CAPÍTULO 22

Encuestas muestrales

CONTENIDO

LA ESTADÍSTICA *en* LA PRÁCTICA

DUKE ENERGY*
CHARLOTTE, CAROLINA DEL NORTE

Duke Energy es una empresa diversificada de energía con un portafolio de negocios en gas natural y electricidad y una empresa inmobiliaria afiliada. En el 2006, Duke Energy se fusionó con Cinergy de Cincinnati, Ohio, y formaron una de las empresas de energía más grandes de Estados Unidos, con un activo que asciende a más de $70 mil millones. En la actualidad, Duke Energy da servicio a más de 5.5 millones de usuarios de gas y electricidad en Carolina del Norte, Carolina del Sur, Ohio, Kentucky, Indiana y Ontario, Canadá.

Para dar un mejor servicio, Duke Energy constantemente está atenta a las necesidades emergentes de sus clientes. En el ejemplo siguiente se verá cómo esta empresa realizó una encuesta acerca de las características de los edificios para conocer mejor los requerimientos de energía de los edificios comerciales en el área de servicio de Cincinnati, Ohio.

La empresa buscó información diversa acerca de los edificios comerciales como arquitectura, cantidad de empleados, uso final dado a la energía, antigüedad del edificio, tipo de materiales de construcción y medidas para la conservación de la energía. Durante los preparativos de la encuesta, los analistas determinaron que en el área de servicio de Cincinnati había aproximadamente 27 000 edificios comerciales en función. De acuerdo con los recursos disponibles y con la precisión deseada para la encuesta, se recomendó tomar una muestra de 616 edificios comerciales.

El tipo de muestreo que se eligió fue un muestreo aleatorio simple estratificado. La empresa contaba con registros sobre el consumo de energía total en los años recientes de cada uno de los edificios en el área de servicio de Cincinnati y, dado que muchas de las características que interesaban de los edificios (tamaño, cantidad de empleados, etc.) estaban relacionadas con el consumo, ése fue el criterio empleado para dividir la población de edificios en seis estratos.

El primer estrato estaba constituido por los edificios que eran los 100 principales consumidores de energía; todos los edificios de este estrato fueron incluidos en la muestra. Aunque estos edificios constituían únicamente el 0.2% de la población, consumían el 14.4% de toda la energía eléctrica. De los otros estrados, el número de edificios muestreados se determinó en función de la obtención de la mayor precisión posible por costo unitario.

La empresa elaboró un cuestionario que se probó antes de realizar la encuesta. Los datos se obtuvieron a través de entrevistas personales. De los 616 edificios comerciales de la muestra se obtuvieron 526 cuestionarios completamente contestados. Esta tasa de respuesta de 85.4% fue excelente. La empresa usó los resultados de la encuesta para pronosticar la demanda de energía y para mejorar el servicio prestado a sus clientes comerciales.

En este capítulo, el lector conocerá los tópicos que consideran los estadísticos para el diseño y realización de una encuesta muestral como la realizada por Duke Energy. Las encuestas muestrales suelen emplearse para obtener perfiles de los clientes de una empresa; también son empleados por los gobiernos y por otras instituciones para conocer diversos segmentos de la población.

En Cincinnati, Ohio, se llevó a cabo una encuesta muestral sobre las necesidades de electricidad en los edificios comerciales. © Getty Images/PhotoDisc.

* Los autores agradecen a Jim Ruddle de Duke Energy por proporcionar este artículo para *La Estadística en la práctica*.

22.1 Terminología empleada en las encuestas muestrales

En el capítulo 1 se dieron las siguientes definiciones de elemento, población y muestra.

- **Elemento** es la entidad de la que se toman los datos.
- **Población** es la colección de todos los elementos que interesan.
- **Muestra** es un subconjunto de la población.

Para ilustrar estos conceptos considere la situación siguiente. Dunning Microsystems, Inc. (DMI), fabricante de computadoras personales y periféricos, desea obtener datos acerca de las características de las personas que le han comprado sus computadoras personales. Para esto, debe realizar una encuesta muestral a los poseedores de una computadora personal DMI. En esta encuesta muestral los *elementos* son cada uno de los individuos que hayan comprado una computadora personal DMI. La *población* es el conjunto de todas las personas que hayan comprado una computadora personal DMI y la *muestra* será el subconjunto de poseedores de una computadora personal que se tome para la encuesta.

En las encuestas muestrales es necesario distinguir entre la población objetivo y la población muestreada. La **población objetivo** es la población acerca de la cual se desean hacer inferencias, mientras que la **población muestreada** es la población de la que, realmente, se toma la muestra. Es importante entender que estas dos poblaciones no siempre son una misma. En el ejemplo de DMI, la población objetivo consta de todas las personas que han comprado una computadora personal DMI. La población muestreada, en cambio, puede que sea, por ejemplo, todos los poseedores de una computadora personal DMI que hayan enviado a DMI la tarjeta de registro para la garantía. No todos los que compran una computadora personal DMI envían la tarjeta de registro para la garantía, de manera que la población muestreada es diferente de la población objetivo.

Las inferencias obtenidas a partir de una muestra son válidas, si la población muestreada es representativa de la población objetivo.

Las conclusiones que se obtienen de una encuesta muestral sólo son válidas para la población muestral. El que estas conclusiones puedan o no ampliarse a la población objetivo depende del criterio del analista. El punto clave es si entre la población muestreada y la población objetivo existe una semejanza suficiente respecto a la característica de interés como para permitir ampliar las conclusiones.

Antes del muestreo, se divide la población en **unidades muestrales**. En algunos casos las unidades muestrales son simplemente los elementos. En otros casos, las unidades muestrales son grupos de elementos. Por ejemplo, suponga que se desea hacer una encuesta a los ingenieros que trabajan en el diseño de sistemas de calefacción y de aire acondicionado para edificios comerciales. Si se tuviera una lista de todos estos ingenieros, las unidades muestrales serían los ingenieros que se desea investigar. Si no se cuenta con tal lista, es necesario hallar otra alternativa. Una alternativa puede ser la lista de las empresas de ingeniería que se dedican al diseño de sistemas de calefacción y aire acondicionado que se encuentran en un directorio telefónico comercial. Dada tal lista se toma una muestra de estas empresas para la encuesta; en cada empresa tomada para la encuesta se entrevista a todos los ingenieros. En este caso las unidades muestrales serán las empresas de ingeniería y los elementos serán los ingenieros entrevistados.

A la lista de las unidades muestrales tomadas para un estudio particular se le conoce como el **marco**. En la encuesta a los ingenieros el marco está definido por todas las empresas de ingeniería enumeradas en el directorio telefónico; el marco no es una lista de todos los ingenieros porque no se cuenta con tal lista. El marco que se elija y, por tanto, la definición de las unidades muestrales, suele estar determinado por la lista de que se disponga y la confiabilidad de la misma. En la práctica la elección del marco suele ser uno de los pasos más difíciles e importantes al realizar una encuesta muestral.

22.2 Tipos de encuestas y métodos de muestreo

Los tres tipos de encuestas muestrales más comunes son las encuestas por correo, las encuestas por teléfono y las encuestas a través de entrevistas personales. Hay otros tipos de investigaciones que se emplean para recabar datos en los que no se emplean cuestionarios. Por ejemplo, para muestrear el inventario de bienes de una empresa con objeto de estimar el valor de inventario en el balance general de la empresa suele contratarse a una empresa de contadores. En estas investigaciones, una persona simplemente cuenta los artículos y anota los resultados.

En las encuestas en que se usan cuestionarios, el diseño del cuestionario es relevante. Al hacer el diseño de un cuestionario hay que resistirse a incluir preguntas que *pueden* ser de interés, ya que cada pregunta agregada al cuestionario lo hace más largo. Los cuestionarios largos no sólo conducen al cansancio del entrevistado, sino también al del entrevistador, en especial cuando se trata de encuestas por correo o por teléfono. Cuando se emplean entrevistas personales, es po-

Los costos de las encuestas por correo o por teléfono son más bajos, pero las entrevistas personales, cuando se cuenta con entrevistadores bien capacitados, suelen dar tasas de respuesta más altas y permitir cuestionarios más largos. En el caso del estudio presentado en el artículo de La estadística en la práctica de este capítulo, dada la cantidad de datos que se deseaba recabar de cada elemento, la única posibilidad de hacerlo era mediante entrevistas personales.

sible hacer cuestionarios más largos y más complejos. Ya existe una gran cantidad de conocimientos sobre la redacción y secuencia de las preguntas para un cuestionario, así como sobre la manera de agruparlas. Estos temas corresponden a libros especializados sobre encuestas muestrales; en la bibliografía se citan varias fuentes en las que se encontrará este tipo de información.

Las encuestas muestrales también se clasifican de acuerdo con el método de muestreo que se utilice. Los **muestreos probabilísticos** permiten calcular la probabilidad de obtener cada una de las posibles muestras; en los **muestreos no probabilísticos** esto no es posible. Los métodos no probabilísticos de muestreo no deben usarse cuando el investigador desea determinar la precisión de las estimaciones. En cambio, los métodos probabilísticos de muestreo se emplean para obtener intervalos de confianza con los que se pueden obtener límites para el error muestral. En las secciones siguientes se estudiarán cuatro de los métodos de muestreo probabilístico más usados: el muestreo aleatorio simple, el muestreo aleatorio simple estratificado, el muestreo por conglomerados y el muestreo sistemático.

Aunque los especialistas en estadística prefieren usar los métodos probabilísticos de muestreo, los métodos no probabilísticos de muestreo también suelen ser necesarios. Las ventajas de los métodos de muestreo no probabilísticos son su bajo costo y su fácil realización. La desventaja es que no se puede decir de una manera estadística válida cuál es la precisión de la estimación. Dos de los métodos probabilísticos más usados son el muestreo de conveniencia y el muestreo subjetivo.

En el **muestreo de conveniencia** las unidades que se toman en la muestra, se toman por su accesibilidad. Por ejemplo, cuando un profesor de una universidad que realiza una investigación suele solicitar alumnos voluntarios que participen en el estudio, estos alumnos participan en la muestra sólo porque son alumnos del profesor. En este caso a la muestra de estudiantes se le conoce como muestra de conveniencia. En algunos casos el muestreo de conveniencia es la única posibilidad práctica. Por ejemplo, para muestrear un cargamento de naranjas, el investigador tomará de manera aleatoria naranjas de varias cajas ya que no sería práctico etiquetar todas las naranjas del cargamento para obtener un marco y emplear un método de muestreo probabilístico. Otros ejemplos de muestreos de conveniencia son los estudios sobre la flora y fauna salvajes y los paneles de voluntarios en las investigaciones de mercado.

Aun cuando el muestreo de conveniencia es una manera relativamente sencilla de seleccionar una muestra y obtener los datos deseados, es imposible estimar la “bondad” de los estadísticos muestrales obtenidos como estimaciones de los parámetros poblacionales que interesan. Un muestreo de conveniencia puede o no dar buenos resultados; no hay ningún procedimiento estadísticamente justificado que permita hacer inferencias muestrales a partir de los resultados muestrales. A pesar de esto, algunas veces, los investigadores aplican métodos probabilísticos, diseñados para muestras probabilísticas, a datos obtenidos mediante un muestreo de conveniencia. En tales casos, el investigador suele argumentar que la muestra de conveniencia puede ser considerada como una muestra aleatoria en el sentido de que es representativa de la población. Pero, hay que cuestionar este argumento; se debe ser muy cuidadoso al emplear muestras de conveniencia para hacer inferencias estadísticas acerca de los parámetros poblacionales.

En la técnica de muestreo conocida como **muestreo subjetivo**, una persona, con conocimientos en la materia de estudio, selecciona las unidades muestrales que considera más representativas de la población. El muestreo subjetivo suele ser una manera relativamente sencilla para tomar una muestra, sin embargo, los usuarios de los resultados de tales encuestas deben aceptar que la calidad de los resultados es dependiente del criterio de la persona que selecciona la muestra. Por tanto, se debe tener mucho cuidado al usar muestras subjetivas para hacer inferencias estadísticas acerca de los parámetros poblacionales. En general, conviene no hacer aseveraciones estadísticas acerca de la precisión de los resultados obtenidos de una muestra subjetiva.

En el caso de los métodos de muestreo no probabilístico, cuando se pueden emplear métodos que garanticen que se ha obtenido una muestra representativa, las estimaciones puntuales basadas en la muestra pueden ser útiles. Sin embargo, incluso en estos casos no se puede conocer la precisión de los resultados.

Para tomar una muestra se pueden usar tanto métodos probabilísticos como no probabilísticos. La ventaja de los métodos no probabilísticos es que, por lo general, no son caros y son fáciles de usar. Pero cuando se necesita indicar la precisión de las estimaciones, será necesario emplear métodos probabilísticos de muestreo. En casi todas las encuestas muestrales grandes se emplean métodos probabilísticos.

