

La representatividad de una muestra en investigaciones educativas

The representativeness of a sample in educational investigations

Artículo de investigación

AUTOR:

Dr.C. Redisber Polo Barrera

Correo electrónico: redisberp@gmail.com

<http://orcid.org/0000-0002-5037-6044>

Investigador del Instituto Central de Ciencias Pedagógicas. La Habana, Cuba.

Recibido: 4 de noviembre de 2021

Aceptado: 14 de diciembre de 2021

Publicado: 10 de enero de 2022

RESUMEN

La selección de las personas que pueden formar parte de una investigación es uno de los pasos más importantes del proceso investigativo que conlleva al investigador a plantear los resultados de un problema científico. En este sentido, son muchos los factores que deben ser considerados para lograr mayor precisión al hacer inferencias de los datos obtenidos y de este modo generalizarlos, con una alta probabilidad, a toda la población en la que se selecciona la muestra utilizada. En este trabajo podrás encontrar una explicación de la representatividad de una muestra a partir de los diferentes tipos de muestreo, su importancia y metodología para la selección en los diferentes tipos de investigaciones científicas.

Palabras clave: muestra; muestreo; representatividad de la muestra; ponderadores

ABSTRACT



Artículo de acceso abierto distribuido bajo los términos de la licencia Creative Commons Reconocimiento-NoComercial 4.0 Internacional (CC BY-NC 4.0), que permite su uso, distribución y reproducción en cualquier medio, siempre que el trabajo original se cite de la manera adecuada.

The selection of people that you/they can be part of an investigation is one of the most important steps in the investigative process that bears the investigator to outline the results of a scientific problem. In this sense, there are many factors that should be considered to achieve bigger precision when making inferences of the obtained data and this way to generalize them, with a high probability, to the whole population in which the used sample is selected. In this work you will be able to find an explanation of the representativeness of a sample starting from the different sampling types, their importance and methodology for the selection in the different types of scientific investigations.

Keywords: sample; sampling; representativeness of the sample; pondered

INTRODUCCIÓN

Una sistematización en el campo del muestreo permitirá encontrar, entre otros aspectos, conceptos relacionados con la muestra, población o universo como es conocido por diferentes autores; tamaño de la muestra, representatividad y procedimientos para su selección. También, los diferentes tipos de muestra: probabilísticas y no probabilísticas, así como las diferentes unidades de análisis de las cuales se recolectan los datos.

Asimismo, resulta importante para un investigador hacer cuestionamientos respecto a cómo determinar el tamaño adecuado de una muestra cuando se pretende generalizar los resultados a una población, y qué procedimientos realizar para obtener la muestra en dependencia del tipo de selección a realizar. Son cuestiones que en las investigaciones educativas y de otras índoles como la Evaluación educativa, generan contradicciones en el momento de presentar los resultados finales de estas.

Un gran campo de la ciencia estadística se preocupa por minimizar los errores muestrales y los que pueden ser cometidos en el período de la recolección de los datos. Se trata de un procedimiento que responde a la necesidad de información estadística precisa sobre la población y los conjuntos de elementos que la conforman; el muestreo trata con investigaciones parciales sobre la población que apuntan a inferir a la población completa.

Un ejemplo al que se puede aludir son los estudios de evaluación de la calidad que realiza el LLECE a los diferentes sistemas educativos en varios países de la región. Para el caso de Cuba, específicamente, los resultados que presenta el proyecto de investigación “Evaluación de la calidad, del Instituto Central



de Ciencias Pedagógicas”, asevera que: “... el estudio ha sido diseñado de forma tal que sus resultados son generalizables, con una alta probabilidad, a toda la población de la que se extrajo la muestra utilizada” (Torres et al., 2022. p. 17)

La idea anterior, refuerza la necesidad de análisis a las cuestiones estadísticas que guardan relación con la representatividad de una muestra. Como ya es conocido, una población es el conjunto de individuos sobre los que hacemos cierto estudio, y que una muestra es el subconjunto de esa población. Según Hernández-Sampieri, Fernández & Baptista, (2010) para seleccionar una muestra, lo primero que hay que hacer es definir la unidad de análisis y una vez definida la unidad de análisis se delimita la población.

Como es evidente los resultados de una investigación tendrán un mayor grado de fiabilidad si los instrumentos aplicados ocupan toda la población; sin embargo, no siempre esto es posible, debido a múltiples razones: no tener todos los materiales al alcance; el tiempo para presentar los resultados se puede extender; el acceso a toda la población puede ser limitado, entre otras. No obstante, esto no sería necesario si la muestra resulta estar bien elegida. De modo que es importante saber elegir bien una muestra de la población, una muestra que represente bien a dicha población.

