que cuando una variable aumenta, la otra también lo hace. Si los datos van hacia abajo, esto sugiere que cuando una variable aumenta, la otra disminuye. Debemos buscar valores extremos, que son puntos que se ubican muy lejos del resto de los puntos.

La figura 10-2 muestra diagramas de dispersión con características diferentes. Los diagramas de dispersión de la figura 10-2a), b) y c) describen un patrón en el que los valores crecientes de y corresponden a valores crecientes de x. Si observa las figuras de la a) a la c), se dará cuenta de que el patrón de puntos se acerca cada vez más a una línea recta, lo que sugiere que la relación entre x y y se vuelve más fuerte. Los diagramas de dispersión de la figura 10-2d), e) y f) describen patrones en los que los valores de y disminuyen cuando los valores de x aumentan. Nuevamente, si observa las figuras de la d) a la f), verá que la relación se vuelve más fuerte. En contraste con las primeras seis gráficas, el diagrama de dispersión de la figura 10-2g) no muestra patrón alguno y sugiere que no existe una correlación (o relación) entre x y y. Por último, el diagrama de dispersión de la figura 10-2h) muestra un patrón muy diferente, que sugiere una relación entre x y y, pero no es un patrón lineal (línea recta). [Las figuras 10-2a) a la g) son de ActivStats, y la figura 10-2h) es de Minitab].

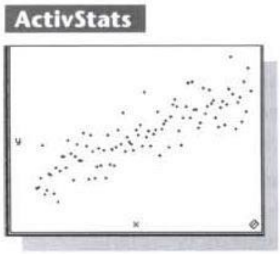
Coeficiente de correlación lineal

Puesto que el examen visual de los diagramas de dispersión es muy subjetivo, necesitamos medidas más precisas y objetivas. Empleamos el coeficiente de correlación lineal r, que sirve para detectar patrones lineales.

Definición

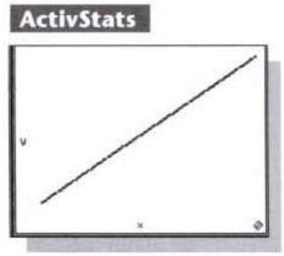
El **coeficiente de correlación lineal** r mide la fuerza de la relación lineal entre los valores cuantitativos apareados x y y en una *muestra*. Su valor se calcula con la fórmula 10-1, incluida en el recuadro de la página 520. [El coeficiente de correlación lineal también se conoce como **coeficiente de correlación producto momento de Pearson**, en honor de Karl Pearson (1857-1936), quien lo desarrolló originalmente].

Puesto que el coeficiente de correlación lineal r se calcula utilizando datos muestrales, se trata de un estadístico muestral empleado para medir la fuerza de la correlación lineal entre x y y. Si tuviéramos cada par de los valores poblacionales de x y y, el resultado de la fórmula 10-1 sería un parámetro poblacional, representado por ρ (rho griega). El recuadro en la página 520 incluye los supuestos requeridos, la notación y la fórmula 10-1.

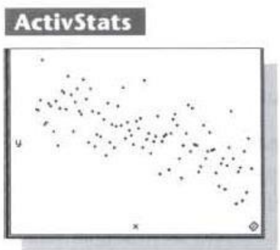


a) **Correlación positiva: $r = 0.851$**

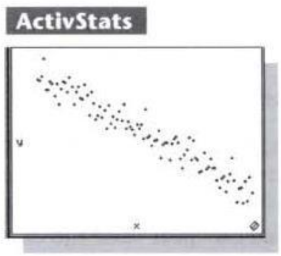
b) **Correlación positiva: $r = 0.991$**



c) **Correlación positiva perfecta: $r = 1$**



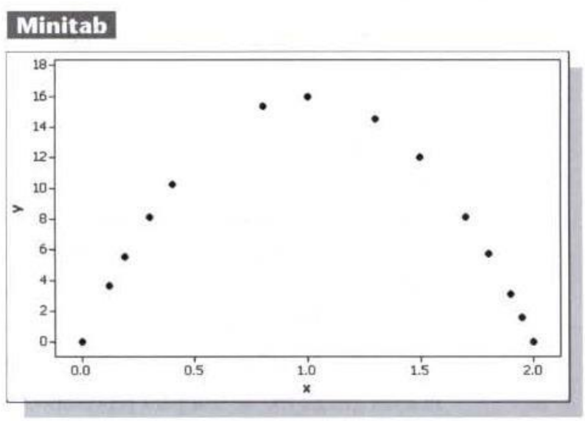
d) **Correlación negativa: $r = -0.702$**



e) **Correlación negativa: $r = -0.965$**

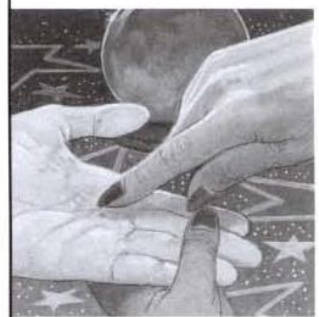
f) **Correlación negativa perfecta: $r = -1$**

g) **Sin correlación: $r = 0$**



h) **Relación no lineal: $r = -0.087$**

Figura 10-2 Diagramas de dispersión



Lectura de las manos

Algunas personas piensan que la longitud de la línea de la vida de la palma de las manos puede utilizarse para predecir la longevidad. En una carta publicada en el *Journal of the American Medical Association*, los autores M. E. Wilson y L. E. Mather refutaron esta creencia con un estudio de cadáveres. Se registraron las edades de los sujetos al morir, junto con las longitudes de la línea de la vida de sus palmas. Los autores concluyeron que no existe una correlación significativa entre la edad al morir y la longitud de la línea de la vida. La quiromancia pierde: ¡Abajo las manos!

Requisitos

Dado cualquier conjunto de datos muestrales apareados, siempre se puede calcular el coeficiente de correlación lineal r, pero se deben satisfacer los siguientes requisitos cuando se prueban hipótesis o cuando se hacen inferencias acerca de r.

1. La muestra de datos apareados (x, y) es una muestra *aleatoria* de datos cuantitativos. (Es importante que los datos muestrales no se hayan reunido por medio de algún método inapropiado, como una muestra de respuesta voluntaria).
2. El examen visual del diagrama de dispersión debe confirmar que los puntos se acercan al patrón de una línea recta.
3. Es necesario eliminar cualquier valor extremo, si se sabe que se trata de un error. Los efectos de cualquier otro valor extremo deben tomarse en cuenta calculando r con y sin el valor extremo incluido.

Nota: Los requisitos 2 y 3 se simplifican al verificar el siguiente requisito formal:

Los pares de datos (x, y) tienen una **distribución normal bivariada**. (Las distribuciones normales se estudiaron en el capítulo 6, pero este supuesto requiere que, para cualquier valor fijo de x, los valores correspondientes de y tengan una distribución con forma de campana, y que para cualquier valor fijo de y, los valores de x tengan también una distribución con forma de campana). Suele ser difícil verificar este supuesto, así que, por ahora, usaremos los requisitos 2 y 3 descritos arriba.

Notación para el coeficiente de correlación lineal

n	representa el número de pares de datos presentes.
Σ	denota la suma de los elementos indicados.
Σx	denota la suma de todos los valores de x.
Σx^2	indica que cada valor de x debe elevarse al cuadrado y después deben sumarse esos cuadrados.
$(\Sigma x)^2$	indica que los valores de x deben sumarse y el total elevarse al cuadrado. Es sumamente importante evitar confundirse entre Σx^2 y $(\Sigma x)^2$
Σxy	indica que cada valor de x debe multiplicarse primero por su valor y correspondiente. Después de obtener todos estos productos, se calcula su suma.
r	representa el coeficiente de correlación lineal de una *muestra*.
ρ	la letra griega rho se usa para representar el coeficiente de correlación lineal de una *población*.

Fórmula 10-1 $$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\,\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

Esta fórmula abreviada simplifica los cálculos manuales. Su formato la hace fácil de usar en una hoja de cálculo o en un programa de cómputo. Consulte otras fórmulas equivalentes que se presentan más adelante en esta sección; también consulte los fundamentos del cálculo de r.

Interpretación de *r* por medio de la tabla A-6: Si el valor absoluto del valor calculado de *r* excede el valor de la tabla de A-6, concluya que existe una correlación lineal significativa. De lo contrario, no existe evidencia suficiente para sustentar la conclusión de una correlación lineal.

Interpretación de *r* por medio de un programa de cómputo: Si el valor *P* calculado es menor o igual que el nivel de significancia, concluya que existe una correlación lineal. De lo contrario, no existe evidencia suficiente para sustentar la conclusión de una correlación lineal.

Redondeo del coeficiente de correlación lineal

Redondee el coeficiente de correlación lineal *r* a tres decimales (de manera que su valor pueda compararse directamente con los valores críticos de la tabla de A-6). Al calcular a mano *r* y otros estadísticos de este capítulo, hacer un redondeo a la mitad de un cálculo suele generar errores importantes, así que, trate de utilizar la memoria de su calculadora para almacenar los resultados inmediatos y redondee sólo al final.

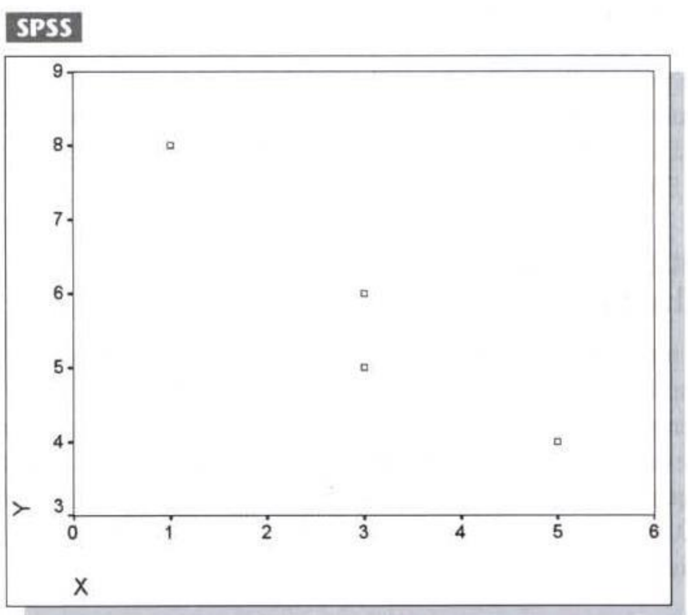
x	3	1	3	5
y	5	8	6	4

EJEMPLO Cálculo de *r* Utilice la muestra aleatoria simple de datos que aparece al margen para calcular el valor del coeficiente de correlación lineal *r*.

SOLUCIÓN

REQUISITO ✔ Los datos constituyen una muestra aleatoria simple. El diagrama de dispersión generado por SPSS que aparece abajo sugiere un patrón de puntos que se acerca al patrón de una línea recta. No hay valores extremos. Los requisitos se satisfacen y podemos continuar con los cálculos del coeficiente de correlación lineal *r*. ✔

continúa



Las evaluaciones de profesores se correlacionan con las calificaciones

Con frecuencia se utilizan las evaluaciones que hacen los estudiantes de los profesores para medir su efectividad. Muchos estudios revelan una correlación de altas calificaciones de los estudiantes con evaluaciones positivas de los profesores. Un estudio realizado en la Universidad Duke incluyó evaluaciones de los estudiantes, recabadas antes y después de la entrega de las calificaciones finales. El estudio reveló que "las expectativas de las calificaciones o las calificaciones recibidas causaron un cambio en la forma en que los estudiantes percibían a los maestros y la calidad de su enseñanza". Se señaló que con las evaluaciones de los estudiantes "aumentan los incentivos de los profesores para manipular sus políticas de calificación con el fin de mejorar sus evaluaciones". Se concluyó que "la consecuencia final de este tipo de manipulaciones es la degradación de la calidad de la educación en Estados Unidos". (Véase "Teacher Course Evaluation and Student Grades: An Academic Tango", de Valen Johnson, *Chance*, vol. 15, núm. 3).

Tabla 10-2 Cálculo de estadísticos empleados para calcular r

	x	y	$x \cdot y$	x^2	y^2
	3	5	15	9	25
	1	8	8	1	64
	3	6	18	9	36
	5	4	20	25	16
Total	**12**	**23**	**61**	**44**	**141**
	↑	↑	↑	↑	↑
	Σx	Σy	Σxy	Σx^2	Σy^2

Para la muestra de datos apareados, $n = 4$, ya que existen cuatro pares de datos. Los otros componentes requeridos para la fórmula 10-1 se obtienen de los cálculos en la tabla 10-2. Observe que este formato vertical facilita los cálculos.

Con los valores calculados y la fórmula 10-1, ahora podemos evaluar r de la siguiente forma:

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}}$$

$$= \frac{4(61) - (12)(23)}{\sqrt{4(44) - (12)^2}\sqrt{4(141) - (23)^2}}$$

$$= \frac{-32}{\sqrt{32}\sqrt{35}} = -0.956$$

Estos cálculos se vuelven confusos con conjuntos grandes de datos, por lo que es una suerte que el coeficiente de correlación lineal pueda calcularse automáticamente por medio de muchas calculadoras y programas de cómputo distintos. Consulte el recuadro "Uso de la tecnología" al final de esta sección, donde encontrará comentarios acerca de STATDISK, Minitab, Excel y la calculadora TI-83/84 Plus.

Interpretación del coeficiente de correlación lineal

Necesitamos interpretar un valor calculado de r, tal como el valor de -0.956 obtenido en el ejemplo anterior. Dada la manera en que la fórmula 10-1 está construida, el valor de r siempre debe estar entre -1 y $+1$, inclusive. Si r se acerca a 0, concluimos que no existe una correlación lineal entre x y y, pero si r se acerca -1 o $+1$, concluimos que hay una correlación lineal entre x y y. Interpretaciones tales como "cercano a" 0, a 1 o a -1 son vagas, por lo que utilizamos el siguiente criterio específico de decisión:

> ***Uso de la tabla A-6:*** **Si el valor absoluto del valor calculado de r excede el valor de la tabla a A-6, se concluye que existe una correlación lineal. De lo contrario, no existe evidencia suficiente para sustentar la conclusión de una correlación lineal.**

Uso de un programa de cómputo: **Si el valor P calculado es menor o igual que el nivel de significancia, se concluye que existe una correlación lineal. De lo contrario, no existe evidencia suficiente para sustentar la conclusión de una correlación lineal.**

Cuando en realidad no existe una correlación lineal entre x y y, la tabla A-6 lista valores que son "críticos" en este sentido: separan valores *comunes* de r de aquellos que son *poco comunes*. Por ejemplo, la tabla de A-6 nos indica que con $n = 4$ pares de datos muestrales, los valores críticos son 0.950 (para $\alpha = 0.05$) y 0.999 (para $\alpha = 0.01$). Los valores críticos y el papel de α se describen detalladamente en los capítulos 7 y 8. He aquí cómo interpretamos estos números: con 14 pares de datos y ninguna correlación lineal entre x y y, existe una probabilidad del 5% de que el valor absoluto del coeficiente de correlación lineal r calculado exceda 0.950. Con $n = 4$ y sin correlación lineal, existe una probabilidad del 1% de que $|r|$ exceda 0.999. Dado $r = -0.956$ calculado en el ejemplo anterior, si usamos un nivel de significancia de 0.05, concluimos que existe una correlación lineal entre x y y (porque $|-0.956|$ excede el valor crítico de 0.950). Sin embargo, si usamos un nivel de significancia de 0.01, no concluimos que existe una correlación lineal (porque $|-0.956|$ no excede el valor crítico de 0.999).

PROBLEMA 10 DEL CAPÍTULO

EJEMPLO Old Faithful Utilice los datos apareados de la duración y el intervalo de tiempo posterior a la erupción que aparecen en la tabla 10-1 para calcular el valor del coeficiente de correlación lineal r. Después remítase a la tabla A-6 para determinar si existe una correlación lineal significativa entre la duración y los intervalos de tiempo posteriores a las erupciones. En la tabla A-6, utilice el valor crítico para $\alpha = 0.05$. (Con $\alpha = 0.05$ concluimos que existe una correlación lineal significativa sólo si la muestra es improbable en este sentido: si no existe una correlación lineal entre dos variables, un valor de r como éste ocurre el 5% de las veces o menos).

SOLUCIÓN

REQUISITO ✔ Los datos provienen de erupciones elegidas al azar. El diagrama de dispersión que aparece en la figura 10-1a) muestra un patrón de puntos que se aproxima a una línea recta. Al parecer, no hay valores extremos. Los requisitos se satisfacen, por lo que procedemos a calcular r y a determinar si existe una correlación lineal entre la duración de los intervalos posteriores a las erupciones. ✔

Al utilizar el mismo procedimiento ilustrado en el ejemplo anterior o empleando algún recurso tecnológico, obtenemos que los ocho pares de datos duración/intervalo después de la erupción que aparecen en la tabla 10-1 dan como resultado $r = 0.926$. Éstos son los resultados de Minitab:

Minitab

```
Pearson correlation of Duration and Interval After = 0.926
P-Value = 0.001
```

Si nos remitimos a la tabla A-6, localizamos el renglón en el que $n = 8$ (porque hay 8 pares de datos). Este renglón contiene los valores críticos de 0.707 (para $\alpha = 0.05$) y 0.8345 (para $\alpha = 0.01$). Con el valor crítico para $\alpha = 0.05$, observamos que existe una probabilidad menor al 5% de que, sin correlación

continúa

lineal, el valor absoluto de r calculado exceda a 0.707. Puesto que $r = 0.926$, su valor absoluto excede a 0.707, por lo que concluimos que existe una correlación lineal entre la duración y los intervalos de tiempo posteriores a la erupción.

Ya hemos señalado que la fórmula 10-1 requiere que el valor calculado de r caiga siempre entre -1 y $+1$, inclusive. Consideramos esa propiedad, junto con otras igualmente importantes, en la siguiente lista.

Propiedades del coeficiente de correlación lineal r

1. El valor de r está siempre entre -1 y $+1$, inclusive. Es decir,
$$-1 \leq r \leq +1$$
2. *El valor de r no cambia si todos los valores de cualquiera de las variables se convierten a una escala diferente.*
3. *El valor de r no se ve afectado por la elección de x o y.* Intercambie todos los valores de x y y, y el valor de r no sufrirá cambios.
4. *r mide la fuerza de una relación lineal.* No está diseñada para medir la fuerza de una relación que no sea lineal.

Interpretación de r: Variación explicada

Si concluimos que existe una correlación lineal entre x y y, podemos obtener una ecuación lineal que exprese y en términos de x, y la ecuación puede emplearse para predecir valores de y a partir de valores dados de x. En la sección 10-3 describiremos un procedimiento para el cálculo de tales ecuaciones y mostraremos cómo predecir valores de y cuando se tienen valores dados de x. Pero un valor predicho de y no será necesariamente el resultado exacto porque, además de x, existen otros factores que afectan a y, como la variación aleatoria y otras características que no están incluidas en el estudio. En la sección 10-4 presentaremos los fundamentos y más detalles acerca de este principio importante:

El valor de r^2 es la proporción de la variación de y que está explicada por la relación lineal entre x y y.

PROBLEMA DEL CAPÍTULO 10

EJEMPLO Old Faithful Con los datos de la duración y los intervalos posteriores a la erupción que se presentan en la tabla 10-1, hemos encontrado que el coeficiente de correlación lineal es $r = 0.926$. ¿Qué proporción de la variación en los intervalos posteriores a la erupción puede explicarse por la variación en las duraciones?

SOLUCIÓN Con $r = 0.926$, obtenemos $r^2 = 0.857$.

INTERPRETACIÓN Concluimos que 0.857 (o aproximadamente el 86%) de la variación de los intervalos posteriores a la erupción puede explicarse por la relación lineal entre los tiempos de duración y los intervalos posteriores a la erupción. Esto implica que cerca del 14% de la variación de los intervalos posteriores a la erupción no pueden explicarse por los tiempos de duración.

Errores comunes en las correlaciones

Ahora identificamos tres de las fuentes más comunes de errores que se cometen al interpretar los resultados de correlaciones:

1. *Un error común es concluir que la correlación implica causalidad*. Con los datos muestrales de la tabla 10-1, podemos concluir que existe una correlación entre la duración y los intervalos de tiempo posteriores a la erupción, pero no podemos concluir que una mayor duración *cause* intervalos más prolongados posteriores a la erupción. Los intervalos de tiempo posteriores a la erupción pueden verse afectados por alguna otra variable interventora en los antecedentes. (Una **variable interventora** es aquella que afecta las variables que se estudian, pero que no está incluida en la investigación). Por ejemplo, la temperatura del aire exterior podría afectar la duración de una erupción y los intervalos posteriores a una erupción. Por lo tanto, la temperatura del aire exterior sería una variable interventora.
2. *Otro error proviene de los datos basados en promedios*. Los promedios eliminan la variación individual y pueden inflar el coeficiente de correlación. Un estudio produjo un coeficiente de correlación lineal de 0.4 para datos apareados que relacionaban el ingreso y la educación de individuos, pero el coeficiente de correlación lineal se convirtió en 0.7 cuando se utilizaron promedios regionales.
3. *Un tercer error implica la propiedad de linealidad*. Puede existir una relación entre x y y, aun cuando no haya una correlación lineal. Los datos presentados en la figura 10-2h) dan por resultado un valor de $r = -0.087$, lo que indica que no existe una correlación *lineal* entre las dos variables. Sin embargo, al observar la figura, con facilidad podemos ver que existe un patrón que refleja una relación *no lineal* muy fuerte. [La figura 10-2h) es un diagrama de dispersión que representa la relación entre la distancia del suelo hacia arriba y el tiempo transcurrido para un objeto lanzado hacia arriba].

Parte 2: Prueba formal de hipótesis (requiere el estudio del capítulo 8)

Presentamos dos métodos (resumidos en el siguiente recuadro y en la figura 10-3) para utilizar una prueba formal de hipótesis con el fin de determinar si existe una correlación lineal significativa entre dos variables. Algunos profesores prefieren el método 1 porque refuerza conceptos estudiados en capítulos anteriores. Otros prefieren el método 2 porque supone cálculos más fáciles. El método 1 emplea la distribución t de Student, con un estadístico de prueba con la forma $t = r/s_r$, donde s_r denota la desviación estándar muestral de valores de r. El estadístico de prueba incluido en el recuadro (para el método 1) refleja el hecho de que la desviación estándar de los valores de r puede expresarse como $\sqrt{(1 - r^2)/(n - 2)}$.

La figura 10-3 indica que el criterio de decisión es el rechazo de la hipótesis nula de $\rho = 0$, si el valor absoluto del estadístico de prueba excede los valores críticos; el rechazo de $\rho = 0$ significa que existe evidencia suficiente para sustentar una aseveración de una correlación lineal entre las dos variables. Si el valor absoluto del estadístico de prueba no excede los valores críticos, entonces no rechazamos $\rho = 0$; es decir, no existe suficiente evidencia para concluir que existe una correlación lineal entre las dos variables.

