

Comprensión que estudiantes universitarios desarrollan sobre el muestreo y variabilidad muestral desde una perspectiva informal

Understanding that university students develop about sampling and sampling variability from an informal perspective

Santiago Inzunza Cazares

Universidad Autónoma de Sinaloa, México

Resumen

En el presente artículo reportamos resultados de un estudio sobre la comprensión que desarrollaron algunos estudiantes universitarios de ciencias sociales sobre el muestreo y variabilidad muestral, desde una perspectiva informal. Partiendo de poblaciones con datos reales y resultados de una encuesta, los estudiantes simulaban muestras de tamaño creciente repetidamente hasta llegar a formar distribuciones con los estadísticos de interés. Los resultados muestran que es posible que desde una perspectiva informal basada en simulación los estudiantes puedan desarrollar un esquema del muestreo y variabilidad con nociones adecuadas, que se pueden consolidar y profundizar en un estudio más formal de estos conceptos. Se observó que muchos estudiantes utilizaron un lenguaje coloquial en las argumentaciones en vez de utilizar un lenguaje estadístico que precisara los conceptos.

Palabras clave: Muestreo, variabilidad muestral, distribuciones muestrales, razonamiento inferencial

Abstract

In this article, we report the results of a study on the understanding developed by university students of social sciences about sampling and sampling variability from an informal perspective. Starting from populations with real data and results of a survey, the students repeatedly simulated samples of increasing size until arriving to the distributions of relevant statistics. The results show that it is possible from an informal perspective based on simulation that students develop a sampling scheme and variability with appropriate notions that can be consolidated in a more formal study of these concepts. It was observed that many students used a colloquial language in the argumentations instead of using a statistical language that would make the concepts precise.

Keywords: Sampling, sampling variability, sampling distributions, inferential reasoning

1. Introduction

El muestreo es un concepto fundamental para comprender la inferencia estadística, pues proporciona los datos a partir de los cuales se realizan las estimaciones o contraste de hipótesis sobre parámetros poblacionales. Heitele (1975) identifica al muestreo como una de las ideas estocásticas fundamentales del currículo de estadística que debe ser objeto de enseñanza desde los niveles escolares básicos hasta el nivel universitario, incrementando gradualmente el nivel de formalización.

El muestreo en apariencia es una idea sencilla, pues una vez seleccionada la muestra se procede al cálculo de resúmenes numéricos de los datos, los cuales nos informan con cierta precisión sobre características de la población de la que fue extraída la muestra. Sin embargo, una propiedad intrínseca del muestreo es la variabilidad, de tal forma que es posible que los resultados que aporta una muestra puedan estar alejados de las características de la población, es decir, una muestra puede no ser representativa de la población aunque sea seleccionada al azar. Estas dos propiedades que subyacen al

muestreo -representatividad y variabilidad muestral-, han sido identificadas por Rubin, Bruce y Tenney (1991) como dos ideas complejas que están en la base de la comprensión de los métodos de inferencia estadística, y constituyen un barrera conceptual para muchos estudiantes.

En la enseñanza de la estadística –al menos en el caso mexicano- ha predominado por mucho tiempo un enfoque descriptivo centrado en la construcción de gráficas y cálculo de medidas descriptivas de un conjunto de datos, con poca referencia a la forma como estos fueron seleccionados. Sin embargo,

La falta de apreciación del muestreo como una parte importante de una investigación estadística hace que el resto del proceso se vuelva totalmente inútil. Los estudiantes necesitan apreciar el muestreo, ya sea para llevar a cabo sus propias investigaciones o para evaluar si otros están utilizando los métodos adecuados (Watson, 2006, p. 28).

El razonamiento sobre muestreo ha adquirido especial importancia en los años recientes, como consecuencia del incremento en el uso de datos provenientes de muestras. Un ejemplo muy recurrente de uso del muestreo son los resultados de encuestas de opinión que publican frecuentemente los medios de comunicación, en las cuales se reportan estimaciones sobre parámetros de una población.

Harradine, Batanero y Rossman (2011) y Watson y Moritz (2000), sostienen que el muestreo y sus propiedades pueden ser comprendidos desde los niveles básicos desde una perspectiva informal, así cuando los estudiantes arriban a los cursos universitarios de inferencia estadística pueden tener mayor éxito al haber desarrollado ideas correctas sobre el muestreo y la red de conceptos que los rodean desde edades tempranas; pues como señala Pfannkuch (2008, p.1) “cuando los estudiantes no son conscientes del muestreo, su razonamiento inferencial informal es limitado”.

En esta perspectiva algunos documentos curriculares como Common Core State Standards for Mathematics (CCSSM, 2010) establecen para el nivel medio y bachillerato: a) usar muestreo aleatorio para hacer inferencias sobre una población, b) comprender que generalizaciones sobre la población desde una muestra son válidas solo si la muestra es representativa de la población, d) comprender que el muestreo aleatorio tiene a producir muestras representativas para generar inferencias válidas, e) generar múltiples muestras (o muestras simuladas) del mismo tamaño de muestra, para calibrar la variabilidad en estimaciones o predicciones.