22.3 Errores en una encuesta

Al realizar una encuesta se pueden presentar dos tipos de errores. Uno, el error muestral, que es la magnitud de la diferencia entre el estimador puntual insesgado obtenido de la muestra y el parámetro poblacional. En otras palabras, el error muestral es el error que se presenta debido a que no se investigan todos los elementos de la población. El segundo error es el error no muestral, que se refiere a todos los demás tipos de errores que se presentan cuando se realiza una encuesta, como errores de medición, errores del entrevistador y errores de procesamiento. Los errores muestrales sólo pueden presentarse en una encuesta muestral; los errores no muestrales ocurren tanto en un censo como en una encuesta muestral.

Errores no muestrales

Uno de los errores no muestrales más comunes se presenta cuando una característica de interés es medida de forma incorrecta. Los errores de medición ocurren tanto en un censo como en una encuesta muestral. En cualquier tipo de encuesta es necesario tener cuidado de que todos los instrumentos de medición (por ejemplo, los cuestionarios) estén adecuadamente calibrados y de que las personas que hagan las mediciones estén debidamente capacitadas. Poner atención a los detalles es la mejor precaución en la mayor parte de las situaciones.

Los errores no muestrales se minimizan mediante una capacitación adecuada de los entrevistadores, un buen diseño de los cuestionarios, que deben ser probados antes de ser empleados en la encuesta, y cuidado en el proceso de codificación y transferencia de los datos a la computadora.

Los errores debidos a la falta de respuestas preocupan tanto al especialista en estadística, que es el responsable del diseño de la encuesta, como al ejecutivo que usará los resultados de la misma. Este tipo de error no muestral se presenta siempre que no es posible obtener, de algunas de las unidades de la encuesta, los datos deseados, o cuando únicamente se obtienen datos parciales. Un problema más serio es cuando se crea un sesgo. Por ejemplo, si se realizan entrevistas para evaluar la opinión de las mujeres respecto de que las mujeres trabajen fuera de casa y se llama a los hogares únicamente durante el día, se creará un sesgo obvio, debido a que las mujeres que trabajan fuera de casa quedarán excluidas de la muestra.

En el censo llevado a cabo en Estados Unidos en 1990, 25.9% de los hogares no respondieron. En el censo de 2000 se hizo un estudio muestral de los que no respondían al censo con objeto de estimar las características de esta porción de la población.

En encuestas técnicas son comunes los errores no muestrales que se deben a falta de conocimientos de los entrevistados. Por ejemplo, suponga que se hace una encuesta entre los administradores de edificios para obtener información detallada acerca del tipo de sistemas de ventilación que se usan en los edificios de oficinas. Los administradores de edificios grandes de oficinas tendrán buenos conocimientos acerca de tales sistemas, ya que es probable que hayan asistido a seminarios y obtengan apoyos para mantenerse informados y al día. En cambio, es posible que los administradores de edificios pequeños tengan menos conocimientos acerca de tales sistemas, debido a la gran variedad de tareas que deben realizar. Esta diferencia en los conocimientos afecta significativamente los resultados de la encuesta.

Otros dos tipos de errores no muestrales son el error de selección y el error de procesamiento. El error de selección se presenta cuando en la muestra se incluye algún elemento que no sea adecuado. Suponga que se diseña una encuesta muestral para obtener el perfil de un hombre con barba; si algunos entrevistadores entienden que entre los "hombres con barba" están comprendidos los hombres con bigote, mientras que otros entienden que no lo están, los datos resultantes serán deficientes. Los errores de procesamiento se presentan cuando los datos son anotados con incorrecciones o cuando son transferidos de manera incorrecta, por ejemplo, de los cuestionarios a la computadora.

Aun cuando algunos de los errores no muestrales se presentan en la mayor parte de las encuestas, es posible minimizarlos mediante una planeación cuidadosa. Debe tener cuidado de que haya una estrecha correspondencia entre la población muestreada y la población objetivo; que se sigan los buenos principios para la formulación de cuestionarios; que los entrevistadores estén bien capacitados, etc. En el informe final de una encuesta es recomendable incluir un análisis sobre el impacto que pueden tener los errores no muestrales sobre los resultados.

Error muestral

Recuerde la encuesta muestral de Dunning Microsystems (DMI). Suponga que DMI desea estimar la edad promedio de las personas que compran una computadora personal. Si se pudiera investigar a toda la población de personas que poseen una computadora DMI (hacer un censo) y no se cometiera ningún error no muestral, se podría determinar esta edad promedio con toda exactitud. Pero, ¿qué pasa si no se puede investigar el 100% de todos los propietarios de una computadora DMI? En este caso, es posible que exista alguna diferencia entre la media muestral y la media poblacional; al valor absoluto de esta diferencia se le conoce como error muestral. En la práctica no es posible determinar cuál es el error muestral en una muestra determinada, ya que no es posible conocer la media poblacional, sin embargo, sí es posible dar una estimación probabilística acerca del tamaño del error muestral.

El error muestral se minimiza al elegir un diseño adecuado para la muestra.

Como ya se dijo, el error muestral se debe a que la encuesta se hace a partir de una muestra y no de toda la población. Aun cuando el error muestral no puede evitarse, sí es controlable. Una manera de controlar este tipo de error es elegir un método o diseño apropiado de muestreo. En las secciones siguientes se verán cuatro métodos de muestreo probabilístico: aleatorio simple, aleatorio estratificado, por conglomerados y sistemático.

22.4 Muestreo aleatorio simple

La definición de muestreo aleatorio simple se presentó en el capítulo 7:

> Una muestra aleatoria simple de tamaño n tomada de una población finita de tamaño N es una muestra que se elige de tal manera que todas las muestras posibles de tamaño n tengan la misma probabilidad de ser elegidas.

Para realizar una encuesta muestral usando el **muestreo aleatorio simple**, se empieza por elaborar un marco o lista de todos los elementos de la población muestral. A continuación se emplea un procedimiento de selección que se basa en el uso de números aleatorios, para garantizar que todos los elementos de la población muestral tengan la misma probabilidad de ser elegidos para la muestra. En esta sección se verá cómo se obtienen estimaciones de la media, del total y de la proporción poblacionales cuando en una encuesta muestral se usa el muestreo aleatorio simple.

Media poblacional

En el capítulo 8 se vio que la media muestral $\bar{x}$ es una estimación de la media poblacional μ y que la desviación estándar muestral s es una estimación de la desviación estándar poblacional σ. Un intervalo de estimación para μ, dada una muestra de tamaño n y empleando la distribución t, es el siguiente.

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} \tag{22.1}$$

En la expresión (22.1), $s/\sqrt{n}$ es la estimación de $\sigma_{\bar{x}}$, el error estándar de la media.

Cuando la muestra aleatoria simple de tamaño n se toma de una población finita de tamaño N, la estimación del error estándar de la media se obtiene mediante la fórmula siguiente

$$s_{\bar{x}} = \sqrt{\frac{N-n}{N}}\left(\frac{s}{\sqrt{n}}\right) \tag{22.2}$$

Al usar $s_{\bar{x}}$ como estimación de $\sigma_{\bar{x}}$ el intervalo de estimación para la media poblacional se convierte en

$$\bar{x} \pm t_{\alpha/2}s_{\bar{x}} \tag{22.3}$$

En las encuestas muestrales se acostumbra emplear el valor $t = 2$ para obtener una estimación por intervalo. Por tanto, cuando se emplea el muestreo aleatorio simple, el intervalo de confianza de aproximadamente 95% para estimar la media poblacional es el dado por la expresión siguiente.

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA ESTIMAR LA MEDIA POBLACIONAL

$$\bar{x} \pm 2s_{\bar{x}} \tag{22.4}$$

Considere, por ejemplo, el caso de la editorial de la revista *Great Lakes Recreation,* una revista regional especializada en navegación y pesca. En la actualidad, la revista cuenta con 8 000 suscriptores. En una muestra aleatoria simple de $n = 484$ suscriptores el ingreso anual medio encontrado fue \$30 500 y la desviación estándar \$7 040. Una estimación insesgada del ingreso anual medio de todos los suscriptores es $\bar{x} = \$30\,500$. Con estos resultados muestrales y la ecuación (22.2) se obtiene la estimación siguiente para el error estándar de la media.

$$s_{\bar{x}} = \sqrt{\frac{8000 - 484}{8000}}\left(\frac{7040}{\sqrt{484}}\right) = 310$$

Por tanto, de acuerdo con la fórmula (22.4), un intervalo de confianza de aproximadamente 95% para el ingreso anual medio de los suscriptores, es

$$30\,500 \pm 2(310) = 30\,500 \pm 620$$

es decir, \$29 880 a \$31 120.

El número que se le suma y se le resta a la estimación puntual para obtener el intervalo de estimación, se conoce como **cota del error muestral**. Por ejemplo, en la encuesta muestral de *Great Lakes Recreation,* una estimación del error estándar del estimador puntual es $s_{\bar{x}} = \$310$, y la cota del error muestral es $2(\$310) = \$620$.

El procedimiento anterior también sirve para calcular intervalos de estimación para otros parámetros poblacionales, como, por ejemplo, para el total poblacional y para la proporción poblacional. En estos casos, un intervalo de confianza de aproximadamente 95% puede expresarse de la manera siguiente

Estimador puntual ± 2(Estimación del error estándar del estimador puntual)

Total poblacional

Considere el problema que se le plantea a la empresa Northeast Electric and Gas (NEG). Como parte de un estudio sobre el consumo de energía, NEG necesita estimar el área *total*, en pies cuadrados, de las 500 escuelas públicas en su área de servicio. Esta área total de las 500 escuelas públicas se denotará como X; en otras palabras, X denota la población total. Observe que si se conociera μ, el promedio en pies cuadrados de las 500 escuelas públicas, al multiplicar N por μ se obtendría el valor de X. Pero, como no se conoce μ, una estimación puntual de X es la que se obtiene al multiplicar N por $\bar{x}$. El estimador puntual de X se denota $\hat{X}$.

ESTIMADOR PUNTUAL DEL TOTAL POBLACIONAL

$$\hat{X} = N\bar{x} \tag{22.5}$$

La estimación del error estándar de este estimador puntual está dada por

$$s_{\hat{X}} = Ns_{\bar{x}} \tag{22.6}$$

donde

$$s_{\bar{x}} = \sqrt{\frac{N - n}{N}}\left(\frac{s}{\sqrt{n}}\right) \tag{22.7}$$

Observe que la ecuación (22.7) es la fórmula obtenida para la estimación de error estándar de la media. Para obtener un intervalo de confianza de aproximadamente 95% para estimar el total poblacional se emplea este error estándar y la ecuación (22.6).

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA ESTIMAR EL TOTAL POBLACIONAL

$$N\bar{x} \pm 2s_{\hat{X}} \tag{22.8}$$

Suponga que en el estudio de NEG se toma una muestra aleatoria de $n = 50$ escuelas públicas de la población de $N = 500$ escuelas; la media muestral es $\bar{x} = 22\,000$ pies cuadrados y la desviación estándar muestral es $s = 4\,000$ pies cuadrados. Mediante la ecuación (22.5) se obtiene

$$\hat{X} = (500)(22\,000) = 11\,000\,000$$

Para obtener una estimación del error estándar de la media se emplea la ecuación (22.7).

$$s_{\bar{x}} = \sqrt{\frac{500 - 50}{500}}\left(\frac{4\,000}{\sqrt{50}}\right) = 536.66$$

Después, con la ecuación (22.6), se obtiene una estimación del error estándar de $\hat{X}$.

$$s_{\hat{X}} = (500)(536.66) = 268\,330$$

Por tanto, con la expresión (22.8), se encuentra que un intervalo de confianza de aproximadamente 95% para el total de pies cuadrados de las 500 escuelas públicas en el área de servicio de NEG es

$$11\,000\,000 \pm 2(268\,330) = 11\,000\,000 \pm 536\,660$$

es decir, 10 463 340 a 11 536 660 pies cuadrados.

Proporción poblacional

Una proporción poblacional p es la fracción de elementos de la población que posee alguna característica de interés. Por ejemplo, en un estudio de investigación de mercado el interés puede ser la proporción de consumidores que prefieren determinada marca de un producto. La proporción muestral $\bar{p}$ es un estimador puntual insesgado de la proporción poblacional. Un estimador del error estándar de la proporción es el dado por

$$s_{\bar{p}} = \sqrt{\left(\frac{N - n}{N}\right)\left(\frac{\bar{p}(1 - \bar{p})}{n - 1}\right)} \tag{22.9}$$

Un intervalo de confianza de aproximadamente 95% para estimar la proporción muestral es el dado por la expresión siguiente.

