UN EXPERIMENTO DE DISEÑO PARA LA INTRODUCCIÓN AL RAZONAMIENTO SOBRE CONTRASTE DE HIPÓTESIS EN EL BACHILLERATO

Ernesto Alonso, Sánchez Sánchez
esanchez@cinvestav.mx
esanchez0155@gmail.com
CINVESTAV-IPN (México)
Eleazar, Silvestre Castro
eleazar.silvestre@gmail.com
CINVESTAV-IPN (México)

Asunto: Uso de recursos didácticos o tecnológicos
Temática: Muestreo

RESUMEN

Se presentan resultados de dos ciclos de un experimento de diseño cuyo objetivo es propiciar el aprendizaje de pruebas de significación y contraste de hipótesis con apoyo de un software educativo (Fathom). Del primer ciclo del experimento se informa de la primera actividad de cuatro realizadas, estas tenían el objetivo de desarrollar el razonamiento sobre pruebas de significación. Del segundo ciclo, se reporta la última actividad de seis diseñadas para desarrollar la noción de distribución muestral y aplicarla al contraste de hipótesis. Las diferencias en el diseño entre actividades se derivaron de decisiones metodológicas para incrementar las probabilidades para un mejor aprendizaje del tema. Los resultados de ambos ciclos permiten bosquejar rasgos importantes del razonamiento sobre contraste de hipótesis, en estudiantes de bachillerato.

PALABRAS CLAVES

Pruebas de significación, Contraste de hipótesis, Distribución muestral, Enfoque informal.

INTRODUCCIÓN

El *contraste de hipótesis* es un concepto fundamental en los cursos introductorios de estadística, y al mismo tiempo existe una considerable cantidad de evidencia que señala que la mayoría de estudiantes tienen dificultades, cometen errores sistemáticos y manifiestan concepciones erróneas cuando realizan tareas sobre el tema (Castro-Sotos, Vanhoof, Van den Noorgate y Onghena, 2007, Vallecillos & Batanero, 1997). Algunos autores atribuyen el origen de tales dificultades a una mezcla de diferentes razones, tales como la gran complejidad epistemológica del concepto (Batanero, 2000) y la prevalencia de prácticas tradicionalistas propensas a reproducir los mismos errores que los profesores cometen (Liu & Thompson, 2009; Harradine, Batanero & Rossman, 2011). En la última década, ha surgido una tendencia en la comunidad de investigadores estadísticos que proponen explorar las posibilidades de que los estudiantes desarrollen un razonamiento inferencial informal antes de estudiar procedimientos más formales (Zieffler, Garfield, delMas & Reading, 2008) y evitar que solamente se promueva el aprendizaje de procedimientos y algoritmos. Las propuestas que se derivan de esta tendencia en lo que concierne a la inferencia clásica, se apoyan fuertemente en el uso de *simulaciones* para construir *distribuciones muestrales* empíricas. 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. Sobre todo permite abrigar 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 (Lipson, 2002). 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 en distribuciones muestrales y contraste de hipótesis.

En este estudio respondemos a las siguientes preguntas que han emergido durante el desarrollo del experimento:

a) ¿Cómo razonan los estudiantes cuando enfrentan un problema sobre pruebas de significación con el apoyo de un programa computacional que simula muestreo repetido?

b) ¿Cómo influye en el razonamiento de los estudiantes un trabajo previo con distribuciones muestrales empíricas cuando enfrentan problemas de contraste de hipótesis?

MARCO DE REFERENCIA

El trabajo que aquí se presenta es una investigación de diseño, también llamado experimento de diseño (Cobb, Confrey, Disessa, Lehrer & Schauble, 2003), cuyo propósito, en el caso particular de la inferencia estadística que aquí nos ocupa, es entender las relaciones entre el diseño y el desarrollo del razonamiento estadístico de los estudiantes mediante la creación de un ambiente que les proporcione oportunidades para que adquieran conocimientos estadísticos y mejoren su razonamiento inferencial sobre contraste de hipótesis. El diseño consiste en secuencias de actividades de aprendizaje para ser realizadas por estudiantes de bachillerato. El aprendizaje que se espera que los estudiantes adquieran es el razonamiento básico de las pruebas de significación, mismo que mediante algunas modificaciones llevan al contraste de hipótesis, aunque en su momento ambos enfoques provocaron fuertes controversias (Batanero 2000).

