

I. DISEÑO Y ANÁLISIS DEL MUESTREO

Una vez que el problema, los objetivos y las hipótesis de una investigación han sido planteados, la etapa siguiente que debe ser atendida es el diseño del muestreo. La parte estadística del muestreo es indispensable en la obtención de datos, es la piedra angular de la investigación. Una muestra que no tenga representación de la población de estudio imposibilita la obtención de datos con-fiables, aún cuando en las etapas posteriores se utilicen las técnica de análisis más sofisticadas.

En este apartado se dan las bases teóricas para atender las tres grandes preguntas que deben ser contestadas en todo proceso de muestreo: *¿cómo* debo tomar las muestras?, *¿dónde* debo muestrear? y, *cuántas* muestras debo tomar? para que mis datos sean representativos de la población que pretendo estudiar.

1

INTRODUCCIÓN AL DISEÑO Y ANÁLISIS DEL MUESTREO DE POBLACIONES FINITAS

Jorge Navarro Alberto*

Muestreo de poblaciones finitas I

Diseño de encuestas por muestreo

A diferencia de los científicos físicos que realizan experimentos, los científicos sociales y de la vida silvestre llevan a cabo encuestas para recolectar una muestra. Reconocemos así, que existen diferencias de un campo de la ciencia a otro, en la naturaleza de las poblaciones y la naturaleza en la que una muestra puede ser extraída. Por ejemplo, la población de animales de una especie particular pueden contener únicamente un número pequeño de elementos.

Las limitaciones en el procedimiento de muestreo también varían de un área de la ciencia a otra. El muestreo en las ciencias físicas frecuentemente pueden ser ejecutadas bajo *condiciones experimentales controladas*. Tal control es prácticamente *imposible* en las ciencias sociales y manejo de recursos naturales. Por ejemplo, un investigador en medicina puede comparar el crecimiento de ratas sometidas a dos fármacos diferentes. Para este experimento los pesos iniciales de las ratas y la ingesta diaria de alimento pueden ser controlados para reducir alguna variación indeseable en el experimento. En contraste, muy pocas variables pueden ser controladas al estudiar el efecto de las condiciones ambientales sobre el número de venados en la Península de Yucatán.

Las técnicas descritas en este capítulo se han aplicado primordialmente en el diseño y análisis de encuestas, por ello el nombre dado a esta área de la estadísti-

* Departamento de Ecología. Facultad de Medicina Veterinaria y Zootecnia. Universidad Autónoma de Yucatán

ca. No obstante, las técnicas han sido ampliadas y mejoradas para dar cabida al muestreo de poblaciones biológicas, en ramas como la Ecología o el Manejo de Recursos Naturales. En este capítulo se presentan métodos básicos para el diseño y análisis del muestreo de poblaciones finitas útiles al manejador de recursos naturales. Al revisar cada tema, tenga presente que el objetivo primordial de cada sección es la *inferencia*. Identifique el procedimiento de muestreo asociado a cada sección, los parámetros poblacionales de interés, sus estimadores y los límites asociados a los errores de estimación.

Inferencia estadística

El objetivo de cualquier encuesta por muestreo es realizar inferencias acerca de una población de interés, partiendo de la información obtenida en una muestra de dicha población. Las inferencias en las encuestas por muestreo usualmente son dirigidas a la *estimación* de ciertas características numéricas de la población, tales como la media, el total o la varianza. Estas medidas descriptivas numéricas de la población se denominan *parámetros*.

Ejemplos de parámetros (parámetros-objetivo).

En esta parte introductoria consideraremos tres parámetros-objetivo que habitualmente se desearían estimar en estudios observacionales: la media, el total y la proporción de “éxito”. Los dos primeros parámetros son útiles cuando el investigador tiene interés en variables cuantitativas. En el cuadro 1 se ejemplifican situaciones en donde estos dos parámetros podrían ser estimados.

Cuadro 1

Variable	Total	Media
Peso de una vaca	Suma de pesos de todas las vacas de un hato	Total dividido por el número de vacas del hato
Superficie de una finca rústica	Suma de superficies en hectáreas	Total dividido por el número de fincas
Número de hijos de una familia	Suma de hijos	Total dividido por el número de familias
Contenido en miligramos de fósforo de una hoja	Suma de miligramos	Total dividido por el número de hojas

Ejemplos de totales y medias poblacionales.

Cuando en lugar de variables cuantitativas, el investigador solamente desea registrar una característica cualitativa *binaria* (es decir, a cada elemento de la

población corresponde una característica o *atributo* y_i que solo puede tomar dos valores, llamados técnicamente “éxito” y “fracaso”), entonces un parámetro-objetivo podría ser la proporción de “éxitos” en la población. Para calcular esta proporción se suma el número de veces que ocurre un “éxito”; entonces la proporción es igual a ese total dividido entre el número de elementos de la población. Por tanto, la proporción puede considerarse un caso particular de la media aritmética cuando el resultado de la medida tiene que ser o “éxito” o “fracaso”. Ejemplos de proporciones se dan en el Cuadro 2.

Desde ahora indicamos al lector que las fórmulas que iremos presentando en el transcurso de este capítulo (que se podrán identificar fácilmente por estar en cuadros etiquetados con la leyenda “CAJA #”), se referirán exclusivamente a los parámetros-objetivo media, total y proporción.

Cuadro 2

Atributo	Total de clase	Proporción
Venado adulto	Número de venados adultos	Total de la clase dividido por el número de animales (adultos y no-adultos)
Cultivo de maíz	Número de padres de familia campesina que cultivan maíz	Total de la clase dividido por el número de padres de familia campesina
Estado civil, soltero	Número de solteros	Total de la clase dividido por el número de individuos
Opinión positiva sobre determinada cuestión	Número de personas con opiniones positivass	Total de la clase dividido por el número de opiniones registradas

Ejemplos de proporciones poblacionales

Estimadores de parámetros

Un *estimador* es una función de variables aleatorias observables y quizás otras constantes conocidas, usado para estimar un parámetro.

Por ejemplo, la media muestral puede ser usada como un estimador de la media poblacional μ . Nótese que es \bar{y} una variable aleatoria y tiene una distribución de muestreo que depende del mecanismo muestral. Algunos de los posibles valores que \bar{y} puede tomar estarán cercanos a μ , y otros pueden estar bastante alejados de μ en cualquiera de los lados, positivo o negativo. Con los métodos que veremos en estas secciones, seleccionaremos un plan de muestreo que nos asegure que $E(\bar{y}) = \mu$ (el valor esperado de \bar{y} es μ) y que la varianza de, $V(\bar{y})$, es “pequeña”.

En general, sea θ un parámetro y sea $\hat{\theta}$ un estimador para θ . Dos propiedades deseables para $\hat{\theta}$ son:

1. $E(\hat{\theta}) = \theta$ (Insesgamiento de)
2. $V(\hat{\theta}) = \sigma_{\hat{\theta}}^2$ es pequeña. (Varianza mínima entre estimadores de θ).

Aunque la distribución de muestreo de $\hat{\theta}$ dependerá del mecanismo de muestreo y los tamaños de muestra y la población, en muchas situaciones se puede usar el Teorema del Límite Central para afirmar que $\hat{\theta}$ es aproximadamente normal. El Teorema del Límite Central y su aplicación a estimadores como \bar{y} o p (probabilidad de éxito de una población binomial) es adecuada si n , el tamaño de la muestra, es grande, digamos, $n \geq 30$.

Definición.

Sea θ un parámetro y $\hat{\theta}$ un estimador de θ . El error de estimación se define como: $|\hat{\theta} - \theta|$.

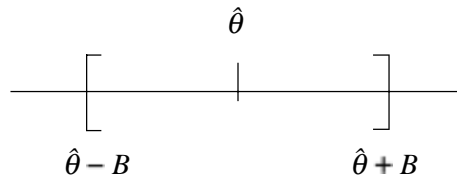
No podemos establecer que un estimador observado estará dentro de una distancia especificada de θ , pero podemos, aproximadamente, encontrar un límite tal que:

$$P(|\hat{\theta} - \theta| \leq B) = 1 - \alpha$$

para cualquier probabilidad deseada, $1 - \alpha$, donde $0 < \alpha < 1$. Si tiene una distribución normal, $B = z_{\alpha/2} \sigma_{\hat{\theta}}$, donde $z_{\alpha/2}$ es el valor que separa un área de $\alpha/2$ en la cola del lado derecho de la distribución normal estándar. Si, $1 - \alpha = .95$ entonces $z_{\alpha/2} = z_{0.025} = 1.96$, o sea, aproximadamente 2. Ya que muchos estimadores que usamos a lo largo de estas sesiones no van a tener una distribución precisamente normal para muchos valores de los tamaños de muestra n y de los tamaños de población N , y ya que el Teorema de Tchebysheff establece que al menos 75% de las observaciones para cualquier distribución de probabilidad estará dentro de dos desviaciones estándar de su media vamos a usar $2\sigma_{\hat{\theta}}$ como límite del error de estimación.

Esto nos da una $P(|\hat{\theta} - \theta| \leq B) = 0.95$ para los casos aproximadamente normales y $P(|\hat{\theta} - \theta| \leq B) \geq 0.75$ en cualquier otro caso, si $B = 2\sigma_{\hat{\theta}}$.

Si $P(|\hat{\theta} - \theta| \leq B) = 1 - \alpha$ entonces, $P(\hat{\theta} - B \leq \theta \leq \hat{\theta} + B) = 1 - \alpha$. En esta forma $[\hat{\theta} - B, \hat{\theta} + B]$, se denomina un *intervalo de confianza para θ con coeficiente de confianza $1 - \alpha$* . La cantidad $\hat{\theta} - B$ se llama el *límite inferior de confianza (LIC)* y $\hat{\theta} + B$ se llama el *límite superior de confianza (LSC)*.



Elementos del problema de muestreo

Consideraremos el problema particular del muestreo de poblaciones finitas, aunque ocasionalmente nos referiremos a poblaciones infinitas.

La cantidad de información obtenida en la muestra para hacer inferencias acerca de la población depende del número de elementos muestreados y de la cantidad de variación de los datos.

El *diseño de la encuesta por muestreo* es el método de selección de la muestra dirigido a controlar la variación de los datos que pudiera afectar la inferencia.

El diseño de la encuesta y el tamaño de la muestra determinan la cantidad de información pertinente a un parámetro poblacional, siempre y cuando se obtengan mediciones exactas en cada elemento muestreado. Como siempre estaremos sujetos a errores, la manera de controlar la exactitud de las mediciones sería mediante métodos adecuados de recolección de datos y por una buena elaboración del instrumento de muestreo (o cuestionario o plan de muestreo).

Ejemplo. En la comunidad de Villa Natura se realizó una encuesta de opinión para determinar la actitud del público hacia la creación de una nueva sección especial para acampar en un parque ecológico. El objetivo de la encuesta fue estimar la proporción del número de personas en la comunidad mayores de 18 años que pudieran hacer uso de la nueva sección para acampar.

Un *ítem o unidad experimental última* es un objeto en el cual se toman las mediciones. En el ejemplo anterior, una unidad experimental última es un habitante de la comunidad, mayor de 18 años.

La medición tomada de esta unidad experimental última es la preferencia del individuo en cuestión respecto a la posibilidad de uso, en el futuro, de la sección para acampar. Podría registrarse la medición como 0 = no lo usará; 1 = la usará.

Una *población* es una colección de elementos acerca de los cuales deseamos hacer una inferencia. La población en el ejemplo de anterior es la colección de los habitantes mayores de 18 años de la comunidad. Otros ejemplos de poblaciones podrían ser todos los hospitales en una ciudad, todos los pacientes con una enfermedad ahora y en el futuro, todos los venados cola blanca en un área, o todas

las órdenes de pago procesadas por una compañía en el período de un año. Algunas poblaciones, tales como los hospitales en una ciudad, son de un tamaño finito de modo que pueden determinarse si es necesario. Otras, como los pacientes con un enfermedad ahora y en el futuro son finitas, pero de un tamaño indeterminado. En algunos casos una población es infinita, como son los resultados que pueden obtenerse repetidamente al lanzar una moneda y ver si el resultado es águila o sol.

Consideraciones importantes

Uno debe definir cuidadosa y completamente la población *antes* de recolectar la muestra. Así, debemos distinguir entre la *población muestreada* y la *población objetivo*, puesto que algunas partes de la población objetivo pueden ser imposibles de alcanzar (p.e. los “niños de la calle” o los “mendigos” en ciudades grandes no están en listas y no tienen residencias permanentes). Sin embargo, ellos son aún parte de la población de la ciudad. De manera similar, al considerar la población de pacientes potenciales con una enfermedad, no hay manera de muestrear a aquellos que no han nacido aun. Los estudios deberán ser diseñados de modo que la correspondencia entre la población objetivo y la población muestreada sea lo más estrecha posible.

