Revista Contacto / ISSN L 2710-7620
Volumen 1, Número 2 / septiembre – diciembre de 2021
Páginas: 75 - 87
Recibido: 25/07/2021 / **Aceptado:** 29/08/2021
Enlace: https://www.revistas.up.ac.pa/index.php/contacto

Fuentes de variabilidad en el muestreo de encuestas económicas y los operadores esperanza matemática y varianza

Sources of variability in sampling economic surveys and operators mathematical hope and variance

René Isaac Bracho Rivera[1]

Universidad de Panamá, Centro Regional Universitario de San Miguelito, Facultad de Economía, Panamá

https://orcid.org/0000-0002-3247-2075

rene.bracho@up.ac.pa

RESUMEN

El siguiente artículo es de revisión y se presentan las nociones generales del muestreo de encuestas, sus implicaciones teóricas y las fuentes de variabilidad que tiene efectos sobre las estimaciones estadísticas derivadas del muestreo. Se mencionan las sutiles diferencias conceptuales inherentes a las poblaciones hipotéticas infinitas y las poblaciones finitas junto a las raíces históricas de la inferencia estadística clásica. Adicionalmente, se exponen las definiciones y propiedades de los operadores esperanza matemática y varianza junto a su utilidad como herramientas de evaluación de los estimadores.

Palabras Clave: Muestreo, variabilidad, aleatorización, encuesta, población

ABSTRACT

The following article is a review article and presents the general notions of survey sampling, its theoretical implications, and the sources of variability that have effects on statistical estimates derived from sampling. The subtle conceptual differences inherent in hypothetical infinite populations and finite populations are mentioned along with the historical roots of

[1] Profesor universitario y economista. Doctorando en Estadística por la Universidad de La Habana, Máster en Operaciones y Tecnología INCAE Business School y Máster en Estadística Económica Universidad de Panamá.

classical statistical inference. Additionally, the definitions and properties of the mathematical expectation and variance operators are presented along with their usefulness as tools for evaluating the estimators.

Key Words: *Sampling, variability, randomization, survey, population*

Introducción

Una característica del muestreo de encuestas es que la construcción de sus formulaciones teóricas ha transitado un camino que va desde el escepticismo hasta alcanzar un estatus riguroso como programa de investigación en la ciencia estadística. Las aplicaciones que aporta este campo son de gran utilidad en la práctica de la investigación científica económica, social y humana, no obstante, sus resultados en ocasiones son incompletos, parciales, fundamentados en supuestos taxativos y condiciones idealizadas.

Debido al hecho que el proceso de realización de encuestas requiere el cumplimiento de varias actividades y operaciones, este conjunto de operaciones se convierte en fuente de errores, dichos errores se traducen en la variabilidad de los estimadores. Se puede decir, que la teoría del muestreo de encuestas busca minimizar los errores derivados de cada operación del proceso de aplicación de encuestas, mientras que la teoría de la estimación es la rama de la inferencia estadística que estudia organizadamente la variabilidad de los estimadores. (Sarndal, Carl., Swesson, Bengt., & Wretman, Jan., 1992).

En este artículo de revisión se abordan nociones elementales de ambos programas de investigación de la ciencia estadística. Por un lado, se presentan algunos conceptos básicos de muestreo de encuestas, en segundo lugar, se exponen los operadores esperanza matemática y varianza junto a su rol en la construcción de estimadores para muestreos de encuestas.

El proceso de encuestar debe ser entendido como una secuencia de operaciones con el objetivo de recolectar datos entrevistando a una persona. Mientras que los operadores esperanza matemática y varianza son herramientas teóricas de la estimación estadística para evaluar las propiedades cuantitativas de las variables aleatorias presentes en la encuesta.

Por su parte, las variables aleatorias son presentadas a través de estimadores. Es decir, a través de "una regla expresada mediante fórmula, que indica cómo calcular una estimación con base a las mediciones obtenidas en base a una muestra". (Wackerly, Mendenhall, & Scheaffer, 2010, p. 391).

