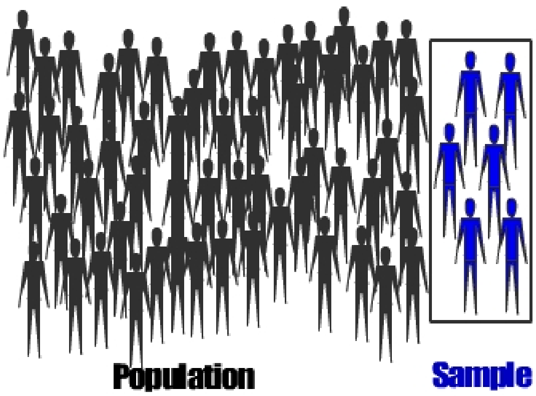
Estrategias de muestreo



Autor: **Prof. Christian Salas Eljatib**

E-mail: **cseljatib@gmail.com**

Índice

Estrategias de muestreo

Con esta unidad buscamos:

proveer un entendimiento sobre el rol de muestreo probabilístico en hacer inferencia cientifífica valida sobre parámetros descriptivos de una población.

1. Población y muestra

Una población es la membresía total o "universo" de una definida clase de gente, objetos, o eventos.

o más simplemente

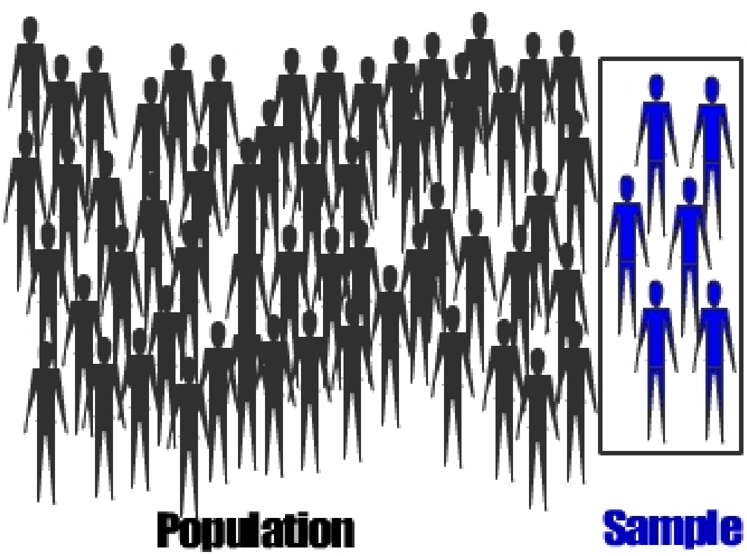
"La población es el set completo de valores o atributos de interés"
Algunos otros conceptos relacionados

- La población objetivo es aquella población que nosotros hemos definido sobre la cual queremos inferir o entender ciertas relaciones
- Cada elemento de la población puede ser indexado por i, donde $i = 1, 2, ..., N$, siendo N el tamaño poblacional.
- $\wp$ representa a la población consiste de N unidades, las cuales pueden denotarse por $U_1, U_1, U_3, ..., U_N$
- Las unidades son también referidas "como elementos".

- Asociada con la i-ésima unidad, U_i, esta el valor de alguna característica (o variable), y_i
- La única forma de conocer una población es realizar un censo, en donde cada elemento de la población es conocido.

La muestra es una porción de la población.

- La muestra es seleccionada mediante algún procedimiento estadístico
- Cada elemento de la muestra puede también ser indexado por i, donde $i = 1, 2, ..., n$, siendo n el tamaño muestral.



2. Algunos ejemplos de contextos de muestreo

a. Muestreos por cuestionarios, por ejemplo: Una muy pequeña fracción de la población es seleccionada y se le pregunta sobre cual es su candidato político favorito, o que detergente prefiere, o actitud sobre política exterior o creencias religiosas.

b. Muestreo que involucre ubicacion de parcelas en terreno y medir variables o plantas u otras "cosas" que esten dentro de cada parcela.