DESARROLLO

Es evidente que la selección de la muestra, en cualquier tipo de investigación hace fiable sus resultados, debido a este supuesto hay que cuidar su elección, para que represente en la medida de lo posible a la población de la que se extrae. Si la muestra está mal elegida, diremos que no es representativa y se pueden producir errores imprevistos e incontrolados. Estos errores se denominan “sesgos”, entonces se diría que la muestra está sesgada.

En esta dirección, Hernández-Sampieri, Fernández & Baptista, (2010), señala el interés de centrar la atención en los participantes, las unidades de análisis, lo cual depende del planteamiento de la investigación y de los alcances del estudio. Por lo tanto, elegir la muestra implica definir la unidad de análisis y requiere delimitar la población para generalizar sus resultados y establecer parámetros.

En este sentido, se reconocen dos tipos de muestras: Probabilística, aquella que requiere precisar su tamaño; Seleccionar elementos muestrales por medio de listados o marcos muestrales con sus debidos procedimientos y, reconoce diferentes clasificaciones como son el caso de la muestra aleatoria simple;



muestra sistemática; muestra estratificada, muestra por conglomerados. En la muestra no probabilística el investigador no pretende que los casos sean representativos de la población.

Teniendo en cuenta los criterios expuestos en el párrafo anterior, entonces podemos afirmar que las distintas formas de elegir una muestra en una población se denominan *muestreos*. Se reconocen dos tipos de muestreos: el muestreo no probabilístico donde el investigador no elige la muestra al azar, sino mediante determinados criterios subjetivos y el muestreo probabilístico cuando la muestra se elige al azar.

Del muestreo probabilístico podemos distinguir varios tipos:

Muestreo aleatorio simple: Aquel en el que cada individuo de la población tiene las mismas posibilidades de ser elegido.

Analicemos este ejemplo:

Nos orientan recoger una muestra de 100 escolares de tercer grado de las 20 instituciones educativas de un municipio de La Habana que tienen una población de 10 000 estudiantes.

Vamos a elegir un alumno al azar. La probabilidad es de elegir $1/10\ 000$ y este alumno lo devolvemos a la población; para continuar con el muestreo vamos a elegir el segundo alumno y entonces la probabilidad de elegirlo continúa siendo de $1/10\ 000$ y así hasta llegar a elegir los 100 alumnos. Como podemos percatarnos si nosotros no devolviésemos al alumno seleccionado, entonces, la probabilidad de escoger al 2º alumno sería ya de $1/9\ 999$ y ya no todos tendrían la misma probabilidad de ser elegidos. El problema consiste en que si devolvemos a los estudiantes entonces existe la probabilidad de que se puedan repetir alumnos.

El muestreo sistemático: es cuando se elige un sujeto al azar y a partir de él, a intervalos constantes, se eligen los demás hasta completar la muestra solicitada.

Como hemos de elegir 100 estudiantes de los 10 000 que hay en el municipio, es decir, un alumno de cada 100, se procede ordenándolos en numeración desde 1 hasta el 10 000, entonces elegimos uno al azar, por ejemplo, el alumno que ocupa la posición 10, y luego los demás estudiantes se eligen a partir de este, a intervalos de 100 alumnos. Escogeríamos por tanto a los alumnos:



10; 110; 210; 310; 410; 510; 610; 710; 810; 910; y el alumno 10 010 ya es otra vez el que ocupaba la posición 10.

El muestreo estratificado: es el que divide la población en clases o estratos y se escoge, aleatoriamente, un número de individuos de cada estrato proporcional al número de componentes de cada estrato. Los estratos no son más que las diferentes aulas o unidades que concentran una población de estudiantes.