Volviendo al ejemplo, si la única población disponible para el muestreo es una lista de residentes de la comunidad, entonces se debe recolectar información acerca de la edad de la persona muestreada y, así, identificar si el individuo tiene más de 18 años.

Las *unidades de muestreo* son colecciones no sobrelapadas de elementos de la población que cubren la población completa.

En el ejemplo anterior, una unidad de muestreo puede ser un habitante de la comunidad mayor de 18 años, visitante potencial o no, de la zona para acampar en el parque. Sin embargo, un proceso más eficiente puede ser muestrear hogares (colecciones de unidades experimentales últimas). Si los hogares son unidades de muestreo, éstos deben definirse de tal manera que ninguna persona mayor de 18 años de la población pueda ser muestreada más de una vez, y que cada unidad experimental última tenga alguna oportunidad de ser seleccionada.

NOTA. Las muestras de parcelas tomadas, por ejemplo, en estudios en una región en donde vive un animal, son frecuentemente circulares. A pesar de sus ventajas, obviamente las parcelas no pueden cubrir un campo sin que ocurra algún sobrelapamiento. Se sugiere que éste sea lo más pequeño posible para lograr un muestreo eficiente.

Un *marco* es una lista de unidades de muestreo. Como ilustración de lo que podría ser un marco, consideremos de nuevo el ejemplo de Villa Natura. Si especificamos al visitante del parque, potencial o no, mayor de 18 años, habitante de la comunidad, como unidad de muestreo = unidad experimental última, una lista de estos elementos puede servir como el marco de esta encuesta. Tomando en cuenta que el marco tiene que tomarse de una lista más amplia, no está claro muchas veces cuáles son los elementos de la población. Además actualizar la lista diariamente es imposible. En otras situaciones, un directorio de la ciudad o una lista de padres de familia obtenida de los datos del censo puede servir como un marco.

Algunos esquemas de muestreo pueden requerir de marcos múltiples. Por ejemplo, la estimación de rendimientos de un cultivo en un estado, puede involucrar una lista de productores para ser entrevistados y una lista de parcelas para ser medidas.

Una *muestra* es una colección de unidades seleccionadas de un marco o varios marcos. En el ejemplo de la encuesta sobre una nueva sección del parque, cierto número de habitantes con las características señaladas (la muestra) va a ser entrevistado para determinar su posibilidad de uso de la nueva sección. Podemos usar la información obtenida de estos habitantes acerca de la opinión de los habitantes mayores de 18 años en toda la comunidad.

• Por qué tomar muestras?

Hay tres razones principales por las que en una población se deben tomar muestras en lugar de realizar censos:

1. Puede ser impráctico un censo completo debido al costo y el esfuerzo involucrados (p.e. un botánico puede no tener suficiente tiempo para muestrear cada planta en un área).
2. El muestreo es más rápido que un conteo completo (p.e. una administración gubernamental puede decidir tomar una muestra del 10% de la población porque los resultados de un censo completo pueden estar parcialmente obsoletos en el momento en que sean procesados).
3. Las muestras pueden ser más exactas que los censos completos. La tercera razón puede ser sorprendente. Esto sucede porque a menudo los errores más grandes en la encuesta no son los *errores de muestreo* (debidos a los efectos casuales al seleccionar las unidades experimentales). Más bien, son los *errores de no-muestreo* que se deben a cosas como un muestreo sesgado, datos

mal registrados, preguntas no entendidas correctamente, registros perdidos, etc. Una muestra relativamente pequeña pero bien organizada puede dar mejores resultados que una encuesta completa o una muestra grande que no puede ser administrada debido a la falta de recursos.

- ¿Cómo seleccionar la muestra? El diseño de la encuesta por muestreo.

Si θ es el parámetro de interés en una población y $\hat{\theta}$ es el estimador de θ , debemos especificar un límite o cota para el error de estimación, B .

$$\text{error de estimación} = |\hat{\theta} - \theta| < B$$

con la condición de que:

$$P(\text{error de estimación} < B) = 1 - \alpha$$

Una selección usual de B es: $B = 2\sigma_{\hat{\theta}}$ ($1 - \alpha \approx 0.95$).

Después de obtener un límite específico con su probabilidad asociada $1 - \alpha$, podemos comparar diseños diferentes para determinar cuál procedimiento proporciona la precisión deseada al mínimo costo. Los diseños básicos se irán tratando en las siguientes secciones.

Muestreo irrestricto aleatorio. Procedimiento y estimaciones de medias, totales y proporciones

Diseño básico: Muestreo irrestricto aleatorio (m.i.a).

Si un tamaño de muestra n es seleccionado de una población de tamaño N de tal manera que cada muestra posible de tamaño n tiene la misma probabilidad de ser seleccionada, el procedimiento de muestreo se denomina *muestreo irrestricto aleatorio*. A la muestra obtenida se llama *muestra irrestricta aleatoria*.

Lo que es importante aquí es el proceso de selección más que el resultado. Por tanto, una muestra puede ser aleatoria aun cuando parezca que no es aleatoria porque sucede que por pura casualidad sus elementos provienen de una parte pequeña de la población. Por ejemplo, si se toma una muestra aleatoria de muje-

res de una comunidad campesina entonces se podría notar que pareciera contener demasiadas mujeres jóvenes. Esto no invalida la muestra puesto que la mayoría de las muestras pequeñas tienen alguna apariencia de no ser representativas.

¿Cómo seleccionar una muestra irrestricta aleatoria? Con la ayuda de tablas de números aleatorios (o calculadoras con números aleatorios). La selección de números corresponderá al número de elementos de la muestra, de entre un total de N elementos de la población. Con base en esta selección, se procede al muestreo mismo.

¿Cómo se analizan los datos generados por un muestreo irrestricto aleatorio? El análisis de los datos generados dependerá del parámetro-objetivo. En las Cajas 1 y 2, se describen los estimadores para la media y el total poblacionales, respectivamente. En estos casos, se supone que la variable medida en cada unidad muestral es continua. Así, y_i representa la variable medida en la i -ésima unidad de muestreo seleccionada por el m.i.a., $i = 1, 2, \dots, n$.

Caja 1

$$\hat{\mu} = \bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (\text{media muestral de } y)$$

$$\text{VARIANZA ESTIMADA DE: } \hat{V}(\bar{y}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$$

$$\text{Donde } s^2 = \frac{\sum y_i^2 - (\sum y_i)^2}{n-1} \quad (\text{varianza muestral de } y)$$

$$\text{LÍMITE PARA EL ERROR DE ESTIMACIÓN: } 2\sqrt{\hat{V}(\bar{y})} = 2\sqrt{\frac{s^2}{n} \left(\frac{N-n}{N} \right)}$$

Estimador de la media poblacional m , muestreo irrestricto aleatorio

NOTAS

- La cantidad $(N-n)/N$ se llama *corrección por población finita* (cpf). Cuando n sea relativamente pequeña con respecto al tamaño de la población, la cpf estará muy cercana a 1. En la práctica, la cpf puede despreciarse si $(N-n)/N \cdot 0.95$, o equivalentemente, si $n \leq (\frac{1}{30})N$. En tal caso, la varianza estimada de es la cantidad s^2/n , conocida como el *error estándar de la media muestral*.
- Las fórmulas dadas arriba se justifican por el *Teorema del Límite Central*.

- En muchos casos el tamaño de la población no está claramente definido o se desconoce. Por ejemplo, generalmente es imposible conocer el número de contribuyentes de un estado, en donde se quisiera seleccionar una muestra de tal población acerca de sus ingresos. Aquí N se supone grande y la cpf se omite.

Ejercicio (Scheaffer *et al.* 1987, pág. 68).

1. Los encargados de administrar los recursos de los terrenos dedicados a la caza silvestre están interesados en el tamaño de las poblaciones de venado y de conejo en los meses de invierno en un bosque particular. Como una estimación del tamaño de la población, los administradores proponen usar el número promedio de grupos densos de conejos y de venados por parcelas de 10 metros por lado. Una muestra irrestricta aleatoria de $n = 500$ parcelas fue seleccionada, y se observó un número de grupos densos de conejos y de venados. Los resultados de este estudio se resumen en el cuadro 3. Estime m_1 y m_2 , el número promedio de grupos densos de venados y de conejos, respectivamente, por parcelas de 10 metros por lado. Establezca los límites para los errores de estimación.

Cuadro 3

Venados	Conejos
Media muestral = 2.30	Varianza muestral = 0.65
Media muestral = 4.52	Varianza muestral = 0.97

Grupos densos de venados y conejos

Caja 1

$$\hat{\tau} = N\bar{y} = \frac{N \sum_{i=1}^n y_i}{n}$$

VARIANZA ESTIMADA DE: $\hat{\tau}$ $\hat{V}(\hat{\tau}) = \hat{V}(N\bar{y}) = N^2 \left(\frac{s^2}{n} \right) \left(\frac{N-n}{N} \right)$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\hat{\tau})} = 2\sqrt{\hat{V}(N\bar{y})} = 2\sqrt{N^2 \left(\frac{s^2}{n} \right) \left(\frac{N-n}{N} \right)}$$

Estimador del total poblacional τ en el m.i.a.

Ejercicio (Modificado de Scheaffer *et al.* 1987, pág. 70).

- Un investigador está interesado en estimar el número total de “árboles marcados” (árboles más grandes que cierto tamaño específico) en una plantación de $N=1500$ mecatres (parcelas de 400 m^2 c/u). Esta información se utiliza para estimar el volumen total de madera aserrada para los árboles de la plantación. Una muestra irrestricta aleatoria de $n=100$ parcelas de 1 mecate fue seleccionada, y cada parcela fue examinada en relación con el número de árboles marcados. El promedio muestral para las $n=100$ parcelas de 1 mecate fue, con una varianza muestral de $s^2=136$. Estime el número total de árboles marcados en la plantación. Establezca un límite para el error de estimación.

Caja 3

$$\hat{\pi} = p = \frac{\sum_{i=1}^n y_i}{n}$$

$$y_i = \begin{cases} 1 & \text{Si la } i\text{-ésima unidad de muestreo posee la característica deseada de otro modo} \\ 0 & \end{cases}$$

VARIANZA ESTIMADA DE $\hat{\pi}$: $\hat{V}(\hat{\pi}) = \hat{V}(p) = \frac{pq}{n-1} \left(\frac{N-n}{N} \right)$; $q = 1 - p$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\hat{\pi})} = 2\sqrt{\hat{V}(p)} = 2\sqrt{\frac{pq}{n-1} \left(\frac{N-n}{N} \right)}$$

Estimador de una proporción poblacional $\hat{\pi}$ en el m.i.a.

$\hat{\pi}$ es simplemente el número de “éxitos” dividido entre el tamaño de muestra, o sea, $\hat{\pi}$ es el número promedio de “éxitos” en la muestra.

Ejercicio (Scheaffer *et al.* 1987, pág. 69).

- El Departamento de Caza y Pesca de cierto estado está interesado en la dirección de sus programas futuros de caza. Para mantener un potencial mayor de caza futura, el departamento desea determinar la proporción de cazadores que buscan cualquier tipo de ave de caza. Se obtuvo una muestra irrestricta aleatoria de $n = 1000$ de los $N = 99,000$ cazadores con permiso. Suponga que

430 indicaron que cazaron aves. Estime $\hat{\pi}$, la proporción de cazadores con permiso que buscan aves de caza. Establezca un límite para el error de estimación.

Muestreo aleatorio estratificado. Procedimiento y estimaciones de medias, totales y proporciones

Diseño básico: Muestreo aleatorio estratificado (m.a.e).

Aunque el diseño de muestreo irrestricto aleatorio suele suministrar buenas estimaciones de cantidades poblacionales a un costo bajo, es posible incrementar la cantidad de información para un costo dado con el muestreo aleatorio estratificado.

Muestra aleatoria estratificada

Una muestra aleatoria estratificada es la obtenida mediante la separación de la población en grupos que no presenten traslapes llamados *estratos* y la selección posterior de una muestra irrestricta aleatoria en cada estrato.

En general, no hay nada que perder al usar este tipo de muestreo más complicado, pero hay algunas ganancias potenciales. Primero, si los individuos dentro de cada estrato son mucho más similares que los individuos en general, entonces la estimación de la media poblacional global tendrá un error estándar menor que el que puede obtenerse con un muestreo irrestricto aleatorio del mismo tamaño muestral. Segundo, puede ser de valor tener estimaciones separadas de los parámetros poblacionales para los diferentes estratos. Tercero, la estratificación hace posible muestrear diferentes partes de una población en diferentes formas, lo cual puede constituir un posible ahorro en los costos.