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA ESTIMAR LA PROPORCIÓN POBLACIONAL

$$\bar{p} \pm 2s_{\bar{p}} \tag{22.10}$$

Como ejemplo, suponga que en el problema de muestreo de Northeast Electric and Gas, también quiere estimar la proporción de las 500 escuelas públicas, en su área de servicio, que emplean gas natural para la calefacción. Si 35 de las 50 escuelas muestreadas indican que usan gas natural, la estimación puntual de la proporción poblacional, en las 500 escuelas de la población que usan gas natural, es $\bar{p} = 35/50 = 0.70$. Mediante la ecuación (22.9), se calcula la estimación del error estándar de la proporción.

$$s_{\bar{p}} = \sqrt{\left(\frac{500 - 50}{500}\right)\left(\frac{0.7(1-0.7)}{50 - 1}\right)} = 0.0621$$

Por tanto, con la expresión (22.10) se encuentra que un intervalo de confianza de aproximadamente 95% para la proporción poblacional es

$$0.7 \pm 2(0.0621) = 0.7 \pm 0.1242$$

es decir, 0.5758 a 0.8242.

Como se ve en este ejemplo, en una estimación de la proporción poblacional la amplitud del intervalo de confianza puede ser bastante grande. En general, para obtener estimaciones precisas de las proporciones poblacionales se necesitan tamaños de muestra grandes. Louis Harris & Associates en un informe de una encuesta muestral de 529 inversionistas de fondos mutualistas, dice: "Los resultados deben tener una exactitud de 4.3 puntos porcentuales." Esto significa que el intervalo de confianza de aproximadamente 95% tiene una amplitud de 0.086. En poblaciones grandes, muestras de $n = 200$, o más, son frecuentes.

Determinación del tamaño de la muestra

Una consideración importante en el diseño de la muestra es la elección de su tamaño. Lo mejor suele ser un compromiso entre costo y precisión. Las muestras mayores permiten una mejor precisión (cotas más estrechas del error muestral) pero son más costosas. Con frecuencia, lo que dicta el tamaño de la muestra es el presupuesto con que se cuenta para el proyecto. En otras ocasiones, el tamaño de la muestra debe ser lo suficientemente grande para que permita obtener un determinado nivel de precisión.

El método que suele emplearse para determinar el tamaño de la muestra es, primero, especificar la precisión deseada y después determinar el menor tamaño de muestra con el que se obtiene esa precisión. En el presente contexto, el término *precisión* se refiere al tamaño del intervalo de confianza aproximado; intervalos de confianza más pequeños proporcionan mayor precisión. Como el tamaño del intervalo de confianza aproximado depende de la cota B del error muestral, elegir un nivel de precisión equivale a elegir un valor para B. A continuación se muestra este método para la elección del tamaño de muestra necesario para estimar la media poblacional.

La ecuación (22.2) indica que la estimación del error estándar de la media es

$$s_{\bar{x}} = \sqrt{\frac{N - n}{N}}\left(\frac{s}{\sqrt{n}}\right)$$

Recuerde que la cota del error muestral es "2 multiplicado por la estimación del error estándar del estimador puntual". Por tanto,

$$B = 2\sqrt{\frac{N - n}{N}}\left(\frac{s}{\sqrt{n}}\right) \tag{22.11}$$

Al despejar n en la ecuación (22.11), se obtiene una cota del error muestral igual a B. De esta manera

$$n = \frac{Ns^2}{N\left(\frac{B^2}{4}\right) + s^2} \tag{22.12}$$

Una vez elegido el nivel de precisión (un valor para B), al aplicar la ecuación (22.12) se obtiene el valor de n que permite obtener la precisión deseada. Pero el empleo de la ecuación (22.12) para elegir el valor de n presenta algunos problemas, pues además de especificar el valor de la cota deseada B del error muestral, se necesita el valor de la varianza muestral s^2, pero s^2 no se puede conocer sino hasta que se tome la muestra.

Cochran,* sugiere varias maneras prácticas para obtener el valor de s^2. Tres de ellas son las siguientes:

1. Tomar la muestra en dos etapas. Usar en la ecuación (22.12) el valor de s^2 hallado en la etapa 1; el valor que se obtenga para n es el tamaño que debe tener la muestra. Después, en la etapa 2 tomar el número de unidades adicionales necesarias para alcanzar el tamaño total de la muestra, determinada en la etapa 1.
2. Usar los resultados de una encuesta piloto o de una prueba preliminar para estimar s^2.
3. Usar la información de una muestra previa.

A continuación se verá un ejemplo en el que se estiman los salarios medios iniciales de los egresados de una determinada universidad. Suponga que hay 5 000 egresados y se quiere obtener un intervalo de confianza de aproximadamente 95% cuya amplitud sea, a lo más, \$1 000. Para obtener este intervalo de confianza se requiere que $B = 500$. Antes de usar la ecuación (22.12) para estimar el tamaño de la muestra, se necesita estimar s^2. Suponga que en un estudio similar realizado el año anterior se encontró $s = \$3\,000$. Para estimar s^2 se pueden emplear los datos de esa muestra anterior. Ahora ya se puede usar la ecuación (22.12), con $B = 500$, $s = 3\,000$ y $N = 5\,000$, para determinar el tamaño de la muestra.

$$n = \frac{5\,000(3\,000)^2}{5\,000\left(\frac{(500)^2}{4}\right) + (3\,000)^2}$$
$$= 139.97$$

Al redondear hacia arriba, se halla que para obtener un intervalo de confianza de aproximadamente 95% cuya amplitud sea \$1 000, se necesita que el tamaño de la muestra sea de 140. Pero hay que tener presente que estos cálculos se hicieron con base en la estimación inicial $s = \$3\,000$. Si en la encuesta muestral de este año, s resulta ser más grande, la amplitud del intervalo de confianza que se obtenga será mayor que \$1 000. En consecuencia, si el presupuesto lo permite, se deberá elegir una muestra de, por ejemplo, 150 para garantizar que el intervalo de confianza que se obtenga tenga una amplitud menor que \$1 000.

La fórmula para determinar el tamaño de muestra necesario para estimar el total poblacional, dada una cota B del error muestral, es la siguiente.

$$n = \frac{Ns^2}{\left(\frac{B^2}{4N}\right) + s^2} \tag{22.13}$$

En el ejemplo de arriba se quería estimar el salario medio inicial con una cota del error muestral $B = 500$. Suponga que también se desea estimar el salario total de los 5 000 egresados y que la cota sea \$2 millones. Con la ecuación (22.13) con $B = 2\,000\,000$ se obtendrá el tamaño de muestra necesario para obtener esa cota para el total poblacional.

*William G. Cochram, *Sampling Techniques,* 3a. ed., Wiley, 1977.

$$n = \frac{5\,000(3\,000)^2}{\dfrac{(2\,000\,000)^2}{4(5\,000)} + (3\,000)^2}$$
$$= 215.31$$

Al redondear hacia arriba se ve que para obtener un intervalo de confianza de aproximadamente 95%, cuya cota sea \$2 millones, el tamaño de la muestra deberá ser de 216. Hay que hacer notar que si en esta misma encuesta se desea tener una cota de \$500 para la media poblacional y una cota de \$2 millones para el total poblacional, será necesario usar una muestra cuyo tamaño sea, por lo menos, 216. Con este tamaño de muestra se obtendrá una cota más estrecha de lo necesario para la media poblacional y la precisión mínima necesaria para el total poblacional.

El tamaño de la muestra para una estimación de la proporción poblacional, se determina con una fórmula similar a la de la media poblacional; sustituya en la ecuación (22.12) s^2 por $\bar{p}(1 - \bar{p})$, con lo que obtiene

$$n = \frac{N\bar{p}(1 - \bar{p})}{N\left(\dfrac{B^2}{4}\right) + \bar{p}(1 - \bar{p})} \tag{22.14}$$

Para usar la ecuación (22.14) hay que especificar la cota B deseada y una estimación de $\bar{p}$. Si no se cuenta con una estimación de $\bar{p}$, se puede usar $\bar{p} = 0.5$; con este valor de $\bar{p}$ se garantiza que el intervalo de confianza que se obtenga tenga una cota del error muestral tan pequeña, por lo menos, como la deseada.

Ejercicios

Métodos

Autoexamen

1. Para obtener una muestra de $n = 50$ de una población de $N = 800$, se empleó el muestreo aleatorio simple. Se halló una media muestral $\bar{x} = 215$ y una desviación estándar muestral $s = 20$.
   a. Estime la media poblacional.
   b. Estime el error estándar de la media.
   c. Obtenga un intervalo de confianza de aproximadamente 95% para la media poblacional.
2. Para obtener una muestra de $n = 80$ de una población de $N = 400$, se empleó el muestreo aleatorio simple. Se halló una media muestral $\bar{x} = 75$ y una desviación estándar muestral $s = 8$.
   a. Estime el total poblacional.
   b. Estime el error estándar del total poblacional.
   c. Obtenga un intervalo de confianza de aproximadamente 95% para el total poblacional.
3. Para obtener una muestra de $n = 100$ de una población de $N = 1\,000$, se empleó el muestreo aleatorio simple. Se halló una proporción muestral de $\bar{p} = 0.30$.
   a. Estime la proporción poblacional.
   b. Estime el error estándar de la proporción.
   c. Obtenga un intervalo de confianza de aproximadamente 95% para la proporción poblacional.
4. Se va a tomar una muestra para obtener un intervalo de confianza de aproximadamente 95% para estimar la media poblacional. La población consta de 450 elementos y en un estudio piloto se encontró $s = 70$. ¿De qué tamaño deberá ser la muestra para que la amplitud del intervalo sea 30?

Aplicaciones

Autoexamen

5. En 1996 la Small Business Administration (SBA) concedió 771 créditos a pequeñas empresas en Carolina del Norte (*The Wall Street Journal Almanac,* 1998). Suponga que en una muestra de 50 pequeñas empresas el promedio de los créditos fue de $149 670 y la desviación estándar de $73 420 y que 18 de las empresas de la muestra hayan sido empresas de fabricación.
   a. Obtenga un intervalo de confianza de aproximadamente 95% para la media de los créditos.
   b. Estime un intervalo de confianza de aproximadamente 95% para el valor total de los 771 créditos en Carolina del Norte.
   c. Obtenga un intervalo de confianza de aproximadamente 95% para la proporción de créditos otorgados a empresas de fabricación.
6. En un condado de California se tienen 724 declaraciones de impuestos corporativos. El ingreso anual medio reportado es de $161 220 con una desviación estándar de $31 300. ¿De qué tamaño deberá ser la muestra el siguiente año para obtener un intervalo de confianza de aproximadamente 95% para el ingreso corporativo anual medio? La precisión deseada es de una amplitud de intervalo no mayor que $5 000.

22.5 Muestreo aleatorio simple estratificado

En el **muestreo aleatorio simple estratificado**, primero se divide la población en H grupos, a los que se llama estratos. A continuación de cada estrato h se toma una muestra aleatoria simple de tamaño n_h. Los datos de las H muestras aleatorias simples se juntan para obtener una estimación del parámetro poblacional de interés como de la media, del total o de la proporción poblacionales.

Si la variabilidad dentro de cada estrato es menor que la variabilidad entre los estratos, la precisión que se obtiene con una muestra aleatoria simple estratificada puede ser muy buena (intervalos de confianza estrechos para los parámetros poblacionales). La base para la formación de los estratos depende del criterio de quien diseña la muestra. De acuerdo con la aplicación, una población puede estratificarse por departamentos, por ubicación, según la edad, el tipo de producto, el tipo de industria, la cantidad de ventas, etcétera.

Por ejemplo, suponga que el College of Business del Lakeland College desea realizar una encuesta a los egresados ese año para conocer sus salarios iniciales. En este College hay cinco áreas principales: contaduría, finanzas, sistemas de la información, marketing y administración de operaciones. De los $N = 1\,500$ estudiantes egresados ese año, $N_1 = 500$ pertenecieron a contaduría, $N_2 = 350$ a finanzas, $N_3 = 200$ a sistemas de la información, $N_4 = 300$ a marketing y $N_5 = 150$ a administración de operaciones. Los datos de análisis previos sugieren que existe mayor variabilidad entre los salarios iniciales de las distintas áreas que dentro de cada área. Por tanto, se toma una muestra aleatoria simple estratificada de $n = 180$ estudiantes; 45 de los 180 estudiantes pertenecen a contaduría ($n_1 = 45$), 40 a finanzas ($n_2 = 40$), 30 a sistemas de la información ($n_3 = 30$), 35 a marketing ($n_4 = 35$) y 30 a administración de operaciones ($n_5 = 30$).

Media poblacional

En los muestreos estratificados para obtener una estimación insesgada de la media poblacional se calcula el promedio ponderado de las medias muestrales de los estratos. Para ponderar se usa la proporción de la población que representa cada estrato. El estimador puntual que se obtiene de esta manera y que se denota $\bar{x}_{\text{st}}$, está definido como sigue.

