

# MEDICIÓN INFORMAL DEL P-VALOR MEDIANTE SIMULACIÓN

## Informal p-value through simulation

García, V. N. y Sánchez, E.

Departamento de Matemática Educativa, Cinvestav-IPN, México

### Resumen

Existe un creciente reconocimiento de la importancia de desarrollar el razonamiento inferencial informal (RII) antes de aprender los conceptos formalmente. No obstante todavía hay poca investigación sobre su desarrollo en el aula en el nivel bachillerato (15-18 años). En situaciones informales, la principal dificultad es medir el p-valor de un estadístico, debido que no se tiene la noción de distribución muestral de forma natural. En el presente estudio se analiza el razonamiento de estudiantes de bachillerato para medir el p-valor de un estadístico al usar una aplicación dinámica que crea una distribución muestral empírica mediante la simulación computarizada. Se encuentra que la mayoría de los estudiantes miden el p-valor adecuadamente con ayuda de la simulación, lo que representa un cambio significativo en el RII.

**Palabras clave:** Razonamiento inferencial informal, prueba de significación, distribución muestral, p-valor

### Abstract

There is growing recognition of the importance of developing the informal inferential reasoning (RII) before learning concepts formally. However there is still little research on its development in the classroom at the high school level (15-18 years). In informal situations, the main difficulty is to measure the p-value of a statistic because students don't have the sampling distribution concept naturally. Reasoning of high school students is analyzed while measuring the p-value using a dynamic applet that creates an empirical sampling distribution by computer simulation. It is found that most students measure the p-value properly when using simulation, which represents a significant change in the RII.

**Keywords:** Informal inferential reasoning, significance test, sampling distribution, p-value

## INTRODUCCIÓN

En la actualidad, las personas se ven en la necesidad de saber interpretar y comprender información sobre una gran diversidad de temas para tomar decisiones. Los medios de comunicación generalmente contienen información numérica, ya sea en forma de tablas, diagramas o gráficas. El estudio de la estadística proporciona a las personas las herramientas y las ideas para enfrentarse inteligentemente a la información numérica que emerge del mundo cotidiano (Garfield y Ben-Zvi, 2008). Dentro de la estadística, la herramienta principal es la inferencia estadística, pues capacita para leer, entender e interpretar de manera objetiva las conclusiones derivadas de los análisis de datos. La habilidad con el manejo de datos disponibles para observar más allá de ellos y hacer estimaciones acerca de un fenómeno o población desconocida es el corazón de la estadística (Makar, Bakker y Ben-Zvi, 2011).

De la literatura sobre la inferencia estadística formal (Castro-Sotos, Vanhoof, Van den Noorgate y Onghena, 2007; Garfield y Ben-Zvi, 2008) se concluye que la inferencia es un tema difícil de aprender, los resultados de muchos estudios empíricos informan que estudiantes, profesores e

incluso profesionales persistentemente cometan errores conceptuales. Una de las razones puede provenir del hecho de que el tratamiento de la inferencia en los cursos de estadística del bachillerato y primer curso universitario, se enfoca hacia el aprendizaje de procedimientos rutinarios (recetas) y/o la reproducción de la exposición formal y no ofrece a los estudiantes la oportunidad de discutir y entender las ideas fundamentales subyacentes en los procedimientos inferenciales antes de formalizarlas con los conceptos matemáticos correspondientes.

Esto ha motivado el interés por estudiar la Inferencia Estadística Informal (IEI) y el Razonamiento Inferencial Informal (RII) como lo confirma la publicación de un número especial de *Statistics Education Research Journal* (Pratt y Ainley, 2008) y *Mathematical Thinking and Learning* (Makar y Ben-Zvi, 2011), con los objetivos de: 1) descubrir y describir formas en que sea posible que los estudiantes desarrollen ideas fundamentales de la inferencia estadística sin utilizar el aparato matemático que las fundamenta, y 2) crear un repertorio de problemas y actividades que jueguen un papel de antecedente o sustrato en el aprendizaje de los estudiantes sobre el cual puedan construir los conocimientos formales de la inferencia estadística. Investigadores han visto a la tecnología, en particular las simulaciones como una opción idónea para llevar a cabo los objetivos anteriores. Sin embargo, se han realizado pocos estudios para el nivel bachillerato (15-18 años).