Generalmente, los tipos de estratificación que deberían ser considerados son aquellos basados en ubicaciones espaciales, regiones dentro de las cuales se espera que la población sea uniforme. Por ejemplo, al muestrear una población de plantas sobre un área grande es natural tomar un mapa y partir el área en unos cuantos estratos aparentemente homogéneos basados en factores tales como altitud y tipo de vegetación. Al muestrear insectos en árboles puede tener sentido estratificar con base en diámetros pequeños, medios y grandes de los árboles. Al muestrear casas-habitación, un poblado puede dividirse en regiones en donde las características de edad y clase social son relativamente uniformes. Usualmente la

elección de qué estratificar es sólo una cuestión de sentido común para la encuesta en cuestión.

• ¿Cómo seleccionar una muestra aleatoria estratificada?

1. Especificar claramente los estratos: cada unidad muestral debe ubicarse en uno y sólo un estrato apropiado
2. Seleccionar una muestra irrestricta aleatoria de cada estrato, mediante la técnica correspondiente
3. Asegurarse que las muestras seleccionadas en los estratos sean independientes

Notación para el muestreo aleatorio estratificado.

L = número de estratos.

N_i = número de unidades muestrales en el estrato i .

N = número de unidades muestrales en la población

$= N_1 + N_2 + \dots + N_L$.

n_i = tamaño de la muestra en el estrato i .

N = tamaño de la muestra combinada

$= n_1 + n_2 + \dots + n_L$.

Puede verse en la Caja 4 que la estimación de la media poblacional en el m.a.e. es un promedio pesado o ponderado de las medias de las muestras irrestrictas aleatorias en cada estrato, donde los pesos son los tamaños proporcionales de los estratos: N_i/N . De manera similar, la varianza de la media muestral en el m.a.e. es una ponderación de las varianzas de las muestras irrestrictas aleatorias de cada estrato. Estas ponderaciones aparecen también en las fórmulas correspondientes a la estimación de un total en el m.a.e.

Caja 4

$$\bar{y}_{est} = \frac{1}{N} [N_1 \bar{y}_1 + N_2 \bar{y}_2 + \dots + N_L \bar{y}_L] = \sum_{i=1}^L \frac{N_i}{N} \bar{y}_i = \frac{1}{N} \sum_{i=1}^L N_i \bar{y}_i$$

VARIANZA ESTIMADA DE: \bar{y}_{est}

$$\begin{aligned} \hat{V}(\bar{y}_{est}) &= \frac{1}{N^2} [N_1^2 \hat{V}(\bar{y}_1) + N_2^2 \hat{V}(\bar{y}_2) + \dots + N_L^2 \hat{V}(\bar{y}_L)] \\ &= \frac{1}{N^2} \left[N_1^2 \left(\frac{N_1 - n_1}{N_1} \right) \left(\frac{s_1^2}{n_1} \right) + \dots + N_L^2 \left(\frac{N_L - n_L}{N_L} \right) \left(\frac{s_L^2}{n_L} \right) \right] \\ &= \frac{1}{N^2} \sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{s_i^2}{n_i} \right) \end{aligned}$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\bar{y}_{est})} = 2\sqrt{\frac{1}{N^2} \sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{s_i^2}{n_i} \right)}$$

Estimador de la media poblacional m para el m.a.e.

Ejemplo (Manly, 1992, págs. 30-31).

En un bosque se disponen 4 bloques, y cada bloque se subdivide en parcelas de 20×20 , haciendo un total de 400 parcelas por bloque. Para estimar el número de árboles en el bosque, se toma una muestra aleatoria de los bloques (estratos) eligiendo un número de renglón aleatorio entre 1 y 20 y un número de columna aleatorio entre 1 y 20, usando un generador de números aleatorios en una computadora. Esto produce los conteos muestrales para el número de árboles por parcela que se muestra en el cuadro 4.

Con estos datos, la media de la población estimada se encuentra usando la ecuación del cuadro correspondiente a la estimación de μ : $\bar{y}_{est} = 5.075$, con un error estándar de $\hat{V}(\bar{y}_{est}) = 0.238$.

Ejercicio (Scheaffer *et al.* 1987, pág. 114)

1. Una escuela desea estimar la calificación promedio que pueden obtener estudiantes de sexto grado en un examen de comprensión de lectura. Los estu-

diantes de la escuela son agrupados en tres estratos, los que aprenden de manera rápida en el estrato I, los que aprenden regular en el estrato II y los que aprenden lento en el estrato III. La escuela decide esta estratificación porque de esta manera se reduce la variabilidad en las calificaciones del examen. El sexto grado contiene 55 estudiantes en el estrato I, 80 en el estrato II y 65 en el estrato III. Una muestra aleatoria estratificada de 50 estudiantes es asignada proporcionalmente y produce muestras irrestrictas aleatorias de $n_1=14$, $n_2=20$ y $n_3=16$ de los estratos I, II y III respectivamente. El examen se aplica a la muestra de estudiantes y se obtienen los resultados que se muestran en el cuadro 5. Estimar: a) La calificación promedio para este grado y b) Un límite para el error de estimación.

Cuadro 4

Unidad Muestral	Estrato			
	1	2	3	4
1	8	3	8	0
2	6	5	3	4
3	8	5	5	3
4	6	6	8	4
5	7	2	2	5
6	7	4	4	4
7	7	6	7	6
8	6	4	6	3
9	8	5	5	3
10	8	3	7	4
Media	7.1	4.3	5.5	3.4
Desviación estándar	0.88	1.34	2.07	1.65

Número de árboles por bloque o estrato

Cuadro 5

Estrato I		Estrato II		Estrato III	
80	92	85	82	42	32
68	85	48	75	36	31
72	87	53	73	65	29
85	91	65	78	43	19
90	81	49	69	53	14
62	79	72	81	61	31
61	83	53	59	42	30
		68	52		
		71	61		
		59	42		

Calificaciones por estratos

Caja 5

$$\hat{t}_{est} = N\bar{y}_{est} = N_1\bar{y}_1 + N_2\bar{y}_2 + \dots + N_L\bar{y}_L = \sum_{i=1}^L N_i\bar{y}_i$$

VARIANZA ESTIMADA DE: $N\bar{y}_{est}$

$$\hat{V}(N\bar{y}_{est}) = N^2\hat{V}(\bar{y}_{est}) = \sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{s_i^2}{n_i} \right)$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\bar{y}_{est})} = 2\sqrt{\sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{s_i^2}{n_i} \right)}$$

Estimador de la media poblacional m para el m.a.e.

Ejemplo.

Respecto al ejemplo referido arriba acerca del muestreo estratificado de árboles en un bosque, y tomando como base los resultados del mismo, obtenemos el número total de árboles estimados en la población como:

$$\hat{t}_{est} = N\bar{y}_{est} = 1600 \cdot 5.075 = 8210$$

El error estándar estimado es:

$\sqrt{\hat{V}(N\bar{y}_{est})} = N\sqrt{\hat{V}(\bar{y}_{est})} = 1600 \cdot 0.238 = 380.8$. De aquí, el intervalo de confianza aproximado del 95% para el total poblacional es 8120 ± 380.8 , o de 7359 a 8882.

Ejercicio (Scheaffer *et al.* 1987, pág. 115)

2. Un guardabosques quiere estimar el número total de acres plantados de árboles en los ranchos de un estado. Ya que el número de acres de árboles varía considerablemente con respecto al tamaño del rancho, decide estratificar con base en el tamaño de los ranchos. Los 240 ranchos en el estado son puestos en una de 4 categorías de acuerdo con el tamaño. Una muestra aleatoria estratificada de 40 ranchos, seleccionada mediante asignación proporcional,

produce los resultados del número de acres plantados de árboles que se muestran en el cuadro 6. Estime el número total de acres plantados de árboles en los ranchos del estado, y fije un límite para el error de estimación.

El ejercicio 6, que se presenta a continuación, se resuelve mediante la estimación de una proporción poblacional para un m.a.e. El resumen de las fórmulas correspondientes se da en la Caja 6. Nuevamente, el estimador del parámetro en cuestión es un promedio pesado de las estimaciones de las proporciones poblacionales en cada estrato, con pesos iguales a N_i/N .

Cuadro 6

Estrato I		Estrato II		Estrato III		Estrato IV	
$N_1 = 86$		$N_2 = 72$		$N_3 = 52$		$N_4 = 30$	
$n_1 = 14$		$n_2 = 12$		$n_3 = 9$		$n_4 = 5$	
97	67	125	155	142	256	167	655
42	125	67	96	310	440	220	540
25	92	256	47	495	510	780	
105	86	310	236	320	396		
27	43	220	352	196			
45	59	142	190				
	53	21					

Número de árboles por tamaño de rancho

Ejercicio (Scheaffer *et al.* 1987, pág. 115)

- Una administración municipal está interesada en ampliar las instalaciones de un centro de atención diurna para niños con retraso mental. La ampliación va a incrementar los costos de asistencia a los niños del centro. Se realiza una encuesta por muestreo para estimar la proporción de familias con niños afectados que utilizarán las instalaciones ampliadas. Las familias viven en la ciudad donde se encuentra localizado el centro, y otras viven en las áreas rurales o suburbanas de los alrededores. Entonces se usa muestreo aleatorio estratificado con personas en la ciudad que usan las instalaciones, personas de los alrededores que las usan, personas en la ciudad que no las usan y personas en los alrededores que no las usan, formando los estratos I, 2, 3 y 4, respectivamente. Registros existentes nos dan $N_1 = 97$, $N_2 = 43$, $N_3 = 145$ y $N_4 = 68$. Se obtiene la siguiente proporción que usarán las nuevas instalaciones: $p_1 = 0.87$, $p_2 = 0.93$, $p_3 = 0.60$, $p_4 = 0.53$. Estime la proporción poblacional, p , y establezca un límite para el error de estimación.

Caja 6

$$\hat{x}_{est} = p_{est} = \frac{1}{N} [N_1 p_1 + N_2 p_2 + \dots + N_L p_L] = \frac{1}{N} \sum_{i=1}^L N_i p_i$$

VARIANZA ESTIMADA DE:

$$\begin{aligned} \hat{V}(p_{est}) &= \frac{1}{N^2} [N_1^2 \hat{V}(p_1) + N_2^2 \hat{V}(p_2) + \dots + N_L^2 \hat{V}(p_L)] \\ &= \frac{1}{N^2} \sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{p_i q_i}{n_i - 1} \right) \end{aligned}$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\bar{y}_{est})} = 2\sqrt{\frac{1}{N^2} \sum_{i=1}^L N_i^2 \left(\frac{N_i - n_i}{N_i} \right) \left(\frac{p_i q_i}{n_i - 1} \right)}$$

Aquí p_i es la estimación de la proporción de éxito en el l -ésimo estrato; $q_i = 1 - p_i$

Estimador de la media poblacional π para el m.a.e.

Contrastes entre el muestreo irrestricto aleatorio y el muestreo estratificado aleatorio. Regla óptima para formar estratos

- La varianza estimada de un parámetro puede ser menor en el muestreo irrestricto aleatorio que en el muestreo estratificado aleatorio. Esto puede deberse a que hay mucha variación dentro de algún estrato (ver ejemplo 5.16, del Scheaffer *et al.* (1987), págs. 103-105).
- Si nuestro objetivo en la estratificación es producir estimadores con varianza pequeña, entonces el mejor criterio para definir los estratos es el conjunto de valores que la respuesta puede tomar. Por ejemplo, supóngase que deseamos estimar el ingreso promedio por hogar en una comunidad. Podríamos estimar este promedio con bastante exactitud si pudiéramos poner todos los hogares de bajos ingresos en un estrato y todos los hogares de altos ingresos en otro estrato antes de muestrear. Pero frecuentemente esto es imposible porque el conocimiento detallado de los ingresos antes del muestreo hace desaparecer en principio el problema estadístico. Sin embargo, en ocasiones tenemos algunas frecuencias de datos en categorías generales de la variable

de interés o de alguna variable altamente correlacionada. Para delimitar los estratos, existe el “*método acumulativo de la raíz cuadrada de la frecuencia*”. Esta regla para encontrar el número adecuado de estratos, L , se aplica como sigue:

1. Tabular los datos disponibles en una distribución de frecuencias basada en la variable de estratificación
2. Calcular la raíz cuadrada de la frecuencia observada y acumula estas raíces cuadradas hacia abajo de la tabla
3. Obtener los límites superiores de los estratos para L estratos de los puntos igualmente espaciados

$$\text{Frontera del } i\text{-ésimo estrato} = \left(\frac{\text{Máxima } \sqrt{f} \text{ acumulada}}{L} \right) \times i$$

Ejercicio (Krebs, 1989, pág. 223-225)

4. Se tomó del Cuadro 6.4 de Krebs (1989). Son las abundancias de almejas de la costa de New Jersey en 1981, de acuerdo a los datos originales publicados por Iachan (1985), dispuestos en orden de la profundidad de la muestra. El objetivo es estimar la abundancia de almejas para esta región. En este caso la estratificación se lleva a cabo con base en una variable auxiliar “profundidad”. Haz la delimitación de las muestras en $L = 5$ estratos y, después, estima la abundancia media global con base en la estratificación generada (Cuadro 7).

