

# Patrones de razonamiento sobre pruebas de significación de estudiantes de bachillerato

## Patterns of reasoning on significance tests of high school students

Ernesto Sánchez Sánchez, Eleazar Silvestre Castro y Víctor N. García Ríos

Cinvestav-IPN, México

### Resumen

La presente investigación tiene el objetivo de explorar los patrones de razonamiento que exhiben estudiantes de bachillerato cuando enfrentan problemas de pruebas de significación con el método de simulación de muestreo repetido por computadora. Dieciocho parejas de estudiantes de 10° grado (16-17 años de edad) que no habían llevado el curso de probabilidad y estadística participaron en un taller de 6 sesiones, en la que abordaron 4 problemas de prueba de significación, generando en cada caso la distribución muestral correspondiente mediante una técnica de simulación. Como resultado, los estudiantes aprendieron la técnica y utilizaron la distribución muestral para tomar una decisión correcta acerca de rechazar o no la hipótesis. No obstante, varias parejas de estudiantes muestran, en sus respuestas, evidencia de concepciones erróneas acerca de la prueba mediante el método de muestreo repetido.

**Palabras clave:** Prueba de significación, razonamiento, simulación, muestreo repetido, distribución muestral.

### Abstract

This research has the objective of exploring the reasoning patterns exhibited by high school students when they face problems of significance tests with the repeated sampling simulation method by computer. Eighteen pairs of 10th grade students (16-17 years-old) who had not taken the probability and statistics course, participated in a 6-session workshop, in which they addressed 4 significance test problems, generating in each case the corresponding sample distribution by means of a simulation technique. As a result, the students learned the technique and used the sample distribution to make a correct decision about whether or not to reject the hypothesis. However, several pairs of students show, in their answers, evidence of misconceptions about the significance test using the repeated sampling method.

**Keywords:** Significance test, reasoning, simulation, repeated sampling, sampling distribution.

## 1. Introducción

El *contraste de hipótesis y las pruebas de significación* son conceptos fundamentales de la inferencia y, al menos el contraste de hipótesis, se incluye en todos los primeros cursos universitarios de estadística. Al igual que otros temas de estadística y de matemáticas, existe múltiple evidencia que señala que muchos estudiantes tienen serias dificultades, cometen errores sistemáticos y manifiestan concepciones erróneas cuando realizan tareas sobre el tema (Castro-Sotos, Vanhoof, Van den Noorgate y Onghena, 2007, Vallecillos y Batanero, 1997). Algunos autores explican que la causa de esta situación es una combinación de diferentes razones, tales como la gran complejidad epistemológica del concepto (Batanero, 2000) y la prevalencia de prácticas tradicionalistas poco eficaces (Harradine, Batanero y Rossman, 2011; Liu y Thompson, 2009). Recientemente se ha propuesto en la investigación en educación estadística la exploración de acercamientos informales para la enseñanza de la inferencia con el

propósito de que los estudiantes construyan las bases para una progresiva formalización de sus conocimientos estadísticos (por ejemplo, Batanero y Díaz, 2015; Zieffler, Garfield, delMas y Reading, 2008) y evitar que solamente se promueva el aprendizaje de procedimientos y algoritmos. Varias propuestas que se derivan de esta tendencia, en lo que concierne a la inferencia estadística, se apoyan fuertemente en el uso de simulaciones para construir distribuciones muestrales empíricas (Chance, delMas y Garfield, 2004; Inzunza, 2015; Lipson, 2002), las cuales, sin embargo, se han llevado a cabo con estudiantes de nivel universitario. La cada vez más frecuente disponibilidad de recursos de tecnología digital en las escuelas y de software educativo para las diferentes áreas de la matemática, implica la necesidad de contar con evidencia que sugiera formas particulares y efectivas de utilizar la tecnología en el aula para desarrollar el razonamiento estadístico de los estudiantes. En particular, se abriga la esperanza de que el razonamiento sobre contraste de hipótesis y los conceptos relacionados pueden desarrollarse con un enfoque informal en niveles previos al universitario.

En particular, la instrucción sobre la relación entre muestreo y estimación parece que puede ser favorecida si se lleva a cabo en ambientes ricos en tecnología (Insunza, 2015; Silvestre y Sánchez, 2017), es por esto que es crucial para la investigación documentar las posibilidades reales del uso de la tecnología para el desarrollo del razonamiento de los estudiantes de bachillerato respecto a las distribuciones muestrales, las pruebas de significación y el contraste de hipótesis. En este sentido, nos interesan las conexiones e inferencias que hacen los estudiantes cuando se les enseña la técnica de simulación de muestreo repetido para realizar una prueba de significación.

