

Muestreo: Herramientas de la investigación en el aula

Svetlana Ivanovna Rudnykh¹, Ramón Antonio Matos Mareño²

Grupo de Investigación Métodos Estadísticos Aplicados (MEA), Facultad de Ciencias Básicas,
Universidad del Atlántico, Km7 Antigua via Puerto, Colombia

Resumen

Este documento resume los temas y conceptos tratados en el curso: muestreo- una herramienta de la estrategia de investigación en el aula. Se presentan en detalle los conceptos de la teoría estadística del muestreo; estos se aplican a un ejemplo hipotético de investigación social cuya solución se logra mediante el muestreo estratificado.

Palabras clave: Investigación en el aula, muestreo estratificado.

Abstract

This document summarizes the subjects and concepts discussed in the course: Sampling- a tool of investigation in the classroom. The concepts of the theory of sampling appear in detail, these are applied on a hypothetical example of social investigation in the school, this obtained by the means of stratified sampling.

Key words: Investigation in the classroom, stratified sampling.

1. Introducción

La investigación realizada en el aula escolar es un espacio de reflexión académica, donde se puede debatir, indagar, cuestionar, proponer, consultar, interpretar, argumentar, compartir, cuestionar, en fin un número indeterminado de competencias que se puede desarrollar

¹M.Sc en Física. Email: svetarudn@hotmail.com

²Ph.D en Estadística. Email: ramonmatos_194@yahoo.es

con los estudiantes, todas ellas con un solo fin conocer los procesos sociales o naturales que se han desarrollado en nuestro entorno y en el mundo en general, y donde el estudiante puede construir y reconstruir los acontecimientos que a diario ocurren en la sociedad y que a veces no nos preguntamos por que ocurren; siendo una de las tantas funciones de las ciencias comprender, analizar y explicar el porque de esos eventos, es entonces la investigación en el aula una estrategia de enseñanza y aprendizaje muy útil para responder a esas preguntas.

Por otra parte, el muestreo es una herramienta de la investigación, cuya función básica es determinar que parte de la realidad en estudio (población o universo) debe examinarse con la finalidad de hacer inferencias sobre dicha población. El error que se comete debido al hecho que se obtienen conclusiones sobre cierta realidad a partir de la observación de sólo una parte de ella, se denomina error de muestreo. Obtener una muestra adecuada significa lograr una versión simplificada de la población, que reproduzca de algún modo sus rasgos básicos.

La intención de este breve curso es motivar a los participantes a profundizar en el estudio de la técnica estadística de muestreo para ser implementada en el marco de la estrategia didáctica de investigación en el aula.

2. Investigación en el aula

La didáctica definida como un proceso general de producción de conocimiento, basado en el tratamiento de problemas, que se apoya tanto en el conocimiento cotidiano como en el científico, que se perfecciona progresivamente en la práctica y que persigue unos fines educativos determinados (Porlán, 1999), promueve una formación de sujetos que se preguntan acerca de lo que sucede en el entorno, sujetos activos y con una actitud científica de búsqueda. Esto último implica saber preguntar, buscar, analizar, sistematizar, explicar, fundamentar. En suma, adquirir un pensamiento reflexivo y autónomo.

Esta concepción de aprendizaje supone que aprender implica pensar en la posibilidad de cambio y de desarrollo del pensamiento como un trabajo más complejo que la mera acumulación de datos e información.

La investigación en el aula facilita que el alumno se reconozca como sujeto participativo del aprendizaje y aprenda que el pensamiento no es una mera acumulación de datos sino un proceso metódico de cambio, dinámico y perfectible que permite ir construyendo estructuras cognoscitivas cada vez más abarcadoras.

Entre los propósitos educativos (referidos a la intencionalidad didáctica) que se plantea la investigación de los alumnos en el contexto del aula se cuentan los siguientes: fortalecer el proceso de aprendizaje, promover la construcción de conceptos, favorecer el orden y la jerarquización de la información, fomentar el pensamiento autónomo y crítico con la elaboración de conclusiones, despertar en los alumnos el deseo de aprender cosas nuevas.