Muestreo de Encuestas y Estructuras Estocásticas de Variabilidad

En ocasiones se recurre al método de selección de muestras debido a los elevados costos económicos, logísticos y de tiempo que limitan una evaluación exhaustiva de la población de estudio. Dado que la muestra es un subconjunto de la población, se derivan de las operaciones de extracción de esta un conjunto de errores que convencionalmente se denominan: *errores muestrales*. Es decir, aquellos provenientes de la observación de una porción de la población. Sin embargo, existen otros errores llamados: *errores no muestrales*, que están vinculados con el proceso de aplicación de encuestas, pero no con la extracción muestral.

Se ha planteado que la aplicación de encuesta involucra varias operaciones estas son: la estrategia muestral, las entrevistas y recolección de datos, el procesamiento de los datos, la estimación y análisis, y la presentación de resultados y evaluación final.

Cuadro 1. Operaciones y Errores del Proceso de Aplicación de Encuestas

Operación	Suboperaciones	Tipo de Error
Estrategia Muestral	Construcción de Marco Muestral Diseño Muestral Definición de Estimadores	Errores de marco Errores de muestreo
Entrevistas y Recolección de Datos	Encuentro con el entrevistado	De Medición No Respuestas
Procesamiento de los Datos	Captura de datos Codificación Imputación	Errores de "Tecla" Codificación Incorrecta Imputación Incorrecta
Estimación y Análisis	Cálculo y Evaluación de la Eficiencia y la Precisión (medidas de variabilidad).	Cuantificación de Imprecisiones derivadas de errores de las operaciones previas.
Presentación de Resultados y Evaluación Final	Preparación de Reportes, Informes o Dashboards	-------

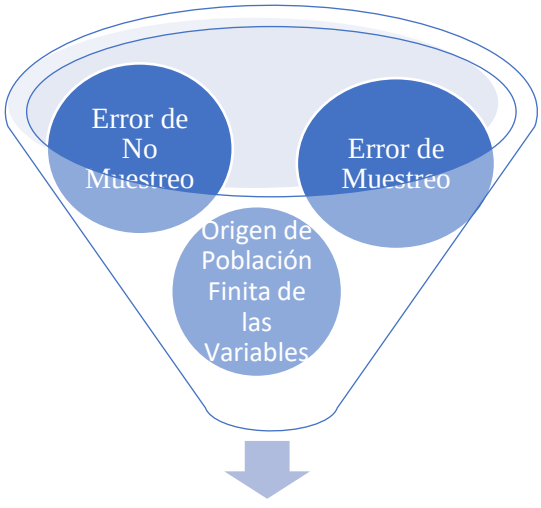
Fuente: Elaboración Propia

Una vez que se ha expuesto que los errores de las estimaciones en las encuestas proceden de la variabilidad introducida durante las operaciones de la aplicación de dichas encuestas, entonces, se puede explicar que la esencia de la estimación estadística consiste en estructurar analíticamente dicha variabilidad para darle precisión a las estimaciones. En otras palabras:

> *"Suponga que se pueden obtener las distribuciones de probabilidad de los diversos errores, si no una especificación completa, al menos algunas características generales.*
> *De ese modo se define la estructura estocástica, y podemos trabajar para obtener los estados de probabilidad sobre el error total en una estimación. Este es el objetivo tradicional de la inferencia estadística.*
> *Para estudiar la variabilidad de las estimaciones de la encuesta, necesitamos especificar con la mayor precisión posible las características estocásticas de los diversos errores". (Sarndal, Carl. et al., 1992, p. 21).*

El trabajo de identificación de las características estocásticas implica conocer las fuentes de aleatoriedad.

Diagrama 1. Estructuras Estocásticas Fuentes de la Variabilidad en Encuestas



Fuente: (Sarndal, Carl. et al., 1992).

En el Cuadro 1 se indican dos de las estructuras estocásticas de las que proviene la variabilidad en las encuestas a partir de las operaciones en las aplicaciones de estas (errores de muestreo y no muestreo).

