

Análisis del razonamiento de estudiantes de bachillerato frente a una tarea introductoria al contraste de hipótesis

Eleazar Silvestre Castro¹ y Manuel Alfredo Urrea Bernal²

e-mail: ¹eleazar.silvestre@unison.mx, ²manuel.urrea@unison.mx

Universidad de Sonora

Resumen

El contraste de hipótesis es una de las principales técnicas de la inferencia estadística que ha probado ser un concepto muy difícil de utilizar e interpretar apropiadamente. En años recientes se ha explorado la posibilidad de introducir algunas nociones acerca del contraste en niveles previos al universitario utilizando acercamientos didácticos que se enfocan en desarrollar la lógica que lo sustenta. Bajo este contexto, en este artículo se describen patrones de respuesta y algunos elementos matemáticos desplegados por un conjunto de estudiantes de bachillerato frente a un problema introductorio al contraste de hipótesis, en donde la distribución muestral del estadístico se construye mediante un enfoque de muestreo repetido. Nuestros resultados sugieren que, en un primer momento, la mayoría de los estudiantes se apoya únicamente en la heurística de la representatividad para tomar la decisión de rechazar o aceptar la hipótesis nula, pero también que, una vez que disponen de la distribución muestral que modeliza la hipótesis nula, son receptivos a utilizar apropiadamente sus conocimientos acerca de la noción de valor muestral típico o atípico para tomar una decisión mayoritariamente acertada ante el problema. Con base en estos resultados finalizamos el reporte con una serie de reflexiones para la enseñanza del tema.

Palabras clave: Contraste de hipótesis, inferencia estadística, distribución muestral, muestreo repetido, razonamiento estadístico, bachillerato.

Recibido 31 de enero de 2021

Aceptado 14 de abril de 2021

1. Introducción

Uno de los objetivos actuales y de mayor relevancia para la educación estadística es que los estudiantes de niveles educativos diversos desarrollen una comprensión apropiada acerca de los vínculos entre las ideas de muestreo e inferencia estadística (Burril y Biehler, 2011). Como parte de este sofisticado sistema de conceptos estadísticos y probabilísticos, el tema del contraste o “prueba” de hipótesis figura como una técnica de la inferencia estadística cuyo propósito es – dicho de forma reduccionista – evaluar la fuerza o consistencia de los datos experimentales respecto a un mecanismo probabilístico que podría haberlos producido. Junto a otras técnicas de la inferencia clásica o frecuentista, el contraste de hipótesis contribuye a avanzar en el problema general sobre cómo desarrollar nuevo conocimiento (una nueva teoría científica) a partir del análisis de casos particulares (inducción empírica).

Podría decirse, entonces, que la técnica del contraste constituye un método particular de la investigación científica que permite justificar el razonamiento inductivo y el alcance de sus extensiones, por lo cual es actualmente utilizada en una amplia variedad de situaciones y disciplinas científicas: nuevas medicinas y tratamientos médicos deben probar su efectividad e inocuidad; científicos sociales se interesan por los hábitos e impactos del uso de las redes sociales; comunicólogos y mercadólogos caracterizan preferencias electorales a través de encuestas o estudios de opinión pública; psicológicos realizan experimentos para caracterizar el desarrollo de capacidades cognitivas en determinados temas o situaciones; entre otras.

Dada la importancia del contraste para la educación estadística, el desarrollo del pensamiento científico e inclusive para la evaluación crítica de información estadística encontrada en el día a día (Gal, 2004), no es

de extrañar que el concepto figure en prácticamente todo programa universitario que contemple una introducción a la inferencia estadística, y, más contemporáneamente, en algunos programas curriculares de matemáticas del nivel bachillerato (ver por ejemplo, Ministerio de Educación, 2007; CCH, 2016). No obstante, la literatura especializada ha documentado serias dificultades para la enseñanza y aprendizaje del concepto tanto en estudiantes, profesores e inclusive investigadores en ejercicio (Batanero, 2000; Castro-Sotos, Vanhoof, Noortgate y Onghena, 2009; Inzunza y Jiménez, 2013; Lee, 2018).

La caracterización de estas dificultades se enfoca principalmente en la interpretación del resultado del contraste y en la de algunos conceptos involucrados; por ejemplo, entre las más documentados se destacan: confundir la identificación de la hipótesis nula con la alternativa; creer que las hipótesis aplican tanto a la población bajo estudio como a la muestra; considerar que el p – valor representa la probabilidad de que el resultado de la prueba sea debido a la aleatoriedad; interpretar el nivel de significancia intercambiando los términos de su probabilidad condicional; interpretar el nivel de significancia como la probabilidad de que alguna hipótesis sea verdadera, entre otras. De la literatura especializada se advierten algunas de las principales causas de esta situación, entre las cuales destacamos:

- *Una elevada complejidad epistémica del concepto.* El concepto del contraste se sostiene en una compleja red de relaciones entre conceptos como muestreo, distribución muestral, distribución de probabilidad, probabilidad condicional, incertidumbre, entre otros, donde cada uno tiene su propia fuente de retos y dificultades respecto a sus procesos de enseñanza y aprendizaje. Los libros de texto suelen presentar el tema como un híbrido entre los enfoques de sus principales contribuidores – la propuesta de Ronald Fisher (1890-1962), y la de Jerzy Neyman (1894-1981) y Egon Pearson (1895-1980) –, ignorando que la interpretación del concepto está influenciada por las diferentes posturas filosóficas entre ambos planteamientos (Batanero, 2000).
- *Una alta propensión a incurrir en razonamientos heurísticos.* El influyente programa de investigación desarrollado por Kahneman, Slovic y Tversky (1982) puso de manifiesto que las personas, en general, se apoyan fuertemente en el uso de mecanismos cognitivos simples (heurísticas) que ayudan a reducir y hacer frente a la complejidad involucrada en la toma de decisiones en situaciones bajo incertidumbre. El apoyo en tales recursos contradice a menudo los principios normativos que rigen al razonamiento estadístico-inferencial y, además, es reacio al cambio. Un ejemplo paradigmático de estas heurísticas es la de la *representatividad*¹; bajo esta heurística las personas evalúan el nivel de atipicidad de un evento, entre otros atributos, con base en qué tanto representa algunos aspectos salientes de la población o proceso del que proviene.
- *Un enfoque tradicionalista aplicado a la enseñanza estadística.* Al mismo tiempo, problemas prototípicos sobre contrastes de hipótesis encontrados en libros de texto suelen presentar de antemano todas las especificaciones de los supuestos necesarios para realizarlos, lo cual inhibe que los estudiantes participen en el proceso de traducir los aspectos contextuales de la situación a los modelos probabilísticos involucrados y las hipótesis que habrán de ser contrastadas. Más aún, la técnica del contraste suele introducirse de forma atomista y descontextualizada porque, generalmente, no hay vínculo directo a la intención de contribuir a resolver un problema más amplio que, en este caso, es de índole estadístico inferencial (i.e., cómo pasar de una muestra a la inferencia sobre una población o proceso).

¹ La falacia más conocida que se deriva de esta heurística es la *Ley de los Números Pequeños*, que consiste en asumir que inclusive muestras de tamaño pequeño deben reflejar de manera exacta las características de la población de donde provienen.



