**Algunas cuestiones básicas sobre muestreo**

*Alejandro M. Fridman* •

Cuando hablamos de muestreo, estamos haciendo referencia a la necesidad de obtener información acerca de una determinadª pobla­ ción de unidades de estudio en base a una selccióñ debidamente representativa de ésta. Este procedimiento, habitualmente necesario por razones de co.sto y tiempo, i!Ilplica una serie de cuestiones teóri­ cas y prácticas a 18$ que es necesario considerar. El presente artículo intentará sintetizarlas partiendo de ciertas definiciones básicas.1

Llamamos poblaci9n a un conjunto de ea.sos o unidades que tie­ nen en común una serie determinftda de características y sobre la que se desea obtener cierta información. De esta población, por me­ dio de ciertos pasos a los que nos referiremos más delante, se toma­ rá un subconjunto de unidades qµe, al ser representativas del todo podrá brindarnos\_, mediante lo que conoceremos como inferencia es­ tadísticª los datos que necesitamos con un nivel de precisión acep­ table y mayormente !Jlanipulable por el investigador con un ªhorro considerable de dinero y tiempo.

Cabe acotar que en los estudios cuantitativos siempre se toman muestras con la excepción de los cnsos nacionales que se hacen cada 10 años en casi todos los países, para los CtJ.ales se cuenta con todo la capacidad fipañciera y organizativa de las estructurg.s estata­ les (y aun así se tardan varios años en con- cer los resultados defini­ tivos). Por lo demás, en el Censo Nacionai de Población Argentino de 1991, además de la medición tradicional, se seleccionó una muestra dentro de ésta de hogares a los que se les ?.plicó un cuestionª1"iO ampliado, más detª1.lado en lo que hace a la información a obtener; Por otro lado, el Estado también utiliza muestras cuando lleva a cabo la encuesta permanente de hogares dos veces por año, todos los año;;, para recabar datos sobre empleo principªlmente. De este modo logra reflejar los cambios en la diná\_mica poblacional (cosa que no ocurre con el censo).

El tamaño y tipo de unidades que terminen constituyendo la mues­ tra deben ser los adecuados para poder llevar 1! cabo la inferencia

\* Lieenciado en Sociología y docente del Departamento de Ciencias Sociales y Humanas de la Universidad Nacional de La Matanza y de la Facultad de Ciencias Sociales (UBA).

1. Aquí se tratará excl!!siva­ mente el muestreo de base es­ tadística, dejando de lado d lla­ mado muestreo cualitativo por saturación teórica, utilizado bá­ sicaente en antropología.

*¡:*

'

Algunas cuestione! básicas .11ohre muestreo

,,

estadística que me b?-nde la informaci6 necesaria para mi investi­ gación, en términos de que ésta puedª syr proyectable a toda la po­ blación. En otras palabras una muestra bien construída me garantiza una adecuada validez externa\_ En este septido incluso se estima que una buena muestra puede brindar inforll1iacióv de mayor calidad que up estudio en el quse toma a todª la polación ya qµe en la primera

el error de muestreo será menor que en «fl último.2

Por otro lado, señalemos que en cienpias sociales se suele hacer

una diferencia entre universo general y pniverso de trabajo3

•

El pri­

mero es la población teórica sobre la qu el investigador desea gene­ ralizar sus resultados mientras q.ue el seg4,I, J.do es la operacionalización concreta de ese universo general del que se va a extraer la muestra, delimitada en términos de la dimensión spacio-tiempo. Esta delimi­

. tación no es excluyente y puede amplia.rSe a otras variables. En este

1

sentido, un ejemplo de universo de trabªjo podría ser "madres con

hijos de hasta seis mese.s que residían en,, mys pasado".

Capital Federal durante el

Cabe distinguir también entre poblaÓióñ fipita e infinita. La dife­ rencia es simple, remite al conoci:giiento o desconocimiento dl ta­

maño del universo de est9dio. Si yo me tefiero a una población com­ puesta por indivjduos desocupados qtle viven en lfl. provincia de Chubut, estoy refiriéndome a una población finita (la cifra exacta se puede conseguir de los organismos del' información estadística del Estado). Si mi universo de estudio es "potenciales usuarios de trenes subterráneos en Capital Federal'', no phedo conocer la cifra exact por lo tanto se tratª de una población infinita.

