

Ander Egg Ezequiel, **Técnicas de investigación social**, Humanitas, Buenos Aires, 1987, pág. 178/190

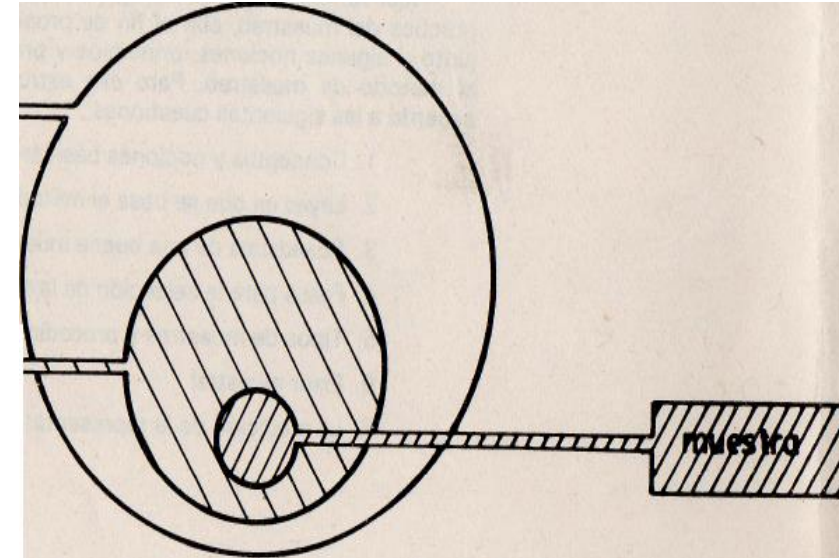
Capítulo 9 El método de muestreo

Para realizar una investigación social, no necesariamente hay que estudiar la totalidad de una población; basta con elegir una muestra representativa de la misma. Esta operación -que es un aspecto auxiliar dentro del proceso general de investigación- es lo que se denomina método o técnica de muestreo.

En sentido lato la palabra muestra designa la parte o elemento representativo de un conjunto. Si usamos el término en sentido estricto, tal como se emplea en Estadística, la muestra es parte de una población objeto de investigación. Esta muestra es elegida por determinados procedimientos y su estudio conduce a conclusiones que son extensivas a la totalidad de la población, con una significativa economía de coste y con una mayor rapidez de ejecución.

Cuando dentro del proceso general de investigación se decide - porque así lo exige la índole del trabajo - elegir una muestra, esta tarea debe planearse y realizarse de manera cuidadosa, puesto que el valor de las conclusiones depende, en buena medida, de la representatividad de la muestra.

Como ya lo indicamos en el capítulo anterior, en las investigaciones de cierta envergadura, esta labor pertenece al estadístico especializado en técnicas de muestreo. Sin embargo, todo investigador social, y, simplemente, toda persona que se inicia en el campo de la investigación, siempre ha de tener una serie de nociones básicas sobre la técnica de muestreo. Y esto por dos razones principales: en unos casos -cuando se trata de investigaciones más simples- porque él mismo podrá elegir la muestra y las unidades de muestreo; en otros, porque deberá acoplar a su trabajo la tarea del estadístico responsable de la determinación de la muestra.



1. Conceptos y nociones básicas

En el método de muestreo se utiliza una serie de conceptos básicos que vamos a precisar en sus aspectos sustanciales y en sus nociones más utilizadas:

Muestra: parte o fracción representativa de un conjunto de una población, universo o colectivo, que ha sido obtenida con el fin de investigar ciertas características del mismo. El problema principal consiste en asegurar que el subconjunto sea representativo de la población, de manera que permita generalizar al universo los resultados obtenidos sobre la muestra.

Muestreo (técnica de): conjunto de operaciones que se realizan para elegir una muestra.

Unidad de la muestra: está constituida por uno o varios de los elementos de la población en las que se subdivide la base de la muestra y que dentro de ella se delimitan de manera bien precisa.

La unidad de la muestra puede ser simple (constituida por un individuo) colectiva (constituida por un grupo, una familia, un pueblo, etcétera). La muestra, en última instancia, comprende un cierto número de estas unidades, obtenidas conforme a un plan de muestreo.

