ESTADÍSTICA Y PROBABILIDAD
con aplicaciones

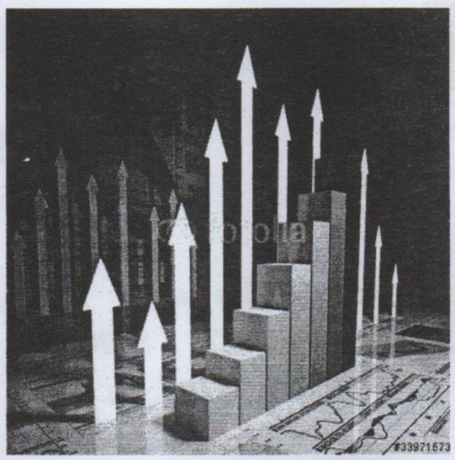
TOMO 2

Lic. Gabriel Leandro Oviedo, MBA

PUBLITEX
Grupo Editorial S.A.

Lic. Gabriel Leandro Oviedo, MBA

ESTADÍSTICA Y PROBABILIDAD CON APLICACIONES



TOMO 2

PUBLITEX
Grupo Editorial S.A.

519.5
L437es Leandro Oviedo, Gabriel
Estadística descriptiva con aplicaciones / Gabriel Leandro Obando -- 1ª. ed. -- San José, C.R.: Publitex, 2014
324 p.; 24.5 x 17 cm

ISBN: 978-9977-987-72-9 Tomo 2

1. Estadística descriptiva - Enseñanza. I. Título.

Edición 2014

Levantado de texto, artes, diagramación y montaje
Publitex Grupo Editorial S.A.
Diseño de portada e ilustraciones
Betty Quirós Marín
Impresión en sistema digital
Publitex Grupo Editorial S.A.
Dirección Editorial
Eduardo Marín Gutiérrez

PUBLITEX
Grupo Editorial S.A.

Agradecimientos

Muchas personas me han brindado su colaboración en la creación de todo este material, pero en forma especial desearía agradecer a mis amigos y compañeros Lipcia Munguía, Erick Torres, Pablo Calderón y Félix Amado por sus valiosos comentarios y sugerencias en el desarrollo de este texto. Igualmente a Natalie Leitón, a doña Higinia Esquivel y don Orlando Saborío por el apoyo que me han dado. Por otro lado a mis estimados compañeros Felipe Masís, Edgar Chaves, Héctor Guerra, Rodolfo Mainieri y don Jorge Acuña. Por supuesto a don Rodrigo Ortiz Astúa, de la Editorial Guayacán, por la gran oportunidad que me ha brindado al abrirme las puertas de esta prestigiosa editorial. También a mi esposa e hijas por su paciencia, apoyo e inspiración, y sobre todo al Señor Jesús que me ha dado la vida y toda capacidad, a quien doy toda honra.

Prefacio

Hay muchos factores que influyen en el desempeño de un estudiante en un curso de estadística. Entre ellos están la clara exposición de los conceptos por parte del profesor, la motivación y empeño por parte del estudiante, los conocimientos previos del alumno, materiales apropiados para el curso, etc. Este texto y todos los materiales digitales que lo acompañan han sido diseñados para ayudar al estudiante y al profesor en todos estos aspectos, pues provee al profesor de materiales que puede emplear en su clase para exponer los conceptos con claridad, presenta al estudiante materiales con ejercicios paso a paso, aplicaciones y amplio uso de la tecnología, de modo que pueda sentirse más motivado al disponer de recursos para adquirir los distintos conceptos y procedimientos, a la vez que se le ofrece gran cantidad de ejercicios resueltos, presentaciones interactivas, videos, entre otras ventajas.

Características didácticas

A lo largo del texto, las definiciones se presentan en recuadros de modo que el estudiante pueda efectuar un repaso posterior de estos conceptos, los cuales son de gran importancia para comprender los ejercicios e interpretar los resultados obtenidos.

> **UNIDAD ESTADÍSTICA:** Es aquella unidad u objeto de interés en la investigación y de ella se desprenden las observaciones, o sea, de ella se derivan los datos para el análisis.

Adicionalmente cada tema incluye gran cantidad de ejemplos. Los ejemplos en el texto impreso se presentan en cuadros que incluyen primero el planteamiento del ejercicio y su solución explicada paso a paso.

Ejemplo:

Suponga que una empresa posee quince vendedores de un determinado producto. Cuatro de los vendedores lograron vender 50 unidades, 6 vendieron 40 unidades, tres vendieron 35 unidades y 2 vendieron 20 unidades. ¿Cuál es el número de unidades promedio de cada vendedor?

Solución:

Dado que existen valores repetidos, entonces se aplica la fórmula:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n} = \frac{4 \cdot 50 + 6 \cdot 40 + 3 \cdot 35 + 2 \cdot 20}{15} = 39$$

Es decir, el número de unidades vendidas por cada vendedor es 39 unidades.

Luego de los ejemplos se incluye un ejercicio de revisión, que es similar al ejemplo. La finalidad es que el estudiante pueda efectuar ese ejercicio luego de leer y realizar por cuenta propia el ejemplo. Las soluciones de los ejercicios de revisión se proveen en la página de internet del texto www.auladeeconomia.com/raeep.html.

Ejercicio de revisión:

En un muelle hay 20 contenedores que pesan 15 toneladas cada uno, 25 que pesan 20 toneladas cada uno. ¿Cuál es el peso promedio de los contenedores?

Ver solución de este ejercicio en la página de internet de este texto www.auladeeconomia.com/raeep.html

Al final de cada capítulo se proveen muchos ejercicios. Estos ejercicios se dividen en dos bloques. Los primeros son ejercicios de desarrollo. Los segundos son ejercicios se selección única y se les llama examen del capítulo. Las soluciones de estos ejercicios se proveen a través de la página de internet del texto.

Ejercicios:

Ejercicios de desarrollo:

1. A continuación se presenta una lista de variables obtenidas en distintas investigaciones, en cada caso indique cuál tipo de gráfico debería efectuarse para representar en forma adecuada los datos.

   a. Tiempo medio de espera de los pacientes de un hospital en el servicio de consulta externa obtenido mes a mes durante un año.

Examen del capítulo

1. Se tiene un grupo de *n* libros. El número de diferentes órdenes posibles de los *n* libros en una mesa no equivale a:

(a) $P(n,n)$ (b) $n!$
(c) $C(n,n)$ (d) Ninguna de las anteriores

2. Suponga que los *n* libros se van a conformar en grupos de 3 libros (n>3). El número de diferentes grupos con distinto orden, equivale a:

(a) $P(n,3)$ (b) $n!/3!$
(c) $C(n,3)$ (d) $P(n,n-3)$

Por otro lado, cabe destacar que Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad es en realidad más que solo un libro, pues se compone de una gran cantidad de recursos en línea de gran utilidad tanto para el estudiante como para el profesor.

Con la factura de compra del texto cada lector podrá recibir una contraseña de ingreso a la página del texto www.auladeeconomia.com/raeep.html y así acceder a una gran cantidad de

- Más de 2000 **diapositivas** que puede ser presentadas por el profesor para impartir su clase, o bien, empleados por el estudiante para repasar posteriormente.
- **Videos** que explican los conceptos y que exponen el uso del software.
- **Ejercicios interactivos**. El texto y las diapositivas contienen más de mil ejercicios con sus soluciones
- **Uso de software Excel y Minitab y uso de la calculadora**. El uso de la tecnología es fundamental en el procesamiento, presentación y análisis de datos y por eso se le concede gran importancia en el texto.

- Además se incluye una gran colección de **otros recursos** disponibles en la web, tales como links a páginas con bases de datos, simulaciones (applets), calculadoras de probabilidades y medidas estadísticas, entre otros recursos disponibles en internet.

El lector debe escribir a info@auladeeconomia.com y así recibirá una contraseña para acceder a los recursos completos.

A lo largo del texto se incluyen cuadros que indican algunos de los principales recursos audiovisuales que pueden ser empleados por estudiantes y profesores.

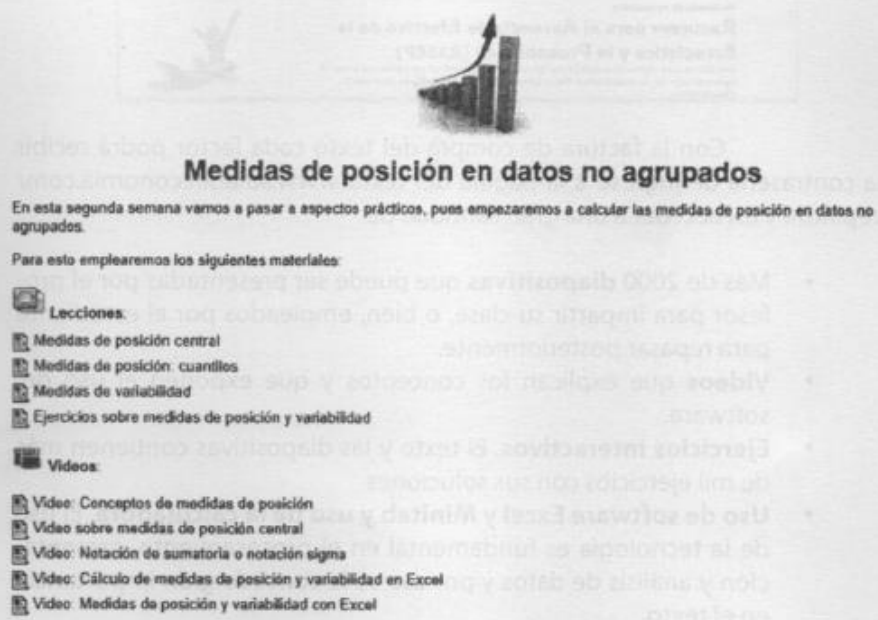
Material audiovisual:

En la página de internet de este texto podrá encontrar una presentación de diapositivas que expone este tema y es un complemento a este texto.

Introducción a la Estadística descriptiva
Por Lic. Gabriel Leandro, MBA
www.auladeeconomia.com

Solicite su usuario y contraseña escribiendo al correo electrónico **info@auladeeconomia.com.** Para ello deberá indicar el número de su factura de compra del texto.

En la página los temas están agrupados de modo similar al que se emplea en el texto, tal como se muestra en la imagen, donde se pueden elegir los diferentes recursos.



Medidas de posición en datos no agrupados

En esta segunda semana vamos a pasar a aspectos prácticos, pues empezaremos a calcular las medidas de posición en datos no agrupados.

Para esto emplearemos los siguientes materiales:

Lecciones:

- Medidas de posición central
- Medidas de posición: cuantiles
- Medidas de variabilidad
- Ejercicios sobre medidas de posición y variabilidad

Videos:

- Video: Conceptos de medidas de posición
- Video sobre medidas de posición central
- Video: Notación de sumatoria o notación sigma
- Video: Cálculo de medidas de posición y variabilidad en Excel
- Video: Medidas de posición y variabilidad con Excel

Adicionalmente, en cada capítulo se presenta uno o varios cuadros de Aplicación, los cuales consisten en algunos ejemplos de aplicaciones de la estadística en distintos ámbitos, como las ciencias económicas, la ingeniería, las ciencias sociales y las ciencias de la salud, entre otras áreas del conocimiento humano.

Aplicación:

Índice Dow Jones
Posiblemente en alguna oportunidad usted habrá escuchado noticias sobre el mundo de los negocios que se relacionen con la Bolsa de Valores de Nueva York y entonces habrá escuchado hablar sobre el Dow Jones, el cual es el indicador bursátil más conocido en el planeta.

En términos generales, este material pretende ser una valiosa ayuda para estudiantes y profesores de modo que puedan seguir un mejor proceso de enseñanza – aprendizaje, ya que al disponer de gran cantidad de recursos a través de internet ofrece una serie de ventajas que los textos tradicionales no proveen:

- **Está siempre disponible**. Es como tener al profesor disponible las 24 horas del día y en cualquier lugar donde haya conexión a internet.
- **Es interactivo.** El texto tradicional no es interactivo, pero Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad es totalmente interactivo. Muestra una explicación paso a paso y ejercicios solucionados paso a paso. Usted da clic y el material lo lleva a su ritmo.
- **Es motivador e innovador**. Actualmente los jóvenes emplean intensivamente las tecnologías de información y la comunicación. De hecho, la mayoría de la población joven emplea internet como su principal herramienta para investigar y estudiar.
- **Se ajusta a distintos estilos de aprendizaje**. Bien es sabido que cada persona posee diferente modo de aprender. Algunos aprenden más por lo que ven, otros aprenden más por lo que oyen, y aprendemos mucho por lo que hacemos. Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad ofrece presentaciones con múltiples imágenes, videos, herramientas tecnológicas, entre otros recursos que ayudan al estudiante a aprender según su estilo de aprendizaje.
- **Fomenta el uso de la tecnología.** Tanto a través del uso de software como Excel y Minitab como por los recursos disponibles en la web se realiza un uso intensivo de las tecnologías de la información y la comunicación.

- **Facilita aprender haciendo**. Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad promueve una modalidad de estudio en la que se aprende haciendo, pues usted puede ir desarrollando los ejercicios en su cuaderno o computadora conforme se presentan en el material.
- **Expone aplicaciones diversas**: Recursos para el Aprendizaje Efectivo de la Estadística y la Probabilidad posee ejemplos y ejercicios que se relacionan con la ingeniería, las ciencias económicas, las ciencias de la salud, las ciencias sociales, etc.
- **Siempre está actualizado**: Gracias a internet los lectores podrán tener acceso a nuevas actualizaciones del material disponible en línea.
- **Apoyo a los docentes**: A través de la página del texto se proveen recursos para que los docentes puedan desarrollar mejor sus cursos, a la vez que se propicia el intercambio de experiencias entre profesores.

CONTENIDOS

Agradecimiento 5
Prefacio 7

Capítulo 1
Distribución de probabilidad de variable discreta
1.1 Distribuciones de probabilidad 16
1.2 Media y varianza de una distribución de probabilidad 18
1.3 Distribución binomial 22
1.4 Distribución hipergeométrica 29
1.5 Distribución de Poisson 35
1.6 Aproximación de la distribución de Poisson a la binomial 40
1.7 Distribución multinomial 42
1.8 Distribución geométrica 44
1.9 Ejercicios 49

Capítulo 2
Distribuciones de probabilidad de variable continua
2.1 Distribuciones continuas de probabilidad 66
2.2 La distribución normal 66
2.3 Aproximación de la distribución normal binomial 82
2.4 Distribución exponencial 86
2.5 Ejercicios 92

Capítulo 3
Estimación por intervalos
3.1 Inferencia estadística 104
3.2 Distribución muestral de la medida si σ es conocida 104
3.3 Teorema de límite central 109
3.4 Distribución muestral de la medida con σ desconocida
(distribución *t* de Student) 109
3.5 Inferencia para la media 111
3.6 Intervalos de confianza para la media y error máximo de la estimación ...
3.7 Inferencia para proporciones 111
3.8 Ejercicios 122
125

Capítulo 4
Muestreo
4.1 Introducción 140
4.2 Necesidad de trabajar con muestras 143
4.3 Muestreo estadístico y no estadístico 145
4.4 Determinación del tamaño de muestra para estimar la media poblacional 147
4.5 Determinación del tamaño de la muestra 146

4.6 Cálculo del tamaño de la muestra ____ 149
4.7 Técnicas de muestreo ____ 158
4.8 Ejercicios ____ 171

Capítulo 5
Pruebas de hipótesis
5.1 Inferencia mediante pruebas de hipótesis ____ 186
5.2 Procedimiento para pruebas de hipótesis sobre la media ____ 193
5.3 Prueba de hipótesis con muestras pequeñas ____ 197
5.4 Prueba de hipótesis para la proporción poblacional ____ 198
5.5 Intervalos de confianza y pruebas de hipótesis ____ 205
5.6 Ejercicios ____ 207

Capítulo 6
Pruebas de hipótesis para la diferencia de dos medias o proporciones poblacionales
6.1 Pruebas de hipótesis para la diferencia entre medias ____ 224
6.2 Diferencia entre medias en poblaciones independientes ____ 224
6.3 Diferencia entre medias (muestras pequeñas) ____ 230
6.4 Observaciones pareadas ____ 235
6.5 Pruebas para la diferencia de dos proporciones ____ 240
6.6 Ejercicios ____ 244

Capítulo 7
Correlación lineal y regresión lineal simple
7.1 Asociación estadística entre dos variables ____ 264
7.2 Correlación final ____ 265
7.3 Correlaciones espurias y causalidad ____ 279
7.4 Regresión lineal simple ____ 280
7.5 Interpolación y extrapolación ____ 289
7.6 Ejercicios ____ 291

Otros temas ____ 309

Apéndice
Tablas y fórmulas
Apéndice 1: Fórmulas de estadística descriptiva ____ 313
Medidas de variabilidad ____ 314
Apéndice 2: Fórmulas de probabilidad ____ 315
Apéndice 3: Fórmulas de distribución de probabilidad ____ 316
Apéndice 4: Fórmulas de inferencia estadística ____ 317
Apéndice 5: Fórmulas de regresión y correlación lineal simple ____ 318
Apéndice 6: Distribución normal estándar acumulada ____ 319
Apéndice 7: Distribución t de Student ____ 320
Apéndice 8: Percentiles de la distribución Chi-Cuadrado ____ 321
Apéndice 9: Valores de F con probabilidad de 5% ____ 322
Apéndice 10: Tabla de números aleatorios ____ 323

CAPÍTULO 1

DISTRIBUCIÓN DE PROBABILIDAD DE VARIABLE DISCRETA

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Calcular la media y la varianza de una distribución de probabilidad
2. Resolver problemas empleando la distribución binomial
3. Resolver problemas empleando la distribución hipergeométrica
4. Resolver problemas empleando la distribución de Poisson
5. Resolver problemas empleando la distribución multinomial
6. Resolver problemas empleando la distribución geométrica

1.1 Distribuciones de probabilidad

Para hablar de distribuciones de probabilidad es necesario presentar el concepto de experimento estadístico.

EXPERIMENTO ESTADÍSTICO Es un proceso mediante el cual se generan observaciones aleatorias.

Es decir, en un experimento estadístico se producen observaciones al azar, y generalmente a esas observaciones se les asigna una descripción numérica, aunque pueden ser cualitativas, tal como se expuso en el capítulo 1. Por ejemplo, lanzar un dado puede ser un experimento aleatorio, pues el resultado es producto del azar. Ahora bien, no solo juegos de ese tipo son experimentos aleatorios, pues también podrían ser el número de piezas defectuosas en un proceso productivo, el número de personas en la fila de una farmacia, el número de llamadas que ingresan en una hora cualquiera en un centro de llamadas, entre muchas otras posibilidades.

VARIABLE ALEATORIA Cantidad numérica cuyo valor se determina a través de un *experimento aleatorio.*

En otras palabras, las variables aleatorias toman valores al azar en cada caso. Las variables aleatorias se pueden clasificar de acuerdo con el número de valores que pueden asumir:

- **Variables aleatorias discretas:** si se puede contar su conjunto de resultados posibles, por ejemplo: número de artículos defectuosos, número de personas en una fila, números de datos procesados etc.
- **Variables aleatorias continuas**: Cuando la variable aleatoria puede tomar valores en una escala continua se llama ***variable aleatoria continua***. Por regla general las variables aleatorias continuas representan datos medidos. Por ejemplo: la velocidad a la que va un grupo de vehículos, estaturas, longitud, temperaturas, diámetros, espesor, intensidad del ruido, etc.

Una distribución de probabilidades permite determinar el valor de la probabilidad para todos y cada uno de los eventos del espacio muestral. La distribución de probabilidad puede expresarse empleando una tabla, un gráfico o una función algebraica.

Ejemplo

Suponga que se lanza al aire una moneda dos veces para ver si cae "cara" (evento A) o "cruz" (evento B). Construya la tabla de la distribución de probabilidad.

Solución

En este caso existen 4 resultados posibles, cada uno con las siguientes probabilidades:

Evento	Probabilidad
AA	0,25
AB	0,25
BB	0,25
Total	0,25

La tabla anterior es la distribución de probabilidad para el experimento "lanzar al aire una moneda dos veces".

Ejemplo

Suponga que se está efectuando el siguiente juego de dados: el jugador hace una apuesta y lanza los dos dados. Si la suma de los puntos es 7 u 11, gana el monto apostado. Pero si sale cualquier otra suma, pierde el monto apostado. Construya la distribución de probabilidad para la suma de los puntos de los dos dados y la distribución de probabilidad para los resultados del juego.

Solución

En este caso existen resultados posibles, cada uno con las siguientes probabilidades:

Evento	Sumas	Probabilidad
2	1 + 1	1/12
3	1 + 2, 2 + 1	2/12
4	1 + 3, 2 + 2, 3 + 1	3/12
5	1 + 4, 2 + 3, 3 + 2, 4 + 1	4/12
6	1 + 5, 2 + 4, 3 + 3, 4 + 2, 5 + 1	5/12
7	1 + 6, 2 + 5, 3 + 4, 4 + 3, 5 + 2, 6 + 1	6/12
8	2 + 6, 3 + 5, 4 + 4, 5 + 3, 6 + 2	5/12
9	3 + 6, 4 + 5, 5 + 4, 6 + 3	4/12
10	4 + 6, 5 + 5, 6 + 4	3/12
11	5 + 6, 6 + 5	2/12
12	6 + 6	1/12
Total	-	**1,00**

La tabla anterior es la distribución de probabilidad para el experimento "lanzar al aire una moneda dos veces".

Ejercicio de revisión

En cada uno de los siguientes casos construya la tabla de la distribución de frecuencias:

1. Una rifa consta de 100 números a un precio de \$20 cada uno. El premio es de \$1000 (premio único) y el jugador compra un número.
2. Una rifa consta de 100 números a un precio de \$20 cada uno. El premio es de \$1000 (premio único) y el jugador compra dos números.
3. En una caja hay 10 bolas, 2 son azules, 3 son verdes y 5 son rojas. Se saca una bola y si la bola es azul se ganan cero puntos, si es verde se gana un punto y si es roja se ganan dos puntos.
4. En una caja hay 10 bolas, 2 son azules, 3 son verdes y 5 son rojas. Se sacan dos bolas y se suman los puntos sabiendo que si la bola es azul se ganan cero puntos, si es verde se gana un punto y si es roja se ganan dos puntos.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video que exponen este tema.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

1.2 Media y varianza de una distribución de probabilidad

La media de una distribución de probabilidad es la esperanza matemática o valor esperado de la variable aleatoria correspondiente. Si las probabilidades de los valores x_i son $P(x_i)$ entonces:

La media es:

$$E(X) = \mu = \sum xP(x)$$

Y la varianza es:

$$\sigma^2 = \sum (x - E(x))^2 P(x)$$

La desviación estándar es la raíz cuadrada de la varianza:

$$\sigma = \sqrt{\sigma^2}$$

Ejemplo

Calcule la media y la desviación estándar de la demanda semanal de cierto artículo en una ferretería. Los datos de demanda y su probabilidad de ocurrencia se dan en la tabla.

Unidades vendidas χ_i	30	35	40	45	50
Probabilidad $P(\chi_i)$	0,20	0,28	0,30	0,15	0,07

Solución

La media o valor esperado es:

$$E(X) = \mu = \sum x_i P(x_i)$$
$$= 30 \cdot 0.2 + 35 \cdot 0.28 + 40 \cdot 0.30 + 45 \cdot 0.15 + 50 \cdot 0.07 = 38.05$$

Y la varianza es:

$$\sigma^2 = \sum (x_i - \mu)^2 P(x_i) = 0.2(30 - 38.05)^2 + 0.28(35 - 38.05)^2$$
$$+0.30(40 - 38.05)^2 + 0.15(45 - 38.05)^2 + 0.07(50 - 38.05)^2 = 33.95$$

La desviación estándar es la raíz cuadrada de la varianza:

$$\sigma = \sqrt{\sigma^2} = \sqrt{33.95} = 5.83$$

Ejercicio de revisión

En cada uno de los siguientes casos, a partir de la tabla de la distribución de frecuencias, calcule el valor esperado (media) y la desviación estándar:

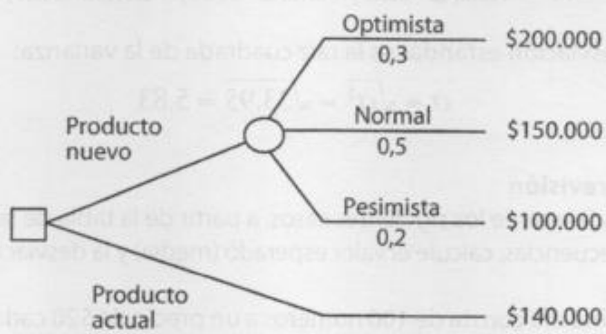
1. Una rifa consta de 100 números a un precio de \$20 cada uno. El premio es de \$1000 (premio único) y el jugador compra un número.
2. Una rifa consta de 100 números a un precio de \$20 cada uno. El premio es de \$1000 (premio único) y el jugador compra dos números.
3. En una caja hay 10 bolas, 2 son azules, 3 son verdes y 5 son rojas. Se saca una bola y si la bola es azul se ganan cero puntos, si es verde se gana un punto y si es roja se ganan dos puntos.
4. En una caja hay 10 bolas, 2 son azules, 3 son verdes y 5 son rojas. Se sacan dos bolas y se suman los puntos sabiendo que si la bola es azul se ganan cero puntos, si es verde se gana un punto y si es roja se ganan dos puntos.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Aplicación

Árboles de decisión

Un árbol de decisión es una herramienta útil para la toma de decisiones. Se basan en probabilidades y en el concepto de valor esperado. Por ejemplo, suponga que una empresa puede lanzar un producto nuevo, y que según sus estudios de mercado y financieros, puede obtener ganancias según tres escenarios diferentes. En un escenario optimista ganarían $200.000, en un escenario normal ganarían $150.000 y en un escenario pesimista las ganancias llegarían a $100.000. Los escenarios anteriores tienen probabilidades de ocurrencia de 0.3, 0.5 y 0.2, respectivamente. La otra alternativa es seguir con su producto actual, el cual genera ganancias de $140.000. La empresa no tiene la posibilidad de producir los dos productos simultáneamente. Entonces, puede construirse un árbol de decisión como el siguiente:



La alternativa del producto nuevo tiene un valor esperado de:

$$200.000 * 0,3 + 150.000 * 0,5 + 100.000 * 0,2 = \$155.000$$

El producto nuevo tiene un rendimiento esperado mayor que el del producto actual, sin embargo el producto nuevo tiene un nivel de riesgo mayor.

Los árboles de decisión pueden ser mucho más complejos, y son útiles en situaciones en las que se tienen secuencias de decisiones, como sería indicar una nueva decisión en caso de que se dé el escenario pesimista.

Medición del riesgo

Suponga que usted invierte en un tipo de activo que paga un rendimiento fijo. Sin duda, esa sería una inversión segura, sin riesgo. Pero si invierte en un proyecto en el cual se podría ganar mucho, pero también con la posibilidad de perder mucho, entonces esa sería una inversión muy riesgosa. Así, el riesgo puede asociarse con el concepto de variabilidad. En la medida en la que la variabilidad sea poca, el riesgo es bajo; pero mientras más elevada sea la variabilidad, más riesgosa será la decisión.

El riesgo puede medirse a través de la varianza o la desviación estándar. Por ejemplo, suponga que se tienen dos inversiones, una con probabilidad de 60% de ganar \$100.000 y 40% probabilidad de perder \$50.000. La otra tiene 50% de probabilidad de ganar \$90.000 y otro 50% de probabilidad de perder \$10.000. Podemos calcular el valor esperado para determinar cuál tendría un mayor rendimiento esperado. Para la primera inversión:

$$E(X) = 100.000 * 0{,}60 + -50.000 * 0{,}40 = 40.000$$

Para la segunda inversión:

$$E(X) = 90.000 * 0{,}50 + -10.000 * 0{,}50 = 40.000$$

Como se observa, ambas tienen el mismo rendimiento esperado, pero vamos a calcular la varianza para ver cuál es más riesgosa. Para la primera inversión:

$$\sigma^2 = (100.000 - 40.000)^2 * 0{,}60 + (-50.000 - 40.000)^2 * 0{,}40 = 5.400.000.000$$

Para la segunda inversión:

$$\sigma^2 = (90.000 - 40.000)^2 * 0{,}50 + (-10.000 - 40.000)^2 * 0{,}50 = 2.500.000.000$$

Vemos que la primera inversión tiene una varianza mayor, o sea, es una inversión más riesgosa que la segunda.

Distribuciones discretas de probabilidad

Tal como se ha mencionado en capítulos anteriores, las variables discretas toman valores que se obtienen por conteo y cada uno de ellos asumirá esos valores con una cierta probabilidad. Generalmente se emplea una fórmula para determinar las probabilidades que cada uno de los posibles valores de estas variables aleatorias discretas. Así, si *x* representa los posible valores numéricos de la variable, entonces se podría disponer de una función *f*(*x*) que permita conocer los valores de cada valor de *x*. Esa función *f*(*x*) se definiría como la distribución de probabilidad de la variable aleatoria discreta en cuestión.

1.3 Distribución binomial

Suponga que un vendedor de un producto sabe que cada cliente que visita puede comprar su producto, o bien, no comprarlo, por lo que solamente hay dos posibles resultados. Por su experiencia sabe que el porcentaje de casos en los que logra la venta permanece constante a lo largo del tiempo y que generalmente cada cliente no tiene contacto con los demás. El vendedor desea saber la probabilidad de lograr 3 ventas si visita 8 clientes. Una situación como esta corresponde a un problema de una distribución binomial de probabilidad.

Un experimento binomial (o sea, un ejercicio en que se emplea la distribución binomial) se da cuando se realiza un experimento aleatorio cuyo resultado es una variable aleatoria discreta y que cumple con las siguientes suposiciones:

1. Existen solamente dos resultados posibles en cada ensayo, llamados, arbitrariamente, *éxitos* y *fracasos*. Por ejemplo, vender y no vender, en el caso mencionado anteriormente.
2. Existe un número fijo n de intentos o ensayos. Por ejemplo, el vendedor visita 8 clientes, es decir, va a realizar 8 intentos de vender su producto.
3. La probabilidad de un éxito, representada por p, permanece constante en todos los intentos. Por ejemplo, suponga que el vendedor logra la venta en el 30% de los casos.
4. Todos los n intentos repetidos son independientes. En el ejemplo se dijo que cada cliente no tiene contacto con los demás, por lo que cada evento es independiente de los otros.

Un experimento de este tipo puede resultar en un éxito, con una probabilidad de p, o en un fracaso, con una probabilidad de $q = 1 - p$. Entonces, la distribución de probabilidad de la variable aleatoria X es:

$$P(X) = C(n,x)\, p^x q^{n-x} \quad \text{para } x = 0, 1, 2, \ldots, n$$

Donde X es el número establecido de éxitos, n el número de ensayos u observaciones, p la probabilidad de éxito y q la probabilidad de fracaso ($q = 1 - p$).

La expresión $C(n, x)$ es conocida como coeficiente binomial y equivale a:

$$C(n,x) = \frac{n!}{x!(n-x)!}$$

Entonces, la fórmula de la distribución binomial puede ser escrita como:

$$P(X) = C(n,x)p^x q^{n-x} = \frac{n!}{x!(n-x)!}\, p^x q^{n-x}$$

Para la distribución binomial, su media y su desviación estándar corresponden a:

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

Ejemplo

Un vendedor de un producto sabe, por su experiencia, que logra la venta en el 30% de los clientes que visita, porcentaje que ha permanecido constante a lo largo del tiempo. Cada cliente no tiene contacto con los demás. El vendedor desea saber la probabilidad de que si visita 8 clientes,

a) logre vender en exactamente 3 casos.
b) logre vender en por lo menos 3 casos.
c) logre vender en menos de 6 casos.
d) no logre vender en a lo más 5 casos.
e) no logre en más de 7 casos.

Solución

a) Se tiene que se realizan 8 intentos de vender el producto, por lo que se tiene que $n = 8$. Además, se desea saber la probabilidad de lograr 3 ventas, o sea que $x = 3$.

En este caso se define éxito como lograr la venta, por tanto $p = 0{,}30$.

La probabilidad de fracaso es $q = 1 - p = 1 - 0{,}30 = 0{,}70$.

Así, sustituyendo en la fórmula de probabilidad:

$$P(X = 3) = \frac{8!}{3!(8-3)!}(0{,}30)^3(0{,}70)^{8-3} = 0.2541$$

b) En este caso se requiere que $x \geq 3$, lo que significa que nos interesa que 3 o más clientes compren el producto, por lo que buscamos:

$$P(x \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7) + P(X = 8)$$

Esto implicaría emplear la fórmula anterior 6 veces y luego sumar los resultados. Una opción que lleva un poco menos de trabajo es calcular lo que no nos interesa, o sea que 0 clientes, o 1 cliente o 2 clientes compren el producto, y luego restar esos valores de uno, que es la probabilidad total. O sea, se puede recurrir a la regla de la complementación para encontrar la probabilidad de $x \geq 3$:

$$P(x \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2)$$

Aplicando la fórmula o la tabla de probabilidades binomiales, se tiene:

$$P(x \geq 3) = 1 - 0{,}0576 - 0{,}1977 - 0{,}2965 = 0{,}4482$$

c) En este caso se requiere que $x < 6$, es decir, nos interesa la probabilidad de que de 0 a 5 clientes compren el producto:

$$P(x < 6) = P(x \leq 5)$$

Obsérvese que no se incluye al 6 mismo, pues se indica menos de 6, así se calculan las probabilidades para los valores entre 0 y 5:

$$P(x \leq 5) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)$$
$$= 0{,}0576 + 0{,}1977 + 0{,}2965 + 0{,}2541 + 0{,}1361 + 0{,}0468 = 0.9887$$

d) Se desea determinar la probabilidad de que a lo más 5 clientes **no** realicen la compra. Aquí se considera éxito no lograr la venta, así que $p = 0{,}70$ y $q = 0{,}30$. Entonces, se debe calcular:

$$P(x \leq 5) = P(x = 0) + P(x = 1) + P(x = 2) + P(x = 3) + P(x = 4) + P(x = 5)$$

$$= 0{,}0001 + 0{,}0012 + 0{,}0100 + 0{,}0467 + 0{,}1361 + 0{,}2541$$
$$= 0{,}4482$$

e) Se desea determinar la probabilidad de que más de 7 clientes **no** compren el producto ($p = 0{,}70$). Es decir, solo interesa que $x = 8$:

$$P(x = 8) = 0{,}0576$$

Ejemplo

Se sabe que la probabilidad de que un cierto tipo de calentador falle ante un sobrecalentamiento es de 15%, calcule la probabilidad de que entre 6 de tales calentadores:

a) fallen entre 2 y 4

b) no fallen como máximo 3

a) Se tiene que $n = 6$ y que éxito es fallar, así que $p = 0{,}15$ y $q = 0{,}85$:

$$P(2 \leq X \leq 5) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$$

Se calcula cada una por separado:

$$P(X = 2) = C(6, 2)\,(0{,}15)^2\,(0{,}85)^{6-2} = 0{,}1762$$
$$P(X = 3) = C(6, 3)\,(0{,}15)^3\,(0{,}85)^{6-3} = 0{,}0415$$
$$P(X = 4) = C(6, 4)\,(0{,}15)^4\,(0{,}85)^{6-4} = 0{,}0055$$
$$P(X = 5) = C(6, 5)\,(0{,}15)^5\,(0{,}85)^{6-5} = 0{,}0004$$

Entonces se suman los resultados anteriores:

$$= 0{,}1762 + 0{,}0415 + 0{,}0055 + 0{,}0004$$

$$= 0{,}2235$$

b) Si éxito es no fallar, entonces $p = 0{,}85$ y $q = 0{,}15$:

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

Se calcula cada una por separado:

$$P(X = 0) = C(6, 0)\ (0{,}85)^0\ (0{,}85)^{6-0} = 0{,}0000$$
$$P(X = 1) = C(6, 1)\ (0{,}85)^1\ (0{,}85)^{6-1} = 0{,}0004$$
$$P(X = 2) = C(6, 2)\ (0{,}85)^2\ (0{,}85)^{6-2} = 0{,}0055$$
$$P(X = 3) = C(6, 3)\ (0{,}85)^3\ (0{,}85)^{6-3} = 0{,}0415$$

Entonces se suman los resultados anteriores:

$$= 0{,}0000 + 0{,}0004 + 0{,}0055 + 0{,}0415$$

$$= 0{,}0473$$

Ejercicio de revisión

Según un estudio aproximadamente tres de cada diez computadoras portátiles falla en un plazo de 3 años o menos. De una muestra de 10 computadoras portátiles calcule la probabilidad de que, en tres años o menos:

- **a.** Fallen exactamente 4 computadoras.
- **b.** Fallen menos de 3 computadoras.
- **c.** Fallen como mínimo 8 computadoras.
- **d.** No fallen a lo sumo 7 computadoras.
- **e.** No fallen entre 3 y 5 computadoras.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video que exponen los conceptos y teoremas de probabilidad.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Uso de Excel y Minitab para la distribución binomial

Ejemplo

Según un estudio, de las muertes de motociclistas en el 2005, el 42% no tenían el casco puesto en el accidente. Calcule, usando Excel y Minitab, la probabilidad de que de una muestra de 12 accidentes ocurridos ese año y seleccionados aleatoriamente:

a. En exactamente 5 de ellos el motociclista no tenía puesto el casco en el accidente.

b. En menos de 5 de ellos el motociclista no tenía puesto el casco en el accidente.

Solución

Se tiene que $n = 12$, el éxito es que no llevara el casco, entonces $p = 0{,}42$ y $q = 0{,}58$.

a. Lo que se desea calcular es:

$$P(X = 5) =$$

Entonces, en **Excel** se emplea la función DISTR.BINOM, cuya sintaxis es:

=DISTR.BINOM(núm_éxito;ensayos;prob_éxito;acumulado)

Así en este caso, se completa la función en la celda en la que se desea el resultado como:

=DISTR.BINOM(5;12;0,42;0)

Se indicó acumulado como 0, para calcular el valor exacto y no el acumulado. El resultado es 0,2285.

Se tiene que $n = 12$, el éxito es que no llevara el casco, entonces $p = 0{,}42$ y $q = 0{,}58$.

a. Lo que se desea calcular es:

$$P(X = 5) =$$

Entonces, en **Excel** se emplea la función DISTR.BINOM, cuya sintaxis es:

=DISTR.BINOM(núm_éxito;ensayos;prob_éxito;acumulado)

Así en este caso, se completa la función en la celda en la que se desea el resultado como:

=DISTR.BINOM(5;12;0,42;0)

Se indicó acumulado como 0, para calcular el valor exacto y no el acumulado. El resultado es 0,2285.

b. Lo que se desea calcular es:

$$P(X < 5) = P(X \leq 4)$$

Entonces, en **Excel** se emplea la función DISTR.BINOM, cuya sintaxis es:

=DISTR.BINOM(núm_éxito;ensayos;prob_éxito;acumulado)

Así en este caso, se completa la función en la celda en la que se desea el resultado como:

=DISTR.BINOM(4;12;0,42;1)

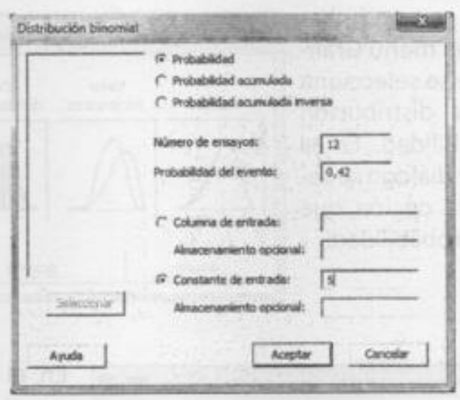
Se indicó acumulado como 1, para calcular el valor acumulado. El resultado es 0,3825.

En **Minitab**, se tiene los mismos datos, o sea, que $n = 12$, el éxito es que no llevara el casco, entonces $p = 0,42$ y $q = 0,58$.

a. Lo que se desea calcular es:

$$P(X = 5) =$$

Entonces, se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Binomial. Se completa el cuadro de diálogo:

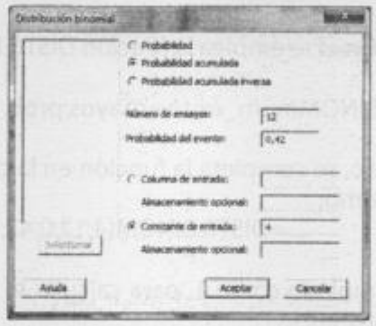


Se selecciona probabilidad para que calcule el valor exacto del número de eventos. El número de ensayos es n y la probabilidad del evento es p. El número establecido de éxitos se puede dar como una columna, y en ese caso de debe elegir columna de entrada, o se puede digitar en el cuadro, en cuyo caso es constante de entrada, que es lo que se muestra en la ilustración anterior. Luego se da clic en Aceptar y se obtiene el resultado 0,2285 en la ventana Sesión.

b. Lo que se desea calcular es:

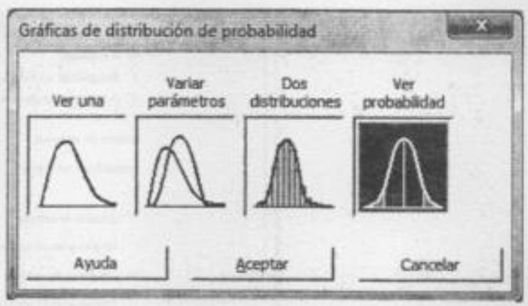
$$P(X < 5) = P(X \leq 4)$$

Entonces, se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Binomial. Se completa el cuadro de diálogo:

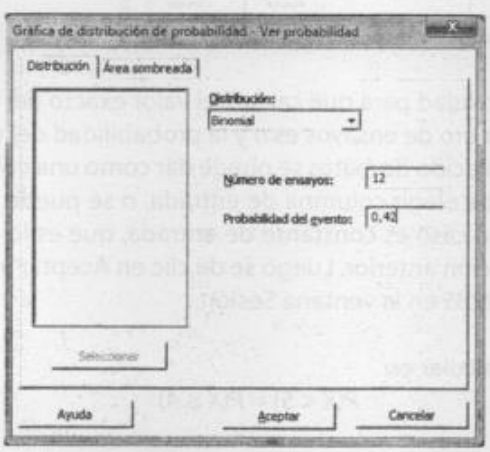


Se selecciona probabilidad acumulada para que calcule el valor acumulado desde $x = 0$ hasta el número establecido de éxitos. El número de ensayos es n y la probabilidad del evento es p. El número establecido de éxitos se puede dar como una columna, y en ese caso de debe elegir columna de entrada, o se puede digitar en el cuadro, en cuyo caso es constante de entrada, que es lo que se muestra en la ilustración anterior. Luego se da clic en Aceptar y se obtiene el resultado 0,3825 en la ventana Sesión.

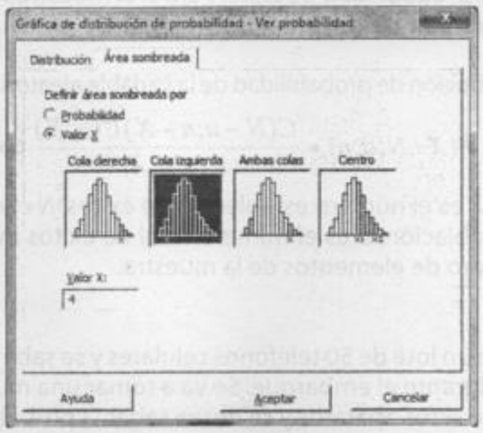
También, se puede hacer uso del menú Gráfica, donde se selecciona Gráfica de distribución de probabilidad. En el cuadro de diálogo se selecciona la opción que dice Ver probabilidad.



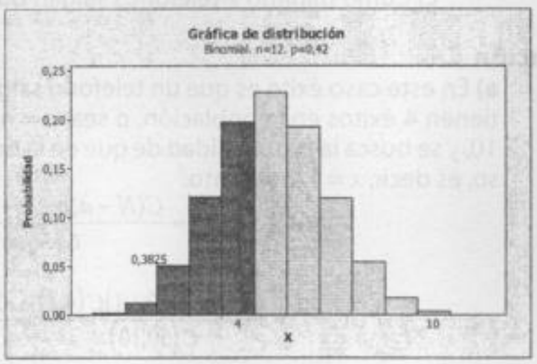
En el cuadro de diálogo se selecciona en la lista la distribución binomial y se introduce el dato del número de ensayos y la probabilidad de éxito:



Posteriormente se da clic en la pestaña Área sombreada. Aquí se elige definir el área sombreada por valor *X* y como en este caso se desea saber la probabilidad de que $x = 4$, entonces se selecciona Cola izquierda y se escribe el valor de *x* en el espacio que aparece:



Al dar clic en Aceptar, Minitab crea un gráfico que indica el valor de la probabilidad:



1.4 Distribución hipergeométrica

Suponga que se tiene un lote de 50 teléfonos celulares y se sabe que 4 de ellos se dañaron durante el embarque. Se va a tomar una muestra sin reemplazo de 10 de estos aparatos y se desea saber la probabilidad de que uno de ellos esté defectuoso. Para esto se realizará un muestreo sin reemplazo, por lo que los eventos en este caso no son independientes, y por tanto, no se puede emplear la distribución binomial. En un caso como este se emplea la distribución hipergeométrica.

La distribución hipergeométrica se utiliza en situaciones similares a los casos en que se emplea la distribución binomial, pero la diferencia principal es que en la binomial los eventos deben ser independientes, mientras que en la hipergeométrica no son independientes. Entonces, si, como en el ejemplo, se realiza un muestreo de un lote, para aplicar la binomial debe hacerse un muestreo *con reemplazo* para obtener la independencia, y de ese modo se mantendrá constante la probabilidad de éxito. Pero si el muestreo se hace *sin reemplazo*, los eventos no son independientes y la probabilidad no es constante.

La distribución de probabilidad de la variable aleatoria hipergeométrica es:

$$P(X/N,a,n)=\frac{C(N-a,n-X)C(a,X)}{C(N,n)} \text{ para } x=0,1,\ldots,n$$

Donde X es el número establecido de éxitos, N el número total de elementos en la población, a es el número total de éxitos incluidos en la población y n el número de elementos de la muestra.

Ejemplo

Se tiene un lote de 50 teléfonos celulares y se sabe que 4 de ellos se dañaron durante el embarque. Se va a tomar una muestra sin reemplazo de 10 de estos aparatos y se desea saber la probabilidad de que:

a) Exactamente un teléfono salga defectuoso.
b) Por lo menos dos teléfonos salgan defectuosos.
c) Como mínimo 7 teléfonos salgan buenas.

Solución

a) En este caso éxito es que un teléfono salga defectuoso, por tanto se tienen 4 éxitos en la población, o sea, $a = 4$ defectuosos, $N = 50$ y $n = 10$, y se busca la probabilidad de que en la muestra haya uno defectuoso, es decir, $x = 1$, por tanto:

$$P(X/N,a,n)=\frac{C(N-a,n-X)C(a,X)}{C(N,n)}$$

$$P(X=1)=\frac{C(50-4,10-1)C(4,1)}{C(50,10)}=\frac{C(46,9)C(4,1)}{C(50,10)}$$

$$=\frac{1101716330\cdot 4}{10272278170}=0{,}4290$$

b) La probabilidad de que por lo menos dos teléfonos salgan defectuosos se puede calcular como:

$$P(X\geq 2)=P(X=2)+P(X=3)+\ldots+P(X=10)$$

Lo anterior lleva aplicar la fórmula de la distribución hipergeométrica 9 veces, por lo que es más rápido calcular del modo siguiente, usando el principio de complementariedad:

$$P(X\geq 2)=1-P(X=0)-P(X=1)$$

$$=1-\frac{C(50-4,10-0)C(4,0)}{C(50,10)}-\frac{C(50-4,10-1)C(4,1)}{C(50,10)}$$

$$=1-0{,}3968-0{,}4290=0{,}1742$$

c) Se define éxito como que un teléfono salga bueno, así que $a = 46$ buenos, por tanto:

$$P(X \geq 7buenos) = P(x = 7) + P(x = 8) + P(x = 9) + P(x = 10)$$

$$P(x = 7) = \frac{C(50-46,10-7)C(46,7)}{C(50,10)} = \frac{C(4,3)C(46,7)}{C(50,10)} = 0{,}0208$$

$$P(x = 8) = \frac{C(50-46,10-8)C(46,8)}{C(50,10)} = \frac{C(4,2)C(46,2)}{C(50,10)} = 0{,}1524$$

$$P(x = 9) = \frac{C(50-46,10-9)C(46,9)}{C(50,10)} = \frac{C(4,1)C(46,9)}{C(50,10)} = 0{,}4290$$

$$P(x = 10) = \frac{C(50-46,10-10)C(46,10)}{C(50,10)} = \frac{C(4,0)C(46,10)}{C(50,10)} = 0{,}3968$$

$$P(x \geq 7) = 0{,}0208 + 0{,}1524 + 0{,}4290 + 0{,}3968 = 0{,}9991$$

Ejemplo

Para evaluar la calidad de los materiales de construcción comprados, el departamento de compras realiza muestreos con cierta frecuencia. Hay un material que se recibe en lotes de 30 unidades. Frecuentemente cada lote tiene 2 unidades con defectos. Aleatoriamente se seleccionan muestras sin reemplazo de 4 unidades y se rechaza el lote completo si se encuentra una o más unidades defectuosas. Determine la probabilidad de aceptación del lote.

Solución

Dado que se realiza un muestreo sin reemplazo, entonces corresponde a un experimento hipergeométrico.

En la población hay 2 defectuosos, o sea, se tiene que $a = 3$, el tamaño de la población es 30, $N = 30$ y se toma una muestra de tamaño 4, $n = 4$.

Para que el lote sea aceptado, en la muestra debe haber cero defectuosos, o sea, $x = 0$, por lo tanto la probabilidad de aceptación del lote corresponde a $P(x = 0)$:

$$P(X / N, a, n) = \frac{C(N-a, n-X)C(a, X)}{C(N, n)}$$

$$P(X = 0) = \frac{C(30-2,4-0)C(2,0)}{C(30,4)} = \frac{20475 \cdot 1}{27405} = 0{,}7471$$

Ejercicio de revisión

Si de un lote de 200 comprimidos de un medicamento se sabe que hay 10 que no satisfacen las especificaciones. Si se toma una muestra de 9 de esos comprimidos, determine la probabilidad de que:

a. Exactamente 2 de ellos no satisfagan las especificaciones.
b. A lo sumo 2 no satisfagan las especificaciones.
c. Al menos 8 satisfagan las especificaciones.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Ejemplo

Se sabe que en un lote de 70 comprimidos para la fiebre hay 8 que no satisfacen las especificaciones solicitadas. Calcule la probabilidad de que en una muestra de 5 de esos comprimidos haya exactamente 2 comprimidos que no satisfagan la especificación:

a) usando la fórmula de la distribución hipergeométrica ,
b) usando la binomial como aproximación y compare los valores.

Solución

a) Se considera éxito si un comprimido no satisface la especificación, por lo que $a = 8$, $N = 70$ y $n = 5$:

$$P(X = 2) = \frac{C(70-8,5-2)C(8,2)}{C(70,5)} = 0{,}0875$$

b) Se puede resolver usando la binomial como aproximación porque $N/10 = 70/10 = 7 > n$. Con una población de tamaño 70, n puede llegar a valer hasta 7 y se puede seguir usando la aproximación por la binomial.

Para usar la binomial se necesita tener la probabilidad poblacional p:
$p = a/N = 8/70 = 0{,}11$

Aplicando la fórmula de la binomial con $n = 5$ y $p = 0{,}11$ se obtiene:

$$P(X = 2) = C(5,2)(0{,}11)^2(0{,}89)^{5-2} = 0{,}0853$$

La diferencia entre el valor real y el valor aproximado es apenas de:

$0{,}0875 - 0{,}0853 = 0{,}0022$

Uso de Excel y Minitab para la distribución hipergeométrica

Ejemplo

En un lote de 200 frascos de un medicamento se sabe que 8 frascos no satisfacen las especificaciones de calidad establecidas para dicho fármaco. Calcule, usando Excel y Minitab, la probabilidad de que de una muestra aleatoria de 12 frascos exactamente 3 de ellos no satisfagan las especificaciones.

Solución

Se tiene que una población $N = 200$ frascos, $a = 8$ éxitos (el éxito sería que no satisfaga la especificación), una muestra $n = 12$ frascos, y se pregunta la probabilidad de que 3 no satisfagan la especificación, o sea, que lo que se desea calcular es:

$$P(X = 3) =$$

Entonces, en Excel se emplea la función DISTR.HIPERGEOM, cuya sintaxis es:

=DISTR.BINOM(muestra_éxito;núm_de_muestra;población_éxito;núm_de_población)

Los argumentos de la función anterior son:

muestra_éxito: número establecido de éxitos (x)
núm_de_muestra: tamaño de muestra (n)
población_éxito: número de éxitos en la población (a)
núm_de_población: tamaño de la población (N)

Así en este caso, se completa la función en la celda en la que se desea el resultado como:

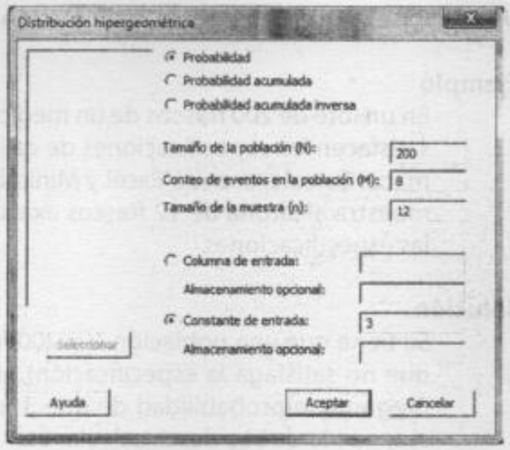
=DISTR.HIPERGEOM(3;12;8;200)

El resultado es 0,0074.

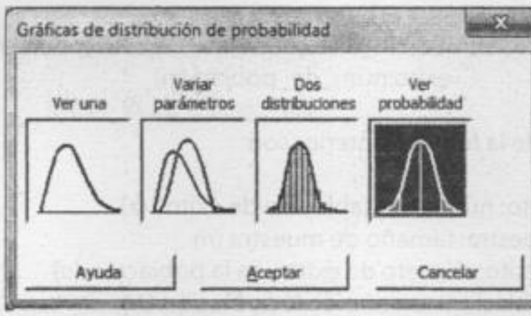
En Minitab, se tiene los mismos datos, una población $N = 200$ frascos, $a = 8$ éxitos (el éxito sería que no satisfaga la especificación), una muestra $n = 12$ frascos, y se pregunta la probabilidad de que 3 no satisfagan la especificación, o sea, que lo que se desea calcular es:

$$P(X = 3) =$$

Entonces, se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Hipergeométrica. Se completa el cuadro de diálogo:

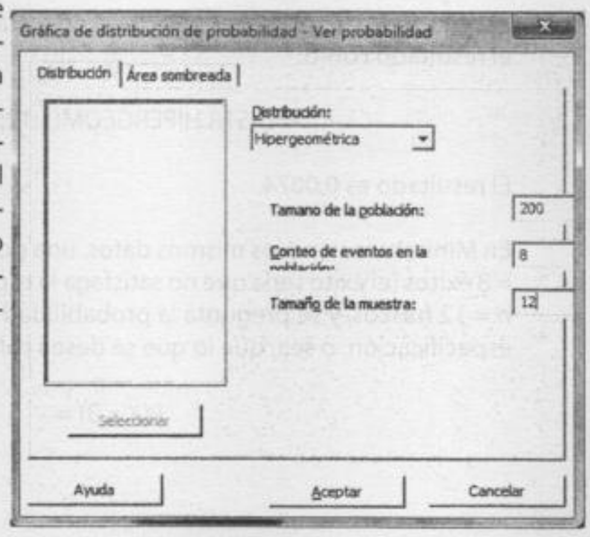


Se selecciona probabilidad para que calcule el valor exacto del número de eventos y se completan los datos tal como se muestra en la imagen. Luego se da clic en Aceptar y se obtiene el resultado 0,0074 en la ventana Sesión.

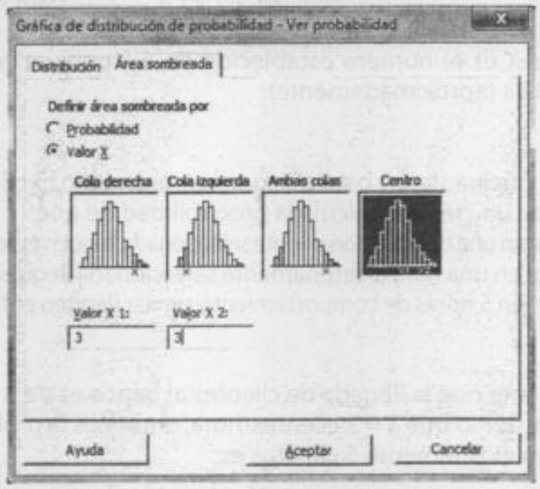


También, se puede hacer uso del menú Gráfica, donde se selecciona Gráfica de distribución de probabilidad. En el cuadro de diálogo se selecciona la opción que dice Ver probabilidad.

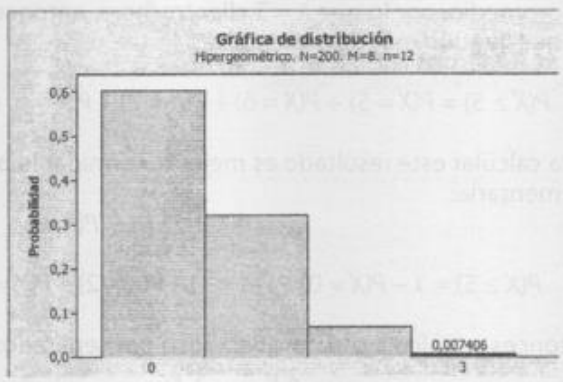
En el cuadro de diálogo se selecciona en la lista la distribución hipergeométrica y se introduce el dato del tamaño de población, del número de éxitos en la población y del tamaño de la muestra:



Posteriormente se da clic en la pestaña Área sombreada. Aquí se elige definir el área sombreada por valor X y como en este caso se desea saber la probabilidad de que $x = 3$, entonces se selecciona Centro y se escribe el valor de x en los dos espacios que aparecen:



Al dar clic en Aceptar, Minitab crea un gráfico que indica el valor de la probabilidad:



1.5 Distribución de Poisson

A una oficina de un banco llegan, en promedio, 3 clientes por hora a solicitar un crédito. Eso quiere decir, que en una hora cualquiera, regularmente llegará una cantidad de clientes cercana a 3. Bajo la capacidad actual, se ha determinado que se puede atender óptimamente hasta un máximo de 5 clientes por hora, por lo que se desea determinar la probabilidad de que en una hora cualquiera lleguen más de 5 clientes. Un problema de este tipo, en el que se desea calcular la probabilidad de que ocurra determinada cantidad de eventos en un intervalo continuo de tiempo, área o volumen, se puede resolver empleando la distribución de Poisson.

Así, la distribución de probabilidad de la variable aleatoria discreta de Poisson permite determinar la probabilidad del número de eventos que ocurren en un intervalo de tiempo dado o en un área o volumen específico. Esta probabilidad se obtiene por medio de:

$$P(X/\lambda) = \frac{\lambda^x e^{-\lambda}}{X!} \text{ para } x = 0, 1, 2, \ldots$$

Donde X es el número establecido de éxitos, λ es la media y e es la constante 2.7183 (aproximadamente).

Ejemplo

A una oficina de un banco llegan, en promedio, 3 clientes por hora a solicitar un crédito. Calcule la probabilidad de que:

a) en una hora aleatoriamente seleccionada lleguen exactamente 5 clientes.

b) en una hora aleatoriamente seleccionada lleguen 5 o más clientes.

c) en 5 horas de comportamiento similar lleguen entre 14 y 17 clientes.

Solución

a) Se tiene que la llegada de clientes al banco es de 3 por hora en promedio, por lo que $\lambda = 3$ clientes/hora, entonces la probabilidad de que lleguen exactamente 5 clientes es:

$$P(X=5) = \frac{\lambda^x e^{-\lambda}}{X!} = \frac{3^5 e^{-3}}{5!} = 0{,}1008$$

b) Se sabe que la tasa de llegada de clientes al banco es de 3 por hora en promedio, por lo que $\lambda = 3$ clientes/hora, entonces la probabilidad de que lleguen más de 5 clientes es:

$$P(X \geq 5) = P(X=5) + P(X=6) + P(X=7) + P(X=8) + P(X=9) + \ldots$$

Para calcular este resultado es mejor determinar la probabilidad complementaria:

$$P(X \geq 5) = 1 - P(X < 5)$$

$$P(X \geq 5) = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) - P(X=4)$$

Entonces se calcula cada probabilidad por separado:

$$P(X=0) = \frac{\lambda^x e^{-\lambda}}{X!} = \frac{3^0 e^{-3}}{0!} = 0{,}0498$$

$$P(X=1) = \frac{\lambda^x e^{-\lambda}}{X!} = \frac{3^1 e^{-3}}{1!} = 01494$$

$$P(X=2) = \frac{\lambda^x e^{-\lambda}}{X!} = \frac{3^2 e^{-3}}{2!} = 0{,}2240$$

$$P(X=3) = \frac{\lambda^x e^{-\lambda}}{X!} = \frac{3^3 e^{-3}}{3!} = 02240$$

$$P(X=4) = \frac{\lambda^x e^{-\lambda}}{X!} = \frac{3^4 e^{-3}}{4!} = 0{,}1680$$

Luego se resta cada resultado de uno:

$$P(X \geq 5) = 1 - 0{,}0498 - 0{,}1494 - 0{,}2240 - 0{,}2240 - 0{,}1680 = 0{,}1847$$

c) Aquí el período de interés es de 5 horas, por lo que $\lambda = 5 \times 3 = 15$ clientes/período de 5 horas, entonces se calcula la probabilidad de que lleguen entre 14 y 17 clientes:

$$P(14 \leq x \leq 17) = P(X = 14) + P(X = 15) + P(X = 16) + P(X = 17)$$

$$= 0{,}1024 + 0{,}1024 + 0{,}0960 + 0{,}0847 = 0{,}3856$$

Ejercicio de revisión

Un promedio de 15 personas por hora ingresa un parque zoológico. Si se selecciona una hora cualquiera, calcule la probabilidad de que:

a. Ingresen entre 12 y 15 personas.
b. Ingresen menos de 8 personas.
c. Ingresen más de 10 personas.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video que exponen los conceptos y teoremas de probabilidad.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Uso de Excel y Minitab para la distribución de Poisson

Ejemplo

A una clínica llega un promedio de 5 pacientes cada hora. Calcule, usando Excel y Minitab, la probabilidad de que en una hora seleccionada en forma aleatoria lleguen exactamente 3 pacientes.

Solución

Se tiene que una media de 5 pacientes por hora y se pregunta la probabilidad de que lleguen 3 por hora, o sea, que lo que se desea calcular es:

$$P(X = 3) =$$

Entonces, en **Excel** se emplea la función POISSON, cuya sintaxis es:

=POISSON(x;media;acumulado)

Los argumentos de la función anterior son:

x: número establecido de éxitos (x)
media: promedio (λ)
acumulado: 0 si no es acumulado o 1 si es acumulado

Así en este caso, se completa la función en la celda en la que se desea el resultado como:

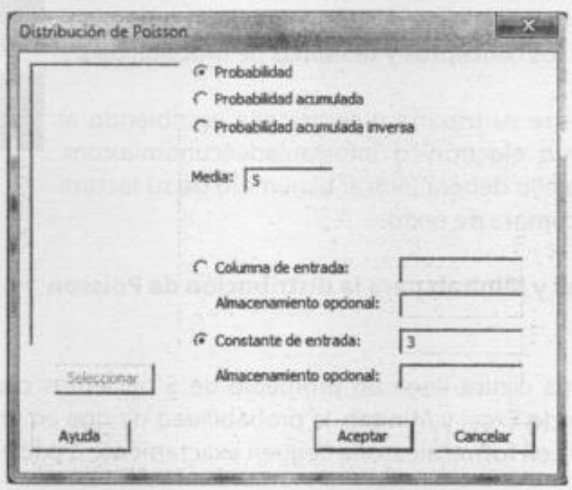
=POISSON(3;5;0)

El resultado es 0,1404.

En **Minitab**, con base en los datos dados, una media de 5 pacientes por hora y se pregunta la probabilidad de que lleguen 3 por hora, o sea, que lo que se desea calcular es:

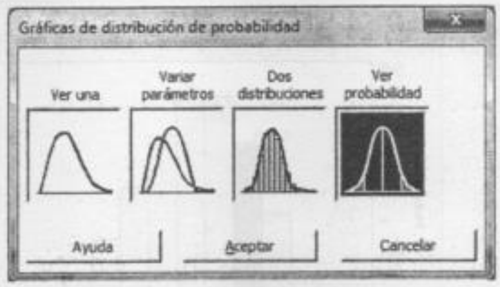
$$P(X = 3) =$$

Entonces, se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Poisson. Se completa el cuadro de diálogo:

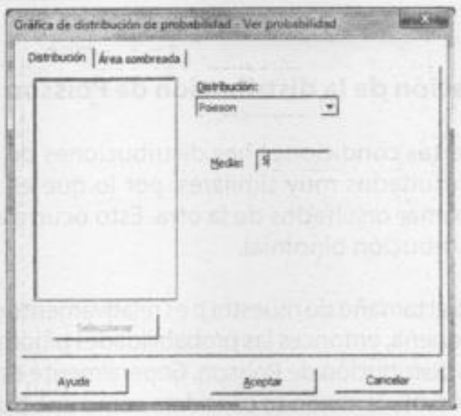


Se selecciona probabilidad para que calcule el valor exacto del número de eventos y se completan los datos tal como se muestra en la imagen. Luego se da clic en Aceptar y se obtiene el resultado 0,1404 en la ventana Sesión.

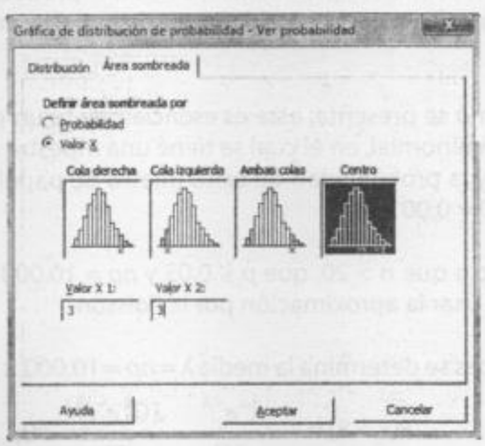
También, se puede hacer uso del menú Gráfica, donde se selecciona Gráfica de distribución de probabilidad. En el cuadro de diálogo se selecciona la opción que dice Ver probabilidad.



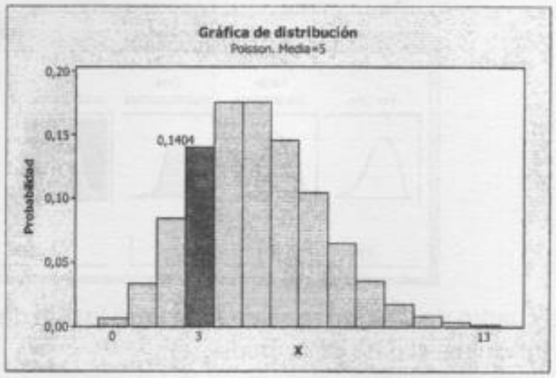
En el cuadro de diálogo se selecciona en la lista la distribución Poisson y se introduce el dato de la media:



Posteriormente se da clic en la pestaña Área sombreada. Aquí se elige definir el área sombreada por valor *X* y como en este caso se desea saber la probabilidad de que $x = 3$, entonces se selecciona Centro y se escribe el valor de *x* en los dos espacios que aparecen:



Al dar clic en Aceptar, Minitab crea un gráfico que indica el valor de la probabilidad:



1.6 Aproximación de la distribución de Poisson a la binomial

Bajo ciertas condiciones hay distribuciones de probabilidad que tienden a ofrecer resultados muy similares, por lo que es posible utilizar una de ellas para aproximar resultados de la otra. Esto ocurre entre la distribución de Poisson y la distribución binomial.

Cuando el tamaño de muestra n es relativamente grande y la probabilidad de éxito p es pequeña, entonces las probabilidades binomiales pueden aproximar por medio de la distribución de Poisson. Generalmente esta aproximación se utiliza si $n > 20$ y $p < 0{,}05$. También se considera como una buena regla si $n \geq 30$, $np < 5$ y $nq < 5$. Si $n > 100$, la aproximación es excelente, siempre que $np < 10$.

Ejemplo

En un proceso de manufactura de papel se encuentra un defecto por cada 1.000 metros producidos. Calcule la probabilidad de que en una muestra aleatoria de 10.000 metros de papel se encuentren 8 defectos.

Solución

Tal como se presenta, este es esencialmente un problema de la distribución binomial, en el cual se tiene una muestra $n = 10.000$ metros de papel y la probabilidad de éxito (metro de papel con defectos) es $p = 1/1000 = 0{,}001$.

Debido a que $n > 20$, que $p \leq 0{,}05$ y $np = 10.000 \times 0{,}001 = 10 \leq 10$ se puede usar la aproximación por la Poisson.

Entonces se determina la media $\lambda = np = 10.000 \times 0{,}001 = 10$, entonces:

$$P(x=8) = \frac{\lambda^x e^{-\lambda}}{X!} = \frac{10^8 e^{-10}}{8!} = 0{,}112599$$

Si este problema se hubiera resuelto empleando la distribución binomial, se tendría n = 10.000, con p = 1/1000 = 0,001, q = 1 – 0,001 = 0,999, entonces:

$$P(x = 8) = C(10.000, 8)\ (0{,}001)^8\ (0{,}999)^{1000-8} = 0{,}112622$$

Se observa claramente que los resultados son sumamente próximos.

Para la distribución de Poisson se cumple que el valor esperado de la distribución es igual a la media l, y que a su vez este valor equivale a la varianza de esta distribución de probabilidad, es decir:

$$E(X) = V(X) = \boldsymbol{\lambda} = \boldsymbol{\sigma}^2$$

Ejemplo

En la tabla se da la distribución de probabilidad del número de delfines (x) que se encuentran por cada cierta área de mar luego de un derrame de petróleo de un barco. Si se sabe que esta variable sigue una distribución de Poisson, muestre que:

$$\lambda = \sigma^2$$

χ	0	1	2	3	4	5	6
$P(\chi)$	0,2465	0,3452	0,2417	0,1128	0,0395	0,0111	0,0032

Solución

Con base en los datos de la tabla se obtiene primero el valor esperado:

$$E(x) = 0 * 0{,}2465 + 1 * 0{,}3452 + 2 * 0{,}2417 + 3 * 0{,}1128 + 4 * 0{,}0395 + 5 * 0{,}0111 + 6 * 0{,}0032 = 1{,}39$$

Luego se calcula la varianza:

$$\sigma^2 = S(x - E(x))^2\ P(x) = (0 - 1{,}3997)^2 * 0{,}2465 + ... + (6 - 1{,}3997)^2 * 0{,}0032$$

$$= 1{,}39$$

Por lo que queda claro que si λ = 1,39, entonces σ^2 = 1,39. Queda comprobado que $\boldsymbol{\lambda} = \boldsymbol{\sigma}^2$.

Aplicación

Seguros
Cuando una persona desea adquirir un seguro es porque desea contar con una compensación ante la eventualidad de que se materialice determinado riesgo, o sea, que se presente un evento desfavorable que le genere alguna necesidad económica. Cuando la persona adquiere el seguro, la compañía de seguros se compromete a pagar la compensación prometida. Para que una empresa esté anuente a participar en este tipo de negocio debe ser capaz de medir el tipo de riesgo que está asegurando y para ello es necesario emplear distribuciones de probabilidad. Así, la teoría del riesgo analiza esos posibles eventos y permite a la aseguradora determinar cuántas reservas debe mantener y cuánto riesgo aceptar en su cartera.

Por ejemplo, cuando una aseguradora conoce cuál podría ser el número máximo de siniestros posibles, puede emplear la distribución binomial para determinar la probabilidad de que tenga que pagar una determinada cantidad de reclamaciones. Cuando no conoce cuál podría ser el máximo de posibles siniestros, entonces podría utilizar la distribución de Poisson, la cual se aproxima a la binomial conforme n tiende a infinito, para determinar el número esperado de reclamaciones o las probabilidades de que se den ciertas cantidades de siniestros.

1.7 Distribución multinomial

Los audífonos fabricados por una empresa son sometidos a un control de calidad en el cual se clasifican como perfectos, con defectos secundarios o con defectos mayores. Generalmente el 85% de los audífonos se clasifican como perfectos, el 10% con defectos secundarios y un 5% con defectos mayores. En una muestra de 8 audífonos se quiere saber la probabilidad de que haya 5 perfectos, 2 con defectos secundarios y uno con defectos mayores. Esta situación es similar en todo a un problema binomial, excepto que hay tres resultados posibles y no dos. En estos casos se emplea la distribución multinomial.

Cuando un experimento binomial tiene más de dos resultados posibles, y no solo dos resultados posibles, se convierte en un experimento multinomial. Se tiene, por ejemplo, la clasificación de calidad de la situación de la fabricación de audífonos, en la cual estos artículos se clasifican como perfectos, con defectos secundarios o con defectos mayores. Cada uno de los eventos es independiente y sus probabilidades se mantienen constantes al realizar los muestreos con reemplazo.

En general, si un ensayo puede resultar en cualquiera de k posibilidades $E_1, E_2, \dots, E_k$, con probabilidades $p_1, p_2, \dots, p_k$, entonces la distribución multinomial dará la probabilidad de que E_1 ocurra x_1 veces, de que E_2 ocurra x_2 veces, ... , y de que E_k ocurra x_k veces, en n intentos independientes.

La distribución multinomial es:

$$P(x_1, x_2, ..., x_k) = \frac{n!}{x_1!x_2!...x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$$

Donde: $n = x_1 + x_2 + ... + x_k$

Ejemplo

Los audífonos fabricados por una empresa son sometidos a un control de calidad en el cual se clasifican como perfectos, con defectos secundarios o con defectos mayores. Generalmente el 85% de los audífonos se clasifican como perfectos, el 10% con defectos secundarios y un 5% con defectos mayores. En una muestra de 8 audífonos se quiere saber la probabilidad de que haya 5 perfectos, 2 con defectos secundarios y uno con defectos mayores.

Solución

Primeramente se plantean los datos del problema:

Perfectos: $p_1 = 0{,}85$
Con defectos secundarios: $p_2 = 0{,}10$
Con defectos mayores: $p_3 = 0{,}05$

Se tiene que $x_1 = 5$, $x_2 = 2$ y que $x_3 = 1$, por lo que $n = 5 + 2 + 1 = 8$. Entonces, se sustituye en la fórmula:

$$P(x_1 = 5, x_2 = 2, x_3 = 1) = \frac{8!}{5! \cdot 2! \cdot 1!}(0{,}85)^5(0{,}10)^2(0{,}05)^1 = 0{,}0372$$

Ejemplo

En una encuesta de intención de voto se obtuvo que el candidato A obtendría el 35% de los votos, el candidato C el 45% y el candidato B el restante 20%.

Si se toma una muestra de 10 personas, ¿cuál es la probabilidad de que la mitad deseen votar por el candidato A, dos quintas partes por el candidato B y el resto por C?

Solución

Primeramente se plantean los datos del problema:

Candidato A: $p_1 = P(A) = 0{,}35$
Candidato B: $p_2 = P(B) = 0{,}20$
Candidato C: $p_3 = P(C) = 0{,}45$

Se tiene que $x_1 = 5$, $x_2 = 4$ y que $x_3 = 1$, por lo que $n = 5 + 4 + 1 = 10$. Entonces, se sustituye en la fórmula:

$$P(x_1 = 5, x_2 = 4, x_3 = 1) = \frac{10!}{5! \cdot 4! \cdot 1!}(0{,}35)^5(0{,}20)^4(0{,}45)^1 = 0{,}0048$$

Ejercicio de revisión

Un equipo de futbol gana el 40% de los partidos que juega, empata el 25% y pierde el resto de los encuentros. Suponiendo que se mantienen estas proporciones, calcule la probabilidad de que en los próximos 6 partidos:

a. Gane 3 veces, empate 2 y pierda 1 juego.

b. Gane o empate 4 partidos y pierda los otros dos.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

1.8 Distribución geométrica

Una empresa de televisión por cable pone a disposición de sus clientes un número telefónico para proveer soporte en caso de que haya problemas con el servicio. Sin embargo la central telefónica pasa ocupada el 90% del tiempo, por lo que los clientes deben hacer más de intento para que su llamada sea contestada. ¿Cuál es la probabilidad de que la llamada de un cliente sea contestada en su tercer intento? Un problema de este tipo satisface todas las condiciones de la distribución binomial, excepto que no hay un número fijo de intentos, por lo que es un problema que se resuelve empleando la distribución geométrica.