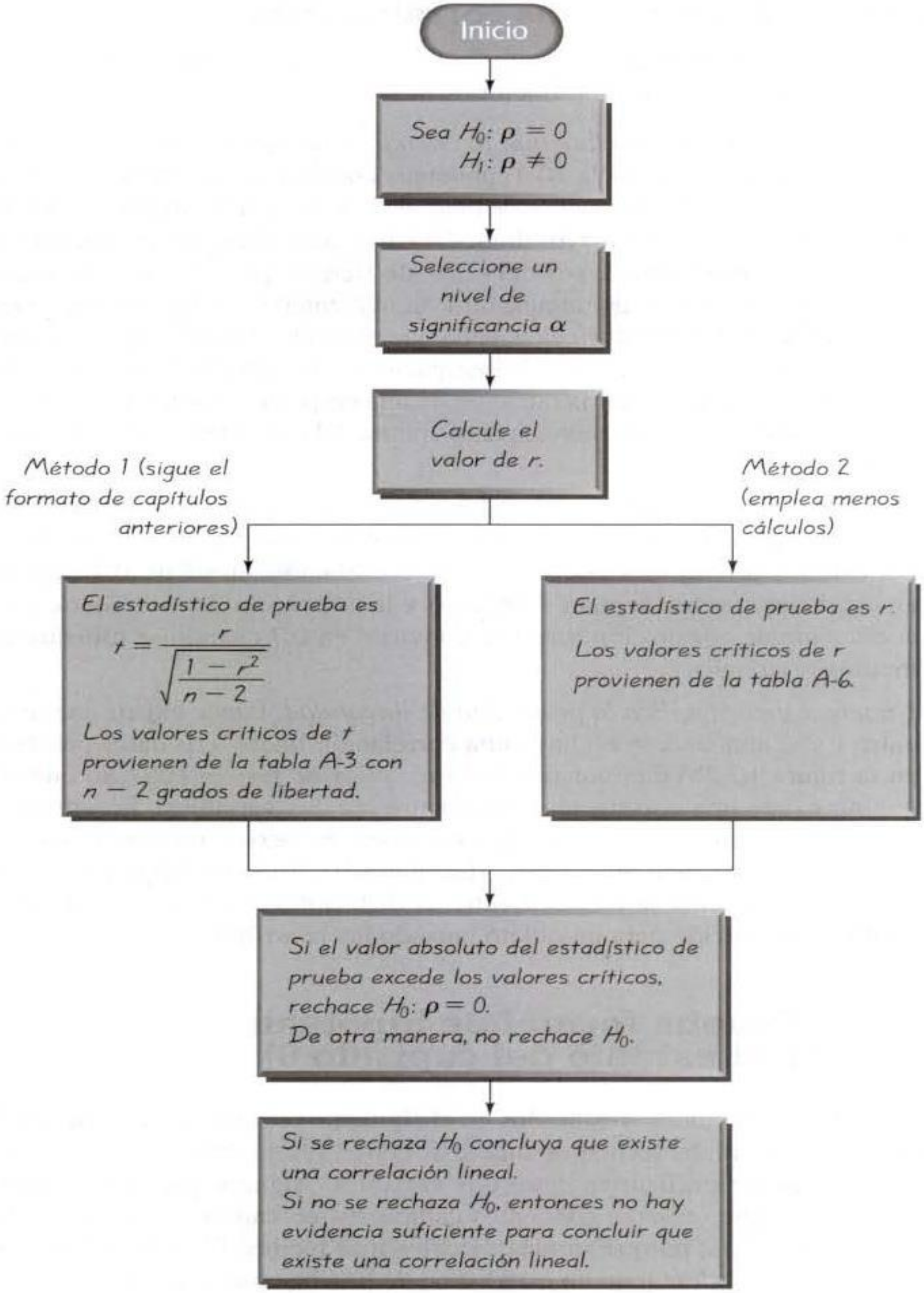


Figura 10-3 Prueba de hipótesis para una correlación lineal

Prueba de hipótesis de correlación (véase la figura 10-3).

H_0: $\rho = 0$ (No existe una correlación lineal).

H_1: $\rho \neq 0$ (Existe una correlación lineal).

Método 1: El estadístico de prueba es *t*

Estadístico de prueba: $$t = \frac{r}{\sqrt{\dfrac{1 - r^2}{n - 2}}}$$

Valores críticos: Utilice la tabla A-3 con $n - 2$ grados de libertad.

Valor *P*: Utilice la tabla A-3 con $n - 2$ grados de libertad.

Conclusión: Si $|t| >$ el valor crítico de la tabla A-3, rechace H_0 y concluya que existe una correlación lineal. Si $|t| \leq$ valor crítico, no rechace H_0; no hay evidencia suficiente para concluir que existe una correlación lineal.

Método 2: El estadístico de prueba es *r*

Estadístico de prueba: r

Valores críticos: Remítase a la tabla A-6.

Conclusión: Si $|r| >$ el valor crítico de la tabla A-6, rechace H_0 y concluya que existe una correlación lineal. Si $|r| \leq$ valor crítico, no rechace H_0; no hay evidencia suficiente para concluir que existe una correlación lineal.

PROBLEMA 10 DEL CAPÍTULO

EJEMPLO Old Faithful Con los datos muestrales de la tabla 10-1 sobre la duración y los intervalos posteriores a las erupciones, pruebe la aseveración de que existe una correlación lineal entre la duración de una erupción y el intervalo de tiempo posterior a la erupción. Para obtener el estadístico de prueba utilice *a*) el método 1 y *b*) el método 2.

SOLUCIÓN

REQUISITO ✔ Un ejemplo previo incluye la verificación de los requisitos. [Los datos muestrales se seleccionaron al azar; el diagrama de dispersión en la figura 10-1a) indica un patrón de puntos de apariencia lineal, sin valores extremos]. Una vez completada esta verificación simplificada de requisitos, procedemos con nuestro análisis. ✔

Esta solución seguirá el procedimiento que se resume en la figura 10-3. Afirmar que existe una correlación lineal equivale a aseverar que el coeficiente de correlación lineal poblacional ρ es diferente de 0. Por lo tanto, tenemos las siguientes hipótesis:

H_0: $\rho = 0$ (No existe una correlación lineal).

H_1: $\rho \neq 0$ (Existe una correlación lineal).

Puesto que no se especificó un nivel de significancia, utilice $\alpha = 0.05$.

continúa

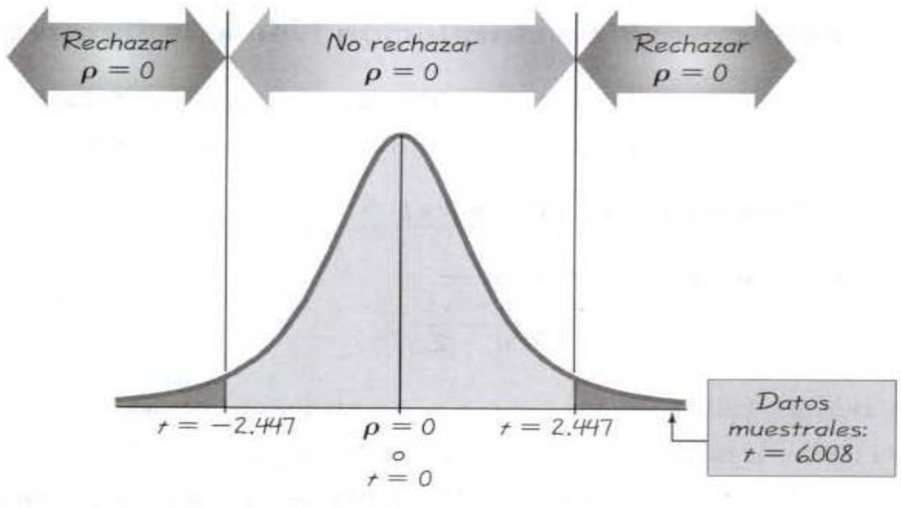


Figura 10-4 **Prueba de H_0: $\rho = 0$ con el método 1**

En un ejemplo previo ya calculamos que $r = 0.926$. Con ese valor ahora calculamos el estadístico de prueba y el valor crítico por medio de los dos métodos descritos.

a. *Método 1:* El estadístico de prueba es

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{0.926}{\sqrt{\frac{1 - 0.926^2}{8 - 2}}} = 6.008$$

Los valores críticos de $t = \pm 2.447$ se encuentran en la tabla A-3, donde 2.447 corresponde a una área de 0.05, dividida entre dos colas, y el número de grados de libertad es $n - 2 = 6$. Observe en la figura 10-4 la gráfica que incluye el estadístico de prueba y los valores críticos.

b. *Método 2:* El estadístico de prueba es $r = 0.926$. Los valores críticos de $r = \pm 0.707$ se encuentran en la tabla A-6, con $n = 8$ y $\alpha = 0.05$. Observe en la figura 10-5 la gráfica que incluye el estadístico de prueba y los valores críticos.

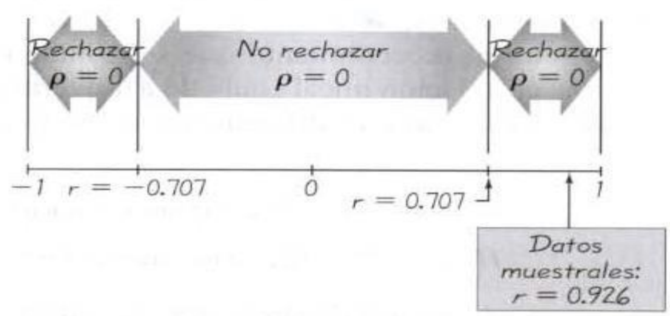


Figura 10-5 **Prueba de H_0: $\rho = 0$ con el método 2**

Al utilizar de cualquiera de los dos métodos, encontramos que el valor absoluto del estadístico de prueba excede el valor crítico (Método 1: $6.008 > 2.447$. Método 2: $0.926 > 0.707$); es decir, el estadístico de prueba cae en la región crítica. Por lo tanto, rechazamos H_0: $\rho = 0$. Existe evidencia suficiente para sustentar la aseveración de una correlación lineal entre las duraciones de las erupciones y los intervalos posteriores a éstas.

Pruebas de una cola: El ejemplo anterior y las figuras 10-4 y 10-5 ilustran una prueba de hipótesis de dos colas. En general, los ejemplos y ejercicios de esta sección implicarán únicamente pruebas de dos colas, pero puede presentarse una prueba de una cola en una aseveración de una correlación lineal positiva o una aseveración de una correlación lineal negativa. En estos casos, las hipótesis serán como las que se presentan a continuación.

Aseveración de correlación *negativa* (prueba de cola izquierda)	**Aseveración de correlación *positiva*** (prueba de cola derecha)
H_0: $\rho = 0$	H_0: $\rho = 0$
H_1: $\rho < 0$	H_1: $\rho > 0$

Con estas pruebas de una cola, el método 1 puede realizarse como se hizo en capítulos anteriores. En el caso del método 2, calcule el valor crítico como se describió en el ejercicio 38, o bien, modifique la tabla A-6 reemplazando los encabezados de columna de $\alpha = 0.05$ y $\alpha = 0.01$ por los niveles unilaterales de significancia de $\alpha = 0.025$ y $\alpha = 0.005$, respectivamente.

Fundamentos: Ya presentamos la fórmula 10-1 para el cálculo de r e ilustramos su uso. A continuación se presenta la fórmula 10-1 junto con algunas otras fórmulas que son "equivalentes" en el sentido de que todas producen los mismos valores. Los distintos autores de los libros prefieren expresiones diferentes por diversas razones, y las fórmulas que se presentan abajo son las expresiones de r más utilizadas. Estas fórmulas simplemente son versiones diferentes de la misma expresión. (Invitamos a los amantes del álgebra a divertirse probando la equivalencia de las fórmulas; los demás pueden confiar en la afirmación del autor de que las diversas fórmulas para r producen los mismos resultados). El formato de la fórmula 10-1 simplifica tanto los cálculos manuales como los que se realizan con hojas de cálculo, y también es fácil de incluir en un programa de cómputo original.

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n(\Sigma x^2) - (\Sigma x)^2}\sqrt{n(\Sigma y^2) - (\Sigma y)^2}} \qquad r = \frac{\Sigma\left[\dfrac{(x-\bar{x})}{s_x}\dfrac{(y-\bar{y})}{s_y}\right]}{n-1}$$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y} \qquad r = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}}$$

Aun cuando la fórmula 10-1 simplifica los cálculos manuales, otras fórmulas para r son mejores para *entender* la manera en que funciona r. Para tratar de comprender los fundamentos que subyacen en el desarrollo del coeficiente de correlación lineal, usaremos la siguiente versión del coeficiente de correlación lineal:

$$r = \frac{\sum \left[\frac{(x - \bar{x})}{s_x} \frac{(y - \bar{y})}{s_y} \right]}{n - 1}$$

De manera temporal utilizaremos esta última versión de la fórmula 10-1, ya que se relaciona de manera más directa con la teoría subyacente. Ahora considere los siguientes datos apareados, que están representados en el diagrama de dispersión de la figura 10-6.

x	1	1	2	4	7
y	4	5	8	15	23

La figura 10-6 incluye el punto $(\bar{x}, \bar{y}) = (3, 11)$, denominado el *centroide* de los puntos muestrales.

Definición

Dado un conjunto de datos apareados (x, y), el punto $(\bar{x}, \bar{y})$ se denomina **centroide**.

El estadístico r, que en ocasiones se llama *producto momento de Pearson*, fue creado por Karl Pearson. Se basa en la sumatoria de los productos $(x - \bar{x})(y - \bar{y})$. El estadístico $\Sigma(x - \bar{x})(y - \bar{y})$ es la expresión básica del fundamento de r, y ahora explicaremos por qué el estadístico es la clave.

En cualquier diagrama de dispersión, las líneas vertical y horizontal que pasan a través del centroide $(\bar{x}, \bar{y})$ dividen el diagrama en cuatro cuadrantes, como se ob-

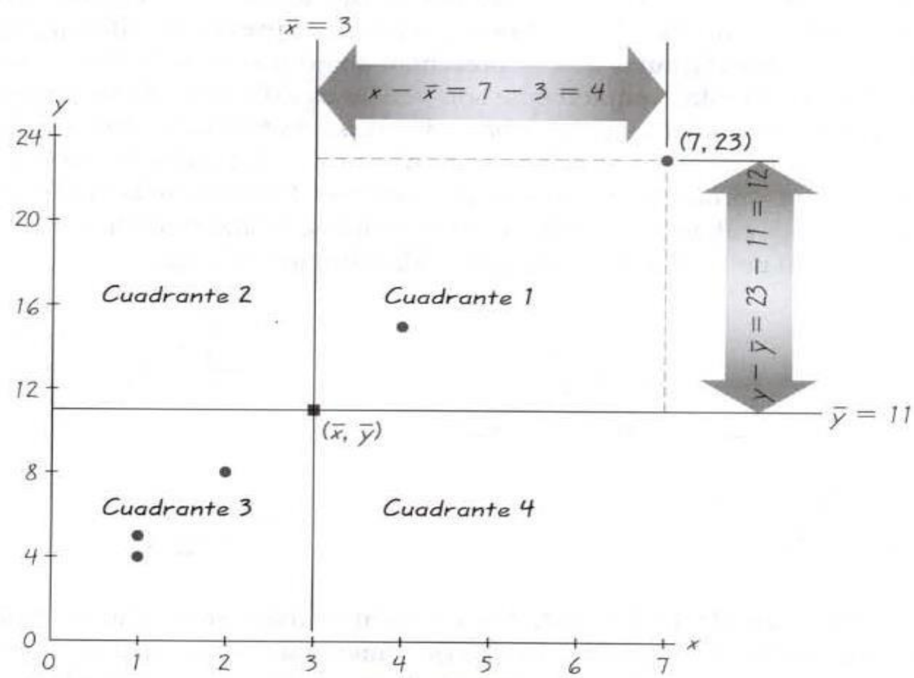


Figura 10-6 Diagrama de dispersión dividido en cuadrantes

serva en la figura 10-6. Si los puntos del diagrama de dispersión tienden a aproximarse a una línea ascendente (como en la figura), los valores individuales del producto $(x - \bar{x})(y - \bar{y})$ tienden a ser positivos ya que la mayoría de los puntos se encuentran en el primer y tercer cuadrantes, donde los productos de $(x - \bar{x})$ y $(y - \bar{y})$ son positivos. Si los puntos del diagrama de dispersión se aproximan a una línea descendente, la mayoría de los puntos se encuentran en el segundo y cuarto cuadrantes, donde $(x - \bar{x})$ y $(y - \bar{y})$ tienen el signo opuesto, de manera que $\Sigma(x - \bar{x})(y - \bar{y})$ es negativo. Los puntos que no siguen un patrón lineal tienden a dispersarse en los cuatro cuadrantes, de manera que el valor de $\Sigma(x - \bar{x})(y - \bar{y})$ tiende a 0. Por lo tanto, podemos usar $\Sigma(x - \bar{x})(y - \bar{y})$ para medir el arreglo de los puntos. Una suma positiva grande sugiere que los puntos están predominantemente en el primer y en el tercer cuadrantes (lo que corresponde a una correlación lineal positiva); una suma negativa grande sugiere que los puntos están predominantemente en el segundo y en el cuarto cuadrantes (lo que corresponde a una correlación lineal negativa), y una suma cercana a cero sugiere que los puntos se dispersan en los cuatro cuadrantes (y no existe una correlación lineal).

Por desgracia, la suma $\Sigma(x - \bar{x})(y - \bar{y})$ depende de la magnitud de los números utilizados. Por ejemplo, si se convierte x de pulgadas a pies, la suma cambiará. Para lograr que r sea independiente de la escala utilizada, incluimos la desviación estándar muestral. En la sección 3-4 aprendimos que podemos "estandarizar" los valores al convertirlos en puntuaciones z, como en $z = (x - \bar{x})/s$. (Aquí utilizamos s_x para denotar la desviación estándar de los valores muestrales x, y utilizamos s_y para denotar la desviación estándar de los valores muestrales y). Empleamos una técnica similar al estandarizar cada desviación $(x - \bar{x})$ dividiéndola entre s_x. También logramos que las desviaciones $(y - \bar{y})$ sean independientes de las magnitudes de los números al dividirlas entre s_y. Ahora tenemos el estadístico

$$\Sigma\left[\frac{(x - \bar{x})}{s_x}\frac{(y - \bar{y})}{s_y}\right]$$

que modificamos aún más al introducir el divisor de $n - 1$, que nos da un tipo de promedio en vez de una suma que aumenta simplemente porque tenemos más datos. (Las razones para dividir entre $n - 1$ y no entre n son básicamente las mismas relacionadas con la desviación estándar). El resultado final es la expresión

$$r = \frac{\Sigma\left[\frac{(x - \bar{x})}{s_x}\frac{(y - \bar{y})}{s_y}\right]}{n - 1}$$

Esta expresión puede manipularse algebraicamente en la forma equivalente de la fórmula 10-1 o cualquiera de las otras expresiones para r.

Intervalos de confianza En capítulos anteriores estudiamos métodos de estadística inferencial al utilizar métodos de prueba de hipótesis y métodos para construir estimados de intervalos de confianza. Se puede emplear un procedimiento similar para calcular intervalos de confianza para ρ. Sin embargo, como la construcción de tales intervalos de confianza implica transformaciones que son hasta cierto punto complicadas, ese proceso se presenta en el ejercicio 39.

Podemos utilizar el coeficiente de correlación lineal para determinar si existe una relación lineal entre dos variables. Con los datos de la tabla 10-1 sobre la duración y los intervalos posteriores a las erupciones, concluimos que existe una correlación

10-3 Regresión

Concepto clave La sección 10-2 nos dio las herramientas para determinar si existe una correlación lineal entre dos variables. El concepto clave de esta sección es describir la relación entre dos variables por medio del cálculo de la gráfica y la ecuación de la recta que representa mejor la relación. Esta recta se conoce como *recta de regresión* y su ecuación como *ecuación de regresión*. A partir de datos muestrales apareados, calcularemos valores estimados de b_0, que es la intersección en y, y la pendiente b_1, de manera que podamos identificar una línea recta con la ecuación $\hat{y} = b_0 + b_1x$. En condiciones adecuadas, esa ecuación resulta útil para hacer predicciones. Existen programas de cómputo y calculadoras para realizar los cálculos aritméticos que son hasta cierto punto engorrosos, de manera que nos enfocaremos en entender los conceptos más que en procesar los datos numéricos. Al igual que en la sección 10-2, esta sección se divide en dos partes: **1.** Conceptos básicos de regresión; **2.** Más allá de los conceptos básicos de regresión. La primera parte incluye conceptos fundamentales que deben quedar muy claros antes de pasar a la segunda parte.

Parte 1: Conceptos básicos de regresión

En algunos casos, dos variables están relacionadas de una forma *determinista*, es decir, dado un valor de una variable, el valor de la otra variable se determina automáticamente sin error. Por ejemplo, el costo total y de un artículo con un precio de lista x y un impuesto de venta del 5% se calcula utilizando la ecuación determinista $y = 1.05x$. Si un artículo tiene un precio de \$50, su costo total será de \$52.50. Este tipo de funciones se estudian ampliamente en los cursos de álgebra. En este capítulo estamos más interesados en los modelos *probabilísticos*, en los que una variable no está determinada por completo por la otra variable. Por ejemplo, la estatura de un niño no está completamente determinada por la estatura del padre (o de la madre). Sir Francis Galton (1822-1911) estudió el fenómeno de la herencia y demostró que cuando parejas altas o bajas tienen hijos, las estaturas de éstos tienden a *regresar* o a revertirse a la estatura media más común de las personas del mismo género. Continuaremos utilizando la terminología de "regresión" de Galton, aun cuando nuestros datos no incluyen el mismo fenómeno de estatura estudiado por Galton.

El recuadro de la página 544 incluye la definición de la ecuación de regresión y de la recta de regresión, así como la notación y las fórmulas que estamos utilizando. La ecuación de regresión expresa una relación entre x (llamada **variable explicativa**, **variable de predicción** o **variable independiente**) y $\hat{y}$ (llamada **variable de respuesta** o **variable dependiente**). La ecuación típica de una línea recta $y = mx + b$ está expresada en la forma $\hat{y} = b_0 + b_1x$, donde b_0 es el intercepto y (o la intersección en y) y b_1 es la pendiente. La notación dada muestra que b_0 y b_1 son estadísticos muestrales utilizados para estimar los parámetros poblacionales β_0 y β_1. Emplearemos datos muestrales apareados para estimar la ecuación de regresión. Si utilizamos únicamente datos muestrales no podemos calcular los valores exactos de los parámetros poblacionales β_0 y β_1, pero podemos emplear los datos muestrales para estimarlos con b_0 y b_1, que se calculan con las fórmulas 10-2 y 10-3.