En este contexto, nos hemos planteado como objetivo diseñar una trayectoria de aprendizaje con el propósito de desarrollar el razonamiento sobre el muestreo y variabilidad muestral de estudiantes universitarios de ciencias sociales y humanidades, previo al estudio de los métodos de inferencia estadística. La trayectoria se caracteriza por la adopción de un enfoque informal mediado por el uso de tecnologías digitales para la simulación del muestreo en contextos de datos reales y encuestas. Pretendemos caracterizar el razonamiento que los estudiantes desarrollaron sobre el muestreo después de completar la trayectoria de aprendizaje con el propósito de identificar dificultades y razonamientos adecuados para apoyar el diseño instruccional de lecciones de inferencia estadística.

2. Antecedentes

Entre las primeras investigaciones sobre comprensión del muestreo destacan las de los psicólogos Kahneman y Tversky (1972), quienes reportan muchas personas evalúan la

probabilidad de que una muestra ocurra basando sus juicios en la similitud de características que la muestra tiene con la población (heurística de representatividad), aunque este sea pequeña (sesgo conocido como creencia en la *ley de los pequeños números*). En una continuación de su trabajo, Kahneman, Slovic y Tversky (1982) reportan que muchas personas, cuando hacen juicios a partir de muestras aleatorias, son proclives a ver las muestras desde una perspectiva singular (muestras individuales) en lugar de una perspectiva de distribución (colección de muestras), se enfocan en las causas que produce un resultado particular y evalúan la probabilidad basándose en la información disponible del caso que tienen a la mano (heurística de disponibilidad).

Torok y Watson (2000) estudian el desarrollo de la comprensión de la variabilidad muestral en estudiantes de nivel elemental y medio. En los resultados identifican cuatro niveles de desarrollo los estudiantes (A, B, C y D) que son progresivos en nivel de complejidad. Los estudiantes de nivel A tienden a responder a preguntas de predicción con números individuales, los de nivel B tienen a proporcionar sub-rangos de variabilidad y los de nivel C y D, utilizaron principalmente valores promedios alrededor del cual se ubicaban los valores. Los autores recomiendan explorar la variabilidad en el salón de clases muchos años antes de la introducción de medidas para el cálculo de la variabilidad como es el caso de la desviación estándar.

Saldanha y Thompson (2002) identifican dos concepciones de muestra y muestreo que surgieron en el contexto de un experimento de enseñanza con estudiantes de bachillerato. Una concepción caracterizada como *aditiva*, donde se concibe a una muestra como una versión a pequeña escala de la población. La otra concepción es caracterizada como *multiplicativa* e implica una red de imágenes interrelacionadas en la cual los resultados muestrales son vistos como un subconjunto de la población, que en la medida que el muestreo es repetido, se ven como casos similares que se acumulan para formar una distribución. Saldanha y Thompson señalan que una concepción multiplicativa del muestreo es deseable para desarrollar un esquema adecuado sobre la inferencia estadística, sin embargo no es una actividad trivial, pues resultó ser una tarea compleja para la mayoría de los estudiantes que participaron en la investigación.

Inzunza (2006) reporta algunas dificultades y concepciones erróneas que estudiantes universitarios tuvieron cuando resolvieron algunas tareas sobre la variabilidad. En los ítems gráficos propuestos por el autor, algunos estudiantes consideraron que la variabilidad depende de la cantidad de datos y está asociada con la irregularidad de la distribución, más que con su dispersión. En los ítems de contexto estadístico -tales como muestras y encuestas- tendieron a ignorar el proceso de muestreo como causa de variabilidad en si misma; mostraron además una comprensión superficial de la desviación estándar como medida de variabilidad al no relacionarla correctamente con propiedades de la distribución de datos.

Begué, Batanero y Gea (2018), analizan la comprensión intuitiva de la relación entre la proporción en una población y el valor esperado de la proporción muestral y de la variabilidad de dicha proporción en función del tamaño de la muestra –conceptos muy ligados a las distribuciones muestrales- con estudiantes españoles de segundo y cuarto curso de la educación secundaria obligatoria. Los resultados indican una buena percepción del valor esperado de la proporción en las situaciones de muestreo propuestas, pero menos de la cuarta parte de los estudiantes parecen percibir correctamente el efecto del tamaño de la muestra sobre la variabilidad muestral. La mayoría de los estudiantes, en especial los de cuarto curso, alcanzan el nivel

proporcional de razonamiento sobre el muestreo, pero pocos llegan al nivel distribucional de razonamiento sobre muestreo.

Orta y Sánchez (2018) por su parte, construyen una jerarquía basada en la taxonomía SOLO para evaluar los niveles de razonamiento sobre variación estadística de estudiantes de nivel medio superior en contextos de riesgo. Los resultados muestran que el desarrollo del razonamiento en sus niveles más bajos solo considera un valor de cada conjunto, y en sus niveles más altos articula varios datos de cada conjunto e integra consideraciones sobre la variabilidad, en particular, en forma de riesgo. En el análisis de las respuestas se revela la importancia de que los estudiantes consideren la media aritmética. Los autores consideran que el contexto de riesgos ofrece la oportunidad a los estudiantes de valorar la importancia de percibir y analizar la dispersión de los datos.

