MICRO DESARROLLO DE LAS CONCEPCIONES DE DISTRIBUCIÓN MUESTRAL SIMULADA DE ESTUDIANTES DE BACHILLERATO

MICRO DEVELOPMENT OF THE CONCEPTIONS OF SIMULATED SAMPLE DISTRIBUTION OF HIGH SCHOOL STUDENTS

Sepúlveda, F. y Sánchez, E.

Departamento de Matemática Educativa, Cinvestav-IPN, México

Resumen

El objetivo de esta investigación fue entender la influencia de las actividades de diseño (uso de software estadístico Fathom, actividades de resolución de problemas en parejas e intervenciones oportunas del instructor), en la conceptualización de la distribución muestral simulada (DMS) en situaciones de problemas de pruebas de significación por parte de los estudiantes. Para lograrlo, se exploran y analizan las respuestas de 18 parejas de estudiantes, obtenidas con apoyo de tecnología, a cuatro problemas de pruebas de significación que les fueron administrados a lo largo de cuatro sesiones de trabajo. Los resultados muestran que poco más de un tercio de las parejas logra, en el curso de las actividades, transitar de una concepción concreta de DMS formada por muestras reales a una concepción de la DMS como un patrón de variación con una métrica para evaluar la inusualidad de una muestra dada; esta es ya una concepción matemática de la DMS.

Palabras clave: *Distribución muestral empírica, distribución muestral simulada, pruebas de significación.*

Abstract

The aim of this research was to understand the influence of design activities (using statistical software Fathom, problem-solving activities in pairs, and timely instructor interventions) on high school students' conceptualization of simulated sampling distribution (SSD) in significance testing problems. To achieve this, the answers of 18 pairs of students, obtained with the support of technology, to four problems of significance tests administered over four work sessions, are explored and analyzed. The results show that just over one-third of the pairs managed, during the activities, to move from a concrete conception of SSD formed by real samples to a conception of the SSD as a pattern of variation with a metric to assess the unusualness of a given sample; this is already a mathematical conception of the SSD.

Keywords: *Empirical sampling distribution, simulated sampling distribution, significance tests.*

INTRODUCCIÓN

El concepto de distribución muestral es crucial para entender los procedimientos de inferencia estadística. Por otro lado, las pruebas de hipótesis y de significación son procedimientos frecuentemente utilizados en la investigación científica. Una prueba de hipótesis o de significación requiere razonar acerca de y con la distribución muestral. Los temas de distribución muestral y pruebas estadísticas suelen ser estudiados en los cursos universitarios de estadística, sin embargo, varios informes de investigación en educación dan cuenta de grandes dificultades de los estudiantes con ambos conceptos (Aquilonius y Brenner, 2015; Chance et al., 2004), por lo que es un problema actual de la Educación Estadística. Esta investigación se restringe al estudio de la distribución muestral y las pruebas de significación sin abordar el tema más general de pruebas de hipótesis.

Teniendo en cuenta las siguientes tres recomendaciones tomadas de la literatura: 1) Introducir a edades más tempranas conceptos fundamentales (Heitele, 1975), 2) explorar variantes informales de

Sepúlveda, F. y Sánchez, E. (2023). Micro desarrollo de las concepciones de distribución muestral simulada de estudiantes de bachillerato. En C. Jiménez-Gestal, Á. A. Magreñán, E. Badillo y P. Ivars (Eds.), *Investigación en Educación Matemática XXVI* (pp. 507–514). SEIEM.

los conceptos y procedimientos estadísticos (Zieffler et al., 2008) , 3) Utilizar recursos tecnológicos (Biehler et al., 2013) nos hicimos la siguiente pregunta: ¿Qué concepciones muestran los estudiantes de bachillerato de la distribución muestral en el contexto de actividades de resolución de problemas de pruebas de significación estadística con ayuda de tecnología? El objetivo de la investigación es describir e interpretar las concepciones sobre la distribución muestral simulada que los estudiantes se forman al buscar resolver problemas de pruebas de significación y cómo tales concepciones se transforman con base en las intervenciones del instructor.