Para incidir en esta problemática, desde hace aproximadamente dos décadas que algunos investigadores comenzaron a explorar la idea de introducir las nociones que subyacen al concepto del contraste en niveles preuniversitarios, y con la intención de que los estudiantes generen las bases conceptuales para una progresiva formalización de sus conocimientos y razonamiento estadísticos. En esta tarea se ha dado un rol clave a los recursos tecnológicos, particularmente el uso de simulaciones aleatorias para simular procesos de muestreo, pues su uso flexibiliza los requerimientos conceptuales y procedimentales que se derivan del aparato matemático que sostiene a la técnica del contraste (Rossman, 2008; Batanero y Díaz, 2015). Sin embargo, el estado actual de la investigación en este problema brinda incipientes evidencias acerca de las verdaderas potencialidades y limitaciones para el aprendizaje del contraste cuando se utilizan este tipo de acercamientos didácticos (Lee, 2018).

Así, en este reporte describimos los resultados de un experimento de enseñanza, que sigue los principios de una investigación de diseño (Cobb, Confrey, diSessa, Lehrer y Schauble, 2003), en el que un grupo de estudiantes de bachillerato resolvió una serie de actividades didácticas que abordaron el concepto de distribución muestral empírica, utilizando simulaciones computarizadas para emular un proceso de muestreo repetido, para entonces enfrentarse a una situación introductoria al contraste de hipótesis. En particular, estamos interesados en documentar y caracterizar los patrones de razonamiento y los elementos matemáticos que despliegan ante la tarea introductoria al contraste, con el objetivo de aportar conocimiento acerca de las posibles causas de las dificultades para su aprendizaje y, potencialmente, dilucidar la influencia de los principales rasgos de nuestro diseño instruccional. De esta manera, nuestra pregunta de investigación consiste en: *¿qué patrones de razonamiento y elementos de significado manifiestan estudiantes de bachillerato ante una actividad diseñada para introducir el concepto del contraste de hipótesis y que utiliza un enfoque de muestreo repetido en un ambiente computacional?*

2. Antecedentes

Algunos de los estudios de más alto impacto han sido realizados con estudiantes del nivel superior, por ello describiremos brevemente algunos de sus resultados y luego otros de estudios realizados con estudiantes de niveles previos.

2.1 Estudios realizados en el nivel superior

La serie de estudios realizados por Vallecillos y colaboradores (Batanero, Godino, Vallecillos, Green & Holmes, 1994) reveló que los estudiantes interpretan conceptos involucrados en el contraste de manera considerablemente distinta a lo esperado normativamente. Por ejemplo, los autores encontraron que los estudiantes a menudo intercambian los dos términos de la probabilidad condicionada implicada en el nivel de significación (i.e., lo interpretan como la probabilidad de que la hipótesis nula sea verdadera, una vez que se tomado la decisión de rechazarla), interpretan el p-valor como la probabilidad de que los datos experimentales sean producto del azar, o asumen que el procedimiento del contraste posibilita calcular la probabilidad de que alguna hipótesis sea verdadera (una interpretación bayesiana de una metodología frecuentista).

Una creencia destacada en estos estudios fue la de considerar al contraste como una demostración matemática de la veracidad de cualquiera de las hipótesis involucradas, la cual también ha sido identificada en estudios más contemporáneos (Liu y Thompson, 2009; Batanero, Vera y Díaz, 2012; Inzunza y Jiménez,



2013). En años posteriores, Castro-Sotos et al. (2009) corroboraron que la propensión a caer en las interpretaciones mencionadas había sufrido un cambio mínimo pero favorable.

Varios investigadores han señalado que profesores y libros de texto promueven prácticas de enseñanza que no parecen corregir este tipo de creencias, sino que, desafortunadamente, más bien parecen reproducirlas (Harradine, Batanero y Rossman, 2011). Por ejemplo, Liu y Thompson (2009) evidenciaron que profesores de matemáticas del nivel bachillerato presentaron dificultades para concebir la atipicidad de un resultado estadístico bajo un enfoque estocástico, además de considerar que el rechazo de la hipótesis nula implicaba probar su falsedad. Por su parte, Batanero, López-Martín, Gea y Arteaga (2019) aplicaron un problema prototípico sobre contrastes de hipótesis utilizado en las pruebas de admisión para la universidad a una muestra de profesores de matemáticas del nivel secundario y de bachillerato, y encontraron que poco menos del 40% logró tomar la decisión correcta respecto al rechazo o aceptación de la hipótesis nula y sólo el 15% brindó respuestas sin errores procedimentales y conceptuales.

2.2 Estudios realizados en niveles preuniversitarios

En vista de las dificultades reportadas en la literatura acerca de la comprensión sobre el contraste de hipótesis, en años recientes se ha explorado la posibilidad introducir el concepto a través de un enfoque informal (Rossman, 2008; Batanero y Díaz, 2015). En general, en este enfoque se diseñan tareas basadas en el esquema de las pruebas de significación de Fisher en conjunto a la sustitución del modelo probabilístico de la distribución muestral implicada por su versión empírica. La intención principal para el aprendizaje es provocar que los estudiantes comiencen a familiarizarse con algunos procedimientos e ideas de la lógica subyacente a la técnica del contraste.

Dentro de esta línea de investigación, Lee, Angiotti y Tarr (2010) realizaron un experimento con estudiantes de secundaria (11 y 12 años de edad) para analizar su razonamiento en la tarea de inferir si un dado virtual estaba trucado o no, partiendo únicamente de analizar evidencias experimentales derivadas de simular su lanzamiento en repetidas ocasiones. Los autores señalan que, desde esta etapa, los estudiantes son capaces de traducir la hipótesis nula de que el dado sea justo a una distribución de probabilidad uniforme, lo cual les permite mapear los datos empíricos a dicho modelo y arribar eventualmente a inferir correctamente si el dado está equilibrado o no.

Por otra parte, Saldanha y Thompson (2002 y 2007) encontraron que estudiantes de bachillerato, al adentrarse en la lógica del contraste utilizando simulaciones computarizadas, tuvieron múltiples dificultades para interpretar un resultado estadístico típico o atípico bajo una concepción estocástica, en particular debido a los retos que supone comprender y utilizar el concepto de distribución muestral. Al igual que Chance, delMas y Garfield (2004), los autores señalan que el estudio de conceptos de la inferencia estadística, como el caso del contraste, se ve tremendamente limitado cuando los estudiantes son ajenos al concepto de distribución muestral.

Esta valoración es compartida por Sánchez, García-Ríos y Mercado (2017), quienes aplicaron una trayectoria de actividades didácticas a estudiantes de bachillerato para introducir el contraste de hipótesis desde una perspectiva informal. Sus resultados dan muestra de que los estudiantes de este nivel pueden mejorar considerablemente su desempeño en reproducir la técnica del contraste usando simulaciones computarizadas, y enfatizan que es necesario trabajar constante y cercanamente con ellos para evitar que



caigan en confusiones acerca del papel de la distribución muestral en el esquema del contraste y discutir el papel de la incertidumbre en la interpretación de su resultado.

3. Marco de referencia

3.1 Razonamiento estadístico y el contraste de hipótesis

Nuestro interés primario reside en el razonamiento de los estudiantes; interesa caracterizar los patrones de razonamiento que son observables en sus respuestas y comentarios a las tareas que configuran un primer acercamiento al concepto del contraste de hipótesis. Por razonamiento nos referimos al proceso de obtener o producir conclusiones sobre la base de evidencia y/o de proposiciones supuestas o ya establecidas (Shaughnessy, Chance y Kranendonk; 2009).

