

TEMA : INTRODUCCIÓN AL MUESTREO

Copyright [Marcelino Cuesta](#) y [Fco.J.Herrero](#)
Dpto. Psicología
Universidad de Oviedo

1.- Definición y conceptos previos

2.- Tipos de muestreo

3.- Cálculo del tamaño de la muestra

1.- Definición y conceptos previos

En la investigación científica es habitual que se empleen muestras como medio de acercarse al conocimiento de la realidad. Sin embargo, para que esto sea posible, para que a través de las muestras sea posible reproducir el universo con la precisión que se requiera en cada caso es necesario que el diseño muestral se atenga a los principios recogidos en las técnicas de muestreo.

Antes de pasar describir algunos de los métodos de muestreo más habituales introduzcamos algunos conceptos importantes en este contexto:

Población: Es todo conjunto de elementos, finito o infinito, definido por una o más características, de las que gozan todos los elementos que lo componen, y sólo ellos.

En muestreo se entiende por población a la totalidad del universo que interesa considerar, y que es necesario que esté bien definido para que se sepa en todo momento que elementos lo componen.

No obstante, cuando se realiza un trabajo puntual, conviene distinguir entre **población teórica**: conjunto de elementos a los cuales se quiere extrapolar los resultados, y **población estudiada**: conjunto de elementos accesibles en nuestro estudio.

Censo: En ocasiones resulta posible estudiar cada uno de los elementos que componen la población, realizándose lo que se denomina un censo, es decir, el estudio de todos los elementos que componen la población.

La realización de un censo no siempre es posible, por diferentes motivos: a) economía: el estudio de todos los elementos que componen una población, sobre todo si esta es grande,

suele ser un problema costoso en tiempo, dinero, etc; b) que las pruebas a las que hay que someter a los sujetos sean destructivas; c) que la población sea infinita o tan grande que exceda las posibilidades del investigador.

Si la numeración de elementos, se realiza sobre la **población accesible o estudiada**, y no sobre la **población teórica**, entonces el proceso recibe el nombre de **marco o espacio muestral**.

Muestra: En todas las ocasiones en que no es posible o conveniente realizar un censo, lo que hacemos es trabajar con una muestra, entendiendo por tal una parte representativa de la población. Para que una muestra sea representativa, y por lo tanto útil, debe de reflejar las similitudes y diferencias encontradas en la población, ejemplificar las características de la misma. Cuando decimos que una muestra es representativa indicamos que reúne aproximadamente las características de la población que son importantes para la investigación.

Por ejemplo, supongamos que deseamos medir el rendimiento académico de los niños escolarizados en España en la segunda etapa de EGB, pero por problemas económicos solo es posible acceder a los niños de zonas urbanas.

- ¿A quién deseo generalizar los resultados? :

Todos los niños españoles de la segunda etapa (población teórica).

- ¿A quién puedo acceder en el estudio? :

Todos los niños escolarizados en zonas urbanas (población estudiada).

- ¿Cómo puedo acceder a ellos? :

Numerando los sujetos accesibles (espacio o marco muestral).

- ¿Quién forma parte del estudio? :

Elegiendo un grupo de los sujetos enumerados (muestra).

2.- Tipos de muestreo

Los autores proponen diferentes criterios de clasificación de los diferentes tipos de muestreo, aunque en general pueden dividirse en dos grandes grupos: métodos de muestreo probabilísticos y métodos de muestreo no probabilísticos.