En resumen, los resultados de investigación descritos ponen de relieve la complejidad del muestreo y la variabilidad muestral para los estudiantes, independientemente de su nivel educativo.

3. Marco conceptual

En el muestreo intervienen una serie de conceptos que se entrelazan para dar lugar a una distribución muestral, que es la base de los métodos de inferencia estadística. Dierdorp Bakker, Maanen y Eijkelhof (2016, p. 39) identifican como parte de la red conceptual que subyace al muestreo, las ideas de tamaño de muestra, proceso aleatorio, distribución, intervalo intuitivo de variabilidad y relación entre muestra y población. En este sentido, los estudiantes requieren:

- Comprender que el incremento en el tamaño de muestra conduce a estimaciones más precisas de la probabilidad y de características poblacionales.
- Comprender que las mediciones repetidas de un mismo fenómeno es un proceso aleatorio que conduce a resultados distintos, y que las inferencias están influenciadas por la muestra seleccionada.
- Comprender que la extracción de muchas muestras permite formar una colección o distribución que puede ser descrita por medidas de centro y variabilidad.
- Tener una idea razonable de variabilidad alrededor de un valor esperado en un proceso aleatorio.
- Comprender que, como consecuencia de la variabilidad, puede proporcionar una imagen distorsionada de la población.

Por su parte, Harradine, Batanero y Rossman (2011) señalan que para la comprensión integral del muestreo se requiere tener claridad sobre tres tipos de distribuciones:

- La distribución de probabilidad que modela los valores de una variable aleatoria y depende de algún parámetro de la población.
- La distribución los valores de una variable en una muestra aleatoria, sobre la cual se calculan estadísticos (por ejemplo: media y desviación estándar) que pueden ser usados para inferir sobre parámetros de la población.
- La distribución muestral del estadístico, que modela todos los valores que puede tomar un estadístico en el conjunto de las muestras posibles de un tamaño dado que pueden ser seleccionadas de una población.

En relación con lo anterior, en el proceso de selección repetida de muestras de una población, se pueden identificar tres niveles o ámbitos que guían el proceso de simulación:

- Definir el modelo o distribución de la población.
- Seleccionar una muestra de la población y calcular el estadístico de interés en forma repetida para muchas muestras de la población.
- Registrar los valores de los estadísticos calculados para generar una colección y representarlos gráficamente para construir la distribución muestral.

Saldanha y Thompson (2002) identifican los pasos anteriores como un proceso de tres niveles que conforman un esquema centrado en imágenes repetidas del proceso de muestreo de la población. Los autores reportan que cuando los estudiantes pueden visualizar el proceso de simulación a través del este esquema pueden lograr una comprensión del proceso y lógica de la inferencia. El primer nivel lo representa la población, el segundo nivel las muestras y los estadísticos calculados en cada una de ellas, y el tercer nivel es representado por la distribución muestral del estadístico. En el último nivel, una muestra cualquiera puede ser comparada con el resto de la distribución para determinar si es una muestra inusual o es una muestra que tiene mucha probabilidad de ocurrir, entre otro tipo de información útil para una inferencia.

4. Método

4.1. Los sujetos de estudio y el escenario de investigación

Los sujetos de estudio fueron 16 estudiantes de ciencias sociales (19-20 años) de una universidad mexicana que estaban tomando el curso de probabilidad y estadística. Sus conocimientos en el tema consistían de estadística descriptiva y probabilidad básica. El estudio se llevó a cabo en un aula en la que los estudiantes individualmente tuvieron acceso a una computadora con acceso a internet en la que se instaló el software Fathom para responder las actividades de la trayectoria de aprendizaje.

4.2. Características de las actividades e instrumentos de recolección

La trayectoria de aprendizaje tiene como propósito central la comprensión de la variabilidad muestral de un estadístico (media y proporción), su relación con el tamaño de muestra y la interpretación de la información que proporciona la distribución muestral. El punto de partida de la trayectoria inicia con la exploración de la relación entre población (caracterizada por un parámetro) y muestras (caracterizadas por un estadístico) variando el tamaño de muestra como un primer paso hacia la identificación y comprensión de la variabilidad muestral y el error de muestreo- Al final se construye empíricamente la distribución muestral de una proporción y se estudian algunas de sus propiedades.

La trayectoria de aprendizaje esta formada por tres actividades. Cada actividad iniciaba con la lectura de algún artículo, video o resultados de una encuesta para introducir a los estudiantes a la temática en cuestión. Los instrumentos de recolección fueron hojas de trabajo para cada actividad con instrucciones precisas y con espacio para sus argumentaciones, así como los archivos del software, que fueron recolectados en cada actividad. En todas las actividades se consideraron poblaciones con parámetros

conocidos y se seleccionaron muestras de tamaño creciente repetidamente, con un número creciente de simulaciones para formar una distribución empírica de al menos 1000 muestras. En las hojas de trabajo se planteaban preguntas orientadoras para que los estudiantes centraran su atención en los conceptos de interés. Los resultados de este artículo están basados en el análisis de las argumentaciones hechas por los estudiantes a las diferentes preguntas planteadas.