Una prueba de significación consiste en decidir si una muestra dada es significativa, es decir, si la muestra aconseja rechazar la hipótesis a contrastar. Ésta es una afirmación acerca de la proporción de elementos de la población que tienen alguna propiedad (*v.g.* La propiedad de que a los individuos de una población de personas les guste o no la Coca Cola). La hipótesis generalmente es conservadora, en el sentido de que, si no hay suficiente evidencia en sentido contrario, se conserva; es decir, se asume que en efecto la hipótesis corresponde a la población en investigación. La evidencia para evaluar la hipótesis proviene de la muestra concreta que forma parte de la información del problema. La pregunta es: ¿La muestra sería muy rara si proviniera de una población cuya proporción es la que afirma la hipótesis? Si la respuesta es positiva, entonces se rechaza la hipótesis, en caso contrario se mantiene. Para determinar si la muestra es rara o no se calcula el p-valor, es decir, la probabilidad de que el estadístico tome el valor dado u otro más extremo suponiendo verdadera la hipótesis. Una estimación del p-valor, y en general de probabilidades de que el estadístico caiga en un conjunto dado, se puede obtener mediante una distribución muestral empírica, generada con el software Fathom. En un contraste de hipótesis el razonamiento es algo diferente a las pruebas de significación, pero la distribución muestral sigue siendo central para realizar el contraste. La decisión se hace comparando el p-valor con el nivel de significación definido de antemano (generalmente 5%). En el nivel informal en que esperamos que los estudiantes avancen, asumimos que no es pertinente incluir como objetivo de la instrucción abarcar otros conceptos como tipos de errores y potencia de la prueba. El procedimiento informal que se espera que los estudiantes lleven a cabo lo describimos en el apéndice 2, para el ejemplo B que se mostrará adelante.

MÉTODO

En un experimento de diseño (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003) hay tres etapas iniciales que constituyen un ciclo en el desarrollo de una investigación: diseño, implementación y análisis retrospectivo. Con base en los resultados del análisis se modifican aspectos de la actividad inicial y el resultado es una actividad mejorada respecto a la primera, misma sirve de base para un segundo ciclo de la investigación. La investigación de la cual mostramos una parte, estuvo formada por dos ciclos; el primero constaba de cuatro actividades y el segundo ciclo de seis actividades. El presente estudio se enfoca en la primera actividad del primer ciclo y la última del segundo. El primer ciclo del experimento se enfocó en la noción de pruebas de significación de Fisher bajo la consideración de que podría ser más accesible para los estudiantes que el método más general de contraste de hipótesis (García & Sánchez, 2017). Se llevó a cabo con 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. En el segundo ciclo el acercamiento fue más hacia el contraste de hipótesis, pero sin desarrollar todos los conceptos que este enfoque presupone (Silvestre & Sánchez, 2017). Participaron 42 estudiantes de quinto semestre de bachillerato (17-18 años de edad) organizados en 21 parejas; ya habían cursado un primer semestre de probabilidad y estadística, cubriendo temas de estadística descriptiva y probabilidad, pero no de inferencia estadística. En Tabla 1 se resumen las diferencias de las características del experimento en los ciclos.

Grupo A	Grupo B
• Enfoque de pruebas de significación. • 36 estudiantes de onceavo grado (16-17 años). • No han llevado curso de Probabilidad y Estadística del Bachillerato. • Resolvieron directamente tareas sobre prueba de significación teniendo como único antecedente algunas experiencias con tópicos de estadística descriptiva.	• Enfoque de contraste de hipótesis. • 44 estudiantes de doceavo grado (17-18 años). • Han cursado un semestre del curso de Probabilidad y Estadística. • Exploraron y resolvieron tareas de muestreo enmarcadas utilizando simulaciones aleatorias en Fathom como antecedente a la actividad de contraste de hipótesis.

Tabla 1. Características de los ciclos: grupos A y B
Fuente. Creación propia

El problema de la primera actividad del primer ciclo (Ejemplo A) y el de la última actividad del segundo ciclo se muestran en la Tabla 2.