Importantes dificultades que suelen tener los estudiantes para hacer inferencias informales se relacionan con su carencia de la idea de distribución muestral y, por tanto, de la imposibilidad de su uso en la inferencia estadística (por ejemplo, Chance, delMas y Garfield, 2004). Esta idea es empleada para medir el p-valor del estadístico (estadístico significativo o no), algo difícil de hacer de forma natural por los estudiantes ya sea por falta de percepción de la variación (García y Sánchez, 2015) o por falta de consideración de los datos (García-Ríos, 2013). El presente estudio pretende mostrar cómo el diseño de una aplicación dinámica e interactiva que simula una distribución muestral empírica tiene el potencial para desarrollar informalmente esta idea fundamental. Se espera que los resultados sirvan para elaborar una estrategia de enseñanza en temas de inferencia estadística enmarcados en el RII. La aplicación está creada en Fathom (Finzer, 2014); un software dinámico y animado creado especialmente para la enseñanza y aprendizaje de la Estadística. Permite explorar libremente y creativamente ideas matemáticas y estadísticas. Diseñado para realizar tareas y procesos estadísticos como crear tablas, gráficas, obtener muestras, definir y calcular estadísticos y crear simulaciones animadas. Una característica importante es que un conjunto de datos se puede representar al mismo tiempo en una tabla y en una gráfica y al cambiar en alguna de ellas un dato, la otra representación se actualiza automáticamente, así como todas las funciones y procedimientos que hayan sido definidos con el conjunto inicial de datos.

El objetivo principal del estudio es analizar el razonamiento de estudiantes de bachillerato para medir el p-valor del estadístico de una muestra al usar una aplicación dinámica que crea una distribución muestral empírica mediante la simulación computarizada y determinar de qué forma contribuye la simulación a desarrollar el RII. Las preguntas de investigación que ayudaran a lograr el objetivo son: ¿Cuál es el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación que llevan a cabo estudiantes de bachillerato? ¿Qué errores y dificultades persisten en el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación?

## ANTECEDENTES

Varios trabajos en los últimos años aluden a los conceptos de IEI y RII; sin embargo, todavía no hay consenso sobre qué significan estos términos. En un intento de combinar distintas perspectivas, Zieffler, Garfield, delMas y Reading (2008) definen RII como “la forma en que los estudiantes usan sus conocimientos informales de estadística para crear argumentos basados en muestras observadas para sustentar las inferencias sobre la población desconocida” (p.44). Estos autores, también proponen un marco conceptual para caracterizar el RII y apoyar el desarrollo de tareas que permitan examinar el RII natural de los estudiantes, así como el desarrollo de tal razonamiento.

Las investigaciones sobre RII se pueden dividir en dos enfoques, a saber, una que contiene estudios sobre la naturaleza del RII y otra que incluye estudios que se refieren al desarrollo del RII. Se analizan trabajos con ambos enfoques, comenzando con los estudios de la naturaleza del RII.

En García-Ríos (2013) se informa que los estudiantes no pueden medir la significatividad del estadístico de la muestra adecuadamente, debido a dos posibles causas: a) Razonan de forma determinista al medir la significatividad del estadístico de la muestra, en el sentido de que el estadístico debe coincidir exactamente con el parámetro poblacional (personal). b) Comparan el estadístico con un modelo probabilístico inapropiado de la población (hipótesis nula), creado por el estudiante con base en sus conocimientos. Además, utilizan muchos prejuicios y creencias en sus argumentos y no reflejan grados de incertidumbre en sus inferencias. García y Sánchez (2014) han observado que la concepción de Fisher de las pruebas de significación es algo natural para los estudiantes, pues establecen una hipótesis nula (modelo personal de la población) para comparar la muestra y medir intuitivamente su significatividad (aunque de manera inapropiada). Consideran que es importante que los estudiantes trabajen con los datos y ubiquen en un segundo plano sus conocimientos informales personales sin descartarlos por completo. Se concluye que para desarrollar inferencias apropiadas antes de su formalización, es crucial que los estudiantes cuenten con un método informal para determinar un criterio numérico para saber cuándo rechazar o aceptar la hipótesis. La simulación es un recurso que suministra dicho método.

