Dificultades en la comprensión del concepto de muestra aleatoria simple en estudiantes universitarios

Difficulties in understanding the simple random sample concept in university students

Amable Moreno

Universidad Nacional de Cuyo

Resumen

En este trabajo analizamos las dificultades en la comprensión del concepto de muestra aleatoria simple en un grupo de estudiantes universitarios, desde el enfoque ontosemiótico del conocimiento y la instrucción matemática. Para responder a la pregunta *qué comprenden los estudiantes* sobre este concepto, aplicamos un cuestionario con ocho ítems. Las dificultades más importantes encontradas fueron: la imposibilidad de identificar a todas las muestras que cumplen cierta condición, la falta de reconocimiento de la muestra aleatoria como variable aleatoria conjunta multivariada, la falta de reconocimiento de la independencia de las variables aleatorias de la muestra y la complejidad que implica pasar de un nivel de generalidad a otro en los conceptos de variable aleatoria y distribución de probabilidad.

Palabras clave: muestra aleatoria simple, teoría ontosemiótica, estudiantes universitarios

Abstract

In this paper we analyse the difficulties in understanding of the concept of simple random sample in a group of university students, from the onto-semiotic approach to mathematical knowledge and instruction. To analyse what the students understand about this concept, we applied a questionnaire with eight items. The most important difficulties encountered were: the inability to identify all samples that meet a certain condition, the lack of recognition of random sample as a multivariate random variable, the lack of recognition of the independence of the random variables in the sample and the involved complexity in the concepts of random variable and probability distribution, going from one level of generality to another.

Keywords: Simple random sample, onto-semiotic theory, university students

1. Introducción

En este trabajo se analizan las dificultades en la comprensión del concepto de muestra aleatoria simple en un grupo de estudiantes universitarios con formación previa en estadística, para determinar hasta qué punto la enseñanza ha sido eficaz. El análisis se realiza desde el enfoque ontosemiótico (EOS) sobre el conocimiento y la instrucción matemática (Godino, 2002; Godino y Batanero, 1994; Godino, Batanero y Font, 2007).

Para comprender la esencia de la inferencia estadística es necesario comprender la naturaleza de una población y de una muestra. Esto nos lleva a considerar a la muestra aleatoria como un concepto fundamental para el aprendizaje de la inferencia estadística. Diversos autores, como Batanero y Díaz (2015) señalan que las técnicas de inferencia estadística fundamentan las formas de obtener un conocimiento general a partir del análisis de un conjunto de datos. Dicha importancia ha sido expresada por Heitele (1975), para quien el muestreo es una de las diez ideas estocásticas fundamentales.

Moreno, A. (2017). Dificultades en la comprensión del concepto demuestra aleatoria simple en estudiantes universitarios. En J. M. Contreras, P. Arteaga, G. R. Cañadas, M.M. Gea, B. Giacomone y M. M. López-Martín (Eds.), *Actas del Segundo Congreso International Virtual sobre el Enfoque Ontosemiótico del Conocimiento y la Instrucción Matemáticos.* Disponible en, enfoqueontosemiotico.ugr.es/civeos.html

Además, en el caso de la enseñanza de la estadística en la universidad, uno de los problemas didácticos principales es la enseñanza de la inferencia (Artigue, Batanero y Kent, 2007; Moore, 1997).

Desde el punto de vista estadístico, la inferencia incluye un conjunto de métodos que permiten generalizar los resultados obtenidos en una muestra a una población. Por lo que resulta de importancia destacar el papel que juega en la inferencia estadística el concepto de muestra aleatoria simple.

Batanero, Godino, Green, Holmes y Vallecillos (1994) citan a Moses (1992), quien indica que la muestra proporciona información sobre la población y de este modo aumenta el conocimiento sobre la misma. En este sentido los autores citados sugieren que la comprensión adecuada de la inferencia estadística requiere de dos ideas complementarias: *la representatividad y variabilidad muestral*. La comprensión de la representatividad implica entender que la muestra es parecida a la población, si ha sido elegida convenientemente. La variabilidad indica el hecho de las muestras son diferentes unas de otras.