Ejemplo A / Actividad 1	Ejemplo B / Actividad 6
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 $\alpha = .05$]	"Una compañía posee cuatro máquinas especializadas en producir tarjetas madre para laptop. De forma irremediable, cada máquina produce aleatoriamente una cantidad de tarjetas defectuosas después de un cierto período de tiempo. El departamento de control de calidad indica que *las máquinas pueden presentar hasta 10% de tarjetas defectuosas en su producción*, de lo contrario éstas deben ser *enviadas a revisión para una posible reparación*. Un equipo de técnicos recolectó en cierto día una muestra aleatoria de 120 tarjetas de cada máquina para analizar el número de tarjetas defectuosas que se presentan, obteniendo así los siguientes resultados: M_A: 42 tarjetas defectuosas, M_B: 21, M_C: 27 y M_D: 15; *¿cuáles máquinas consideras que deben ser enviadas a una inspección?* – [H_0: $P \leq .1$ es aceptada únicamente para M_D al nivel de $\alpha = .025$] (Ver soluciones en el Apéndice 1)

Tabla 2. Problemas de las actividades elegidas
Fuente. Creación propia

En ambos ciclos, los estudiantes leían y discutían en parejas las preguntas alrededor de los problemas, para responderlas individualmente por escrito en hojas de trabajo. 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 (Glaser & Strauss, 1967/2008; Birks & Mills, 2011*)*, que es una metodología general (Holton, 2008) de investigación en ciencias sociales cuyo objetivo es elaborar teorías locales que emerjan de datos y no de deducciones lógicas o de otras teorías. 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, las respuestas a cada pregunta se compararon entre sí para codificarlas. Mediante este proceso se determinaron patrones generales que son los que se exponen en seguida. 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 cultura de la clase; se han omitido variaciones particulares que son intrínsecas a la complejidad de los razonamientos de los estudiantes.

RESULTADOS

Con relación a la primera pregunta de investigación, se puede afirmar que, en general, un estudiante de bachillerato percibe un problema de decisión sobre la hipótesis acerca de una proporción, como un problema simple consistente en comparar la hipótesis con la proporción de la muestra y, a partir de esto, tomar la decisión de rechazar la hipótesis cuando son diferentes. La totalidad de estudiantes del estudio A percibió de esta forma el problema del ejemplo A; la siguiente respuesta de una pareja de estudiantes es un ejemplo de este tipo de percepción:

> […] 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", recalcando 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 × 60 = 35.4), podemos concluir que no erran en lo que presumen ya que están en todo lo correcto […]

La situación cambia sólo de manera parcial cuando los estudiantes previamente han aprendido a generar e interpretar una distribución muestral, pues en el Estudio B, 75% de los estudiantes respondió de manera parecida; por ejemplo, teniendo en cuenta el ejemplo B, una pareja de estudiantes respondió:

> Mandaríamos a reparar todas las máquinas ya que el límite permitido de tarjetas defectuosas es el 10% y todas las máquinas rebasaron ese porcentaje, pues el máximo de tarjetas defectuosas que deberían salir serían 12 de una muestra de 120.

Pero el 15% de los estudiantes del estudio B no se fueron por este camino fácil y notaron que la situación encierra variabilidad; quizá influidos por su trabajo previo con distribuciones muestrales consideraron que la pretensión de obtener la igualdad de la hipótesis con la proporción de la muestra es excesiva y sugirieron considerar un criterio de aproximación entre ellas, este criterio es lo que llama la pareja de estudiantes del siguiente ejemplo "un margen de error":

> Se mandarían a revisión las máquinas A, B y C de manera inmediata debido a que exceden el 10% que se permitía como margen de error. De la misma manera se mandaría a revisar la máquina D, pero esa, se haría dicha operación en un tiempo posterior, debido a que no hay mucha variación en el margen de error.

En ambos estudios (A y B) el instructor promueve una discusión con el fin de que aquellos estudiantes que sobre-simplifican la situación perciban la variabilidad muestral. En el grupo A se comenzó a introducir el proceso de muestro repetido para confrontar sus creencias. En el grupo B se propició que se hiciera un análisis con base en la distribución muestral.