Dado que las actividades intelectuales o cognitivas tienen, en última instancia, el propósito de formular y establecer conclusiones verdaderas sobre determinada situación problemática (i.e. proposiciones), éstas implican la generación de razonamiento, por lo que, al conectar de diversas maneras los enunciados y evidencias nuevos con conocimientos ya establecidos, el razonamiento resulta en la *producción de sentido* o de *significado* sobre un objeto o concepto, el cual varía de individuo a individuo y en términos de diferentes niveles de complejidad cognitiva, argumentativa, de formalidad y de abstracción.

Asimismo, el objeto central sobre el que interesa desarrollar significado es el del contraste de hipótesis. A diferencia de la propuesta de Fisher, el concepto del contraste se concibe como una regla de decisión entre dos hipótesis, la nula o alternativa, como el modelo probabilístico más plausible de haber producido los datos experimentales. Para ello se definen, en la distribución muestral del estadístico (también llamado “estadístico de prueba”), regiones de rechazo y aceptación de la hipótesis nula, el nivel de significancia, los errores tipo I y II y la probabilidad de cometer cada uno.

En el esquema del contraste, el nivel de significancia se selecciona previo a realizar la experimentación, lo cual delimita las regiones de rechazo y no rechazo de la hipótesis nula. Si el valor del estadístico cae en la región de rechazo, la hipótesis nula es rechazada; en caso contrario, la hipótesis nula no es rechazada y se acepta su complemento (la hipótesis alternativa). Para introducir a los estudiantes de bachillerato al concepto del contraste, hemos planteado una situación problema en la que deben tomar una decisión contextual basada en la selección (o aceptación) de la hipótesis nula o alternativa, además de enfocarse (o no) en la noción de región de rechazo / de aceptación de la hipótesis nula como un recurso clave dentro de la propuesta de solución.

3.2 Elementos teóricos del Enfoque Ontosemiótico de la Cognición Matemática (EOS)

Nuestro segundo componente teórico proviene de la teoría EOS (Godino, 2002), la cual parte de supuestos pragmáticos (los objetos matemáticos emergen de la actividad de resolución de problemas) y realistas (los objetos tienen un carácter ostensible y se sitúan en determinada realidad cultural) acerca del significado de los objetos matemáticos. En el EOS, el significado de un objeto matemático se constituye progresivamente a partir del *sistema de prácticas* en las que se involucra. Tales sistemas, así como los campos de problemas de los que se desprenden, varían entre los considerados grupos de expertos y las instituciones educativas que llevan a cabo procesos de enseñanza y aprendizaje matemático (*faceta institucional*). A su vez, se reconoce



que cada estudiante desarrolla significado en función de su propia experiencia con los sistemas de práctica y campos de problema que haya enfrentado hasta determinado momento de su formación escolar (*faceta personal*).

Para el EOS, distintos elementos intervienen en la actividad de resolución de problemas matemáticos; en particular, utilizaremos los llamados *elementos primarios del significado*, que plantean una tipología de cinco componentes para dar cuenta sobre el significado (*personal*) que los estudiantes desarrollan, en este caso, acerca del objeto del contraste de hipótesis. Tales elementos son desplegados, consciente o inconscientemente, al resolver un problema matemático determinado y permiten identificar el bagaje y/o desarrollo de nuevo conocimiento matemático en función de los siguientes objetos:

- *Situaciones problema* (elementos fenomenológicos). Se refiere al campo de problemas del cual emerge progresivamente el concepto, es decir, las situaciones que inducen a la activación del empleo del razonamiento matemático con los objetos de interés y que ultimadamente configuran su significado. Por ejemplo, el contraste de hipótesis puede realizarse para el caso de una media o proporción muestral, o tomando en consideración sólo una de las dos colas de la distribución muestral del estadístico.
- *Lenguaje* (elementos representacionales). Se refiere a las representaciones ostensivas de los conceptos involucrados. En el contraste se manipula y opera con diversas representaciones simbólicas (\hat{p} , H_0 y H_1 , α , β), gráficas (curva suave para la distribución muestral del estadístico, región de rechazo y aceptación) y verbales (“rechazar / aceptar H_0 ”).
- *Acciones* (elementos procedimentales). Se refiere a los procedimientos específicos realizados por el estudiante al resolver la situación problema; estos pueden ir desde operaciones sencillas hasta algoritmos y procedimientos de cálculo que se realizan en varias etapas y que relacionan distintos dominios matemáticos. Por ejemplo, una primera acción a realizar para resolver un contraste es la determinación de las hipótesis nula y alternativa, así como la selección de una variable aleatoria que se corresponda con la distribución muestral del estadístico de prueba.
- *Argumentos* (elementos de validación). Se refiere a las argumentaciones utilizadas para justificar y explicar las acciones y procedimientos que se siguieron para arribar a una determinada solución. Por ejemplo, en el contraste se rechaza la hipótesis nula *justificando* que el valor del estadístico se considera demasiado atípico de obtener bajo el supuesto de que es verdadera (i.e., se encuentra en la región de rechazo).
- *Conceptos y propiedades* (elementos conceptuales). El primer elemento se refiere a los conceptos y definiciones que se utilizan para resolver la situación problema, considerando en algunos casos estos son referidos de manera indirecta; para el caso del contraste, ejemplos son los conceptos de muestra, hipótesis estadística, distribución muestral, errores tipo I y II, etcétera. El segundo elemento se refiere a las *proposiciones teóricas* que sirven como sustento para llevar a cabo una determinada acción, o bien, que emergen de la producción de sentido; por ejemplo, una propiedad de las hipótesis estadísticas dentro del esquema del contraste es que se representan mediante una distribución muestral, o que el resultado de la prueba puede alterarse si se modifica el tamaño de muestra.

Es importante considerar que el significado personal (*logrado*) de los estudiantes no necesariamente se corresponde con el institucional (*pretendido*), pero es precisamente el análisis de esta diferencia lo que posibilita identificar errores en el razonamiento matemático de los estudiantes y, potencialmente, informar y reorientar el proceso del diseño instruccional.



4. Método

4.1 Participantes

La metodología del estudio es de carácter cualitativo. En el experimento participaron 42 estudiantes, agrupados en 21 parejas, que se encontraban cursando un segundo curso de Estadística y Probabilidad II (CCH, 2016) durante su último semestre de bachillerato (17-18 años de edad). Los antecedentes estadísticos y probabilísticos más recientes de los estudiantes se corresponden con los contenidos del curso de Estadística y Probabilidad I (introducción al análisis exploratorio de datos e introducción a los enfoques de la probabilidad), llevado a cabo en el semestre previo. El profesor responsable del grupo y del experimento de enseñanza fue el primer autor de este reporte.

Los estudiantes trabajaron con el concepto de distribución muestral empírica durante las semanas previas a enfrentarse a la situación del contraste. Dicho concepto, que se construye mediante simulaciones computarizadas que emulan un proceso de muestreo repetido (Chance et al., 2004), fue el objeto central de aprendizaje durante una trayectoria de seis actividades didácticas, las cuales a continuación se enlistan de manera somera:

- Introducción al concepto de distribución muestral: (1) partiendo de un parámetro p desconocido, construir una distribución muestral empírica de la variable “número de éxitos en la muestra” (distribución binomial) a través del muestreo repetido usando simulaciones computarizadas, para caracterizarla a través de su forma, centro y nivel de variabilidad; (2) estimar probabilidades usando la distribución muestral cuando se conoce p .
- Rol del tamaño de muestra en la distribución muestral: (1) partiendo de un parámetro p conocido, identificar los efectos de modificar el tamaño de muestra en la distribución muestral (acercamiento al Teorema del Límite Central); (2) estimar probabilidades usando la distribución muestral y modificando el tamaño de muestra.
- Acercamiento a la formalización del concepto de distribución muestral: (1) partiendo de un parámetro p desconocido y una distribución muestral generada con dicho parámetro, estimar probabilidades y el valor de p ; (2) partiendo de un parámetro p conocido, bosquejar la distribución muestral correspondiente.