LA ESTADÍSTICA EN LAS NOTICIAS

Error de pronóstico de 1°F = mil millones de dólares

A pesar de que el pronóstico de las temperaturas puede parecer una ciencia inexacta, muchas compañías están trabajando con fervor para obtener estimaciones más exactas. El reportero de *USA Today*, Del Jones, escribió que "el costo anual de la electricidad podría disminuir por lo menos $1000 millones si se mejorara la exactitud de las predicciones del tiempo en 1 grado Fahrenheit". Al referirse a las autoridades de Tennessee Valley, afirma que "los pronósticos sobre sus 80,000 millas cuadradas han fallado un promedio de 2.35 grados durante los últimos 2 años, que es bastante representativo de los pronósticos que se hacen a nivel nacional. Si se mejorara en 1.35 grados, esto ahorraría al Tennessee Valley tanto como $100,000 diarios y tal vez más". El pronóstico de temperaturas se utiliza para determinar la distribución de la energía proveniente de generadores, plantas nucleares, plantas hidroeléctricas, de carbón, de gas natural y eólicas. Las técnicas de pronóstico estadístico se encuentran en proceso de refinamiento, de manera que se pueda ahorrar dinero y recursos naturales.

Requisitos

1. La muestra de datos apareados (x, y) es una muestra *aleatoria* de datos cuantitativos.
2. El examen visual del diagrama de dispersión indica que los puntos se aproximan al patrón de una línea recta.
3. Se debe eliminar cualquier valor extremo, si se sabe que es un error. Es importante tomar en cuenta los efectos de cualquier valor extremo que no sea un error conocido.

Nota: Los requisitos 2 y 3 representan una verificación simplificada de los siguientes requisitos formales del análisis de regresión:

- Para cada valor fijo de x, los valores correspondientes de y tienen una distribución en forma de campana.
- Para los distintos valores fijos de x, las distribuciones de los valores correspondientes de y tienen la misma varianza. (Esto se viola si parte del diagrama de dispersión presenta puntos muy cercanos a la línea de regresión, mientras otra porción del diagrama presenta puntos que se alejan mucho de la línea de regresión. Consulte la explicación de los puntos residuales casi al final de esta sección).
- Para los distintos valores fijos de x, las distribuciones de los valores correspondientes de y tienen medias que se ubican en la misma línea recta.
- Los valores de y son independientes.

Los resultados no se ven muy afectados si la distribución no se aleja demasiado de la normalidad y si las varianzas no son demasiado diferentes.

Definiciones

Dado un conjunto de datos muestrales apareados, la **ecuación de regresión**

$$\hat{y} = b_0 + b_1 x$$

describe algebraicamente la relación entre las dos variables. La gráfica de la ecuación de regresión se denomina **recta de regresión** (o *recta del mejor ajuste o recta de mínimos cuadrados*).

Notación para la ecuación de regresión

	Parámetro poblacional	Estadístico muestral
Intercepto y de la ecuación de regresión	β_0	b_0
Pendiente de la ecuación de regresión	β_1	b_1
Ecuación de la recta de regresión	$y = \beta_0 + \beta_1 x$	$\hat{y} = b_0 + b_1 x$

Cálculo de la pendiente b_1 y de b_0 (el intercepto y) en la ecuación de regresión $\hat{y} = b_0 + b_1 x$

Fórmula 10-2 **Pendiente:** $$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Fórmula 10-3 **intercepto y:** $$b_0 = \bar{y} - b_1\bar{x}$$

El intercepto y, b_0, también puede calcularse por medio de la siguiente fórmula, pero es mucho más fácil utilizar la fórmula 10-3.

$$b_0 = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Tal vez las fórmulas 10-2 y 10-3 resulten intimidantes, pero están programadas en muchas calculadoras y programas de cómputo, de manera que los valores de b_0 y b_1 se calculan con facilidad. (Véase el recuadro "Uso de la tecnología" al final de esta sección). Una vez que hemos evaluado b_1 y b_0, podemos identificar la ecuación estimada de regresión, que tiene la siguiente propiedad especial: *La recta de regresión es la que se ajusta mejor a los puntos muestrales.* (El criterio específico utilizado para determinar cuál recta se ajusta "mejor" es la propiedad de los mínimos cuadrados, que se describirá más adelante). Ahora estudiaremos brevemente el redondeo y después ejemplificaremos el procedimiento del cálculo y la aplicación de la ecuación de regresión.

Redondeo de la pendiente b_1 y de b_0 (el intercepto y)

Redondeo de b_1 y b_0 a tres dígitos significativos. Es difícil dar una regla universal sencilla para redondear los valores de b_1 y b_0, pero esta regla servirá en la mayor parte de las situaciones de este libro. Dependiendo de la forma de redondeo, las respuestas a los ejemplos y ejercicios de este libro pueden variar un poco de las respuestas de usted.

EJEMPLO Cálculo de la ecuación de regresión En la sección 10-2 empleamos la muestra aleatoria simple de los valores listados abajo para calcular el coeficiente de correlación lineal de $r = -0.956$. (Al usar los métodos de la sección 10-2, concluimos que existe una correlación lineal entre x y y). Use los datos muestrales indicados para calcular la ecuación de regresión.

x	3	1	3	5
y	5	8	6	4

SOLUCIÓN

REQUISITO ✔ Se trata de una muestra aleatoria simple. El diagrama de dispersión generado por SPSS, que se presenta en la sección 10-2, sugiere un patrón de puntos similar al patrón de una recta. No hay valores extremos. Procedemos a calcular la pendiente y el intercepto de la recta de regresión. ✔

Calculamos la ecuación de regresión por medio de las fórmulas 10-2 y 10-3, y los siguientes valores ya se encuentran en la tabla 10-2 de la sección 10-2:

$$n = 4 \qquad \Sigma x = 12 \qquad \Sigma y = 23$$
$$\Sigma x^2 = 44 \qquad \Sigma y^2 = 141 \qquad \Sigma xy = 61$$

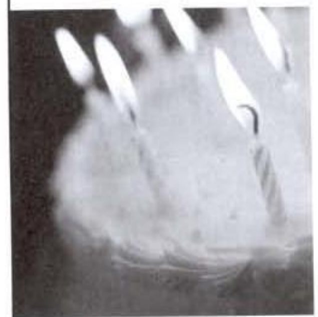
Primero calcule b_1 usando la fórmula 10-2:

$$b_1 = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$
$$= \frac{4(61) - (12)(23)}{4(44) - (12)^2} = \frac{-32}{32} = -1$$

Luego, encuentre b_0, el intercepto en y, utilizando la fórmula 10-3 (con $\bar{y} = 23/4 = 5.75$ y $\bar{x} = 12/4 = 3$):

$$b_0 = \bar{y} - b_1\bar{x}$$
$$= 5.75 - (-1)(3) = 8.75$$

continúa



Posposición de la muerte

Varios estudios tratan sobre la capacidad de las personas para retrasar su muerte hasta después de un suceso importante. Por ejemplo, el sociólogo David Phillips analizó las tasas de mortalidad de hombres judíos que murieron cerca de la Pascua judía, y descubrió que la tasa de mortalidad disminuía drásticamente una semana antes de esa festividad, pero que aumentaba la semana posterior a ella. Un estudio más reciente sugiere que las personas no tienen esta capacidad de posponer la muerte. Con base en los registros de 1.3 millones de muertes, este estudio no encontró una relación entre el momento de la muerte y la Navidad, el Día de Acción de Gracias o el cumpleaños del individuo. El doctor Donn Young, uno de los investigadores, dijo que "el hecho es que, la muerte no toma en cuenta el calendario. Usted no puede anotar en su Palm Pilot 'Bien, cenemos el viernes y agendaré mi muerte para el domingo'". David Phillips rebatió estos resultados y dijo que el estudio se enfocó en pacientes con cáncer, quienes tienen menos probabilidades de presentar efectos psicosomáticos.

Conociendo la pendiente b_1 y b_0 (el intercepto y), ahora podemos expresar la ecuación estimada de la recta de regresión como

$$\hat{y} = 8.75 - 1x$$

Debemos estar conscientes de que esta ecuación es un *estimado* de la verdadera ecuación de regresión $y = \beta_0 + \beta_1 x$. Este estimado se basa en un conjunto particular de datos muestrales, pero otra muestra obtenida de la misma población quizá generaría una ecuación ligeramente diferente.

PROBLEMA 10 DEL CAPÍTULO

EJEMPLO Old Faithful Con los datos de las duraciones y los intervalos posteriores a las erupciones de la tabla de 10-1, calculamos que el coeficiente de correlación lineal es $r = 0.926$. Utilice los mismos datos muestrales y calcule la recta de regresión.

SOLUCIÓN

REQUISITO ✔ La figura 10-1a) es un diagrama de dispersión de esos puntos, el cual muestra un patrón que se aproxima a una línea recta. No hay valores extremos y los datos muestrales se obtuvieron al azar. Ahora podemos continuar con el cálculo de la ecuación de la recta de regresión. ✔

Si utilizamos el mismo procedimiento ilustrado en el ejemplo anterior o empleamos herramientas tecnológicas, podemos calcular que los 8 pares de datos duración/intervalos posteriores a la erupción de la tabla 10-1 dan como resultado $b_0 = 34.8$ y $b_1 = 0.234$. Abajo se presentan los resultados de Minitab. Sustituyendo los valores calculados para b_0 y b_1, expresamos la ecuación de regresión como $\hat{y} = 34.8 + 0.234x$. A continuación se presenta también el diagrama de dispersión generado por Minitab, con la recta de regresión incluida. Podemos ver que la recta de regresión se ajusta bien a los datos.

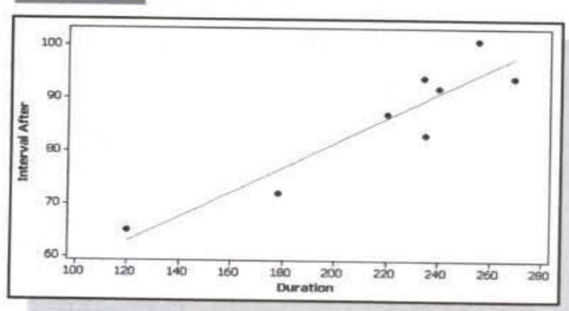
Minitab

```
The regression equation is
Interval After = 34.8 + 0.234 Duration

Predictor      Coef   SE Coef     T      P
Constant     34.770     8.732  3.98  0.007
Duration    0.23406   0.03908  5.99  0.001

S = 4.97392   R-Sq = 85.7%   R-Sq(adj) = 83.3%
```

Minitab



Uso de la ecuación de regresión para hacer predicciones

Las ecuaciones de regresión a menudo se utilizan para *predecir* el valor de una variable, dado algún valor particular de la otra variable. Si la recta de regresión se ajusta bastante bien a los datos, entonces es sensato utilizar su ecuación para hacer predicciones, siempre y cuando no vayamos más allá del alcance de los valores disponibles. No haga predicciones con base en valores que rebasen las fronteras de los datos muestrales conocidos. Por ejemplo, si utilizamos los datos muestrales de la tabla 10-1, un diagrama de dispersión sugiere que las duraciones y los intervalos de tiempo posteriores a las erupciones se ajustan bastante bien al patrón de una línea recta, de manera que si observamos una erupción de 180 segundos, podemos predecir el intervalo posterior a la erupción (hasta la siguiente erupción) al sustituir $x = 180$ en la ecuación de regresión $\hat{y} = 34.8 + 0.234x$. El siguiente resultado indica que si una erupción tiene una duración de 180 segundos, el mejor intervalo de tiempo predicho después de la erupción es de 76.9 minutos. (Recuerde, las duraciones están en segundos y los intervalos posteriores a las erupciones están en minutos).

$$\hat{y} = 34.8 + 0.234x$$
$$= 34.8 + 0.234(180) = 76.9 \text{ min} \quad \textbf{(redondeado)}$$

Sin embargo, *debemos utilizar la ecuación de la recta de regresión sólo si la ecuación de regresión es un buen modelo para los datos.* La adecuación de la ecuación de regresión se puede juzgar al probar la significancia del coeficiente de correlación lineal r. Observe el siguiente procedimiento en el que se utiliza la ecuación de regresión para hacer predicciones. Vea también el siguiente ejemplo, que formaliza el procedimiento para utilizar una duración de 180 segundos para predecir el intervalo posterior a la erupción.

Parte 2: Más allá de los conceptos básicos de regresión

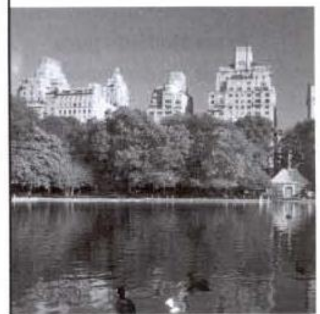
Uso de la ecuación de regresión para hacer predicciones (*continuación*)

Ya señalamos que debemos utilizar la ecuación de la recta de regresión para hacer predicciones sólo si la ecuación de regresión es un buen modelo para los datos. Para ser más precisos, *debemos utilizar la ecuación de regresión para hacer predicciones sólo si existe una correlación lineal. En ausencia de una correlación lineal, no debemos usar la ecuación de regresión para proyectar o predecir; en vez de ello, nuestro mejor estimado de la segunda variable es simplemente su media muestral.*

Al predecir un valor de y con base en algún valor dado de x . . .

1. **Si *no* existe una correlación lineal, el mejor valor predicho de y es $\bar{y}$.**
2. **Si existe una correlación lineal, el mejor valor predicho de y se calcula sustituyendo el valor de x en la ecuación de regresión.**

¿Qué tan buena es la ecuación de regresión como modelo para los datos poblacionales? En la figura 10-7 se resume el proceso para hacer predicciones, el cual se comprende con mayor facilidad si pensamos en r como una medida de lo bien que se ajusta la recta de regresión a los datos muestrales. Además de reflexionar sobre la presencia o ausencia de una correlación lineal, también debemos considerar lo siguiente: las ecuaciones de regresión obtenidas de datos muestrales apareados con r muy cercana a -1 o $+1$ (porque los puntos en el diagrama de dispersión se acercan mucho a la recta de regresión) tienen mayores probabilidades de ser mejores modelos para los datos poblacionales que las ecuaciones de regresión de conjuntos de datos con valores de r que no son tan cercanos a -1 o $+1$ (aun cuando r resulte significativa). Con algunas muestras muy grandes de datos apareados, podríamos encontrar que r es significativa, aun cuando tenga un valor relativamente pequeño como 0.200. En este caso, la ecuación de regresión podría ser un modelo aceptable, pero las predicciones no serían tan exactas porque r no se acerca tanto a -1 o $+1$. Si r no es significativa, entonces la recta de regresión no se ajusta bien a los datos, y la ecuación de regresión no debe utilizarse para hacer predicciones.

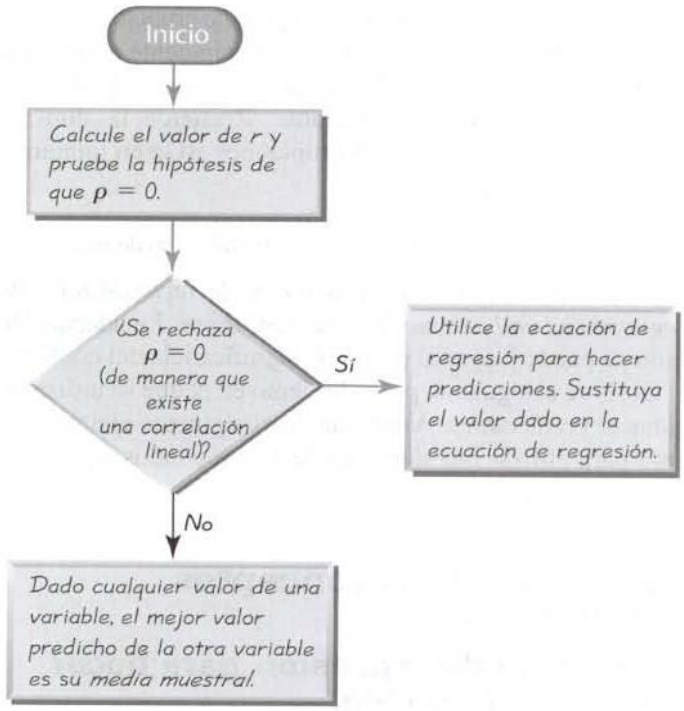


Predicción de precios de condominios

Un estudio masivo incluyó 99,491 operaciones de venta de condominios y cooperativas en Manhattan. El estudio utilizó 41 variables diferentes para predecir el valor del condominio o de la cooperativa. Las variables incluían la condición de la vivienda, el vecindario, la antigüedad, el tamaño y si contaba con portero. He aquí algunas conclusiones: si todos los factores permanecen igual, un condominio vale 15.5% más que una cooperativa; una chimenea incrementa el valor de un condominio en un 9.69% e incrementa el valor de una cooperativa en un 11.36%; una recámara adicional en un condominio aumenta su valor en un 7.11% y en una cooperativa en un 18.53%. Este uso de los métodos estadísticos permite a los compradores y a los vendedores estimar el valor con mucha mayor exactitud. Los métodos de regresión múltiple (sección 10-5) se utilizan cuando existe más de una variable de predicción, como en este estudio. (Según datos de "So How Much Is That . . . Worth", de Dennis Havesi, *New York Times*).

Figura 10-7

Procedimiento para hacer predicciones



PROBLEMA 10 DEL CAPÍTULO

EJEMPLO Predicción de las erupciones del Old Faithful Con los datos de las duraciones y los intervalos posteriores a las erupciones de la tabla de 10-1, encontramos que existe una correlación lineal entre las dos variables y también que la ecuación de regresión es $\hat{y} = 34.8 + 0.234x$. Suponiendo que la erupción actual tiene una duración de $x = 180$ segundos, calcule el mejor valor de predicción de y, el intervalo posterior a esta erupción (que es el tiempo predicho hasta la siguiente erupción).

SOLUCIÓN

REQUISITO ✔ Quizá nos sintamos tentados a apresurarnos y sustituir x por 180 en la ecuación de regresión, pero primero debemos considerar si existe una correlación lineal que justifique el uso de esa ecuación. En este ejemplo, el diagrama de dispersión indica que los puntos se aproximan al patrón de una línea recta, por lo que sí tenemos una correlación lineal (con $r = 0.926$), de manera que el valor predicho se calcula como se indica a continuación. ✔

Puesto que existe una correlación lineal, el valor predicho se calcula sustituyendo en la ecuación de regresión como sigue:

$$\begin{aligned}\hat{y} &= 34.8 + 0.234x \\ &= 34.8 + 0.234(180) = 76.9 \text{ min} \quad \textbf{(redondeado)}\end{aligned}$$

El intervalo de tiempo predicho desde la erupción actual hasta la siguiente erupción es de 76.9 minutos. (Si no hubiera una correlación lineal, el mejor valor predicho sería $\bar{y} = 688/8 = 86.0$ min).

EJEMPLO Medida de sombrero y CI Obviamente no existe una correlación lineal entre la medida de sombrero de las personas y sus puntuaciones de CI. Dado que un individuo utiliza un sombrero de tamaño 7, calcule el mejor valor predicho de la puntuación de CI de esta persona.

SOLUCIÓN
REQUISITO ✔ Puesto que no existe una correlación lineal, no empleamos la ecuación de la recta de regresión para hacer predicciones. ✔

Puesto que no hay una correlación lineal, no usamos una ecuación de regresión. No hay necesidad de reunir datos muestrales apareados consistentes en las medidas de sombrero y las puntuaciones de CI de una muestra de adultos seleccionados al azar. En vez de ello, la mejor puntuación de CI predicha es sencillamente la media del CI de todos los adultos, que es de 100.

Compare con cuidado las soluciones de los dos ejemplos anteriores y note que utilizamos la ecuación de regresión cuando existía una correlación lineal; sin embargo, en ausencia de esa correlación, el mejor valor predicho de y es sencillamente el valor de la media muestral $\bar{y}$. Un error común es el uso de la ecuación de regresión para hacer una predicción cuando no existe una correlación lineal. Este error viola el primero de los siguientes lineamientos.

Lineamientos para el uso de la ecuación de regresión

1. *Si no existe una correlación lineal, no utilice la ecuación de regresión para hacer predicciones.*
2. *Cuando utilice la ecuación de regresión para hacer predicciones, permanezca en el ámbito de los datos muestrales disponibles.* Si usted calcula una ecuación de regresión que relaciona la estatura y el número de calzado de mujeres, es absurdo predecir el número de calzado de una mujer que mide 10 pies de estatura.
3. *Una ecuación de regresión que está basada en datos antiguos no necesariamente es válida ahora.* La ecuación de regresión que relaciona precios de automóviles usados con la antigüedad de los automóviles ya no es útil si está basada en datos de la década de 1990.
4. *No haga predicciones acerca de una población distinta de la población de donde se obtuvieron los datos muestrales.* Si reunimos datos muestrales de hombres y desarrollamos una ecuación de regresión que relaciona la edad con el uso del control remoto del televisor, los resultados no necesariamente se aplican a las mujeres. Si empleamos *promedios* estatales para desarrollar una ecuación de regresión que relaciona las calificaciones de matemáticas del SAT con las calificaciones verbales del SAT, los resultados no necesariamente se aplican a los *individuos*.

Interpretación de la ecuación de regresión: Cambio marginal

Podemos utilizar la ecuación de regresión para observar el efecto en una variable, cuando la otra variable cambia una cantidad específica.

Definición

Cuando se trabaja con dos variables relacionadas por una ecuación de regresión, el **cambio marginal** en una variable es la cantidad que cambia cuando la otra variable cambia exactamente una unidad. La pendiente b_1 en la ecuación de regresión representa el cambio marginal que ocurre en y cuando x cambia una unidad.

Para los datos de la tabla 10-1 sobre las duraciones y los intervalos de tiempo posteriores a las erupciones del géiser Old Faithful, la recta de regresión tiene una pendiente de 0.234, lo que demuestra que si incrementamos x (las duraciones) en un segundo, el intervalo posterior a la erupción que se predijo se incrementará 0.234 minutos. Es decir, por cada segundo adicional de duración, esperamos que el intervalo posterior a la erupción (hasta la siguiente erupción) aumente 0.234 minutos más.

Valores extremos y puntos de influencia

Un análisis de correlación/regresión de datos bivariados (apareados) debe incluir la investigación de *valores extremos* y *puntos de influencia*, que se definen a continuación.

Definiciones

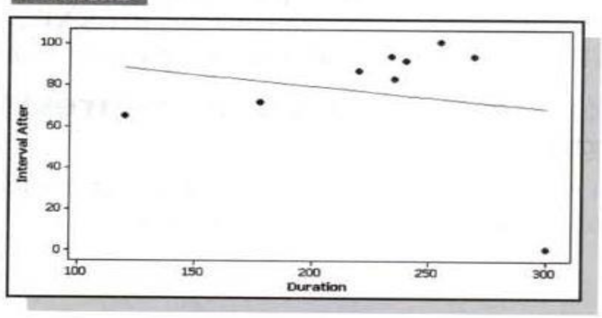
En un diagrama de dispersión, un **valor extremo** es un punto que aparece muy lejos de los otros puntos de datos.

