

EL MUESTREO Y SU PROBLEMÁTICA EN LAS EVALUACIONES DE PROGRAMAS INSTITUCIONALES

José Luis Gaviria Soto

Universidad Complutense de Madrid

I. INTRODUCCIÓN

Cuando se aborda la evaluación de resultados en los programas institucionales, especialmente en los niveles básicos de la educación, se siente siempre la tentación de evaluar a todos y cada uno de los individuos sobre los que actúa el sistema. Esta posibilidad puede estar en ocasiones justificada. Se trata del caso en el que la población de referencia es tan pequeña que la satisfacción de ciertos niveles de precisión impone una muestra cuyo tamaño es no mucho más pequeño.

Sin embargo por regla general la extracción de muestras presenta numerosas ventajas que hacen recomendable y beneficioso el esfuerzo. Siguiendo a Som (1996), podemos afirmar que las técnicas de muestreo tienen la ventajas de:

- mayor economía,
- intervalos de tiempo más cortos entre las distintas recogidas de información,
- más amplio espectro de la información,
- mayor calidad del trabajo,
- y posibilidad de la valoración de la fiabilidad de los resultados

Ciertamente si nos centramos en un grupo comparativamente pequeño respecto de la población, podremos recoger más información de cada individuo, asegurarnos que los datos obtenidos de cada sujeto son fiables, llevar a cabo la recogida de información más frecuentemente, y todo ello a un menor coste.

Pero por supuesto los procedimientos de muestreo pueden estar contraindicados en algunas ocasiones. Por ejemplo, si como resultado de la evaluación han de tomarse decisiones que repercuten directamente en la población evaluada, no es correcto

basar decisiones respecto a individuos en informaciones de grupo. En algunos casos esto resulta evidente. No podemos afirmar nada respecto de la competencia profesional de un docente, o sobre los conocimientos de un alumno, basándonos en los resultados en la evaluación de la competencia profesional de los docentes de una región, o en la evaluación de los conocimientos de todos los alumnos de un país. Ni siquiera aunque tuviésemos los resultados específicos de ese sujeto estaría justificada dicha práctica. No ya sólo por cuestiones morales de respeto a la confidencialidad, sino sobre todo porque lo que puede representar una muestra muy informativa de las conductas o conocimientos de un sujeto de una muestra, puede no serlo de las conductas o conocimientos del sujeto «per se». Esto es especialmente cierto en los diseños «matrix sampling» en los que de cada sujeto se obtiene sólo una parte de la información relevante.

Pero aunque no es tan evidente, tampoco pueden tomarse decisiones sobre un centro basándonos en los resultados de uno de sus grupos que fue seleccionado para pertenecer a una muestra nacional. Ese elemento muestral es adecuado para la evaluación nacional, y aporta una cierta cantidad de información a ese propósito. Sin embargo no es representativa del centro del que procede, ya que el dominio no coincide con el mismo. Ésta es una de las muchas tentaciones en las que pueden caer los representantes políticos o los administradores cuando se encuentran con la información que proporciona una evaluación. Son muchas las posibilidades de uso indebido de la información procedente de los muestreos, y por eso es tan importante que desde su inicio estén bien definidos los objetivos de la evaluación y que los sujetos y las instituciones evaluadas sepan de antemano cuál es el uso que se va a dar a esa información, y sobre todo, qué uso no se va a dar jamás a la misma.

Ciertamente no es este el lugar para desarrollar la teoría del muestreo, ni los distintos procedimientos. Sin embargo podemos destacar aquellos conceptos relevantes al muestreo para la evaluación, y aclarar algunos problemas que en la misma se presentan.

Nos vamos a referir por tanto a la población objetivo, a los dominios, a los estratos, a los conglomerados, a la correlación intra-clase, al muestreo de los conglomerados, al tamaño de la muestra o las ponderaciones de los datos y sus efectos, y a un caso especial en el muestro.

2. LA POBLACIÓN OBJETIVO

Cuando se planifica una evaluación, una cuestión que aparentemente está siempre clara es cuál es la población objetivo, es decir, cuál es el grupo de sujetos, o en general unidades de análisis, a las que la información recabada por la evaluación va a hacer referencia. Por lo general una primera definición permite acercarse bastante al objetivo buscado. Por ejemplo, «todos los alumnos correspondientes al curso modal de 14 años». Esta primera definición determina cuál es la **población implícitamente excluida**. Se trata de todos los alumnos, independientemente de su edad, que correspondan a cursos cuya edad modal no sea 14 años.