ESTIMADOR PUNTUAL DE LA MEDIA POBLACIONAL

$$\bar{x}_{\text{st}} = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right) \bar{x}_h \tag{22.15}$$

donde

H = número de estratos
$\bar{x}_h$ = media muestral del estrato h
N_h = número de elementos en el estrato h
N = número total de elementos en la población; $N = N_1 + N_2 + \cdots + N_H$

En el muestreo aleatorio simple estratificado, la fórmula para obtener una estimación del error estándar de la media es función de s_h, la desviación estándar muestral del estrato h.

$$s_{\bar{x}_{st}} = \sqrt{\frac{1}{N^2}\sum_{h=1}^{H} N_h(N_h - n_h)\frac{s_h^2}{n_h}} \tag{22.16}$$

Con esta fórmula el intervalo de confianza de aproximadamente 95% para estimar la media poblacional está dado por la siguiente expresión.

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA ESTIMAR LA MEDIA POBLACIONAL

$$\bar{x}_{st} \pm 2s_{\bar{x}_{st}} \tag{22.17}$$

Suponga que en la encuesta realizada a los 180 egresados del College Business del Lakeland College se obtuvieron los resultados muestrales que se presentan en la tabla 22.1. Las medias muestrales en cada área o estrato son \$35 000 para contaduría, \$33 500 para finanzas, \$41 500 para sistemas de la información, \$32 000 para marketing y \$36 000 para administración de operaciones. Con estos resultados y la ecuación (22.15) se obtiene una estimación puntual para la media poblacional.

$$\begin{aligned}\bar{x}_{st} &= \left(\frac{500}{1\,500}\right)(35\,000) + \left(\frac{350}{1\,500}\right)(33\,500) + \left(\frac{200}{1\,500}\right)(41\,500)\\ &\quad + \left(\frac{300}{1\,500}\right)(32\,000) + \left(\frac{150}{1\,500}\right)(36\,000) = 35\,017\end{aligned}$$

Los cálculos necesarios para estimar el error estándar se presentan en la tabla 22.2; observe que

$$\sum_{h=1}^{5} N_h(N_h - n_h)\frac{s_h^2}{n_h} = 42\,909\,037\,698$$

TABLA 22.1 ENCUESTA MUESTRAL SOBRE LOS SALARIOS INICIALES DE LOS EGRESADOS DEL LAKELAND COLLEGE

Área (h)	$\bar{x}_h$	s_h	N_h	n_h
Contaduría	\$35 000	2000	500	45
Finanzas	\$33 500	1700	350	40
Sistemas de la información	\$41 500	2300	200	30
Marketing	\$32 000	1600	300	35
Administración de operaciones	\$36 000	2250	150	30

TABLA 22.2 CÁLCULOS PARCIALES PARA LA ESTIMACIÓN DEL ERROR ESTÁNDAR DE LA MEDIA EN LA ENCUESTA MUESTRAL DEL LAKELAND COLLEGE

Área	h	$N_h(N_h - n_h)\dfrac{s_h^2}{n_h}$
Contaduría	1	$500(500 - 45)\dfrac{(2\,000)^2}{45} = 20\,222\,222\,222$
Finanzas	2	$350(350 - 40)\dfrac{(1\,700)^2}{40} = 7\,839\,125\,000$
Sistemas de la información	3	$200(200 - 30)\dfrac{(2\,300)^2}{30} = 5\,995\,333\,333$
Marketing	4	$300(300 - 35)\dfrac{(1\,600)^2}{35} = 5\,814\,857\,143$
Administración de operaciones	5	$150(150 - 30)\dfrac{(2\,250)^2}{30} = 3\,037\,500\,000$
		$42\,909\,037\,698$

$\sum_{h=1}^{5} N_h(N_h - n_h)\dfrac{s_h^2}{n_h}$

Por tanto,

$$s_{\bar{x}_{st}} = \sqrt{\left(\frac{1}{(1\,500)^2}\right)(42\,909\,037\,698)} = \sqrt{19\,070.68} = 138$$

De esta manera, con la ecuación (22.17), un intervalo de confianza de aproximadamente 95% para la estimación de la media poblacional es 35 017 $\pm$ 2(138) = 35.017 $\pm$ 276, es decir, \$34 741 a \$35 293.

Total poblacional

La estimación puntual del total poblacional (X) se obtiene al multiplicar N por $\bar{x}_{st}$.

ESTIMADOR PUNTUAL DEL TOTAL POBLACIONAL

$$\hat{X} = N\bar{x}_{st} \tag{22.18}$$

Una estimación del error estándar de este estimador puntual es

$$s_{\hat{X}} = Ns_{\bar{x}_{st}} \tag{22.19}$$

Por tanto, un intervalo de confianza de aproximadamente 95% es el dado por la expresión siguiente.

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA ESTIMAR EL TOTAL POBLACIONAL

$$N\bar{x}_{st} \pm 2s_{\hat{X}} \tag{22.20}$$

Ahora suponga que el College of Business del ejemplo anterior desea estimar también el ingreso total de los 1 500 egresados con objeto de estimar su impacto en la economía. Mediante la ecuación (22.18), se obtiene una estimación insesgada del total de ingresos.

$$\hat{X} = (1\,500)35\,017 = 52\,525\,500$$

Una estimación del error estándar del total poblacional se obtiene con la ecuación (22.19).

$$s_{\hat{X}} = 1\,500(138) = 207\,000$$

En conclusión, con la ecuación (22.20), se determina que un intervalo de confianza de aproximadamente 95% para estimar los ingresos totales de los 1 500 egresados es 52 525 500 $\pm$ 2(207 000) = 52 525 500 $\pm$ 414 000, es decir \$52 111 500 a \$52 939 500.

Proporción poblacional

Una estimación insesgada de la proporción poblacional, p, cuando se emplea el muestreo aleatorio simple estratificado, es un promedio ponderado de las proporciones de cada estrato. Para ponderar se usan las fracciones de la población que corresponden a cada estrato. El estimador puntual que se obtiene, denotado $\bar{p}_{\text{st}}$, se define como sigue.

ESTIMADOR PUNTUAL DE LA PROPORCIÓN POBLACIONAL

$$\bar{p}_{\text{st}} = \sum_{h=1}^{H} \left(\frac{N_h}{N}\right) \bar{p}_h \tag{22.21}$$

donde

H = número de estratos
$\bar{p}_h$ = proporción muestral del estrato h
N_h = número de elementos de la población que pertenecen al estrato h
N = número total de elementos en la población: $N = N_1 + N_2 + \cdots + N_H$

Una estimación del error estándar de $\bar{p}_{\text{st}}$ es la dada por

$$s_{\bar{p}_{\text{st}}} = \sqrt{\frac{1}{N^2} \sum_{h=1}^{H} N_h(N_h - n_h)\left[\frac{\bar{p}_h(1 - \bar{p}_h)}{n_h - 1}\right]} \tag{22.22}$$

Por tanto, un intervalo de confianza de aproximadamente 95% para estimar la proporción poblacional es el dado por la expresión siguiente.

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA ESTIMAR LA PROPORCIÓN POBLACIONAL

$$\bar{p}_{\text{st}} \pm 2 s_{\bar{p}_{\text{st}}} \tag{22.23}$$

Suponga que en este ejemplo se desea conocer la proporción de recién egresados que obtuvo un salario inicial de \$36 000 o más. En los resultados de la encuesta muestral de los 180 recién egresados se observa que 63 de ellos obtuvieron un salario inicial de \$36 000 o más y que 16 de los

63 corresponden a contaduría, 3 a finanzas, 29 a sistemas de la información, 0 a marketing y 15 a administración de operaciones.

Con la ecuación (22.21) se calcula la proporción de egresados que obtuvo un salario inicial de $36 000 o más.

$$\begin{aligned}\bar{p}_{\text{st}} &= \left(\frac{500}{1\,500}\right)\left(\frac{16}{45}\right) + \left(\frac{350}{1\,500}\right)\left(\frac{3}{40}\right) + \left(\frac{200}{1\,500}\right)\left(\frac{29}{30}\right) + \left(\frac{300}{1\,500}\right)\left(\frac{0}{35}\right) + \left(\frac{150}{1\,500}\right)\left(\frac{15}{30}\right)\\ &= 0.3149\end{aligned}$$

Los cálculos necesarios para estimar el error estándar se muestran en la tabla 22.3; observe que

$$\sum_{h=1}^{5} N_h(N_h - n_h)\left[\frac{\bar{p}_h(1-\bar{p}_h)}{n_h - 1}\right] = 1570.6913$$

Por tanto,

$$\begin{aligned}s_{\bar{p}_{\text{st}}} &= \sqrt{\frac{1}{(1\,500)^2}(1570.6913)}\\ &= 0.0264\end{aligned}$$

Con la expresión (22.23), se encuentra que un intervalo de confianza de aproximadamente 95% para la proporción de egresados que tienen un salario inicial de $36 000 o más es 0.3149 ± 2(0.0264) = 0.3149 ± 0.0528, es decir 0.2621 a 0.3677.

Determinación del tamaño de la muestra

Cuando se emplea un muestreo aleatorio simple estratificado, la elección del tamaño de la muestra se entiende como un proceso de dos pasos. Primer paso, se elige un tamaño total n para la muestra. Segundo paso, se decide cuántas unidades muestrales tomar de cada estrato. Otra alternativa es decidir primero de qué tamaño se tomará la muestra de cada estrato y después sumar

TABLA 22.3 CÁLCULOS PARCIALES PARA LA ESTIMACIÓN DEL ERROR ESTÁNDAR DE $\bar{p}_{\text{st}}$ EN LA ENCUESTA MUESTRAL DE LOS ESTUDIANTES DEL LAKELAND

Área	h	$N_h(N_h - n_h)\left[\frac{\bar{p}_h(1-\bar{p}_h)}{n_h - 1}\right]$
Contaduría	1	$500(500 - 45)\left[\frac{(16/45)(29/45)}{45 - 1}\right] = 1\,184.7363$
Finanzas	2	$350(350 - 40)\left[\frac{(3/40)(37/40)}{40 - 1}\right] = 193.0048$
Sistemas de la información	3	$200(200 - 30)\left[\frac{(29/30)(1/30)}{30 - 1}\right] = 37.7778$
Marketing	4	$300(300 - 35)\left[\frac{(0/35)(35/35)}{35 - 1}\right] = 0.0000$
Administración de operaciones	5	$150(150 - 30)\left[\frac{(15/30)(15/30)}{30 - 1}\right] = 155.1724$
	$\sum_{h=1}^{5} N_h(N_h - n_h)\left[\frac{\bar{p}_h(1-\bar{p}_h)}{n_h - 1}\right]$	1 570.6913

los tamaños muestrales de los estratos para obtener el tamaño total de la muestra. Con frecuencia interesa obtener estimaciones de la media, el total y la proporción de cada estrato; por tanto se suele emplear una combinación de estos dos métodos. Se determina un tamaño general n de la muestra y una asignación con la que se obtenga la precisión necesaria para los parámetros poblacionales generales de interés. Después, si los tamaños de las muestras de algunos de los estratos no son lo suficientemente grandes para obtener la precisión necesaria en las estimaciones para el estrato, se aumentan los tamaños de las muestras de esos estratos. En esta subsección se verán algunos de los aspectos relacionados con la asignación de toda la muestra a los diferentes estratos y se presentará un método para elegir el tamaño total de la muestra y hacer la asignación.

La asignación consiste en decidir qué fracción del total de la muestra le será asignada a cada estrato. Esta fracción determina cuán grande será la muestra aleatoria simple de cada estrato. Los factores más importantes para hacer la asignación son los siguientes:

1. El número de elementos en cada estrato.
2. La varianza de los elementos dentro de cada estrato.
3. El costo de la selección de los elementos de cada estrato.

En general, se asignan muestras más grandes a los estratos más grandes y a los estratos que tienen mayor varianza. En sentido inverso, para obtener la mayor información por un costo dado, a los estratos que tienen un costo por unidad muestreada mayor se les asignan muestras más pequeñas.

Las varianzas de cada uno de los estratos suelen ser muy diferentes. Por ejemplo, suponga que en un determinado estudio se desee determinar la cantidad media de empleados por edificio; como la variabilidad será mayor en un estrato que tenga edificios grandes que en un estrato que tenga edificios pequeños, en tales estratos se tomará una muestra proporcionalmente mayor. El costo de la selección puede ser una consideración importante cuando el entrevistador tiene que hacer recorridos significativos entre las unidades muestrales de unos estratos, pero no entre las de otros; esta situación suele surgir cuando algunos de los estratos comprenden zonas rurales y otras ciudades.