**Tipos de mues«as**

* 1. Ver García Ferrando, M., *Socioestadística. lntroducci6n a la estadútica en sociología,* Alianza, Madrid, 1985, pp. 132

E.s sabido que existen distintos tipo de muestras y que esas dife­ rencias dependen del tipo de diseño d investigción (y por lo tanto de los objetivos ge la misma) y del tiempo y dinero disponibles, entre otros factores.

Laprimera distinción que se hace est.entre muestras probabilísticas

y no probabilísticas. En el muestreo prbbabilístico todos los elemen­ tos tienen una probabilidad conocida, igual y distinta de 1y O de ser

y 133.

* 1. Ver fdem, p. 133.Más allá del

11

seleccionados pa la muestra. Aquí ,la

idea es construir modelos

hecho de que hay autores que distinguen ent,re universo y po­ blación (Capriglioni, 1996), en

reducidos de la pobl.ción total para I9grar resultados extrapolables

al universo del que se extraen. , 1

este trabajo utilizaremos ambos conceptos como sinónimos.

194 - Propuesta s

En este tipo de muestreo la obtenciqn de las unidades es realizada

al azar, por lo que la inclusión de cadé( una de ellas en la muestra no

¡;

#### ,¡

'.I

p:

# Por Alejandro M. Fridman

depende del sujeto que se encarga de realizarla. En estos casos, ade­ más, se puede calcular el error de muestreo que es la diferencia entre el valor de muestreo y el valor real de la población, pudiendo deter­ minarse la precisió\_q de esa proyección que se denomina estimación4 • En este sentido, el estimador obtenido por muestra debe ser, si está bien construida, insesgado y preciso, lo que en otras palabras quiere decir que los valores estimados deben estar n torno a los valores reales (pohlaeionales) y su posi}?Je desviación no debe ser muy gran­ de.

En el caso de las muestras polietápicas, aquellas en las que se deben llevar a cabo varios pasos (etapas) para selecéionar la unidad final, estas pueden ser probabilístioas en algunas de es:as etapas y no en otras. Esto naturalmente perjudicará la calidad de los datos fina­ les en términos de representatividad, o sea a la hora de la inferencia estadística. Según Rodríguez Osuna esto suele ocurrir en la etapa de recoleceión de datos en campo: "es aquí don:de las dificultades son mayores y se pueden dar importantes sesgos por no respetar (...) las instrucciones referidas a la selección de individuos, por recurrir, fá­ cilmente, a las sustituciones o por la existencia de un alto porcentaje de no respuestas" .5

Por todo lo dicho, siempre es preferible que el método de muestreo a utilizar sea probabilístico, sin embargo, no sierqpre es posible. Esto ocurre por límitaciones de orden técnico (ausencia de un marco muestral, listado exhaustivo de las unidades a seleccionar), de tiem-= po e incluso presupuestario. En estas ocasiones debemos apelar a muestras no probabilístieas, en las cuales la selección se haee par­ cia} o totalmente según criterios fijados por el investigador (o incluso por el mismo encuestador, como se verá en el muestreo por cuotas por ejemplo) qüe, al desconocerse la probabilidad de ingreso en la muestra por parte de las unidades, la representatividad de los datos permanece indeterminada. En estos casos, como es obvio, no puedo

estimar los márgenes de error ni, por lo tanto, proyectar los resulta­ dos obtenidos ai unj.verso. Habitualmente este tipo de muestreo sé

utilira cuando se está en la etapa exploratoria del diseño o por los motivos ya expuestos.

Podemos mencionar cüátro tipos de muestreo probabilisticm

1. aleatorio simple,

b. aleatorio sistemático,

*c.* estratificado proporcional y no proporcional y

d. por conglomerados o áreas.

En el primer caso, por ejemplo, si tuviera que extraer uná muestra

* 1. La estimación tiene sustento teórico y estadístico a partir de la tepña de las probabilidades cuyo desarrollo excede el mar· co del presente artículo.

S. Rodáguez Osuna, J., *Cuaíler­*

*nos metodol6gicos: Métodos muestreo.* Cuadernos del CIS, Madrid, 1993, p. 13.