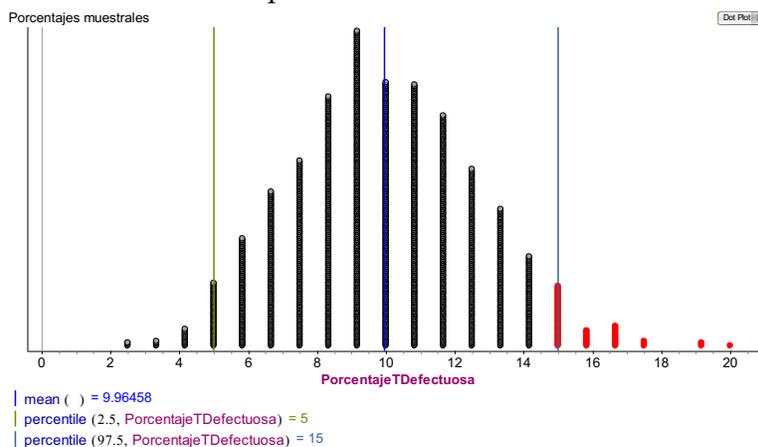
Los estudiantes trabajaron con el software *Fathom* para resolver los problemas de la trayectoria, ya que el ambiente y herramientas virtuales que proporciona, en particular las que permiten simular procesos de muestreo de manera intuitiva y en un entorno amigable, son ampliamente recomendadas para asistir procesos de educación estocástica (Biehler, Ben-Zvi, Bakker y Makar; 2013). En general, estos recursos fueron empleados en las actividades de la trayectoria para promover que los estudiantes asociaran la noción de variabilidad muestral (el estadístico puede tomar distintos valores) con el concepto de la distribución muestral (pero unos son más probables de ocurrir que otros). En la actividad que aquí reportamos no se requirió que los estudiantes manipularan el programa, pero aun así requirieron apoyarse en su experiencia con el mismo dado que fue en este entorno donde construyeron su significado del concepto de distribución muestral (empírica).

Adicionalmente, un antecedente relevante para los estudiantes al momento de enfrentarse a la situación introductoria al contraste fue un ítem que se presentó en los tres bloques de actividades y que consistió en, dada una distribución muestral empírica, clasificar un valor muestral específico como probable/típico o



atípico de obtener. Los estudiantes, en conjunto con el profesor, acordaron realizar esta clasificación delimitando un rango de valores alrededor del promedio de la distribución, de tal manera que abarcara el 95% de todos los valores que la conforman. La siguiente imagen ilustra este rango de valores probables o típicos al centro de la colección para una distribución muestral empírica de 2 mil valores del estadístico del porcentaje muestral, con parámetros $n = 120$ y $p = .1$:

Figura 1. Distribución empírica de 2 mil valores construida en Fathom



4.2 Instrumentos y procedimiento de ejecución

El experimento se llevó a cabo en una sola sesión de clase de aproximadamente una hora y media de duración. En el diseño del experimento se contemplan tres momentos; en el primero, agrupados los estudiantes en parejas, se les planteó la siguiente situación problema:

Una fábrica dedicada a la generación de componentes para computadora posee cuatro máquinas especializadas en producir tarjetas madre para laptop. De manera irremediable, cada máquina produce, de forma *aleatoria*, una cierta cantidad de tarjetas defectuosas en determinado período de tiempo. El departamento de control y 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*. Ahora bien, un equipo de técnicos recolectó en cierto día una muestra aleatoria de 120 tarjetas de cada máquina, para estimar el % de tarjetas defectuosas que produce, obteniéndose los siguientes resultados: máquina A - 42 tarjetas defectuosas; máquina B - 21; máquina C - 27 y máquina D - 15. Con base en esta información, ¿cuáles máquinas consideras son necesarias mandar a revisión²?

Cada pareja elaboró una propuesta de solución al problema y sin ningún tipo de ayuda por parte del profesor-investigador; sus respuestas fueron registradas en una hoja de papel que entregaron al profesor antes de pasar al segundo momento de la actividad.

La respuesta esperada al problema debería iniciar, según el punto de vista normativo, asumiendo que cada máquina produce un 10% de tarjetas defectuosas y entonces preguntarse por la *atipicidad* de obtener el porcentaje muestral correspondiente o algo más extremo. No obstante, anticipando que los estudiantes tendrían un abordaje de la situación relacionado con la heurística de la representatividad (i.e., su decisión se

² Sólo para el caso de la máquina D se acepta la hipótesis nula de que $p = .1$ con un $\alpha = .025$, en el resto de los casos se rechaza.

basaría únicamente en la cercanía del estadístico con el 10% permitido), en el segundo momento de la actividad el profesor utilizó un simulador de porcentajes muestrales, donde el parámetro poblacional era conocido, para que visualizaran (o recordaran) que, debido a la variabilidad muestral, el porcentaje muestral puede ser considerablemente distinto al 10% permitido inclusive si $p = .1$. El propósito de esta acción era inducir a los estudiantes a que relacionaran la variabilidad muestral del estadístico con el concepto de distribución muestral, para entonces medir su grado de atipicidad asumiendo la hipótesis nula de que la máquina producía 10% de tarjetas defectuosas. Hecho esto, los estudiantes tuvieron la oportunidad de replantear sus respuestas a la situación del contraste, registrando sus respuestas en otra hoja de papel que fue entregada al profesor.

En el tercer momento de la actividad, el profesor mostró a los estudiantes una distribución muestral empírica correspondiente a una máquina Z que *se sabe con certeza* produce un 10% de tarjetas defectuosas (ver Figura 1). En esta distribución se grafican y presentan dos valores del estadístico que delimitan el rango que contiene el 95% de los valores simulados y que están alrededor del promedio de la distribución. Debido a sus experiencias recientes con el concepto de distribución muestral empírica, los estudiantes estaban familiarizados con este rango de valores y por esta razón no fue modificado³. A partir de esta distribución muestral, los estudiantes realizaron una tercera y última propuesta de solución a la situación del contraste, registraron sus respuestas en una nueva hoja de trabajo y la entregaron al profesor.

4.3 Datos y procedimiento de análisis

La fuente primaria de evidencias se constituye de las hojas de trabajo que los estudiantes entregaron en cada momento de la actividad. El análisis de datos se realizó en dos etapas; en la primera se realizó una codificación abierta de las respuestas de los estudiantes tomando como categorías de análisis sus acciones y argumentos frente a la situación problema y de acuerdo con cada momento de la actividad. A partir de este análisis se caracterizaron algunos patrones de respuesta que aluden, desde nuestra perspectiva, a modos de razonamiento intrínsecos a las capacidades de nuestros estudiantes y la micro-cultura de la clase. En la segunda etapa de análisis se utilizaron los elementos primarios del significado para analizar las respuestas de los estudiantes y así inferir su significado personal acerca del objeto del contraste.