Ejercicio (Krebs, 1989, pág. 235)

5. Tabular el número observado de almejas (X) en el cuadro 7 en una distribución de frecuencias acumulativa. Estimar los límites óptimos de los estratos para esta variable, basado en tres estratos, usando el procedimiento acumulativo de la raíz cuadrada de la frecuencia. ¿Qué tanto difieren los resultados de esta estratificación de los obtenidos estratificando con la variable profundidad.

Cuadro 7

Clase	Profundidad, Y (m)	No. de <i>f</i> muestras	Número de almejas observadas, X
1	14	4	34, 128, 13, 0
2	15	1	27
3	18	2	361, 4
4	19	3	0, 5, 363
5	20	4	176, 32, 122, 41
6	21	1	21
7	22	2	0, 0
8	23	5	9, 112, 255, 3, 65
9	24	4	122, 102, 0, 7
10	25	2	18, 1
11	26	2	14, 9
12	27	1	3
13	28	2	8, 30
14	29	3	35, 25, 46
15	30	1	15
16	32	1	11
17	33	4	9, 0, 4, 19
18	34	2	11, 7
19	35	3	2, 10, 97
20	36	2	0, 10
21	37	3	2, 1, 10
22	38	2	4, 13
23	40	3	0, 1, 2
24	41	4	0, 2, 2, 15
25	42	1	13
26	45	2	0, 0
27	49	1	0
28	52	1	0

Frecuencia de almejas observadas a diferentes profundidades

Muestreo sistemático. Procedimiento y estimaciones de medias, totales y proporciones

Diseño básico: *Muestreo sistemático (m.s.)*

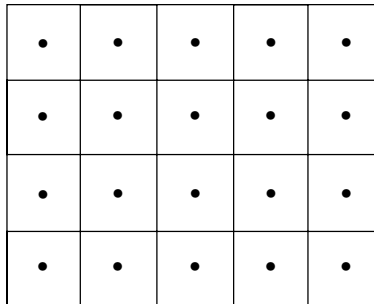
El muestreo irrestricto aleatorio y el muestreo aleatorio estratificado requieren un trabajo muy detallado en el proceso de selección de la muestra. Las unidades de muestreo en un marco adecuado deben ser numeradas de modo que un mecanismo de aleatorización, tal como una tabla de números aleatorios, pueda utilizarse para seleccionar unidades específicas de la muestra.

Un diseño de muestreo usado a menudo debido principalmente a que simplifica el proceso de selección de la muestra se denomina *muestreo sistemático*.

Ejemplo. (Krebs, 1989, págs. 227-229.)

Los ecólogos usan a menudo muestreo sistemático en el campo. Así, las trampas para ratones pueden ser puestas sobre una línea o en una retícula en intervalos a 50 metros. O el llamado método de la distancia del punto-cuarto pudiera aplicarse a lo largo de una línea siguiendo una brújula con 100 metros entre puntos. De hecho, el tipo de muestreo sistemático más común en ecología es el *área de muestreo sistemático céntrico* ilustrado en la Figura 1. El área de estudio se subdivide en cuadrados iguales y una unidad muestral se toma del centro de cada cuadro. Las muestras a lo largo del límite exterior están a la mitad de la distancia de la frontera como lo están de la muestra más cercana. Nótese que una vez que ha sido especificado el número de muestras, solamente hay una muestra céntrica para cualquier área –todas las otras muestras serán excéntricas–.

Figura 1



Desde el punto de vista de la teoría del muestreo por encuesta de poblaciones finitas, el muestreo sistemático, se define como sigue:

Una muestra obtenida al seleccionar aleatoriamente un elemento de los primeros k elementos en el marco y después seleccionar cada k -ésimo elemento, se denomina *muestra sistemática de 1 en k* .

El muestreo sistemático proporciona una opción útil para el muestreo irrestricto aleatorio por las siguientes razones:

1. El muestreo sistemático es fácil de llevar a cabo en el campo, y por tanto, a diferencia de las muestras irrestrictas aleatorias y las muestras aleatorias estratificadas, está menos expuesto a errores de selección que cometen los investigadores en el campo.

2. El muestreo sistemático puede proporcionar mayor información que la que puede proporcionar el muestreo irrestricto aleatorio por unidad de costo. Esto puede ser consecuencia del deseo de muestrear de manera “pareja” a lo largo del área de estudio completa (por ejemplo en ecología, todo un hábitat).

• ¿Cómo seleccionar una muestra sistemática?

1. Se selecciona algún “punto de arranque”, entre 1 y k , para localizar en el marco. Se requiere conocer el tamaño de la población N para poder seleccionar exactamente k . En este caso, para una muestra sistemática de n elementos de una población de tamaño N , k debe ser un número aleatorio menor o igual que N/n ($k \leq N/n$).

2. Se selecciona una unidad al azar entre 1 y k , y después se selecciona cada k -ésimo elemento desde el “punto de arranque”.

Ejemplo (Scheaffer *et al.*, 1987, pág. 171).

Un investigador en medicina está interesado en obtener información acerca del número promedio de veces en que 15000 especialistas recetaron un cierto medicamento en el año anterior ($N = 15000$). Si se quisiera seleccionar una muestra sistemática con $n = 100$ elementos de la población, elegimos el valor de $k \leq N/n = 15,000/100 = 150$. Supóngase que el valor de k elegido es $k = 9$. Entonces podríamos seleccionar un nombre (especialista) al azar de entre los primeros $k = 9$ nombres que aparecen en la lista y luego seleccionar cada noveno nombre hasta que el tamaño de muestra 100 es seleccionado. Esta muestra sistemática es de 1 en 9.

NOTA. En las estimaciones de m y p , si N es desconocida, eliminamos la corrección por población finita (c.p.f.) a saber, $(N-n)/N$, en las ecuaciones correspondientes.

Caja 7

$$\hat{\mu} = \bar{y}_{\text{sist}} = \frac{\sum_{i=1}^n y_i}{n}$$

VARIANZA ESTIMADA DE \bar{y}_{sist} : $\hat{V}(\bar{y}_{\text{sist}}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$

donde $s^2 = \frac{\sum y_i^2 - (\sum y_i)^2}{n-1}$

LÍMITE PARA EL ERROR DE ESTIMACIÓN: $2\sqrt{\hat{V}(\bar{y}_{\text{sist}})} = 2\sqrt{\frac{s^2}{n} \left(\frac{N-n}{N} \right)}$

Estimador de la media poblacional m en el muestreo sistemático

Caja 8

(Requisito: Conocer N , el tamaño de la población)

$$\hat{t}_{\text{sist}} = N\bar{y}_{\text{sist}}$$

VARIANZA ESTIMADA DE \hat{t}_{sist} : $\hat{V}(\hat{t}_{\text{sist}}) = \hat{V}(N\bar{y}_{\text{sist}}) = N^2 \left(\frac{s^2}{n} \right) \left(\frac{N-n}{N} \right)$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\hat{t}_{\text{sist}})} = 2\sqrt{\hat{V}(N\bar{y}_{\text{sist}})} = 2\sqrt{N^2 \left(\frac{s^2}{n} \right) \left(\frac{N-n}{N} \right)}$$

Estimador de un total poblacional t en el m.s.

¿Es correcto usar las fórmulas para el m.i.a en las estimaciones de parámetros basados en una m.s.? Reconocerás que la varianza estimada de \bar{y}_{sist} , \hat{t}_{sist} y p_{sist} es idéntica a la varianza estimada de usando m.i.a. Este resultado no implica que las varianzas poblacionales sean iguales.

La varianza de \bar{y}_{mia} es: $V(\bar{y}_{\text{mia}}) = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right)$. Mientras que en el muestreo

sistemático es: $V(\bar{y}_{\text{sist}}) = \frac{\sigma^2}{n}(1 + (n-1)\rho)$, donde es una medida de correlación entre pares de elementos dentro de la misma muestra sistemática.

Caja 9

$$\hat{\pi}_{\text{sist}} = p_{\text{sist}} = \frac{\sum_{i=1}^n y_i}{n},$$

$y_i = \begin{cases} 1 & \text{Si la } i\text{-ésima unidad de muestreo posee la característica deseada de otro modo} \\ 0 & \end{cases}$

VARIANZA ESTIMADA DE p_{sist} :

$$\hat{V}(\hat{\pi}_{\text{sist}}) = \hat{V}(p_{\text{sist}}) = \frac{p_{\text{sist}} q_{\text{sist}}}{n-1} \left(\frac{N-n}{N} \right); \quad q_{\text{sist}} = 1 - p_{\text{sist}}$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\hat{\pi}_{\text{sist}})} = 2\sqrt{\hat{V}(p_{\text{sist}})} = 2\sqrt{\frac{p_{\text{sist}} q_{\text{sist}}}{n-1} \left(\frac{N-n}{N} \right)}$$

Estimador de una proporción poblacional p en el m.s.

- Si $\rho \geq 1$, entonces los elementos dentro de la muestra son bastante similares con respecto a la característica que se está midiendo y el muestreo sistemático producirá una varianza de la media muestral mayor que la obtenida por el m.i.a.
- Si ρ es negativo, entonces el muestreo sistemático puede ser mejor que el m.i.a. La correlación puede ser negativa si los elementos dentro de la muestra sistemática tienden a ser extremadamente diferentes.
- ρ no puede alcanzar un valor negativo muy grande como para que la expresión de la varianza llegue a ser negativa.
- Para ρ cercano a cero y N bastante grande, el muestreo sistemático y el m.i.a. son aproximadamente iguales.

En general, cuando el muestreo sistemático es equivalente al m.i.a. podemos tomar $V(\hat{\theta}_{\text{sist}})$ (por ejemplo, $V(\bar{y}_{\text{sist}})$) como aproximadamente igual a la varianza estimada de $\hat{\theta}$ en el m.i.a. La posibilidad de que ocurra esto tiene que considerar el tipo de población objetivo:

1. Población aleatoria. Una población es aleatoria si sus elementos están ordenados al azar
2. Población ordenada. Una población está ordenada si sus elementos dentro de la población están ordenados de acuerdo a algún esquema
3. Población periódica. Una población es periódica si los elementos de la población tienen variación cíclica

Relaciones entre el m.s. y el m.i.a.

1. Los elementos de una muestra sistemática de una población aleatoria se espera que sean heterogéneos con $\rho \approx 0$. Por tanto, cuando N es grande $V(\bar{y}_{\text{sis}}) \approx V(\bar{y}_{\text{mia}})$ y, en este caso, el muestreo sistemático es equivalente al muestreo irrestricto aleatorio.

Ejemplo (Scheaffer *et al.* 1987, págs. 175-176)

Un investigador desea determinar la calidad del jarabe de arce contenido en la savia de los árboles de una finca. El número total de árboles es desconocido; por lo tanto, es imposible realizar un m.i.a. de árboles. Como un procedimiento alternativo, el investigador decide usar una muestra sistemática de 1 en 7. Si quisiera estimar m , el contenido de azúcar promedio de los árboles de arce en la finca, es razonable que él suponga que la población de árboles es aleatoria. Según esta suposición, la $V(\bar{y}_{\text{sis}})$ estimada estaría dada por la fórmula correspondiente

para el m.i.a.,
$$\hat{V}(\bar{y}_{\text{mia}}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$$

2. Una muestra sistemática extraída de una población ordenada es generalmente heterogénea con $\rho \leq 0$. Puede demostrarse que para una población ordenada, $V(\bar{y}_{\text{mia}}) \leq V(\bar{y}_{\text{sis}})$.

Por tanto, una muestra aleatoria de una población ordenada proporciona más información que una m.i.a. por unidad de costo. Ya que no podemos obtener una estimación de m con base en los datos de la muestra, una estimación conservadora (una que es mayor que lo que se esperaría) de m está dada por:

$$\hat{V}(\bar{y}_{\text{mia}}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$$

Ejemplo.

Si el marco del cual se extrae una muestra sistemática está listado de acuerdo al valor numérico ascendente de la variable de interés, entonces la población está ordenada.