Entonces, si en una sucesión de pruebas o ensayos se desea saber la probabilidad del número de la prueba en que ocurre el primer éxito, y si además se cumplen todas las demás condiciones de la distribución binomial, o sea que existen solamente dos resultados posibles en cada ensayo (*éxitos* y *fracasos*), la probabilidad de un éxito, representada por p permanece constante en todos los intentos, y todos los n intentos repetidos son independientes. En resumen, se cumplen las mismas suposiciones fundamentales de la distribución binomial, excepto que n no es fijo.

De ese modo, si en pruebas independientes repetidas puede resultar un éxito con una probabilidad p y en un fracaso con una probabilidad $q = 1 - p$, entonces la distribución de probabilidad de la variable aleatoria X, el número de la prueba en la cual ocurre el primer éxito es la distribución geométrica:

$$P(x, p) = p(1-p)^{x-1} \quad \text{para } x = 1, 2, 3, \ldots$$

La media de esta distribución es $\boldsymbol{\mu} = 1/p$

Ejemplo

Una empresa de televisión por cable pone a disposición de sus clientes un número telefónico para proveer soporte en caso de que haya problemas con el servicio. Sin embargo la central telefónica pasa ocupada el 90% del tiempo, por lo que los clientes deben hacer más de intento para que su llamada sea contestada. ¿Cuál es la probabilidad de que la llamada de un cliente sea contestada en su tercer intento?

Solución

En este problema se busca la probabilidad de que la llamada ingrese, pero si la central telefónica pasa ocupada el 90% del tiempo, esta probabilidad es de solo 10%. Esa es la probabilidad de éxito $p = 0{,}10$.

Sustituyendo en la fórmula de la distribución geométrica:

$$P(x=3) = 0{,}10(1-0{,}10)^{3-1} = 0{,}10(0{,}90)^{2} = 0{,}081$$

Observe, apoyándose en el ejemplo anterior, la diferencia entre la distribución binomial y la geométrica. Si este problema hubiera dicho que va a hacer 5 llamadas y se va a calcular la probabilidad de que entren 3 llamadas, entonces sería un problema de distribución binomial. Pero el problema no plantea un número de intentos determinado, solo desea saberse la probabilidad de que en el tercer intento entre la llamada, por lo que corresponde a la distribución geométrica.

Ejemplo

En un establecimiento de producción de lana se sabe que el 40% de los animales poseen algún tipo de lunar que produce fibras pigmentadas, las cuales reducen el valor del producto. Si se empiezan a examinar los animales, ¿cuál es la probabilidad de que la quinta oveja inspeccionada sea la primera en poseer algún tipo de lunar que produzca fibras pigmentadas?

Solución

Si la primera oveja que posee algún tipo de lunar que produzca fibras pigmentadas es la quinta ($x = 5$), quiere decir que las primeras 7 no poseen este tipo de lunares ($x - 1 = 4$). La probabilidad de obtener una oveja con este tipo de lunares es $p = 0{,}40$, por tanto, aplicando la fórmula:

$$P(x=5) = 0{,}40(1-0{,}40)^{5-1} = 0{,}40(0{,}60)^{4} = 0{,}0518$$

Ejercicio de revisión

Un basquetbolista encesta el 60% de los tiros libres que lanza. Calcule la probabilidad de que:

a. El primer tiro que enceste sea el tercero.

b. El primer tiro que falle sea el cuarto.

c. Si el jugador lanza 6 veces, ¿cuál es la probabilidad de que enceste en exactamente tres ocasiones?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Minitab para la distribución geométrica

Ejemplo

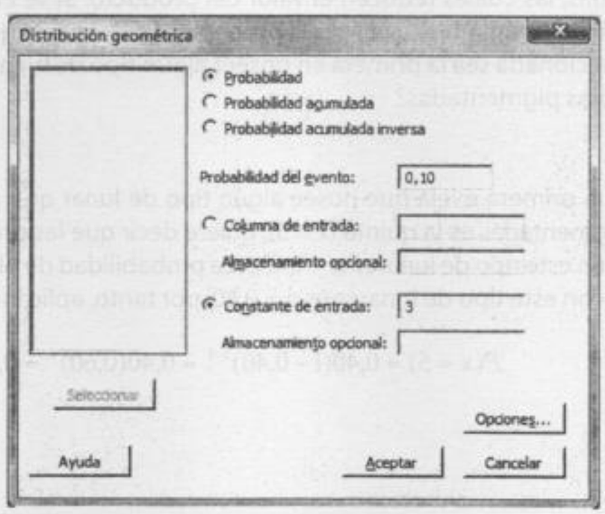
El 10% de las llamadas que ingresan al centro de servicio telefónico de una empresa son para reportar averías. Calcule, usando Minitab, la probabilidad de que la primera llamada que ingresa para reportar averías sea la tercera.

Solución

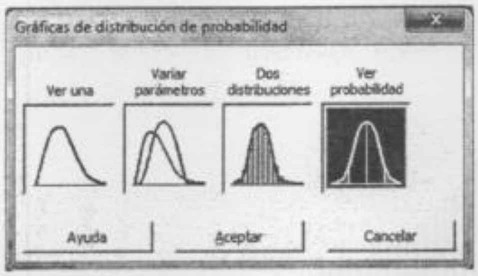
Se tiene que una probabilidad de éxito $p = 0.10$ y se pregunta la probabilidad de que la primera llamada que ingresa para reportar averías sea la tercera, o sea, que lo que se desea calcular es:

$$P(X = 3) =$$

En **Minitab**, se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Geométrica. Se completa el cuadro de diálogo

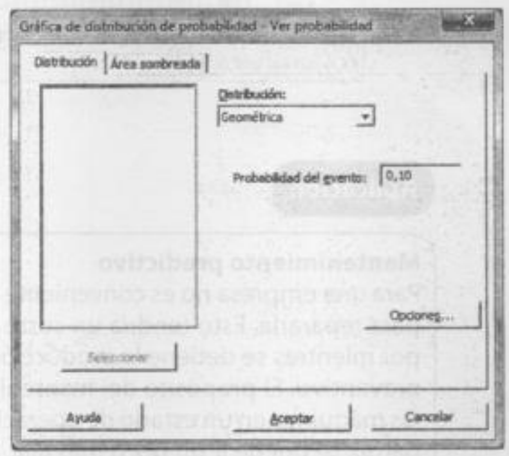


Se selecciona probabilidad para que calcule el valor exacto del número de eventos y se completan los datos tal como se muestra en la imagen. Luego se da clic en Aceptar y se obtiene el resultado 0,081 en la ventana Sesión.

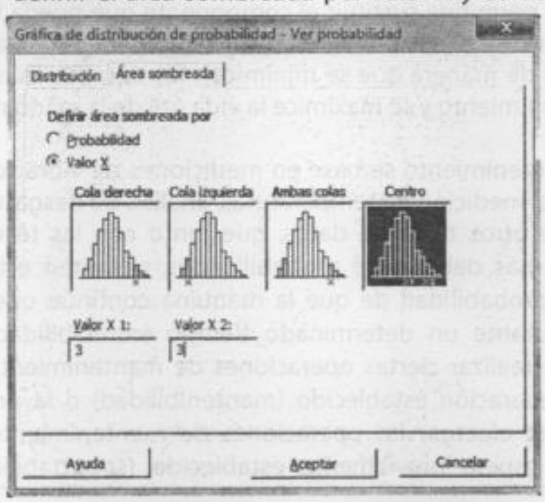


También, se puede hacer uso del menú Gráfica, donde se selecciona Gráfica de distribución de probabilidad. En el cuadro de diálogo se selecciona la opción que dice Ver probabilidad.

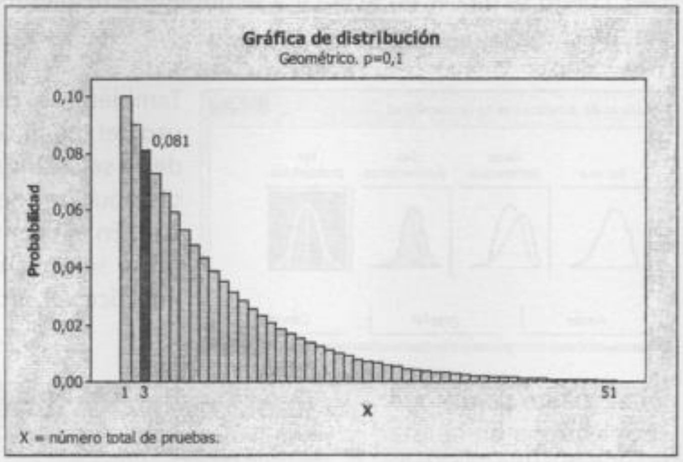
En el cuadro de diálogo se selecciona en la lista la distribución geométrica y se introduce el dato de la probabilidad de éxito:



Posteriormente se da clic en la pestaña Área sombreada. Aquí se elige definir el área sombreada por valor *X* y como en este caso se desea saber la probabilidad de que $x = 3$, entonces se selecciona Centro y se escribe el valor de *x* en los dos espacios que aparecen:



Al dar clic en Aceptar, Minitab crea un gráfico que indica el valor de la probabilidad:



Aplicación

Mantenimiento predictivo

Para una empresa no es conveniente esperar a que una máquina falle para repararla. Esto tendría un costo elevado, especialmente porque por mientras se detiene la producción. Es mejor emplear un enfoque preventivo. El propósito del mantenimiento preventivo es mantener las máquinas en un estado de operación que sea satisfactorio. Sin embargo, se puede ir un poco más allá, y aplicar un enfoque predictivo. Esto es tratar de pronosticar un punto futuro de falla de una máquina, o sea, determinar los puntos óptimos para realizar un mantenimiento preventivo. Esto permitiría sustituir el componente que pueda fallar justo antes de que falle. Así, puede construirse un plan para realizar estos remplazos de manera que se minimicen los tiempos muertos debido al mantenimiento y se maximice la vida útil de la máquina.

Este tipo de mantenimiento se base en mediciones de vibraciones, análisis de aceite, medición de temperaturas, análisis de desgaste de partículas, entre otros tipos de datos, que junto con las técnicas apropiadas, además del uso de probabilidades, permiten estimar datos como la probabilidad de que la máquina continúe operando sin fallas durante un determinado tiempo (confiabilidad), la probabilidad de realizar ciertas operaciones de mantenimiento en el tiempo de reparación establecido (mantenibilidad) o la probabilidad de poder efectuar las operaciones de mantenimiento en un tiempo de espera previamente establecido (soportabilidad).

Material audiovisual

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

1.9 Ejercicios

Ejercicios de desarrollo

Conteste las preguntas que se formulan a continuación (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

Distribuciones de probabilidad

1. A continuación se muestra la tabla de la distribución de probabilidad para el número de accidentes por día que se presentan por día en una fábrica (cuyo máximo es 4). ¿Cuál es la probabilidad de que se presente en un día cualquiera dos o más accidentes?

x	0	1	2	3	4
$P(x)$	0,40	0,30		0,10	0,05

2. La cantidad de pacientes que llegan a una clínica cada hora se comporta de acuerdo con la tabla siguiente, la cual muestra de distribución de probabilidad de esta variable discreta. Nunca llegan más de 6 pacientes. Obtenga e interprete el valor esperado y la desviación estándar.

x	0	1	2	3	4	5	6
$P(x)$	0,05	0,10	0,15		0,20	0,15	0,05

3. La tabla siguiente muestra el número de quejas por día que recibe una compañía de televisión por cable:

Número de quejas x	Probabilidad $P(x)$
0	0
1	1
2	2
3	3
4	4
5	5
6	6
Total	Total

a. ¿Cuál es el número esperado de quejas que recibirá compañía de televisión por cable en un día cualquiera?

b. ¿Cuál es la desviación estándar del número de quejas?

4. Se lanzan dos monedas, si sale escudo gana un punto, si sale corona, gana cero puntos, y X es la suma de los puntos. Construya la tabla de la distribución de probabilidad de X, calcule el valor esperado y la desviación estándar de la distribución.

5. Una compañía de internet ha observado el número de veces que sus operaciones se han visto interrumpidas por la caída de alguno de sus servidores. En 200 días analizados, se determinó que en 120 no hubo caída alguna de los servidores, pero que en 35 días el servicio se interrumpió una vez por esa causa, en 25 días los servidores se cayeron dos veces, en 15 días se presentaron tres caídas de los servidores y que en 5 días el servicio se vio afectado cuatro veces. Nunca se presentaron más de 4 caídas de los servidores en un mismo día.

a. Construya la distribución de probabilidad para la variable número caídas diarias de los servidores de la empresa.

b. Calcule el valor esperado y la desviación estándar de la variable.

6. En una encuesta aplicada a nivel nacional a exportadores se les preguntó sobre el efecto que ha tenido la crisis económica sobre su actividad. El 40% indicó que la crisis le había afectado mucho en su nivel de exportaciones. Si se toma una muestra de 500 empresas exportadoras, ¿cuántas de esas empresas se esperaría que hayan sido muy afectadas por la crisis? Calcule e interprete la desviación estándar para esta distribución de probabilidad.

7. A continuación se muestra la función de distribución de probabilidad para el número de accidentes por día que se presentan por día en una fábrica. ¿Cuál es la desviación estándar de la distribución de probabilidad?

8. En la tabla se muestra la probabilidad de que una red de cómputo se halle fuera de operación durante un cierto número de horas por semana en su fase de operación normal.

Número de horas	0	1	2	3	4	5
Probabilidad	0,45	0,20	0,15	0,10	0,05	0,05

a. Calcule el número esperado de horas por semana en que la red esté fuera de operación y la desviación estándar de esta variable.

b. Si usted tiene que escoger entre este tipo de red y otro tipo de red cuyo número esperado de horas por semana fuera de operación es 1,05 y con desviación estándar de 2,12, ¿cuál tipo de red seleccionaría y por qué?

9. Calcule la media y la desviación estándar de la demanda semanal de cierto artículo en un almacén. Los datos de demanda y su probabilidad de ocurrencia se dan en la tabla.

Unidades vendidas	0	10	20	30	40
Probabilidad	0,05	0,25	0,50	0,15	0,05

10. Un ingeniero conoce los siguientes datos relacionados con el número de fallas semanales que un sistema informático ha experimentado en el transcurso de tres años:

# fallas/semana	0	1	2	3	4	5
Frecuencia	35	61	32	21	12	2

Este ingeniero afirma que es mejor otro sistema, ya que el valor esperado del número de fallas semanales es 1,9 con una desviación estándar de 3,5 veces por semana. ¿Cambiaría usted el sistema actual por el sugerido por este ingeniero o mantendría el sistema actual? Explique basándose en los resultados del valor esperado y la desviación estándar, para esto primero calcule ambos datos primero.

Distribución binomial

11. En un esfuerzo para fomentar la creatividad y la innovación entre su personal, una empresa financiera ha implementado un sistema para captar propuestas de sus empleados. Un elemento del sistema es un formulario que luego es depositado en un buzón. Se ha observado que de cada 4 formularios entregados, uno posee una propuesta que podría generar un gran impacto en la empresa y sus actividades. Con base en estos datos, calcule la probabilidad de que, de una muestra de:

a. 5 formularios, 3 posean ideas de impacto.

b. 8 formularios, la mitad o menos tengan ideas de impacto.

c. 10 formularios, 6 o más no tengan ideas de impacto.

12. El jefe de un departamento de recursos humanos de una empresa grande, estudia con frecuencia el grado de satisfacción de los trabajadores dentro de la empresa, y ha encontrado que 5 de cada 11 empleados se siente insatisfecho con su salario. Esta proporción se ha mantenido constante durante mucho tiempo. Si se seleccionan aleatoriamente 8 personas,

    a. ¿Cuál es la probabilidad de que exactamente 3 de ellas se sientan insatisfechas con su salario?
    b. ¿Cuál es la probabilidad de que la mitad de ellas se sientan satisfechas con su salario?
    c. ¿Cuál es la probabilidad de que a lo sumo 3 de ellas se sientan insatisfechas con su salario?

    ¿Cuál es la probabilidad de que ninguna de ellas se sienta satisfecha con su salario?

13. Se considera que el 20% de los votantes de un país están a favor de la política económica seguida por el actual gobierno. Si se toma una muestra de 12 votantes, ¿cuál es la probabilidad de que del menos 25% de ellos esté a favor de las políticas del gobierno actual?

14. Considere la siguiente información publicada en el sitio http://www.nutriologo.net/2009/09/02/mexico-lider-en-diabetes/:

México líder en diabetes

Un mal que silenciosamente ha permeado nuestro país es la Diabetes Melitus y sus consecuencias son mortales. Es el resultado de una falta de balance en nuestros hábitos alimenticios y puede prevenirse fácilmente.
La Federación Mexicana de Diabetes nos presenta estadísticas alarmantes, como por ejemplo que en México el 10% de la población tiene esta enfermedad y algunos ni si quiera saben que la tienen.
La diabetes tiene mayor incidencia en países en vías de desarrollo y su causa principal es la obesidad, en nuestro país 7 de cada 10 personas tiene sobrepeso u obesidad. México se encuentra en 2° lugar de obesidad en el mundo.

Si se selecciona una muestra al azar de 10 mexicanos, calcule la probabilidad de que:

a. Tres o menos padezcan diabetes
b. Exactamente 6 tengan sobrepeso u obesidad
c. Al menos ocho no padezcan diabetes
d. A lo sumo dos no sean obesos

15. Según una noticia publicada en días recientes "El 4% de los medicamentos que los proveedores entregan a la Caja Costarricense de Seguro Social (CCSS) se rechazan porque no cumple con los estándares de calidad internacionales que la institución tiene establecidos". Si se toma una muestra aleatoria de 10 de los medicamentos comprados por esta institución, calcule la probabilidad de que:

    a. Al menos 5 de ellos se rechazado por no cumplir los estándares de calidad.
    b. Menos de 3 sean rechazados por no cumplir los estándares de calidad.
    c. Si se realiza dos veces la experiencia de seleccionar un medicamento empleando un muestreo con reemplazo, ¿cuál es la probabilidad de que en ambos casos exactamente 9 de sean aceptados por satisfacer los estándares de calidad.

16. El jefe de un departamento de recursos humanos de una empresa grande, estudia con frecuencia el grado de satisfacción de los trabajadores dentro de la empresa, y ha encontrado que 5 de cada 12 empleados se siente insatisfecho con su salario. Esta proporción se ha mantenido constante durante mucho tiempo. Si se seleccionan aleatoriamente 8 personas ¿Cuál es la probabilidad de que exactamente 5 de ellas se sientan insatisfechas con su salario?

Distribución hipergeométrica

17. Para probar la calidad de un tipo de concreto se prepararon 60 probetas de la mezcla y se obtuvo que 5 de ellas no presentaban una resistencia a la compresión adecuada. Si se toma una muestra aleatoria de 8 de esas 60 probetas, calcule la probabilidad de que:

    a. Al menos 3 no satisfagan el requerimiento de resistencia a la compresión.
    b. A lo sumo 55 satisfagan el requerimiento de resistencia a la compresión.
    c. Entre 2 y 6 no satisfagan el requerimiento de resistencia a la comprensión.

18. En un embarque se recibieron 250 cajas con unos componentes electrónicos llamados termistores. Los termistores pueden ser positivos o negativos. De la cantidad total de cajas se sabe que 120 solo contienen termistores negativos. Con base en estos datos calcule la probabilidad de que al seleccionar 10 cajas de componentes al azar se encuentre:

    a. Exactamente 3 cajas de termistores negativos.
    b. Al menos 4 cajas de termistores negativos.
    c. A lo sumo 6 cajas de termistores positivos.
    d. Entre 4 y 7 cajas de termistores negativos.

19. En una encuesta reciente se obtuvo que 900 de una muestra 2400 personas se había visto afectadas por la gripe en al menos una ocasión en los últimos seis meses. Suponga que se selecciona una muestra aleatoria de 12 personas, determine la probabilidad de que en los últimos 6 meses:

    a. Exactamente 7 personas hayan tenido gripe al menos una vez.
    b. Más de 5 personas hayan tenido gripe al menos una vez.
    c. Como máximo 8 no hayan tenido gripe.
    d. Entre 7 y 11 no hayan tenido gripe.

Distribución de Poisson

20. Al servicio de emergencias de un hospital llegan, en promedio, 5 pacientes por hora. Con base en esta información calcule la probabilidad de que en una hora seleccionada al azar:

    a. Lleguen exactamente 2 pacientes
    b. Lleguen menos de 3 pacientes
    c. Lleguen más de 2 pacientes

21. Usted ha observado que la carretera por la que se dirige desde la universidad hasta su casa tiene un promedio de tres grietas por kilómetro. ¿Cuál es la probabilidad de que en un kilómetro seleccionado al azar de esa carretera:

    a. no haya grieta alguna?
    b. hayan al menos cinco grietas?

22. Un estudio reveló que a la fila de una caja de un supermercado llegan en promedio 15 clientes cada 20 minutos. Con base en esta información determine la probabilidad de:

    a. Que en una hora aleatoriamente seleccionada lleguen exactamente 50 clientes.
    b. Que en un plazo cualquiera de 10 minutos lleguen como máximo 10 clientes.
    c. Que en un plazo cualquiera de media hora lleguen como mínimo 20 clientes.

23. Según estudios realizados en una autopista del país, en un determinado tramo ocurren, en promedio, 5 accidentes por semana. Con esta información, calcule la probabilidad de que en una semana seleccionada aleatoriamente:

    a. Ocurran más de 7 accidentes.
    b. Ocurran menos de 4 accidentes.
    c. Ocurran entre 6 y 9 accidentes.
    d. No ocurran accidentes.

24. En la central telefónica de un hospital privado de la ciudad capital, en el servicio de medicina general, se reciben 2 llamadas en promedio cada media hora para realizar una cita con alguno de los médicos. Con esta información calcule las siguientes probabilidades de que:

   a. Se presenten menos de 3 llamadas en media hora.
   b. Se reciban más de 5 llamadas en un lapso de 45 minutos.
   c. Se reciban menos de 6 llamadas en un lapso de una hora.

Distribución multinomial

25. En una encuesta de intención de voto se obtuvo que el candidato A obtendría el 35% de los votos, el candidato C el 45% y el candidato B el restante 20%. Si se toma una muestra de 12 personas, ¿cuál es la probabilidad de que la mitad deseen votar por el candidato A, una cuarta partes por el candidato B y el resto por C?

26. Una empresa clasifica a su clientes como promotores, si están muy dispuestos a recomendar sus productos, como pasivos, si están poco dispuestos a recomendar sus productos, y como detractores si no tienen anuencia a recomendar los productos de la empresa. El 50% de los clientes de la empresa se han catalogado como promotores, el 40 como pasivos y el 10% como detractores. Calcule la probabilidad de que en una muestra de 10 clientes se obtenga 6 promotores, 3 pasivos y un detractor.

27. En una caja hay 20 bolas, 8 rojas, 6 verdes, 4 azules y 2 blancas. Si se seleccionan 8 bolas al azar, ¿cuál es la probabilidad de que 3 sean rojas, dos verdes, dos azules y una blanca?

28. En el control de calidad de una empresa los productos de una empresa se clasifican como perfectos, aceptables y como rechazados. Si el 80 de los productos se consideran perfectos, el 15% como aceptables y el 5% como rechazados, calcule la probabilidad de que en una muestra de 10 productos haya 8 perfectos, uno aceptable y otro rechazado.

Distribución geométrica

29. En el proceso de ensamblado de un camión de juguete a baterías se sabe que se obtiene una unidad defectuosa con una probabilidad de 4%. Periódicamente se efectúan pruebas para controlar la calidad del producto. Determine la probabilidad de encontrar el primer juguete defectuoso en:

   a. Exactamente la quinta prueba.
   b. Antes de la sexta prueba.
   c. Después de la tercera prueba.

30. Una central telefónica de una empresa está ocupada todo el tiempo. Si la probabilidad de lograr hacer una llamada en un momento de alta congestión es del 25%, calcule la probabilidad de que se tengan que hacer 6 intentos para lograr comunicarse.

31. Un tirador experto acierta en el blanco con una probabilidad del 90%. Calcule la probabilidad de que al hacer 8 disparos falle por primera vez en el octavo intento.

32. Una compañía tiene la costumbre de ofrecer descuentos a los radioescuchas de un popular programa de radio, los cuales llaman por teléfono para participar en el sorteo. Durante el lapso del programa la central telefónica tiende a saturarse, por lo que la probabilidad de que la llama de una persona ingrese es de solo 5%. Suponga que una persona desea participar en el sorteo, calcule la probabilidad de que tenga que:

   a. hacer 10 llamadas para participar.
   b. realizar más de 5 llamadas para participar.
   c. logre participar en el sétimo intento.

Ejercicios sobre varias distribuciones

33. La isla Pinta es la isla más al norte en las Islas Galápagos. Hace algunos años había gran cantidad de tortugas de la especie Chelonoidis abingdoni, una especie nativa de la isla Pinta pero que se extinguió hace un tiempo. El último espécimen fue conocido como el Solitario George. Un grupo de investigadores tomaron muestras de 1600 tortugas gigantes en las cercanías de un volcán en las Galápagos y encontraron que 17 tortugas (nueve hembras, tres machos y cinco jóvenes) eran "híbridas", es decir, que podrían ser descendientes de Chelonoidis abingdoni.

   a. Si se toma una muestra aleatoria de 50 tortugas, ¿cuál es la probabilidad de que a lo sumo dos sean descendientes de Chelonoidis abingdoni?
   b. Si se tienen 100 tortugas y se sabe que 3 de ellas son descendientes de Chelonoidis abingdoni, ¿cuál es la probabilidad de que en una muestra de 12 de ellas exactamente una sea descendiente de Chelonoidis abingdoni?
   c. Suponga que se puede encontrar un promedio de una tortuga descendiente de Chelonoidis abingdoni por cada hectárea de una cierta zona de la isla Pinta, ¿cuál es la probabilidad de que en una hectárea aleatoriamente seleccionada se encuentren exactamente dos tortugas descendientes de Chelonoidis abingdoni?

d. Si se examinan tortugas en la zona de las Galápagos, ¿cuál es la probabilidad de que la primera tortuga descendiente de Chelonoidis abingdoni sea la décimo novena?

e. Considere que se tienen las 17 tortugas descendientes de Chelonoidis abingdoni que se encontraron y se toma una muestra de 6 de ellas, ¿cuál es la probabilidad de que dos sean hembras, una macho y tres sean jóvenes?

34. Un empresa posee 500 empleados. Se sabe que 100 de ellos trabajan en el área administrativa, un 50 laboran en ventas y el resto en producción. Calcule la probabilidad de:

a. Que al seleccionar 9 empleados aleatoriamente, 2 sean del área administrativa en un muestreo con reemplazo.

b. Que al seleccionar 9 empleados aleatoriamente, 2 sean del área administrativa en un muestreo sin reemplazo.

c. Que al seleccionar 9 empleados aleatoriamente, 2 sean del área administrativa, uno de ventas y el resto de producción, en un muestreo con reemplazo.

d. Que al seleccionar varios empleados, el primero que pertenezca al área de ventas sea el quinto.

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. La distribución de probabilidad que se aplica en un experimento de acuerdo con un proceso de Bernoulli y tiene más de dos resultados posibles se llama:
(a) Binomial (b) Hipergeométrica
(c) Multinomial (d) Poisson

2. La distribución de probabilidad que representa el número de resultados que ocurren en un intervalo de tiempo dado o en un área o volumen específico se llama:
(a) Binomial (b) Hipergeométrica
(c) Multinomial (d) Poisson

3. La distribución de probabilidad que se emplea en una sucesión de pruebas y se quiere saber el número de la prueba en que ocurre el primer éxito se llama:
(a) Binomial (b) Hipergeométrica
(c) Geométrica (d) Poisson

4. La distribución de probabilidad que se emplea en una situación similar a un proceso de Bernoulli, pero con un muestreo sin reemplazo, se llama:
   (a) Binomial (b) Hipergeométrica
   (c) Geométrica (d) Poisson

5. A continuación se muestra la función de distribución de probabilidad para el número de accidentes por día que se presentan en una fábrica (nunca se presentan más de 4 accidentes).

x	0	1	2	3	4
P(x)	0,40	0,30		0,10	0,05

¿Cuál es la probabilidad de que se presente en un día cualquiera dos o más accidentes?
   (a) 0,85 (b) 0,15
   (c) 0,30 (d) Ninguna de las anteriores

6. Con base en la tabla del ejercicio 4, en el largo plazo, el número esperado de accidentes diarios en esa fábrica es de:
   (a) 0,8 (b) 2
   (c) 1,1 (d) Ninguna de las anteriores

7. Con base en la tabla del ejercicio 4, la desviación estándar de la distribución de probabilidad es:
   (a) 1,18 (b) 0,1215
   (c) 1,39 (d) 3,68

8. A continuación se muestra la función de distribución de probabilidad para el número de accidentes por día que se presentan en una fábrica (nunca se presentan más de 4 accidentes).

x	0	1	2	3	4
P(x)	0,30		0,20	0,10	0,02

¿Cuál es la probabilidad de que se presente en un día cualquiera dos o menos accidentes?
   (a) 0,78 (b) 0,88
   (c) 1 (d) Ninguna de las anteriores

9. Con base en la tabla del ejercicio 8, en el largo plazo, el número esperado de accidentes diarios en esa fábrica es de:
   (a) 0 (b) 1,6
   (c) 2 (d) 1,16

10. Con base en la tabla del ejercicio 8, la desviación estándar de la distribución de probabilidad es:
   (a) 1,05 (b) 1,08
   (c) 1,03 (d) Ninguna de las anteriores

11. Las acciones de la empresa A tienen una probabilidad de 0,7 de devolver una ganancia de $200. También tienen una probabilidad de 0,3 de tener una pérdida de $600. En el largo plazo, ¿cuál es la mejor opción de las siguientes que se puede hacer para maximizar su beneficio, y por qué?
   (a) Invertir en las acciones porque hay una mayor probabilidad de ganar dinero que perder dinero.
   (b) No invertir en las acciones debido a la cantidad de dinero por cada pérdida es mayor que el monto en dólares para cada ganancia.
   (c) Invertir en las acciones porque la inversión tiene un valor esperado positivo.
   (d) No invertir en las acciones debido a que el valor esperado es una pérdida.

12. Las acciones de la empresa A tienen una probabilidad de 0,7 de devolver una ganancia de $200. También tienen una probabilidad de 0,3 de tener una pérdida de $600. Las acciones de la empresa B tienen una probabilidad de 0,3 de devolver una ganancia de $600 y una probabilidad de 0,7 de tener una pérdida de $200. En el largo plazo, usando la desviación estándar como medida del riesgo, es cierto que:
   (a) Las acciones de la empresa A son más riesgosas que las acciones de la empresa B
   (b) Las acciones de la empresa A son menos riesgosas que las acciones de la empresa B
   (c) Las acciones de la empresa A son igualmente riesgosas que las acciones de la empresa B
   (d) Falta información para determinar la desviación estándar

13. Si usted toma una muestra de 15 artículos con reemplazo, para conocer si se presentan unidades con algún defecto, entonces se emplea la distribución:
   (a) Binomial (b) Hipergeométrica
   (c) Multinomial (d) Geométrica

14. En un proceso de producción se genera una unidad defectuosa por cada 10 unidades producidas. Si usted desea saber la probabilidad de que, en un muestra de 20 unidades sin reemplazo, se presenten 2 defectuosas, debería emplear la distribución:
   (a) Binomial (b) Hipergeométrica
   (c) Multinomial (d) Ninguna de las anteriores

15. La tasa media de llegadas de vehículos a un peaje es de 10 por minuto. Si usted desea saber la probabilidad de que en una hora seleccionada aleatoriamente lleguen menos de 50 vehículos, entonces usaría:
   (a) Binomial (b) Exponencial
   (c) Poisson (d) Normal

16. La tasa media de llegadas de vehículos a un peaje es de 15 por minuto. Si usted desea saber la probabilidad de que pasen 4 minutos entre la llegada de dos vehículos en una hora seleccionada aleatoriamente, entonces usaría:
(a) Binomial (b) Exponencial
(c) Poisson (d) Normal

17. Si se sabe que, en un problema binomial, la probabilidad de éxito es 0.60, en 10 ensayos, la probabilidad de obtener exactamente 4 fracasos es, aproximadamente:
(a) 0,1115 (b) 0,5630
(c) 0,2508 (d) Ninguna de las anteriores

18. Si se sabe que, en un problema hipergeométrico, hay 5 éxitos en una población de 12 unidades, en una muestra de 4 unidades la probabilidad de obtener exactamente 3 fracasos es, aproximadamente:
(a) 0,1414 (b) 0,6465
(c) 0,3535 (d) Ninguna de las anteriores

19. Si se sabe que, en un problema geométrico, la probabilidad de éxito es 0.1, entonces la probabilidad de que el primer éxito sea el tercero es, aproximadamente:
(a) 0,919 (b) 0,271
(c) 0,081 (d) Ninguna de las anteriores

20. Si se sabe que, en un problema hipergeométrico, hay 7 éxitos en una población es de 10 unidades, en una muestra de 4 unidades la probabilidad de obtener al menos 3 fracasos es, aproximadamente:
(a) 0,7381 (b) 0,9762
(c) 0,0333 (d) 0,2381

21. Si se sabe que, en un problema hipergeométrico, hay 4 éxitos en una población es de 9 unidades, en una muestra de 4 unidades la probabilidad de obtener a lo sumo 2 fracasos es, aproximadamente.
(a) 0,1667 (b) 0,3571
(c) 0,6429 (d) 0,8333

22. Si se sabe que *x* sigue una distribución de Poisson con media igual a 3, la probabilidad de *x* sea mayor que 2 es:
(a) 0,4232 (b) 0,8009
(c) 0,5768 (d) Ninguna de las anteriores

23. Si se sabe que *x* sigue una distribución de Poisson con media igual a 5, la probabilidad de *x* sea cuando mucho 1 es:
(a) 0,9933 (b) 0,0337
(c) 0,0404 (d) Ninguna de las anteriores

24. Si los resultados del análisis de un producto pueden ser bueno, regular o malo, y se conoce que las probabilidades de dichos resultados son 0.6, 0.3 y 0.1, respectivamente, entonces en una muestra de 5 unidades, la probabilidad de que una de ellas sea clasificada como regular, 1 como mala y 3 como buenas es:
    (a) 0,0036 (b) 0,0324
    (c) 0,1296 (d) Ninguna de las anteriores

25. Si los resultados del análisis de un producto pueden ser bueno, regular o malo, y se conoce que las probabilidades de dichos resultados son 0.6, 0.3 y 0.1, respectivamente, entonces en una muestra de 5 unidades, la probabilidad de que 4 de ellas sean clasificadas como buenas es:
    (a) 0,9222 (b) 0,7408
    (c) 0,2592 (d) Ninguna de las anteriores

26. Si se sabe que, en un problema binomial, la probabilidad de éxito es 0.60, en 10 ensayos, la cantidad esperada de fracasos es:
    (a) 0,0016 (b) 6
    (c) 4 (d) Ninguna de las anteriores

27. Si usted controla la calidad de las piezas compradas a un proveedor y desea calcular la probabilidad de que en un muestreo sin reemplazo se rechace el lote por contener más de 3 piezas defectuosas, entonces se emplea la distribución:
    (a) Binomial (b) Hipergeométrica
    (c) Multinomial (d) Geométrica

28. Se tiene un cargamento de 60 alarmas contra robo el cual contiene 9 defectuosas. La probabilidad de que salgan exactamente 2 defectuosas en una muestra de 5 alarmas es:
    (a) 0,8627 (b) 0,1886
    (c) 0,1373 (d) Ninguna de las anteriores

29. Un fabricante de medicamentos sostiene que cierta medicina cura una enfermedad para la sangre en el 80% de los casos. Para verificarlo los inspectores del gobierno utilizan una muestra de 100 individuos y deciden aceptar la afirmación si se curan por lo menos 75 de ellos. La probabilidad de que lo que dice sea rechazado, si efectivamente la probabilidad de curación es del 80%, es:
    (a) 0,9162 (b) 0.0838
    (c) 0,4567 (d) Ninguna de las anteriores

30. En un proceso de manufactura se sabe que la probabilidad de obtener una pieza defectuosa es de 2%. ¿Cuál es la probabilidad de que la octava pieza inspeccionada sea la primera defectuosa?
    (a) 0,9826 (b) 0,9800
    (c) 0,0174 (d) Ninguna de las anteriores

31. Un fabricante sabe que cierto tipo de refrigeradores tienen una probabilidad de 0,8 de clasificarse como aceptable, una probabilidad de 0,15 de ser clasificados como con defectos secundarios y de 0,05 de ser clasificados como con defectos mayores. Si se revisan seis refrigeradores, escogidos al azar, la probabilidad de que tres sean aceptables, 2 tengan defectos menores y 1 tenga defecto mayor es:
    (a) 0,9654 (b) 0,7645
    (c) 0,0346 (d) Ninguna de las anteriores

32. Una empresa de mercadeo por internet tiene una promoción por e–mail que produce una respuesta de 15%. Si se hace un envío de dicha promoción a 10 clientes (independientes), la probabilidad de que nadie responda es:
    (a) 0,0000 (b) 0,8031
    (c) 0,1969 (d) Ninguna de las anteriores

33. Una empresa de mercadeo por internet tiene una promoción por e–mail que produce una respuesta de 15%. Si se hace un envío de dicha promoción a 10 clientes (independientes), la probabilidad de que exactamente dos personas respondan es:
    (a) 0,0000 (b) 0,8241
    (c) 0,2759 (d) Ninguna de las anteriores

34. Una empresa de mercadeo por internet tiene una promoción por e–mail que produce una respuesta de 15%. Si se hace un envío de dicha promoción a 10 clientes (independientes), la probabilidad de que más de la mitad respondan es:
    (a) 0,0000 (b) 0,9986
    (c) 0,0014 (d) Ninguna de las anteriores

35. Una empresa de mercadeo por internet tiene una promoción por e–mail que produce una respuesta de 15%. Si se hace un envío de dicha promoción a 10 clientes (independientes), la probabilidad de que más de 4 no respondan es:
    (a) 0,0099 (b) 0,0014
    (c) 0,9986 (d) Ninguna de las anteriores

36. Considere la siguiente información: "En una encuesta aplicada a 700 hogares a nivel nacional, de los cuales la mitad tienen actualmente acceso al servicio de telefonía celular, se obtuvieron los siguientes datos: ante la apertura en el mercado de telecomunicaciones, el 68% los usuarios actuales de telefonía celular estaría dispuesto a cambiar de operador. Entre la población que aun no posee celular, solo un 38% optará por el operador actual, mientras que el resto escogerá un nuevo proveedor de servicio". Si se toma una muestra de 6 hogares que ya poseen servicio celular, la probabilidad de que todos conserven el operador actual es:
    (a) 0,0011 (b) 0,0989
    (c) 0,0030 (d) 0,0568

37. Utilizando la misma información del ejercicio 36, si se toma una muestra de 6 hogares que ya poseen servicio celular, la probabilidad de que 2 o menos hogares cambien su proveedor actual de telefonía celular es:
(a) 0,0011 (b) 0,0875
(c) 0,7064 (d) 0,1527

38. Utilizando la misma información del ejercicio 36, si se toma una muestra de 6 hogares que aun poseen servicio celular, la probabilidad de que 4 o más hogares utilicen el proveedor actual en el mercado de telefonía celular es:
(a) 0,7064 (b) 0,1202
(c) 0,3201 (d) 0,1527

39. Utilizando la misma información del ejercicio 36, si se toma una muestra de 6 hogares que aun poseen servicio celular, la probabilidad de que 4 o más hogares utilicen el proveedor actual en el mercado de telefonía celular es:
(a) 0,7064 (b) 0,1202
(c) 0,3201 (d) 0,1527

40. El jefe de un departamento de recursos humanos de una empresa grande, estudia con frecuencia el grado de satisfacción de los trabajadores dentro de la empresa, y ha encontrado que 4 de cada 20 empleados se siente insatisfecho con su salario. Esta proporción se ha mantenido constante durante mucho tiempo. Si se seleccionan aleatoriamente 8 personas, la probabilidad de que exactamente 3 de ellas se sientan insatisfechas con su salario es:
(a) 0,7064 (b) 0,1202
(c) 0,3201 (d) 0,1468

41. Es un proceso de Bernoulli es falso que:
   (a) Existen solamente dos resultados posibles en cada ensayo
   (b) La probabilidad de un éxito permanece constante en todos los intentos.
   (c) Todos los intentos repetidos son independientes.
   (d) Ninguna de las anteriores

42. La probabilidad de que cierto componente falle ante una carga axial específica es de 5%. La probabilidad de que entre 16 de tales componentes fallen entre 2 y 5:
(a) 0,1891 (b) 0,8109
(c) 0,95 (d) Ninguna de las anteriores

43. Con respecto a las láminas de zinc esmaltadas que se emplearán en el techo de un edificio nuevo, se sabe que el 95% no tienen defecto alguno, que el 4% tienen, en promedio, un defecto menor en el esmalte por cada dos metros cuadrados de lámina, y el resto poseen huecos u otros defectos mayores, y por tanto serán devueltas al proveedor. La probabilidad de que al seleccionar una muestra aleatoria de 8 láminas haya que devolver a lo sumo una lámina es:
(a) 0,9926 (b) 0,0027
(c) 0,0074 (d) Ninguna de las anteriores

44. En relación a la misma situación de la pregunta 43, la probabilidad de que al seleccionar una muestra aleatoria de 10 láminas haya al menos 8 en perfecto estado es:
(a) 0,0861 (b) 0,9238
(c) 0,0115 (d) Ninguna de las anteriores

45. En relación a la misma situación de la pregunta 43, si se selecciona una lámina al azar correspondiente a las que tienen un defecto en el esmalte, entonces la probabilidad de que posea una superficie continua de 1,5 metros cuadrados sin defecto alguno es:
(a) 0,5276 (b) 0,3679
(c) 0,4724 (d) Ninguna de las anteriores

46. En relación a la misma situación de la pregunta 43, si un empleado está inspeccionando las láminas, entonces la probabilidad de que la primera lámina con defectos en el esmalte sea la sexta es:
(a) 0,0340 (b) 0,0326
(c) 0,0311 (d) Ninguna de las anteriores

47. En relación a la misma situación de la pregunta 43, suponga que se han comprado 100 láminas. Si se toma una muestra aleatoria sin reemplazo de 6 láminas, la probabilidad de que exactamente 5 estén en perfecto estado es:
(a) 0,2430 (b) 0,2709
(c) 0,2649 (d) Ninguna de las anteriores

48. En relación a la misma situación de la pregunta 43, si se toma una muestra aleatoria de 8 láminas, la probabilidad de que 6 estén en perfecto estado, que una tenga un defecto en el esmalte y otra un defecto mayor es:
(a) 0,0261 (b) 0,0138
(c) 0,0315 (d) Ninguna de las anteriores

49. En una distribución binomial se cuenta la probabilidad de obtener un número establecidos de éxitos cuando:
(a) el número de intentos es constante pero la probabilidad de fracaso no
(b) el número de intentos no es constante ni la probabilidad de fracaso
(c) el número de intentos es constante y la probabilidad de fracaso también
(d) el número de intentos no es constante pero la probabilidad de fracaso sí

50. Considere las dos siguientes afirmaciones:
A. En una distribución de Poisson es verdadero que $E(X) = V(X) = \lambda = \sigma^2$
B. Cuando n es relativamente grande y p pequeña, las probabilidades binomiales a menudo se aproximan por medio de la distribución de Poisson.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Son verdaderas ambas (b) Solo B es verdadera
(c) Son falsas ambas (d) Solo A es verdadera

CAPÍTULO 2

DISTRIBUCIONES DE PROBABILIDAD DE VARIABLE CONTINUA

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Resolver problemas empleando la distribución normal
2. Resolver problemas empleando la distribución exponencial

2.1 Distribuciones continuas de probabilidad

En el capítulo anterior se trabajó con distribuciones de probabilidad en las cuales la variable aleatoria era discreta, es decir, se obtenía por conteo, pero en las distribuciones de probabilidad de variables aleatorias continuas el valor de la variable se obtiene por medición, como es el caso del peso de paquetes de harina, el diámetro del anillo de un pistón, la longitud de una pieza, la vida útil de un producto, el nivel de glucosa en la sangre, la temperatura de un horno, el tiempo de atención de un paciente, entre muchas otras variables que se obtienen a través de la medición, y por tanto se van a expresar en unidades de medida, como los gramos, milímetros, metros, años, segundos, etc.

Otra diferencia con respecto a las distribuciones de variable discreta, es que al calcular las probabilidades de variables aleatorias continuas no se puede hablar de la probabilidad de que tome un valor en particular, por ejemplo, calcular la probabilidad de que el tiempo de espera en una fila sea exactamente 4 minutos; sino que se calcula la probabilidad de que la variable tome valores en un intervalo, como sería que el tiempo de espera esté entre 3 y 5 minutos. Entonces, se encontrará la probabilidad de que la variable aleatoria x tome valores entre a y b: $P(a \leq x \leq b)$, como en el ejemplo, que sería obtener $P(3 \leq x \leq 5)$.

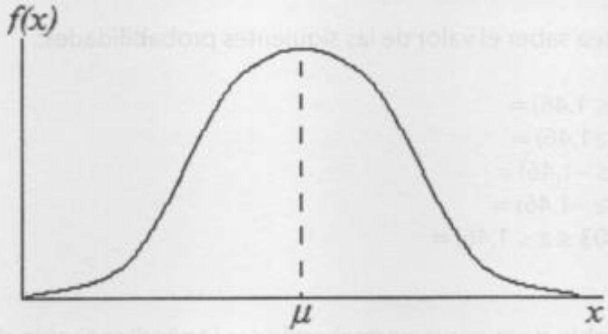
A las funciones de probabilidad de las variables continuas se les llama funciones densidad y se integran para obtener las probabilidades buscadas.

2.2 La distribución normal

La distribución de probabilidad de variable continua más importante es la distribución normal. Esta distribución, cuya curva tiene forma de campana, mide en forma muy aproximada muchos fenómenos que ocurren en la naturaleza, los negocios, la industria y la investigación, como por ejemplo:

- En una fábrica, las mediciones sobre las partes manufacturadas.
- Características físicas de las personas y otros seres vivos, tales como estatura, masa corporal, etc.
- Algunas variables financieras (principalmente el logaritmo de índices de precios, tasas de cambio, etc.).
- El cociente intelectual.
- Efecto de un fármaco en el organismo.
- Nivel de ruido en telecomunicaciones.
- La media muestral se distribuye normalmente, tal como se expondrá en el capítulo 8.

Esta distribución fue estudiada a partir del siglo XVII, cuando el matemático francés Abraham DeMoivre desarrolló la ecuación matemática de la distribución de probabilidad. Posteriormente el alemán Karl Friedrich Gauss también derivó la ecuación en un estudio de errores en mediciones repetidas de la misma cantidad.



La curva de la distribución, a la que se le llama campana de Gauss o gaussiana, satisface las siguientes propiedades básicas:

- Asintótica con respecto al eje horizontal.
- El área bajo la curva es igual a uno (se trata de una función de densidad).
- Es simétrica con respecto a la media aritmética (μ).
- Queda perfectamente determinada si se conocen μ y σ.

La función densidad de la distribución normal es:

$$f(x,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{para } -\infty < x < \infty$$

La probabilidad de que *x* tome valores en el intervalo entre *a* y *b*, es:

$$P(a \le x \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Se mencionó que la curva normal queda perfectamente determinada si se conocen m y s, lo cual indica que para cada par de valores de m y s existe una curva normal diferente, o sea, existe una cantidad infinita de curvas normales y dada la complejidad de integrar la expresión anteriormente dada, el trabajo con la curva normal sería muy complicado. Para resolver este problema se emplea una curva normal que tiene μ = 0 y σ = 1, llamada *distribución normal estándar*, así que, no es necesario integrar para obtener las probabilidades. La variable aleatoria de la distribución normal estándar, denotada por *z*, es:

$$z = \frac{x-\mu}{\sigma}$$

Los valores de esta distribución normal estándar se obtienen de una tabla, la cual da el valor de probabilidad para cada valor de *z*. La fórmula anterior se emplea para convertir de la variable *x* a *z* y viceversa, según se requiera.

El uso de la tabla de la distribución normal estándar acumulada (ver Apéndice 6) es muy sencillo.

Ejemplo

Se desea saber el valor de las siguientes probabilidades:

a) $P(z \leq 1{,}46) =$
b) $P(z \geq 1{,}46) =$
c) $P(z \leq -1{,}46) =$
d) $P(z \geq -1{,}46) =$
e) $P(1{,}03 \leq z \leq 1{,}46) =$

Solución

a) La tabla de la curva normal estándar (Apéndice 6) solo da probabilidades para valores acumulados hasta el número buscado, en otras palabras, la probabilidad de que la variable z sea menor o igual que cierto valor. Entonces, si se busca la probabilidad de que z sea menor o igual que 1,46, la tabla va a dar directamente el resultado.

Se desea conocer $P(z \leq 1{,}46)$, entonces en la tabla se busca el entero y el primer decimal, o sea, 1,4, en la primera columna, y luego el segundo decimal, en este caso 6, se busca en la primera fila, tal como se ilustra a continuación:

Segundo decimal

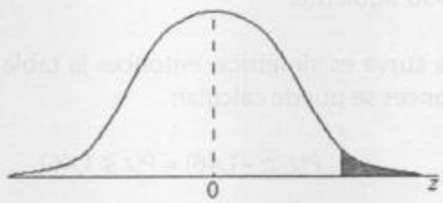
Distribución normal estándar acumulada

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952

Entero y primer decimal

El número que aparece donde se cruza esa fila con esa columna es el valor de la probabilidad, que en este caso es 0,9279.

b) Tal como se señaló anteriormente, la tabla de la distribución normal estándar del Apéndice 6 solo da la probabilidad de que la variable z sea menor o igual que cierto valor, pero en este caso se busca la probabilidad de que z sea mayor que 1,46. Gráficamente este problema se vería del modo siguiente:



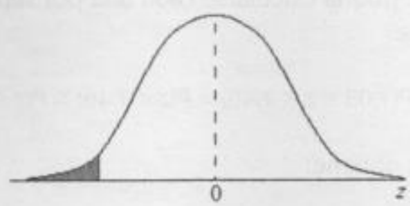
La tabla daría el área en blanco, no el área sombreada, pero sabiendo que el área total bajo la curva es igual a 1, entonces se puede calcular:

$$P(z \geq 1{,}46) = 1 - P(z \leq 1{,}46)$$

De la tabla se obtiene que $P(z \leq 1{,}46) = 0{,}9279$, por lo que:

$$P(z \geq 1{,}46) = 1 - P(z \leq 1{,}46) = 1 - 0{,}9279 = 0{,}0721$$

c) Como se señaló en los casos anteriores, la tabla de la distribución normal estándar (Apéndice 6) da la probabilidad de que la variable z sea menor o igual que cierto valor positivo, pero en este caso se busca la probabilidad de que z sea menor que $-1{,}46$. Gráficamente este problema se vería del modo siguiente:



Dado que la curva es simétrica, entonces la tabla daría el área en blanco, no el área sombreada, pero sabiendo que el área total bajo la curva es igual a 1, entonces se puede calcular:

$$P(z \leq -1{,}46) = 1 - P(z \leq 1{,}46)$$

De la tabla se obtiene que $P(z \leq 1{,}46) = 0{,}9279$, por lo que:

$$P(z \leq -1{,}46) = 1 - P(z \leq 1{,}46) = 1 - 0{,}9279 = 0{,}0721$$

d) Nuevamente sabemos que la tabla de la distribución normal estándar (Apéndice 6) da la probabilidad de que la variable z sea menor o igual que cierto valor positivo, pero en este caso se busca la probabilidad de que z sea mayor que −1,46. Gráficamente este problema se vería del modo siguiente:

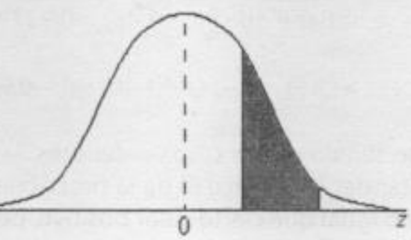
Dado que la curva es simétrica, entonces la tabla daría el área sombreada, entonces se puede calcular:

$$P(z \geq -1{,}46) = P(z \leq 1{,}46)$$

De la tabla se obtiene que $P(z \leq 1{,}46) = 0{,}9279$, por lo que:

$$P(z \geq -1{,}46) = P(z \leq 1{,}46) = 0{,}9279$$

e) Este problema se vería en forma gráfica del modo siguiente:



La tabla da el área acumulada hasta 1,46 y da el área acumulada hasta 1,03, por que podría calcularse cada una por separado y luego restar los resultados:

$$P(1{,}03 \leq z \leq 1{,}46) = P(z \leq 1{,}46) = P(z \leq 1{,}03) =$$

De la tabla se obtiene:

$$= 0{,}9279 - 0{,}8485 = 0{,}0794$$

Ejercicio de revisión

Calcule el valor de las siguientes probabilidades:

a) $P(z \leq 2{,}38) =$
b) $P(z \geq 3{,}01) =$
c) $P(z \leq -0{,}96) =$
d) $P(z \geq -2{,}81) =$
e) $P(-0{,}19 \leq z \leq 2{,}71) =$

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Ahora bien, en la práctica muchas variables se distribuyen normalmente, pero no tienen media igual a 0 y desviación estándar igual a 1. Es necesario entonces estandarizar estas variables, tal como se indicó anterior, para lo cual se emplea la fórmula siguiente:

$$z = \frac{x-\mu}{\sigma}$$

Ejemplo

La cantidad de refresco envasada por una empresa está normalmente distribuido con una media de un litro (1000 ml) y tiene desviación estándar de 30 ml. Calcule las probabilidades de que una botella aleatoriamente seleccionada tenga una cantidad de refresco:

a) De menos de 1010 ml.
b) Mayor que 1050 ml.
c) Por lo menos de 990 ml.
d) Como máximo de 1090 ml.
e) Entre 980 y 1040 ml.
f) ¿Cuál es el valor máximo del 20% de las botellas con menor cantidad de líquido?
g) ¿Cuál es el valor mínimo del 40% de las botellas con mayor cantidad de líquido?

Solución

Se tiene $\mu = 1000$ y $\sigma = 30$, y los valores de la probabilidad de z se obtienen de la tabla.

a) La probabilidad que se busca es $P(x \leq 1010)$. Para las distribuciones continuas menor o igual es lo mismo que estrictamente menor. Lo primero que se hace es aplicar la fórmula de estandarización para convertir x en z:

$$P(x \leq 1010) = P\left(z \leq \frac{1010-1000}{30}\right) = P(z \leq 0{,}33)$$

Para obtener dicha área se aplica la tabla de distribución normal estándar (Apéndice 6), de donde se obtiene:

$$P(Z \leq 0{,}33) = 0{,}6293$$

b) Se busca la probabilidad $P(x \geq 1050)$. Aplicando la fórmula de estandarización y la tabla normal estándar:

$$P(x \geq 1050) = P\left(z \geq \frac{1050-1000}{30}\right)$$
$$= P(z \geq 1{,}67) = 1 - 0{,}9525 = 0{,}0475$$

c) Se busca la probabilidad $P(x^{3}\ 990)$. Siguiendo los pasos señalados:

$$P(x \geq 990) = P\left(z \geq \frac{990 - 1000}{30} \right)$$

$$= P(z \geq -0{,}33) = 0{,}6293$$

d) Se requiere encontrar $P(x \leq 1090)$, entonces:

$$P(x \leq 1090) = P\left(z \leq \frac{1090 - 1000}{30} \right)$$

$$= P(z \leq 3) = 0{,}9987$$

e) En este caso la probabilidad buscada es $P(980 \leq x \leq 1040)$, por lo que a cada valor se aplica la fórmula de estandarización y luego la tabla normal estándar:

$$P(980 \leq x \leq 1040)$$

$$= P\left(\frac{980 - 1000}{30} \leq z \leq \frac{1040 - 1000}{30} \right)$$

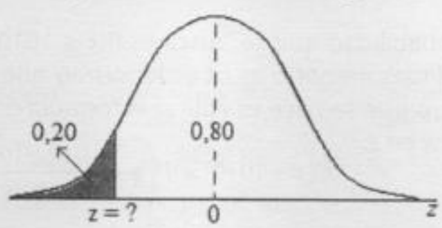
$$= P(-0{,}67 \leq z \leq 1{,}33)$$

$$= P(z \leq 1{,}33) - P(z \leq -0{,}67)$$

$$= 0{,}9082 - (1 - 0{,}7486)$$

$$= 0{,}9082 - 0{,}2514 = 0{,}6568$$

f) El valor máximo del 20% de las botellas con menor cantidad de líquido se encuentra al lado izquierdo de la curva, en el cual los valores de z son negativos, por estar a la izquierda de $z = 0$ ($\mu = 0$). Gráficamente el problema queda representado del modo siguiente:



Al buscar en la tabla el valor de z que corresponde a una probabilidad máxima de 0,20 se encuentra en la tabla del Apéndice 6 que solo aparecen valores positivos, y no negativos, pero esto no es problema dado que la curva es simétrica. También se observa que los valores de probabilidad de la tabla son iguales o mayores que 0,5, y no menores, por lo que el valor de 0,20 no va a aparecer, así que se busca su complemento $1 - 0{,}20 = 0{,}80$. De ese modo se busca el valor de probabilidad (no de z) más cercano a 0,80:

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Véase que el valor de probabilidad de 0,7995 es el más cercano a 0,80, por lo que en la primera columna se obtiene el entero y el primer decimal del valor de *z*, y en la primera fila el segundo decimal. Así, se obtiene que $z = 0{,}84$, pero se dijo que este valor debía ser negativo por encontrase del lado izquierdo de la gráfica, así que $z = -0{,}84$.

Ahora se sustituye y se despeja el valor de *x* de la fórmula de *z*:

$$z = \frac{x - \mu}{\sigma}$$

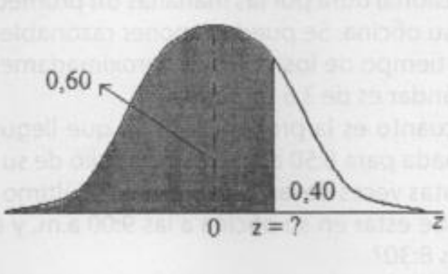
$$-0{,}84 = \frac{x - 1000}{30}$$

$$x = -0{,}84 * 30 + 1000$$

$$x = 974{,}8$$

Esto indica que 974,8 ml es el valor máximo del 20% de las botellas con menor cantidad de líquido.

g) El valor mínimo del 40% de las botellas con mayor cantidad de líquido se encuentra al lado derecho de la curva, en el cual los valores de *z* son positivos, por estar a la derecha de $z = 0$ ($\mu = 0$). Gráficamente el problema queda representado del modo siguiente:



Se observa que los valores de probabilidad de la tabla de la distribución normal estándar (Apéndice 6) son iguales o mayores que 0,5, y no menores que 0,5, por lo que el valor de 0,40 no va a aparecer, así que se debe buscar su complemento 1 – 0,40 = 0,60. De ese modo se busca el valor de probabilidad (no de *z*) más cercano a 0,60:

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

Véase que el valor de probabilidad de 0,5987 es el más cercano a 0,60, por lo que en la primera columna se obtiene el entero y el primer decimal del valor de *z*, y en la primera fila el segundo decimal. Así, se obtiene que $z = 0,25$.

Ahora se sustituye y se despeja el valor de *x* de la fórmula de *z*:

$$z = \frac{x - \mu}{\sigma}$$

$$0,25 = \frac{x - 1000}{30}$$

$$x = 0,25 * 30 + 1000$$

$$x = 1007,5$$

Esto indica que 1007,5 ml es el mínimo del 40% de las botellas con mayor cantidad de líquido.

Ejemplo

Un profesional dura por las mañanas un promedio de 26 minutos para llegar a su oficina. Se puede suponer razonablemente que la distribución del tiempo de los viajes es aproximadamente normal. La desviación estándar es de 3,5 minutos.

a. ¿De cuánto es la probabilidad de que llegue tarde a una reunión programada para 8:50 a.m. si ese día salió de su casa a las 8:35?

b. ¿Cuántas veces de las 120 que viajó el último semestre llegó a tiempo, si debe estar en su oficina a las 9:00 a.m. y acostumbra salir de su casa a las 8:30?

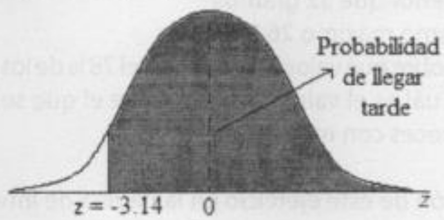
c. Encuentre el tiempo máximo que le tomó el 62% de los viajes más rápidos.

Solución

Se tiene que $\mu = 26$ minutos y $\sigma = 3{,}5$ minutos.

a. Si salió de su casa a las 8:35 y la reunión es a las 8:50 llegará tarde si el viaje el toma más de los 15 minutos con que cuenta.

$$P(\text{llegar tarde}) = P(x \geq 15) = P(Z \geq -3{,}14) = 0{,}9992$$

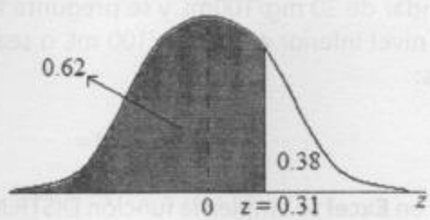


b. Si sale a las 8:30 y tiene que estar en la oficina a las 9 cuenta con 30 minutos para llegar.

$$P(\text{llegar a tiempo}) = P(x \leq 30) = P(Z \leq 1{,}14) = 0{,}8729$$

De esta forma, el número de veces que llegó a tiempo = 120 × 0,8729 = 104,75. Por tanto, llegará a tiempo entre 104 y 105 veces de las 120 del semestre.

c. Los viajes más rápidos son los que toman menos tiempo, por lo tanto, el área es el 62% del lado izquierdo.



Usando la fórmula de estandarización:

$$x = 26 + 0{,}31 * 3{,}5 = 27{,}09$$

De ese modo se tiene que 27,09 minutos es el tiempo máximo que toma el 62% de los viajes más rápidos.

Ejercicio de revisión

Un biólogo ha determinado que el peso promedio de los alevines de cierta especie de tilapia se distribuye normalmente con media de 30 gramos a los 120 días de cultivo y una desviación estándar de 4,5 gramos. Calcule la probabilidad de que al seleccionar una de estas tilapias al azar tenga un peso:

a) Mayor que 34 gramos
b) Menor que 32 gramos
c) Como máximo 26,8 gramos
d) ¿Sobre qué valor se encuentra el 78% de los peces de mayor peso?
e) ¿Cuál es el valor más alto sobre el que se encuentra el 35% de los peces con menor peso?

Ver solución de este ejercicio en la página de internet de este texto

Uso de Excel y Minitab para la distribución normal

Ejemplo

Se conoce que el nivel de colesterol en sangre en una población adulta entre 50 y 60 años se distribuye normalmente con una media de 180 mg/100 ml de sangre y que la desviación estándar es de 30 mg/100 ml. Calcule, usando Excel y Minitab, la probabilidad de que uno de esos adultos entre 50 y 60 años tenga un nivel inferior a 200 mg/100 ml de sangre.

En **Excel**: Se tiene que una media de 180 mg/100ml con una desviación estándar de 30 mg/100ml, y se pregunta la probabilidad de que tenga un nivel inferior a 200 mg/100 ml, o sea, que lo que se desea calcular es:

$$P(X < 200) =$$

Entonces, en **Excel** se emplea la función DISTR.NORM, cuya sintaxis es:

= DISTR.NORM(x;media;desv_estándar;acum)

Los argumentos de la función anterior son:

x: número establecido de la variable (*x*)
media: promedio (μ)
desv_estándar: desviación estándar (s)
acumulado: 0 si no es acumulado o 1 si es acumulado

Así en este caso, se completa la función en la celda en la que se desea el resultado (se indica al final 1 para que dé el resultado acumulado):

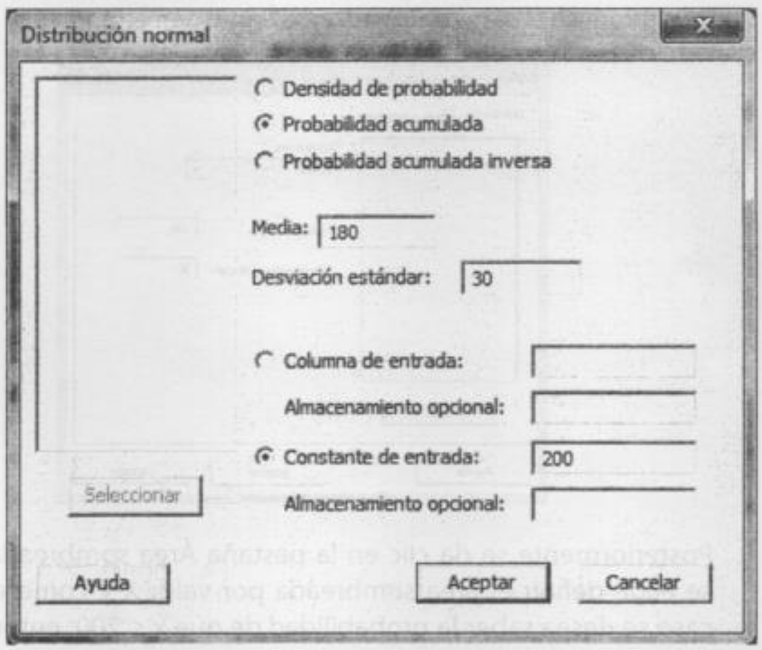
=DISTR.NORM(200;180;30;1)

El resultado es 0,7475.

En **Minitab**: Se tiene que una media de 180 mg/100ml con una desviación estándar de 30 mg/100ml, y se pregunta la probabilidad de que tenga un nivel inferior a 200 mg/100 ml, o sea, que lo que se desea calcular es:

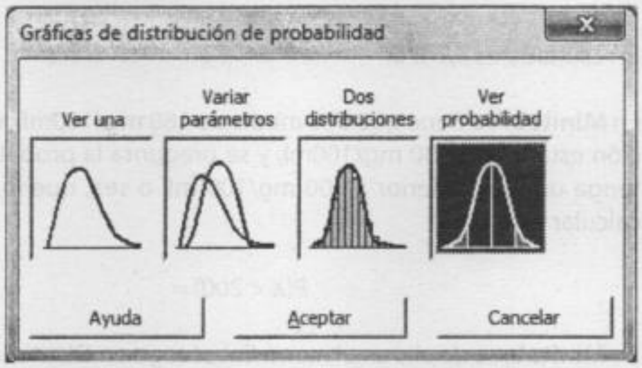
$$P(X < 200) =$$

En **Minitab**, se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Normal. Se completa el cuadro de diálogo:

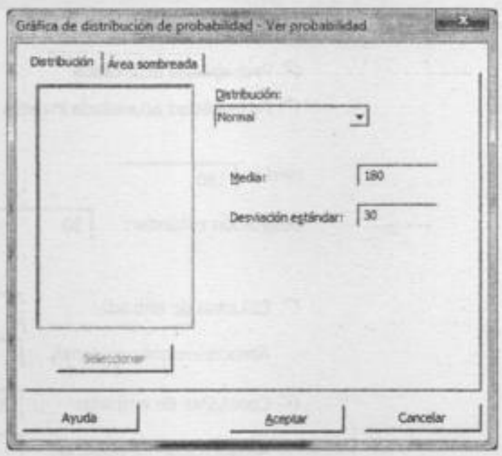


Se selecciona probabilidad acumulada para que calcule el valor de que *x* sea menor que 200 y se completan los datos tal como se muestra en la imagen. Luego se da clic en Aceptar y se obtiene el resultado 0,7475 en la ventana Sesión.

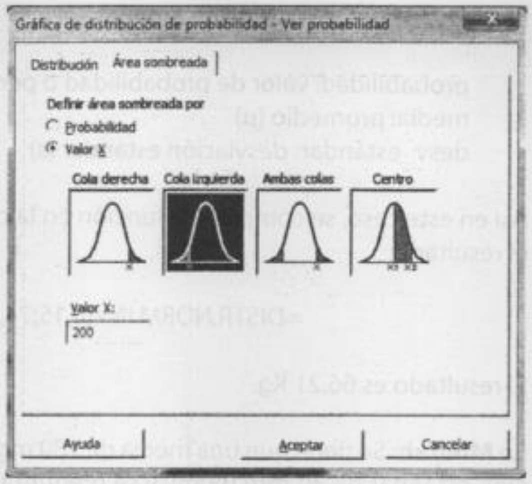
También, se puede hacer uso del menú Gráfica, donde se selecciona Gráfica de distribución de probabilidad. En el cuadro de diálogo se selecciona la opción que dice Ver probabilidad.



En el cuadro de diálogo se selecciona en la lista la distribución normal y se introduce el dato de la media y la desviación estándar:

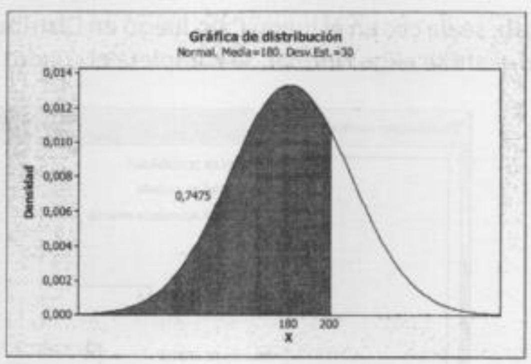


Posteriormente se da clic en la pestaña Área sombreada. Aquí se elige definir el área sombreada por valor X y como en este caso se desea saber la probabilidad de que $x \leq 200$, entonces se selecciona Cola izquierda y se escribe el valor de x en el espacio que aparece:



Al dar clic en Aceptar, Minitab crea un gráfico que indica el valor de la probabilidad:

Solución



Ejemplo:

Si el peso promedio de un hombre adulto es 74,8 kilogramos con una desviación estándar de 8 kilogramos. Si las medidas se distribuyen según una distribución normal, calcule, usando Excel y Minitab, el peso que separa el 15% de los hombres adultos con menor peso.

En **Excel**: Se tiene que una media de 74,5 Kg. con una desviación estándar de 8 Kg., y se pregunta el peso que separa el 15% de los hombres adultos con menor peso. Entonces, en **Excel** se emplea la función DISTR.NORM.INV, cuya sintaxis es:

= DISTR.NORM.INV(probabilidad;media;desv_estándar)

Los argumentos de la función anterior son:

probabilidad: valor de probabilidad o percentil dado
media: promedio (μ)
desv_estándar: desviación estándar (s)

Así en este caso, se completa la función en la celda en la que se desea el resultado:

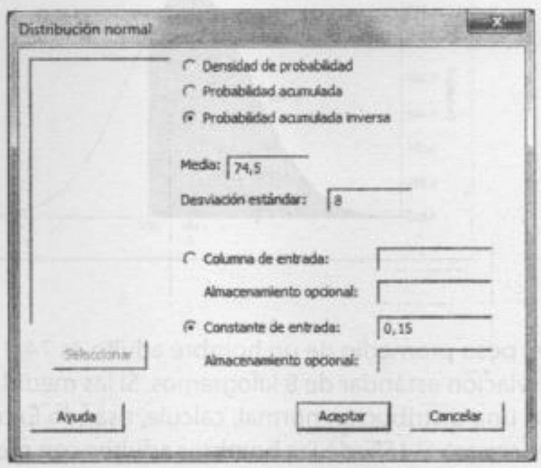
=DISTR.NORM.INV(0,15;74,5;8)

El resultado es 66,21 Kg.

En **Minitab**: Se tiene que una media de 180 mg/100ml con una desviación estándar de 30 mg/100ml, y se pregunta la probabilidad de que tenga un nivel inferior a 200 mg/100 ml, o sea, que lo que se desea calcular es:

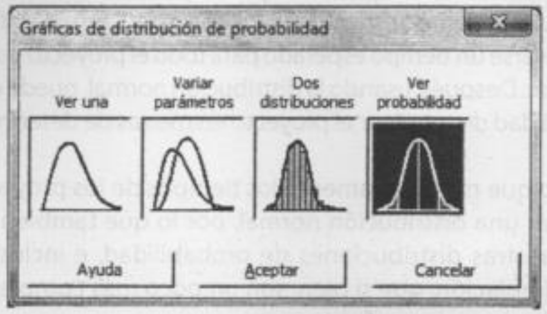
$$P(X < 200) =$$

En **Minitab**, se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Normal. Se completa el cuadro de diálogo:



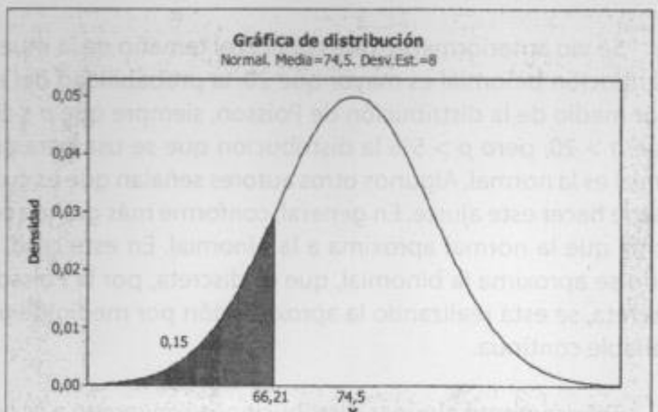
Se selecciona probabilidad acumulada inversa para que devuelva el valor de la variable en vez de calcular la probabilidad y se completan los datos tal como se muestra en la imagen. Luego se da clic en Aceptar y se obtiene el resultado 66,21 Kg. en la ventana Sesión.

También, se puede hacer uso del menú Gráfica, donde se selecciona Gráfica de distribución de probabilidad. En el cuadro de diálogo se selecciona la opción que dice Ver probabilidad.



En el cuadro de diálogo se selecciona en la lista la distribución normal y se introduce el dato de la media y la desviación estándar:

Solución



Aplicación

Administración de proyectos

Cuando se administra un proyecto se tiene que planear un conjunto de actividades y la duración de las mismas. Hay muchos factores que pueden provocar que una actividad dure más tiempo, o menos tiempo, que lo que se planea. Esto genera un riesgo para los encargados del proyecto, es decir, no poder concluir el proyecto en el plazo comprometido.

Se han desarrollado técnicas que permiten determinar la probabilidad de terminar el proyecto en un determinado plazo. Esto es fundamental y valioso para tomar decisiones en la gestión del proyecto.

Una técnica bastante conocida consiste en calcular los tiempos de terminación de las distintas actividades bajo tres escenarios diferentes, uno optimista, otro pesimista y uno que sea el más probable. Luego se calcula un tiempo esperado y una varianza. Con esta información podrá calcularse un tiempo esperado para todo el proyecto y su desviación estándar. Después, usando la distribución normal, puede estimarse la probabilidad de terminar el proyecto en menos de determinado tiempo.

Claro que no estrictamente los tiempos de los proyectos tienen que seguir una distribución normal, por lo que también es posible emplear otras distribuciones de probabilidad, e incluso usar técnicas de simulación, que si bien, son un poco más complejas, pueden dar resultados más confiables.

2.3 Aproximación de la distribución normal a la binomial

Se vio anteriormente que cuando el tamaño de la muestra tomada en la distribución binomial es mayor que 20, la probabilidad del evento se obtenía por medio de la distribución de Poisson, siempre que $p \leq 5\%$. En los casos en que $n > 20$, pero $p > 5\%$ la distribución que se usa para aproximarse a la binomial es la normal. Algunos otros autores señalan que es cuando $n \geq 30$ que se puede hacer este ajuste. En general, conforme más grande sea n, mejor será la forma que la normal aproxima a la binomial. En este caso, a diferencia de cuando se aproxima la binomial, que es discreta, por la Poisson que también es discreta, se está realizando la aproximación por medio de una distribución de variable continua.

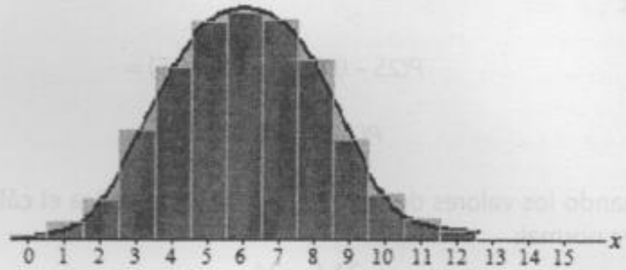
Teóricamente algunas distribuciones convergen a la normal a medida que sus parámetros se aproximan a ciertos límites. La distribución binomial se aproxima bastante bien con la normal.

Teorema. Si X es una variable aleatoria binomial con media $\mu = np$ y varianza $\sigma^2 = npq$, y si:

$$z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}$$

entonces la expresión límite de la función de distribución de esa variable aleatoria estandarizada cuando $n \to \infty$ es la distribución normal estándar.

Cuando n es pequeña la aproximación es aún bastante buena si p es razonablemente cercana a 0,5. La figura que se presenta muestra la aproximación por la normal de una distribución binomial con $n = 15$ y $p = 0{,}4$. Las barras muestran las probabilidades según la binomial y la campana corresponde a la distribución normal.



Ejemplo

Un ingeniero de sistemas cree que el 30% de las empresas estarían dispuestas a actualizar el sistema operativo de sus equipos de cómputo a la nueva versión que va a ser lanzada al mercado dentro de poco tiempo. De acuerdo con ese dato, calcule la probabilidad de que de una muestra de 80 empresas:

a) entre 25 y 35 actualicen su sistema operativo.
b) por lo menos 20 actualicen su sistema operativo.
c) menos de 60 no actualicen su sistema operativo.

