
ARTÍCULO

Diseño y Evaluación de una Trayectoria Hipotética de Aprendizaje orientada a introducir la Prueba de Hipótesis desde un Acercamiento Informal

Design and Evaluation of a Hypothetical Learning Trajectory aimed at introducing hypothesis testing from an informal approach

Eleazar **Silvestre** Castro*

 ORCID iD 0000-0002-9472-483X

Maricela **Armenta** Castro**

 ORCID iD 0000-0002-4992-5763

Santiago **Inzunza** Cazares***

 ORCID iD 0000-0003-4014-6031

Resumen

Aunque algunos estudios sugieren que un acercamiento informal a la inferencia estadística puede ser benéfico para el aprendizaje de conceptos abstractos y complejos, se requiere más investigación enfocada en el diseño didáctico que ayude a concretar las distintas sugerencias y recomendaciones en una trayectoria de aprendizaje. Como una contribución, en este reporte se presenta y analiza el primer ciclo de una trayectoria hipotética de aprendizaje orientada a introducir la prueba de hipótesis en el nivel universitario, desde un enfoque informal basado en el enfoque de Fisher y simulaciones del muestreo. La trayectoria, aplicada a estudiantes de licenciaturas de ciencias sociales, constó de tres tareas que exploran distintos componentes de la prueba vía un acercamiento informal. Analizamos las respuestas de los estudiantes ante las tareas distinguiendo sus aciertos y dificultades en términos del tránsito entre las perspectiva real e hipotética y los niveles conceptuales implicados en la prueba. Nuestros resultados sugieren que la trayectoria favorece la emergencia y desarrollo del razonamiento con el concepto, pero también que adoptar una perspectiva hipotética para llevar a cabo la prueba y su exploración constituye un reto de alta complejidad, especialmente a la hora de explorar un modelo nulo que difiere del de no diferencia.

* Doctor en Ciencias en la especialidad de Matemática Educativa por el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN). Profesor e Investigador de Tiempo Completo en el Departamento de Matemáticas (Campus Hermosillo) de la Universidad de Sonora (UNISON). E-mail: eleazar.silvestre@unison.mx.

** Maestra en Ciencias en la especialidad de Matemática Educativa por la Universidad de Sonora (UNISON). Profesora de Tiempo Completo en el Departamento de Matemáticas (Campus Hermosillo) de la Universidad de Sonora (UNISON). E-mail: maricela.armenta@unison.mx.

*** Doctor en Ciencias en la especialidad de Matemática Educativa por el Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (CINVESTAV-IPN). Profesor e Investigador de Tiempo Completo en la Facultad de Informática (Campus Culiacán) de la Universidad Autónoma de Sinaloa (UAS). E-mail: sinzunza@uas.edu.mx.

Palabras clave: Enfoque informal a la inferencia estadística. Trayectoria hipotética de aprendizaje. Prueba de hipótesis.

Abstract

Although some studies suggest that an informal approach to statistical inference can be beneficial for the learning of abstract and complex concepts, more research focused on the didactic design is needed to help concretize the different suggestions and recommendations into a learning trajectory. As a contribution, this report presents and analyzes the first cycle of a hypothetical learning trajectory aimed at introducing hypothesis testing at the university level, from an informal approach based on Fisher's approach and sampling simulations. The trajectory, applied to undergraduate students in social sciences, consisted of three tasks that explore different test components via an informal approach. We analyze students' responses to the tasks, distinguishing their successes and difficulties in terms of the transition between the real and hypothetical perspective and the conceptual levels involved in the test. Our results suggest that the trajectory favors the emergence and development of reasoning with the concept. However, adopting a hypothetical perspective to carry out the test and its exploration constitutes a highly complex challenge, especially when exploring a null model that differs from the no-difference.

Keywords: Informal approach to statistical inference. Hypothetical learning trajectory. Hypothesis testing.