Los datos muestrales apareados pueden incluir uno o más **puntos de influencia**, que son puntos que afectan fuertemente la gráfica de la recta de regresión.

Es fácil identificar un valor extremo: examine el diagrama de dispersión e identifique un punto que se aleja de los otros puntos. He aquí cómo determinamos un punto de influencia: grafique la recta de regresión que resulta de los datos con el punto incluido, después grafique la recta de regresión resultante de los datos sin incluir el punto. Si la gráfica cambia de forma considerable, se trata de un punto de influencia. Los puntos de influencia a menudo se encuentran al identificar los valores extremos que están alejados *horizontalmente* de los demás puntos.

Por ejemplo, observe el diagrama de dispersión de la figura 10-1a) en la página 516. Suponga que incluimos el siguiente par adicional de datos: $x = 300$, $y = 1$ (una erupción dura 300 segundos y el intervalo posterior a la erupción es de 1 minuto). Este punto adicional sería un punto de influencia porque la gráfica de la recta de regresión cambiaría considerablemente, tal como se observa en la siguiente pantalla de Minitab. Compare esta recta de regresión con la que se presentó en la imagen previa de Minitab, y observará con claridad que la incorporación de ese par de valores tiene un efecto drástico en la recta de regresión.

Minitab



Residuales y la propiedad de los mínimos cuadrados

Hemos establecido que la ecuación de regresión representa la recta que se ajusta "mejor" a los datos, y ahora describiremos el criterio utilizado para determinar

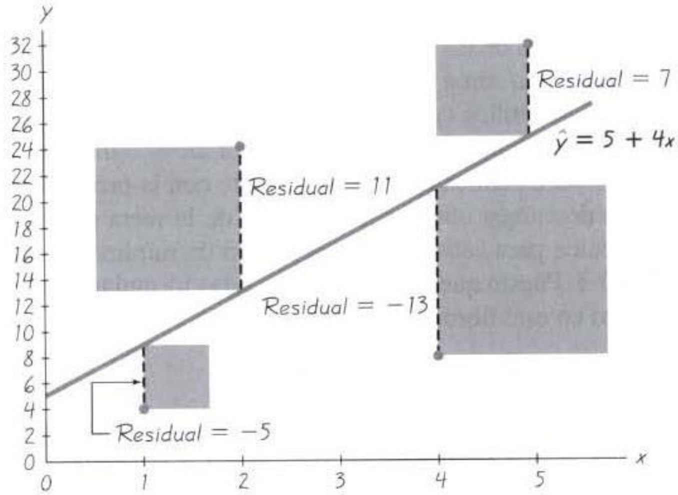


Figura 10-8

Residuales y cuadrados de los residuales

cuál recta es mejor que todas las demás. Este criterio se basa en las distancias verticales entre los puntos de datos originales y la recta de regresión. Tales distancias se denominan *residuales*.

Definición

Para una muestra de datos apareados (x, y), un **residual** es la diferencia $(y - \hat{y})$ entre un valor y muestral observado y el valor de $\hat{y}$, que es el valor de y predicho por medio de la ecuación de regresión. Es decir,

$$\text{residual} = \text{observada } y - \text{predicha } y = y - \hat{y}$$

Esta definición puede parecer tan clara como las instrucciones de un formulario fiscal, pero usted comprenderá fácilmente los residuales si se remite a la figura 10-8, que corresponde a los datos muestrales apareados que se presentan al margen. En la figura 10-8, los residuales están representados por las líneas punteadas. Para tener un ejemplo específico, observe el residual indicado como 7, que se encuentra directamente por arriba de $x = 5$. Si sustituimos $x = 5$ en la ecuación de regresión $\hat{y} = 5 + 4x$, obtenemos un valor predicho de $\hat{y} = 25$. Cuando $x = 5$, el valor *predicho* de y es $\hat{y} = 25$, pero el valor muestral real *observado* es $y = 32$. La diferencia $y - \hat{y} = 32 - 25 = 7$ es un residual.

x	1	2	4	5
y	4	24	8	32

La ecuación de regresión representa la recta que se ajusta "mejor" a los puntos, de acuerdo con la siguiente *propiedad de mínimos cuadrados*.

Definición

Una recta satisface la **propiedad de mínimos cuadrados** si la suma de los cuadrados de los residuales es la menor suma posible.

En la figura 10-8 podemos observar que los residuales son −5, 11, −13 y 7, de manera que la suma de sus cuadrados es

$$(-5)^2 + 11^2 + (-13)^2 + 7^2 = 364$$

Podemos visualizar la propiedad de mínimos cuadrados si nos remitimos a la figura 10-8, donde los cuadrados de los residuales están representados por las áreas de los cuadrados sombreados. La suma de las áreas sombreadas cuadradas es 364, que es la menor suma posible. Utilice cualquier otra recta y los cuadrados sombreados se combinarán para producir una área mayor que el área sombreada combinada de 364.

Por fortuna, no necesitamos lidiar directamente con la propiedad de mínimos cuadrados cuando deseamos obtener la ecuación de la recta de regresión. Ya se realizaron los cálculos para satisfacer la propiedad de mínimos cuadrados en las fórmulas 10-2 y 10-3. Puesto que la obtención de estas fórmulas requiere del cálculo, no las incluimos en este libro.

Gráficas residuales

En esta sección, al igual que en la anterior, describimos requisitos simplificados para el análisis eficaz de los resultados de correlación y regresión. Indicamos que siempre debemos comenzar con un diagrama de dispersión y que debemos verificar que el patrón de puntos se aproxime a una línea recta. También es necesario tomar en cuenta los valores extremos. La *gráfica residual* puede ser otra herramienta útil para analizar resultados de correlación y regresión, así como para verificar los requisitos necesarios para hacer inferencias sobre una correlación y una regresión.

Definición

Una **gráfica residual** es un diagrama de dispersión de los valores (x, y) una vez que cada uno de los valores de la coordenada y han sido reemplazados por el valor residual $y - \hat{y}$ (donde $\hat{y}$ denota el valor predicho de y). Es decir, una gráfica residual es una gráfica de los puntos $(x, y - \hat{y})$.

Para construir una gráfica residual, utilice el mismo eje x que en el diagrama de dispersión, pero use un eje vertical de valores residuales. Dibuje una línea horizontal de referencia a partir del valor residual de 0 y luego grafique los valores apareados de $(x, y - \hat{y})$. Puesto que la construcción manual de gráficas residuales podría convertirse en lo que los matemáticos denominan "tedioso", se recomienda el uso de un programa de cómputo. Cuando analice una gráfica residual, busque un patrón en la configuración de los puntos y utilice los siguientes criterios:

Si una gráfica residual no revela ningún patrón, la ecuación de regresión es una buena representación de la asociación entre las dos variables.

Si una gráfica residual revela algún patrón sistemático, la ecuación de regresión no es una buena representación de la asociación entre las dos variables.

Considere los siguientes ejemplos de pantallas de Minitab. En el caso 1 todo está bien; los puntos se aproximan a la recta de regresión, de manera que la ecuación de regresión es un buen modelo para describir la asociación entre las dos variables. La gráfica residual correspondiente no revela un patrón diferente.

El caso 2 genera un diagrama de dispersión que señala una asociación entre las dos variables, pero la relación no es lineal. La gráfica residual correspondiente indica un patrón diferente, lo que confirma que el modelo lineal no es un buen modelo en este caso.

El caso 3 tiene un diagrama de dispersión en el que los puntos se van alejando de la recta de regresión, y la gráfica residual revela un patrón de variación creciente, lo cual viola el requisito de que, para diferentes valores de x, las distribuciones de los valores de y deben tener la misma varianza. En este caso, probablemente la ecuación de regresión no sea un buen modelo.

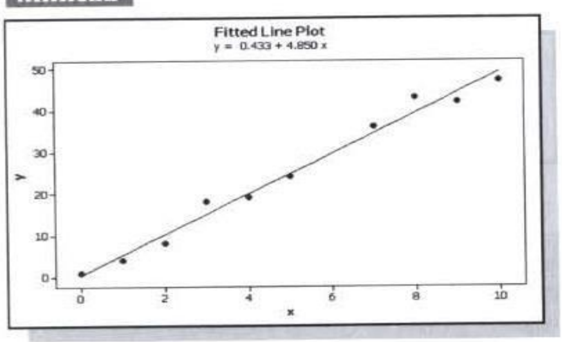
Después de adquirir la habilidad de calcular y analizar gráficas residuales junto con diagramas de dispersión, estamos más preparados para verificar los requisitos necesarios para asegurarnos de la validez de las inferencias que hacemos a partir de los procedimientos de correlación y regresión.

Caso 1

x	0	1	2	3	4	5	7	8	9	10
y	1	4	8	18	19	24	36	43	42	47

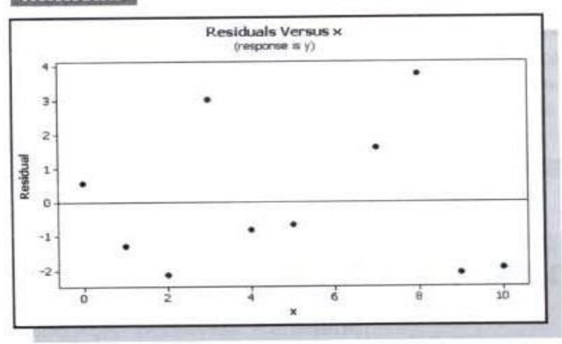
La recta de regresión se ajusta bien a los datos.

Minitab



La gráfica residual no revela ningún patrón.

Minitab

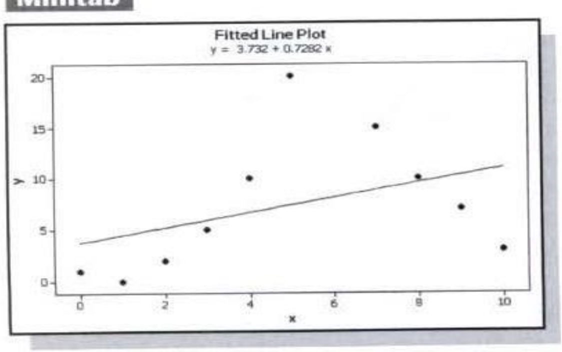


Caso 2

x	0	1	2	3	4	5	7	8	9	10
y	1	0	2	5	10	20	15	10	7	3

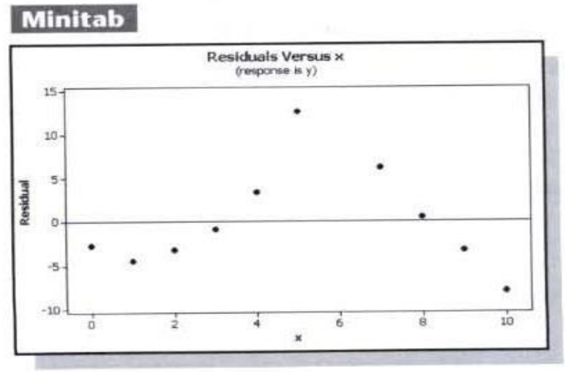
El diagrama de dispersión muestra que la asociación no es lineal.

Minitab



La gráfica residual revela un patrón diferente.

Minitab



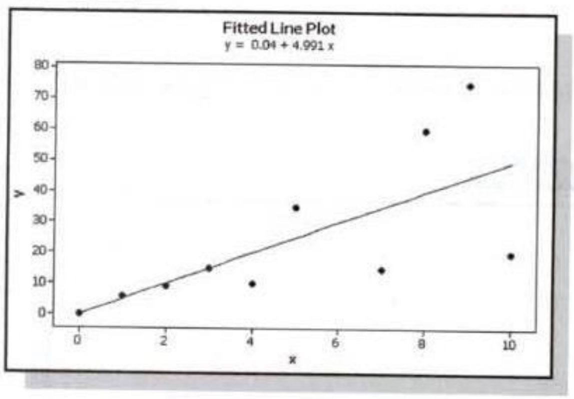
Caso 3

x	0	1	2	3	4	5	7	8	9	10
y	0	6	9	15	10	35	15	60	75	20

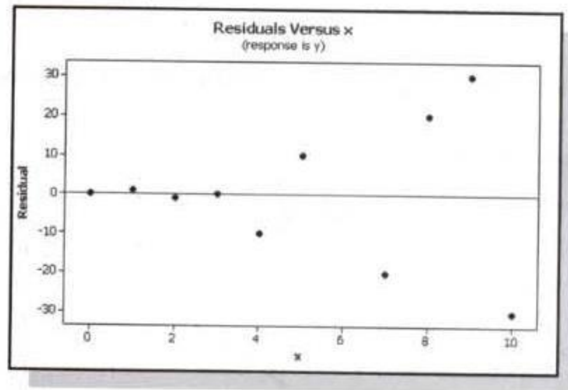
El diagrama de dispersión muestra la variación creciente de los puntos, alejándose de la recta de regresión.

La gráfica residual revela este patrón: de izquierda a derecha los puntos están más dispersos. (Esto va en contra del requisito de que, para diferentes valores de x, las distribuciones de los valores de y deben tener la misma varianza).

Minitab



Minitab



10-4 Variación e intervalos de predicción

Concepto clave Utilizando datos apareados (x, y) describiremos la variación que puede explicarse por la correlación lineal entre x y y y la variación que no puede explicarse. Luego, procederemos a considerar un método para construir un *intervalo de predicción*, que es un estimado del intervalo de un valor predicho de y. (Los estimados de intervalos de parámetros se conocen como *intervalos de confianza*, en tanto que los estimados de intervalos de variables suelen denominarse *intervalos de predicción*).

Variación explicada y sin explicar

Ahora examinaremos medidas de *desviación* y *variación* para un par de valores (x, y). En vez de contemplar abstracciones, consideremos el caso específico descrito en la figura 10-9. Imagine una muestra de datos apareados (x, y) que incluye (5, 19). Suponga que utilizamos esta muestra de datos apareados para calcular los siguientes resultados:

- Existe una correlación lineal (con r significativamente diferente de 0).
- La ecuación de la recta de regresión es $\hat{y} = 3 + 2x$.
- La media de los valores de y está dada por $\bar{y} = 9$.
- Uno de los pares de datos muestrales es $x = 5$ y $y = 19$.
- El punto (5, 13) es uno de los puntos sobre la recta de regresión, ya que la sustitución de $x = 5$ en la ecuación de regresión produce $\hat{y} = 13$.

La figura 10-9 indica que el punto (5, 13) está sobre la recta de regresión, pero el punto (5, 19) del conjunto de datos original no se ubica en la recta de regresión. Si ignoramos por completo los conceptos de correlación y regresión, y deseamos predecir un valor de y dado un valor de x y un conjunto de datos apareados (x, y), nuestra mejor conjetura sería la media $\bar{y}$. Pero en este caso, con una correlación lineal significativa, la forma de predecir el valor de y cuando $x = 5$ consiste en usar la ecuación de regresión para obtener $\hat{y} = 13$. Podemos explicar la discrepancia entre $\bar{y} = 9$ y $\hat{y} = 13$ al señalar que existe una relación lineal mejor descrita por medio de la recta de regresión. Como consecuencia, cuando $x = 5$, el valor predicho de y es 13 y no el valor medio de 9. Para $x = 5$, el valor predicho de y es 13, pero el valor muestral observado de y es en realidad 19. La discrepancia entre $\hat{y} = 13$ y $y = 19$ no puede explicarse por medio de la recta de regresión y se le denomina *desviación sin explicación* o *residual*. Esta desviación sin explicar se expresa en símbolos como $y - \hat{y}$.

El Súper Bowl como pronosticador del mercado de valores

La "superstición del Súper Bowl" afirma que una victoria por parte de un equipo de la NFL original es seguida por un año en el que aumenta el índice de cotizaciones del mercado de valores de Nueva York; si no resulta victorioso un equipo que originalmente formó parte de la NFL, el índice cae. (En 1970, la NFL y la AFL se unieron para formar la NFL actual). Después de los primeros 29 juegos del Súper Bowl, la predicción fue correcta el 90% de las veces, pero ha tenido mucho menos éxito en años recientes. Hasta el momento en que se escribe esto, ha sido correcta en 29 de 38 juegos del Súper Bowl, con una tasa de éxito del 76%. Los pronósticos y las predicciones son objetivos importantes de los especialistas en estadística y de los consejeros de inversiones, pero el sentido común sugiere que nadie debe basar sus inversiones en el resultado de un juego de fútbol. Otros indicadores que se utilizan para pronosticar el desempeño del mercado de valores incluyen el aumento del tamaño del dobladillo de las faldas, las ventas de aspirina, las limosnas en Wall Street, los pedidos de cajas de cartón, las ventas de cerveza contra las ventas de vino y el tránsito de los elevadores en la Bolsa de Valores de Nueva York.

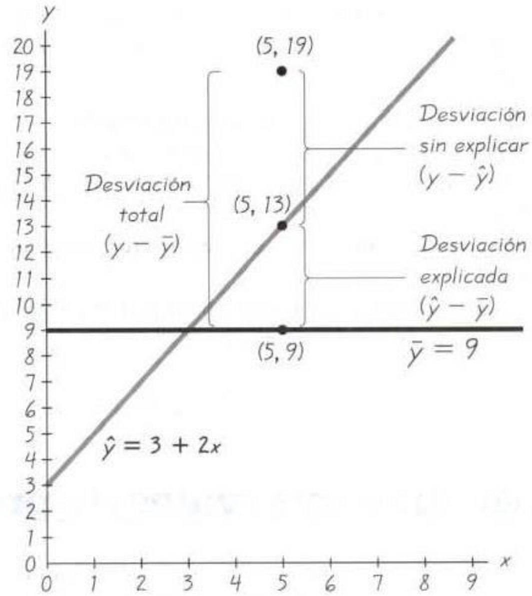


Figura 10-9 **Desviación sin explicar, explicada y total**

Igual que en la sección 3-3, donde definimos la desviación estándar, nuevamente considere que una *desviación* es la diferencia entre un valor y la media. (En este caso, la media es $\bar{y} = 9$.) Examine con atención la figura 10-9 y observe las siguientes desviaciones específicas a partir de $\bar{y} = 9$:

Desviación total (a partir de $\bar{y} = 9$) del punto $(5, 19) = y - \bar{y} = 19 - 9 = 10$

Desviación explicada (a partir de $\bar{y} = 9$) del punto $(5, 19) = \hat{y} - \bar{y} = 13 - 9 = 4$

Desviación sin explicar (a partir de $\bar{y} = 9$) del punto
$$(5, 19) = y - \hat{y} = 19 - 13 = 6$$

Estas desviaciones a partir de la media se generalizan y definen formalmente como sigue.

Definiciones

Suponga que tenemos un conjunto de datos apareados que contienen el punto muestral (x, y), que $\hat{y}$ es el valor predicho de y (obtenido por medio de la ecuación de regresión), y que la media de los valores y muestrales es $\bar{y}$.

La **desviación total** de (x, y) es la distancia vertical $y - \bar{y}$, que es la distancia entre el punto (x, y) y la recta horizontal que pasa por la media muestral $\bar{y}$.

La **desviación explicada** es la distancia vertical $\hat{y} - \bar{y}$, que es la distancia entre el valor predicho y y la recta horizontal que pasa por la media muestral $\bar{y}$.

La **desviación sin explicar** es la distancia vertical $y - \hat{y}$, que es la distancia vertical entre el punto (x, y) y la recta de regresión. (La distancia $y - \hat{y}$ también se conoce como *residual*, tal como se definió en la sección 10-3).

En la figura 10-9 podemos apreciar la siguiente relación:

$$\begin{array}{ccccc} \text{(desviación total)} & = & \text{(desviación explicada)} & + & \text{(desviación sin explicar)} \\ (y - \bar{y}) & = & (\hat{y} - \bar{y}) & + & (y - \hat{y}) \end{array}$$

Esta última expresión implica desviaciones a partir de la media y se aplica a cualquier punto (x, y) particular. Si sumamos los cuadrados de las desviaciones utilizando todos los puntos (x, y), obtenemos cantidades de *variación*, y la misma relación se aplica a las sumas de cuadrados que se muestran en la fórmula 10-4, aunque esta última expresión no es algebraicamente equivalente a la fórmula 10-4. En esa fórmula, la **variación total** se expresa como la suma de los cuadrados de los valores de desviación totales, la **variación explicada** es la suma de los cuadrados de los valores de desviación explicados, y la **variación sin explicar** es la suma de los cuadrados de los valores de desviación sin explicar.

Fórmula 10-4

$$\begin{array}{cccccc} & \text{(variación total)} & = & \text{(variación explicada)} & + & \text{(variación sin explicar)} \\ \text{o} & \Sigma(y - \bar{y})^2 & = & \Sigma(\hat{y} - \bar{y})^2 & + & \Sigma(y - \hat{y})^2 \end{array}$$

En la sección 10-2 vimos que el coeficiente de correlación lineal r se utiliza para calcular la proporción de la variación total en y que puede explicarse por medio de la correlación lineal. En la sección 10-2 hicimos la siguiente afirmación:

El valor de r^2 es la proporción de la variación en y que está explicada por la relación lineal entre x y y.

Esta aseveración sobre la variación explicada se formaliza en la siguiente definición.

Definición

El **coeficiente de determinación** es la cantidad de variación en y que está explicada por la recta de regresión. Se calcula como

$$r^2 = \frac{\text{variación explicada}}{\text{variación total}}$$

Podemos calcular r^2 por medio de la definición dada con la fórmula 10-14 o podemos simplemente elevar al cuadrado el coeficiente de correlación lineal r.

PROBLEMA 10 DEL CAPÍTULO

EJEMPLO Old Faithful En la sección 10-2 utilizamos las duraciones y los intervalos posteriores a las erupciones de la tabla 10-1 para calcular que $r = 0.926$. Calcule el coeficiente de determinación. También, obtenga el porcentaje de la variación total en y (intervalo posterior a la erupción) que puede explicarse por medio de la relación lineal entre la duración y el intervalo posterior a una erupción.

SOLUCIÓN El coeficiente de determinación es $r^2 = 0.926^2 = 0.857$. Como r^2 es la proporción de la variación total que está explicada, concluimos que aproximadamente el 86% de la variación total de los intervalos posteriores a las erupciones se pueden explicar por las duraciones. Esto significa que alrededor del 14% de la variación total de los intervalos posteriores a las erupciones está explicada por otros factores y no por las duraciones. Sin embargo, recuerde que esos resultados son estimaciones que se basan en los datos muestrales con que se cuenta. Es probable que otros datos muestrales produzcan estimaciones diferentes.