Con respecto a los estudios para desarrollar el RII, Rossman (2008) sugiere una introducción informal siguiendo los pasos siguientes: (a) Partir de una hipótesis dada acerca de los datos (b) Uso de la simulación para concluir que los datos observados son muy poco probables si el modelo fuera cierto (cálculo intuitivo de un p-valor), y (c) Rechazar la hipótesis inicial basado en el p-valor muy pequeño. Este proceso de razonamiento, parece natural para los estudiantes, y de hecho sigue la concepción de Fisher de pruebas de significación.

Pfannkuch (2005) considera que el enfoque de “re-muestreo” tiene el potencial de hacer una fuerte conexión entre la probabilidad y el análisis de datos, pues está más en sintonía con una conceptualización de un p-valor, y Fathom permite este enfoque. Hofmann, Maxara, Meyfarth y Prommel (2014) ven en Fathom un gran potencial didáctico para la enseñanza y aprendizaje de la estadística con simulaciones. Ellos proponen que para tener éxito en la enseñanza con Fathom es necesario diseñar tareas adecuadas, contar con fases de familiarización de la tarea, de presentación, de trabajo de los estudiantes y de discusión.

Lane y Pres (2006) muestran que incluso una simulación bien diseñada es poco probable que sea una herramienta de enseñanza eficaz a menos que la interacción de los estudiantes con ella se estructure cuidadosamente. Recomiendan evitar el uso de simulaciones que haga de los estudiantes observadores pasivos. Lipson, Kokonis, y Francis (2003) afirman que es necesario que en una actividad de simulación vaya acompañada de preguntas que garanticen que los estudiantes presten atención y entiendan todos los componentes de la pantalla, los que cambian y los que no durante el muestreo. Recomiendan alentar la formación del concepto de distribución muestral y facilitar la formación de enlaces de dicho concepto con los contrastes de hipótesis y la estimación.

## **MARCO CONCEPTUAL**

En este trabajo, se entiende por marco conceptual a una red de conceptos o categorías relacionados entre sí que en conjunto proporcionan una comprensión global de un fenómeno o fenómenos (Jabareen, 2009). Es posible construir un marco conceptual a partir de los datos del estudio mediante la metodología teoría fundamentada (Birks y Mills, 2011).

## **Pruebas de significación**

Dentro de la inferencia estadística hay dos concepciones sobre los contrastes de hipótesis: (a) las pruebas de significación, que fueron introducidas por Fisher y (b) los contrastes como reglas de

decisión entre dos hipótesis, que fue la concepción de Neyman y Pearson. La enseñanza ignora estas diferencias y presenta los contrastes de hipótesis como si se tratase de una única metodología (Batanero, 2000). En el razonamiento de una prueba de significación de Fisher se mide la fuerza de la evidencia proporcionada por los datos observados en contra de una hipótesis llamada nula. Esta fuerza de la evidencia es capturado en el p-valor, que mide la probabilidad de haber obtenido los datos observados (o más extremos) si la hipótesis nula fuera verdadera. Se empleará esta concepción para introducir el contraste de hipótesis.

### **Razonamiento inferencial informal**

La inferencia estadística informal (IEI) es una generalización probabilística de los patrones que son revelados por los datos disponibles, y esta generalización es el producto final de un razonamiento inferencial informal (RII) (Makar et al., 2011). Una IEI se representa mediante un enunciado, mientras que el RII es el proceso mediante el cual se descubren y establecen dichos enunciados. Es decir, la forma en que los estudiantes usan sus conocimientos para hacer y sustentar inferencias estadísticas sobre una población desconocida basadas en muestras observadas y sin utilizar los métodos o técnicas formales de la estadística inferencial, como el uso de distribución muestral teórica, desviación estándar, puntuación estándar, etc.