Como parte de la instrucción, se enseña a los estudiantes a realizar simulaciones de muestreo repetido con el software Fathom y, en consecuencia, tienen la posibilidad de generar muchos valores de un estadístico que pueden organizar en una distribución muestral empírica. Se les sugiere utilizar la distribución para estimar probabilidades pertinentes a la prueba, como el  $p$ -valor o determinar la zona crítica. Bajo la lógica de las pruebas de significación de Fisher, con dicha estimación o con base en la zona delimitada, los estudiantes pueden tomar una decisión racional y rechazar, o no, la hipótesis. Sin embargo, surgen preguntas como: ¿realmente toman la decisión racional?, ¿entienden lo que están haciendo?, ¿qué es lo que entienden? Estas inquietudes llevan a formular la siguiente pregunta de investigación:

*¿Qué patrones de razonamiento muestran los estudiantes al enfrentar problemas de pruebas de significación cuando se les enseña la técnica de simulación de muestro repetido para resolverlos?*

Nos interesa percibir y caracterizar los patrones de razonamiento sobre pruebas de significación que emerjan durante el desarrollo de las actividades y dilucidar la influencia que en dichos patrones tienen los rasgos del diseño de la actividad. En particular, conviene destacar el papel del recurso tecnológico Fathom como un elemento crucial del diseño.

## **2. Marco de referencia**

El *contenido estadístico* del presente estudio se centra en un concepto de Fisher conocido como *prueba de significación*. Ésta consiste en determinar si una muestra contiene suficiente evidencia para rechazar una hipótesis conservadora acerca de la población: “Fischer vio el objetivo de una prueba de significación como la confrontación de una hipótesis nula con las observaciones, y para él, un  $p$ -valor indica la

fuerza de la evidencia contra la hipótesis” (Batanero, 2000, p. 87). En efecto, si un evento es muy raro bajo el supuesto de que la hipótesis nula es verdadera, y al hacer una observación se presenta ese evento, es más razonable poner en duda la veracidad de la hipótesis nula que aceptar que se ha presentado un evento extraordinario. Este razonamiento resulta muy útil en situaciones experimentales en donde se quiere determinar si la manipulación de una variable produce un cambio significativo. Suponemos que el razonamiento sobre pruebas de significación es más asequible para los estudiantes que el del contraste de hipótesis, sobre todo si en la instrucción se hace uso de un recurso tecnológico como Fathom.

El objeto de estudio es el *razonamiento* de los estudiantes, por lo que se busca caracterizar los patrones de razonamiento observables en sus respuestas y comentarios a las tareas que involucran la prueba de significación. Dentro del análisis, se formulan conjeturas sobre si tal o cual patrón fue producido por algunos de los elementos del diseño. Se entiende por *razonamiento* el proceso de obtener o producir conclusiones sobre la base de evidencia y/o de proposiciones supuestas o ya establecidas (Shaughnessy, 2008). Gran parte de las actividades intelectuales o cognitivas tienen, en última instancia, el propósito de formular y establecer conclusiones verdaderas (i.e. proposiciones) y, por tanto, implican la generación de razonamiento. A su vez, éste se liga a la *Producción de sentido*, que es la actividad de conectar de diversas maneras los enunciados y evidencias nuevos con conocimientos ya establecidos, de modo que se produce la sensación de comprensión a diferentes niveles de la situación. Un enfoque informal a las pruebas de significación puede permitir la producción de sentido sobre la inferencia estadística.

La elaboración de actividades de clase para lograr el aprendizaje de conceptos estadísticos implica cierta *posición sobre la enseñanza*, toda vez que no hay una normatividad cuya eficacia esté garantizada. Asumimos, entonces, tres recomendaciones actuales sobre la didáctica de la matemática: organizar la clase en torno a la actividad de resolución de problemas, promover el trabajo en equipo para propiciar la discusión y el desarrollo del discurso matemático y estadístico, y el uso de la tecnología computacional como un recurso mediador entre los conceptos y la comprensión. Se considera que es posible favorecer el desarrollo del razonamiento y la producción de sentido matemático (y estadístico) de los estudiantes cuando estos se comprometen en la solución de problemas matemáticos (Francisco y Maher, 2005). Así, aquellos procesos se pueden disparar por el afán de alcanzar un objetivo (resultado matemático) cuyo camino para lograrlo no se conoce de antemano. El trabajo en equipo permite que los estudiantes den forma a sus razonamientos mediante el intercambio y el contraste de ideas, y el uso de tecnología les permite construir en forma concreta conceptos abstractos, como lo es el de distribución muestral.

El trabajo que aquí se presenta es una investigación de diseño, también llamado experimento de diseño (Cobb, Confrey, Disessa, Lehrer y Schauble 2003), cuyo propósito es formular hipótesis y entender las relaciones entre el diseño y el aprendizaje; en nuestro caso, consiste en comprender con más precisión las relaciones entre el desarrollo del razonamiento estadístico de los estudiantes y la creación de un ambiente para que adquieran conocimientos estadísticos y mejoren su razonamiento inferencial. Las actividades, con características que reflejan nuestra posición sobre la enseñanza, incluyen hojas de trabajo donde los estudiantes escriben sus respuestas y justificaciones a los problemas, mismas que recoge el profesor y que constituyen nuestra fuente primaria de datos; el análisis de éstos se realiza siguiendo las

recomendaciones de la Teoría fundamentada (Glaser y Strauss, 1967/2008). Esta metodología recomienda comenzar con un procedimiento de codificación libre a partir de una comparación entre las respuestas de los estudiantes. A partir de lo anterior se encuentran categorías, que en nuestro caso son patrones de razonamiento de los estudiantes o características relevantes de su razonamiento, como falsas concepciones o sesgos.