La primera actividad tenía como contexto los niveles máximos diarios de contaminación en la ciudad de México (2012-2016), los cuales constituyen la población objetivo con parámetros medibles: *media diaria de ozono y proporción de días con buena calidad del aire* (ver Figura 1).

	Fecha	Ozono	Dioxido_Azufre	Dioxido_Nitrogeno	Carbono	Particulas	Calidad
1793	11/27/2016	101	9	34	12	102	Mala
1794	11/28/2016	37	6	34	23	102	Buena
1795	11/29/2016	47	4	39	30	81	Buena
1796	11/30/2016	73	4	41	18	83	Regular
1797	01/12/2016	67	4	31	18	97	Regular

Figura 1. Niveles máximos de contaminación diarios en zona centro de la ciudad de México (2012-2016).

La segunda y tercera actividad tenían como contexto una encuesta realizada por Gallup sobre la opinión de los ciudadanos de Estados Unidos acerca de la construcción de un muro en la frontera con México (<https://news.gallup.com/opinion/polling-matters/209384/building-wall-sync-american-public-opinion.aspx>). Se requería en este caso construir un modelo simbólico con el software, teniendo como información que la *proporción de ciudadanos que están a favor del muro* era de 36%.

5. Descripción y discusión de los resultados

En la Actividad 1 los estudiantes se dieron a la tarea primeramente de calcular los parámetros de la población -así se consideró a un gran conjunto de datos- que les fue proporcionada en un fichero (media diaria de ozono y proporción de días con buena calidad del aire) y representaron gráficamente los resultados (ver Figura 2). A partir de la población se seleccionaron muestras de 50, 100 y 200 días en forma repetida para cada caso y se calcularon los estadísticos correspondientes a los parámetros descritos, se representaron en tablas y gráficas los resultados muestrales. La actividad promovía la visualización del comportamiento entre parámetros y estadísticos conforme se repetía el muestreo e incrementaba el tamaño de muestra.

Las preguntas orientadoras de la actividad fueron: ¿qué sucede con los estadísticos (media de ozono y proporción de días con buena calidad del aire) cuando se selecciona una y otra muestra?, ¿qué efecto tiene incrementar el tamaño de la muestra en la variabilidad de los estadísticos?, ¿qué efecto tiene incrementar el tamaño de la muestra en la estimación de los parámetros y en el error de muestreo?

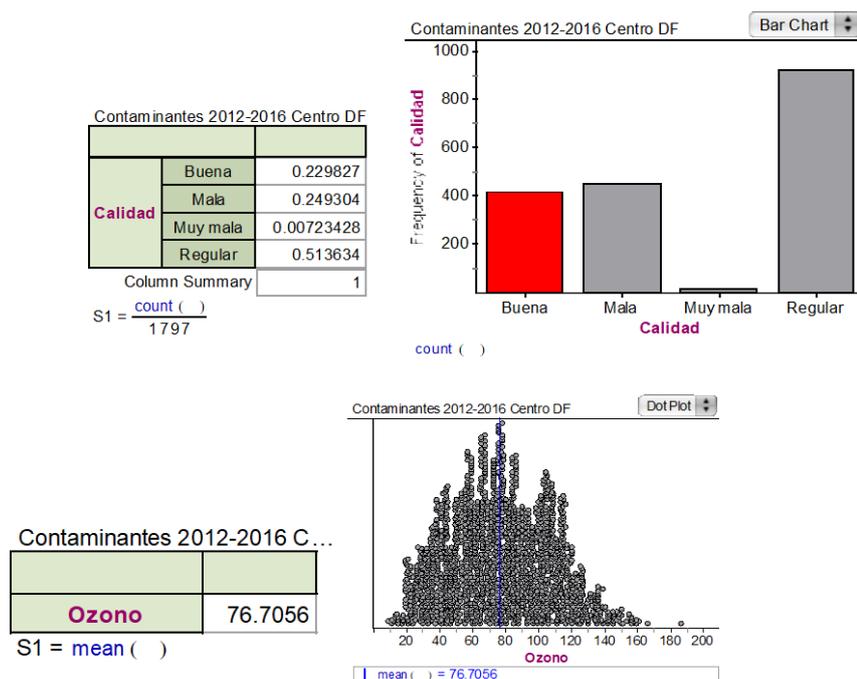


Figura 2. Cálculo de parámetros y representaciones gráficas de las variables poblacionales

En la primera actividad se observó que muchos estudiantes mostraron una comprensión intuitiva de la variabilidad muestral y establecieron una relación correcta con el incremento del tamaño de muestra. Sin embargo, el lenguaje que utilizaron algunos estudiantes en sus argumentaciones es coloquial y carece de los términos estadísticos que den precisión a sus afirmaciones. Una situación análoga es reportada por Johnson et al. (2010) en una investigación sobre margen de error y confianza con estudiantes de posgrado. Algunos ejemplos siguen:

Los resultados son más representativos (Carolina).

