
LA TECNICA DE MUESTREO: EJEMPLO PRACTICO DE SU APLICACION EN LAS EVALUACIONES EDUCATIVAS

Mariano Alvaro Page
Servicio de Evaluación del CIDE

RESUMEN. Este artículo es un ejemplo eminentemente práctico sobre la forma de llevar a cabo un muestreo dentro del campo educativo, en el que la unidad muestral es generalmente, al menos en una primera fase, un conglomerado, el centro. Entre los temas tratados, los más interesantes podrían ser la definición de las poblaciones, el efecto del diseño en los muestreos aleatorios por conglomerados y el procedimiento utilizado para elegir un centro con probabilidad directamente proporcional al tamaño del mismo. Con la lectura de este artículo puede conocerse un método de muestreo que es común, en gran parte, a los utilizados en las diferentes instituciones internacionales dedicadas a la evaluación educativa.

El contenido de este escrito hará referencia de manera global a los muestreos de las diferentes evaluaciones que está llevando a cabo el Servicio de Evaluación del CIDE o en las que está implicado, haciendo especial hincapié en dos de ellas, puesto que pueden considerarse paradigmáticas: evaluación de la Reforma del segundo ciclo de EE. MM. y evaluación de la Reforma del ciclo superior de EGB. Se han escogido estas dos, ya que son de ámbito español únicamente y, por consiguiente, somos absolutamente autónomos a la hora de realizar el muestreo. En las evaluaciones internacionales tenemos que atenernos a las directrices que nos marquen las instituciones promotoras.

Los puntos a desarrollar en estas líneas van a ser los siguientes:

- Definición de poblaciones.
- Determinación del número de sujetos de la muestra (N).
- Procedimiento de muestreo.

1. *Definición de poblaciones*

En algunos de los trabajos del Servicio de Evaluación, definir la población es un procedimiento algo complejo, debido a que, por una parte, trabajamos con colectivos diversos y heterogéneos y, por otra, porque los objetivos de los estudios son también diversos. En las próximas líneas se expondrán estas situaciones o hechos al referirnos de manera concreta a los dos trabajos mencionados poco más arriba:

A) Empezamos por el más complejo: *La evaluación de la Reforma del segundo ciclo de las Enseñanzas Medias*. En el segundo ciclo educativo de nuestro país hay dos subculturas, dos colectivos claramente diferenciados: BUP y FP.

En FP, además, hay varias ramas y un enorme número de especialidades; cada una de las cuales tiene un *currículum* específico y muy o totalmente diferente a las demás. Teniendo en cuenta que la medición del rendimiento es uno de los objetivos básicos de nuestros estudios, y que al pretender evaluar el plan experimental deben compararse alumnos de dos planes, puede comprenderse la dificultad del problema.

Con este trabajo, además de evaluar la Reforma, se pretende también estudiar cómo está la situación educativa —en algunos aspectos obviamente— al finalizar la enseñanza media.

Por consiguiente, las dos *poblaciones* de este estudio han estado constituidas por:

a) Alumnos de centros de Bachillerato de todo el Estado español que están cursando, en el momento de la aplicación, COU y segundo curso del Bachillerato Experimental del segundo ciclo de EE. MM.

b) Alumnos de centros de Formación Profesional de todo el Estado español que están cursando el segundo curso de Bachillerato Experimental del segundo ciclo de EE. MM. y las siguientes ramas y especialidades del segundo curso de FP II:

Ramas FP (2.º)

Especialidades

— Administrativa y Comercial	— Administrativa — Secretariado
— Electricidad y Electrónica	— Electrónica Industrial — Equipos Informática
— Metal	— Calderería en chapa estructural — Máquinas y Herramientas — Matricería y Moldes

Los criterios para considerar sólo estas ramas y especialidades han sido:

— En cuanto a las ramas, se han escogido las más numerosas (con mayor número de matriculados) y las que tenían aspectos curriculares comunes con el Bachillerato Experimental.