2.1.- Métodos de muestreo probabilísticos

Los métodos de muestreo probabilísticos son aquellos que se basan en el principio de equiprobabilidad. Es decir, aquellos en los que todos los individuos tienen la misma probabilidad de ser elegidos para formar parte de una muestra y, consiguientemente, todas las posibles muestras de tamaño n tienen la misma probabilidad de ser elegidas. Sólo estos métodos de muestreo probabilísticos nos aseguran la representatividad de la muestra extraída y son, por tanto, los más recomendables. Dentro de los métodos de muestreo probabilísticos encontramos los siguientes tipos:



2.1.1.- Muestreo aleatorio simple: El procedimiento empleado es el siguiente: 1) se asigna un número a cada individuo de la población y 2) a través de algún medio mecánico (bolas dentro de una bolsa, tablas de números aleatorios, números aleatorios generados con una calculadora u ordenador, etc) se eligen tantos sujetos como sea necesario para completar el tamaño de muestra requerido.

Este procedimiento, atractivo por su simpleza, tiene poca o nula utilidad práctica cuando la población que estamos manejando es muy grande.

2.1.2.- Muestreo aleatorio sistemático: Este procedimiento exige, como el anterior, numerar todos los elementos de la población, pero en lugar de extraer n números aleatorios sólo se extrae uno. Se parte de ese número aleatorio i , que es un número elegido al azar, y los elementos que integran la muestra son los que ocupan los lugares $i, i+k, i+2k, i+3k, \dots, i+(n-1)k$, es decir se toman los individuos de k en k , siendo k el resultado de dividir el tamaño de la población entre el tamaño de la muestra: $k=N/n$. El número i que empleamos como punto de partida será un número al azar entre 1 y k .

El riesgo de este tipo de muestreo está en los casos en que se dan periodicidades en la población ya que al elegir a los miembros de la muestra con una periodicidad constante (k) podemos introducir una homogeneidad que no se da en la población. Imaginemos que estamos seleccionando una muestra sobre listas de 10 individuos en los que los 5 primeros son varones y los 5 últimos mujeres, si empleamos un muestreo aleatorio sistemático con $k=10$ siempre seleccionaríamos o sólo hombres o sólo mujeres, no podría haber una representación de los dos sexos.

2.1.3.- Muestreo aleatorio estratificado: Trata de obviar las dificultades que presentan los anteriores ya que simplifican los procesos y suelen reducir el error muestral para un tamaño dado de la muestra. Consiste en considerar categorías típicas diferentes entre sí (estratos) que poseen gran homogeneidad respecto a alguna característica (se puede estratificar, por ejemplo, según la profesión, el municipio de residencia, el sexo, el estado civil, etc). Lo que se pretende con este tipo de muestreo es asegurarse de que todos los estratos de interés estarán representados adecuadamente en la muestra. Cada estrato funciona independientemente, pudiendo aplicarse dentro de ellos el muestreo aleatorio simple o el estratificado para elegir los elementos concretos que formarán parte de la muestra. En ocasiones las dificultades que plantean son demasiado grandes, pues exige un conocimiento detallado de la población. (tamaño geográfico, sexos, edades,...).

La distribución de la muestra en función de los diferentes estratos se denomina afijación, y puede ser de diferentes tipos:

Afijación Simple: A cada estrato le corresponde igual número de elementos muestrales.

Afijación Proporcional: La distribución se hace de acuerdo con el peso (tamaño) de la población en cada estrato.

Afijación Óptima: Se tiene en cuenta la previsible dispersión de los resultados, de modo que se considera la proporción y la desviación típica. Tiene poca aplicación ya que no se suele conocer la desviación.

Supongamos que estamos interesados en estudiar el grado de aceptación que la implantación de la reforma educativa ha tenido entre los padres de una determinada provincia. A tal efecto seleccionamos una muestra de 600 sujetos. Conocemos por los datos del ministerio que de los 10000 niños escolarizados en las edades que nos interesan, 6000 acuden a colegios públicos, 3000 a colegios privados concertados y 1000 a colegios privados no concertados. Como estamos interesados en que en nuestra muestra estén representados todos los tipos de colegio, realizamos un muestreo estratificado empleando como variable de estratificación el tipo de centro.