5. Resultados

5.1 Patrones de razonamiento según los momentos de la tarea

En el primer momento de la actividad se espera que los estudiantes asuman que cada máquina presenta un 10% de tarjetas defectuosas y entonces se pregunten por el grado de atipicidad en la muestra respecto al resto de los valores del estadístico que podrían haberse obtenido; si se considera al valor del estadístico como atípico se acepta provisionalmente la hipótesis de que la máquina excede el 10% de tarjetas defectuosas. No obstante, los patrones de respuesta caracterizados en esta etapa evidencian que la mayoría de los estudiantes no parece considerar que el porcentaje muestral, de repetirse el experimento, podría modificarse hasta quedar por debajo del 10% independientemente del verdadero porcentaje de tarjetas defectuosas que produzca:

³ La solución normativa considera sólo el extremo superior de la distribución muestral. Es decir, la situación corresponde a una prueba de una cola porque sólo interesa analizar si los valores muestrales son atípicamente elevados asumiendo H_0 como $p = .1$.



Tabla 1: Patrones de respuesta en el primer momento de la tarea

Código	Acción y argumento	Frecuencia
Determinista: no identifica variabilidad muestral;	Todas las máquinas a reparación porque exceden el 10%;	14 parejas
Semideterminista: intuye variabilidad muestral	A y C a revisión, exceden por mucho el 10%; casos B y D son ambiguos	7 parejas

Esto es, 14 parejas evidenciaron una indiferencia hacia la variabilidad muestral del estadístico, por lo cual se considera siguieron un razonamiento determinista ya que tomaron la decisión de enviar todas las máquinas a *reparación* bajo el argumento de que el valor muestral excedió el 10% permitido de tarjetas defectuosas. Por ejemplo, la pareja P8 señaló: “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”.

Las siete parejas restantes siguieron un razonamiento semideterminista al reconocer la existencia de un cierto *margen de error* en el que puede oscilar el valor del estadístico respecto al 10% esperado; este razonamiento implica asumir como hipótesis nula que la máquina en cuestión presenta un máximo de 10% de tarjetas defectuosas. Esta línea de argumentación se aplicó en los casos de las máquinas D y B por haber presentado los porcentajes muestrales más cercanos al 10% permitido. Por ejemplo, la pareja P2 mencionó que:

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 el segundo momento de la actividad, los estudiantes tuvieron la oportunidad de reformular su respuesta a partir de observar cómo varía una proporción muestral cuando se conocen el tamaño de muestra y la proporción poblacional de donde se extraen (pero sin aludir al concepto de distribución muestral). Finalizada esta actividad, las respuestas a la situación del contraste se reformularon como:

Tabla 2: Patrones de respuesta en el segundo momento de la actividad

Código	Acción y argumento	Frecuencia
Determinista: no identifica variabilidad muestral;	Todas las máquinas a reparación porque exceden el 10%;	3 parejas
No determinista: identifica variabilidad muestral	B y/o D dentro del “margen de error aceptable”, no se envían a reparar	18 parejas

La casi totalidad de los estudiantes, con 18 de las 21 parejas, consideró ahora que las máquinas con el menor número de tarjetas defectuosas cumplían con estar dentro de un cierto margen de error o variación aceptable con respecto al 10% permitido, preservando así la idea de asumir temporalmente que cualquier máquina producía un 10% de tarjetas defectuosas. Esta consideración de la variabilidad muestral llevó a clasificar este razonamiento como no determinista; adicionalmente, como puede apreciarse en la respuesta de P20, ningún estudiante evocó de forma alguna el concepto de distribución muestral para modelar la forma en la que varía el estadístico del porcentaje muestral:

Sí, en la máquina B no la mandaríamos a arreglar porque el simulador nos hizo recordar la variabilidad que puede haber en la muestra y no sobrepasa por mucho el límite establecido, ya que puede bajar o subir el % de



tarjetas defectuosas siempre y cuando no sobresalga mucho. En la D considerando lo anterior no la mandaríamos a arreglar.

Estos patrones de respuesta sugieren que, hasta este punto de la actividad, el razonamiento de los estudiantes parece apoyarse en la heurística de la representatividad como recurso único para enfrentarse el problema. En otras palabras, bajo esta aproximación, los estudiantes se limitan a juzgar la representatividad de las muestras con base en su parecido o diferencia con el 10% esperado (según la hipótesis nula).

En el tercer momento de la actividad, los estudiantes se confrontaron por última vez con la pregunta del contraste disponiendo de la simulación de una distribución muestral empírica de porcentajes obtenida de la máquina Z, la cual se sabe con certeza presenta 10% de artículos defectuosos (ver Figura 1). En este escenario los patrones de respuesta se caracterizaron como:

Tabla 3: Patrones de respuesta en el tercer momento de la actividad

Código	Acción y argumento	Frecuencia
Acercamiento intuitivo al contraste	D no se envía a reparar, es un valor probable de obtener bajo la hipótesis nula;	10 parejas
	D no se envía a reparar, es un valor probable de obtener bajo la hipótesis nula y es, además, cercano al valor esperado	4 parejas
Razonamiento a la inversa	Sólo D se envía a reparar, es un valor probable de obtener bajo la hipótesis nula;	3 parejas
Heurística de representatividad	D no se envía a reparar, su valor es cercano al esperado	4 parejas

10 de las 21 parejas tomaron la decisión correcta de rechazar la hipótesis nula de que la máquina correspondiente producía un 10% de tarjetas defectuosas. Para ello se basaron en identificar en qué región se encontraban los distintos valores del estadístico: si el valor se encontraba fuera del rango de valores considerados como probables o usuales de obtener (en este caso en la cola derecha de la distribución), entonces la hipótesis nula del 10% fue rechazada. Un ejemplo de estas respuestas es la de P10:

Mandaríamos a reparar las máquinas A, B y C porque los % de tarjetas defectuosas de estas máquinas quedan fuera de los valores críticos (representan sólo el 2.5% de todos los valores obtenidos). La máquina D sería la única que no mandaríamos a reparar debido a que su % de tarjetas defectuosas entra en los valores críticos que representan el 95% de todos los valores obtenidos.

Este tipo de razonamiento se toma como indicio de que los estudiantes fueron receptivos a razonar intuitivamente con la lógica del contraste en (por lo menos) esta etapa del proceso. Por otro lado, cuatro parejas aplicaron este razonamiento y complementaron su respuesta con la heurística de la representatividad. Esto es, para decantarse por una u otra hipótesis, los estudiantes evaluaron la distancia que hay entre el valor del estadístico y el valor esperado según la hipótesis nula: si el porcentaje muestral se considera como demasiado lejano del esperado (10%), entonces la hipótesis de que produce un 10% de tarjetas defectuosas es rechazada. La respuesta de P3 ilustra este tipo de razonamiento:



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.

Pese al uso de esta heurística, consideramos que este tipo de respuestas preservan un acercamiento intuitivo a la lógica del contraste porque se valen de argumentos basados – al menos parcialmente – en la distribución muestral empírica del estadístico y la idea de considerar si su valor es probable de obtener bajo la hipótesis nula. Aunque la decisión de tres parejas más fue errónea por haber invertido la decisión de aceptar la hipótesis nula para la máquina D y viceversa en los casos restantes, clasificamos sus respuestas dentro de la misma categoría por haberse enfocado en la distribución muestral empírica y en el rango de valores probables o usuales de obtener.