— En cuanto a las especialidades, se han escogido aquellas que tenían parte de su *currículum* común con el Bachillerato Experimental, lo que permite hacer comparaciones.

Después de la lectura de estas líneas se comprenderá perfectamente lo que repiten los manuales sobre «técnicas de muestreo», en la introducción a las mismas, cuando afirman que la definición de la población y, por consiguiente, la muestra de un estudio depende de los objetivos del estudio. Pues en el caso que nos ocupa si su finalidad primordial no hubiera sido evaluar la Reforma, que exige comparar diferentes planes, la población deseada en términos de Ross (1987, pp. 58 y ss.) hubiera sido otra, en cuanto a las ramas y especialidades de FP.

B) En el estudio que realizamos para evaluar la *Reforma del ciclo superior de la Educación General Básica*, la población deseada está constituida por alumnos de 8.º de EGB, tanto de enseñanza vigente como experimental, que están siguiendo sus estudios en centros del territorio MEC.

Hasta aquí se ha hablado de población «deseada», en términos de Ross. Frente a este término existen otros dos: población «definida» y población «excluida». La población deseada es aquella que es requerida en términos ideales; aquella que, si no hubiera ningún impedimento circunstancial, sería la evaluada o investigada. Ahora bien, puede darse el caso que por razones administrativas, económicas, etc., se reduzca, en parte, la población a estudiar. Así, por ejemplo, en nuestro caso hemos dejado de considerar, por razones económicas, a los alumnos de BUP y EGB que siguen sus estudios en centros con menos de 30 alumnos en los cursos evaluados (COU, 2.º de Bachillerato Experimental y 8.º de EGB) y a los alumnos de FP de los centros con menos

de 20 alumnos en los cursos considerados (2.º de FP II-2.º de Bachillerato Experimental). La población *definida* es aquella que va a ser *realmente* estudiada y la *excluida* es aquella que se incluye en la deseada, pero no en la definida. En nuestro caso, la definida estaría constituida por la deseada, menos los alumnos que siguen sus estudios en centros con menos de 30 ó 20 alumnos en los cursos considerados. La población excluida está formada, precisamente, por estos alumnos de los centros con menos de 30 ó 20 unidades en los cursos a evaluar.

2. Determinación del número de sujetos de la muestra

Bajo el punto de vista estadístico puede decirse que el tamaño (N) de una muestra depende, básicamente, de:

- Tamaño de la población.
- Intervalo de confianza.
- Margen de error.
- Tipo de diseño.

También depende del número de casos favorables esperados, aunque siempre se considera, de manera sistemática, la proporción del 50 por 100, ya que es la que origina el producto PQ mayor y, en consecuencia, el mayor tamaño muestral, teniendo en cuenta la fórmula:

$$N = \frac{S^2 PQ}{E^2}$$

Como se verá más tarde, el tamaño también puede venir condicionado, de manera funcional, por los objetivos del estudio y por el modelo o, mejor, por las variables no dependientes incluidas en el mismo y, de manera operativa, por las condiciones presupuestarias.

En los trabajos del Servicio de Evaluación, casi siempre trabajamos con grandes tamaños muestrales que sobrepasan, con creces, el número de elementos (estudiantes en nuestro caso) requeridos por los niveles de confianza y márgenes de error más exigentes para muestreos aleatorios simples, sistemáticos y estratificados que tienen como unidad muestral un solo elemento. Sin embargo, dado que el procedimiento de muestreo es por conglomerados, siempre comprobamos, de manera sistemática, cuál es el número mínimo de conglomerados exigido, y subsidiariamente el de alumnos, cuando se trabaja con un nivel de confianza del 95,5 por 100 (± 2 sigmas) y con un margen de error del 5 por 100.