Este es un problema de distribución binomial, pero que por tener $n > 20$ se resuelve por aproximación. En este caso no se puede aplicar la Poisson porque $p > 5\%$, por lo tanto se resuelve usando la normal como aproximación.

Al aplicarse la normal se debe realizar una corrección por continuidad debido a que se está resolviendo un problema de variable discreta con una distribución de variable continua, para lo cual se restará 0,5 y se sumará 0,5 a los valores de x en el cálculo de la probabilidad, tal como se explicará más adelante.

a) La probabilidad de que las empresas actualicen su sistema operativo es de 30%, por lo tanto:

$$\mu = np = 80 \times 0{,}3 = 24$$

$$\sigma = \sqrt{npq} = \sqrt{80 \cdot 0{,}3 \cdot 0{,}7} = 4{,}10$$

Se requiere calcular:

$$P(25 \leq x \leq 35) =$$

Ahora se va a aplicar la corrección por continuidad, que es de media unidad (0,5) hacia atrás y media unidad (0,5) hacia delante en el intervalo:

$$P(25-0{,}5 \leq x \leq 35+0{,}5) =$$

$$P(24{,}5 \leq x \leq 35{,}5) =$$

Tomando los valores de m = 24 y s = 4,10, se aplica el cálculo por la curva normal:

$$z_1 = \frac{24{,}5-24}{4{,}10} = 0{,}12$$

$$z_2 = \frac{35{,}5-24}{4{,}10} = 2{,}80$$

Entonces:

$$P(25 \leq x \leq 35) = P(24{,}5 \leq x \leq 35{,}5) = P(0{,}12 \leq z \leq 2{,}80)$$

Aplicando la tabla de la curva normal estándar (Apéndice 6):

$$= 0{,}9974 - 0{,}5478 = 0{,}4496$$

Si se utiliza Minitab para hacer el cálculo con la distribución binomial con valores de $n = 80$ y $p = 0{,}3$, se obtiene una probabilidad de 0,4419, lo cual indica que la aproximación por la normal tiene un resultado bastante cercano.

b) En este segundo caso se quiere calcular:

$$P(x \geq 20)$$

Al aplicar la corrección por continuidad se recomienda poner los valores en un intervalo, en este caso desde 20 hasta 80, ya que el tamaño de muestra es 80, y luego corregir:

$$P(x \geq 20) = P(19{,}5 \leq x \leq 80{,}5)$$

Luego se estandariza:

$$z_1 = \frac{19{,}5-24}{4{,}10} = -1{,}10$$

$$z_2 = \frac{80{,}5-24}{4{,}10} = 13{,}78$$

Calculando con la distribución normal:

$$P(x \geq 20) = P(19{,}5 \leq x \leq 80{,}5) = P(-1{,}10 \leq z \leq 13{,}78)$$

$$= P(z \leq 13{,}78) - P(z \leq -1{,}10) = 1{,}0000 - 0{,}1357 = 0{,}8643$$

c) En este caso éxito es que las empresas no deseen actualizar su sistema operativo, por lo que $p = 0{,}70$. Entonces:

$$\mu = np = 80 \times 0{,}7 = 56$$

$$\sigma = \sqrt{npq} = \sqrt{80 \cdot 0{,}7 \cdot 0{,}3} = 4{,}10$$

Luego se tiene que se busca:

$$P(x < 60)$$

Solución

Primero se expresa como el problema equivalente pero empleando el signo ≤ en vez de <, ya que menos de 60 es lo mismo que menor o igual a 59, esto para aplicar el mismo procedimiento visto anteriormente:

$$P(x < 60) = P(x \leq 59)$$

Ahora se convierte al intervalo, pues menor o igual que 59 equivale al intervalo de 0 a 59 y se aplica la corrección por continuidad:

$$P(x < 60) = P(x \leq 59) = P(0 \leq x \leq 59) = P(-0{,}5 \leq x \leq 59{,}5)$$

Ahora se estandariza:

$$z_1 = \frac{-0{,}5 - 56}{4{,}10} = -13{,}78$$

$$z_2 = \frac{59{,}5 - 56}{4{,}10} = 0{,}85$$

Planteando el problema completo y aplicando la tabla de normal estándar:

$$P(x < 60) = P(x \leq 59) = P(0 \leq x \leq 59) = P(-0{,}5 \leq x \leq 59{,}5)$$

$$= P(-13{,}78 \leq z \leq 0{,}85) = P(z \leq 0{,}85) - P(z \leq -13{,}78)$$

$$= 0{,}8023 - 0 = 0{,}8023$$

Aplicación

Garantía de los productos
Cuando usted adquiere un producto, como un televisor, un teléfono o una computadora, siempre espera que el fabricante ofrezca un plazo de garantía razonable. Posiblemente usted no compraría un aparato de estos si le ofrecen una garantía de un mes. Así, un producto con una garantía por un plazo amplio es visto como un producto de calidad, es un producto confiable.

Entonces, para la empresa fabricante la garantía es un factor para competir contra otras compañías. Ahora bien, las empresas ofrecen un plazo de garantía determinado cuando saben que el número de unidades que podrían fallar en dicho plazo es mínimo, pues cada falla significa un costo y significa que hubo algún problema a lo largo del proceso de producción.

Las empresas tratan de calcular intervalos de confianza de la tasa de reclamaciones por garantía y su varianza, además de estimar el posible efecto sobre sus costos y entonces poder pronosticar el impacto futuro de las reclamaciones por garantía. En el cálculo de estas estimaciones se requiere gran cantidad de datos y el uso de modelos estadísticos, entre ellos distribuciones de probabilidad, tales como la distribución de Poisson, la distribución normal, la distribución exponencial, la distribución lognormal y la distribución de Weibull, entre otras.

2.4 Distribución exponencial

A una oficina de un banco llegan, en promedio, 3 clientes por hora a solicitar un crédito. Se desea saber la probabilidad de que transcurran 30 minutos entre la llegada de un cliente y el siguiente. En un problema de este tipo, en el que se desea calcular la probabilidad de que se dé un determinado tiempo entre la ocurrencia de dos eventos sucesivos, muy similar a un problema en el que se pueda aplicar la distribución de Poisson, se puede resolver empleando la distribución exponencial.

Así, si un determinado evento ocurre en el mismo contexto de un proceso de Poisson, según se explicó anteriormente, entonces el intervalo de tiempo o espacio entre dos eventos sucesivos sigue la distribución exponencial. Así, la distribución exponencial se utiliza para calcular la probabilidad de ocurrencia de un cierto evento en una extensión de tiempo, en un área o en un volumen específico.

Un proceso de Poisson es estacionario, por lo que la probabilidad de ocurrencia del evento es igual a todo lo largo del espacio o periodo relevante, de modo que la distribución exponencial permite calcular la probabilidad de, por ejemplo:

- Que un evento ocurra dentro de un determinado plazo o espacio.
- Que transcurra determinado tiempo entre dos eventos sucesivos.
- Que transcurra determinado tiempo desde un determinado punto temporal hasta un primer evento.

Si l es el número promedio de ocurrencias en el plazo de interés, entonces la variable *x* está exponencialmente distribuida si su función de densidad es:

$$P(x) = \lambda e^{-\lambda x} \quad \text{para } x \geq \lambda \geq 0.$$

La probabilidad exponencial de que el primer evento ocurra dentro de un determinado intervalo temporal o especial es:

$$P(X \leq x) \quad 1 - e^{-\lambda x}$$

Por otro lado, la probabilidad exponencial de que el primer evento no ocurra dentro de un intervalo temporal o especial especificado es:

$$P(X > x) = e^{-\lambda x}$$

Ejemplo

A una oficina de un banco llegan, en promedio, 3 clientes por hora a solicitar un crédito. Se desea saber la probabilidad de:

a. Que transcurran 30 minutos entre la llegada de un cliente y el siguiente.

b. Que tras la salida de un cliente el próximo llegue en el curso de los 20 minutos siguientes.

Solución

Se tiene que $\lambda = 3$ clientes por hora:

a. Dado que la pregunta se refiere al periodo 30 minutos, entonces:

$$\lambda = 3/2 = 1{,}5$$

Dado que se busca la probabilidad de que el primer evento no ocurra dentro de un intervalo temporal dado, entonces se sustituye en la fórmula:

$$P = e^{-\lambda} = e^{-1{,}5} = 0{,}2231$$

b. En un lapso de 20 minutos, l = 1. Dado que se busca la probabilidad de que el primer evento ocurra dentro de un intervalo temporal dado, entonces se sustituye en la fórmula:

$$P = 1 - e^{-\lambda} = 1 - e^{-1} = 1 - 0{,}3679 = 0{,}6321$$

Ejemplo

El tiempo requerido para que ocurra una reacción química está exponencialmente distribuido con un tiempo esperado de 4 minutos:

a. ¿Qué proporción de la sustancia se formará dentro de dos minutos?
b. ¿Qué proporción de la sustancia se formará entre 3 y 8 minutos?

Solución

Se pueden emplear intervalos de un minuto, dado que la reacción se hace, en promedio, en 4 minutos, entonces el número esperado de sustancia formada en un minuto será $\lambda = 1/4 = 0{,}25$ (este sería la cantidad media de ocurrencias por minuto), entonces:

a. Dado que la pregunta se refiere al periodo 2 minutos, entonces $\lambda = 0{,}5$:

$$P = e^{-\lambda} = e^{-0{,}5} = 0{,}6065$$

b. Entre 3 y 8 minutos, se usa l = 0,75 y l = 2, respectivamente:

$$P(4 < x < 8) = e^{-0{,}75} - e^{-2} = 0{,}4724 - 0{,}1353 = 0{,}3370$$

Ejercicio de revisión

Se sabe que la vida útil de cierto tipo de bujías sigue una distribución exponencial con media de 160.000 km. ¿Cuál es la probabilidad de que una bujía seleccionada aleatoriamente dure:

a) a lo sumo 180.000 km?
b) entre 150.000 y 200.000 km?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para la distribución exponencial

Ejemplo

Los clientes de una tienda llegan en promedio de 20 por hora. Utilice Excel y Minitab para determinar la probabilidad de que transcurran a lo sumo 6 minutos después de la llegada del último cliente y el próximo.

Solución

En **Excel**: Se tiene que una media $\lambda = 20$ clientes por hora y se pregunta la probabilidad de que transcurran a lo sumo 6 minutos después de la llegada del último cliente y el próximo, por lo que $x = 0{,}1$, pues equivale a 6 minutos de una hora que tiene 60 minutos, o sea, $x = 6/60 = 0{,}1$.

Entonces, en **Excel** se emplea la función DISTR.EXP, cuya sintaxis es:

= DISTR.EXP(x;lambda;acum)

Los argumentos de la función anterior son:

x: número establecido de la variable (x)
lambda: promedio (λ)
acumulado: 0 si no es acumulado o 1 si es acumulado

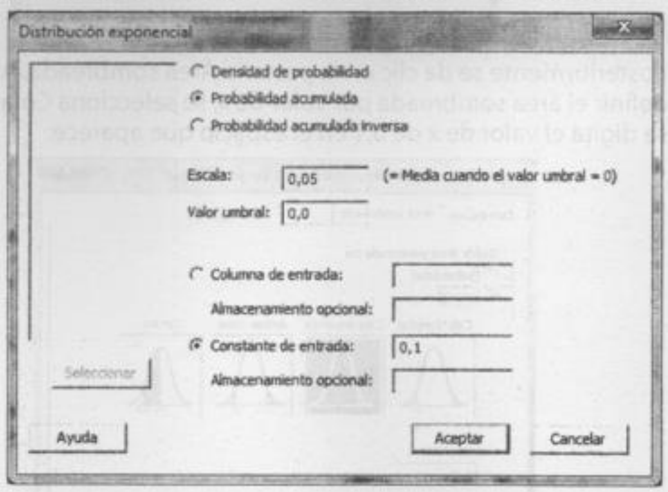
Así en este caso, se completa la función en la celda en la que se desea el resultado colocando el valor de x como 0,1, la media de 20 y (se indica al final 1 para que dé el resultado acumulado):

=DISTR.EXP(0,1;20;1)

El resultado es 0,8647.

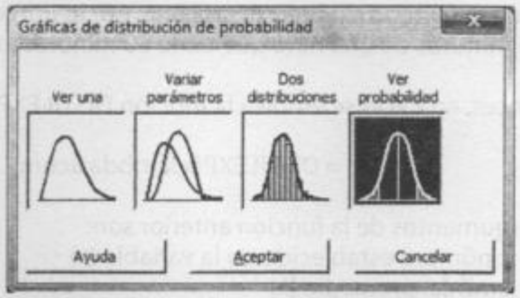
En **Minitab**: Se tiene que una media de 1/20, y se pregunta la probabilidad de que transcurran a lo sumo 6 minutos después de la llegada del último cliente y el próximo, por lo que $x = 0{,}1$, pues equivale a 6 minutos de una hora que tiene 60 minutos, o sea, $x = 6/60 = 0{,}1$.

Se da clic en el menú Calc, luego en Distribuciones de probabilidad, y ahí se elige Exponencial. Se completa el cuadro de diálogo:

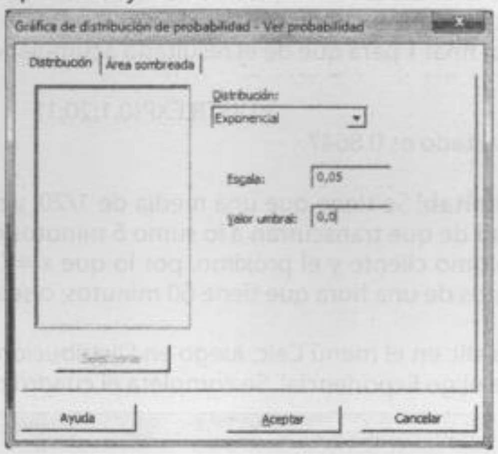


Se selecciona probabilidad acumulada y se completan los datos tal como se muestra en la imagen. Luego se da clic en Aceptar y se obtiene el resultado 0,8647 en la ventana Sesión.

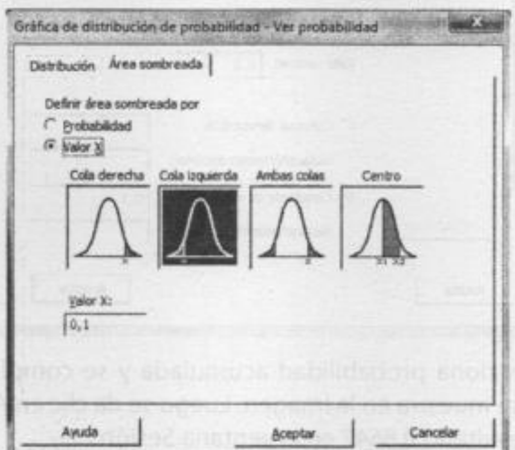
También, se puede hacer uso del menú Gráfica, donde se selecciona Gráfica de distribución de probabilidad. En el cuadro de diálogo se selecciona la opción que dice Ver probabilidad.



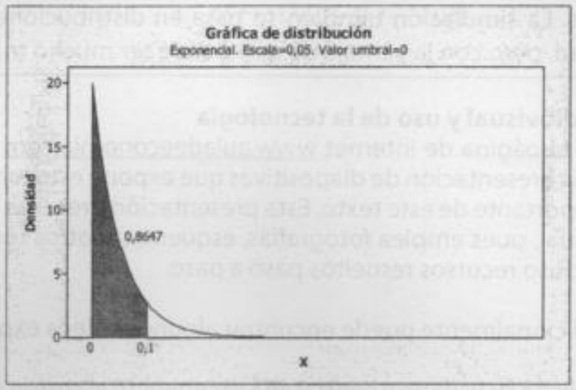
En el cuadro de diálogo se selecciona en la lista la distribución exponencial y se introduce el dato de la escala, que es la media, y el dato del valor umbral se puede dejar en cero:



Posteriormente se da clic en la pestaña Área sombreada. Aquí se elige definir el área sombreada por Valor de *x*, se selecciona Cola izquierda y se digita el valor de *x* de 0,1 en el espacio que aparece:



Al dar clic en Aceptar, Minitab crea un gráfico que indica el valor de la variable *x* en el eje horizontal:



El valor esperado y la varianza de una distribución exponencial de probabilidad, donde la variable es *x* y l corresponde a una unidad de tiempo o espacio, son:

$$E(x) = \frac{1}{\lambda}$$

$$\sigma^2 = \frac{1}{\lambda^2}$$

Aplicación

Líneas de espera

Todos conocemos la frase "el tiempo es oro". Para algunos negocios esa frase es clave para su competitividad, como en el caso de los restaurantes de comida rápida o los bancos. En la medida en la cual el cliente espere poco tiempo, sea en la fila o en el servicio mismo, esos clientes tendrán una razón más para continuar empleando sus servicios.

Los ingenieros han desarrollado modelos de teoría de colas o líneas de espera basados en distribuciones de probabilidades (como la distribución de Poisson o la distribución exponencial) que permiten determinar qué tan larga podría llegar a ser la cola o cuánto sería el tiempo promedio de espera, de manera que es posible diseñar un sistema en el que el cliente no espere demasiado y a un costo razonable para el proveedor del servicio. Por ejemplo, en el área de cajas, podría tomarse la decisión de que los clientes hagan filas independientes para cada caja (varios cajero cada uno con su propia cola), o si es deben formarse en una única fila y que luego pasen al cajero que se desocupe primero (una cola con varios servidores).

Una técnica alternativa a la teoría de colas es la simulación, la cual puede ser empleada para evaluar el desempeño del sistema de colas. La simulación también se basa en distribuciones de probabilidad, pero con la ventaja de que puede ser mucho más flexible.

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Otras distribuciones de variable continua

Existen muchas otras distribuciones de variable continua como la distribución uniforme, triangular, Gamma, *t*, Weibull, entre muchas otras. En el material impreso de este texto no se presentan estas otras distribuciones, pero en el material digital en la página de internet de este libro sí se desarrollan otras distribuciones de variable continua.

2.5 Ejercicios

Ejercicios de desarrollo

Conteste las preguntas que se formulan a continuación (las respuestas a los ejercicios se encuentran en la página de internet del texto:
www.auladeeconomia.com/raeep.html).

Distribución normal

1. Si se tiene que $m = 35$, $s = 5$, halle las siguientes probabilidades normales:
   a. $P(x \geq 28) =$
   b. $P(x \geq 40) =$
   c. $P(26 \leq x \leq 34) =$

2. Se sabe que las concentraciones de colesterol total en la sangre para cierta población se distribuyen normalmente con promedio 210 mg/100ml y desviación estándar de 18 mg/100ml. Con base en estos datos, conteste
   a. ¿Cuál es la probabilidad de que al seleccionar una de estas personas al azar tenga una concentración de colesterol entre 175 y 220 mg/100ml?
   b. Si hay 25000 personas en esa población, ¿cuántos de ellos se espera que tengan los niveles de colesterol superiores a 230 mg/100 ml?

3. Suponga que el nivel de glucosa en la sangre de una población se distribuye normalmente con media de 101 mg/dl con desviación estándar de 29 mg/dl, entonces calcule la probabilidad de que al seleccionar una de esas personas al azar, tenga un nivel de glucosa en la sangre:
   a. Inferior a 90 mg/dl
   b. Superior a 108 mg/dl
   c. Entre 95 y 120 mg/dl

4. Un laboratorio farmacéutico prepara pastillas circulares. Se sabe que el diámetro se distribuye normalmente con media es 9 mm con una desviación estándar de 0,5 mm. Si se sabe que las pastillas deben satisfacer un diámetro mínimo de 8 mm y máximo de 10,2 mm, calcule la probabilidad de que una pastilla seleccionada al azar:
   a. Tenga un diámetro fuera de la especificación
   b. ¿A partir de qué diámetro se encuentra el 10% de las pastillas con mayor diámetro?
   c. ¿Qué porcentaje de las pastillas tiene un diámetro en un rango de más/menos una vez la desviación estándar con respecto a la media?

5. En un estudio sobre las alpacas se encontró que los adultos machos tenían, en promedio, un peso vivo de 64,78 Kg., con una desviación estándar de 12,87 Kg. Con base en estos datos determine la probabilidad de que un macho adulto seleccionado tenga un peso:
   a. Superior a 71 Kg.
   b. Inferior a 56 Kg.
   c. Entre 60 y 70 Kg.
   d. ¿Cuál es el peso mínimo que marca el límite del 35% de los machos adultos de mayor peso?
   e. ¿Cuál es el peso máximo que marca el límite del 15% de los machos adultos de menor peso?

6. Una empresa fabrica pantallas LCD para cámaras digitales. Las pantallas deben tener una dimensión de 76 mm. En promedio las pantallas fabricadas miden 75,95 mm con una desviación estándar de 0,08 mm distribuidas normalmente. Con base en estos datos, calcule la probabilidad de obtener al azar una pantalla con dimensión:
   a. Mayor que 76,1 mm.
   b. Menor que 75,0 mm.
   c. Entre 75,2 y 76,9 mm.
   d. ¿Cuál es la dimensión del 10% de las pantallas de mayor dimensión?

7. En una determinada localidad se va a construir un puente. Luego del análisis de distintos escenarios en los que se efectuaría cada una de las etapas del proyecto, desde la delimitación y preparación de la zona, la construcción de los apoyos, montaje de la estructura, entre otras, hasta su conclusión se tomaría un tiempo esperado de 39 semanas con una desviación estándar de 7 semanas. Calcule la probabilidad de que el proyecto:
   a. ¿Se comprometería usted a entregar el proyecto en 42 semanas?
   b. ¿Cuánto tiempo propondría usted para entregar el proyecto de modo que se tenga un 95% de probabilidad de cumplir con ese plazo estipulado?

8. Suponga que el proceso de empacado de frijoles tiene una media de 896 gramos por bolsa con una desviación estándar de 12 gramos. Se considera que los pesos se distribuyen normalmente. Con base en los datos anteriores determine la probabilidad de que al seleccionar aleatoriamente una bolsa:
   a. Tenga un peso superior a 1 kilogramo.
   b. Tenga un peso inferior a 890 gramos.
   c. Tenga un peso entre 0,90 y 0,95 kilogramos.
   d. Si se decide volver a llenar la cuarta parte de las bolsas con menor peso, ¿a partir de qué peso, en gramos, se deberán volver a llenar estas bolsas?

9. Un analista ha observado las comisiones que ganan los corredores de bolsa durante los últimos años y se ha dado cuenta que siguen una distribución de probabilidad normal. El monto anual medio percibido es de \$30000, y la desviación estándar de \$7000.
   a. ¿Qué porcentaje de los corredores de bolsa percibe entre \$27500 y \$40000 al año?
   b. Si hay 200 corredores de bolsa, ¿cuánto de ellos se espera que tengan comisiones superiores a \$42000?
   c. Si se va a otorgar un incentivo al 10% de los corredores que logran mayores comisiones, ¿a partir de cuál monto de comisiones se ofrecerá este incentivo?

10. El departamento de recursos humanos de una empresa ha desarrollado un novedoso sistema de entrenamiento. Luego de probarlo por unas semanas se ha observado que el tiempo para completar el entrenamiento se comporta normalmente y que los empleados lo concluyen en un promedio de 265 horas con una desviación estándar de 55 horas. Encuentre la probabilidad de que un empleado seleccionado al azar termine el programa:
    a. En menos de 285 horas.
    b. En menos de 240 horas.
    c. En más de 230 horas.
    d. En más de 300 horas.
    e. Entre 270 y 310 horas.
    f. Entre 200 y 260 horas.
    g. Entre 250 y 290 horas.
    h. ¿Cuál es el tiempo que separa el 20% de los empleados que duran más en la terminación del programa?
    i. ¿Cuál es el tiempo que separa el 15% de los empleados que duran menos en la terminación del programa?

11. Un viaje en autobús tiene una duración distribuida normalmente con media de 23 minutos y desviación estándar 3,4 minutos. Todos los días se requiere llegar al destino a las 9.00 a.m.
    a. ¿A qué hora debe iniciarse para tener una probabilidad del 95% de estar a tiempo?
    b. Si se inicia a las 8.30 a.m., ¿Cuál es la probabilidad de que no llegue a tiempo?
    c. Si se ha decidido iniciar a las 8.30 todos los días, ¿Cuántas veces se llegará a tiempo en un periodo de un año (suponga que se realiza 360 veces en ese periodo)?

12. Un hotel registra datos sobre el número de días de estancia de cada uno de sus huéspedes. Esta variable se distribuye normalmente con media 6 y desviación estándar 3,1. Según una proyección se esperan 150 huéspedes. Con base en estos datos, calcule cuántos huéspedes se espera que se hospeden:
    a. Menos de 4 días?
    b. Más de 8 días?
    c. Entre 5 y 9 días?
    d. ¿Cuál es el tiempo mínimo que se prolonga la estancia del 75% de los huéspedes que se alojan más tiempo en el hotel?

13. Las puntuaciones de una prueba para medir el nivel de inglés de personas de habla hispana se distribuyen normalmente con media de 520 puntos y desviación estándar de 110 puntos. Calcule la probabilidad de obtener una nota:
    a. Menor de 630 puntos. Primero indique cuál debería ser el resultado (empleando la regla empírica) y luego efectúe los cálculos para comprobar el resultado.
    b. De por lo menos 740 puntos.
    c. De más de 850 puntos.
    d. Como máximo 345 puntos.
    e. De más de 475 puntos.
    f. De entre 450 y 800 puntos.
    g. ¿Cuántas personas, de un grupo de 400, se esperaría que obtengan notas entre 420 y 735 puntos?
    h. ¿Cuál fue la calificación obtenida por el 10% de los que salieron con mejor nivel de inglés?
    i. ¿Cuál fue la calificación máxima obtenida por el 20% de los de más bajo nivel de inglés?

14. Algunas personas desarrollan adicción al trabajo, lo cual les resta calidad de vida y afecta sus relaciones familiares y sociales, e incluso afecta su salud. Un equipo especializado ha elaborado una prueba para determinar si una persona es adicta al trabajo. La prueba mide la adicción al trabajo en una escala y si una persona obtiene entre 120 y 150 puntos está en riesgo de ser adicto al trabajo, y si obtiene más de 150 puntos, entonces es adicto al trabajo. Luego de aplicar la prueba a varios miles de personas, se encontró que las puntuaciones se distribuyen normalmente con una media de 105 puntos y desviación estándar de 28 puntos. Calcule la probabilidad de que una persona seleccionada al azar:
    a. Esté en riesgo de ser adicto al trabajo.
    b. Sea adicto al trabajo.
    c. No sea adicto ni esté en riesgo de ser adicto al trabajo.
    d. ¿A partir de cuál puntaje se encuentra el 10% de las personas con mayor nivel de adicción al trabajo?

15. Si el número de horas que los funcionarios de una empresa dedican a iniciativas de voluntariado se distribuye normalmente, ¿cuál es la probabilidad de que un funcionario seleccionado al azar dedique una cantidad de hora que se ubique en un rango de dos veces la desviación estándar (hacia arriba o hacia abajo) con respecto a la media?

16. Una institución realizó un plan de capacitación entre mujeres emprendedoras de una zona marginal del país. Se logró determinar que el ingreso de estas mujeres se distribuye normalmente con una media de $450 al mes. Si el percentil 70 de ingreso es de $650, ¿cuál es la desviación estándar del ingreso de estas mujeres?

17. En un estudio se determinó que los gastos en publicidad de las universidades privadas del país se distribuye normalmente con una desviación estándar de $400 mil al año. Si el monto máximo del gasto en publicidad del 30% de las empresas que menos gastan en publicidad es $100 mil dólares al año, calcule el gasto anual promedio de estas grandes compañías.

18. En una investigación sobre las iniciativas de voluntariado que realizan los empleados de distintas empresas del país involucradas en programas sociales se estimó que el tiempo que cada funcionario dedica a estas actividades se distribuye normalmente con media de 40 horas/año y desviación estándar de 22 horas/año. Si en una empresa participan en estos programas 50 colaboradores:
    a. Calcule la probabilidad de que un funcionario seleccionado al azar participe menos de 30 horas/año.
    b. ¿Cuántos empleados se esperaría que participen más de 60 horas/año?
    c. ¿Cuál es el número de horas mínimo correspondiente al 15% de los funcionarios que más tiempo participan?

19. Suponga que el proceso de empacado de detergente en polvo tiene una media de 996 gramos por bolsa con una desviación estándar de 42 gramos. Se considera que los pesos se distribuyen normalmente. Con base en los datos anteriores determine la probabilidad de que al seleccionar aleatoriamente una bolsa:

    a. Tenga un peso superior a 1,1 kilogramos.
    b. Tenga un peso superior a 900 gramos.
    c. Tenga un peso inferior a 1000 gramos.
    d. Tenga un peso inferior a 920 gramos.
    e. Tenga un peso entre 900 y 950 gramos.
    f. Si se decide volver a llenar el 10% de las bolsas con menor peso, ¿a partir de qué peso, en gramos, se deberán volver a llenar estas bolsas?

20. Si los diámetros del cable coaxial RG-174 fabricado por una empresa se distribuyen normalmente con media 2,55 mm. y desviación estándar 0,8 mm., determinar el porcentaje de cables con diámetros:

    a. entre 2,5 y 2,6 mm.
    b. menores o iguales a 2,44 mm.
    c. mayores o iguales a 2,62 mm.
    d. ¿Cuál debe ser el tamaño mínimo que debe tener el 10% de los cables con el mayor diámetro?

21. Los diámetros interiores de las arandelas hechas por cierto fabricante se distribuyen normalmente con una media de 11,2 mm y una desviación estándar de 0,15 mm.

    a. ¿Cuántas arandelas de un lote de 1500 tienen un diámetro interno mayor a 11,5 mm?
    b. ¿Cuál es la probabilidad de que una arandela tenga un diámetro inferior a 11,4 mm?
    c. ¿Cuál es la probabilidad de que una arandela tenga un diámetro entre 11,05 y 11,59 mm?
    d. ¿Cuál es el diámetro mínimo del 70% de las arandelas con mayor diámetro interno?
    e. ¿Cuál es el valor del diámetro sobre el que está el 15% de las arandelas de menor diámetro interno?
    f. ¿Cuál es el valor del diámetro bajo el que está el 25% de las arandelas?
    g. ¿Cuál es el diámetro máximo del 82,5% de las arandelas?

22. Los rendimientos de un portafolio de inversiones se distribuyen normalmente con una media anual de 4,5% con una desviación estándar de 1,7%.

    a. ¿Cuál es el rendimiento porcentual que está por debajo del 95% de los rendimientos del portafolio?
    b. ¿Cuál es la probabilidad de tener pérdidas con este portafolio?

23. Los resultados de una prueba para determinar el cociente intelectual se distribuyen normalmente con una media de 101 puntos y desviación estándar de 10 puntos. Si se selecciona al azar una persona, determine la probabilidad de que obtenga en la prueba:
   a. Más de 120 puntos.
   b. Menos de 85 puntos.
   c. Entre 90 y 110 puntos.
   d. ¿Cuál es el puntaje mínimo del 6% de las personas con mayor cociente intelectual según este test?

24. Una empresa que emisora de tarjetas de crédito sabe, por experiencia, que el saldo mensual promedio de los clientes es de \$258 con una desviación estándar de \$136. Calcule la probabilidad de que el saldo mensual sea:
   a. Más de \$220.
   b. Menos de \$300.
   c. Entre \$188 y \$227.
   d. ¿Cuánto será el monto máximo del 70% de los saldos más bajos?
   e. ¿Sobre qué valor se encuentra el 23% de los saldos?

Distribución exponencial

25. Como administrador de un restaurante de comidas rápidas sabe que, en promedio, llega un cliente cada 2,5 minutos y desea conocer la probabilidad exponencial de que pasen más de 50 segundos entre la llegada de dos clientes.

26. Un fabricante de discos duros indica que los discos que fabrica tiene una vida media de 1 400 000 horas. Si esta vida útil varía exponencialmente, determine la probabilidad de que uno de estos discos duros seleccionados al azar tenga una vida útil:
   a. Menos a 1 000 000 de horas.
   b. Inferior a 10 años funcionando 24 horas al día.
   c. Si la esperanza de vida en un país es de 78 años, ¿cuál es la probabilidad de que uno de estos discos duros dure más que la vida esperada de una persona?

27. Se sabe que el tiempo de espera de una persona que llama a un centro de llamadas de una empresa de soporte tecnológico es una variable aleatoria exponencial con media 4,5 minutos. Encuentre la probabilidad de que un cliente que llame en un momento seleccionado al azar tenga que esperar:
   a. A lo sumo 6 minutos.
   b. Como máximo 3 minutos.
   c. Entre 2 y 5 minutos.

28. Se sabe que un cierto tipo de motor debería tener una vida útil distribuida exponencialmente de 400000 kilómetros. Con base en estos datos determine:

    a. La probabilidad de que un motor seleccionado al azar tenga una vida de a lo sumo la mitad de lo esperado.

    b. Si se tiene una flota de 200 vehículos empleando este tipo de motor, ¿cuántos se esperaría que se que tenga una vida superior a 500 000 km?

29. Los tiempos de espera en la fila de los clientes de un banco se distribuyen exponencialmente con una media de 15 minutos. Calcule la probabilidad de que un cliente seleccionado al azar espere:

    a. Más de 20 minutos.

    b. Entre 17 y 23 minutos.

    c. Menos de 25 minutos.

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Si se sabe que *b* es una variable normal estándar, ¿Cuál es la probabilidad de que *b* sea mayor que 2,5?

   (a) 0,0000 (b) 0,9938
   (c) 0,0062 (d) Falta información

2. Si se sabe que *b* es una variable normal estándar, ¿Cuál es el valor de que *b* que se ubica en el percentil 19? Usando la tabla de la distribución normal estándar acumulada:

   (a) 0,88 (b) 0,7910
   (c) −0,88 (d) Ninguna de las anteriores

3. Si se sabe que *x* es una variable normal con media 12 y varianza 9, ¿Cuál es la probabilidad de que *x* sea menor que 10? Usando la tabla de la distribución normal estándar acumulada:

   (a) 0,4121 (b) 0,7486
   (c) 0,2514 (d) Ninguna de las anteriores

4. Si se sabe que *x* es una variable normal con media 12 y varianza 9, ¿Cuál es la probabilidad de que *x* sea como mínimo igual a 8? Usando la tabla de la distribución normal estándar acumulada:

   (a) 0,6716 (b) 0,0918
   (c) 0,9082 (d) Ninguna de las anteriores

5. Si se sabe que *x* es una variable normal, ¿Cuál es la probabilidad de que *x* tome valores en un intervalo de ±2 veces la desviación estándar con respecto a la media?

(a) 0,997 (b) 0,683
(c) 0,954 (d) Falta información

6. Si se sabe que *x* es una variable normal con media 12 y varianza 9, ¿Cuál es el valor de *x* que separa el 15% superior de los valores posibles de *x*?

(a) 0,85 (b) 8,891
(c) 15,11 (d) Ninguna de las anteriores

7. La distribución exponencial permite obtener respuesta a preguntas relacionadas con la probabilidad de que un evento ocurra en determinado plazo, el tiempo entre dos eventos sucesivos o el tiempo que transcurre desde un determinado punto temporal hasta un primer evento, se llama:

(a) Exponencial (b) Normal
(c) Geométrica (d) Poisson

8. Una empresa ha comprado un equipo para su producción que requiere una pieza especial. Según el proveedor esa pieza especial posee una vida esperada de 8 meses. Si usted desea saber la probabilidad de tener que reemplazar esta pieza antes de 6 meses, debería emplear la distribución:

(a) Binomial (b) Normal
(c) Multinomial (d) Geométrica

9. La tasa media de llegadas de clientes a un restaurante de comida rápida es de 4 por minuto. Si usted desea saber la probabilidad de que en una hora seleccionada aleatoriamente lleguen menos de 200 clientes, entonces usaría:

(a) Binomial (b) Exponencial
(c) Poisson (d) Normal

10. La tasa media de llegadas de clientes a un restaurante de comida rápida es de 4 por minuto. Si usted desea saber la probabilidad de que pasen 30 segundos entre la llegada de dos clientes en una hora seleccionada aleatoriamente, entonces usaría:

(a) Binomial (b) Exponencial
(c) Poisson (d) Normal

11. Una fábrica de cemento empaca su producto en sacos que tienen una media de 51,9 kilogramos, con una desviación estándar de 350 gramos, de acuerdo con una distribución normal. La especificación es que cada saco pese exactamente 52 kilogramos. La probabilidad de que un saco seleccionado al azar tenga un exceso en el peso de un kilogramo o más con respecto al peso especificado es:

(a) 0,0008 (b) 0,9992
(c) 0,0021 (d) Ninguna de las anteriores

12. Tomando los mismos datos del problema 11, la probabilidad de que un saco seleccionado al azar tenga un peso en un rango de ±2 veces la desviación estándar con respecto al promedio es:

(a) 0,9540 (b) 0,6830
(c) 0,9970 (d) Ninguna de las anteriores

13. Tomando los mismos datos del problema 11, la probabilidad de que un saco seleccionado al azar tenga un peso entre 50 y 52 kilogramos es:

(a) 0,6125 (b) 0,3875
(c) 0,9999 (d) Ninguna de las anteriores

14. Tomando los mismos datos del problema 11, la probabilidad de que un saco seleccionado al azar tenga un peso inferior a 51,1 kilogramos es:

(a) 0,4991 (b) 0,5009
(c) 0,0111 (d) Ninguna de las anteriores

15. Tomando los mismos datos del problema 11, en un lote de 200 sacos, el número de sacos que se esperaría que tengan un peso superior a 52,5 kilogramos es:

(a) 12,5 (b) 8,64
(c) 0,0432 (d) Ninguna de las anteriores

16. Tomando los mismos datos del problema 11, si se considera que si un saco tiene un peso en el 10% inferior debe reprocesarse, entonces el valor que marca el peso en kilogramos a partir del cual los sacos deben reprocesarse es:

(a) 51,95 (b) 51,45
(c) 0,5235 (d) Ninguna de las anteriores

17. Tomando los mismos datos del problema 11, el valor que marca el percentil 85 de los pesos, en gramos, es:

(a) 52,26 (b) 52260
(c) 0,5154 (d) Ninguna de las anteriores

18. En una distribución exponencial:

A. La media de la distribución equivale a uno entre lambda
B. La varianza de la distribución equivale a uno entre lambda al cuadrado

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Son verdaderas ambas (b) Solo B es verdadera
(c) Son falsas ambas (d) Solo A es verdadera

19. ¿Cuál de las siguientes es falsa acerca de los datos que sigue la distribución normal?

(a) El promedio es el mismo que el modo
(b) La desviación estándar es la misma que la media
(c) La mediana es el mismo que el modo
(d) La mayoría de los datos está dentro de 3 desviaciones estándar de la mediana

20. ¿Cuál de las siguientes no es cierto acerca de la distribución normal?

(a) la media, la mediana y la moda son iguales
(b) la curva es sesgada a la derecha
(c) la curva nunca toca el eje x
(d) el área bajo la curva es uno

21. En los problemas binomiales en que $n > 20$ y $p > 5\%$, es falso que:

(a) se puede emplear la normal para aproximar la binomial
(b) los resultados de la binomial y la normal son iguales
(c) los resultados de la binomial y la normal convergen con forme n tiende a infinito
(d) efectuar el cálculo usando la distribución binomial lleva más trabajo

Capítulo 3

Estimación por intervalos

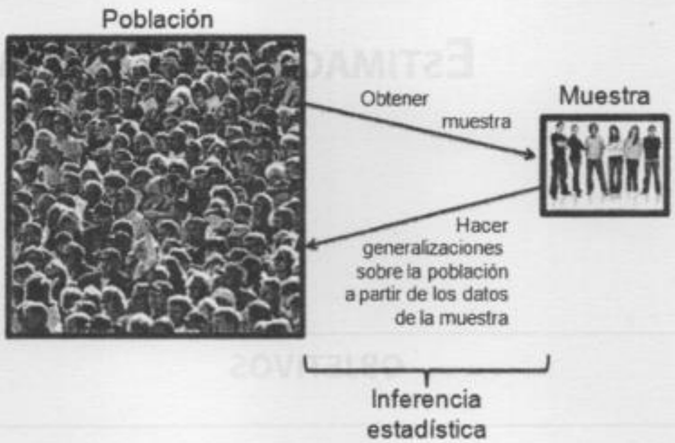
OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Explicar el concepto de inferencia estadística.
2. Explicar el teorema del límite central.
3. Calcular intervalos de confianza para la media poblacional.
4. Calcular intervalos de confianza para la proporción poblacional.

3.1 Inferencia estadística

El objetivo de una investigación estadística por muestreo es poder realizar inferencias acerca de la población pero obtenidas a partir de la información contenida en una muestra tomada aleatoriamente. Por ejemplo, si en una investigación se desea conocer la ingesta promedio de calorías en una determinada población, entonces se toma una muestra representativa y se espera que esos datos muestrales permitan efectuar conclusiones relacionadas con toda la población, pues el objetivo es conocer la población, no la muestra.



Las poblaciones son "representadas" por descripciones numéricas llamadas *parámetros*, por ejemplo, la media (μ) o la desviación estándar (σ), así, la inferencia estadística consiste en hacer inferencias sobre parámetros a partir de *estadísticos* como $\overline{x}$ y s, los cuales son calculados con base en las observaciones de las muestras. Para efectos de este texto, estas inferencias se relacionarán principalmente con la estimación por intervalos de una media poblacional y de una proporción poblacional, las cuales se tratarán en este capítulo, y en un capítulo posterior se trabajará lo relacionado con las pruebas de hipótesis.

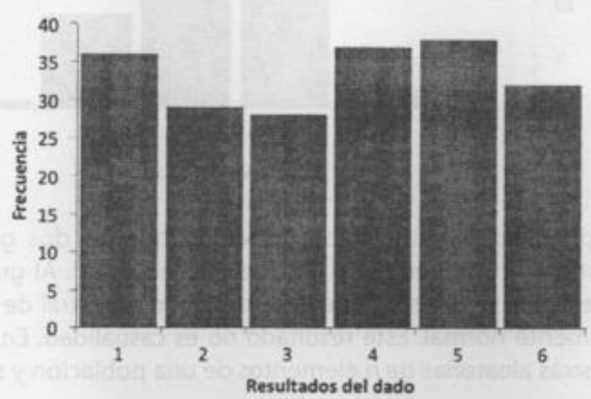
3.2 Distribución muestral de la medida si σ es conocida

Antes de empezar a exponer propiamente cómo se realiza una estimación por intervalos, es necesario presentar algunos aspectos que son fundamentales para poder comprender en qué se basa la teoría que sustenta la determinación de los intervalos de confianza.

Suponga que se realizan 4 lanzamientos seguidos de un dado. En cada caso se van a apuntar los 4 resultados. En la tabla se presentan los resultados de simular 50 series de 4 lanzamientos del dado. La simulación se efectuó usando la función de Excel =ALEATORIO.ENTRE(1;6) y los promedios se:

	# lanzamiento del dado					# lanzamiento del dado			
Serie	**1**	**2**	**3**	**4**	**Serie**	**1**	**2**	**3**	**4**
1	1	5	3	6	26	5	2	1	6
2	4	5	5	2	27	2	5	6	1
3	4	5	5	2	28	1	3	4	6
4	4	5	3	6	29	3	1	6	6
5	4	4	6	2	30	4	5	3	1
6	2	4	2	4	31	4	6	6	1
7	2	2	6	5	32	5	1	5	1
8	3	5	1	4	33	6	4	3	4
9	5	3	1	1	34	5	3	2	3
10	2	2	6	2	35	3	6	1	4
11	2	3	3	2	36	4	1	2	4
12	2	6	5	6	37	4	5	3	2
13	4	5	3	1	38	1	6	4	4
14	4	6	1	5	39	5	1	5	1
15	1	5	3	5	40	4	1	1	6
16	3	1	2	4	41	6	6	5	5
17	4	2	5	2	42	1	6	2	2
18	6	1	1	6	43	3	1	5	4
19	5	4	4	6	44	3	5	4	1
20	4	6	6	4	45	5	4	4	5
21	2	1	2	3	46	5	1	6	4
22	1	2	6	4	47	4	4	6	3
23	2	1	1	3	48	6	5	3	3
24	5	5	3	3	49	2	1	6	1
25	2	4	5	5	50	3	1	5	3

Si se toman los 200 lanzamientos en total (50 series de 4 lanzamientos cada una) y se construye una gráfica representando los resultados, es de esperar que todas las barras sean de una altura similar, porque cada uno de los 6 posibles resultados tiene la misma probabilidad de aparecer (probabilidad que en cada caso es 1/6), pero con pequeñas diferencias debidas al azar, tal como se muestra en el siguiente histograma:



Ahora, para cada una de las 50 serie se va a obtener la media, es decir, en cada caso se suman los 4 valores y se divide entre 4 para cada una de las series de 4 lanzamientos. Esto va a dar una lista de 50 promedios:

Serie	$\overline{x}$
1	3,75
2	4,00
3	4,00
4	4,50
5	4,00
6	3,00
7	3,75
8	3,25
9	2,50
10	3,00
11	2,50
12	4,75
13	3,25

Serie	$\overline{x}$
14	4,00
15	3,50
16	2,50
17	3,25
18	3,50
19	4,75
20	5,00
21	2,00
22	3,25
23	1,75
24	4,00
25	4,00
26	3,50

Serie	$\overline{x}$
27	3,50
28	3,50
29	4,00
30	3,25
31	4,25
32	3,00
33	4,25
34	3,25
35	3,50
36	2,75
37	3,50
38	3,75
39	3,00

Serie	$\overline{x}$
40	3,00
41	5,50
42	2,75
43	3,25
44	3,25
45	4,50
46	4,00
47	4,25
48	4,25
49	2,50
50	3,00

Observe que cada serie tiene una media distinta, a pesar de que se origina del mismo proceso aleatorio. Esas diferencias se deben al azar. Ahora, en vez de emplear los datos de los 200 lanzamientos, se tomarán las 50 medias obtenidas en cada serie, y se volverá a construir la gráfica, la cual corresponde al histograma siguiente.

En principio se pudo haber esperado que las dos gráficas debieran tener forma similar, pero es claro que esto no es así. Al graficar los promedios muestrales se observa que la distribución muestral de las medias es aproximadamente normal. Este resultado no es casualidad. En general, si se toman muestras aleatorias de n elementos de una población y se calculan los promedios $\overline{x}_i$ de cada una de las muestras, es de esperar que los valores de las medias no sean todos iguales, sino que presentan diferencias debidas al azar, pero que estas medias tienden a distribuirse según una curva normal, independientemente de la forma de la población original, o sea, que no importa si los datos originales siguen o no una distribución normal. Todo esto conduce al siguiente teorema:

Teorema: Si se extrae una muestra aleatoria de tamaño *n* de una población que tiene media μ y varianza finita σ^2 entonces *x* es el valor de una variable aleatoria con media μ y desviación estándar:

Para muestras tomadas de poblaciones infinitas:

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

Para muestras tomadas de poblaciones finitas de tamaño *N*:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

A la desviación estándar de las medias muestrales se le conoce como error estándar de la media y se denota por $\sigma_{\bar{x}}$.

ERROR ESTÁNDAR DE LA MEDIA Desviación estándar de las medias muestrales. Se denota por $\sigma_{\bar{x}}$.

Se puede demostrar que para *N* grande en comparación al tamaño *n* de la muestra el factor de corrección para muestras $\frac{N-n}{N-1}$ es aproximadamente igual a 1 por lo que las dos fórmulas del error estándar darían prácticamente el mismo valor.

Ejemplo

El nivel de glucosa en la sangre de una cierta población compuesta por 5000 miembros tiene una desviación estándar de 29 mg/dl. Se toma una muestra de 40 personas, ¿cuál es el factor de corrección y el error estándar de la media?

Solución

El factor de corrección es:

$$\frac{N-n}{N-1} = \frac{5000-40}{5000-1} = 0{,}9922$$

Si la desviación estándar σ = 29, entonces el error estándar de la media es:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{29}{\sqrt{40}} = 4{,}59$$

y aplicando el factor de corrección:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{29}{\sqrt{40}} \cdot \sqrt{\frac{5000-40}{5000-1}} = 4{,}57$$

Ejercicio de revisión

La prueba de admisión de una universidad tiene una desviación estándar de 250 puntos. Si se toma una muestra de 60 estudiantes que han aplicado la prueba, ¿cuál es el error estándar?

¿Cómo cambia el resultado anterior si se sabe que un total de 6000 estudiantes han realizado la prueba?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Material audiovisual

En la página de internet de este texto podrá encontrar una presentación y un video que exponen el cálculo de intervalos de confianza.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Puede verse en el ejemplo que la diferencia entre el error estándar con el factor de corrección y el primero que se calculó que no lo incorporaba es mínima, ya que en ese caso el factor de corrección es muy cercano a uno. Si el tamaño de la muestra es más grande o si la población es más pequeña, entonces el factor de corrección marcará una diferencia más importante. Por ejemplo, con una población más pequeña de tamaño 500 (y con tamaño de muestra 40), el factor de corrección es:

$$\frac{N-n}{N-1} = \frac{500-40}{500-1} = 0{,}9218$$

y entonces el error estándar sería:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{29}{\sqrt{40}} \cdot \sqrt{\frac{500-40}{500-1}} = 4{,}40$$

3.3 Teorema del límite central

Con base en los conceptos anteriores, puede enunciarse un importante teorema para la inferencia estadística, conocido como el teorema del límite central:

Teorema: Si $\overline{x}$ es la media de una muestra aleatoria de tamaño n extraída de una población que tiene media μ y varianza finita σ^2, entonces dicha media muestral $\overline{x}$ tendrá una distribución normal con media μ y varianza σ^2/n conforme aumenta el tamaño de la muestra n.

En otras palabras se tendrá que z, dada por: $z = \dfrac{x - \mu}{\sigma / \sqrt{n}}$

es el valor de una variable aleatoria cuya función de distribución se aproxima a la distribución normal estándar cuando n tiende a ∞. Esto se cumple independientemente de la forma de la población, es decir, la distribución de x es aproximadamente normal con media μ y varianza σ^2/n cuando n es grande.

3.4 Distribución muestral de la media con σ desconocida (distribución *t* de Student)

Ahora bien, muchas veces la desviación estándar poblacional σ no es conocida, entonces no se cumple el teorema anterior, pero si es posible emplear el teorema siguiente:

Teorema: Si $\overline{x}$ es la media de una muestra aleatoria de tamaño n extraída de una población normal que tiene media que tiene media μ y varianza σ^2, entonces:

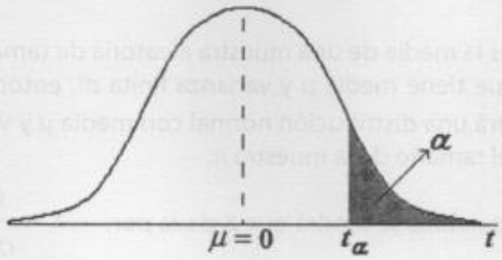
$$t = \frac{x - \mu}{s / \sqrt{n}}$$

es el valor de una variable aleatoria con distribución t y $gl = n - 1$ grados de libertad.

Para usar la distribución t no se requiere conocer σ, pero se debe poder suponer razonablemente que la población de donde se tomó la muestra es normal.

La distribución t tiene forma de campana y es simétrica con respecto al origen, igual que la distribución normal, e, igual que la distribución normal estándar, tiene media $\mu = 0$, sin embargo, su desviación estándar depende de los grados de libertad. A medida que la muestra es más grande la distribución t se aproxima a la normal estándar, es decir, σ tiende a 1 cuando n tiende a ∞ y, por lo tanto, gl tiende a ∞. Se considera que la distribución normal estándar es una buena aproximación a la distribución t para muestras mayores o iguales a 30.

Los valores de t se encuentran en la tabla de la distribución t en el Apéndice 7 de este texto, y pueden corresponder a las probabilidades de una cola o de dos colas. Por ejemplo, para el caso de la cola derecha se tendría gráficamente:



La tabla que aparece en el Apéndice 7 se muestra del modo siguiente (aquí aparece solo una parte de la tabla para ilustrar su uso):

Distribución T de Student

	Nivel de significancia para pruebas de una cola					
	0,1	0,05	0,025	0,01	0,005	0,0005
	Nivel de significancia para pruebas de dos colas					
gl	0,2	0,1	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,599
3	1,638	2,353	3,182	4,541	5,841	12,924
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,408
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965

Los siguientes son ejemplos de valores de t tomados de la tabla t (Apéndice 7). Si se tiene que el valor $\alpha = 0{,}01$, y que $n = 2$, entonces $gl = n - 1 = 1$, por lo que se busca el $\alpha = 0{,}01$ en los niveles de significancia para pruebas de una cola, o sea, en la cuarta columna, y en el renglón de $gl = 1$, así que se tendría que $t = 31{,}821$. Si se buscara en el valor $\alpha = 0{,}01$ para pruebas de dos colas, entonces se busca en la quinta columna, y en el renglón de $gl = 1$, así que se tendría que $t = 63{,}657$. Luego, en este capítulo y los siguientes, se indicará cuándo emplear los valores para la prueba de una cola y cuándo los de la prueba de dos colas.

Si $\alpha = 0{,}01$ con $n = 12$, entonces $gl = n - 1 = 11$, entonces se busca el $\alpha = 0{,}01$ en los niveles de significancia para pruebas de una cola, o sea, en la cuarta columna, y en el renglón de $gl = 11$, así que se tendría que $t = 2{,}718$. Otros ejemplos son:

Si $n = 22$, entonces $gl = n - 1 = 21$ y $t = 2{,}518$
Si $n = 30$, entonces $gl = n - 1 = 29$ y $t = 2{,}462$, también $z = 2{,}325$

Claramente se observa que conforme n crece, y por lo tanto gl también crece, el valor de t se aproxima al de z en la curva normal estándar.

3.5 Inferencia para la media

La inferencia para la media aritmética se realiza principalmente mediante intervalos de confianza y mediante pruebas de hipótesis. En este capítulo se tratará solamente lo correspondiente a los intervalos de confianza, pero en un capítulo posterior se trabajará el tema de las pruebas de hipótesis.

3.6 Intervalos de confianza para la media y error máximo de la estimación

Cuando se usa una media muestral $\bar{x}$ para estimar el valor de la media poblacional μ no se puede esperar que los dos valores sean iguales. La probabilidad de que el estimador (la media muestral) y el parámetro (la media poblacional) sean diferentes es muy alta. Esa diferencia entre estos dos valores se le conoce como error de la estimación.

ERROR DE ESTIMACIÓN Diferencia, en valor absoluto, entre el valor de la media muestral y la media poblacional.

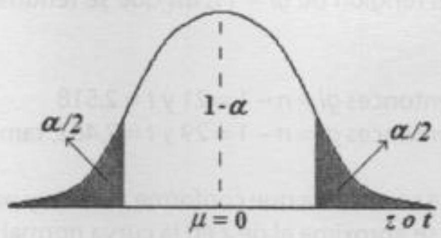
En términos de los símbolos de cada uno de los valores el error de estimación se expresaría:

ERROR DE ESTIMACIÓN $E = |\bar{x} - \mu|$

Cuando se utiliza la media muestral $\bar{x}$ para construir un intervalo de confianza para estimar la media poblacional μ, la probabilidad de que esta estimación no falle es como máximo de $1 - \alpha$. Esta probabilidad se le conoce como nivel de *confianza*.

CONFIANZA Probabilidad de que la estimación por intervalos de la media poblacional μ no falle.

Gráficamente se puede expresar, en términos de la distribución normal o la distribución t, como sigue:



Según el teorema del límite central el valor de z en la distribución normal es:

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

Solo interesan los valores de Z que están entre en el intervalo de confianza $-Z_{\alpha/2}$ y $Z_{\alpha/2}$:

$$-Z_{\alpha/2} \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \le Z_{\alpha/2}$$

Si se toman los valores en los extremos de las desigualdades y el valor absoluto de $\bar{x} - \mu$ (que es el error de estimación) al sustituir en las desigualdades de arriba se tiene:

$$Z_{\alpha/2} = \frac{|\bar{x} - \mu|}{\sigma / \sqrt{n}} = \frac{E}{\sigma / \sqrt{n}}$$

Despejando el valor de E se obtiene el error máximo de la estimación de μ por $\bar{x}$:

ERROR MÁXIMO DE ESTIMACIÓN

$$E = Z \cdot \frac{\sigma}{\sqrt{n}} \text{ para } n \ge 30$$

$$E = Z \cdot \frac{\sigma}{\sqrt{n}} \text{ para } n < 30$$

Retomando la expresión anterior:

$$-Z_{\alpha/2} \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \le Z_{\alpha/2}$$

se obtienen las fórmulas para el intervalo de confianza. Se toma primero sólo el lado izquierdo de la desigualdad:

$$-Z_{\alpha/2} \le \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

y se despeja:

$$-Z_{\alpha/2} \cdot \sigma / \sqrt{n} \le \bar{x} - \mu$$

$$\mu \le \bar{x} + Z_{\alpha/2} \cdot \sigma / \sqrt{n}$$

Despejando el lado derecho se obtiene:

$$\mu \leq \bar{x} - Z_{\alpha/2} \cdot \sigma / \sqrt{n}$$

Y juntando los lados se obtiene en intervalo de confianza para la media poblacional μ:

$$\bar{x} - Z_{\alpha/2} \cdot \sigma / \sqrt{n} \leq \mu \leq \bar{x} + Z_{\alpha/2} \cdot \sigma / \sqrt{n}$$

La fórmula resumida para el intervalo de confianza es: $\bar{x} \pm E$

A modo de resumen, los límites de estimación o intervalos de confianza corresponden a:

INTERVALOS DE CONFIANZA:

Cuando $n \geq 30$ y σ conocida: $\bar{x} \pm Z \cdot \sigma / \sqrt{n}$

Cuando $n < 30$ y σ conocida: $\bar{x} \pm Z \cdot \sigma / \sqrt{n}$

Cuando $n \geq 30$ y σ desconocida: $\bar{x} \pm Z \cdot s / \sqrt{n}$

Cuando $n < 30$ y σ desconocida: $\bar{x} \pm t \cdot s / \sqrt{n}$
si se puede suponer que la población de donde se tomó la muestra es normal

Cuando $n < 30$ y σ desconocida: $\bar{x} \pm k \cdot s / \sqrt{n}$
si la población no es normal

En el último caso la confianza del intervalo es $1 - 1/k^2$ (se aplica el teorema de Chebyshev).

Intervalos de confianza cuando n ≥ 30 y σ conocida

Tal como se mostró en el cuadro anterior, si el tamaño de muestra es mayor o igual a 30 y se conoce la desviación estándar poblacional, entonces al calcular los intervalos de confianza se emplea la distribución normal, o sea, un valor de *z* y el valor conocido de la desviación estándar.

INTERVALO DE CONFIANZA:

Cuando $n \geq 30$ y σ conocida: $\bar{x} \pm Z \cdot \sigma / \sqrt{n}$

Ejemplo

Durante una semana se toma una muestra aleatoria de 50 empleados de una empresa, y se obtiene un salario promedio de $206. Se conoce que la desviación estándar poblacional de $40.

Determine los intervalos de confianza del 95% para la media de los salarios de esta empresa.

Solución

Se tiene que $n = 50$, $\overline{x} = \$206$, $\sigma = 40$ y una confianza $1 - \alpha = 0{,}95$.

Dado que la confianza es: $1 - \alpha = 0{,}95$, entonces $\alpha = 0{,}05$, o sea, que se tendría $\alpha/2 = 0{,}025$, por lo que $1 - \alpha/2 = 1 - 0{,}025 = 0{,}975$. Como $n \geq 30$ y σ conocida, se debe usar z. De la tabla de la distribución normal estándar z con $\alpha/2$ equivale a $z = 1{,}96$.

Luego se sustituye en la fórmula del intervalo de confianza cuando $n \geq 30$ y σ conocida:

$$\bar{x} \pm Z \cdot \sigma / \sqrt{n}$$

$$= 206 \pm 1{,}96 \cdot 40 / \sqrt{50}$$

Para obtener el límite inferior se resta:

$$L_i = 206 - 1{,}96 \cdot 40 / \sqrt{50} = 194{,}91$$

Y para obtener el límite superior se suma:

$$L_s = 206 + 1{,}96 \cdot 40 / \sqrt{50} = 217{,}09$$

En conclusión, se tiene una confianza de que la media de los salarios de esta empresa se encuentra entre \$194,91 y \$217,09.

Ejercicio de revisión

En una muestra de 50 hectáreas tomadas al azar de diferentes fincas productoras de papa se rendimiento promedio de 40 toneladas por hectárea al emplear un cierto tipo de abono orgánico. Se conoce, por un estudio previo, que la desviación estándar poblacional es de 8 toneladas/ha. Obtenga un intervalo de confianza del 95% para la media del rendimiento de papa por hectárea.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para calcular intervalos de confianza

Ejemplo

Utilice Excel y Minitab para resolver el problema: Durante una semana se toma una muestra aleatoria de 50 empleados de una empresa, y se obtiene un salario promedio de \$206. Se conoce que la desviación estándar poblacional de \$40.

Determine los intervalos de confianza del 95% para la media de los salarios de esta empresa.

Solución
Se tiene que $n = 50$, $\overline{x}$ = \$206, $\sigma = 40$ y una confianza: $1 - \alpha = 0{,}95$.

En **Excel** se emplea la función INTERVALO.CONFIANZA, la cual da el error máximo de estimación, o sea, el resultado de calcular $z \cdot \sigma / \sqrt{n}$, por lo que luego es necesario tomar el promedio obtenido en la muestra y restar y sumar el valor dado por la función para obtener los límites de confianza inferior y superior, respectivamente. La función tiene la siguiente sintaxis:

=INTERVALO.CONFIANZA(alfa;desv_estándar;tamaño)

Los argumentos de la función anterior son:
alfa: es el valor a dado
desv_estándar: es la desviación estándar (s o *s*)
tamaño: es el tamaño de muestra (*n*)

Luego se sustituyen los valores:

=INTERVALO.CONFIANZA(0,05;40;50)

Así se obtiene el valor 11,09. Para obtener el límite inferior se resta:

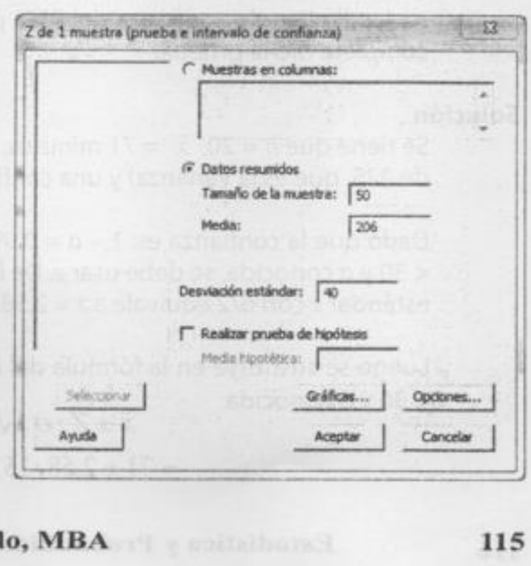
= 206 - 11,09 = 194,91

Y para obtener el límite superior se suma:

= 206 + 11,09 = 217,09

En conclusión, se tiene una confianza de que la media de los salarios de esta empresa se encuentra entre \$194,91 y \$217,09.

En **Minitab** se da clic en el menú Estadística, se elige Estadística básica y luego se selecciona Z de 1 muestra. Ahí se completa el cuadro de diálogo siguiente:



Se marca la opción de datos resumidos y se completan los datos tal como se muestra en la imagen. En el botón Opciones se indica el nivel de confianza, que en este caso es 95%. La opción Hipótesis alterna debe dejarse como "no es igual a". Luego se da clic en Aceptar y el resultado se obtiene en la ventana Sesión:

```
Sesión

Z de una muestra

La desviación estándar supuesta = 40

                   Error
                estándar
                   de la
 N   Media         media        IC de 95%
50  206,00          5,66  (194,91. 217,09)
```

En la salida en la ventana Sesión se observa IC de 95%, que corresponde al intervalo de confianza del 95%, y que este es 194,91 a 217,09.

Intervalos de confianza cuando n < 30 y σ conocida

Cuando el tamaño de muestra es menor que 30 y se conoce la desviación estándar poblacional, entonces al calcular los intervalos de confianza se emplea la distribución normal, o sea, un valor de z y el valor conocido de la desviación estándar.

INTERVALO DE CONFIANZA:

Cuando $n < 30$ y σ conocida: $\bar{x} \pm Z \cdot \sigma / \sqrt{n}$

Ejemplo

Se sabe que el tiempo que toma completar una prueba psicométrica tiene una varianza de 225 minutos. Una muestra de 20 estudiantes es sometida a la prueba obteniéndose una media de 71 minutos. Obtenga los límites de confianza del 99% para el tiempo medio en que se completa dicha prueba.

Solución

Se tiene que $n = 20$, $\bar{x} = 71$ minutos, $\sigma = 15$ minutos (la raíz cuadrada de 225, que es la varianza) y una confianza: $1 - \alpha = 0{,}99$.

Dado que la confianza es: $1 - \alpha = 0{,}99$, entonces $\alpha/2 = 0{,}005$. Como $n < 30$ y σ conocida, se debe usar z. De la tabla de la distribución normal estándar z con $\alpha/2$ equivale a $z = 2{,}58$.

Luego se sustituye en la fórmula del intervalo de confianza cuando $n < 30$ y σ conocida:

$$\bar{x} \pm Z \cdot \sigma / \sqrt{n}$$
$$= 71 \pm 2{,}58 \cdot 15 / \sqrt{20}$$

Para obtener el límite inferior se resta:

$$L_i = 71 - 2{,}58 \cdot 15 / \sqrt{20} = 62{,}36$$

Y para obtener el límite superior se suma:

$$L_s = 71 + 2{,}58 \cdot 15 / \sqrt{20} = 79{,}64$$

En conclusión, se tiene una confianza de el tiempo de terminación de la prueba se encuentra entre 62,36 y 79,64 minutos.

Ejercicio de revisión

Se desea estimar el consumo promedio de leche de los habitantes de un pueblo rural. En una muestra de 15 pobladores se obtuvo un consumo medio por día de 288 ml y una desviación estándar de 52 ml. Determine los intervalos de confianza del 90% para el verdadero promedio del consumo diario de leche de esta población.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Material audiovisual

En la página de internet de este texto podrá encontrar un video que expone el uso de Minitab para el cálculo de intervalos de confianza.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Intervalos de confianza cuando n ≥ 30 y σ desconocida

Cuando el tamaño de muestra es mayor o igual a 30 y no se conoce la desviación estándar poblacional, sino que solo se conoce la desviación estándar muestral, entonces al calcular los intervalos de confianza se emplea la distribución normal (por el teorema del límite central), o sea, se utiliza un valor de *z* y el valor de la desviación estándar muestral.

INTERVALO DE CONFIANZA:

Cuando $n \geq 30$ y σ desconocida: $\bar{x} \pm Z \cdot s / \sqrt{n}$

Ejemplo

En una muestra de 42 personas que se han sometido a un trasplante de corazón se ha obtenido un tiempo medio de sobrevivencia (en años) de 5,25 años con una desviación estándar muestral de 1,75 años. Hallar un intervalo de confianza del 95 por ciento para el promedio de vida de todas las personas que se han sometido a un trasplante de corazón.

Solución

Se tiene que $n = 42$ personas, $\overline{x} = 5{,}25$ años, $s = 1{,}75$ años y una confianza: $1 - \alpha = 0{,}95$.

Dado que la confianza es: $1 - \alpha = 0{,}95$, entonces $\alpha/2 = 0{,}025$. Como n^{3} 30 y σ desconocida se debe usar z. De la tabla de la distribución normal estándar z con $\alpha/2$ equivale a $z = 1{,}96$.

Luego se sustituye en la fórmula del intervalo de confianza cuando n^{3} 30 y σ desconocida:

$$\bar{x} \pm z \cdot s / \sqrt{n}$$
$$= 5{,}25 \pm 1{,}96 \cdot 1{,}75 / \sqrt{42}$$

Para obtener el límite inferior se resta:

$$L_i = 5{,}25 \pm 1{,}96 \cdot 1{,}75 / \sqrt{42} = 4{,}72$$

Y para obtener el límite superior se suma:

$$L_s = 5{,}25 \pm 1{,}96 \cdot 1{,}75 / \sqrt{42} = 5{,}78$$

En conclusión, se tiene una confianza de que el promedio de vida de todas las personas que se han sometido a un trasplante de corazón se encuentra entre 4,72 y 5,78 años.

Ejercicio de revisión

Una empresa productora de harina de trigo empaca paquetes que deben contener un kilogramo de producto. En una muestra de 60 paquetes se obtuvo un peso medio de 992 gramos y una desviación estándar muestral de 44 gramos. Calcule los intervalos de confianza del 98% para el peso medio de los paquetes de harina.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Intervalos de confianza cuando n < 30 y σ desconocida

Cuando el tamaño de muestra es menor o igual a 30 y no se conoce la desviación estándar poblacional, sino que solo se dispone de la desviación estándar muestral, entonces al calcular los intervalos de confianza no se emplea la distribución normal, sino que se utiliza la distribución t (con grados de libertad $gl = n - 1$) y el valor conocido de la desviación estándar. En estos casos se supone que la variable que se estima se distribuye normalmente.

INTERVALO DE CONFIANZA:

Cuando $n < 30$ y σ desconocida: $\bar{x} \pm t \cdot s / \sqrt{n}$

Cuando $n < 30$ y σ desconocida se puede emplear $\bar{x} \pm k \cdot s / \sqrt{n}$ si la población no es normal, donde la confianza del intervalo es $1 - 1/k^2$, aplicando el teorema de Chebyshev.

Ejemplo

El ciclo medio de vida de una muestra aleatoria de 12 focos es de 2000 horas, con una desviación estándar muestral de 200 horas. Se supone que la vida media de los focos se distribuye normalmente. Determine los intervalos de confianza del 95% para la vida media de los focos.

Solución

Se tiene que $n = 10$, $\bar{x} = 2000$, $s = 200$ y una confianza: $1 - \alpha = 0{,}95$.

Dado que la confianza es: $1 - \alpha = 0{,}95$, entonces $\alpha/2 = 0{,}025$. Como $n < 30$ y σ desconocida, se debe usar t. Después se busca en la tabla de la distribución t de Student, con una significancia de 0,05, con dos colas y grados de libertad $gl = n - 1 = 10 - 1 = 9$, el valor de t equivale a $t = 2{,}262$.

Luego se sustituye en la fórmula del intervalo de confianza cuando $n \leq 30$ y σ desconocida:

$$\bar{x} \pm t \cdot s / \sqrt{n}$$
$$= 2000 \pm 2{,}262 \cdot 200 / \sqrt{10}$$

Para obtener el límite inferior se resta:

$$L = 2000 - 2{,}262 \cdot 200 / \sqrt{8} = 1856{,}94$$

Y para obtener el límite superior se suma:

$$L_s = 2000 + 2{,}262 \cdot 200 / \sqrt{8} = 2143{,}06$$

En conclusión, se tiene una confianza de 95% de que la vida media de los focos se encuentra entre 1856,94 y 2143,06 horas.

Ejercicio de revisión

En una encuesta aplicada a 25 personas residentes de la ciudad capital se encontró que, por semana, dedicaban un promedio de 4,8 horas a la lectura, tanto de libros, revistas, periódicos y otros materiales. Se conoce que la desviación estándar poblacional es de 3,5 horas/semana. Determine los intervalos de confianza del 99% para el número de horas promedio que las personas dedican a la lectura.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Grados de libertad

Los grados de libertad corresponden a una medida del número de observaciones independientes entre los elementos de una muestra, es decir, es el número de datos que se podrían cambiar de modo que, dado un total fijo, se puede obtener ese total. Por ejemplo, la media aritmética tiene n - 1 grados de libertad, ya que, si tuviéramos una muestra de 10 datos y conocemos el valor de la media, entonces podríamos modificar 9 datos, o sea, n - 1 datos, y el décimo quedaría determinado. Si se tuviera una tabla de 5 filas y 4 columnas, o sea, 5 x 4, y se conocen los totales de cada fila y columna, entonces en cada fila se podría modificar 3 datos (y el quinto quedaría determinado por la diferencia de los 4 primeros con el total) y en cada columna se podrían cambiar 4 datos, de modo que los grados de libertad de esta tabla 5 x 4 serían (5 - 1) x (4 - 1) = 4 x 3 = 12 grados de libertad.

Uso de Minitab para calcular intervalos de confianza usando la distribución t

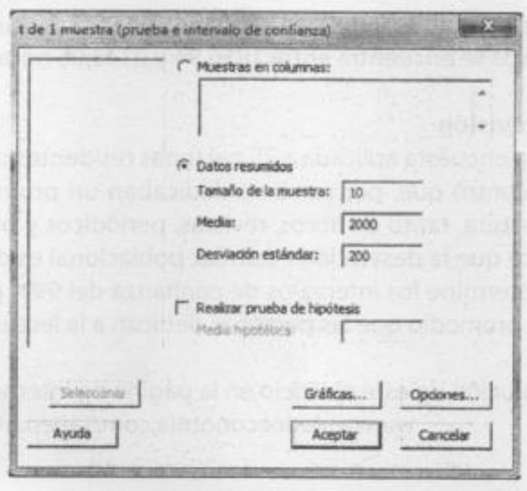
Ejemplo

Utilice Minitab para resolver el problema: El ciclo medio de vida de una muestra aleatoria de 12 focos es de 2000 horas, con una desviación estándar muestral de 200 horas. Se supone que la vida media de los focos se distribuye normalmente. Determine los intervalos de confianza del 95% para la vida media de los focos.

Solución

Se tiene que $n = 10$, $\overline{x} = 2000$, $s = 200$ y una confianza: $1 - \alpha = 0{,}95$.

En **Minitab** se da clic en el menú Estadística, se elige Estadística básica y luego se selecciona t de 1 muestra. Ahí se completa el cuadro de diálogo siguiente:



Se marca la opción de datos resumidos y se completan los datos tal como se muestra en la imagen. En el botón Opciones se indica el nivel de confianza, que en este caso es 95%. La opción Hipótesis alterna debe dejarse como "no es igual a". Luego se da clic en Aceptar y el resultado se obtiene en la ventana Sesión:

```
Sesión

T de una muestra

                          Error
                        estándar
                          de la
 N    Media  Desv.Est.    media      IC de 95%
10   2000,0      200,0     63,2   (1856,9. 2143,1)
```

En la salida en la ventana Sesión se observa IC de 95%, que corresponde al intervalo de confianza del 95%, y que este es 1856,9 a 2143,1.

Ejemplo

Se sabe que 20 fusibles que fueron sometidos a una sobrecarga del 20% se fundieron en un tiempo promedio de 10,63 minutos, con desviación estándar de 2,48 minutos.

a) Si se utiliza $\overline{x} = 10{,}63$ como estimación puntual de la media de tiempo poblacional, ¿de cuánto es el error máximo si se desea con una confianza del 95%?

b) Determine un intervalo de confianza del 95% para el promedio verdadero del tiempo de fusión.

Solución

Se tiene que $n = 20$, $\overline{x} = 10{,}63$ minutos, $s = 2{,}48$ minutos y una confianza: $1 - \alpha = 0{,}95$.

a) Dado que la confianza es: $1 - \alpha = 0{,}95$, entonces $\alpha/2 = 0{,}025$. Como $n < 30$ y σ desconocida se debe usar t en vez de Z, con $gl = 20 - 1 = 19$. De la tabla t con $\alpha/2 = 0{,}025$ y $gl = 19$, se obtiene $t_{\alpha/2} = 2{,}093$.

De ahí el error: $E = t_{\alpha/2} \cdot s/\sqrt{n} = 2{,}093 \times 2{,}48/\sqrt{20} = 1{,}16$ minutos. Se puede afirmar con una confianza del 95% que la media de la muestra se aparta de la media poblacional a lo sumo en 1,16 minutos.

b) Si se supone que la población de donde se tomó la muestra es normal, el intervalo de confianza está dado por $\overline{x} \pm t_{\alpha/2} \cdot s/\sqrt{n}$ porque $n < 30$ y σ desconocida.

En la parte (a) ya se obtuvo el valor del error de la estimación por lo que el intervalo es 1.16, por lo que los intervalos estarán dados por $10{,}63 \pm 1{,}16$, de donde se obtiene $10{,}63 - 1.16 = 9{,}47$ y $10{,}63 + 1{,}16 = 11{,}79$. Así se puede concluir que:

$$P(9{,}47 < \mu < 11{,}79) = 0{,}95$$

Es decir, 95 de cada 100 promedios calculados con muestras de 20 elementos tendrán un valor de entre 9,47 y 11,79 minutos.