ANTECEDENTES

Las investigaciones sobre distribuciones muestrales empíricas y simuladas a nivel bachillerato y con apoyo de tecnología son recientes y escasas, por ejemplo: Bakker y Gravemeijer (2004), van Dijke-Droogers et al. (2019). No obstante, en el nivel universitario se encuentran varios estudios que cubren aspectos de las distribuciones muestrales con y sin tecnología. Un buen recuento de ellas se puede consultar en Saldanha y Thompson (2014). Sobre pruebas estadísticas con profesores de secundaria encontramos el trabajo de López-Martín et al. (2018) y en el nivel universitario podemos mencionar a Aquilonius y Brenner (2015) y Lane-Getaz (2017), por citar sólo estudios relativamente recientes.

La presente comunicación se apoya en los datos generados en la investigación de García-Ríos (2017) que realizó con estudiantes de bachillerato. En dicho trabajo se informa que un tercio de sus sujetos logra, gracias a su participación en una secuencia de cuatro problemas de pruebas de significación, resolver los dos últimos utilizando adecuadamente una distribución muestral simulada. Pero no responden a la pregunta formulada en este estudio; para hacerlo, nosotros retomamos los datos de García-Ríos y los analizamos desde una nueva perspectiva.

MARCO CONCEPTUAL

La distribución de los valores de un estadístico tomados sobre todas las muestras posibles de un tamaño dado de la población constituye la *distribución muestral* del estadístico. Una distribución muestral se puede aproximar de dos formas: Una distribución muestral teórica (DMT) se obtiene con métodos de la teoría de la probabilidad, mientras que una distribución muestral simulada (DMS) se obtiene mediante métodos de simulación computacional (Moore, 2000). Una distribución muestral empírica (DME) es obtenida calculando el estadístico sobre un conjunto de muestras efectivamente tomadas de una población real (material y finita). Si el número de muestras es suficientemente grande la DME será próxima a una DMS, no obstante, la diferencia entre muestras real y simuladas es fundamental y se resalta en la presente comunicación.

Una *hipótesis nula* es la hipótesis de que el fenómeno que se quiere demostrar está ausente, es la hipótesis que se quiere nulificar o rechazar. Cuando se tiene una muestra aleatoria de la población se define el *p-valor*; éste consiste en la probabilidad condicional de obtener un valor del estadístico igual o más extremo que el valor de la muestra-dato, suponiendo cierta la hipótesis nula. El razonamiento de una prueba de significación es: *Si una muestra es inusual suponiendo cierta la hipótesis nula entonces se rechaza la hipótesis nula*. Para determinar si la muestra-dato es usual o inusual se acuerda un *nivel de significación* α y el siguiente criterio: Si el p-valor es menor que α decimos que la muestra es inusual, en caso contrario es usual. El valor preciso de α es convencional y se establece por los científicos antes de llevar a cabo la prueba; un valor recomendado es: $\alpha=0.05$, es decir de 5%. Una prueba más fuerte podría utilizar $\alpha=0.01$ o 1%. (Lecoutre y Poitevineau, 2014)

Nociones de concepción e inferencia estadística informal

El término *concepción* es difícil de definir (Ruiz-Higueras,1994), pero para los fines de este trabajo, informalmente se considera como una idea o representación mental que un individuo se forma sobre un concepto o fenómeno. Siguiendo a la autora citada, una concepción suele ser una forma embrionaria del conocimiento de un concepto o fenómeno, aunque, en ocasiones, puede representar

un obstáculo para su formación y desarrollo. En cualquier caso, la didáctica estudia las concepciones que los estudiantes se forman de contenidos específicos para entender su proceso de aprendizaje. Aquí describimos tres concepciones de los estudiantes acerca de la noción de DMS.

En las últimas décadas se ha propuesto que los estudiantes tengan experiencias con un enfoque informal en la inferencia estadística. Sin embargo, la noción de inferencia informal es relativa, ya que la informalidad depende de las tareas, los conceptos y el nivel escolar de los sujetos (Pratt y Ainley, 2008). Para abordar la transición de lo informal a lo formal en la inferencia, nos centramos en el nivel de abstracción de conceptos básicos de muestreo, que van desde una concepción concreta de ellos hasta una concepción matemática. Inferimos las concepciones de los estudiantes a partir de sus respuestas a los problemas y las clasificamos en tres fases: práctico, informal y matemático-formal. Las respuestas de nuestros estudiantes no alcanzan el nivel matemático-formal.