Si empleamos una afijación simple elegiríamos 200 niños de cada tipo de centro, pero en este caso parece más razonable utilizar una afijación proporcional pues hay bastante diferencia en el tamaño de los estratos. Por consiguiente, calculamos que proporción supone cada uno de los estratos respecto de la población para poder reflejarlo en la muestra.

Colegios públicos: $6000/10000=0.60$

Colegios privados concertados: $3000/10000=0.30$

Colegios privados no concertados: $1000/10000=0.10$

Para conocer el tamaño de cada estrato en la muestra no tenemos más que multiplicar esa proporción por el tamaño muestral.

Colegios públicos: $0.60 \times 600 = 360$ sujetos

Colegios privados concertados: $0.30 \times 600 = 180$ sujetos

Colegios privados no concertados: $0.10 \times 600 = 60$ sujetos

2.1.4.- Muestreo aleatorio por conglomerados: Los métodos presentados hasta ahora están pensados para seleccionar directamente los elementos de la población, es decir, que las unidades muestrales son los elementos de la población. En el muestreo por conglomerados la unidad muestral es un grupo de elementos de la población que forman una unidad, a la que llamamos conglomerado. Las unidades hospitalarias, los departamentos universitarios, una caja de determinado producto, etc, son conglomerados naturales. En otras ocasiones se pueden utilizar conglomerados no naturales como, por ejemplo, las urnas electorales. Cuando los conglomerados son áreas geográficas suele hablarse de "muestreo por áreas".

El muestreo por conglomerados consiste en seleccionar aleatoriamente un cierto número de conglomerados (el necesario para alcanzar el tamaño muestral establecido) y en investigar después todos los elementos pertenecientes a los conglomerados elegidos.

En una investigación en la que se trata de conocer el grado de satisfacción laboral los profesores de instituto necesitamos una muestra de 700 sujetos. Ante la dificultad de acceder

individualmente a estos sujetos se decide hacer una muestra por conglomerados. Sabiendo que el número de profesores por instituto es aproximadamente de 35, los pasos a seguir serían los siguientes:

1. Recoger un listado de todos los institutos.
2. Asignar un número a cada uno de ellos.
3. Elegir por muestreo aleatorio simple o sistemático los 20 institutos ($700/35=20$) que nos proporcionarán los 700 profesores que necesitamos.

Para finalizar con esta exposición de los métodos de muestreo probabilísticos es necesario comentar que ante lo compleja que puede llegar a ser la situación real de muestreo con la que nos enfrentemos es muy común emplear lo que se denomina *muestreo polietápico*. Este tipo de muestreo se caracteriza por operar en sucesivas etapas, empleando en cada una de ellas el método de muestreo probabilístico más adecuado.

2.2.- Métodos de muestreo no probabilísticos

A veces, para estudios exploratorios, el muestreo probabilístico resulta excesivamente costoso y se acude a métodos no probabilísticos, aun siendo conscientes de que no sirven para realizar generalizaciones, pues no se tiene certeza de que la muestra extraída sea representativa, ya que no todos los sujetos de la población tienen la misma probabilidad de ser elegidos. En general se seleccionan a los sujetos siguiendo determinados criterios procurando que la muestra sea representativa.

2.2.1.- Muestreo por cuotas: También denominado en ocasiones "accidental". Se asienta generalmente sobre la base de un buen conocimiento de los estratos de la población y/o de los individuos más "representativos" o "adecuados" para los fines de la investigación. Mantiene, por tanto, semejanzas con el muestreo aleatorio estratificado, pero no tiene el carácter de aleatoriedad de aquél.

En este tipo de muestreo se fijan unas "cuotas" que consisten en un número de individuos que reúnen unas determinadas condiciones, por ejemplo: 20 individuos de 25 a 40 años, de sexo femenino y residentes en Gijón. Una vez determinada la cuota se eligen los primeros que se encuentren que cumplan esas características. Este método se utiliza mucho en las encuestas de opinión.