¿Qué pasaría si se supiera que la población no es normal? Entonces se aplican los intervalos dados por $\bar{x} \pm k \cdot s / \sqrt{n}$, donde la confianza está dada por $1 - 1/k^2$. Es decir:

$$1 - 1/k^2 = 0,95$$

Despejando k:

$$1/k^2 = 0,05$$
$$1/0,05 = 20 = k^2$$
$$k = 4,472$$

Calculando los límites:

$$\bar{x} \pm k \cdot s / \sqrt{n} = 10,63 \pm 4,472 \cdot 2,48 / \sqrt{20} = 10,63 \pm 2,48 = \begin{cases} 8,15 \\ 13,11 \end{cases}$$

Entonces:

$$P(8,15 < \mu < 13,11) = 0,95$$

3.7 Inferencia para proporciones

Para realizar la estimación de una proporción poblacional se cuenta con la proporción muestral, $p = \frac{x}{n}$, donde x es el número de veces que ha ocurrido un evento en n pruebas. Por ejemplo, si en un lote de 3000 piezas salen 30 defectuosas, $p = \frac{x}{n} = \frac{30}{3000} = 0,01$ es la proporción muestral que podrá usarse como estimador del valor poblacional. Este valor puede darse también en forma porcentual, así es que se puede afirmar que el porcentaje de defectuosos es de 1%. También, si el valor se repite en gran cantidad de muestras, se puede decir que la probabilidad de defectuosos en el proceso es de 0,01.

Al trabajar con proporciones se supondrá siempre que las situaciones satisfacen las condiciones de la distribución binomial, por lo tanto, la distribución de muestreo en la que se basarán los métodos es la binomial con $\mu = np$ y $\sigma = \sqrt{npq}$. Se sabe además que cuando $n > 20$ y $p > 5\%$ la binomial se puede aproximar por la distribución normal con:

$$z = \frac{\bar{x} - \mu}{\sigma} = \frac{\bar{x} - nP}{\sqrt{nPQ}} = \frac{p - P}{\sqrt{PQ/n}}$$

El valor $p - P$ es la diferencia entre la proporción muestral p, y la poblacional P. Si se toma en valor absoluto se tiene el error de la estimación de la proporción $E = |p - P|$.

En los intervalos de confianza interesan los valores de Z que estén en $-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}$, los cuales se obtienen por medio de:

$$p \pm z \sqrt{pq/n}$$

Debido a que no se tienen la proporción poblacional, en todas estas fórmulas se usa el valor de *p* y *q* de la muestra.

Ejemplo

Se toma una muestra de 500 varones adultos y se encuentra que 156 son fumadores. Encuentre los límites de confianza del 99% para la proporción de fumadores varones.

Solución

Se tiene que $x = 156$ fumadores de una muestra de $n = 500$ varones adultos, así que la proporción muestral *p* sería:

$$p = x/n = 156/500 = 0{,}312$$

por lo que $q = 1 - p = 1 - 0{,}312 = 0{,}688$.

La confianza del 99%, es decir, $1 - \alpha = 0{,}99$, $\alpha = 0{,}01$, $\alpha/2 = 0{,}005$, así que de la tabla se obtiene $z = 2{,}58$, según la distribución normal.

Calculando el intervalo con $p = 0{,}312$, $q = 0{,}688$, $z = 2{,}58$ y $n = 500$:

$$p \pm z \sqrt{pq/n} = 0{,}312 \pm 2{,}58 \sqrt{0{,}312 \cdot 0{,}688/500} = \begin{cases} 0{,}2586 \\ 0{,}3653 \end{cases}$$

Se tiene una confianza del 99% de que la proporción de fumadores está entre 25,86% y 36,53%.

Ejemplo

El departamento de ventas de una empresa sostiene que se entregan en la fecha fijada con el cliente el 95% de los pedidos. Si al revisar las fechas de entrega de 200 órdenes se encontró que 184 fueron entregadas a tiempo, con los datos de la muestra encuéntrese un intervalo del 95% de confianza para la proporción verdadera de pedidos entregados a tiempo. Debe señalarse el error de la estimación.

Solución

Se pide el intervalo para la proporción poblacional $p \pm z \sqrt{pq/n}$ con una confianza del 95%, es decir, $1 - \alpha = 0{,}95$, $\alpha = 0{,}05$, $\alpha/2 = 0{,}025$, así que de la tabla se obtiene $z = 1{,}96$, según la distribución normal.

Además se tiene que $x = 184$ entregas a tiempo de una muestra de $n = 200$ entregas, así que la proporción muestral *p* sería:

$$p = x/n = 184/200 = 0{,}92$$

por lo que $q = 1 - p = 1 - 0{,}92 = 0{,}08$.

Calculando el intervalo con $p = 0{,}92$, $q = 0{,}08$, $z = 1{,}96$ y $n = 200$:

$$p \pm z\ \sqrt{pq/n} = 0{,}92 \pm 1{,}96\ \sqrt{0{,}92 \cdot 0{,}08/200} = 0{,}92 \pm 0{,}038$$

El error es 0,038 y el intervalo queda:

$$\mathrm{P}(0{,}882 \le p \le 0{,}958) = 0{,}95$$

Es decir, se tiene una confianza de 95% de que la proporción de pedidos entregados a tiempo se encuentra entre 88,2% y 95,8%.

Ejercicio de revisión

Una empresa desea lanzar un nuevo servicio por internet al mercado y para ello requiere conocer la proporción de hogares de la zona que posee acceso a internet. En una muestra 120 hogares, 70 indicaron que poseían algún tipo de conexión a la red. Determine los intervalos de confianza del 99% para la proporción de hogares de la zona con acceso a internet.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

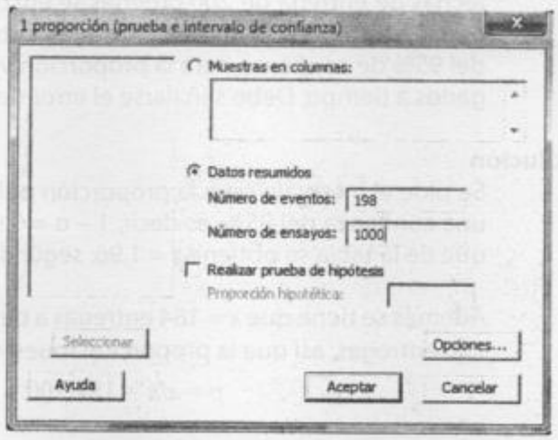
Uso de Minitab para calcular intervalos de confianza para proporciones

Ejemplo

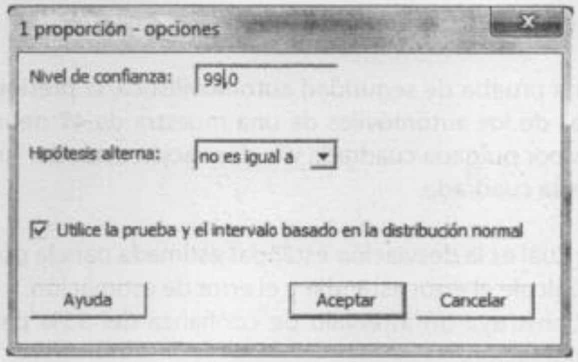
En una muestra de 1000 adultos y se encuentra que 198 estarán de acuerdo con la despenalización de la marihuana. Encuentre los límites de confianza del 99% para la proporción de adultos que apoyarían la despenalización de la marihuana.

Solución

Se tiene que $x = 198$ eventos de una muestra de $n = 1000$ adultos, así que en el menú Estadísticas / Estadística básica / 1 Proporción se completa el cuadro, seleccionando la opción Datos resumidos con 198 eventos y 1000 ensayos:



Luego en el botón Opciones se indica el nivel de confianza del 99% y se debe marcar la opción que dice Utilice la prueba y el intervalo basado en la distribución normal:



El resultado se obtiene en la ventana Sesión. Se concluye que se tiene una confianza del 99% de que la proporción de adultos que está de acuerdo con la despenalización de la marihuana está entre 16,55% y 23,05%.

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@ auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

3.8 Ejercicios

Ejercicios de desarrollo

Resuelva los ejercicios siguientes (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. En un estudio se logró determinar que el aclaramiento sistémico promedio del diclofenaco en plasma es de 263 ml/min con una desviación estándar de 56 ml/min. Se empleó un tamaño de muestra de 350 observaciones. Obtenga los intervalos de confianza del 90% para aclaramiento sistémico promedio.

2. En una muestra de 250 metros de cable coaxial RG-174 fabricado por una empresa se obtuvo un diámetro medio de 2,55 mm. Se sabe que la desviación estándar es 0,5 mm. Determine los intervalos de confianza del 99% para el diámetro promedio.

3. En una prueba de seguridad automovilística, la presión promedio en las llantas de los automóviles de una muestra de 47 neumáticos fue de 28 libras por pulgada cuadrada y la desviación estándar fue de 2,7 libras por pulgada cuadrada.

   a. ¿Cuál es la desviación estándar estimada para la población?
   b. Calcule el error estándar y el error de estimación.
   c. Construya un intervalo de confianza del 95% para la media de la población.

4. En una feria de empleo se tomó una muestra al azar de 40 aplicantes. Se les aplicó una prueba para determinar su nivel de inglés y se obtuvo que 32 tenían un nivel de 600 puntos o más y un promedio de 680 puntos. Se sabe que la prueba tiene una desviación estándar de 150 puntos. Con base en estos datos:

   a. Determine los intervalos de confianza del 95% para el nivel medio de inglés de las personas que buscaban empleo en esa la feria.
   b. Si el nivel de 600 puntos se considera como aceptable para las empresas que buscan un alto dominio del inglés, determine los intervalos de confianza del 99% para la proporción de personas que tienen un elevado dominio del idioma inglés.

5. En una nuestra de 40 mujeres se determinó que su pulso cardíaco promedio era de 76,3 latidos por minuto. Se conoce que la desviación estándar poblacional es de 12,5 latidos por minuto. Con base en los datos anteriores, determine:

   a. Calcule el error estándar del pulso cardíaco promedio (en latidos por minuto).
   b. Calcule los intervalos de confianza del 95% para el pulso cardíaco promedio de las mujeres.

6. El gerente de una empresa procesadora de café está preocupado porque sus proveedores le entregan con frecuencia café verde. Se tomaron como muestras 36 probetas de 250 ml cada una y encontró que en promedio se entregó 30 ml de grano verde por probeta con una desviación de 10 ml. Calcule el intervalo del 98% para la cantidad de grano verde por probeta.

7. Un fabricante de hornos de microondas quiere saber la proporción de familias que preferirían usar su marca. Toman al azar una muestra de 75 amas de casa y 30 de ellas respondieron que la usarían. Calcule un intervalo de confianza del 95% para la verdadera proporción de amas de casa que preferirían dicha marca de lavadora.

8. Según un estudio en 25 familias de estratos socio económicos medios y altos, los niños inician alguna práctica de cuidado de su salud buco dental a los 15,6 meses. Se conoce que la desviación estándar es 8,5 meses. Calcule los intervalos de confianza del 99% para la edad de inicio de las prácticas cuidado buco dental de los niños de familias de estratos socio económicos medios y altos.

9. Un arquitecto considera que actualmente las casas tienen un área construida inferior a la de hace dos o más décadas. En una muestra de 15 viviendas construidas en el transcurso de los dos últimos años se encontró un área construida promedio de 92 metros cuadrados y una desviación estándar de 30 metros cuadrados.

   a. Obtenga los intervalos de confianza del 95% para el área construida de las viviendas nuevas.
   b. Si se sabe, por otro estudio, que las viviendas construidas hace dos décadas o más tenían un área construida promedio de 130 metros cuadrados, ¿podría considerarse, con base en el intervalo calculado, que efectivamente el área de las viviendas es menor actualmente?

10. En una muestra de 12 motores para automóvil de cierto tipo se obtuvo una vida útil promedio de 300000 kilómetros. La desviación estándar muestral es de 60000 kilómetros. Obtenga los intervalos de confianza del 90% para la vida útil promedio de estos motores.

11. Una compañía ha desarrollado un nuevo motor de gasolina. Para valorar su consumo de combustible efectúa 15 recorridos y el motor consumió en promedio 8,7 galones de gasolina por minuto con una desviación estándar de 1,2 galones. ¿Qué se podría afirmar, con un 99% de confianza, acerca del tamaño máximo del error de estimación?

12. Un jefe de producción de una maquiladora desea estimar en forma rápida el tiempo medio que requieren las 520 costureras de la planta para realizar cierta tarea. Con ese fin selecciona una muestra de 15 de ellas y cronometra sus tiempos, obteniendo los siguientes resultados (tiempos en segundos):

52	59	63	49	54	56	54	50	66	70	61	57	53	55	56

Efectúe una estimación por intervalos, con una confianza del 95% para el tiempo medio de realización de dicho trabajo.

13. Cierta compañía desea determinar el tiempo medio de horas extra laboradas cada semana por sus trabajadores del área de producción. Se ha tomado una muestra de 15 trabajadores con las siguientes cantidades de horas para la semana pasada:

5	8	7	2	1	6	0	4	3	11	6	13	7	2	4

Construya un intervalo de confianza del 98% para la media poblacional.

14. Los siguientes datos corresponden a una muestra de personas que aplicaron un test para medir su cociente intelectual: 102, 98, 95, 112, 125, 85, 83, 129, 78, 116, 101, 96, 100, 103, 97, 102, 95, 101, 98, 99, 103, 101. Determine los intervalos de confianza del 90% para el cociente intelectual promedio.

15. Una institución realizó un plan de capacitación entre mujeres emprendedoras de una zona marginal del país. El objetivo es que estas mujeres lograran incrementar los ingresos de sus microempresas. Luego de implementadas las capacitaciones se compararon los ingresos de una muestra de 12 microempresas y se registraron los siguientes ingresos (en dólares al mes):

Participantes	1	2	3	4	5	6	7	8	9	10	11	12
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

a. Determine los intervalos de confianza del 90% para los ingresos de las participantes antes de la capacitación.
b. Determine los intervalos de confianza del 90% para los ingresos de las participantes después de la capacitación.
c. ¿A qué conclusión puede llegarse luego de analizar los dos resultados anteriores?

16. Un ingeniero ha recopilado datos sobre la vida útil de 20 filtros de gasolina del mismo tipo y marca. Los datos son los siguientes (en miles de kilómetros): 12, 14, 16, 15, 10, 20, 13, 15, 16, 14, 12, 13, 11, 13, 15, 16, 13, 14, 14, 12. Calcule los intervalos de confianza del 95% para la vida media de un filtro de gasolina.

17. Un contador está realizando una auditoría de los cheques extendidos por una compañía. En una muestra de 20 cheques se obtuvieron los siguientes montos, en miles dólares:

15, 17, 22, 8, 4, 2, 16, 6, 8, 14, 15, 25, 20, 5, 6, 8, 12, 19, 11, 12

Construya un intervalo del 95% de confianza para el monto medio de los cheques extendidos por la compañía.

18. Dada la alta incidencia de la diabetes, en una zona del país se ha implementado un programa de dieta y ejercicio para la población que padece esta enfermedad. Para determinar la efectividad del programa se han reunido los siguientes datos, que corresponden a los niveles de glucosa en la sangre de una muestra de 5 pacientes diabéticos, tomadas antes del programa de dieta y ejercicio y luego de dicho programa.

Paciente	Antes	Después
1	142	127
2	115	109
3	132	117
4	91	102
5	170	158

Con base en los datos anteriores:

a. Calcule los intervalos de confianza del 95% para los niveles de glucosa en la sangre antes del programa de dieta y ejercicio.
b. Calcule los intervalos de confianza del 95% para los niveles de glucosa en la sangre después del programa de dieta y ejercicio.
c. ¿Considera usted que existe evidencia estadística suficiente para considerar que hubo una disminución en los niveles de glucosa de los pacientes?

19. Se sabe que el tiempo que toma completar una prueba psicométrica tiene desviación estándar de 10 minutos. Una muestra de 30 estudiantes son sometidos a la prueba obteniéndose una media de 91 minutos. Obtenga los límites de confianza del 97% para la media poblacional.

20. Se desea conocer la incidencia del cáncer de estómago en una cierta zona del país. Si en una muestra de 2000 personas de esa zona se encontraron 26 con dicho padecimiento, obtenga los límites de confianza del 95% para la proporción de personas con cáncer de estómago en esa zona.

21. Una reciente encuesta incluyó a 1220 adultos elegidos al azar y se les preguntó si consideraban que debería permitirse la clonación de seres humanos. Los resultados mostraron que 912 de los encuestados dijeron que no debe permitirse la clonación. Se desea determinar si estos resultados constituyen una fuerte evidencia de que la mayoría de las personas (más del 50%) se oponen a este tipo de clonación. Para dar su respuesta,

a. Construya un intervalo de confianza del 99% de la proporción de adultos que considera que no debe permitirse este tipo de clonación.
b. ¿Considera usted que efectivamente la mayoría de la gente se opone a la clonación de seres humanos? Explique basándose en los resultados del punto a).

22. Según un estudio el 25% de las viviendas de zonas urbanas del país están en estado malo o regular. Si se empleó un tamaño de muestra de 50 viviendas, obtenga los intervalos de confianza del 98% para la proporción de viviendas de zonas urbanas del país están en estado malo o regular.

23. Una empresa pauta publicidad en televisión todos los meses. Recientemente ha lanzado una campaña muy agresiva, y se espera que al menos el 55% de los consumidores de menos de 40 años recuerden el anuncio de la empresa. Se ha tomado una muestra de 120 consumidores de menos de 40 años y el 48% dijeron que recordaban el anuncio de la empresa. Determine los intervalos de confianza del 95% para la proporción de consumidores meta que recuerda el anuncio. ¿Puede considerarse que se ha alcanzado la meta?

24. En una muestra de 250 periodistas, 120 indicaron que no percibían que hubiera importantes amenazas a la libertad de prensa en el país. Estime los intervalos de confianza del 90% para la proporción de periodistas que no percibían que hubiera importantes amenazas a la libertad de prensa en el país.

25. Un artículo publicado en la British Medical Journal (http://www.bmj.com/content/345/bmj.e4737) cuestiona los estudios que respaldan el efecto supuestamente positivo de las bebidas hidratantes como Gatorade, Powerade y otras. En el artículo se indica que el 76% de los estudios tenía algún tipo de problemas de metodología. Estime los intervalos del 95% de confianza para la proporción de estudio con problemas metodológicos si se empleó una muestra de 106 estudios.

26. En una población de 850 familias, se llevó a cabo una encuesta para estimar el consumo medio de leche. La muestra consistió en 44 familias seleccionadas por muestreo simple al azar. El consumo mensual en litros fue el siguiente:

15	45	75	90	45	68	41	12	16	52	53
8	28	35	63	54	90	47	35	41	49	38
30	15	53	43	39	54	21	43	38	32	45
36	32	47	29	41	40	40	42	52	30	44

a. Estime puntualmente la media de consumo mensual de leche en la población. Interprete el resultado.
b. Calcule e interprete el intervalo de confianza del 95% para la estimación del consumo mensual medio de leche por familia. Se conoce que $s^2 = 324$ litros. Interprete los resultados.
c. Si se conoce que cada familia, en promedio, tiene 4.4 miembros, calcule el intervalo de confianza del 95% para estimar el consumo per cápita de leche, en la población.
d. Obtenga e interprete el intervalo de confianza del 98%, para la proporción de familias que consumen menos de 35 litros mensuales de leche.

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Un ejemplo de inferencia estadística es:
   (a) Elaborar gráficas para un conjunto de datos muestrales
   (b) Calcular la media de la muestra
   (c) Estimar un parámetro poblacional a partir de datos muestrales
   (d) Calcular la media de una variable a partir de datos poblacionales

2. Un valor que describe una población se denomina:
   (a) Parámetro (b) Estadístico
   (c) Estimador (c) Observación

3. Luis está tratando de estimar el gasto promedio en alimentación de las familias de su país. Para resolver este problema:
   (a) Se puede entrevistar a todas y cada una de las familias del país
   (b) Seleccionar algunas familias "modelo" según el criterio de Luis
   (c) Seleccionar una muestra aleatoria de familias de todo el país
   (d) Seleccionar una muestra de familias cercanas al lugar donde Luis vive

4. De los siguientes, no es un ejemplo de un parámetro:
   (a) Media m (c) Desviación estándar s
   (b) Proporción P (d) Varianza s2

5. Un buen estimador debe ser insesgado, lo cual consiste en:
   (a) El valor esperado del estadístico es igual al valor del parámetro que se estima
   (b) Se utiliza toda la información proporcionada por la muestra en lo que se refiere al parámetro
   (c) La distribución del estimador está concentrada alrededor del parámetro
   (d) La precisión del estimador será mayor para tamaños de muestra grandes

6. Un buen estimador debe ser consistente, lo cual consiste en:
   (a) El valor esperado del estadístico es igual al valor del parámetro que se estima
   (b) Se utiliza toda la información proporcionada por la muestra en lo que se refiere al parámetro
   (c) La distribución del estimador está concentrada alrededor del parámetro
   (d) La precisión del estimador será mayor para tamaños de muestra grandes

7. Un buen estimador debe ser de varianza mínima, lo cual consiste en:
   (a) El valor esperado del estadístico es igual al valor del parámetro que se estima
   (b) Se utiliza toda la información proporcionada por la muestra en lo que se refiere al parámetro
   (c) La distribución del estimador está concentrada alrededor del parámetro
   (d) La precisión del estimador será mayor para tamaños de muestra grandes

8. Si se toman muestras aleatorias de n elementos de una población y se calculan los promedios es de esperar que:
   (a) El valor en cada caso sea igual a valor poblacional
   (b) Los valores de las medias no sean todos iguales
   (c) Los valores de las medias sean todos iguales
   (d) La diferencia entre una media y otra no se atribuya al azar

9. Si se toman muestras aleatorias de n elementos de una población, se calculan los promedios, se ponen los promedios muestrales en una tabla de frecuencia y se hace un histograma es de esperar que el gráfico:
   (a) Se parezca a una curva normal solo si la población original es normal
   (b) Se parezca a una curva normal aun cuando la población original no sea normal
   (c) No se parezca a una curva normal, excepto por casualidad
   (d) Se parezca a la distribución original de los datos de la población

10. El error estándar consiste en:
   (a) La media de los errores muestrales
   (b) La desviación estándar de los errores de muestreo
   (c) La media de los datos estandarizados
   (d) La desviación estándar de las medias muestrales

11. Se sabe que una variable x tiene una desviación estándar de 10. Si se toma una muestra de 16 unidades, entonces el error estándar equivale a:
   (a) 2,5 (c) 0,625
   (b) 4 (d) Ninguna de las anteriores

12. Se sabe que una variable x tiene una desviación estándar de 10. Si se toma una muestra de 16 unidades de una población de 70, entonces el error estándar equivale a:
   (a) 2,5 (c) 0,5529
   (b) 2,21 (d) Ninguna de las anteriores

13. La diferencia en, valor absoluto, entre el valor de la media muestral y la media poblacional se conoce como:
(a) Error estándar (b) Error de la estimación
(c) Error absoluto medio (d) Ninguna de las anteriores

14. Cuando se utiliza la media muestral como estimación de la media poblacional μ, la probabilidad de que esta estimación no falle es:
(a) La media poblacional
(b) El error estándar
(c) El error estimado
(d) El nivel de confianza

15. Se desea estimar la media poblacional de una variable *x* cuya desviación estándar poblacional es de 5 unidades. En una muestra de tamaño 45 se obtiene una media de 63 unidades, entonces el valor de z necesario para obtener los intervalos de confianza del 95% es:
(a) 1,645 (b) 0,95
(c) 1,96 (d) 2,58

16. Se desea estimar la media poblacional de una variable *x* cuya desviación estándar poblacional es de 5 unidades. En una muestra de tamaño 45 se obtiene una media de 63 unidades, entonces el al obtener los intervalos de confianza del 95%, el límite inferior es:
(a) 61,77 (b) 61,54
(c) 64,46 (d) Ninguna de las anteriores

17. Se desea estimar la media poblacional de una variable *x* cuya desviación estándar poblacional es de 5 unidades. En una muestra de tamaño 45 se obtiene una media de 63 unidades, entonces el al obtener los intervalos de confianza del 95%, se concluye que:
(a) Con una confianza del 95% la media poblacional es 63 unidades
(b) Con una confianza del 95% la media poblacional está entre 61,77 y 64,23 unidades
(c) Con una confianza del 95% la media poblacional es mayor que 61,77 unidades
(d) Con una confianza del 95% la media poblacional está entre 61,54 y 64,46 unidades

18. Se desea estimar la media poblacional de una variable *x* cuya desviación estándar poblacional es de 15 unidades. En una muestra de tamaño 200 se obtiene una media de 87 unidades, entonces al obtener los intervalos de confianza del 90%, el límite superior es:
(a) 87 (b) 85,26
(c) 88,74 (d) Ninguna de las anteriores

19. Se desea estimar la media poblacional de una variable *x* distribuida normalmente cuya desviación estándar poblacional es de 20 unidades. En una muestra de tamaño 12 se obtiene una media de 125 unidades, al obtener los intervalos de confianza del 99%, un investigador realizó las siguientes dos afirmaciones:
    A. Se debe usar un valor de z de 2,58.
    B. El límite inferior es 107,07.

    Con respecto a las dos afirmaciones anteriores, es correcto que:
    (a) Ambas son verdaderas (b) Solo A es verdadera
    (c) Ambas son falsas (d) Solo B es verdadera

20. Se desea estimar la media poblacional de una variable x distribuida normalmente. En una muestra de tamaño 12 se obtiene una media de 125 unidades y una desviación estándar de 20 unidades, al obtener los intervalos de confianza del 99%, un investigador realizó las siguientes dos afirmaciones:
    A. Se debe usar un valor de t de 3,11.
    B. El límite superior es 142,93.

    Con respecto a las dos afirmaciones anteriores, es correcto que:
    (a) Ambas son verdaderas (b) Solo A es verdadera
    (c) Ambas son falsas (d) Solo B es verdadera

21. Se desea estimar la media poblacional de una variable *x* distribuida normalmente. En una muestra de tamaño 20 se obtiene una media de 3200 unidades y una desviación estándar de 450 unidades, al obtener los intervalos de confianza del 95%, un investigador realizó las siguientes dos afirmaciones:
    A. Se debe obtener el valor de *t* con 21 grados de libertad.
    B. Los límites de confianza son 2989,39 y 3410,61 unidades.

    Con respecto a las dos afirmaciones anteriores, es correcto que:
    (a) Ambas son verdaderas (b) Solo A es verdadera
    (c) Ambas son falsas (d) Solo B es verdadera

22. Se desea estimar la media poblacional de una variable *x*. En una muestra de tamaño 80 se obtiene una media de 30 unidades y una desviación estándar de 4,5 unidades, al obtener los intervalos de confianza del 99%, un investigador realizó las siguientes dos afirmaciones:
    A. Se debe usar un valor de z de 2,58.
    B. El valor de la media poblacional es superior a 28,70.

Con respecto a las dos afirmaciones anteriores, es correcto que:
    (a) Ambas son verdaderas (b) Solo A es verdadera
    (c) Ambas son falsas (d) Solo B es verdadera

23. Al estimar la media poblacional de una variable *x*, en una muestra de tamaño 50 se obtiene una media de 1500 unidades y una desviación estándar de 250 unidades. Al obtener los intervalos de confianza del 95%, es verdadero que:
    (a) Con certeza la media está entre 1430,70 y 1569,30
    (b) Con una confianza del 95% la media poblacional es mayor que 1430,70 unidades
    (c) Con una confianza del 95% la media poblacional está alrededor de 1500 unidades
    (d) Con una confianza del 95% la media poblacional está entre 1430,70 y 1569,30 unidades

24. Al estimar la media poblacional de una variable *x*, en una muestra de tamaño 500 se obtiene una media de 2150 unidades y una desviación estándar de 600 unidades. Al obtener los intervalos de confianza del 90% se obtuvo como límite inferior 2105,86 y como límite superior 2194,14, entonces es verdadero que:
    (a) La media está entre 2105,86 y 2194,14
    (b) Con una confianza del 90% la media poblacional es menor que 2194,14 unidades
    (c) La media poblacional será mayor que 2194,14 con una probabilidad de 5%
    (d) La media poblacional estará entre 2105,86 y 2194,14 unidades en 90 de cada 100 muestras

25. Si *x* es el número de veces que ha ocurrido un evento en una muestra *n* pruebas, entonces el cociente *x/n* representa:
    (a) La proporción poblacional (b) La proporción muestral
    (c) La probabilidad de fracaso (d) Ninguna de las anteriores

26. Se desea estimar una proporción poblacional de una cierta variable. En una muestra de tamaño 120 se obtiene un conteo de 90 eventos. Al obtener los intervalos de confianza del 99%, un investigador realizó las siguientes dos afirmaciones:
    A. Se debe usar un valor de z de 2,58.
    B. No se pueden calcular los intervalos porque no se tiene la desviación estándar.

    Con respecto a las dos afirmaciones anteriores, es correcto que:
    (a) Ambas son verdaderas (b) Solo A es verdadera
    (c) Ambas son falsas (d) Solo B es verdadera

27. Se desea estimar una proporción poblacional de una cierta variable. En una muestra de tamaño 120 se obtiene un conteo de 90 eventos. Al obtener los intervalos de confianza del 99%, un investigador realizó las siguientes dos afirmaciones:

    A. La proporción muestral es 0,75.
    B. Los intervalos de confianza del 99% son 0,6482 y 0,8518.

    Con respecto a las dos afirmaciones anteriores, es correcto que:
    (a) Ambas son verdaderas (b) Solo A es verdadera
    (c) Ambas son falsas (d) Solo B es verdadera

28. Se desea estimar una proporción poblacional de una cierta variable. En una muestra de tamaño 12 se obtiene un conteo de 5 eventos. Al obtener los intervalos de confianza del 90%, un investigador realizó las siguientes dos afirmaciones:

    A. Se emplea un valor de t con 11 grados de libertad.
    B. El límite superior es de 65,08.

    Con respecto a las dos afirmaciones anteriores, es correcto que:
    (a) Ambas son verdaderas (b) Solo A es verdadera
    (c) Ambas son falsas (d) Solo B es verdadera

29. Se desea estimar una proporción poblacional de una cierta variable. En una muestra de tamaño 1200 se obtiene un conteo de 750 eventos. Al obtener los intervalos de confianza del 95%, un investigador realizó las siguientes dos afirmaciones:

    A. La proporción poblacional es 62,5%.
    B. El límite inferior es de 59,76%.

    Con respecto a las dos afirmaciones anteriores, es incorrecto que:
    (a) Ambas son verdaderas
    (b) Solo A es verdadera
    (c) Ambas son falsas
    (d) Solo B es verdadera

30. Se desea estimar el peso promedio de las galletas que se elaboran en una fábrica. En una muestra de tamaño 1100 paquetes de galletas se obtiene una media de 195 gramos con una desviación estándar de 45 gramos. La empresa ha especificado que el peso de cada paquete de galletas debe ser 200 gramos. Al obtener los intervalos de confianza del 95%, el encargado del control del proceso realizó las siguientes dos afirmaciones:

    A. No hay problema con el peso de las galletas, el 95% de las galletas tiene un peso de 195 grs.
    B. El peso especificado de 200 gramos está fuera del intervalo de confianza del 95%.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas
(b) Solo A es verdadera
(c) Ambas son falsas
(d) Solo B es verdadera

31. Se desea estimar el peso promedio de las galletas que se elaboran en una fábrica. En una muestra de tamaño 10 paquetes de galletas se obtienen los siguientes pesos (en gramos):

190	210	201	196	197	185	176	208	200	191

La empresa ha especificado que el peso de cada paquete de galletas debe ser 200 gramos. Al obtener los intervalos de confianza del 95%, el encargado del control del proceso realizó las siguientes dos afirmaciones:

A. El 95% de las galletas tiene un peso entre 188,01 y 202,79 gramos.
B. El peso especificado de 200 gramos está dentro del intervalo de confianza del 95%.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

32. Se desea estimar el peso promedio de las galletas que se elaboran en una fábrica. Se sabe que el peso medio de los paquetes de galletas se distribuye normalmente y que tiene una desviación estándar de 15 gramos. En una muestra de tamaño 10 paquetes de galletas se obtienen los siguientes pesos (en gramos):

190	210	201	196	197	185	176	208	200	191

La empresa ha especificado que el peso de cada paquete de galletas debe ser 200 gramos. Al obtener los intervalos de confianza del 95%, el encargado del control del proceso realizó las siguientes dos afirmaciones:

A. El 95% de las galletas tiene un peso entre 186,10 y 204,70 gramos.
B. El peso especificado de 200 gramos está fuera del intervalo de confianza del 95%.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

33. Se desea estimar proporción de las galletas que se elaboran en una fábrica cuyo peso está por debajo de la especificación. En una muestra de tamaño 10 paquetes de galletas se obtienen los siguientes pesos (en gramos):

190	210	201	196	197	185	176	208	200	191

La empresa ha especificado que el peso de cada paquete de galletas debe ser 200 gramos. Al obtener los intervalos de confianza del 95%, el encargado del control del proceso realizó las siguientes dos afirmaciones:

A. El límite superior del 95% es un peso de 90,36 gramos.
B. Con una confianza del 95% entre 29,6% y 90,4% de las galletas pesan menos de 200 grs.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

34. En un periódico se presentan los resultados de una encuesta aplicada a una muestra aleatoria de 1200 adultos, de los cuales 610 indicaron que la labor del gobierno es buena o muy buena. El estudio se hizo con una confianza del 95%. Según el autor del artículo la mayoría de los ciudadanos consideran que la labor del gobierno es buena o muy buena. Con respecto a esa afirmación del autor del artículo un crítico realizó la siguiente aseveración: "El autor se ha equivocado, ya que, 1. Con una confianza del 95% la proporción de ciudadanos de ciudadanos que aprueban la gestión del gobierno podría estar entre 48% y 53,6%, con lo cual es muy probable que el porcentaje de ciudadanos que están de acuerdo con la gestión del gobierno sea inferior al 50%". Con respecto a esta situación es correcto que:

(a) El autor está en lo correcto y el crítico está equivocado
(b) El autor está equivocado y el crítico también
(c) El autor está equivocado y el crítico está en lo correcto
(d) Falta información para indicar quién está equivocado y quién no

CAPÍTULO 4

MUESTREO

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Explicar la importancia y necesidad de trabajar con muestras para conocer información sobre la población
2. Describir algunas aplicaciones empresariales del muestreo
3. Calcular el tamaño de muestra necesario para estimar la media poblacional y proporción poblacional
4. Describir las principales técnicas de muestreo probabilístico

4.1 Introducción

En los primeros capítulos de este texto se han presentado distintas técnicas que se emplean en estadística para describir un conjunto de datos, tales como las distribuciones de frecuencias y el cálculo de diferentes medidas numéricas, como la media, la mediana, la moda, la varianza, la desviación estándar, entre otras medidas. Sin embargo, en la práctica rara vez se cuenta con los datos ya recolectados y, por tanto, es necesario efectuar un proceso de recolección de la información, en el cual, con gran frecuencia se debe tomar la decisión de trabajar con la población o con una muestra.

Así, por ejemplo, cuando una compañía está estudiando la proporción de sus clientes que están satisfechos con sus servicios podría aplicar un cuestionario a todos sus clientes, o bien, emplear una muestra. Aplicar el cuestionario a todos y cada uno de los clientes, o sea, a la población, puede requerir mucho tiempo y recursos, ya que, la cantidad de clientes podría ser de muchos miles, lo que podría no ser viable para la organización. Ahora bien, empleando una muestra la compañía podría conseguir resultados muy confiables, que le ayuden a alcanzar su propósito de medir la satisfacción de los clientes, pero a un costo mucho más bajo.

La decisión de trabajar o no con una muestra es apenas una primera decisión. Si se decide trabajar con una muestra, debe determinar el tamaño de dicha muestra. Cabe resaltar que esa muestra que se va a emplear puede ser de apenas unos cientos de clientes, aun cuando la compañía tenga miles de clientes. Además, se debe emplear una técnica que permita seleccionar cuáles clientes van a ser incluidos en la muestra y cuáles no. Tanto un tamaño adecuado de muestra como la aplicación de la técnica adecuada de muestreo son importantes para obtener datos confiables, y por tanto, útiles para la toma de decisiones.

En este capítulo se expondrá la importancia y necesidad de trabajar con muestras, a la vez que se describirán algunas aplicaciones frecuentes del muestreo en el entorno empresarial. Luego se presentarán los distintos factores que determinan el tamaño de la muestra cuando se desea estimar la media de la población o la proporción poblacional, como en el ejemplo anterior, que se podría desear conocer la proporción de clientes satisfechos. Finalmente, se presentarán distintas técnicas de muestreo, que son las que permiten establecer cuáles elementos de la población se van a incluir en la muestra.

Conceptos de población y muestra

Tal como se mencionó en el capítulo 1, con frecuencia la palabra población se asocia con la cantidad de personas que habitan en un país o en una región, pero en estadística este concepto abarca muchos otras posibles unidades de estudio, que no son exclusivamente personas.

En toda investigación estadística hay un objeto de estudio, que es la unidad de interés y sobre la cual recae la observación. Por ejemplo, en el control de la calidad de una fábrica de llantas, se podrían emplear pruebas de rendimiento en laboratorio bajo determinadas condiciones y también pruebas en carreteras, para medir el desgaste en diferentes rutas, entre otras variables de interés. A esta unidad de estudio también s ele puede llamar unidad elemental o unidad estadística.

> **UNIDAD ESTADÍSTICA** Unidad de interés en un estudio estadístico.

Así, la población está constituida por el conjunto total de los elementos de interés. A este conjunto de elementos que son objeto de estudio le llamamos población o universo.

> **POBLACIÓN** Conjunto de individuos u objetos de interés o medidas obtenidas a partir de todos los individuos u objetos de interés.

Así, podríamos tener que la población está constituida por los clientes de una empresa, por productos de una empresa, como las llantas producidas en un determinado periodo; por documentos, tales como los cheques emitidos por un el departamento financiero de una compañía; los estudiantes de la carrera de administración de negocios de una universidad; los empleados de una organización; familias bajo condiciones de pobreza; entre muchos otros.

Al efectuar el estudio estadístico se definen las características o variables que se desea observar en cada unidad de estudio. Por ejemplo, en el caso de que la unidad elemental sea un hogar, entonces estas variables pueden el ingreso mensual del hogar, el número de miembros del hogar, el nivel académico del jefe del hogar, entre muchas otras.

Tal como se señaló anteriormente, existen diversas razones por las que sería imposible trabajar con toda la población, o bien, demasiado costoso, y por tanto es necesario emplear una muestra. Una muestra es una parte de la población de interés, y si es bien selecciona es representativa de esa población, y las conclusiones que en ella se obtengan luego podrán ser generalizadas al resto de la población.

> **MUESTRA** Porción o parte de la población de interés.

En los ejemplos anteriores mencionamos que se podía obtener una muestra de clientes para determinar la proporción de clientes satisfechos con los servicios de la empresa, o bien, emplear una muestra para determinar el desgaste promedio de determinado tipo de llantas bajo ciertas condiciones. Cuando se desea lanzar al mercado un nuevo producto, como parte del estudio de mercado, generalmente, se emplea una muestra de los potenciales consumidores para estimar la demanda del producto. También un auditor puede tomar una muestra de las cuentas por cobrar de una compañía, para verificar el cumplimiento de ciertos requisitos.

Ahora bien, cuando se habla de muestreo se hace referencia a todo un proceso mediante el cual se va a establecer un tamaño de muestra, es decir, la cantidad de elementos que se van a tomar en la muestra; se va a seleccionar una técnica de muestreo, o sea, una técnica que permita establecer cuáles elementos de la población se van a incluir en la muestra; y se va a medir la confianza que tienen las estimaciones realizadas con base en esa muestra.

MUESTREO Conjunto de técnicas que se utilizan para seleccionar una muestra de una población de interés y medir la confianza de las estimaciones realizadas.

En algunos casos el diseño de la muestra no es muy importante, como en los exámenes de sangre, dado que la sangre es muy homogénea, pero en la mayoría de las situaciones, el diseño de la muestra es clave para garantizar la validez y confiabilidad de los resultados. El problema del muestreo estadístico es lograr seleccionar muestras representativas en poblaciones que no son homogéneas.

Además de definir la unidad estadística, es necesario también establecer la unidad de muestreo.

UNIDAD DE MUESTREO Unidad básica en términos de la cual se aplica una técnica de muestreo y se seleccionan los elementos incluidos en la muestra.

Tal como se verá más adelante, existen diferentes técnicas de muestreo, y según la técnica seleccionada, además de las características del estudio, y de los recursos e información disponible se define la unidad de muestreo. Por ejemplo, si se va a efectuar un estudio sobre la satisfacción de los empleados de una empresa, entonces la unidad de muestreo y la unidad de estudio coinciden, y serían un empleado de la compañía, en este caso.

En otros casos la unidad de estudio y la unidad de muestreo no coinciden. Suponga que se va a efectuar un estudio entre estudiantes de primer año de secundaria de todo el país. Entonces el investigador podría tomar inicialmente una muestra de centros educativos, luego dentro de cada colegio escoge una muestra de grupos de primer año y, finalmente, dentro de cada grupo, selecciona una muestra de estudiantes. Entonces, podemos decir que la unidad de muestreo es aquella que es objeto de escogencia aleatoria a la hora de seleccionar la muestra.

UNIDAD DE INFORMACIÓN Unidad que proporciona los datos relacionados con la unidad de estudio.

Ahora bien, cuando se efectúa el estudio, muchas veces la persona que proporciona la información no es la misma unidad de muestreo. Por ejemplo, en una encuesta de hogares, donde la unidad de muestreo puede ser el hogar, la unidad de información podría ser el jefe del hogar. O bien, en un estudio sobre empresas, la unidad de información o informante puede ser un gerente de cada una de las empresas seleccionadas en la muestra.

4.2 Necesidad de trabajar con muestras

Es posible mencionar distintas razones por las cuales es necesario trabajar con muestras. Entre las razones más importantes puede mencionarse las siguientes:

- **Estudiar la población sería muy costoso.** Esta es una de las principales razones para emplear muestras. Suponga que un candidato presidencial desea conocer la proporción de votantes que votarían por él. Para dar un ejemplo, en Costa Rica el padrón electoral superó los 2.800.000 electores para las elecciones presidenciales del 2010. Imagine el costo de entrevistar a tal cantidad de personas distribuidas a lo largo y ancho de todo el país.
- **Estudiar la población requeriría demasiado tiempo.** Si una empresa desea lanzar un nuevo producto al mercado, y cuenta con un equipo de 5 personas para efectuar el estudio de mercado. Si hubiera un millón de potenciales clientes, ¿cuánto tiempo durarían las 5 personas en encuestar a ese millón de personas? El estudio duraría años, de modo que cuando obtengan los resultados, la información ya no sería útil.
- **La población se destruye al ser observada.** Por ejemplo, si una fábrica de bombillos desea estimar la vida útil de su producto, si decidiera emplear la población para hacer esta estimación, tendría que probar todos los bombillos que produce, lo cual significaría destruir toda su producción, y por tanto, sería imposible estudiar la población.
- **La población es infinita.** Si en un proceso de producción industrial se requiere establecer la proporción de artículos que satisfacen una cierta especificación, entonces se tiene que la producción tiende a infinito a lo largo del tiempo, por lo que sería imposible estudiar la población.
- **Conveniencia.** Esta es otra importantísima razón para emplear muestras. Si se selecciona adecuadamente la muestra, los resultados que se obtienen a partir de ella pueden ser suficientemente

buenos, de modo que aun cuando se tuvieran suficientes recursos y tiempo, no necesariamente valdría la pena estudiar la población. Se sabe que conforme se incrementa el tamaño de la muestra, el error de la estimación se disminuye, así que si se tiene un nivel de error aceptable con un alto grado de confianza, los resultados de una muestra pueden ser muy satisfactorios.

- **Mayor calidad de la información recopilada.** Al trabajar con muestras es más fácil controlar el trabajo de recolección de la información, lo cual permite una mejor supervisión y verificación, de modo que se asegura que la información obtenida es de la calidad requerida.
- **Imposibilidad.** En algunos casos simplemente es imposible efectuar un estudio empleando toda la población. Suponga que se desea determinar el nivel de cierta sustancia contaminante en el agua de un río. Es imposible hacer el estudio con toda el agua del río. La única opción es tomar una muestra.

Muestras e inferencia

La finalidad de hacer un estudio estadístico es conocer información sobre la población. Tal como se ha señalado, se emplea la muestra para efectuar la estimación de los parámetros poblacionales. Entonces, si en un estudio se desea conocer el desgaste promedio de cierto tipo de neumáticos bajo determinadas condiciones, entonces el valor de ese promedio para la población es un parámetro poblacional.

PARÁMETRO Medida descriptiva de la población de interés.

Dado que normalmente no se conocen los parámetros, entonces se emplea la muestra para estimarlos. Por ejemplo, se toma una muestra de neumáticos y se determina el desgaste promedio. Ese valor muestral es un estimador que se emplea para estimar el promedio poblacional.

ESTIMADOR Medida descriptiva de la muestra y que sirve como una estimación del parámetro poblacional correspondiente.

Cuando se han determinado los estimadores, entonces lo que se debe hacer es un proceso mediante el cual se generalizan los resultados de la muestra a la población. Esto es lo que se conoce como la inferencia estadística.

INFERENCIA ESTADÍSTICA Métodos empleados para determinar una propiedad de una población con base en la información de la muestra.

Al realizar la inferencia, si la muestra es representativa y aleatoria, entonces se puede establecer el nivel de confianza de la estimación y el nivel de error implícito en dicha estimación. Para poder realizar la inferencia es fundamental que la muestra sea tomada aleatoriamente.

4.3 Muestreo estadístico y no estadístico

Puede hablarse de dos tipos básicos de muestras, de acuerdo con el método que se emplee para seleccionar las unidades incluidas en la muestra:

- **Muestreo aleatorio**: Si cada elemento de la muestra ha sido seleccionado siguiendo un procedimiento que concede a cada unidad de la población una probabilidad conocida de ser incluido en la muestra, entonces se dice que la muestra es aleatoria. Una muestra aleatoria es aquella que se selecciona al azar; también se le llama muestra estadística.
- **Muestreo no aleatorio**. Dentro de este tipo de muestreo se distinguen varios casos: la selección por conveniencia, que corresponde a aquellas muestras seleccionadas de acuerdo con la conveniencia del investigador. Este tipo de muestreo es muy sesgado y no puede ser aceptado por su naturaleza. Un segundo caso es la selección intencional o de juicio, en la cual la selección la hace un experto con conocimiento amplio de la población bajo estudio. También se puede dar un muestreo voluntario, que se da cuando los individuos deciden participar por su propia decisión en el estudio.

Dentro del muestreo aleatorio hay distintas técnicas de muestreo, las cuales se diferencian por sus distintos procedimientos aleatorios, o sea, que asignan diferentes probabilidades a las unidades bajo estudio de ser seleccionadas en la muestra. Así, se tiene el muestreo simple al azar, el muestreo estratificado, el muestreo en etapas, el muestreo sistemático y el muestreo por conglomerados.

La principal ventaja del muestreo aleatorio es que limita el sesgo de selección y permite la cuantificación y control del error de muestreo. Las muestras no aleatorias no permiten medir ni controlar el error de muestreo. Es por eso que en términos generales se va a preferir el muestreo aleatorio sobre los muestreos no aleatorios. Solo en ciertos casos especiales, se considera que el muestreo de juicio puede lograr muestras más representativas. Esto cuando se toma una muestra pequeña seleccionada de una poblaciones muy heterogénea, pues el juicio del experto podría ayudar a obtener una muestra representativa.

Errores de muestreo y sesgos

En los estudios estadísticos por muestreo aleatorio se pueden presentar dos tipos de error. Uno es el error de muestreo y el otro corresponde a los llamados sesgos. Tal como el nombre lo sugiere, el error de muestreo se presenta solo en aquellos casos en los cuales se ha tomado una muestra, mientras que los sesgos pueden presentarse tanto en los casos en los cuales se trabaje con una muestra como en aquellos en que se efectúa la enumeración total de la población.

Suponga que se realiza un estudio de intención de voto para conocer la proporción de electores que están decididos a votar por determinado candidato. Entonces se toma una muestra aleatoria para realizar dicha estimación. Aunque esta muestra se haya obtenido empleando procedimientos adecuados, la proporción obtenida en la muestra no es exactamente igual a la población debido a la naturaleza aleatoria de la muestra. Esa diferencia entre el verdadero valor poblacional y el valor muestral estimado es el error de muestreo.

ERROR DE MUESTREO Discrepancia, debida al azar, entre la estimación de una característica obtenida a través de una muestra y su verdadero valor en la población.

El error de muestreo tiene la gran ventaja de que puede ser medido y controlado. Por ejemplo, si se desea reducir la magnitud del error de muestreo, se puede incrementar el tamaño de la muestra.

La naturaleza aleatoria de la muestra permite conocer la probabilidad de que la muestra sea representativa, sin embargo siempre existe la posibilidad de que la muestra no sea representativa, y que, por tanto, el error de muestreo sea muy alto.

Por otro lado, cualquier error que no se deba al azar también ocasiona que el valor estimado no sea igual al valor verdadero, pero no es un error de muestreo, sino un sesgo.

SESGO Error sistemático, no debido al azar, y que ocasiona diferencias entre el valor estimado a través de la muestra y el valor verdadero.

Los sesgos son de naturaleza sistemática y no pueden ser medidos. Los sesgos deben ser prevenidos a través de distintas medidas administrativas. Algunas de las principales fuentes de sesgos son las siguientes:

- **Selección inadecuada de la muestra**. Si al seleccionar la muestra influye algún factor subjetivo, como el criterio o conveniencia de alguna persona, es decir, cuando la muestra, o parte de la muestra, no es obtenida al azar. Por ejemplo, si un encuestador debe aplicar un cuestionario telefónicamente a una muestra seleccionada al azar, pero no logra localizar a uno de los individuos seleccionados, y entonces decide llamar a cualquier otra persona. También esto ocurre en los llamados muestreos voluntarios, como los "sondeos" realizados por algunos canales de televisión y que las personas llaman para dar su opinión sobre algún tema.

- **Sesgos introducidos por el encuestador.** Cuando el encuestador no hace bien su trabajo, puede generar una serie de sesgos. Por ejemplo, si entrevistar a las personas, no registra apropiadamente las respuestas o si introduce sus propios juicios al interpretar las respuestas de los entrevistados. Para esto es fundamental que el encuestador esté bien capacitado y que su trabajo sea supervisado apropiadamente.
- **Inadecuado método de estimación.** Cada parámetro que se desea estimar posee su propio método de estimación. Si no se emplea el método adecuado, también se introduce un sesgo.
- **No se logra respuesta de todas las unidades.** Esto ocurre cuando muchos elementos no responden la encuesta, como en las encuestas aplicadas por correo postal o correo electrónico. Cuando por alguna razón una unidad no responde el cuestionario, entonces debe efectuarse la sustitución de modo adecuado.
- **Otras causas.** Pueden presentar otras causas de sesgos, la mayoría de ellas se relaciona con la ejecución errónea de alguna de las fases del proceso de investigación estadística.

4.4 Determinación del tamaño de muestra para estimar la media poblacional

Uno de los parámetros que con mayor frecuencia se requiere estimar es la media poblacional. Este es el caso en el que se desea estimar el ingreso promedio de los hogares, el tiempo promedio en que los operarios de una fábrica realizan una actividad, el peso promedio de las latas de leche condensada envasadas en una empresa, entre muchos otros posibles casos.

Es por eso que expondremos los determinantes del tamaño de muestra cuando se desea estimar la media poblacional y la forma en que se calcula el tamaño de muestra, tanto cuando la población es infinita como cuando es finita.

4.5 Determinación del tamaño de la muestra

Cuando se desea estimar el tamaño de muestra cuando se desea estimar la media poblacional, hay que considerar tres variables principales:

- **Variabilidad de la población**: La variabilidad es, tal vez, el principal determinante del tamaño de la muestra. Por ejemplo, al efectuar un examen de sangre para determinar el tipo de sangre de una persona, basta con una pequeña cantidad de sangre, pues la sangre es homogénea, o sea, su variabilidad es muy baja. Pero, por ejemplo, cuando el Instituto Nacional de Estadística y Censos

(INEC) realiza la encuesta de hogares de propósitos múltiples, se estudian variables con una elevada dispersión, como pobreza, ingreso, desempleo, entre otras. Podemos ver que hay hogares con niveles de ingresos muy bajos y otros con ingresos muy altos. Es decir, el ingreso no es una variable tan homogénea como la sangre. En el 2009 el INEC aplicó la encuesta a un total de 15.242 hogares. Tal como se estudió en el capítulo 3, la variabilidad o dispersión de un conjunto de datos se puede medir a través de la desviación estándar poblacional, denotada por σ.

- **Precisión deseada o nivel máximo de error permitido**: Anteriormente se indicó que en las estimaciones a partir de una muestra se presenta un error, pues debido al azar se va a presentar una discrepancia entre la estimación de una característica obtenida a través de una muestra y su verdadero valor poblacional. Aunque el investigador no conoce el valor de ese error, sí puede establecer un nivel máximo de error permitido en la estimación. Conforme se establezca un nivel de error menor, más precisa será la estimación y se requerirá una muestra de mayor tamaño. Si, por el contrario, se establece un nivel de error mayor, menos precisa será la estimación y, en consecuencia, podrá trabajarse con una muestra de menor tamaño. A este nivel de error se le denotará por E.
- **Nivel de confianza que se desea tener**: El investigador debe indicar el nivel de confianza deseado para la estimación. Los niveles de confianza del 95% y del 99% son los que se emplean con mayor frecuencia. Cuando se estima la media poblacional, por el llamado teorema de límite central (tema que se desarrollará más adelante), se toma la distribución de probabilidad normal como la distribución muestral de la media. Entonces, luego de establecer el nivel de confianza, se busca el valor de z correspondiente en la curva normal. Por ejemplo, para un nivel de confianza del 95%, el valor de z es de 1,96, y para un nivel de confianza del 99%, el valor de z es de 2,58. El nivel de confianza generalmente lo denotamos $1 - \alpha$, y entonces buscamos un valor z en la curva normal acumulada correspondiente a $1 - a/2$.

Tal como se comentó en la sección Estadística en acción, en realidad el tamaño de la población no es un factor tan relevante. Esto explica por qué en los estudios de intención de voto, el tamaño de la muestra puede ser muy similar en países como Estados Unidos, México o costa Rica, a pesar de que tienen poblaciones de muy distinto tamaño, se pueden emplear tamaños de muestra similares. Sin embargo, más adelante se mostrará que en poblaciones finitas se efectúa un ajuste por el tamaño de la población.

En la práctica existe un factor que es muy importante, y aunque las fórmulas no lo toman en cuenta debe ser valorado por cualquier investigador, el cual es el costo. Efectuar un estudio en una muestra grande puede ser muy caro y es posible no contar con los recursos suficientes. Por supuesto que si por cuestiones de costo se reduce el tamaño de la muestra, se podría estar sacrificando la confianza y la precisión de las estimaciones realizadas posteriormente. Esta es una decisión que debe ser tomada con mucho cuidado.

Del mismo modo, es importante tener en cuenta el número esperado de no respuestas que se puedan obtener. Por ejemplo, si por experiencia previa se pudiera estimar que el 25% de los encuestados no contesta el cuestionario, entonces sería necesario tomar una muestra un 25% más grande con el objetivo de poder entrevistar al número adecuado de personas.

4.6 Cálculo del tamaño de la muestra

Para el cálculo del tamaño de muestra para estimar la media poblacional se tomarán en cuenta los factores descritos anteriormente los cuales conforman las fórmulas que a continuación se presentan.

Caso de poblaciones infinitas En el caso de poblaciones infinitas, tal como se mencionó, el tamaño de la muestra depende de:

Desviación estándar de la población (σ)
Precisión deseada o nivel máximo de error permitido (E)
Valor de z correspondiente al nivel de confianza que se desea tener (z)

Entonces estos tres elementos se incorporan en la siguiente fórmula:

FÓRMULA DEL TAMAÑO DE MUESTRA $$n = \left(\frac{\sigma \cdot z}{E}\right)^2$$

Para explicar la aplicación de la fórmula se desarrolla el siguiente ejemplo.

Ejemplo

Suponga que se desea estimar el gasto promedio diario que realizan los turistas estadounidenses cuando visitan el país. Por un estudio anterior se sabe que esta variable tiene una desviación estándar de $46,6. Además, se desea que la estimación tenga un error máximo de $10 y con una confianza del 95%. ¿Qué tamaño de muestra se necesita?

Solución

Con base en los datos anteriores, se tiene que para el nivel de confianza del 95% corresponde un valor de z de 1,96. Así que se plantea:

Desviación estándar de la población: $\sigma = \$46{,}6$
Precisión deseada o nivel máximo de error permitido: $E = \$10$
Valor de z correspondiente al nivel de confianza del 95%: $z = 1{,}96$

Sustituyendo los valores en la fórmula:

$$n = \left(\frac{\sigma \cdot z}{E}\right)^2 = \left(\frac{46{,}6 \cdot 1{,}96}{10}\right)^2 = 83{,}42 \approx 84$$

Generalmente cuando se determine el tamaño de muestra se va a redondear *hacia arriba*.

De acuerdo con el resultado anterior, se requiere una muestra de 84 turistas estadounidenses para efectuar una estimación del gasto promedio diario en el país con una confianza del 95% y con una discrepancia máxima entre el valor estimado y el valor real de \$10.

Ejercicio de revisión

Se desea estimar el salario promedio de los operarios industriales del país. Se conoce que la desviación estándar de estos salarios es de \$236. Se requiere una estimación con un error máximo de \$50 y una confianza del 99%. ¿De qué tamaño debe ser la muestra?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Caso de poblaciones finitas

La fórmula dada anteriormente se aplica a poblaciones infinitas, pero en poblaciones finitas cuando se trabaja con reemplazo o en poblaciones finitas cuando el tamaño de la muestra n es muy pequeño con respecto al tamaño de la población N puede realizarse un ajuste.

Entonces, en un muestreo sin reemplazo, cuando la población es finita y el tamaño de la muestra n es relativamente grande con respecto al tamaño de la población N, se puede aplicar el siguiente factor de corrección:

FACTOR DE CORRECCIÓN PARA POBLACIONES FINITAS

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

donde n_0 es el resultado dado por la fórmula $n = \left(\frac{\sigma \cdot z}{E}\right)^2$ y N es el tamaño de la población.

Ejemplo

Una empresa posee un total de 800 camiones que se emplean para repartir sus productos a nivel nacional. Se desea estimar mediante una muestra aleatoria de los camiones para determinar la cantidad de kilómetros recorridos mensualmente. Por otro estudio realizado hace un tiempo, se conoce que esta variable tiene una desviación estándar de 380 kilómetros. La estimación debe tener un error máximo de 30 kilómetros y una confianza del 95%. ¿Qué tamaño de muestra se necesita?

Solución

Dado que la población es finita, entonces la determinación del tamaño de muestra se efectuará en dos etapas. Primero se calculará el tamaño de muestra como si la población fuera infinita. Luego se aplicará el factor de corrección para poblaciones finitas.

Con base en los datos anteriores, se tiene que para el nivel de confianza del 95% corresponde un valor de z de 1,96. Así que se plantea:

Desviación estándar de la población: $\sigma = 380$ km.
Precisión deseada o nivel máximo de error permitido: $E = 30$ km.
Valor de z correspondiente al nivel de confianza del 95%: $z = 1{,}96$
Tamaño de la población: 800 camiones

Sustituyendo los valores en la fórmula:

$$n = \left(\frac{\sigma \cdot z}{E}\right)^2 = \left(\frac{380 \cdot 1{,}96}{30}\right)^2 = 616{,}36 \approx 617$$

Ahora se aplica el factor de corrección tomando $n_0 = 617$ y $N = 800$:

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{617}{1 + \frac{617}{800}} = 348{,}34 \approx 349$$

De acuerdo con el resultado anterior, se requiere una muestra de 349 camiones para estimar la cantidad de kilómetros recorridos mensualmente con una confianza del 95% y con un error máximo de 30 km.

Ejercicio de revisión

El departamento de compras de una empresa grande desea estimar qué porcentaje de sus 600 proveedores ha actualizado su información, pues el mes pasado se envió una solicitud a todos los proveedores enviando un formulario para mantener actualizados todos los datos. ¿De qué tamaño debe ser la muestra si se quiere una confianza en la estimación del 95% y que el error no exceda el 5%?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Minitab para determinar el tamaño de la muestra

El paquete estadístico Minitab posee la capacidad de determinar el tamaño para estimar la media, la desviación estándar, la proporción y otros parámetros. Para emplear esta función debe dar clic en el menú **Estadísticas**. En ese debe darse clic a la opción **Potencia y tamaño de muestra**, y luego en el submenú se debe dar clic a **Tamaño de muestra para estimación**.

En el cuadro de diálogo se completan los campos. Primero se selecciona el parámetro, en este caso la media, después se indica la desviación estándar y el margen de error permitido. En el botón opciones se selecciona el nivel de confianza y se debe chequear la opción **Suponer que la desviación estándar de la población se conoce.**

Ejemplo

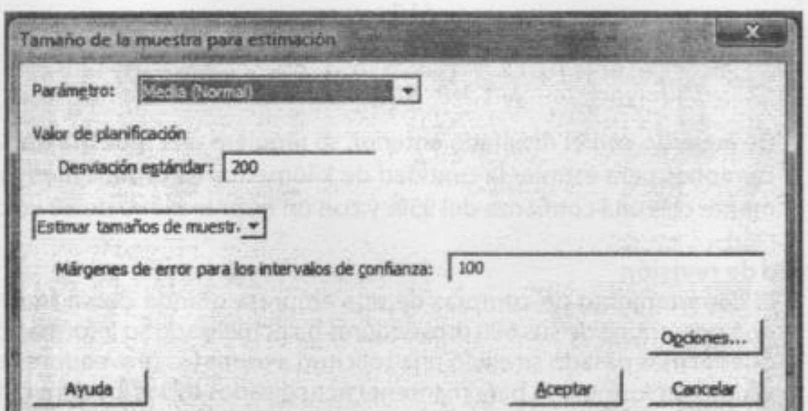
Un fabricante de impresoras desea estimar la cantidad promedio semanal de hojas de papel que se imprimen en distintas oficinas públicas del país. Por un estudio anterior se sabe que esta variable tiene una desviación estándar de 200 hojas. Además, se desea que la estimación tenga un error máximo de 100 hojas y con una confianza del 95%. ¿Qué tamaño de muestra se necesita?

Solución

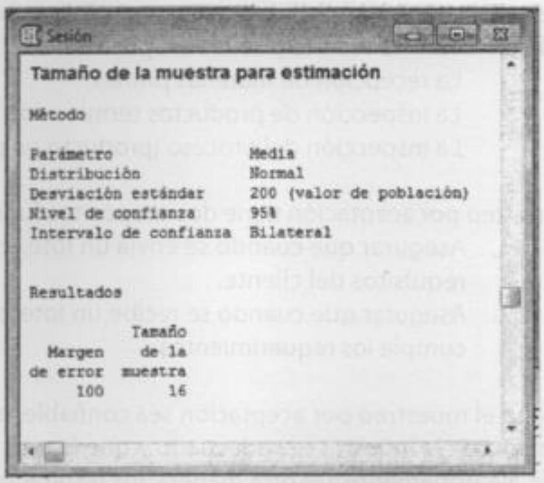
Con base en los datos anteriores, se plantea:

Desviación estándar de la población: $\sigma = 200$ hojas
Precisión deseada o nivel máximo de error permitido: $E = 100$ hojas
Nivel de confianza: 95%

Aplicando los valores en Minitab:



El software genera en resultado en Sesión:


Sesión

Tamaño de la muestra para estimación

Método

Parámetro	Media
Distribución	Normal
Desviación estándar	200 (valor de población)
Nivel de confianza	95%
Intervalo de confianza	Bilateral

Resultados

Margen de error	Tamaño de la muestra
100	16

De acuerdo con el resultado anterior, se requiere una muestra de 16 oficinas para efectuar la estimación con una confianza del 95% y con una discrepancia máxima entre el valor estimado y el valor real de 100 hojas.

Aplicación

Control de calidad

El aseguramiento de la calidad es un tema que ha tenido una gran evolución en las últimas décadas. Inicialmente el enfoque que las compañías le deban era la simple inspección final del producto, para luego tomar las acciones correctivas requeridas. Con el tiempo se adoptó un enfoque de proceso, ya que se observó que era necesario un enfoque más preventivo que correctivo, y que de esta manera eran más efectivos los esfuerzos de buscar e identificar las causas y los efectos de los problemas de calidad.

Posteriormente se va más allá, y el enfoque es el control total de la calidad y el aseguramiento de la calidad. Este enfoque busca que el producto o servicio se ajuste a su diseño, y este diseño responde a las necesidades y expectativas del consumidor.

Para lograr este aseguramiento de la calidad se requiere obtener y procesar información desde el diseño del producto o servicio, hasta su proceso de producción y su resultado final. Para esto es necesario tomar muestras. El muestreo estadístico es un valiosísimo recurso para conocer los procesos y para mejorarlos.

Una técnica común en el control de la calidad es el llamado muestreo por aceptación. Esta es una técnica estadística que permite determinar la calidad de un lote de materiales o de productos, basándose en una o más características de calidad que se estudian en una o más muestras de un lote. Esta técnica puede ser aplica en:

- La recepción de materias primas.
- La inspección de productos terminados.
- La inspección del proceso (producto en proceso).

El muestreo por aceptación tiene dos propósitos fundamentales:

1. Asegurar que cuando se envía un lote, este satisfaga los requisitos del cliente.
2. Asegurar que cuando se recibe un lote, se acepte solo si cumple los requerimientos.

Para que el muestreo por aceptación sea confiable, es necesario que el tamaño de la muestra sea adecuado y que la muestra se haya obtenido aleatoriamente. De otra manera, los resultados no serían adecuados para la toma de decisiones.

Otra técnica de uso muy frecuente en el control de la calidad es el gráfico de control. Un gráfico de control es una especie de intervalo de confianza que se elabora a lo largo del tiempo y que permite descubrir desajustes del proceso, identificar tendencias o puntos fuera de control, y buscar las causas de estos problemas de calidad. Cuando se efectúan mejoras en los procesos, se espera que el mismo gráfico presente su impacto.

La aplicación de los gráficos de control implica primero su construcción, para lo cual es necesario tomar muestras. Con base en las muestras se recolectan los datos y se elabora el gráfico. Una vez que se tiene el gráfico es necesario analizarlo y darle seguimiento para poder saber si realmente el proceso está controlado o no.

Determinación del tamaño de muestra para estimar la proporción poblacional

Muchas variables se miden en una escala nominal. Por ejemplo, cuando una empresa aseguradora desea saber sus clientes estarían dispuestos a adquirir un nuevo seguro de gastos médicos. En este caso se tiene una variable que no se puede medir en kilómetros, dólares, gramos u otra unidad de medida, sino que en una situación como esta lo que se hace es calcular una proporción. Es decir, la empresa estará interesada en saber qué proporción o porcentaje de sus clientes tendrán interés en adquirir el nuevo producto.

PROPORCIÓN Fracción, razón o porcentaje que indica la parte de la muestra o de la población que posee un rasgo de interés particular.

Si, por ejemplo, de una muestra de 50 clientes, 30 afirman estar dispuestos a adquirir el nuevo producto, entonces la proporción se obtendrá como el número de éxitos entre el número de elementos en la muestra. En esta situación esta proporción será 30/50, que equivale a 0,60 o 60%. Si se denota por p la proporción de la muestra, X es el número de éxitos y n es el tamaño de la muestra, entonces la proporción muestral se calcula como:

PROPORCIÓN MUESTRAL $$p = \frac{X}{n}$$

También se puede definir la proporción poblacional como el número de éxitos en la población, denotado por p y que, si se denota por P la proporción de la población, X es el número de éxitos en la población y N es el tamaño de la población, se calcula como:

PROPORCIÓN POBLACIONAL $$\pi = \frac{X}{N}$$

Determinantes del tamaño de la muestra

La determinación del tamaño de muestra para estimar la proporción poblacional es similar a la determinación del tamaño de muestra para estimar la media poblacional. Entonces hay que considerar tres elementos:

- **Precisión deseada o nivel máximo de error permitido**: Del mismo modo que en la determinación del tamaño, el investigador debe establecer un nivel de error máximo permitido. En este caso sería expresado como un porcentaje de divergencia entre el verdadero valor de la proporción p y el valor estimado p. A este nivel de error se le denotará por E.
- **Nivel de confianza que se desea tener**: Del mismo modo que cuando se determina el tamaño de muestra para estimar la media, el investigador debe indicar el nivel de confianza deseado para la estimación. Los niveles de confianza más usuales son del 95% y del 99%, cuyos valores de z son 1,96 y 2,58, respectivamente.
- **Aproximación de la proporción poblacional**: Se requiere conocer un valor aproximado de la proporción poblacional. Cuando no se conozca dicho valor, puede emplearse el valor de 0,5, pues es el valor que maximiza el tamaño de la muestra, dado un nivel de precisión y de confianza, y es mejor emplear una muestra más grande que una pequeña.

Cálculo del tamaño de la muestra

Para el cálculo del tamaño de muestra para estimar la proporción poblacional se tomarán en cuenta los factores descritos anteriormente los cuales conforman las fórmulas que a continuación se presentan, tanto para el caso de poblaciones infinitas como poblaciones finitas.

Caso de poblaciones infinitas En el caso de poblaciones infinitas, tal como se mencionó, el tamaño de la muestra depende de:

Precisión deseada o nivel máximo de error permitido (E)
Valor de z correspondiente al nivel de confianza que se desea tener (z)
Aproximación de la proporción poblacional (p)

Entonces estos tres elementos se incorporan en la siguiente fórmula:

FÓRMULA DEL TAMAÑO DE MUESTRA PARA LA PROPORCIÓN POBLACIONAL $$n = p(1-p)\left(\frac{z}{E}\right)^2$$

Ejemplo

Una compañía desea conocer el porcentaje de consumidores de ingresos medios y altos que estarían dispuestos a efectuar compras por internet en el transcurso de los próximos 6 meses. No se conoce ninguna estimación previa de este valor y se desea que la estimación tenga un error máximo de 3% y una confianza del 99%. ¿Qué tamaño de muestra se necesita?

Solución

Con base en los datos anteriores, se tiene que para el nivel de confianza del 99% corresponde un valor de z de 2,58. Además, como no se tiene una estimación de p, se empleará el valor de 0,5. Entonces se plantea:

Precisión deseada o nivel máximo de error permitido: $E = 0{,}03$
Valor de z correspondiente al nivel de confianza del 99%: $z = 2{,}58$
Aproximación de la proporción poblacional: $p = 0{,}5$

Sustituyendo los valores en la fórmula:

$$n = p(1-p)\left(\frac{z}{E}\right)^2 = 0{,}5(1-0{,}5)\left(\frac{2{,}58}{0{,}03}\right)^2 = 1.849$$

Es decir, es necesaria una muestra de 1.849 personas de ingresos medios y altos para efectuar una estimación del porcentaje de consumidores que estarían dispuestos a efectuar compras por internet en el transcurso de los próximos 6 meses, estimación que se realizará con una confianza del 99% y con un error máximo de 3%.

Ejercicio de revisión

Un candidato político requiere una estimación del porcentaje de electores que votaría por él en las próximas elecciones presidenciales. Desea que el error no exceda el 2,8% y una confianza del 95%. ¿Cuál debe ser el tamaño de la muestra?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Caso de poblaciones finitas

Del mismo modo que en la determinación del tamaño de muestra para estimar la media, la fórmula dada anteriormente se aplica a poblaciones infinitas, pero en poblaciones finitas cuando se trabaja con reemplazo o en poblaciones finitas cuando el tamaño de la muestra n es muy pequeño con respecto al tamaño de la población N puede realizarse un ajuste empleando el mismo factor de corrección:

FACTOR DE CORRECCIÓN PARA POBLACIONES FINITAS	$n = \dfrac{n_0}{1 + \dfrac{n_0}{N}}$

donde n_0 es el resultado dado por la fórmula $n = p(1-p)\left(\dfrac{z}{E}\right)^2$ y N es el tamaño de la población.

Ejemplo

Una empresa desea conocer la proporción de sus empleados que estarían de acuerdo en un nuevo programa de beneficios. La compañía tiene un total de 350 colaboradores y quiere hacer la estimación con un error máximo de 5% y una confianza del 95%. Se estima, por un estudio piloto, que esta proporción podría ser del 40%. ¿Qué tamaño de muestra se necesita?

Solución

Dado que la población es finita, entonces la determinación del tamaño de muestra se efectuará en dos etapas. Primero se calculará el tamaño de muestra como si la población fuera infinita. Luego se aplicará el factor de corrección para poblaciones finitas.

Con base en los datos anteriores, se tiene que para el nivel de confianza del 95% corresponde un valor de z de 1,96. Los datos del problema son:

Precisión deseada o nivel máximo de error permitido: $E = 0{,}05$
Valor de z correspondiente al nivel de confianza del 95%: $z = 1{,}96$
Aproximación de la proporción poblacional: $p = 0{,}4$
Tamaño de la población: 350 empleados

Sustituyendo los valores en la fórmula:

$$n = p(1-p)\left(\frac{z}{E}\right)^2 = 0{,}4(1-0{,}4)\left(\frac{1{,}96}{0{,}05}\right)^2 = 368{,}79 \approx 369$$

Ahora se aplica el factor de corrección tomando $n_0 = 369$ y $N = 350$:

$$n = \frac{n_0}{1+\dfrac{n_0}{N}} = \frac{369}{1+\dfrac{369}{350}} = 179{,}58 \approx 180$$

Es necesaria una muestra de 180 empleados para tener una estimación de la proporción de empleados que estarían de acuerdo en un nuevo programa de beneficios con una confianza del 95% y con un error máximo del 5%.

Ejercicio de revisión

Un investigador está investigando la prevalencia de diabetes en adultos mayores de 30 años en una población de 2000 personas. Desea un nivel de confianza de 95% y un error máximo de 3,5% en su estimación. ¿De qué tamaño debe ser su muestra?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

4.7 Técnicas de muestreo

Tal como se ha expuesto a lo largo de este capítulo, cuando se toma una muestra es necesario tomar ciertas decisiones. Una de ellas es el tamaño de la muestra, pero otra muy importante es establecer el modo para escoger cuáles elementos se van a incluir en la muestra. Es por eso que se habla de diferentes técnicas de muestreo.

La decisión sobre la técnica de muestreo a emplear depende de varios factores, entre los que están la existencia de un marco muestral bien definido, la conveniencia, el costo, la ubicación geográfica de las unidades de estudio, entre otros.

La selección de la técnica apropiada es tan importante como la determinación del tamaño de la muestra, pues de nada valdría tener un tamaño de muestra adecuado si la escogencia de los elementos que integran la muestra genera algún tipo de sesgo. Por ejemplo, si se va a efectuar un estudio de mercado para determinar la demanda de una nueva crema dental, sería un grave problema que todos los miembros de la muestra, o una gran mayoría, sean del mismo sexo, o de la misma zona geográfica, o de la misma clase socioeconómica, etc.

A continuación se explicarán algunas de las principales técnicas de muestreo aleatorias. Aunque el más sencillo es el muestreo simple al azar, los más empleados son el muestreo sistemático, el muestreo estratificado y el de conglomerados, entre otras técnicas importantes.

Importancia del marco muestral

Cuando se desea seleccionar una muestra aleatoria es muy importante contar con un marco muestral apropiado. El marco muestral es una lista, mapa u otro tipo de herramienta que permite identificar y seleccionar cada una de las unidades de muestreo. Por ejemplo, el marco muestral puede ser una lista con todos los empleados de una empresa, de modo que al seleccionar la muestra se pueda escoger a una cantidad de ellos.

En otros casos el marco muestral puede ser un mapa de una determinada zona geográfica, el cual se emplea para identificar y seleccionar viviendas, esto en un estudio sobre viviendas, o en un estudio donde la unidad estadística sea la familia, asumiendo que en cada vivienda va a habitar una familia.

En otras situaciones el marco muestral puede ser una guía telefónica, listados de distintos tipos, archivos, registros, entre otros.

Muestreo simple al azar

La técnica de muestreo más sencilla es la del muestreo simple al azar. Esta técnica es importante por su facilidad y porque es la base para otras técnicas de muestreo.

MUESTREO ALEATORIO SIMPLE Técnica de muestreo en la que cada miembro de la población tiene la misma probabilidad de ser seleccionado en la muestra.

Suponga que una compañía desea conocer cuál es la opinión de sus empleados acerca de la posibilidad de impartir cursos de capacitación a través de internet. Para esto desea aplicar una encuesta a una muestra de 60 empleados. El departamento de recursos humanos ha elaborado una lista de los 670 empleados que laboran en la empresa a la fecha, la cual servirá de marco muestral. En la lista a cada empleado se le ha otorgado un número, iniciando en 1 y terminando en 670. Para elegir a las 60 personas de la muestra, empleando un muestreo simple al azar, se requiere un mecanismo que le dé a cada uno de los 670 empleados la misma probabilidad de ser incluido en la muestra.

Una alternativa echar en una caja 670 papelitos, cada uno con un número distinto entre 1 y 670, los cuales se mezclan bien, y sin mirar el número, se seleccionan 60 papelitos. Después se toma el número indicado en cada papelito y de ese modo se selecciona la persona correspondiente en la lista elaborada por recursos humanos.

Una alternativa más conveniente es emplear una **tabla de números aleatorios**. Las tablas de números aleatorios, como la del apéndice 10 de este libro, es un cuadro con números que han sido generados mediante un proceso aleatorio. El cuadro siguiente muestra un ejemplo.