Los resultados son más exactos y cercanos a la población (Fernanda).

Se ha concentrado más los valores de la media de ozono y la proporción de días con buena calidad del aire (Gustavo).

Se concentran más los datos hacia la media del parámetro (Hascibe).

El efecto que tiene al incrementar el tamaño de la muestra es que los resultados tienden a ser más representativos (Marcela).

Conforme se incrementa el tamaño de la muestra se van aglomerando los datos, y la cantidad de la media se acerca más a la original (María).

Las respuestas de estos estudiantes contienen frases como *representativos*, *más exactos* y *cercanos*, *se concentran más*, *tienden a ser más representativos*, *se van aglomerando*, las cuales muestran que los estudiantes han adquirido la noción de que la variabilidad se reduce al aumentar el tamaño de muestra, que muestras más grandes tienden a ser más representativas que las muestras pequeñas y que los resultados son más cercanos a los de la población. Sin embargo no todos los estudiantes lograron este nivel de razonamiento en su primera actividad, como fue el caso de Fabricio y Miguel Angel entre otros:

Conforme se incrementa el tamaño de la muestra, la variabilidad en los datos estadísticos será cada vez más frecuente (Fabricio).

Los valores de la media disminuyen (Miguel Angel).

Se observó una apreciación del tamaño de muestra en la precisión de la estimación por parte de muchos estudiantes.

Los resultados tienden a ser más precisos (Carolina).

La estimación se vuelve más correcta (Daniela).

Más precisión en los resultados, tenemos más datos para calcular por lo que es más exacto (Fernanda).

Mientras se incrementa el tamaño de la muestra existe menos variabilidad y se acerca a los datos reales (Fabricio).

Que la estimación se vuelve más correcta y el error muestral se reduce (Gustavo).

Se acerca más al resultado original, y es más representativo (María).

Ahora encontramos frases como *más exactos*, *más correcta*, *más representativa*. En este caso Carolina habla de precisión y no de representatividad como en la situación anterior lo cual muestra una idea correcta del incremento del tamaño de muestra en una estimación. Ahora Fabricio es mucho más claro que en el caso anterior y da muestra que tiene una noción correcta de la relación entre el tamaño de muestra y la variabilidad, agrega además que *se acerca a los datos reales*, quizá refiriéndose con ello a la población. Gustavo por su parte señala que el error muestral se reduce.

Respuestas de estudiantes que no logran comprender la relación entre el tamaño de muestra y la estimación son las siguientes:

El resultado puede ser más favorable para la investigación (Janeth).

El resultado tiende a ser favorable para los días con buena calidad del aire ya que aumentan los días (Jesus Manuel).

Ambos estudiantes utilizan el término *favorable*, quizá queriendo referirse a la precisión.

Se observa una apreciación correcta en casi todos los estudiantes sobre el efecto del tamaño de muestra en el error de muestreo.

El error de muestreo se hace menor (Carolina).

Conforme se incrementa el tamaño de la muestra, el error de muestreo disminuye (Marcela).

Se reduce el error muestral (Hascibe).

El error de muestreo se reduce (Daniela).

En la Actividad 2, cuyo contexto provenía de una encuesta acerca de la opinión de ciudadanos de Estados Unidos sobre la construcción del muro en la frontera con México, se requería la construcción de un modelo para generar los datos (ver Figura 3), a diferencia de la actividad anterior en la cual los datos fueron proporcionados en un fichero. La variable de interés era la *proporción de personas a favor del muro*. En la encuesta se empleó una muestra aleatoria con alcance nacional de 800 ciudadanos, y el 36% respondieron estar a favor del muro.

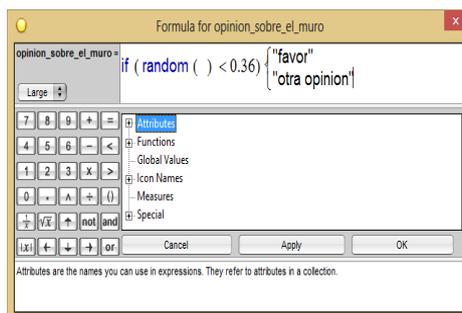


Figura 3. Modelo construido en Fathom para generar datos en actividades 2 y 3.

Se les indicó a los estudiantes que seleccionaran 10 muestras de tamaño 100 y que determinarían para cada muestra las proporciones de personas a favor del muro, después que calcularían la media y el rango de la distribución de proporciones. El mismo proceso debía hacerse para muestras de tamaño 500 y 1000. Las preguntas orientadoras fueron: a) identifica qué pasa con la media y el rango conforme se incrementa el tamaño de muestra, b) ¿para que tamaño de muestra observas más diferencia entre el estadístico muestral y el parámetro poblacional?, ¿qué tamaño de muestra (100, 500 o 1000) puede resultar más adecuado para estimar la proporción de personas a favor del muro?