Las cuatro parejas restantes no utilizaron argumentos como los anteriores y, pese a tomar la decisión correcta de aceptar o rechazar la hipótesis nula en cada caso, su razonamiento se basó únicamente en la heurística de la representatividad; la respuesta de P13 ilustra este tipo de argumentación:

Mandaríamos a revisar las máquinas A, B y C ya que el % que se observa en las muestras es muy elevado lejano al 10% permitido. La máquina D no la mandaríamos a revisar ya que el 12.5% que se observa en la muestra es muy cercano al 10%. Después de realizar más muestras el valor se mantendría muy cercano a este 10%, caso contrario de las A, B y C que está muy alejado.

Finalmente, como parte del análisis de la formulación de la conclusión, conviene destacar que a partir del segundo momento de la actividad, los estudiantes intercambiaron el verbo “revisar” por “reparar”, sugiriendo el haber asumido (implícitamente) que si el porcentaje muestral se encuentra fuera del rango de valores probables o usuales de obtener entonces se *asegura* que el porcentaje de tarjetas defectuosas de la máquina excede el 10% permitido. Más aún, el análisis de las respuestas en el último momento de la actividad reveló una falta de expresiones que aludan a una incertidumbre respecto al hecho de que las máquinas enviadas a reparación no necesariamente presentan una descompostura; análogamente, tampoco se encontraron expresiones que denotaran la consideración de que la máquina D podría exceder el 10% de tarjetas defectuosas. Una vez que se entregaron las últimas hojas de trabajo, el profesor-investigador preguntó a los estudiantes: “¿no sería posible que su conclusión (enviar máquinas A, B y C a “reparar” porque exceden el 10% permitido), sea incorrecta? Inclusive si no hay “fallas” en el procedimiento...” Todos los estudiantes afirmaron que el procedimiento que habían seguido permitiría *determinar* si la máquina tenía una descompostura o no.

5.2 Elementos del significado

5.2.1 Acciones y lenguaje

En el primer momento de la actividad, los estudiantes primero traducen el estadístico “número de tarjetas defectuosas en la muestra de $n=120$ ” al estadístico del porcentaje muestral, después comparan cada porcentaje muestral con el porcentaje límite de tarjetas defectuosas (10%) y descartan la hipótesis nula para los casos en que se considera demasiado alejado del límite permitido (máquinas A y C). En este punto, sólo una tercera parte de los estudiantes realiza una acción adicional, de naturaleza analítica, en la que consideran que la producción de tarjetas defectuosas tiene un carácter aleatorio; ello permite flexibilizar su abordaje determinista para analizar los casos de las máquinas B y C. Además de los cálculos y representaciones



numéricas que se realizan en la obtención de los porcentajes muestrales, el lenguaje utilizado por los estudiantes es únicamente verbal.

En el segundo momento de la actividad, en donde los estudiantes traen de vuelta la idea de que los porcentajes muestrales no necesariamente reflejan con exactitud el parámetro que los genera, la casi totalidad del grupo realiza la acción, también de naturaleza analítica, de concebir un “margen de error” en el que el porcentaje muestral podría presentarse (i.e., el rango de valores que podría tomar el estadístico) asumiendo implícitamente que la máquina en cuestión podría producir un 10% de tarjetas defectuosas. Quienes se apoyan en la heurística de la representatividad no parecen realizar acciones adicionales a las descritas en el momento anterior de la actividad. Hasta este punto el lenguaje utilizado por los estudiantes continúa siendo mayoritariamente de carácter verbal.

En el tercer momento de la actividad, la primera acción que realizan los estudiantes es enfocarse en la interpretación de la distribución muestral empírica que representa la hipótesis nula de que la máquina produce un 10% de tarjetas defectuosas (máquina Z), en tanto que sólo un grupo minoritario de estudiantes se apoya exclusivamente en la heurística de la representatividad para tomar la decisión de rechazar o aceptar la hipótesis nula (evalúa subjetivamente qué tan distante es el porcentaje muestral del 10% permitido). Como segunda acción, del grupo de estudiantes que se enfoca en la distribución muestral, la mayoría identifica en qué región de la distribución se encuentra cada porcentaje muestral (i.e., dentro o fuera de la región de valores probables o usuales de obtener) para tomar la decisión, en su mayoría correcta, de aceptar la hipótesis nula sólo en el caso de la máquina D. Algunos estudiantes complementan la acción de identificar en qué región se encuentra el porcentaje muestral con la de aplicar la heurística de representatividad mientras que otros, pese a haber identificado que el valor muestral es probable o usual de obtener, invierten la decisión de rechazar la hipótesis nula y viceversa para los otros casos. De una u otra forma, todos los estudiantes continúan apoyándose en un lenguaje verbal que incorpora escasas representaciones numéricas y simbólicas (porcentajes) como parte de su argumentación.

5.2.2 Argumentos y propiedades

La siguiente tabla exhibe nuestras conjeturas acerca de los principales argumentos y propiedades que los estudiantes utilizaron, de acuerdo con cada momento de la actividad, para justificar su valoración acerca de cuáles máquinas podrían producir más del 10% de tarjetas defectuosas:

Tabla 4: Argumentos y propiedades para cada momento de la actividad

Momento	Argumento	Propiedad
1	El porcentaje muestral excede el 10%, entonces hay descompostura	Heurística de la representatividad: las muestras presentan valores iguales (o muy cercanos) al parámetro que las genera
2	El porcentaje muestral puede moverse en cierto rango de valores;	Debido a la aleatoriedad, las muestras varían alrededor del parámetro que las genera;
	El porcentaje muestral se sale (o no) del “margen de error”	Heurística de la representatividad



3	<p>El porcentaje muestral se encuentra (o no) en el rango de valores probables / usuales de obtener;</p> <p>El porcentaje muestral se sale (o no) del “margen de error”</p>	<p>Si el porcentaje muestral es atípico bajo la hipótesis nula, entonces debe ser falsa y viceversa;</p> <p>Heurística de la representatividad</p>
---	---	--

Respecto al primer momento de la actividad, consideramos que el argumento primario de los estudiantes se nutre directamente de la heurística de la representatividad. Al comparar el valor del porcentaje muestral con el de la hipótesis que se quiere refutar, y “verificar” que ésta se cumple, los estudiantes asumen implícitamente la creencia de que las muestras presentan valores iguales (o muy cercanos) al parámetro que las genera; esto conduce a que ignoren el escenario en el que el porcentaje muestral podría obtenerse inclusive si la máquina no excede el 10% de tarjetas defectuosas.

En el segundo momento de la actividad, la mayoría de los estudiantes son receptivos a traer de vuelta la noción de variabilidad muestral al contexto de la situación y por tanto argumentan que la máquina D no necesitaría una reparación, siempre y cuando su porcentaje muestral “no se aleje demasiado de un margen de error (variabilidad) aceptable alrededor del 10%” permitido. Dicha argumentación evidencia la creencia de los estudiantes en la propiedad de que, al tomar una muestra aleatoriamente, el valor del estadístico debería encontrarse alrededor del parámetro poblacional que la genera, relajando entonces la expectativa de que ambos valores deberían ser idénticos.

Es importante observar que hasta este punto de la actividad los estudiantes no consideran necesario precisar los límites del margen de variabilidad en el estadístico que sería razonable de obtener. Aunque el razonamiento de los estudiantes ya no es tan determinista como en el primer momento de la tarea, sus argumentos no aluden de forma alguna al concepto de distribución muestral para especificar cómo es que se comporta la variabilidad del estadístico, por lo que el argumento que alude a la heurística de la representatividad sigue siendo de fuerte influencia.