El error de muestreo está íntimamente ligado a las probabilidades de inclusión de los individuos en la muestra. Esto es inherente al estudio no exhaustivo de la población. Sin embargo, este error puede ser cuantificado mediante las medidas de probabilidad generadas por la aleatorización artificial que supone el diseño muestral.

El error de no muestreo puede ser tratado con algunos modelos que tratan de cuantificar y minimizar las fallas de no respuestas o respuestas faltantes. Sin embargo, en el diagrama 1 se introduce una estructura estocástica adicional, que tiene que ver con la fuente desde donde se generan los valores de las variables aleatorias. Entendiendo, que las variables aleatorias son funciones que conectan a los elementos del espacio muestral (Ω) con los números reales ($\mathbb{R}$). (Evans & Rosenthal, 2005).

Desde el punto de vista epistemológico, el entendimiento de la forma como se originan los valores de las variables de la población está marcado históricamente, por el paradigma de la inferencia estadística clásica, fundado por Ronald Fisher, que considera "cualquier conjunto de mediciones independientes como una muestra aleatoria de una población hipotética infinita" (Fisher, 1922, p. 313).

Es decir, la muestra observada se entiende como un conjunto de realizaciones de un fenómeno, proceso o sistema aleatorio reportadas por las variables aleatorias independientes e idénticamente distribuidas (i.i.d.). Siendo la fuente de los valores reportados una población hipotética infinita.

En donde dicho fenómeno aleatorio constituye un espacio de probabilidad; que puede representarse mediante un modelo de probabilidad definido por parámetros. Entonces, un objetivo relevante es la estimación de dichos parámetros poblacionales desconocidos. Pero, los parámetros habitan en el espacio abstracto paramétrico, mientras que la muestra

observada (en cuanto colección de variables aleatorias i.i.d.) habita en el espacio de la información; en la zona de lo observable. (Vallejos, 2020).

La noción fisheriana de una población hipotética infinita y la muestra observada como una colección de variables aleatorias bajo el supuesto que son registros independientes e idénticamente distribuidos permite aplicar el teorema del límite central y la ley de los grandes números.

Esta visión corresponde a la naturaleza conceptual de algunas poblaciones, que pueden considerarse "teóricamente infinitas" y a los estudios de tipo experimental y semi-empírico en los que se realiza muestreo con variables bajo control; que son comunes en ciencias como: la biología, física, biotecnología, fisiología y similares. Sin embargo, cuando se realizan encuestas en otras ciencias como: la economía, la medicina, la psicología, las ciencias sociales y políticas o en técnicas empresariales como el marketing los estudios son de tipo observables y las poblaciones finitas. Es decir, las inferencias se pueden hacer por diseños muestrales sobre las poblaciones finitas. (Bautista, Palacio, & Delfín, 2011).

No obstante, el hecho que las observaciones de las variables en muestreo de encuestas se originan en poblaciones finitas, implica que las muestras no son colecciones de variables aleatorias independientes, en sentido exacto y riguroso.

La explicación sigue a continuación. La población y su muestra tienen idéntica distribución, suponiendo que no hay sesgo en la selección. Esta distribución de la población puede ser especificada por la función de distribución acumulada ($F_x: R^1 \rightarrow [0,1]$).

Dicha función de distribución poblacional acumulada puede ser evaluada en un valor x. Obteniendo así que: $P(X_{w1} \leq x) = F_x(x)$ y $P(X_{w2} \leq x) = F_x(x)$; de modo tal que: $P(X_{w1} \leq x) = \frac{|w: X_w \leq x|}{N} = F_x(x)$; en consecuencia, si se retira un elemento de la población (de tamaño N), cuyo valor sea x_1, entonces el número de elementos que quedan en la población con $X_w \leq x_1$ es $NF_x(x) - 1$. (Evans & Rosenthal, 2005).