### **La tecnología en la enseñanza de la estadística**

La tecnología ha ampliado la gama de técnicas gráficas y de procedimientos para proporcionar nuevas y poderosas formas de ayudar a los estudiantes en el análisis de datos, lo que les permite centrarse en la interpretación de los resultados, en la comprensión y utilidad de conceptos e ideas estadísticas fundamentales, y no en la mecánica computacional. Erickson (2006) cree que los conceptos difíciles y el proceso del contraste de hipótesis se hacen más visibles y comprensibles al simular con Fathom, ya que la velocidad de la computadora hace que sea posible hacer muchos más ensayos y permite centrarse en la distribución muestral, ayudando a elevar el nivel de abstracción. Además, estudiante puede variar parámetros y describir y explicar el comportamiento que observan en lugar de confiar exclusivamente en las discusiones teóricas de probabilidad, pues estas, a menudo son contrarias a la intuición de los estudiantes (delMas, Garfield y Chance, 1999). El entorno visual, interactivo y dinámico de las aplicaciones ayuda a lograr estos objetivos (Chance, Ben-Zvi, Garfield y Medina, 2007).

Sin embargo, la tecnología tiene un gran impacto sólo si se usa apropiadamente, por esta razón es importante enfocarse en buscar formas eficientes de usar la herramienta tecnológica en el salón de clase. Sugerencias para el uso de la herramienta es el empleo del aprendizaje colaborativo y facilitar la interacción y la accesibilidad a la tecnología, manteniendo el enfoque en los conceptos estadísticos en lugar de la herramienta (Biehler, Ben-Zvi, Bakker y Makar, 2013).

### **METODOLOGIA**

El acopio de datos se llevó a cabo en una sesión mediante un problema de prueba de hipótesis de proporciones con cuatro preguntas, aplicado a 19 parejas de estudiantes de nivel bachillerato (16 y 17 años) en una escuela pública, los cuales no han llevado curso de estadística. Los estudiantes respondieron las preguntas en el ordenador con la posibilidad de usar la aplicación.

Antes de aplicar el problema se dio una sesión de introducción a la aplicación para explicar su funcionamiento con un problema introductorio (fase de familiarización y presentación). Además se contó con hojas de trabajo orientadas a los procesos y se trabajó en un aprendizaje cooperativo con 2 estudiantes por computadora. Esta sesión fue de una hora y la llevó a cabo uno de los autores de la comunicación. Después de la sesión introductoria se les presentó el problema y se dio un tiempo de una hora para responder las preguntas. Durante el cuestionario los estudiantes podían comentar sus dudas con el profesor (fase de trabajo con estudiantes) y discutir entre ellos.

El problema para la recogida de los datos contiene cuatro preguntas: “Propaganda de la Coca Cola presume que el 90% de la población de México prefiere su refresco que Pepsi. Para comprobarlo se les dieron dos vasos de refresco (uno con Coca y otro con Pepsi) a 200 personas escogidas al azar y decidían cuál le gustaba más. De los 200 participantes 188 personas prefirieron Coca Cola. ¿El resultado del experimento es suficiente para decir que el 90% de la población de México prefiere Coca Cola? ¿Por qué? ¿Será posible que la conclusión sea incorrecta aun haciendo bien todo el proceso? ¿Por qué?”. En este trabajo presentamos resultados relativos a las dos primeras preguntas.