Por ejemplo, en una escuela todas las aulas de 3° grado constituyen un estrato cada una. Si deseamos elegir 100 estudiantes de este grado y lograr que la muestra seleccionada sea representativa, lo mejor sería conocer cuántos alumnos de cada aula de 3° hay, es decir, si hay 100 alumnos de 3° A; 150 alumnos de 3° B, 85 alumnos de 3° C y 180 alumnos de 3° D, entonces procedemos a realizar los cálculos siguientes:

¿Cuántos alumnos hay en total, en los cuatro grupos (estratos)? Son 515 estudiantes en total. Como de 515 en total se van a elegir 100 estudiantes, entonces de los 100 del aula 3° A ¿Cuántos debemos seleccionar? Vamos a establecer la proporción siguiente:

Regla de tres:

$$100/515=X/100 \approx 19 \text{ alumnos}$$

Y de este modo, utilizando esta regla matemática, podemos calcular los alumnos correspondientes a los demás grupos. Cuando tengamos la cantidad de estudiantes a seleccionar por cada grupo (estrato), se procede a realizar un muestreo aleatorio simple en cada grupo. Estas técnicas permiten la proporcionalidad del número de componentes de cada estrato y así disminuir el margen de error.

El muestreo por conglomerados: en este tipo de muestreo, en lugar de elegir individuos directamente, se eligen unidades más amplias donde se clasifican los elementos de la población, llamados “conglomerados”. Se realiza en dos etapas, (bietápico). En cada etapa del muestreo en lugar de seleccionar elementos al azar, se eligen conglomerados. La importancia reviste en conocer bien las características de estos. Deben ser tan heterogéneos como la población a estudiar, para que la represente bien.



El otro paso sería elegir algunos de los conglomerados al azar, y dentro de éstos, analizar todos sus elementos para luego tomar una muestra aleatoria simple.

Es importante saber diferenciar los estratos y los conglomerados. Un estrato es homogéneo, sus elementos tienen las mismas características, mientras que un conglomerado es heterogéneo, aspecto que permite representar bien a la población. Según Gutiérrez, (2016) el conglomerado es la agrupación de elementos cuya característica principal es que son homogéneos dentro de sí, y heterogéneos entre sí.

Por ejemplo, si deseamos conocer en qué medida los estudiantes del nivel educativo Preuniversitario del país, tienen una correcta orientación y formación vocacional pedagógica, nos llevaría mucho tiempo y gastos en recursos, así considerado. Para ello podemos extraer una muestra aleatoria de los estudiantes del nivel educativo Preuniversitarios del país.

Si procedemos a aplicar alguno de los tres tipos de muestreos anteriores, necesitaríamos una lista con el nombre de todos ellos, lo cual es muy difícil de conseguir. Sin embargo, los estudiantes están clasificados por Preuniversitarios, grados (10°; 11° y 12°) y aulas o grupos. De este modo, podemos seleccionar en una primera etapa algunos Preuniversitarios, después algunos grados al azar, dentro de los grados algunos grupos y dentro de los grupos, algunos estudiantes por muestreo aleatorio simple. Los conglomerados en cada etapa serían los distintos tipos de Preuniversitarios, los diferentes grados y los diferentes grupos. Como se puede apreciar los conglomerados son unidades amplias y heterogéneas.

Hay muchas maneras de elegir una muestra de una población. La esencia radica en su representatividad. En el muestreo aleatorio simple y en el muestreo aleatorio sistemático, todos los sujetos tienen la misma probabilidad de ser seleccionados; en cambio, para un diseño bietápico, la probabilidad de selección es la probabilidad de selección combinada de cada etapa.

El objetivo principal de la estadística inferencial es el estudio de la población, realizar predicciones a cerca de ella, pero a partir de una muestra de esta, no del total de la población; es decir, predecir, sacar conclusiones, para una población tomando como base una parte de ella y como todas las predicciones, siempre han de hacerse bajo cierto grado de fiabilidad o confianza.

Pero no basta con una correcta selección de la muestra y caracterización de la población. Existen otros factores a considerar, como son los ponderadores, estos aseguran que las estimaciones describan a la



población y no solo a la muestra. Si no utilizamos los *ponderadores*, simplemente obtendremos estimaciones sesgadas.

¿Qué aspectos son considerados?:

- las escuelas que pertenecen a estratos explícitos diferentes pueden representar un número diferente de escuelas en la población;
- los estudiantes de diferentes grupos representan diferentes números de estudiantes en la población;
- los maestros de diferentes escuelas representan a número diferente de maestros en la población.

Los ponderadores ajustan estas diferencias cuando se hacen estimaciones con datos provenientes de la muestra. Digamos que, al seleccionar una muestra de estudiantes en el país, están sub-representados los que asisten a escuelas graduadas (de mayor matrícula), estos estudiantes difieren sistemáticamente de los que asisten a escuelas pequeñas. Aquí se manifiesta una estimación sesgada de la población.