La mayoría de los estudiantes mostraron en este nuevo contexto una comprensión intuitiva de la relación entre el tamaño de muestra y la variabilidad. Algunos estudiantes desarrollaron razonamientos más complejos que otros, incorporando el tamaño de muestra, la variabilidad de las proporciones y su centralidad relacionándola con el parámetro poblacional. Se identificaron los siguientes razonamientos de acuerdo a su nivel de complejidad:

Los estudiantes no comprenden la relación entre tamaño de muestra, variabilidad y centralidad (relación con el parámetro). Algunos consideran que a mayor tamaño de muestra existe mayor variabilidad, al igual que ocurrió en Begué, Batanero y Gea (2018):

El rango se hace más grande conforme la muestra se hace más grande; la media entre la muestra de 500 y 1000 es más parecida entre sí; varía menos con la muestra de 1000 (Fernanda).

Conforme se incrementa el tamaño de la muestra, la media y el rango van en aumento de igual forma (María).

En cuanto al rango, observamos que este varía, por tanto, conforme el tamaño de la muestra aumenta, los valores se acercan a la realidad (Kassandra).

Fernanda identifica poca variación en la media de la muestra al aumentar el tamaño de 500 a 1000, lo cual es correcto, pero incorrectamente señala que el rango se hace más grande para muestras más grandes; al referirse a la muestra en lugar de las muestras, está viendo la colección de 10 muestras como si fuera una sola muestra, por lo que se piensa que tiene problemas para ver las muestras en forma de distribución. Esta dificultad está reportada en estudios de Saldanha y Thompson (2002). Kassandra por su parte señala que conforme el tamaño de la muestra aumenta los valores se acercan a la realidad, quedando la duda a qué se refiere con la frase *los valores se acercan a la realidad*.

Los estudiantes relacionan el tamaño de muestra correctamente con la variabilidad o con la centralidad, como se muestra en el siguiente ejemplo:

Se va acercando más al resultado final de la probabilidad, 0.36, es decir que se va haciendo más exacto y se reduce el margen de variabilidad (Ana Lucía).

En la muestra de 100 tiende a variar más y su rango es mayor a comparación de las otras muestras. En la muestra de 500 los resultados tienen un rango menor y están cercanos unos con el otro. En la muestra de 1000 los resultados obtenidos son muy cercanos (Marcela).

Mientras más se amplía el número de muestras, se cierran más las proporciones y la media se acerca más a 0.36 (Daniela).

Ana Lucía acierta cuando señala que aumentando el tamaño de muestra se reduce el margen de variabilidad, pero para referirse al parámetro utiliza la palabra probabilidad. Por su parte Daniela señala que *mientras más se amplía el número de muestras se cierran más las proporciones*, refiriéndose quizá al tamaño de muestra. Obsérvese el uso de la frase *se cierran* para referirse las proporciones toman valores cercanos y por ende a la variabilidad.

Los estudiantes relacionan el tamaño de muestra correctamente con la variabilidad o con la centralidad y relacionan los valores muestrales con la proporción poblacional, como se observa a continuación:

Mientras más grande es la muestra menor variabilidad hay entre los resultados. En la de 500 observamos que el rango es mucho mayor, pero conforme hay más muestras, el rango se va disminuyendo. También en la de tamaño 1000 observamos que los valores de la proporción se parecen cada vez más a los dados por Gallup (Ana Cristina).

Conforme se incrementa el tamaño de la muestra, la media se acerca más a la proporción del número de personas a favor de la construcción del muro (36%) y el rango se hace más pequeño, puesto que el nivel de variabilidad se reduce (Gustavo)

La respuesta de Gustavo contiene un lenguaje estadístico adecuado y muestra que comprende la relación entre el tamaño de muestra, el centro y variabilidad de la distribución. Ana Cristina también recurre al uso del término *mas muestras* para referirse a un mayor tamaño de muestra como lo hicieron otros estudiantes.

Respecto a la relación entre el tamaño de muestra y el error de muestreo (diferencia entre el valor del estadístico y el valor del parámetro), solo dos estudiantes no mostraron comprensión de ello. Ejemplos de razonamientos correctos:

Con la más pequeña (tamaño 100) hay más error porque hay mayor variabilidad (Ana Cristina).

La muestra de tamaño 100, porque la variación se da entre el 0.25 y el 0.44, es más amplia que en los otros casos (Hascibe).

Para la primera muestra (tamaño 100) el error muestral es mayor porque la cantidad de casos de esta población es muy pequeña (Nicole).

Con la de 100 porque tiene menos gente, por lo tanto, más rango de error (Mónica).

Ana Cristina relaciona el error con la variabilidad, Hascibe por su parte identifica un intervalo de variabilidad más amplio para muestras pequeñas que para muestras grandes. Nicole refiere a la población cuando en realidad se debe referir a la muestra.