MÉTODO

El presente es un estudio retrospectivo de los datos generados en una investigación de doctorado (García-Ríos, 2017). Tres elementos formaron parte del diseño: 1) uso de software estadístico Fathom, 2) actividades de resolución de problemas abordados en parejas y 3) intervenciones oportunas del instructor. La interpretación de datos generados permitió inferir las concepciones de los estudiantes sobre la DMS, que es el objetivo de esta investigación.

Los participantes fueron un grupo de 36 estudiantes (16-17 años), organizados en 18 parejas, de un sistema de bachillerato de México, en el que García-Ríos (2017) era el profesor regular de la asignatura de Matemáticas, y diseñó e implementó las actividades en su grupo como parte de sus estudios de doctorado.

Para la simulación se utilizó una aplicación del software Fathom. Este es un software dinámico para aprender y hacer estadística en el nivel bachillerato y universitario que permite a los estudiantes explorar y analizar datos tanto mediante el cálculo como de manera visual. Brinda a los estudiantes las posibilidades de arrastrar datos para formar automáticamente distribuciones; visualizar en tiempo real cambios en las gráficas cuando se cambian parte de los datos; vincular múltiples representaciones de datos y crear simulaciones (Biehler et al., 2013).

Procedimiento y organización de la secuencia

Figura 1. Problemas

Problema 1. La propaganda de Coca cola presume que la mayoría (más del 50%) de la población que consume refresco prefiere su refresco en lugar de Pepsi. Para comprobarlo se hizo un experimento donde se les dio dos vasos de refresco (uno con Coca y otro con Pepsi) a 60 personas escogidas al azar y decidían cuál le gustaba más. De los 60 participantes 35 personas prefirieron Coca Cola. Con estos resultados ¿es correcta la hipótesis "más del 50% de la población prefiere Coca Cola que Pepsi"?
Problema 2. El mismo contexto del problema 1 con las siguientes condiciones: Se hace el experimento a 180 personas elegidas al azar y de ellas 104 prefieren Coca Cola.
Problema 3. Si la producción diaria de la máquina de una fábrica tiene más del 10% de artículos defectuosos es necesario repararla. Para revisar la calidad de la maquina el supervisor toma una muestra aleatoria de 120 piezas del día y contiene 16 piezas defectuosas. Con esos datos ¿debe el supervisor solicitar reparar la maquina?
Problema 4. El fabricante de una medicina afirmó que su producto es 80% eficaz para aliviar una alergia. Para verificar esta hipótesis se realizó un experimento con 100 personas que padecían de la alergia, y la medicina alivio a 74. ¿Se puede rechazar la información del fabricante?

La actividad se organizó en torno de cuatro problemas de pruebas de proporciones (Figura 1) aplicados en cuatro sesiones de 90 minutos cada una. Se formaron parejas de estudiantes a los que, en cada sesión, se les pidió que resolvieran un problema con apoyo del software. Al final de cada sesión debían redactar un reporte escrito en Word orientado por consignas como: "Describan y expliquen es su conclusión" "Digan qué tan seguros están de dicha conclusión" etc.

En la primera sesión se instruyó a los estudiantes para simular en Fathom muestreo repetidos y obtener la DMS. Se les administró entonces el primer problema para que lo resolvieran en parejas durante la sesión y al final entregaran un informe capturado en un archivo de Word con imágenes del software y las explicaciones pertinentes. En la segunda sesión, el instructor revisó e hizo observaciones a los procedimientos de solución que las parejas de estudiantes entregaron en la sesión previa. Les introdujo el método del p-valor enfatizando en la importancia de distinguir entre muestras típicas o usuales y muestra atípicas o inusuales. En la tercera sesión, el instructor hizo indicaciones acerca de las soluciones al problema 2 y volvió a exponer el método del p-valor enfatizando ahora en los conceptos de p-valor y nivel de significación del 5%. Administró entonces el problema 3. En la cuarta sesión sólo administró el problema 4 como evaluación final.

Procedimiento de análisis

El stock de datos consiste en las respuestas de los estudiantes a los problemas y preguntas-guía. Las respuestas fueron codificadas y comparadas entre sí, con la intención de observar patrones de respuestas que, desde el punto de vista de los autores, revelen el razonamiento de los estudiantes. Se identificaron varias categorías que dan cuenta de aspectos importantes del razonamiento de los estudiantes; estas se exponen a continuación.