Sin embargo se presentan dudas que es preciso resolver. ¿Deben incluirse los alumnos que están implicados en programas complementarios de educación especial? En

ese caso debe acudir al objetivo de la evaluación para responder a la cuestión. Por ejemplo, podría ser el objetivo del estudio la evaluación de la acción del sistema excluido el subsistema de educación especial, cuya evaluación se realizaría en fase distinta. Eso entonces nos lleva a determinar la **población explícitamente excluida**. Así, en el estudio del INCE (1998) «se determinó como población excluida los alumnos con necesidades educativas especiales tales que impidiesen contestar la prueba en igualdad de condiciones espaciales y temporales con los demás alumnos del grupo. Se instruyó a los aplicadores para que aplicasen a estos alumnos las pruebas igual que al resto de sus compañeros, y sólo *a posteriori* deberían identificar dichas pruebas para su exclusión».

3. LOS DOMINIOS

Un dominio se refiere a una subdivisión de la población para la que se han planeado, designado y seleccionado diferentes muestras, con el objeto de proporcionar resultados específicos para la misma con un conocido margen de error. (Kish, 1987). Esto quiere decir que los resultados se proporcionarán con una precisión determinada para cada dominio, y que el resultado global procederá de una combinación adecuada de los resultados de todos los dominios. Lo habitual es que los dominios coincidan con ciertas unidades político-administrativas, tales como regiones, comunidades, provincias, etc.

Si se decide la utilización de dominios, entonces el muestreo en la práctica se realiza como si cada uno de los mismos fuese una población de la que hay que extraer una muestra representativa. Naturalmente tiene que ocurrir que la muestra agregada de todos los dominios a su vez cumpla los requisitos de representatividad establecidos para la población total.

4. LOS ESTRATOS

Es conocido que cuando se clasifica a una población según ciertas características demográficas, los grupos así formados tienden a ser más homogéneos respecto a otras variables de interés que el conjunto de la población. Así por ejemplo, la varianza en altura de los sujetos de alto nivel socio-económico familiar es menor que la varianza en altura de la población en general. Los grupos así formados se denominan estratos, y su utilización permite obtener mayor cantidad de información de la muestra utilizando menos sujetos. A esto se le llama efecto diseño, y consiste en la disminución o el aumento del tamaño de la muestra respecto al muestreo aleatorio simple debido al tipo de muestreo utilizado.

La utilización de estratos es muy recomendable en la evaluación, aunque también es posible que un número excesivo de estratos haga la extracción de la muestra excesivamente compleja, compensando negativamente las ventajas obtenidas por la disminución de la varianza. En teoría los estratos dividen horizontalmente a los dominios. Es posible incluso que cada dominio tenga una estratificación distinta, reflejando los intereses diversos de los administradores. Pero también es posible que

una cierta combinación de estratos se convierta en un subdominio. Por ejemplo, es frecuente que los administradores deseen conocer con un nivel predeterminado de precisión los rendimientos en los centros públicos y en los centros privados en una región determinada. En ese caso los alumnos de centros públicos y de centros privados de dicha región se convierten en subdominios de los que habrá que extraer muestras representativas.

¿Cuántos estratos pueden definirse en un procedimiento de muestreo? La tentación vuelve a ser la de definir el mayor número posible. Sin embargo, como regla práctica, diremos que el número máximo de estratos definidos no debe ser mayor que la mitad del número de unidades primarias de muestreo. El sentido de esta regla consiste en que debe ser posible calcular el error muestral, y para ello es necesario al menos dos de estas unidades por cada estrato.

Las variables de estratificación más habituales son la clasificación administrativa (provincia, estado, región, comunidad autónoma), ámbito rural o urbano, nivel socioeconómico del distrito escolar o del entorno de la escuela, naturaleza pública o privada del centro, tamaño del centro, etc.