A ese mínimo exigido en los intervalos de confianza y márgenes de error

señalados, casi siempre añadimos un número considerable teniendo en cuenta que, posteriormente, en los análisis realizados con los datos comparamos, simultáneamente, los resultados obtenidos por varios subgrupos y que, por lo tanto, puede correrse el riesgo de quedarnos sin sujetos en las casillas de alguno de los subgrupos. He aquí la explicación de por qué se decía anteriormente que el tamaño de la muestra también debe depender del modelo utilizado. El modelo de una investigación trata de las variables consideradas y de las relaciones que se establecen, *a priori*, entre ellas. La situación, por lo que al muestreo se refiere, debe ser diferente si las variables independientes o de control (moduladoras) son una o varias o si tiene uno o varios estratos cada una de ellas.

El tamaño de la muestra también depende del tipo de *diseño muestral*. En los estudios del Servicio de Evaluación del CIDE se utiliza un tipo de muestreo muy común en el mundo educativo, muy posiblemente el más común: el *muestreo por conglomerados*.

En nuestro caso, cada centro es un conglomerado y en cada uno de estos conglomerados se elige un número fijo de alumnos: 20 ó 30.

El tamaño de la muestra en un muestreo por conglomerados depende directamente del llamado *efecto del diseño*, además del «N» de la muestra en un muestreo aleatorio simple (Kish, 1965, p. 162; Ross, *op. cit.*, pp. 66 y ss.; Azorín y Sánchez-Crespo, 1986, p. 166; Cohran, 1985, p. 299).

El efecto del diseño es la proporción de la varianza del muestreo aleatorio simple explicada por la varianza del muestreo por conglomerados:

$$ED = \frac{S^2_{MC}}{S^2_{MAS}}$$

- ED = efecto del diseño
- S^2_{MC} = varianza del muestreo por conglomerados
- S^2_{MAS} = varianza del muestreo aleatorio simple

El efecto del diseño también se expresa así:

$ED = 1 + (b-1) \rho$, donde b es el tamaño del conglomerado y

$$\rho = \frac{bS^2_c - S^2_T}{(b-1)S^2_T}$$

Es decir, el coeficiente ρ está en función de la varianza de las puntuaciones medias de las escuelas, de la varianza total o varianza del conjunto global de elementos (estudiantes en nuestro caso) considerados y del tamaño de los conglomerados, que es fijo (30 ó 20, según los casos en los trabajos del Servicio de Evaluación). Como es fácilmente deducible de la fórmula del ρ ,

a mayor varianza, heterogeneidad, entre los centros, mayor coeficiente ρ y, por consiguiente, a mayor ρ , mayor número de conglomerados y mayor número de ρ sujetos tiene que tener la muestra (Ross y Postlethwaite, 1988, p. 12; Cochran, *op. cit.*, p. 300).

La estimación del coeficiente ρ puede hacerse en base a los datos recogidos en un estudio piloto o en aplicaciones anteriores a poblaciones similares a las del estudio en cuestión.

El tamaño de la muestra por conglomerados (N_{MC}) depende directamente de la muestra del muestreo aleatorio simple (N_{MAS}) y del efecto del diseño ED , como decíamos.

$$N_{MC} = N_{MAS} \times ED$$

$$N_{MC} = N_{MAS} [1 + \rho (b-1)]$$

En el caso de los estudios del Servicio de Evaluación del CIDE, que trabaja a un nivel de confianza del 95,5 por 100, con un margen de error del 5 por 100 y que tiene unos coeficientes ρ entre .2 y .3, las muestras serían:

- a) COU: $N_{MC} = 400 [1 + .25 (30-1)] = 400 (1 + 7.25) = 400 (8.25) = 3.300$
- b) FP: $N_{MC} = 400 [1 + .25 (20-1)] = 400 (1 + 4.75) = 400 (5.75) = 2.300$
- c) EGB: $N_{MC} = 400 [1 + .25 (30-1)] = 400 (1 + 7.25) = 400 (8.25) = 3.300$

En suma, los N_s mínimos de nuestras muestras deberían ser:

- COU: 3.300 sujetos y 110 conglomerados.
- FP: 2.300 sujetos y 115 conglomerados.
- EGB: 3.300 sujetos y 110 conglomerados.