En muchas encuestas el costo por unidad de muestreo es aproximadamente el mismo en todos los estratos (por ejemplo, en las encuestas por correo o por teléfono); en tales casos el costo del muestreo puede ignorarse al hacer la asignación. Aquí se presentan las fórmulas apropiadas para elegir el tamaño de la muestra y para hacer la asignación en tales casos. En los libros más avanzados sobre muestreo se proporcionan las fórmulas para el caso en el que los costos de muestreo varían significativamente entre los estratos. Las fórmulas que se presentan en esta sección minimizan el costo total de muestreo dado un nivel de precisión. Este método conocido como *asignación de Neyman* asigna el total de la muestra n a los diversos estratos como sigue.

$$n_h = n\left(\frac{N_h s_h}{\sum_{h=1}^{H} N_h s_h}\right) \tag{22.24}$$

La ecuación (22.24) indica que el número de unidades asignadas a un estrato aumenta con el tamaño del estrato y la desviación estándar. Observe que para hacer esta asignación primero se necesita determinar el tamaño total n de la muestra. Dada una determinada precisión B, para elegir el tamaño de la muestra para la determinación de la media poblacional y del total poblacional se emplean las fórmulas siguientes.

TAMAÑO DE LA MUESTRA PARA ESTIMAR LA MEDIA POBLACIONAL

$$n = \frac{\left(\sum_{h=1}^{H} N_h s_h\right)^2}{N^2\left(\frac{B^2}{4}\right) + \sum_{h=1}^{H} N_h s_h^2} \tag{22.25}$$

TAMAÑO DE LA MUESTRA PARA ESTIMAR EL TOTAL POBLACIONAL

$$n = \frac{\left(\sum_{h=1}^{H} N_h s_h\right)^2}{\frac{B^2}{4} + \sum_{h=1}^{H} N_h s_h^2} \tag{22.26}$$

Por ejemplo, suponga que un vendedor de Chevrolet desea hacer una encuesta a los clientes que compran un Corvette o un Cavalier o un Geo Prizm para obtener información que él considera puede ser útil para la publicidad futura. Suponga que este vendedor desea estimar el ingreso medio mensual de estos clientes y que la cota del error muestral sea \$100. Los 600 clientes de este vendedor se dividen en tres estratos: 100 que poseen un Corvette, 200 que poseen un Geo Prizm y 300 que poseen un Cavalier. Para estimar la desviación estándar de cada estrato se empleó una encuesta piloto; los resultados son $s_1 = \$1\,300$, $s_2 = \$900$ y $s_3 = \$500$ para los poseedores de un Corvette, un Geo Prizm y un Cavalier, respectivamente.

El primer paso para elegir el tamaño de la muestra es usar la ecuación (22.25) para determinar el tamaño total de la muestra que se necesita para tener una cota $B = \$100$ en la estimación de la media poblacional. Primero, se calcula

$$\sum_{h=1}^{3} N_h s_h = 100(1\,300) + 200(900) + 300(500) = 460\,000$$

Después se calcula

$$\sum_{h=1}^{3} N_h s_h^2 = 100(1\,300)^2 + 200(900)^2 + 300(500)^2 = 406\,000\,000$$

Al sustituir estos valores en la ecuación (22.25), se obtiene el tamaño total de la muestra que se requiere para que la cota del error muestral sea $B = \$100$.

$$n = \frac{(460\,000)^2}{\frac{(600)^2(100)^2}{4} + 406\,000\,000} = 162$$

Por tanto, el tamaño total de la muestra para obtener la precisión deseada es 162. Para la asignación del total de la muestra a los tres estratos se emplea la ecuación (22.24).

$$n_1 = 162\left(\frac{100(1\,300)}{460\,000}\right) = 46$$

$$n_2 = 162\left(\frac{200(900)}{460\,000}\right) = 63$$

$$n_3 = 162\left(\frac{300(500)}{460\,000}\right) = 53$$

De manera que la recomendación será: 46 propietarios de Corvette, 62 de Geo Prizm y 53 de Cavalier que hacen un tamaño total de la muestra de 162 clientes.

Para determinar el tamaño de la muestra necesario para la estimación de la proporción poblacional, simplemente se sustituye en la ecuación (22.25) s_h por $\sqrt{\bar{p}_h(1 - \bar{p}_h)}$; el resultado es

$$n = \frac{\left(\sum_{h=1}^{H} N_h \sqrt{\bar{p}_h(1 - \bar{p}_h)}\right)^2}{N^2\left(\frac{B^2}{4}\right) + \sum_{h=1}^{H} N_h \bar{p}_h(1 - \bar{p}_h)} \qquad \textbf{(22.27)}$$

Una vez que se ha determinado el tamaño total de la muestra necesario para la estimación de la proporción poblacional, la asignación a los varios estratos se hace otra vez con la ecuación (22.25) al sustituir s_h por $\sqrt{\bar{p}_h(1 - \bar{p}_h)}$.

NOTAS Y COMENTARIOS

1. Una ventaja del muestreo aleatorio simple estratificado es que automáticamente, como consecuencia del procedimiento de muestreo, se obtienen estimaciones de los parámetros poblacionales de cada estrato. Por ejemplo, además de obtener una estimación del salario inicial promedio en el problema de los egresados de administración, se obtuvo también una estimación del salario inicial promedio de los egresados de cada área. Como cada una de las estimaciones de los salarios iniciales se hizo con base en una muestra aleatoria simple de cada estrato, se puede emplear el procedimiento para obtener un intervalo de confianza cuando se toma una muestra aleatoria simple (véase la ecuación (22.4)) para calcular una estimación de la media de cada estrato mediante un intervalo de confianza de aproximadamente 95% para estimar la media de cada estrato. De manera similar, se obtienen intervalos de confianza para estimar el total poblacional y la proporción poblacional de cada estrato con las ecuaciones (22.8) y (22.10), respectivamente.

2. Otro tipo de asignación que se usa cuando se hace un muestreo aleatorio simple estratificado es la *asignación proporcional.* Con este método el tamaño de la muestra asignada a cada estrato está dado por la fórmula siguiente.

$$n_h = n\left(\frac{N_h}{N}\right) \qquad \textbf{(22.28)}$$

La asignación proporcional debe usarse cuando las varianzas de los estratos son todas aproximadamente iguales y el costo por unidad muestreada es casi el mismo en todos los estratos. En los casos en que las varianzas de los estratos son iguales, la asignación proporcional y el procedimiento de Neyman dan las mismas asignaciones.

Ejercicios

Métodos

7. Los resultados obtenidos de una muestra aleatoria simple estratificada fueron los siguientes.

Auto examen

Estrato (h)	$\bar{x}_h$	s_h	$\bar{p}_h$	N_h	n_h
1	138	30	0.50	200	20
2	103	25	0.78	250	30
3	210	50	0.21	100	25

a. Proporcione una estimación de la media poblacional de cada estrato.
b. Dé un intervalo de confianza de aproximadamente 95% para la media poblacional de cada estrato.
c. Encuentre un intervalo de confianza de aproximadamente 95% para la media poblacional de toda la población.

8. Reconsidere los resultados muestrales del ejercicio 7.
   a. Encuentre una estimación del total poblacional de cada estrato.
   b. Dé una estimación puntual del total de los 550 elementos de la población.
   c. Proporcione un intervalo de confianza de aproximadamente 95% para el total poblacional.
9. Regrese a los resultados muestrales del ejercicio 7.
   a. Proporcione un intervalo de confianza de aproximadamente 95% para la proporción de cada estrato.
   b. Dé una estimación puntual de la proporción poblacional para los 550 elementos de la población.
   c. Estime el error estándar de la proporción poblacional.
   d. Encuentre un intervalo de aproximadamente 95% para la proporción poblacional.
10. Una población fue dividida en tres estratos $N_1 = 300$, $N_2 = 600$ y $N_3 = 500$. De una encuesta anterior se tienen las estimaciones siguientes de la desviación estándar en cada uno de los estratos: $s_1 = 150$, $s_2 = 75$ y $s_3 = 100$.
    a. Suponga que se necesita una estimación de la media poblacional con una cota del error de estimación de $B = 20$. ¿De qué tamaño deberá ser la muestra? ¿Cuántos elementos deberán ser tomados de cada estrato?
    b. Admita que se requiere que la cota sea $B = 10$ ¿De qué tamaño deberá ser la muestra? ¿Cuántos elementos deberán ser tomados de cada estrato?
    c. Suponga que se quiere tener una estimación del total poblacional y que la cota sea $B =$ 15 000. ¿De qué tamaño deberá ser la muestra? ¿Cuántos elementos deberán ser tomados de cada estrato?

Aplicaciones

11. Una cadena de farmacias tiene tiendas en cuatro ciudades: 38 tiendas en Indianápolis, 45 en Louisville, 80 en St. Louis y 70 en Memphis. Las ventas en las cuatro ciudades varían considerablemente debido a la competencia. De una encuesta muestral se tienen los datos siguientes (en miles de dólares). Cada una de las ciudades se consideró como un estrato y se tomó una muestra aleatoria simple estratificada.

Indianápolis	Louisville	St. Louis	Memphis
50.3	48.7	16.7	14.7
41.2	59.8	38.4	88.3
15.7	28.9	51.6	94.2
22.5	36.5	42.7	76.8
26.7	89.8	45.0	35.1
20.8	96.0	59.7	48.2
	77.2	80.0	57.9
	81.3	27.6	18.8
			22.0
			74.3

    a. Estime la media de las ventas en cada ciudad (estrato).
    b. Obtenga un intervalo de confianza de aproximadamente 95% para la media de las ventas en cada ciudad.
    c. Estime la proporción de farmacias cuyas ventas son de $50 000 o más.
    d. Obtenga un intervalo de confianza de aproximadamente 95% para la proporción de farmacias cuyas ventas son de $50 000 o más.
12. Reconsidere los resultados de la encuesta muestral del ejercicio 11.
    a. Estime el total poblacional de las ventas en St. Louis.
    b. Valore el total poblacional de las ventas en Indianápolis.
    c. Obtenga un intervalo de confianza de aproximadamente 95% para la media de las ventas de la cadena de farmacias.

d. Encuentre un intervalo de confianza de aproximadamente 95% para el total de las ventas de la cadena de farmacias.

13. Una empresa de contadores tiene clientes en la industria bancaria, en la de seguros y en la de corretaje: $N_1 = 50$ bancos, $N_2 = 38$ empresas de seguros y $N_2 = 35$ empresas de corretaje. Esta empresa ha contratado a una empresa que se dedica al marketing para que realice una encuesta entre sus clientes en estas tres industrias. En la encuesta se harán diversas preguntas relacionadas, tanto con el negocio de los clientes como con su satisfacción con el servicio que reciben de la empresa de contadores. Suponga que se desea un intervalo de confianza de aproximadamente 95% para estimar el número promedio de empleados en los 123 clientes con una cota en el error de estimación $B = 30$.
    a. Suponga que en un estudio piloto se encuentra $s_1 = 80$, $s_2 = 150$ y $s_3 = 45$. Determine el tamaño total de la muestra y explique los tamaños de las muestras asignados a los tres estratos.
    b. Suponga que se duda del estudio piloto y que para determinar el tamaño de la muestra se decide suponer que todos los estratos tienen la misma desviación estándar de 100. Determine el tamaño de la muestra y diga cuántos elementos deben tomarse de cada estrato.