Propuestas - 195

Algunas cuestiones básici!\_s sobre muestreo

6. En teoría de muestreo el tér­ mino sesgo remite a la diferen­ cia entre el valor del estimador y el parámetro poblacfonal real. La magnitnd del sesgo se hace inaceptable cuando de algún modo se abandona la alea" toriedad en la selección de las unidades (por ejemplo, dejando la misma a cri terio del en­ cuestador quien tenderá natu­ ralmente a seleccionar a "cono­ cidos").

196 - Propuestas

J:

de médicos matriculados, selecciono aqullos cuyos números de ma- trícula coinciden con los obtenidos con uni¡i. tabla de números aleatorios (al azar) o por otros medios electrón\_icos o mecánicos. De este modo se le asigna un númerQ a cada individuq con la misma cantidad de dígitos qu a la totalidad de la población (al comienzo se utiliza el cero a la izquierda, ej. ?i hay 900 usar Odl pata el primero). Con este fin se utilizan tablas de números al ª2.ar (caracterizadas por el hecho de que el conocimiento de uno no sirve fpará predecir cuál es el si­ guieIJte o el previo). Se selecciona una p4gina cualquiera de la tabla\_, como los número§ están puestos en columnas, se consideran tantas columnas como dígitos del total (aquf res). Empezamos por cual­ quier colum.na y tomamos el individuo ae tres dígitos que salió (si sale más de 900, en este caso, lo saltean:10s) y así hasta completar el número de unidades estipulado como tamaño de la muestra.

Este tipo de muestreo requiere que [a población a estudiar sea relativamente reducida y no esté muy dispersa geográficamente, si no puede ser costoso e insumir más tiempo del aconsejable. Además es requisito la existencia de un marco, muestra!. Por otro lado, el muestreo aleatorio simple es sencillo de construir; y fácil también de Qalcular, su error de muestreo. Esto se ebe a que es unietápico y a que todas las unidªdes tienen la misma probabilidad de ingresar en la muestra.

En el muestreo sistemático varía la forma de seleccionar las uni­ dades muestrales. Se calcula primero el coeficiente de elevación *Nin* y se elige al azar un número no mayor al ismo. Este será el punto de arranque de la selección a partir del cuhl se le suma sucesivamente el coeficiente de elevación hasta complf tar el tamaño de la muestra determinado. ·

Por ejemplo si un·determinado univprso tiene una población \_ " .

1.000 unidades y quiero tomar una muestra de 50 tendré: *Nin =* 1.000 *I* 50: 20, se elegirá 1de cada 20, lueg se elige al azar un número entre el 1y el 20, por ej. el 17 y al resto se le suma 20 hasta comple­ tar el tamaño de l muestra, es decir 50¡, unidades (17, 37, 57, etc.).

Algunos metod6logos objetan el carácter probabilístico de esta muestra ya que sólo el primer componnte es aleatorio, el resto se selecciona en forma sistemática. De esty modo a partir de la primera selección habrá unidades con probabilidad 1de ingresar en la mues­ tra y otras con prob'!hiiidad O. Más allá de la discusión, se la consi­ dera una muestra probabilística y se la utiliza como tal. También se puede introducir algún sesgo6 si el listdo est\_á ordenado en función de determinados criterios (por ejempl en orden alfabéti90): si los primeros tiene una determin.!lda caracerística y los segundos otra,

1

# Por Akjandro M. Frldman

podría no contemplarse esta heterogeneidad si el tamaño de la mues­ tra se agota en la primera parte.

Este muestreo se utiliza también cuando no hay marco muestral preciso. En este caso, a partir de un plano, se le indica al encuestador que debe entrevistar a la persona que viva en la k-ésima vivienda contando (por ejemplo) desde una esquina sorteada y en el sentido opuesto al de las agujas del reloj (por ejemplo). Lo importante aquí es reducir al mínimo los márgenes de libertad del encuestador y unifi­ car criterios de ingreso final en la muestra para garantizar la aleatoriedad de la misma.

El muestreo estratificado es aplicable cuando la población puede dividirse en clases o estratos (por ej. sexo, edad, clase, nivel de estu­ dios, etc.). Una vez determinados los estratos, se aplica a cada uno un muestreo aleatorio simple (o sistemático) como si fuera una pobla­ ción independiente. La población se divide en estratos exhaustivos (lo que implica que todas las unidades deben poder ser incluidos en algún estrato) y excluyentes (ninguna unidad debe poder estar en dos estratos).