De hecho, las muestras reales han sido aumentadas considerablemente en COU y FP y se ha mantenido en EGB, quedando de la siguiente forma:

COU		FP		EGB	
Sujetos	Centros	Sujetos	Centros	Sujetos	Centros
6.000	200	3.000	150	3.300	110

3. Procedimiento de muestreo

Como es bien sabido por quienes siguen nuestros estudios, los muestreos realizados en el Servicio de Evaluación son *bietápicos*. En una primera fase la unidad muestral es el centro y en una segunda la unidad muestral es el alumno.

En la *primera fase* se utiliza un muestreo *aleatorio estratificado con afijación proporcional y por conglomerados*.

En esta fase se empieza definiendo lo que llamamos *Unidades Primitivas de Muestreo*, que están delimitadas por la transferencia o no de las competencias educativas, por lo que las Unidades Primitivas de Muestreo son: territorio MEC, Andalucía, Canarias, Cataluña, Comunidad de Valencia, Euskadi y Galicia.

Dentro de cada Unidad Primitiva se llevan a cabo muestreos aleatorios estratificados con afijación proporcional. Las variables tenidas en cuenta a la hora de la estratificación son el tamaño del hábitat, y la titularidad de los centros. El hábitat se divide en:

- 10.000 hab.
- de 10.001 a 50.000 hab.
- de 50.001 a 500.000 hab., incluyendo las capitales de provincia, aunque no lleguen a 50.001 hab.
- + 500.000 hab.
- Madrid y Barcelona.

Los *tipos de centro* utilizados son: públicos y privados.

También se utiliza el *tamaño del centro* a la hora de asignar probabilidades a cada elemento de la muestra. Es decir, la probabilidad de que un centro salga elegido en la muestra depende no sólo del estrato de hábitat al que pertenece y del tipo de centro que sea, sino también de su tamaño. Las dos primeras variables se consideran de manera estratificada y la segunda, no. La probabilidad asignada según el tamaño depende directamente del propio tamaño de cada centro.

Por consiguiente, el proceso seguido en cada Comunidad Autónoma es el siguiente:

a) Asignación proporcional por estratos según hábitat y tipo de centro, de tal forma que en el territorio MEC, por ejemplo, quedaría una tabla como la 1.

b) A continuación se eligen los centros de cada casilla de la tabla, dependiendo la probabilidad de ser elegidos de su tamaño. Es decir:

$$p_i = \frac{n_i}{N_{Ei}}$$

- p_i = probabilidad de cada centro
- n_i = tamaño de cada centro
- N_{Ei} = tamaño del estrato correspondiente

TABLA 1

Población y muestra de COU en territorio MEC

	PUBLICOS			PRIVADOS		
	<i>Población</i>	<i>Muestra alumnos</i>	<i>Muestra centros</i>	<i>Población</i>	<i>Muestra alumnos</i>	<i>Muestra centros</i>
Menos de 10.000	7.572	180	6	1.603	30	1
10.001-50.000	13.408	330	11	2.220	30	1
50.001-500.000	38.003	960	32	12.841	240	8
Más de 500.000	3.533	90	3	2.322	60	2
Madrid	19.598	480	16	18.024	360	12
TOTALES	82.114	2.040	68	37.010	720	24

La forma de llevar a cabo este proceso implica la siguiente secuencia:

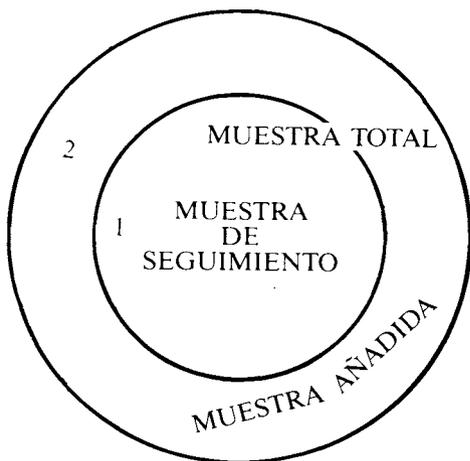
1.º Se averigua el número de alumnos que tiene cada centro del curso a evaluar (tamaño del centro).