Es bueno recalcar que la investigación en el aula apunta a fines educativos determinados, que distan mucho de los que persiguen los científicos eruditos. La ciencia es el ámbito de referencia del cual podremos extraer aquellos conocimientos que se consideran válidos para ser transmitidos en la escuela, y cierta metodología para llegar a alcanzarlos. El ámbito de la ciencia es el de la producción del conocimiento; el ámbito de la escuela es el de la reproducción del conocimiento (en términos sociales y no individuales). Sin embargo, las formas que se definan para que los alumnos lleguen a ese conocimiento determinarán, por un lado, el desarrollo del pensamiento formal y crítico que los estudiantes alcancen, y la real comprensión de ese conocimiento, por otro. (Furman & Zysman, 2001).

Considerando al aprendizaje como acción, reflexión, interacción y teniendo en cuenta los propósitos y fines que se de ella se derivan se comprende que el trabajo de investigación en el aula se inicia con el planteamiento de un problema a resolver, continua con la elaboración de una o varias hipótesis de trabajo. El profesor tiene que facilitar a sus alumnos el acceso al manejo de técnicas cualitativas y/o cuantitativas acordes a los objetivos planteados en la investigación.

Plantear un trabajo de investigación en el aula implica cierta rigurosidad en el cumplimiento de determinados pasos, a saber:

1. La elaboración de una o más hipótesis: todo trabajo de investigación requiere de una hipótesis que se habrá de probar.
2. La búsqueda de antecedentes de la investigación: es necesario realizar un trabajo de

relevamiento de las distintas fuentes de información e indagar acerca de su calidad. Conocer datos sobre el tema de investigación contribuirá a la elaboración del marco teórico.

3. Desarrollo de la investigación: la metodología que se seleccione dependerá del área y del tema a investigar. Puede ser experimental o mediante entrevistas, encuestas, por ejemplo. Será necesario conocer determinadas metodologías de manejo de datos.
4. La publicación de los resultados: una vez desarrollado el plan de la investigación, los alumnos tendrán que elaborar un informe o monografía que contenga los pasos que siguieron hasta llegar a la conclusión correspondiente.

3. Conceptos básicos del muestreo

POBLACIÓN. Es el conjunto formado por individuos o elementos que le podemos observar, medir una característica o atributo.

Ejemplos de población:

- El conjunto formado por todos los estudiantes de una escuela o colegio de Barranquilla.
- El conjunto de personas fumadoras de costa Caribe Colombiana.
- El conjunto de colegios del departamento del Atlántico.

VARIABLES. Son las características medibles u observables de cada elemento por ejemplo, su estatura, su peso, edad, sexo, número de estudiantes, etc.

Supongamos que nos interesa conocer el ingreso mensual promedio de las familias de los estudiantes de una escuela. Si el número de familias de la escuela es 537, bastaría encuestar cada familia de los estudiantes, sumar los 537 ingresos mensuales y dividirlo entre 537.

Pero este proceso puede presentar dificultades dentro de las que podemos mencionar:

- localizar y pesar con precisión cada estudiante,
- escribir todos los datos sin equivocaciones en una lista,
- efectuar los cálculos.

Las dificultades son mayores si en número de elementos de la población es infinito (familias en este caso), si los elementos se destruyen, si sufren daños al ser medidos o están muy dispersos, si el costo para realizar el trabajo es muy costoso.

Una solución a este problema consiste en medir solo una parte de la población que llamaremos **MUESTRA** y preguntar sobre el ingreso mensual medio en la muestra como una aproximación del verdadero valor del peso medio de la población.

EL TAMAÑO DE LA POBLACIÓN es la cantidad de elementos de esta y el **tamaño de la muestra** es la cantidad de elementos de la muestra. Las poblaciones pueden ser finitas e infinitas.

PARÁMETRO. Son las medidas o datos que se obtienen sobre la distribución de probabilidades de la población, tales como la media, la varianza, la proporción, etc.

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

ERROR ESTANDAR. 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 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.

INFERENCIA ESTADÍSTICA. Trata el problema de la extracción de la información sobre la población contenida en las muestras.

Para que los resultados obtenidos de los datos muestrales se puedan extender a la población, la muestra debe ser representativa de la población en lo que se refiere a la característica en estudio, o sea, la distribución de la característica en la muestra debe ser aproximadamente igual a la distribución de la característica en la población.

La representatividad en estadística se logra con el tipo de muestreo adecuado que siempre incluye la aleatoriedad en la selección de los elementos de la población que formaran la muestra. No obstante, tales métodos solo nos garantizan una representatividad muy probable pero no completamente segura.