> **Pregunta**:
> *¿Son las anteriores una muestra necesariamente estadística o científica?*
>
> No! Esto depende de como los elementos son seleccionados para estar en la muestra y luego, que se hace con ellos para estimar las características de la población de interes.

c. muestreo de industrias para entregar datos sobre la cual podriamos estimar, e.g., (i) % de mujeres que trabaja, o (ii) sueldo por clase de edad y genero.

d. evaluar la efectividad de iniciativas educacionales, programas de gobierno o de ONGs.

e. muestreo de pesca de arrastre para estimar el tamaño, composición, y extensión geográfica de peces. Ejemplo el “Maine - New Hampshire Inshore Trawl Survey” https://www.maine.gov/dmr/science-research/projects/trawlsurvey/index.html

f. muestreo de marcado/re-marcado para estimar el tamaño de animales terrestres. Esta técnica también se ha extendido a otros contextos.

g. inventarios forestales para estimar el stock y estatus temporal de recursos forestales.

h. inventarios de recreación para estimar el número de hikers y largo y calidad de rutas.

i. muestrear cursos de agua para estimar temperatura media del agua, o cantidad de sedimentos, compuestos contaminantes.

j. muestreo por imagenes satelitales o laser para estimar la biomasa aerea o superficie impermeable sobre grandes regiones

k. muestrear el suelo de un desierto para estimar la abundancia de madera petrificada

l. muestrear la profundidad de una placa de hielo para estimar el volumen de nieve que cubre una cuenca durante el invierno, o el peso de un glaciar.

¿Por qué muestrear y no medir todas las unidades de la población?

recordemos que

- Normalmente la fracción

$$\frac{n}{N} \tag{1}$$

  es muy baja. Eq (1) se conoce como la *intensidad de muestro.*

- Cuando $n = N$, tenemos un censo.

Ejemplo en el contexto de muestreo de bosques:
Si un rodal tiene 8 hectáreas, y estuvieramos planeando en muestrearlo con parcelas de muestreo que tienen una superficie (representado como a) de 500 m2, si pudieramos cubrir el rodal completo, necesitaríamos 160 parcelas, es decir $N = 160$. ¿Cómo calculamos lo anterior?

- Primero, transformemos la superficie de la parcela en hectareas como sigue

$$\frac{a}{10000} = \frac{500}{10000} = 0{,}05 \text{ ha} \tag{2}$$

- Segundo, determinemos cuantas parcelas de dicha superficie (en hectareas) caben en un rodal de 8 ha

$$\frac{8}{0{,}05} = 160 \text{ parcelas} \tag{3}$$

a. costo: al medir una fracción de la población se gastara menos en identificar, ubicar, medir cada unidad en la población de interes;

b. tiempo: los resultados se pueden compilar y publicar mas rápidamente;

c. practicabilidad: el esfuerzo de gente necesaria y de equipos es mucho menos;

d. el error de medición podría ser considerablmente menor (identificación, ubicación, transcripción).

para pensar...:

"Muestreo no es tan sólo la simple sustitución de una cobertura parcial por un cobertura total. Muestreo es la ciencia y el arte de controlar y medir la seguridad (confiabilidad) de información estadística útil a través de la teoría de probabilidades."

3. Muestreo probabilístico

Características esenciales de una muestra probabilística

- probabilidad deducible de seleccionar cada posible muestra
- la probabilidad relacionada de incluir cada elemento de la población en una o más muestras
- el contexto de muestreo desde el cual los elementos son seleccionados para una muestra

Interés en estimar...

- el valor promedio de una (o más) características que poseen cada elemento (unidad) de la población;
- la cantidad total de "algo" en la población (e.g., peso total de manzanas cosechadas, volumen de madera de un bosque)
- la proporción de la población que tiene una cierta característica, e.g., la proporción de ranas de lagos con deformaciones físicas; o

- la relación entre dos características, e.g., el cuociente entre la superficie de follaje y el peso del follaje de un vegetal.

3.1. Paramétro y estadístico: recapitulando

El promedio, total, o proporción de algo para la población es un parámetro, o un parámetro descriptivo de la población.

- Supongamos que conocemos la variable y para cada uno de los i-elementos de una población de tamaño N. Un **parámetro** es una constante que describe a una población.
- Un parámetro puede ser obtenido (o calculado) sólo si todos los elementos de la población son empleados.
- **Parámetros son siempre simbolizados con letras griegas**, como por ejemplo; $\alpha, \beta, \gamma, \mu, \tau, \phi, \sigma, \theta$, etc.