### 3. Método

*Participantes.* Los resultados que aquí se presentan corresponden a los datos que se obtuvieron con un grupo que estuvo formado por 36 estudiantes de tercer semestre de bachillerato (16-17 años de edad) agrupados en 18 parejas, con una computadora por pareja. Estos estudiantes aún no habían cursado la asignatura de probabilidad y estadística que se suele impartir en el ciclo escolar del bachillerato. Se llevaron a cabo seis sesiones en las que se abordaron 4 problemas de pruebas de significación; los problemas fueron elegidos de libros de texto de estadística y se les hicieron modificaciones para adaptarlos a las características del grupo. En las actividades los estudiantes llenaron hojas de trabajo que constituyeron la principal fuente de datos.

*Instrumentos y datos.* El siguiente es uno de los cuatro problemas que abordaron los estudiantes; las evidencias que se presentan corresponden a las respuestas a este problema, pero cabe aclarar que los datos fueron las respuestas a los cuatro problemas y los patrones que se percibieron corresponden a un análisis de las respuestas a todos ellos.

**Ejemplo A / Actividad 1:** Una propaganda de la Coca Cola asegura que la mayoría (más del 50%) de la población que consume bebidas de cola en México prefiere su refresco sobre los que prefieren Pepsi. Para comprobarlo, se realizó un experimento en donde se eligieron a 60 personas aleatoriamente de los consumidores de refresco de cola, se les dieron dos vasos de refresco no etiquetados (uno con Coca y otro con Pepsi) para luego señalar cuál les gustó más. De los 60 participantes, 35 prefirieron Coca Cola. ¿Consideras que la hipótesis “más de 50% de la población prefiere Coca Cola sobre Pepsi es correcta? – [ $H_0: P \leq .5$  es aceptada al nivel de  $\alpha = .05$ ]

El profesor recogía las hojas de trabajo para acumular datos para su posterior análisis. En el primer ciclo, los problemas y las simulaciones de distribuciones muestrales en Fathom eran presentados y discutidos con los estudiantes durante la primera hora. En la segunda hora se les permitía trabajar libremente en parejas para resolver el problema y hacer un reporte de sus conclusiones (respuestas) en la computadora; el profesor sólo intervenía para resolver dudas atendiendo a cada equipo individualmente: éste nunca daba la solución ni guiaba a los equipos para llegar a ella.

Para el análisis de los datos se consideraron los principios y procedimientos de la teoría fundamentada (Birks y Mills, 2011; Glaser y Strauss, 1967/2008), que es una metodología general de investigación en ciencias sociales (Holton, 2008) cuyo objetivo es elaborar teorías locales que emerjan de los datos y no de deducciones lógicas o de otras teorías previamente establecidas. Los datos que se analizaron son las respuestas y conclusiones que los estudiantes escribieron en sus hojas de trabajo. Estas se transcribieron en archivos electrónicos para poder manipularlas, y para cada pregunta, éstas fueron comparadas entre sí para codificarlas. Mediante este proceso se determinaron patrones generales que son los que se exponen a continuación; cabe mencionar que los resultados y observaciones se refieren a patrones generales que aparecieron con cierta frecuencia y que se consideraron los más representativos de la

micro-cultura de la clase; se han omitido variaciones particulares que son intrínsecas a la complejidad de los razonamientos de los estudiantes.

#### 4. Resultados

Los resultados consisten en patrones de respuesta o de razonamiento que se presentaron en las soluciones de, al menos, dos parejas de estudiantes. No hemos brindado atención a las frecuencias con las que se presentan los diferentes tipos de respuesta o de razonamiento, sino sólo en la constatación de que se presentase en dos o más equipos, independientemente uno del otro. Una clasificación de las respuestas y sus frecuencias se puede encontrar en García y Sánchez (2017). Hemos organizado dichos patrones en tres momentos o etapas del proceso de llevar a cabo una prueba de significación con ayuda de tecnología, además, una consideración adicional. A continuación se describe de manera general lo que se esperaba que hicieran los estudiantes en cada una de ellas.

- *Hipótesis y simulación.* Los estudiantes leen y analizan el problema, establecen la hipótesis nula y el tamaño de la muestra y generan la distribución muestral correspondiente en Fathom.
- *Uso de la distribución muestral.* Utilizan la distribución muestral para estimar probabilidades con relación al estadístico; por ejemplo, el *p-valor*.
- *Formulación de la conclusión.* Con base en los resultados de las estimaciones formulan una conclusión acerca de la hipótesis nula (se rechaza o no).

Como un complemento a las etapas del proceso, es conveniente incluir:

- *Consideración de la incertidumbre.* Saber que existe la posibilidad de error, es decir, rechazar una hipótesis verdadera o de no rechazar una hipótesis falsa.