5. LOS CONGLOMERADOS Y LA CORRELACIÓN INTRA-CLASE

Una particularidad que el sistema educativo comparte con otras organizaciones sociales es que los sujetos forman grupos que comparten tiempo y actividades. De hecho acceder a un solo individuo es prácticamente tan costoso como acceder a todo el grupo primario, en este caso el aula. Lógicamente conviene entonces recabar información no de un solo sujeto, sino de los sujetos de toda el aula. Este grupo primario es un conglomerado. El hecho de que todos los individuos del conglomerado compartan tiempo, y actividades, además de que probablemente proceden del mismo entorno social, etc., hace que estos sujetos sean entre sí más homogéneos que con relación a los demás individuos de la población. Pero el efecto de esta mayor homogeneidad en los conglomerados tiene un efecto contrario al que se producía en los estratos.

Supongamos que para obtener una muestra representativa de una población necesitamos una muestra de 300 sujetos por muestreo aleatorio simple. Si el investigador accede a centros en los que hay 30 sujetos por aula, podría parecer que visitando 10 centros obtendría toda la información necesaria.

Pero imaginemos por un momento que la acción del sistema educativo es tan fuerte que todos los alumnos de un mismo centro fuesen exactamente iguales en las variables de interés, es decir, que la varianza dentro de los centros fuese cero, y que todas las diferencias en la población fuesen debidas a la diferente influencia de cada centro. En ese caso, la información que proporciona un centro se obtiene con un solo sujeto, y el resto de los individuos proporcionaría información redundante. Esto quiere decir que cuando el investigador evalúa a los treinta sujetos de un centro la información obtenida no es mayor que la que proporciona uno sólo. Por tanto después de visitar 10 centros, y aunque haya evaluado a sus trescientos sujetos, tiene información correspondiente sólo a 10 de los 300 individuos necesarios para que su muestra sea representativa.

En el otro extremo, si los sujetos fuesen completamente heterogéneos dentro de las aulas, y los centros fuesen completamente equivalentes unos y otros, es decir, si toda la varianza en la población fuese debida sólo a las diferencias intra-aula, bastaría con recoger la información de esos 10 centros para completar la muestra.

El índice que mide la mayor o menor homogeneidad de los sujetos en los conglomerados es la correlación intra-clase. De hecho la correlación intra-clase es una medida de la proporción de la varianza total que es debida a las diferencias dentro de los conglomerados.

En el muestreo por conglomerados el efecto diseño tiene un efecto negativo. Esto quiere decir que cuanto mayor sea la correlación intra-clase, mayor deberá ser el tamaño de la muestra final respecto de lo que hubiera sido si se tratase de muestreo aleatorio simple.

6. EL MUESTREO DE LOS CONGLOMERADOS

Que la muestra sea representativa supone que todos los sujetos de la población objetivo, o en su caso del dominio, tengan la misma probabilidad de pertenecer a la muestra. Cuando las unidades primarias de muestreo son los conglomerados, es decir, las aulas, si todas las aulas tienen la misma probabilidad de formar parte de la muestra, entonces los sujetos que están en aulas más pequeñas tienen mayor probabilidad de formar parte de la muestra que la que en puridad les corresponde. Esto quiere decir entonces que las aulas no deben tener la misma probabilidad de formar parte de la muestra. Las aulas pequeñas deben tener menor probabilidad y las grandes mayor. Es decir, el muestreo debe realizarse con probabilidad proporcional al tamaño del aula.

7. EL TAMAÑO DE LA MUESTRA

Es una creencia muy generalizada entre los no especialistas que lo que determina esencialmente el tamaño de la muestra es el tamaño de la población a que esa muestra tiene que representar. Según esto, para una población de 100.000 sujetos, correspondería una muestra de doble tamaño que para una población de 50.000 individuos. Sin embargo, no es así. Para un nivel de confianza dado, los parámetros que más afectan al tamaño de la muestra son el error muestral máximo que se está dispuesto a asumir, la desviación típica de la variable medida, y la correlación intra-clase en el caso del muestreo por conglomerados. En la gráfica 1 vemos la relación que existe entre el tamaño de la desviación típica de la población y el tamaño necesario de la muestra para un error muestral máximo de 1,5 puntos a un nivel de confianza del 95%, en dos poblaciones distintas, una con 12.000 sujetos, y otra mucho más grande de 450.000. Como vemos para una sigma de 9,5 puntos por ejemplo, el tamaño de la muestra que garantiza un error no mayor de 1,5 es de 158 sujetos en una población de 12.000 sujetos, y de 160 en una población de 450.000.

Por tanto, garantizando un cierto valor de error muestral máximo el tamaño de la población no va a afectar de forma sustancial al tamaño de la muestra.