Por ejemplo, la Consejería de Sanidad desea estudiar la incidencia de las drogas en la adolescencia. Lo que deberíamos hacer sería: conocer por los informes de la Consejería de Educación cuales son los centros más afectados por el problema, fijar un número de sujetos a entrevistar proporcional a cada uno de los estratos (cuotas) y finalmente dejar en manos de los responsables del trabajo de campo a que sujetos concretos se deberá entrevistar.

2.2.2.- Muestreo opinático o intencional: Este tipo de muestreo se caracteriza por un esfuerzo deliberado de obtener muestras "representativas" mediante la inclusión en la muestra de grupos supuestamente típicos. Es muy frecuente su utilización en sondeos preelectorales de zonas que en anteriores votaciones han marcado tendencias de voto.

2.2.3.- Muestreo casual o incidental: Se trata de un proceso en el que el investigador selecciona directa e intencionadamente los individuos de la población. El caso más frecuente de este procedimiento el utilizar como muestra los individuos a los que se tiene fácil acceso (los profesores de universidad emplean con mucha frecuencia a sus propios alumnos). Un caso particular es el de los voluntarios.

2.2.4.- Bola de nieve: Se localiza a algunos individuos, los cuales conducen a otros, y estos a otros, y así hasta conseguir una muestra suficiente. Este tipo se emplea muy frecuentemente cuando se hacen estudios con poblaciones "marginales", delincuentes, sectas, determinados tipos de enfermos, etc.

3.- Cálculo del tamaño de la muestra

A la hora de determinar el tamaño que debe alcanzar una muestra hay que tomar en cuenta varios factores: el tipo de muestreo, el parámetro a estimar, el error muestral admisible, la varianza poblacional y el nivel de confianza. Por ello antes de presentar algunos casos sencillos de cálculo del tamaño muestral delimitemos estos factores.

Parámetro. Son las medidas o datos que se obtienen sobre la población.

Estadístico. Los datos o medidas que se obtienen sobre una muestra y por lo tanto una estimación de los parámetros.

Error Muestral, de estimación o standard. Es la diferencia entre un estadístico y su parámetro correspondiente. Es una medida de la variabilidad de las estimaciones de muestras repetidas en torno al valor de la población, nos da una noción clara de hasta dónde y con qué probabilidad una estimación basada en una muestra se aleja del valor que se hubiera obtenido por medio de un censo completo. Siempre se comete un error, pero la naturaleza de la investigación nos indicará hasta qué medida podemos cometerlo (los resultados se someten a error muestral e intervalos de confianza que varían muestra a muestra). Varía según se calcule al principio o al final. Un estadístico será más preciso en cuanto y tanto su error es más pequeño. Podríamos decir que es la desviación de la distribución muestral⁽¹⁾ de un estadístico y su fiabilidad.

Nivel de Confianza. Probabilidad de que la estimación efectuada se ajuste a la realidad. Cualquier información que queremos recoger está distribuida según una ley de probabilidad

(Gauss o Student), así llamamos nivel de confianza a la probabilidad de que el intervalo construido en torno a un estadístico capte el verdadero valor del parámetro.

Varianza Poblacional. Cuando una población es más homogénea la varianza es menor y el número de entrevistas necesarias para construir un modelo reducido del universo, o de la población, será más pequeño. Generalmente es un valor desconocido y hay que estimarlo a partir de datos de estudios previos.

3.1.- Tamaño de muestra para estimar la media de la población

Veamos los pasos necesarios para determinar el tamaño de una muestra empleando el muestreo aleatorio simple. Para ello es necesario partir de dos supuestos: en primer lugar el nivel de confianza al que queremos trabajar; en segundo lugar, cual es el error máximo que estamos dispuestos a admitir en nuestra estimación. Así pues los pasos a seguir son:

1.- Obtener el tamaño muestral imaginando que $N \rightarrow \infty$:

$$n_{\infty} = \frac{z_{\alpha/2}^2 \sigma^2}{e^2}$$

donde:

$z_{\alpha/2}$: z correspondiente al nivel de confianza elegido

σ^2 : varianza poblacional

e: error máximo

2.- Comprobar si se cumple

$$N > n_{\infty}(n_{\infty} - 1)$$

si esta condición se cumple el proceso termina aquí, y ese es el tamaño adecuado que debemos muestrear.