  - El total poblacional para la variable y se obtiene mediante

    $$\tau_y = \sum_{i=1}^{N} y_i \tag{4}$$

    por lo tanto, τ_y es un parámetro.
  - El promedio poblacional para la variable y se obtiene mediante

    $$\mu_y = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{5}$$

por lo tanto, μ_y es un parámetro. Note que

$$\mu_y = \frac{\tau_y}{N} \tag{6}$$

Ejercicio:

Asumiendo la siguiente población de valores de altura de árboles (medida en m)

`25.7 30 53.3 33.9 34.7`

Conteste a las siguientes preguntas

- ¿Cual es el tamaño de la población?
- Calcule el valor del parámetro descrito en Ec. (4). ¿Qué nombre recibe este parámetro?
- Calcule el valor del parámetro descrito en Ec. (5). ¿Qué nombre recibe este parámetro?

- Supongamos que conocemos los valores de la variable y para una muestra de la población, dicha muestra es de tamaño n. Un **estadístico** es un número que describe a una muestra.
- Un estadístico puede ser obtenido (o calculado) sólo si una muestra de los elementos de la población son empleados.
- El promedio muestral para la variable y se obtiene mediante

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{7}$$

por lo tanto, $\bar{y}$ es un estadístico.

Ejercicio:
Asumiendo la misma población dada anteriormente, seleccione una muestra aleatoria de tres elementos de la población y conteste a las siguientes preguntas:

- ¿Cual es la intensidad de muestreo (Ec. 1)?
- Calcule el valor del estadístico descrito en Ec. (7). ¿Qué nombre recibe este estadístico?
- Calcule la diferencia entre el parámetro descrito en Ec. (5) y el valor del estadístico ¿A que se podría deber esa diferencia? ¿Podría indicar si esta diferencia es alta/baja? Justifique.

4. Diseño de muestreo, estimación e inferencia

La "trilogía" del muestreo:

Diseño de muestreo, estimación e inferencia

- Seleccionar la muestra es importante, pero debe ser seguido de mediciones y luego análisis
- recordemos que la estimación es tratar de "adivinar" el valor de los parámetros que describen a la población en estudio
- hay muchas formas para diseñar y ejecutar una muestra, esto es, seleccionar elementos probabilísticamente en una muestra. (nosotros cubriremos sólo un par)
- hay muchas formas (innumerables) para estimar los parámetros poblacionales empleando datos obtenidos en una muestra en particular.
- Estimación es fácil (e.g., 16), pero estimar parámetros con confiabilidad y eficientemente es más díficil.
- Inferencia va un paso más adelante que estimación de un parámetro con tal de responder la pregunta: ¿Cuan bueno es el estimador que tengo?

5. Muestra y probabilidad de los elementos

- En muestreo probabilístico el diseño de muestreo determina la probabilidad de inclusión de cada elemento de la población.

- la probabilidad de inclusión la simbolizaremos por π_i, siguiendo la simbología que emplea Gregoire y Valentine (2008) en su libro de muestreo.

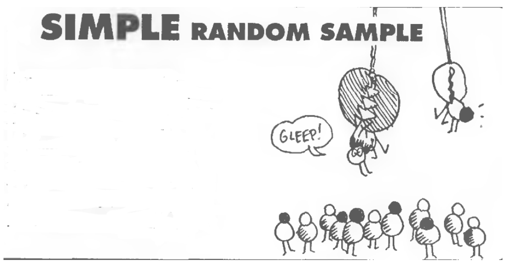
  $\pi_i =$ probabilidad (U_i es incluida en la muestra **s** seleccionada con el diseño de muestreo estipulado)

  $\pi_i = Prob(U_i \in \mathbf{s})$

- Para diseños de muestreo de igual probabilidad (“*equal sampling probability*”), las probabilidades de inclusión son iguales para todos los elementos de la población. Lo contrario ocurre para los diseños de muestreo de diferente probabilidad (“*unequal sampling probability*”).

- Para una misma población, π_i puede tomar valores diferentes bajo diferentes diseños.