Por esta razón las diferencias en el tamaño de las muestras utilizadas en los distintos dominios definidos en una evaluación, provincia, región, comunidad autónoma, no vienen determinadas por el tamaño relativo de la población de cada una de ellos, sino por el error muestral máximo que se esté dispuesto a asumir en cada una de ellos.

Por distintos motivos, especialmente cuando los dominios definidos coinciden con unidades político-administrativas distintas, suele ocurrir que se definen distintos márgenes de error para cada uno. Esto depende del tipo de información en que la burocracia de cada una de estas unidades tiene mayor interés, del presupuesto disponible y de la influencia que pueda tener sobre el órgano que gobierna la planificación y realización de la evaluación.

Para la determinación del tamaño muestral es necesario no sólo el error muestral máximo, sino también el valor de la correlación intra-clase. La estimación de los valores de este parámetro pueden obtenerse, bien por referencias previas, o bien por la aplicación de una prueba piloto. A modo de ejemplo, en la evaluación de la secundaria realizada por el INCE en 1997, la aplicación de la prueba piloto permitió obtener una estimación de los valores de la correlación intra-clase para los grupos de 14 y de 16 años, siendo esos valores de 0,13 y 0,22 respectivamente.

La práctica habitual consiste en fijar un determinado nivel de confianza, y realizar una primera estimación del muestreo que sería necesario en caso de Muestreo Aleatorio Simple (M.A.S.). A continuación se calcula la corrección necesaria por el efecto diseño debido al Muestreo Aleatorio por Conglomerados, según se explica en el apartado siguiente.

8. PROCEDIMIENTO DE MUESTREO

El procedimiento de selección de la muestra se basa en el supuesto de equiprobabilidad de selección de todos los sujetos. Si no atendiésemos a la estructura natural de los datos, es decir, agrupados en aulas, a su vez agrupadas en colegios, determinaríamos el máximo error muestral que estaríamos dispuestos a admitir para cierto nivel de confianza, y a partir de estos parámetros calcularíamos el tamaño de la muestra mediante la expresión

$$n = \frac{N \frac{k^2}{e^2}}{N + \frac{k^2}{e^2}}$$

donde N es el tamaño de la población de referencia, k es la puntuación típica correspondiente al nivel de confianza seleccionado, y e es el error muestral expresado en desviaciones típicas. Esta expresión contiene ya la corrección por poblaciones finitas.

Esta fórmula es de aplicación para cada uno de los dominios.

La muestra total será la agregación de las muestras de todos los dominios.

Dado que es de interés proporcionar información acerca de algunos estratos de la población, se dividirá a ésta en varios estratos. Es el caso por ejemplo de las provincias, de líneas curriculares, de la titularidad de los centros, o cualesquiera otros. Por tanto, para proporcionar datos acerca de cada uno de los estratos, el número de sujetos que se determina para un dominio se reparte proporcionalmente entre los distintos estratos en que se divide su población, de manera proporcional al tamaño de estos.

Pero los sujetos están agrupados, dentro de cada estrato, en conglomerados, constituidos por las aulas, y estas aglomeradas en colegios.

Por esta razón cuando los sujetos forman conglomerados, el tamaño de la muestra en comparación con el muestreo aleatorio simple, debe aumentar en función del tamaño del conglomerado y de la autocorrelación. A ese incremento en el tamaño se le llama efecto diseño, y para el caso que nos ocupa, viene dado por $F=1+(B-1)\rho$, donde B es el tamaño del conglomerado, y ρ es la autocorrelación.

Por tanto el tamaño muestral definitivo será $n'=Fn$. Como los sujetos están agrupados de B en B , se seleccionarán por tanto n'/B aulas distintas.

Para ilustrar este punto tomaremos un caso de la mencionada evaluación de la secundaria (INCE 1998). Para la comunidad autónoma de La Rioja los cálculos serían los siguientes. En el grupo de 14 años hay 3.541 alumnos en la población. Si queremos que el nivel de confianza sea del 95,5%, y el error muestral máximo sea de 0,12 sigmas, obtenemos un tamaño para muestreo aleatorio simple de 257,57 alumnos. Como el valor estimado de rho para los 14 años era de 0,13, el efecto diseño era igual a 4,12. Por tanto en un muestreo por conglomerados le correspondería un tamaño de muestra de 1.061 alumnos.