El fundamento del muestreo estratificado es que las categorías pueden variar mucho entre sí respecto al parámetro de interés y, pese a ello, exhibir una gran semejanza dentro de cada categoría (estrato). En este sentido la/s variable/s que se toman como base de la estrati­ ficación deben ser relevantes para la investigación o deben estar al­ tamente correlacionadas con éstas.

En este procedimiento:

l.El universo sujeto a muestreo está subdividido (o estratificado) en grupos que se excluyen entre sí pero incluyen a todos los elementos del universo.

1. A continüación se e.scoge de cada grupo, o estrato, una muestra aleatoria simple. Este método de muestreo difiere del muestreo simple aleatorio porque en este último los elementos de la muestra son escogidos al azar del universo entero, mientras que en el muestreo al azar estratificado la muestra es determi­ nada de tal manera que se escoge un número determinado de elementos de cada estrato.

De este modo obtengo mayor confiabilidad ya que la estratifica­ ción es empleada con frecuencia porque generalmente permite llegar a estimaciones de las muestras que resultan más confiables que las obtenidas con otros medios. Además obtengo información referente a las partes del universo: pudiendo extraerse una muestra aleatoria sim­ ple para cada estrato.

Propuestas - 197

Algunas cuestiones básicas sobre muestreo

* 1. Lo dicho para la cantidad de variables es aplicable de modo idéntico para las cªtegorias de las mismas ya que el efecto es igual: la conformación de un número demasiado grande de submuestras (estratos).

198 - Propuestas

En general la división de estratos se basa en la información dispo­ nible o en los juicios de expertos sobre la pertinencia de las variables geográficas, demográficas, económicas ::y otras variables con el parámetro de la población que estamos traando de estimar.La estra­ tegia general consiste en tener estratos dÓnde los miembros sean lo más semejante posible, pero que difieran máximo de los integran­ tes de otros estratos en cuanto al parámetro (por ej. varones y muje­ res, conservadores, progresistas e indiferntes, etc.).

En cuanto al número de estratos, es claro que cuantos más estra­ tos se utilicen mayor homogeneidad habrá dentro de ellos, pero esto puede ocurrir a expensas de un costo exorbitante que no se compen­ sa necesariamente con la mayor exactitud de la estimación del parámetro de la población. Además una cantidad excesiva de estra­ tos puede complicar el análisis hasta estt1'rilizarlo.7

Al poder calcular la dispersión de los stratos por separado, tam­ b!én puedo utilizar un número menor dtd unidades de muestreo en aquellos estratos que son más homogéne1s así como un número ma­ yor en aquellos con una varianza mayor. ·

Las muestras estratificadas pueden ser, de ponderación proporcio­ nal, óptima o no proporcional. En el primer caso, el número de uni­ dades es porcentualmente igual al que se da en el universo, para lo cual obviamente debo contar con información fiable de éste. En el caso de la ponderación óptima, además de tener en cuenta la distri­ bución porcentual del universo, también vonsidero la homogeneidad de cada estrato ya que cuanto mayor sea sta, menor número de uni­ dades necesitaré para que las estimaciones se aproximen a la reali­ dad d la población. Aquí se toma el valqr del desvío típico de cada estrato como medida de dispersión a la qe se multiplica por el por­ centg.je que éste representa en el universb.

Los muestreos no proporcionales, en cambio, no respetan el por­ centaje distribuido a imitación del universo. Esto puede obedecer a diversas razones. La primera y más obvia es el desconocimiento de las características de mi población. Otro caso que ya fue mencionado es cuando ciertos agregados tienen desviación *stantard* (varianza) mayor y al ser más heterogéneos requierp. un mayor número de uni­ dades de muestreo para reducir la variaqza. También puede ocurrir que interesen más ciertos grupos, por lo 9ua! los sobrerrepresento en algun/os estrato/s de la muestra aun sabii;mdo que la distribución no es un reflejo del todo fiel de la distribucion real de la población. Por otro lado, puede que no tenga un grupo dé interés prioritario, y por lo tanto seleccione idéntica cantidad de unidades en cada estrato, lo cuª1. también daría lugar a muestreo no proporcional.