6. Muestreo aleatorio simple (sin reemplazo)



- Se denota por SRSwoR ("*Simple random sampling without replacement*") o SRS
- En este diseño, π_i son iguales para todos los elementos
- Bajo SRSwoR, el número total de muestras (Ω) de tamaño fijo de n se determina de acuerdo a la formula de combinaciones,

$$\Omega = \binom{N}{n} = \frac{N!}{(N-n)!\, n!} \tag{8}$$

  donde, n es el tamaño muestral y N es el tamaño poblacional. Por ejemplo, si $N = 5$ y $n = 3$

$$\Omega = \binom{5}{3} = \frac{5!}{(5-3)!\, 3!} = 10$$

  hay 10 posibles (i.e., diferentes) combinaciones en donde 3 elementos pueden ser asignados a una muestra.

Ejercicio:

Enumere todas las posibles configuraciones de muestras que se podrían formar para un tamaño poblacional de 5 elementos, si el tamaño muestral es 3. Por ejemplo, una primera muestra potencial es la cual donde los tres primeros elementos de la población son seleccionados. Es decir,

Muestra No. 1: elementos 1, 2, 3

Continua tu entonces describiendo todas las posibles restantes muestras

Muestra No. 2: elementos ...

$\vdots \qquad \vdots$

Muestra No. Ω: elementos ...

- El ejercicio anterior de enumerar y describir todas las posibles muestras sólo es "realizable" para Ω pequeños, pues Ω tiende a ser muy grande aun para moderados tamaños poblacionales. Por ejemplo: si $N = 100$ y $n = 10$ con SRSwoR, y empleando (8), tenemos que

$$\Omega = \binom{100}{10} = \frac{100!}{(100-10)!\;10!} = 17{,}310{,}309{,}456{,}440 \tag{9}$$

  Entonces, existen más de 17 trillones de posibles muestras distintas de una población con tamaño de población 100 y con un tamaño muestral de 10 unidades.

- La probabilidad de seleccionar una particular muestra es

$$p(s) = \frac{1}{\Omega} \tag{10}$$

- La probabilidad de inclusión de una unidad (o elemento), U_i, es

$$\pi_i = \frac{n}{N} \tag{11}$$

  este resultado proviene de

$$\pi_i = \frac{\text{Numero de muestras que incluyen } U_i}{\Omega} \tag{12}$$

6.1. Paramétros de estimadores vs. estimadores de los parámetros estimados: ojo con el "trabalengua"

Es importante entender la diferencia entre un parámetro y un estimador de un parámetro, ya que el primero es una constante, mientras que el segundo es una variable aleatoria.

Por ejemplo, la varianza de la variable aleatoria y se calcula mediante

$$\sigma_y^2 = \frac{\sum_{i=1}^{N} \left(y_i - \mu_y\right)^2}{N} \tag{13}$$

donde: N es el tamaño de la población; la sumatoria de las diferencias cuadráticas (numerador de la expresión anterior) se realiza con todos los elementos de la población; y

$$\mu_y = \frac{1}{N} \sum_{i=1}^{N} y_i, \tag{14}$$

es la media aritmética de la población. Note entonces que μ_y y σ_y^2 son parámetros.

Algo muy diferente es la siguiente expresión que permite estimar la varianza de y a partir de una muestra aleatoria de tamaño n, como sigue debemos calcular la desviación estándar de la variable y en la muestra, definida como

$$\widehat{\sigma}_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}, \tag{15}$$

donde

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \tag{16}$$

es la media aritmética de la muestra. Note entonces que $\bar{y}$ y $\widehat{\sigma}_y^2$ son estimadores de parámetros, y por lo tanto son variables aleatorias, a igual que y.

El estimador $\bar{y}$, que podemos reemplazar por $\widehat{\mu}_y$, al ser una variable aleatoria, es posible entonces calcular su variabilidad. La varianza del estimador $\widehat{\mu}_y$, es decir $\mathrm{V}[\widehat{\mu}_y]$ se puede calcular mediante

$$\mathrm{V}\left[\widehat{\mu}_y\right] = \mathrm{V}\left[\bar{y}\right] = \frac{\sigma_y^2}{n}\left(1 - \frac{n}{N}\right), \tag{17}$$

donde σ_y^2 es el parámetro de la varianza de la variable aleatoria y (Ec. 13)

Es importante destacar que a partir de la expresión anterior, si se podemos conocer el valor de la varianza de la población y el tamaño poblacional (N), es posible determinar en cuanto se puede disminuir la varianza de mi estimador, al aumentar el tamaño muestral (n).