22.6 Muestreo por conglomerados

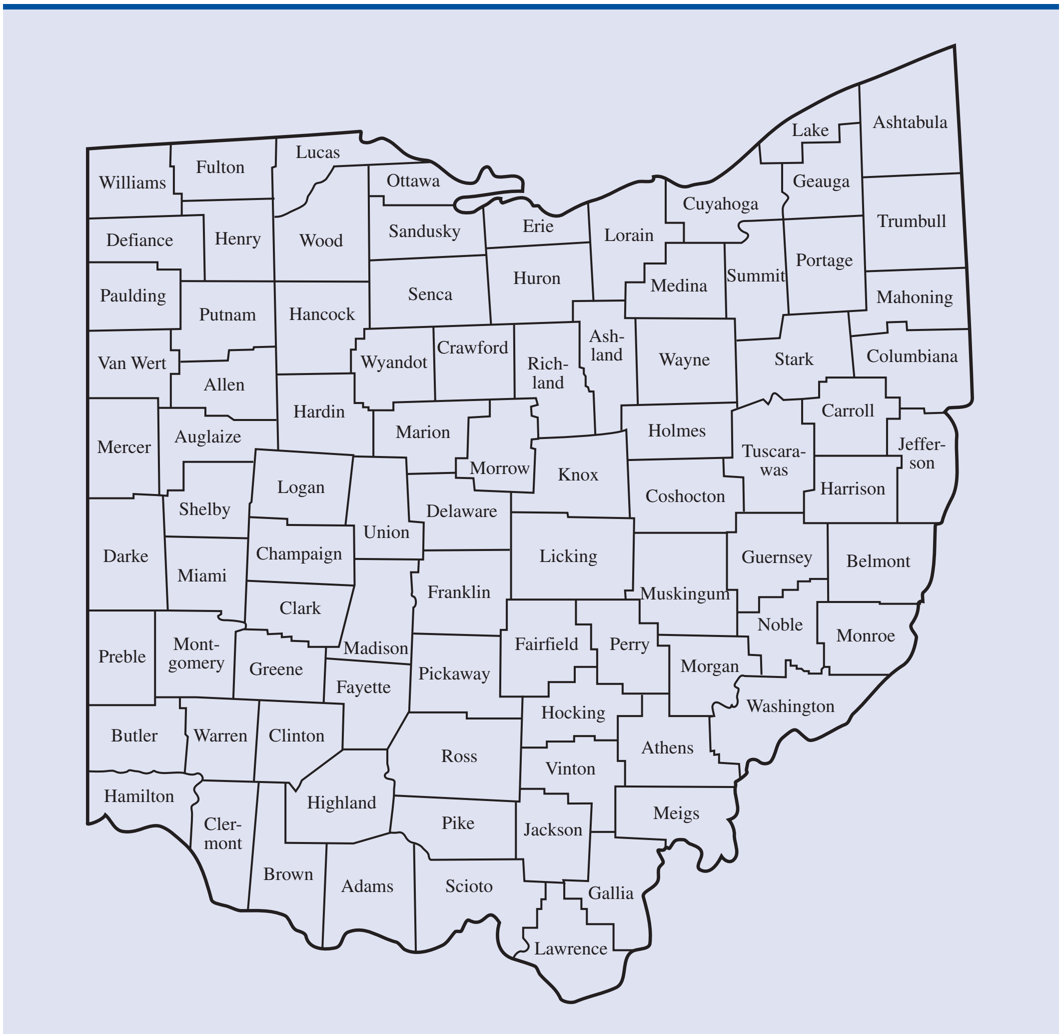
En el **muestreo por conglomerados** se requiere que la población se divida en N grupos de elementos llamados conglomerados, de manera que cada elemento de la población pertenezca a uno y sólo un conglomerado. Por ejemplo, suponga que se quiere investigar a los votantes registrados del estado de Ohio. Un método de muestreo puede ser elaborar un marco con todos los votantes registrados del estado de Ohio y, de este marco, tomar una muestra aleatoria simple. De manera alternativa, para el muestreo por aglomerados se define el marco como la lista de los $N = 88$ condados del estado (véase figura 22.1). Con este método, cada condado o conglomerado consiste en un grupo de votantes registrados y cada votante registrado del estado pertenece a uno y sólo un conglomerado.

Suponga que de los 88 condados se toma una muestra aleatoria simple de $n = 12$ condados. Ahora se pueden recolectar los datos de *todos* los votantes registrados de cada uno de los 12 conglomerados, método que se conoce como *muestreo por conglomerados en una sola etapa* o se puede tomar una muestra aleatoria simple de los votantes registrados en cada uno de los 12 conglomerados muestreados, método que se conoce como *muestreo por conglomerados en dos etapas*. En cualquier caso existen fórmulas para usar los resultados muestrales en la obtención de estimaciones puntuales o por intervalo de estimación de parámetros poblacionales como la media, el total o la proporción poblacionales. En este capítulo se considerará únicamente el muestreo por conglomerados en una sola etapa; en libros más avanzados sobre muestreo se encuentra la información sobre el muestreo por conglomerados en dos etapas.

El muestreo estratificado y el muestreo por conglomerados son similares, ya que en los dos se divide a la población en grupos de elementos. Sin embargo, las razones para elegir el muestreo por conglomerados son diferentes de las razones para elegir el muestreo estratificado. Con el muestreo por conglomerados se obtienen mejores resultados cuando los elementos dentro de los conglomerados son heterogéneos (no son parecidos). Lo ideal es que los conglomerados sean una versión a pequeña escala de la población. Cuando es así, al muestrear una pequeña cantidad de conglomerados se obtiene una buena información de las características de toda la población.

Una de las principales aplicaciones del muestreo por conglomerados es el muestreo de áreas, en donde los conglomerados son condados, poblaciones, manzanas u otras secciones geográficamente bien definidas de la población. Como sólo se recolectan datos de una muestra o conglomerado de toda un área geográfica y como los elementos dentro de un conglomerado se suelen encontrar cerca, uno de otros, cuando un recolector de datos o entrevistador se envía a una unidad muestreada se obtiene un considerable ahorro de tiempo y costos. Por tanto, aun cuando se requiera que el tamaño total de la muestra sea grande, el muestreo por conglomerados puede resultar menos costoso que el muestreo aleatorio simple o el muestreo aleatorio simple estratificado. Además,

FIGURA 22.1 CONDADOS DEL ESTADO DE OHIO USADOS COMO CONGLOMERADOS DE VOTANTES REGISTRADOS



No es necesario hacer una lista de todos los elementos de la población. En el muestreo por conglomerados sólo se necesita una lista de los elementos en los conglomerados muestreados.

el muestreo por conglomerados minimiza el tiempo y el costo de la elaboración de un marco o lista de los elementos a ser muestreados, dado que en el muestreo por conglomerados no se necesita una lista de todos los elementos de la población. Sólo se necesita una lista de los elementos de los conglomerados muestreados.

Con el fin de ilustrar el muestreo por conglomerados se verá una encuesta realizada por la CPA Society (Certified Public Account Society) de los 12 000 integrantes en servicio en un determinado estado. Como parte de la encuesta, la CPA Society recolectó información sobre el ingreso, el género y sobre factores relacionados con el estilo de vida de los contadores. Dado que para obtener toda la información deseada era necesario realizar entrevistas, para minimizar los gastos de desplazamiento y de las entrevistas, la CPA Society realizó un muestreo por conglomerados. El marco consistió en todas las empresas con contadores registradas en el estado. Suponga que el número de empresas registradas en el estado haya sido $N = 1\,000$ y que se haya tomado una muestra aleatoria simple de $n = 10$ empresas.

Para las fórmulas que se requieren en el muestreo por conglomerados para la obtención de intervalos de confianza de aproximadamente 95% para estimar la media, el total o la proporción poblacionales, se usará la notación siguiente.

N = número de conglomerados en la población
n = número de conglomerados tomados para la muestra
M_i = número de elementos en el conglomerado i
M = número de elementos en la población; $M = M_1 + M_2 + \cdots + M_N$
$\bar{M} = M/N$ = número promedio de elementos en un conglomerado
x_i = número total de observaciones en el conglomerado i
a_i = número de observaciones con una determinada característica en el conglomerado i

En el caso de la encuesta muestral de la CPA Society, la información que se tiene es la siguiente.

$$N = 1\,000$$
$$n = 10$$
$$M = 12\,000$$
$$\bar{M} = 12\,000/1\,000 = 12$$

En la tabla 22.4 se dan los valores de M_{i} y x_i correspondientes a cada uno de los conglomerados muestreados, así como el número de mujeres de la CPA Society en las empresas muestreadas.

Media poblacional

El estimador puntual de la media poblacional en un muestreo por conglomerados está dado por la fórmula siguiente.

ESTIMADOR PUNTUAL DE LA MEDIA POBLACIONAL

$$\bar{x}_{\text{c}} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} M_i} \tag{22.29}$$

TABLA 22.4 RESULTADOS DE LA ENCUESTA MUESTRAL SOBRE LOS CPA

Empresa (i)	CPA (M_i)	Salario total (miles de $) en la empresa i (x_i)	Mujeres CPA (a_i)
1	8	384	2
2	25	1350	8
3	4	148	0
4	17	857	6
5	7	296	1
6	3	131	2
7	15	761	2
8	4	176	0
9	12	577	5
10	33	1880	9
Totales	128	6560	35

Una estimación del error estándar de este estimador puntual es

$$s_{\bar{x}_c} = \sqrt{\left(\frac{N - n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}(x_i - \bar{x}_c M_i)^2}{n - 1}} \qquad \textbf{(22.30)}$$

Por tanto, con la expresión siguiente se obtiene un intervalo de aproximadamente 95% como estimación de la media poblacional.

INTERVALO DE APROXIMADAMENTE 95% COMO ESTIMACIÓN DE LA MEDIA POBLACIONAL

$$\bar{x}_c \pm 2s_{\bar{x}_c} \qquad \textbf{(22.31)}$$

Al emplear los datos de la tabla 22.4 se obtiene una estimación del salario medio de los contadores públicos certificados.

$$\bar{x}_c = \frac{6\,560}{128} = 51.250$$

Los datos de la tabla 22.4, dados en miles de dólares, indican que una estimación del salario medio de los contadores públicos certificados del estado es \$51 250.

En la tabla 22.5 se presenta parte de los cálculos que se necesitan para estimar el error estándar, observe que

$$\sum_{i=1}^{n}(x_i - \bar{x}_c M_i)^2 = 51{,}281.378$$

Por tanto,

$$s_{\bar{x}_c} = \sqrt{\left[\frac{1\,000 - 10}{(1\,000)(10)(12)^2}\right]\frac{51\,281.378}{10 - 1}} = 1.979$$

Por tanto, el error estándar es \$1 979. Con la expresión (22.31) se encuentra que un intervalo de confianza de aproximadamente 95% para la estimación del salario anual medio es 51 250 $\pm$ 2(1 979) = 51 250 $\pm$ 3 958, es decir, \$47 292 a \$55 208.

Total poblacional

El estimador puntual para el total poblacional se obtiene multiplicando M por $\bar{x}_c$.

ESTIMADOR PUNTUAL PARA EL TOTAL POBLACIONAL

$$\hat{X} = M\bar{x}_c \qquad \textbf{(22.32)}$$

Una estimación del error estándar de este estimador es

$$s_{\hat{X}} = Ms_{\bar{x}_c} \qquad \textbf{(22.33)}$$

TABLA 22.5 CÁLCULOS PARCIALES PARA LA ESTIMACIÓN DEL ERROR ESTÁNDAR DE LA MEDIA EN LA ENCUESTA MUESTRAL DE CPA

Empresa (i)	M_i	x_i	$(x_i - 51.250M_i)^2$
1	8	384	$[384 - 51.250(8)]^2 =$ 676.000
2	25	1350	$[1350 - 51.250(25)]^2 =$ 4 726.563
3	4	148	$[148 - 51.250(4)]^2 =$ 3 249.000
4	17	857	$[857 - 51.250(17)]^2 =$ 203.063
5	7	296	$[296 - 51.250(7)]^2 =$ 3 937.563
6	3	131	$[131 - 51.250(3)]^2 =$ 517.563
7	15	761	$[761 - 51.250(15)]^2 =$ 60.063
8	4	176	$[176 - 51.250(4)]^2 =$ 841.000
9	12	577	$[577 - 51.250(12)]^2 =$ 1 444.000
10	33	1880	$[1880 - 51.250(33)]^2 =$ 35 626.563
Totales	128	6560	51 281.378

$\sum_{i=1}^{n}(x_i - \bar{x}_c M_i)^2$ → 51 281.378

Por tanto, un intervalo de confianza de aproximadamente 95% para estimar el total poblacional está dado por la expresión siguiente.

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA EL TOTAL POBLACIONAL

$$M\bar{x}_c \pm 2s_{\hat{X}} \quad \textbf{(22.34)}$$

En la encuesta muestral de la CPA.

$$\hat{X} = M\bar{x}_c = 12\,000(51\,250) = \$615\,000\,000$$
$$s_{\hat{X}} = Ms_{\bar{x}_c} = 12\,000(1\,979) = \$23\,748\,000$$

En consecuencia, con la expresión (22.34), se encuentra que un intervalo de confianza de aproximadamente 95% es \$615 000 000 ± 2(\$23 784 000) = \$615 000 000 ± \$47 496 000, es decir, \$567 504 000 a \$662 496 000.

Proporción poblacional

A continuación se da el estimador puntual para la proporción poblacional cuando se hace un muestreo por conglomerados.

ESTIMADOR PUNTUAL PARA LA PROPORCIÓN POBLACIONAL

$$\bar{p}_c = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} M_i} \quad \textbf{(22.35)}$$

donde

a_i = número de elementos, con la característica de interés, en el conglomerado i.

Una estimación de la desviación estándar de este estimador puntual es

$$s_{\bar{p}_c} = \sqrt{\left(\frac{N-n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}(a_i - \bar{p}_c M_i)^2}{n-1}} \tag{22.36}$$

Entonces un intervalo de confianza de aproximadamente 95% para estimar la proporción poblacional es el dado por la expresión siguiente.

INTERVALO DE CONFIANZA DE APROXIMADAMENTE 95% PARA ESTIMAR LA PROPORCIÓN POBLACIONAL

$$\bar{p}_c \pm 2 s_{\bar{p}_c} \tag{22.37}$$

En el caso de la encuesta muestral de los contadores certificados con la ecuación (22.35) y los datos de la tabla 22.4 se obtiene una estimación de la proporción de contadores certificados que son mujeres.

$$\bar{p}_c = \frac{2 + 8 + \cdots + 9}{8 + 25 + \cdots + 33} = \frac{35}{128} = 0.2734$$

En la tabla 22.6 se presenta parte de los cálculos necesarios para estimar el error estándar; observe que

$$\sum_{i=1}^{n}(a_i - \bar{p}_c M_i)^2 = 15.2098$$

Por tanto,

$$s_{\bar{p}_c} = \sqrt{\left[\frac{1\,000 - 10}{(1\,000)(10)(12)^2}\right]\frac{15.2098}{10 - 1}} = 0.0341$$

De esta manera, con la expresión (22.37) se halla que un intervalo de confianza de aproximadamente 95% para la proporción de mujeres que son contadoras certificadas es 0.2734 ± 2(0.0341) = 0.2734 ± 0.0682, es decir, 0.2052 a 0.3416.