Para emplear la tabla de números aleatorios se requiere un punto de inicio aleatorio. Para obtenerlo usted puede simplemente cerrar los ojos y dejar caer el lápiz sobre la tabla y tomar como punto inicial el que marque la punta del lápiz.

Suponga en la tabla siguiente que el lápiz cayó en la segunda columna y la quinta fila, donde se encuentra el número 61722. Como usted necesita números entre 1 y 670, entonces puede tomar los primeros 3 dígitos de ese número, o sea, 617. Por tanto, selecciona al empleado #617 en la lista. A partir de ahí puede avanzar en la tabla hacia la derecha o hacia abajo, pero siempre de la misma manera. Entonces, si se avanza hacia la derecha, el segundo número seleccionado es el 291, el tercero es el 013, es decir, 13, el cuarto es el 130 (nótese que el elemento 891 está fuera del rango buscado, pues la empresa solo posee 670 empleados), y el quinto es el 357. Del mismo modo se continúa en la siguiente línea hasta completar los 60 miembros de la muestra.

49854	79596	24816	84416	32069	07091	99351
37273	14128	01981	93529	25802	50894	23045
52783	53435	55180	36802	99787	86515	17740
97021	08077	11064	81625	11897	79822	75449
81844	61722	29167	01368	89148	13062	35705
56669	33492	74076	77256	45475	30746	00852
32116	37619	53054	24127	00963	29832	70602
44715	90371	01674	25413	70020	68679	34848

Los números aleatorios también pueden generarse en Excel, Minitab u otros paquetes de software, e incluso calculadoras científicas.

Ejemplo

Un auditor desea verificar si todos los cheques emitidos en una compañía satisfacían los requerimientos de control establecidos por la gerencia. Durante el mes pasado se emitieron 81 cheques y la muestra debe contener 10 cheques. ¿Cuáles cheques se seleccionarían si se aplica un muestreo simple al azar y usando la tabla de números aleatorios anterior (tome como punto inicial la primera columna y segundo renglón de la tabla)?

Solución

Para seleccionar la muestra de 10 cheques de acuerdo con un muestreo simple al azar, se toman 10 números aleatorios. De acuerdo con lo establecido en el ejercicio, el punto inicial sería el número 37273. Como solo se extendieron 98 cheques, que es un número de 2 dígitos, entonces se requieren 10 números de 2 cifras entre 1 y 81. En la tabla dada estos número serían 37, 14, 01, 25, 50, 23, 52, 53, 55 y 36 (note que el 93 está fuera del rango requerido).

Ahora se buscan los cheques con los 10 números seleccionados y el auditor realiza su verificación.

Ejemplo

Utilice Excel para generar una muestra simple al azar de 5 unidades de una población total de 20 unidades.

Solución

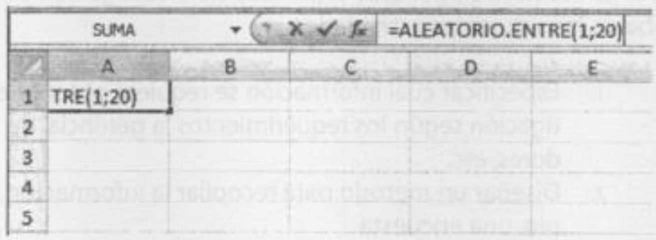
Tal como se ha mencionado en el muestreo es necesario generar números aleatorios. Las funciones ALEATORIO y ALEATORIO.ENTRE se pueden emplear para generar números aleatorios para realizar el muestreo.

Utilice Excel para generar números aleatorios para seleccionar una muestra de tamaño 5 de una población total de tamaño 20.

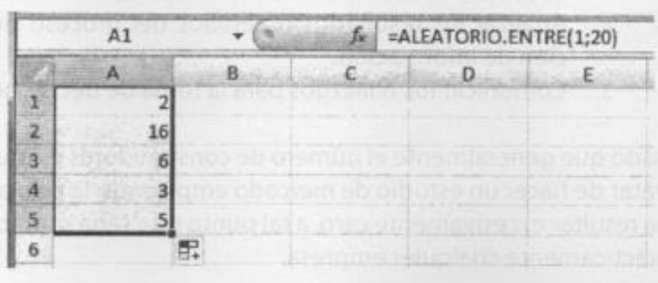
Para resolver este ejercicio se va a emplear la función ALEATORIO.ENTRE la cual genera números aleatorios entre dos valores, un límite inferior y otro superior, que en este caso serían 1 y 20, respectivamente, pues se desea obtener números al azar entre 1 y 20, porque la población es de tamaño 20.

En una celda de la hoja de Excel, por ejemplo, la celda A1 introduce la función:

= ALEATORIO.ENTRE(1;20)



Y presiona la tecla Intro (Enter). Como se quiere una muestra de tamaño 5, entonces se copia la fórmula 5 veces:



En este caso se seleccionarían, según la imagen, los elementos 2, 16, 6, 3 y 5 de la población para conformar la muestra.

Aplicación

Investigación de mercados

Cada día las empresas adoptan un mayor enfoque hacia el cliente, ya que los gerentes de estas organizaciones se han dado cuenta de lo necesario e importante que es adoptar un enfoque hacia el cliente, de modo que estos clientes logren una muy alta satisfacción con los productos y servicios de la empresa, y por tanto, se logre cierto grado de lealtad o fidelidad del cliente hacia la empresa. En épocas de crisis y alta competencia esta lealtad del cliente es clave para el éxito de un negocio, ya que los estudios muestran que es más económico retener a los clientes actuales que alcanzar nuevos clientes. Para lograr esto es necesario conocer a fondo a los consumidores, específicamente saber cuáles son sus necesidades y cómo las satisfacen, y cuáles son los factores que les motivan a adquirir determinados bienes y servicios.

La investigación de mercados es una de las principales *herramientas* a disposición de las empresas para obtener información que sirve para identificar y definir las acciones de marketing de una compañía u organización. En general, el proceso de investigación de mercados abarca las siguientes actividades:

1. Especificar cuál información se requiere obtener en la investigación según los requerimientos la gerencia, de los vendedores, etc.
2. Diseñar un método para recopilar la información, por ejemplo, una encuesta.
3. Administrar e implementar el proceso de recolección de información (que podría consistir en la aplicación de la encuesta a una muestra de consumidores).
4. Analizar los resultados obtenidos del proceso de recolección de información.
5. Comunicar los hallazgos para la toma de decisiones.

Dado que generalmente el número de consumidores es muy grande, tratar de hacer un estudio de mercado empleando la población puede resultar excesivamente caro, a tal punto que sería prohibitivo para prácticamente cualquier empresa.

Muestreo aleatorio sistemático

Este tipo de muestreo es muy sencillo de efectuar. Se emplea principalmente porque en algunos casos el muestreo simple al azar produce ciertas *concentraciones* en algunas partes del marco muestral, o bien, espacios de la población sin representación. También se puede emplear en casos en que el marco muestral no es completo o no se conoce exactamente la población.

MUESTREO ALEATORIO SISTEMÁTICO Técnica de muestreo que selecciona un punto aleatorio de inicio y posteriormente se elige cada k-ésimo elemento de la población.

Para emplear este tipo de muestreo es necesario determinar una constante de espaciamiento k:

CONSTANTE DE ESPACIAMIENTO $$k = \frac{N}{n}$$

Donde N es el tamaño de la población y n es el tamaño de muestra seleccionado.

Entonces, el muestreo sistemático consiste en seleccionar a la k–ésima unidad de muestreo después de un punto de *inicio* aleatorio entre 1 y k. En otras palabras, se selecciona al azar una unidad entre las primeras k unidades, y luego, después de esa unidad, se toma a la k–ésima unidad, y así sucesivamente hasta completar las n unidades de la muestra.

Ejemplo

Una empresa tiene 700 empleados y se desea tomar una muestra de 20 de ellos para aplicar un cuestionario sobre la opinión de los colaboradores sobre los resultados obtenidos luego de la implementación de un nuevo sistema informático. ¿Cómo se seleccionarían los miembros de la muestra si se emplea el muestreo aleatorio sistemático?

Solución

Para poder seleccionar la muestra es necesario que previamente se haya preparado una lista con los nombres de los 700 empleados, la cual servirá de marco muestral.

Como la población es de 700 personas, $N = 700$, y se tomará una muestra de 20 empleados, $n = 20$, entonces:

$$k = \frac{N}{n} = \frac{700}{20} = 35$$

Luego se busca el punto de inicio. Para esto se busca un número aleatorio entre 1 y 35. Suponga que se ha empleado una tabla de números aleatorios y que se ha obtenido el 8. Entonces, se selecciona al octavo empleado de la lista. Ese sería el primer integrante de la muestra.

Para obtener el segundo elemento en la muestra, al 8 se le suma la constante *k*, es decir, se le suma 35, por lo que el segundo miembro de la muestra será el número 43 de la lista, pues 8 + 35 = 43.

De manera similar se obtendrá el tercer elemento en la muestra. A 43 se le suma la constante *k*, de modo que se seleccionará al empleado número 78 de la lista, ya que 43 + 35 = 78.

Del mismo modo se seleccionarán los siguientes miembros de la muestra:

Elemento de la muestra	Elemento seleccionado de la población	Elemento de la muestra	Elemento seleccionado de la población
1	8	11	323 + 35 = 358
2	8 + 35 = 43	12	358 + 35 = 393
3	43 + 35 = 78	13	393 + 35 =428
4	78 + 35 = 113	14	428 + 35 = 463
5	113 + 35 = 148	15	463 + 35 = 498
6	148 + 35 = 183	16	498 + 35 = 533
7	183 + 35 = 218	17	533 + 35 = 568
8	218 + 35 = 253	18	568 + 35 = 603
9	253 + 35 = 288	19	603 + 35 = 638
10	288 + 35 = 323	20	638 + 35 = 673

De ese modo vemos como se ha seleccionado a los 20 miembros de la muestra. Luego se buscan los nombres respectivos en la lista y se aplica el cuestionario a cada uno de ellos.

Cuando la constante de espaciamiento *k* no da un número entero, puede recurrirse al redondeo. Por ejemplo, si de una población de 700 personas se desea seleccionar una muestra de tamaño 40, entonces $k = 700/40 = 17{,}5$. Como 17,5 no es entero, entonces se puede seguir el siguiente procedimiento. Se multiplica *k* por 10, para obtener un entero (si *k* tuviera 2 decimales, se multiplica por 100 para que quede un entero), que en este caso es 175. Después se elige un número aleatorio entre 1 y 175. Suponga que en la tabla de números aleatorios se obtiene el 83. Ahora se repiten los mismos pasos que se efectuaron cuando *k* es entero; es decir, se va sumando la constante de espaciamiento *k* para ir obteniendo cada nuevo elemento de la muestra:

83
83 + 175 = 258
258 + 175 = 433
433 + 175 = 608
...
6650 + 175 = 6825

Posteriormente se divide cada uno de estos números entre 10 (ya que se había multiplicado por 10) y se redondea cada resultado:

Primer elemento de la muestra:	83 , 10 = 8,3 » 8
Segundo elemento de la muestra:	258 , 10 = 25,8 » 26
Tercer elemento de la muestra:	433 , 10 = 43,3 » 43
Cuarto elemento de la muestra:	608 , 10 = 60,8 » 61
…	
Último elemento de la muestra:	6825 , 10 = 682,5 » 683

Entonces, se selecciona a los elementos número 8, 26, 43, 61, …, 683 de la población para ser incluidos en la muestra.

Puede observarse que el muestreo sistemático es muy sencillo, especialmente si el tamaño de muestra es muy grande. En general, puede considerarse que es fácil tomar al *k*–ésimo nombre de una lista, seleccionar al *k*–ésimo artículo en una línea de producción, la *k*–ésima persona esperando en una fila en un aeropuerto, etc.

Ahora bien, hay que tener también algunas precauciones. Si la lista que sirve de marco muestral sigue algún patrón o ciclo particular, entonces la muestra obtenida podría no ser representativa. Si, por ejemplo, la lista de los 700 empleados ha sido ordenada siguiendo el orden hombre – mujer, entonces podría obtenerse una muestra de solo mujeres, lo cual haría que la muestra no fuera representativa, pues la población también contiene hombres.

Muestreo aleatorio estratificado

En ocasiones es posible agrupar las unidades de estudio que se *parecen* en algún aspecto. Estos grupos o subpoblaciones se denominan estratos. Suponga que una empresa puede dividir a sus empleados por departamentos, por ejemplo, personal del área de finanzas, del área de mercadeo y del área de operaciones; o bien, por su antigüedad de laborar en la empresa, donde se podría tener un grupo de empleados con menos de 5 años de trabajar para la compañía, otros con 5 años o más pero menos de 10 años en la empresa, y un tercer grupo correspondiente a los empleados con 10 o más años en la empresa. Entonces, en un muestreo estratificado se elabora un diseño de muestra en el que se selecciona una muestra al azar de cada uno de los estratos establecidos. En cada estrato se va a estimar la característica de interés y luego se combinarán dichas estimaciones para obtener la estimación para toda la población.

MUESTREO ALEATORIO ESTRATIFICADO Técnica de muestreo que divide la población en subgrupos, denominados estratos, y selecciona al azar una muestra de cada estrato.

Este tipo de muestreo es muy útil en poblaciones muy heterogéneas, las cuales pueden dividirse en estratos más homogéneos gracias al conocimiento previo que el investigador tiene acerca de la población.

Tal como se mencionó, los estratos pueden ser de muchos tipos. Por ejemplo, en un estudio realizado en una universidad, cada uno de los estratos podría corresponder a los estudiantes de cada una de las carreras que se imparten, y además, otro estrato podría estar compuesto por el personal docente y otro por el personal administrativo. Este tipo de muestreo aseguraría que en la muestra va a haber representación para cada uno de los estratos, lo cual no necesariamente ocurre con el muestreo simple al azar.

Como se ha indicado, en esta técnica de muestreo se procura formar estratos más homogéneos que la población (o sea, con una desviación estándar dentro de cada estrato menor que la poblacional), lo cual permitirá realizar estimaciones con un error estándar inferior, y por tanto, se lograría una mayor precisión que la que se obtendría con un muestreo simple al azar con el mismo tamaño de muestra.

Las principales ventajas del muestreo estratificado son:

- Se puede obtener estimaciones más precisas de la característica que se desea medir en el tanto que los estratos sean más homogéneos con relación a esa característica de interés.
- Los estratos pueden ser importantes por sí mismos, es decir, la estratificación no solo es útil para estimar el parámetro poblacional buscado, sino también para conocer a profundidad cada estrato, e incluso, poder hacer comparaciones entre ellos.
- En algunos casos este tipo de muestreo puede ser muy conveniente desde el punto de vista administrativo, pues la recolección de la información puede facilitarse a través de la estratificación. Por ejemplo, en el caso anterior de la universidad, se pueden emplear distintos métodos para cada estrato, así, para obtener la información entre los estudiantes se puede utilizar un método, pero entre los profesores y personal administrativo otro distinto.

Luego de que se han definido los estratos, entonces es necesario efectuar la afijación. La afijación es la distribución del total de la muestra entre cada uno de los estratos establecidos.

La afijación puede hacerse de distintas maneras. La más común es la denominada afijación proporcional. Esto se hace distribuyendo la muestra entre los estratos proporcionalmente al número de unidades de la población que corresponden a cada estrato.

Cabe aclarar que esta imagen solo ilustra la conformación de la muestra a partir de los estratos ya establecidos. No ilustra el modo en que deben ser establecidos los estratos en la población.

Ejemplo

Los empleados de una empresa se pueden dividir en estratos por su antigüedad de laborar en la compañía. Del total de 1.000 empleados, hay 200 empleados con menos de 5 años de trabajar para la compañía, hay 500 con una antigüedad de 5 años o más pero menos de 10 años en la empresa, y 300 con una antigüedad de 10 o más años. Se va a seleccionar una muestra de 50 empleados para conocer la opinión de los empleados sobre la posibilidad de implementar la modalidad del teletrabajo en la empresa. ¿Cuántos empleados deben seleccionarse de cada estrato?

Solución

Para establecer cuántos empleados deben seleccionarse de cada uno de los estratos establecidos, si se emplea la afijación proporcional, primero se debe determinar la frecuencia relativa de cada uno de los estratos en la población:

Estrato	Antigüedad	Número de empleados	Frecuencia relativa	Muestra por estrato
1	Menos de 5 años	200	0,20	10
2	De 5 a 10 años	500	0,50	25
3	10 años o más	300	0,30	15
	Total	1.000	1,00	50

Tal como se observa en la tabla, para obtener la frecuencia relativa de cada estrato se divide el número de elementos del estrato entre el total de la población:

Estrato 1: 200/1.000 = 0,20
Estrato 2: 500/1.000 = 0,50
Estrato 3: 300/1.000 = 0,30

Observe que la suma de las frecuencias relativas debe ser exactamente uno.

Luego para determinar el número de empleados que se incluirán en la muestra por cada estrato se multiplica cada frecuencia relativa por el tamaño de muestra, que en este caso es 50:

Estrato 1: $0{,}20 \times 50 = 10$
Estrato 2: $0{,}50 \times 50 = 25$
Estrato 3: $0{,}30 \times 50 = 15$

La suma de los tamaños de muestra por estrato debe ser igual al tamaño de la muestra total, que en este caso es 50.

Podemos decir que se requiere incluir en la muestra a 10 empleados con una antigüedad de menos de 5 años, a 25 con una antigüedad de más de 5 años pero menos de 10 años en la empresa, y a 15 con una antigüedad de 10 o más años de laborar para la empresa.

La afijación proporcional no es la única forma de establecer la distribución de la muestra entre los distintos estratos, pero sí es la más sencilla y práctica, además de que concede a todos los miembros de la población la misma probabilidad de ser incluidos en la muestra, tal como en el muestreo simple al azar. Además, los cálculos son más sencillos al no requerirse factores de ponderación.

Muestreo aleatorio por conglomerados

Con mucha frecuencia se tiene una población dispersa geográficamente y no se posee un marco muestral bien establecido. Por ejemplo, si se desea efectuar una investigación entre estudiantes de secundaria en todo el país, los cuales se encuentran dispersos en todo el territorio nacional y, además, no se posee una lista de todos los estudiantes matriculados en todos los colegios del país. Entonces, el muestreo por conglomerados es una buena alternativa para seleccionar la muestra.

> **MUESTREO ALEATORIO POR CONGLOMERADOS** Técnica de muestreo que divide la población en conglomerados a partir de sus límites naturales geográficos o de otra clase, para luego seleccionar aleatoriamente conglomerados y tomar una muestra al azar con elementos de cada uno de ellos.

Observe que en este tipo de muestreo inicialmente no se seleccionan las unidades de estudio, sino que se seleccionan *unidades primarias*, es decir, se seleccionan primero regiones, o barrios, o colegios, entre otros posibles conglomerados. Cabe señalar que no todos los conglomerados son muestreados en esta técnica.

Este tipo de muestreo es empleado por distintas razones:

- En ciertos casos no se posee un marco muestral que indique todas las unidades de estudio, pero sí se puede contar con una lista de los conglomerados. Por ejemplo, tal vez no se tenga una lista de todos los empleados de las micro y pequeñas empresas del país, pero sí podría tenerse una lista de dichas empresas. Así podría seleccionarse al azar una muestra de las empresas, y en cada empresa obtener una muestra al azar de algunos de sus empleados.
- Este tipo de muestreo permite disminuir los costos, especialmente en cuanto al recurso humano. El hecho de que las unidades estén agrupadas dentro de los conglomerados es lo que facilita la reducción de los costos.
- Las unidades individuales dentro de cada "conglomerado", por lo general, tienden usualmente a ser similares. Por ejemplo, las familias de altos ingresos viven en el mismo barrio, mientras que las familias de clase media vivirían en otra zona.

Empleando el muestreo de conglomerados, generalmente, se produce un menor nivel de precisión (mayor error muestral) que en una muestra simple al azar del mismo tamaño. Frecuentemente la variación entre los elementos obtenidos de los conglomerados seleccionados es mayor que en las muestras aleatorias simples. Esta debilidad puede enfrentarse si se incrementa el tamaño de la muestra de cada conglomerado.

Ejemplo

En un proyecto de investigación se desea conocer el grado de satisfacción laboral de los profesores universitarios del país. Se requiere una muestra total de 300 profesores, pero no se posee una lista de todos los profesores de las universidades del país. ¿Cómo podría obtenerse la muestra en este estudio?

Solución

Dado que ya está establecido el tamaño de la muestra, el problema consiste en seleccionar los 300 miembros de la muestra. Como no se cuenta con un marco muestral, podría emplearse el muestreo por conglomerados. Para ello se toma una lista de las universidades del país. Cada universidad será una unidad primaria, es decir, se seleccionará una muestra aleatoria de varias universidades, para luego tomar una muestra de profesores de cada una de ellas. Cada muestra de profesores se puede obtener por muestreo simple al azar o muestreo sistemático.

Muestreo no aleatorio

Existen otras técnicas de muestreo aleatorio como el muestreo por etapas (multietápico), muestreo proporcional al tamaño, entre otros. De igual manera existen diversos tipos de muestreo no aleatorio, tales como:

- **Muestreo de juicio**: En este tipo de muestreo las unidades son seleccionadas mediante el juicio personal del investigador. Esta persona usualmente es un experto en la materia relacionada con la investigación. Este tipo de muestra no es una muestra probabilística, pues este método está basado en el punto de vista subjetivo de una persona, lo cual ocasiona que la teoría de la probabilidad no pueda ser aplicada para medir el error de muestreo.
- **Muestreo por conveniencia**: En este caso el investigador selecciona para conformar la muestra aquellas unidades que sean más fáciles de acceder. En otras palabras, la muestra se elige de acuerdo con la comodidad del investigador. En este tipo de muestreo la representatividad la determinaría el investigador de modo subjetivo, por lo que tampoco se puede emplear la teoría de la probabilidad para determinar el nivel de error.
- **Otros tipos de muestreo no aleatorio**. En ocasiones se habla del muestreo por cuotas, de las muestras autogeneradas, entre otras formas de obtener una muestra no probabilística. Todos estos métodos tienen la desventaja de que no es posible establecer la representatividad de la muestra de forma objetivo, es decir, no se puede determinar un nivel de confianza y medir el margen de error de la estimación.

Aplicación

Auditoría

Aunque existen muchos tipos de auditoría, en este capítulo se hará énfasis en la auditoría financiera. La información financiera es fundamental para la toma de decisiones tanto para usuarios internos de la empresa, la gerencia, por ejemplo, como para usuarios externos, como los acreedores, el fisco e inversionistas, entre otros. Es por eso que es fundamental que los estados financieros de la empresa sean fiel reflejo de la realidad económica y financiera de una empresa.

La auditoría juega un papel muy importante como un instrumento para garantizar, hasta donde sea posible, que la información financiera sea veraz y confiable. En el ámbito financiero la auditoría consiste en una revisión objetiva de los estados financieros que han sido elaborados en una organización, sea esta una empresa privada, una entidad sin fines de lucro o una institución pública, entre otras. Entonces, dentro del trabajo del auditor está el obtener las evidencias suficientes para poder ofrecer una opinión objetiva sobre la veracidad y confiabilidad de la información financiera presentada.

Dado que una empresa efectúa miles, e incluso hasta millones de transacciones, un auditor no puede examinar todas y cada una de estas operaciones que son las que forman los saldos finales. Es por eso que debe aplicar los procedimientos de auditoría a una muestra representativa de estas transacciones o partidas para obtener la evidencia que requiere.

Entonces, el muestreo en la auditoría es el proceso de selección de una muestra entre un grupo más grande de partidas (población), de modo que se emplean las características de la muestra para llegar a conclusiones sobre las características de la población.

La normativa que rige el trabajo del auditor hace énfasis en la necesidad del uso de muestreo estadístico. Las evidencias se recolectan a partir de muestras representativas, de modo que las inferencias que se efectúen posean altos niveles de confianza. Por eso debe tenerse cuidado en la determinación del tamaño de la muestra y la selección de las unidades de observación que se seleccionan. Todo el procedimiento estadístico empleado debe justificarse y debe permitir al auditor valorar el nivel de riesgo que existe en su interpretación de los datos muestrales y un cierto grado de confiabilidad en la interpretación de los resultados.

Entre las principales ventajas del muestreo en la auditoría están las siguientes:

1. Permite que el criterio del auditor se base en cifras objetivas, no en juicios de valor.
2. Reduce el costo de la auditoría al no tener que examinar el 100% de los datos, registros, documentos, etc.
3. Reduce el riesgo de la auditoría, pues al trabajar con una muestra el auditor puede asegurarse de controlar mejor su trabajo.
4. En general puede decirse que el proceso de la auditoría es más eficaz y eficiente, pues puede cumplir su propósito y a un costo menor.

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

4.8 Ejercicios

Ejercicios de desarrollo

Resuelva los ejercicios siguientes (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Un investigador está interesado en conocer ciertos datos relacionados con la calidad de vida de un pueblo en una zona rural del país. Entre las variables que desea estimar está el nivel de calorías que en promedio consumen los niños de 7 a 12 años de esa población por día. Conoce, por un estudio anterior, que la desviación estándar de la citada variable es 185 calorías. Además sabe que habitan en ese poblado 745 niños. Según datos del gobierno local el consumo medio de calorías de los niños de esa edad es de 1170 calorías por día, pero el investigador duda de la veracidad de dicha información. El desea que sus estimaciones tengan altos niveles de confiabilidad (por lo menos el 99%), entonces decide realizar un estudio estadístico. Toma una muestra de 27 niños y evalúa su situación nutricional.
   a) ¿Cuál podría ser el marco muestral a emplear?
   b) ¿Cuál sería la unidad de estudio, la unidad de muestreo y la información?
   c) ¿Por qué necesita este investigador emplear una muestra?
   d) ¿Cuál es el parámetro que el investigador desea estimar?
   e) ¿Qué tipo de muestreo debería emplear, un muestreo aleatorio o no aleatorio?
   f) ¿Tomó el tamaño de muestra adecuado? Explique.

2. Una asociación de profesionales desea fomentar el uso de las tecnologías entre sus afiliados. Para lograr dicho objetivo, se quiere establecer un convenio con alguna empresa vendedora de equipo de cómputo, para posteriormente poder un sistema de financiamiento para los miembros de la asociación. A fin de poder realizar la negociación, primero se quiere estimar la proporción de profesionales afiliados que estarían interesados en adquirir una computadora, tanto para el trabajo como para el hogar. Por esta razón, solicita un estudio por muestreo.
   a) ¿Cuál podría ser el marco muestral a emplear?
   b) ¿Cuál sería la unidad de estudio, la unidad de muestreo y la información?
   c) ¿Por qué necesita este investigador emplear una muestra?
   d) ¿Cuál es el parámetro que el investigador desea estimar?
   e) ¿Qué tipo de muestreo debería emplear, un muestreo aleatorio o no aleatorio?

3. En los siguientes casos, ¿cuál puede ser la razón principal por la cual se va a emplear una muestra?
   a) Una fábrica de refrescos desea conocer si el nivel de azúcar en los refrescos que se producen está dentro de los niveles aceptables.
   b) Un banco desea evaluar la satisfacción de sus clientes.
   c) Una tienda en línea desea saber la proporción de usuarios del internet que creen que es seguro efectuar pagos en línea.
   d) Un gestor ambiental desea conocer el nivel de determinados contaminantes que se encuentran en las aguas de un río.

4. En los siguientes casos, ¿cuál puede ser la razón principal por la cual se va a emplear una muestra?
   a) Un supervisor de calidad desea saber la cantidad de piezas defectuosas que salen de un proceso de producción.
   b) Un distribuidor de ropa femenina desea conocer la proporción de mujeres entre 20 y 40 años que reconocen las marcas que vende.
   c) Una universidad desea conocer cuáles son las destrezas principales que deben poseer los profesionales en administración de negocios que contratan las empresas.
   d) Una compañía de seguros desea saber la proporción de personas que poseen algún tipo de seguro para vehículos.

5. Se desea realizar un estudio para determinar los gastos familiares mensuales promedio en educación que efectúan los empleados de una compañía. El gerente desea tener un 95% de confianza de que los gastos no difieran en más de $40 con respecto al valor verdadero.
   a) En una encuesta previa se determinó que la varianza es de $6.400. ¿De qué tamaño deberá ser la muestra?
   b) ¿De qué tamaño debe ser la muestra si la empresa tiene un total de 5.500 empleados?
   c) ¿Cuáles especificaciones del estudio podrían cambiarse, si se deseara un tamaño de muestra menor? (Conteste sin cálculos, sólo con palabras)

6. Una asociación de profesionales desea fomentar el uso de las tecnologías entre sus afiliados. Para lograr dicho objetivo, se quiere establecer un convenio con alguna empresa vendedora de equipo de cómputo, para posteriormente poder un sistema de financiamiento para los miembros de la asociación. A fin de poder realizar la negociación, primero se quiere estimar la proporción de profesionales afiliados que estarían interesados en adquirir una computadora, tanto para el trabajo como para el hogar. Por esta razón, solicita un estudio por muestreo. ¿Cuál debería ser el tamaño de muestra, si el margen de error máximo permitido es 3,5%, la asociación posee 5.000 afiliados actualmente y se quiere un nivel de confianza del 95%?

7. El Presidente de la República desea una estimación del porcentaje de la población que apoya su política económica actual. El mandatario espera que la estimación esté dentro del 0.05 de la proporción verdadera, utilizando un nivel de confianza del 95%. ¿De qué tamaño debe ser la muestra?

8. El encargado de capacitación de una institución con 3.200 empleados desea hacer una encuesta que le sirva para detectar diversas necesidades de capacitación en el área de idiomas.
   a) Si desea que la estimación de la proporción de empleados que requieran cursos de inglés tenga un error máximo de 2 puntos porcentuales y una confianza del 99%, ¿de qué tamaño debe ser la muestra para llevar a cabo la encuesta?
   b) Si considerara el tamaño de muestra anterior muy grande con relación a los recursos que dispone, ¿en qué dirección (aumentar o disminuir) tiene que cambiar el nivel de confianza o el error muestral para disminuir el tamaño de la muestra?

9. Se realiza un estudio para estimar la proporción de residentes de una zona rural del país que están a favor de la construcción de una mina. ¿De qué tamaño debe ser la muestra, si se requiere una confianza de al menos de 95%, de que la estimación estará dentro del 0.04 de la proporción real de residentes de esta zona que están a favor de la construcción de la mina?

10. Un agente de seguros está realizando una encuesta entre gerentes de empresas públicas y privadas para determinar la proporción de ellos que estarían dispuestos a comprar seguros de gastos médicos y cuánto es el nivel de gastos mensuales promedio en salud que realizan. En una encuesta realizada hace un año se determinó que un 15% estarían anuentes a comprar seguros de gastos médicos y que la desviación estándar de los gastos mensuales en salud era de $145. Si se desea una confianza del 95% en las estimaciones:
    a) Calcule el tamaño de muestra para estimar la media de los gastos mensuales en salud con un margen de error máximo de $100.
    b) Calcule el tamaño de muestra para estimar la proporción de gerentes que comprarían un seguro de gastos médicos con un margen de error máximo de 3 puntos porcentuales.

11. Se desea realizar una encuesta a entre los 10.500 profesionales en mercadeo que hay en el país para preguntarles si desean asistir a un congreso sobre marketing relacional. Si en la encuesta se desea un margen de error máximo del 4% y si no se conoce algún estudio previo, determine el tamaño de muestra necesario para hacer la encuesta.

12. El comedor de un albergue infantil quiere estimar el peso promedio de los niños. Un estudio anterior de diez niños mostró que la desviación estándar de sus pesos es de 6,05 kilogramos. ¿Qué tan grande debe ser una muestra para que el administrador tenga un 95% de confianza de que el error de estimación es a lo más de 2 kilogramos?

13. El gerente de una empresa desea estimar las horas mensuales que trabajan los empleados de la empresa bajo presión, ya que la entrega del entregable de los proyectos que se realizan tiene que estar listo al final del día. Un estudio piloto reveló que la desviación estándar es de 24 horas.
    a) ¿De qué tamaño se necesita una muestra si se desea tener 96% de confianza que la media real esté dentro de un margen de error de 10 horas de la media real?
    b) ¿De qué tamaño debería ser la muestra si en lugar de tener un error de estimación de 10 horas sólo se requiere un error de 5 horas?
    c) Suponga que se tiene una población de 300 empleados, ¿de qué tamaño debe de ser la muestra, si se desea una confianza de 96% y un error máximo de 5 horas?

14. El gerente de recursos humanos de una organización está analizando una serie de nuevas políticas con el fin de mejorar el desempeño e incrementar la productividad. Por tal razón desea conocer el rendimiento de los 230 operarios de la empresa según la más reciente evaluación del desempeño. Para ello debe seleccionar una muestra y hacer una estimación del desempeño medio de los empleados. ¿De qué tamaño debe ser la muestra si se quiere tener una confianza de 90% de que el error no será mayor de 4 puntos, conociendo que la desviación estándar es de 22 puntos?

15. Una organización desea conocer cuántos de sus 3000 empleados están satisfechos en su puesto de trabajo. Para ello debe seleccionar una muestra y hacer una estimación de la proporción de empleados que si se sienten satisfechos. ¿De qué tamaño debe ser la muestra si se quiere tener una confianza de 95% de que el error no será mayor de 3 puntos?

16. En un estudio anterior entre 500 familias que tienen televisores en una ciudad, se encontró que 340 están suscritas a algún servicio de televisión por cable. ¿Qué tan grande se requiere que sea una muestra si se quiere tener 95% de confianza de que la estimación de la proporción de familias que posee algún servicio de cable esté dentro de 0,02?

17. La tabla siguiente contiene los nombres de 30 empleados del departamento de inversiones de una aseguradora. Se va a tomar una muestra de 6 de empleados. Con base en la información anterior, determine, si se emplea un muestreo simple al azar:
    a) ¿Cuáles empleados se seleccionarían si los números aleatorios que genera un paquete de software son 13, 22, 16, 08, 02 y 24?
    b) Ahora utilice la tabla de números aleatorios del apéndice 10 para generar su propia muestra.

#	Nombre del empleado	#	Nombre del empleado
1	Magally Castillo	16	Ricardo López
2	Andrés González	17	Andrés Méndez
3	Sergio Vargas	18	Daniela Rojas
4	Patricia Chinchilla	19	Javier Hidalgo
5	José Cordero	20	Alberto Aguilar
6	Melisa Pérez	21	Norman Monge
7	Ana Cristina Campos	22	Rolando Madriz
8	Victoria Fonseca	23	Felipe Salas
9	Sandra Gamboa	24	Verónica González
10	Vanesa Gómez	25	Laura Solano
11	Rolando Rojas	26	María Vargas
12	Esteban González	27	Gustavo Mejía
13	Mario Leiva	28	Álvaro Bustamante
14	Tatiana Fuentes	29	Luis Zúñiga
15	María Gamboa	30	Susan Alfaro

18. Tomando como base el ejercicio de 17, se extraerá una muestra de tamaño 6 empleando la técnica de muestreo aleatorio sistemático:
    a) ¿Cuáles elementos seleccionaría en la muestra si el punto inicial es el quinto elemento de la población?
    b) Ahora utilice la tabla de números aleatorios del apéndice 10 para generar su propia muestra.

19. Tomando como base el ejercicio de 17, se extraerá una muestra de tamaño 8 empleando la técnica de muestreo aleatorio sistemático:
    a. ¿Cuáles elementos seleccionaría en la muestra si el punto inicial es el segundo elemento de la población?
    b. Ahora utilice la tabla de números aleatorios del apéndice 10 para generar su propia muestra.

20. Una empresa tiene 2.000 empleados y desea tomar una muestra de 200 de colaboradores para aplicar un cuestionario para conocer su disposición a participar en comités de educación y gestión ambiental. Explique cómo se seleccionarían los miembros de la muestra si se emplea el muestreo aleatorio sistemático.

21. Un investigador desea conocer el nivel nutricional de los jóvenes que ingresan a las secundarias del país. Se requiere una muestra total de 3.600 estudiantes de primer año de secundaria (no se posee una lista de todos los estudiantes de secundaria del país). Explique cómo podría obtenerse la muestra en este estudio.

22. A continuación se presenta una lista de investigaciones que se realizarán empleando un muestreo estadístico (aleatorio), en cada caso indique cuál considera usted que es el diseño muestral más adecuado (simple al azar, sistemático, estratificado, conglomerados):
    a. En un hospital se desea determinar el tiempo medio de espera de los pacientes en el servicio de consulta externa.
    b. Se desea conocer la efectividad de un programa de prevención de enfermedades cardiovasculares aplicado en la población de la ciudad capital de julio a noviembre del año 2012.
    c. Un laboratorio farmacéutico ha desarrollado un nuevo medicamento y desea evaluar la posible interacción con otras sustancias, para lo cual posee una muestra de ratas en las cuales va a efectuarse el experimento.
    d. Un investigador está preocupado por el continuo uso de computadoras, televisores y otros aparatos y su efecto en la agudeza visual de los niños de 6 a 12 años. El estudio se realizará a nivel nacional.
    e. Como consecuencia del calentamiento global se considera que cierto tipo de ranas han ido desapareciendo, por lo que se requiere conocer cuántas ranas habitan aun en las zonas montañosas del país.
    f. Se requiere conocer qué porcentaje de la población del país alcanza ciertos niveles de peso, para saber cuánta es la prevalencia de la obesidad en el país durante el último año.
    g. Se va a determinar el estado de las viviendas del país. El estudio se efectuará de enero a marzo del próximo año.
    h. En un proceso de producción se requiere saber la proporción de artículos defectuosos, además de conocer las posibles causas de los defectos.
    i. Se desea saber la resistencia de determinados componentes electrónicos ante ciertas circunstancias extremas (sobrecargas, calentamiento, etc.).
    j. Se va a realizar un estudio sobre las medidas de seguridad informática que las empresas financieras implementan. Para esto se va a efectuar una encuesta a los gerentes de tecnologías de información de bancos, cooperativas, mutuales, etc.
    k. Ante el lanzamiento al mercado de un nuevo sistema operativo se va a estimar la proporción de empresas que están dispuestas a implementar este nuevo sistema en el transcurso de los próximos doce meses.
    l. Un inversionista está buscando nuevas opciones para colocar su capital, por lo que decide invertir en la Bolsa de Valores de Nueva York. Ha reunido información de distinta empresas para conocer cómo han variado los precios de sus acciones y los pagos de dividendos.
    m. Una nueva empresa de telecomunicaciones está efectuando un estudio de mercado para conocer las preferencias de los consumidores nacionales en lo que respecta al uso del teléfono celular, uso de internet, etc.

n. El encargado de recursos humanos de una empresa va a estimar el nivel de satisfacción de los empleados de la empresa para la cual labora.
o. En una auditoría se van a examinar las cuentas por cobrar de una empresa.
p. Se requiere conocer cómo afectaría un nuevo paquete de impuestos a los diferentes sectores productivos del país. Para tal fin se efectuará una encuesta entre empresarios, líderes de organizaciones empresariales y expertos en el tema.
q. El gobierno desea saber qué tan efectiva ha sido una campaña para promocionar al país como destino turístico. Para ello se aplicará un cuestionario a turistas que ingresan al país por los distintos aeropuertos y fronteras terrestres.
r. Una educadora está investigando el impacto de las tecnologías de la información y la comunicación en los hábitos de estudio de los estudiantes de secundaria del país.
s. Un psicólogo en una empresa aplica un test para identificar distintos tipos de personalidad de los candidatos a una serie de puestos vacantes en la empresa.
t. Se está investigando el efecto que tienen las redes sociales en internet en las relaciones familiares, para esto se tomará una muestra de familias del área metropolitana durante los próximos 6 meses.
u. Un investigador de una institución desea conocer el estado general (salud, emocional, etc.) de los niños dados en adopción durante el último año en el país.
v. El gobierno está analizando el efecto que han tenido algunas manifestaciones estudiantes en su imagen. Para esto va a realizar una encuesta a nivel de las zonas urbanas del país.
w. Un periodista está preocupado porque considera que algunas acciones del gobierno afectan la libertad de prensa. Va a realizar una encuesta aplicando a una muestra representativa de sus colegas para valorar la percepción que tiene de esos mismos sucesos.
x. Para determinar la efectividad de una campaña publicitaria, se va a evaluar la exposición de los consumidores a los comerciales pautados y el volumen de ventas de la compañía en el mismo periodo. Se tomará una muestra de consumidores del área metropolitana.

23. Suponga que se desea estimar la prevalencia de diabetes en la comunidad de Nuevo Volcán para luego poder someter a algunos de ellos en un plan especial del Seguro Social. ¿Cuál es el tamaño de la muestra requerido para estimar la prevalencia de diabetes en una población de 2.000 habitantes? Se sabe por un estudio previo que prevalencia de diabetes en esta población fue de 8%. Se desea que la proporción estimada no difiera en más de un 1% con respecto a la verdadera y una confianza del 99%.

24. ¿Cuál es el tamaño de la muestra requerido para estimar la hemoglobina promedio en gestantes en una población de 1800 gestantes? Se conoce que la desviación estándar en un estudio previo fue de 1,2 mg/dl y se desea una precisión de 0,7mg/dl.

25. Según una revista sobre salud, en una encuesta a 20 familias, se encontró que las primas anuales de seguros promediaron $10.979 con una desviación estándar de $1.000. ¿De qué tamaño debe ser la muestra debe ser la muestra para que la media poblacional se encuentre dentro de un margen de $250 con 99% de confianza?

26. Un artículo publicado en la British Medical Journal (http://www.bmj.com/content/345/bmj.e4737) cuestiona los estudios que respaldan el efecto supuestamente positivo de las bebidas hidratantes como Gatorade, Powerade y otras. En el artículo se indica que en 106 estudios realizados, el tamaño de muestra promedio era de 9 participantes y que en solo un estudio la muestra era superior a 100 deportistas.
    a. Si se quisiera estimar la proporción de deportistas para los cuales una bebida hidratante genera un efecto positivo con una confianza de 90% y error máximo de 10%, ¿de qué tamaño debería ser la muestra?
    b. Si se empleara una muestra de 100 deportistas, a un 95% de confianza, ¿cuál sería el nivel del error permitido en la estimación?

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. La unidad estadística es:
    (a) Una porción o parte de la población de interés
    (b) La unidad de interés en un estudio estadístico
    (c) La unidad básica en términos de la cual se aplica una técnica de muestreo
    (d) La unidad que proporciona los datos relacionados con la unidad de estudio
2. La unidad de información en un estudio es:
    (a) Una porción o parte de la población de interés
    (b) La unidad de interés en un estudio estadístico
    (c) La unidad básica en términos de la cual se aplica una técnica de muestreo
    (d) La unidad que proporciona los datos relacionados con la unidad de estudio
3. La unidad de muestreo es:
    (a) Una porción o parte de la población de interés
    (b) La unidad de interés en un estudio estadístico
    (c) La unidad básica en términos de la cual se aplica una técnica de muestreo
    (d) La unidad que proporciona los datos relacionados con la unidad de estudio

4. Un banco está estudiando el nivel de satisfacción de los clientes con sus servicios y para tal fin realizará un estudio por muestreo. Al respecto el investigador a cargo expresó que:
   A. La unidad de estudio y la unidad de información son las mismas en este caso.
   B. La unidad de muestreo es un cliente del banco.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

5. Una trabajadora social desea investigar algunos aspectos relacionados con la calidad de vida de los adultos mayores que residen en hogares de ancianos. Para tal fin selecciona una muestra aleatoria y visita varios hogares de ancianos para valorar si dichas organizaciones poseen planes e infraestructura adecuados. Al respecto la trabajadora social considera que:
   A. La unidad de información corresponde a los ancianos que residen en el hogar visitado.
   B. La unidad de muestreo corresponde al director del hogar visitado.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

6. Una nutricionista está investigando la calidad de la nutrición que reciben los niños de una escuela. Para este fin selecciona una muestra aleatoria y visita los hogares de los niños y entrevista a sus padres. Con relación a esta situación la nutricionista considera que:
   A. La unidad de información corresponde a los niños de la escuela.
   B. La unidad de muestreo corresponde a los padres de cada uno de los niños seleccionados.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

7. Una nutricionista está investigando la calidad de la nutrición que reciben los niños de una escuela. Para este fin selecciona una muestra aleatoria y visita los hogares de los niños y entrevista a sus padres. Con relación a esta situación la nutricionista considera que:
   A. El marco muestra es una lista de todos los niños de la escuela.
   B. La unidad de estudio corresponde a los padres de cada uno de los niños seleccionados.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

8. Una nutricionista está investigando la calidad de la nutrición que reciben los niños de una escuela. Para este fin selecciona una muestra aleatoria y visita los hogares de los niños y entrevista a sus padres. Con relación a esta situación la nutricionista considera que es necesario emplear una muestra porque:

   A. Visitar todos los hogares de todos los niños de la escuela requiere demasiado tiempo.

   B. El estudio de la variable en cuestión implica la destrucción de la unidad de interés.

Con respecto a las dos afirmaciones anteriores, es incorrecto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

9. De las siguientes, no es una razón para trabajar con muestras:
   (a) Se mejora la calidad de la información recopilada
   (b) Se reducen los costos
   (c) En ocasiones la población se destruye al ser observada
   (d) Se eliminan el riesgo de definir mal la población

10. Una fábrica de fusibles prueba la calidad de su producto terminado. El ingeniero a cargo afirma que es estrictamente necesario emplear un muestreo porque:
   (a) Estudiar la población requeriría demasiado tiempo
   (b) Estudiar la muestra es más barato
   (c) La prueba del producto es destructiva
   (d) La población es infinita

11. Una __________ es una colección de todos los elementos de un grupo. Una colección de algunos de esos elementos es una __________. Las opciones que mejor completan la frase anterior son:
   (a) muestra, población
   (b) población, muestra por conveniencia
   (c) población, muestra aleatoria
   (d) población, muestra

12. Con respecto al tamaño de la muestra es verdadero que:
   (a) Depende del tamaño de la población
   (b) El nivel de confianza en la estimación no es importante
   (c) La variabilidad de la característica que se estima influye fuertemente
   (d) Ninguna de las anteriores

13. Con respecto al tamaño de muestra un investigador realizó las siguientes dos afirmaciones:

A. El costo es determinante del tamaño de muestra, aunque no esté en la fórmula.

B. El nivel de precisión se refiere al nivel de error permitido en la estimación.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

14. Con respecto al uso de muestras un investigador realizó las siguientes dos afirmaciones:

A. Toda investigación estadística requiere la utilización del muestreo.

B. Cuando se estudia la población completa, se dice que se realiza un censo.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (d) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

15. Se desea estimar la media poblacional de una variable x. Se conoce que la desviación estándar es de 87 unidades. Se requiere una confianza en la estimación del 90% y que el error no sea mayor que 20 unidades. Entonces, el tamaño de muestra requerido es:

(a) 520 (b) 73
(c) 52 (d) Ninguna de las anteriores

16. Se desea estimar la media poblacional de una variable x. Se conoce que la desviación estándar es de 87 unidades y que la población está compuesta por 200 unidades. Se requiere una confianza en la estimación del 95% y que el error no sea mayor que 15 unidades. Entonces, el tamaño de muestra requerido es:

(a) 130 (b) 92
(c) 79 (d) Ninguna de las anteriores

17. Se desea estimar una proporción poblacional para una cierta variable. Se cuenta con una estimación previa del 20%. Se requiere una confianza en la estimación del 95% y que el error no sea mayor que 5%. Entonces, el tamaño de muestra requerido es:

(a) 174 (b) 246
(c) 385 (d) Ninguna de las anteriores

18. Se desea estimar una proporción poblacional para una cierta variable. Se requiere una confianza en la estimación del 99% y que el error no sea mayor que 3,5%. Entonces, el tamaño de muestra requerido es:
(a) 1105 (b) 1355
(d) 867 (d) Ninguna de las anteriores

19. Se desea estimar una proporción poblacional para una cierta variable. La población tiene un tamaño de 220 individuos. Se requiere una confianza en la estimación del 90% y que el error no sea mayor que 3%. Entonces, el tamaño de muestra requerido es:
(a) 171 (b) 752
(c) 457 (d) 149

20. En un estudio por muestreo se desea estimar la talla promedio de las 400 mujeres que laboran en una empresa. Por un estudio previo se sabe que la desviación estándar es de 22 cm y se ha establecido una precisión de 5 cm con una confianza del 99%. Entonces, el tamaño de muestra requerido es, en número de mujeres:
(a) 98 (b) 129
(c) 105 (d) 84

21. En un estudio por muestreo se desea estimar la proporción de las 400 mujeres que laboran en una empresa que poseen hijos menores de 10 años. Se ha establecido una precisión de 5% con una confianza del 95%. Entonces, el tamaño de muestra requerido es, en número de mujeres:
(a) 385 (b) 271
(c) 162 (d) 197

22. Un noticiero en la televisión decide realizar una encuesta sobre la calidad de la educación en el país. Los televidentes que desean participar llaman a un número telefónico y votan por la opción que consideran más apropiada. Durante el noticiero votaron 5.500 personas, y el 75% considera que la educación del país debe mejorarse. El tipo de muestreo empleado por este noticiero es:
(a) Aleatorio
(b) Por conveniencia
(c) Voluntario
(d) De juicio

23. Un noticiero en la televisión decide realizar una encuesta sobre la calidad de la educación en el país. Los televidentes que desean participar llaman a un número telefónico y votan por la opción que consideran más apropiada. Durante el noticiero votaron 5.500 personas, y el 75% considera que la educación del país debe mejorarse. Con respecto a este muestreo es falso que:
(a) Es una muestra representativa por ser muy grande
(b) Se presenta un posible sesgo de selección
(c) Es un muestreo no aleatorio
(d) La muestra no es representativa a pesar de su tamaño

24. La principal ventaja de un muestreo aleatorio es que:
    (a) Elimina los sesgos de selección
    (b) Permite la cuantificación y control del error de muestreo
    (c) Reduce los costos del estudio
    (d) Emplea muestras de menor tamaño

25. La discrepancia, debida al azar, entre la estimación de una característica obtenida a través de una muestra y su verdadero valor en la población corresponde al concepto de:
    (a) Sesgo de selección
    (b) Error de muestreo
    (c) Sesgo de medición
    (d) Aleatoriedad

26. El error sistemático, no debido al azar, y que ocasiona que diferencias entre el valor estimado a través de la muestra y el valor verdadero corresponde al concepto de:
    (a) Sesgo
    (b) Error de muestreo
    (c) Variabilidad
    (d) No aleatoriedad

27. Un gerente está haciendo un estudio de mercado. Ha seleccionado una muestra aleatoria de 385 consumidores, pero hubo 50 de ellos que no contestaron el cuestionario. Esta situación:
    (a) No es problema porque la mayoría sí lo contestaron
    (b) Es un problema porque el tamaño de la muestra efectivamente tomada es menor
    (c) Es un problema, pero se resuelve sustituyendo los valores faltantes por sus valores esperados
    (d) No es un problema porque no fue causado intencionalmente por el investigador

28. Con respecto al muestreo un investigador realizó las siguientes dos afirmaciones:

    A. La selección de la técnica apropiada no es tan importante como la determinación del tamaño de la muestra para lograr una muestra representativa.

    B. La existencia de un marco muestral bien definido es clave para seleccionar la técnica de muestreo.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

29. El departamento de recursos humanos de una empresa va a efectuar un estudio por muestreo sobre la satisfacción de los empleados. La empresa tiene 2500 empleados y el tamaño de la muestra es de 250 empleados. El encargado tomó una lista de todos los funcionarios de la empresa y con ayuda de un software obtuvo 250 números aleatorios, los cuales empleó para seleccionar a los empleados que conformarían la muestra. En esta situación el tiempo de muestreo empleado es:
    (a) Muestreo sistemático
    (b) Muestreo simple al azar
    (c) Muestreo estratificado
    (d) Muestreo por conglomerados

30. El departamento de recursos humanos de una empresa va a efectuar un estudio por muestreo sobre la satisfacción de los empleados. La empresa tiene 2500 empleados y el tamaño de la muestra es de 250 empleados. El encargado calculo un valor $k = 2500/250 = 10$, entonces tomó una lista de todos los funcionarios de la empresa, con ayuda de un software obtuvo un número aleatorio entre uno y diez, y a ese número empezó a sumar 10 una y otra vez hasta completar 250 números, los cuales empleó para seleccionar a los empleados que conformarían la muestra. En esta situación el tiempo de muestreo empleado es:
    (a) Muestreo sistemático
    (b) Muestreo simple al azar
    (c) Muestreo estratificado
    (d) Muestreo por conglomerados

31. El departamento de recursos humanos de una empresa va a efectuar un estudio por muestreo sobre la satisfacción de los empleados. La empresa tiene 2500 empleados y el tamaño de la muestra es de 250 empleados. El encargado dividió la empresa en sus distintos departamentos, por considerar que los empleados en cada uno de ellos tienden a ser más homogéneos entre sí con respecto a la variable estudiada. Luego tomó una muestra de cada uno de estos subgrupos, de modo que la muestra total resultante refleje en forma proporcional la cantidad de empleados que hay en cada departamento. En esta situación el tiempo de muestreo empleado es:
    (a) Muestreo sistemático
    (b) Muestreo simple al azar
    (c) Muestreo estratificado
    (d) Muestreo por conglomerados

32. Con respecto a una muestra sea representativa de una población es correcto que:
    (a) Basta con que sea del tamaño apropiado
    (b) Debe ser obtenida al azar sin importar su tamaño
    (c) Debe al menos el 20% de la población
    (d) Ninguna de las anteriores

CAPÍTULO 5

PRUEBAS DE HIPÓTESIS

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Plantear las hipótesis nula y alternativa en problemas de decisión con respecto a la media o la proporción poblacional
2. Identificar los posibles errores que se pueden cometer al tomar decisiones con base en muestras
3. Describir los pasos del procedimiento de prueba de hipótesis
4. Calcular los estadísticos de prueba adecuados según el tipo de problema
5. Tomar decisiones con base en el procedimiento de prueba de hipótesis

5.1 Inferencia mediante pruebas de hipótesis

Conceptos generales

Muchas veces la inferencia que se debe realizar no se refiere a la estimación de un parámetro, sino que se deben tomar decisiones sobre afirmaciones hechas sobre un parámetro. Esto es, se debe decidir, con base en evidencia experimental, si una afirmación (hipótesis) hecha acerca de un parámetro es falsa o verdadera. Por ejemplo un ingeniero tendría que decidir, basándose en datos muestrales, si existe diferencia en la precisión de dos diferentes aparatos de medición o si la media de un proceso de llenado ha cambiado de tal manera que la máquina debe ser calibrada ya que el proceso está fuera de control.

Digamos que en el ejemplo de la media del proceso, ésta debe ser 250 gramos, por lo tanto para tratar de confirmarlo el ingeniero toma la información de una muestra de 40 bolsas salidas de esta máquina. Supóngase que la media de la muestra es de 255 gramos por lo que se decide que la máquina requiere ajustes. Dado que la decisión se basa en una muestra podrá ocurrir que, aunque la media muestral sea mayor, la media de la población sea realmente de 250 gramos. Incluso podría ser que la media del proceso sea mayor de 255, por ejemplo de 260 gramos. Para evitar la toma de decisiones erróneas, en vez de tomar la decisión basándose en los resultados de una muestra se debe realizar un proceso formal mediante una *prueba de hipótesis*. Aunque nunca se sabe con certeza absoluta la verdad o falsedad de una hipótesis, a menos que se examine toda la población, el proceso es mucho más seguro que si la decisión se basa en una muestra.

PRUEBA DE HIPÓTESIS Procedimiento basado en evidencia de la muestra y la teoría de probabilidades para determinar si la hipótesis es una afirmación razonable.

Algunas definiciones

Los siguientes son algunos de los términos usados con más frecuencia en el contexto de las pruebas de hipótesis:

HIPÓTESIS ESTADÍSTICA Es una afirmación o conjetura acerca de una o más poblaciones.

HIPÓTESIS NULA Es cualquier hipótesis que se desea probar. Se denota por H_0.

HIPÓTESIS ALTERNATIVA Es la hipótesis que se acepta cuando la hipótesis nula es rechazada. Se denota por H_1.

Tomando el ejemplo anterior, el ingeniero cree que la máquina podría estar llenando las bolsas con una cantidad superior a la especificada, así que esa es su hipótesis de investigación, y para decidir si debe realizar ajustes, entonces toma la muestra. El valor especificado constituirá su hipótesis nula. La hipótesis nula es aquello que se desea probar. En este caso el ingeniero desea probar si el peso promedio de las bolsas es efectivamente igual a 250 gramos. Esta hipótesis nula podría ser aceptada o rechazada. En caso de rechazarla, entonces tomará aceptará la hipótesis alternativa. La evidencia que posee sugiere que las bolsas tienen un peso superior a 250 gramos, por lo que su hipótesis alternativa sería que el peso de las bolsas es mayor que 250 gramos.

Ejemplo

Se sabe por estudios previos que los recién nacidos de cierta población tienen una talla promedio de 49,5 cm. Una enfermera estudió un grupo de 40 recién nacidos, y obtuvo una media de 53,4 cm.

La enfermera desea saber si estos resultados apoyan los estudios previos. ¿Cuáles serían sus hipótesis nula y alternativa?

Solución

En esta situación la enfermera tiene un valor poblacional establecido, que es que los recién nacidos miden, en promedio, una talla de 49,5 cm. Por tanto, su hipótesis nula será:

H_0: La talla media de los recién nacidos es 49,5 cm.

Pero los datos recopilados sugieren que este promedio podría ser mayor que 49,5 cm, por lo que, de descartar la hipótesis nula anterior, se aceptaría la hipótesis alternativa:

H_1: La talla media de los recién nacidos es mayor que 49,5 cm.

Generalmente las hipótesis se expresan en términos de símbolos:

$H_0: \mu = 49,5$
$H_1: \mu > 49,5$

Ejemplo

En cada uno de los siguientes casos plantee la hipótesis nula y la alternativa:

1. Un cierto material viene en cajas de peso promedio 17 libras y desviación estándar 0,4 libras. Se recibe un cargamento grande y se tiene la sospecha de que el peso promedio de las cajas es inferior al usual. Para verificar la sospecha se toma una muestra al azar de 86 cajas y se pesan, obteniéndose un promedio de 16,5 libras. ¿Se puede afirmar que efectivamente el peso de las cajas es inferior al acostumbrado?

2. En una granja bastante grande se producen pollos. Según los estándares establecidos, el peso medio de los pollos debe ser de 4,2 Kg. con varianza 1,96. Se desea determinar si es cierta la queja de un grupo de clientes de que el peso medio ha disminuido durante las últimas semanas. Para verificar tal afirmación se contrata un ingeniero avícola, el cual toma una muestra de 65 pollos, y encuentra un peso medio de 3,86 Kg. ¿Significa esto que efectivamente el peso medio es inferior al usual?
3. De acuerdo con datos de un estudio realizado en un país europeo la edad promedio de diagnóstico del cáncer de próstata es 75 años. Un investigador nacional considera que en nuestro país esa edad de diagnóstico es menor. Se tomó una muestra de 80 casos diagnosticados y encontró una edad promedio de 69 años con una desviación estándar muestral de 9 años. ¿Qué puede concluirse con base en estos datos?
4. Según un estudio los niños de los estratos socio económicos medios y altos inician alguna práctica de cuidado de su salud buco dental a los 15,6 meses. En una muestra de 35 niños de familias de estratos bajos se encontró una edad media de inicio de la higiene bucal a los 18,2 meses, con una desviación estándar de 8,5 meses. ¿Puede considerarse que la edad de los niños de familias de estratos bajos es mayor que 15,6 meses?

Solución

1. En esta situación se indica que el peso promedio de las cajas en que viene el material es 17 libras, por tanto se querrá verificar que se satisface esta especificación, de manera que la hipótesis nula será que el peso promedio es 17 libras. Por otro lado, en la muestra de 86 cajas se obtuvo un peso promedio inferior, lo cual también sugiere la pregunta, entonces la hipótesis nula será que la media es inferior a 17 libras. En resumen:

$H_0: \mu = 17$
$H_1: \mu < 17$

2. De acuerdo con este problema existe un estándar de 4,2 Kg. en promedio por animal, por lo que la hipótesis nula es que el promedio sea igual a 4,2 Kg. En la muestra se encuentra un peso medio inferior a 4,2 Kg., de modo que la hipótesis nula es que el peso medio es inferior:

$H_0: \mu = 4,2$
$H_1: \mu < 4,2$

3. De acuerdo con los datos la edad promedio de diagnóstico del cáncer de próstata es 75 años, de manera que se plantea la hipótesis nula

de que el promedio es igual a 75. En la muestra se obtiene una edad promedio menor que 75, así que la hipótesis alternativa sería que la media es menor que 75 años:

H_0: $\mu = 75$
H_1: $\mu < 75$

4. Según el estudio la media es 15,6 meses, de modo que la hipótesis nula será que la media es igual a 15,6. En la muestra se obtuvo un valor más alto, de manera que la hipótesis alternativa será que la media es mayor que 15,6:

H_0: $\mu = 15{,}6$
H_1: $\mu > 15{,}6$

Ejercicio de revisión

Un ingeniero está estudiando la vida útil de distintos proyectos construidos con cierto tipo de pavimento. Se sabe que los camiones pesados producen un daño elevado y que reducen la vida útil de las vías. Se realiza un estudio para saber qué proporción de los camiones llevan una carga excesiva. En el caso de los camiones de 3 ejes se cuenta con un estudio previo en el que se indica que el 10% de estos vehículos portaban un peso superior al permitido. En una muestra de 40 de estos camiones, se encontró que 6 de ellos portaba una carga excesiva. ¿Cuáles es la hipótesis nula y cuáles es la hipótesis alternativa de este problema?

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Las pruebas de hipótesis pueden hacer referencia a una media de una población, como en el ejemplo anterior, pero también podrían referirse a una proporción. Este sería el caso en el cual lo que se desea probar es una afirmación que se relaciona con el porcentaje de ciudadanos que votarían por un determinado candidato, la tasa de prevalencia de una enfermedad, la proporción de artículos defectuosos en una producción, la tasa de desempleo, entre otros casos de distintas proporciones poblacionales.

Ejemplo

Un empresario es el único distribuidor de electrodomésticos y productos tecnológicos de su zona. Leyó en un medio que hasta un 74% de los internautas ha realizado alguna compra por internet en el transcurso de los últimos 3 meses. Para comprobar si en su zona esta proporción es similar a la publicada, aplicó un cuestionario a una muestra de 50 personas que fueran residentes de la zona y que usaran internet regularmente, y les preguntó si habían realizado compras en línea en el último trimestre. La encuesta reveló que 30 internautas de la zona han realizado compras por internet en ese periodo. O sea, que solo el 60% de los entrevistados respondió afirmativamente. ¿Cuál sería la hipótesis nula y alternativa en este caso?

Solución

El empresario desea probar que si es cierto que el 74% de los usuarios de internet han realizado compras por internet en el último trimestre, por tanto, su hipótesis nula será:

H_0: La proporción de usuarios que internet que ha realizado compras por internet es igual a 74%.

Pero los datos recopilados indican que ese porcentaje podría ser menor, por lo que, de descartar la hipótesis nula anterior, se aceptaría la hipótesis alternativa:

H_1: La proporción de usuarios que internet que ha realizado compras por internet es menor que 74%.

Generalmente las hipótesis se expresan en términos de símbolos:

$$H_0: P = 0,74$$
$$H_1: P < 0,74$$

También, las pruebas de hipótesis pueden referirse a dos o más medias poblacionales. En estos casos se cuenta con dos o más medias, o bien, dos o más proporciones, las cuales se desea comparar para determinar si presentan diferencia significativa. El caso específico de pruebas de hipótesis con dos medias o dos proporciones se trata en el capítulo siguiente. También cabe señalar que las pruebas pueden referirse, además de medias y proporciones, a otros parámetros, como varianzas, coeficientes de correlación, coeficientes de ecuaciones de regresión, entre muchos otros.

Ahora bien, en las pruebas de hipótesis las decisiones se toman con base en datos muestrales. Suponga que se toman varias muestras aleatorias de una misma población y se calcula cada una de las medias muestrales. La media en cada muestra va a ser distinta. Estas diferencias se deben al azar. Entonces, simplemente por el azar, a veces el valor de muestra va a estar muy cercano al verdadero valor poblacional, y otras veces, también por el azar, el valor muestral estará alejado del valor verdadero. Esto significa que en las pruebas de hipótesis existe el riesgo de tomar una decisión equivocada, cuando sea alguno de esos casos en los cuales la media muestral queda muy alejada del valor verdadero. Es decir, podría cometerse un error en la decisión a la cual conduce la prueba. El único modo de eliminar dicha posibilidad sería estudiar toda la población, lo cual podría ser demasiado costo, e incluso, imposible.

Entonces, al realizar pruebas de hipótesis se pueden cometer dos tipos de errores, llamados error tipo I y error tipo II.

ERROR TIPO I Es el error que se comete cuando se rechaza una hipótesis que es correcta y la probabilidad de cometer este error se denota por α.

ERROR TIPO II Es el error que se comete cuando se acepta una hipótesis que es incorrecta y la probabilidad de cometer este error se denota β.

Ejemplo

Una empresa fabrica bombillos. Cada bombillo tiene una vida esperada de 1000 horas, pero algunos clientes se han quejado de que los bombillos se queman antes de las 1000 horas. La gerencia decide tomar una muestra y probar la hipótesis nula de que los bombillos tienen una vida media de 1000 horas, contra la hipótesis alterna de que la vida media de los bombillos es menor que dicha especificación. ¿Cómo podrían darse y qué significan los errores tipo I y tipo II en esta situación?

Solución

En esta situación los errores tipo I y tipo II podrían darse si la muestra no representa bien a la población. Esto puede darse de los modos siguientes:

El proceso de producción de la empresa está bien controlado, y la vida media de los bombillos es 1000 horas, pero en la muestra usada en la prueba de hipótesis se seleccionaron, por cuestión del azar, muchos bombillos con una vida media inferior a 1000 horas, por lo que se rechazó la hipótesis nula de que la vida media de los bombillos es 1000 horas, a pesar de que era verdadera. Este es el **error tipo I**. Este error llevaría a la empresa a tratar de mejorar su producción innecesariamente, lo cual le generaría costos adicionales.

El proceso de producción de la empresa no está bien controlado, por lo que, efectivamente, la vida media de los bombillos es inferior a 1000 horas, como lo han indicado los clientes que se han quejado, pero en la muestra, por cuestión del azar, se seleccionaron muchos bombillos con una media cercana a 1000 horas, por lo que no se rechazó la hipótesis nula, a pesar de que era falsa. Este es el **error tipo II**. Este error llevaría a la empresa a no mejorar una producción que sí requiere mejoras, por lo cual sus clientes podrían dejar de comprar sus productos.

Ejercicio de revisión

Una empresa realiza un estudio de mercado en una muestra de 150 consumidores y se plantea probar la hipótesis de que al menos el 30% de ellos compraría su producto. ¿Cuáles serían las hipótesis nula y alternativa? ¿En qué consistirían los errores tipo I y tipo II?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Material audiovisual

En la página de internet de este texto podrá encontrar varias presentaciones que explican el tema de las pruebas de hipótesis.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

Aplicación

Métodos analíticos

Actualmente se realizan en laboratorios de todo el mundo análisis de la composición de diferentes productos o materiales, y para ello se emplean diversas técnicas analíticas, como por ejemplo, un análisis de aguas. Los resultados de estos análisis son de gran importancia en diferentes industrias, e incluso para algunas oficinas del gobierno, pues, además de impactar la actividad empresarial, podrían afectar la salud y la vida de las personas. Esto quiere decir que los errores que se comentan en estos estudios pueden producir graves consecuencias para la sociedad.

Por supuesto, que al realizar estos análisis son necesarias muestras representativas, la aplicación de las técnicas apropiadas y el uso de equipo adecuado. Aun así, va a haber un cierto nivel de incertidumbre en los resultados.

Cuando se efectúa un análisis químico existe la posibilidad de cometer un falso positivo, o sea, un error tipo I. También es posible cometer un falso negativo, es decir, un error tipo II. En muchísimos casos se requiere que estas probabilidades sean mínimas, por lo que se emplean altos niveles de confianza $(1 - \alpha)$ y elevadas potencias de las pruebas de hipótesis $(1 - \beta)$.

5.2 Procedimiento para pruebas de hipótesis sobre la media

El siguiente es un procedimiento que puede seguir cuando se desea efectuar una prueba de hipótesis con respecto a la media aritmética.

- **Paso 1.** Plantear las hipótesis. Se establece la hipótesis nula y la hipótesis alternativa.
- **Paso 2.** Determinar el nivel de significancia. Se selecciona un nivel de significancia a para la prueba.
- **Paso 3.** Identificar el estadístico de prueba. Dependiendo del tipo de prueba, del tamaño de la muestra y otras condiciones se selecciona el estadístico apropiado para realizar la prueba.
- **Paso 4.** Se formula una regla para tomar la decisión. Con base en el estadístico de prueba se establece el criterio específico para aceptar o rechazar la hipótesis nula.
- **Paso 5.** Se toma una muestra y se llega a una decisión: se acepta o se rechaza la hipótesis nula.

Se explicará el procedimiento general para realizar una prueba de hipótesis realizando una prueba para una media aritmética mediante el siguiente ejemplo:

Ejemplo

Una empresa fabrica bombillos. Cada bombillo tiene una vida esperada de 1000 horas, pero algunos clientes se han quejado de que los bombillos se queman antes de las 1000 horas. La gerencia decide tomar una muestra de 50 bombillos y desea probar que los bombillos tienen una vida media de 1000 horas. La media obtenida a partir de la muestra es de 970 horas. Se conoce que la desviación estándar es 60 horas. Determine, a un nivel de significación del 5%, si la media poblacional de estos bombillos es efectivamente de 1000 horas.

Solución

Paso 1. Plantear las hipótesis. Toda prueba inicia planteando las hipótesis. La hipótesis nula se plantea como $H_0: \mu = \mu_o$, donde μ_o es el valor a probar (en este caso 1000 horas), y la hipótesis alternativa podría ser como alguna de las siguientes:

$$H_1: \mu > \mu_o$$
$$H_1: \mu < \mu_o$$
$$H_1: \mu \neq \mu_o$$

La hipótesis alternativa se formula dependiendo del valor obtenido en la muestra o de lo que se desee plantear como hipótesis alternativa. Es decir, si en vez de querer saber si $\mu > \mu_o$ o $\mu < \mu_o$, se desea simplemente saber si $\mu \neq \mu_o$.

En este ejemplo se desea probar que la media verdadera es de 1000 horas ($\mu_o = 1000$), por lo tanto la hipótesis nula es:

$$H_0: \mu = 1000$$

Como $\overline{x}$ (valor muestral que "representa" a μ) es igual a 970, que es un valor menor que 1000, entonces la hipótesis alternativa lógica sería que la media es menor que 1000, o sea, $H_1: \mu < 1000$. En resumen se tiene que las hipótesis son:

$$H_0: \mu = 1000$$
$$H_1: \mu < 1000$$

Paso 2. Especificar el nivel de significación α (la probabilidad de cometer el error tipo I) con que se desea trabajar. Los valores usualmente usados son 5% y 1%. Si se escoge una probabilidad de error tipo I muy pequeña esto hace que la probabilidad de error tipo II sería muy grande. En el ejemplo se especifica un valor de α de 0,05.

Paso 3. Se usa el estadístico de prueba apropiado. En el caso de la media, dependiendo del tamaño de la muestra y si se conoce o no la desviación estándar poblacional, se usa:

$$z_c = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \text{ con } n \geq 30 \text{ con } \sigma \text{ conocida o con } n < 30 \text{ y } \sigma \text{ conocida}$$

$$z_c = \frac{\overline{x} - \mu}{s / \sqrt{n}} \text{ con } n \geq 30 \text{ con } \sigma \text{ desconocida}$$

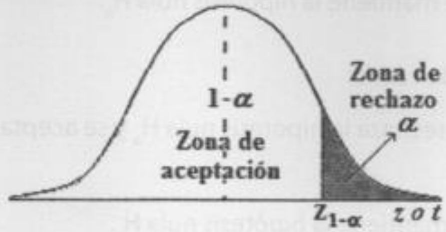
$$t_c = \frac{\overline{x} - \mu}{s / \sqrt{n}} \text{ con } n < 30 \text{ y } \sigma \text{ desconocida}$$

A este valor se le llamará "*z* calculada" o "*t* calculada", según el caso. En el ejemplo se tiene $n > 30$ y σ conocida, pues $n = 50$ y $\sigma = 60$ horas, por lo que se calcula z (según el problema se tiene que $\overline{x} = 970$ y de la hipótesis nula se obtiene que m = 1000):

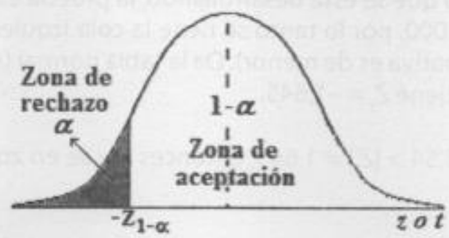
$$z_c = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} = \frac{970 - 1000}{60 / \sqrt{50}} = -3{,}54$$

Paso 4. Se especifica un criterio de aceptación o rechazo de la hipótesis nula según el estadístico de prueba usado en el paso anterior. En las hipótesis para la media el criterio es:

Prueba de una cola: Cuando se plantea la hipótesis alternativa como $H_1: \mu > \mu_0$



Prueba de una cola: Cuando se plantea la hipótesis alternativa como $H_1: \mu < \mu_0$



Prueba de dos colas: Cuando se plantea la hipótesis alternativa como $H_1: \mu \neq \mu_0$

Puede observarse que cuando la hipótesis alternativa se ha planteado como $H_1: \mu < \mu_0$ o como $H_1: \mu < \mu_0$, entonces se dice que la prueba es de una cola, y la zona de aceptación queda definida por el valor de $1 - \alpha$.

Cuando la hipótesis alternativa se ha planteado como $H_1: \mu \neq \mu_0$, entonces se dice que la prueba es de dos colas, y la zona de aceptación queda definida por el valor de $1 - \alpha$

El valor de Z_α o de t_α se obtiene de la tabla respectiva con una probabilidad igual a $1 - \alpha$ en el caso de Z y α en el caso de t en las pruebas de una cola y con una probabilidad igual a $1 - \alpha$ en el caso de Z y $\alpha/2$ en el caso de t en las pruebas de dos colas. A este valor de z se le llamará "z tabular" o "z de la tabla" (por ser obtenida de la tabla de la distribución normal), o en el caso de t, "t de la tabla".

Puede establecerse la regla siguiente en términos de z:

Si $|z_c| > |z_t|$ se rechaza la hipótesis nula H_o y se acepta la hipótesis alternativa H_1.

Si $|z_c| \leq |z_t|$ se mantiene la hipótesis nula H_o.

En términos de *t* sería:

Si $|t_c| > |t_t|$ se rechaza la hipótesis nula H_o y se acepta la hipótesis alternativa H_1.

Si $|t_c| \leq |t_t|$ se mantiene la hipótesis nula H_o.

En el ejemplo que se está desarrollando, la prueba es de una cola, porque H_1: $\mu < 1000$, por lo tanto se tiene la cola izquierda (porque la hipótesis alternativa es de menor). De la tabla normal (Apéndice 6) con $\alpha = 0{,}05$, se obtiene $Z_t = -1{,}645$.

Como $|Z_c| = 3{,}54 > |Z_t| = 1{,}645$, entonces Z_c cae en zona de rechazo de la H_0.

Paso 5. Se acepta o se rechaza la H_0 y se toma la decisión. En este ejemplo se rechaza la hipótesis nula H_0. Es decir, se rechaza que $\mu = 1000$ y se acepta la H_1: $\mu \leq 1000$.

La conclusión es que a un nivel de significación del 5% se rechaza la hipótesis nula de que la vida media de los bombillos es de 1000 horas y se considera que existe evidencia estadística para aceptar la hipótesis alternativa de que la vida útil de los focos es menor de 1000 horas.

Ejercicio de revisión

Una institución del gobierno periódicamente verifica que las empresas y los comercios no realicen prácticas abusivas contra los consumidores. Recientemente ha verificado una muestra de 200 latas de atún cuya etiqueta indica que contienen 130 grs. como peso escurrido. El promedio en la muestra fue 112 grs. como peso escurrido Por un estudio anterior se conoce que la desviación estándar es 20,5 grs.

¿Constituyen estos datos muestrales evidencia suficiente para considerar que las latas de atún poseen un peso escurrido inferior al ofrecido?

Use un nivel de significancia de 5%.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

5.3 Prueba de hipótesis con muestras pequeñas

Tal como se indicó en el ejemplo anterior, si el tamaño de la muestra es menor que 30 y se desconoce la desviación estándar poblacional no se conoce, entonces no se emplea la distribución normal para realizar la prueba, sino que se utiliza la distribución *t*, según el caso con una o dos colas, y $n - 1$ grados de libertad.

En el siguiente ejemplo se muestra la prueba empleado la distribución *t*.

Ejemplo

Una muestra aleatoria de frascos de mantequilla de maní presentaron pesos de (en gramos):

252, 251, 249, 253, 250, 255, 248, 258

La empresa ha tratado de ajustar el proceso de llenado para que cada frasco contenga 250 gramos. Verifique, a un nivel de significación del 5% si ese valor esperado se mantiene sin cambio.