Continuando con las ideas expresadas hasta acá, el estimador de la varianza del estimador $\widehat{\mu}_y$, es decir $\widehat{\mathrm{V}}[\widehat{\mu}_y]$ se puede calcular mediante

$$\widehat{\mathrm{V}}[\widehat{\mu}_y] = \widehat{\mathrm{V}}\left[\bar{y}\right] = \frac{\widehat{\sigma}_y^2}{n}\left(1 - \frac{n}{N}\right), \tag{18}$$

donde $\widehat{\sigma}_y^2$ es el estimador del parámetro de la varianza de la variable aleatoria y (Ec. 15)

El error estándar del estimador $\widehat{\mu}_y$ se obtiene simplemente como la raíz cuadrada de la varianza de la Ec. (18)

$$\mathrm{SE}(\widehat{\mu}_y) = \sqrt{\mathrm{V}(\widehat{\mu}_y)}. \tag{19}$$

El error estándar porcentual, o error relativo, del estimador $\widehat{\mu}_y$ se obtiene mediante

$$\mathrm{SE}(\widehat{\mu}_y)(\,\%) = 100\frac{\mathrm{SE}(\widehat{\mu}_y)}{\widehat{\mu}_y} \tag{20}$$

6.2. Ejemplo de muestreo aleatorio simple

El archivo `pelicanEgg.csv` contiene los datos de 92 huevos de pelicanos cafes. Dos variables fueron medidas para cada cascara de huevo en este archivo. Uno es la concentración de PCB en ppm, y el otro es el grosor de la cascara. El parámetro de interes para este ejemplo es la media poblacional (μ_y) de la variable PCB, por lo tanto la variable aleatoria y es PCB. Como estimador del parámetro mu_y emplearemos $\bar{y}$, i.e., media aritmética de la muestra, definido como

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{21}$$

Por lo tanto $\widehat{\mu}_y = \bar{y}$

En este ejemplo, seleccionaremos mediante SRSwoR una muestra de $n = 10$ huevos a partir de esta "población" de 92 hueveos, y luego estimar en base a esa muestra el valor del parámetro de interes.

a. La intensidad de muestreo es

$$f = \frac{n}{N} = \frac{10}{92} = 0{,}1086 \tag{22}$$

lo cuales es 10.9 %.

b. En el cuadro 1, se detalla la muestra seleccionada aleatoriamente para ambas variables, aunque nos centraremos en la variable PCB

Elemento en la muestra	Huevo (No.)	**PCB**	**Thickness**
1	30	212	0.37
2	49	236	0.47
3	62	144	0.39
4	87	110	0.4
5	4	175	0.24
6	55	265	0.29
7	74	270	0.29
8	63	232	0.41
9	48	185	0.42
10	84	530	0.39

Cuadro 1: Muestra aleatoria de 10 elementos de la población descrita en el archivo `pelicanEgg.csv`

c. Ahora calculamos el estimador del parámetro de interes, mediante Eq. (21) como sigue,

$$\bar{y} = \frac{1}{10}(212 + 236 + ... + 530) = 235{,}9 \tag{23}$$

d. $\widehat{\mathrm{SE}}\left[\widehat{\mu}_y\right]$: Error estándar estimado de $\widehat{\mu}_y$

Para estimar el error estándar de $\widehat{\mu}_y$, primero debemos calcular la desviación estándar de la variable y en la muestra, definida como

$$\widehat{\sigma}_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}} \tag{24}$$

Notesé que $\widehat{\sigma}_y$ también a veces se representa por S_y. La desviación estándar se calcula como sigue

$$\widehat{\sigma}_y = \sqrt{\frac{(212 - 235{,}9)^2 + ... + (530 - 235{,}9)^2}{10-1}} = 115{,}32 \tag{25}$$