Determinación del tamaño de la muestra

Una vez formados los conglomerados, lo primero para determinar el tamaño de la muestra es elegir el número n de conglomerados. Este procedimiento en el caso del muestreo por conglomerados es parecido al procedimiento empleado con los otros métodos de muestreo. El nivel de precisión se especifica al elegir un valor para B, la cota del error muestral. Después se elabora la fórmula para obtener el valor de n que permitirá lograr la precisión deseada.

Para decidir cuántos conglomerados incluir en la muestra, los factores decisivos son el tamaño promedio de los conglomerados y la varianza entre los conglomerados. Si los conglomerados son parecidos, la varianza entre ellos será pequeña y el número de conglomerados que se muestree puede ser pequeña. Las fórmulas para determinar exactamente el tamaño de la muestra se encuentran en libros más avanzados sobre muestreo.

TABLA 22.6 PARTE DE LOS CÁLCULOS PARA LA ESTIMACIÓN DEL ERROR ESTÁNDAR DE $\bar{p}_c$ EN EL ESTUDIO MUESTRAL DE LA CPA, DONDE $\bar{p}_c = 0.2734$

Empresa (i)	M_i	a_i	$(a_i - 0.2734M_i)^2$
1	8	2	$[2 - 0.2734(8)]^2 =$ 0.0350
2	25	8	$[8 - 0.2734(25)]^2 =$ 1.3572
3	4	0	$[0 - 0.2734(4)]^2 =$ 1.1960
4	17	6	$[6 - 0.2734(17)]^2 =$ 1.8284
5	7	1	$[1 - 0.2734(7)]^2 =$ 0.8350
6	3	2	$[2 - 0.2734(3)]^2 =$ 1.3919
7	15	2	$[2 - 0.2734(15)]^2 =$ 4.4142
8	4	0	$[0 - 0.2734(4)]^2 =$ 1.1960
9	12	5	$[5 - 0.2734(12)]^2 =$ 2.9556
10	33	9	$[9 - 0.2734(33)]^2 =$ 0.0005
Totales	128	35	15.2098

$$\sum_{i=1}^{n}(a_i - \bar{p}_c M_i)^2$$

Ejercicios

Métodos

Autoexamen

14. De una población que tiene $N = 25$ conglomerados y $M = 300$ elementos se va a tomar una muestra de cuatro conglomerados. En la tabla siguiente se presentan los valores de M_i, x_i y a_i en cada conglomerado.

Conglomerado (i)	M_i	x_i	a_i
1	7	95	1
2	18	325	6
3	15	190	6
4	10	140	2
Totales	50	750	15

a. Dé estimaciones puntuales de la media, del total y de la proporción poblacionales.
b. Estime el error estándar de las estimaciones del inciso a.
c. Obtenga un intervalo de confianza de aproximadamente 95% para la media poblacional.
d. Proporcione un intervalo de confianza de aproximadamente 95% para el total poblacional.
e. Obtenga un intervalo de confianza de aproximadamente 95% para la proporción poblacional.

15. De una población que tiene $N = 30$ conglomerados y $M = 600$ elementos se va a tomar una muestra de cuatro conglomerados. En la tabla siguiente se presentan los valores de M_i, x_i y a_i en cada conglomerado

Conglomerado (i)	M_i	x_i	a_i
1	35	3 500	3
2	15	965	0
3	12	960	1
4	23	2 070	4
5	20	1 100	3
6	25	1 805	2
Totales	130	10 400	13

a. Dé estimaciones puntuales de la media, del total y de la proporción poblacionales.
b. Obtenga un intervalo de confianza de aproximadamente 95% para la media poblacional.
c. Proporcione un intervalo de confianza de aproximadamente 95% para el total poblacional.
d. Obtenga un intervalo de confianza de aproximadamente 95% para la proporción poblacional.

Aplicaciones

16. Una empresa de servicio público realiza una encuesta a los ingenieros mecánicos para conocer los factores que influyen en la elección del equipo de calefacción, ventilación y aire acondicionado (HVAC, por sus siglas en inglés) en los nuevos edificios comerciales. En el área de acción de esta empresa de servicio público hay 120 empresas relacionadas con el diseño de sistemas HVAC. La idea es hacer un muestreo por conglomerados en el que cada empresa represente un conglomerado. Todos los ingenieros mecánicos de cada empresa que se tome en la muestra serán entrevistados. Se cree que en las 120 empresas están empleados aproximadamente 500 ingenieros mecánicos. Se toma una muestra de 10 empresas. Entre otras cosas se anota la edad de cada entrevistado y si estudió en la universidad local.

Conglomerado (i)	M_i	Total de las edades de los entrevistados	Número que estudió en la universidad local
1	12	520	8
2	1	33	0
3	2	70	1
4	1	29	1
5	6	270	3
6	3	129	2
7	2	102	0
8	1	48	1
9	9	337	7
10	13	462	12
Totales	50	2000	35

a. Estime la edad promedio de los ingenieros mecánicos que trabajan en esta actividad.
b. Halle la proporción de ingenieros mecánicos en esta actividad que estudió en la universidad local.
c. Dé un intervalo de confianza de aproximadamente 95% para la edad promedio de los ingenieros mecánicos que trabajan en esta actividad.
d. Proporcione un intervalo de confianza de aproximadamente 95% para la proporción de ingenieros mecánicos en el área de acción de la empresa pública que estudiaron en la universidad local.

17. Una empresa inmobiliaria nacional acaba de adquirir una empresa más pequeña que cuenta con 150 oficinas y 6 000 agentes en Los Ángeles y en otros sitios del sur de California. La empresa nacional realizó una encuesta muestral para conocer las actitudes y otras características de sus nuevos empleados. En una muestra de ocho oficinas, todos los empleados llenaron los cuestionarios. A continuación se presentan los resultados de esta encuesta en las ocho oficinas.

Oficina	Agentes	Edad promedio	Título universitario	Agentes del sexo masculino
1	17	37	3	4
2	35	32	14	12
3	26	36	8	7
4	66	30	38	28
5	43	41	18	12
6	12	52	2	6
7	48	35	20	17
8	57	44	25	26

a. Estime la edad promedio de los agentes.
b. Estime la proporción de agentes que tienen un título universitario y la proporción de agentes del sexo masculino.
c. Dé un intervalo de confianza de aproximadamente 95% para la edad promedio de los agentes.
d. Dé un intervalo de confianza de aproximadamente 95% para la proporción de agentes que tiene un título universitario.
e. Dé un intervalo de confianza de aproximadamente 95% para la proporción de agentes del sexo masculino.

22.7 Muestreo sistemático

El **muestreo sistemático** suele emplearse como alternativa al muestreo aleatorio simple. En algunas situaciones, en especial cuando las poblaciones son grandes se lleva mucho tiempo tomar una muestra aleatoria simple en la que primero hay que hallar un número aleatorio y después contar o buscar en el marco el elemento correspondiente. En tales casos el muestreo sistemático es una alternativa al muestreo aleatorio simple. Por ejemplo, si se quiere una muestra de tamaño 50 de una población que contiene 5 000 elementos, se toma en la muestra un elemento cada 5 000/50 = 100 elementos de la población. En este caso un muestreo sistemático consistirá en elegir en forma aleatoria uno de los primeros 100 elementos del marco. Los demás elementos para la muestra se encuentran contando a partir del primer elemento tomado en la muestra hasta encontrar el centésimo elemento que sigue en el marco. En efecto, la muestra de 50 elementos se obtiene avanzando sistemáticamente a través de la población y tomando el centésimo elemento después del último elemento tomado en la muestra. De esta manera suele ser mas fácil tomar la muestra de 50 elementos de lo que sería si se usara un muestreo aleatorio simple. El primer elemento se elige aleatoriamente, lo que permite suponer que una muestra sistemática tiene las propiedades de una muestra aleatoria simple. Esta suposición suele ser correcta cuando el marco es un ordenamiento aleatorio de los elementos de la población.

Resumen

Se le presentó una breve introducción al campo de muestreo de encuestas. El propósito del muestreo de encuestas es recolectar datos con el objeto de hacer estimaciones de parámetros poblacionales como la media, el total o la proporción poblacional. El muestreo de encuestas es comparable con la realización de experimentos para generar datos. Cuando se usa el muestreo de encuestas, el diseño del plan de muestreo tiene una importancia crítica en la determinación de qué datos ya existentes se recolectarán. Cuando se emplean experimentos, el diseño experimental tiene una importancia vital en la determinación de cuáles son los datos que serán generados o creados.

En el muestreo de encuestas pueden presentarse dos tipos de errores: el error muestral y los errores no muestrales. El error muestral es el tipo de error que se presenta debido a que para estimar un parámetro poblacional se emplea una muestra y no toda la población. Los errores no muestrales se refieren a *todos* los otros tipos de errores que pueden presentarse, tales como errores de medición, del entrevistador, de falta de respuesta y errores de procesamiento. Los errores no muestrales se controlan a través del diseño del cuestionario, de la capacitación de los entrevistadores, de la verificación cuidadosa de los datos, etc. El error muestral se minimiza mediante la elección adecuada del diseño y del tamaño de la muestra.

En este capítulo se habló de cuatro diseños muestrales muy empleados: muestreo aleatorio simple, muestreo aleatorio simple estratificado, muestreo por conglomerados y muestreo sistemático. El objetivo del diseño muestral es obtener la estimación más precisa al mínimo costo. Cuando la población es dividible en estratos, de manera que los elementos dentro de cada estrato sean relativamente homogéneos, el muestreo aleatorio simple estratificado permitirá obtener mayor precisión (intervalos de confianza más estrechos) que el muestreo aleatorio simple. Cuando los elementos se pueden agrupar en conglomerados de manera que todos los elementos de un conglomerado se encuentren cercanos unos de otros geográficamente, el muestreo por conglomerados suele reducir el costo del entrevistador; en estos casos, el muestreo por conglomerados proporcionará mayor precisión a menor costo. El muestreo aleatorio sistemático se presentó como alternativa al muestreo aleatorio simple.

Glosario

Elemento Entidad de la que se recolectan los datos.
Población El conjunto de todos los elementos de interés.
Muestra Un subconjunto de la población.
Población objetivo Población acerca de la cual se quieren hacer las inferencias.
Población muestreada Población de la que se toma la muestra.
Unidades muestrales Las unidades que se toman para la muestra.
Marco Lista de las unidades muestrales para un estudio. La muestra es aleatoria para seleccionar unidades del marco.
Muestreo probabilístico Todo método de muestreo en el que se puede calcular la probabilidad que tiene cada posible muestra.
Muestreo no probabilístico Todo método de muestreo en el que no se puede calcular la probabilidad de seleccionar una determinada muestra.
Muestreo de conveniencia Un método no probabilístico de muestreo en el que los elementos se toman con base en la conveniencia.
Muestreo subjetivo Un método no probabilístico de muestreo en el que los elementos se seleccionan con base en el criterio de la persona que diseña el estudio.
Error muestral El error que se presenta debido a que se emplea una muestra y no toda la población para estimar un parámetro poblacional.
Error no muestral Todos los tipos de errores que no son un error muestral, como errores de medición, errores del entrevistador y errores de procesamiento.
Muestra aleatoria simple Muestra que se toma de tal manera que toda muestra de tamaño n tiene la misma probabilidad de ser elegida.
Cota del error muestral Número que se suma o que se resta de una estimación puntual para obtener un intervalo de confianza de aproximadamente 95%. La cota del error muestral es igual al doble del error estándar del estimador puntual.
Muestreo aleatorio simple estratificado Método probabilístico para tomar una muestra en el que, primero, se divide la población en estratos y después de cada estrato se toma una muestra aleatoria simple.
Muestreo por conglomerados Método probabilístico de muestreo en el que primero se divide la población en conglomerados y después se selecciona uno o más de los conglomerados para la muestra. En el muestreo por conglomerados en una sola etapa, se toman en la muestra todos los elementos de cada uno de los conglomerados elegidos; en el muestreo por conglomerados en dos etapas se toma una muestra de los elementos de cada uno de los conglomerados elegidos.
Muestreo sistemático Método para tomar una muestra en el que el primer elemento se toma aleatoriamente y después se toma cada k-ésimo elemento.