RESULTADOS

En esta sección exhibimos textos de las respuestas de los estudiantes que ofrecen indicios de su razonamiento frente a los problemas. Hay tres fuentes que nutren sus razonamientos, una consiste en sus intuiciones y conocimientos previos sobre muestreo, la segunda la constituye la información que obtienen de operar con el software y la tercera, son las indicaciones del instructor que incluyen correcciones y la introducción de conceptos nuevos de muestreo para los estudiantes. Las parejas (Ei, Pj) es una abreviatura de (Pareja de estudiantes i, Problema j) y significa que es un texto que escribió la pareja i en sus hojas de trabajo en respuesta al problema j.

Una intuición de base

Es conocido que muchos estudiantes tienen la intuición errónea de que toda muestra es representativa, así de manera automática resuelven los problemas de estimación con la idea de que: *La proporción de la muestra-dato es "igual" a la proporción de la población*. A partir de esta, los estudiantes derivan opiniones como la siguiente:

> (E8, P3). [en el problema 3 la muestra-dato es de 120 artículos con 16 defectuosos], si nos dicen que la máquina no puede trabajar si produce más del 10% de piezas defectuosas, entonces tendríamos que, a partir de 13 piezas defectuosas (más del 10%), la máquina debe ser reparada.

La mayoría de los que tienen dicha intuición no tarda en sustituirla por la que sigue, en la que se reconoce cierta variabilidad: *La proporción de la muestra-dato es "aproximada" a la proporción de la población*. Por ejemplo, la siguiente pareja se basa en esta intuición:

> (E9, P2): Sabemos que el 50% exacto de 180 es 90, sin embargo, podemos tomar un rango en el cual determinar hasta donde es el 50% [Es decir, un rango alrededor de 90 en el que caen todas las proporciones que pueden considerarse provenientes de una población con 50% de preferencias de ambos refrescos]

Esta intuición es similar a lo que Saldanha y Thompson (2002, p. 257) llaman una *concepción multiplicativa de muestra*, la cual definen como: "una versión cuasi proporcional a pequeña escala de la población". No es una intuición primaria, sino el resultado de la educación básica que Watson y Moritz (2000) describen como un desarrollo que culmina con una concepción representativa de muestra. Pero en este desarrollo los estudiantes tienen como referente muestras reales, con toda la imaginería que esto conlleva.

Consideración de la DMS como si fuera una DME

Dos características en las respuestas de los estudiantes nos hacen pensar que éstos interpretan a la DMS como si fuera una DME: 1) que se refieren a las muestras simuladas como muestras reales y 2) que utilizan un *criterio de doble mayoría* para hacer la inferencia. Ambas se describen a continuación.

Un rasgo general en las respuestas de los estudiantes a los problemas es que describen aspectos de la simulación que llevaron a cabo utilizando términos del contexto del problema que intentan resolver. Esta manera de expresarse sugiere que imaginan que están operando con resultados obtenidos de muestras reales. Por ejemplo:

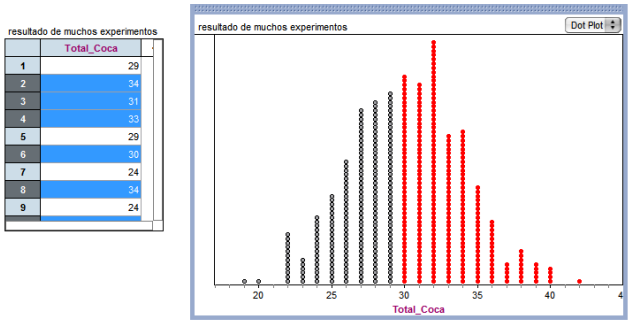
> (E3, P1): …se realizaron 500 encuestas de 60 personas cada una. En la gráfica podemos observar que la mayoría prefiere la coca cola, es decir, que la mayoría de los encuestados inclinan su gusto por este refresco

> (E9, P1): Se realizaron 500 encuestas de 60 personas, por lo tanto, se encuestaron a 30 000 personas.

En las respuestas de los estudiantes al primer problema, no sólo el lenguaje indica que piensan que la DMS está formada por muestras reales, sino que el criterio *de doble mayoría* que utilizaron para resolverlo lo corrobora.