En el tercer momento de la actividad, la mayoría de los estudiantes basa su argumentación en propiedades que extraen de la distribución muestral empírica que representa la hipótesis nula. La propiedad que les permite tomar una decisión sobre aceptar o rechazar la hipótesis nula, para la mayoría, consiste en asociar un evento poco probable (casos de las máquinas A, B y C) con un evento imposible; esto les lleva a asumir que si el porcentaje muestral se considera poco probable de ocurrir, entonces la hipótesis nula de que la máquina produce 10% de tarjetas defectuosas debe ser falsa y por tanto su complemento sería verdadero. Por otra parte, en este punto de la actividad sólo una minoría de los estudiantes presenta argumentos que se desprenden de la heurística de la representatividad (i.e., que ignoran la información que provee la distribución muestral y continúan asumiendo de manera intuitiva que las muestras presentan nula o muy poca variabilidad muestral).

5.2.3 Conceptos y situación problema

La siguiente tabla exhibe una serie de conceptos utilizados por los estudiantes y de acuerdo con cada momento de la actividad. Algunos de estos conceptos son referenciados, desde nuestra perspectiva, de manera implícita o indirecta:

Tabla 5: Conceptos referenciados durante la actividad



Momento	Conceptos
1	Muestra, proporción y porcentaje muestral, margen de error (variabilidad muestral);
2	Muestra, parámetro, proporción y porcentaje muestral, margen de error, simulación del muestreo;
3	Muestra, parámetro, proporción y porcentaje muestral, margen de error, simulación del muestreo, distribución muestral empírica, porcentaje muestral probable/atípico, región de rechazo/aceptación

En el primer momento de la actividad, la casi totalidad de los estudiantes hace un uso muy limitado de los conceptos estadísticos y probabilísticos que se esperarían observar en un contraste de hipótesis. En efecto, verificar que el porcentaje muestral se corresponde con el porcentaje esperado del 10% requiere, para la mayoría, una cantidad muy limitada de esfuerzo cognitivo y procedimental, en tanto que el grupo minoritario considera indirectamente la presencia del componente aleatorio en la producción de tarjetas y por ello incorpora el concepto de margen de error (variabilidad muestral).

Para el segundo momento de la actividad, el número de conceptos evocados por los estudiantes sufre un cambio mínimo respecto al anterior ya que del simulador identifican y abstraen, de forma más explícita, los conceptos de parámetro, variabilidad muestral (i.e., variabilidad del estadístico) y simulaciones del muestreo. Pero una diferencia clave entre ambos momentos es que, a diferencia del primero, la mayoría de los estudiantes se aleja de abordar la situación desde un punto de vista meramente determinista.

Para el tercer momento de la actividad, la mayoría de los estudiantes incorpora a su razonamiento los conceptos de distribución muestral, rango de valores probables o usuales de obtener, porcentaje muestral probable y atípico y regiones de aceptación y rechazo. Ningún estudiante parece intuir el concepto de error tipo I o II, entendido como aquel que se derivaría de aceptar o rechazar una hipótesis cuando esta puede ser ultimadamente falsa o verdadera según sea el caso.

De esta forma, podemos inferir que la interpretación de los estudiantes acerca de la situación problema cambia dependiendo del momento de la actividad en la que se encuentren, pero preservando la idea de que el procedimiento tiene la intención última de establecer la veracidad de las hipótesis involucradas. Esto es, en el primer momento, es más probable que los estudiantes conciban la situación problema como una que puede presentarse en un curso de matemáticas en general, ya que sobreentienden que se solicita realizar una operación aritmética sencilla correspondiente a la verificación de la hipótesis que se quiere contrastar. En el segundo momento, toda vez que los estudiantes han incorporado la noción de variabilidad muestral al tratamiento de la situación, es más probable la conciban como una especie de prueba de tolerancia hacia la variabilidad del porcentaje muestral respecto al 10% esperado, pues se valen del recurso de la heurística de la representatividad para tomar la decisión de aceptar o rechazar (implícitamente) la hipótesis nula de que la máquina presente descompostura. Y en el tercer momento, aunque la mayoría de los estudiantes toma la decisión correcta de aceptar o rechazar la hipótesis nula de que la máquina produce 10% de tarjetas defectuosas – inclusive apoyándose en una versión informal del concepto de regiones de rechazo y aceptación dentro de la distribución muestral –, la mayoría termina por desterrar la noción de incertidumbre de su razonamiento pues concibe que el procedimiento y la lógica que se han seguido hasta el momento, de manera similar a cualquier procedimiento atendido en una clase de matemáticas, permiten dar una respuesta



correcta y certera al problema del contraste (i.e., permiten establecer la veracidad o falsedad de cualquiera de las hipótesis involucradas).

6. Conclusiones y discusión

En respuesta a nuestra pregunta de investigación, un primer patrón de razonamiento a destacar es la tendencia de nuestros estudiantes a valerse mayoritariamente de un razonamiento idiosincrático acerca de la relación parámetro-estimador cuando resuelven el problema de manera independiente. Bajo este modo de razonamiento, el apoyo preponderante en la heurística de la representatividad les conduce a basar su decisión acerca de aceptar o rechazar la hipótesis nula con base en el parecido del valor muestral con el del valor esperado. Pero arribado al punto en el que los estudiantes disponen de la distribución muestral que materializa la hipótesis nula, un segundo patrón que se destaca en su razonamiento es la alta propensión para utilizar su concepción de valor muestral atípico para asociarla con la imposibilidad de que la hipótesis nula sea cierta.

En lo que respecta a los elementos del significado, nuestros resultados señalan que éstos dan un salto cualitativo en el razonamiento de los estudiantes sólo hasta que se les sitúa en el contexto de interpretar y utilizar la distribución muestral empírica que modeliza la hipótesis nula. Previo a este punto, los elementos del significado que los estudiantes despliegan se rigen principalmente por la manipulación de los conceptos de muestra, parámetro y margen de error, que se encuentran desligados de los conceptos de distribución muestral y los de hipótesis nula y alternativa, y con los cuales se razona bajo la lógica de la heurística de la representatividad. Por esta razón, el lenguaje, acciones y propiedades que utilizan se limitan a los necesarios para valorar la representatividad de una muestra con base en su parecido al parámetro supuesto, algo muy distinto de valorarla en términos de la comparación hacia un conjunto de valores del estadístico (Saldanha y Thompson, 2007).

A partir de que los estudiantes disponen de la distribución muestral que modeliza la hipótesis nula, incorporan nociones conceptuales relacionadas con los conceptos de hipótesis nula y alternativa, valor muestral típico o atípico y las regiones de rechazo y aceptación de la hipótesis nula. En particular, el utilizar una proto-versión de las regiones de rechazo y aceptación para tomar una decisión sobre las hipótesis, en su mayoría correcta, permitió a los estudiantes conjeturar una nueva proposición acerca de la relación muestra-parámetro: un valor muestral inusual o atípico hace – por lo menos – insostenible el supuesto de que la hipótesis nula es verdadera, en tanto que, análogamente, uno típico lleva a la decisión de – por lo menos – no rechazarla.