Después de estos preliminares imprescindibles es posible pasa a tratar algunas de las formas que desde el punto de vista científico se puede extraer una muestra.

Al realizar un muestreo en una población podemos hablar de muestreos probabilísticos y no probabilísticos, en nuestro caso nos referiremos a los muestreos probabilísticos.

Existen dos formas de extraer una muestra de una población: con reposición y sin reposición.

MUESTREO CON REEMPLAZO: Es aquel en que un elemento puede ser seleccionado más de una vez en la muestra para ello se extrae un elemento de la población se observa y se devuelve a la población, por lo que de esta forma se pueden hacer infinitas extracciones de la población aun siendo esta finita.

MUESTREO SIN REEMPLAZO: No se devuelve los elementos extraídos a la población hasta que no se hallan extraídos todos los elementos de la población que conforman

la muestra.

DISEÑOS MUÉSTRALES BÁSICOS

Los métodos de muestreo probabilístico considerados básicos para las poblaciones finitas pueden combinarse, originando de esta manera los métodos polietápicos, que son más complejos de estudiar que los primeros. Entre los diseños muestrales básicos se encuentran:

a) Muestreo aleatorio simple.

En este método todas las muestras posibles, de un mismo tamaño n , tienen la misma probabilidad de ser seleccionadas, lo cual conlleva a que todos los elementos de la población también tengan la misma probabilidad de ser seleccionados. Cada elemento sólo puede ser seleccionado una vez como máximo. La importancia del diseño radica en que sirve de base para la definición de muchos otros.

b) Muestreo estratificado aleatorio simple

Este procedimiento consiste en subdividir previamente la población en subpoblaciones o estratos. Una vez hecho esto se toma una muestra aleatoria simple de cada estrato. La asignación del tamaño de muestra a seleccionar de cada uno de ellos se hace teniendo en cuenta distintos factores como son, por ejemplo, el tamaño de la subdivisión, la homogeneidad de los elementos dentro de ella, el costo de seleccionar cada unidad por estrato y/o la importancia que pueden tener las estimaciones para los diferentes estratos.

c) Muestreo sistemático aleatorio

En este caso se seleccionan las unidades a intervalos iguales pero tomando un punto de partida (selección del primer elemento) aleatorio. Es de gran utilidad para poblaciones que están originalmente ordenadas ya que ello facilita la recolección de la información disminuyendo los costos. Por otra parte, el procedimiento de selección permite una distribución espacial de la muestra conveniente en muchas investigaciones.

d) Muestreo por conglomerados.

Básicamente consiste de seleccionar, antes que unidades elementales o individuos, grupos de ellas o “conglomerados”. Se utilizan cuando los individuos constituyen

agrupaciones naturales, por ejemplo los alumnos del mismo curso, las familias nucleares, etc. Esto es recomendable cuando existe cercanía de algún tipo entre las unidades finales. Esta cercanía es generalmente de tipo físico o administrativo. (Ospina, 2.001)

4. Aplicación del muestreo estratificado

FORMULACIÓN DEL PROBLEMA:

El consumo de tabaco es la principal causa de defunción previsible en el mundo (1). Estudios recientes informan que su inicio se da en poblaciones cada vez más jóvenes y que de seguir así, ocasionará el fallecimiento de 8.4 millones de personas en el mundo en los siguientes 20 años (2,3).

En Colombia para 1998, el Ministerio de Salud (4) comunicó que la prevalencia de fumadores adultos fue 18.9% y la de ex-fumadores 12.4%; además, 11% de los adultos menores de 24 años fumaban y 4% eran ex-fumadores; entre los adolescentes, 19% habían fumado alguna vez y de ellos, 11% eran fumadores habituales, 19% fumaban ocasionalmente y 70% lo hicieron por probar.

Las directivas de un colegio (hipotético) de bachillerato de la ciudad de Barranquilla al que llamaremos Liceo Santander, están interesadas en estimar el porcentaje de estudiantes que tienen hábito de fumar (variable de interés). Se conoce la siguiente información sobre el colegio:

El número total de estudiantes matriculados en el año de realización de la investigación es de 455. Estos estudiantes están repartidos en grados, que serán considerados estratos (método de escogencia de las unidades el muestreo simple estratificado con afijación igual):

- Estrato I: Estudiantes matriculados en el grado 6,
- Estrato II: Estudiantes matriculados en el grado 7,
- Estrato III: Estudiantes matriculados en el grado 8,
- Estrato IV: Estudiantes matriculados en el grado 9,
- Estrato V: Estudiantes matriculados en el grado 10,
- Estrato VI: Estudiantes matriculados en el grado 11.