2.º Se hace una acumulación de frecuencias por centro, con lo que cada centro se queda con tantos números asignados de manera correlativa, como sujetos tenga. Así, por ejemplo, supongamos que hay tres centros con 100, 75 y 150 alumnos, respectivamente. El centro 1 tendría asignados los números comprendidos entre 1 y 100, el centro 2 los comprendidos entre 101 y 175 y el centro 3 entre 176 y 325.

3.º Por último, se generan números aleatorios y aquel centro que incluya, entre los números asignados en el paso anterior, el número generado aleatoriamente pasa a formar parte de la muestra.

En el caso de las muestras de COU y FP se pretende, como ya se ha señalado, por una parte, evaluar el segundo ciclo de la Reforma de las EE. MM. y, por otra, hacer un diagnóstico de la situación educativa al final del segundo ciclo. Para cumplir el primer objetivo nos ha parecido muy importante continuar el estudio longitudinal que comenzamos con los mismos alumnos

en 1985. Por esta razón, el procedimiento muestral se ha complicado en parte. Se ha querido, desde un primer momento, utilizar a los alumnos de seguimiento como parte de la muestra total que nos sirve para obtener resultados de la situación al final de la EM. De esta forma, se abarataban los costes, como puede colegirse fácilmente. De manera gráfica, esta situación podría representarse así:



Para resolver el problema originado de la situación mencionada, se procedió de la siguiente forma: se eligió una muestra total teórica; a continuación de esa muestra teórica, según tamaño del hábitat y del tipo de centro, se fueron deduciendo los centros de seguimiento, de forma que sólo hemos elegido de manera aleatoria los centros necesarios para completar la muestra teórica; es decir, los pertenecientes al círculo 2 del gráfico anterior. Si, por ejemplo, en la casilla de los centros públicos del estrato de más de 50.000 habitantes del territorio MEC se necesitaban 32 centros de la muestra teórica y teníamos 10 de la de seguimiento, sólo hemos tenido que elegir aleatoriamente 22, para completar la muestra.

En la *segunda fase* se ha considerado como unidad muestral el alumno. El tipo de muestreo utilizado ha sido el *aleatorio sistemático*, a partir del coeficiente de elevación.

$$C = \frac{N_i}{n_i}$$

- N_i = tamaño de la población de cada centro
- n_i = tamaño de la muestra de cada centro

En nuestro caso, el tamaño de la muestra es el mismo para todos los centros de la misma población (BUP, FP II, EGB).

Por tanto, en un centro de BUP con 180 alumnos de COU y Bachillerato Experimental se elegirían de manera sistemática los alumnos múltiplos de 6 en las listas del colegio, ya que $6 = \frac{180}{30}$.

En los centros de seguimiento los alumnos ya están elegidos previamente, porque forman parte del estudio longitudinal, pero en su momento, al medir la línea base, pretest, se utilizó exactamente el mismo procedimiento.

REFERENCIAS BIBLIOGRAFICAS

- AZORÍN, Francisco, y SÁNCHEZ-CRESPO, José Luis (1986): *Métodos y aplicaciones del muestreo*, Alianza Editorial, Madrid.
- COCHRAN, William G. (1985): *Técnicas de muestreo*, CECSA, México.
- KISH, L. (1965): *Survey Sampling*, Wiley, Nueva York.
- ROSS, Kenneth N. (1987): «Sample Design», en *International Journal of Educational Research*, vol. 11, núm. 1, Oxford.
- ROSS, Kenneth N., y POSTLETHWAITE, T. Neville (1988): *Sample Design Procedures for the IEA International Study of Reading Literacy*, IEA.

TEXTOS CLASICOS