Si comparamos ese resultado con el tamaño de muestra que correspondería a una comunidad proporcionalmente mucho más grande como Cataluña, podemos comprobar lo poco que el tamaño de la población afecta al tamaño de la muestra. Así, manteniendo los demás parámetros iguales, es decir, el nivel de confianza y el error muestral máximo, dado que Cataluña tiene 73.293 alumnos de 14 años, en un muestreo aleatorio simple le correspondería un total de 267,72 alumnos (véase la poca diferencia con los 257,57 que correspondían a La Rioja). Aplicando la corrección del efecto diseño, con el mismo rho, tenemos que en el muestreo por conglomerados le correspondería un total de 1.103 alumnos de 14 años. (En la realidad la comunidad autónoma de Cataluña decidió extraer la muestra de 14 y la de 16 años con menor error que otras comunidades autónomas, lo que explica el mayor tamaño de su muestra de 14 años. Lo mismo se aplica a la muestra de 16 años).

Un caso muy habitual en la evaluación de resultados educativos es el muestreo en dos etapas. En primer lugar se seleccionan n'/B centros, con probabilidad proporcional a su tamaño. Y dentro de cada centro, se selecciona en la siguiente etapa un aula completa con probabilidad también proporcional a su tamaño.

Este procedimiento debe repetirse para cada población objetivo distinta. No importa que dos poblaciones objetivo estén localizadas en las mismas unidades primarias de muestreo, en nuestro caso colegios. Se trata de mantener el supuesto de equiprobabilidad. Si la selección de un centro se hace con probabilidad proporcional al número de

sujetos de 14 años que tiene, si seleccionásemos también a otra muestra de sujetos de 16 años ocurriría que en esta población no todos los sujetos tendrían la misma probabilidad de ser incluidos en la muestra. El procedimiento de muestreo debe por tanto repetirse para cada población objetivo distinta.

Con relación a la muestra global para toda la población evaluada, se constituye por la agregación de las muestras de los dominios. Esta forma de muestreo implica un muestreo no proporcional, ya que el número de sujetos de cada dominio puede no ser proporcional a N_i/N . Esto lleva de forma inmediata a la necesidad de calcular la media global con ciertos coeficientes de ponderación, que hacen que cada dominio esté representado en la media total de forma proporcional a su tamaño relativo de la población.

Las pequeñas diferencias entre las muestras planificadas y las muestras obtenidas dentro de los subestratos de cada dominio también aconsejan utilizar ciertas ponderaciones para lograr que en la media de cada dominio estén respresentadas proporcionalmente las poblaciones de cada subestrato.

Para más información puede consultarse Blalock (1972) y Kalton (1983).

9. DETERMINACIÓN DE LAS PONDERACIONES

Como consecuencia del muestreo por dominios, aquéllos que son más pequeños quedan sobrerrepresentados, mientras que los más grandes quedan subrepresentados.

Para corregir ese efecto es preciso que a cada dominio se le asigne un peso de forma que su muestra tenga la importancia que le corresponde en el total. Para que las estimaciones sean correctas ese peso tiene que ser tal que su media total sea igual a 1.

Hay otra fuente de desviaciones que hacen necesario el uso de otras ponderaciones. Entre la muestra planificada y la muestra prevista existen ciertas discrepancias producidas por las contingencias propias del proceso de toma de datos. Recordemos que se trata de un muestreo por conglomerados. Sólo aproximadamente puede anticiparse el número de sujetos en cada conglomerado. Dentro de cada dominio, se definen cuáles son los estratos de interés. Dado que en última instancia el tamaño final de la muestra depende de los verdaderos tamaños de los conglomerados, se dan ciertas pequeñas discrepancias entre el tamaño planificado y el tamaño obtenido. Como consecuencia los elementos de cada estrato, dentro de un mismo dominio, deben ser también ponderados para que ninguno de ellos tenga un mayor peso que aquél que le corresponde. Estos pesos intradominios, que deben ser utilizados siempre que se realicen comparaciones entre distintos estratos de un mismo dominio, deben cumplir también la condición señalada anteriormente, por la cual la media de estos pesos debe ser igual a uno.

Ciertamente cuando se realizan las estimaciones de las medias globales, debe tenerse en cuenta tanto las correcciones debidas a las desviaciones de los estratos dentro de cada dominio, como las ponderaciones atribuibles a cada dominio respecto del conjunto.