Base de la muestra: es el conjunto de unidades individualizadas que forman un universo o población. Puede ser un censo, un padrón electoral, un catálogo, una lista, un mapa, un plano, etcétera. Sin embargo, no todos los universos están censados o catalogados; en esos casos la base de la muestra no existe y hay que elegirla por algún procedimiento empírico, por ejemplo, el encuestar a una de cada diez personas que se encuentran en un determinado lugar o dejando que el encuestador elija como mejor le parezca cada unidad, una vez que se le asigna una serie de encuestas para cada una de las categorías elegidas, ya sea nivel de ingresos, edad, sexo, etcétera.

Universo, población o colectivo: constituye la totalidad de elementos, seres u objetos que se desea investigar y de la cual se estudiará una fracción (la muestra) que se pretende que reúna las mismas y en igual proporción (*).

Fracción de muestreo: es el porcentaje que representa la muestra respecto al universo; se obtiene dividiendo la muestra por el universo. Así por ejemplo, si en un universo de 10.000 unidades hemos elegido una muestra de 1.200 unidades, la fracción de muestreo será 0,12 o el 12%.

Coefficiente de elevación: Indica las veces que la muestra está contenida en el universo; en otras palabras, es la cantidad por la que hay que elevar la muestra para obtener el universo. Se trata de una relación inversa a la fracción de muestreo. Si tomamos el ejemplo anterior, tendríamos que el coeficiente de elevación es 8'33.

2. Leyes en que se basa el método de muestreo

Salvo en el caso de muestras empíricas o no aleatorias, el método de muestreo se basa en ciertas leyes que le otorgan fundamentación científica: la ley de los grandes números y el cálculo de probabilidades.

La ley de los grandes números, formulada principalmente por Jacques Bernoulli, se expresa del siguiente modo:

«Si en una prueba la probabilidad de un acontecimiento o suceso es p , y si éste se repite una gran cantidad de veces, la relación entre las veces que se produce el suceso y la cantidad total de pruebas -es decir, la frecuencia f del suceso-tiende a acercarse cada vez más a la probabilidad p . Más exactamente, si el número de pruebas es suficientemente grande, resulta totalmente improbable que la diferencia entre f y p supere cualquier valor prefijado por pequeño que sea».

La probabilidad de un hecho o suceso es la relación entre el número de casos favorables (p) a este hecho con la cantidad de casos posibles, suponiendo que todos los casos son igualmente posibles. El modo de establecer la probabilidad es lo que se denomina cálculo de probabilidad.

De estas dos leyes fundamentales de la estadística se infieren aquéllas que sirven de bases más directamente al método de muestreo, a saber:

Ley de la regularidad estadística, según la cual un conjunto de n unidades, tomadas al azar de un conjunto N , es casi seguro que tiene las características del grupo más grande.

Ley de la inercia de los grandes números, corolario de la anterior, se refiere al hecho de que en la mayoría de los fenómenos cuando una parte varía en una dirección, es probable que una parte igual del mismo grupo varíe en dirección opuesta.

Ley de la permanencia de los números pequeños, que los estadísticos formulan de la siguiente manera: si una muestra suficientemente numerosa es representativa de la población, una segunda muestra de igual magnitud deberá ser semejante a la primera; y si en la primera muestra se encuentran pocos individuos con características raras, es de esperar encontrar igual proporción en la segunda muestra.

3. Cualidades de una buena muestra

Para que la muestra tenga validez técnico-estadística, debe cumplir con los siguientes requisitos:

- ser representativa o reflejo general del conjunto o universo estudiado, reproduciendo lo más exactamente las características del mismo
- que su tamaño sea estadísticamente proporcionado a la magnitud del universo
- que el error muestral se mantenga dentro de los límites adoptados como permitidos

4. Fases para la selección de una muestra (plan de muestreo)

La tarea de determinación de una muestra se inserta dentro de las diferentes etapas de la investigación, de ahí que la tarea propiamente dicha del muestreo -diseño de la muestra y recolección de los datos que constituyen la muestra- forma parte de otras labores anteriores y posteriores que llamamos fases para la selección de la muestra, a saber:

- * tener en cuenta los objetivos y finalidad de la investigación
- * la disponibilidad de información sobre el conjunto del que se extraerá la muestra (censos, material cartográfico, etcétera)
- * nivel de confianza adoptado
- * error de estimación permitido
- * los recursos humanos y equipo mecánico disponible
- * los medios financieros
- * los métodos y técnicas a utilizar en la investigación
- * recopilación y análisis de los datos
- * medición de la representatividad de la muestra

Ahora bien, el diseño de la muestra y plan de muestreo es una operación que exige especial capacitación y para la cual los investigadores en muchos casos requieren de un especialista en muestreo; las tareas o fases anteriores y las posteriores -salvo la medición de la representatividad de la muestra- forman parte de las etapas generales de la investigación pero también deben ser bien conocidas por el estadístico responsable de la selección de la muestra.

5. Tipos de muestras y procedimientos de selección

Tomando en cuenta la estructura y los procedimientos de selección, se pueden distinguir dos tipos de muestras y dentro de cada uno de ellos diferentes clases.

- * El primer tipo -muestreo aleatorio o probabilístico- se basa en ciertas

stratificación geográfica, en la cual las unidades de muestreo elementales se sustituyen por un conjunto de unidades físicas situadas en un territorio determinado. No se sorteán individuos sino áreas determinadas en las que se escogen unidades primarias y se recopila información dentro de esas unidades, que pueden incluir un submuestreo de familias.

El procedimiento consiste en dividir un mapa geográfico (de escala adecuada para este trabajo) o bien un relevamiento aerofotográfico, en distintas áreas de cierto grado de homogeneidad (áreas unidad), de las que se extraerán por azar las «áreas muestras». En algunos casos, una ciudad pequeña por ejemplo, puede ser más conveniente recoger la información de toda el área muestra; en otras circunstancias, de esta área muestra -también por azar- se extraen manzanas y dentro de cada manzana se sorteán las casas o establecimientos (según la índole de la investigación) que formarán la muestra.

Cuando se trata de una investigación sobre problemas rurales, hay que dividir el país, región o provincia, en zonas homogéneas o de relativa homogeneidad, para escoger aquéllas dentro de las cuales se sorteán los establecimientos agropecuarios o familias que constituirán las unidades de muestreo. Los mapas catastrales resultan imprescindibles para establecer las zonas o sub-zonas en las que se escogerán las unidades de muestreo.

Este tipo de muestra reviste particular importancia para ciertos estudios de comunidad, razón por la cual explicaremos brevemente la muestra de área realizada por el Instituto de Sociología de la Universidad de Buenos Aires. Para construir esta muestra tuvieron que suplir la falta de información básica, para lo cual se complementaron los datos que se poseían con los del empadronamiento para las elecciones. Además se utilizaron mapas de la Capital y del Gran Buenos Aires, y un relevamiento aerofotográfico completo de la zona. El conjunto -Capital Federal y partidos del Gran Buenos Aires- se dividió en zonas de unas 30 manzanas cada una. Por lo general estas zonas correspondían a los distritos electorales, pero en el Gran Buenos Aires hubo que subdividir los distritos demasiado grandes. Se sortearon luego 78 de estas zonas de 30 manzanas aproximadamente cada una, haciéndose el sorteo de manera proporcional a la población de cada una de ellas. Así por ejemplo, el circuito 130, con 554 electores empadronados por manzana, tuvo casi cinco veces más posibilidades de ser sorteado que el circuito 13, con 115 empadronados por manzana. De este modo, a cada circuito se le dieron tantos números en el sorteo como empadronados tenía. Sorteadas las 78 zonas, se procedió a la confección de una lista de todas las unidades de vivienda existentes en cada una de ellas. Se recorrieron todas las manzanas de cada zona, y se tomó nota de lo que allí había: casas de familia, negocios, fábricas, casas de departamentos y número de los mismos, etc. Es así como se dispuso de listas completas, incluyendo todas las viviendas de cada una de las 78 zonas. Por ejemplo, para obtener una muestra de algo más de 2.000, se sortearon 28 viviendas por cada una de las 78 zonas, obteniéndose 2.184 unidades de vivienda, con su respectiva dirección. Estas direcciones constituyeron la muestra para el estudio en realización (3).