El número de unidades (estudiantes) en cada estrato es el siguiente:

Estrato I: $N_1 = 95$; Estrato II: $N_2 = 82$; Estrato III: $N_3 = 71$; Estrato IV: $N_4 = 65$,
Estrato V: $N_5 = 75$; Estrato VI: $N_6 = 67$.

SOLUCIÓN DEL PROBLEMA.

Tamaño de muestra La proporción de unidades en cada uno de los estratos poblacionales es:

$$W_1 = \frac{N_1}{N} = \frac{95}{455} = 0,2088 \cong 0,21 \text{ ó } 21 \%$$

$$W_2 = \frac{N_2}{N} = \frac{82}{455} = 0,1802 \cong 0,18 \text{ ó } 18 \%$$

$$W_3 = \frac{N_3}{N} = \frac{71}{455} = 0,1560 \cong 0,16 \text{ ó } 16 \%$$

$$W_4 = \frac{N_4}{N} = \frac{65}{455} = 0,1429 \cong 0,14 \text{ ó } 14 \%$$

$$W_5 = \frac{N_5}{N} = \frac{75}{455} = 0,1648 \cong 0,16 \text{ ó } 16 \%$$

$$W_6 = \frac{N_6}{N} = \frac{67}{455} = 0,1473 \cong 0,15 \text{ ó } 15 \%$$

Para determinar el tamaño de la muestra para la estimación del porcentaje de estudiantes con hábito de fumar aplicamos la siguiente fórmula (Martínez, 1984):

$$n_0 = \sum \frac{W_h P_h Q_h}{V}$$

$$\text{siendo: } V = \frac{d^2}{z^2} = \left(\frac{d}{z}\right)^2 \text{ y } n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Donde

P_h =proporción de unidades que presentan la característica en el estrato h ,

Q_h =proporción de unidades que presentan la característica en el estrato h .

d =Error máximo admisible; este valor lo postula el investigador.

Z =Valor del cantil normal conseguido por tabla a partir de un nivel de confianza.

Para la aplicación de estas fórmulas en primer lugar, si se considera que no existe forma alguna de obtener las varianzas poblacionales para cada estrato, hay necesidad de realizar

una encuesta preliminar.

Supongamos, arbitrariamente, que el tamaño de la muestra preliminar es del 4% del tamaño de la población.

$$n_{preliminar} = 0,04(455) = 18,2$$

Aplicando el método de asignación igual, donde $n_1 = n_2 = n_3 = n_4 = n_5 = n_6$ se tendrá:

$$n_{\frac{preliminar}{h}} = \frac{18,2}{6} = 3,03 \cong 3$$

$n_1 = 3; n_2 = 3; n_3 = 3; n_4 = 3; n_5 = 3; n_6 = 3; n_{preliminar} = 18$. Con los anteriores tamaños para la encuesta preliminar, seleccionamos, en forma aleatoria, 3 estudiantes en cada uno de los estratos, utilizando la tabla de números aleatorios o computador.

Se supone, que una vez seleccionados los estudiantes en cada grado, se obtuvieron los siguientes resultados:

Estrato	Estudiantes	Fumar	Coeficientes		
			p_h	$p_h q_h$	$W_h p_h q_h$
I	1	Si	0.66	0.2178	0.045
	2	Si			
	3	No			
II	1	No	0.33	0.2178	0.039
	2	No			
	3	Si			
III	1	Si	0.33	0.2178	0.035
	2	No			
	3	No			
IV	1	No	0.66	0.2178	0.030
	2	Si			
	3	Si			
V	1	No	0.33	0.2178	0.035
	2	No			
	3	Si			

Estrato	Estudiantes	Fumar	Coeficientes		
			p_h	$p_h q_h$	$W_h p_h q_h$
VI	1	Si	0.66	0.2178	0.033
	2	No			
	3	Si			

La proporción estratificada se calcula mediante la formula:

$$\bar{p}_{st} = \sum W_h p_h = 0,21(0,66) + 0,18(0,13) + 0,16(0,33) + 0,14(0,66) + 0,16(0,33) + 0,15(0,66) = 0,9456$$