Si no se cumple, pasamos a una tercera fase:

3.- Obtener el tamaño de la muestra según la siguiente fórmula:

$$n = \frac{n_{\infty}}{1 + \frac{n_{\infty}}{N}}$$

Veamos un ejemplo: La Consejería de Trabajo planea un estudio con el interés de conocer el promedio de horas semanales trabajadas por las mujeres del servicio doméstico. La muestra

será extraída de una población de 10000 mujeres que figuran en los registros de la Seguridad Social y de las cuales se conoce a través de un estudio piloto que su varianza es de 9.648. Trabajando con un nivel de confianza de 0.95 y estando dispuestos a admitir un error máximo de 0,1, ¿cuál debe ser el tamaño muestral que empleemos?.

Buscamos en las tablas de la curva normal el valor de $z_{\alpha/2}$ que corresponde con el nivel de confianza elegido: $z_{\alpha/2} = \pm 1.96$ y seguimos los pasos propuestos arriba.

1.-

$$n_{\alpha} = \frac{1.96^2 \cdot 9.648}{0.1^2} = 3706$$

2.- Comprobamos que no se cumple $N > n_{\alpha}(n_{\alpha} - 1)$, pues en este caso

$$10000 < 3706(3706 - 1); 10000 < 13730730$$

3.-

$$n = \frac{3706}{1 + \frac{3706}{10000}} = 2704$$

3.2.- Tamaño de muestra para estimar la proporción de la población

Para calcular el tamaño de muestra para la estimación de proporciones poblacionales hemos de tener en cuenta los mismos factores que en el caso de la media. La fórmula que nos permitirá determinar el tamaño muestral es la siguiente:

$$n = \frac{N z_{\alpha/2}^2 P(1-P)}{(N-1)e^2 + z_{\alpha/2}^2 P(1-P)}$$

donde

$z_{\alpha/2}$: z correspondiente al nivel de confianza elegido

P: proporción de una categoría de la variable

e: error máximo

N: tamaño de la población

Siguiendo con el estudio planteado en el punto anterior, supongamos que tratamos de estimar la proporción de mujeres que trabajan diariamente 10 horas o más. De un estudio piloto se dedujo que $P=0.30$, fijamos el nivel de confianza en 0.95 y el error máximo 0.02.

$$n = \frac{10000 + 1.96^2 \cdot 0.30(1-0.30)}{(10000-1) \cdot 0.02^2 + 1.96^2 \cdot 0.30(1-0.30)} = 1678$$

(1) Por distribución muestral se entiende la distribución de frecuencias de los valores de un estadístico en infinitas muestras iguales.

Bibliografía

- ABAD DE SERVIN, A. y SERVIN ANDRADE, L.A. (1978). *Introducción al muestreo*. Mexico: Limusa.
- AZORIN, F. (1972). *Curso de muestreo y aplicaciones*. Madrid: Aguilar.
- AZORIN, F. y SANCHEZ CRESPO, J.L. (1986). *Métodos y aplicaciones de muestreo*. Madrid: Alianza.
- MARTINEZ ARIAS, M.R.; MACIA, A. y PEREZ RUY-DIAZ, J. (1989). *Psicología Matemática II*. Madrid: UNED.
- PARDO, A. y SAN MARTIN, R. (1994). *Análisis de datos en psicología II*. Madrid: Pirámide.
- RODRIGUEZ OSUNA, J. (1991). *Métodos de muestreo*. Madrid: CIS.
- RODRIGUEZ OSUNA, J. (1993). *Métodos de muestreo. Casos prácticos*. Madrid: CIS.
-