## Por Alij'andro **M.** Fridman

En el muestreo por conglomerados o áreas, la motivación princi­ pal es resolver la excesiva dispersión de las unidades muestrales finales para lo cual se toman otros agregados primarios, secundarios, el número que sea necesario, en etapas previas. Lo más habitual es que se trate de **un** determinado espacio geográfico, coñ lo que en cada etapa se subdividirá dicho determinado territorio en áreas geo­ gráficas, de las que a la vez se toman aleatoriamente otras y así suce,,­ sivamente hasta llegar a la unidad de muestreo final que también se seleccionará al aza]". Una secuencia muy común de este tipo sería: país, ciudad, barrio, manzana, edificio, individuo. En este sentido, este muestreo es polietª-pico.

Esta modalidad resuelve, como se dijo, la excesiva dispersión de las unidades muestr s en **un** ampiio territorio ya que se toman sólo algunos conglomerados (áreas) que seleccionadas aleatoriamente re­ presentan a las que no entraron en la muestra. Cabe señalar que así como los estratos de la muestra correspoñdiéntes deben ser lo más homogéneo posible, los conglomerados, por el contrario, deben pre­ sentar una máxima heterogeneidad ya que cada uno de ellos debe reflejar la diversidad social, política y económica existente además en los otros conglomerados que representa.

Se debe agregar sin embargo que este tipo de muestreo se puede utilizar también cuando se trata de agregados no geográficos, siendo el procedimiento idéntico, por ejemplo; las unidades hospitalarias, los departamentos universitarios, las urnas electorales o las cajas de productos terminados (eventualmente a ser muestreados para un con­ trol de calidad).8

Los diseños muestrales no probabilísticos tienen las limitaciones ya mencionadas más arriba pero de todos modos mencionaremos los más utilizados. Señalamos en primer lugar el muestreo de conve­ niencia o intencional donde el investigador elige a los elementos que más se adaptan o que están más a su alcance (por ej. estudiantes). Esta modalidad puede ser útil en la fase exploratoria del proyecto de investigación.

Otro caso es el muestreo por cuotas en el que la muestra se selec-

- ciona en base a la división de la población según las cuotas estable-' ciclas que hacen que la composición de la muestra sea proporcional­ mente semejante a la de la población respecto al núm'ero de miem­ bros según un grupo de características consideradas relevantes para **la** investigación, por ejemplo sexo, escolaridad, edad, etc. En este punto se asemeja al muestreo estratificado, sin embargo, la diferen­ cia clave está en el modo de selección de los integrantes de la mues­ tra en cada uno de los grupos (o cuotas). Mientras que aquí laselec-

* 1. Rodríguez Osuna hace una distinción entre conglomerados naturales (como las unidades hospitalarias) y otros conglome­ rados no naturales como las ur­ nas electorales. No vemos cuál puede ser la relevancia de di­ cha distinción. Ver *op. cit.,* p. 31.

Propu,estas - 199

Algunas cuestiones básicas sobre muestreo

ción depende del criterio del encuestado (con el consecuente ses­ go), con la condición de que se cumplan CQ[l determinados requisitos fijados por las cuotas, en el estr:;i.tificado la selección es al azar, con todas las implicancias que hacen que sea un diseño probabilístico.

También podemos mencionar el muestreo de bola de nieve por el cual se selecciona a los individuos en base a referencias de entrevis­ tados inicíales. Se utiliza para muestrear poblaciones extrañas, difí­ ciles y/o de baja incidencia.

### Breves consideraciones acerca de:!... la inferencia estadística !

* 1. Cuando se trata de estimación de parámetros poblacionales se debe tener en cuenta que no to­ das las distribuciones se dispo­ nen del mismo modo, por ejem plo: si el parámetro a estimar es la media, suponemos una distri"'­ bución normal si cbnocemos el valor de la varianzá poblacionál, en caso contrario, utilizamos la distribución t de Student. Si es­ timamos la varianza, la distribu­ ción a utilizar es la de ji o chi cuadrado, muy distinta a las anteriores ya que no contiene valores negativos. No desarrolla­ remos más el tema en este artí­ culo.
	2. Kishe, L., *Diseño estadísti­*