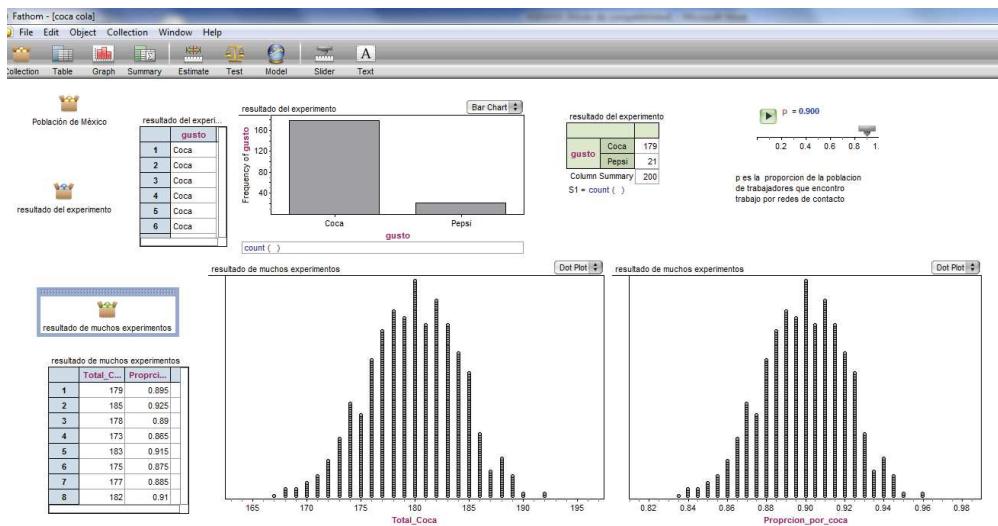


Figura 1. Aplicación interactiva para la prueba de significación

La aplicación utilizada (Figura 1) simula encuestas de 200 personas tomadas de una población hipotética, en este caso el parámetro de la población era de 90%. El parámetro de la población hipotética es manejado por un deslizador, ayudando a los estudiantes a observar efectos al mover parámetros y con ello construir conceptos y justificar conclusiones. Las muestras de tamaño 200 se representan en una gráfica de barras y una tabla. La simulación es usada para generar una distribución muestral empírica y poder medir la probabilidad de los datos observados con el método frecuencial (empírico), esto es, el cálculo informal de un p-valor mediante frecuencias (Rossman, 2008). La distribución muestral se exhibe en una tabla y en dos gráficas de puntos, tanto en valores absolutos como en proporciones.

El diseño de la aplicación pretende ser accesible y facilitar la interacción con los estudiantes, tratando de mantenerlos enfocados en los conceptos estadísticos, sin que la elaboración de la herramienta (aplicación) los perturbe. La animación hace que el estudiante observe todo el proceso de re-muestreo simultáneamente, a saber: obtener una muestra, obtener el valor del estadístico de la muestra y graficar.

## RESULTADOS

Ante la pregunta, ¿El resultado del experimento es suficiente para decir que el 90% de la población de México prefiere Coca Cola?, primero debe surgir la pregunta de si 188 de 200 bajo la hipótesis de  $p = 0.9$  es un resultado plausible o raro. Luego se debe inspeccionar en la distribución muestral empírica la posición 188 y ver que tantos valores (considerando sus frecuencias) de la distribución están por arriba de ese valor. Debe calcularse la frecuencia relativa acumulada de los valores que están por arriba de 188 (p-valor) y compararlo con 0.05. Si el resultado es menor o igual a este valor (resultado raro) se rechaza la hipótesis; si es mayor (resultado normal) no se rechaza.

La metodología teoría fundamentada (Birks y Mills, 2011) establece que es posible elaborar categorías con base en los datos que se recopilan y analizan sistemáticamente y exhaustivamente a través de una variedad de estrategias. Las categorías de análisis emergentes en el presente estudio que explican el RII de los estudiantes se presentan en la tabla 1 y son: Simulación, datos y creencias. Las respuestas se pueden dividir en dos conjuntos: Hay 12 parejas que se basan en la aplicación y 7 parejas que no la utilizan. Estos últimos probablemente no ven relaciones entre el funcionamiento de la aplicación y el problema que se les plantea. Los códigos empleados para identificar a las parejas se denotan con la letra E y un número, así, E4 significa la pareja número 4.

Tabla 1. Categorías de análisis

Medición del p-valor	Argumentación	Conclusión
Simulación 12 (de 19 parejas)	Muestra rara	Rechazar hipótesis
	Muestra común	No rechazar hipótesis
Datos 4 (de 19 parejas)	Proporción diferente a la hipótesis (mayor o menor)	Rechazar hipótesis
		No rechazar hipótesis
Creencias 3 (de 19 parejas)	Población y publicidad	Rechazar hipótesis
		No rechazar hipótesis

Simulación. En esta categoría se clasifican las respuestas de los estudiantes que utilizan la simulación y cuentan las frecuencias en las que se obtienen los datos para medir informalmente un p-valor y determinar si los datos son resultados normales o raros suponiendo una población donde 90% prefiere Coca Cola.