3. Los elementos de una muestra sistemática extraída de una población periódica pueden ser homogéneos ($\rho > 0$). Cuando N es grande y $\rho > 0$,

$$V(\bar{y}_{\text{sist}}) > V(\bar{y}_{\text{mia}}).$$

Por tanto, en este caso, el muestreo sistemático proporciona menos información que el m.i.a. por unidad de costo. Como en las situaciones anteriores, $V(\bar{y}_{\text{sist}})$ no puede ser estimada directamente mediante una sola muestra sistemática. Podemos aproximar su valor usando $V(\bar{y}_{\text{mia}})$. En general, este estimador subestimará la varianza verdadera de \bar{y}_{sist} . Para evitar este problema que ocurre con el muestreo sistemático de una población periódica, el investigador podría cambiar varias veces el punto de inicio aleatorio. Este procedimiento reducirá la posibilidad de seleccionar observaciones con la misma posición relativa en una población periódica. Estos cambios del punto \bar{y}_{sist} inicio tendrán el efecto de mezclar los elementos de la población y al mismo tiempo el de seleccionar una muestra sistemática. Así, la muestra obtenida es equivalente a una muestra sistemática de una población aleatoria y, por tanto, la varianza de puede ser entonces aproximada usando

$$\hat{V}(\bar{y}_{\text{mia}}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right)$$

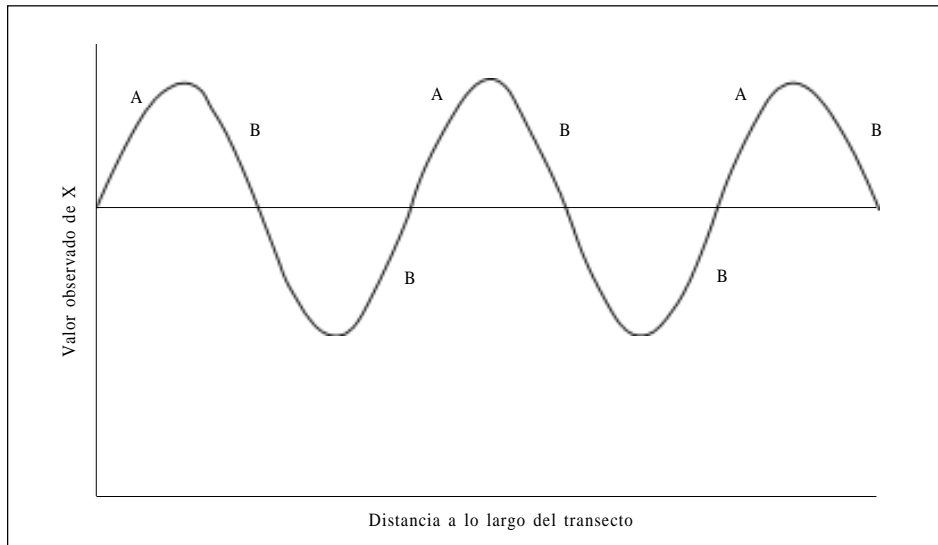
Ejemplo (Krebs, 1989, pág. 228).

La figura 2 ilustra un ejemplo hipotético en la cual una variable ambiental (por ejemplo, contenido de humedad del suelo) varía en una forma periódica, como una sinusoidal sobre el área de estudio. El efecto de usar muestreo sistemático para estimar la media de esta variable puede ser de muy diversa índole. Si no se muestreara en A, siempre se obtendría la misma medición y se obtendría una estimación altamente sesgada de la media. Si se tiene suerte y se muestreara en B, se obtendría la misma media y varianza como si se hubiera usado muestreo aleatorio. ¿Qué tan probable es que estos problemas como la variación periódica vista en los puntos de muestreo A ocurran en datos de campo reales? Milne (1959)

intentó responder esta pregunta observando muestras sistemáticas tomadas sobre poblaciones biológicas que habían sido enumeradas completamente (de manera que eran conocidas la media y la varianza verdaderas). Él analizó datos de 50 poblaciones y encontró que, en la práctica, no había un error introducido al suponer que una muestra sistemática céntrica era una muestra aleatoria simple, usando todas las fórmulas de la teoría de muestreo irrestricto aleatorio.

La variación periódica como la de la figura 2 no parece ocurrir en sistemas ecológicos. Más aun, la mayoría de los patrones ecológicos son altamente aglutinados e irregulares, de modo que la preocupación de los estadísticos acerca de influencias periódicas, parece ser una preocupación de más. La recomendación práctica es así: *se puede usar muestreo sistemático pero cuidar posibles tendencias periódicas*. Krebs termina la sección correspondiente al muestreo sistemático **en ecología diciendo que, si tienes para elegir entre tomar una muestra aleatoria o una muestra sistemática, escoge siempre un muestreo aleatorio**, debido a que siempre puede haber preocupación por la presencia de efectos periódicos que pueden influir en las estimaciones. Pero si el costo e inconveniencia de la aleatorización son demasiado grandes, puedes perder muy poco al muestrear de manera sistemática.

Figura 2



Variación periódica de una variable X en función de la distancia a lo largo de un transecto.

• **Observaciones**

Las consideraciones acerca de los tipos de población (aleatoria, ordenada y periódica) y la relación de los estimadores de la varianza de la media muestral para el muestreo sistemático y el m.i.a., también son válidas para cuando los estimadores de interés son un total, \hat{t}_{sist} y una proporción, P_{sist} .

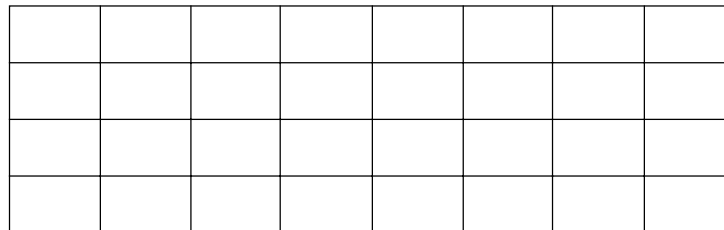
Si la estratificación de la población es ventajosa, el muestreo sistemático puede usarse dentro de cada estrato en lugar del m.i.a.

Como en ciertos casos el muestreo aleatorio sistemático no es equivalente al m.i.a., un método alternativo para estimar la varianza de un parámetro estimado, $\hat{\theta}_{sist}$, digamos, $\hat{V}(\hat{\theta}_{sist})$, es el *muestreo sistemático repetido*. Como el nombre lo indica, este método de muestreo requiere de la selección de más de una muestra sistemática. Los detalles de este método pueden encontrarse en Scheaffer *et al.* (1987), págs. 182-186.

Ejercicio. (Scheaffer *et al.* 1987, págs. 188-189).

6. Edafólogos quieren estimar la cantidad de calcio intercambiable (en ppm) en una parcela de terreno. Para simplificar el esquema de muestreo, en el terreno se sobrepone un malla rectangular. En cada punto de intersección en la malla se toman muestras de suelo (Figura 3). Determina la cantidad promedio de calcio intercambiable en la parcela de terreno. Establecer un límite para el error de estimación. Los datos son: $n = 45$. $\sum y_i = 90,320$ calcio intercambiable, $\sum y_i^2 = 148,03000$

Figura 3



Muestreo sistemático sobre una rejilla

Ejercicio (Scheaffer *et al.* 1987, págs. 190-191)

- Los funcionarios de un museo están interesados en el número total de personas que visitan el lugar durante el periodo de 180 días en que una colección de antigüedades está en exhibición. Puesto que llevar el control diario de visitantes es muy costoso, los funcionarios deciden obtener estos datos cada décimo día. La información de esta muestra sistemática de 1- en -10 se resume en la tabla adjunta. Use estos datos para estimar τ , el número total de personas que visitan el museo durante el período específico. Establezca un límite para el error de estimación.

Cuadro 6

Día	Número de personas que visitaron el museo
3	160
13	350
23	225
⋮	⋮
173	290

$$\sum_{i=1}^{18} y_i = 4,868$$

$$\sum_{i=1}^{18} y_i^2 = 1,321,450$$

Número de visitantes a un museo

Ejercicio (Scheaffer *et al.* 1987, pág. 191)

- Los funcionarios de cierta sociedad profesional desean determinar la proporción de miembros que apoyan varias enmiendas propuestas en las prácticas de arbitraje. Los funcionarios conducen una muestra sistemática de 1- en -10, a partir de una lista en orden alfabético de los $N = 650$ miembros registrados. Sea $y_i = 1$, si la i -ésima persona favorece los cambios propuestos y $y_i = 0$ si se opone a los cambios. Usar los siguientes datos de la muestra para estimar π , la proporción de miembros en favor de cambios propuestos. Establecer un límite para el error de estimación.

$$\sum_{i=1}^{65} y_i = 48$$

Muestreo por conglomerados. Procedimiento y estimaciones de medias, totales y proporciones

Diseño básico: *Muestreo por conglomerados (m.c.)*

Una *muestra por conglomerados* es una muestra aleatoria en la cual la unidad de muestreo es una colección o conglomerado de unidades experimentales últimas. Se recomienda su uso cuando no se encuentra disponible o es muy costoso obtener un buen marco para la población, mientras que se puede lograr fácilmente un marco que liste conglomerados. También es idóneo cuando el costo por obtener datos se incrementa con la distancia que separa las unidades experimentales últimas.

Generalmente, del muestreo aleatorio por conglomerados no puede esperarse que dé la misma precisión que una muestra aleatoria simple con el mismo número total de elementos, puesto que los elementos cercanos tienden a ser más similares que los elementos en general. Por tanto, una muestra conglomerada es equivalente a una muestra de unidades experimentales con un tamaño de muestra, en cierta forma, más pequeño. Sin embargo, los ahorros en los costos pueden hacer que una muestra por conglomerados sea considerablemente más grande que lo que pueda ser una muestra irrestricta aleatoria. De aquí que una muestra por conglomerados puede dar una mejor precisión que una muestra aleatoria simple por el mismo costo de muestreo.

Ejemplos.

1. Las manzanas de la ciudad son usadas frecuentemente como conglomerados de hogares o personas.
2. Un naranjo forma un conglomerado de naranjas para la investigación de infestación por insectos.
3. Una parcela en el bosque contiene un conglomerado de árboles para la estimación de volúmenes de madera o proporción de árboles enfermos.

• ¿Cómo seleccionar una muestra por conglomerados?

1. Especificar los conglomerados apropiados
Las unidades experimentales últimas tienden a presentar características similares dentro de un conglomerado, así que la medición de una unidad experimental última en un conglomerado puede estar altamente correlacionado con la de otra unidad experimental última. Por tanto, si se selecciona un conglomerado de gran tamaño, la cantidad de información al tomar todas las

medidas de cada unidad experimental última puede no incrementarse sustancialmente.

El problema de elegir el tamaño de conglomerado puede aun ser más complicado cuando se dispone de un número infinito de posibles tamaños de conglomerados, como en la selección de parcelas forestales para la estimación de la proporción de árboles enfermos. Si existe variabilidad en la densidad de árboles enfermos a lo largo y ancho del bosque, entonces muchos conglomerados pequeños localizados de manera aleatoria o sistemática, pueden ser lo deseable. Sin embargo, localizar aleatoriamente una parcela consume mucho tiempo y, una vez localizada, el muestreo de muchos árboles es económicamente inconveniente. Entonces muchas parcelas pequeñas son ventajosas para controlar la variabilidad, pero pocas parcelas grandes son económicamente recomendables. Se debe de encontrar un equilibrio entre el número y el tamaño de las parcelas. Aunque no hay reglas generales para tomar esta decisión, las encuestas piloto pueden ayudar al investigador a encontrar la dirección.

2. Conformar un marco que liste todos los conglomerados.
3. Seleccionar una muestra irrestricta aleatoria de este marco.
4. Los elementos o unidades experimentales últimas muestreadas serán todos los elementos de los conglomerados seleccionados.

NOTACIÓN EN EL MUESTREO POR CONGLOMERADOS (m.c.).

N = número de conglomerados (población); n = número de conglomerados (m.i.a.).

m_i = número de elementos en el conglomerado i , $i = 1, \dots, N$.

$\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$ = tamaño promedio del conglomerado en la muestra.

$M = \sum_{i=1}^N m_i$ = número de unidades experimentales últimas en la población.

$\bar{M} = \frac{M}{N}$ = tamaño promedio de los conglomerados en la población.

y_i = total de todas las observaciones en el i -ésimo conglomerado.

Caja 10

$$\bar{y}_{mc} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i}$$

VARIANZA ESTIMADA DE \bar{y}_{mc} :
$$\hat{V}(\bar{y}_{mc}) = \left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}_{mc} m_i)^2}{n-1}$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(\bar{y}_{mc})} = 2\sqrt{\left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}_{mc} m_i)^2}{n-1}}$$

Aquí \bar{M} puede ser estimada por \bar{m} si se desconoce M .