Como se señaló anteriormente, el objef o de extraer una muestra de una población\_ es describir y/o explicar características de la se­ gunda a partir de la primera de un modo r4zonablemente exacto. Di­ cho en otras palabras, se pretende hacer inferencias acerca de una población por medio de distribuciones muestrales similares a ésta, las inferencias serán más exactas cuanto más representativas sean las distribuciones 9• Pero, como dice Kishe, "el plural de la palabra distribuciones se refiere no solo a un (parámetro) estadístico (...), ni solamente a una única variable sino (a) las distribuciones multivariadas de todas las variables (...)".10

Sin embargo, nunca se debe perder de: vista que las muestras se construyen a partir de observaciones que están sujetas a errores di­ versos y a fluctuaciones aleatorias, sobre todo si tenemos en cuenta que representan fenómenos que se encuad m eq el marco de las cien­ cias sociales. Naturalmente estos inconvenientes no invalidan las técnicas de muestreo desde el punto de vista epistemológico ni metodológico, sino que nos obligan a tenetlos en cuenta a la hora de aplicarlas y considerar sus resultados. ¡

En este sentido, la precisión de los m;ismos y la posibilidad de extrapolarlos al universo dependerá del tamaño de la muestra, de los procesos de selección de las unidades y de estimación que se utili­ cen, lo que a la vez dependerá de factores tales como presupuesto y tiempo disponible (particularmente en el caso de los últimos). A la vez, juegan un rol importante ciertas características estructurales de la población como su homogeneidad, que jugará un papel importante en el error de muestreo (y por lo tanto en el tamaño de la muestra) y

hasta su distribución geográfica (en términos de dispersión o con-

*co para la investigación,* Centro

de investigaciones Sociológicas (CIS), Siglo XXI, s/f, p. ,?9.

200 - Propuestas

centración) como se explicó aldescribir fica&.

1

el;,muesteo

por áreas geográ-

## Por A!Jandro M. Frldman

Más allá de estas consideraciones, el ejercicio típico de inferen cia estadística es el de estimaci6n de un" parámetro poblacional a partir de un parámetro muestral obviamente más sencillo de calcu­ lar. Naturalmente esta estimación no es totalmente certera sino que es probabilística y, en este sentido, el grado de incertidumbre es manejable si la muestra se toma utilizando métodos que garanticen aleatoriedad. Desde este punto de partida y en virtud de lo expuesto, se descartan las muestras no probabilísticas a la hora de llevar a cabo estimaciones mínimamente confiables.

Hay dos tipos (complementarios) de estimación: la estimación puntual y la estimaci6n por intervalo. El primero consiste en calcular un valor numérico único que asume el estimador luego de tomar la muestra y realizar las mediciones correspondientes. En otras pª-1-a­ bras, es el valor que asume el parámetro poblacional a estimar en la muestra. Este valor numérico recibe el nombre de punto de estima­ ción y si la misma está razonablemente bien construida, tendrá algún grado de similitud con ese valor del parámetro universal.

La estimación *por* intervalo añade algún grado mayor de seguri­ dad aunque también de imprecisión. En efecto, aquí lo que se obtie­ ne no es un valor numérico "exacto" sino un intervalo llamado inter­ valo de confianza cuyos límites resultan de sumar y restar el error de muestreo al punto de estimación. Si se trabaja con un error de muestreo mayor, el intervalo se "-agranda" y la estimción pierde precisión pero gana en seguridad (la probabilidad de que el intervalo de confianza contenga al valor poblacional estimado es mayor), en caso contrario obviamente se da la situación inversa, con un intervalo más acotado, lo que se gana en precisión se pierde en certeza (de que ese intervalo efectivamente contenga al parámetro poblacional estimado). A esto se refiere Osuna cuando manifiesta que ''cuando el intervalo de con­ fianza en que se sitúa la estimación es muy amplio, el valor de la

estimación es escaso (...)".11

### El tamaño de la muestra

El sentido común nos dice de entrada que cuanto mayor es el número de unidades que ingresan (aleatoriamente) en una muestra, más representativa será ésta de la población de la que fue extraída. En otras pª1.abras, cuanto mayor sea el número de elementos consi­ derados, más seguro será el ,resultado.

La ley de cálculo de probabilidadesT en la que se basa la teoría del muestreo, no desmiente la anterior aprecición, pero la relativiza. En

* 1. Rodríguez Osuna, J., *op.cit.,*

p.14.