Ahora podemos calcular el error estándar estimado del estimador $\bar{y}$, el cual representamos por $\widehat{\mathrm{SE}}\left(\bar{y}\right)$, empleando la siguiente expresión

$$\begin{aligned}\widehat{\mathrm{SE}}\left[\bar{y}\right] &= \sqrt{\frac{\widehat{\sigma}_y^2}{n}\left(1-f\right)} \\ &= \sqrt{\frac{\widehat{\sigma}_y^2}{n}\left(1-\frac{n}{N}\right)}\end{aligned} \tag{26}$$

Notesé que $\widehat{\mathrm{SE}}\left[\bar{y}\right]$ también a veces se representa por $S_{\bar{y}}$. Ahora apliquemos (26) como sigue

$$\widehat{\mathrm{SE}}\left[\bar{y}\right] = \sqrt{\frac{13298{,}5}{10}\left(1-0{,}1086\right)} = 34{,}4 \tag{27}$$

e. Error de muestreo o margen de error para un 90 % de confianza estadística.

El error de muestreo (EM) se calcula como sigue

$$\text{EM}_{(\bar{y},\alpha)} = \widehat{\text{SE}}\,[\bar{y}]\, t_{(1-\frac{\alpha}{2},n-1)}, \tag{28}$$

donde $t_{(1-\frac{\alpha}{2},n-1)}$ es el valor de la distribución de t-student empleando un nivel de significancia α. Para obtener el valor de t podemos ocupar R o una tabla de valores para la distribución de t-student

```
> n <- 10
> df <- n - 1
> conf <- 90
> alpha <- 1 - (conf/100)
> alpha.2 <- alpha/2
> t.value <- abs(qt(1 - alpha.2,
     df))
> t.value

[1] 1.8331
```

Entonces $t_{(1-\frac{0,1}{2},10-1)} = 1{,}8331$. Ahora podemos aplicar (28) y calcular el error de muestreo

$$\text{EM}_{(\bar{y},0,1)} = 34{,}43 \times 1{,}8331 = 63{,}11 \tag{29}$$

este valor se expresa en las mismas unidades que la variable aleatoria y. Este error de muestreo se puede expresar en porcentaje al dividirlo por la media aritmética, como sigue

$$\text{EM}_{(\bar{y},0,1)}\,\% = 100 \times \frac{63{,}111}{235{,}9} = 26{,}8\,\% \tag{30}$$

f. Estimar un intervalo de confianza al 90 % estadístico para el estimador $\widehat{\mu}_y$.

Un intervalo de confianza para $\widehat{\mu}_y$ se calcula como sigue

$$\bar{y} \pm \widehat{\text{SE}}\,[\bar{y}]\, t_{(1-\frac{\alpha}{2},n-1)} \tag{31}$$

$$\pm\, EM \tag{32}$$

lo cual es

$$235{,}9 \pm 1{,}833 \times 34{,}4 \tag{33}$$

$$\pm\, 63{,}111 \tag{34}$$

$$[172{,}8\ ;\ 299{,}0], \tag{35}$$

Ejercicio:

- Para la muestra aleatoria anterior (cuadro 1), calcule el error de muestreo y el intervalo de confianza, pero ahora utilizando un nivel de confianza estadístico de 95 % y 99 %. Posteriormente, compare los tres intervalos confidenciales calculados y responda a las siguientes preguntas: ¿Cual de ellos es el más amplio, el más reducido?, ¿En cuál de ellos tienes más confianza?, ¿Cuál de ellos es más preciso?.
- Si sólo seleccionas las primeras cinco unidades de muestra del cuadro (1), calcule el error de muestreo y el intervalo de confianza, para un nivel de confianza estadístico de 90 %, 95 % y 99 %. Compara tus resultados con los obtenidos en la pregunta anterior.

6.3. Estimación del total τ_y

- Hasta ahora hemos centrado nuestro interes en estimar el parámetro poblacional μ_y, así como en la estimación de su variabilidad
- Sin embargo, también podriamos estar interesados en estimar τ_y
- Tal como detalladamente lo explican Gregoire y Valentine (2008), un estimador insesgado del parámetro del total de y τ_y es el denominado estimador de Horvitz-Thompson o "HT", el cual corresponde a

$$\widehat{\tau}_y = \sum_{i}^{n} \frac{y_i}{\pi_i} \tag{36}$$

el cual para el muestreo aleatorio simple se transforma en

$$\begin{aligned} \widehat{\tau}_y &= \frac{N}{n} \sum_{i}^{n} y_i \\ &= N\bar{y} \end{aligned} \tag{37}$$

- La varianza del estimador $\widehat{\tau}_y$ se calcula mediante

$$\mathrm{V}[\widehat{\tau}_y] = N^2 \mathrm{V}[\widehat{\mu}_y], \tag{38}$$

donde note que $\mathrm{V}[\widehat{\mu}_y]$ es el parámetro de la varianza del estimador $\widehat{\mu}_y$.