Estimación de la media poblacional m para el m.c.

Ejercicio (Scheaffer *et al.* 1987, págs. 226-227).

9. Un guardabosques desea estimar la altura promedio de los árboles de una plantación. La plantación se divide en parcelas de un cuarto de acre. Se selecciona una muestra irrestricta aleatoria de 20 parcelas de un total de 386 parcelas en la plantación. Se miden todos los árboles en las parcelas muestreadas, con los resultados que se muestran en el cuadro 9. Estime la altura promedio de los árboles en la plantación y establecer un límite para el error de estimación. (*Sugerencia:* el total para el conglomerado I se puede encontrar tomando m_i veces el promedio del conglomerado).

El ejercicio 13 se puede resolver mediante la estimación de un total poblacional (Caja 11). Nótese que el estimador $\hat{\tau} = M\bar{y}_{mc}$ es útil únicamente si se conoce el número de elementos M en la población.

Cuadro 9

No. de árboles	Altura promedio (pies)	No. de árboles	Altura promedio (pies)
42	6.2	60	6.3
51	5.8	52	6.7
49	6.7	61	5.9
55	4.9	49	6.1
47	5.2	57	6.0
58	6.9	63	4.9
43	4.3	45	5.3
59	5.2	46	6.7
48	5.7	62	6.1
41	6.1	58	7.0

Número de árboles y altura promedio de una plantación

Ejercicio (Scheaffer *et al.* 1987, pág. 228)

10. Un gran embarque de mariscos congelados es empaquetado en cajas, conteniendo cada una veinticuatro paquetes de 5 libras. Hay cien cajas en el embarque. Un inspector del gobierno determina el peso total (en libras) de mariscos dañados para cada una de cinco cajas muestreadas. Los datos son como sigue: 9, 6, 3, 10, 2. Estimar el peso total de mariscos dañados en el embarque y establecer un límite para el error de estimación.

Caja 11

$$\hat{\tau} = M\bar{y}_{mc} = M \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i}$$

VARIANZA ESTIMADA DE $\hat{\tau} = M\bar{y}_{mc}$:

$$\hat{V}(M\bar{y}_{mc}) = M^2 \hat{V}(\bar{y}_{mc}) = N^2 \left(\frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}_{mc} m_i)^2}{n-1}$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(N\bar{y}_i)} = 2\sqrt{N^2 \left(\frac{N-n}{Nn} \right) \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n-1}}$$

Estimación de un total poblacional τ en el m.c. ($\tau = Mm$).

Cuando M se desconoce, se tendrá que usar otro estimador de τ :

Caja 12

$\hat{\tau} = N\bar{y}_t$, donde $\bar{y}_t = \frac{\sum_{i=1}^n y_i}{n}$ es el promedio de los totales de conglomerados para los n conglomerados muestrados

VARIANZA ESTIMADA DE $\hat{\tau} = N\bar{y}_t$:

$$\hat{V}(N\bar{y}_t) = N^2\hat{V}(\bar{y}_t) = N^2\left(\frac{N-n}{Nn}\right)\frac{\sum_{i=1}^n (y_i - \bar{y}_t)^2}{n-1}$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(N\bar{y}_t)} = 2\sqrt{N^2\left(\frac{N-n}{Nn}\right)\frac{\sum_{i=1}^n (y_i - \bar{y}_t)^2}{n-1}}$$

Estimador de τ en el m.c. cuando M se desconoce

$\hat{V}(\bar{y}_{mc})$ es un buen estimador de $\hat{V}(\bar{y}_{mc})$ si n fuera grande ($n \cdot 20$). El sesgo desaparece si $m_1 = m_2 = \dots = m_N$. En el caso de τ , el estimador $N\bar{y}_t$ puede ser menos preciso que $M\bar{y}_{mc}$ pues no usa la información de los tamaños de conglomerados m_1, m_2, \dots, m_n . Pero si $m_1 = m_2 = \dots = m_N$, los dos estimadores son equivalentes.

Ejercicio (Scheaffer *et al.* 1987, pág. 226).

- Los comercios solicitan frecuentemente a los contadores la realización de inventarios. Ya que un inventario completo es costoso, a través del muestreo se pueden realizar inventarios cada cuatro meses. Supóngase que una empresa abastecedora de artículos de plomería desea un inventario para muchos artículos pequeños en existencia. La obtención de una muestra aleatoria de artículos es muy difícil. Sin embargo, los artículos se encuentran dispuestos en anaqueles, y las selección de una muestra aleatoria de anaqueles es relativamente facil, considerando a cada anaquel como un conglomerado de artículos. Una muestra de 10 anaqueles de un total de 48 dio los resultados que se muestran en la tabla siguiente. Estimar la cantidad de dólares de los artículos en los anaqueles y establecer un límite para el error de estimación.

Cuadro 10

Conglomerado	Número de artículos, m_i	Cantidad real de dólares, y_i
1	42	83
2	27	62
3	38	45
4	63	112
5	72	96
6	12	58
7	24	75
8	14	58
9	32	67
10	41	80

Número de artículos y su cantidad en dólares

Caja 13

Sea a_i = número total de elementos en el conglomerado i que poseen la característica de interés. Entonces:

$$\hat{\pi}_{mc} = p_{mc} = \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n m_i}$$

VARIANZA ESTIMADA DE $\hat{\pi}_{mc} = p_{mc}$:
$$\hat{V}(p_{mc}) = \left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (a_i - p_{mc}m_i)^2}{n-1}$$

LÍMITE PARA EL ERROR DE ESTIMACIÓN:

$$2\sqrt{\hat{V}(p_{mc})} = 2\sqrt{\left(\frac{N-n}{Nn\bar{M}^2} \right) \frac{\sum_{i=1}^n (a_i - p_{mc}m_i)^2}{n-1}}$$

Estimación de una proporción poblacional p para el mc.

La fórmula de $\hat{V}(p_{mc})$ es un buen estimador de p únicamente cuando n es grande ($n \cdot 20$). Si $m_1 = m_2 = \dots = m_N$, entonces p es un estimador insesgado de p , y la $\hat{V}(p_{mc})$ expresada en el cuadro de arriba es un estimador insesgado de la varianza real de p , para cualquier tamaño de muestra.

12. Una empresa está considerando la revisión de su política de jubilación y quiere estimar la proporción de empleados que apoyan la nueva política. La industria consiste de 87 plantas separadas localizadas en todo el país. Ya que los resultados deben ser obtenidos rápidamente y con poco dinero, la industria deci-

de usar muestreo por conglomerados, con cada planta como un conglomerado. Se selecciona una muestra irrestricta aleatoria de 15 plantas y se obtienen las opiniones de los empleados en estas plantas a través de un cuestionario. Los resultados se presentan en el cuadro 11. Estimar la proporción de empleados de la industria que apoyan la nueva política de jubilación y establecer un límite para el error de estimación.

Ejercicio (Scheaffer *et al.* 1987)

Cuadro 11

Planta	Número de empleados	Número de empleados que apoyan la nueva política	Planta	Número de empleados	Número de empleados que apoyan la nueva política
1	51	42	9	73	54
2	62	53	10	61	45
3	49	40	11	58	51
4	73	45	12	52	29
5	101	63	13	65	46
6	48	31	14	49	37
7	65	38	15	55	42
8	49	30			

Muestra de empleados de diferentes plantas de una empresa a favor de una nueva política.

Comentarios finales acerca del muestreo de conglomerados

1. El muestreo de conglomerados puede ser combinado con el muestreo estratificado, con objeto de que la población quede dividida en L estratos y se pueda seleccionar una muestra por conglomerados en cada estrato. Ver Scheaffer *et al.* (1987) páginas 215-216.
2. El muestreo sistemático puede interpretarse como un muestreo por conglomerados en donde el tamaño de cada conglomerado es $m_i = 1$, para cada $i = 1, 2, \dots, n$.

Consideraciones generales acerca de otros diseños de muestreo de poblaciones finitas. Muestreos de encuentro y muestreo secuencial

Dentro de la teoría de poblaciones finitas, hemos omitido presentar algunos métodos de estimación de parámetros basados en diferentes diseños de muestreo. Esto no implica que no sean importantes, pero por el espacio dedicado en esta

primera unidad al t3pico del dise1o estad3stico de muestreo por encuesta, se decidi3 limitarse a las t3cnicas b3sicas.

Estimaciones de raz3n, regresi3n y diferencia combinadas con los dise1os de muestreo b3sicos

De los m3todos no vistos, tres de ellos hacen uso de una variable auxiliar, x , para realizar la estimaci3n de par3metros respecto a una variable de respuesta y , sobre la misma unidad muestral. Estos m3todos son: estimaci3n de raz3n, de regresi3n y de diferencia. Lo importante de estas t3cnicas es que diversos dise1os de muestreo pueden ser empleados en uni3n con la estimaci3n. En Scheaffer *et al.* (1987) y en Cochran (1976), entre otros, se explica con detalle la conjugaci3n de estos m3todos de estimaci3n con el muestreo irrestricto aleatorio y con el muestreo aleatorio estratificado.

Muestreo por conglomerados en dos etapas y poliet3pico

Estos m3todos son modificaciones del muestreo por conglomerados o 3reas. En el muestreo en dos etapas o biet3pico no todos los elementos o unidades de los conglomerados forman parte de la muestra, sino que, una vez seleccionados 3stos, se efect3a una nueva selecci3n o submuestreo dentro de cada uno. La generalizaci3n del muestreo por conglomerados biet3pico es el poliet3pico, en donde los conglomerados seleccionados en la primera etapa pueden estar constituidos por nuevos conglomerados incluidos en ellos, de modo que pueda procederse a un submuestreo de unidades de conglomerados dentro de conglomerados y as3 sucesivamente, en varias etapas (Azor3n Poch, 1972).

Submuestras interpenetrantes

Este m3todo consiste en la divisi3n aleatoria de una muestra, generalmente tambi3n aleatoria, en un cierto n3mero de grupos del mismo n3mero de unidades. Se emplea para medir la concordancia entre los resultados de muestras sucesivas, en la obtenci3n r3pida de resultados provisionales y en la comparaci3n de resultados obtenidos por diferentes entrevistadores (Azor3n Poch, 1972).

Muestras de encuentro

Las secciones anteriores se refirieron a lo que puede describirse como la teoría clásica de muestreo de poblaciones finitas. Estos problemas cubren mucho de los problemas de muestreo que son probables que surjan en una investigación, pero hay situaciones en donde se requieren enfoques diferentes. En particular, algunas veces surgen casos en donde no es posible decidir de manera previa dónde y cuándo se muestrearán los elementos de la población. En lugar de un esquema de muestreo, debe vislumbrarse qué permite a los elementos ser encontrados con una cierta probabilidad. El análisis de los datos debe tomar en cuenta la naturaleza de este esquema de muestreo. Como se verá frecuentemente en este curso, varios de los métodos de muestreo de organismos pueden considerarse métodos de muestreo de encuentro. Sin embargo, desde una perspectiva biológica o ecológica, es difícil evaluar la probabilidad de encuentro de un determinado organismo o elemento de interés. A continuación damos una lista de diseños de muestreo que son útiles en el estudio de procesos biológicos o ecológicos, mismos que serían tratados en las secciones correspondientes al muestreo de organismos en este curso. Aquel lector que lo desee puede acudir a los libros de Krebs (1988), Manly (1992) y Thompson (1992), en donde se abordan estos métodos de forma breve y sin demasiado formalismo.

- Muestreo de marcaje-recaptura
- Muestreo de remoción
- Muestreo en transectos (en cualquiera de sus versiones particulares: muestreo de intercepto de línea, de transecto en banda y de transecto en línea)
- Muestreo con sesgo en el tamaño
- Muestreo de posiciones de objetos en el espacio (patrones espaciales) en sus versiones particulares: métodos basados en distancias, métodos basados en cuadrantes
- Muestreo adaptativo.
- Muestreo de conjunto ordenado.

Muestreo secuencial

En el libro de Krebs (1989), en el capítulo 7, y en el libro de Azorín Poch (1972), en los capítulos 23 y 24, se trata con detalle otro método de muestreo conocido como muestreo secuencial, cuya cualidad más importante es que los tamaños de las muestras no se establecen a priori. En lugar de ello, se hacen observaciones o

mediciones, una en cada tiempo, y después de que se haga cada observación “se pregunta” a los datos acumulados si puede alcanzarse o no una conclusión. Así, el tamaño de muestra es minimizado, y en algunos casos solamente la mitad del número de observaciones requeridas con el muestreo clásico es necesario con el muestreo secuencial.