Propuestas • 201

AJgunas cuestiones básicas sobre muestreo

efecto, si el tamaño de la muestra es mayor, la probabilidad de obte-

I'

ner resultados más certeros aumenta, pelo también aument su costo

y tiempo de procesamiento. En las muestras probabilísticas, más allá de un cierto tamaño, las ganancias en precisi611 son demasiado pe­ queñas como para seguir aumentándolas Por este motivo se puede y se debe hallar un equilibrio entre los m4rgenes de exactitud que se pretende obtener de los resultados de la muestrª y el costo de la misma, siendo que generalmente el presupuesto (y el tiempo) dispo- nible es limitado. 11

Por otro lado, también puede influir n la determinación del ta­ maño de la muestra el tipo de muestreo a utilizar (el cual depende a su vez, en última instancia, de su .!ldecuación a los objetivos de la investigación). Si f!ecesito subdividi# la muestra porque voy a estratificar, es posible que necesite un húmero mayor de unidades seleccionadas ya que debe haber un número razonablemente grande de éstas al interior de cada estrato. ¡;

Desde el punto de vista de las fórmuJas, para calcular el tamaño

de la muestra tenemos la siguiente síntesis en el cuadro A:

Cuadro A: calculo de tamaño de mustra según tipo (tamaño) de población y parámetrb a estimar 12

---

|  |  |  |
| --- | --- | --- |
|  | Población infinita (con *CJ* conocida).... -1'• | Población infinita (con *g* desconocida)- |
| Parª estimarµ | z2 ·o2 -N z2.G2I2 (N I) + Z2 ·o2 -J2-¡, | t2·s2*T* |
| *n*Proporcióno porcentaje | z2 ·(pq)N 1,.¡2(N l) + Z2 (pq) | z2(pq)-12 |

-- - --

#### -

Í2. Todas estas fórmulas se apli­ can solo a muestreo por azar simple que es el más utilizado.

202 - Propuestas

En el caso del cálculo de la población finita, nosotros conocemos el tamaño de la población y lo tenemos en cuenta en la fórmula (N); como es obvio, este valor no está presen,te en las fórmulas de tamaño del universo desconocidas (población irlfinita). Cabe destacar que si el tamaño del universo es muy grande (más de 10.000 unidades aun­ que esto varía según los autores que se ornen, pero lo importante es el concepto) los resyltados utilizados 9on la fórmula de población finita serán muy similares a los de la fórmula de población infinita. En lo que hace al parámetro a estimar, s1se trata de la media pueden darse tres situaciones:

# Por Alejandro M. Frldman

l. Varianza conocida en pobJación finita, donde, como se dijo, tomo en cuenta el tamaño de la población y aplico el llamado factor de corrección para poblaciones finitas. Este último re­ presenta, en términos teóricos, a la parte de la población que no ha ingresado en la muestra.

* + 1. Varianza conocida en población infinita, donde no hay factor de corrección porque no conozco el tamaño del universo.
		2. Varianza desconocid{l en población infinita, en cuyo caso tra­ bajo con la varianza muestra! mediante la técnica de muestreo iterativo (no lo desarrollaremos aquí, lo planteamos para dejar en claro que se puede calcular el tamaño de la muestrª para estimar la media aun desconociendo la varianza de la pobla­ ción).

En el caso de necesitar estimar una proporción o porcentaje de la población que reúne alguña característica en particuiar, las posibili­ dades son las siguientes:

l. Población finita a la cual le e.abe lo dicho en el caso anterior.

2. Población infinita, análoga al caso 2 anterior.

Cabe acotar que salvo en el e.aso tres, en los otros cuatro casos planteados, se supone CJlle las unidades tienen una distribución aproxi­ mada a la normal. En el caso tres se trabaja con la distribución t de student que es similar a la normal, pero para un número menor de valores (por eso se suele utilizar cuando me manejo con parámetros muestrales).

Uno de los componentes importantes de la fórmula es la varianza, que es una medida que registra la dispersión del universo del que se toma la muestra. n este sentido es importante, ya que cuanto mayor es la heterogeneidad, mayor será el tªmaño de la muestra que necesi­ te para que ésta me dé resultados confiables en té.gninos de la infe­ rencia a desarrollar. En virtud de esto, naturalmente que cuanto más homogénea es una distribución, menor será el tamaño de la muestra requerido para obtener resultados confiables.