Por lo expresado podemos concluir que un estudio sobre las dificultades, que obstaculizan la comprensión del concepto de muestra aleatoria simple, puede proporcionar información que favorezca la enseñanza de la inferencia estadística.

2. Marco Teórico

El marco teórico que se ha considerado en este trabajo se corresponde con el enfoque ontosemiótico (EOS) sobre el conocimiento y la instrucción matemática (Godino 2002; Godino y Batanero, 1994; Godino, Batanero y Font, 2007).

Consideramos la idea de situación-problema de acuerdo con este enfoque; es decir, como cualquier actividad, pregunta o cuestión que requiera de una actividad de matematización. Luego, los objetos matemáticos surgen de las prácticas realizadas en la resolución de problemas (Godino y Batanero, 1994). Desde este enfoque el significado de un objeto matemático se concibe como el sistema de prácticas que realiza una persona (significado personal) o se realizan en una institución (significado institucional) para resolver un tipo de situación-problema de las que surge el objeto.

Por otra parte, en relación con la enseñanza de la inferencia, como expresan Harradine, Batanero y Rossman (2011) y Kadijevich, Kokol-Voljc y Lavicza (2008), se debe trabajar con tres distribuciones que están presentes durante todo el proceso; la distribución poblacional, la distribución de la muestra aleatoria y la distribución del estadístico en el muestreo. Además, de acuerdo con Batanero, Díaz, Contreras y Roa (2013) y Harradine et al. (2011), el razonamiento inferencial requiere de la coordinación de distintos conceptos: los datos, la población de donde se tomaron los datos y las posibles muestras de la misma.

En nuestro caso particular, el interés recae en el objeto matemático muestra aleatoria simple. Nos focalizamos en la comprensión sobre el concepto de muestra aleatoria y su distribución de probabilidad, en una muestra de estudiantes universitarios. Entendiendo la comprensión en el sentido de Godino (1996), es decir, se pretende estudiar la correspondencia entre los significados personales y los significados institucionales. Para este autor la comprensión tiene dos ejes, uno descriptivo, que indica los aspectos o componentes de los objetos a comprender, y el otro procesual que indica las fases o niveles necesarios en el logro de la correcta comprensión. En este trabajo nos centramos en el eje descriptivo.

3. Análisis conceptual y antecedentes

Siguiendo a Harradine et al. (2011), consideramos que la inferencia está constituida por tres elementos fundamentales: proceso de razonamiento, los conceptos y los cálculos; entre los conceptos que mencionan se encuentra la aleatoriedad y el muestreo.

Además, consideramos importante destacar lo indicado por Batanero et al. (1994), en relación con la necesidad de establecer un equilibrio adecuado entre la *representatividad muestral* y la *variabilidad muestral*. Para asegurar la primera los elementos deben ser extraídos de la manera más objetiva posible, a fin de no influir en la selección de cada uno de ellos, lográndose la representatividad cuando la extracción se realiza al azar. Si esto ocurre, decimos que la muestra es probabilística o aleatoria. Este procedimiento permitirá conocer en términos de probabilidad el error que se comete al utilizar la muestra como reflejo de la población.

En este trabajo consideramos la noción matemática de población como el conjunto de todos los posibles resultados de un fenómeno o experimento aleatorio. Se sabe que si de una población se extrae una segunda muestra, el resultado que se obtiene puede ser o no distinto del obtenido en la primera muestra. Esto nos indica que las inferencias de la muestra a la población tienen un carácter inductivo y son inciertas. Ante esta situación, la Teoría de la inferencia estadística debe dar rigor matemático a tal incertidumbre, mediante la aplicación de la Teoría de las probabilidades.