Muestras sucesivas. Este tipo de muestras presenta dos modalidades: la medición «antes-después» (*before-after*), y el estudio de tendencias (*trends*). La medición «antes-después» es un tipo de muestra utilizado para probar los efectos de una acción, un estímulo, un acontecimiento o un cambio introducido en dos mediciones sucesivas. Es un procedimiento muy adecuado para estudiar los efectos de programas de acción social, sobre un grupo o comunidad, y, en general, es útil para estudiar cualquier tipo de fenómenos sociales que interese investigar en su evolución.

El estudio de tendencias difiere del anterior en que se hacen más de dos sondeos, sin que estén determinados por un acontecimiento específico.

Muestras por etapas (multistage sampling): bietápicas y polietápicas.

De algún modo son muestras sucesivas, pero las técnicas son completamente diferentes puesto que las distintas muestras se efectúan sobre fracciones o conglomerados variables del conjunto. Se obtiene una muestra amplia y sobre ella se efectúa una nueva selección o submuestreo realizándose un estudio más profundo. Según se realice en dos o más etapas, se denomina bietápico o polietápico. En diferentes etapas se van seleccionando las fracciones o con-

glomerados hasta obtener la muestra sobre la que se efectuará el estudio: cada una de las etapas da base para la selección de otra más pequeña.

Esta técnica es muy adecuada para estudios sobre costos de vida, hábitos alimentarios, consumos y todos aquéllos que exigen investigaciones muy detenidas y profundas y que por razones prácticas sólo pueden aplicarse a una muestra restringida.

Las muestras de varios grados. Podría considerarse una modalidad de la anterior, pero la técnica es diferente. Son sucesivas y en etapas, pero difieren de las muestras sucesivas en que la recopilación de información no se reitera sobre la misma muestra, y es diferente a la muestra en etapas pues no se utiliza el procedimiento de ésta, consistente en que en cada etapa se efectúa una encuesta, cada vez más profunda de acuerdo con el objetivo de la investigación.

Se utiliza en estudios de gran escala, en donde resulta difícil aplicar una muestra aleatoria simple o una muestra estratificada. El procedimiento consiste en escoger del universo o población «racimos» o conjuntos seleccionados por muestreo simple; sin efectuar ninguna encuesta en la población que constituye el racimo, se efectúa un nuevo sorteo dentro del mismo estableciéndose los «subracimos». En este caso tenemos una muestra de segundo grado; si hacemos nuevos sorteos, será de 3.º, 4.º, *n* grados. El ejemplo que señalamos precedentemente del procedimiento utilizado por el Instituto de Sociología de la Universidad de Buenos Aires es una muestra de áreas de varios grados.

Muestras a partir de listas. Como su nombre lo indica, este tipo de muestra se obtiene partiendo de listas confeccionadas previamente, con lo cual se ahorra tiempo y trabajo. Consiste en escoger los elementos que forman la muestra tomándolos de *k* en *k* a partir de uno escogido al azar. Para que este tipo de muestra tenga validez, se ha de tener la precaución de asegurarse que no hay en la lista características periódicas ligadas al intervalo de sondeo.

b. Muestras no aleatorias o empíricas

La nota característica de estas muestras consiste en que no se basan en una teoría matemática-estadística sino que dependen del juicio del investigador. En relación con las muestras aleatorias, este método posee ventajas en lo que se refiere a costos y a tiempo, pero es más difícil controlar la validez de los resultados.

Se dan dos modalidades diferentes en las muestras no aleatorias: el muestreo intencional u opinático y el muestreo errático, circunstancial y sin norma; el primero de ellos ofrece más garantías en cuanto a los resultados.

Muestreo intencional u opinático: La persona selecciona la muestra

procurando que sea representativa, pero haciéndolo de acuerdo a su intención u opinión. En el muestreo intencional u opinático podemos distinguir: las muestras por cuotas y las muestras razonadas o intencionadas.