Solución

Se tiene que hay un peso especificado para los frascos de mantequilla de maní de 250 gramos, por lo que $\mu_0 = 250$ y además $n = 8$.

De los datos de la muestra se obtiene una media $\bar{x} = 252$ y una desviación estándar $s = 3,3$. Como la media muestral $\bar{x}$ que "representa" a la media poblacional μ es mayor que μ_0, entonces se planteará una hipótesis alternativa de $\mu > \mu_0$.

Paso 1. Planteamiento de las hipótesis:

$$H_0: \mu = 250$$
$$H_1: \mu > 250$$

Paso 2. Como $n < 30$ y σ desconocida, se calcula t_c:

$$t_c = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{252 - 250}{3{,}3 / \sqrt{8}} = 1{,}72$$

Paso 3. De la tabla, con una cola, para un nivel de significancia $\alpha = 0{,}05$ y grados de libertad $gl = n - 1 = 8 - 1 = 7$, se obtiene $t_\alpha = 1{,}895$.

Paso 4. Como $|t_c| < |t_t|$, se acepta H_0 con $\alpha = 0{,}05$.

Paso 5. Se concluye que no hay evidencia suficiente para considerar que el peso promedio de los frascos de mantequilla de maní es mayor que 250 gramos.

Ejercicio de revisión

Una compañía de tarjetas de crédito desea probar si el saldo promedio de sus clientes es superior a \$500. En una muestra de 15 tarjetahabientes se obtuvo un saldo promedio de \$535 con una desviación estándar de \$215. ¿Qué puede concluirse a un nivel de significación del 5%?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

5.4 Prueba de hipótesis para la proporción poblacional

En estas pruebas de hipótesis se debe comprobar, con base en los resultados obtenidos en una muestra, si el valor verdadero de una proporción es igual a una constante determinada o si las proporciones de dos o más poblaciones son iguales. En este capítulo se abordará solo el primer caso, en el cual se prueba si una proporción es igual a una constante establecida.

Los pasos para la realización de estas pruebas de hipótesis son los mismos expuestos anteriormente.

Entonces, en el caso de una sola proporción P la hipótesis nula sería:

$$H_0: P = p_0$$

donde p_0 es la constante determinada. La hipótesis alternativa tiene las siguientes tres posibilidades:

$$H_1: P > p_0$$
$$H_1: P < p_0$$
$$H_1: P \neq p_0$$

El estadístico de prueba es:

ESTADÍSTICO DE PRUEBA PARA LA PROPORCIÓN POBLACIONAL	$z = \dfrac{x - nP}{\sqrt{nPQ}}$

Donde x es el número de eventos observados en la muestra, n es el tamaño de la muestra o número de ensayos, P es la proporción poblacional que se estableció en la hipótesis nula, y Q es el complemento de P, es decir, $Q = 1 - P$.

El criterio de aceptación o rechazo de la hipótesis nula H_0 es igual al de la prueba de las medias, es decir, si:

- Si $|z_c| > |z_t|$ se rechaza la hipótesis nula H_0 y se acepta la hipótesis alternativa H_1.
- Si $|z_c| \leq |z_t|$ se mantiene la hipótesis nula H_0.

Ejemplo

Pruebe la aseveración de que la proporción de adultos que realizaron algún tipo de ejercicio físico al menos una vez durante la semana pasada es menor de 20%, si se tomó una muestra de 1.200 personas, de los cuales 215 dicen que realizaron actividad física la semana pasada. Use $\alpha = 0.01$.

Solución

Hay que distinguir claramente que en los problemas de pruebas de hipótesis relacionados con proporciones no aparece una variable métrica, es decir, no aparece un promedio que se pueda medir en centímetros, gramos, dólares, minutos u otra unidad de medida. En este caso el problema se relaciona con un porcentaje supuesto de adultos que realizaron ejercicio físico y el conteo de esas personas en la muestra. En todos los casos de pruebas de hipótesis sobre una proporción se va a presentar esta situación, no hay una variable medible y se presentan datos de una variable que se obtiene por conteo y que se relaciona con respecto a un total poblacional o muestral (una proporción).

Una vez que se tiene bien definida la naturaleza del problema, entonces se siguen los mismos 5 pasos expuestos para el caso de las pruebas de hipótesis sobre la media poblacional.

Paso 1. Planteamiento de las hipótesis: El problema señala que se desea probar si el 20% de los adultos realizaron ejercicio físico al menos una vez durante la semana pasada, por lo que la hipótesis nula será:

$$H_0: P = 0{,}20$$

Por otro lado, los datos muestrales indican que de los 1200 adultos encuestados, 215 realizaron ejercicio físico la semana pasada, por lo que se tendría una proporción muestral equivalente a:

$$p = 215 / 1200 = 0{,}1792$$

Este dato muestral sugiere que la proporción de adultos que realizaron ejercicio físico es menor que 0,20, por lo que las hipótesis se plantearían como:

$$H_0: P = 0{,}20$$
$$H_1: P < 0{,}20$$

Paso 2. El problema indica que la prueba debe realizar a un nivel de significancia de un 1%.

Paso 3. Como el problema es una prueba de una proporción se calcula z_c:

$$z = \frac{x - nP}{\sqrt{nPQ}} = \frac{215 - 1200 \cdot 0{,}20}{\sqrt{1200 \cdot 0{,}20 \cdot 0{,}80}} = -1{,}80$$

De la tabla de la curva normal, para un nivel de significancia α = 0,01, con una cola, o sea, una confianza de 0,99, se obtiene $z_\alpha = -2{,}33$.

Paso 4. Como $|z_c| \leq |z_t|$, se acepta H_0 con $\alpha = 0{,}01$.

Paso 5. Se concluye que no se tiene evidencia estadística suficiente para rechazar la hipótesis de que la proporción de adultos que realizaron ejercicio físico al menos una vez durante la semana pasada es menor que 20%, a un nivel de significación del 1%.

Ejercicio de revisión

Un laboratorio farmacéutico considera que uno de sus fármacos alcanza en el 80% de los casos su máxima absorción en un plazo de 2 horas. En una muestra de 120 personas se obtuvo el resultado esperado en 80 casos. ¿Puede sostenerse la afirmación de la empresa a un nivel de significancia del 95%?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Minitab para realizar la prueba de hipótesis y uso del valor *P*

Un modo alternativo para decidir sobre la aceptación o rechazo de una hipótesis es emplear el valor *P* o valor de probabilidad. El valor *P* es la probabilidad de obtener un estadístico de prueba (sea *z*, *t* u otro, según la prueba) que sea tan extremo como el que se obtuvo a partir de los datos muestrales, bajo el supuesto de que la hipótesis nula es verdadera. En otras palabras, el valor *P* es el menor nivel de significación al que se puede rechazar la hipótesis nula cuando esta sea verdadera. La hipótesis nula se rechaza si el valor *P* es muy pequeño, inferior al nivel de significancia (a). En caso de que el valor *P* sea mayor o igual que el nivel de significancia, entonces se acepta la hipótesis nula. Veamos que esto es así porque, si suponemos que tenemos un valor *P* pequeño, como 0,005, entonces este valor nos indica que la probabilidad de que la hipótesis nula sea verdadera es muy baja. Al ser una probabilidad, este valor *P* siempre tomará valores entre 0 y 1.

La mayoría de los paquetes de cómputo calculan el valor *P*, además de los otros estadísticos de prueba, por lo que su uso en la investigación y en la industria es frecuente. Un ejemplo de un software útil para realizar la prueba de una hipótesis es Minitab.

Ejemplo

Una empresa fabrica bombillos. Cada bombillo tiene una vida esperada de 1000 horas, pero algunos clientes se han quejado de que los bombillos se queman antes de las 1000 horas. La gerencia decide tomar una muestra de 50 bombillos y desea probar que los bombillos tienen una vida media de 1000 horas. La media obtenida a partir de la muestra es de 970 horas. Se conoce que la desviación estándar es 60 horas. Utilice Minitab para determinar, a un nivel de significación del 5%, si la media poblacional de estos bombillos es efectivamente de 1000 horas.

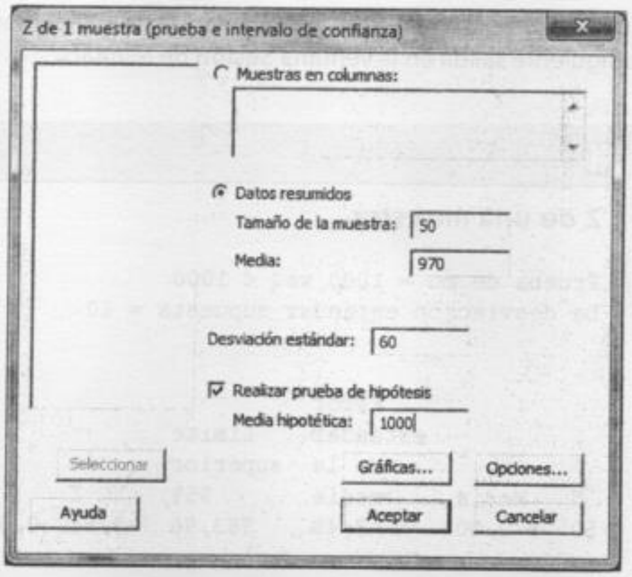
Solución

Como en cualquier prueba de hipótesis, se inicia por plantear las hipótesis. Tal como se expuso anteriormente, las hipótesis son:

$H_0: \mu = 1000$
$H_1: \mu < 1000$

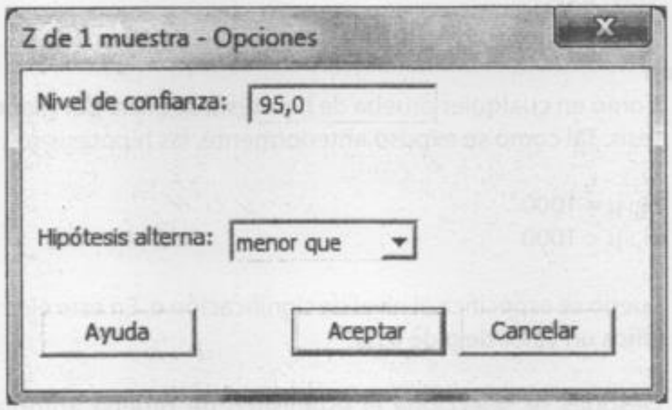
Luego se especifica el nivel de significación α. En este ejemplo se especifica un valor de α de 0,05.

Después se selecciona el estadístico de prueba apropiado. En este ejemplo se tiene $n > 30$ y σ conocida, pues $n = 50$ y $\sigma = 60$ horas, por lo que se calcula z. Así, en Minitab se debe dar clic en el menú **Estadísticas**, luego en el submenú **Estadística básica**, y ahí se elige la opción **Z de 1 Muestra**. Ahora se completa el cuadro de diálogo siguiente:



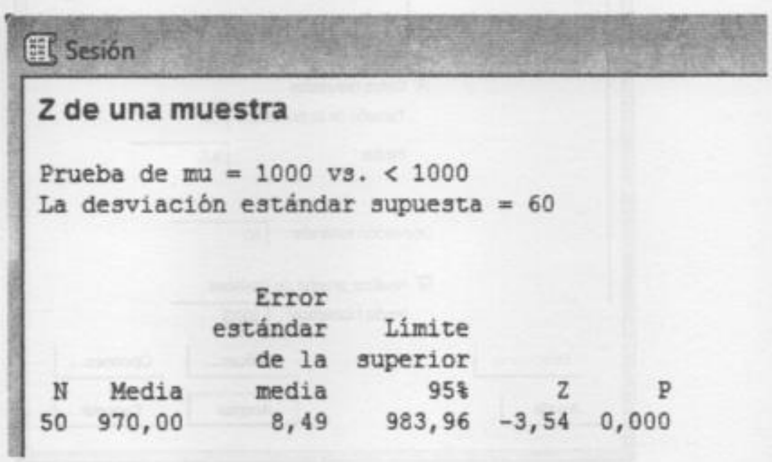
Se selecciona **Muestras en columnas** cuando se tiene la serie original de datos muestrales, pero en este caso ya se tiene calculada la media muestral, por lo que se escoge **Datos resumidos**, y se digita el tamaño de la muestra y la media muestral. Debe marcarse la casilla **Realizar prueba de hipótesis**, pues de otro modo Minitab solo dará el intervalo de confianza. En la celda se digita la media poblacional indicada en la hipótesis nula.

Luego debe darse clic en el botón **Opciones**, pues es ahí donde se indica el nivel de significancia y se selecciona la hipótesis alternativa:



Dado que el nivel de significancia de este ejercicio es 5%, entonces el nivel de confianza será de 95%. En la opción de hipótesis alterna se elige la que dice **menor que**, pues la hipótesis alternativa indicada anteriormente fue H_1: $\mu < 1000$.

Luego se da clic en Aceptar, y nuevamente clic en Aceptar, y se obtiene la siguiente salida en la ventana **Sesión** de Minitab:



Sesión

Z de una muestra

```
Prueba de mu = 1000 vs. < 1000
La desviación estándar supuesta = 60


            Error
          estándar    Límite
             de la  superior
 N   Media   media       95%      Z      P
50  970,00    8,49    983,96  -3,54  0,000
```

Puede verse que Minitab indica que el valor del estadístico de prueba Z_c es –3,54, que coincide con el valor calculado anteriormente en este capítulo. De la tabla de la curva normal, o bien, del mismo Minitab se calcula el valor Z_t, que es –1,645, por lo que Z_c cae en zona de rechazo de la H_0.

Además, observe que Minitab calculó el valor *P*, que en este caso es 0,000, un valor inferior al nivel de significancia del 5%, por lo que se rechazaría la hipótesis nula.

Por cualquiera de los dos criterios (*z* o valor *P*), la conclusión es la misma, que a un nivel de significación del 5% se rechaza la hipótesis nula de que la vida media de los bombillos es de 1000 horas y se considera que existe evidencia estadística para aceptar la hipótesis alternativa de que la vida útil de los bombillos es menor de 1000 horas.

Cuando el estadístico de prueba sea *t*, entonces la opción del menú **Estadísticas** > **Estadística básica** que se emplea es **t de una Muestra**, y cuando la prueba se refiere a una proporción, entonces se elige **1 Proporción**.

Ejemplo

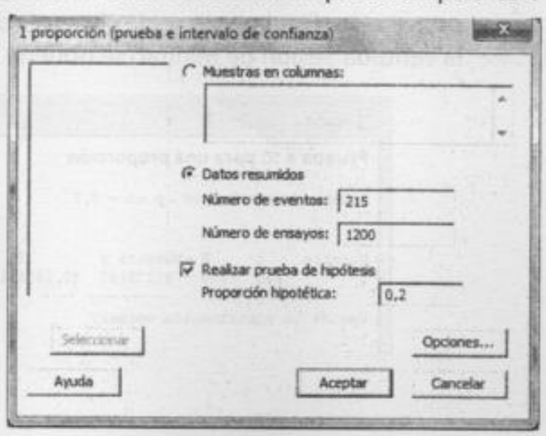
Pruebe la aseveración de que la proporción de adultos que realizaron algún tipo de ejercicio físico al menos una vez durante la semana pasada es menor de 20%, si se tomó una muestra de 1.200 personas, de los cuales 215 dicen que realizaron actividad física la semana pasada. Use $\alpha = 0.01$.

Las hipótesis se plantearían como (pues este ejemplo ya se explicó anteriormente en este capítulo):

$$H_0: P = 0,20$$
$$H_1: P < 0,20$$

El problema indica que la prueba debe realizar a un nivel de significancia de un 1%.

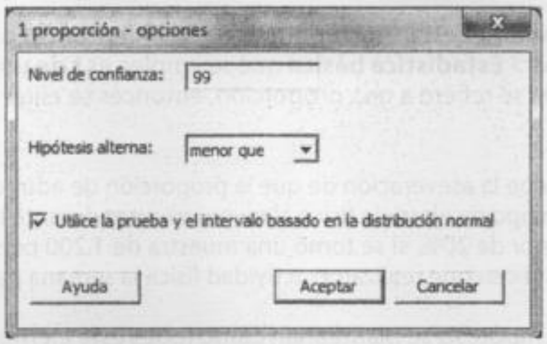
Como el problema es una prueba de una proporción, entonces la opción del menú **Estadísticas** > **Estadística básica** que se emplea es **1 Proporción**. Se debe completar el cuadro de diálogo siguiente:



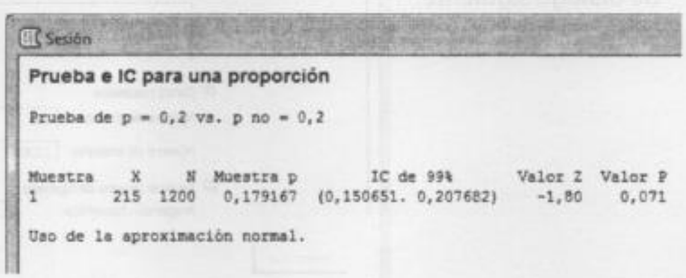
Se selecciona **Muestras en columnas** cuando se tiene la serie original de datos muestrales, pero en este caso ya se tiene calculada la cantidad de eventos o éxitos de la muestra, por lo que se escoge **Datos resumidos**, y se digita el número de eventos, que en este caso es 215, y el tamaño de la muestra o número de ensayos, que es 1200 en este caso. Debe marcarse la casilla **Realizar prueba de hipótesis**, pues de otro modo Minitab solo dará el intervalo de confianza. En la celda se digita la proporción hipotética, que es la proporción indicada en la hipótesis nula.

Solución

Luego debe darse clic en el botón **Opciones**, pues es ahí donde se indica el nivel de significancia y se selecciona la hipótesis alternativa:



En este caso, como se indicó un nivel de significancia del 1%, entonces se digita el nivel de confianza del 99%. En la hipótesis alternativa se había establecido que era $H_1: P < 0{,}20$, por lo que se elige **menor que**, y finalmente se marca la casilla **Utilice la prueba y el intervalo basado en la distribución normal**, pues así Minitab va a utilizar la aproximación normal para la distribución binomial para calcular la prueba de hipótesis y el intervalo de confianza, tal como se expuso en ese capítulo en la teoría relacionada con las pruebas de hipótesis sobre la proporción. Finalmente se da clic en Aceptar, y luego en Aceptar, y en la ventana Sesión de Minitab se obtiene el resultado siguiente:



En esta ventana se observa que Minitab ha calculado el valor del estadístico de prueba *z*, que es −1,80, y el valor *P*, que es 0,071. Por cualquiera de los dos criterios se acepta la hipótesis nula (ya que de la tabla de la curva normal, para un nivel de significancia $\alpha = 0{,}01$, con una cola, o sea, una confianza de 0,99, se obtiene $z_{\alpha} = -2{,}33$, o bien, el valor *P* de 0,071 es mayor que el de significancia $\alpha = 0{,}01$. Se concluye que no se tiene evidencia estadística suficiente para rechazar la hipótesis de que la proporción de adultos que realizaron ejercicio físico durante la semana pasada es menor que 20%, a un nivel de significación del 1%.

5.5 Intervalos de confianza y prueba de hipótesis

Los intervalos de confianza calculados anteriormente en este texto también se pueden usar para realizar la prueba de una hipótesis. Así, si el valor hipotético indicado en la hipótesis nula se encuentra dentro de los límites establecidos por el intervalo de confianza, entonces podría aceptarse la hipótesis nula. Pero si el valor hipotético está fuera de dicho intervalo, entonces podría rechazarse la hipótesis nula.

Ejemplo

La nueva directora de desarrollo de sistemas de una empresa consideró que el tiempo medio de 28 días para resolver los requerimientos de sus usuarios era demasiado. Ante esta situación optó por implementar una serie de cambios para acelerar el proceso. Seis meses después, en una muestra de 27 nuevos requerimientos se obtuvo que el tiempo promedio para resolverlos fue de 26,9 días, con una desviación estándar de 8 días. Sin embargo, algunos empleados se han quejado, y piensan que los cambios más bien retrasan el proceso. Utilizando un 1% de significancia, evalué si el tiempo medio para resolver los requerimientos de los usuarios ha cambiado.

Solución

Se inicia por plantear las hipótesis. Se desea probar que el tiempo medio para resolver los requerimientos de los usuarios es de 28 días, por lo que esa será la hipótesis nula. Por otro lado, la evidencia muestral indica que dicho tiempo se ha disminuido, pero algunos empleados opinan lo contrario, por lo que se podría plantear la hipótesis alternativa como que el tiempo medio es diferente de 28 días. En resumen, las hipótesis son:

$$H_0: \mu = 28$$
$$H_1: \mu \neq 28$$

Luego se especifica el nivel de significación α. En este ejemplo se especifica un valor de α de 0,01.

Después se selecciona el estadístico de prueba apropiado. Se tiene $n < 30$ y que la desviación estándar poblacional σ es desconocida, por lo que se calculará el intervalo de confianza usando t. Tomando una media muestral $\bar{x} = 26{,}9$ días, $n = 27$, $s = 8$ y $\alpha = 0{,}01$ (t con dos colas y 26 grados de libertad es 2,779), por lo que el intervalo de confianza será:

$$\bar{x} \pm t \cdot s/\sqrt{n} = 26{,}9 \pm 2{,}779 \cdot 8/\sqrt{27} = \begin{cases} 22{,}62 \\ 31{,}18 \end{cases}$$

La media planteada en la hipótesis nula es $\mu = 28$ días, valor que se encuentra dentro del intervalo de confianza calculado, por lo que no podría rechazarse la hipótesis nula.

Si se calculara el estadístico t se obtendría:

$$t_c = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{26{,}9 - 28}{8/\sqrt{27}} = -0{,}714$$

Ese valor calculado de t es inferior que el valor crítico de 2,779, confirmando que la hipótesis nula se acepta.

Usando Minitab se obtendría un valor $P = 0{,}481$, mayor que el nivel de significancia, por lo que se acepta la hipótesis nula.

Por cualquiera de los criterios se llega a la misma conclusión, de que se acepta la hipótesis nula, por lo que no hay evidencia suficiente para concluir que los tiempos medios para resolver los requerimientos de los usuarios haya cambiado.

Aplicación

Farmacocinética y farmacometría

Cuando usted emplea un medicamento, esa sustancia ingresa a su cuerpo y es absorbida. Luego se distribuye por el organismo y es metabolizada, es decir, el organismo detecta que es una sustancia extraña, de modo que busca eliminarla. Todo ese proceso que realiza el fármaco a su paso por su cuerpo es estudiado por la farmacocinética. Estos efectos del fármaco requieren ser medidos. De esto último se encarga la farmacometría, la cual se ocupa de desarrollar y aplicar métodos matemáticos y estadísticos para comprender y predecir los resultados que un fármaco pueda tener. Para esto se realizan diversas mediciones y análisis que permiten obtener herramientas útiles para los profesionales en el área, como las curvas dosis–respuesta, e identificar el impacto de los factores que pueden influir sobre la acción farmacológica

El uso de las probabilidades es importante, pues distintos individuos pueden tener diferentes respuestas ante la misma dosis. Así, es necesario efectuar pruebas de hipótesis sobre los eventuales efectos farmacológicos. Los errores tipo I y tipo II y el poder estadístico del estudio son muy importantes, dadas las posibles repercusiones que estos tendrían sobre un problema o una enfermedad en un paciente o en un conjunto de pacientes.

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

5.6 Ejercicios

Ejercicios de desarrollo

Resuelva los ejercicios que a continuación se presentan (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Un cierto material viene en cajas de peso promedio 17 libras y desviación estándar 0,4 libras. Se recibe un cargamento grande y se tiene la sospecha de que el peso promedio de las cajas es inferior al usual. Para verificar la sospecha se toma una muestra al azar de 86 cajas y se pesan, obteniéndose un promedio de 16,5 libras. ¿Se puede afirmar que efectivamente el peso de las cajas es inferior al acostumbrado? Use un nivel de significancia de 5%.

2. En una granja bastante grande se producen pollos. Según los estándares establecidos, el peso medio de los pollos debe ser de 4,2 Kg. con varianza 1,96. Se desea determinar si es cierta la queja de un grupo de clientes de que el peso medio ha disminuido durante las últimas semanas. Para verificar tal afirmación se contrata un ingeniero avícola, el cual toma una muestra de 65 pollos, y encuentra un peso medio de 3,86 Kg. ¿Significa esto que efectivamente el peso medio es inferior al usual? Use un nivel de significancia de 0,01.

3. De acuerdo con datos de un estudio realizado en un país europeo la edad promedio de diagnóstico del cáncer de próstata es 75 años. Un investigador nacional considera que en nuestro país esa edad de diagnóstico es menor. Se tomó una muestra de 80 casos diagnosticados y encontró una edad promedio de 69 años con una desviación estándar muestral de 9 años. ¿Qué puede concluirse a un nivel de significancia del 6%?

4. Según un estudio los niños de los estratos socio económicos medios y altos inician alguna práctica de cuidado de su salud buco dental a los 15,6 meses. En una muestra de 35 niños de familias de estratos bajos se encontró una edad media de inicio de la higiene bucal a los 18,2 meses, con una desviación estándar de 8,5 meses. ¿Puede considerarse, a una significancia del 1%, que la edad de los niños de familias de estratos bajos es mayor que 15,6 meses?

5. El jefe de producción de una planta de productos electrónicos desea estimar la producción diaria promedio de un cierto producto. La producción diaria durante los últimos 40 días da una media de 418 unidades y una desviación estándar de 45 unidades. Con estos datos pruebe la hipótesis de que la producción diaria promedio es 450 unidades, con un nivel de significancia de 0,025.

6. El gerente de una empresa procesadora de café está preocupado porque sus proveedores le entregan con frecuencia café verde. Se tomaron como muestras 36 probetas de 250 ml cada una y encontró que en promedio se entregó 30 ml de grano verde por probeta con una desviación de 10 ml. A un nivel de significación de 98%, ¿se cumple la especificación de que la cantidad de grano verde por probeta no supere 25 ml?

7. Un investigador está interesado en conocer ciertos datos relacionados con la calidad de vida de una población, entre ellos se interesa por el nivel de calorías que en promedio consumen los niños de 7 a 12 años de esa población por día. Conoce, por un estudio anterior, que la desviación estándar de la citada variable es 185 calorías, además sabe que habitan en ese poblado 745 niños. Según datos del gobierno local el consumo medio de calorías de los niños de esa edad es de 1170 calorías por día, pero el investigador duda de la veracidad de dicha información. El desea que sus estimaciones tengan altos niveles de confiabilidad (por lo menos el 97%), entonces decide realizar un estudio estadístico. Toma una muestra de 27 niños y evalúa su situación nutricional. Si una vez realizado el estudio en muestra de 27 niños, obtuvo una media de 1100 calorías diarias. ¿Es estadísticamente correcta la afirmación gubernamental, o existe evidencia estadística para rechazarla, a un nivel de significancia del 5%?

8. Se sabe que para cierto tipo de bombillo eléctrico, que su duración media es de 1250 horas. Se introduce una nueva técnica que pretende mejorar su duración, y luego se somete a una prueba una muestra al azar de 60 bombillos fabricados con el nuevo procedimiento, obteniéndose una duración media de 1271 horas y varianza de 3600. ¿Existe base para pensar que el nuevo procedimiento aumenta la duración de los bombillos?

9. El gerente de una empresa procesadora de café está preocupado porque considera que la mitad de las veces sus proveedores le entregan demasiado café verde. Se tomaron como muestras 36 probetas de 250 ml cada una y encontró que en 125 de ellas había una cantidad de grano verde que superaba lo especificado. A un nivel de significación de 99%, ¿puede considerarse que la mitad de las veces los proveedores entregan más grano verde del establecido?

10. En una muestra de 10 distintas acciones de empresas tecnológicas elegidas al azar entre todas las acciones que se negocian en la Bolsa de Valores de Nueva York se obtuvo una razón promedio de precio/utilidades (p/u) de \$12,89 con una desviación estándar de \$6,98. Si se sabe que la tasa precio/utilidades promedio para todas las empresas que cotizan en la bolsa fue de \$10,97. ¿Es esta evidencia suficiente para concluir, a un nivel de significancia del 5%, que la tasa precio/utilidades promedio para las empresas tecnológicas es superior al resto de acciones negociadas en la Bolsa?

11. Un laboratorio afirma que el tiempo promedio de eliminación del ibuprofeno en sujetos sanos es de 2,3 horas. Para determinar la validez de esa afirmación se tomó una muestra de 15 personas y se obtuvo un tiempo medio de 2,9 horas. Se conoce que la desviación estándar es 1,1 horas. ¿A qué conclusión se llega a un nivel de significancia del 5%?

12. Una persona considera que los habitantes de nuestro país tienen un cociente intelectual superior a 100 puntos, pues los niveles educativos han mejorado en años recientes. Los siguientes datos corresponden a una muestra de personas que aplicaron un test para medir su cociente intelectual: 102, 99, 96, 100, 99, 112, 125, 85, 83, 105, 98, 129, 78, 109, 116, 101, 112, 101, 96, 100, 103, 97, 104, 97, 98, 99, 113, 102, 95, 121, 98, 100, 98, 99, 111, 102, 99, 103, 108. ¿Indicar estos datos que esa persona tiene razón, al nivel de significancia de 5%?

13. En un artículo reciente de una revista se comparó el costo de adopción de niños en China y en Rusia. En una muestra de 16 adopciones en China, el costo medio fue de \$11045, con desviación estándar de \$835. En una muestra de 18 adopciones de niños en Rusia, el costo medio fue de \$12840, con una desviación estándar de \$1545. ¿Se puede concluir que el costo medio de adoptar un niño es mayor en Rusia? Utilice el nivel de significancia del 5%.

14. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es menor que 3,13?

15. Según un estudio la duración promedio de la cirugía de corrección de Incontinencia de Orina de Esfuerzo (IOE) es de 39,3 minutos. Para comprobar estos datos se tomó una muestra de 10 de estos tipos de cirugía obteniendo los siguientes tiempos operatorios (en minutos): 25, 38, 42, 74, 68, 58, 31, 21, 26 y 32. ¿Son estos datos evidencia suficiente para apoyar que el tiempo medio operatorio es de 39,3 minutos? Utilice una significancia de 1%.

16. Un arquitecto considera que actualmente las casas tienen un área construida inferior a la de hace dos o más décadas. En una muestra de 15 viviendas construidas en el transcurso de los dos últimos años se encontró un área construida promedio de 92 metros cuadrados y una desviación estándar de 30 metros cuadrados. A un nivel de significancia de 5%, ¿qué puede concluirse si se sabe, por otro estudio, que las viviendas construidas hace dos décadas o más tenían un área construida promedio de 130 metros cuadrados?

17. Un ingeniero ha recopilado datos sobre la vida útil de una muestra de filtros de gasolina del mismo tipo y marca. Los datos son los siguientes (en miles de kilómetros): 12, 14, 16, 15, 10, 12, 13, 15, 16. Pruebe la hipótesis de que la vida media de los filtros es de 15000 kilómetros. Use un nivel de significancia del 5%.

18. El cable coaxial RG-174 debe tener un diámetro de 2,6 mm. En una muestra de 25 metros de cable coaxial RG-174 fabricado por una empresa se obtuvo un diámetro medio de 2,65 mm. Se sabe que la desviación estándar es 0,9 mm. ¿Está cumpliéndose la especificación? Use un nivel de significancia del 5%.

19. Un contador está realizando una auditoría de los cheques extendidos por una compañía. En una muestra de 20 cheques se obtuvieron los siguientes montos, en miles dólares:

    15, 17, 22, 8, 4, 2, 16, 6, 8, 14, 15, 25, 20, 5, 6, 8, 12, 19, 11, 12

20. El director financiero considera que el monto promedio de los cheques de la compañía es \$11 mil dólares y el jefe del departamento contable cree que la proporción de cheques con montos superiores a \$12 mil dólares no es mayor que 45%.
    a. Pruebe, al 5% de significancia, la afirmación del director financiero contra la alternativa de que es diferente de ese monto.
    b. Pruebe, al 1% de significancia, la afirmación del jefe del departamento contable contra la alternativa de que la proporción es diferente del 45%.

21. El ingreso medio de los habitantes del país es $10.420 y la distribución del ingreso se comporta normalmente. Una muestra aleatoria de 10 residentes de una ciudad presenta una media de $8.540 con una desviación estándar de $3.560. A un nivel de significancia del 5%, ¿existe evidencia suficiente para considerar que los habitantes de esta ciudad tienen ingresos inferiores a la media nacional?

22. La desnutrición continúa siendo un problema importante en los países en desarrollo especialmente en niños. Según las recomendaciones de organismos internacionales la ración diaria no debería ser inferior a 2000 Kcal en edades de 8 a 10 años. En una zona de bajos ingresos del país se realizó un estudio entre 25 escolares y se determinó una ingesta promedio de 1875 Kcal/día. Se supone una desviación estándar de 315 Kcal. ¿Puede concluirse que se presenta algún grado de desnutrición a un nivel de significancia de 1%?

23. En una muestra de 12 motores para automóvil de cierto tipo se obtuvo una vida útil promedio de 320000 kilómetros. La desviación estándar muestral es de 90000 kilómetros. Si el vendedor ofrece una vida útil de 350000 kilómetros, ¿podría considerarse que engaña a los clientes? Use un nivel de significancia de 1%.

24. Una institución realizó un plan de capacitación entre mujeres emprendedoras de una zona marginal del país. El objetivo es que estas mujeres lograran incrementar los ingresos de sus microempresas. Luego de implementadas las capacitaciones se compararon los ingresos de una muestra de 12 microempresas y se registraron los siguientes ingresos (en dólares al mes):

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

¿Hubo un aumento significativo en los niveles de ingresos de las microempresarias de la zona? Utilice el nivel de significancia del 5%.

25. El gerente de una empresa considera que el 70% de sus clientes estarían dispuestos a efectuar compras a través de internet. Para comprobar esta afirmación se efectúa una encuesta entre 20 clientes, de los cuales 11 dijeron que no estaban dispuestos a comprar por este medio. ¿Se puede decir que el gerente tenía razón o que el resultado de la encuesta se debe a fluctuaciones del muestreo? Use $\alpha = 1\%$.

26. El encargado de reclutamiento y selección de personal de una empresa considera que las habilidades sociales, como la buena comunicación y un adecuado manejo de los conflictos, son fundamentales para lograr un buen desempeño en la compañía, y considera que menos del 25% de las personas han desarrollado estas destrezas a un nivel suficiente para laborar en la organización. En las entrevistas y pruebas que aplican a los candidatos para los puestos que demanda la empresa se evalúan estas destrezas. En una muestra aleatoria de 45 candidatos, se encontró que 19 habían obtenido una evaluación suficientemente buena. ¿Puede concluirse que el encargado de reclutamiento y selección está equivocado? Use un nivel de significancia del 5%.

27. Un candidato de un partido político considera que el 54% de los electores indecisos tienen una opinión favorable acerca de sus planteamientos. Para comprobar si su expectativa es acertada, se decidió efectuar una encuesta entre un grupo de electores, de los cuales 210 dijeron no haber definido su candidato para las próximas elecciones, y se logró determinar que 80 de esas personas tienen cierta simpatía por este candidato. ¿Puede el candidato mantener vigente su expectativa, o más bien esta información muestra lo contrario? Use $\alpha = 0{,}05$.

28. De acuerdo con un artículo publicado en una revista científica la vasectomía es exitosa en el 99% de los casos. Un investigador seleccionó 2400 casos realizados recientemente en el país y encontró 29 fallos. ¿Puede considerarse que la tasa de éxito es inferior a 99%? Use una significancia de 5%.

29. La aminotriptilina se ha aprobado para el tratamiento de la depresión. Un laboratorio indica que este fármaco es efectivo en el 85% de los casos. En una muestra de 60 pacientes con un cuadro depresivo se encontró que 45 de ellos afirmaron que la aminotriptilina les había resultado efectiva. ¿Son estos datos evidencia suficiente, al 3% de significancia, para respaldar la afirmación del laboratorio?

30. Según un artículo publicado en una revista científica la miopía podría afectar al 25% de la población de un país como Estados Unidos. Un investigador desea saber si a nivel nacional esa tasa podría ser similar. En una muestra de 120 adultos residentes en nuestro país se encontró que 18 personas tenían miopía. ¿Son estos datos evidencia suficiente de que la población de nuestro país se ve menos afectada por la miopía? Use una significancia del 5%.

31. El encargado de seguridad ocupacional de una empresa considera que en su organización ocurren menos accidentes que en la mayoría de las empresas del mismo sector económico. Según sus datos, en el país, 8 de cada 100 trabajadores se ven afectados por algún tipo de accidente. Para tratar de comprobar su afirmación se toma una muestra de 21 empleados de la empresa y se determina que 2 de ellos han sufrido algún accidente laboral este año. ¿Confirman estos datos las aseveraciones del encargado de seguridad ocupacional de la empresa? Utilice un nivel de significancia del 5%.

32. Según un estudio solo el 40% de los niños entre los 3 y 5 años de edad tienen hábitos de cuidado de su salud bucal. Luego de una campaña se tomó una muestra de 25 niños y se determinó que 13 de ellos poseía buenos hábitos en este sentido. ¿Podría concluirse que la campaña fue exitosa? Utilice un nivel de significancia de 1%.

33. Un consultor en el área de tecnologías de información considera que al menos la mitad de los proyectos informáticos no se concluyen a tiempo. En una muestra de 120 proyectos realizados en distintas empresas, se encontró que 56 no se habían concluido a tiempo. ¿Qué puede concluirse a partir de esta información? Utilice un nivel de significancia de 1%.

34. Una institución de asistencia social indicó que el año pasado el 44% de los niños de sexto grado que asisten a la escuela de una localidad solicitaron una beca para continuar sus estudios de secundaria. A inicios de este mes se seleccionó una muestra de 200 niños de sexto grado de la misma escuela y se descubrió que 96 solicitarían la beca. ¿Hubo un incremento significativo en el porcentaje de niños que solicitarían la beca en la secundaria? Lleve a cabo la prueba al 5%.

35. Una empresa pauta publicidad en televisión todos los meses. Recientemente ha lanzado una campaña muy agresiva, y se espera que al menos el 55% de los consumidores de menos de 40 años recuerden el anuncio de la empresa. Se ha tomado una muestra de 120 consumidores de menos de 40 años y el 48% dijeron que recordaban el anuncio de la empresa. ¿Puede considerarse, al 5% de significancia, que se ha alcanzado la meta?

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Cuando se debe decidir, con base en evidencia experimental, si una afirmación hecha acerca de un parámetro es falsa o verdadera, es necesario realizar:
   (a) Una estimación por intervalos
   (b) Una prueba de hipótesis
   (c) Un análisis de correlación
   (d) Un estudio por muestreo

2. Una __________ es una afirmación acerca de un __________ de una o más poblaciones y que está sujeta a verificación. La opción que mejor completa la frase anterior es:
   (a) hipótesis; parámetro
   (b) prueba de hipótesis; estimador
   (c) prueba de hipótesis; parámetro
   (d) hipótesis; estimador

3. Una prueba de hipótesis es un procedimiento basado en evidencia de la ________ y la teoría ________ para determinar si la hipótesis es una afirmación razonable. La opción que mejor completa la frase anterior es:
   (a) población; estadística
   (b) muestra; de probabilidades
   (c) probabilidad; de muestreo
   (d) población; de probabilidades

4. En una prueba de hipótesis:
   A. La hipótesis alternativa es cualquier hipótesis que se desea probar.
   B. La hipótesis nula es la hipótesis que se acepta cuando la hipótesis nula es rechazada.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

5. Un investigador desea probar la hipótesis de que la media de una determinada variable *x* es igual a 500. En una muestra obtuvo una media de 350, entonces debe:
   (a) rechazar la hipótesis nula porque la diferencia con respecto a la media muestral es muy grande
   (b) rechazar la hipótesis nula porque la media muestral es menor que la media hipotética
   (c) aceptar la hipótesis nula porque la diferencia encontrada es muy pequeña
   (d) ninguna de las anteriores

6. El nivel de significancia es la probabilidad de:
   (a) rechazar la hipótesis nula cuando es falsa
   (b) rechazar la hipótesis nula cuando es verdadera
   (c) aceptar la hipótesis nula cuando es falsa
   (d) aceptar la hipótesis nula cuando es verdadera

7. El error tipo II se comete cuando se:
   (a) rechaza la hipótesis nula cuando es falsa
   (b) rechaza la hipótesis nula cuando es verdadera
   (c) acepta la hipótesis nula cuando es falsa
   (d) acepta la hipótesis nula cuando es verdadera

8. El error tipo I se comete cuando se:
   (a) rechaza la hipótesis nula cuando es falsa
   (b) rechaza la hipótesis nula cuando es verdadera
   (c) acepta la hipótesis nula cuando es falsa
   (d) acepta la hipótesis nula cuando es verdadera

9. El gerente de una empresa espera que el 60% de sus clientes actuales estén dispuestos a efectuar compras a través de internet. Para comprobar esta afirmación se efectúa una encuesta entre una muestra de clientes, en la cual se encuentra que solo el 40% de los clientes efectuarían compras por internet. Según los estudios financieros de la empresa, se requiere que al menos el 60% de los clientes actuales realicen compras en línea para que valga la pena implementar dicha modalidad de negocios. Con respecto a esta situación se han realizado dos afirmaciones:

    A. Cometer el error tipo I significaría perder una buena oportunidad de negocios.

    B. Cometer el error tipo II significaría enfrentarse a pérdidas económicas en un sistema que no es rentable.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

10. Un candidato de un partido político considera que al menos el 40% de los electores tienen una opinión favorable acerca de sus planteamientos y, por tanto, votarían por él en las próximas elecciones. Para comprobar si su expectativa es acertada, decidió efectuar una encuesta entre un grupo de 200 electores, de los cuales 70 dijeron tener simpatía por este candidato. Las elecciones se ganan con al menos el 40% de los votos y el candidato participará solo si posee posibilidades de contar con al menos el 40% de los votos. Con respecto a esta situación se han realizado dos afirmaciones:

    A. Cometer el error tipo II significaría no participar en una elección que pudo haber ganado.

    B. Cometer el error tipo I significaría gastar muchos recursos en propaganda en una elección que no ganaría.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

11. Las autoridades sanitarias consideran que los recién nacidos procedentes de zonas rurales deberían pesar al menos 2500 gramos al nacer, en promedio. De presentarse una media inferior, valdría la pena implementar un programa para la mejora de la atención prenatal en las zonas rurales. Se decide hacer un estudio por muestreo para valorar esta decisión. Con respecto a esta situación se han realizado dos afirmaciones:

    A. Cometer el error tipo II significaría un deterioro de las condiciones de salud de una población.

    B. Cometer el error tipo I significaría destinar recursos a un programa innecesario.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

12. Con respecto al nivel de significancia usado en las pruebas de hipótesis se han realizado dos afirmaciones:
    A. Generalmente es de 1% o de 5%.
    B. Representa la posibilidad de aceptar una hipótesis incorrecta.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

13. Un contrato laboral exige los operarios realicen una producción diaria no menor de 50 unidades. Una muestra de 150 días de producción revela una media de 47,3 unidades, con una desviación estándar de 5,7 unidades, ¿se cumple con la disposición del contrato?

En este problema la hipótesis nula es:

(a) La producción media diaria es de 50 unidades.
(b) La producción media diaria es menor que 50 unidades.
(c) La producción media diaria es de 47,3 unidades.
(d) La producción media diaria es mayor que 50 unidades.

14. Un contrato laboral exige los operarios realicen una producción diaria no menor de 50 unidades. Una muestra de 150 días de producción revela una media de 47,3 unidades, con una desviación estándar de 5,7 unidades, ¿se cumple con la disposición del contrato?

En este problema la hipótesis alternativa es:

(a) La producción media diaria es de 47,3 unidades.
(b) La producción media diaria es menor que 50 unidades.
(c) La producción media diaria es de 47,3 unidades.
(d) La producción media diaria es mayor que 47,3 unidades.

15. En un colegio se estima que, cuando mucho, 25% de los estudiantes se traslada a clases en bicicleta. ¿Parecería esta ser una estimación válida si, en una muestra aleatoria de 180 estudiantes, se encuentra que 60 utilizan este transporte?

En este problema la hipótesis nula es:

(a) Una proporción de 33,33% de los estudiantes se traslada en bicicleta a clases.
(b) Una proporción de 25% de los estudiantes se traslada en bicicleta a clases.
(c) Una media de 25% de los estudiantes se traslada en bicicleta a clases.
(d) Una proporción mayor que 25% de los estudiantes se traslada en bicicleta a clases.

16. En un colegio se estima que, cuando mucho, 25% de los estudiantes se traslada a clases en bicicleta. ¿Parecería esta ser una estimación válida si, en una muestra aleatoria de 180 estudiantes, se encuentra que 60 utilizan este transporte?

En este problema la hipótesis alternativa es:

(a) Una proporción de 33,33% de los estudiantes se traslada en bicicleta a clases.
(b) Una proporción menor que 33,33% de los estudiantes se traslada en bicicleta a clases.
(c) Una media mayor de 25% de los estudiantes se traslada en bicicleta a clases.
(d) Una proporción mayor que 25% de los estudiantes se traslada en bicicleta a clases.

17. Con respecto al procedimiento de prueba de hipótesis se han realizado dos afirmaciones:

A. La prueba de hipótesis solo indica si la hipótesis es apoyada o no por los datos disponibles.

B. Cuando no se rechaza la hipótesis nula, no se dice que sea verdadera, sino que probablemente es verdadera.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

18. Con respecto al valor *P* (que ofrecen la mayoría de los programas de computación como parte de la prueba de hipótesis) se han realizado dos afirmaciones:

A. El valor P es la probabilidad de obtener un valor muestral más extremo que el observado cuando la hipótesis nula es falsa.

B. El valor P es el menor nivel de significación al que se puede rechazar la hipótesis nula cuando sea verdadera.

Con respecto a las dos afirmaciones anteriores, es correcto que:

(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

19. Si al realizar una prueba de hipótesis sobre una media de una población, se cuenta con datos de una muestra de 58 observaciones y se conoce el valor de la desviación estándar poblacional, entonces se emplea como estadístico de prueba:

(a) $z_c = \dfrac{\overline{x} - \mu}{\sigma / \sqrt{n}}$ (b) $z_c = \dfrac{\overline{x} - \mu}{s / \sqrt{n}}$

(c) $t_c = \dfrac{\overline{x} - \mu}{s / \sqrt{n}}$ (d) $z = \dfrac{x - nP}{\sqrt{nPQ}}$

20. Si al realizar una prueba de hipótesis sobre una media de una población, se cuenta con datos de una muestra de 17 observaciones y se conoce el valor de la desviación estándar poblacional, entonces se emplea como estadístico de prueba:

(a) $z_c = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ (b) $z_c = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$

(c) $t_c = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$ (d) $z = \dfrac{x - nP}{\sqrt{nPQ}}$

21. Si al realizar una prueba de hipótesis sobre una media de una población, se cuenta con datos de una muestra de 17 observaciones y no se conoce el valor de la desviación estándar poblacional, entonces se emplea como estadístico de prueba:

(a) $z_c = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ (b) $z_c = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$

(c) $t_c = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$ (d) $z = \dfrac{x - nP}{\sqrt{nPQ}}$

22. Si al realizar una prueba de hipótesis sobre una media de una población, se cuenta con datos de una muestra de 90 observaciones y no se conoce el valor de la desviación estándar poblacional, entonces se emplea como estadístico de prueba:

(a) $z_c = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ (b) $z_c = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$

(c) $t_c = \dfrac{\bar{x} - \mu}{s / \sqrt{n}}$ (d) $z = \dfrac{x - nP}{\sqrt{nPQ}}$

23. Si al realizar una prueba de hipótesis sobre una media de una población, se dice que la prueba es de dos colas, entonces es verdadero que la hipótesis alternativa puede ser (donde μ_0 es el valor hipotético de la media poblacional):

(a) $H_1: \mu > \mu0$ (b) $H_1: \mu < \mu0$

(c) $H_1: \mu \neq \mu_0$ (d) $H_1: \mu = \mu_0$

24. Si al realizar una prueba de hipótesis sobre una media de una población, se sabe que no se rechazó la hipótesis nula, entonces puede ser verdadero que:

(a) $|z_c| > |z_t|$ (b) $|t_c| > |t_t|$

(c) $|t_c| \leq |t_t|$ (d) Ninguna de las anteriores

25. Si al realizar una prueba de hipótesis sobre una media de una población, se sabe que se rechazó la hipótesis nula, entonces puede ser verdadero que:

(a) $|z_c| > |z_t|$ (b) $|t_c| \leq |t_t|$

(c) $|t_c| \leq |t_t|$ (d) Ninguna de las anteriores

26. Si al realizar una prueba de hipótesis sobre una media de una población, se sabe que se rechazó la hipótesis nula, entonces puede ser verdadero que:

(a) $|z_c| < |z_t|$ (b) valor P < α
(c) $|t_c| \leq |t_t|$ (d) valor P > α

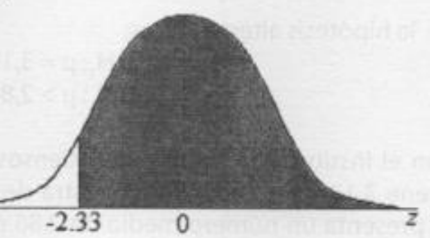
27. Si se realiza una prueba de hipótesis de dos colas, con un nivel de significancia del 5%, entonces el valor *z* crítico es:

(a) 2,00 (b) 1,645
(c) 1,28 (d) 1,96

28. Si se realiza una prueba de hipótesis de dos colas, con un tamaño de muestra de 10 observaciones y un nivel de significancia del 5%, entonces el valor *t* crítico es:

(a) 1,96 (b) 2,262
(c) 1,833 (d) 2,228

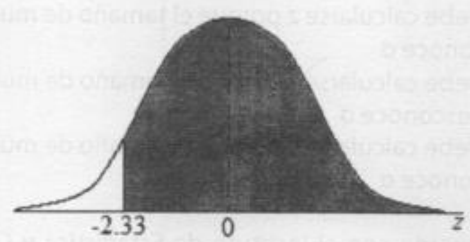
29. Observe la gráfica:



Con respecto a la gráfica anterior, es falso que:

(a) Si z_c es −2,56, se rechaza la hipótesis nula.
(b) Si z_c es −1,88, se acepta la hipótesis nula.
(c) Si z_c es −3,02, el valor P es menor que el nivel de significancia.
(d) Si $|z_c|$ es 2,33, se rechaza la hipótesis nula.

30. Observe la gráfica:



Con respecto a la gráfica anterior, es falso que:

(a) La prueba es de una cola.
(b) El nivel de significancia es 1%.
(c) La hipótesis nula puede ser H_1: μ < 50.
(d) Ninguna de las anteriores.

31. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, la hipótesis nula es:

(a) $H_0: \mu = 2,86$ (b) $H_0: \mu = 3,13$
(c) $H_0: \mu < 3,13$ (d) $H_0: \mu = 1,2$

32. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, la hipótesis alternativa es:

(a) $H_1: \mu = 2,86$ (b) $H_1: \mu = 3,13$
(c) $H_1: \mu < 3,13$ (d) $H_1: \mu > 2,86$

33. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, es cierto que:

(a) Debe calcularse z porque el tamaño de muestra es menor que 30 y se desconoce σ
(b) Debe calcularse z porque el tamaño de muestra es menor que 30 y se conoce σ
(c) Debe calcularse t porque el tamaño de muestra es menor que 30 y se desconoce σ
(d) Debe calcularse t porque el tamaño de muestra es menor que 30 y se conoce σ

34. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, es cierto que:

(a) Es una prueba de una cola porque en la hipótesis nula se emplea el signo <

(b) Es una prueba de una cola porque en la hipótesis alternativa se emplea el signo <

(c) Es una prueba de una cola porque en la hipótesis alternativa se emplea el signo ≠

(d) Es una prueba de dos colas porque en la hipótesis alternativa se emplea el signo <

35. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, el valor del estadístico de prueba es:

(a) $z = -1{,}13$ (b) $t = -1{,}13$

(c) $t = 1{,}711$ (d) Ninguna de las anteriores

36. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, al calcular el valor tabular crítico para hacer la prueba es cierto que:

(a) Los grados de libertad son 25

(b) Los grados de libertad son 24

(c) Los grados de libertad son 26

(d) No se necesita determinar los grados de libertad

37. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, el valor tabular crítico para hacer la prueba es:

(a) $z = -1{,}645$ (b) $t = 1{,}711$

(c) $t = -1{,}711$ (d) $t = 2{,}064$

38. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentes. A un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, al hacer la prueba es verdadero que:

(a) Se acepta la hipótesis nula porque $tc < t_t$
(b) Se acepta la hipótesis nula porque $|tc| < |t_t|$
(c) Se rechaza la hipótesis nula porque $tc > t_t$
(d) Se acepta la hipótesis nula porque $|z_c| < |z_t|$

39. De acuerdo con el Instituto de Estadística y Censos, un hogar típico en nuestro país tiene 3,13 miembros. Una muestra de 25 hogares del área metropolitana presenta un número medio de 2,86 miembros. La desviación estándar de esta muestra era de 1,2 residentesvA un nivel de significancia del 5%, ¿es razonable concluir que el número medio de residentes de esta ciudad es igual a la media nacional?

Para este problema, al hacer la prueba se puede concluir, con respecto al tamaño medio de los hogares del área metropolitana, que:

(a) El tamaño medio es 2,86 miembros
(b) El tamaño medio es 3,13 miembros
(c) No hay evidencia suficiente para decir que el tamaño medio es menor que 3,13 miembros
(d) Hay evidencia suficiente para decir que el tamaño medio es menor que 3,13 miembros

CAPÍTULO 6

PRUEBAS DE HIPÓTESIS PARA LA DIFERENCIA DE DOS MEDIAS O PROPORCIONES POBLACIONALES

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Identificar los distintos tipos de problemas para la diferencia de dos medias poblacionales
2. Aplicar el procedimiento de prueba de hipótesis para problemas de medias de dos poblaciones independientes con distintos tamaños de muestra
3. Aplicar el procedimiento de prueba de hipótesis para problemas de medias con datos apareados
4. Aplicar el procedimiento de prueba de hipótesis para problemas de diferencia de dos proporciones

6.1 Prueba de hipótesis para la diferencia entre medias

Una empresa posee operaciones en dos países distintos y en cada país posee una planta de producción. En uno de los países se tienen 2000 empleados y en el otro 3000. En ambas plantas se realizan los mismos procesos, pero se ha observado que, según algunos datos muestrales, la productividad de los operarios tiende a ser mayor en uno de los países que en el otro. Dado que los datos provienen de muestras, es necesario saber si las diferencias observadas entre ambas plantas son significativas, o si pueden ser atribuidas al azar. Para resolver este problema se necesita realizar una prueba de hipótesis para la media de dos poblaciones distintas.

Tal como en el ejemplo, muchas veces es necesario decidir si la diferencia entre dos medias muestrales se puede atribuir al azar, o si en realidad las dos muestras provienen de poblaciones con medias diferentes. Otros ejemplos de problemas en que se comparan dos medias se pueden referir a comparar dos métodos de soldadura para encontrar cuál es más resistente, comparar el rendimiento de llantas radiales versus llantas con neumático, comparar dos métodos de enseñanza, etc. A continuación se expone el procedimiento que se utilizará para probar si la diferencia observada entre las dos medias muestrales es estadísticamente significativa.

6.2 Diferencia entre medias en poblaciones independientes

Si $\bar{x}_1$ y $\bar{x}_2$ son las medias de dos muestras aleatorias *independientes*, entonces la distribución muestral de la diferencia $\bar{x}_1 - \bar{x}_2$ tiene la media $\mu_1 - \mu_2$ y la desviación estándar $\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$ donde μ_1, μ_2, σ_1, σ_2 son las medias y las desviaciones estándar de las dos poblaciones muestreadas.

En estos casos se plantea la hipótesis nula como:

$$H_0: \mu_1 - \mu_2 = \delta$$

donde δ es la diferencia que se desea probar. Este valor δ es una constante determinada con base en muestras aleatorias e independientes de tamaño n_1 y n_2. La mayor parte de las veces se tendrá que $\delta = 0$, es decir, se prueba si las dos medias son iguales, por lo que la hipótesis nula podría plantearse como:

$$H_0: \mu_1 = \mu_2$$

La hipótesis alternativa se puede formular como cualquiera de las tres siguientes:

$$H_1: \mu_1 - \mu_2 > \delta$$
$$H_1: \mu_1 - \mu_2 < \delta$$
$$H_1: \mu_1 - \mu_2 \neq \delta$$

El estadístico de prueba es:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

cuando las muestras son grandes y las varianzas poblacionales conocidas.

Los criterios de aceptación o rechazo son iguales a los de la prueba para una media.

Ejemplo

Una empresa posee operaciones en dos países distintos y en cada país posee una planta de producción. En uno de los países se tienen 2000 empleados y en el otro 3000. En ambas plantas se realizan los mismos procesos, pero se ha observado que, según algunos datos muestrales, la productividad de los operarios tiende a ser mayor en uno de los países que en el otro.

Los datos recopilados se muestran en la tabla (la media y la desviación estándar se expresan en número de unidades producidas correctamente por hora):

Planta de producción	Tamaño de muestra n	Media $\bar{x}$	Desviación estándar σ
En el país 1	40	22	3,1
En el país 2	50	31	4,2

Determine, a un nivel de significación del 5%, si se presenta diferencia entre los dos promedios.

Solución

Se tiene que $n_1 = 40$, $n_2 = 50$, $\bar{x}_1 = 22$, $\bar{x}_2 = 31$, $\sigma_1 = 3,1$ y $\sigma_2 = 4,2$.

El ejercicio busca determinar si existe diferencia, por lo que se tendrá que probar si $\delta = 0$.

Además, se indica que $\alpha = 0,05$.

Entonces, se plantea la hipótesis nula como:

$$H_0: \mu_1 - \mu_2 = 0$$

Como $\bar{x}_1 = 22 < \bar{x}_2 = 31$, entonces se formula la hipótesis alternativa como:

$$H_1: \mu_1 - \mu_2 < 0$$

Se usa Z porque aunque las desviaciones estándar poblacionales son desconocidas, se tienen muestras grandes ($n \geq 30$):

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(22 - 31) - 0}{\sqrt{\frac{3{,}1^2}{40} + \frac{4{,}2^2}{50}}} = -11{,}69$$

De la tabla normal con un nivel de significación del 5% se obtiene $Z_\alpha = 1{,}645$. El valor de Z calculado con la fórmula es menor que el Z_α por lo tanto cae en la zona de rechazo de la hipótesis nula. Se concluye que se rechaza H_o con $\alpha = 0{,}05$. Se puede decir que existe evidencia suficiente para creer que la productividad en el primer país es menor que en el segundo.

Ejemplo

Un vendedor de equipo tecnológico quiere determinar si hay diferencias en el consumo de este tipo de productos por parte de entre los profesionales en ciencias económicas y profesionales en ingeniería, pues ha tenido muy buena experiencia vendiendo equipos para el primer profesional mencionado.

Seleccionó una muestra al azar de 80 profesionales en ciencias económicas y 70 ingenieros, encontrando que los primeros gastaron un promedio de $1.250 en productos tecnológicos durante el último año, con una desviación estándar de $400. Los ingenieros gastaron en promedio $980, con una desviación estándar de $620. ¿Existe diferencia significativa, al 1% de significancia entre ambas poblaciones?

Solución

En esta situación se tienen los datos para los dos grupos de profesionales, las cuales se pueden resumir del modo siguiente:

	1	2
Grupo	Ciencias económicas	Ingeniería
Tamaño de muestra	80	70
Promedio	$1.250	$980
Desviación estándar	$400	$620

Se plantean la hipótesis nula como la igual de las dos medias, o sea, que la diferencia es cero:

$$H_0: \mu_1 - \mu_2 = 0$$

También podrían plantearse la hipótesis nula como:

$$H_0: \mu_1 = \mu_2$$

De acuerdo con la evidencia de la muestra, el promedio para los profesionales en ciencias económicas es mayor, por lo que la hipótesis alternativa podría plantearse como:

$$H_1: \mu_1 > \mu_2$$

Así, las hipótesis serían:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

Dado que se cuenta con tamaños de muestra superiores a 30 unidades, y se conocen las desviaciones estándares poblacionales, entonces se aplica el estadístico de prueba z:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(1250 - 980) - 0}{\sqrt{\frac{400^2}{80} + \frac{620^2}{70}}} = 3{,}12$$

La prueba se realiza a un nivel de significancia del 1%, por lo que de la tabla de la curva normal estándar se obtiene $z_t = 2{,}33$.

Dado que el valor crítico $|z_c| = 3{,}12$ es mayor que $|z_t| = 2{,}33$, entonces se rechaza la hipótesis nula. Se concluye que a un nivel de significancia del 1% existe evidencia suficiente para considerar que el consumo de productos tecnológicos por parte de entre los profesionales en ciencias económicas es mayor que el de los profesionales en ingeniería.

Ejercicio de revisión

Un analista de inversiones requiere asesorar a un cliente con respecto a los riesgos de invertir en las acciones de dos compañías distintas llamadas MuchMoney y VeryRich. Para ello toma una muestra de 40 variaciones diarias en los precios de MuchMoney y obtiene un promedio de \$2,8 con una desviación estándar de \$1,2; y una muestra de 50 variaciones diarias de los precios de VeryRich, las cuales dan una media de \$3,5 con una desviación estándar de \$1,8. ¿Es esta evidencia suficiente para considerar que el comportamiento de ambas acciones es el mismo o son diferentes?

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Minitab para realizar la prueba de dos medias

Ejemplo

Utilice Minitab para resolver el problema: Un vendedor de equipo tecnológico quiere determinar si hay diferencias en el consumo de este tipo de productos por parte de entre los profesionales en ciencias económicas y profesionales en ingeniería, pues ha tenido muy buena experiencia vendiendo equipos para el primer profesional mencionado. Seleccionó una muestra al azar de 80 profesionales en ciencias económicas y 70 ingenieros, encontrando que los primeros gastaron un promedio de $1.250 en productos tecnológicos durante el último año, con una desviación estándar de $400. Los ingenieros gastaron en promedio $980, con una desviación estándar de $620. ¿Existe diferencia significativa, al 1% de significancia entre ambas poblaciones?

Solución

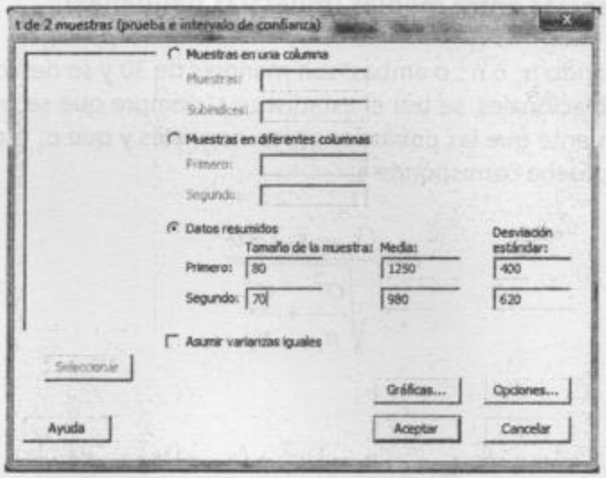
En esta situación se tienen los datos para los dos grupos de profesionales, las cuales se pueden resumir del modo siguiente:

	1	2
Grupo	Ciencias económicas	Ingeniería
Tamaño de muestra	80	70
Promedio	$1.250	$980
Desviación estándar	$400	$620

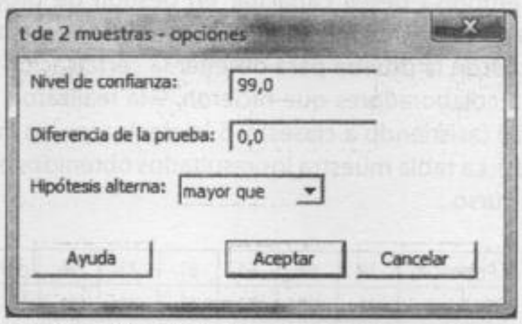
Se plantean la hipótesis nula y la hipótesis alternativa:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

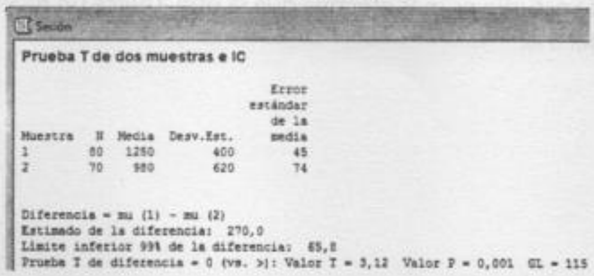
Dado que se cuenta con tamaños de muestra superiores a 30 unidades, y se conocen las desviaciones estándares poblacionales, entonces se aplica el estadístico de prueba *z*, pero en Minitab no aparece en el menú Estadísticas / Estadística básica una prueba "*z* de 2 muestras", sino que solo aparece "*t* de 2 muestras". Sin embargo, la distribución normal y la distribución *t* convergen conforme se incrementa el tamaño de la muestra, por lo que usando el menú "*t* de 2 muestras" se obtendrán resultados bastante aproximados. Entonces, se da clic al menú Estadísticas, luego Estadística básica y se selecciona *t* de 2 muestras, y se completa el cuadro de diálogo siguiente:



En el cuadro anterior se marcó la opción datos resumidos, pues ya se cuenta con los cálculos de la media y la desviación estándar en cada caso. En el botón opciones se indica el nivel de confianza, que en este caso sería de 99%, pues la significancia es de 1%. La diferencia de la prueba es cero, ya que se prueba la hipótesis nula de que ambas medias son iguales. Y la hipótesis alternativa corresponde a que la primera media es mayor que la segunda, por lo que se indica "mayor que":



Después se da clic en Aceptar, y luego Aceptar en el primer cuadro de diálogo, y en la ventana Sesión se obtiene:



Se observa el valor de $T = 3,12$, que en este caso es igual al valor calculado de z, y además se cuenta con el valor $P = 0,001$. Por cualquiera de los dos criterios se rechaza la hipótesis nula.

6.3 Diferencias entre medias (muestras pequeñas)

Cuando n_1 o n_2, o ambas, son menores de 30 y se desconocen las varianzas poblacionales, se usa el estadístico t, siempre que se pueda suponer razonablemente que las poblaciones son normales y que $\sigma_1 = \sigma_2 = \sigma$. El estadístico de prueba corresponde a:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}}$$

donde:

$$\sigma^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Los grados de libertad son $gl = n_1 + n_2 - 2$. Se emplea la tabla *t* (Apéndice 7) para obtener los valores *t* tabulares.

Ejemplo

Una empresa desea capacitar en gestión de proyectos a todos sus profesionales. Una muestra de 15 empleados realizó la capacitación y efectuaron la prueba para obtener la certificación en dicho tema. De los 15 colaboradores que hicieron, 9 la realizaron en modalidad presencial (asistiendo a clases) y 6 la efectuaron en línea (a través de internet). La tabla muestra los resultados obtenidos en la prueba final de cada curso.

Presencial	79	88	54	81	73	56	79	64	58
En línea	70	80	72	52	70	61			

El departamento de recursos humanos desea saber si una modalidad de estudio es más efectiva que la otra. Utilice un nivel de significación del 5%.

Solución

Se tienen los datos:

Modalidad presencial: $n_1 = 9$, $\bar{x}_1 = 70{,}2$, $s_1 = 12{,}5$

Modalidad en línea: $n_2 = 6$, $\bar{x}_2 = 67{,}50$, $s_2 = 9{,}71$

Además, $\alpha = 0{,}05$.

Como no se especifica el valor de la diferencia, puede suponerse que va a ser cero, por lo que $\delta = 0$. Además como $\bar{x}_1 > \bar{x}_2$, se plantean las hipótesis como:

$$H_0: \mu_1 - \mu_2 = 0$$
$$H_1: \mu_1 - \mu_2 > 0$$

También, podrían plantearse las hipótesis como:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

Debido a que las desviaciones estándar poblacionales son desconocidas y se tienen muestras pequeñas ($n < 30$) se usa t. Para esto se supone que las poblaciones son normales y que $\sigma_1 = \sigma_2$. Se calcula:

$$\sigma^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(9-1)12{,}5^2 + (6-1)9{,}71^2}{9+6-2} = 132{,}42$$

Luego se calcula t:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}} = \frac{(70{,}2 - 67{,}5) - 0}{\sqrt{\dfrac{132{,}42}{9} + \dfrac{132{,}42}{6}}} = 0{,}45$$

Aplicando la distribución t:

$$gl = n_1 + n_2 - 2 = 9 + 6 - 2 = 13$$

Entonces de la tabla con $\alpha = 0{,}05$, se obtiene $t_\alpha = 1{,}771$.

El valor de t calculado con la fórmula es menor que el t_α, por lo tanto, cae en la zona de aceptación de la hipótesis nula. Se acepta H_o con $\alpha = 0{,}05$. Se puede decir que la diferencia entre los promedios de ambas modalidades de estudio no es estadísticamente significativa.

Ejemplo

Los datos corresponden a los tiempos, en minutos, requeridos por una muestra de empleados de dos departamentos distintos de una empresa en un simulacro de evacuación de las instalaciones, esto como parte de la preparación que se realiza ante eventuales situaciones de emergencia, como terremotos o incendios.

Depto. 1	5	3	4	1	3	4	9	2
Depto. 2	4	2	5	4	6	3	2	

Se desea saber a un nivel de significación del 5% si la diferencia de los tiempos promedio de los dos grupos es significativa.

Solución

Se tienen los datos: $n_1 = 8$, $\bar{x}_1 = 3{,}875$, $s_1 = 2{,}416$, $n_2 = 7$, $\bar{x}_2 = 3{,}714$, $s_2 = 1{,}496$, $\alpha = 0{,}05$. Como no se especifica el valor de la diferencia, puede suponerse que va a ser cero, por lo que $\delta = 0$. Además como $\bar{x}_1 > \bar{x}_2$, se plantean las hipótesis como:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

Debido a que las desviaciones estándar poblacionales son desconocidas y se tienen muestras pequeñas ($n < 30$) se usa t. Para esto se supone que las poblaciones de los tiempos son normales y que $\sigma_1 = \sigma_2$. Se calcula:

$$\sigma^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(8-1)2{,}416^2 + (7-1)1{,}496^2}{8+7-2} = 4{,}176$$

Luego se calcula t:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} = \frac{(3{,}875 - 3{,}714) - 0}{\sqrt{\frac{4{,}176}{8} + \frac{4{,}176}{7}}} = 0{,}152$$

Aplicando la distribución t:

$$gl = n_1 + n_2 - 2 = 8 + 7 - 2 = 13$$

Entonces $t_\alpha = 1{,}771$ de la tabla con $\alpha = 0{,}05$.

El valor de t calculado con la fórmula es menor que el t_α, por lo tanto, cae en la zona de aceptación de la hipótesis nula. Se acepta H_o con $\alpha = 0{,}05$. Se puede decir que la diferencia entre los promedios de los tiempos de evacuación de los dos departamentos no es estadísticamente significativa.

Ejercicio de revisión

Un investigador desea determinar si la tasa de mortalidad anestésica se ha incrementado en los hospitales del país durante el último año. Toma una muestra de 15 casos de pacientes anestesiados durante este último mes, de los cuales fallecieron por anestesia dos de ellos, y una muestra de 13 casos de pacientes anestesiados para el mismo mes del año pasado, y los registros indican que falleció solamente uno. ¿Son estos datos evidencia suficiente para concluir que la mortalidad anestésica se ha incrementado? Use un nivel de significancia del 1%.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Minitab para realizar la prueba de dos medias con n pequeña

Ejemplo

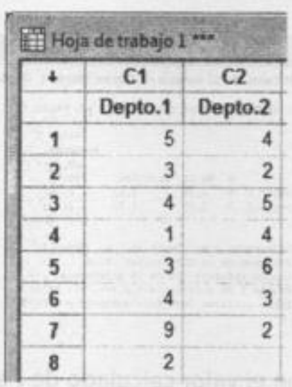
Utilice Minitab para resolver el problema: Los datos corresponden a los tiempos, en minutos, requeridos por una muestra de empleados de dos departamentos distintos de una empresa en un simulacro de evacuación de las instalaciones, esto como parte de la preparación que se realiza ante eventuales situaciones de emergencia, como terremotos o incendios.

Depto. 1	5	3	4	1	3	4	9	2
Depto. 2	4	2	5	4	6	3	2	

Se desea saber a un nivel de significación del 5% si la diferencia de los tiempos promedio de los dos grupos es significativa.

Solución

Primero que todo se introducen los datos en columnas de la hoja de trabajo de Minitab:

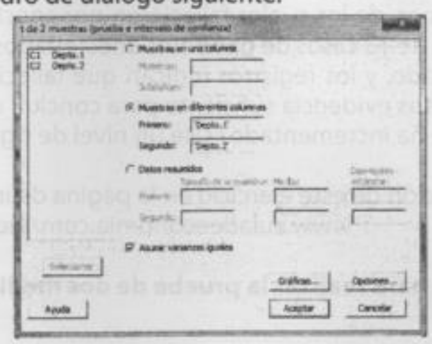

Hoja de trabajo 1 ***

↓	C1	C2
	Depto.1	Depto.2
1	5	4
2	3	2
3	4	5
4	1	4
5	3	6
6	4	3
7	9	2
8	2	

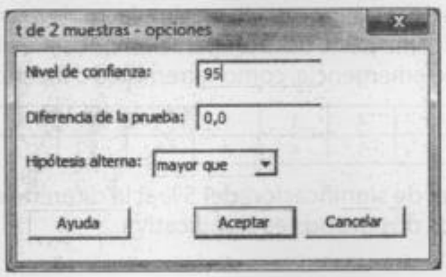
Resulta útil calcular cada una de las medias, para saber que $\bar{x}_1 = 3,875$ y que $\bar{x}_2 = 3,714$, y se plantean las hipótesis como:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 > \mu_2$$

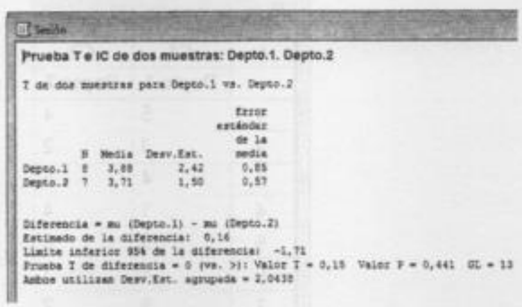
Debido a que las desviaciones estándar poblacionales son desconocidas y se tienen muestras pequeñas ($n < 30$) se usa t. Para esto se da clic en el menú Estadísticas / Estadística básica / t de 2 muestras, y se completa el cuadro de diálogo siguiente:



En el cuadro de diálogo anterior se marca la opción de Muestras en diferentes columnas, pues los datos de cada departamento se introdujeron en una columna separada. Además es necesario marcar la casilla Asumir varianza iguales. Luego en el botón opciones se indica el nivel de confianza, que sería de 95%, la diferencia de la prueba, que es cero, y el signo de la hipótesis alternativa, que es mayor que:



Al dar clic en Aceptar se obtiene en la ventana Sesión la solución siguiente:



Sesión

Prueba T e IC de dos muestras: Depto.1. Depto.2

```
T de dos muestras para Depto.1 vs. Depto.2

                                 Error
                              estándar
                                 de la
         N  Media  Desv.Est.     media
Depto.1  8   3,88       2,42      0,85
Depto.2  7   3,71       1,50      0,57


Diferencia = mu (Depto.1) - mu (Depto.2)
Estimado de la diferencia:  0,16
Límite inferior 95% de la diferencia:  -1,71
Prueba T de diferencia = 0 (vs. >): Valor T = 0,15  Valor P = 0,441  GL = 13
Ambos utilizan Desv.Est. agrupada = 2,0438
```

Se observa que el valor calculado de t es 0,15, menor que el t_{α}, por lo tanto, cae en la zona de aceptación de la hipótesis nula. O bien, se usa el valor $P = 0{,}441$. Se acepta H_o con $\alpha = 0{,}05$. Se puede decir que la diferencia entre los promedios de los tiempos de evacuación de los dos departamentos no es estadísticamente significativa.

6.4 Observaciones pareadas

Suponga la situación siguiente. Los empleados de un departamento de una empresa han realizado un simulacro de evacuación de las instalaciones, esto como parte de la preparación que se realiza ante eventuales situaciones de emergencia, como terremotos o incendios, y se obtuvo, en una muestra de 8 empleados un tiempo medio de evacuación de 5,25 minutos. Se considera que ese tiempo es muy alto, por lo que se implementa un plan para informar al personal sobre los planes de emergencias de la empresa. Luego de estas medidas se vuelve a realizar el simulacro, y los mismos 8 empleados promedian 4,5 minutos. Aunque se presenta una mejora, existe la duda de si esa diferencia es significativa estadísticamente. Para resolver un problema de este tipo también debe realizarse una prueba de hipótesis para la diferencia de dos medias, como en las secciones anteriores, pero con la diferencia de que en este caso las observaciones son pareadas, es decir, la segunda muestra de datos corresponde a las mismas unidades de estudio de la primera muestra, por lo que los datos de la primera muestra y la segunda no son independientes.