Criterio de "doble mayoría"

Figura 2. DMES del equipo E7



Este criterio consiste en el siguiente razonamiento: Si en la mayoría de las muestras hay una mayoría de personas que prefiere Coca, entonces en la población hay más personas que prefieren Coca. Doce parejas aplicaron este razonamiento sobre la DMS que obtuvieron en el problema 1. Los estudiantes llaman "encuestas" a cada punto muestral de la distribución y creen que en la DMS hay muestras reales, de modo que creen obtener información de la población. Por ejemplo, La respuesta del equipo E7 es representativa del razonamiento subyacente a dicho procedimiento. Generan una DMS con p=0.5 y N=60, y argumentan su conclusión de la siguiente forma:

> (E7, P1): [La hipótesis de que hay más preferencia por la Coca-Cola es correcta] debido a que en la simulación se muestran que en 285 encuestas de 500 hubo más del 50% que prefirió Coca-Cola de 60 personas entrevistadas en cada una de las encuestas. En cambio, en 215 encuestas realizadas se mostró un gusto en menos del 50% de 60 personas (Figura 2) [Nota: consideraron erróneamente las muestras con exactamente 50% de éxitos como favorables a la hipótesis de que Coca-Cola tiene mayor preferencia]

Muestras simuladas y la hipótesis del simulacro

El uso del lenguaje del contexto del problema para referirse a aspectos del proceso de simulación y los procedimientos que proponen, muestran que los estudiantes actúan como si la DMS fuera una representación de muestras reales, por ejemplo, una DME. No obstante, podemos observar que en algunas respuestas a las preguntas que se les pidió responder tales como: ¿Será posible que sus conclusiones sean incorrectas aun haciendo bien todo el proceso? ¿Por qué? ¿Qué tan seguros están

de su conclusión? los estudiantes señalan que un problema es que la solución que proponen se realiza con base en muestras simuladas y no reales. Por ejemplo:

> (E5, P1): …el resultado no es absoluto, ya que esto es solo una simulación y no refleja la estadística de las personas que conforman la población total mexicana actual

> (E4, P2): No mucho ya que este programa es únicamente ilustrativo, no podemos saber a ciencia cierta si de verdad la coca cola es mayormente consumida en nuestra población

Es decir, los estudiantes tratan a la DMS como si estuviera compuesta por muestras reales, pero reconocen que son simuladas ¿cómo resuelven este conflicto? Nosotros proponemos la hipótesis de que los estudiantes creen que la actividad que hacen en clase es sólo un simulacro de lo que se debe hacer en el campo cuando se tienen que enfrentar a un problema real. De esta manera, no hay conflicto. Una consecuencia de esta concepción ya había sido detectada por Hogdson y Burke (2000) quienes observan que después de un curso basado en simulación los estudiantes creen que para hacer una inferencia tienen que obtener muchas muestras (reales).

Cambio de perspectiva: *márgenes de tolerancia*

Al analizar las soluciones al problema 1, el instructor se percata de que los estudiantes toman a las muestras simuladas como muestras reales dándose cuenta de que los estudiantes creen que pueden estimar la proporción de la población analizando la DMS. Para contrarrestar esta idea les advierte que la DMS es sólo hipotética y no contiene información de la población. Les enseña que la DMS sirve para identificar muestras típicas (usuales) y muestras atípicas (inusuales). Esto propicia que los estudiantes busquen soluciones alternativas.

Para el segundo problema, y gracias a la intervención del instructor, ocho parejas de estudiantes (Tabla 1) comienzan a enfocar el problema desde una perspectiva diferente; buscan darle sentido a la idea, propuesta por el instructor, de que pueden identificar con la DMS conjuntos de valores del estadístico (proporciones) que ocurren con mucha frecuencia y otros que ocurren con poca frecuencia. Por ejemplo, la pareja E2 menciona que dividen la distribución en tres partes:

> (E2, P2): …se tomaron tres rangos de referencia para garantizar los resultados y de esta manera decir que la hipótesis puede llegar a ser correcta […] se realizaron las 500 encuestas (simuladas) para graficar los resultados; una vez obtenida la gráfica se tomaron los siguientes rangos de las personas que prefirieron Coca-Cola: 1) menor a 50%: 0–79 personas. 2) Igual a 50%: 80-100 personas. 3) mayor a 50%: 101-180 personas.