##### 4.1. Hipótesis y simulación

Cuando los estudiantes se enfrentan por primera vez a un problema de prueba de significación, inicialmente suelen seguir un procedimiento que revela una *indiferencia a la variabilidad muestral*. El procedimiento consiste en calcular la proporción en la muestra y simplemente compararla con la hipótesis de la población que se quiere afirmar. El criterio para determinar si una hipótesis sobre la población es verdadera consiste en verificar que se presenta la misma proporción en la muestra. Por ejemplo, una pareja de estudiantes argumenta a favor de la hipótesis “*más de 50% de la población prefiere Coca Cola sobre Pepsi*”:

“La propaganda está en lo correcto al poder presumir que la mayoría de la población prefiere su refresco en lugar de Pepsi ya que mencionan que “LA MAYORÍA de la población prefiere su refresco”, recalando esto podemos entender que como dicen, hacen referencia a más de la mitad de ésta, es decir a partir de un 51% de la población ya podemos comprender que es la MAYORÍA. Y debido a que los resultados del experimento arrojaron que 35 personas de 60 prefieren el refresco Coca Cola, lo que es igual a un 59% del total ( $0.59 \times 60 = 35.4$ ), podemos concluir que no erran en lo que presumen ya que están en todo lo correcto...”

Todas las parejas de estudiantes dan este tipo de respuesta al enfrentarse por primera vez al problema, pero cuando diseñan y realizan una simulación física, la mayoría percibe que el valor del estadístico cambia con la muestra y modifica su opinión.

*Elección de una hipótesis inapropiada.* Cuando los estudiantes entienden que pueden avanzar en la prueba simulando un muestreo repetido, se enfrentan al problema de determinar la hipótesis a partir del contexto del problema. Se nota una tendencia a determinarla a partir de los datos explícitos del problema sin poner en juego un criterio

adecuado para determinar la hipótesis nula. De esta manera, suelen elegir hipótesis inapropiadas. Por ejemplo, el siguiente estudiante genera distribuciones comenzando con 0.54, pues interpreta que en el problema se quiere saber si es más preferida la coca:

“Al momento de utilizar el programa para realizar las encuestas utilizando una proporción de 0.54 podemos notar que efectivamente la Coca es más consumida que la Pepsi”.

Este razonamiento circular revela que, aunque este estudiante puede generar distribuciones muestrales, aún no lo hace con un propósito claro. En otros casos se toma como hipótesis nula la proporción de la muestra, es decir, el estadístico. Posiblemente, la dificultad para establecer la hipótesis se ve agravada por la creencia errónea de que una prueba de significación es para ofrecer evidencia a favor de una hipótesis y no en contra de la hipótesis nula. En un sentido diferente al normativo, creerían que las pruebas de significación son para confirmar o consolidar hipótesis y no para refutarlas.

Vallecillos (1996) encontró entre los errores en el contraste de hipótesis la “falta de apreciación de la variabilidad del estadístico” (p. 171) y la “confusión de hipótesis nula y alternativa” (p. 169), los cuales son análogas a los patrones de respuesta que hemos descrito arriba. En la investigación presente, la detección de las dificultades se combina con la realización de actividades por parte del estudiante que le permitan desarrollar un razonamiento normativo y, por tanto, superar tales dificultades. En el desarrollo de las actividades, sobre todo las de simulación, se observó que los estudiantes superan fácilmente la indiferencia a la variabilidad, no así la dificultad relacionada con el establecimiento correcto de la hipótesis nula. En consecuencia, es necesario trabajar más en el diseño de las actividades para encontrar formas que les permitan desarrollar la habilidad de identificar la hipótesis nula; en especial, propiciar la reflexión acerca de que a menudo esta información no se encuentra explícitamente en el enunciado del problema.

#### 4.2. Uso de la distribución muestral

Los estudiantes aprenden con facilidad a generar distribuciones muestrales con el apoyo del software; no obstante, esto no significa que se apropien del concepto de distribución muestral, pues varios estudiantes, una vez generada, no la interpretan adecuadamente ni derivan de ella las consecuencias necesarias para la prueba de significación. En la primera ocasión que los estudiantes obtienen la distribución muestral del problema, ninguno de ellos estima el  $p$ -valor o determina la zona crítica con base en ella, en su lugar, hacen interpretaciones como los dos ejemplos que siguen.

*Localización en la “zona de mayoría”.* Dos parejas de estudiantes obtiene la distribución muestral de proporciones para muestras de tamaño 60 y  $p = \frac{1}{2}$ , dividen la distribución muestral en dos zonas, una menor que 30 (teniendo en cuenta que  $60 \times 0.5 = 30$ ) y la otra mayor. Cuentan entonces las frecuencias de cada zona y con base en esto toman la decisión. Observen el siguiente ejemplo:

“Debido a que en la simulación se muestran que en 285 encuestas de 500 se mostró un gusto por Coca Cola en más de 50% de 60 personas entrevistadas en cada una de las encuestas. En cambio en 215 encuesta realizadas se mostró un gusto en menos del 50% de 60 personas” (ver Figura 1).