Si las poblaciones de donde se toman las muestras no son independientes, como en el caso de experimentos de "antes" y "después" y muchas otras situaciones en las que los datos están naturalmente apareados, se usa la prueba *t* para la diferencia de los datos apareados.

La hipótesis nula es $H_0: \mu_d = 0$, porque si μ_d que es el promedio poblacional de las diferencias entre las dos poblaciones es cero, es porque, en promedio, las dos poblaciones son iguales.

Si se denota como $\overline{x}_d$ al promedio de la diferencia de las dos muestras, la hipótesis alternativa puede ser:

$$H_1: \mu_d > 0 \text{ si } \overline{x}_d > 0$$

$$H_1: \mu_d < 0 \text{ si } \overline{x}_d < 0$$

$$H_1: \mu_d \neq 0$$

El estadístico de prueba es:

$$t = \frac{\overline{x}_d}{s_d / \sqrt{n}}$$

Y los grados de libertad son $gl = n - 1$.

Ejemplo

Los empleados de un departamento de una empresa han realizado un simulacro de evacuación de las instalaciones, esto como parte de la preparación que se realiza ante eventuales situaciones de emergencia, como terremotos o incendios, y se obtuvo, en una muestra de 8 empleados un tiempo medio de evacuación de 5,25 minutos. Se considera que ese tiempo es muy alto, por lo que se implementa un plan para informar al personal sobre los planes de emergencias de la empresa. Luego de estas medidas se vuelve a realizar el simulacro, y los mismos 8 empleados promedian 4,5 minutos. La tabla muestra los tiempos antes y después de las medidas implementadas. Aunque se presenta una mejora, existe la duda de si esa diferencia es significativa estadísticamente.

Antes	7	4	5	3	4	5	10	4
Después	5	3	5	4	6	4	6	3

Se desea saber a un nivel de significación del 5% si la diferencia de los tiempos promedio es significativa.

Solución

Primero se calculan las diferencias, d_r, entre el "antes" y el "después" para cada una de las observaciones, o sea, se resta el dato "antes" menos el dato "después":

Antes	7	4	5	3	4	5	10	4
Después	5	3	5	4	6	4	6	3
Diferencia	2	1	0	-1	-2	1	4	1

Con estas diferencias se calcula la media de las diferencias y su desviación estándar:

$$\bar{x}_d = 0{,}75$$
$$s_d = 1{,}832$$

Se plantean las hipótesis:

$$Ho: \mu_d = 0$$
$$H_1: \mu_d > 0$$

Se calcula t:

$$t = \frac{\bar{x}_d}{s_d/\sqrt{n}} = \frac{0{,}75}{1{,}832/\sqrt{8}} = 1{,}158$$

Se tienen $gl = n - 1 = 8 - 1 = 7$, con $\alpha = 0{,}05$, con una cola, por lo que, de la tabla, $t_\alpha = 1{,}895$.

El valor de t calculado es menor que el t_α, por lo tanto, cae en la zona de aceptación de la hipótesis nula. Se rechaza H_o con $\alpha = 0{,}05$. Se puede decir que las medidas implementadas no han sido efectivas.

Ejemplo

La tabla muestra las cantidades producidas por hora elaboradas por 8 operarios antes de recibir un entrenamiento y las cantidades producidas luego de la misma.

Antes	8	8	9	6	9	7	12	12
Después	6	10	7	11	9	12	14	8

Pruebe la afirmación de que la capacitación ha sido efectiva, al nivel de significancia de 0,05.

Solución

Primero se calculan las diferencias, d_i: 2, –2, –2, –5, 0, –5, –2, 4.

Con estas diferencias se calcula: $\bar{x}_d = -0,75$ y $s_d = 3,33$.

Se plantean las hipótesis:

$$\text{Ho: } \mu_d = 0$$
$$H_1: \mu_d < 0$$

Se calcula t:

$$t = \frac{\bar{x}_d}{s_d/\sqrt{n}} = \frac{-0,75}{3,33/\sqrt{8}} = -0,637$$

Se tienen $gl = n - 1 = 8 - 1 = 7$, con $\alpha = 0,05$, por lo que, de la tabla de la distribución t se obtiene $t_\alpha = 1,895$.

El valor de t calculado, en valor absoluto, es menor que el t_α, por lo tanto, cae en la zona de aceptación de la hipótesis nula. Se acepta H_o con $\alpha = 0,05$. Se puede decir que no hay evidencia estadística suficiente para concluir que el entrenamiento ha sido efectivo.

Ejercicio de revisión

Un ingeniero desea probar la hipótesis de que los neumáticos para automóviles fabricados en el país son de tanta calidad como los importados. Con este fin toma una muestra de 7 vehículos los cuales serán acelerados hasta 100 km/h y luego serán frenados en seco y en cada caso se medirá la distancia de frenado. La prueba será aplicada a los mismos 7 vehículos, primero con los neumáticos nacionales y luego con los importados. Después de realizar las pruebas se obtuvieron los siguientes datos (distancias de frenado en metros):

Automóvil	1	2	3	4	5	6	7
Neumático nacional	142	138	144	146	150	137	141
Neumático importado	140	139	142	139	141	137	135

Pruebe la hipótesis al 5% de significancia.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Minitab para realizar la prueba con observaciones pareadas

Ejemplo

Utilice Minitab para resolver el siguiente problema. La tabla muestra las cantidades producidas por hora elaboradas por 8 operarios antes de recibir un entrenamiento y las cantidades producidas luego de la misma.

Antes	8	8	9	6	9	7	12	12
Después	6	10	7	11	9	12	14	8

Pruebe la afirmación al nivel de 0,05, de que la capacitación ha sido efectiva.

Solución

En Minitab lo primero que se realiza es la introducción de los datos en dos columnas distintas de la hoja de trabajo:


Hoja de trabajo 1 ***

↓	C1	C2	C3
	Antes	Después	
1	8	6	
2	8	10	
3	9	7	
4	6	11	
5	9	9	
6	7	12	
7	12	14	
8	12	8	

Se plantean las hipótesis:

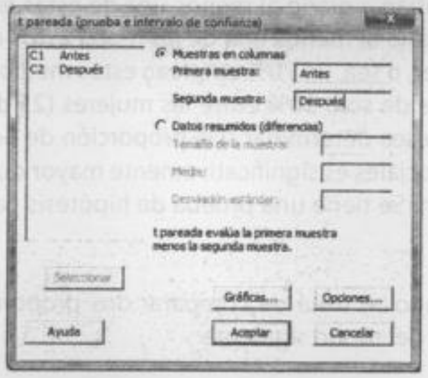
$$Ho: \mu_d = 0$$
$$H_1: \mu_d < 0$$

Para esto se da clic en el menú Estadísticas / Estadística básica / t pareada, y se completa el cuadro de diálogo siguiente, indicando en Muestras en columnas las columnas en que se hallan los datos:

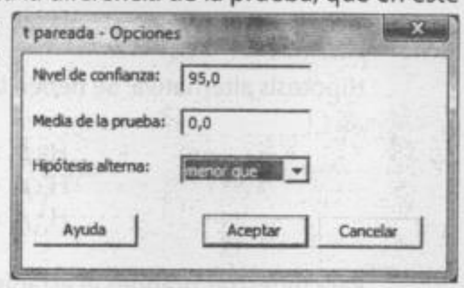
Se plantean las hipótesis:

$$H_o: \mu_d = 0$$
$$H_1: \mu_d < 0$$

Para esto se da clic en el menú Estadísticas / Estadística básica / t pareada, y se completa el cuadro de diálogo siguiente, indicando en Muestras en columnas las columnas en que se hallan los datos:



En el botón Opciones se indica la diferencia de la prueba, que en este caso es cero, y el signo de la prueba, que es el menor que de la hipótesis alternativa:



Luego se da clic en Aceptar y Minitab da el resultado en la ventana Sesión:

En esta salida se observa el valor de t calculado de -0,64, que es necesario comparar con el valor t tabular. También se puede hacer la prueba empleando el valor $P = 0,272$. En cualquier caso, se acepta H_o con $\alpha = 0,05$. Se puede decir que no hay evidencia estadística suficiente para concluir que el entrenamiento ha sido efectivo.

6.5 Pruebas para la diferencia de dos proporciones

Un investigador en el área de tecnología tiene la hipótesis de que los hombres tienden a emplear las tecnologías de la información y comunicación más que las mujeres. Específicamente, quiere determinar si hay diferencias entre hombres y mujeres en el uso de las redes sociales en internet. Para este fin toma una muestra de 40 hombres y 50 mujeres, y obtuvo que de ellos 28 hombres empleaban a diario al menos una de estas redes y 25 mujeres también usaban a diario al menos una de las redes. Estos datos muestran que 28 de los 40 hombres, o sea, un 70% emplean estos medios a diario, mientras que ese porcentaje es de solo 50% entre las mujeres (25 de 50). En un problema de este tipo se desea determinar si la proporción de hombres que emplean a diario las redes sociales es significativamente mayor que la de las mujeres que también lo hacen. Se tiene una prueba de hipótesis para la diferencia de dos proporciones.

Así, cuando se trata de comparar dos proporciones, las hipótesis se pueden plantear del modo siguiente:

Hipótesis nula:

$$H_o: p_1 = p_2$$

Hipótesis alternativa: Se tienen las siguientes tres posibilidades:

$$H_1: p_1 > p_2$$
$$H_1: p_1 < p_2$$
$$H_1: p_1 \neq p_2$$

Para muestras grandes el estadístico de prueba es:

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

donde se tiene que:

$$p = \frac{p_1 + p_2}{n_1 + n_2}$$

$$q = 1 - p.$$

El criterio de aceptación o rechazo de la prueba es igual a los casos anteriores estudiados.

Ejemplo

Un investigador en el área de tecnología quiere determinar si hay diferencias en el uso de las redes sociales en internet entre hombres y mujeres. Para este fin toma una muestra de 40 hombres y 50 mujeres, y obtuvo que de ellos 28 hombres empleaban a diario al menos una de estas redes y 25 mujeres también usaban a diario al menos una de las redes. Con base en esos datos y a un nivel de significancia de 5%, ¿puede concluirse que existe diferencia significativa entre hombres y mujeres e n cuanto a su frecuencia de uso de las redes sociales en internet?

Solución

Se cuenta con la siguiente información:

$$\text{Hombres: } p_1 = 28/40 = 0{,}70$$
$$\text{Mujeres: } p_2 = 25/50 = 0{,}50$$

Se plantean las hipótesis:

$$H_o: p_1 = p_2$$
$$H_1: p_1 > p_2$$

Primero se calculan p y q:

$$p = \frac{p_1 + p_2}{n_1 + n_2} = \frac{28 + 25}{40 + 50} = 0{,}59$$

$$q = 1 - 0{,}5889 = 0{,}41$$

Luego se calcula z:

$$z = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0{,}70 - 0{,}50}{\sqrt{0{,}59 \cdot 0{,}41\left(\frac{1}{40} + \frac{1}{50}\right)}} = 1{,}92$$

De la tabla se obtiene $Z_{\alpha/2} = 1{,}96$. Como puede verse en el gráfico, el valor de Z_c cae en la zona de aceptación de H_o, por lo tanto se acepta H_0 con $\alpha = 0{,}05$. La diferencia no es estadísticamente significativa. La proporción de hombres que emplea a diario las redes sociales en internet no es significativamente diferente de la proporción de mujeres que realizan esta actividad.

Ejemplo

En un lote de 500 piezas fabricadas esta semana en una línea de ensamblado se obtuvieron 18 con defectos. En otro lote de 400 piezas tomadas de otra línea de ensamblado se obtuvieron 25 defectuosas. Determine si las líneas producen la misma proporción de piezas con defectos, al nivel de significación de 5%.

Se cuenta con la siguiente información:

$$p_1 = 18/500 = 0{,}036$$
$$p_2 = 25/400 = 0{,}0625$$

Entonces se calcula p:

$$p = \frac{p_1 + p_2}{n_1 + n_2} = \frac{18 + 25}{500 + 400} = 0{,}0478$$

$$q = 1 - 0{,}0478 = 0{,}9522$$

Se plantean las hipótesis:

$$\text{Ho}: p_1 = p_2$$
$$H_1: p_1 \neq p_2$$

Luego se calcula z:

$$z_c = \frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0{,}036 - 0{,}0625}{\sqrt{0{,}0478 \cdot 0{,}9522\left(\frac{1}{500} + \frac{1}{400}\right)}} = -1{,}852$$

De la tabla se obtiene $Z_{\alpha/2} = 1{,}96$. El valor de Z_c cae en la zona de aceptación de H_0, por lo tanto se acepta H_0 con $\alpha = 0{,}05$. La diferencia no es estadísticamente significativa. La proporción de piezas con defectos mayores es igual en las dos líneas de ensamble.

Ejercicio de revisión

Un investigador cree que las mujeres emplean la tarjeta de crédito más que los hombres. Para probar su hipótesis toma una muestra de 90 mujeres y encuentra que 64 de ellas emplea regularmente la tarjeta de crédito. Por otro lado, una muestra de 120 hombres arrojó que 76 empleaban la tarjeta de crédito con regularidad. ¿Tiene razón el investigador? Utilice un nivel de significancia del 1%.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Minitab para la prueba de dos proporciones

Ejemplo

Utilice Minitab para resolver el problema siguiente: En un lote de 500 piezas fabricadas esta semana en una línea de ensamblado se obtuvieron 18 con defectos. En otro lote de 400 piezas tomadas de otra línea de ensamblado se obtuvieron 25 defectuosas. Determine si las líneas producen la misma proporción de piezas con defectos, al nivel de significación de 5%.

Solución

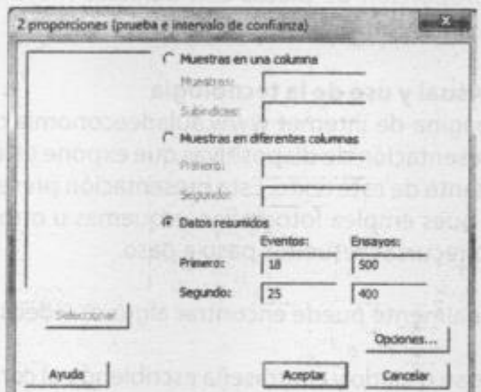
Se cuenta con la siguiente información:

$$p_1 = 18/500 = 0{,}036$$
$$p_2 = 25/400 = 0{,}0625$$

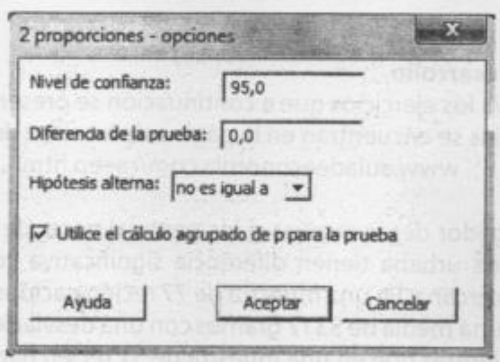
Entonces, se plantean las hipótesis:

$$\text{Ho}: p_1 = p_2$$
$$\text{H}_1: p_1 \neq p_2$$

Ahora se da clic en el menú Estadísticas / Estadística básica / 2 proporciones, y se completa el cuadro de diálogo siguiente, indicando en Datos resumidos los valores correspondientes:



Luego en el botón Opciones se indica el nivel de confianza, el signo de la prueba (mayor, menor o diferente en la hipótesis alternativa) y se debe marcar la casilla Utilice el cálculo agrupado de p para la prueba:



Después se da clic en Aceptar y la salida se obtiene en la ventana sesión:

```
Sesión

Prueba e IC para dos proporciones

Muestra   X    N  Muestra p
1        18  500   0,036000
2        25  400   0,062500


Diferencia = p (1) - p (2)
Estimado de la diferencia:  -0,0265
IC de 95% para la diferencia:  (-0,0552983. 0,00229829)
Prueba para la diferencia = 0 vs. no = 0:  Z = -1,85  Valor P = 0,064

Prueba exacta de Fisher: Valor P = 0,083
```

Se observa el valor de *z* de -1,85, que cae en la zona de aceptación de H_0, o bien, se emplear el valor $P = 0{,}064$, que es mayor que el nivel de significancia de 0,05. Por lo tanto, se acepta concluye que la diferencia en la proporción de piezas con defectos mayores no es estadísticamente significativa.

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

6.6 Ejercicios

Ejercicios de desarrollo

Resuelva los ejercicios que a continuación se presentan (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Un investigador desea evaluar si los pesos al nacer de los recién nacidos en una zona urbana tienen diferencia significativa con respecto a una zona rural cercana. En una muestra de 77 recién nacidos de la zona urbana se obtuvo una media de 3317 gramos con una desviación estándar de 415 gramos, mientras que en una muestra de 51 recién nacidos de la zona rural se obtuvo una media de 3257 gramos con una desviación estándar de 508 gramos. ¿A qué conclusión puede llegarse a un nivel de significancia del 5%?

2. La exposición al sílice se puede dar en distintas actividades productivas y puede generar ciertos problemas en los pulmones. Se ha tomado una muestra de 60 trabajadores que han laborado en minería. Se ha determinado que 40 de ellos se han visto expuestos al sílice. Además, de los 60 trabajadores se encontró que 36 presentaban fibrosis pulmonar. De esos 36, 22 habían tenido exposición al sílice. Con respecto a los trabajadores que presentan fibrosis pulmonar, ¿puede concluirse, al 1% de significancia, que la proporción de trabajadores que han sido expuestos al sílice es la misma que la de los trabajadores que no han sido expuestos a esa sustancia?

3. Un investigador desea conocer cómo evoluciona el peso de las mujeres luego de la gestación. Para ello toma una muestra de 10 mujeres y determina su peso un año antes de la gestación, obteniendo los siguientes resultados (en kilogramos): 60, 56, 80, 67, 62, 65, 52, 59, 51 y 53. Posteriormente determina su peso un año después de la gestación y obtiene los valores siguientes (en kilogramos, en el mismo orden anterior): 62, 59, 89, 62, 60, 72, 61, 66, 58 y 57. ¿Puede concluirse, a un nivel de significancia del 5%, que el peso promedio de las mujeres es mayor un año después de la gestación?

4. Una compañía de productos nutricionales desea determinar qué tan efectivo es un suplemento para reducir de peso. Una muestra de 16 mujeres ha participado en el estudio. Antes de empezar a consumir el suplemento los pesos de las mujeres fueron (en kilogramos): 60, 75, 82, 96, 105, 116, 79, 83, 90, 81, 65, 77, 84, 85, 101 y 99. Luego de tres meses de consumo del producto y de seguimiento por parte de los profesionales de la compañía se volvieron a tomar los pesos, que fueron los siguientes (en kilogramos, en el mismo orden anterior): 56, 74, 80, 86, 107, 120, 74, 80, 80, 78, 63, 79, 84, 82, 99 y 98. ¿Puede concluirse, a un nivel de significancia del 1%, que realmente el producto ayuda a reducir el peso de las mujeres?

5. Un arquitecto considera que actualmente las casas tienen un área construida inferior a la de hace dos o más décadas. En una muestra de 15 viviendas construidas en el transcurso de los dos últimos años se encontró un área construida promedio de 92 metros cuadrados y una desviación estándar de 30 metros cuadrados. En otra muestra de 20 viviendas construidas hace dos o más décadas se encontró un área construida promedio de 132 metros cuadrados y una desviación estándar de 45 metros cuadrados. A un nivel de significancia de 5%, ¿qué puede concluirse que ha habido un cambio significativo en las dimensiones de las viviendas?

6. El jefe de producción de una planta de productos electrónicos desea estimar la producción diaria promedio de un cierto producto. La producción diaria durante 7 días elegidos al azar durante el trimestre pasado es (en unidades): 415, 425, 398, 456, 384, 410, 440. La gerencia considera que con los mismos recursos podrían producirse al menos 450 unidades al día, por lo que implementa algunos cambios para incrementar la productividad en este trimestre. En una nueva muestra de 9 días elegidos al azar se obtuvieron los siguientes niveles de producción diarios: 425, 435, 375, 486, 394, 440, 460, 455, 410. Con estos datos, con un nivel de significancia de 0,025, ¿podría decirse que los cambios fueron efectivos?

7. Un ingeniero ha recopilado datos sobre la vida útil de filtros de gasolina de dos marcas distintas. Los datos son los siguientes (en miles de kilómetros):

Marca 1	12	14	16	15	10	12	13	15	16
Marca 2	13	15	17	17	12	14	13	18	15

Pruebe la hipótesis de que la marca 2 tiene una vida útil que excede en al menos 2 mil kilómetros a la vida útil de la marca 1. Use un nivel de significancia de 5%.

8. Dos profesores de estadística discuten sobre el efecto del uso de un software en la enseñanza de la estadística. Uno de ellos cree que es mejor no utilizar el software, pues así los estudiantes se concentrarán más en los conceptos y los procedimientos. El segundo profesor piensa que empleando un software es mejor, pues así los estudiantes se esfuerzan más en la formulación de los problemas y en la interpretación de los resultados. Para determinar quién tiene la razón, deciden aplicar el mismo examen final, primero a un grupo sin el uso del software y luego a un segundo que sí lo utilice. El primer grupo estaba compuesto por 14 estudiantes, los cuales obtuvieron una calificación promedio de 73 puntos y una desviación estándar de 12 puntos. El segundo grupo estaba compuesto por 15 estudiantes y obtuvieron un promedio de 77 puntos y una desviación estándar de 17 puntos. ¿A qué conclusión puede llegarse a un nivel de significancia del 10%?

9. En una muestra de 10 distintas acciones de empresas tecnológicas elegidas al azar entre todas las acciones que se negocian en la Bolsa de Valores de Nueva York se obtuvo una razón promedio de precio/utilidades (p/u) de \$12,89 con una desviación estándar de \$6,98. En otra muestra de 12 distintas acciones de empresas industriales elegidas al azar entre todas las acciones que se negocian en la Bolsa de Valores de Nueva York se obtuvo una razón promedio de precio/utilidades (p/u) de \$9,92 con una desviación estándar de \$7,15. ¿Es esta evidencia suficiente para concluir, a un nivel de significancia del 5%, que la tasa precio/utilidades promedio para las empresas tecnológicas es superior a la razón obtenida por las empresas industriales?

10. Un contador está realizando una auditoría de los cheques extendidos por una compañía. En una muestra de 15 cheques para agosto de este año se obtuvieron los siguientes montos, en miles dólares:

15, 17, 22, 8, 4, 2, 16, 6, 8, 14, 15, 25, 20, 5, 6

11. En una muestra de 10 cheques para agosto del año pasado se obtuvieron los montos siguientes:

8, 12, 9, 11, 12, 10, 9, 4, 6, 8

Pruebe, al 5% de significancia, si el monto medio de los cheques se ha mantenido al comparar estos dos periodos.

12. Una persona considera que los asiáticos poseen un cociente intelectual promedio superior al de los habitantes de nuestro país. Los siguientes datos corresponden a una muestra de asiáticos que aplicaron un test para medir su cociente intelectual: 102, 100, 97, 101, 100, 113, 126, 86, 85, 106, 99, 130, 79, 110, 117, 102, 113, 102. Los siguientes datos corresponden a una muestra de nacionales que aplicaron un test para medir su cociente intelectual: 95, 99, 102, 96, 103, 96, 97, 98, 112, 101, 94, 120, 97, 99, 97, 98, 110, 101, 98, 102, 107. ¿Indicar estos datos que esa persona tiene razón, al nivel de significancia de 5%?

13. Los siguientes datos muestran el número promedio semanal de horas de uso internet de estudiantes de secundaria para hacer trabajos académicos en una muestra de 8 estudiantes de colegios públicos y de 7 estudiantes de colegios privados:

Colegios públicos	2,0	2,5	4,0	3,5	1,0	4,0	5,0	4,5
Colegios privados	1,5	2,5	5,0	3,5	4,0	2,0	3,5	5,5

¿Puede concluirse, a un nivel de significancia del 5%, que existen diferencias entre los niveles de uso de internet de estudiantes de secundaria para hacer trabajos académicos según el tipo de colegio?

14. Una investigadora ha aplicado un instrumento estandarizado para la medición de los aprendizajes en preescolares de 5 años. La prueba fue desarrollada por la investigadora y utiliza una selección de aprendizajes esperados y cubre distintos ámbitos, como la formación social, la comunicación, la relación con el medio natural, entre otros. Al final se obtiene un puntaje total luego de la aplicación de todos los ítems de la prueba. Posteriormente se aplicó a la misma muestra una segunda prueba para la medición de los aprendizajes en preescolares de 5 años. Los siguientes son los resultados de ambas pruebas:

Niño	1	2	3	4	5	6	7	8	9
Prueba 1	60	53	47	29	38	45	56	49	42
Prueba 2	110	85	88	66	60	67	72	78	76

¿Puede considerarse que existe diferencia significativa entre los dos resultados? Use un nivel de significancia del 1%.

15. Una institución realizó un plan de capacitación entre mujeres emprendedoras de una zona marginal del país. El objetivo es que estas mujeres lograran incrementar los ingresos de sus microempresas. Luego de implementadas las capacitaciones se compararon los ingresos de una muestra de 12 microempresas y se registraron los siguientes ingresos (en dólares al mes):

Empresaria	1	2	3	4	5	6	7	8	9	10	11	12
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

¿Hubo un aumento significativo en los niveles de ingresos de las microempresarias de la zona? Utilice el nivel de significancia del 5%.

16. La cefixima es un antibiótico de cefalosporina usado para tratar infecciones causadas ciertas bacterias. En un estudio se analizó la concentración sérica máxima (Cmax) alcanzada con una solución oral y se comparó con la concentración sérica máxima alcanzada con cápsulas. El estudio incluyó a 16 voluntarios masculinos sanos con edades de entre 21 y 43 años, los cuales no habían consumido ningún fármaco durante los 7 días previos al estudio. Ocho de ellos recibieron durante 4 semanas, con un intervalo de una semana entre cada dosis, 200 mg del fármaco como solución oral. Los otros ocho recibieron en el mismo régimen 200 mg de la droga como una cápsula única. En cada fase y a intervalos regulares luego de administrada la dosis se tomaron muestras de sangre y orina, en las cuales se determinó la concentración de cefixima mediante cromatografía líquida de alta resolución. Para la solución oral se encontró un promedio de 3,22 μg/ml con una desviación estándar de 0,88 μg/ml. Para las cápsulas el promedio fue 2,92 μg/ml con una desviación estándar de 0,86 μg/ml. ¿Puede considerarse que existe diferencia en la concentración sérica máxima entre ambas presentaciones del antibiótico? Use un nivel de significancia del 5%.

17. Una compañía ha desarrollado dos nuevos motores de gasolina. Para valorar su consumo de combustible efectúa 15 recorridos para cada motor. El primer motor consumió en promedio 8,7 galones de gasolina por minuto con una desviación estándar de 1,2 galones. El segundo motor consumió en promedio 9,6 galones de gasolina por minuto con una desviación estándar de 2,1 galones. ¿Qué se podría afirmar, con un 1% de significancia, acerca de la diferencia en el consumo medio de combustible de estos dos motores?

18. En una muestra de 6 distintas acciones de empresas tecnológicas elegidas al azar entre todas las acciones que se negocian en la Bolsa de Valores de Nueva York se obtuvieron las siguientes razones precio/utilidades (p/u) antes de la crisis financiera (en dólares): 12, 15, 17, 18, 16, 10. Luego de la crisis financiera se tomaron los datos de la razón precio/utilidad (p/u) para esas mismas 6 empresas obteniendo los siguientes datos (en dólares y en el mismo orden anterior): 9, 19, 12, 7, 25, 8. ¿Es esta evidencia suficiente para concluir, a un nivel de significancia del 5%, que la tasa precio/utilidades promedio después de la crisis es igual a la que se tenía antes de la crisis?

19. Según un estudio el 25% de las viviendas de zonas urbanas del país están en estado malo o regular y que ese porcentaje se incrementa a 35% en las zonas rurales. Si los tamaños de muestra fueron de 40 y 50 viviendas, respectivamente, ¿pueden considerarse que hay diferencias significativas en el estado de la vivienda entre ambas zonas? Use un nivel de significancia de 1%.

20. Un consultor en el área de tecnologías de información considera que las empresas pequeñas pueden ser más ágiles que las empresas grandes a la hora de adoptar nuevas tecnologías. Para probar su hipótesis toma una muestra de 12 empresas pequeñas y determina que 3 estarían en capacidad de reemplazar sus equipos tecnológicos en un plazo corto. En otra muestra de 17 empresas grandes encuentra que 3 podrían reemplazar sus equipos en ese mismo plazo. ¿Respaldan estos datos la creencia del consultor o podría argumentarse que no son evidencia suficiente? Utilice un nivel de significancia del 2%.

21. El encargado de reclutamiento y selección de personal de una empresa considera que las habilidades sociales, como la buena comunicación y un adecuado manejo de los conflictos, son fundamentales para lograr un buen desempeño en la compañía, y considera que hombres y mujeres no desarrollan por igual estas destrezas. En las entrevistas y pruebas que aplican a los candidatos para los puestos que demanda la empresa se evalúan estas destrezas. En una muestra aleatoria de 35 hombres que aplicaron para algún puesto en la empresa se obtuvo que 10 de ellos tuvieron una evaluación satisfactoria y en una muestra aleatoria de 50 mujeres que aplicaron para algún puesto en la empresa 18 presentaron una evaluación satisfactoria en este sentido. ¿Apoyan estos datos la apreciación del encargado de reclutamiento y selección de personal de esta empresa? Utilice un nivel de significancia de 1%.

22. En un país dos candidatos presidenciales aparecen en las encuestas con preferencias entre los electores muy similares. En una encuesta realizada entre 1000 electores uno de los candidatos obtuvo la preferencia de 480 personas, mientras que en otra encuesta con una muestra de 1200 personas obtuvo la intención de voto de 500 votantes. ¿Existe diferencia significativa entre los resultados de las dos encuestas? Utilice un nivel de significancia de 5%.

23. Hace un año 120 periodistas de una muestra de 250 indicaron que no percibían que hubiera importantes amenazas a la libertad de prensa en el país. Luego de la aprobación de una ley que regula las publicaciones en medios digitales 315 periodistas de una muestra de 400 percibían la existencia de amenazas a la libertad de prensa. ¿Podría considerarse que la opinión de los periodistas ha cambiado? Use un nivel de significancia del 2%.

24. Una empresa pauta publicidad en televisión todos los meses. Recientemente ha lanzado una campaña muy agresiva, sin embargo la gerencia cree que esta campaña está impactando principalmente a los consumidores más jóvenes. Para comprobar esta creencia de la gerencia se han tomado dos muestras, una de 80 consumidores de menos de 40 años y otra de 90 consumidores de 40 años o más. El 58% de los consumidores de menos de 40 años dijeron que recordaban el anuncio de la empresa, mientras que solo el 42% de los consumidores de 40 años o más recordaban el anuncio. ¿Puede considerarse, al 1% de significancia, que existe diferencia entre los dos segmentos de consumidores?

25. Un asesor empresarial considera que los proyectos empresariales que implican la participación de miembros de más de un departamento de la empresa fracasan con más frecuencia que aquellos que se realizan a lo interno de un departamento de la empresa. En una muestra de 40 proyectos del primer tipo fracasaron 15, y en una muestra de 30 proyectos del segundo tipo, fracasaron 9. A un nivel de significancia del 5%, ¿a qué conclusión podría llegarse?

26. Un fabricante de helados desarrolló hace un tiempo una línea de productos con menor contenido de calorías, pues pensaba abarcar el mercado de personas con problemas de obesidad. Sin embargo, muchas personas que no poseen problemas de obesidad también compran este tipo de producto con regularidad. La empresa se cuestiona si su participación de mercado en el mercado de personas con problemas de obesidad es realmente mayor que en el mercado de personas que no poseen este problema. Se efectuó un estudio en el que se entrevistó a 80 personas con obesidad y 35 dijeron que consumían este producto con frecuencia. También se entrevistó a 130 personas sin obesidad y 40 de ellas indicaron que consumía el producto con frecuencia. ¿A qué conclusión puede llegarse al 2% de significancia?

27. Un economista sostiene que el monto del ingreso del deudor no es un factor que influya en una mayor morosidad a la hora de pagar las deudas. Para comprobar su hipótesis toma una muestra de deudores de un banco. Entre 60 deudores de bajo ingreso encontró que 5 de ellos estaban atrasados en el pago de su deuda. En una muestra de 70 deudores de alto ingreso se encontró 8 tenían algún grado de morosidad en sus obligaciones con el banco. ¿Son estos datos evidencia suficiente para darle la razón al economista? Use un nivel de significación del 1%.

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de empleados de la primera empresa, que oyen música, y se midió su productividad. También se tomó una muestra de empleados de la segunda empresa, que no oyen música, y se midió la productividad empleando los mismos métodos que en la primera empresa.

En un problema como este, la hipótesis nula se podría expresar como, si μ_1 es la productividad media en la primera empresa y μ_2 es la productividad media en la segunda empresa:

(a) $H_1: \mu_1 = \mu_2$ (b) $H_0: \mu_1 > \mu_2$
(c) $H_0: \mu_1 - \mu_2 = 0$ (d) $H_0: \mu_1 \geq \mu_2$

2. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de empleados de la primera empresa, que oyen música, y se midió su productividad. También se tomó una muestra de empleados de la segunda empresa, que no oyen música, y se midió la productividad empleando los mismos métodos que en la primera empresa.

En un problema como este, la hipótesis alternativa se podría expresar como, si μ_1 es la productividad media en la primera empresa y μ_2 es la productividad media en la segunda empresa:

(a) $H_0: \mu_1 > \mu_2$ (b) $H_1: \mu_1 - \mu_2 > 0$
(c) $H_1: \mu_1 = \mu_2$ (d) $H_1: \mu_1 \geq \mu_2$

3. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de empleados de la primera empresa, que oyen música, y se midió su productividad. También se tomó una muestra de empleados de la segunda empresa, que no oyen música, y se midió la productividad empleando los mismos métodos que en la primera empresa.

En este problema, si las muestras son grandes y las varianzas poblacionales conocidas, se emplea el siguiente estadístico de prueba:

(a) $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}}$ (b) $t = \dfrac{\bar{x}_d}{s_d / \sqrt{n}}$

(c) $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ (d) $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

4. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de empleados de la primera empresa, que oyen música, y se midió su productividad. También se tomó una muestra de empleados de la segunda empresa, que no oyen música, y se midió la productividad empleando los mismos métodos que en la primera empresa.

En este problema, si las muestras son pequeñas y las varianzas poblacionales desconocidas, se emplea el siguiente estadístico de prueba:

(a) $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}}$ (b) $t = \dfrac{\bar{x}_d}{s_d / \sqrt{n}}$

(c) $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ (d) $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

5. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de 50 empleados de la primera empresa y se obtuvo una productividad media de 40 unidades elaboradas por hora por operario. Se tomó una muestra de 60 empleados de la segunda empresa y se obtuvo una productividad media de 36 unidades elaboradas por hora. Se conoce que las desviaciones estándar poblacionales son de 8 y 12 unidades por hora para la primera y la segunda empresa, respectivamente.

Con base en estos datos se puede calcular el estadístico de prueba:

(a) z = 2,09 (b) z = 1,96
(c) t = 2,09 (d) Ninguna de las anteriores

6. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de 50 empleados de la primera empresa y se obtuvo una productividad media de 40 unidades elaboradas por hora por operario. Se tomó una muestra de 60 empleados de la segunda empresa y se obtuvo una productividad media de 36 unidades elaboradas por hora. Se conoce que las desviaciones estándar poblacionales son de 8 y 12 unidades por hora para la primera y la segunda empresa, respectivamente.

En este problema, el valor crítico o tabular para hacer la prueba es, al 5% de significancia:

(a) $z = 2,09$ (b) $z = 1,645$
(c) $t = 1,96$ (d) Ninguna de las anteriores

7. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de 50 empleados de la primera empresa y se obtuvo una productividad media de 40 unidades elaboradas por hora por operario. Se tomó una muestra de 60 empleados de la segunda empresa y se obtuvo una productividad media de 36 unidades elaboradas por hora. Se conoce que las desviaciones estándar poblacionales son de 8 y 12 unidades por hora para la primera y la segunda empresa, respectivamente.

En este problema, es correcto que, al 5% de significancia:

(a) Se rechaza la hipótesis alternativa porque $|z_c| > |z_t|$
(b) Se acepta la hipótesis nula porque $|z_c| > |z_t|$
(c) Se rechaza la hipótesis nula porque $|z_c| > |z_t|$
(d) Ninguna de las anteriores

8. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de 50 empleados de la primera empresa y se obtuvo una productividad media de 40 unidades elaboradas por hora por operario. Se tomó una muestra de 60 empleados de la segunda empresa y se obtuvo una productividad media de 36 unidades elaboradas por hora. Se conoce que las

desviaciones estándar poblacionales son de 8 y 12 unidades por hora para la primera y la segunda empresa, respectivamente. Al realizar la prueba de hipótesis el gerente de la primera empresa indica que "hay evidencia muestral suficiente para considerar que la música sí tiene efecto positivo sobre la productividad", y el gerente de la segunda empresa expresa que "la evidencia muestral señala que la diferencia entre las productividades medias entre las dos empresas es significativa". Con respecto a estas dos afirmaciones, es correcto que, al 5% de significancia:

(a) Ambas son verdaderas	(b) Ambas son falsas
(c) Solo el primer gerente tiene razón	(d) Solo el segundo gerente tiene razón

9. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de 10 empleados de la primera empresa y se obtuvo una productividad media de 40 unidades elaboradas por hora por operario. Se tomó una muestra de 15 empleados de la segunda empresa y se obtuvo una productividad media de 36 unidades elaboradas por hora. Se conoce que las desviaciones estándar poblacionales son de 8 y 12 unidades por hora para la primera y la segunda empresa, respectivamente.

Con base en estos datos se puede calcular el estadístico de prueba:

(a) $t = 2{,}09$	(b) $z = 2{,}09$
(c) $t = 0{,}92$	(d) Ninguna de las anteriores

10. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de 10 empleados de la primera empresa y se obtuvo una productividad media de 40 unidades elaboradas por hora por operario. Se tomó una muestra de 15 empleados de la segunda empresa y se obtuvo una productividad media de 36 unidades elaboradas por hora. Se conoce que las desviaciones estándar poblacionales son de 8 y 12 unidades por hora para la primera y la segunda empresa, respectivamente.

En este problema, el valor crítico o tabular para hacer la prueba es, al 5% de significancia:

(a) $t = 1{,}714$	(b) $z = 1{,}96$
(c) $t = 1{,}645$	(d) Ninguna de las anteriores

11. Considere la siguiente situación: El gerente de producción de una fábrica considera que los operarios que realizan tareas repetitivas son más productivos cuando oyen música empleando algún aparato con audífonos. El gerente de producción de otra empresa no cree que la música genere ese efecto positivo. Para determinar quién tiene la razón se tomó una muestra de 10 empleados de la primera empresa y se obtuvo una productividad media de 40 unidades elaboradas por hora por operario. Se tomó una muestra de 15 empleados de la segunda empresa y se obtuvo una productividad media de 36 unidades elaboradas por hora. Se conoce que las desviaciones estándar poblacionales son de 8 y 12 unidades por hora para la primera y la segunda empresa, respectivamente.

En este problema, es correcto que, al 5% de significancia:

(a) Se acepta la hipótesis alternativa porque $|t_c| < |t_t|$
(b) Se acepta la hipótesis nula porque $|t_c| < |t_t|$
(c) Se rechaza la hipótesis nula porque $|t_c| < |t_t|$
(d) Ninguna de las anteriores

12. Con respecto a los problemas cuando n_1 o n_2, o ambas, son menores de 30 y se desconocen las varianzas poblacionales, se afirma que:

A. Se usa el estadístico $z = \left((\bar{x}_1 - \bar{x}_2) - \delta\right) \Big/ \sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}$

B. Se usa el estadístico t si se puede suponer que las poblaciones son normales y que $\sigma_1 = \sigma_2 = \sigma$.

Con respecto a estas dos afirmaciones, es correcto que:

(a) Ambas son verdaderas
(b) Ambas son falsas
(c) Solo la afirmación A es verdadera
(d) Solo la afirmación B es verdadera

13. Con respecto a los problemas cuando n_1 o n_2, o ambas, son menores de 30 y se desconocen las varianzas poblacionales, se afirma que:

A. Se calcula la varianza como $\sigma^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$

B. Los grados de libertad son $gl = n_1 + n_2 - 2$.

Con respecto a estas dos afirmaciones, es correcto que:

(a) Ambas son verdaderas
(b) Ambas son falsas
(c) Solo la afirmación A es verdadera
(d) Solo la afirmación B es verdadera

14. Este mes se ha estrenado una nueva película de dibujos animados en los cines del país. Se desea saber si los adultos y los niños valoran de igual manera la película. Por lo tanto, se pidió a una muestra de adultos evaluar la película en una escala de 0 a 10, donde 0 es el mínimo y 10 el máximo. Lo mismo se aplicó a una muestra de niños. Los resultados obtenidos fueron:

Adultos	8	5	6	4	5	6	7	3	
Niños	9	10	7	8	9	6	8	6	8

En este problema, la hipótesis nula se podría expresar como, si μ_1 es la evaluación media de los adultos y μ_2 es la evaluación media de los niños:

(a) $H_1: \mu_1 = \mu_2$ (b) $H_0: \mu_1 > \mu_2$
(c) $H_0: \mu_1 - \mu_2 = 0$ (d) $H_0: \mu_1 \geq \mu_2$

15. Este mes se ha estrenado una nueva película de dibujos animados en los cines del país. Se desea saber si los adultos y los niños valoran de igual manera la película. Por lo tanto, se pidió a una muestra de adultos evaluar la película en una escala de 0 a 10, donde 0 es el mínimo y 10 el máximo. Lo mismo se aplicó a una muestra de niños. Los resultados obtenidos fueron:

Adultos	8	5	6	4	5	6	7	3	
Niños	9	10	7	8	9	6	8	6	8

En este problema, la hipótesis alternativa se podría expresar como, si μ_1 es la evaluación media de los adultos y μ_2 es la evaluación media de los niños:

(a) $H_1: \mu_1 < \mu_2$ (b) $H_1: \mu_1 = \mu_2$
(c) $H_0: \mu_1 - \mu_2 < 0$ (d) $H_1: \mu_1 \leq \mu_2$

16. Este mes se ha estrenado una nueva película de dibujos animados en los cines del país. Se desea saber si los adultos y los niños valoran de igual manera la película. Por lo tanto, se pidió a una muestra de adultos evaluar la película en una escala de 0 a 10, donde 0 es el mínimo y 10 el máximo. Lo mismo se aplicó a una muestra de niños. Los resultados obtenidos fueron:

Adultos	8	5	6	4	5	6	7	3	
Niños	9	10	7	8	9	6	8	6	8

En este problema, se puede calcular el estadístico de prueba:

(a) $|z| = 3{,}32$ (b) $t = 1{,}771$
(c) $t = -3{,}32$ (d) Ninguna de las anteriores

17. Este mes se ha estrenado una nueva película de dibujos animados en los cines del país. Se desea saber si los adultos y los niños valoran de igual manera la película. Por lo tanto, se pidió a una muestra de adultos evaluar la película en una escala de 0 a 10, donde 0 es el mínimo y 10 el máximo. Lo mismo se aplicó a una muestra de niños. Los resultados obtenidos fueron:

Adultos	8	5	6	4	5	6	7	3	
Niños	9	10	7	8	9	6	8	6	8

En este problema, se puede calcular el valor crítico o tabular, al 5% de significancia:

(a) $z = -1{,}645$ (b) $t = -1{,}753$
(c) $t = -1{,}746$ (d) Ninguna de las anteriores

18. Este mes se ha estrenado una nueva película de dibujos animados en los cines del país. Se desea saber si los adultos y los niños valoran de igual manera la película. Por lo tanto, se pidió a una muestra de adultos evaluar la película en una escala de 0 a 10, donde 0 es el mínimo y 10 el máximo. Lo mismo se aplicó a una muestra de niños. Los resultados obtenidos fueron:

Adultos	8	5	6	4	5	6	7	3	
Niños	9	10	7	8	9	6	8	6	8

En este problema, es correcto que, al 5% de significancia:

(a) Se rechaza la hipótesis alternativa porque $|t_c| > |t_t|$
(b) Se acepta la hipótesis nula porque $|t_c| > |t_t|$
(c) Se rechaza la hipótesis nula porque $|t_c| > |t_t|$
(d) Ninguna de las anteriores

19. Un instituto que trabaja en la investigación de riesgos analizó el tiempo que las personas duran en cruzar una calle cuando hablan por teléfono celular o envían mensajes de texto y cuando no lo hacen, pues se considera que distraerse puede incrementar la probabilidad de ser atropellado. Se seleccionó una calle y una muestra de 10 personas hicieron la prueba de cruzar la calle usando su celular y luego volvieron a hacer la prueba sin emplear ese dispositivo. Los resultados obtenidos son los siguientes (tiempo en segundos para cruzar la calle):

Persona	1	2	3	4	5	6	7	8	9	10
Usando celular	6	8	10	9	7	12	8	9	15	9
Sin usar celular	5	6	7	6	5	8	7	7	12	8

En este problema se puede plantear la hipótesis nula, donde μ_d es la media de las diferencias entre los tiempos con y sin uso del celular:

(a) $H_o: \mu_d = 0$ (b) $H_o: \mu_d \geq 0$
(c) $H_1: \mu_d = 0$ (d) Ninguna de las anteriores

20. Un instituto que trabaja en la investigación de riesgos analizó el tiempo que las personas duran en cruzar una calle cuando hablan por teléfono celular o envían mensajes de texto y cuando no lo hacen, pues se considera que distraerse puede incrementar la probabilidad de ser atropellado. Se seleccionó una calle y una muestra de 10 personas hicieron la prueba de cruzar la calle usando su celular y luego volvieron a hacer la prueba sin emplear ese dispositivo. Los resultados obtenidos son los siguientes (tiempo en segundos para cruzar la calle):

Persona	1	2	3	4	5	6	7	8	9	10
Usando celular	6	8	10	9	7	12	8	9	15	9
Sin usar celular	5	6	7	6	5	8	7	7	12	8

En este problema no se puede plantear la hipótesis alternativa del modo siguiente, donde μ_d es la media de las diferencias entre los tiempos con y sin uso del celular:

(a) H1: $\mu_d < 0$ (b) H1: $\mu_d \geq 0$
(c) H1: $\mu_d > 0$ (d) Ninguna de las anteriores

21. Un instituto que trabaja en la investigación de riesgos analizó el tiempo que las personas duran en cruzar una calle cuando hablan por teléfono celular o envían mensajes de texto y cuando no lo hacen, pues se considera que distraerse puede incrementar la probabilidad de ser atropellado. Se seleccionó una calle y una muestra de 10 personas hicieron la prueba de cruzar la calle usando su celular y luego volvieron a hacer la prueba sin emplear ese dispositivo. Los resultados obtenidos son los siguientes (tiempo en segundos para cruzar la calle):

Persona	1	2	3	4	5	6	7	8	9	10
Usando celular	6	8	10	9	7	12	8	9	15	9
Sin usar celular	5	6	7	6	5	8	7	7	12	8

La desviación estándar de las diferencias es:

(a) 1,033 (b) 1,067
(c) 2,2 (d) Ninguna de las anteriores

22. Un instituto que trabaja en la investigación de riesgos analizó el tiempo que las personas duran en cruzar una calle cuando hablan por teléfono celular o envían mensajes de texto y cuando no lo hacen, pues se considera que distraerse puede incrementar la probabilidad de ser atropellado. Se seleccionó una calle y una muestra de 10 personas hicieron la prueba de cruzar la calle usando su celular y luego volvieron a hacer la prueba sin emplear ese dispositivo. Los resultados obtenidos son los siguientes (tiempo en segundos para cruzar la calle):

Persona	1	2	3	4	5	6	7	8	9	10
Usando celular	6	8	10	9	7	12	8	9	15	9
Sin usar celular	5	6	7	6	5	8	7	7	12	8

En este problema se emplea el siguiente estadístico de prueba:

(a) $z = 6{,}74$ (b) $t = 2{,}12$
(c) $t = 6{,}74$ (d) Ninguna de las anteriores

23. Un instituto que trabaja en la investigación de riesgos analizó el tiempo que las personas duran en cruzar una calle cuando hablan por teléfono celular o envían mensajes de texto y cuando no lo hacen, pues se considera que distraerse puede incrementar la probabilidad de ser atropellado. Se seleccionó una calle y una muestra de 10 personas hicieron la prueba de cruzar la calle usando su celular y luego volvieron a hacer la prueba sin emplear ese dispositivo. Los resultados obtenidos son los siguientes (tiempo en segundos para cruzar la calle):

Persona	1	2	3	4	5	6	7	8	9	10
Usando celular	6	8	10	9	7	12	8	9	15	9
Sin usar celular	5	6	7	6	5	8	7	7	12	8

En este problema, es correcto que, al 5% de significancia:

(a) Se rechaza la hipótesis alternativa porque $|t_c| > |t_t|$
(b) Se acepta la hipótesis nula porque $|t_c| > |t_t|$
(c) Se rechaza la hipótesis nula porque $|t_c| > |t_t|$
(d) Ninguna de las anteriores

24. Un instituto que trabaja en la investigación de riesgos analizó el tiempo que las personas duran en cruzar una calle cuando hablan por teléfono celular o envían mensajes de texto y cuando no lo hacen, pues se considera que distraerse puede incrementar la probabilidad de ser atropellado. Se seleccionó una calle y una muestra de 10 personas hicieron la prueba de cruzar la calle usando su celular y luego volvieron a hacer la prueba sin emplear ese dispositivo. Los resultados obtenidos son los siguientes (tiempo en segundos para cruzar la calle):

Persona	1	2	3	4	5	6	7	8	9	10
Usando celular	6	8	10	9	7	12	8	9	15	9
Sin usar celular	5	6	7	6	5	8	7	7	12	8

En este problema, se puede concluir que, al 5% de significancia:

(a) No existe diferencia significativa entre los tiempos promedio de los peatones que cruzan la calle usando el teléfono celular y los que no lo hacen.
(b) Los tiempos promedio de los peatones que cruzan la calle usando el teléfono celular y los que no lo hacen son iguales.
(c) Los tiempos promedio de los peatones que cruzan la calle usando el teléfono celular son menores que los tiempos de los que no lo hacen.
(d) Ninguna de las anteriores

25. Un fabricante de teléfonos celulares líder en el mercado ha anunciado que pronto lanzará un nuevo modelo de su principal producto. La empresa realizó un estudio en el que descubrió que, en una muestra de 70 usuarios actuales de sus productos, que 20 comprarían el nuevo modelo en la misma semana del lanzamiento. En una muestra de 50 consumidores que no son usuarios de sus productos, 10 comprarían el nuevo modelo en la misma semana del lanzamiento. ¿Constituyen estos datos evidencia suficiente para considerar que los usuarios actuales tienen mayor disposición para comprar el nuevo modelo?

En este problema, la hipótesis nula se puede plantear como:

(a) $H_1: P_1 = P_2$ (b) $H_0: P_1 > P_2$
(c) $H_0: P_1 - P_2 = 0$ (d) $H_0: P_1 \geq P_2$

26. Un fabricante de teléfonos celulares líder en el mercado ha anunciado que pronto lanzará un nuevo modelo de su principal producto. La empresa realizó un estudio en el que descubrió que, en una muestra de 70 usuarios actuales de sus productos, que 20 comprarían el nuevo modelo en la misma semana del lanzamiento. En una muestra de 50 consumidores que no son usuarios de sus productos, 10 comprarían el nuevo modelo en la misma semana del lanzamiento. ¿Constituyen estos datos evidencia suficiente para considerar que los usuarios actuales tienen mayor disposición para comprar el nuevo modelo?

En este problema, si P_1 es la proporción de usuarios actuales que comprarían el nuevo modelo en la semana del lanzamiento y P_2 es la proporción de no usuarios actuales que comprarían el nuevo modelo en la semana del lanzamiento, la hipótesis alternativa se puede plantear como:

(a) $H_1: P_1 = P_2$ (b) $H_1: P_1 > P_2$
(c) $H_1: P_1 - P_2 < 0$ (d) $H_1: P_1 \geq P_2$

27. Un fabricante de teléfonos celulares líder en el mercado ha anunciado que pronto lanzará un nuevo modelo de su principal producto. La empresa realizó un estudio en el que descubrió que, en una muestra de 70 usuarios actuales de sus productos, que 20 comprarían el nuevo modelo en la misma semana del lanzamiento. En una muestra de 50 consumidores que no son usuarios de sus productos, 10 comprarían el nuevo modelo en la misma semana del lanzamiento. ¿Constituyen estos datos evidencia suficiente para considerar que los usuarios actuales tienen mayor disposición para comprar el nuevo modelo?

En este problema, si las muestras son grandes, se emplea el siguiente estadístico de prueba:

(a) $t = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}}$ (b) $t = \dfrac{\bar{x}_d}{s_d / \sqrt{n}}$

(c) $z = \dfrac{p_1 - p_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ (d) $z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

28. Un fabricante de teléfonos celulares líder en el mercado ha anunciado que pronto lanzará un nuevo modelo de su principal producto. La empresa realizó un estudio en el que descubrió que, en una muestra de 70 usuarios actuales de sus productos, que 20 comprarían el nuevo modelo en la misma semana del lanzamiento. En una muestra de 50 consumidores que no son usuarios de sus productos, 10 comprarían el nuevo modelo en la misma semana del lanzamiento. ¿Constituyen estos datos evidencia suficiente para considerar que los usuarios actuales tienen mayor disposición para comprar el nuevo modelo?

En este problema se emplea el siguiente estadístico de prueba:

(a) $z = 1{,}07$ (b) $z = 1{,}96$
(c) $z = 1{,}10$ (d) Ninguna de las anteriores

29. Un fabricante de teléfonos celulares líder en el mercado ha anunciado que pronto lanzará un nuevo modelo de su principal producto. La empresa realizó un estudio en el que descubrió que, en una muestra de 70 usuarios actuales de sus productos, que 20 comprarían el nuevo modelo en la misma semana del lanzamiento. En una muestra de 50 consumidores que no son usuarios de sus productos, 10 comprarían el nuevo modelo en la misma semana del lanzamiento. ¿Constituyen estos datos evidencia suficiente para considerar que los usuarios actuales tienen mayor disposición para comprar el nuevo modelo?

En este problema el cálculo agrupado de p da por resultado:

(a) 0,25 (b) 0,2429
(c) 0,4857 (d) Ninguna de las anteriores

30. Un fabricante de teléfonos celulares líder en el mercado ha anunciado que pronto lanzará un nuevo modelo de su principal producto. La empresa realizó un estudio en el que descubrió que, en una muestra de 70 usuarios actuales de sus productos, que 20 comprarían el nuevo modelo en la misma semana del lanzamiento. En una muestra de 50 consumidores que no son usuarios de sus productos, 10 comprarían el nuevo modelo en la misma semana del lanzamiento. ¿Constituyen estos datos evidencia suficiente para considerar que los usuarios actuales tienen mayor disposición para comprar el nuevo modelo?

En este problema, al 1% de significancia, se puede concluir con respecto a la diferencia entre la proporción de clientes actuales que comprarían el nuevo modelo y la proporción de los que no son usuarios actuales que también comprarían el nuevo modelo que:

(a) Existe diferencia significativa entre ambas proporciones.
(b) Ambas proporciones son iguales.
(c) La evidencia muestral no indica que haya diferencia significativa.
(d) Ninguna de las anteriores

CAPÍTULO 7

CORRELACIÓN LINEAL Y REGRESIÓN LINEAL SIMPLE

OBJETIVOS

Al concluir el capítulo, será capaz de:

1. Aplicar el concepto de correlación para analizar la relación dos variables
2. Calcular e interpretar el coeficiente de correlación lineal simple
3. Distinguir los conceptos de correlación y causalidad
4. Calcular e interpretar los coeficientes de la recta de regresión lineal simple
5. Calcular e interpretar el coeficiente de determinación
6. Emplear la ecuación de la recta de regresión para interpolar y extrapolar nuevos valores de las variables del modelo

7.1 Asociación estadística entre dos variables

Con frecuencia aparecen en los distintos medios de comunicación noticias sobre estudios realizados en diferentes partes del mundo. Muchas veces estos estudios analizan si dos o más variables tienen algún tipo de relación entre ellas. Por ejemplo:

- Se investiga si las personas que tienen un elevado consumo de sal en las comidas también tienen una presión arterial más elevada.
- Se trata de encontrar si un fenómeno como la desigualdad social se relaciona con mayores niveles de inseguridad y criminalidad en los países.
- Se examina si el ingreso de la economía está relacionado con mayores niveles de consumo y ahorro.

Ahora bien, a nivel de la empresa y la industria también hay muchas relaciones que los tomadores de decisiones desean estudiar:

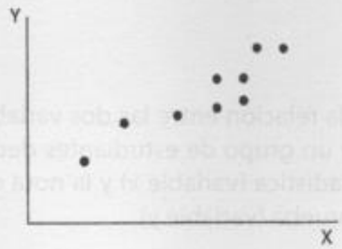
- La relación entre el nivel de un puesto (medida en puntos según el grado de responsabilidad del cargo, los riesgos, etc.) y la remuneración que en el mercado laboral se ofrece.
- El efecto que tienen los cambios en los precios sobre la cantidad demandada de los productos que una empresa vende.
- El impacto que tienen las tasas de interés sobre las ventas de bienes como los vehículos y las viviendas, que generalmente las personas compran a crédito.
- La relación entre la cantidad de horas extra que trabajan los operarios de una fábrica y el número de defectos encontrados en la producción que ellos realizan.

Según el nivel de medición de las variables en cuestión así se mide el grado de la asociación estadística entre ellas. En este capítulo se trabajará únicamente con variables cuantitativas, ya sea que se midan en escala de intervalo o de razón, y no variables cualitativas.

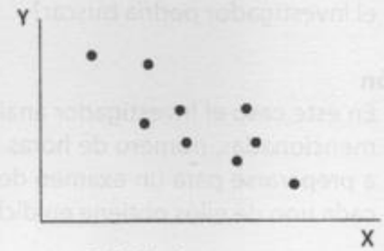
Cuando se tienen dos variables cuantitativas resulta necesario conocer cómo es la relación establecida en cuanto a dos aspectos principales:

1. **La naturaleza o dirección de la asociación entre las variables**, o sea, determinar si cuando una variable se incrementa, la otra también aumenta (relación directa), o si cuando una se incrementa, la otra disminuye (relación inversa). Observe a continuación en la gráfica de la izquierda, como conforme el valor de la variable x es mayor, también se incrementa la variable y, lo cual corresponde a una relación directa, como puede ser el caso entre el ingreso y el consumo, pues a mayor ingreso, también se da un mayor consumo, y viceversa. Del lado izquierdo, se observa que conforme

aumenta la variable *x* se reduce la variable *y*, lo cual corresponde a una relación inversa, como puede ser el caso entre el precio y la cantidad demandada de un producto, pues a mayor precio, se espera una cantidad demandada menor, y viceversa.

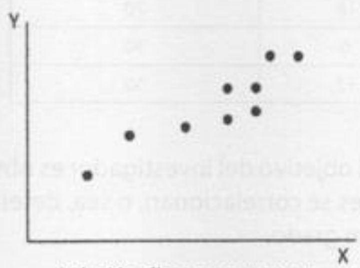


Relación directa entre *x* y *y*.

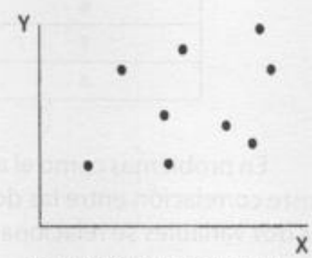


Relación inversa entre *x* y *y*.

2. **La fuerza o intensidad de la asociación entre las variables**, es decir, si generalmente mayores cambios en los valores de una de las variables se asocian con cambios mayores (o menores) en valores de la otra variable (relación fuerte), o si los cambios en una de las variables no se relacionan en forma consistente con cambios en la otra variable (relación débil). A continuación, en la gráfica de la izquierda se observa que los puntos siguen una clara relación lineal, pues si se trazara una recta en medio de ellos, estos quedarían bastante cercanos a dicha recta, por lo que hay una fuerte asociación lineal entre las variables. En el caso de la gráfica de la derecha los puntos están mucho más dispersos, por lo que la relación entre las dos variables es débil.



Relación directa entre *x* y *y*.



Relación inversa entre *x* y *y*.

7.2 Correlación final

Tal como se expuso anteriormente, es posible establecer qué tanto es el grado de asociación estadística entre dos variables. Generalmente se trata de determinar el grado de correlación lineal entre dos variables, es decir, qué tanto se aproxima la relación entre las variables a una línea recta. Para ilustrar este tema emplearemos un caso hipotético.

Ejemplo

Un investigador desea analizar la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba. ¿Cuáles son las variables de este problema y cuáles son los datos que el investigador podría buscar?

Solución

En este caso el investigador analiza la relación entre las dos variables mencionadas, número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística (variable x) y la nota que cada uno de ellos obtiene en dicha prueba (variable y).

El investigador debe tomar una muestra de estudiantes y registrar los valores de ambas variables. Suponga que los resultados de observar ocho estudiantes se resumen en la tabla (las notas están expresadas en una escala de 0 a 100):

Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

En problemas como el anterior el objetivo del investigador es observar si existe correlación entre las dos variables se correlacionan, o sea, determinar estas dos variables se relacionan en algún grado.

CORRELACIÓN Relación recíproca entre dos variables.

Una herramienta útil para observar si existe algún grado de correlación entre dos variables cuantitativas es el diagrama o gráfico de dispersión. En este gráfico cada eje representa una variable y se dibujan puntos que asocian cada valor de x con su correspondiente valor de y.

DIAGRAMA DE DISPERSIÓN Gráfico que utiliza las coordenadas cartesianas para mostrar la relación entre dos variables cuantitativas.

El comportamiento de los puntos en el diagrama de dispersión es un indicador gráfico de la presencia de correlación entre las variables.

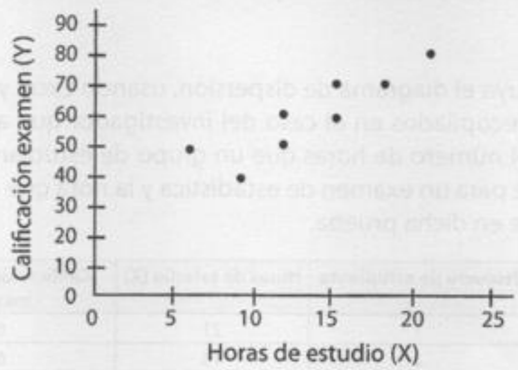
Ejemplo

Construya el diagrama de dispersión para los datos recopilados en el caso del investigador que analiza la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba.

Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Solución

Para construir el diagrama de dispersión se trazan primero los dos ejes cartesianos, y luego cada par de valores (*x*, *y*) se representa como un punto en el gráfico. En este caso, por ejemplo, el punto que se encuentra más arriba a la derecha representa al estudiante número 1, que estudió 21 horas para su examen y obtuvo una calificación de 80 puntos. El punto que se encuentra más a la izquierda representa al estudiante número 7, que estudió solo 6 horas y obtuvo una nota de 50.



Ejercicio de revisión

En un estudio se desea determinar si existe relación entre el ingreso familiar mensual y los gastos mensuales en esparcimiento de las familias. La tabla muestra los datos para una muestra de 12 familias:

Número de familia	Ingreso familiar mensual (X, en $)	Gasto mensual en esparcimiento (Y, en $)
1	500	60
2	1200	100
3	1800	150
4	2500	300
5	750	50
6	800	30
7	900	80
8	1000	75
9	400	25
10	650	60
11	825	95
12	750	60

Construya el diagrama de dispersión.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para la construcción del diagrama de dispersión

Los paquetes de software, tales como Excel o Minitab, construyen los diagramas de dispersión. Se mostrará a continuación cómo construir las gráficas de dispersión empleando Excel y Minitab.

Ejemplo

Construya el diagrama de dispersión, usando Excel y Minitab, para los datos recopilados en el caso del investigador que analiza la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba.

Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Solución

En Excel se introducen los datos de X y Y cada uno en una columna separada:

	A	B
1	X	Y
2	21	80
3	15	60
4	15	70
5	9	40
6	12	60
7	18	70
8	6	50
9	12	50

Luego se seleccionan los datos, se da clic en la pestaña Insertar, se selecciona en la sección Gráficos y se elige la primera opción de Dispersión:

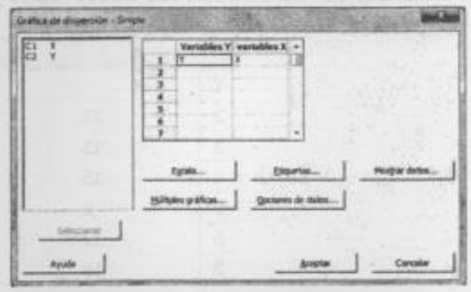
Así, aparecerá en la hoja de Excel el gráfico construido:

En Minitab se introducen los datos de X y Y cada uno en una columna separada de la hoja de trabajo:

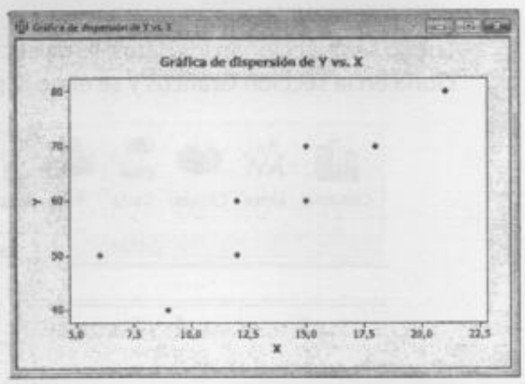
Hoja de trabajo 1 ***

↓	C1	C2
	X	Y
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Luego se da clic en el menú Gráfica, y se elige Dispersión. En el cuadro de diálogo se escoge la opción Simple y se completa el cuadro de diálogo siguiente:

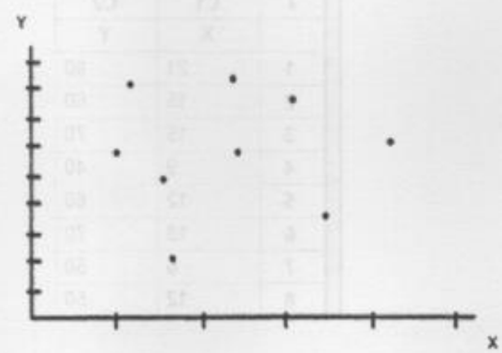


Al dar clic en el botón Aceptar se obtiene el gráfico en una ventana separada en Minitab:

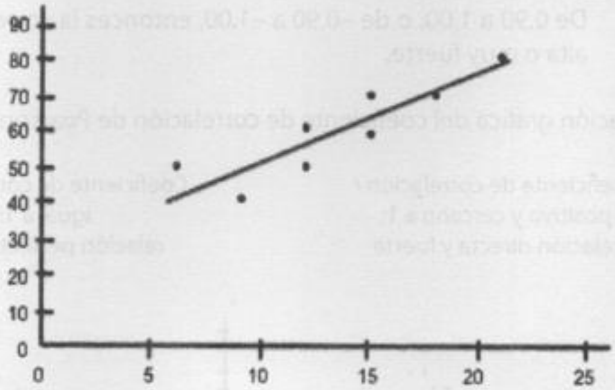


En el gráfico anterior es claro que mientras más horas dedica a prepararse cada uno de los estudiantes, mayor tiende a ser su calificación, por lo que se podría decir que sí hay correlación entre las dos variables.

Si se observa el gráfico siguiente resulta evidente que las dos variables *x* y *y* representadas prácticamente no se correlacionan, pues no se muestra que sistemáticamente haya algún tipo de variación en la variable *x* que se asocie con su correspondiente variación en la variable *y*.



Ahora bien, muchas veces se observa que los puntos en el diagrama de dispersión se acercan a una línea recta, tal como se muestra en la gráfica siguiente. En estos casos puede decirse que existe correlación lineal.



La correlación lineal puede medirse a través del coeficiente de correlación lineal de Pearson, denotado por *r*.

COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON Es un coeficiente que mide la fuerza o intensidad de la relación lineal entre dos variables cuantitativas.

Este coeficiente solo toma valores en el intervalo de −1 a 1. El coeficiente de correlación lineal indica, a través de su signo, la naturaleza o dirección de la relación, pues cuando sea positivo denota una relación directa entre las dos variables, pero cuando sea negativo es porque esta relación es inversa entre las dos variables. Pero también el coeficiente *r* señala la fuerza o intensidad de la relación, la cual será fuerte en la medida en que *r* sea cercano a 1 o −1, pero la relación es débil si es cercano a cero. En el caso de que sea exactamente igual a 1, indicará una relación lineal perfecta, y los puntos del diagrama de dispersión estarán sobre una misma recta con pendiente positiva. En el caso de que sea exactamente igual a −1, también indicará una relación lineal perfecta, pero en esta situación los puntos del diagrama de dispersión estarán sobre una misma recta con pendiente negativa.

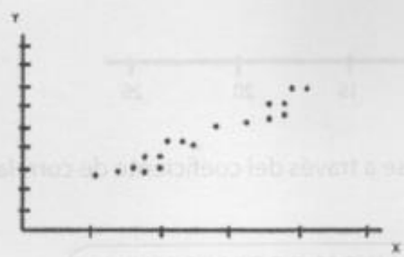
Los siguientes intervalos pueden ser empleados para definir el grado de intensidad de la relación, aunque no existe un criterio único para determinar a partir de cuál valor la relación es fuerte o débil:

- De 0 a 0.20, o de 0 a −0.20, entonces la correlación es muy baja o muy débil.
- De 0.20 a 0.40, o de −0.20 a −0.40, entonces la correlación es baja o débil.

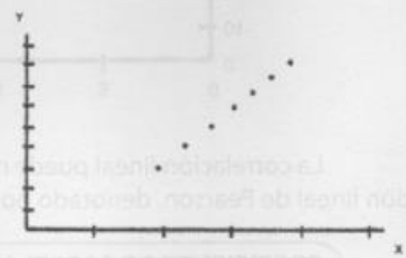
- De 0.40 a 0.70, o de −0.40 a −0.70, entonces la correlación es moderada.
- De 0.70 a 0.90, o de −0.70 a −0.90, entonces la correlación es alta o fuerte.
- De 0.90 a 1.00, o de −0.90 a −1.00, entonces la correlación es muy alta o muy fuerte.

Interpretación gráfica del coeficiente de correlación de Pearson:

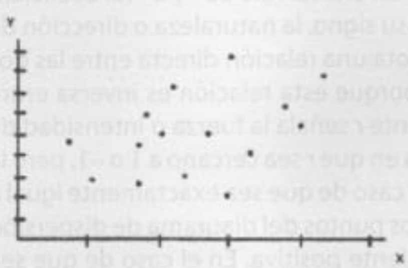
Coeficiente de correlación *r*
positivo y cercano a 1:
relación directa y fuerte



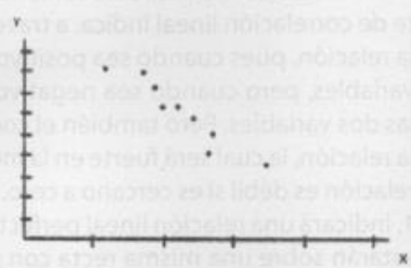
Coeficiente de correlación *r*
igual a 1:
relación perfecta lineal



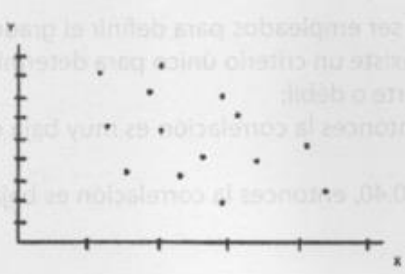
Coeficiente de correlación *r*
positivo y cercano a 0:
relación directa y débil



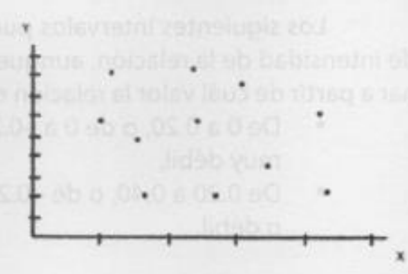
Coeficiente de correlación *r*
negativo y cercano a −1:
relación inversa y fuerte



Coeficiente de correlación *r*
negativo y cercano a 0:
relación inversa y débil

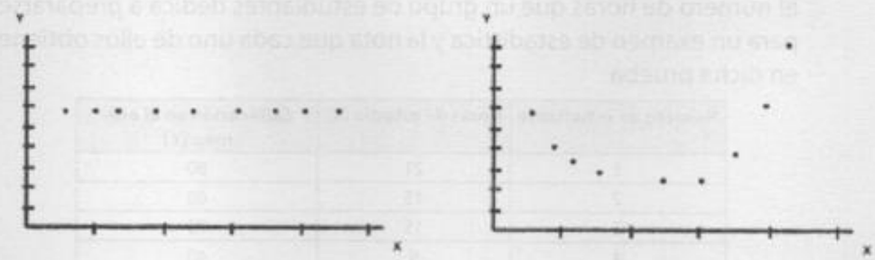


Coeficiente de correlación *r*
muy cercano a 0:
ausencia de relación



Coeficiente de correlación *r* muy igual a 0: ausencia de relación

Coeficiente de correlación *r* muy cercano a 0: ausencia de relación lineal



La siguiente figura resume los conceptos expuestos anteriormente sobre la interpretación del coeficiente de correlación lineal de Pearson *r*:

Cálculo del coeficiente de correlación lineal de Pearson

Existen diferentes maneras de calcular el coeficiente de correlación lineal de Pearson. Tal vez la más sencilla es emplear la fórmula de cálculo siguiente:

FÓRMULA DEL COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}}$$

donde *N* corresponde al número de pares de valores *X* y *Y* observados, S*XY* es la sumatoria de los productos de cada valor de *X* por cada valor de *Y*, S*X* es la sumatoria de todos los valores de *X*, S*Y* es la sumatoria de todos los valores de *Y*, SX^2 es la sumatoria de los cuadrados de cada uno de los valores de *X*, y SY^2 es la sumatoria de los cuadrados de cada uno de los valores de *Y*.

Ejemplo

Calcule el coeficiente de correlación lineal de Pearson para los datos recopilados en el caso del investigador que analiza la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba.

Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Solución

Para el cálculo del coeficiente conviene preparar una tabla como la siguiente, en la cual se han agregado tres columnas más, una para los productos de cada valor de *X* por cada valor de *Y*, otra para calcular los cuadrados de cada valor de *X*, y una más para cada calcular los cuadrados de cada valor de *Y*. Al final se agregó una línea más para calcular las sumatorias de cada una de las columnas.

	X	Y	XY	X^2	Y^2
1	21	80			
2	15	60			
3	15	70			
4	9	40			
5	12	60			
6	18	70			
7	6	50			
8	12	50			
Total Σ					

Ahora se completan las tres columnas nuevas. La columna XY se completa multiplicando cada X por cada Y. Por ejemplo, se multiplica 21 por 80, y así obtiene 1680, después multiplica 15 por 60, que es 900, y así sucesivamente. La columna X^2 se completa elevando al cuadrado cada valor de X. Por ejemplo, se eleva al cuadrado 21, y así obtiene 441, después eleva al cuadrado 15, que es 225, y así sucesivamente se completa la columna. La columna Y se completa elevando al cuadrado cada Y. Por ejemplo, se eleva 80 al cuadrado, y así obtiene 6400, después eleva al cuadrado 60, que es 3600, y así sucesivamente.

	X	Y	XY	X^2	Y^2
1	21	80	1680	441	6400
2	15	60	900	225	3600
3	15	70	1050	225	4900
4	9	40	360	81	1600
5	12	60	720	144	3600
6	18	70	1260	324	4900
7	6	50	300	36	2500
8	12	50	600	144	2500
Total Σ					

Luego se calculan las sumatorias o totales de cada una de las columnas:

	X	Y	XY	X^2	Y^2
1	21	80	1680	441	6400
2	15	60	900	225	3600
3	15	70	1050	225	4900
4	9	40	360	81	1600
5	12	60	720	144	3600
6	18	70	1260	324	4900
7	6	50	300	36	2500
8	12	50	600	144	2500
Total Σ	108 ↑ ΣX	480 ↑ ΣY	6870 ↑ ΣXY	1620 ↑ ΣX^2	30000 ↑ ΣY^2

Finalmente se sustituyen los valores en la fórmula del coeficiente de correlación lineal:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}}$$

Los valores a sustituir son:
$N = 8$, $SX = 108$, $SY = 480$, $SXY = 6870$, $SX^2 = 1620$, $SY^2 = 30000$:

$$r = \frac{8 \cdot 6870 - 108 \cdot 480}{\sqrt{(8 \cdot 1620 - (108)^2)(8 \cdot 30000 - (480)^2)}}$$

$$r = 0{,}885$$

Ejercicio de revisión

En un estudio se desea determinar si existe relación entre el ingreso familiar mensual y los gastos mensuales en esparcimiento de las familias. La tabla muestra los datos para una muestra de 12 familias:

Número de familia	Ingreso familiar mensual (X, en $)	Gasto mensual en esparcimiento (Y, en $)
1	500	60
2	1200	100
3	1800	150
4	2500	300
5	750	50
6	800	30
7	900	80
8	1000	75
9	400	25
10	650	60
11	825	95
12	750	60

Calcule el coeficiente de correlación lineal de Pearson.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para el cálculo del coeficiente de correlación lineal

Los paquetes de software, tales como Excel o Minitab, e incluso la mayoría de las calculadoras científicas, calculan el coeficiente de correlación lineal. Se mostrará a continuación cómo calcular el coeficiente de correlación empleando Excel y Minitab.