Intervalos de predicción

En la sección 10-3 empleamos los datos muestrales de la tabla 10-1 para calcular la ecuación de regresión $\hat{y} = 34.8 + 0.234x$, donde $\hat{y}$ representa el intervalo de tiempo predicho (en minutos) después de una erupción y hasta la siguiente, en tanto que x representa la duración de una erupción (en segundos). Después utilizamos esa ecuación para predecir el valor de y, dado que una erupción tiene una duración de $x = 180$ segundos. Encontramos que el mejor intervalo de tiempo predicho después de una erupción es de 76.9 minutos. Puesto que 76.9 es un único valor, se le conoce como *estimado puntual*. En el capítulo 7 aprendimos que los estimados puntuales tienen la grave desventaja de no darnos ninguna información acerca de su exactitud. Aquí, sabemos que 76.9 es el mejor valor predicho, pero no sabemos qué tan exacto es este valor. En el capítulo 7 elaboramos estimados del intervalo de confianza para superar esa desventaja, y en esta sección seguiremos el mismo método; utilizaremos un **intervalo de predicción**, que es un estimado del intervalo de un valor predicho de y. [Un estimado del intervalo de un *parámetro* (como la media de todos los intervalos posteriores a las erupciones) suele denominarse *intervalo de confianza*, mientras que el estimado de un intervalo de una *variable* (como el intervalo estimado posterior a una erupción con una duración de 180 segundos) se conoce generalmente como *intervalo de predicción*].

La creación de un intervalo de predicción requiere una medida de la dispersión de los puntos muestrales alrededor de la recta de regresión. Recuerde que la desviación sin explicar (o residual) es la distancia vertical entre un punto muestral y la recta de regresión, tal como se ilustra en la figura 10-9. El *error estándar del estimado* es una medida colectiva de la dispersión de los puntos muestrales alrededor de la recta de regresión, y se define de manera formal como sigue.

Definición

El **error estándar del estimado**, denotado por s_e, es una medida de las diferencias (o distancias) entre los valores muestrales observados de y y los valores predichos $\hat{y}$ que se obtienen por medio de la ecuación de regresión. Está dado por

$$s_e = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - 2}} \quad \textbf{(donde } \hat{y} \textbf{ es el valor predicho de y)}$$

o por medio de la siguiente fórmula equivalente:

Fórmula 10-5 $$s_e = \sqrt{\frac{\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy}{n - 2}}$$

STATDISK, Minitab, Excel y la calculadora TI-83/84 Plus están diseñados para calcular de manera automática el valor de s_e. Consulte el recuadro "Uso de la tecnología" al final de esta sección.

El cálculo del error estándar del estimado s_e se asemeja mucho al de la desviación estándar ordinaria que se explicó en la sección 3-3. Así como la desviación estándar es una medida de la desviación de los valores a partir de su media, el error estándar del estimado s_e es una medida de la desviación de los puntos de los datos muestrales a partir de su recta de regresión. La lógica que subyace en la división entre $n - 2$ es similar a la lógica que condujo a la división entre $n - 1$ para la desviación estándar ordinaria. Es importante señalar que valores relativamente pequeños de s_e reflejan puntos que están cercanos a la recta de regresión, y los valores relativamente grandes se presentan cuando hay puntos que se alejan de la recta de regresión.

La fórmula 10-5 es algebraicamente equivalente a la otra ecuación en la definición, pero la fórmula 10-5 suele ser más fácil ya que no requiere que calculemos cada uno de los valores predichos $\hat{y}$ por medio de sustitución en la ecuación de regresión. Sin embargo, la fórmula 10-5 sí requiere que calculemos el intercepto y, b_0, y la pendiente b_1 de la recta de regresión estimada.

EJEMPLO Cálculo de s_e Utilice la fórmula 10-5 para calcular el error estándar del estimado s_e para los datos de las duraciones y los intervalos posteriores a las erupciones del géiser Old Faithful, que aparecen en la tabla 10-1.

SOLUCIÓN Con los datos muestrales de la tabla 10-1, calculamos estos valores:

$$n = 8 \qquad \Sigma y^2 = 60{,}204 \qquad \Sigma y = 688 \qquad \Sigma xy = 154{,}378$$

En la sección 10-3 empleamos los datos muestrales de la tabla 10-1 para obtener el intercepto y y la pendiente de la recta de regresión. Esos valores se presentan aquí con más decimales para una mayor precisión.

$$b_0 = 34.7698041 \qquad b_1 = 0.2340614319$$

Ahora podemos usar estos valores en la fórmula 10-5 para calcular el error estándar del estimado s_e.

$$s_e = \sqrt{\frac{\Sigma y^2 - b_0\Sigma y - b_1\Sigma xy}{n-2}}$$

$$= \sqrt{\frac{60{,}204 - (34.7698041)(688) - (0.2340614319)(154{,}378)}{8-2}}$$

$$= 4.973916052 = 4.97 \quad \textbf{(redondeado)}$$

Es posible medir la dispersión de los puntos muestrales alrededor de la recta de regresión con el error estándar del estimado $s_e = 4.97$. Podemos emplear el error estándar del estimado s_e para construir estimados de intervalo que nos ayuden a ver qué tan confiables son realmente nuestros estimados puntuales de y. Suponga que para cada valor fijo de x, los valores muestrales correspondientes de y se distribuyen normalmente alrededor de la recta de regresión, y que estas distribuciones normales tienen la misma varianza. El siguiente estimado del intervalo se aplica a un valor y *individual*. (Consulte el ejercicio 26 para ver un intervalo de confianza utilizado para predecir la *media* de todos los valores de y, para algún valor dado de x).

Intervalo de predicción para una y individual

Dado el valor fijo x_0, el intervalo de predicción para una y individual es

$$\hat{y} - E < y < \hat{y} + E$$

donde el margen de error E es

$$E = t_{\alpha/2}\, s_e \sqrt{1 + \frac{1}{n} + \frac{n(x_0 - \bar{x})^2}{n(\Sigma x^2) - (\Sigma x)^2}}$$

y x_0 representa el valor dado de x, $t_{\alpha/2}$ tiene $n - 2$ grados de libertad, y s_e se calcula a partir de la fórmula 10-5.

TEXTO N° 19

Regresión multiple

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 566 a 571

10-5 Regresión múltiple

Concepto clave Aunque las secciones previas de este capítulo se aplican a una relación entre *dos* variables, en esta sección presentamos un método para analizar una relación lineal que incluye *más de dos* variables. Nos enfocamos en tres elementos fundamentales: **1.** la ecuación de regresión múltiple, **2.** el valor de R^2 ajustada y **3.** el valor P. Debido a la naturaleza tan compleja de las operaciones requeridas, los cálculos manuales son poco prácticos y constituyen una amenaza para la salud mental; así que en esta sección se destaca el uso y la interpretación de los resultados obtenidos con un programa estadístico de cómputo o una calculadora TI-83/84 Plus.

Ecuación de regresión múltiple

Al igual que en las secciones anteriores de este capítulo, sólo estudiaremos relaciones *lineales*. Utilizamos la siguiente *ecuación de regresión múltiple* para describir relaciones lineales que incluyen más de dos variables.

Definición

Una **ecuación de regresión múltiple** expresa una relación lineal entre una variable de respuesta y y dos o más variables de predicción $(x_1, x_2, \ldots, x_k)$. La forma general de una ecuación de regresión múltiple es

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k.$$

Emplearemos la siguiente notación, que surge de manera natural de la notación utilizada en la sección 9-3.

Notación

$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$ (forma general de la ecuación de regresión múltiple estimada)

n = tamaño de la muestra

k = número de variables de predicción. (Las variables de predicción también se conocen como *variables independientes* o variables x).

$\hat{y}$ = valor predicho de y (se calcula por medio de la ecuación de regresión múltiple)

$x_1, x_2, \ldots, x_k$ son las variables de predicción

β_0 = intercepto y, o el valor de y cuando todas las variables de predicción son 0. (Este valor es un parámetro poblacional).

b_0 = estimado de β_0 basado en los datos muestrales (b_0 es un estadístico muestral).

$\beta_1, \beta_2, \ldots, \beta_k$ son los coeficientes de las variables de predicción $x_1, x_2, \ldots, x_k$

$b_1, b_2, \ldots, b_k$ son estimados muestrales de los coeficientes $\beta_1, \beta_2, \ldots, \beta_k$

Para cualquier conjunto específico de valores de x, la ecuación de regresión está asociada con un error aleatorio que suele denotarse por ε, y suponemos que estos errores se distribuyen normalmente, con una media de 0 y una desviación estándar de σ, y que los errores aleatorios son independientes. Es difícil verificar estos supuestos. A lo largo de esta sección suponemos que los requisitos necesarios se satisfacen.

Los cálculos que se requieren para la regresión múltiple son tan complicados que *debe* utilizarse un programa de cómputo de estadística, así que nos concentraremos en *interpretar* las pantallas de resultados de los programas de cómputo. Al final de esta sección se incluyen instrucciones para el uso de STATDISK, Minitab, Excel y la calculadora TI-83/84 Plus.

PROBLEMA DEL CAPÍTULO 10

EJEMPLO Old Faithful En la sección 10-3 analizamos métodos para predecir el intervalo posterior a una erupción del géiser Old Faithful, pero en esa sección sólo se incluyó una variable de predicción. Utilice los datos muestrales de la tabla 10-1, incluida en el problema del capítulo, y calcule la ecuación de regresión múltiple en la que la variable de respuesta (y) es el intervalo posterior a una erupción y las variables de predicción (x) son la duración y la altura de la erupción. A continuación se presentan los resultados de Minitab.

continúa

Minitab

```
The regression equation is
Interval After = 45.1 + 0.245 Duration - 0.098 Height

Predictor      Coef   SE Coef       T      P
Constant      45.10     19.41    2.32  0.068
Duration    0.24464   0.04486    5.45  0.003
Height      -0.0983    0.1623   -0.61  0.571

S = 5.25937   R-Sq = 86.7%   R-Sq(adj) = 81.3%

Analysis of Variance

Source           DF       SS      MS      F      P
Regression        2   897.69  448.85  16.23  0.007
Residual Error    5   138.31   27.66
Total             7  1036.00
```

Factores de predicción de éxito

Cuando una universidad acepta a un nuevo estudiante, desearía tener algunos indicadores positivos de que éste será exitoso en sus estudios. Los decanos universitarios de admisiones toman en cuenta las calificaciones de la prueba SAT, las pruebas estándar de aprovechamiento, el lugar que ocupa el alumno en la clase, la dificultad de los cursos de preparatoria, las calificaciones de preparatoria y las actividades extracurriculares. En un estudio de las características que suelen ser buenos factores de predicción de éxito en la universidad, se encontró que el lugar que se tiene en la clase y las puntuaciones en pruebas estándar de aprovechamiento son mejores factores de predicción que las calificaciones del SAT. Una ecuación de regresión múltiple con el promedio general en la universidad, que se predijo por el lugar que ocupa el alumno en la clase y su puntuación en pruebas de aprovechamiento, no mejoró al incluir la calificación de la prueba SAT como otra variable. Este estudio en particular sugiere que las calificaciones de la prueba SAT no deben incluirse entre los criterios de admisión, aunque otros argumentan que las calificaciones de esta prueba son útiles para comparar estudiantes de diferentes lugares y de distintas preparatorias de procedencia.

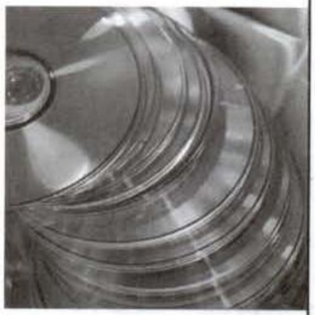
SOLUCIÓN Con Minitab obtenemos los resultados que se presentan a continuación. La ecuación de regresión múltiple que aparece en los resultados anteriores de Minitab es

```
Interval After = 45.1 + 0.245 Duration - 0.098 Height
```

Si utilizamos la notación presentada anteriormente en esta sección, podemos escribir esta ecuación de la siguiente forma

$$\hat{y} = 45.1 + 0.245x_1 - 0.098x_2$$

Si una ecuación de regresión múltiple se ajusta bien a los datos muestrales, se puede emplear para hacer predicciones. Por ejemplo, si determinamos que la ecuación es adecuada para hacer predicciones, y tenemos una erupción con una duración de 180 segundos y una altura de 130 pies, podemos predecir el intervalo posterior a la erupción sustituyendo esos valores en la ecuación de regresión, para obtener una duración predicha de 76.5 minutos. (Recuerde, las duraciones están en segundos, las alturas en pies y los intervalos posteriores a las erupciones en minutos). Además, los coeficientes $b_1 = 0.245$ y $b_2 = -0.098$ pueden emplearse para determinar el cambio marginal, como se describió en la sección 10-3. Por ejemplo, el coeficiente $b_1 = 0.245$ indica que, cuando la altura de una erupción permanece constante, el intervalo posterior a la erupción predicho aumenta 0.245 minutos por cada incremento de un segundo en la duración de la erupción.



Fabricación de CD con regresión múltiple

Sony fabrica millones de discos compactos en Terre Haute, Indiana. En un punto del proceso de fabricación, se expone una placa fotográfica a un láser, de manera que una señal musical es transferida a una señal digital codificada con ceros y unos. Este proceso se analizó estadísticamente para identificar los efectos de diferentes variables, tales como el tiempo de exposición y el grosor de la emulsión fotográfica. Métodos de regresión múltiple demostraron que, entre todas las variables consideradas, cuatro eran las más significativas. El proceso fotográfico se ajustó, con base en estas cuatro variables, para obtener resultados óptimos. Esto permitió disminuir el número de discos defectuosos y mantener la calidad. El uso de métodos de regresión múltiple condujo a costos más bajos de producción y a un mejor control del proceso de fabricación.

R^2 ajustada

R^2 denota el **coeficiente múltiple de determinación**, que es una medida de lo bien que se ajusta la ecuación de regresión múltiple a los datos muestrales. Un ajuste perfecto daría como resultado $R^2 = 1$, y un ajuste muy bueno daría por resultado un valor cercano a 1. Un ajuste muy deficiente se relaciona con un valor de R^2 cercano a 0. El valor de $R^2 = 86.7\%$ en los resultados de Minitab, indica que el 86.7% de la variación de los intervalos posteriores a las erupciones puede explicarse por la duración x_1 y la altura x_2. Sin embargo, el coeficiente múltiple de determinación R^2 tiene una grave desventaja: a mayor número de variables incluidas, R^2 se incrementa. (R^2 podría permanecer igual, pero suele incrementarse). La R^2 más grande se obtiene por el simple hecho de incluir todas las variables disponibles, pero la mejor ecuación de regresión múltiple no necesariamente utiliza todas las variables disponibles. A causa de esta desventaja, la comparación de diferentes ecuaciones de regresión múltiple se logra mejor con el coeficiente ajustado de determinación, que es R^2 ajustada para el número de variables y el tamaño de la muestra.

Definición

El **coeficiente ajustado de determinación** es el coeficiente múltiple de determinación R^2 modificado para justificar el número de variables y el tamaño de la muestra. Se calcula por medio de la fórmula 10-6.

Fórmula 10-6 $$R^2 \text{ ajustada} = 1 - \frac{(n-1)}{[n-(k+1)]}(1-R^2)$$

donde n = tamaño muestral

k = número de variables de predicción (x)

Los resultados anteriores de Minitab para los datos indican que el coeficiente ajustado de determinación es R-sq(adj) = 81.3%. Si utilizamos la fórmula 10-6 con el valor de $R^2 = 0.867$, $n = 8$ y $k = 2$, encontramos que el valor ajustado de R^2 es 0.813, lo que confirma el valor de 81.3% de los resultados de Minitab. (En realidad obtuvimos 0.814, pero obtenemos 0.813 si utilizamos más dígitos para minimizar el error de redondeo). El valor de R^2 de 86.7% indica que el 86.7% de la variación de los intervalos posteriores a las erupciones puede explicarse por la duración x_1 y la altura x_2, pero cuando comparamos esta ecuación de regresión múltiple con otras, es mejor utilizar la R^2 ajustada de 81.3% (o 0.813).

Valor *P*

El valor *P* es una medida de la significancia general de la ecuación de regresión múltiple. El valor *P* de 0.007 de los resultados de Minitab es pequeño, lo que indica que la ecuación de regresión múltiple tiene una buena significancia general y es útil para hacer predicciones. Es decir, tiene sentido predecir intervalos posteriores a las erupciones con base en las duraciones y las alturas de las erupciones. Al igual que la R^2 ajustada, este valor *P* es una buena medida de qué tan bien se ajusta la ecuación a los datos muestrales. El valor de 0.007 resulta de una prueba de la hipótesis nula de que $\beta_1 = \beta_2 = 0$. El rechazo de $\beta_1 = \beta_2 = 0$ implica que al menos uno de β_1 y β_2 no es 0, lo que indica que esta ecuación de regresión es eficaz para determinar los intervalos posteriores a las erupciones. Un análisis completo de los resultados de Minitab podría llevarnos a incluir otros elementos importantes, como la significancia de los coeficientes individuales, pero limitaremos nuestra explicación a los tres componentes principales: la ecuación de regresión múltiple, la R^2 ajustada y el valor *P*.

Cálculo de la mejor ecuación de regresión múltiple

El resultado anterior de Minitab se basa en el uso de las variables de predicción de duración y altura con los datos muestrales de la tabla 10-1. Pero si deseamos predecir el intervalo posterior a una erupción, ¿existe alguna otra combinación de variables que podría ser mejor que la duración y la altura? La tabla 10-3 lista

Tabla 10-3 Búsqueda de la mejor ecuación de regresión múltiple

← **R^2 ajustada más alta y valor *P* más bajo**

	R^2	R^2 ajustada	Significancia general
DURACIÓN	0.857	0.833	0.001
INTERVALO PREVIO	0.011	0.000	0.802
ALTURA	0.073	0.000	0.519
DURACIÓN e INTERVALO PREVIO	0.872	0.820	0.006
DURACIÓN y ALTURA	0.867	0.813	0.007
INTERVALO PREVIO y ALTURA	0.073	0.000	0.828
DURACIÓN e INTERVALO PREVIO y ALTURA	0.875	0.781	0.028

Salarios de la NBA y desempeño

El investigador Matthew Weeks estudió la correlación entre los salarios de la NBA y las estadísticas del juego de básquetbol. Además del salario (S), consideró los minutos jugados (M), las intervenciones (I), los rebotes (R) y los puntos anotados (P); utilizó datos de 30 jugadores. La ecuación de regresión múltiple es $S = -0.716 - 0.0756M - 0.425I + 0.0536R + 0.742P$ con $R^2 = 0.458$. Debido a una alta correlación entre los minutos jugados (M) y los puntos anotados (P), y puesto que estos últimos tuvieron una alta correlación con el salario, la variable de minutos jugados se eliminó de la ecuación de regresión múltiple. Además, no se encontró que las variables de intervenciones (I) y rebotes (R) fueran significativas, por lo que también se eliminaron. La variable de los puntos anotados pareció ser la mejor elección para predecir los salarios de la NBA, pero se encontró que las predicciones no eran muy exactas ya que no se consideraron otras variables, tales como la popularidad del jugador.

distintas combinaciones de variables. Ahora nos enfrentamos al importante objetivo de calcular la *mejor* ecuación de regresión múltiple. Puesto que la determinación de la mejor ecuación de regresión múltiple requiere de una buena dosis de juicio, no existe un procedimiento exacto y automático para esto. *La determinación de la mejor ecuación de regresión múltiple suele ser bastante difícil y rebasa los alcances de este libro*, pero los siguientes lineamientos servirán de guía.

Lineamientos para el cálculo de la mejor ecuación de regresión múltiple

1. *Utilice el sentido común y consideraciones prácticas para incluir o excluir variables.* Por ejemplo, podríamos excluir la variable de la estatura después de saber que esta variable es un estimado visual y no una medida exacta.
2. *Considere el valor P.* Seleccione una ecuación que tenga significancia general, tal como lo determina el valor P indicado en los resultados del programa de cómputo. Por ejemplo, observe los valores de significancia general en la tabla 10-3. Los valores P de 0.802, 0.519 y 0.828 corresponden a las combinaciones de variables que no dan como resultado una significancia general, de manera que esas combinaciones deben excluirse.
3. *Considere ecuaciones con valores altos de R^2 ajustada y trate de incluir sólo unas cuantas variables.* En vez de incluir casi todas las variables disponibles, trate de incluir relativamente pocas variables de predicción (x). Utilice los siguientes lineamientos:
   - Seleccione una ecuación que tenga un valor de R^2 ajustada con esta propiedad: si se incluye una variable independiente adicional, el valor de R^2 ajustada no se incrementa de manera sustancial.
   - Para un número dado de variables de predicción (x), seleccione la ecuación con el valor más grande de la R^2 ajustada.
   - Para eliminar las variables de predicción (x) que no tienen mucho efecto sobre la variable de respuesta (y), sería útil calcular el coeficiente de correlación lineal r para cada par de variables en consideración. Si dos valores de predicción tienen un coeficiente de correlación lineal muy alto, no es necesario incluir a ambos, y debemos excluir la variable con el valor de r más bajo.

Si seguimos estos lineamientos al intentar calcular la mejor ecuación para predecir los intervalos posteriores a las erupciones del géiser Old Faithful, encontramos que, para los datos de la tabla 10-3, la mejor ecuación de regresión utiliza únicamente la variable de predicción (x) de duración. Parece que la mejor ecuación de regresión es

$$\text{INTERVALO POSTERIOR} = 34.8 + 0.234\ \text{DURACIÓN}$$

o

$$\hat{y} = 34.8 + 0.234x_1$$

Los lineamientos anteriores se basan en la R^2 ajustada y en el valor P, pero también podemos hacer pruebas de hipótesis individuales, basadas en los valores de los coeficientes de regresión. Considere el coeficiente de regresión de β_1. Una prueba de la hipótesis nula $\beta_1 = 0$ nos puede indicar si la variable de predicción correspondiente debe incluirse en la ecuación de regresión. El rechazo de $\beta_1 = 0$ sugiere que β_1 tiene un valor diferente de cero y, por lo tanto, sirve para predecir el valor de la variable de respuesta. En el ejercicio 17 se describen procedimientos para este tipo de pruebas.

Algunos programas estadísticos de cómputo incluyen un programa para realizar la **regresión por pasos**, en la cual los cálculos se realizan con distintas combinaciones de variables de predicción (x), aunque existen graves problemas asociados con ella, incluyendo los siguientes: la regresión por pasos no necesariamente produce el mejor modelo si algunas variables de predicción tienen una alta correlación; produce valores inflados de R^2; utiliza demasiado papel y no nos permite *pensar* acerca del problema. Como siempre, debemos ser cuidadosos de emplear los resultados de las computadoras como una herramienta que nos ayude a tomar decisiones inteligentes; no debemos permitir que la computadora tome las decisiones. En vez de confiar únicamente en los resultados de una regresión por pasos realizada por un programa de cómputo, considere los factores anteriores cuando trate de identificar la mejor ecuación de regresión múltiple.

Si corremos el programa de regresión por pasos de Minitab utilizando los datos de la tabla 10-1, obtendremos una pantalla de resultados que sugiere que la mejor ecuación de regresión es aquella en la que la duración es la única variable de predicción. Parece que podemos estimar el intervalo posterior a una erupción, y la ecuación de regresión nos conduce a esta regla: se estima que el intervalo (en minutos) posterior a una erupción es 34.8 más 0.234 veces la duración (en segundos).