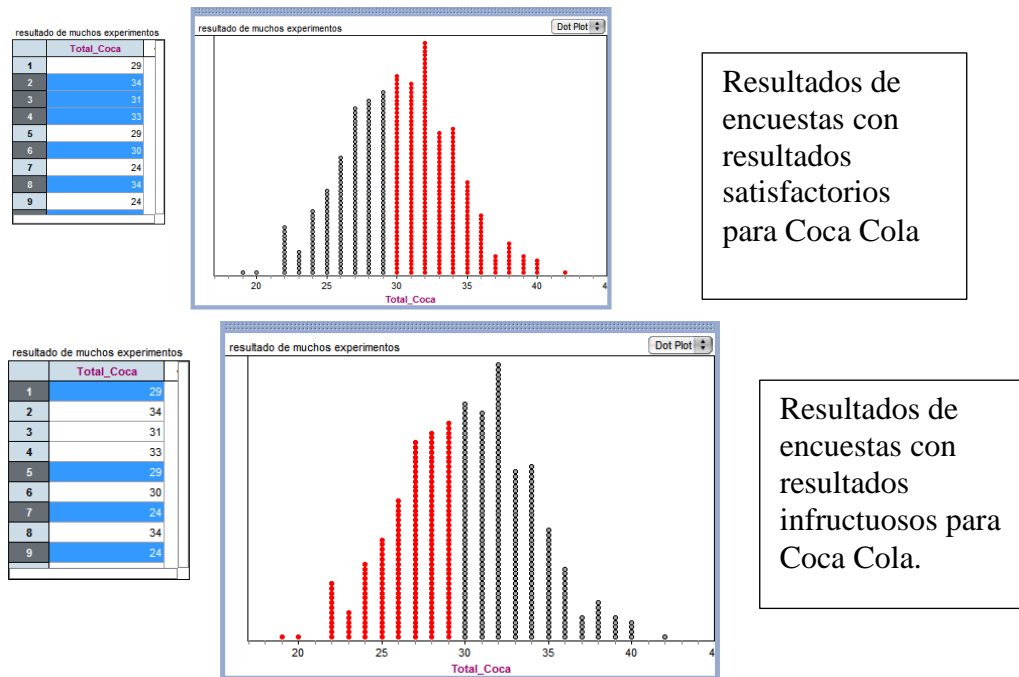


Figura 1. Gráficas de “Localización en la zona de mayoría (minoría)”

*Moda de la distribución muestral.* Otras parejas observan la moda y de acuerdo a su posición en la distribución rechaza o no la hipótesis; el criterio es “si la moda está por debajo de 30 se considera que los que prefieren coca son minoría, si está por arriba entonces los que prefieren coca es mayoría”. Una pareja que siguió este razonamiento lo expresó de la siguiente manera:

“...vimos que nuestro valor más grande en una encuesta fue de 28 personas de 60 que les gusto más Coca-Cola, entonces a partir de esto vemos que dentro de esa muestra a menos del 50% les gusta Coca-Cola, pero estamos considerando exactamente al 50% de la población total, entonces por eso concluimos que la hipótesis es falsa”.

La moda de la distribución muestral en la que se basó la pareja que dio la anterior respuesta fue 28 como se muestra en la Figura 2.

Pareciera que los estudiantes que utilizan los criterios de “la mayoría” o de “la moda” toman a la distribución muestral como si reprodujeran las características de la distribución poblacional. Esas respuestas reflejan una dificultad más general que consiste en no percibir que una distribución muestral es un patrón de variación independiente del contexto concreto en el que se formula el problema; esto se refleja al expresar que “en 285 encuestas de 500 se mostró un gusto por coca cola” o “28 personas de 60 que les gustó más coca cola”, es decir, los estudiantes interpretan los puntos de la distribución como “encuestas” o “personas”, cuando en realidad la distribución muestral sólo documenta variabilidad numérica, abstracta de objetos y de personas. El diseño puede influir en la persistencia de esta dificultad en la medida en que para cada problema se genera una distribución muestral, a diferencia del método clásico que orilla a ver el carácter abstracto de la distribución pues se encuentra en tablas fijas y pre-elaboradas.