Muestras por cuotas o proporcionales. Este tipo de muestreo se utiliza principalmente para realizar sondeos de opinión pública e investigaciones de mercado. Consiste fundamentalmente en establecer cuotas para las diferentes categorías del universo, que son réplicas del conjunto, quedando a disposición del encuestador la selección de unidades. Si en una población hay un 60% de hombres y un 40% de mujeres, la muestra debe reflejar esa población; pero si además se tienen en cuenta categorías socio-profesionales: agricultores, obreros, técnicos, profesionales, etc., la muestra debe tener en cuenta estas proporciones. De ordinario una muestra por cuota comprende la combinación de varias categorías.

Si bien cada encuestador elige según su criterio cada elemento de la encuesta, éste debe ceñirse a la cuota que se le ha asignado. Por ejemplo, en una cuota de 100 pueden ser 55 hombres y 45 mujeres; entre los 55 hombres, 25 serán obreros, 15 agricultores, 9 técnicos y 6 profesionales, y así según las categorías a utilizar.

Digamos por último que el muestreo por cuota pretende ser representativo del total, reflejando las diferentes categorías que se dan en la sociedad, pero su mayor ventaja está en su simplicidad, lo que permite ahorro de tiempo y dinero. Su «debilidad» o limitación se deriva de las distorsiones que pueden provenir de los encuestadores en la selección de las personas, quienes con frecuencia escogerán las que le son afines y las que son más accesibles.

Muestras razonadas o intencionadas. Este tipo de muestra supone o exige un cierto conocimiento del universo a estudiar; su técnica consiste en que el investigador escoge -intencionadamente y no al azar- algunas categorías que él considera típicas o representativas del fenómeno a estudiar. En el estudio de comunidades rurales, por ejemplo, se puede elegir algunas chacras o fincas que se estiman «típicas» o representativas del conjunto.

Sin embargo, la elección intencionada no es absolutamente representativa sino sólo bajo ciertos puntos de vista que el investigador considera importantes o típicos. Como es obvio, este juicio puede estar distorsionado por motivaciones subjetivas, o porque lo que era típico ha dejado de serlo.

Muestreo errático

En un muestreo de este tipo, llamado también circunstancial y sin norma, simplemente se toman los casos que se tienen a mano o que se seleccionan arbitrariamente. Así, por ejemplo, en el estudio de una comunidad en el cual se quieren averiguar las necesidades sentidas por la población, una muestra de este tipo sería aquella que se obtiene recogiendo información de las personas que vienen a mano, sin plan alguno.

Los métodos mixtos o combinaciones de muestreo

En la práctica se combinan con frecuencia diversos procedimientos, no sólo entre las diferentes clases de muestras probabilísticas sino también entre el muestreo probabilístico y el no probabilístico. Así es posible seleccionar áreas que se consideran típicas y luego efectuar en ellas un sorteo de base probabilístico, o a la inversa, puede hacerse una elección aleatoria y seleccionar luego por cuotas los elementos de la muestra.

Las diferentes muestras aleatorias se pueden combinar también de muy diversas maneras y dar como resultado diseños más o menos complejos. ¿Qué es lo más oportuno? ¿Qué se puede cambiar? ¿En qué circunstancias? Todo ello se deja al buen criterio del investigador.

Muestra tipo (master sample)

Es una aplicación combinada y especial de los tipos de muestras que ya hemos explicado. Consiste en seleccionar una muestra «para ser usada»; como se dispone de tiempo, la muestra se establece empleando procedimientos más refinados; una vez establecida, ella constituye el módulo general del cual se extraerá la muestra definitiva conforme a la necesidad específica de cada investigación en concreto.

Como ejemplo suele señalarse la muestra del Bureau of Census de Estados Unidos, cuyo *master samples* ha sido compuesto sobre fotos aéreas y fragmentos de planos de ciudades. La muestra tipo rural está formada por 67.000 «áreas unidades» y comprende 1/18 del territorio agrícola del país. Sobre estas áreas se han seleccionado tres muestras: una comprende 1/54 de las tierras, la otra 1/54 de las explotaciones y la última 1/54 de la población rural.

6. Error muestral

Aun cuando se utilicen los mejores procedimientos, ninguna muestra puede dar una garantía absoluta de ser réplica exacta del universo o población que representa. En último término, los errores son inevitables en toda muestra. Esta diferencia entre el universo y la muestra se denomina error de muestreo o error muestral.