En este trabajo, entendemos una muestra como la observación de *n* elementos de una población en un cierto orden, o el resultado de repetir n veces un experimento aleatorio. Así, el conjunto de todas las muestras posibles de tamaño *n* de una población constituye el espacio muestral o espacio de muestras. La idea de espacio muestral se debe a Von Mises (1928/1952); esta idea hizo posible construir una teoría estrictamente matemática basada en la teoría de la medida, que le permitió a Kolmogorov (1933/1950) desarrollar un tratamiento axiomático, que representa el desarrollo moderno de la Teoría de la probabilidad.

Designaremos a una *muestra aleatoria* de tamaño *n* por un vector aleatorio $(X_1, X_2, ..., X_n)$ y sea una muestra observada:$(x_1, x_2, ... , x_n)$, donde x_i es el elemento que corresponde a la i-ésima observación, y como cada muestra viene definida por un conjunto de n valores $(x_1 , x_2 , ... , x_n)$. Entonces podemos pensar en el espacio muestral, como incluido en un espacio n-dimensional, en el que las coordenadas de cada punto son los correspondientes elementos muestrales, $(x_1, x_2 , ... , x_n)$ de cada una de ellas. Por lo tanto, es posible contemplar la muestra aleatoria como una variable aleatoria n-dimensional, cuya distribución de probabilidad depende de la *distribución de probabilidad de la variable aleatoria poblacional, F(x), del procedimiento de selección y del tamaño n de la muestra*.

En el muestreo probabilístico se distinguen dos tipos dependiendo de cuál sea el procedimiento aleatorio de extracción utilizado, pues influye en la probabilidad de obtención de los elementos de la muestra. Se pueden extraer los elementos muestrales de dos maneras: con reemplazamiento y sin reemplazamiento. El muestreo con reemplazamiento conduce a que los elementos de la muestra sean probabilísticamente independientes en poblaciones finitas o infinitas; mientras que el muestreo sin reemplazamiento no lo son en poblaciones finitas. En este trabajo las muestras serán aleatorias con reemplazamiento, a las que llamamos *muestras aleatorias simples.*

Por lo tanto, consideraremos la muestra aleatoria simple de tamaño "*n*" de una población con densidad *f*, como el vector de "*n*" variables aleatorias independientes y cada una con idéntica distribución que la correspondiente a la población de la cual se extrajo. Consideremos *X* es una variable aleatoria discreta con una cierta función densidad de probabilidad *f(x)* y una muestra aleatoria simple concreta (x_1, x_2, ..., x_n). La probabilidad de obtener esta muestra se obtiene como $f(x_1, x_2, ..., x_n) = f(x_1)$ $f(x_2)$..................$f(x_n)$.

4. Metodología

En este trabajo se realizó un estudio exploratorio sobre la comprensión del concepto de muestra aleatoria simple, mediante la aplicación de un cuestionario con ocho ítems a una muestra de 42 estudiantes de segundo año de una de las carreras de la facultad de ciencias económicas, de la provincia de Mendoza, Argentina. En el cuestionario (que se presenta como anexo) se plantea una situación-problema con información estadística relativa a la población de Mendoza, con la finalidad de evaluar la comprensión de muestra aleatoria simple y de su distribución de probabilidad conjunta. Las respuestas de los estudiantes fueron analizadas buscando la correspondencia entre los significados personales y los significados institucionales, lo que permitió identificar conflictos semióticos entre dichos significados, y el análisis de los mismos nos permitió describir las dificultades de los estudiantes.

5. Resultados

Las respuestas que dieron los estudiantes a cada uno de los ítems se presentan en la Tabla 1. En el ítem 1, el estudiante debe expresar la función densidad de la variable aleatoria discreta *X* que plantea el enunciado del problema. De las respuestas que dieron los estudiantes, 35 (83,3%) estudiantes lo hicieron de manera correcta. Mientras que 6 (14,3%) dieron respuestas incompletas debido a que indicaron la probabilidad de cada uno de los valores de la variable discreta *X*, pero omitieron el conjunto soporte de la variable discreta *X*; y solamente 1 (2,4%) estudiante dio una respuesta incorrecta por considerar el valor de la variable en lugar de la probabilidad de dicho valor.