En general, para calcular el peso que corresponde a un cierto estrato i , procederemos de la siguiente manera.

Si llamamos

a_i = tamaño del estrato i en la muestra

A_i = tamaño del estrato i en la población.

n = tamaño de toda la muestra

N = tamaño de la población.

Entonces en un dominio con k estratos, la media total del dominio es

$$\bar{x}_t = \frac{\sum_i^k A_i \frac{\sum_j^{a_i} x_{ji}}{a_i}}{N} = \sum_i^k \frac{A_i}{a_i N} \sum_j^{a_i} x_{ji} \quad \text{Luego,}$$

$$p_i = \frac{A_i}{a_i N}$$

Aquí la media se obtiene como la suma ponderada de todas las puntuaciones. Pero, nos conviene un peso para cada puntuación, de forma que la media, no la suma, de todas las puntuaciones nos permita obtener la media total. Así que para dividir por n todo, tenemos que multiplicar por n el peso también.

De esta forma obtenemos un peso tal que multiplicando el cociente del tamaño del estrato en la muestra respecto a la muestra, éste es proporcional al mismo cociente en la población.

$$\frac{a_i}{A_i} p_i = \frac{n}{N}$$

$$p_i = \frac{nA_i}{Na_i}$$

Estos pesos tienen la propiedad de que su media es 1. Si las ponderaciones no tienen como media 1, entonces los niveles de significación de que informan los procedimientos estadísticos no son válidos.

10. LA PONDERACIÓN Y LA MEDIA DE LOS ESTRATOS

Cuando se calculan las medias de los estratos individualmente utilizando los pesos, las medias de los estratos no varían.

$$\bar{x}_i = \frac{\sum_j^{a_i} p_i x_{ji}}{\sum_j^{a_i} p_i} = \frac{p_i \sum_j^{a_i} x_{ji}}{a_i p_i} = \frac{\sum_j^{a_i} x_{ji}}{a_i}$$

En cuanto a la media total usando los pesos,

$$\text{y como } a_i p_i = \frac{A_i}{N} n$$

$$\begin{aligned} \bar{x}_t &= \frac{\sum_i^k \sum_j^{a_i} p_i x_{ji}}{\sum_i^k \sum_j^{a_i} p_i} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\sum_i^k a_i p_i} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\sum_i^k \frac{A_i}{N} n} = \\ &= \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\frac{n}{N} \sum_i^k A_i} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{\frac{n}{N} N} = \frac{\sum_i^k p_i \sum_j^{a_i} x_{ji}}{n} \end{aligned}$$

11. EFECTO DE LA PONDERACIÓN SOBRE LA VARIANZA

En los procedimientos estadísticos utilizados habitualmente la variable de ponderación se utiliza para «replicar» los casos. De esta manera tenemos para un estrato.

$$\sigma_{p_i x_i}^2 = \frac{\sum_j^{a_i} (x_{ji} - \bar{x}_i)^2 p_i}{\sum_j^{a_i} p_i - 1} = \frac{p_i \sum_j^{a_i} (x_{ji} - \bar{x}_i)^2}{a_i p_i - 1} = \frac{\sum_j^{a_i} (x_{ji} - \bar{x}_i)^2}{a_i - \frac{1}{p_i}}$$

Esta es la estimación de la varianza dentro del subestrato i . Así cuando $p_i > 1$, la varianza de las puntuaciones ponderadas es inferior a la varianza de las puntuaciones sin ponderar. Sin embargo esa diferencia es muy pequeña cuando los valores de la ponderación son los que proceden de los pequeños errores de muestreo en los subestratos.

12. ALGUNOS CASOS ESPECIALES

Por último nos vamos a referir a algunos casos especiales que surgen en la evaluación. Un problema muy interesante en el muestreo aparece cuando tratamos de evaluar sujetos que pertenecen a un grupo muy minoritario, como por ejemplo aquéllos que fueron definidos como parte de la población excluida. En ocasiones es difícil

determinar cuántos son los sujetos que están en esas condiciones en la población de referencia. Es el mismo caso cuando tratamos de determinar por ejemplo cuál es, en una población, la proporción de sujetos que presentan algunas características especiales, tales como una alta capacidad matemática, o especial habilidad para la música, o tienen problemas de aprendizaje, etc.