Lo importante es poder determinar el orden o margen de los errores y su frecuencia dentro del conjunto, puesto que ello permite establecer el intervalo de confianza dentro del que nos movemos. Suelen distinguirse dos tipos de errores:

- * sistemáticos
- * por azar o accidentales

Los errores sistemáticos provienen por lo general de diferentes causas ajenas a la muestra misma, pero produce distorsiones o sesgos en la muestra que hace que los resultados obtenidos varíen en una dirección particular.

sustituciones inadecuadas: el encuestador tropieza con una dificultad para obtener la información y la sustituye con el elemento que está más fácilmente a su alcance, no siempre el más adecuado.

- * errores por omisión debidos a la insuficiencia en la recopilación de datos: en este caso la distorsión proviene directamente de la falta de cooperación o de «no-respuesta», ya sea por ignorancia, omisión de datos relativos a elementos incluidos, ausencia, etc.; una proporción importante de «no-respuestas» puede hacer que una muestra probabilística deje de serlo. Este error suele producirse más acentuadamente en los cuestionarios enviados por correo (es probable que quienes responden tengan características diferentes de quienes no responden) errores de observación o sesgos del agente: se trata de distorsiones derivadas del mismo encuestador causadas por sus prejuicios, sus intereses y todo aquello que constituye su «ecuación personal»
- * sesgos de selectividad: errores de cobertura a causa de que no se han incluido elementos importantes y significativos para la investigación que se realiza, o bien hubo una inclusión errónea de unidades.

Los errores accidentales o por azar: cualquiera que sea el procedimiento utilizado y la perfección del método, los valores medios obtenidos de la muestra diferirán de los valores medios reales del universo. Estas diferencias se producen en la asignación al azar o en los instrumentos de medición.

El error de muestreo depende de dos factores:

- * del tamaño de la muestra: a mayor fracción de muestreo, menor será el error de la muestra
- * de la dispersión o desviación típica de la muestra: a mayor dispersión, mayor error.

¿Cómo se puede controlar la representatividad de una muestra, conociendo la extensión probable o margen de error? ¿En qué medida se pueden rectificar los errores?

7. La medición de la representatividad de una muestra

La determinación de la representatividad de una muestra sólo es posible cuando se trata de un muestreo aleatorio. Esta medición no puede establecer en modo alguno en qué medida una muestra es representativa del universo -si así fuese conoceríamos la realidad que se estudia y no necesitaríamos efectuar ninguna investigación-; lo que se puede establecer es la probabilidad de no rebasar ciertos márgenes de

error y de verificar la significatividad de los resultados muestrales para la comprobación de determinadas hipótesis.

Para lo primero se han elaborado métodos de cálculo directo y el abaco de S.S. Wilks; para lo último las pruebas más conocidas son la distribución χ^2 de Pearson y la distribución t Student.

Los métodos de cálculo directo permiten establecer qué dimensiones debe obtener la muestra, para no exceder ciertos límites de error.

Por lo que respecta al abaco de Wilks, en cambio, para cada dimensión de la muestra y los resultados obtenidos, determina la probabilidad de que no se excedan los límites de la proporción real.

La distribución χ^2 de Pearson resulta apropiada para comprobar la correspondencia entre los efectivos teóricos y los efectivos observados; en otras palabras: permite juzgar las relaciones entre teoría y realidad.

Por último, la distribución t Student se utiliza para resolver problemas de pequeñas muestras, permitiendo determinar si una muestra de media \bar{X}_m proviene de un universo de media X . Vista su utilización desde otra perspectiva, la distribución Student permite averiguar la significatividad de la desviación de la media muestral con respecto al valor hipotético del valor de la media del conjunto del universo.

BIBLIOGRAFÍA CITADA

1. YULE, Y. y KENDALL, M. Introducción a la estadística matemática. Aguilar, Madrid, 4Ed.1
2. GOODE, W. y HATT, P. Métodos de investigación social, Trillas, México, 1972
3. GERMANI, Gino Manual del encuestador, Departamento de Sociología, Facultad de Filosofía y Letras, Universidad de Buenos Aires, Publicación interna nro.22, s/f