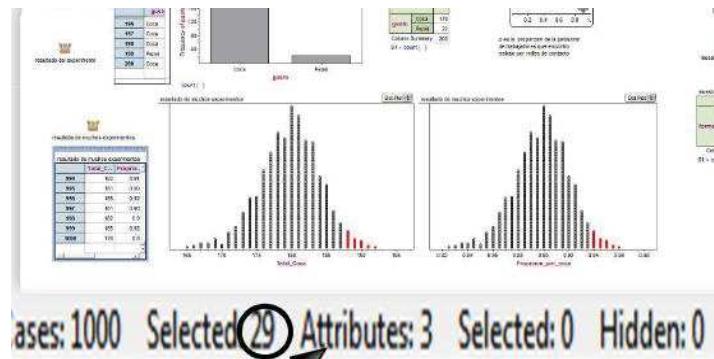


Figura 1. Distribución muestral empírica de E12

La Tabla 1 muestra una mayoría de estudiantes utilizando la simulación para medir el p-valor (12 de 19 parejas) y determinar si la muestra es rara o no. Un ejemplo de respuesta en esta categoría corresponde al equipo E12, quienes argumentan “Como ha ocurrido un resultado raro suponiendo que el 90% de la población prefirió la coca cola, concluimos que la suposición es muy rara y que el resultado de la encuesta se debió a algo raro” y agregan la imagen que se muestra en la Figura 1.

Otro ejemplo en esta categoría es la respuesta de E9, donde explican “el resultado de la encuesta nos dice que son 37 de 1000, por lo cual nos da a entender que es un resultado raro”, y presenta la Figura 2 en su respuesta donde señalan las frecuencias.

La pareja E14 es otro ejemplo en esta categoría. Responden “por cada 1000 por lo menos debe de haber 50 casos, concluimos que como en esta encuesta fue 28 por cada 1000, es un resultado poco probable y que por lo tanto tampoco se representa el 90% de la población”, y presentan la Figura 3.

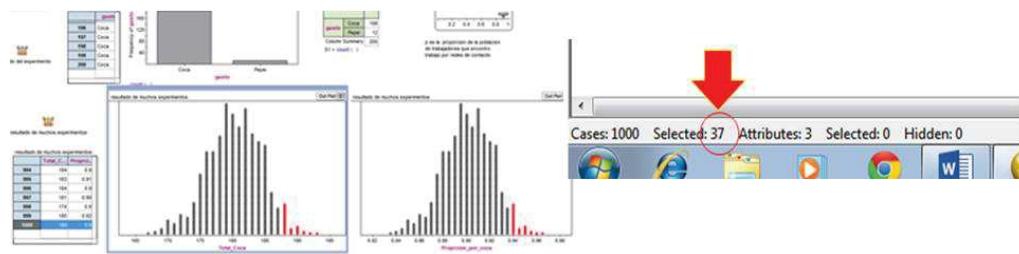


Figura 2. Distribución muestral empírica de E9

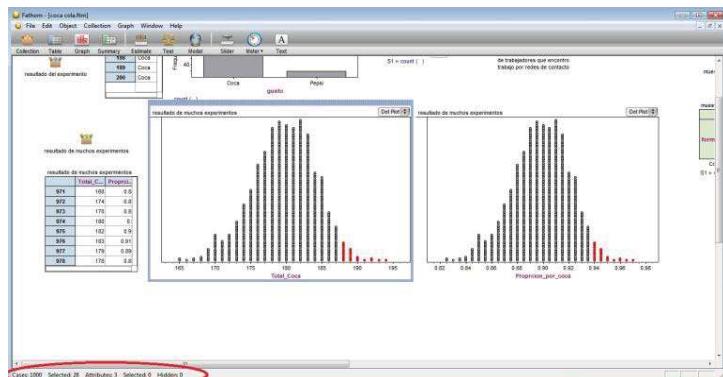


Figura 3. Distribución muestral empírica de E14

A pesar de que la mayoría utiliza la simulación hay otros que se basan en sus creencias o no van más allá de los datos (García-Ríos, 2013). Estos dos tipos de razonamiento son considerados como intuitivos (García y Sánchez, 2015) y es necesario desarrollar la noción de distribución muestral de una manera informal para alcanzar un nivel mayor en el RII. Estas respuestas se clasificaron en “Datos” y “Creencias” respectivamente.

**Datos.** Dentro de esta categoría se ubican las respuestas donde el razonamiento se basa en los datos de las muestras y no consideran que hay variación del estadístico de muestra a muestra; así, si la proporción de la muestra es mayor que la hipótesis entonces concluyen que es más del 90% y rechazan la hipótesis. E8 responde “porque el 90% de 200 es 180 por lo tanto la mayoría de las personas prefieren coca porque esta superior que la Pepsi según los participante de la encuesta”.