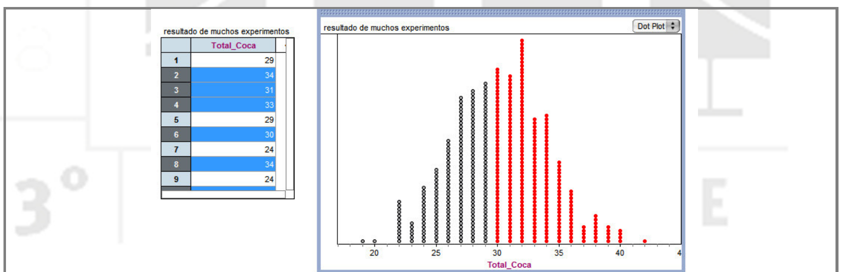
Con relación a la pregunta ¿qué hacen los estudiantes cuando son introducidos a la distribución muestral? En ambos estudios el instructor intervino para enseñar a los estudiantes a generar distribuciones muestrales y a analizar la situación mediante el razonamiento de pruebas de significación de Fischer (grupo A) o de contraste de hipótesis (grupo B); en el primer caso se hizo paralelamente al desarrollo de las actividades, en el segundo caso se hizo previamente a las sesiones de resolución del problema de contraste.

Aunque con los procesos de instrucción los estudiantes aprendieron con relativa facilidad a generar distribuciones muestrales con el software, los estudiantes del grupo A, no construyeron el concepto de distribución muestral pues aún no identifican y ni utilizan sus propiedades básicas, mientras que los del grupo B tuvieron avances relativamente importantes en la construcción del concepto.

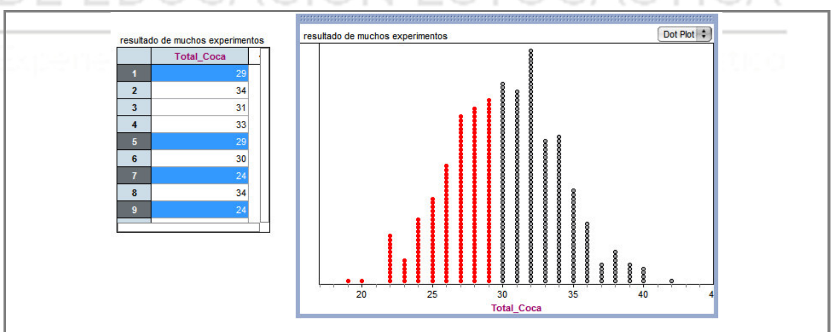
Una propiedad fundamental de la distribución muestral es permitir el cálculo de la probabilidad de que el valor del estadístico se encuentre en una determinada zona de su dominio. Ninguno de los estudiantes del grupo A en su uso de la distribución muestral utilizó en absoluto enunciados de probabilidad. En su lugar, creían que la distribución les daba información sobre la cantidad de "personas que les gusta la Coca-Cola" o el número de encuestas favorables a "personas que les gusta la Coca-Cola". En las Gráficas 1 y 2 se muestran los conjuntos que consideran para hacer una inferencia (errónea).

> Debido a que en la simulación se muestra 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 encuestas realizadas se mostró un gusto en menos del 50% de 60 personas"

Otros se fijan en la moda o en conjuntos cuya frecuencia tiene más del 50%.



Gráfica 1. Resultados de encuestas con resultados satisfactorios para Coca Cola
Fuente. Creación de los estudiantes



Gráfica 2. Resultados de encuestas con resultados infructuosos para Coca Cola
Fuente. Creación de los estudiantes

De la manera en que usan la distribución muestral se deduce que los estudiantes no alcanzan a comprender que hay una diferencia entre una muestra real y una muestra simulada. Se pueden observar dos tipos de razonamiento que adolecen del mismo defecto, a saber, creen que el proceso de simulación aporta información adicional sobre la población en investigación. Los razonamientos (erróneos) subyacentes se podría expresar de así:

- Si mediante un proceso de simulación se encuentra que más personas prefieren Coca-Cola entonces se rechaza la hipótesis de igualdad.
- Si mediante un proceso de simulación se encuentra que en la mayoría de muestras hay mayor preferencia por la Coca-Cola entonces se rechaza la hipótesis de igualdad.