Donde:

F_x: función de distribución poblacional acumulada

R^1: conjunto de los números reales en una dimensión

$[0,1]$: intervalo donde se ubican las probabilidades de la función de distribución

X_{w1}: medida de la variable aleatoria $w1$

X_{w2}: medida de la variable aleatoria $w2$

$F_x(x)$: proporción de elementos de la población cuyo valor medido es menor o igual a x.

N: tamaño de la Población

Es decir, cuando la población es finita y extraemos un elemento, el número de elementos de la población de característica con un determinado valor x_1 cambia ($X_w \leq x_1$), deja de ser $F_x(x)$ y pasa a ser N$F_x(x) - 1$. Pero cuando el tamaño de la población (N) es lo suficientemente grande los valores observados se hacen aproximadamente i.i.d. y cuando la muestra es del tamaño n adecuado la función de distribución empírica se aproxima a la función de distribución acumulada de la población. (Evans & Rosenthal, 2005).

El carácter finito de las poblaciones tiene efectos sobre nuestra capacidad de estimar la función de distribución acumulada de la población de forma exacta. Sin embargo, este efecto sólo es significativo si el tamaño N de la población finita es pequeño. Pero a pesar de estos efectos (en sentido estricto y riguroso) para fines concretos, las consecuencias no son un problema para la aplicación del muestreo y la inferencia estadística en encuestas. La siguiente cita es tranquilizadora:

> *"Una diferencia entre la teoría de la encuesta por muestreo y la teoría de muestreo clásico es que las poblaciones en el trabajo de encuestas contienen un número finito de unidades. Los métodos para probar los teoremas son diferentes y los resultados son ligeramente más complicados cuando el muestreo es de una población finita en lugar de una infinita. Para propósitos*

> *prácticos, estas diferencias en los resultados para las poblaciones finitas e infinitas rara vez son de importancia. Todas las ocasiones en las que la extensión de la muestra es pequeña (en términos del número de unidades del muestreo primario), relativa a la extensión de la población, los resultados derivados de una población infinita son totalmente adecuados." (Cochran, 1971, p. 30).*

Las poblaciones finitas permiten la aplicación del muestreo probabilístico, de forma tal que cada posible muestra de tamaño (n) tenga una probabilidad igual a $\frac{1}{\binom{N}{n}}$ de ser extraída, también conociendo las probabilidades de inclusión de los individuos en la muestra.

Otro resultado de transitar a un enfoque de inferencia de poblaciones finitas implica "*asumir que los valores observados corresponden parámetros fijos poblacionales*".(Gutierrez, 2016).

Muestreo Probabilístico y los Operadores Esperanza Matemática y Varianza

En ocasiones es de utilidad presentar en lenguaje formal estadístico-matemático los mecanismos de operación de la esperanza matemática y la varianza en el marco de una estrategia de muestreo definida para la evaluación de una variable de interés.

Se sabe que en el muestreo e inferencia para poblaciones finitas genera una circunstancia metodológica en la que se requiere cambiar la visión de la inferencia estadística clásica en la cual se asume una población hipotética infinita y pasar a una visión de inferencia de poblaciones finitas.

Dicho cambio de enfoque implica suponer que la aleatoriedad deja de estar en las realizaciones de la muestra aleatoria (como se presupone en la inferencia estadística clásica) provenientes de una población teórica y pasa a descansar sobre la aleatorización artificial del diseño muestral. En esto consiste el enfoque de la inferencia de poblaciones finitas.

En consecuencia, es útil un tipo especial de muestra aleatoria, la muestra probabilística. Ésta última reemplaza a la muestra aleatoria debido a que la aleatorización artificial que introduce el diseño muestral permite estimar las probabilidades de selección a cada posible muestra del conjunto de todas las muestras de un marco de muestreo y las probabilidades de inclusión de un elemento en la muestra.