Esta pareja considera que los resultados dentro del rango de 80 a 100 son típicos y fuera de él atípicos. Esta manera de hacer el análisis es una variante de la intuición: "La proporción de la muestra es aproximada a la proporción de la población", pero ahora utilizan la DMS para representar lo que entienden por "aproximado". Seis de estos estudiantes realizan este procedimiento, pero no lo utilizan para valorar la muestra-dato (los otros dos si lo hacen), sino que lo combinan con un método de comparación del número de muestras en el rango de la izquierda con el número de muestras en el rango de la derecha. Estos seis estudiantes todavía actúan como en un simulacro y no abandonan la idea de que la DMS está formada por muestras reales.

Cálculo del *p-valor*

Al corregir las respuestas al problema 2, el instructor subraya la importancia de observar en la DMS la frecuencia del conjunto de proporciones con igual proporción o con proporción más extrema que la proporción de la muestra-dato (es decir, una estimación del p-valor) y la regla de que si el p-valor es menor o igual al 5% entonces la muestra-dato se considera atípica, en caso de que el p-valor sea mayor al 5% entonces la muestra dato se considera común o típica. Después de estas instrucciones se les administra el problema 3. En los problemas 3 y 4 hubo 8 y 7 parejas de estudiantes respectivamente (Tabla 1) que siguieron el procedimiento utilizando la noción de *p-valor*. Por ejemplo:

(E9, P3): La máquina no debe repararse, ya que el resultado no es atípico [inusual], es decir no es más del 10% como lo presenta el problema. Tomando la regla del 5% nos debe salir un resultado menor o igual a 5% de las piezas, pero en esta ocasión nos salió 15.4% (77 experimentos de 500) y esto refleja que es un resultado típico [usual] de las muestras que nosotros tomamos, por lo tanto, si nos hubiese salido un porcentaje menor o igual a 5% la máquina debería repararse. [Los corchetes fueron agregados por nosotros para mayor claridad]

(E2, P4): Se tomó un rango de 74 hacia abajo para comprobar si el resultado es atípico o no; esto se hizo realizando una simulación de 500 pruebas de las cuales 42 de estas fueron de 74 personas o menos y al sacar el porcentaje de esto nos dio 8.4%, entonces usando la teoría: "menos del 5% es atípico y más del 5% es normal", podemos concluir que 8.4% es mayor al 5% por lo que entra en el 80% de eficacia del medicamento hacia la población debido a que el resultado es normal.

Tabla 1. Frecuencias de cada tipo de procedimiento por problema

Procedimiento	Problema				
	1	2	3	4	Total
1. Doble mayoría	12	2	1	0	15
2. Márgenes de tolerancia	0	8	3	4	15
3. Cálculo del p-valor	0	0	8	7	15

DISCUSIÓN Y CONCLUSIÓN

Sugerimos un micro desarrollo de la concepción de la DMS de los estudiantes en tres fases. En la primera, ven a la DMS como una representación de muestras reales de la población. La atención de doce parejas de estudiantes se centra en la cardinalidad de los puntos muestrales a izquierda o derecha de la hipótesis. El procedimiento no incluye a la muestra-dato, pues, para ellos, ésta no es más que una más de todas las que organiza la DMS. No ven una contradicción en la evidencia de que las muestras son simuladas y su creencia de que son reales debido al fenómeno de simulacro. Un fenómeno similar ha sido mencionado por Hodgson y Burke (2000) y Case y Jacobbe (2018).

En la segunda fase, ocho parejas de los estudiantes (Tabla 1) le dan una estructura tripartita a la DMS mediante la cual prefiguran un patrón de variabilidad; piensan que las muestras en la sección central de la distribución son abundantes, mientras que las muestras en los extremos son pocas comparadas con las centrales. Este procedimiento se relaciona con la concepción multiplicativa de muestra de Saldanha y Thompson (2002). No obstante, al hacer la inferencia seis parejas aplican un método similar al de doble mayoría y sólo dos utilizan la distribución para evaluar la muestra dato. El instructor refuta el método de los estudiantes y aprovecha para dirigir su atención hacia las muestras inusuales de un extremo de la DMS enfatizando en el p-valor y ofreciendo el criterio de un nivel de significación de 5%.