En caso de no conocer la población (lo cual es el caso real), podemos utilizar el valor estimado de $\widehat{\mathrm{V}}[\widehat{\mu}_y]$, y de esa forma poder estimar el valor del parámetro $\mathrm{V}[\widehat{\tau}_y]$. Para lo anterior, se puede utilizar (26) y reemplazarlo en (38), obteniendo la siguiente expresión

$$\widehat{\mathrm{V}}[\widehat{\tau}_y] = N^2 \left(\widehat{\mathrm{SE}}\left[\widehat{\mu}_y\right]\right)^2. \tag{39}$$

- Dado lo anterior, el error estándar estimado del estimador $\widehat{\tau}_y$ sería entonces

$$\widehat{\mathrm{SE}}\left[\widehat{\tau}_y\right] = N \frac{\widehat{\sigma}_y}{\sqrt{n}} \sqrt{\left(1 - \frac{n}{N}\right)} \tag{40}$$

- Un intervalo de confianza para $\widehat{\tau}_y$ se calcula como sigue

$$\widehat{\tau}_y \pm \widehat{\mathrm{SE}}\left[\widehat{\tau}_y\right] t_{(1-\frac{\alpha}{2}, n-1)} \tag{41}$$

$$\pm \mathrm{EM}_{(\widehat{\tau}_y, \alpha)} \tag{42}$$

donde, $\mathrm{EM}_{(\widehat{\tau}_y, \alpha)}$ es el error de muestreo del estimador $\widehat{\tau}_y$ con un nivel de significancia α

Ejemplo de estimador del total poblacional:

Para monitorear el uso de agua semanal en una ciudad, la municipalidad realizó un muestreo aleatorio simple de 100 casas ($n = 100$). Esta muestra fue seleccionada a partir de los registros de la ciudad que indican que existen 5392 propiedades con medidores de agua potable ($N = 5392$).

Producto del muestreo, se calculó un consumo medio semanal de 4,5 m^3 ($\bar{y} = 4{,}5$), mientras que la varianza muestral fue de 170,3 $\left(m^3\right)^2$ ($\widehat{\sigma}_y^2 = 170{,}3$).

Usando Ec. (26) se obtiene el estimador del error estándar como sigue

$$\widehat{\mathrm{SE}}\left[\bar{y}\right] = \sqrt{\frac{170{,}3}{10}\left(1 - \frac{100}{5392}\right)} = 1{,}3 \text{ m}^3 \tag{43}$$

Ahora, nos interesaría expandir estos resultados a nivel total, es decir, para toda la ciudad. Empleamos entonces la Ec. 37 como sigue

$$\begin{aligned} \widehat{\tau}_y &= 5392\,(4{,}5) \\ &= 24264 \text{ m}^3 \end{aligned} \tag{44}$$

Este valor note que es para toda la ciudad, es decir, el consumo semanal de la ciudad.

Usando Ec. (39), se tiene

$$\widehat{\mathrm{V}}[\widehat{\tau}_y] = 5392^2\,(1{,}3)^2 \tag{45}$$

$$= 48594192{,}192 \tag{46}$$

y por lo tanto el error estándar del estimador del total es

$$\widehat{\mathrm{SE}}\,[\widehat{\tau}_y] = \sqrt{\widehat{\mathrm{V}}[\widehat{\tau}_y]} = 6971 \text{ m}^3 \tag{47}$$

El intervalo de confianza para $\widehat{\tau}_y$ (Ec. 41) es

$$24264 \pm 6971\,(1{,}8331)$$

$$[11485{,}5\ ;\ 37042{,}5] \tag{48}$$

Referencias

Gregoire TG, HT Valentine. 2008. Sampling Strategies for Natural Resources and the Environment. New York, USA. Chapman & Hall/CRC. 474 p.