**Creencias.** En esta categoría se clasifican las respuestas que no consideran los datos para argumentar su conclusión. E2 responde “Coca Cola es una empresa mundial y por lo cual tiene mayor publicidad, mayor calidad probablemente”. Una respuesta por demás interesante es la dada por E6 “Porque los resultados son al azar, entonces alguna otra parte de la población que no fue tomada en la encuesta prefiere Pepsi, así que los porcentajes cambian y no se puede saber con seguridad”. Se tiene presente la variación, sin embargo en un grado excesivo, al no considerar que una muestra pueda ser representativa.

## DISCUSIÓN

Como parte de la discusión se dará respuesta a las preguntas de investigación planteadas en la introducción y comentarios sobre los resultados encontrados.

¿Cuál es el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación que llevan a cabo estudiantes de bachillerato? Los resultados muestran que varios estudiantes comprenden el proceso de construcción de la distribución muestral por medio de la simulación y que pueden determinar el p-valor del estadístico de la muestra comparándola con la distribución muestral empírica (12 de 19 parejas). Este resultado es importante pues se ha documentado que los estudiantes tienen mucha dificultad para medir la significatividad del estadístico (p-valor) y lo hacen de forma intuitiva: si el estadístico es diferente a la hipótesis entonces se rechaza la hipótesis

(García y Sánchez, 2015). Otros estudiantes requieren que la muestra esté extremadamente lejos para rechazar la hipótesis (García-Ríos, 2013). En nuestra opinión, la dificultad principal que enfrentan los estudiantes para rechazar o no una hipótesis de manera intuitiva es la de carecer de un criterio para determinar las fronteras de la zona crítica; por lo que resulta conveniente que desarrollen un método informal que les permita hacerlo.

Se han observado otras limitaciones en la forma intuitiva de razonar de los estudiantes. Una consiste en la búsqueda de la respuesta en los datos pero sin tener un modelo de probabilidad (simulación), es decir, solo se enfocan en el estadístico de la muestra y lo comparan con el parámetro de la población (García y Sánchez, 2015). La otra limitación consiste en ofrecer argumentaciones con base en conocimientos o creencias personales acerca del contexto, sin llevar a cabo procedimientos que utilicen los datos. Aunque el papel del contexto en las inferencias es delicado (en ciertos momentos del análisis se debe ignorar y en otros se ha de tener en cuenta) la consideración de los datos es una necesidad.

¿Qué errores y dificultades persisten en el razonamiento para hacer inferencias estadísticas informales con el uso de la simulación? Una primera dificultad en el razonamiento es: No tomar en cuenta la distribución muestral empírica ya sea por basarse en creencias para explicar los datos o solo enfocarse en el estadístico de la muestra sin un modelo de probabilidad. También es posible que no se comprendiera el proceso de construcción de la distribución muestral empírica, provocando la falta de percepción de la variación natural de las muestras y su representatividad.

Un dato que no se analiza en esta comunicación es que a pesar de que varios estudiantes parecen comprender el proceso de construcción de la distribución muestral por medio de la simulación y que pueden determinar el p-valor del estadístico, algunos llegan a diferentes conclusiones. Estudiantes que encuentran una muestra normal rechazan la hipótesis cuando no debían; es posible que algunas consideraciones sobre el contexto de la situación influyan en esa decisión. Esta inconsistencia muestra que no son suficientes los resultados de la simulación (Lipson et al., 2003) y que es necesario que los estudiantes desarrollen un esquema general de cómo funcionan las inferencias.

Un posible camino para atacar estas dificultades es proponer más actividades bien diseñadas con variantes en los parámetros y discusiones finales en cada actividad para hacer notar por qué algunos argumentos no son adecuados. Lo presentado aquí es solo una parte de una actividad entre varias actividades que se han planeado para investigar en el tema. Además, se ha centrado en un solo componente del proceso del contraste de hipótesis (muestreo, variación, incertidumbre, lenguaje, conclusión, etc.) pero los resultados aquí obtenidos aportan información para refinar las actividades previstas y para construir un modelo de desarrollo del RII.