Variables ficticias o indicadoras y regresión logística

En esta sección todas las variables han sido de naturaleza continua. El intervalo posterior a una erupción puede ser cualquier valor en un rango continuo de minutos, por lo que es un buen ejemplo de una variable continua. Sin embargo, muchas aplicaciones incluyen una **variable dicotómica**, que sólo tiene *dos* valores discretos posibles (como hombre/mujer, vivo/muerto o sano/no sano). Un procedimiento común consiste en representar las dos variables discretas posibles con 0 y 1, donde 0 representa un "fracaso" (como muerte) y 1 representa un éxito. A una variable dicotómica con los dos valores posibles de 0 y 1 se le llama **variable ficticia** o **indicadora.**

Los procedimientos de análisis difieren de manera drástica, dependiendo de si la variable ficticia es una variable de predicción (x) o la variable de respuesta (y). Si incluimos una variable ficticia como otra variable de predicción (x), podemos utilizar los métodos de esta sección, como se ilustra en el siguiente ejemplo.

TEXTO N° 20

Elaboración de modelos

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 576 a 579

10-6 Elaboración de modelos

Concepto clave En esta sección se introducen algunos conceptos básicos para el desarrollo de un **modelo matemático**, el cual es una función matemática que se "ajusta" o describe datos del mundo real. Por ejemplo, podríamos buscar un modelo matemático consistente en una ecuación que relaciona una variable del tamaño poblacional con otra variable que representa el tiempo. A diferencia de la sección 10-3, no estamos restringidos a un modelo que deba ser lineal. Además, en vez de utilizar datos muestrales seleccionados al azar, consideraremos datos reunidos periódicamente a través del tiempo o alguna otra unidad básica de medición. Existen algunos métodos estadísticos poderosos que podemos estudiar (tales como las *series de tiempo*), pero el principal objetivo de esta sección es describir brevemente la manera en que puede utilizarse la tecnología para obtener un buen modelo matemático.

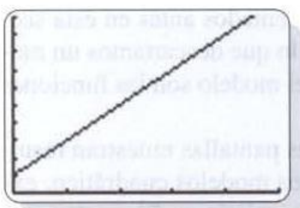
A continuación se presentan algunos modelos genéricos como aparecen en un menú de la calculadora TI-83/84 Plus (presione **STAT** y luego seleccione **CALC**):

Lineal:	$y = a + bx$	Cuadrático:	$y = ax^2 + bx + c$
Logarítmico:	$y = a + b \ln x$	Exponencial:	$y = ab^x$
Potencia:	$y = ax^b$		

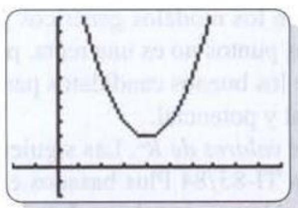
El modelo particular que usted seleccione depende de la naturaleza de los datos muestrales, y un diagrama de dispersión resultará muy útil para tomar esta determinación. Las ilustraciones que aparecen a continuación son gráficas de algunos modelos comunes elaborados en una calculadora TI-83/84 Plus.

TI-83/84 Plus

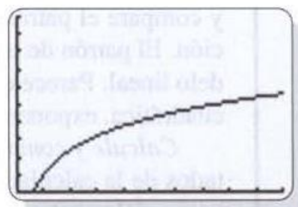
Lineal: $y = 1 + 2x$



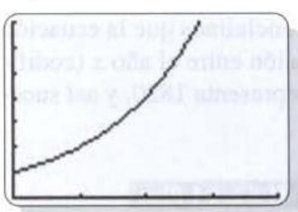
Cuadrático: $y = x^2 - 8x + 18$



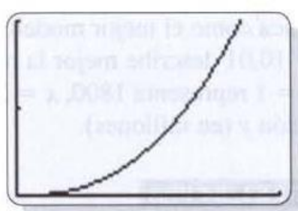
Logarítmico: $y = 1 + 2 \ln x$



Exponencial: $y = 2^x$



Potencia: $y = 3x^{2.5}$



Éstas son las reglas básicas para la creación de un buen modelo matemático:

1. *Busque un patrón en la gráfica.* Examine la gráfica con los puntos y compare el patrón básico de las gráficas genéricas conocidas de una función lineal, una función cuadrática, una función exponencial, una función potencial, etcétera. (Remítase a las gráficas de los resultados de la calculadora TI-83/84 Plus que se presentan en los ejemplos). Cuando trate de seleccionar un modelo, considere sólo aquellas funciones que parecen ajustarse visualmente a los puntos observados, de una forma razonablemente adecuada.
2. *Calcule y compare valores de R^2.* Para cada modelo que considere, utilice programas de cómputo o una calculadora TI-83/84 Plus para obtener el valor del coeficiente de determinación R^2. Los valores de R^2 se pueden interpretar aquí de la misma forma que se interpretaron en la sección 10-5. Al delimitar sus posibles modelos, seleccione funciones que den como resultado valores más grandes de R^2, ya que corresponden a funciones que se ajustan mejor a los puntos observados. Sin embargo, no dé demasiada importancia a las diferencias pequeñas, como la diferencia entre $R^2 = 0.984$ y $R^2 = 0.989$. (Otra medición utilizada para evaluar la calidad de un modelo es la suma de cuadrados de los residuales. Véase el ejercicio 15).

Estadística: Empleos y empleadores

A continuación se describe una muestra pequeña de anuncios de empleos en el campo de la estadística: pronosticador del tiempo, analista de bases de datos, científico de marketing, gerente de riesgos de crédito, investigador y evaluador del cáncer, analista de riesgos de seguros, investigador de pruebas educativas, bioestadístico, estadístico para productos farmacéuticos, criptólogo, programador estadístico.

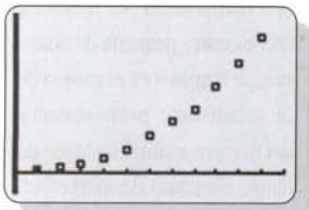
La siguiente es una muestra pequeña de empresas que ofrecen empleos en el campo de la estadística: Centers for Disease Control and Prevention, Cardiac Pacemakers, Inc., National Institutes of Health, National Cancer Institute, CNA Insurance Company, Educational Testing Service, Roswell Park Cancer Institute, Cleveland Clinic Foundation, National Security Agency, Quantiles, 3M, IBM, Nielsen Media Research, AT&T Labs, Bell Labs, Hewlett Packard, Johnson & Johnson, Smith Hanley.

3. *Reflexione*. Aplique el sentido común. No utilice un modelo que conduzca a valores predichos que se sabe son poco realistas. Utilice el modelo para calcular valores futuros, valores pasados y valores de años faltantes; luego determine si los resultados son realistas.

EJEMPLO La tabla 10-4 lista la población de Estados Unidos en diferentes años. Encuentre un buen modelo matemático para el tamaño poblacional, después haga una predicción del tamaño de la población de Estados Unidos para el año 2020.

SOLUCIÓN Primero "codificamos" los valores del año utilizando 1, 2, 3 . . . , en vez de 1800, 1820, 1840. . . . La razón de esta codificación es que, de esta forma, los valores de x son mucho más pequeños y tienen menos posibilidades de causar problemas de cálculo como los que ocurrirían con valores realmente grandes de x.

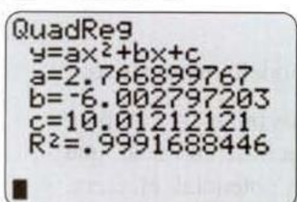
TI-83/84 Plus



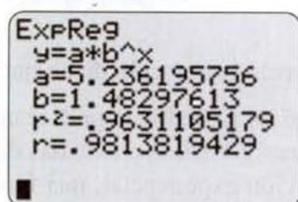
Busque un patrón en la gráfica. Examine el patrón de los valores de los datos en los resultados de la calculadora TI-83/84 Plus (mostrados al margen) y compare el patrón con los modelos genéricos presentados antes en esta sección. El patrón de estos puntos no es una recta, por lo que descartamos un modelo lineal. Parece que los buenos candidatos para el modelo son las funciones cuadrática, exponencial y potencial.

Calcule y compare valores de R^2. Las siguientes pantallas muestran resultados de la calculadora TI-83/84 Plus basados en los modelos cuadrático, exponencial y potencial. Al comparar los valores del coeficiente R^2, parece que el modelo cuadrático es el mejor, ya que tiene el valor más alto de 0.9992, aunque los otros valores mostrados también son bastante altos. Si seleccionamos la función cuadrática como el mejor modelo, concluimos que la ecuación $y = 2.77x^2 - 6.00x + 10.01$ describe mejor la relación entre el año x (codificado de manera que $x = 1$ representa 1800, $x = 2$ representa 1820, y así sucesivamente) y la población y (en millones).

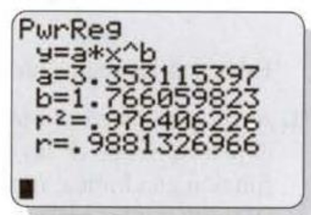
TI-83/84 Plus



TI-83/84 Plus



TI-83/84 Plus



Para predecir la población de Estados Unidos para el año 2020, primero observe que el año 2020 está codificado como $x = 12$ (véase la tabla 10-4). Sustituyendo $x = 12$ en el modelo cuadrático de $y = 2.77x^2 - 6.00x + 10.01$, obtenemos el resultado $y = 337$, lo cual indica que se estima que la población de Estados Unidos será de 337 millones en el año 2020.

Tabla 10-4 Población (en millones) de Estados Unidos

Año	1800	1820	1840	1860	1880	1900	1920	1940	1960	1980	2000
Año codificado	1	2	3	4	5	6	7	8	9	10	11
Población	5	10	17	31	50	76	106	132	179	227	281

Reflexione. El resultado predicho de 337 millones en 2020 parece razonable (una proyección del U.S. Bureau of the Census sugiere que la población en 2020 será de alrededor de 325 millones). Sin embargo, existe un gran riesgo al hacer estimados de tiempos que están más allá del alcance de los datos disponibles. Por ejemplo, el modelo cuadrático sugiere que en 1492 la población de Estados Unidos era de 671 millones, un resultado absurdo. El modelo cuadrático parece ser bueno para los datos disponibles (1800-2000), pero otros modelos podrían ser mejores si es absolutamente necesario hacer estimados poblacionales más allá de este periodo.

En su artículo "Modeling the U.S. Population" (*AMATYC Review*, vol. 20, núm. 2), Sheldon Gordon emplea más datos que los de la tabla 10-4 y utiliza técnicas mucho más avanzadas para obtener mejores modelos poblacionales. En ese artículo, comenta algo importante:

"La mejor opción (de un modelo) depende del conjunto de datos que se analiza y requiere no sólo de cálculos, sino también de ejercitar el juicio".

TEXTO N° 21

Graficas de control para la variación y la media

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 734 a 744

14-1 Panorama general

En el capítulo 2 señalamos que al describir, explorar y comparar conjuntos de datos, es sumamente importante tomar en cuenta las características centrales, la variación, la distribución, los valores extremos y los aspectos que cambian con el paso del tiempo. El principal objetivo de este capítulo es estudiar el último aspecto: las características cambiantes de los datos a lo largo del tiempo. Cuando se investigan características tales como el centro y la variación, es importante saber si se trata de una población estable o de una que está cambiando conforme transcurre el tiempo.

Actualmente existe una fuerte tendencia a tratar de mejorar la calidad de los bienes y servicios, y un número creciente de empresas utilizan los métodos que se presentan en este capítulo. La evidencia de la creciente importancia de la calidad se encuentra en la publicidad y en el gran número de libros y artículos que se enfocan en el tema de la calidad. En muchos casos, quienes solicitan empleo (¿usted?) poseen una ventaja definitiva cuando pueden decir a los empleadores que estudiaron estadística y métodos de control de calidad. Este capítulo presentará algunas de las herramientas básicas que se utilizan comúnmente para controlar la calidad.

Minitab, Excel y otros paquetes estadísticos de cómputo incluyen programas para generar automáticamente el tipo de gráficas que se estudian en este capítulo: incluiremos diversos ejemplos de esas representaciones gráficas. Las gráficas de control son un buen ejemplo de los maravillosos recursos gráficos que nos permiten *ver* y *comprender* algunas propiedades de los datos que, de otra forma, serían muy difíciles o imposibles de entender. El mundo necesita más personas capaces de construir e interpretar gráficas importantes, tales como las gráficas de control descritas en este capítulo.

14-2 Gráficas de control para la variación y la media

Concepto clave El principal objetivo de esta sección es la construcción de gráficas de rachas, gráficas R y gráficas $\bar{x}$ para controlar características importantes de datos a lo largo del tiempo. Usaremos este tipo de gráficas para determinar si un proceso es estadísticamente estable (o si está bajo control estadístico).

La siguiente definición describe formalmente el tipo de datos que manejaremos en este capítulo.

Definición

Los **datos de proceso** son datos ordenados de acuerdo con alguna secuencia de tiempo. Son mediciones de una característica de bienes o servicios que resultan de alguna combinación de equipo, personas, materiales, métodos y condiciones.

Por ejemplo, la tabla 14-1 incluye datos de proceso que consisten en el error medido (en pies) de las lecturas de altímetros durante 20 días consecutivos de producción. Cada día se seleccionaron cuatro altímetros al azar y se probaron. Puesto que los datos en la tabla 14-1 están ordenados de acuerdo con el momento en que se seleccionaron, se trata de datos de proceso. Es muy importante reconocer este punto:

> **Las características importantes de datos de proceso pueden cambiar a lo largo del tiempo.**

Al producir altímetros, el fabricante empleará personal competente y bien capacitado, además de máquinas correctamente calibradas, pero si el personal es reemplazado o si las máquinas se estropean con el uso, los altímetros podrían empezar a resultar defectuosos. Existen compañías que han ido a la bancarrota por haber permitido, involuntariamente, que el proceso de fabricación se deteriorara al no ejercer un control constante.

Gráficas de rachas

Existen varios métodos que pueden emplearse para controlar un proceso y así asegurar que las características deseadas importantes no cambien; el análisis de una *gráfica de rachas* es un método de este tipo.

Definición

Una **gráfica de rachas** es una gráfica secuencial de valores de datos *individuales* a lo largo del tiempo. Un eje (generalmente el eje vertical) se utiliza para los valores de los datos, y el otro eje (generalmente el eje horizontal) se emplea para la secuencia de tiempo.

PROBLEMA 14 DEL CAPÍTULO

EJEMPLO Fabricación de altímetros para aviones Trate los 80 errores de los altímetros de la tabla 14-1 como una secuencia de mediciones consecutivas, construya una gráfica de rachas utilizando el eje vertical para los errores y el eje horizontal para identificar el orden de los datos muestrales.

SOLUCIÓN La figura 14-1 es la gráfica de rachas generada por Minitab para los datos de la tabla 14-1. La escala vertical se diseñó para ajustarse a los errores

continúa

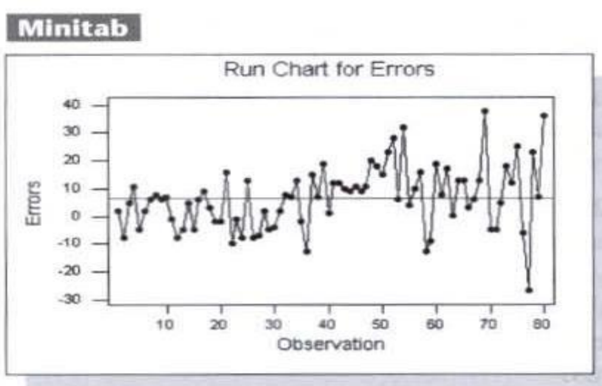


Figura 14-1 Gráfica de rachas de los errores individuales de altímetros de la tabla 14-1

El *efecto* Flynn: tendencia a la alza en puntuaciones de CI

Una gráfica de rachas o gráfica de control de las puntuaciones de CI revelaría que éstas exhiben una tendencia a incrementarse, ya que las puntuaciones de CI han estado aumentando de forma estable desde que empezaron a utilizarse hace casi 70 años. Esta tendencia es mundial y es igual en los distintos tipos de pruebas de inteligencia, incluso en aquellas que se basan casi por completo en el razonamiento abstracto y no verbal, con mínima influencia de la cultura. A esta tendencia al incremento se le ha llamado *efecto Flynn*, ya que el científico político James R. Flynn descubrió esta tendencia en sus estudios con reclutas del ejército de Estados Unidos. La cantidad del incremento es muy sustancial: con base en la puntuación media del CI de 100, se estima que el CI medio en 1920 era de cerca de 77. Por lo tanto, el estudiante común actual es brillante, si se le compara con sus bisabuelos. Hasta ahora no hay una explicación aceptable para el efecto Flynn.

de los altímetros, que van desde −27 pies hasta 38 pies, que son los valores mínimo y máximo de la tabla 14-1. La escala horizontal está diseñada para incluir los 80 valores ordenados en secuencia. El primer punto representa el primer valor de 2 pies, el segundo punto representa el segundo valor de −8 pies y así sucesivamente.

En la figura 14-1 la escala horizontal identifica el número de muestra, de manera que el número 20 indica el vigésimo artículo. La escala vertical representa el error del altímetro (en pies). Ahora examine la figura 14-1 y trate de identificar cualquier *patrón* que resalte a la vista. La figura 14-1 revela este problema: conforme el tiempo avanza de izquierda a derecha, las alturas de los puntos parecen mostrar un patrón de variación creciente. Observe cómo los puntos a la izquierda fluctúan mucho menos que los puntos de la derecha. Las normas de la Federal Aviation Administration exigen errores menores de 20 pies (o que estén entre 20 pies y −20 pies), de tal manera que los altímetros representados por puntos a la izquierda son aceptables, mientras que varios de los puntos de la derecha corresponden a altímetros que no cumplen con las especificaciones requeridas. Parece que el proceso de fabricación empezó bien, pero se deterioró con el paso del tiempo. Si se deja así, este proceso de fabricación provocará que la empresa tenga que cerrar.

Interpretación de las gráficas de rachas Los datos de un proceso se pueden tratar como si provinieran de una población con una media, desviación estándar, distribución y otras características constantes únicamente cuando el proceso es *estadísticamente estable*.

Definición

Un proceso es **estadísticamente estable** (o está **bajo control estadístico**) si sólo varía de forma natural, sin patrones, ciclos o puntos fuera de lo común.

La figura 14-2 consiste en gráficas de rachas que ilustran los patrones típicos que indican formas en las cuales el proceso de llenado de latas de bebida de cola de 12 onzas puede no ser estadísticamente estable.

- **Figura 14-2a):** Hay una evidente *tendencia creciente*, que corresponde a valores que se incrementan con el paso del tiempo. Si el proceso de llenado continúa con este tipo de patrón, las latas se llenarían con más y más bebida de cola hasta que el líquido empezara a derramarse y los empleados terminarían nadando en bebida de cola.
- **Figura 14-2b):** Hay una evidente *tendencia descendente* que corresponde a valores que disminuyen de manera estable. Las latas se llenarían con menos y menos bebida de cola hasta que estarían casi vacías. Un proceso como éstos requeriría de una revisión completa de las latas, con la finalidad de llenarlas con suficiente cantidad para distribuirlas a los consumidores.
- **Figura 14-2c):** Hay un *cambio hacia arriba*. Una gráfica de rachas como ésta podría resultar de un ajuste en el proceso de llenado, provocando que los valores posteriores sean más altos.

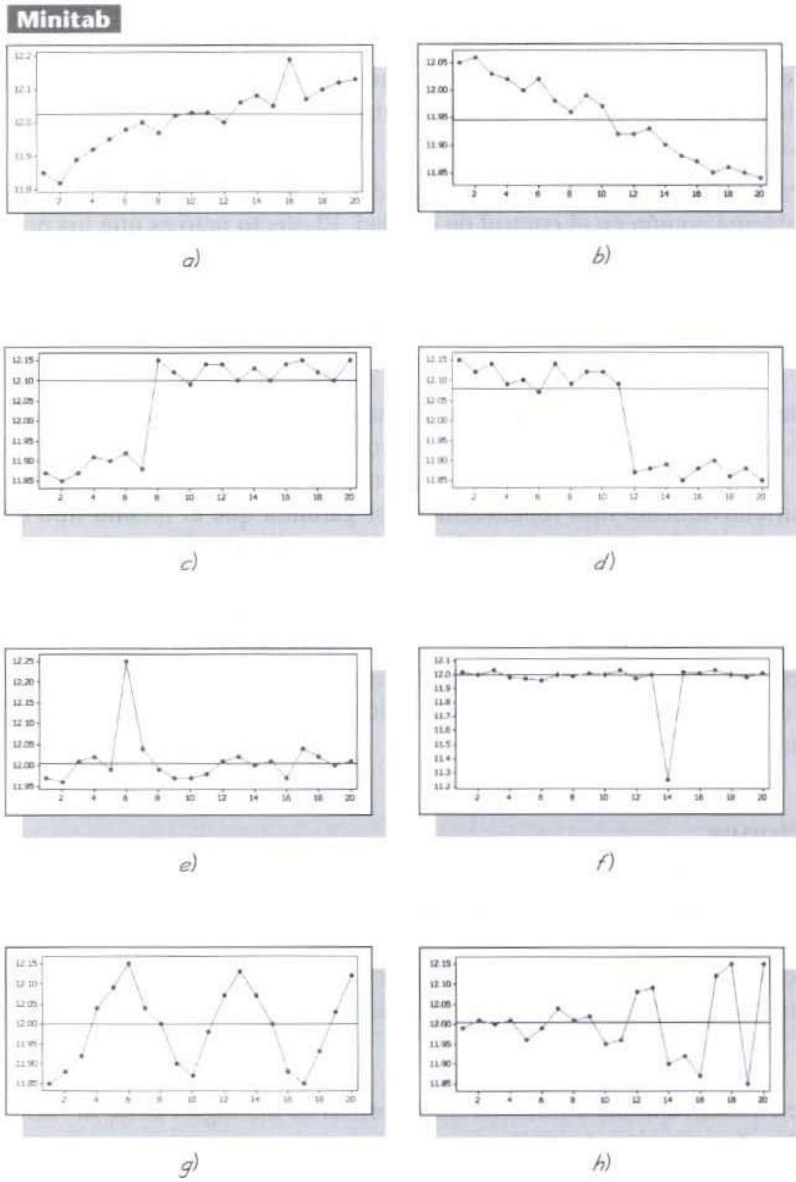


Figura 14-2

Procesos que no son estadísticamente estables

- **Figura 14-2d):** Existe un *cambio hacia abajo*. Los primeros valores son relativamente estables, pero después algo sucede, ya que los últimos valores son relativamente estables, aunque en un nivel mucho más bajo.
- **Figura 14-2e)**: El proceso es estable, salvo por un *valor excepcionalmente alto*. La causa de un valor tan fuera de lo común debe investigarse. Tal vez las latas se atascaron temporalmente y una lata en particular se llenó dos veces en lugar de una.
- **Figura 14-2f):** Existe un *valor excepcionalmente bajo*.