Ahora, sea que el error máximo admisible es del 5 % de \bar{p}_{st}

$$d = 0,1(0,9456) = 0,09456$$

$$V = \left(\frac{d}{z}\right)^2 = \left(\frac{0,09456}{1,96}\right)^2 = 0,00232757$$

Para obtener el tamaño de muestra se realizan los siguientes cálculos auxiliares:

$$\begin{aligned} n_0 &= \frac{1}{V} \sum W_h p_h q_h \\ &= \left(\frac{1}{0,00232757}\right) (0,045 + 0,039 + 0,035 + 0,030 + 0,035 + 0,033) \\ &= \frac{0,217}{0,00232757} = 93,23 \end{aligned}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}} = \frac{93,23}{1 + \frac{93,23}{455}} \approx 78$$

$$n_1 \approx n_2 \approx n_3 \approx n_4 \approx n_5 \approx n_6 \approx 13; n \approx 18.$$

La muestra.

N	I Estrato	II Estrato	III Estrato	IV Estrato	V Estrato	VI Estrato
1	Si	No	No	Si	No	No
2	No	No	Si	Si	No	Si
3	No	No	Si	No	Si	Si
4	No	No	No	No	Si	No
5	Si	Si	No	No	No	No
6	No	No	Si	No	Si	No
7	No	No	Si	Si	Si	Si
8	No	Si	No	No	Si	Si
9	No	No	No	No	No	Si
10	No	No	No	Si	No	Si
11	Si	No	Si	Si	Si	Si
12	No	Si	No	Si	No	No
13	No	Si	No	No	Si	Si
N_h	95	82	71	65	75	67
n_h	13	13	13	13	13	13
$N_h(N_h - n_h)$	7790	5658	4118	3380	4650	3618
p_h	0.2308	0.3077	0.3846	0.4615	0.5385	0.6154
q_h	0.7692	0.6923	0.6154	0.5385	0.4615	0.3846
$p_h q_h$	0.17753136	0.21302071	0.23668284	0.24851775	0.24851775	0.23668284
$(p_h q_h)/n_h$	0.01366	0.01639	0.01821	0.01912	0.01912	0.01821
$N_h(N_h - n_h)$ $(p_h q_h)/n_h$	106.3822	92.7131	74.9438	64.6146	88.8928	65.8706
$W_h p_h$	0.048468	0.05539	0.06154	0.06461	0.08616	0.09231

Las estimaciones.

Aplicando las siguientes formulas para las estimaciones de la proporción estratificada y error estándar de estimación:

$$\hat{V}(\bar{p}_{st}) = \frac{1}{N^2} \sum N_h(N_h - n_h) \frac{p_h q_h}{n_h}$$

$$P_{st} = \bar{p}_{st} \pm \sqrt{\frac{1}{N^2} \sum N_h(N_h - n_h) \frac{p_h q_h}{n_h}},$$

Se obtiene:

$$0,31279 < P_{st} < 0,550415,$$

$$VAR(P_{st}) = 0,002383$$

5. Conclusiones

- El muestreo es una útil herramienta en solución de problemas de investigación en el aula.
- La investigación que usa muestreo estadístico permite la participación colectiva para su realización.
- La investigación que usa muestreo estadístico es taller para alcanzar competencias cognitivas.
- La investigación que usa muestreo estadístico es una herramienta para la caracterización de variables en las poblaciones estudiantiles, y así contribuir en la búsqueda de soluciones a problemas que estas tengan.

Referencias

- [1] Furman, M. & Zysman, A. (2001), *Ciencias naturales: aprender a investigar en la escuela*, Novedades Educativas, Buenos Aires.
- [2] Martínez Bencardino, C. (1984), *Muestreo. Algunos Métodos y sus Aplicaciones Prácticas*, Ecoe Ediciones, Bogotá.
- [3] Ospina Botero, D. (2001), *Introducción al Muestreo*, Unibiblos, Universidad Nacional de Colombia, Bogotá.
- [4] Porlán, R. (1999), *Hacia un modelo de enseñanza-aprendizaje de las ciencias por investigación*, en Kaufman, M. y Fumagalli, L. *Enseñar Ciencias Naturales*, Paidós, Buenos Aires.