Muestreo de poblaciones finitas II (selección del tamaño de muestra)

Enfoque general en el diseño de encuestas por muestreo

El número de observaciones necesarias para estimar un parámetro θ requerirá que se establezca un límite para el error de estimación igual a B e igual también a $2\sqrt{V(\hat{\theta})}$. Como habitualmente $V(\hat{\theta})$ contiene en su expresión al tamaño de la muestra, ésta se puede hallar resolviendo para n :

$$2\sqrt{V(\hat{\theta})} = B$$

Por tanto, en las sub-secciones siguientes se encontrarán las fórmulas de los tamaños de muestra requeridos para estimar el parámetro indicado con un límite para el error de estimación de B ; las fórmulas variarán según el parámetro-objetivo y según el diseño de muestreo.

Selección del tamaño de muestra para estimar medias, totales y proporciones poblacionales en el m.i.a.

Ejercicio (Scheaffer *et al.* 1987, pág. 69)

13. Un psicólogo desea estimar el tiempo de reacción promedio para un estímulo entre 200 pacientes de un hospital especializado en trastornos nerviosos. ¿De qué tamaño debe ser la muestra irrestricta aleatoria que debe tomar para estimar m con un límite para el error de estimación de 1 segundo? Use 1.0 segundos como una aproximación de la desviación estándar poblacional.

Ejercicio (Scheaffer *et al.* 1987, pág. 70)

14. Usando los resultados de la encuesta descrita en el Ejercicio 2, determine el tamaño de muestra requerido para estimar t , el número total de árboles en plantación, con un límite para el error de estimación de magnitud $B = 1500$.

Caja 14

$$\mu: n = \frac{N\sigma^2}{(N-1)D + \sigma^2}, \text{ donde } D = \frac{B}{4N^2}$$

Nota: s^2 puede aproximarse con s^2 , a partir de una muestra piloto, o un estudio previo similar, o bien $\sigma^2 = \frac{\text{Rango}}{4}$. Entonces, en cualquiera de estos casos n será un valor aproximado.

$$\tau: n = \frac{N\sigma^2}{(N-1)D + \sigma^2}, \text{ donde}$$

$$\pi: n = \frac{N\pi(1-\pi)}{(N-1)D + \pi(1-\pi)}, \text{ donde}$$

Nota: π puede aproximarse con p (y, así, $(1-\pi)$ se estima con $q = 1-p$; ver sección 1.1.2) a partir de una muestra piloto, o un estudio previo similar, o bien, en el peor de los casos $p=0.5$ (n será conservador en este caso). Entonces, en cualquiera de estas situaciones n será un valor aproximado. En todas estas fórmulas, se requiere conocer N , el tamaño de la población.

Selección del tamaño de muestra para estimar medidas totales y proporciones poblacionales en el m.i.a.

Ejercicio (Scheaffer *et al.* 1987, pág. 69)

- Usando los datos del Ejercicio 3, determine el tamaño de muestra que el Departamento debe obtener para estimar la proporción de cazadores con permiso que buscan aves de caza. Establezca un límite para el error de estimación de magnitud $B = 0.02$.

Selección del tamaño de muestra para estimar medias, totales y proporciones poblacionales en el m.a.e. Asignación de la muestra

Para el caso de la selección de n en el m.a.e. para estimar m o t o p , se presenta un problema al querer despejar n_i (el tamaño de la muestra para cada estrato) en la expresión

$$V(\hat{\theta}_{st}) = \frac{B^2}{4} \text{ para el caso en que } \hat{\theta}_{st} = \bar{y} \text{ o } \hat{\theta}_{st} = p, \text{ o}$$

$$V(\hat{\theta}_{st}) = \frac{B^2}{4N^2} \text{ para el caso de } \hat{\theta}_{st} = \hat{\tau}.$$

Una solución es asignar una proporción del tamaño de muestra, w_i , en cada estrato i . Esta fracción entonces nos lleva a que $n_i = nw_i$, $i = 1, 2, \dots, L$. Luego, ya es posible despejar n de las dos expresiones anteriores y, al mismo tiempo, conocer n_i , los tamaños de muestra para cada estrato. Nótese que es necesario asignar previamente las fracciones w_i para cada estrato. Ejercicio (Scheaffer *et al.* 1987, págs. 87 y 88)

Caja 15

$$n = \frac{\sum_{i=1}^L N_i^2 \sigma_i^2 / w_i}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}$$

donde w_i es la fracción de las observaciones asignadas al estrato i , σ_i^2 es la varianza poblacional para el estrato i , cuyo valor se puede aproximar como se indica en la Caja

14, y $D = \frac{B^2}{4}$ para estimar μ ; $D = \frac{B^2}{4N^2}$ para estimar τ .

Tamaño de muestra aproximado que se requiere para estimar m o t con un límite para el error de estimación en el m.a.e.

16. Una empresa publicitaria está interesada en determinar qué tanto debe enfatizar la publicidad televisiva en un determinado municipio, y decide realizar una encuesta por muestreo para estimar el número promedio de horas por semana que se ve televisión en los hogares del municipio. Éste comprende dos pueblos, pueblo A y pueblo B, y un área rural. El pueblo A circunda una fábrica, y la mayoría de los hogares son de trabajadores fabriles con niños en edad escolar. El pueblo B es un suburbio exclusivo de una ciudad vecina y consta de habitantes más viejos con pocos niños en casa. Existen 155 hogares en el pueblo A, 62 en el pueblo B y 93 en el área rural. Una encuesta anterior sugiere que las varianzas del número de horas por semana que las familias del municipio dedican a ver televisión por estrato son, aproximadamente, $\sigma_1^2 = 25$, $\sigma_2^2 = 225$, $\sigma_3^2 = 100$. Deseamos estimar la

media poblacional mediante \bar{y}_{est} . Seleccione el tamaño de muestra para obtener un límite en el error de estimación igual a 2 hrs, si las fracciones asignadas son $w_1 = \frac{1}{3}$, $w_2 = \frac{1}{3}$ y $w_3 = \frac{1}{3}$. En otras palabras, se debe tomar un número igual de observaciones de cada estrato.

Ejercicio (Scheaffer *et al.* 1987, pág. 88)

17. Con relación al Ejercicio 19., supóngase que las varianzas son aproximadas por $\sigma_1^2 = 25$, $\sigma_2^2 = 225$, $\sigma_3^2 = 100$. Deseamos ahora estimar el total poblacional τ con un límite de 400 horas para el error de estimación. Seleccione el tamaño de muestra apropiado, si se va a tomar el mismo número de observaciones en cada estrato.

Caja 16

$$n = \frac{\sum_{i=1}^L N_i^2 \pi_i (1 - \pi_i) / w_i}{N^2 D + \sum_{i=1}^L N_i \pi_i (1 - \pi_i)}$$

donde w_i es la fracción de las observaciones asignadas al estrato i , π_i es la proporción poblacional para el estrato i , cuyo valor se puede aproximar como se indicó en la sección 2.2.1. y $D = \frac{B^2}{4}$. Nota: Como es casi seguro que la proporción poblacional de cada estrato i , π_i , sea desconocida, en las fórmula anterior se cambiarían los términos p_i y $(1-p_i)$ por sus correspondientes estimaciones π_i y $q_i=1-\pi_i$, respectivamente. Este origina que la estimación del tamaño de muestra sea aproximada.

Tamaño de muestra aproximado que requiere para estimar p con un límite para el error de estimación en el m.a.e.

Asignación de la muestra

Después de elegir el tamaño de muestra n , existen muchas maneras para dividir entre los tamaños de muestra de los estratos individuales n_1, n_2, \dots, n_L . Cada división puede originar una varianza diferente para la media muestral. Por lo que el objetivo será usar una asignación de modo que se tengan estimadores con varianzas pequeñas al menor costo posible.

En términos del objetivo, el mejor esquema de asignación está influido por tres factores.

1. El número total de elementos en cada estrato
2. La variabilidad de las observaciones dentro de cada estrato
3. El costo por obtener una observación de cada estrato

Veamos cómo se hace la asignación según estos criterios

Caja 17

$$n_i = n \left(\frac{N_i \sigma_i / \sqrt{c_i}}{N_1 \sigma_1 / \sqrt{c_1} + N_2 \sigma_2 / \sqrt{c_2} + \dots + N_L \sigma_L / \sqrt{c_L}} \right)$$

$$= n \left(\frac{N_i \sigma_i / \sqrt{c_i}}{\sum_{k=1}^L N_k \sigma_k / \sqrt{c_k}} \right)$$

Aquí c_i = costo por obtener una observación individual en el i -ésimo estrato. Para poder usar la fórmula, se debe obtener previamente a la realización del muestreo, una aproximación de la varianza de cada estrato. Las aproximaciones pueden ser obtenidas de estudios anteriores o conociendo la amplitud de variación de las mediciones de cada estrato.

Para la asignación óptima con la varianza de \bar{y}_{est} fija en D , nos da el tamaño de la

muestra n , sustituyendo: $w_i = n_i/n$:
$$n = \frac{\left(\sum_{k=1}^L N_k \sigma_k \quad c_k \right) \left(\sum_{i=1}^L N_i \sigma_i \quad c_i \right)}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}$$

Asignación aproximada que minimiza el costo para el valor fijo de $V(\bar{y}_{est})$ o que minimiza $V(\bar{y}_{est})$ para un costo fijo (Asignación óptima para estimar m y t).

- Para aplicar las fórmulas anteriores se deben conocer, entonces, los costos c_i , $i = 1, 2, \dots, L$, y las desviaciones estándar aproximadas σ_i , $i = 1, 2, \dots, L$.
- Primero se evalúa la expresión para n , después se estiman los n_i 's que tendrán a n como incógnita, y por último se sustituye n en estas expresiones para n_i .

Ejercicio (Scheaffer *et al.* 1987, pág. 114)

18. Una corporación desea obtener información acerca de la efectividad de una máquina comercial. Se va a entrevistar por teléfono a un número de jefes de división, para pedirles que califiquen la maquinaria con base en una escala

numérica. Las divisiones están localizadas en Norteamérica, Europa y Asia. Es por eso que se usa muestreo estratificado. Los costos son mayores para las entrevistas de los jefes de división localizados fuera de Norteamérica. El cuadro 12 proporciona los costos por entrevista, varianzas aproximadas de las calificaciones, y los N_i que se han establecido. La corporación quiere estimar la calificación promedio con $V(\bar{y}_{est}) = 0.1$. Elegir el tamaño de muestra n que obtiene este límite y encontrar la asignación apropiada.

Cuadro 12

Estrato I (Norteamérica)	Estrato II (Europa)	Estrato III (Asia)
$c_1 = \$9$	$c_2 = \$25$	$c_3 = \$36$
$\sigma_1^2 = 2.25$	$\sigma_2^2 = 3.24$	$\sigma_3^2 = 3.24$
$N_1 = 112$	$N_2 = 68$	$N_3 = 39$

Varianzas de las calificaciones de maquinaria para tres divisiones (estratos) de una corporación y costos de entrevista por estrato.

Caja 18

Si el costo por obtener información es el mismo para todos los estratos o si los costos son desconocidos, entonces $c_1 = c_2 = \dots = c_L$. Esto trae como consecuencia que los términos de costo se cancelan en las dos ecuaciones de arriba y:

$$n_i = n \left(\frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right)$$

Este método de seleccionar n_1, n_2, \dots, n_L se denomina *asignación de Neyman*. En la asignación de Neyman, el tamaño de muestra total es:

$$n = \frac{\left(\sum_{i=1}^L N_i \sigma_i \right)^2}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2}$$

Asignación de Neyman para estimar μ y τ

Ejercicio (Scheaffer *et al.* 1987, pág. 115)

19. El estudio del ejercicio 5 se va a hacer anualmente, con el límite para el error de estimación de 5000 acres. Encuentre un tamaño de muestra aproximado para obtener este límite si se usa la asignación de Neyman. Use los datos del Ejercicio 5.

El ejercicio 23 ilustra la aplicación de la asignación del tamaño de muestra a estratos, suponiendo que tanto los costos por estrato como las varianzas por estrato son iguales. Esto da lugar a la asignación proporcional del tamaño de muestra, cuyas fórmulas se presentan en la Caja 19.

Ejercicio (Scheaffer *et al.* 1987, pág. 115)

20. Utilizando los datos del Ejercicio 4, encuentre el tamaño de muestra requerido para estimar la calificación promedio, con un límite de 4 puntos para el error de estimación. Use asignación proporcional.