Por otra parte, algunos de nuestros resultados coinciden con otros ya reportados en la literatura especializada; en particular, Sánchez y colaboradores (2017) también habían reportado la tendencia de los estudiantes a brindar respuestas iniciales a la situación del contraste que consisten en verificar la hipótesis que se desea contrastar, además de asumir que la técnica del contraste puede ser vista como una demostración matemática de la veracidad de las hipótesis (Batanero et al., 1994; Castro-Sotos et al., 2009). A un nivel más general, la prevalencia de este tipo de razonamientos puede ser explicada, en parte, por la falta de experiencias escolarizadas de los estudiantes con los conceptos de muestreo e inferencia durante su formación inicial, además de haber pasado por años de instrucción matemática que difícilmente fomenta el razonamiento con las principales ideas estadísticas y probabilísticas que subyacen a los conceptos y técnicas de la inferencia frecuentista (Burril y Biehler, 2011; Harradine et al., 2011).



En este sentido, una de las implicaciones principales de nuestro estudio para la enseñanza es el reconocimiento de que cuando los estudiantes se familiarizan con el concepto de distribución muestral empírica, pueden estar mejor equipados en el momento específico de utilizarla para tomar una decisión apropiada dentro del mecanismo y lógica del contraste. Esto es, si bien es notorio que un desarrollo de significado más robusto acerca del concepto requiere de más experiencias y prácticas con el mismo, la actuación de nuestros estudiantes sugiere que el recurso que les ha resultado más útil, respecto a la solución normativa del contraste, fue el ver a la distribución muestral empírica como un vehículo para clasificar los resultados como probables o no probables de obtener, prefigurando así la noción de regiones de rechazo y aceptación de la hipótesis nula. Conjeturamos que es probable que el paso por la trayectoria de actividades que promueve prácticas con esta noción les haya empujado a utilizarla para enfrentarse a la situación introductoria al contraste.

Nuestros resultados también evidencian que las etapas inicial y final de la técnica del contraste presentan retos potencialmente ineludibles al razonamiento de los estudiantes, ya que requieren del uso de un razonamiento hipotético-deductivo para el planteamiento de la hipótesis nula, su apropiada modelización a través de una distribución muestral, así como la consideración del rol de la noción de la incertidumbre especialmente a la hora de interpretar correctamente el resultado del contraste. La prevalencia de errores conceptuales en estas etapas sugiere que podrían constituir verdaderos conflictos epistemológicos para los estudiantes en general (Batanero, 2000); por consiguiente, se esperaría que la serie de habilidades y modos de razonamiento necesarios para enfrentarlos se desarrolle, probablemente, en años, no semanas.

Finalmente, consideramos que algunas vertientes apropiadas para continuar la presente investigación son:

- Caracterizar y profundizar en el razonamiento de los estudiantes cuando se enganchan en trayectorias didácticas más longitudinales que buscan desarrollar conocimiento acerca del contraste de hipótesis.
- Diseñar y evaluar un currículum escolar que contemple el desarrollo gradual y sistemático de las principales ideas subyacentes a las técnicas de la inferencia estadística, en lugar de dar un peso preponderante o casi exclusivo a contenidos del análisis exploratorio de datos y el cálculo (descontextualizado) de probabilidades.
- En estrecha cercanía al punto anterior, estudiar los conocimientos didácticos y estocásticos de los profesores de matemáticas en relación con el contraste de hipótesis y otros objetos de la inferencia estadística.

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.), *Granada: Didáctica de la Estadística, Probabilidad y Combinatoria 2*, (pp. 207-214).
- Batanero, C., López-Marín, M. M., Gea, M. M., y Arteaga, P. (2019). Conocimiento del contraste de hipótesis por futuros profesores de educación secundaria y bachillerato. *Publicaciones*, 48(2), 73-95.



- Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R. y Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal of Mathematics Education in Science and Technology*, 25 (4), 527–547.
- Batanero, C., Vera, O. D., y Díaz, C. (2012). Dificultades de estudiantes de Psicología en la comprensión del contraste de hipótesis. *Números*, 80, 91-101.
- Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). Technology for enhancing statistical reasoning at the school level. En A. Bishop, K. Clement, C. Keitel, J. Kilpatrick, & A. Y. L. Leung (Eds.), *Third international handbook on mathematics education* (pp. 643-689). Nueva York: Springer.
- Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. En C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics - Challenges for teaching and teacher education: A joint ICMI/IASE Study* (pp. 57-69). Nueva York: Springer.
- Castro-Sotos, A. E., Vanhoof, S., Noortgate, W. V. y Onghena, P. (2009) How confident are students in their misconceptions about hypothesis tests? *Journal of Statistical Education*, 17(2), 18. DOI: [10.1080/10691898.2009.11889514](https://doi.org/10.1080/10691898.2009.11889514)
- CCH (2016). *Programas de estudio. Área de Matemáticas. Estadística y Probabilidad I-II*. Colegio de Ciencias y Humanidades-UNAM. Recuperado de https://www.cch.unam.mx/sites/default/files/programas2016/ESTADISTICA_PROBABILIDAD_I_II.pdf
- Chance, B., delMas, R. y Garfield, J. (2004). Reasoning about sampling distributions. En D. Ben-Zvi y J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*. 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.
- Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. En D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer.
- Godino, J. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, 22(2), 237-284.
- Harradine, A., Batanero, C. y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. En 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.

- Inzunza, S. y Jiménez, J. (2013). Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. *Revista Latinoamericana de Investigación en Matemática Educativa*, 16 (2), 179-211.
- Kahneman, D., Slovic, P. y Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Nueva York: Cambridge University Press.
- Lee, H. (2018). Probability Concepts Needed for Teaching a Repeated Sampling Approach to Inference. En: Batanero C., Chernoff E. (eds) *Teaching and Learning Stochastics. ICME-13 Monographs*. Springer, Cham. https://doi.org/10.1007/978-3-319-72871-1_6
- Lee, H., Angotti, R., & Tarr, J. (2010). Making comparisons between observed data and expected outcomes: Students' informal hypothesis testing with probability simulation tools. *Statistics Education Research Journal*, 9(1), 68–96.
- Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of protohypothesis testing. *Pedagogies*, 4(2), 126-138.
- Ministerio de Educación (2007). *The New Zealand Curriculum*. Wellington, Nueva Zelanda: Learning Media Limited.
- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19.
- Saldanha, L. y Thompson, P. (2002). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 51(3), 257-270.
- Saldanha, L. y Thompson, P. (2007). Exploring Connections between Sampling Distributions and Statistical Inference: an analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education*, 2(3), 270-297.
- Sánchez, E., García-Ríos, V.N. y Mercado, M. (2017). Desarrollo del razonamiento sobre pruebas de significación de estudiantes de bachillerato en un ambiente tecnológico. En J.M. Muñoz-Escolano, A. Arnal-Bailera, P. Beltrán-Pellicer, M.L. Callejo y J. Carrillo (Eds.), *Investigación en Educación Matemática XXI* (pp. 447-456). Zaragoza: SEIEM
- Shaughnessy, M. J., Chance, B. y Kranendonk, H. (2009). *Focus in High School Mathematics: Reasoning and Sense Making*. Reston, VA: National Council of Teachers of Mathematics.

Cómo citar este artículo: SILVESTRE CASTRO, E., & URREA BERNAL (2021), M. A. Análisis del razonamiento de estudiantes de bachillerato frente a una tarea introductoria al contraste de hipótesis. *SAHUARUS. REVISTA ELECTRÓNICA DE MATEMÁTICAS*. (5), 1, pp. 16-34.