Tabla 1. Frecuencia (y porcentaje) de respuestas a cada pregunta, según corrección

Ítem	Correcta	Parcialmente Correcta	Incorrecta	No Contesta
Ítem 1	35 (83,3)	6 (14,3)	1 (2,4)	0 (0)
Ítem 2	22 (52,4)	0 (0)	20 (47,6)	0 (0)
Ítem 3	8 (19)	29 (69,1)	5 (11,9)	0 (0)
Ítem 4	37 (88,1)	0 (0)	4 (9,1)	1 (2,4)
Ítem 5	12 (28,6)	0 (0)	18 (42,8)	12 (28,6)
Ítem 6	16 (38,1)	0 (0)	15 (35,7)	11 (26,2)
Ítem 7	3 (7,1)	0 (0)	25 (59,5)	14 (33,3)
Ítem 8	1 (2,4)	0 (0)	23 (54,8)	18 (42,9)

En el ítem 2, se solicita que el estudiante indique la definición de probabilidad aplicada, las respuestas indicaron que 24 estudiantes (57,1%) reconocieron de manera correcta la definición de probabilidad que se había considerado; que en este caso es la definición frecuencial; y el resto de los estudiantes que fueron 18 (42,9%) contestaron de manera incorrecta aplicando en algunos casos la definición clásica y en otros la axiomática.

En el ítem 3, el estudiante debe indicar el número de muestras aleatorias simples de tamaño 2 que se pueden obtener de la población presentada, en este caso son 9, y además enumerar cada una de ellas. Encontramos que solamente ocho estudiantes (19%) contestaron de manera correcta; cinco estudiantes (11,9%) lo hicieron de manera incorrecta; y 29 (69,1%) de manera parcialmente correcta; de éstos uno (2,4%) indica correctamente el número de muestras pero no consigue enumerarlas, mientras que 28 (66,7%), si bien enumeran correctamente todas las muestras no indican el número total de muestras distintas que se obtienen; y un estudiante que había enumerado correctamente todas las muestras, indica que sólo se pueden obtener tres muestras aleatorias distintas: (1,2); (1,3); (2,3) considerando que el término "distintas" hace referencia a los elementos que integran la muestra aleatoria simple.

En el ítem 4, se solicita que asocie cada muestra con la probabilidad de ocurrencia de la misma. Los resultados obtenidos indican que 37 de los estudiantes (88,1%) dan respuestas correctas, cuatro (9,5%) contestan de manera incorrecta y uno (2,4%) no contesta. De las respuestas incorrectas, algunos estudiantes consideran que el espacio muestral correspondiente a la variable aleatoria conjunta bivariada (X_1, X_2) es equiprobable, por lo que le asignan a cada una de las muestras probabilidad 1/9 dado que se pueden obtener 9 muestras distintas; dando lugar al sesgo de equiprobabilidad (Lecoutre, 1992), debido a que no consideran la información contenida en el enunciado.

En el ítem 5, el estudiante debe considerar el evento que cumple la condición establecida en el enunciado, es decir el conjunto {(1,3); (3,1)} y luego calcular su probabilidad. En este ítem, seis estudiantes (14,3%) no contestan, 12 (28,6%) dan respuestas correctas, y 24 (57,1%) lo hacen de manera incorrecta al no considerar al conjunto de todas las muestras que cumplen con la condición establecida. Entre las respuestas incorrectas encontramos que 15 (35,7%) estudiantes aplican un razonamiento aditivo en el sentido que consideran a la muestra con un subconjunto de la población y suman las probabilidades de cada una de las componentes de la muestra; 3 (7,1%) estudiantes consideran solamente una muestra y además dicha muestra no cumple con la condición establecida, y calculan la probabilidad de dicha muestra de manera correcta; 6 (11,9%) estudiantes consideran solamente una muestra y calculan de manera correcta la probabilidad de la misma, es decir que, reconocen que la muestra aleatoria de tamaño 2 es una variable aleatoria conjunta bivariada cuyas componentes son variables aleatorias independientes.