Igualmente respecto al tamaño de muestra más adecuado para estimar la proporción de personas a favor del muro se observó que los estudiantes tienen la idea correcta que a mayor tamaño la estimación es más precisa:

La muestra de 1000 resultados porque mientras más grande la muestra mayor es la probabilidad de tener más porcentaje de una población y los números se estiman mejor (Ana Cristina).

Conforme se incrementa el tamaño de la muestra, el error de muestreo disminuye (Marcela).

Conforme se incremente el tamaño de la muestra el error se va reduciendo cada más hasta llegar al margen indicado (cerca del 3.5%).

La de 1000, debido a que mientras más grande el tamaño de la muestra esta es más precisa (Jesús).

La muestra de 1000 personas, porque a mayor cantidad de sujetos en la muestra, el error muestral es menor y la posibilidad de estimar con mayor certeza es más certera (Nicole).

En suma hay muchas imprecisiones en el lenguaje de los estudiantes antes la abundancia de términos que requieren manejar en sus argumentaciones

Tamaño de muestra	Gráfica de la distribución muestral	Media y desviación estándar de la distribución muestral				
500		<p>Measures from Collection 3</p> <table border="1"> <tr> <td>Proporción_de_FAVOR</td> <td>0.0220936</td> </tr> <tr> <td></td> <td>0.358908</td> </tr> </table> <p>S1 = stdDev () S2 = mean ()</p>	Proporción_de_FAVOR	0.0220936		0.358908
Proporción_de_FAVOR	0.0220936					
	0.358908					
800		<p>Measures from Collection 3</p> <table border="1"> <tr> <td>Proporción_de_FAVOR</td> <td>0.359753</td> </tr> <tr> <td></td> <td>0.0175716</td> </tr> </table> <p>S1 = mean () S2 = stdDev ()</p>	Proporción_de_FAVOR	0.359753		0.0175716
Proporción_de_FAVOR	0.359753					
	0.0175716					
1000		<p>Measures from Collection 3</p> <table border="1"> <tr> <td>Proporción_de_FAVOR</td> <td>0.359804</td> </tr> <tr> <td></td> <td>0.0156247</td> </tr> </table> <p>S1 = mean () S2 = stdDev ()</p>	Proporción_de_FAVOR	0.359804		0.0156247
Proporción_de_FAVOR	0.359804					
	0.0156247					

Figura 4. Gráficas y cálculos realizados por Gustavo

En la actividad 3, que tenía el mismo contexto que la actividad anterior, nos propusimos que los estudiantes formaran colecciones con los estadísticos (proporción de personas a favor del muro) que calcularon en cada una de las muestras (al menos 1000), para así formar empíricamente distribuciones muestrales para muestras de tamaño 500, 800 y 1000 respectivamente. La idea era que desarrollaran una comprensión intuitiva sobre propiedades de las distribuciones ligadas al teorema del límite central y que interpretaran información de las distribuciones muestrales (por ejemplo: inusualidad de una muestra e intervalos de intuitivos para capturar un cierto porcentaje de muestras).

La parte operativa de la actividad se puede condensar en la Figura 4, donde se les pide: Analiza el centro (media), variabilidad (desviación estándar) y la forma de las distribuciones muestrales y el efecto que tuvo en ellas cambiar de tamaño de muestra (500, 800, 1000). Algunas respuestas son las siguientes:

La desviación estándar se hace mas pequeña entre más grande es la muestra, la variabilidad va disminuyendo (Ana Lucía).

El centro aparece en 0.36 en todas las gráficas, la distribución muestral pasó de estar ancha en la gráfica de 500 a irse haciendo más angosta conforme el tamaño de la muestra crecía (Carolina).

La media se mantuvo en los tres tamaños de muestra en 35%, más parecía a la media de la población en total. Pero en cuanto a la desviación estándar, ésta si va cambiando y haciéndose más pequeña mientras vamos aumentando el tamaño de la muestra. En resumen, podríamos decir que la variabilidad se reduce y la media se queda estática (Gustavo).

El centro sigue siendo 0.36 en todas las muestras, la variabilidad va disminuyendo conforme la muestra se va haciendo más grande. La forma de distribuciones muestrales vemos que se van haciendo más y mas angostas conforme la muestra fue aumentando y hay menos datos a los lados (Cristina).

En cuanto al centro (media) el resultado señala que a mayor tamaño de muestra, esta tiende a cerrarse más y así acercándose mucho más al centro. La variabilidad (desviación estándar) nos muestra que va en decremento dependiendo del tamaño de la muestra; reflejando que al haber un mayor tamaño de la muestra esta disminuye. Lo que significa que hay menor dispersión de la población. En sí el impacto que tuvo cambiar el tamaño de la muestra fue que se disminuyó el margen de error y hubo una menor dispersión de la población (Daniela).