La siguiente cita es precisa para entender la noción de muestra probabilística:

> *"Una muestra es de tipo probabilística sí:*
>
> *Es posible construir (o al menos definir teóricamente) un soporte Q, tal que* $Q = \{s_1, \ldots s_q, \ldots s_Q\}$*, de todas las muestras posibles obtenidas por un método de selección. En donde* s_q*,* $q = 1, \ldots, Q$*, es una muestra perteneciente al soporte Q.*
>
> *Las probabilidades de selección que el proceso aleatorio le otorga a cada posible muestra perteneciente al soporte son conocidas de antemano a la selección de la muestra final". (Gutiérrez, 2016, p. 25).*

Lo anterior nos lleva a describir cómo funciona la inferencia para poblaciones finitas en base a muestras probabilísticas. Que es la lógica que se aplica en el enfoque usado en este artículo.

$$U = \{u_1, u_2, \ldots u_N\} \rightarrow Q = \{s_1, s_2, \ldots s_Q\} \rightarrow p(\cdot) \rightarrow I_k \rightarrow \hat{\theta}_\pi$$

Donde:

U: es la Población finita (conjunto de elementos identificados con etiquetas).

Q: es el conjunto de todas las muestras posibles de la población objetivo.

$p(\cdot)$: Diseño de muestreo (distribución de probabilidad aplicada a Q)

I_k: probabilidad de inclusión de un elemento a la muestra.

$\hat{\theta}_\pi$: Estimador ajustado por un factor de expansión que asegura la representatividad de la muestra.

a) Primero, se elabora un marco de muestreo en el cual los elementos de la población están identificados mediante etiquetas.
b) Segundo, a partir del marco de muestreo se reconoce el conjunto de todas las posibles muestras extraíbles de la población objetivo.

c) Tercero, se construye un diseño muestral que permita calcular las probabilidades de selección de la muestra y las probabilidades de inclusión de los individuos en la muestra.
d) Cuarto, una vez conocidas las probabilidades de inclusión de un individuo en la muestra se pueden establecer los estimadores y estudiar sus propiedades estadísticas.

Los operadores: *esperanza matemática y varianza* permiten el estudio de estas propiedades estadísticas.

La *esperanza matemática* es una aplicación sobre una función de una variable aleatoria para ubicar su centro de gravedad. Es decir, la esperanza matemática mide la propiedad estadística llamada promedio o media aritmética de la función denominada distribución de probabilidad de una variable aleatoria.

La estructura de la esperanza matemática es lineal esto se visualiza en la fórmula para el caso de variables aleatorias discretas y para el caso de variables aleatorias continuas. Según el álgebra lineal una estructura lineal implica una suma de múltiplos.
Este atributo se cumple en la formulación discreta de la esperanza
Fórmula:

$$E(X) = \sum_{x} xf(x)$$

Y en el caso continuo:
Fórmula:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

Se entiende que la integral es una generalización de la suma para infinitos.

Por definición teórico-matemática *un operador es una aplicación entre dos conjuntos*. Cuando dichos conjuntos están constituidos por elementos de diverso tipo (por ejemplo: un vector y un escalar) a dicho conjunto se le denomina: espacio vectorial. Los espacios vectoriales pueden estar compuestos por funciones. Este es el caso de las funciones g(x) de una variable aleatoria. Incluso, la variable aleatoria en sí misma es una función que

parte del espacio muestral hacia el conjunto de los números reales y una vez que ocurre una realización de la variable, queda determinado el número asociado a la misma.

A la esperanza matemática se le denomina operador cuando se generaliza para cualquier función de una variable aleatoria. Y como mencionamos anteriormente las distribuciones de probabilidad son funciones de las variables aleatorias.

La esperanza matemática posee algunas *propiedades* importantes que permiten que podamos operar con esta medida sobre constantes y variables. (Evans & Rosenthal, 2005).

Dichas *propiedades* son:

$E(C) = C$

$E(CX) = C\,E(X)$

Si $X \geq 0$ entonces $E(X) \geq 0$

$E(X + Y) = E(X) + E(Y)$

Si X y Y son independientes entonces $E(XY) = E(X)\,E(Y)$

Al aplicar la esperanza al muestreo estamos ubicando el centro de gravedad de la distribución muestral dada a partir de una estrategia de muestreo definida con la finalidad de evaluar si el estimador cumple con la propiedad de insesgamiento.