## Referencias

- Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, 2(1-2), 75-98.
- Biehler, R., Ben-Zvi, D., Bakker, A. y Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. En A. Bishop, K. Clement, C. Keitel, J. Kilpatrick y A. Y. L. Leung (Eds.), *Third international handbook on mathematics education* (pp. 643-689). Nueva York: Springer
- Birks, M. y Mills, J. (2011). *Grounded theory: A practical guide*. Thousand Oaks, CA: 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, 98-113.
- Chance, B. L., Ben-Zvi, D., Garfield, J. y Medina, E. (2007). The role of technology in improving student learning. *Technology Innovations in Statistics Education*, 1(1), 1-26.

- Chance, B., delMas, R. C. y Garfield, J. (2004). Reasoning about sampling distributions. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295-323). Amsterdam: Kluwer Academic Publishers.
- delMas, R., Garfield, J. y Chance, B. (1999). A model of classroom research in action: Developing simulation activities to improve student's statistical reasoning. *Journal of Statistics Education*, 7(3), <http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm>
- Erickson, T. (2006). Using simulation to learn about inferences. *International Conference on Teaching Statistics*, 7, 1-6, [https://www.stat.auckland.ac.nz/~iase/publications/17/7G2\\_ERIC.pdf](https://www.stat.auckland.ac.nz/~iase/publications/17/7G2_ERIC.pdf)
- Finzer, W. (2014). *Fathom Dynamic Data Software* (Version 2.2) [Software], <http://concord.org/fathom-dynamic-data-softwa>
- García-Ríos, N. (2013). Inferencias estadísticas informales en estudiantes mexicanos. En J. M. Contreras, G. R. Cañadas, M. M. Gea y P. Arteaga (Eds.), *Actas de las Jornadas Virtuales en Didáctica de la Estadística, Probabilidad y Combinatoria* (pp. 343-357). Granada: Universidad de Granada.
- García, V. N. y Sánchez, E. (2014). Razonamiento inferencial informal: El caso de la prueba de significación con estudiantes de bachillerato. En M. T. González, M. Codes, D. Arnau y T. Ortega (Eds.), *Investigación en Educación Matemática XVIII* (pp. 345-357). Salamanca: SEIEM.
- García V. N. y Sánchez E. (2015). Dificultades en el razonamiento inferencial intuitivo. En J. M. Contreras y otros (Eds.), *Didáctica de la Estadística, Probabilidad y Combinatoria 2* (pp. 207-214). Granada: Universidad de Granada.
- Garfield, J. y Ben-Zvi, D. (2008). *Developing student's statistical reasoning*. Nueva York: Springer.
- Hofmann, T., Maxara, C., Meyfarth, T. y Prömmel, A. (2014). Using the Software FATHOM for learning and teaching statistics in Germany –A review on the research activities of Rolf Biehler's working group over the past ten years. En T. Wassong y otros (Eds.), *Using tools for learning mathematics and statistics* (pp. 283-304). Wiesbaden, Alemania: Springer.
- Jabareen, Y. (2009). Building conceptual framework: Philosophy, definitions and procedure. *International Journal of Qualitative Method*, 8(4), 49-62.
- Lane, D. M. y Peres, S.C. (2006). Interactive simulations in the teaching of statistics: Promise and pitfalls. En A. Rossman y B. Chance (Eds.). *Proceedings of the 7<sup>th</sup> International Conference on Teaching Statistics*. Voorburg, Holanda: International Statistical Institute.
- Lipson, K., Kokonis, S. y Francis, G. (2003). Investigation of statistics students experiences with a web-based computer simulation. En L. Weldon y J. Engel (Eds.). *Proceedings of the International Association of Statistical Education Satellite Conference on Statistics Education*, <https://www.stat.auckland.ac.nz/~iase/publications/6/Lipson.pdf>
- Makar, K., Bakker, A. y Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1-2), 152-173.
- Makar, K. y Ben-Zvi, D. (2011). The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, 13(1-2), 1-4.
- Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? En G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267-294). Nueva York: Springer.
- Pratt, D. y Ainley, J. (2008). Introducing the special issue on informal inference. *Statistics Education Research Journal*, 7(2), 3-4.
- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.
- 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.