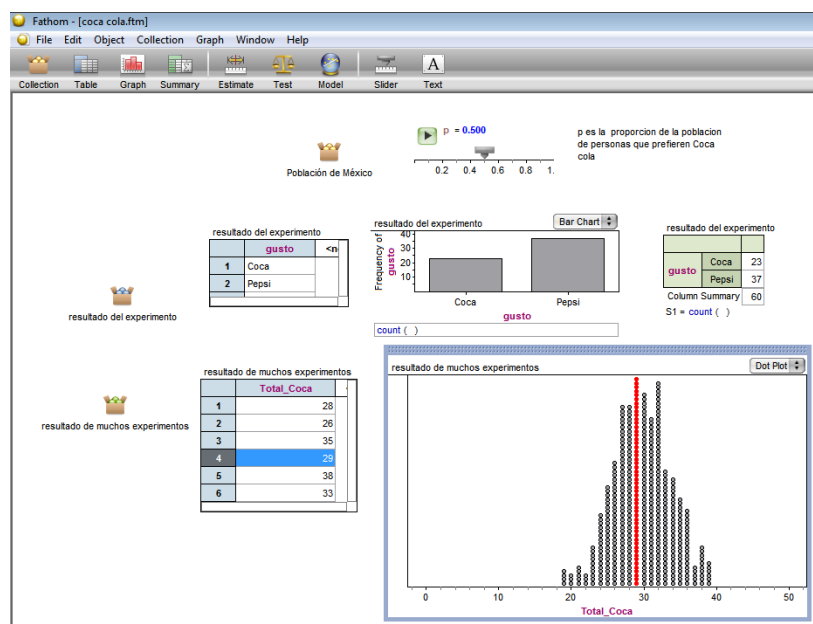


Figura 2. Pantalla de Fathom: Se generó una distribución muestral y se señala la moda.

### 4.3. Formulación de la conclusión

Debido a que las conclusiones de los estudiantes no se reducen a aceptar o no la hipótesis sino que las acompañan con una breve argumentación, a través de ellas se revelan, en parte, los patrones de razonamiento que usan los estudiantes al enfrentar las tareas. A partir de estas argumentaciones se han identificado dos patrones.

*Crecimiento de la muestra.* Los estudiantes cuyas respuestas recibieron este código consideran que el procedimiento de simulación de una muestra, de algún modo posiblemente misterioso, equivale a realizar una nueva encuesta en la población y, por tanto, agrega información relevante a la muestra dada. Es decir, ven que la acumulación de varias muestras se convierte en una gran muestra, que entonces se puede considerar como representativa. Veamos un fragmento de la explicación de una pareja en que se expresa la idea de que el proceso de simulación es para agrandar la muestra:

“...se realizaron 500 encuestas de 60 personas, por lo tanto se encuestaron a 30 000 personas. en la gráfica se determina un rango seleccionando un número de encuestas procurando que no sobrepasen la mitad de las 500 encuestas”.

*Multiplicación de muestras.* Otro tipo de respuestas siguen un razonamiento ligeramente distinto. Los estudiantes que dieron respuestas bajo este código, creen que la muestra ofrece nueva información acerca de la población, pero no piensan en una gran muestra, sino en analizar si cada muestra es favorable a la hipótesis, razonando que si la mayoría de las encuestas favorecen la hipótesis entonces la hipótesis es correcta. Veamos un fragmento en el que se expresa este razonamiento:

“Es verdadera la hipótesis, ya que tras simular en varias ocasiones la opinión de la población tomando muestras de hasta 100 personas y realizando hasta 500 encuestas; en la mayoría de los casos más del 50% de la muestra aceptó que prefiere la Coca-Cola por sobre la Pepsi. Siendo en la mayoría de los casos, los resultados de las personas que prefieren la Coca, cercano o mayor a 50%...”

Estos patrones, aparte de reforzar la observación de que los estudiantes creen que el proceso de simulación que da lugar a una distribución muestral, de algún modo, recoge información de la población a la que se refiere el problema, muestran que asumen el



proceso de prueba por simulación como una expresión de sus ideas del sentido común, es decir, que para conocer o aproximarse a la verdadera proporción se debe agrandar la muestra u obtener muchas de ellas.

#### 4.4. Reconocimiento de la incertidumbre

Conviene tener cuidado acerca de cómo los estudiantes consideran la incertidumbre en una prueba de significación. Por esta razón se les hizo la pregunta: ¿Será posible que tus conclusiones sean incorrectas aun haciendo bien todo el proceso?, ¿Por qué? Excepto una respuesta en blanco, el resto de los estudiantes respondieron afirmativamente, es decir, reconocieron un grado de incertidumbre en el resultado o conclusión. No obstante, lo hicieron por diferentes razones; veamos los patrones de respuesta más importantes que encontramos:

*Influencia de factores contextuales, sociales y/o humanos.* En este tipo de respuestas, se argumenta que las conclusiones podrían ser incorrectas debido a la complejidad del fenómeno, es decir, a los factores contextuales y humanos. Por ejemplo:

“Porque depende mucho de la región donde se realice la encuesta, ya que las diferentes costumbres y tradiciones hacen que las personas dependiendo del lugar donde se encuentren o donde hayan sido educadas tienen gustos diferentes” (Origen en el muestreo).

*La certeza se obtiene con un censo.* En otro patrón de respuestas, se argumenta que mientras la conclusión se haga con base en una (o muchas) muestras y no sobre la población total, habrá entonces posibilidades de cometer errores. Un caso así es el siguiente:

“El resultado jamás será absoluto a menos que vayamos con cada persona de la población mexicana para preguntarle cuál prefiere” (La idea de que sólo el censo ofrece una respuesta segura, pero no valora que hay probabilidad pequeña de error).