Por ejemplo, si yo tuviera que seleccionar individuos para una muestra de estudiantes, ésta sería mucho más homogénea si la tomo de "colegios privados de la zona norte de Capital Federal" (al menos en términos socioculturales y socioeconómicos).que si yo tomara sim­ plemente alun.mos de "colegios de Capital Federal" sin restringir por público - privado ni por área geogrica. Como en el segundo lª varianza es seguramente mayor, el tamaño requerido para poder ex­ trapolar los resultados con alguna seriedad va a ser mayor sin duda.

Propu.estas • 203

Algunas cuestiones bliscas sobre muestreo

1. Rodríguez Osuna J ..*op. cit.•*

p.51

1. No desarrollaremos aquí las características de dicha distri­ bución, que se pueden tomar de cualquier libro de estadística básica.

15.Explicado en p. 3.

204 - Prop.uestas

1

Cuando no se han desarrollado estudios previos que me permitan conocer la varianza poblacional (y dejando de lado el caso tres visto más arriba) o cuando se pretende estiviar parámetros referidos a múltiples variables características de( universo se utilizan las pro­ porciones p . (1- p) que representan lia dispersión del universo a muestrar. Salvo que tengamos más inormación (c·osa que no suele ocurrir), siempre se supone máxima d\spersión o sea 0,-5 . (0,5). En ese caso, se trata de. valores hipotéticos, que no hacen referencia a ninguna variable concreta. Osuna señkla que "esto exige un mayor tamaño de la muestra (...), enc51rece lop costos, pero da garantías de no equivocarse por defecto al determi9-ar el tamaño de una muestra que se va a utilizar para estimar parámetros de variables muy dispa- res".13 /'

Otro de los componentes del cálculp del tamano de la muestra es el nivel de confianza. Éste, en estadís,tica, se refiere a la probabili­ d1:1d de que una estimación de íntervalos, en ausencia de sesgos, se ajuste a la realidad del parámetro estmado. En otras palabras, a la probabilidad de que ei intervalo en cuestión contenga a dicho parámetro (como ya se dijo, promedio,, proporción o varianza según el caso)- De ahí que sea de capital importancia para el proceso de infe­ rencia estadística.

Cuando se trata de estimar media *p* proporción se asume que la distribución de dichos parámetros es aproximada ala norma,l 14 sobre todo cuando la muestra y el universo sÓn grandes. Cuanto mayor es el nivel de confianza mayores son las p:robabilidades de acertar en la estimación en los términos antes despriptos y, como contrapartida, mayor será el tamaño de la muestra. En ciencias sociales los niveles de confianza más utilizados son de 0,9,5 (o 95%) y, en menor medida, 0,99 (o 99%). En la fórmula figuran cpn la notación "Z".

A partir de lo conceptualizado y teniendo en cuenta que la fórmu­ la más utilizada es la de estimación He proporción o porcentaje en poblaciones infinitas, a la hora de martipular el tamaño de la muestra sólo podemos determinar (aumentar º:reducir) el error de muestreo15 o el valor (porcentaje) del nivel de confianza.

En efecto, si mi tamaño de muestra es demasiado elevado para mi presupuesto por ejemplo, para reducirlo puedo optar por disminuir mi valor de Z (nivel de confianza) o aumentar el error de muestreo admitido o, también, hacer ambas coas (tal vez en menor medida). Estas opciones se dan de esta manera[:en virtud del lugar que ocupan el nivel de confianza y el error en la fprmula: la primera en el nume­ rador y el segundo en el denominador. Por otro lado la dispersión es un dato dado o supuesto, no es manipulable.

!;

*Por Alejandro M. Fridmari*

**Bibliografía**

Caprigliorú, C., *Ficha de estadística. lntroducci6n al mues\_treo. Intervalo de confianzq,,* 3C,

1996.

Garcfa Ferrando, M., *So.cioestadística. Introducción a la estadística en sociologCa,* Alianza, Madrid, 1985.

Kishe, L., *Diseño estad(stico para la investigaci6n,* Centro de Investigaciones Sociol6gicas

(CIS), Siglo XXI, s/f.

Rodríguez Osuna, J., *Cuadernos metodol6gico's: Métodos de muestreo.* Cuadernos del CIS,

Madrid, 1993.

Propuestas - 205