En la tercera fase, ocho parejas utilizan el p-valor y el nivel de significación para rechazar o no la hipótesis. En los hechos, estas parejas actúan como si hubieran abandonado su concepción de que la DMS está formada por muestras reales y la conciben como un instrumento para evaluar la inusualidad de la muestra, que prefigura un proceso de inter–comparación (Stigler, 2017); es decir, aunque no lo sabemos con certeza, interpretamos que conciben la DMS como un patrón de variación con la métrica del 5% para evaluar la muestra-dato. Ahora los estudiantes están en la frontera de concebir a la DMS a un nivel matemático, como un objeto independiente del contexto y en mejores condiciones de entender la distribución muestral y su función en la inferencia.

Referencias

Aquilonius, B. C. y Brenner, M. E. (2015). Students' reasoning about p-values. *Statistics Education Research Journal, 14*(2),7-27.

Bakker, A. y Gravemeijer, K. P. E. (2004). Learning to reason about distribution. En D. Ben-Zvi, y J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 147–168*).* Kluwer.

Biehler, R., Ben-Zvi, D., Bakker, A. y Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. En M. A. (Ken) Clements et al. (Eds.), *Third International Handbook of Mathematics Education* (pp. 643-689). Springer.

Case K. y Jacobbe, T. (2018). A framework to characterize student difficulties in learning inference from a simulation-based approach. *Statistics Education Research Journal, 17*(2), 9-29.

Chance, B., del Mas, R. y Garfield, J. (2004). Reasoning about sampling distributions. En: D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 295-323). Springer. https://doi.org/10.1007/1-4020-2278-6_13

García-Ríos, V. N. (2017). *Diseño de una trayectoria hipotética de aprendizaje para la introducción y desarrollo del razonamiento sobre el contraste de hipótesis en el nivel medio superior*. [Tesis de doctorado no publicada]. Departamento de Matemática Educativa.

Heitele, D. (1975). An epistemological view on fundamental stochastic ideas. *Educational Studies in Mathematics, 6* (2), 187-205.

Hodgson, T. y Burke, M. (2000). On simulation and the teaching of statistics. *Teaching Statistics, 22*(3), 91-96.

Lane-Getaz, S. J. (2017). Is the p-value really dead? Assessing inference learning outcomes for social science students in an introductory statistics course. *Statistics Education Research Journal, 16*(1), 357-399.

Lecoutre, B. y Poitevineau, J. (2014). *The Significance Test Controversy. The Fiducial Bayesian Controversy*. Springer.

López-Martín, M., Batanero, C., y Gea, M. (2018). La faceta cognitiva en el conocimiento de futuros profesores sobre el contraste de hipótesis. En L. Rodríguez-Muñiz, L. Muñíz-Rodríguez, A. Aguilar-González, P. Alonso, F. Garcí-García, y A. Bruno (Eds.), *Investigación en Educación Matemática XXII* (pp. 300-309). Gijón: SEIEM.

Moore, D. S. (2000). *Estadística Aplicada Básica.* Antonio Bosch Editor.

Pratt, D. y Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistical Education Research Journal, 7*(2), 3-4.

Ruiz-Higueras, L. (1994). *Concepciones de los alumnos de Secundaria sobre la noción de función: Análisis epistemológico y didáctico*. [Tesis de Doctorado]. Universidad de Granada.

Saldanha, L. y Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics, 51* (3), 257–270.

Saldanha, L. y Thompson, P (2014). Conceptual issues in understanding the inner logic of statistical inference: Insights from two teaching experiments. *The Journal of Mathematical Behavior 35*(1), 1-30.

Stigler, S. M. (2017). *Los Siete Pilares de la Sabiduría Estadística.* Grano de Sal.

Van Dijke-Droogers, M., Drijvers, P. y Bakker, A. (2019). Repeated sampling with a black box to make informal statistical inference accessible. *Mathematical Thinking and Learning, 22*(2), 116-138. https://doi.org/10.1080/10986065.2019.1617025

Watson, J. y Moritz, J. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning, 2*(1-2), 11-15. https://doi.org/10.1207/S15327833MTL0202_2

Zieffler, A., Garfield, J., DelMas, R. y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal, 7*(2), 40-58.