En teoría de la estimación estadística *el insesgamiento* es una característica deseada para un estimador debido que nos permite aproximarnos al valor real del parámetro poblacional con mayor precisión. Esta característica consiste en que el valor esperado del estimador debe ser igual al parámetro poblacional. (Cochran, 1971).

$$E(\hat{\theta}) = \theta$$

Por otro lado, tenemos a la varianza que la podemos definir como un número dado por el valor esperado del cuadrado de una diferencia.

$$VAR(x) = (x - \mu)^2$$

Esta definición se traduce en la formulación discreta de la varianza así,

Fórmula:

$$VAR(X) = \sum_{X} (x - \mu)^2 f(x)$$

Y para el caso continuo,

Fórmula:

$$VAR(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) d(x)$$

En el ámbito del muestreo y la teoría de la estimación este operador sirve para medir la eficiencia del estimador, dada por la variabilidad de x con respecto a la esperanza de la distribución; si el estimador es insesgado entonces esta esperanza coincide con la media poblacional (parámetro). En consecuencia, se trata de ubicar el estimador insesgado de mínima varianza.

Las propiedades de la varianza permiten su operativa y aplicación en los estimadores para alcanzar una alta calidad de los mismo en términos de insesgamiento y eficiencia. Estas propiedades son:

- "VAR (X) ≥ 0
- Si a y b son dos números reales, VAR (aX+b) = a^2 VAR (X)
- VAR (X) = E(X^2) - $(\mu_x)^2$ = E$(X)^2$ - E(X)
- VAR (X) ≤ E(X2)" (Evans & Rosenthal, 2005)

Conclusión

La noción de muestreo de encuestas implica la aplicación de un mecanismo de aleatorización artificial llamado diseño muestral sobre una población finita. Este diseño muestral permite la estimación de las probabilidades de inclusión de un elemento en la muestra lo que sirve para dotar a la investigación por encuesta de objetividad y representatividad.(Gutierrez, 2016).

Por otra parte, los operadores esperanza y varianza sirven para evaluar la calidad de los estimadores en términos de insesgamiento y eficiencia. Ya que contar con un estimador de alta calidad es una condición necesaria para realizar una buena inferencia del parámetro poblacional.

Referencias Bibliográficas

Bautista, F., Palacio, J. L., & Delfín, H. (2011). *Técnicas de Muestreo para Manejadores de Recursos Naturales* (2da.). México, DF: UNAM, CIGA-UNAM, IG-UNAM. Recuperado de https://www.ciga.unam.mx/publicaciones/images/abook_file/tmuestreo.pdf

Cochran, W. (1971). *Técnicas de Muestreo* (1era Español de la 2da en Inglés). México, DF: Compañía Continental.

Evans, M., & Rosenthal, J. (2005). *Probabilidad y Estadística. La Ciencia de la Incertidumbre* (en Español). Barcelona, España: Reverté.

Fisher, R. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *222*(594-604), 309-368. https://doi.org/10.1098/rsta.1922.0009

Gutierrez, H. A. (2016). *Estrategias de Muestreo* (Primera Edición). Bogotá, Colombia: Ediciones de la U.

Sarndal, Carl., Swesson, Bengt., & Wretman, Jan. (1992). *Model Assisted Survey Sampling*. New York, United States of America: Springer-Verlag.

Wackerly, D., Mendenhall, W., & Scheaffer, R. (2010). *ESTADÍSTICA MATEMÁTICA con Aplicaciones* (Séptima Edición). México, DF: Cengage Learning Editores, S.A.

Conflicto de interés

El autor declara no tener conflicto de interés.

Información adicional

La correspondencia y las solicitudes de materiales de este escrito deben dirigirse al autor.

Las impresiones y la información sobre permisos están disponibles en el siguiente enlace: https://www.revistas.up.ac.pa/index.php/contacto/acceso_reuso