¿Cuál sería la solución? Estudiantes de escuelas grandes, graduadas tendrían un ponderador de las escuelas pequeñas. La ponderación de diseño (W_d), se puede describir como el número promedio de unidades dentro de la población del estudio representadas por la unidad muestreada. Se determina a través del diseño de la muestra y se conoce como: la inversa de su probabilidad de inclusión (p).

Citemos un ejemplo:

Si la probabilidad de inclusión es 1 en 50, entonces cada unidad seleccionada representa en promedio 50 unidades de la población. Por tanto, la ponderación de diseño es 50 ($W_d = 50$)

En un muestreo aleatorio simple, la probabilidad de inclusión es $p = n/N$ para cada unidad, y la ponderación de diseño es $W_d = 1/p$.

En el caso de los muestreos aleatorios sistemáticos

$p = 1/k$ y k es el número entero $k = [n/N]$, el que representa la etapa de muestreo; por consiguiente, para cada unidad, la ponderación de diseño es $W_d = 1/p$.

digamos los alumnos de 3° grado de Cuba 27 654, de ellos se extrajo una muestra aleatoria simple de 400 alumnos. La probabilidad de inclusión de cada alumno es:



$$p = n/N = 400/27.654$$

La ponderación de diseño sería:

$$Wd = 1/p = 27.654/400 = 69,135$$

Para un muestreo estratificado consideremos que el marco muestral del estudio, tienen una población N= 1 000 escuelas. Se divide en dos estratos, urbano y rural.

Urbano (N1= 400) escuelas.

Rural (N2= 600) escuelas.

El tamaño total de la muestra (n= 200) para ambos estratos, tiene una asignación uniforme para cada estrato (100 de cada uno)

En el estrato urbano la probabilidad de selección:

$$p=n/N = 100/400 = 1/4 = 0,25$$

En el estrato rural la probabilidad de selección:

$$p= n/N = 100/600 = 1/6 = 0,167$$

Como se puede apreciar, cada escuela del estrato urbano tiene una ponderación de diseño de (wd,1 = 4), y cada escuela del estrato rural es (wd,2 = 6)

En el muestreo de etapas múltiples, la ponderación general de diseño se calcula tomando el inverso de la probabilidad de selección de cada etapa o fase y multiplicándolos.

Consideremos que una muestra por conglomerados en dos etapas selecciona una muestra aleatoria simple de n1 = 10 de N1 = 100 escuelas en la primera etapa y una muestra aleatoria simple de n2 = 30 alumnos de cada escuela (conglomerado). En la segunda etapa, la cantidad de unidades de cada conglomerado es N2 = 60. La probabilidad de selección en la primera etapa es:

$$P=n1/N1 = 10/100 = 1/10 \text{ (el inverso)}$$

La probabilidad de selección en la segunda etapa es:

$$P=n2/N2 = 30/60 = 1/2$$



Entonces la ponderación de diseño de cada alumno seleccionado es:

$$W_d = 1/w_1 * 1/w_2 = 10 * 2 = 20$$

La estimación es una técnica utilizada para generar información acerca de una población de interés a partir de datos recopilados de una muestra de esa misma población. Para obtener estimaciones correctas ERCE utiliza métodos específicos que toman en cuenta la estructura compleja del muestreo (Balanced Repeated Replication).

CONCLUSIONES

Para obtener una alta precisión en un muestreo y hacer más confiable los resultados e inferencias a la población, todas las áreas deben estar representadas. La calidad de un estudio de esta naturaleza se establece a través de la precisión de las estimaciones.

El esquema de muestreo combina múltiples estratificaciones, conglomeraciones, métodos de muestreo y probabilidades variables y los estimadores de interés son complejos; incluyen múltiples variables de interés.

REFERENCIAS BIBLIOGRÁFICAS

- Gutiérrez. A, (2016). *Estrategias de muestreo Diseño de encuestas y estimación de parámetros*. Material de consulta. En conferencia del Taller Internacional de Muestreo, dado en Ciudad México.
- Hernández-Sampieri, R., Fernández Collado, C. & Baptista, M. P. (2010). *Metodología de la investigación (quinta edición)*. Ciudad de México, México: McGraw-Hill.
- Torres. P et al., (2022). *Informe Nacional I del Cuarto Estudio Regional Comparativo y Explicativo (ERCE-2019)*

Declaración de conflicto de interés y conflictos éticos

El autor declara que este manuscrito es original y no se ha enviado a otra revista. El autor es responsable del contenido recogido en el artículo y en él no existen plagios ni conflictos de interés ni éticos.