Fórmulas clave

Muestreo aleatorio simple

Estimación por intervalo de la media poblacional

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} \tag{22.1}$$

Estimación del error estándar de la media

$$s_{\bar{x}} = \sqrt{\frac{N-n}{N}}\left(\frac{s}{\sqrt{n}}\right) \tag{22.2}$$

Intervalo de estimación para la media poblacional

$$\bar{x} \pm t_{\alpha/2} s_{\bar{x}} \tag{22.3}$$

Intervalo de confianza de aproximadamente 95% para estimar la media poblacional

$$\bar{x} \pm 2 s_{\bar{x}} \tag{22.4}$$

Estimación puntual del total poblacional

$$\hat{X} = N\bar{x} \tag{22.5}$$

Estimación del error estándar de $\hat{X}$

$$s_{\hat{X}} = N s_{\bar{x}} \tag{22.6}$$

Intervalo de confianza de aproximadamente 95% para estimar el total poblacional

$$N\bar{x} \pm 2 s_{\hat{X}} \tag{22.8}$$

Estimación del error estándar de la proporción

$$s_{\bar{p}} = \sqrt{\left(\frac{N-n}{N}\right)\left(\frac{\bar{p}(1-\bar{p})}{n-1}\right)} \tag{22.9}$$

Intervalo de confianza de aproximadamente 95% para estimar la proporción poblacional

$$\bar{p} \pm 2 s_{\bar{p}} \tag{22.10}$$

Tamaño de la muestra en una estimación de la media poblacional

$$n = \frac{Ns^2}{N\left(\frac{B^2}{4}\right) + s^2} \tag{22.12}$$

Tamaño de la muestra en una estimación del total poblacional

$$n = \frac{Ns^2}{\left(\frac{B^2}{4N}\right) + s^2} \tag{22.13}$$

Tamaño de la muestra en una estimación de la proporción poblacional

$$n = \frac{N\bar{p}(1-\bar{p})}{N\left(\frac{B^2}{4}\right) + \bar{p}(1-\bar{p})} \tag{22.14}$$

Muestreo aleatorio simple estratificado

Estimación puntual de la media poblacional

$$\bar{x}_{st} = \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)\bar{x}_h \tag{22.15}$$

Estimación del error estándar de la media

$$s_{\bar{x}_{st}} = \sqrt{\frac{1}{N^2}\sum_{h=1}^{H} N_h(N_h - n_h)\frac{s_h^2}{n_h}} \tag{22.16}$$

Intervalo de confianza de aproximadamente 95% para la media poblacional

$$\bar{x}_{st} \pm 2s_{\bar{x}_{st}} \tag{22.17}$$

Estimación puntual del total poblacional

$$\hat{X} = N\bar{x}_{st} \tag{22.18}$$

Estimación del error estándar de $\hat{X}$

$$s_{\hat{X}} = Ns_{\bar{x}_{st}} \tag{22.19}$$

Intervalo de confianza de aproximadamente 95% para estimar el total poblacional

$$N\bar{x}_{st} \pm 2s_{\hat{X}} \tag{22.20}$$

Estimación puntual de la proporción poblacional

$$\bar{p}_{st} = \sum_{h=1}^{H}\left(\frac{N_h}{N}\right)\bar{p}_h \tag{22.21}$$

Estimación del error estándar de $\bar{p}_{st}$

$$s_{\bar{p}_{st}} = \sqrt{\frac{1}{N^2}\sum_{h=1}^{H} N_h(N_h - n_h)\left[\frac{\bar{p}_h(1-\bar{p}_h)}{n_h - 1}\right]} \tag{22.22}$$

Intervalo de confianza de aproximadamente 95% para la proporción poblacional

$$\bar{p}_{st} \pm 2s_{\bar{p}_{st}} \tag{22.23}$$

Asignación del total n de la muestra a los estratos: asignación de Neyman

$$n_h = n\left(\frac{N_h s_h}{\sum_{h=1}^{H} N_h s_h}\right) \tag{22.24}$$

Tamaño de la muestra cuando se estima la media poblacional

$$n = \frac{\left(\sum_{h=1}^{H} N_h s_h\right)^2}{N^2\left(\frac{B^2}{4}\right) + \sum_{h=1}^{H} N_h s_h^2} \tag{22.25}$$

Tamaño de la muestra cuando se estima el total poblacional

$$n = \frac{\left(\sum_{h=1}^{H} N_h s_h\right)^2}{\frac{B^2}{4} + \sum_{h=1}^{H} N_h s_h^2} \quad \textbf{(22.26)}$$

Tamaño de la muestra para la estimación de la proporción poblacional

$$n = \frac{\left(\sum_{h=1}^{H} N_h \sqrt{\bar{p}_h(1 - \bar{p}_h)}\right)^2}{N^2\left(\frac{B^2}{4}\right) + \sum_{h=1}^{H} N_h \bar{p}_h(1 - \bar{p}_h)} \quad \textbf{(22.27)}$$

Asignación proporcional de la muestra *n* a los estratos

$$n_h = n\left(\frac{N_h}{N}\right) \quad \textbf{(22.28)}$$

Muestreo por conglomerados

Estimación puntual de la media poblacional

$$\bar{x}_c = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} M_i} \quad \textbf{(22.29)}$$

Estimación del error estándar de la media

$$s_{\bar{x}_c} = \sqrt{\left(\frac{N - n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}(x_i - \bar{x}_c M_i)^2}{n - 1}} \quad \textbf{(22.30)}$$

Intervalo de confianza de aproximadamente 95% para la media poblacional

$$\bar{x}_c \pm 2s_{\bar{x}_c} \quad \textbf{(22.31)}$$

Estimación puntual del total poblacional

$$\hat{X} = M\bar{x}_c \quad \textbf{(22.32)}$$

Estimación del error estándar de $\hat{X}$

$$s_{\hat{X}} = Ms_{\bar{x}_c} \quad \textbf{(22.33)}$$

Intervalo de confianza de aproximadamente 95% para el total poblacional

$$M\bar{x}_c \pm 2s_{\hat{X}} \quad \textbf{(22.34)}$$

Estimación puntual de la proporción poblacional

$$\bar{p}_c = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} M_i} \quad \textbf{(22.35)}$$

Estimación del error estándar de $\bar{p}_c$

$$s_{\bar{p}_c} = \sqrt{\left(\frac{N-n}{Nn\bar{M}^2}\right)\frac{\sum_{i=1}^{n}(a_i - \bar{p}_c M_i)^2}{n-1}} \tag{22.36}$$

Intervalo de confianza de aproximadamente 95% para estimar la proporción poblacional

$$\bar{p}_c \pm 2 s_{\bar{p}_c} \tag{22.37}$$

Ejercicios complementarios

18. Para evaluar la aceptación que tiene ante los consumidores una nueva publicidad de la cerveza Millar Lite Beer, Louis Harris realizó una encuesta nacional con 363 adultos que habían visto la nueva publicidad de Millar Lite (*USA Today,* 17 de noviembre de 1997). Algunas de las respuestas obtenidas en la encuesta fueron las que se presentan a continuación (*Nota*: Como en esta encuesta sólo se tomó en la muestra una pequeña fracción de todos los adultos, en todas las fórmulas en las que se use el error estándar suponga que $((N - n)/N = 1)$.
    a. Diecinueve por ciento de los entrevistados indicó que la nueva publicidad les gustaba mucho. Dé un intervalo de confianza de 95% para esta proporción poblacional.
    b. A 31% de los entrevistados no les gustó la nueva publicidad. Dé un intervalo de confianza de 95% para esta proporción poblacional.
    c. Diecisiete por ciento de los entrevistados indicó que la nueva publicidad les parecía muy efectiva. Dé un intervalo de confianza de 95% para la proporción de adultos que encuentra que la nueva publicidad es muy efectiva.
    d. Louis Harris informó que “el margen de error era de cinco puntos porcentuales”. ¿Qué significa esto y cómo cree usted que llegó a esta cifra?
    e. ¿Cómo pueden sesgar los errores no muestrales los resultados de un estudio de este tipo?
19. Mediante una encuesta entre los suscriptores de su edición interactiva, *The Wall Street Journal* realizó una investigación. Una de las preguntas que se hizo a los 504 encuestados era si usaban su laptop cuando viajaban; 55% respondió que sí. Otra de las preguntas era si cuando viajaban empleaban un servicio exprés o un servicio de paquetería; 31% respondió que sí (*The Wall Street Journal Interactive Edition Subscriber Survey*, 2000).
    a. Estime el error estándar para la proporción que usa la laptop.
    b. Estime el error estándar de la proporción que usa un servicio exprés o un servicio de paquetería.
    c. ¿Son iguales las estimaciones del error estándar dadas en el inciso a y en el inciso b? Si son diferentes explique por qué.
    d. Dé un intervalo de confianza de 95% para la proporción que usa la laptop.
    e. Dé un intervalo de confianza de 95% para la proporción que usa un servicio exprés o un servicio de paquetería.
20. Mediante una encuesta se realizó un estudio sobre la calidad de vida de los empleados de una fábrica. A 300 de los 3 000 empleados de la fábrica se les envió un cuestionario. La tasa de respuesta fue de 67%, lo que corresponde a 200 cuestionarios contestados.
    a. En la muestra el salario anual medio fue $\bar{x} = \$23\,200$ con $s = \$3\,000$. Dé un intervalo de confianza de aproximadamente 95% para el salario anual medio de la población.
    b. Use la información del inciso a para obtener un intervalo de confianza de aproximadamente 95% para el total de los salarios de los 3 000 empleados.

c. Setenta y tres por ciento de los entrevistados informó estar "en general satisfechos" con su trabajo. Proporcione un intervalo de confianza de aproximadamente 95% para esta proporción poblacional.
d. Haga un comentario sobre si usted considera que el resultado del inciso c pueda estar sesgado. ¿Cambiaría su opinión si supiera que a los entrevistados se les garantizó el anonimato?

21. En un informe del Comité Judicial del Senado de Estados Unidos se presenta el número de homicidios en cada estado. En Indiana, Ohio y Kentucky, el número de homicidios fue, respectivamente, 380, 760 y 260. Suponga que se tomó una muestra aleatoria estratificada para conocer más acerca de las víctimas y de la causa de su muerte, los resultados se presentan a continuación

Estrato	**Tamaño de la muestra**	**Disparo**	**Golpiza**	**Víctima urbana**
Indiana	30	10	9	21
Ohio	45	19	12	34
Kentucky	25	7	11	15

a. Dé un intervalo de confianza de aproximadamente 95% para la proporción de muertes por disparo con arma de fuego en Indiana.
b. Estime el número total de muertes por disparo con arma de fuego en Ohio.
c. Dé un intervalo de confianza de aproximadamente 95% para la proporción de muertes por disparo con arma de fuego en Ohio.
d. Dé un intervalo de confianza de aproximadamente 95% para la proporción de muertes por disparo con arma de fuego en los tres estados.

22. Remítase a los datos del ejercicio 21.
a. Estime la cantidad total de muertes (en los tres estados) por golpizas.
b. Dé un intervalo de confianza de aproximadamente 95% para la proporción de muertes por golpizas en los tres estados.
c. Dé un intervalo de confianza de aproximadamente 95% para la proporción de víctimas urbanas.
d. Estime la cantidad total de víctimas urbanas.

23. Se va a tomar una muestra aleatoria simple estratificada de los clientes de un banco para tener información sobre actitudes y datos demográficos. La estratificación se basará en el estado de cuenta al 30 de junio de 2001. A continuación se presenta una distribución de frecuencias para cada estrato junto con las desviaciones estándar de los estados de cuenta por estrato.

Estrato ($)	**Cuentas**	**Desviación estándar de los estados de cuenta**
0.00–1 000.00	3000	80
1 000.01–2 000.00	600	150
2 000.01–5 000.00	250	220
5 000.01–10 000.00	100	700
más de 10 000.00	50	3000

a. Si el costo por unidad muestreada es aproximadamente el mismo en todos los estratos, determine el número total de personas que deberán incluirse en la muestra. Suponga que se desea una cota del error de estimación de la media poblacional de los estados de cuenta de $B = \$20$.
b. Utilice el procedimiento de asignación de Neyman para determinar el número que debe ser muestreado de cada estrato.

24. Un organismo público está interesado en conocer más acerca de las personas que viven en casas de reposo en una determinada ciudad. En esa ciudad hay en total 100 casas de reposo que atienden a 4 800 personas y se ha tomado una muestra por conglomerados de seis casas de reposo.

Casa	**Residentes**	**Edad promedio de los residentes**	**Residentes inválidos**
1	14	61	12
2	7	74	2
3	96	78	30
4	23	69	8
5	71	73	10
6	29	84	22

a. Estime la edad promedio de los residentes en las casas de reposo en esa ciudad.
b. Dé un intervalo de confianza de aproximadamente 95% para la proporción de personas inválidas en las casas de reposo de esa ciudad.
c. Estime el número total de personas inválidas en las casas de reposo de esa ciudad.

Apéndices