*Variabilidad muestral.* Este tipo de respuesta señala como causa de la incertidumbre el hecho de que con una muestra diferente se podrían obtener resultados distintos, por ejemplo:

“los resultados pueden llegar a variar y de ahí que se presente un margen de error”.

*Porque los datos son simulados y no reales.* Algunos no confían en los resultados, porque se dan cuenta que la simulación no puede proporcionar información real de la población. Un ejemplo de este tipo de respuesta es:

“El rango de error se puede derivar de los datos simulados; en cambio el rango de aceptación es el dato de la encuesta real, por lo cual con este dato tenemos una menor probabilidad de equivocarnos”.

En esta respuesta hay probablemente una clave oculta, consistente en que algunos estudiantes pueden entender el proceso de simulación como un simulacro, algo semejante a decirles “Así como aquí, en clase, para estimar el parámetro se obtienen 1000 muestras simuladas, en los problemas reales, también se tienen que sacar 1000 muestras para hacer la verdadera estimación”. Esta manera de entender es, por supuesto, contraria a la idea de la inferencia y no es lo que se quiere que aprendan los estudiantes.

## 5. Conclusión y discusión

Algunos con menos y otros con más dificultades y al cabo de las 6 sesiones, todos los estudiantes logran reproducir el procedimiento de muestreo repetido para llevar a cabo una prueba de significación. No obstante, hay evidencias que muestran que algunos

estudiantes entienden el proceso de una u otra forma que no necesariamente coincide con la lógica de la prueba. Destacamos tres falsas concepciones sobre una prueba de significación mediante muestreo repetido, que explican algunas de las respuestas observadas de los estudiantes.

*Interpretar la inferencia como “hacer crecer la muestra”.* Una idea sobre muestreo que suelen aceptar los estudiantes es que, si una muestra se hace más y más grande, el estimador en la muestra estará cada vez más cerca del parámetro de la población; de manera alternativa, si se obtienen cada vez más muestras y se promedian los valores del estadístico de las muestras, este promedio se acercará el valor del parámetro. Con esta base, es fácil interpretar (erróneamente) que una prueba de significación con muestreo repetido mediante simulaciones consiste precisamente en desarrollar la idea de que para obtener una buena estimación del parámetro hay que hacer crecer la muestra u obtener muchas muestras. Para resaltar lo equivocado de esta concepción, conviene recordar que la prueba de significación, en su versión normativa o formal, se realiza a partir de una sola muestra.

*Cada situación tiene su distribución muestral.* Esta concepción consiste en creer que la distribución muestral que se genera mediante muestreo repetido de la situación de un problema contiene, de algún modo misterioso, información específica de la situación y por tanto depende de ésta. Así, los estudiantes se refieren a los puntos de la distribución muestral como “encuestas” “personas”, etc. Bajo esta concepción no se entiende que la distribución muestral es un patrón de variación independiente del contenido concreto del problema, y que mantiene con éste relaciones puramente aritméticas. Es pues una manifestación del problema de entender la diferencia entre lo contextual y lo estadístico (Wild y Pfannkuch, 1999); resolverlo implicaría concebir a una distribución muestral como un objeto abstracto.

*El proceso de simulación como un simulacro.* En cierta manera esta concepción es contraria a la anterior, ya que se clarifica que la simulación no aporta información específica sobre la situación concreta, pero esto no es suficiente para que se entienda la lógica de la prueba mediante simulación de muestreo repetido, pues algunos estudiantes piensan que el proceso de simulación que se lleva a cabo en el aula es sólo un ejemplo “de pizarra”, no asumida como una prueba auténtica dado que ésta consistiría en obtener muestras de manera similar a la realizada en el software, pero llevada a cabo en la población real. Esta concepción es similar a la anterior en el sentido de que se cree que la distribución muestral debe estar ligada al contenido concreto, es decir, no se asume como un objeto abstracto.

*Consideración de la incertidumbre.* Aunque todos reconocen que el resultado de una prueba no asegura tomar la decisión correcta, no todos ven de igual manera el origen o causa de la incertidumbre. Los que atribuyen la incertidumbre a factores sociales no reconocen la hipótesis sobre el muestreo aleatorio, mediante el cual las características contextuales dejan de influir; por otro lado, tienen razón los que atribuyen la incertidumbre al hecho de que una muestra siempre correrá el peligro de perder información, sin embargo, no parecen valorar el hecho de que es posible medir la probabilidad de cometer un error. Finalmente, los que piensan que la incertidumbre proviene del hecho de que en las simulaciones no se considera la realidad, refleja, como hemos dicho, que no conciben a una distribución muestral como un patrón general de variación.

Para concluir, consideramos que la introducción de las pruebas de significación con ayuda de tecnología, en nuestro caso utilizando Fathom, podría ser una alternativa al

enfoque procedimental y memorístico en el que se suele introducir el tema de contraste de hipótesis en el bachillerato. A través de simulaciones físicas y computacionales y la posibilidad de que los estudiantes comenten y discutan los resultados, pueden comenzar a entender aspectos relacionados con la variabilidad muestral y el papel e importancia de la elección de las hipótesis; sin embargo, en este trabajo se muestra que el dominio del procedimiento para generar la distribución muestral y su uso para calcular probabilidades pertinentes a la prueba no asegura, por sí mismo, el desarrollo del razonamiento o la lógica que sostiene las pruebas de significación. Existe el riesgo de que asimilen dicho procedimiento a concepciones alejadas de la verdadera naturaleza de la inferencia.