Caja 19

Si además de suponer $c_1 = c_2 = \dots = c_L$ también $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_L^2$, entonces las ecuaciones para las n_i 's presentadas en la Caja 18, quedan así:

$$n_i = n \left(\frac{N_i}{\sum_{i=1}^L N_i} \right) = n \left(\frac{N_i}{N} \right)$$

Este método de asignación de la muestra a los estratos se llama *asignación proporcional* porque los tamaños de muestra n_1, n_2, \dots, n_L son proporcionales a los tamaños de los estratos N_1, N_2, \dots, N_L . En la asignación proporcional, el valor de n que produce $V(\bar{y}_{est}) = D$, se convierte en:

$$n = \frac{\sum_{i=1}^L N_i \sigma_i^2}{ND + \frac{1}{N} \sum_{i=1}^L N_i \sigma_i^2}$$

Esta última fórmula, aparte de ser válida cuando se quiere estimar μ con $D = \frac{B^2}{4}$, se

usaría cuando interesa estimar τ , pero con el valor de $D = \frac{B^2}{4N^2}$.

Asignación proporcional par estimar μ y τ

La fórmula correspondiente para n cuando interesa estimar π en la asignación proporcional, se presentará en una de la siguiente serie de fórmulas para asignar el tamaño de muestra en cada estrato, para estimar π .

NOTA. Como es casi seguro que la proporción poblacional de cada estrato i , π_i , sea desconocida, en las fórmulas siguientes se han cambiado los términos π_i y $(1-\pi_i)$ por sus correspondientes estimaciones p_i y $q_i=1-p_i$, respectivamente. Este origina que la estimación del tamaño de muestra sea aproximado.

Caja 20

$$n_i = n \left(\frac{N_i \sqrt{p_i q_i / c_i}}{N_1 \sqrt{p_1 q_1 / c_1} + N_2 \sqrt{p_2 q_2 / c_2} + \dots + N_L \sqrt{p_L q_L / c_L}} \right)$$

$$= n \left(\frac{N_i \sqrt{p_i q_i / c_i}}{\sum_{k=1}^L N_k \sqrt{p_k q_k / c_k}} \right)$$

$$\text{donde } n = \frac{\left(\sum_{k=1}^L N_k \sqrt{p_k q_k / c_k} \right) \left(\sum_{i=1}^L N_i \sqrt{p_i q_i c_i} \right)}{N^2 D + \sum_{i=1}^L N_i p_i q_i}$$

Asignación aproximada que minimiza el costo para un valor fijo de $V(p_{est})$ o minimiza $V(p_{est})$ para un costo fijo. (Asignación óptima para estimar π)

Ejercicio (Scheaffer *et al.* 1987, págs. 115-116)

21. Con relación al Ejercicio 6 acerca de la encuesta por muestreo que realiza un ayuntamiento municipal para estimar la proporción de familias con niños discapacitados que utilizarán ciertas instalaciones ampliadas, supóngase que se van a formar cuatro estratos como se indicó en ese Ejercicio, y que aproximadamente el 90% de los que usan las instalaciones y 50% de los que no las usan van a utilizar las nuevas instalaciones. Los costos por efectuar la observación de un cliente actual es de \$4.00 y de \$8.00 para uno que no lo es. La diferencia en el costo resulta de la dificultad para localizar a quienes no usan las instalaciones. Registros existentes nos dan $N_1=97, N_2=43, N_3=145, N_4=68$ (como en el Ejercicio 6). Encuentre el tamaño de muestra aproximado y la asignación necesaria para estimar la proporción poblacional con un límite de 0.05 para el error de estimación.

Caja 21

Cuando el costo por obtener información es el mismo para todos los estratos o los costos son desconocidos, entonces $c_1 = c_2 = \dots = c_L$. En este caso, la asignación de Neyman es:

$$n_i = n \left(\frac{N_i \sqrt{p_i q_i}}{\sum_{k=1}^L N_k \sqrt{p_k q_k}} \right)$$

donde

$$n = \frac{\left(\sum_{i=1}^L N_i \sqrt{p_i q_i} \right)^2}{N^2 D + \sum_{i=1}^L N_i p_i q_i}$$

Asignación de Neyman para estimar π

22. Repita el Ejercicio 24, ahora usando la asignación de Neyman (i.e. costos iguales). Compare los resultados con la respuesta del Ejercicio 24.

Caja 22

Como se recordará, en la asignación proporcional, el tamaño de la muestra en cada estrato i está dado por:

$$n_i = n \left(\frac{N_i}{N} \right)$$

En el caso del tamaño de la muestra n para estimar p , se tiene:

$$n = \frac{\sum_{i=1}^L N_i p_i q_i}{ND + \frac{1}{N} \sum_{i=1}^L N_i p_i q_i}$$

Asignación proporcional para estimar p

Ejercicio

26. Ahora repita el ejercicio 24 usando asignación proporcional. Compare los resultados con los de los Ejercicios 24 y 25.

Selección del tamaño de muestra para estimar medias, totales y proporciones poblacionales en el m.s.

El tamaño de muestra requerido se encuentra despejando n de la ecuación:

$$2\sqrt{V(\hat{\theta}_{ms})} = B$$

para cualquiera de los estimadores $\hat{\theta}_{ms}$ de una media, un total o una proporción, basados en un muestreo sistemático. La solución para esta ecuación involucra, en general al parámetro ρ y, o bien a σ_{ms}^2 , o a π_{ms} . En estas notas, en lugar de usar estimaciones específicas de estos parámetros, usamos la fórmula para n de un muestreo irrestricto aleatorio. La señal de alerta se daría en el caso de poblaciones ordenadas (la muestra podría ser extragrande) o de poblaciones periódicas (la muestra podría ser muy pequeña). Por tanto, la validez de las fórmulas para n , como si se supusiera un muestreo irrestricto aleatorio, solamente se presentará si se muestrean sistemáticamente poblaciones aleatorias. En este caso, remitimos al lector a las fórmulas que aparecen en la sección 2.1.2., Cajas 1, 2 y 3.

Ejercicio (Scheaffer *et al.* 1987, pág. 189)

27. La patrulla de caminos está interesada en la proporción de automovilistas que portan licencias. En una carretera principal se detuvo a cada séptimo conductor. En una verificación similar con 400 automovilistas, se encontró que el 81% portaban licencia. La patrulla de caminos espera que pasen cuando menos $N = 3000$ automóviles por el puesto de verificación. Determinar el tamaño de muestra para estimar p con aproximación de $B = 0.015$ unidades.

Selección del tamaño de muestra para estimar medias, totales y proporciones poblacionales en el m.c.

La cantidad de información en una muestra por conglomerados es afectada por dos factores, el número y el tamaño relativo de los conglomerados. No se ha presentado el último factor en ninguno de los procedimientos de muestreo ya analizados. El tamaño del límite para el error de estimación depende crucialmente de la variación entre los *totales* de los conglomerados. Entonces, al intentar obtener límites pequeños para el error de estimación, debemos seleccionar conglomerados con la menor variación posible entre estos totales. Suponer que el tamaño del conglomerado (unidad de muestreo) ha sido elegido y se va a considerar únicamente el problema de seleccionar el número de conglomerados, n .

Caja 23

$$n = \frac{N\sigma_{mc}^2}{ND + \sigma_{mc}^2}$$

donde σ_{ms}^2 se estima con $s_{mc}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_{mc} m_i)^2}{n-1}$. Esta estimación estaría disponible con una encuesta previa, o bien seleccionado una muestra preliminar de n' elementos. Algo similar podría hacerse para estimar, el tamaño medio del conglomerado. Así, las estimaciones de σ_{ms}^2 y pueden calcularse de la muestra preliminar y utilizarse para obtener un tamaño de muestra total aproximado, n , dado por la fórmula de arriba. Aquí. $D = \frac{B^2 \bar{M}^2}{4}$

Tamaño de muestra aproximado para estimar m con un límite B para el error de estimación

Se obtienen resultados similares cuando se usa $M\bar{y}_{mc}$ para estimar τ .

Caja 24

$$n = \frac{N\sigma_{mc}^2}{ND + \sigma_{mc}^2}$$

donde σ_{mc}^2 se estima con $s_{mc}^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_{mc} m_i)^2}{n-1}$. σ_{mc}^2 se estimaría con una encuesta previa, o bien seleccionado una muestra preliminar de n' elementos. Entonces, la estimación puede utilizarse para obtener un tamaño de muestra total aproximado, n , dado por la fórmula de arriba. Aquí se supone $D = \frac{B^2}{4N^2}$.

Tamaño de muestra aproximado para estimar τ , usando $M\bar{y}_{mc}$, con un límite B para el error de estimación, M conocida

Cuando M es desconocido, se usa el estimador $N\bar{y}_i$ para τ (ver sección 3.1.6). En este caso, se puede despejar n de la ecuación $2\sqrt{V(N\bar{y}_i)} = B$, en donde $V(N\bar{y}_i) = N^2V(\bar{y}_i) = N^2((N - n/N)\sigma_i^2)$. Aquí σ_i^2 es la varianza poblacional de los totales de conglomerados, y_i , la cual se estima con:

$$s_1^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_t)^2}{n-1}$$

Caja 25

$$n = \frac{N\sigma_t^2}{ND + \sigma_t^2}$$

donde σ_t^2 se estima con s_1^2 , y $D = \frac{B^2}{4N^2}$

Tamaño de muestra aproximado para estimar τ , usando con un límite B para el error de estimación

De manera análoga a como se deduce la fórmula para n requerida para estimar m , se obtendría la fórmula correspondiente cuando se desea estimar p .

Caja 26

$$n = \frac{N\sigma_{mc}^2}{ND + \sigma_{mc}^2}$$

donde $D = \frac{B^2 \bar{M}^2}{4}$, y σ_{mc}^2 se estima por $s_{mc}^2 = \frac{\sum_{i=1}^n (a_i - p_{mc} m_i)^2}{n-1}$ (sección 1.1.5).

Tamaño de muestra aproximado para estimar π con un límite B para el error de estimación

Ejercicio (Scheaffer *et al.* 1987, pág. 189)

28. La industria del Ejercicio 15 modificó su política de jubilación después de obtener los resultados de la encuesta. Ahora se quiere estimar la proporción de empleados a favor de la política modificada. ¿Cuántas plantas deben ser muestreadas para tener un límite de 0.08 para el error de estimación? Use los datos del Ejercicio 15 para aproximar los resultados de la nueva encuesta.

Referencias

- Azorín Poch, F. (1972) *Curso de Muestreo y Aplicaciones*. Aguilar, Madrid.
- Cochran, W.G. (1976) *Técnicas de Muestreo*. CECSA, México.
- Iachan, R. (1985) Optimum strata boundaries for shellfish surveys. *Biometrics* 41: 1053–1062.
- Krebs, C. (1989) *Ecological Methodology*. Harper Collins Publishers, New York.
- Manly, B.F.J. (1992) *The Design and Analysis of Research Studies*. Cambridge University Press, Cambridge.
- Milne, A. (1959) The centric systematic area-sample treated as a random sample. *Biometrics* 15: 270-297.
- Scheaffer, R. L., W. Mendenhall y L. Ott (1987) *Elementary Survey Sampling*. Duxbury, Belmont.
- Thompson, S. (1992) *Sampling*. Wiley, Nueva York.

II. EL AMBIENTE

Se incluyen los lineamientos del muestreo de recursos no renovables (suelo y agua) con los que los manejadores de recursos han de contender, y de otras herramientas ambientales, como el clima o monolitos de suelo, con que cuentan como elemento de juicio.

La importancia del suelo a nivel mundial ha sido recientemente revalorada, ya que la edafósfera es una capa más del planeta tierra. Esta capa puede considerarse como una geomembrana que permite el intercambio de materia y energía entre la litosfera y la atmósfera, sirviendo de protección a la litósfera ante los efectos climáticos. Además, se le califica como un recurso natural no renovable debido a sus bajas tasas de formación. En los ecosistemas terrestres, es soporte y fuente de nutrimentos para los microorganismos, invertebrados y plantas. Su importancia a nivel agronómico es ampliamente reconocida.

En la esfera ambiental, los recursos hídricos están llamados a ser el gran problema de la humanidad para el presente siglo. Es por esto que el tema “cuerpos de agua superficiales” es de suma importancia, tanto por el consumo de grandes cantidades de este recurso por las comunidades humanas a nivel doméstico y por la industria, como por la descarga de aguas residuales, contaminadas en diversos grados, a los cuerpos de agua.

La toma de datos climáticos no es menos importante. Este aspecto tiene al menos dos vertientes de relevancia para los manejadores de recursos naturales: el clima como factor que puede limitar u orientar la distribución y abundancia de los seres vivos, y el clima como elemento cambiante (calentamiento global) que puede modificar estos patrones y que debe ser considerado en la toma de decisiones en materia ambiental.