En el ítem 6, el estudiante debe considerar al evento {(1,1); (1,2); (2,1); (2,2)} y luego calcular su probabilidad. En este ítem 15 estudiantes (35,7%) responden de manera correcta, 11 (26,2%) no contestan, y el resto 16 estudiantes (38,1%) dan respuestas incorrectas. Las causas que los llevan a cometer errores son similares a las encontradas en el ítem anterior. Así, 4 estudiantes (9,5%) no interpretan de forma correcta la información que se les proporcionó; 11 (26,2%) consideran una única muestra que cumple con la condición establecida y aplican un razonamiento aditivo en el cálculo de la probabilidad de la misma y finalmente uno (2,4%) interpreta incorrectamente la información, considera una única muestra y aplica un razonamiento aditivo en el cálculo de la muestra.

En el ítem 7 se debe considerar el evento {(2,2): (2,3); (3,2); (3,3)} y calcular su probabilidad; solamente tres estudiantes (7,1%) responden de manera correcta; 13

(30,9%) no responden; y 11 (26,2%) solamente consideran dos muestras y calculan de manera correcta la probabilidad de cada una de ellas; dos estudiantes (4,8%) consideran solamente dos muestras y calculan la probabilidad de cada una de ellas sin dar una respuesta única; nueve (21,4%) consideran una única muestra y calculan la probabilidad de la misma aplicando un razonamiento aditivo, es decir suman las probabilidades de cada una de las componentes; y finalmente cuatro estudiantes (9,5%) no interpreta la información que plantea el problema.

En el ítem 8, solamente un estudiante (2,4%) plantea la probabilidad condicional que se le solicita, aplica correctamente la definición de probabilidad condicional y encuentra el resultado correcto, pero no considera que las variables aleatorias que integran la muestra son independientes; 18 (42,9%) no responden y el resto, 23 (54,8%) dan respuestas incorrectas; de los cuales 19 (45,2%) no logran determinar la intersección de dos conjuntos de muestras. Además, entre las respuestas incorrectas se encuentra que cuatro (9,5%) plantean la probabilidad condicional correctamente, calculan la probabilidad del evento condicionante de manera correcta pero no logran determinarla intersección dos conjuntos de muestras y en consecuencia el cálculo es incorrecto; 13 estudiantes (31%) no logran determinar la intersección de dos conjuntos de muestras y además no logran determinar correctamente la probabilidad del evento condicionante y no reconocen la independencia de las variables; dos estudiantes (4,8%) consideran a la probabilidad condicional como la probabilidad de la intersección de dos eventos; cuatro (9,5%) plantean la probabilidad condicional, reconocen la independencia de las variables aleatorias pero cuando realizan los cálculos aplican un razonamiento aditivo de la muestra.

5. Conclusiones y discusión

Los conflictos semióticos explícitos en las actividades realizadas por los estudiantes revelan las principales dificultades que afrontan los estudiantes cuando deben determinar la probabilidad de una muestra. La mayoría de los estudiantes aplican un razonamiento aditivo al sumar las probabilidades de cada uno de los componentes de la misma; este tipo de razonamiento ya ha sido encontrado por Shaughnessy, Ciancetta y Canada (2004), Saldanha y Thompson (2002) y Saldanha (2004). En efecto, los estudiantes tienen una imagen aditiva de la muestra porque tienden a concebir la muestra como un subconjunto de una población. Otra dificultad surge cuando deben determinar la intersección de dos conjuntos de muestras, debido a que el estudiante no percibe la existencia de dos espacios muestrales distintos, el correspondiente a la variable aleatoria en estudio y el correspondiente a la muestra aleatoria. En el primer caso, a cada unidad de análisis, seleccionada aleatoriamente, se le asigna un número real mediante la función variable aleatoria. Mientras que en el muestreo, se seleccionan aleatoriamente n unidades de análisis, a los que un vector de n variables aleatorias les asigna un vector de n números reales. Luego, en el primer caso tenemos un espacio muestral unidimensional, y en el segundo caso otro espacio muestral multidimensional. Esto da lugar a que los estudiantes confundan con frecuencia estos distintos niveles de concreción del concepto de variable aleatoria y su distribución de probabilidad; idea que ya ha sido expresada por Schuyten (1991).