Nos parece que algunos investigadores no perciben o descuidan enfatizar en el hecho de que el enfoque de muestreo repetido mediante simulaciones puede generar concepciones contrarias a la naturaleza de la inferencia. En el trabajo de van Dijke, Drijvers y Bakker (2018) se asegura lo siguiente:

Nuestra teoría preliminar es que el fortalecimiento del razonamiento probabilístico con la distribución de frecuencias de un experimento físico es un paso esencial hacia el razonamiento con distribución de muestreo. Los resultados mostraron que estos estudiantes podían imaginar y dibujar una distribución de frecuencia de muestras repetidas con la mayoría de los resultados de la muestra cerca de la proporción de la población y en la que apenas ocurren desviaciones fuertes. *En consecuencia, se ha fortalecido la comprensión de los estudiantes sobre el papel de la variación y la incertidumbre en la obtención de inferencias estadísticas informales.* (Las cursivas son nuestras) (van Dijke, Drijvers y Bakker, 2018)

Podemos estar de acuerdo con los autores si se interpreta el comentario en cursiva como la afirmación de que cuando los estudiantes están equipados con una buena comprensión de la variación muestral y son conscientes de la incertidumbre, están en mejores condiciones para entender y llevar a cabo inferencias informales, pero no estaríamos de acuerdo si se cree que por sí misma implica la comprensión de un acercamiento informal a la lógica de las pruebas de significación o del contraste de hipótesis, puesto que, como hemos visto, el aprendizaje de muestreo repetido y sus aplicación a las pruebas de significación puede generar concepciones contrarias.

**Agradecimiento:** Proyecto EDU2016-74848-P (AEI, FEDER).

## Referencias

- Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, 2(1-2), 75-98.
- Batanero C. y Díaz 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.), *Didáctica de la Estadística, Probabilidad y Combinatoria*, 2 (pp. 207-214). Granada: Departamento de Didáctica de la Matemática.
- Birks, M. y Mills, J. (2011). *Grounded theory: A practical guide*. California: Sage.
- Castro-Sotos, A. E., Vanhoof, S., Van den Noortgate, W., y Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, 2(2), 98-113.
- Chance, B., delMas, R. C., 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). Dordrecht, The Netherlands: Kluwer.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R. y Schauble, L. (2003). The role of design in educational research. *Educational Researcher*, 32(1), 9-13.
- van Dijke, M., Drijvers, P. y Bakker, A. (2018). Repeated sampling as a step towards informal statistical inference. En M. A. Sorto, A. White, y L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics*, Kyoto: IASE.
- Francisco, J. M., y Maher, C.A. (2005). Conditions for promoting reasoning in problem solving: Insights from a longitudinal study. *Journal of Mathematical Behavior*, 24, 361-372.
- García V. N. y Sanchez, E. (2017). Exploring high school students beginning reasoning about significance tests with technology. En E. Galindo, y J. Newton (Eds.), *Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, (pp. 1032-1039). Indianapolis, IN: Hoosier Association of Mathematics Teacher Educators.
- Glaser, B. G. y Strauss, A. L. (2008). *Discovery of grounded theory: Strategies for qualitative research*. New Brunswick: Aldine Transaction (trabajo original publicado en 1968).
- 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.
- Holton, J. A. (2008). Grounded theory as a general research methodology. *The Grounded Theory Review*, 7(2), 67-93.
- Inzunza, S. (2015). Dificultades en el desarrollo de una concepción estocástica de las distribuciones muestrales utilizando un ambiente computacional. En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria*, 2 (pp. 197-205). Granada: Departamento de Didáctica de la Matemática.
- Liu, Y. y Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies*, 4(2), 126-138.
- Lipson, K. (2002). The role of computer based technology in developing understanding of the concept of sampling distribution. En B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town.: International Statistical Institute.
- Silvestre, E. y Sánchez, E. (2017). High school students' first experiences with the Sampling Distribution: toward a distributive perspective of sampling and inference. En T. Dooley y G. Gueudet (Eds.), *Proceedings of the Tenth Congress of European Research in Mathematics Education (CERME-10)*, Dublín: ERME.
- Vallecillos, A. (1996). *Inferencia estadística y enseñanza: un análisis didáctico del contraste de hipótesis estadísticas*. Granada: Comares.
- Vallecillos, A. y Batanero, C. (1997). Conceptos activados en el contraste de hipótesis estadísticas y su comprensión por estudiantes universitario. *Recherches en Didactique des Mathématiques*, 17, 29-48.
- Zieffler, A., Garfield, J., delMas, R. y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistical Education Research Journal*, 7(2), 40- 58.