Los estudiantes del estudio B, trabajaron previamente en la construcción e interpretación de distribuciones muestrales por lo que la manera de utilizarla en el problema de la revisión de las máquinas (ejemplo B) se acerca más a su propiedad fundamental. Una vez que mediante la discusión promovida por el profesor los estudiantes superaron la idea de comparar sólo la hipótesis con el valor de la proporción de la muestra, la mayoría de ellos utilizó la distribución muestral generada ex profeso para el problema, con el propósito de hacer una valoración acerca de la posibilidad de revisar cada máquina identificando los rangos de valores que contienen el 90% de las frecuencias alrededor de la media. Esto les permitió hablar de "valores críticos" para ver si dentro de ellos se encuentra o no el valor del estadístico. Por ejemplo, una pareja de estudiantes expresan de manera clara cómo una vez identificados los "valores críticos" pueden tomar una decisión con relación a las cuatro máquinas:

> *A:* Sí la mandaríamos a arreglar debido a que no se encuentra dentro de los valores críticos inferior y superior.
> *B:* Sí, la mandaríamos a arreglar debido a que también se encuentra fuera de los valores englobados por el valor crítico inferior y superior y por lo tanto no es un valor muy frecuente de tarjetas defectuosas.
> *C:* Sí, porque al igual que la máquina A y B este valor no está englobado dentro del valor crítico inferior y superior y por lo tanto es un valor con poca frecuencia de aparición.
> *D:* No, porque este valor sí se encuentra englobado dentro del valor crítico inferior y superior por lo cual es un valor frecuente de tarjetas defectuosas, además es un valor cercano al promedio.

Un rasgo importante en algunas respuestas consiste en el uso de enunciados de probabilidad, mismos que estuvieron totalmente ausentes en el grupo A; por ejemplo:

> Es seguro que se mandarían a reparar las máquinas A, B y C ya que son las que presentan mayor cantidad de tarjetas defectuosas y se encuentran fuera del rango de valor crítico superior (el rango 10-15%), por lo que *la máquina D tiene menos probabilidad de producir tarjetas defectuosas*. (Subrayado nuestro).

Como se esperaba, los estudiantes del grupo B aprenden un método informal para hacer el contraste de hipótesis utilizando la distribución muestral; para determinar si el valor del estadístico es frecuente o poco frecuente (raro) y con base en este resultado decir si la hipótesis es "correcta" o no. No obstante, la instrucción y las actividades no fueron suficientes para que la mayoría llegara a ser consciente de que el método de contraste no puede determinar si la hipótesis es correcta o no; ellos creen que siguiendo correctamente el método determinan la verdad o falsedad de la hipótesis. Al analizar con detenimiento el lenguaje que usan los estudiantes al formular sus conclusiones sobre la prueba utilizan la expresión "enviar a reparar" en lugar de "enviar a revisar" con lo que se sugiere que mediante la prueba se puede saber si las máquinas están descompuestas o no. Este no es un detalle menor, ya que abre la sospecha de que todavía persiste la creencia de que de algún modo oculto el proceso de simulación ofrece información sobre la población real que se investiga.

CONCLUSIONES

La noción intuitiva inicial que los estudiantes tienen acerca del muestreo y la inferencia es que entre más grande es una muestra mejor informa sobre los rasgos de la población. Al ser introducidos al proceso para generar una distribución muestral y utilizarla para hacer una inferencia, estos estudiantes interpretan el nuevo conocimiento sobre pruebas de significación en términos de su concepción previa, por lo que ven el proceso de simulación de muestras un procedimiento para tener una muestra grande de la población o contar con información de muchas muestras. Creen que el método de generación de la distribución muestral (es decir, de simular muestras en la computadora), de manera misteriosa, proporciona nueva información sobre la población; así, para ellos, la distribución muestral es la distribución de una gran muestra o el resumen de los resultados de múltiples muestras de la población en estudio.

Cuando reciben instrucción previa sobre distribuciones muestrales antes de enfrentar un problema de contraste de hipótesis comienza a entender la función de la distribución muestral. Esto se revela cuando se basan en ésta para determinar zonas de rechazo (o no) de la hipótesis en la distribución muestral con base en el nivel de significación. En consecuencia, comienzan a entender la lógica de las pruebas del contraste de hipótesis. No obstante, no tenemos evidencias suficientes para afirmar que conciben a la distribución muestral como un modelo teórico que describe la variabilidad muestral y que es independiente de los problemas específicos de los que se está tratando; aún no sabemos si tienen clara la idea de que es un instrumento para evaluar la verosimilitud de un valor del estadístico bajo el supuesto de que la hipótesis es verdadera y no un procedimiento que añade información sobre la población real de que proviene el problema.