La interpretación de la muestra aleatoria como subconjunto de la población, conlleva a la falta de consideración de la independencia de las variables aleatorias de la muestra, siendo ésta propiedad un elemento constitutivo del concepto de muestra aleatoria simple. Otra dificultad se presenta cuando el enunciado del problema no especifica que se deben considerar todas las muestras posibles que cumplen con la condición pedida, en estos casos se observa que los estudiantes no consideran la totalidad de las muestras, como se puede apreciar al comparar los resultados del ítem 4 con los resultados de los ítems 5, 6, 7 y 8.

En conclusión, la mayoría de los estudiantes no han logrado realizar las distintas prácticas presentadas en el cuestionario, y en consecuencia no lograron la comprensión del concepto de muestra aleatoria simple. Por lo tanto, se debería evaluar la idoneidad didáctica del proceso de enseñanza y aprendizaje de la inferencia; para que los estudiantes logren la apropiación de las ideas fundamentales de la estadística y también desarrollar un razonamiento estadístico con el fin de dar sentido a la información estadística, como lo sugieren Garfield y Gal (1999).

Por lo tanto, para completar el estudio sobre la comprensión del concepto "muestra aleatoria simple" se requiere una mayor investigación para dar respuesta al *cómo se logra su comprensión.*

Referencias

Artigue, M., Batanero, C. y Kent, P. (2007). Mathematics thinking and learning at postsecondary level. En F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1011-1049). Greenwich, CT: Information Age Publishing, Inc., & NCTM.

Batanero, C. y Diaz, C. (2015). Aproximación informal al contraste de hipótesis. En J. M. Contreras, C. Batanero, J. D. Godino, G. R. Cañadas, P. Arteaga, E. Molina, M. M. Gea y M. M. López (Eds.), *Actas de las segundas jornadas virtuales de didáctica de la estadística, probabilidad y combinatoria: Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 135-144). Granada: Grupo de Investigación en Didáctica de la Estadística, Probabilidad y Combinatoria de la Sociedad Española de Investigación en Educación Matemática (SEIEM).

Batanero, C., Díaz, C., Contreras, J. M. y Roa, R. (2013). El sentido estadístico y su desarrollo. *Números*, *83,* 7-18.

Batanero, C., Godino, J. D., Green, D. R., Holmes, P. y Vallecillos, A. (1994). Errores y dificultades en la comprensión de los conceptos estadísticos elementales. *International Journal of Mathematics Education in Science in and Technology, 25*(4), 527-547.

Garfield, J. y Gal, I. (1999). Teaching and assessing statistical reasoning. En L. Stiff (Ed.), *Developing mathematical reasoning in grades K-12* (pp. 207-219). Reston VA: National Council Teachers of Mathematics.

Godino, J. D. (1996). Significado y comprensión de los conceptos matemáticos. En L. Puig y A. Gutierrez (Eds), *Proceedings of the 20th Conference of the International Group for Psychology of Mathematics Education* (pp. 417-424). Valencia: Universidad de Valencia.

Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathematiques*, *22* (2-3), 237-248.

Godino, J. D. y Batanero, C. (1994). Significado institucional y personal de los objetos matemáticos. *Recherches en Didactique des Mathématiques, 14* (3), 325-355.

Godino, J. D., Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education. *Zentralblattfür Didaktik der Mathematik, 39*(1-2), 127-135.

Harradine, A., Batanero, C. y Rossman, A. (2011). Students and teachers´ knowledge of sampling and inference. En C. Batanero, G. Burrill y C. Reading (Eds.). *Teaching statistics in school mathematics. Challenges for teaching and teachers education* (pp. 235-246).New York: Springer.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, *6*, 187-205.