La determinación de la proporción en la población, es decir, el parámetro π por los medios tradicionales puede ser muy costoso en términos de tamaños de la muestra o de error de estimación. El tamaño de la muestra viene dado por

$$n = \sigma_p^2 \frac{z_{\alpha/2}^2}{E^2}$$

como puede verse el tamaño de la muestra depende radicalmente del inverso de la precisión que deseamos obtener. Si queremos tener una precisión de 1 por mil al 95% de confianza, el tamaño de la muestra será 4000000 de veces la varianza de p . Para un fenómeno que ocurra alrededor de 5 casos cada mil sujetos de la población, la muestra necesaria para estimar ese parámetro es de cerca de 20.000 sujetos. Esto hace inviable la mayoría de estudios de este tipo.

Existen sin embargo alternativas muy interesantes. Por ejemplo los procedimientos de captura-recaptura, muestreo inverso para proporciones, muestreo de redes. Mencionaremos el segundo, por ser el más adecuado al tema tratado. Los otros dos pueden consultarse en Som (1996).

En el muestreo inverso se procede a seleccionar sujetos para la muestra hasta que se obtiene un número n' determinado de casos buscados. Si n es el tamaño final de la muestra, la estimación del parámetro buscado es

$$p = \frac{n' - 1}{n - 1}$$

la varianza del estimador viene dada por $s_p^2 = \frac{p(1-p)}{n-2} \left(1 - \frac{n-1}{N}\right)$, siendo N el tamaño de la población.

Existen estimadores también para variables continuas en los mismos casos.

13. BIBLIOGRAFÍA

- Blalock, H.M. (1972). *Social Statistics*, New York: McGraw-Hill.
- INCE (1998) 2. *Los Resultados escolares*. Ministerio de Educación y Cultura. Madrid.
- Kalton, G. (1983). *Introduction to Survey Sampling*, Sage, London.
- Kish, L. (1987). *On the future of survey sampling*. En N.K. Namboordi (Ed.) *Survey sampling and measurement*. Academic Press. N.Y.
- Som, R.K. (1996). *Practical Sampling Techniques*. Marcel Dekker, Inc. N.Y.

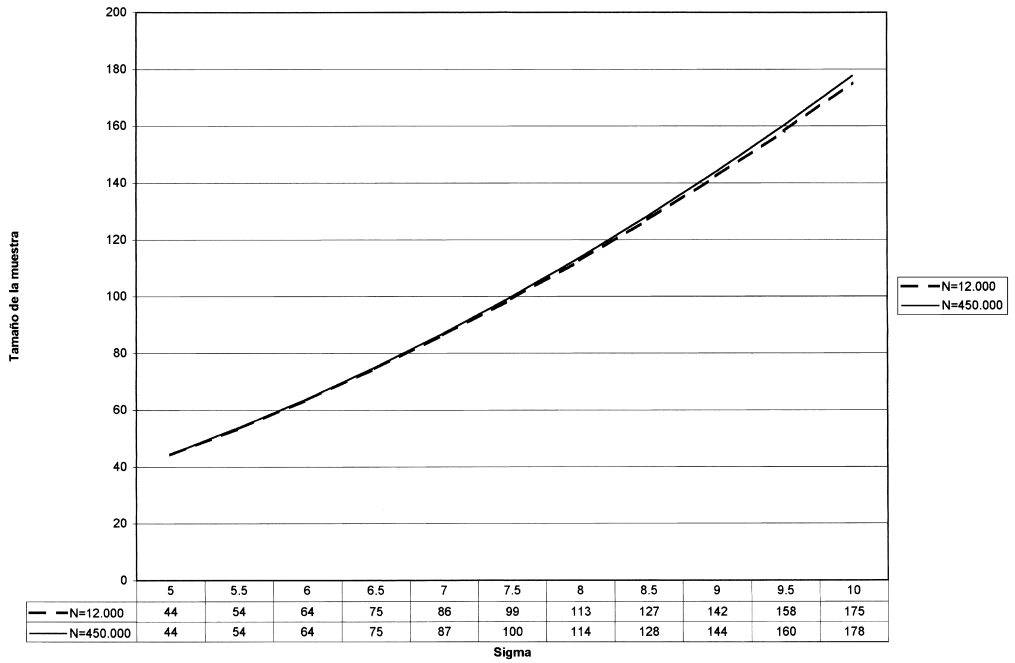


Gráfico 1
Relación entre desviación típica y tamaño de la muestra