ÁPENDICE A

El procedimiento para responder la pregunta del Ejemplo A, comienza con la formulación de la hipótesis nula $\mathbf{H_0: p = 0.5}$, es decir, que el número de personas que prefieren Coca-Cola es igual al número de personas que prefiere Pepsi. Hay que preguntarse si bajo esta hipótesis el resultado de 35 de 60 es plausible o, por el contrario, raro. Esto se determina con el cálculo del p-valor. Éste es la probabilidad de que el estadístico tome el valor 35 o más, suponiendo verdadera la hipótesis nula. Si $\hat{\mathbf{p}}$ es el estadístico "la proporción de personas que prefieren Coca de la muestra", se puede afirmar que el estadístico $\hat{\mathbf{p}}$ se distribuye normalmente ya que se puede aprovechar la aproximación normal a la binomial, pues $\mathbf{0.5 \times 60 = 30 > 15}$. El error estándar es $\mathbf{se_0} = \sqrt{\frac{\mathbf{0.5\times0.5}}{\mathbf{60}}} = \mathbf{0.064}$. Entonces $\mathbf{z_0} = \frac{\mathbf{0.58-0.5}}{\mathbf{0.064}} = \mathbf{1.25}$. Con $\hat{\mathbf{p}} = \frac{\mathbf{35}}{\mathbf{60}}$.
El $\mathbf{p}$-valor es $\mathbf{P(z \geq 1.25\ |z \sim N(0,1)) = 0.10565}$; como es mayor que 0.05, se considera que 35 de 60 es normal y no se puede rechazar la hipótesis nula. Por lo que se considera que la afirmación de que la Coca es más preferida que la Pepsi no está suficientemente sustentada por los datos.

Para responder a las preguntas del ejemplo B, se considera la hipótesis nula $\mathbf{H_0: p = 0.10}$. Como $\mathbf{0.10 \times 120 = 12 < 15}$, no conviene suponer que el estadístico $\hat{\mathbf{p}}$ se distribuye normalmente. Consideramos entonces la variable aleatoria X tal que $\mathbf{X \sim b(x, 120, 0.1)}$.Se deben calcular para cada máquina el p-valor.

$$\mathbf{p_A = P(X \geq 42) \cong 0;\ p_B = P(X \geq 21) \cong 0.008,}$$
$$\mathbf{p_C = P(X \geq 27) \cong 0.00005, p_D = P(X \geq 15) \cong 0.218}$$

De donde las máquinas A, B y C deben ser revisadas mientras la máquina D no.

Como se puede apreciar el procedimiento tradicional es abstracto y está lejos de transparentar las razones que justifican los diferentes pasos que llevan a la solución. En el curso de la actividad se recurrió a un enfoque informal alternativo que utiliza las posibilidades de procesamiento rápido de la tecnología y permite evitar problemas analíticos del enfoque tradicional. Vamos a ilustrar este enfoque para el problema del ejemplo B.

ÁPENDICE 2

Para evaluar el desempeño de los estudiantes, se han distinguido cuatro etapas en el proceso de avance en la solución de los problemas: 1. Hipótesis y simulación; 2. Determinación del p-valor y zona crítica; 3. Formulación de la conclusión; 4. Reconocimiento de la incertidumbre.

Simulación de la distribución muestral del problema (ejemplo B). Se realizan las siguientes acciones en el software Fathom:

Hipótesis y simulación:

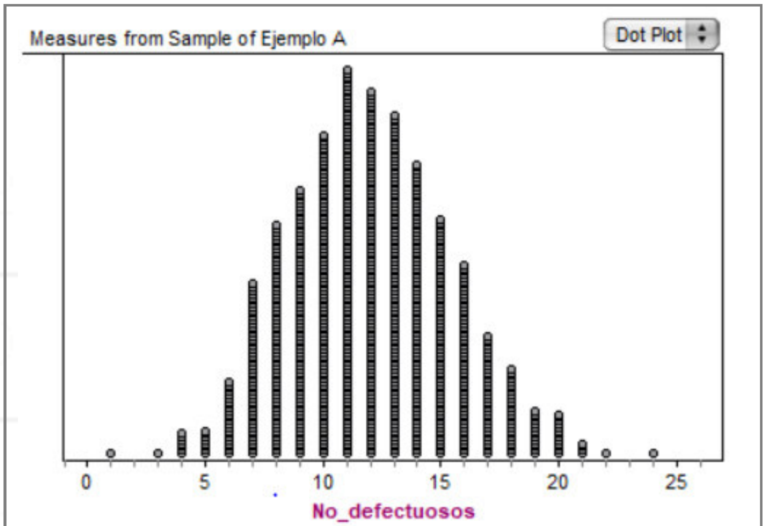
1. Sean N y D las representaciones de un artículo No-defectuoso y Defectuoso respectivamente.
2. Llevar a cabo un sorteo en el que se obtenga una N con 0.9 de probabilidad y un D con 0.1 de probabilidad.
3. Repetir el sorteo 120 veces y observar la variable aleatoria "el número de D's obtenidas en la muestra"
4. Registrar y graficar la distribución (empírica) de la variable aleatoria del punto anterior llevando a cabo 1000 sorteos.
5. Una realización de los pasos anteriores en el software Fathom lleva a la distribución empírica de la Gráfica 3.
6. Determinación de p-valor y/o zona crítica
7. De acuerdo a la distribución muestral empírica o simulada ($H_0 = 0.10, n = 120$), las probabilidades estimadas de obtener los valores muestrales o más (p-valores) son a) cero, b) 0.002, c) cero, d) 0.156. La zona crítica $P(X > z_0) = 0.5$ se determina a partir de $z_0 \epsilon (18, 19)$.

Formulación de la conclusión

Los resultados raros bajo la hipótesis nula son los de las máquinas A, B y C, pues en estos casos el p-valor resultó menos que 0.05; sólo en el caso de la máquina D, el resultado es plausible ya que el p-valor fue de 0.156 mayor a 5%. De donde la conclusión es que las máquinas A, B y C requieren de una revisión, mientras que la máquina D está funcionando adecuadamente.

Reconocimiento de la incertidumbre

El resultado de que las máquinas A, B y C deben revisarse y la D no, no necesariamente quiere decir que las máquinas A, B, C están descompuestas y la D no lo está. Una máquina puede mantener el 10% de probabilidad de artículos defectuosos y no obstante el número de artículos defectuosos de una muestra particular podría caer en la zona crítica. Esto se ve en la misma simulación en la que bajo la hipótesis nula, hay muestras que están en la zona crítica; es decir que pueden ocurrir. No obstante, la lógica del contraste es asumir que es prudente cuestionar la hipótesis nula cuando los resultados de la muestra caen en la zona crítica.



Gráfica 2. Distribución muestral del problema del ejemplo B
Fuente. Creación propia

REFERENCIAS

Batanero, C. (2000). Controversies around significance tests. *Mathematical Thinking and Learning*, *2*(1-2), 75-98.

Birks, M., & Mills, J. (2011). *Grounded theory: A practical guide*. California: Sage.

Castro-Sotos, A. E., Vanhoof, S., Van den Noortgate, W., & 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.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). The role of design in educational research. *Educational Researcher*, *32*(1), 9-13.

García, V. N. & Sanchez, E. (2017). Exploring high school students beginning reasoning about significance tests with technology. In Galindo, E., & Newton, J., (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., & Strauss, A. L. (1967/2008). *Discovery of grounded theory: Strategies for qualitative research*. New Brunswick, USA: Aldinee Transaction.

Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching*

Statistics in School-Mathematics-Challenges for Teaching and Teacher Education (pp. 235- 246). A Joint ICMI/IASE Study.

Holton, J. A. (2008). Grounded Theory as a general research methodology. *The Grounded Theory Review, 7*(2), 67-93.

Liu, 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. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.

Silvestre, E., & Sánchez, E. (2017). High school students' first experiences with the Sampling Distribution: toward a distributive perspective of sampling and inference. En *Proceedings of the Tenth Congress of European Research in Mathematics Education (CERME-10).* Dublín, Irlanda.

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