Kadijevich, D., Kokol-Voljc, V., y Lavicza, Z. (2008). Towards a suitable designed instruction on statistical reasoning: Understanding sampling distribution with technology. En C. Batanero, G. Burrill, C. Reading y A. Rossman (Eds.), *Proceedings of the ICMI Study 18 Conference and IASE 2008 Round Table Conference.* Monterrey: International Statistical Institute. Disponible en, http://www.stat.auckland.ac.nz/~iase/publications/rt08/T4P9_Kadijevich.pdf

Kolmogorov, A. (1950). *Foundations of probability´s calculation.* New York: Chelsea Publishing Company (Original work published in 1933).

Mises, R. (1952). *Probability, statistics and truth*. London: William Hodge. (Original work published 1928).

Lecoutre, M. P. (1992). Cognitive models and problem spaces in “purely random” situations. *Educational Studies in Mathematics, 23*(6), 557-568.

Moore, D. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review*, *65* (2), 123-155.

Moses, L. E. (1992). The reasoning of statistical inference. En D. C. Hoaglin y D. S. Moore (Eds.), *Perspectivas on contemporary statistics* (pp. 107-122). Washington: Mathematical Association of America.

Saldanha, L. A. (2004). *“Is this sample unusual?”: An investigation of students exploring connections between sampling distributions and statistical inference*. Tesis Doctoral. Vanderbilt University.

Saldanha, L. A. y Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51*, 257-270.

Shaughenessy, J.M., Ciancetta, M. y Canada, D. (2004). Types of Student reasoning on sampling tasks. En M.J. Hoines y A.B. Fuglestad (Eds.), *Proceedings of the 28th.Conferenceof the International Group for the Psychology of Mathematics Education* (Vol.4, pp. 177-184). Bergen, Noruega: Mathematics Education.

Schuyten, G. (1991). Statistical thinking in psychology and education. En D. Vere Jones (Ed.), *Proceedings of the III International Conference on Teaching Statistics* (Vol. 2, pp. 486-489). Dunedin, Australia: Universidad de Otago.

Apéndice. Cuestionario

En nuestra población, un estudio estadístico reveló que el 40% de los hogares tienen un ingreso mensual de a lo sumo $12.000; el 35% tienen un ingreso entre $12.000 y $20.000 y el 25% tienen un ingreso superior a $20.000. A los efectos de realizar un análisis probabilístico, se ha categorizado la variable continua "ingreso mensual por hogar", como se indica en la siguiente tabla. De esta manera queda clasificado el ingreso mensual en tres niveles"

Ingreso Mensual por Hogar	X
Menor o igual a $12.000	1
Mayor a $12.000 y menor o igual a $20.00	2
Mayor a $20.000	3

Item 1: Expresa la función densidad de probabilidad de la variable aleatoria discreta X

Ítem 2: ¿Qué definición de probabilidad aplicó en el ítem anterior?

Ítem 3: Indique la cantidad de muestras aleatorias posibles de tamaño 2 que se pueden obtener de la población en estudio y enumere cada una de ellas.

Ítem 4: Indique la probabilidad de ocurrencia de cada una de las muestras, completando la siguiente tabla.

Muestra Aleatoria	Probabilidad de la Muestra Aleatoria	Muestra Aleatoria	Probabilidad de la Muestra Aleatoria

Ítem 5: Si se elige una muestra aleatoria de tamaño 2; encuentre la probabilidad de que uno de los hogares seleccionados tenga un ingreso de a lo sumo $12.000 y el otro de por lo menos $20.000.

Ítem 6: Si se elige una segunda muestra de tamaño 2, ¿qué probabilidad hay de que ninguno de los hogares tenga un ingreso superior a $20.000?

Ítem 7: En otra selección aleatoria de una muestra de tamaño2, ¿qué probabilidad se tiene de que los hogares tengan un ingreso superior a $20.000 o tengan un ingreso entre $12.000 y $20.000?

Ítem 8: Si de la selección de una aleatoria de tamaño 2, se sabe que el primer hogar seleccionado tuvo un ingreso superior a $12.000, ¿qué probabilidad tienes de que el segundo hogar seleccionado tenga a lo sumo $20.000?