Se observa una comprensión intuitiva del efecto que tiene aumentar el tamaño de muestra en la variabilidad de la distribución muestral, sin embargo se hace notar la falta de lenguaje estadístico para expresar las relaciones conceptuales y la confusión entre algunos conceptos. Por ejemplo, Daniela señala que hay menor dispersión en la población cuando aumenta el tamaño de muestra, en vez de referirse a la muestra. Esta confusión entre muestra, población y distribuciones muestrales ha sido muy reportada en otros estudios (Harradine et al., 2011; Inzunza, 2006;)

En cuanto a la pregunta: si partimos que el 36% de los ciudadanos de EU está a favor del muro, ¿qué valores consideras que son poco probables de ocurrir en una muestra considerando el caso de la distribución para muestras de tamaño 1000? Un total de 11 estudiantes identificaron bien que en las colas de la distribución aparecen estos valores. En general respondieron que entre 30%-31% (cola izquierda) y 40%-41% (cola derecha), es decir a un distancia de 4% o 5% aproximadamente, por debajo y por encima del parámetro ($P=36\%$) ocurren para los estudiantes muestras consideradas inusuales.

En esta misma idea de planteó la pregunta: ¿en cuál de las distribuciones una muestra puede tener más error de muestreo? Un total de 14 estudiantes respondieron correctamente señalando que en la muestra de tamaño 500, lo que muestra una buena comprensión de que menores tamaños de muestras inducen a mayores errores de muestreo.

6. Conclusiones

El razonamiento informal que los estudiantes desarrollaron sobre el muestreo después de completar la trayectoria de aprendizaje contiene elementos conceptuales que forman parte de un esquema adecuado del muestreo, como lo define Dierdrop *et al.* (2016). Sin embargo el lenguaje estadístico, que ha sido utilizado por muchos estudiantes en sus

razonamientos, es un tanto coloquial para su nivel de estudios, lo cual podría deberse a su débil conocimiento de estadística que tenían antes del estudio. También ser observado en muchos casos el uso de un término por otro, como el caso de muestra por población y viceversa, nuestra conjetura es que la confusión se debe a la diversidad de conceptos que se involucran y los estudiantes no logran tener clara la relación entre ellos.

En suma, los resultados del estudio muestran que una enseñanza del muestreo y la variabilidad muestral desde una perspectiva informal basada en simulación puede ayudar a los estudiantes a construir intuiciones correctas que son deseables antes de abordar los conceptos desde una perspectiva más formal en el caso de estudiantes universitarios.

Referencias

- Begué, N., Batanero, C. y Gea, M. (2018). Comprensión del valor esperado y variabilidad de la proporción muestral por estudiantes de educación secundaria obligatoria. *Enseñanza de las Ciencias* 36(2), 63-79.
- Dierdorff, A., Bakker, A., Maanen, J. y Eijkelhof, H. (2016). Supporting students to develop concepts underlying sampling and to shuttle between contextual and statistical spheres. En D. Ben-Zvi y K. Makar (Eds.), *The teaching and learning statistics: International perspectives* (pp. 37-48). New York: Springer.
- 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 teacher education* (pp. 235-246). New York: Springer.
- Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics*, 6, 187-205.
- Inzunza, S. (2006). Some conceptions and difficulties of university students about variability. En S. Alatorre, J.L. Cortina, M. Sáiz y A. Méndez, (Eds). *Proceedings of the 28th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 244-250). Mérida, México,.
- Johnson, E., Redmon, S., Bartlo, J. y Noll, J. (2010). *Tertiary Students' Informal Conceptions of Margin of Error and Confidence*. Disponible en: http://sigmaa.maa.org/rume/crume2009/EJohnson_SHORT.pdf
- Kahneman D. y Tversky, A. (1972). Subjective probability: A Judgment of representativeness. *Cognitive Psychology* 3, 430-454.
- Kahneman, D., Slovic, P. y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, Inglaterra: Cambridge University Press.
- National Governors Association Center for Best Practices and Council of Chief State School Officers (2010). *Common Core State Standards for Mathematics*. Disponible en: http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf
- Orta, A. y Sánchez, E. (2018). Niveles de razonamiento sobre variación estadística de estudiantes de nivel medio superior al resolver problemas en un contexto de riesgo. *Educación Matemática*, 30(1), 47-71.
- Pfannkuch, M. (2008). Building sampling concepts for statistical inference: A case study. Trabajo presentado en el *11 International Congress on Mathematics Education*. Monterrey, México: ICMI. Disponible en:

<https://pdfs.semanticscholar.org/b294/f20f83537eaa010351ed1133ee379fcde5bb.pdf>

- Rubin, A., Bruce, B. y Tenney, Y.: 1991, Learning about sampling: Trouble at the core of statistics. En D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (Vol. 1, pp. 314-319), Dunedin, New Zealand: ISI.
- Saldanha, L. A. y Thompson, P. W. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics* 51(3), 257-270.
- Torok, R. y Watson, J.M. (2000). Development of the concept of statistical variation: An exploratory study. *Mathematics Education Research Journal*, 12, 147-169.
- Watson, J. (2006). *Statistical literacy at school: Growth and goals*. New Jersey. Lawrence Erlbaum.
- Watson, J. & Moritz, J. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44-70.