Ejemplo

Calcule el coeficiente de correlación lineal de Pearson, usando Excel y Minitab, para los datos recopilados en el caso del investigador que analiza la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba.

Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Solución

En Excel se introducen los datos de *X* y *Y* cada uno en una columna separada:

	A	B
1	X	Y
2	21	80
3	15	60
4	15	70
5	9	40
6	12	60
7	18	70
8	6	50
9	12	50

Luego en una celda separada se introduce la función:

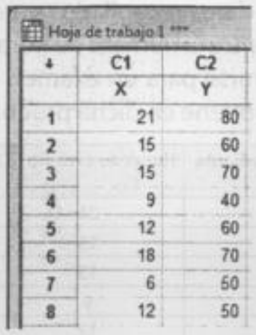
=COEF.DE.CORREL(matriz1;matriz2)

Como los valores de *X* se encuentran en el rango A2:A9, y los valores de *Y* se encuentran en el rango B2:B9, entonces la función se completa del modo siguiente:

=COEF.DE.CORREL(A2:A9;B2:B9)

Al presionar la tecla Enter (o Intro) se obtiene el valor del coeficiente de correlación $r = 0{,}885$.

En Minitab se introducen los datos de *X* y *Y* cada uno en una columna separada de la hoja de trabajo:

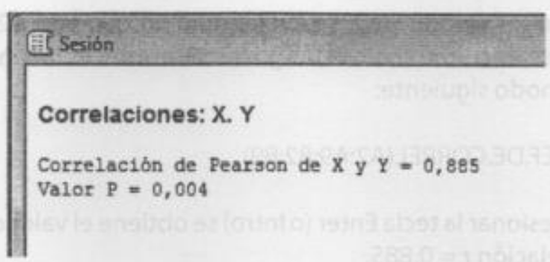


Hoja de trabajo 1 ***

↓	C1 X	C2 Y
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Luego se da clic en el menú Estadísticas, se selecciona Estadística básica y ahí se busca la opción Correlación. Se debe completar el siguiente cuadro de diálogo seleccionando las variables de la lista de la izquierda (debe dar doble clic sobre cada una):

Al dar clic en el botón Aceptar se obtiene el valor del coeficiente de correlación $r = 0{,}885$ en la ventana Sesión de Minitab:

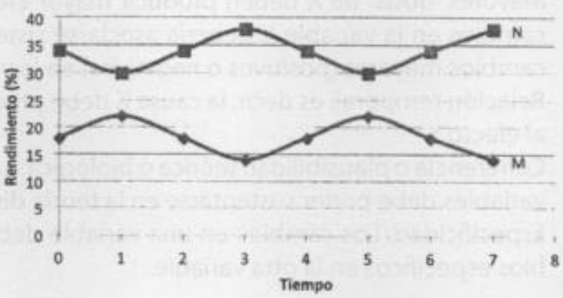


Aplicación

Diversificación financiera

Cuando se decide invertir en activo financieros es muy importante diversificar la cartera o portafolio de inversiones. La idea de diversificar es "no poner todos los huevos en la misma canasta", o sea, distribuir el riesgo entre activos diferentes. Así, si se tiene pérdida en alguno de los activos se esperaría "compensar" con ganancias en los otros. Ahora bien, para que la diversificación sea efectiva la correlación de los rendimientos de los activos en que se invierte debe ser negativa, y no positiva.

Tal como se muestra en la gráfica, la correlación entre los rendimientos de los activos M y N es negativa, pues cuando el rendimiento de uno de ellos aumenta, el del otro disminuye, por lo que una cartera con estos dos activos correspondería a una adecuada diversificación.



Por supuesto que si se tiene una gran cantidad de activos es más difícil establecer las correlaciones y conformar una cartera con una adecuada diversificación, pero existen sistemas informáticos que ayudan a realizar esta tarea.

7.3 Correlaciones espurias y causalidad

En muchas ocasiones es posible encontrar un elevado coeficiente de correlación entre dos variables que no tienen relación alguna, es decir, variables que no presentan relación justificada a través de alguna teoría específica presentan altos coeficientes de correlación. Un ejemplo de esto se presentó en estudios realizados por Neyman en 1952, quien analizó la relación entre la tasa de nacimientos y la población de cigüeñas en varias regiones de Europa, y encontró un alto coeficiente de correlación entre estas variables, a pesar de que todos sabemos que las cigüeñas no traen a los niños. Cuando sucede esto, se dice que la correlación estadística existente entre estas variables es una correlación espuria o sin sentido. Esto se puede dar cuando ocurre la presencia de un tercer factor y no debido a la existencia de una relación con sentido entre las variables analizadas.

Es importante indicar que algunas investigaciones son de nivel correlacional, y por tanto solo buscan establecer la existencia de correlación entre las variables de estudio. Pero otras investigaciones son de nivel explicativo, y por tanto buscan determinar no solo correlación, sino la existencia de relaciones causa – efecto. Cuando se trata de establecer causalidad entre las variables resulta importante emplear algunos criterios que permitan diferencias aquellas correlaciones debidas al azar o debidas a otras variables intervinientes. Algunos criterios de causalidad son:

1. La asociación entre X y Y debe ser fuerte. Esto es que haya un coeficiente de correlación alto entre las variables X y Y.
2. La asociación entre X y Y debe ser consistente y replicable. La relación entre las variables debe poder constatarse en diferentes épocas y lugares.
3. Mayores "dosis" de X deben producir mayor efecto en Y. Mayores cambios en la variable X debería asociarse sistemáticamente con cambios mayores (positivos o negativos) en la variable Y.
4. Relación temporal, es decir, la causa X debe preceder en el tiempo al efecto Y.
5. Coherencia o plausibilidad teórica o biológica. La relación entre las variables debe poder sustentarse en la teoría disponible.
6. Especificidad. Los cambios en una variable deben provocar cambios específicos en la otra variable.

Diversos autores han planteado otros criterios para establecer relaciones causales, sin embargo los 6 anteriores son algunos de los más utilizados por los investigadores.

7.4 Regresión lineal simple

Cuando se logra encontrar variables con un nivel de correlación aceptable y que sea razonable plantear un modelo que permita establecer la relación lineal entre ellas, entonces se puede emplear la técnica de la regresión lineal simple para determinar la ecuación de una recta que permita pronosticar el comportamiento de *y* en términos de *x*.

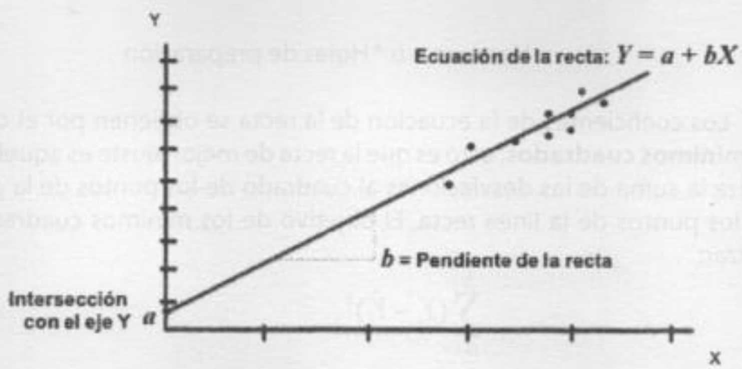
REGRESIÓN LINEAL SIMPLE Es un método matemático que modela la relación lineal entre una variable dependiente *Y* y una variable independiente *X*.

Cabe señalar que en este capítulo solo se expone la regresión lineal simple, es decir, la variable *Y* en términos de solo una variable independiente *X*, pero es posible expresar a *Y* en términos de varias variables independientes. En ese caso se habla de regresión múltiple. Además, la relación entre *Y* y *X* podría ser cuadrática, cúbica u otras formas matemáticas, en cuyo caso se hablaría de regresión no lineal.

Para el caso de la regresión lineal simple, la ecuación de regresión es la ecuación de una línea recta tiene una forma algebraica dada por la expresión:

$$y = a + bx$$

donde a es la constante o intersección con el eje vertical, y b es el coeficiente de pendiente, que da la inclinación de la recta, o bien, el cambio ocurrido en la variable y cuando x varía en una unidad. Gráficamente la recta de regresión se muestra del modo siguiente:



Ejemplo

Para el caso del investigador que desea analizar la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba. ¿Cuál sería el modelo que se podría plantear?

Solución

En este caso el investigador analiza la relación entre las dos variables mencionadas, número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística (variable x) y la nota que cada uno de ellos obtiene en dicha prueba (variable y).

El investigador tomó una muestra de estudiantes y registró los valores de ambas variables y obtuvo un coeficiente de correlación lineal de

Pearson de $r = 0,885$, que indica una correlación lineal directa y fuerte entre las dos variables. Conceptualmente es razonable considerar que la variable número de horas que un estudiante dedica a prepararse para un examen de estadística (variable x) pueda ser determinante de la nota que obtiene en dicha prueba (variable y), por lo que podría formularse un modelo lineal del tipo:

$$y = a + bx$$

donde y es la nota obtenida en el examen, y x es el número de horas dedicadas a la preparación para el examen.

La constante a indicaría la nota que se obtendría si no se estudiara para el examen (cero horas de preparación) y la pendiente b indicaría lo que se esperaría que aumente la nota en el examen por cada hora adicional dedicada a la preparación para esta prueba. También podría expresarse:

$$\text{Nota} = a + b * \text{Horas de preparación}$$

Los coeficientes de la ecuación de la recta se obtienen por el criterio de los **mínimos cuadrados**, esto es que la recta de mejor ajuste es aquella que minimiza la suma de las desviaciones al cuadrado de los puntos de la gráfica desde los puntos de la línea recta. El objetivo de los mínimos cuadrados es minimizar:

$$\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$

y así obtener la recta que mejor se ajusta a los pares (x, y) dados. Los coeficientes de la ecuación de la recta a y b se obtienen por las fórmulas siguientes:

FÓRMULA DEL COEFICIENTE DE PENDIENTE

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

FÓRMULA DEL COEFICIENTE CONSTANTE

$$a = \frac{\sum Y}{N} - b\frac{\sum X}{N}$$

La fórmula para el coeficiente a también podría expresarse como:

FÓRMULA DEL COEFICIENTE CONSTANTE $a = \overline{Y} - b\overline{X}$

Ejemplo

Construya el modelo de regresión lineal para los datos recopilados en el caso del investigador que analiza la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba

Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Solución

Las fórmulas de los coeficientes de la recta de regresión emplean los mismos datos utilizados en el cálculo del coeficiente de correlación lineal, por lo que se empleará la misma tabla que se construyó cuando se calculó *r*. Entonces los datos disponibles son los siguientes:

	X	Y	XY	X^2	Y^2
1	21	80	1680	441	6400
2	15	60	900	225	3600
3	15	70	1050	225	4900
4	9	40	360	81	1600
5	12	60	720	144	3600
6	18	70	1260	324	4900
7	6	50	300	36	2500
8	12	50	600	144	2500
Total Σ	108 ↑ ΣX	480 ↑ ΣY	6870 ↑ ΣXY	1620 ↑ ΣX^2	30000 ↑ ΣY^2

Primero se sustituyen los valores en la fórmula del coeficiente de pendiente:

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

$$b = \frac{8 \cdot 6870 - 108 \cdot 480}{8 \cdot 1620 - (108)^2}$$

$$b = 2{,}41$$

Después se sustituye en la fórmula de la constante *a*:

$$a = \frac{\sum Y}{N} - b\frac{\sum X}{N}$$

$$a = \frac{480}{8} - 2{,}41\frac{108}{8}$$

$$a = 27{,}5$$

Así, el modelo de regresión es:

$$y = 27{,}5 + 2{,}41x$$

O bien, Nota= 27,5 + 2,41 * Número de horas de preparación.

Ejercicio de revisión

En un estudio se desea determinar si existe relación entre el ingreso familiar mensual y los gastos mensuales en esparcimiento de las familias. La tabla muestra los datos para una muestra de 12 familias:

Número de familia	Ingreso familiar mensual (X, en $)	Gasto mensual en esparcimiento (Y, en $)
1	500	60
2	1200	100
3	1800	150
4	2500	300
5	750	50
6	800	30
7	900	80
8	1000	75
9	400	25
10	650	60
11	825	95
12	750	60

Calcule la ecuación de regresión.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Uso de Excel y Minitab para el cálculo de la ecuación de regresión

Ejemplo

Construya, usando Excel y Minitab, el modelo de regresión lineal para los datos recopilados en el caso del investigador que analiza la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba.

Número de estudiante	Horas de estudio (X)	Calificación en el examen (Y)
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

En Excel se introducen los datos de *X* y *Y* cada uno en una columna separada:

	A	B
1	X	Y
2	21	
3	15	
4	15	
5	9	
6	12	
7	18	
8	6	

Luego en una celda separada se introduce la función:

=INTERSECCION.EJE(conocido_y;conocido_x)

Como los valores de *X* se encuentran en el rango A2:A9, y los valores de *Y* se encuentran en el rango B2:B9, entonces la función se completa del modo siguiente:

=INTERSECCION.EJE(B2:B9;A2:A9)

Al presionar la tecla Enter (o Intro) se obtiene el valor del coeficiente de intersección $a = 27{,}5$.

Después, en otra celda se introduce la función:

=PENDIENTE(conocido_y;conocido_x)

Dado que los valores de *X* se encuentran en el rango A2:A9, y los valores de *Y* se encuentran en el rango B2:B9, entonces la función se completa del modo siguiente:

=PENDIENTE(B2:B9;A2:A9)

Al presionar la tecla Enter (o Intro) se obtiene el valor del coeficiente de pendiente $b = 2{,}407 \approx 2{,}41$.

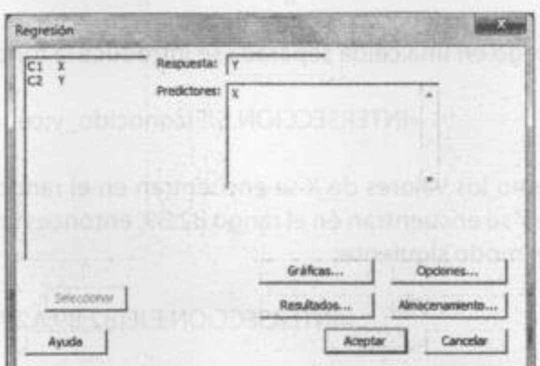
Solución

En Minitab se introducen los datos de *X* y *Y* cada uno en una columna separada de la hoja de trabajo:

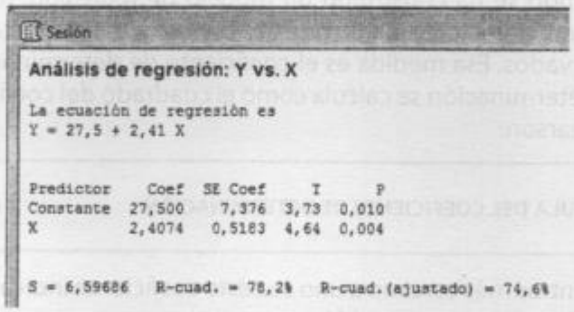
Hoja de trabajo 1 ***

↓	C1	C2
	X	Y
1	21	80
2	15	60
3	15	70
4	9	40
5	12	60
6	18	70
7	6	50
8	12	50

Luego se da clic en el menú Estadísticas, se selecciona Regresión y ahí se busca la opción Regresión. Se debe completar el siguiente cuadro de diálogo seleccionando la variable *y* como Respuesta y la variable *x* como Predictor:



Luego, en la ventana Sesión de Minitab se obtiene:



```
Sesión

Análisis de regresión: Y vs. X

La ecuación de regresión es
Y = 27,5 + 2,41 X


Predictor    Coef   SE Coef     T      P
Constante  27,500    7,376   3,73  0,010
X          2,4074   0,5183   4,64  0,004


S = 6,59686   R-cuad. = 78,2%   R-cuad.(ajustado) = 74,6%
```

En la salida de Minitab aparece claramente la ecuación y otros datos sobre el análisis de regresión, principalmente en lo relacionado con la significancia estadística del modelo. A continuación, en este capítulo se expone el valor que Minitab llama R–cuad.

Aplicación

Econometría

Muchas veces vemos en los medios de comunicación noticias que hablan de diversas proyecciones económicas. ¿Se ha preguntado usted cómo se realizan esas proyecciones? Los economistas se basan en la teoría económica, pero para poder aplicarla se requieren datos sobre las variables que esa teoría abarca. Por ejemplo, se desea conocer cuáles factores influyen sobre el consumo de bienes y servicios del país. La teoría indica que el ingreso disponible puede ser un factor determinante y que la relación entre ambas variables se puede describir a través de una función lineal. Entonces se reúne información sobre ambas variables y se emplea la regresión lineal para determinar si efectivamente se presenta el comportamiento esperado.

Vemos, a través de este sencillo ejemplo, que la econometría busca expresar las distintas teorías económicas a través de modelos matemáticos, y que se utilizan métodos estadísticos para verificar esas teorías. En caso de que las teorías se validen, entonces pueden ser utilizadas para desarrollar pronósticos de las variables y determinar el posible efecto de una variable sobre otra, y así poder realizar recomendaciones de política económica.

La econometría emplea métodos como la regresión lineal, pero no solo la regresión lineal simple, como la estudiada en este capítulo, sino también la regresión lineal múltiple, la regresión no lineal, regresiones de sistemas de ecuaciones, entre otros temas relacionados.

Bondad de ajuste: el coeficiente de determinación

Cuando se ha construido un modelo de regresión, resultado importante contar con una medida de qué tan bien se ajusta la recta estimada a los datos observados. Esa medida es el coeficiente de determinación R^2. El coeficiente de determinación se calcula como el cuadrado del coeficiente de correlación de Pearson:

FÓRMULA DEL COEFICIENTE DE DETERMINACIÓN	$R^2 = r^2$

Mientras más cercano a uno sea este coeficiente indica un mejor ajuste de la recta. Es por eso, que también se puede interpretar este coeficiente como la proporción de la variabilidad explicada por el modelo.

Ejemplo

Para el caso del investigador que desea analizar la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba, y para el cual se conoce que el coeficiente de correlación lineal de Pearson es $r = 0{,}885$. Calcule e interprete el coeficiente de determinación R^2.

Solución

Dado que ya se conoce que el coeficiente de correlación lineal de Pearson es $r = 0{,}885$, basta con elevar ese número al cuadrado para obtener el coeficiente de determinación R^2:

$$R^2 = r^2 = (0{,}885)^2 = 0{,}7832$$

Este resultado quiere decir que el modelo de regresión planteado explica el 78,32% de la variabilidad de y, o sea, que la relación lineal entre la nota en el examen de estadística y el número de horas de preparación explica el 78,32% de la variabilidad de las notas.

Este valor de R^2 indicaría que es un modelo bastante bueno, pues posee un poder explicativo alto.

Ejercicio de revisión

En un estudio se desea determinar si existe relación entre el ingreso familiar mensual (X, en dólares) y los gastos mensuales en esparcimiento de las familias (Y, en dólares). La ecuación de regresión que se ha obtenido es $Y = -29{,}0 + 0{,}119X$ y se obtuvo el coeficiente de correlación lineal de Pearson $r = 0{,}951$. Calcule el coeficiente de determinación.

Ver solución de este ejercicio en la página de internet de este texto:
www.auladeeconomia.com/raeep.html

7.5 Interpolación y extrapolación

La ecuación de regresión se puede emplear para obtener valores de *Y* correspondientes a nuevos valores de *X* que no fueron observados en la muestra. Cuando se obtienen esos nuevos valores de *Y* se dice que se está interpolando o extrapolando. La interpolación es cuando, para calcular el nuevo valor de *Y*, se toma un valor de *X* que se encuentra en el rango de valores de *X* observados. La extrapolación es cuando, para calcular el nuevo valor de *Y*, se toma un valor de *X* que no se encuentra en el rango de valores de *X* observados.

Ejemplo

Para el caso del investigador que desea analizar la relación entre el número de horas que un grupo de estudiantes dedica a prepararse para un examen de estadística y la nota que cada uno de ellos obtiene en dicha prueba, en el cual se había obtenido el modelo $y = 27,5 + 2,41x$, donde y es la nota en el examen, y x es el número de horas de preparación para la prueba:

a. Interpole la calificación de un estudiante que haya estudiado 13 horas.

b. Extrapole la calificación de un estudiante que haya estudiado 25 horas.

Solución

a. En este primer ejercicio se habla de interpolación ya que el rango de valores observados de *X*, los cuales, si se observa en la tabla de datos de las dos variables, el menor valor de x fue 6 y el mayor 21, por lo que 13 se encuentra dentro del rango observado. Entonces, para hallar y se sustituye el valor $x = 13$ en la ecuación:

$$y = 27,5 + 2,41x$$

$$y = 27,5 + 2,41 * 13$$

$$y = 58,83$$

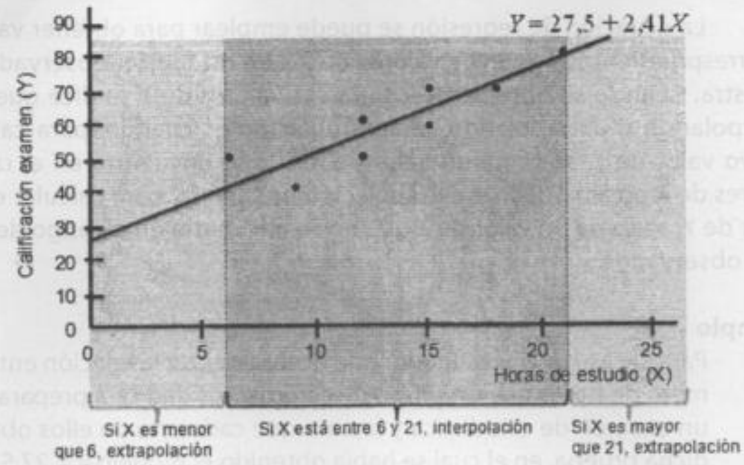
b. En este segundo ejercicio se habla de extrapolación ya que $x = 13$ se encuentra dentro del rango observado. Entonces, para hallar y se sustituye el valor $x = 25$ en la ecuación:

$$y = 27,5 + 2,41x$$

$$y = 27,5 + 2,41 * 25$$

$$y = 87,75$$

Lo anterior se ilustra en la gráfica siguiente:



Ejercicio de revisión

En un estudio se desea determinar si existe relación entre el ingreso familiar mensual (X, en dólares) y los gastos mensuales en esparcimiento de las familias (Y, en dólares). La ecuación de regresión que ha obtenido es $Y = -29{,}0 + 0{,}119X$. Las familias estudiadas tenían ingresos que varían entre \$400 y \$2500.

a. Interpole el gasto mensual en esparcimiento para una familia con un ingreso mensual de \$800.
b. Extrapole el gasto mensual en esparcimiento para una familia con un ingreso mensual de \$3000.

Ver solución de este ejercicio en la página de internet de este texto: www.auladeeconomia.com/raeep.html

Apoyo audiovisual y uso de la tecnología

En la página de internet www.auladeeconomia.com podrá encontrar una presentación de diapositivas que expone este tema y es una parte importante de este texto. Esta presentación presenta el tema en forma visual, pues emplea fotografías, esquemas u otros recursos visuales, e incluso recursos resueltos paso a paso.

Adicionalmente puede encontrar algunos videos explicativos.

Solicite su usuario y contraseña escribiendo al correo electrónico info@auladeeconomia.com. Para ello deberá indicar el número de su factura de compra de texto.

7.6 Ejercicios

Ejercicios de desarrollo:

Resuelva los ejercicios que a continuación se presentan (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. A continuación se presenta una lista de investigaciones que se han correlacionado dos variables, en cada caso indique si considera que se presenta una correlación positiva, negativa o nula entre las dos variables indicadas:
   a. Satisfacción de los pacientes de un hospital en el servicio de consulta externa y su tiempo medio de espera.
   b. Tasa mensual de prevalencia de enfermedades cardiovasculares y el número de actividades de prevención realizadas cada mes en la ciudad capital de julio a noviembre del año 2012.
   c. Frecuencia cardiaca de las ratas en un laboratorio según distintas dosis de un nuevo medicamento.
   d. Número de horas promedio diarias de uso continuo uso de computadoras y agudeza visual de los niños de 12 años.
   e. Temperatura atmosférica media anual en las zonas montañosas del país y población estimada de ranas cuyo hábitat son las zonas montañosas del país.
   f. Peso en kilogramos de adultos y número promedio mensual de horas de actividad física.
   g. Área construida de viviendas de la ciudad capital y precio de la vivienda.
   h. Número de artículos defectuosos por día y número de horas extra pagadas en la fábrica.
   i. Tiempo de fusión de determinados componentes electrónicos y nivel de sobrecarga.
   j. Presupuesto destinado a seguridad informática de las empresas financieras y número de empleados de la empresa.
   k. Número de licencias de un nuevo sistema operativo que las empresas estaría dispuestas a adquirir y el número de empleados de la empresa.
   l. Rendimiento sobre la inversión promedio de las empresas que cotizan en la Bolsa de Valores de Nueva York y nivel de endeudamiento de la empresa.
   m. Número de mensajes mensuales enviados por los usuarios de teléfonos y edad del usuario del teléfono.
   n. Nivel de satisfacción de los empleados de una empresa y números de años de servicio en la empresa.
   o. Monto de cada cuenta por cobrar y antigüedad, en días, de la cuenta por cobrar.
   p. Número de turistas que visitan el país por mes y monto gastado por el gobierno para promocionar al país como destino turístico.
   q. Número de horas de uso internet de estudiantes de secundaria y edad del estudiante.
   r. Volumen mensual de ventas de una compañía y la exposición mensual de los consumidores a los comerciales pautados por la misma empresa.

2. Un equipo de investigadores está realizando un estudio entre la población adulta mayor del país. En el Hogar de Ancianos La Florida, que solo posee 12 ancianos actualmente, se ha recopilado la siguiente información:

# caso	Edad (años)	Presión arterial sistólica (mmHg)
1	66	147
2	72	165
3	72	160
4	86	168
5	73	199
6	97	178
7	85	170
8	89	175
9	78	145
10	82	192
11	68	152
12	90	195

Tomando como base la información anterior se desea relacionar la presión arterial con la edad:

a. Construya el diagrama de dispersión que muestre la relación entre las dos variables.
b. Calcule el coeficiente de correlación lineal de Pearson.
c. Interprete el resultado del coeficiente de correlación lineal de Pearson.
d. Determine la ecuación de regresión que permita predecir el nivel de presión arterial según la edad del anciano.
e. Si uno de estos ancianos tiene una edad de 85 años, ¿cuál sería su nivel esperado de presión sanguínea?
f. Calcule el coeficiente de determinación.
g. Interprete el coeficiente de determinación.

3. En un hospital se evalúa la satisfacción de los pacientes en el servicio de consulta externa. Los administradores del hospital consideran que el tiempo de espera afecta negativamente la satisfacción del paciente. Los siguientes datos corresponden a una muestra de 10 pacientes. La satisfacción se mide a través de un cuestionario en una escala de 0 a 20 puntos y el tiempo de espera está medido en minutos.

Paciente	1	2	3	4	5	6	7	8	9	10
Satisfacción	11	12	5	8	15	16	18	12	15	14
Tiempo espera	30	45	60	55	20	25	30	40	20	25

a. Construya el diagrama de dispersión y analícelo.
b. Calcule el coeficiente de correlación lineal e indique su significado en esta situación.
c. Calcule la ecuación de regresión e interprete sus coeficientes.
d. ¿Cuál sería el nivel de satisfacción esperado para un paciente que esperó 30 minutos por su servicio?
e. Calcule el coeficiente de determinación e interprételo.

4. En una muestra de 12 personas con miopía elevada superior a 12,00 D se midieron algunos parámetros biométricos de sus ojos, como la longitud axial y el equivalente esférico. Los resultados fueron los siguientes:

Paciente	Longitud axial (mm)	Equivalente esférico (D)
1	24	-18
2	28	-13
3	34	-22
4	30	-14
5	28	-11
6	27	-14
7	25	-16
8	29	-17
9	31	-21
10	30	-19
11	31	-18
12	33	-22

5. Con base en estos datos se desea determinar qué tanta relación existe entre la longitud axial y el grado de miopía de estos pacientes. Para esto efectúe lo siguiente:
   a. Construya el diagrama de dispersión y observe qué tipo de correlación se presenta entre las dos variables.
   b. Calcule e interprete el coeficiente de correlación lineal entre la longitud axial y el equivalente esférico.
   c. Calcule la ecuación de regresión entre la longitud axial y el equivalente esférico.
   d. Calcule e interprete el coeficiente de determinación.

6. En un experimento se ha tratado de medir la relación entre el porcentaje de frecuencia cardíaca de reserva (%HRR) y el máximo volumen de oxígeno en la sangre (medido como %VO2 máximo) en deportistas. En un grupo de 12 deportistas que se ejercitó en el escalador elíptico se obtuvieron los siguientes resultados:

%HRR	50	55	60	65	70	75	90	45	80	75	85	45
%VO2max.	55	50	65	75	75	60	95	40	90	80	80	50

   a. Construya el diagrama de dispersión.
   b. Calcule e interprete el coeficiente de correlación lineal de Pearson.
   c. Determine la ecuación de regresión que permita predecir el nivel de HRR según el nivel de VO2.
   d. Calcule e interprete el coeficiente de determinación.

7. Los siguientes datos muestran el área construida de una muestra de 8 viviendas nuevas en la ciudad capital y su respectivo precio. El área se expresa en metros cuadrados y el precio en miles de dólares.

Vivienda	1	2	3	4	5	6	7	8
Área construida	180	250	100	120	75	150	300	210
Precio	200	350	90	145	80	120	300	250

a. Construya el diagrama de dispersión y analícelo.
b. Calcule el coeficiente de correlación lineal e indique su significado en esta situación.
c. Calcule la ecuación de regresión e interprete sus coeficientes.
d. ¿Cuál sería el precio esperado de una vivienda de 160 metros cuadrados de construcción?
e. Calcule el coeficiente de determinación e interprételo. ¿Cuáles otros factores también explicarían la variabilidad del precio de la vivienda?
f. Si una vivienda tiene un precio de $190 mil, ¿cuál sería, según este modelo, su área construida?

8. Los siguientes datos muestran el número de artículos defectuosos por día obtenidos en una fábrica en una muestra de 8 días y el número de horas extra laboradas ese día en la fábrica.

Día	1	2	3	4	5	6	7	8
Unidades defectuosas	5	12	8	2	16	24	13	4
Horas extra	6	20	16	6	25	30	20	10

a. Construya el diagrama de dispersión y analícelo.
b. Calcule el coeficiente de correlación lineal e indique su significado en esta situación.
c. Calcule la ecuación de regresión e interprete sus coeficientes.
d. Calcule el coeficiente de determinación e interprételo.

9. Los siguientes datos muestran el presupuesto destinado a seguridad informática (en miles de dólares) en una muestra de 8 bancos y el número de empleados de la empresa (en miles).

Banco	1	2	3	4	5	6	7	8
Presupuesto	100	145	80	180	125	130	75	240
Número de empleados	3,5	7,0	2,4	12,1	6,2	5,7	4,8	10,6

a. Construya el diagrama de dispersión y analícelo.
b. Calcule el coeficiente de correlación lineal e indique su significado en esta situación.
c. Calcule la ecuación de regresión e interprete sus coeficientes.
d. Calcule el coeficiente de determinación e interprételo.

10. Los siguientes datos muestran el número de mensajes mensuales enviados por los usuarios de teléfonos celulares en una muestra de 8 personas y edad del usuario del teléfono.

Usuario	1	2	3	4	5	6	7	8
Número de mensajes	342	249	128	61	12	37	98	166
Edad	15	23	34	48	62	40	29	19

a. Construya el diagrama de dispersión y analícelo.
b. Calcule el coeficiente de correlación lineal e indique su significado en esta situación.
c. Calcule la ecuación de regresión e interprete sus coeficientes.
d. Calcule el coeficiente de determinación e interprételo.

11. La tabla presenta los reportes de ventas anuales y los años de experiencia de 10 vendedores de una prestigiosa empresa.

Años experiencia	1	3	4	4	6	8	10	10	11	13
Ventas (miles \$/año)	80	97	92	102	103	111	119	123	117	136

a. Establezca e interprete la ecuación de regresión que permita predecir las ventas con base en la experiencia del vendedor.
b. ¿En qué proporción este modelo explica las variaciones en las ventas?
c. Estime las ventas de un vendedor con 7,5 años de experiencia.
d. ¿Cuántos años de experiencia se esperaría que tenga un vendedor que logre vender \$100.000?

12. Un banco que se especializa en créditos para la vivienda intenta analizar el mercado, midiendo el poder explicativo que las tasas de interés tienen sobre el número de casas vendidas en el área. Se compilaron los datos para un período de seis años:

Semestre	1	2	3	4	5	6	7	8	9	10
Interés	12,3	10,5	15,6	9,5	10,5	9,3	8,7	14,2	15,2	12,0
Casas	196	285	125	225	248	303	265	102	105	114

a. Haga un diagrama de dispersión para los datos.
b. Calcule el coeficiente de correlación lineal e interprételo.
c. Calcule e interprete el modelo de regresión. ¿Qué dice este modelo sobre la relación entre las tasas de interés y las ventas de vivienda?
d. Calcule el coeficiente de determinación e interprételo.
e. Si la tasa de interés es del 9.5%, ¿cuántas casas se venderían de acuerdo con el modelo?
f. Obtenga una ecuación de regresión que permita predecir la tendencia de las ventas de casas (o sea, solo en función del tiempo). ¿Qué puede decirse de la calidad de los pronósticos elaborados con base en esa ecuación?

13. Un mercadólogo ha logrado obtener la siguiente información relacionada con el nivel medio de ingreso de los consumidores y la cantidad demandada de un cierto producto.

Año	Cantidad demandada por consumidor (en miles de unidades por año)	Nivel de ingreso medio (en dólares por mes)
2002	48	1592
2003	52	1600
2004	60	1607
2005	84	1620
2006	92	1689
2007	98	1734
2008	96	1830
2009	106	1835
2010	116	1741
2011	120	1755
2012	122	1867

El mercadólogo desea conocer lo siguiente:

a. ¿Qué tanta relación existe entre ambas variables? Conteste en términos del coeficiente de correlación lineal. Interprete claramente los resultados.
b. Trace el diagrama de dispersión.
c. Determinar una ecuación para poder estimar la cantidad demandada del artículo en función del ingreso del consumidor. Interprete los componentes de la ecuación de regresión. Trace la recta de la ecuación calculada.
d. ¿Qué tan bien se ajustan los datos obtenidos a la ecuación obtenida en el inciso c?
e. Si se espera que para el 2013 el nivel de ingreso sea de \$1895 por consumidor, ¿Cuál sería la cantidad demandada según este modelo?
f. ¿Cuánto es el valor del error de la regresión para 2010?
g. ¿A qué nivel tendría que aumentar el ingreso para poder vender 132.000 unidades en el 2013?
h. Interpole la cantidad demandada si el nivel de ingreso es de \$1730.
i. Extrapole la cantidad demandada si el nivel de ingreso es de \$1950.
j. Obtenga una ecuación de regresión que permita predecir la tendencia del ingreso de medio de los consumidores (o sea, solo en función del tiempo). ¿Qué puede decirse de la calidad de los pronósticos elaborados con base en esa ecuación?

14. Se ha establecido que la relación entre el número de años de experiencia de un vendedor (X) y los montos que logra vender por mes (Y, en miles de \$) está dada por la ecuación $Y = 0{,}3 + 2X$, entonces:

(1) Por cada año de experiencia se espera que sus ventas aumenten en \$2 mil al mes.
(2) Si tuviera cero experiencia, se esperaría que venda \$0,3 mil.

¿Son verdaderas ambas afirmaciones?

15. Un contador está analizando el comportamiento de los costos de producción con respecto al volumen total producido. Ha reunido información para varios periodos según la tabla siguiente:

Unidades producidas (miles/periodo)	5	9	14	11	17	20	17
Costos totales (miles $/periodo)	125	200	375	300	425	450	385

a. Construya el diagrama de dispersión. Utilice el número de unidades producidas en el eje horizontal.
b. Calcule e interprete el coeficiente de correlación lineal entre estas dos variables.
c. Calcule e interprete la ecuación de regresión entre estas dos variables.
d. ¿Cuál sería el nivel de costo total esperado al nivel de producción de 15 mil unidades?

16. Usted ha logrado obtener la siguiente información relacionada con el nivel medio de ingreso de los consumidores, la cantidad demandada de un cierto producto y las tasas de interés.

Año	Cantidad demandada por consumidor (en miles de unidades por año)	Nivel de ingreso medio (en dólares por mes)	Tasa de interés (%)
2000	48	892	18
2001	52	900	16
2002	60	1007	17
2003	84	1020	15
2004	92	1129	14
2005	98	1134	17
2006	96	1230	16
2007	106	1335	12
2008	116	1401	10
2009	120	1455	13
2010	122	1567	15

Usted desea conocer lo siguiente:

a. ¿Qué tanta relación lineal existe entre las variables demanda y tasas de interés? Conteste en términos del coeficiente de correlación lineal. Interprete claramente los resultados. Trace el diagrama de dispersión.
b. Determinar una ecuación lineal para poder estimar la cantidad demandada del artículo en función del ingreso del consumidor. Interprete los componentes de la ecuación de regresión. Trace la recta de la ecuación calculada.
c. ¿Qué tan bien se ajustan los datos obtenidos a la ecuación obtenida en el inciso b?
d. Con base en la ecuación del inciso b, si se espera que para el 2011 el nivel de ingreso sea de ¢1575 mil por consumidor, ¿Cuál sería la cantidad demandada según este modelo?
d. ¿A qué nivel tendría que aumentar el ingreso para poder vender 130.000 unidades en el 2011?
e. Obtenga una ecuación de regresión que permita predecir la tendencia de la cantidad demandada del producto (o sea, solo en función del tiempo). ¿Qué puede decirse de la calidad de los pronósticos elaborados con base en esa ecuación?

17. Un analista considera que la demanda de los diferentes tipos de seguros (de vida, automóviles, etc.) se relaciona con el producto interno bruto (PIB) del país. Para comprobar su hipótesis reunió datos de ambas variables para varios años (PIB en miles de millones de dólares y ventas de seguros en millones de dólares):

Año	1	2	3	4	5	6	7	8	9
PIB	1438	1575	1642	1739	1891	2041	2097	2076	2176
Ventas	1238	1382	1512	1680	1705	1914	1945	1965	2010

a. Calcule el coeficiente de correlación lineal e interprételo.
b. Calcule la ecuación de regresión lineal.
c. Si para el año 10 se espera un PIB de 2264 miles de millones de dólares, ¿cuánto se esperaría que sea la venta total de seguros en el país?
Obtenga una ecuación de regresión que permita predecir la tendencia de las ventas de seguros (o sea, solo en función del tiempo). ¿Qué puede decirse de la calidad de los pronósticos elaborados con base en esa ecuación?

18. Se ha establecido que la relación entre el gasto en publicidad de una empresa (X) y los montos que logra vender por mes (Y, en millones de $), está dada por $Y = 4{,}3 + 1{,}5X$, entonces:
a. Interpole el valor de las ventas cuando el gasto en publicidad sea de $5 millones.
b. Extrapole el valor de las ventas cuando el gasto en publicidad sea de $7,5 millones.

19. Una investigadora ha aplicado un instrumento estandarizado para la medición de los aprendizajes en preescolares de 5 años. La prueba fue desarrollada por la investigadora y utiliza una selección de aprendizajes esperados y cubre distintos ámbitos, como la formación social, la comunicación, la relación con el medio natural, entre otros. Al final se obtiene un puntaje total luego de la aplicación de todos los ítems de la prueba. Posteriormente se aplicó a la misma muestra una segunda prueba para la medición de los aprendizajes en preescolares de 5 años. Los siguientes son los resultados de ambas pruebas:

Niño	1	2	3	4	5	6	7	8	9
Prueba 1	60	53	47	29	38	45	56	49	42
Prueba 2	110	85	88	66	60	67	72	78	76

Determine qué tanta correlación hay entre los resultados de las dos pruebas. Calcule para ello el coeficiente de correlación de Pearson e interprete su resultado.

20. Los siguientes datos muestran el número promedio semanal de horas de uso internet de estudiantes de secundaria para hacer trabajos académicos en una muestra de 8 estudiantes y edad del estudiante.

Estudiante	1	2	3	4	5	6	7	8
Horas de uso	2,0	2,5	4	3,5	1,0	4,0	5,0	4,5
Edad	13	15	17	16	14	12	13	15

a. Construya el diagrama de dispersión y analícelo.
b. Calcule el coeficiente de correlación lineal e indique su significado en esta situación.
c. Calcule el coeficiente de determinación e interprételo.

21. Un investigador ha logrado obtener la siguiente información relacionada con el nivel medio de ingreso de hogares de diferentes barrios, la cantidad de hijos por hogar y el número de años de estudios formales del jefe del hogar.

Hogar #	Cantidad de hijos	Nivel de ingreso mensual ($)	Años de educación formal del jefe del hogar
1	5	1200	11
2	2	1534	15
3	3	585	8
4	4	560	6
5	1	1930	16
6	2	910	18
7	4	435	6
8	3	510	9
9	2	866	13
10	1	1190	17

El investigador desea conocer lo siguiente:

a. ¿Qué tanta relación lineal existe entre las variables ingreso y años de educación formal del jefe del hogar? ¿Se puede afirmar que existe una alta correlación entre las variables o no? ¿Qué tipo de relación se da entre las variables? Trace el diagrama de dispersión.
b. Una ecuación para poder estimar la el número de hijos del hogar en función de su ingreso.
c. ¿Qué tan bien se ajustan los datos obtenidos a la ecuación obtenida en el inciso b?
d. Si se analiza una familia cuyo ingreso es de $630, ¿cuántos hijos se esperaría que tuviera según este modelo?
e. Comente sobre la relación existente entre el nivel de ingreso y la cantidad de hijos habidos en el hogar. ¿Qué otras variables influyen y que pueden afectar la validez del modelo? Explique por qué.

22. Una persona considera que mientras más años se estudia, mayor será el cociente intelectual. Para probar esta relación se toma una muestra de 12 personas, que han estudiado las siguientes cantidades de años (años de estudios formales): 11, 9, 8, 15, 18, 10, 14, 16, 20, 5, 17, 16. Estas mismas personas realizaron un test para medir su cociente intelectual (los resultados corresponden al mismo orden anterior): 102, 100, 97, 101, 110, 113, 116, 96, 115, 106, 99, 104. Con base en estos datos:

a. Construya el diagrama de dispersión.
b. Calcule el coeficiente de correlación lineal de Pearson e interprételo.
c. ¿Vale la pena construir una ecuación de regresión para pronosticar el cociente intelectual de una persona con base en sus años de estudios formales?

23. Se ha establecido que la relación entre el gasto en publicidad de una empresa (X) y los montos que logra vender por mes (Y, en millones de \$), está dada por $Y = 4{,}3 + 1{,}5X$, entonces:
   a. Si la empresa no gasta en publicidad, entonces sus ventas serían de \$1,5 millones.
   b. Si la empresa gasta \$1 millón más en publicidad, se esperaría que sus ventas aumenten en \$4,3 millones.

   ¿Son verdaderas ambas afirmaciones?

24. Los siguientes datos muestran el volumen mensual de ventas de una compañía (en millones de dólares) en una muestra de 9 meses. Cada mes la empresa mide la exposición de los consumidores a los comerciales pautados por medio de una encuesta y determina el porcentaje de ellos que indican recordar los anuncios de la compañía.

Mes	1	2	3	4	5	6	7	8	9
Ventas	18	19	23	32	27	22	19	26	30
Exposición	40	42	45	46	43	45	38	40	42

   a. Construya el diagrama de dispersión y analícelo.
   b. Calcule el coeficiente de correlación lineal e indique su significado en esta situación.
   c. Calcule el coeficiente de determinación e interprételo.

25. El gobierno del país está preocupado por las críticas que las personas realizan a través de las redes sociales en Internet, por lo que decide tratar de contrarrestarlas publicando en esos mismos medios información relacionada con sus actividades y logros. Cada tres meses se realiza una encuesta en la que se evalúa la imagen del gobierno. La tabla muestra el número de publicaciones realizadas a su favor por el gobierno en las redes sociales en cada trimestre (*X*) y la opinión de los ciudadanos sobre su gestión (*Y*, medida como porcentaje de opiniones favorables en la encuesta al final de cada trimestre):

Trimestre	1	2	3	4	5	6	7	8	9	10
X	12	18	36	42	77	112	130	100	92	150
Y	29	36	40	42	40	46	38	35	31	24

   a. Construya el diagrama de dispersión y analice el comportamiento de las opiniones favorables con respecto al número de publicaciones efectuadas por el gobierno en las redes sociales.
   b. Calcule el coeficiente de correlación lineal de Pearson entre las dos variables e indique qué significa este coeficiente en este problema.
   c. ¿Vale la pena calcula la ecuación de regresión entre estas dos variables? ¿Sería razonable pronosticar la opinión de los ciudadanos sobre el gobierno a través del número de sus publicaciones en las redes sociales? Explique.

Examen del capítulo

En cada caso seleccione la opción que mejor contesta cada pregunta (las respuestas a los ejercicios se encuentran en la página de internet del texto: www.auladeeconomia.com/raeep.html).

1. Si un investigador descubre que conforme aumenta el número de usuarios de Facebook que son casados, también aumenta el número de divorcios, entonces podría considerar que:
   (a) El mayor uso de Facebook podría ser causante del aumento en el número de divorcios.
   (b) Existe una relación causa – efecto entre las dos variables.
   (c) Estas dos variables podrían correlacionarse.
   (d) La relación entre las dos variables es fuerte y directa.

2. Con relación a la determinación del grado de asociación estadística entre dos variables, un investigador efectuó las siguientes dos afirmaciones:
   A. Solo se trata de establecer la fuerza o intensidad de la relación.
   B. Se determina la naturaleza o dirección de la relación, pero no su intensidad.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

3. Con relación a la determinación del grado de asociación estadística entre dos variables, un investigador efectuó las siguientes dos afirmaciones:
   A. Una fuerte relación entre dos variables implica que exista causalidad.
   B. Una fuerte relación entre dos variables es condición necesaria de la existencia de causalidad entre ellas.

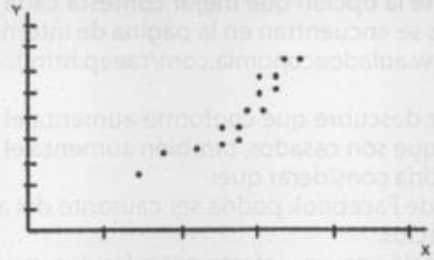
Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

4. Con relación a la determinación del grado de asociación estadística entre dos variables, un investigador efectuó las siguientes dos afirmaciones:
   A. Es posible encontrar un elevado coeficiente de correlación entre dos variables que no tienen relación alguna.
   B. Un alto coeficiente de correlación entre dos variables es espurio si éste se explica por la presencia de un tercer factor.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

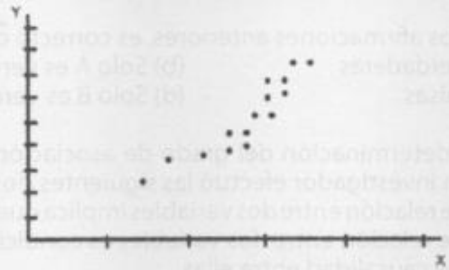
5. Con relación a la determinación de la causalidad entre dos variables, sería falso que la variable x causa a la variable y, si:
   (a) El coeficiente de correlación entre x y y es cercano a -1.
   (b) Las variaciones en x en un periodo podrían asociarse con las variaciones de y en el periodo siguiente.
   (c) Cambios de mayor magnitud en x no se asocian con cambios mayores en y.
   (d) Existe teoría que respalda la relación causal entre x y y.

6. Al observar la gráfica, podría afirmarse que es verdadero que:



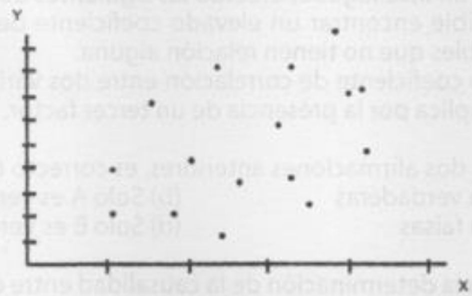
(a) El coeficiente de correlación de Pearson entre x y y es cercano a −1.
(b) El coeficiente de correlación de Pearson entre x y y es cercano a 1.
(c) El coeficiente de correlación de Pearson entre x y y es cercano a 0.
(d) El coeficiente de correlación de Pearson entre *x* y *y* es mayor que 1.

7. Al observar la gráfica, podría afirmarse que es verdadero que:



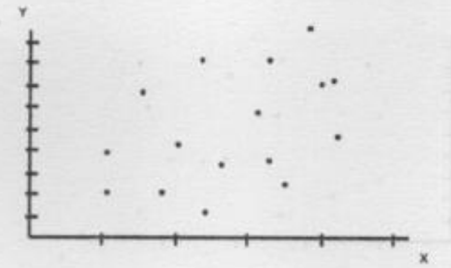
(a) La relación entre las variables x y y es fuerte e inversa.
(b) La relación entre las variables x y y es débil e inversa.
(c) La relación entre las variables x y y es fuerte y directa.
(d) La relación entre las variables *x* y *y* es débil y directa.

8. Al observar la gráfica, podría afirmarse que es verdadero que:



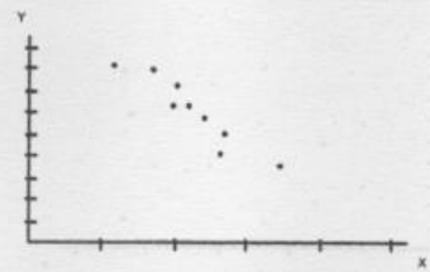
(a) El coeficiente de correlación de Pearson entre x y y es cercano a −1.
(b) El coeficiente de correlación de Pearson entre x y y es cercano a 1.
(c) El coeficiente de correlación de Pearson entre x y y es positivo y cercano a 0.
(d) El coeficiente de correlación de Pearson entre *x* y *y* es negativo y cercano a 0.

9. Al observar la gráfica, podría afirmarse que es verdadero que:



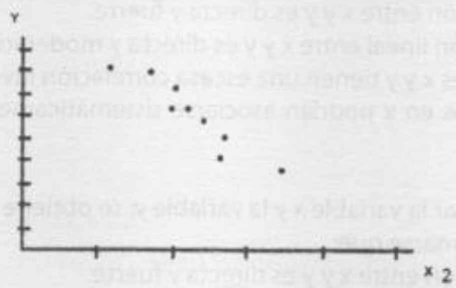
(a) La relación entre las variables x y y es fuerte e inversa.
(b) La relación entre las variables x y y es débil e inversa.
(c) La relación entre las variables x y y es fuerte y directa.
(d) La relación entre las variables *x* y *y* es débil y directa.

10. Al observar la gráfica, podría afirmarse que es verdadero que:



(a) El coeficiente de correlación de Pearson entre x y y es cercano a –1.
(b) El coeficiente de correlación de Pearson entre x y y es cercano a 1.
(c) El coeficiente de correlación de Pearson entre x y y es positivo y cercano a 0.
(d) El coeficiente de correlación de Pearson entre *x* y *y* es negativo y cercano a 0.

11. Al observar la gráfica, podría afirmarse que es verdadero que:

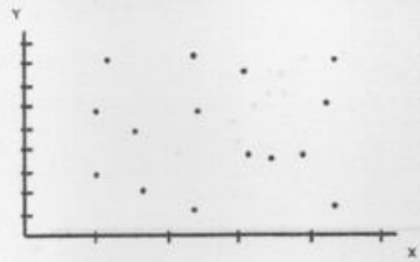


(a) La relación entre las variables x y y es fuerte e inversa.
(b) La relación entre las variables x y y es débil e inversa.
(c) La relación entre las variables x y y es fuerte y directa.
(d) La relación entre las variables x y y es débil y directa.

12. Al observar la gráfica, podría afirmarse que es verdadero que:

(a) El coeficiente de correlación de Pearson entre *x* y *y* es cercano a −1.
(b) El coeficiente de correlación de Pearson entre *x* y *y* es cercano a 1.
(c) El coeficiente de correlación de Pearson entre *x* y *y* es cercano a 0.
(d) Ninguna de las anteriores.

13. Al observar la gráfica, podría afirmarse que es verdadero que:



(a) La relación entre las variables *x* y *y* es fuerte e inversa.
(b) La relación entre las variables *x* y *y* es débil e inversa.
(c) La relación entre las variables *x* y *y* es fuerte y directa.
(d) No hay relación entre *x* y *y*.

14. Si al correlacionar la variable *x* y la variable *y*, se obtiene un coeficiente $r =$ $-0{,}87$, puede afirmarse que:
(a) La correlación entre x y y es directa y fuerte.
(b) La asociación lineal entre x y y es directa y moderada.
(c) Las variables x y y tienen una escasa correlación inversa.
(d) Incrementos en x podrían asociarse sistemáticamente con disminuciones en *y*.

15. Si al correlacionar la variable *x* y la variable *y*, se obtiene un coeficiente $r =$ $0{,}16$, puede afirmarse que:
(a) La correlación entre x y y es directa y fuerte.
(b) La asociación lineal entre x y y es directa y moderada.
(c) Las variables x y y tienen una escasa correlación inversa.
(d) Un aumento fuerte en x no podría asociarse sistemáticamente con un aumento en *y*.

16. Si se correlacionan las tasas de interés de los préstamos con la cantidad de viviendas vendidas por periodo, entonces se esperaría que el coeficiente de correlación entre estas dos variables sea:
   (a) Cercano a cero.
   (b) Positivo y cercano a uno.
   (c) Negativo.
   (d) Ninguna de las anteriores.

17. Si se correlaciona el ingreso disponible de un país con el nivel de consumo agregado, entonces se esperaría que el coeficiente de correlación entre estas dos variables sea:
   (a) Cercano a cero.
   (b) Positivo y cercano a uno.
   (c) Negativo.
   (d) Ninguna de las anteriores.

18. Suponga que se cuenta con los siguientes datos sobre dos variables *x* y *y*:

X	11	15	18	22	14	18	17	24
Y	61	68	73	78	69	71	74	76

Entonces el coeficiente de correlación lineal de Pearson equivale a:

(a) 0,83 (b) 1,16
(c) 0,911 (d) Ninguna de las anteriores

19. Suponga que se cuenta con los siguientes datos sobre dos variables, la humedad relativa en distintas zonas, y el número de casos de neumonía que se presentaron en un determinado periodo:

X	86	88	93	91	90	87	88	90
Y	11	9	15	17	10	13	16	17

Entonces el coeficiente de correlación lineal de Pearson equivale a:

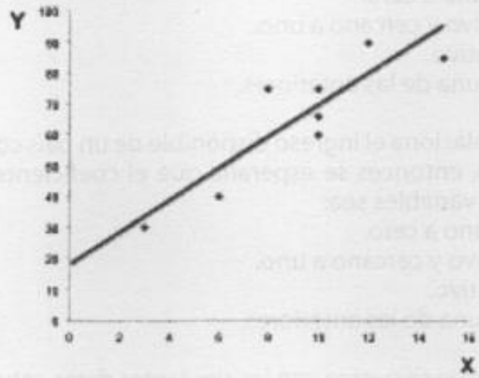
(a) 0,456 (b) 0,208
(c) 0,637 (d) Ninguna de las anteriores

20. Suponga que se cuenta con los siguientes datos sobre dos variables, la humedad relativa en distintas zonas, y el número de casos de neumonía que se presentaron en un determinado periodo:

Humedad relativa	86	88	93	91	90	87	88	90
Casos de neumonía	11	9	15	17	10	13	16	17

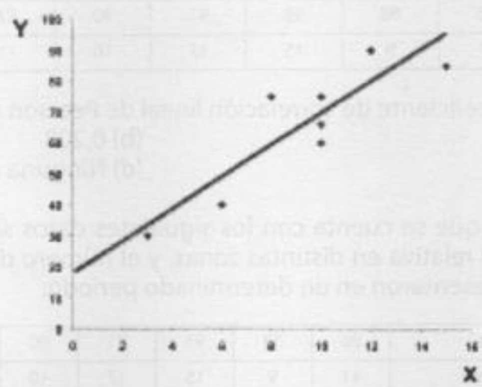
Entonces puede concluirse que:
   (a) La correlación entre la humedad relativa y el número de casos de neumonía es alta
   (b) La humedad relativa es claro determinante del número de casos de neumonía
   (c) Los puntos del diagrama de dispersión estarán muy cercanos a una línea recta
   (d) Ninguna de las anteriores

21. Al observar la gráfica, donde la línea corresponde a la recta de regresión obtenida por el método de mínimos cuadrados, podría afirmarse que es falso que:



(a) El coeficiente de correlación de Pearson entre x y y es positivo.
(b) La pendiente de la ecuación de regresión es positiva.
(c) La suma de los cuadrados de los residuos es máxima.
(d) El intercepto de la recta es cercano a 20.

22. Al observar la gráfica, donde la línea corresponde a la recta de regresión obtenida por el método de mínimos cuadrados, la variable *x* es el número semanal de unidades producidas en una fábrica y la variable *y* corresponde a los costos totales de producción, entonces es falso que:



(a) La pendiente de la recta es el costo incremental de una unidad producida.
(b) La recta estima los costos totales de la fábrica a distintos niveles de producción.
(c) La pendiente de la recta da el costo unitario de producción.
(d) El intercepto de la recta equivale a los costos fijos de producción.

23. Se ha establecido que la relación entre el número de años de experiencia de un vendedor (X) y los montos que logra vender por mes (Y, en miles de \$) está dada por $Y = 0,3 + 2X$, entonces no es verdadero que:
   (a) Por cada año de experiencia se espera que sus ventas aumenten en \$2 mil al mes.
   (b) Si tuviera cero experiencia, se esperaría que venda \$0,3 mil.
   (c) Si tuviera dos años de experiencia, se esperaría que venda \$4,6 millones.
   (d) Si tuviera un año de experiencia, se esperaría que venda \$2,3 millones.

24. Se ha establecido que la relación entre el gasto en publicidad de una empresa (X) y los montos que logra vender por mes (Y, en millones de \$), está dada por $Y = 4,3 + 1,5X$, entonces es verdadero que:
   (a) Si la empresa no gasta en publicidad, entonces sus ventas serían de \$1,5 millones.
   (b) Si la empresa gasta \$1 millón más en publicidad, se esperaría que sus ventas aumenten en \$4,3 millones.
   (c) El coeficiente de correlación lineal entre el gasto en publicidad y las ventas de la empresa es positivo.
   (d) Ninguna de las anteriores.

25. Suponga que se cuenta con los siguientes datos sobre dos variables, la humedad relativa en distintas zonas, y el número de casos de neumonía que se presentaron en un determinado periodo:

Humedad relativa	86	88	93	91	90	87	88	90
Casos de neumonía	11	9	15	17	10	13	16	17

Entonces puede concluirse que:
   (a) La pendiente de la ecuación de regresión es –43,3.
   (b) La pendiente de la ecuación de regresión es 0,637.
   (c) La pendiente de la ecuación de regresión es 0,456.
   (d) Ninguna de las anteriores.

26. Suponga que se cuenta con los siguientes datos sobre dos variables x y y:

X	11	15	18	22	14	18	17	24
Y	61	68	73	78	69	71	74	76

Entonces la ecuación de regresión lineal, tomando a x como variable independiente, es:
   (a) $y = 51,2 - 1,16x$
   (b) $y = -33,8 + 0,718x$
   (c) $y = 1,16x + 51,2$
   (d) Ninguna de las anteriores

27. Si al relacionar la variable *x* y la variable *y*, se obtiene un coeficiente R^2 = 0,87, entonces es falso con certeza que:
   (a) La correlación entre *x* y *y* es fuerte.
   (b) El modelo lineal entre *x* y *y* explica el 93,3% de la variabilidad de *y*.
   (c) El modelo lineal entre *x* y *y* explica el 87% de la variabilidad de *y*.
   (d) El coeficiente de correlación lineal entre las dos variables 0,933.

28. Al relacionar la variable *x* y la variable *y*, se obtiene un coeficiente R^2 = 0,96. Un investigador efectuó las siguientes dos afirmaciones:
   A. El modelo lineal entre *x* y *y* no es un buen modelo, porque tiene escaso poder explicativo.
   B. El ajuste de la recta es muy bueno.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

29. Suponga que se cuenta con los siguientes datos sobre dos variables *x* y *y*:

X	11	15	18	22	14	18	17	24
Y	61	68	73	78	69	71	74	76

Un investigador efectuó las siguientes dos afirmaciones:
   A. Al realizar la interpolación del valor *x* = 10, se obtiene *y* = 62,8.
   B. Al realizar la extrapolación del valor *x* = 12, se obtiene *y* = 65,12.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

30. Suponga que se cuenta con los siguientes datos sobre dos variables *x* y *y*:

X	14	18	11	20	23	14	17	19	15	16
Y	65	72	60	75	80	63	70	74	66	64

Un investigador efectuó las siguientes dos afirmaciones:
   A. Al realizar la extrapolación del valor *x* = 10, se obtiene *y* = 57.
   B. Al realizar la interpolación del valor *x* = 12, se obtiene *y* = 60,54.

Con respecto a las dos afirmaciones anteriores, es correcto que:
(a) Ambas son verdaderas (b) Solo A es verdadera
(c) Ambas son falsas (d) Solo B es verdadera

OTROS TEMAS

Los siguientes temas sólo se encuentran desarrollados en la página de internet de este texto:

- **Análisis de varianza** (ANOVA)
- **Chi-cuadrado**: Pruebas de bondad de ajuste, Pruebas de independencia y Pruebas de homogeneidad
- **Regresión múltiple**
- **Modelos de series de tiempo** y métodos de pronóstico
- **Números índice**
- **Indicadores estadísticos**: Indicadores económicos, indicadores sociales y de salud
- **Correlación en variables cualitativas**: coeficiente Q y coeficiente de correlación de Spearman

Es necesario ingresar a www.auladeeconomia.com/raeep.html y emplear su usuario y contraseña para acceder a estos materiales.

APÉNDICE

TABLAS Y FÓRMULAS

Apéndice 1: Fórmulas de estadística descriptiva

Medidas de posición:

Datos sin agrupar	Datos agrupados
Media aritmética: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$	**Media aritmética:** $\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$
Media aritmética ponderada: a. Caso de valores repetidos: $\bar{x} = \frac{\sum_{i=1}^{k} x_i f_i}{n}$ b. Caso de valores con diferente importancia: $\bar{x} = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i}$	**Mediana:** $Med = L_i + c\left[\frac{n/2 - F_{i-1}\downarrow}{f_i}\right]$ **Moda:** $M_o = L_i + c\frac{d_1}{d_1 + d_2}$
Media geométrica: $Mg = \sqrt[n]{x_1 \cdot x_2 \cdot ... \cdot x_n}$ **Moda:** Valor que se repite más veces **Mediana:** $P_{Med} = \frac{N+1}{2}$ **Percentiles:** $P_m = \left[\frac{m}{100}(n+1)\right]$	**Percentiles:** $P_m = L_i + c\left[\frac{\frac{m}{100}n - F_{i-1}\downarrow}{f_i}\right]$

Medidas de variabilidad:

Datos sin agrupar

Varianza:

– Población

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$$

– Muestra

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$$

Desviación estándar:

– Población

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} = \sqrt{\sigma^2}$$

– Muestra

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}} = \sqrt{s^2}$$

Fórmula alternativa para calcular la desviación estándar:

$$s = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - \frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1}}$$

Coeficiente de variación:

Población:

$$CV = \frac{\sigma}{\mu}\times 100$$

Muestra:

$$CV = \frac{s}{\overline{x}}\times 100$$

Datos agrupados

Varianza:

– Población

$$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i-\mu)^2 f_i}{N}$$

– Muestra

$$s^2 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2 f_i}{n-1}$$

Fórmula alternativa para calcular la varianza:

$$s^2 = \frac{\sum_{i=1}^{n}x_i^2 f_i - \frac{\left(\sum_{i=1}^{n}x_i f_i\right)^2}{n}}{n-1}$$

Desviación estándar:

– Población

$$\sigma = \sqrt{\sigma^2}$$

– Muestra

$$s = \sqrt{s^2}$$

Coeficiente de variación:

Población:

$$CV = \frac{\sigma}{\mu}\times 100$$

Muestra:

$$CV = \frac{s}{\overline{x}}\times 100$$

Apéndice 2: Fórmulas de probabilidad

Factorial:

$$n! = n \cdot (n-1)\,(n-2) \cdot \ldots \cdot 2 \cdot 1$$

Permutaciones:

$$P(n,r) = \frac{n!}{(n-r)!}$$

Combinaciones:

$$C(n,r) = \frac{n!}{r!\,(n-r)!}$$

Permutaciones con repetición:

$$\frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$$

Definición clásica de probabilidad:

$$P(A) = \frac{a}{N}$$

Frecuencia relativa como probabilidad:

$$P(A) = Fr(A) = \frac{F(A)}{N}$$

Teoremas de probabilidad:

$$0 \leq P(A) \leq 1$$

$$P(E^C) = 1 - P(E)$$

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

Reglas de adición de probabilidad:

$$P(A \cup B) = P(A) + P(B)$$

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

Reglas de multiplicación de probabilidad:

$$P(AB) = P(A) \cdot P(B)$$

$$P(AB) = P(A) \cdot P(B/A)$$

Teorema de Bayes:

$$P(A) = \sum \left(P(B_i)P(A/B_i)\right)$$

$$P(B/A) = \frac{P(B)P(A/B)}{P(B)P(A/B) + P(B')P(A/B')}$$

Apéndice 3: Fórmulas de distribuciones de probabilidad

Valor esperado:	Varianza:	Desviación estándar:
$E(X)=\mu=\sum x_i P(x_i)$	$\sigma^2=\sum (x_i-E(X))^2 P(x_i)$	$\sigma=\sqrt{\sigma^2}$

Distribuciones discretas:

Distribución binomial: $P(X/n,p)=C(n,x)p^x q^{n-x}$ $=\frac{n!}{x!(n-x)!}p^x q^{n-x}$ $\mu=np$ $\sigma=\sqrt{npq}$	**Distribución hipergeométrica:** $P(X/N,a,n)=\frac{\binom{N-a}{n-X}\binom{a}{X}}{\binom{N}{n}}$ **Distribución geométrica:** $g(x,p)=p(1-p)^{x-1}$
Distribución de Poisson: $P(X/\lambda)=\frac{\lambda^x e^{-\lambda}}{X!}$	**Distribución multinomial:** $f(x_1,x_2,\ldots,x_k)=\frac{n!}{x_1!x_2!\ldots x_k!}p_1^{x_1}p_2^{x_2}\ldots p_k^{x_k}$

Distribuciones continuas:

Distribución exponencial: $P=1-e^{-\lambda}$ $P=e^{-\lambda}$ $E(x)=\frac{1}{\lambda}$ $V(x)=\frac{1}{\lambda^2}$	**Distribución normal:** $z=\frac{x-\mu}{\sigma}$ Aproximación de la binomial por la normal: $z=\frac{x-\mu}{\sigma}=\frac{x-np}{\sqrt{npq}}$ $\mu=np$ $\sigma^2=npq$

Apéndice 4: Fórmulas de inferencia estadística

Error estándar	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$	$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
Intervalos de confianza para n^3 30 o s conocida	$\bar{x} \pm Z_{\alpha/2} \cdot \sigma / \sqrt{n}$	$\bar{x} \pm Z \cdot s / \sqrt{n}$
Intervalos de confianza para n < 30 y s desconocida	$\bar{x} \pm t \cdot s / \sqrt{n} =$	
Intervalos de confianza para una proporción	$p \underline{+} Z\sqrt{pq/n}$	
Tamaño de muestra en poblaciones infinitas	$n = \left(\frac{\sigma \cdot Z_{\alpha/2}}{E}\right)^2$	$n = P(1 - P)\,(Z/E)^2$
Tamaño de muestra en poblaciones finitas	$n_0 = \frac{n}{1+\frac{n}{N}}$	
Prueba de hipótesis para la media con n^3 30 o s conocida	$z_c = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$	$z_c = \frac{\bar{x}-\mu}{s/\sqrt{n}}$
Prueba de hipótesis para la media con n < 30 y s desconocida	$t_c = \frac{\bar{x}-\mu}{s/\sqrt{n}}$	
Prueba de hipótesis para una proporción	$z = \frac{x-nP}{\sqrt{nPQ}}$	
Prueba de hipótesis para la diferencia de dos medias con n^3 30 o s conocida	$z = \frac{(\bar{x}_1-\bar{x}_2)-\delta}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$	
Prueba de hipótesis para la diferencia de dos medias con n < 30 y s desconocida	$t = \frac{(\bar{x}_1-\bar{x}_2)-\delta}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}$	$\sigma^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
Prueba de hipótesis para la diferencia de dos medias con datos pareados	$t = \frac{\bar{x}_d}{s_d/\sqrt{n}}$	
Prueba de hipótesis para la diferencia de dos proporciones	$z = \frac{p_1-p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$	$p = \frac{x_1+x_2}{n_1+n_2}$

Apéndice 5: Fórmulas de regresión y correlación lineal simple

Coeficiente de correlación lineal de Pearson:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}}$$

Intersección de la recta de regresión:

$$a = \frac{\sum Y}{N} - b\frac{\sum X}{N}$$

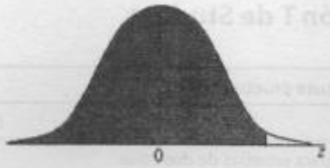
$$a = \overline{Y} - b\overline{X}$$

Pendiente de la recta de regresión:

$$b = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

Coeficiente de determinación:

$$R^2 = r^2$$



Apéndice 6: Distribución normal estándar acumulada

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998

Apéndice 7: Distribución T de Student

	Nivel de significancia para pruebas de una cola					
	0,1	0,05	0,025	0,01	0,005	0,0005
	Nivel de significancia para pruebas de dos colas					
gl	0,2	0,1	0,05	0,02	0,01	0,001
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,599
3	1,638	2,353	3,182	4,541	5,841	12,924
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,408
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,437
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1,333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,101	2,552	2,878	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2,086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,768
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,310	1,697	2,042	2,457	2,750	3,646

Apéndice 8: Percentiles de la distribución Chi-Cuadrado

	Valores de alfa									
gl	0,995	0,990	0,975	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88
2	0,01	0,02	0,05	0,21	1,39	4,61	5,99	7,38	9,21	10,60
3	0,07	0,11	0,22	0,58	2,37	6,25	7,81	9,35	11,34	12,84
4	0,21	0,30	0,48	1,06	3,36	7,78	9,49	11,14	13,28	14,86
5	0,41	0,55	0,83	1,61	4,35	9,24	11,07	12,83	15,09	16,75
6	0,68	0,87	1,24	2,20	5,35	10,64	12,59	14,45	16,81	18,55
7	0,99	1,24	1,69	2,83	6,35	12,02	14,07	16,01	18,48	20,28
8	1,34	1,65	2,18	3,49	7,34	13,36	15,51	17,53	20,09	21,95
9	1,73	2,09	2,70	4,17	8,34	14,68	16,92	19,02	21,67	23,59
10	2,16	2,56	3,25	4,87	9,34	15,99	18,31	20,48	23,21	25,19
11	2,60	3,05	3,82	5,58	10,34	17,28	19,68	21,92	24,72	26,76
12	3,07	3,57	4,40	6,30	11,34	18,55	21,03	23,34	26,22	28,30
13	3,57	4,11	5,01	7,04	12,34	19,81	22,36	24,74	27,69	29,82
14	4,07	4,66	5,63	7,79	13,34	21,06	23,68	26,12	29,14	31,32
15	4,60	5,23	6,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80
16	5,14	5,81	6,91	9,31	15,34	23,54	26,30	28,85	32,00	34,27
17	5,70	6,41	7,56	10,09	16,34	24,77	27,59	30,19	33,41	35,72
18	6,26	7,01	8,23	10,86	17,34	25,99	28,87	31,53	34,81	37,16
19	6,84	7,63	8,91	11,65	18,34	27,20	30,14	32,85	36,19	38,58
20	7,43	8,26	9,59	12,44	19,34	28,41	31,41	34,17	37,57	40,00
21	8,03	8,90	10,28	13,24	20,34	29,62	32,67	35,48	38,93	41,40
22	8,64	9,54	10,98	14,04	21,34	30,81	33,92	36,78	40,29	42,80
23	9,26	10,20	11,69	14,85	22,34	32,01	35,17	38,08	41,64	44,18
24	9,89	10,86	12,40	15,66	23,34	33,20	36,42	39,36	42,98	45,56
25	10,52	11,52	13,12	16,47	24,34	34,38	37,65	40,65	44,31	46,93
26	11,16	12,20	13,84	17,29	25,34	35,56	38,89	41,92	45,64	48,29
27	11,81	12,88	14,57	18,11	26,34	36,74	40,11	43,19	46,96	49,64
28	12,46	13,56	15,31	18,94	27,34	37,92	41,34	44,46	48,28	50,99
29	13,12	14,26	16,05	19,77	28,34	39,09	42,56	45,72	49,59	52,34
30	13,79	14,95	16,79	20,60	29,34	40,26	43,77	46,98	50,89	53,67

Apéndice 9: Valores de F con probabilidades de 5%

gl (denominador)	gl (numerador) 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161,4	199,5	215,7	224,6	230,2	234	236,8	238,9	240,5	241,9	243	243,9	244,7	245,4	245,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18	2,15	2,13
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01

Apéndice 10: Tabla de números aleatorios

	00-04	05-08	09-12	13-16	17-20	21-24	25-28	29-32
1	7726	9142	0509	3417	6443	7468	5143	7968
2	8801	6002	5259	9768	4499	2696	0877	9978
3	6009	5547	8606	9887	1964	8545	5722	7317
4	3481	2006	3831	9038	2583	1540	4953	0512
5	7797	4644	4293	9271	6062	7156	7326	8991
6	8752	2020	5904	7433	8420	6919	6927	2910
7	2791	4235	0722	8137	7215	3097	4187	4456
8	9875	9842	1686	5214	0793	6291	9259	0477
9	1221	5661	9184	0084	3971	1472	9627	1873
10	7123	8801	8455	6712	0954	4398	0655	9439
11	4899	0310	2029	7702	2743	9840	6441	0177
12	6606	5232	9366	9322	2790	2250	8859	822
13	1940	5929	7398	2336	2025	1655	8210	8216
14	1588	2678	4960	7533	5305	5142	1695	7840
15	7480	0040	2576	5797	9105	8726	9761	3024
16	7975	4093	5796	7168	8284	6291	9852	3314
17	5431	3722	3443	8111	7779	5324	8626	8356
18	1122	5219	8457	1967	8211	2308	7645	8414
19	9966	3483	1878	0115	7024	6604	7065	7372
20	5004	3073	1012	1344	9928	5689	6501	8544
21	2676	4479	1260	9629	1788	8160	3228	4977
22	8884	8927	736	8064	5324	9759	9538	3602
23	3098	9329	6283	6973	9107	0693	1650	6601
24	8199	2133	2267	5986	4630	0478	5626	9605
25	7005	5128	2380	4769	6022	0098	6130	0601

Este texto y todos los materiales digitales que lo acompañan han sido diseñados para ayudar al estudiante y al profesor en todos estos aspectos, pues provee al profesor de materiales que puede emplear en su clase para exponer los conceptos con claridad, presenta al estudiante materiales con ejercicios paso a paso, aplicaciones y amplio uso de la tecnología, de modo que pueda sentirse más motivado al disponer de recursos para adquirir los distintos conceptos y procedimientos, a la vez que se le ofrece gran cantidad de ejercicios resueltos, presentaciones interactivas, videos, entre otras ventajas.

Sobre el autor
Gabriel Leandro Oviedo

Es Licenciado en Economía y Máster en Administración de Empresas. También realizó estudios de Enseñanza de la Matemática.

Por más de 20 años se ha desempeñado como docente en diversas universidades del país.

Ha ocupado diversos cargos relacionados con la investigación, la estadística y la gestión de riesgos financieros. En ULACIT ocupó el cargo de Director de Investigación y Director de la Revista Universitaria Rhombus. Se ha desempeñado como experto en las auditorías de la Norma Nacional de los Principios de Inversión y Administración de Riesgos Financieros de los fondos de pensiones reguladas por la SUPEN.

Se ha desempeñado como consultor de la Cámara de Industrias de Costa Rica, gestor de riesgos en una entidad bancaria y director de AulaDeEconomia.com. Los materiales publicados a través de la página www.auladeeconomia.com son empleados por miles de estudiantes, profesores y profesionales en América Latina, España y Estados Unidos.

ISBN: 978-9977-987-72-9

PUBLITEX
Grupo Editorial S.A.