- **Figura 14-2g):** Hay un *patrón cíclico* (o ciclo repetitivo). Como es evidente, este patrón no es aleatorio y, por lo tanto, revela un proceso estadísticamente inestable. Quizá se hayan hecho reajustes periódicos a la maquinaria con el fin de buscar continuamente algún valor deseado, pero sin éxito.
- **Figura 14-2h):** La *variación aumenta con el paso del tiempo.* Éste es un problema común en el control de calidad. El efecto neto es que los productos varían más y más hasta que casi todos son defectuosos. Por ejemplo, algunas latas de bebida de cola se derramarán, desperdiciando el producto, y otras no se llenarán por completo y no podrán distribuirse a los consumidores.

Una meta común de muchos métodos diferentes de control de calidad es la siguiente: *reducir la variación* de un producto o servicio. Por ejemplo, la Ford se preocupó por la variación cuando se dio cuenta de que sus transmisiones requerían significativamente más reparaciones por garantía que el mismo tipo de transmisiones fabricadas por Mazda en Japón. Un estudio reveló que las transmisiones de Mazda tenían mucho menos variación en las cajas de velocidades; es decir, las medidas cruciales en las cajas de velocidades variaban mucho menos en las transmisiones Mazda. Aun cuando las transmisiones Ford estaban construidas dentro de los límites permitidos, las transmisiones Mazda eran más confiables gracias a su menor variación. La variación en un proceso puede resultar por dos causas.

Definiciones

La **variación aleatoria** se debe al azar; es el tipo de variación inherente a cualquier proceso que no es capaz de producir un bien o servicio exactamente de la misma forma cada vez.

La **variación asignable** resulta de causas identificables (como maquinaria defectuosa, empleados sin capacitación adecuada, etcétera).

Más adelante en este capítulo, consideraremos formas de distinguir entre la variación asignable y la variación aleatoria.

La gráfica de rachas es una herramienta para supervisar la estabilidad de un proceso. Ahora estudiaremos las *gráficas de control*, que también son sumamente útiles para los mismos propósitos.

Gráfica de control para supervisar la variación: Gráfica *R*

En el artículo "The State of Statistical Process Control as We Proceed into the 21st Century" (Stoumbos, Ryan y Woodall, *Journal of the American Statistical Association*, vol. 95, núm. 451), los autores afirman que "las gráficas de control son una de las herramientas más importantes y más utilizadas en estadística. Sus aplicaciones han pasado de los procesos de fabricación a la ingeniería, las ciencias ambientales, la biología, la genética, la epidemiología, la medicina, las finanzas e incluso al cumplimiento de la ley y los deportes". Iniciamos con la definición de una gráfica de control.

Definición

Una **gráfica de control** de una característica de proceso (como la media o la variación) consiste en valores graficados en secuencia a lo largo del tiempo e incluye una **línea central** así como un **límite de control inferior** (LCI) y un **límite de control superior** (LCS). La línea central representa un valor central de las mediciones características, mientras que los límites de control son las fronteras utilizadas para separar e identificar cualesquiera puntos considerados *fuera de lo común*.

Supondremos que desconocemos la desviación estándar poblacional σ, mientras consideramos únicamente dos de diversos tipos de *gráficas de control*: **1.** las gráficas R (o gráficas de rangos), que se utilizan para hacer un seguimiento de la variación, y **2.** las gráficas $\bar{x}$ que se emplean para verificar medias. Al utilizar gráficas de control para supervisar un proceso, es común que se consideren las gráficas R y las gráficas $\bar{x}$ en conjunto, ya que un proceso estadísticamente inestable puede ser el resultado de un aumento en la variación, de cambios en las medias o de ambos.

Una **gráfica R** (o **gráfica de rangos**) es una gráfica de los rangos muestrales, en vez de valores muestrales individuales, y se aplica para verificar la *variación* en un proceso. (Podría parecer más sensato utilizar desviaciones estándar, pero las gráficas de rangos se emplean con mayor frecuencia en la práctica. Esto es una consecuencia de los tiempos en que no se disponía de calculadoras ni de computadoras. Véase el ejercicio 17, en el que se incluye una gráfica de control basada en desviaciones estándar). Además de graficar los valores de los rangos, incluimos una línea central localizada en $\bar{R}$, que denota la media de todos los rangos muestrales, así como otra línea para el límite de control inferior y una tercera línea para el límite de control superior. A continuación se presenta un resumen de la notación de los componentes de la gráfica R.

Notación

Se sabe que los datos de proceso consisten en una secuencia de muestras, todas del mismo tamaño n, y la distribución de los datos de proceso es esencialmente normal.

n = tamaño de cada muestra o *subgrupo*

$\bar{R}$ = media de los rangos muestrales (es decir, la suma de los rangos muestrales dividida entre el número de muestras)

Verificación de un proceso de variación: Gráfica de control para R

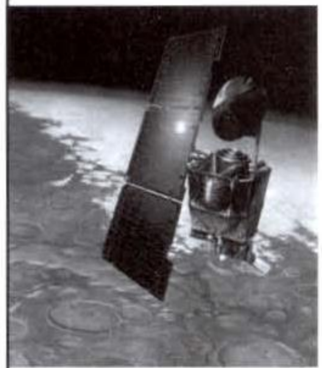
Puntos graficados: rangos muestrales

Línea central: $\bar{R}$

Límite de control superior (LCS): $D_4\bar{R}$ (donde D_4 se encuentra en la tabla 14-2)

Límite de control inferior (LCI): $D_3\bar{R}$ (donde D_3 se encuentra en la tabla 14-2)

Los valores D_4 y D_3 fueron calculados por expertos en control de calidad y sirven para simplificar los cálculos. Los límites de control superior e inferior de $D_4\bar{R}$ y $D_3\bar{R}$ son valores casi equivalentes a los límites de un intervalo de confianza del 99.7%. Por lo tanto, es muy poco probable que los valores de un proceso estadísticamente estable caigan más allá de estos límites. Si un valor cae fuera de esos límites, es muy probable que el proceso no sea estadísticamente estable.



Variación asignable costosa

El Mars Climate Orbiter fue lanzado por la NASA y enviado a Marte, pero se destruyó cuando voló demasiado cerca de su planeta de destino. La pérdida se calculó en \$125 millones. Se descubrió que la causa de la colisión fue la confusión en el empleo de las unidades utilizadas para realizar cálculos. Los datos de la aceleración estaban en las unidades inglesas de libras de fuerza, pero el Jet Propulsion Laboratory supuso que las unidades utilizadas eran "newtons" métricos en vez de libras. Quienes dirigían la nave espacial dieron consecutivamente cantidades erróneas de la fuerza para ajustar la posición de la nave. Los errores causados por la discrepancia fueron muy pequeños al principio, pero el error acumulado a lo largo de los meses de travesía de la nave espacial fue la causa de su fracaso.

En 1962 la nave espacial que transportaba al satélite Mariner I fue destruida por controladores en Tierra, cuando se salió de curso debido a la falta de un signo menos en un programa de cómputo.

¡Cuidado con los ajustes excesivos!

La empresa Nashua Corp. tuvo problemas con su máquina para recubrimiento de papel y consideró gastar millones de dólares para reemplazarla. La máquina estaba funcionando bien y con un proceso estable, pero las muestras se empezaron a tomar con mucha frecuencia; en consecuencia, con base en esos resultados, se le hicieron ajustes. Estos ajustes excesivos, denominados *alteración indebida*, causaron desviaciones de la distribución que hasta entonces había sido buena. El efecto fue un incremento en los defectos. Cuando el experto en estadística y control, W. Edwards Deming, estudió el proceso, recomendó que no se le hicieran ajustes, a menos que hubiera una señal de que el proceso había cambiado o se había vuelto inestable. La compañía funcionó mejor sin ajustes que con la alteración realizada.

Tabla 14-2 Constantes de una gráfica de control

n: Número de observaciones en subgrupo	$\bar{x}$		s		R	
	A_2	A_3	B_3	B_4	D_3	D_4
2	1.880	2.659	0.000	3.267	0.000	3.267
3	1.023	1.954	0.000	2.568	0.000	2.574
4	0.729	1.628	0.000	2.266	0.000	2.282
5	0.577	1.427	0.000	2.089	0.000	2.114
6	0.483	1.287	0.030	1.970	0.000	2.004
7	0.419	1.182	0.118	1.882	0.076	1.924
8	0.373	1.099	0.185	1.815	0.136	1.864
9	0.337	1.032	0.239	1.761	0.184	1.816
10	0.308	0.975	0.284	1.716	0.223	1.777
11	0.285	0.927	0.321	1.679	0.256	1.744
12	0.266	0.886	0.354	1.646	0.283	1.717
13	0.249	0.850	0.382	1.618	0.307	1.693
14	0.235	0.817	0.406	1.594	0.328	1.672
15	0.223	0.789	0.428	1.572	0.347	1.653
16	0.212	0.763	0.448	1.552	0.363	1.637
17	0.203	0.739	0.466	1.534	0.378	1.622
18	0.194	0.718	0.482	1.518	0.391	1.608
19	0.187	0.698	0.497	1.503	0.403	1.597
20	0.180	0.680	0.510	1.490	0.415	1.585
21	0.173	0.663	0.523	1.477	0.425	1.575
22	0.167	0.647	0.534	1.466	0.434	1.566
23	0.162	0.633	0.545	1.455	0.443	1.557
24	0.157	0.619	0.555	1.445	0.451	1.548
25	0.153	0.606	0.565	1.435	0.459	1.541

Fuente: Adaptado del *ASTM Manual on the Presentation of Data and Control Chart Analysis*, © 1976 ASTM, pp. 134-136. Reproducido bajo permiso de American Society of Testing and Materials.

EJEMPLO Fabricación de altímetros para aviones Remítase a los errores de los altímetros de la tabla 14-1. Empleando muestras de tamaño $n = 4$, reunidas cada día de fabricación, construya una gráfica de control para R.

SOLUCIÓN Iniciamos con el cálculo del valor de $\bar{R}$, la media de los rangos muestrales.

$$\bar{R} = \frac{19 + 13 + \cdots + 63}{20} = 21.2$$

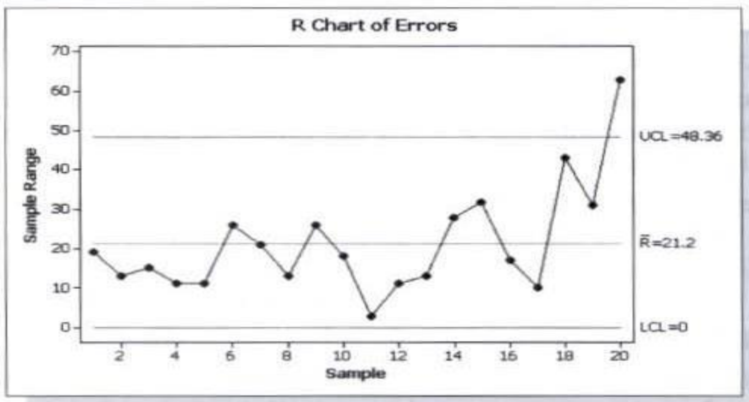
Por lo tanto, la línea central de la gráfica R está ubicada en $\bar{R} = 21.2$. Para calcular los límites de control superior e inferior, debemos obtener los valores de

D_3 y D_4. Si nos remitimos a la tabla 14-2, para $n = 4$, obtenemos $D_3 = 0.000$ y $D_4 = 2.282$, de manera que los límites de control son los siguientes:

$$\text{Límite de control superior: } D_4\overline{R} = (2.282)(21.2) = 48.4$$
$$\text{Límite de control inferior: } D_3\overline{R} = (0.000)(21.2) = 0.0$$

Con un valor de línea central de $\overline{R} = 21.2$ y límites de control de 48.4 y 0.0, ahora procedemos a graficar los rangos muestrales. Los resultados se presentan en la pantalla de Minitab.

Minitab



Interpretación de las gráficas de control

Al interpretar las gráficas de control, el siguiente punto es extremadamente importante:

> **Los límites de control superior e inferior de una gráfica de control están basados en el comportamiento *real* del proceso, no en el comportamiento *deseado*. Los límites de control superior e inferior se desvinculan totalmente de cualesquiera *especificaciones* del proceso estipuladas por el fabricante.**

Al investigar la calidad de algún proceso, hay comúnmente dos preguntas importantes que deben plantearse:

1. Con base en el comportamiento actual del proceso, ¿podemos concluir que el proceso está bajo control estadístico?
2. ¿Cumplen con las especificaciones del diseño los bienes y servicios del proceso?

Los métodos de este capítulo se dirigen a responder la primera pregunta, pero no la segunda. Es decir, nos enfocamos en el comportamiento del proceso con el objetivo de determinar si éste se encuentra bajo control estadístico. El hecho de que el proceso dé como resultado bienes y servicios que cumplen con algunas especificaciones establecidas es otro aspecto que no se cubre con los métodos este capítulo. Por ejemplo, la anterior gráfica R de Minitab incluye límites de control superior e inferior de 48.36 y 0, los cuales resultan de los valores muestrales que se incluyen en la tabla 14-1. Las normas gubernamentales consideran aceptables los errores de los altímetros entre −20 pies y 20 pies, pero tales especificaciones deseadas (o requeridas) no se incluyen en la gráfica de control de *R*.

Además, debemos comprender con claridad los criterios específicos para determinar si un proceso está bajo control estadístico (es decir, si es estadísticamente estable). Hasta ahora, hemos señalado que un proceso no es estadísticamente estable si su patrón se asemeja a cualquiera de los presentados en la figura 14-2. Este criterio se incluye con algunos otros de la siguiente lista.

Criterios para determinar cuando un proceso no es estadísticamente estable (es decir, cuando está fuera de control estadístico)

1. Hay un patrón, una tendencia o un ciclo que evidentemente no es aleatorio (tales como los que se incluyen en la figura 14-2).
2. Hay un punto que está fuera de la región entre los límites superior e inferior. (Es decir, existe un punto por encima del límite de control superior o por debajo del límite de control inferior).
3. *Regla de la racha de 8:* Existen ocho puntos consecutivos, todos por encima o por debajo de la línea central. (En un proceso estadísticamente estable existe una probabilidad de 0.5 de que un punto esté por encima o por debajo de la línea central, de manera que es muy poco probable que ocho puntos consecutivos aparezcan por encima o por debajo de la línea central).

Únicamente utilizaremos los tres criterios listados antes para establecer una falta de control, pero algunas empresas emplean criterios adicionales como éstos:

- Existen seis puntos consecutivos, todos crecientes o decrecientes.
- Hay 14 puntos consecutivos alternantes que se incrementan o disminuyen (tales como incremento, decremento, incremento, decremento y así sucesivamente).
- Dos de cada tres puntos consecutivos están más allá de los límites de control que se encuentran a dos desviaciones estándar de la línea central.
- Cuatro de cada cinco puntos consecutivos están más allá de los límites de control que están a una desviación estándar de la línea central.

PROBLEMA DEL CAPÍTULO 14

EJEMPLO Control estadístico de procesos Examine la gráfica *R* del ejemplo anterior, que aparece en la pantalla de Minitab, y determine si la variación del proceso está bajo control estadístico.

SOLUCIÓN Podemos interpretar gráficas de control de *R* aplicando los tres criterios para establecer una falta de control que listamos anteriormente. Si aplicamos los tres criterios a la gráfica *R* de la pantalla de resultados de Minitab, concluimos que la variación de este proceso está fuera de control estadístico. No hay ocho puntos consecutivos por encima o por debajo de la línea central, de manera que no se viola la tercera condición, pero las primeras dos condiciones no se cumplen.

1. Existe un patrón, tendencia o ciclo que evidentemente no es aleatorio: de izquierda a derecha existe un patrón de tendencia creciente, como en la figura 14-2a).
2. Hay un punto (el punto a la extrema derecha) que está por arriba del límite de control superior.

INTERPRETACIÓN Concluimos que la variación (no necesariamente la media) del proceso está fuera de control estadístico. Como parece que la variación se incrementa con el tiempo, debe tomarse una medida correctiva de inmediato para fijar la *variación* entre los errores de los altímetros.

Gráfica de control para hacer un seguimiento de medias: la gráfica $\bar{x}$

Una **gráfica** $\bar{x}$ es una gráfica de las medias muestrales y se utiliza para llevar un control del *centro* en un proceso. Además de graficar las medias muestrales, incluimos una línea central localizada en $\bar{\bar{x}}$, que denota la media de todas las medias muestrales (igual a la media de todos los valores muestrales combinados), así como otra línea para el límite de control inferior y una tercera para el límite de control superior. Utilizando el método común en los negocios y en la industria, la línea central y los límites de control están basados en rangos y no en desviaciones estándar. Véase el ejercicio 18, que incluye una gráfica $\bar{x}$ basada en desviaciones estándar.

Seguimiento de la media del proceso: gráfica de control de $\bar{x}$

Puntos graficados: medias muestrales

Línea central: $\bar{\bar{x}}$ = media de todas las medias muestrales

Límite de control superior (LCS): $\bar{\bar{x}} + A_2\bar{R}$ (donde A_2 se encuentra en la tabla 14-2)

Límite de control inferior (LCI): $\bar{\bar{x}} - A_2\bar{R}$ (donde A_2 se encuentra la tabla 14-2)

PROBLEMA 14 DEL CAPÍTULO

EJEMPLO Fabricación de altímetros para aviones Remítase a los errores de los altímetros en la tabla 14-1. Empleando las muestras de tamaño $n = 4$, reunidas cada día laboral, construya una gráfica de control de $\bar{x}$. Con base únicamente en la gráfica de control de $\bar{x}$ determine si la media del proceso está bajo control estadístico.

SOLUCIÓN Antes de graficar los 20 puntos correspondientes a los 20 valores de $\bar{x}$, primero debemos calcular el valor de la línea central y los valores de los límites de control. Obtenemos

$$\bar{\bar{x}} = \frac{2.50 + 2.75 + \cdots + 9.75}{20} = 6.45$$

$$\bar{R} = \frac{19 + 13 + \cdots + 63}{20} = 21.2$$

Si nos remitimos a la tabla 14-2, encontramos que para $n = 4$, $A_2 = 0.729$. Conociendo los valores de $\bar{\bar{x}}$, A_2, y $\bar{R}$, ahora podemos evaluar los límites de control.

Límite de control superior: $\bar{\bar{x}} + A_2\bar{R} = 6.45 + (0.729)(21.2) = 21.9$

Límite de control inferior: $\bar{\bar{x}} - A_2\bar{R} = 6.45 - (0.729)(21.2) = -9.0$

INTERPRETACIÓN La gráfica de control de $\bar{x}$ resultante sería como la que aparece en la pantalla de Excel. El examen de la gráfica de control indica que la media del proceso está fuera de control estadístico, porque al menos uno de los tres criterios para establecer una falta de control no se satisface. Específicamente,

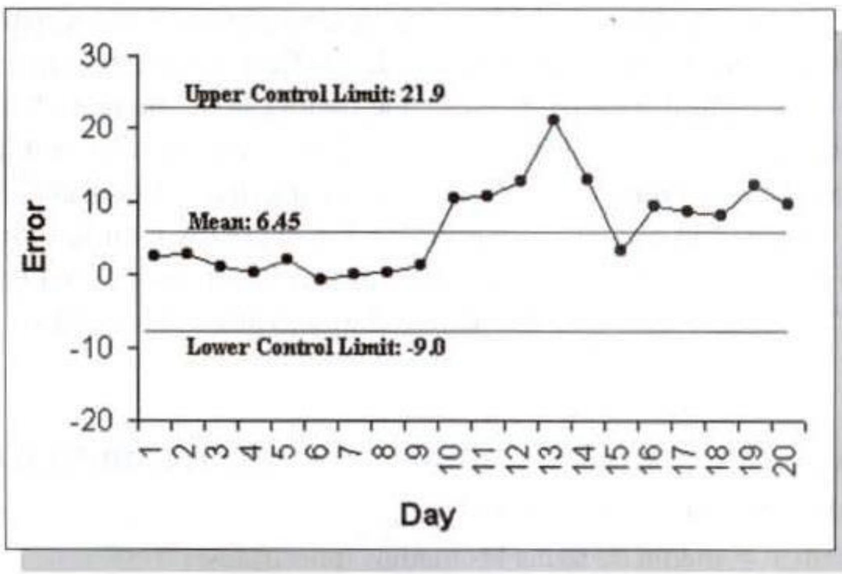
continúa

Detección de soborno con gráficas de control

Las gráficas de control se utilizaron para enviar a prisión a una persona que sobornaba a jugadores de jai alai de Florida para que perdieran. (Véase "Using Control Charts to Corroborate Bribery in Jai Alai", de Charnes y Gitlow, *The American Statistician*, vol. 49, núm. 4). El auditor de una cancha de jai alai notó que cantidades anormalmente grandes de dinero se jugaban en ciertos tipos de apuestas, y que algunos participantes no ganaban tanto como se esperaba cuando se realizaban tales apuestas. Se utilizaron gráficas R y $\bar{x}$ en la Corte como evidencia de patrones sumamente inusuales de apuestas. El examen de las gráficas de control señala claramente puntos que se encuentran muy lejos de límite de control superior, lo que indica que el proceso de apuestas estaba fuera de control estadístico. El especialista en estadística fue capaz de identificar una fecha en la cual la variación asignable parecía detenerse y los fiscales supieron que se trataba de la fecha del arresto del acusado.

el tercer criterio no se satisface porque existen ocho (o más) puntos consecutivos que están por debajo de la línea central. Además, parece existir un patrón de tendencia creciente. Nuevamente, se requieren acciones para corregir el proceso de producción.

Excel

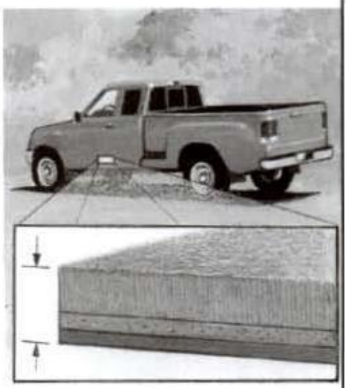


TEXTO N° 22

Graficas de control por atributos.

Compilado y adaptado de Mario Triola Estadística
PEARSON; 2009

Págs. De 748 a 750



Control de calidad en Perstorp

Perstorp Components, Inc. utiliza una computadora que genera automáticamente gráficas de control para verificar el grosor del aislamiento para el piso que fabrica para las camionetas Ford Ranger y Jeep Grand Cherokee. El costo de la computadora de $20,000 se pagó con los ahorros de $40,000 del primer año de operaciones, que se habían empleado para generar gráficas de control manuales que aseguraban que el grosor del aislamiento cumpliera con las especificaciones establecidas entre 2.912 y 2.988 mm. Por medio de gráficas de control y de otros métodos de control de calidad, Perstorp redujo sus mermas en más de dos tercios.

14-3 Gráficas de control para atributos

Concepto clave En esta sección se presenta un método para construir una gráfica de control con el fin de llevar control de la proporción p de algún *atributo* y así saber si el servicio o artículo manufacturado está defectuoso o es discordante. (Un bien o servicio es discordante si no cumple con ciertas especificaciones o requisitos; en ocasiones los bienes discordantes se desechan, se reparan o se denominan "de segunda" y se venden a precios más bajos). La gráfica de control se interpreta usando los mismos tres criterios de la sección 14-2 para determinar si el proceso es estadísticamente estable.

En la sección 14-2 estudiamos gráficas de control para datos *cuantitativos*, pero en esta sección se describe la construcción de gráficas de control para datos *cualitativos*. Como en la sección 14-2, seleccionamos muestras de tamaño n a intervalos de tiempo regulares y dibujamos los puntos en una gráfica secuencial con una línea central y límites de control. (Hay formas de manejar muestras con tamaños diferentes, pero no las estudiaremos aquí).

Definición

La **gráfica de control de p** (o **gráfica p**) es una gráfica de que se dibuja en secuencia en función del paso del tiempo y que incluye una línea central, un límite de control inferior (LCI) y un límite de control superior (LCS).

La notación y los valores de la gráfica de control son los siguientes (donde el atributo de "defectuoso" se puede reemplazar por cualquier otro atributo relevante).

Notación

$\overline{p}$ = estimado agrupado de la proporción de artículos defectuosos en el proceso

$$= \frac{\text{número total de defectos encontrados en todos los artículos muestreados}}{\text{número total de artículos muestreados}}$$

$\overline{q}$ = estimado agrupado de la proporción de artículos del proceso que *no* son defectuosos

$= 1 - \overline{p}$

n = tamaño de cada muestra (no el número de muestras)

Gráfica de control de p

Línea central: $\overline{p}$

Límite de control superior: $\overline{p} + 3\sqrt{\frac{\overline{p}\,\overline{q}}{n}}$

Límite de control inferior: $\overline{p} - 3\sqrt{\frac{\overline{p}\,\overline{q}}{n}}$

(Si el cálculo del límite de control inferior da como resultado un valor negativo, utilice 0 en su lugar. Si el cálculo del límite de control superior excede a 1, utilice 1 en su lugar).

Utilizamos $\overline{p}$ para la línea central porque es el mejor estimado de la proporción de defectos del proceso. Las expresiones de los límites de control corresponden a límites de un intervalo de confianza del 99.7%, como se describió en la sección 7-2.

PROBLEMA DEL CAPÍTULO 14

EJEMPLO Altímetros defectuosos para aviones El problema del capítulo describe el proceso de fabricación de altímetros para aviones. La sección 14-2 incluye ejemplos de gráficas de control para verificar errores en las lecturas de los altímetros. Se considera que un altímetro está defectuoso si no es posible calibrarlo o corregirlo para que proporcione lecturas exactas (que no difieran más de 20 pies de la altitud verdadera). La Altigauge Manufacturing Company produce altímetros en lotes de 100, y cada altímetro se prueba para determinar si está o no defectuoso. A continuación se lista el número de altímetros defectuosos en lotes sucesivos de 100. Construya una gráfica de control para la proporción p de altímetros defectuosos y determine si el proceso está bajo control estadístico. Si no es así, identifique cuál de los tres criterios de control se aplica.

Defectos: 2 0 1 3 1 2 2 4 3 5 3 7

SOLUCIÓN La línea central de la gráfica de control se localiza en el valor de $\overline{p}$:

$$\overline{p} = \frac{\text{número total de defectos de todas las muestras combinadas}}{\text{número total de altímetros muestreados}}$$

$$= \frac{2 + 0 + 1 + \cdots + 7}{12 \cdot 100} = \frac{33}{1200} = 0.0275$$

Puesto que $\overline{p} = 0.0275$, se infiere que $\overline{q} = 1 - \overline{p} = 0.9725$. Al utilizar $\overline{p} = 0.0275$, $\overline{q} = 0.9725$ y $n = 100$, calculamos los límites de control de la siguiente manera:

Límite de control superior:

$$\overline{p} + 3\sqrt{\frac{\overline{p}\,\overline{q}}{n}} = 0.0275 + 3\sqrt{\frac{(0.0275)(0.9725)}{100}} = 0.0766$$

Límite de control inferior:

$$\overline{p} - 3\sqrt{\frac{\overline{p}\,\overline{q}}{n}} = 0.0275 - 3\sqrt{\frac{(0.0275)(0.9725)}{100}} = -0.0216$$

continúa

Seis Sigma en la industria

Seis Sigma es el término utilizado en la industria para describir un proceso que da por resultado una tasa de no más de 3.4 defectos en un millón. La referencia a Seis Sigma sugiere seis desviaciones estándar a partir del centro de una distribución normal, pero el supuesto de un proceso perfectamente estable se reemplaza por el supuesto de un proceso que cambia ligeramente, de manera que la tasa de defectos no es mayor que 3 o 4 por millón.

Los programas Seis Sigma, iniciados alrededor de 1985 en Motorola, ahora intentan mejorar la calidad e incrementar las ganancias al reducir la variación de los procesos. Motorola ahorró más de $940 millones en tres años. Allied Signal reportó ahorros de $1,500 millones. GE, Polaroid, Ford, Honeywell, Sony y Texas Instruments son otras compañías grandes que han adoptado la meta Seis Sigma.

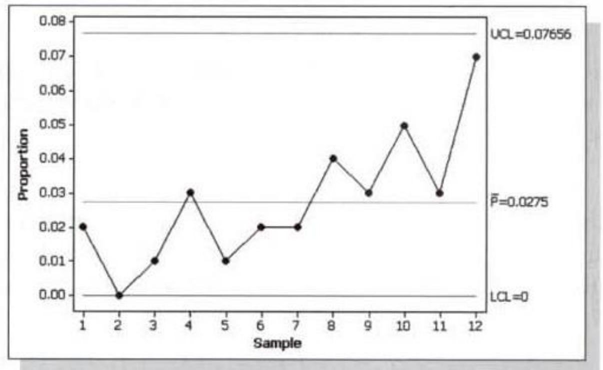
LA ESTADÍSTICA EN LAS NOTICIAS

El alto costo de la baja calidad

La Federal Drug Administration (FDA) recientemente llegó a un acuerdo en el que una compañía farmacéutica, la Schering-Plough Corporation, pagaría la cantidad récord de $500 millones por no lograr corregir problemas en la producción de fármacos. Según un artículo del *New York Times*, de Melody Petersen, "algunos de los problemas se relacionan con la falta de controles que identifican medicamentos defectuosos, mientras otros provienen de equipos demasiado viejos. Tales problemas están relacionados con unos 200 medicamentos, incluyendo el Claritin, el fármaco para alergias que es el producto de mayor venta de Schering".

Puesto que el límite de control inferior es menor que 0, usamos el 0 en su lugar. Una vez que calculamos los valores de la línea central y los límites de control, procedemos a graficar la proporción de altímetros defectuosos. La gráfica de control de p de Minitab se presenta en la siguiente pantalla.

Minitab



INTERPRETACIÓN Podemos interpretar la gráfica de control de p considerando los tres criterios para establecer una falta de control que se listan en la sección 14-2. Con esos criterios podemos concluir que este proceso está fuera de control estadístico por la siguiente razón: parece existir una tendencia creciente. La empresa debe tomar acciones inmediatas para corregir la proporción creciente de defectos.

Práctica de Estadística II - N° 13

TEMA N°13 : PRUEBA DE HIPOTESIS DE REGRESION-INTERVALO DE PREDICCION DE REGRESION

Sección : ……………………..………………..	Apellidos : ……………………..………………
Docente : Escribir el nombre del docente	Nombres : ……………………………………………
	Fecha : ……/……/2017 Duración: 45 minutos
	Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

1.- Presupuestos e ingresos brutos de películas. En la siguiente tabla muestran los presupuestos (en millones de dólares)y los ingresos brutos (en millones de dólares) de películas
Seleccionadas al azar (según datos de la Motion Picture Association of America).

Presupuesto	62	90	50	35	200	100	90
Ingresos brutos	65	64	48	57	601	146	47

Determinar la prueba de Hipotesis teniendo $H_0 : \beta_1=2$, n.s=0.05

2.- Mediciones de presión sanguínea. Catorce estudiantes de medicina del segundo año midieron la presión sanguínea del mismo paciente, y los resultados se presentan a continuación .

Sistólica	138	130	135	140	120	125	130	144
Diastólica	82	90	100	100	80	90	105	95

Determinar la prueba de Hipotesis teniendo $H_0 : \beta_1=0$, n.s= 0.05

3 .- Grillos y temperatura. Una aplicación clásica de la correlación es la asociación entre la temperatura y el número de veces que un grillo chirría en un minuto. A continuación se listan los números de chirridos en un minuto y las temperaturas correspondientes en grados Fahrenheit.

Chirridos en un minuto	882	1188	1104	864	1200	1032	960	900
Temperatura (en °F)	69.7	93.3	84.3	76.3	88.6	82.6	71.6	79.6

Determinar la prueba de Hipotesis teniendo $H_0 : \beta_1=2$ n.s=0.05

4.- Efecto de la variación en un intervalo de predicción. Con los datos que se muestran y suponga que se cumplen las condiciones necesarias de normalidad y varianza.

Peso	3175	3450	3225	3985	2440	2500
Consumo de combustible	27	29	27	24	37	34

a).Calcule la tasa de consumo de combustible predicha para un automóvil que pesa 3700 lb.
b). Calcule un estimado de un intervalo de predicción del 95% de la proporción del consumo de combustible para un automóvil que pesa 3700 lb.

5.- Cálculo del valor predicho y del intervalo de predicción. Con datos que se muestran en la tabla, suponga que se cumplen las condiciones necesarias de normalidad y varianza.

Presupuesto	62	90	50	35	200	100	90
Ingresos brutos	65	64	48	57	601	146	47

a). Calcule el ingreso bruto predicho para una película que tiene un presupuesto de $100 millones.
b). Calcule un estimado de un intervalo de predicción del 95% del ingreso bruto para una película con un presupuesto

de $100 millones.

6.-. Cálculo del valor predicho y del intervalo de predicción. Teniendo los datos que se muestran y suponga que se cumplen las condiciones necesarias de normalidad y varianza.

Sistólica	138	130	130	140	120	144	130
Diastólica	82	91	100	100	80	90	80

a). Calcule la lectura diastólica predicha, dado que la lectura sistólica es de 120.
b). Calcule un estimado de un intervalo de predicción del 95% de la lectura diastólica, dado que la lectura sistólica es de 120.

7.- Cálculo del valor predicho y del intervalo de predicción. Con los datos siguientes suponga que se cumplen las condiciones necesarias de normalidad y varianza.

Chirridos en un minuto	882	1188	1104	864	1200	1032	960	900
Temperatura (en °F)	69.7	93.3	84.3	76.3	88.6	82.6	71.6	79.6

a). Calcule la temperatura predicha cuando un grillo chirría 1000 veces en un minuto.
b). Calcule un estimado de un intervalo de predicción del 99% de la temperatura, cuando un grillo chirría 1000 veces en un minuto.

9.- Teniendo datos sobre las duraciones y los intervalos posteriores a las erupciones, encontramos que para una duración de 180 segundos el mejor intervalo posterior a la erupción predicho es 76.9 min. Construya un intervalo de predicción β_0 al 99% de confianza para el intervalo posterior a una erupción, dado que la duración de la erupción es de 180 segundos (de manera que x= 180).

Duración de la erupción	92	65	72	94	83	94	101	87
Intervalo de tiempo de erupción	204	120	178	234	235	269	255	220

10.- Sobre las duraciones y los intervalos posteriores a las erupciones, encontramos que para una duración de 180 segundos el mejor intervalo posterior a la erupción predicho es 76.9 min. Construya un intervalo de predicción de β_1 al 90% de confianza para el intervalo posterior a una erupción, dado que la duración de la erupción es de 180 segundos (de manera que x= 180).

Duración de la erupción	92	65	72	94	83	94	101	87
Intervalo de tiempo de erupción	204	120	178	234	235	269	255	220

11.- Sobre las duraciones y los intervalos posteriores a las erupciones de un volcán se muestra en la tabla, encontramos que para una duración de 180 segundos el mejor intervalo posterior a la erupción predicho es 76.9 min. Construya un intervalo de predicción de β_1 al 95% de confianza para el intervalo posterior a una erupción, dado que la duración de la erupciones de 200. (de manera que X_0= 200)

Duración de la erupción	92	65	72	94	83	94	101	87
Intervalo de tiempo de erupción	204	120	178	234	235	269	255	220

12.- La tabla que muestra sobre las duraciones y los intervalos posteriores a las erupciones, encontramos que para una duración de 180 segundos el mejor intervalo posterior a la erupción predicho es 76.9 min. Construya un intervalo de predicción de β_1 al 95% de confianza para el intervalo posterior a una erupción, dado que la duración de la erupciones de 120 segundos (de manera que x= 120).

Duración de la erupción	92	65	72	94	83	94	101	87
Intervalo de tiempo de erupción	204	120	178	234	235	269	255	220

Preguntas de gráficas de control de atributos

1. El área de créditos de un banco de Huancayo se encarga de ingresar cada transacción al estado de cuenta mensual de los clientes. La exactitud es fundamental y los errores causarían el descontento de los clientes. Para evitar equivocaciones, cada empleado que ingresa los datos teclea una muestra de 1200 de su lote de trabajo una segunda vez, y un software verifica que los números concuerden. El software imprime además un informe acerca del número y tamaño de cualquier discrepancia. Seis empleados trabajaron durante la última hora cuyos resultados se muestran:

No	Nº inspeccionado	Nº que no concuerdan
1	1200	5
2	1200	3
3	1200	6
4	1200	4
5	1200	14
6	1200	4
Total		

Elabore un diagrama de porcentaje de defectuosos para este proceso. ¿Cuáles son los límites de control superior e inferior? Interprete los datos.
¿Parecería que algunos de los encargados de ingresar los datos están “fuera de control” Explique

2. El editor de un diario está interesado en determinar el número de palabras mal escritas que se publican en su diario. Para controlar el problema y promover la necesidad de una escritura correcta, se utilizará un diagrama de control. El número de palabras con errores en la edición final del periódico durante los últimos 7 días son: 4; 2; 5; 6; 0; 3 y 1. Determine los límites de control adecuados e interprete el diagrama. ¿Hubo algunos días de ese periodo en los que el número de palabras mal escritas haya estado fuera de control?. El número total de palabras mal escritas en el lapso de 7 días es de 21.

3. Un fabricante de triciclos selecciona diariamente al azar 8 armazones y determina la cantidad de defectos. El número de armazones defectuosos encontrado en los últimos 10 días es: 2; 1; 3; 2; 7; 2; 0; 3; 5 y 4. Elabore un diagrama de control para este proceso y determine si está “bajo control”.

4. Una cadena de supermercados evalúa el trabajo de sus cajeros examinando al azar las boletas de venta impresos para verificar si hay errores. Los siguientes datos describen el número de errores en cada boleta que se obtuvo el lunes de la semana pasada: 1; 1; 0; 2; 1; 3; 0; 2; 1. Elabore el diagrama de control para el proceso y determine si está o no “bajo control”.

5. Una empresa papelera pone a prueba el papel higiénico que produce sometiendo 15 rollos a una prueba de esfuerzo en humedad, y comprobando si el papel se rompe durante la prueba y con qué frecuencia. A continuación se muestra el número de rollos defectuosos encontrados en los últimos 10 días: 3; 2; 1; 2; 2; 1; 0; 3; 2 y 0. Elabore el diagrama de control para el proceso y determine si está o no “bajo control”

Práctica de Estadística II - N° 14

TEMA N° 14: REGRESIÓN MÚLTIPLE

Sección : ……………………………..………………..	Apellidos : ……………………….………………
Docente : Escribir el nombre del docente	Nombres : ……………………………………………
	Fecha : ……/……/2017 Duración: 45 minutos
	Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

01.Obtención del mejor modelo. En los ejercicios, construya un diagrama de dispersión e identifique el modelo matemático que se ajusta mejor a los datos dados. Suponga que el modelo se va a emplear únicamente para el alcance que tienen los datos, y considere sólo los modelos lineales, cuadrático, logarítmico, exponencial y potencial.

1.

x	1	2	3	4	5	6
y	8	2	0	2	8	18

2.

x	1	2	3	4	5	6
y	3	8	13	18	23	28

3.

x	1	2	3	4	5	6
y	3	9	27	80	245	725

4.

x	1	2	3	4	5	6
y	2.000	2.828	3.464	4.000	4.472	4.899

01. La tabla lista las cantidades de los incrementos semanales de los salarios y (en dólares), especificadas en un contrato laboral negociado con empleados de la corporación Telektronic.

Año	1	2	3	4	5
Incremento (y)	10	12	14	16	18

02. La tabla lista el valor y (en dólares) de $100 invertidos en un certificado de depósito del MetLife Bank.

Año	1	2	3	4	5	6	7
Valor	100	105	110.25	115.76	121.55	127.63	134.01

03. La tabla lista la distancia d (en pies) por encima del suelo para el caso de un objeto que se deja caer en el vacío desde una altura de 500 pies. El tiempo t (en segundos) es el tiempo que transcurre desde que se suelta el objeto.

t	1	2	3	4	5
d	484	436	356	244	100

04. La tabla lista los costos y (en dólares) de la compra de cierto volumen de tierra vegetal, donde dicho volumen es un cubo en el que cada lado tiene una longitud de x pies.

x	1	2	3	4	5
Costo	1.10	8.80	29.70	70.40	137.50

05. Un experimento para una clase de física implica dejar caer una pelota de golf y registrar la distancia (en metros) que cae en diferentes tiempos (en segundos) después de ser soltada. Los datos se incluyen en la siguiente tabla. Proyecte la distancia para un tiempo de 12 segundos, considerando que la pelota de golf se dejó caer de un edificio con una altura de 50 m.

Tiempo	0	0.5	1	1.5	2	2.5	3
Distancia	0	1.2	4.9	11.0	19.5	30.5	44.0

06. A continuación, se presentan las temperaturas medias globales (en °C) de la superficie de la Tierra para los años 1950, 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000 y 2005. Calcule la temperatura predicha para el año 2010.

13.8	13.9	14.0	13.9	14.1	14.0	14.3	14.1	14.5	14.5	14.4	14.8

07. A continuación, se presentan las concentraciones de dióxido de carbono (en partes por millón) en la atmósfera de la Tierra para los años 1950, 1955, 1960, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000 y 2005. Calcule la concentración predicha de dióxido de carbono para el año 2010.

311	314	317	320	326	331	339	346	354	361	369	381

08. A continuación, se presenta el número de muertes en Estados Unidos, como resultado de choques de vehículos automotores. Utilice el mejor modelo y el segundo mejor modelo con la finalidad de calcular el número proyectado de este tipo de muertes para el año 2010. ¿Difieren mucho las dos estimaciones?

Año	1975	1980	1985	1990	1995	2000	2005
Muertes	44,525	51,091	43,825	44,599	41,817	41,945	43,443

Práctica de Estadística II - N° 15-16
TEMA 15-16 : Control de Calidad

Sección : …………………………..……………...

Docente : Escribir el nombre del docente

Apellidos : ………………………..……………….
Nombres : …………………………………………….
Fecha : ……/……/2017 Duración: 45 minutos
Tipo de Práctica: Individual ( ) Grupal (X)

Instrucciones: Responda a cada una de las preguntas de acuerdo a lo solicitado

Verificación de consumo de energía doméstica. En los ejercicios, utilice la siguiente información sobre el consumo de energía eléctrica (en kilowatt-hora) en una casa hospedaje al norte Huancayo, durante intervalos de dos meses por cuatro años. Los resultados se listan en la tabla:

	Ene.-Feb.	Mar.-Abr.	May.-Jun.	Jul.-Ago.	Sep.-Oct.	Nov.-Dic.
Año 1	4762	3875	2657	4358	2201	3187
Año 2	4504	3237	2198	2511	3020	2857
Año 3	3952	2785	2118	2658	2139	3071
Año 4	3863	3013	2023	2953	3456	2647

1. **Consumo de energía construcción de una gráfica de rachas:** construye una gráfica de rachas para los 24 valores. ¿Parece haber un patrón que sugiera que el proceso no está bajo control estadístico? ¿Existe algún patrón de variación que pueda explicarse?
2. **Consumo de energía. construcción de una gráfica R:** Utilice muestras de tamaño 3, combinando los primeros tres valores de cada año y los últimos tres valores de cada año. Con las ocho muestras de tamaño 3, construya una gráfica R y determine si la variación del proceso está bajo control estadístico. Si no es así, identifique cuál de los tres criterios para establecer una falta de control conduce al rechazo de una variación estadísticamente estable.
3. **Consumo de energía: construcción de una gráfica x** : Utilice muestras de tamaño 3, combinando los primeros tres valores de cada año y los últimos tres valores de cada año. Con las ocho muestras de tamaño 3, construya una gráfica xy determine si la media del proceso está bajo control estadístico. Si no es así, identifique cuál de los tres criterios para establecer una falta de control conducen al rechazo de una media estadísticamente estable. ¿Cuál es un efecto práctico de no tener bajo control estadístico tal proceso? Dé un ejemplo de una causa que pondría a un proceso fuera de control estadístico.

4. Gráfica p para las tasas de matrimonios Utilice gráficas p para comparar la estabilidad estadística de la tasa de matrimonios de Japón y Estados Unidos. De cada año, se seleccionaron aleatoriamente 10,000 personas de cada país; el número de matrimonios que se obtuvo corresponde a ocho años consecutivos recientes (datos de las Naciones Unidas).

Japón:	58	60	61	64	63	63	64	63
Estados Unidos:	98	94	92	90	91	89	88	87

BIBLIOGRAFÍA

- Triola, Mario F. Estadística. Pearson Educación, México 2013.
-

COMPLEMENTARIA

- Berenson, Mark y Levine, David. Estadística para Administración. Pearson. México 2014.
- Jorge Inafuko, Jorge Rubio. Estadística Aplicada. Universidad del Pacífico. 2014.
- Newbold Paúl, Carlson William y Thorne Betty. Estadística para Administración y Economía. Pearson Prentice Hall. España 2013.
- Lind, Douglas, Marchal, William; Wathen, Samuel. Estadística Aplicada a los Negocios y a la Economía. Mc,Graw Hill. México 2012.
- Alfredo Díaz Mata. Estadística Aplicada a la Administración y la Economía. Mc Graw Hill. 2013.
- Ronald M. Weiers. Introducción a la Estadística para negocios. Cengage Learning 2011.
-

RECURSOS DIGITALES

- Instituto Nacional de Estadística e Informática. En: http://www.inei.gob.pe/
- Sistema Integrado de Información de Comercio Exterior del Perú. SIICEX. En
- http://www.siicex.gob.pe/siicex/portal5ES.asp?_page_=160.00000
- Gráficos Interactivos Hans Rosling. www.bit.ly/cApMnI y www.bit.ly/angl9S
- Data Mining Institute, En: http://www.webmining.cl/2011/04/que-es-data-mining/
- Estadística Descriptiva con una variable. Material de Consulta. En:
- http://www.uam.es/personal_pdi/psicologia/carmenx/MaterialD.html.