1 Introducción

Pruebas de hipótesis e intervalos de confianza son dos técnicas de la inferencia estadística que se incluyen en un curso introductorio de estadística inferencial. Sin embargo, se han documentado por investigaciones en educación estadística las múltiples dificultades en su enseñanza y aprendizaje; en particular, se señalan diversos errores tanto conceptuales como procedimentales al tratar con la prueba de hipótesis. Por ejemplo, a menudo se interpreta el nivel de significancia como la probabilidad de que la hipótesis nula sea verdadera tras haber tomado la decisión de rechazarla (Vallecillos, 1999; Harradine; Batanero; Rossman, 2011); se interpreta el p-valor como la probabilidad de que alguna hipótesis (nula o alternativa) sea verdadera (Batanero, 2000; Lane-Getaz, 2017); se confunde el papel de las distribuciones involucradas (la de la población, la muestra y la distribución muestral) (Inzuna y Jiménez, 2013); se omite la interpretación y metodología frecuentista en la que se basa la prueba (López-Martín; Batanero; Gea, 2019) o se concibe como un procedimiento matemático que ofrece certeza total acerca de la veracidad de alguna hipótesis involucrada (Vallecillos, 1999; Batanero *et al.*, 2018).

Tales dificultades han sido encontradas en profesores en formación y en ejercicio; Liu y Thompson (2009) identificaron que las concepciones de profesores de bachillerato acerca de la probabilidad y atipicidad – dos elementos centrales en la lógica de la prueba de hipótesis – no estaban ancladas al concepto de distribución, además de considerar que el rechazo de la hipótesis nula implicaba probar su falsedad. López Martín, Batanero y Gea (2019, p. 688) analizaron el conocimiento de un grupo de profesores en formación, de nivel secundaria y

bachillerato, acerca de los posibles errores que cometerían sus estudiantes al enfrentarse a las pruebas de hipótesis y los intervalos de confianza, e identificaron que los futuros profesores “fueron capaces de identificar un número apreciable de errores” en ambos temas, pero dejaron de lado algunos relacionados con el nivel de significancia y el p-valor. Además, ellos mismos cometieron errores procedimentales y de interpretación de resultados al resolver un problema de prueba de hipótesis.

Batanero (2000) y Harradine, Batanero y Rossman (2011) señalan a la prueba de hipótesis como un concepto multifacético y de alta complejidad epistémica y cognitiva, pues involucra diversos conceptos estadísticos y probabilísticos que se entrelazan por relaciones complejas que involucran al razonamiento hipotético-deductivo. La enseñanza tradicional sobre el tema presenta una metodología híbrida entre los enfoques de Fisher y Neyman y Pearson, los cuales parten de supuestos filosóficos distintos y que puede añadir confusión a la interpretación de resultados (Batanero, 2000), además de privilegiar un enfoque *formal* de la inferencia estadística, que es prácticamente inasequible para la mayoría de los estudiantes de bachillerato e inclusive de nivel superior. En este enfoque, se privilegia la teoría de variables aleatorias y distribuciones de probabilidad para introducir los temas de inferencia. Diversos estudios han probado que este enfoque produce, a lo más, conocimiento inerte que se traduce en la memorización y aplicación de fórmulas cuyos resultados son frecuentemente interpretados de manera incorrecta.

Tanto las dificultades de aprendizaje como los retos de enseñanza han llevado a diversos profesores e investigadores a proponer un *acercamiento informal* a la inferencia estadística (Zieffler *et al.*, 2008; Makar; Bakker; Ben-Zvi, 2011; Batanero; Díaz, 2015; Inzunza e Islas, 2019b), caracterizado principalmente por reducir el uso del aparato matemático que se utiliza en el enfoque de enseñanza formal y a menudo utilizando recursos tecnológicos digitales que facilitan el acceso a conceptos complejos como el de distribución muestral. El objetivo – y reto – de esta aproximación didáctica es hacer accesible las grandes ideas que subyacen a las técnicas de inferencia estadística y que su vez permitan a los estudiantes realizar inferencias razonables, inclusive dentro del entorno laboral (Bakker *et al.*, 2008). Investigadores y educadores han sugerido que la inferencia estadística sea primero introducida por esta vía antes de su versión formal, pues hace asequible desarrollar el razonamiento inferencial en estudiantes con bagajes matemáticos diversos.

A la fecha, distintos trabajos han mostrado que es posible desarrollar un razonamiento apropiado sobre técnicas de inferencia a través del enfoque informal. Inzunza e Islas (2019a) y Van Dijke-Droogers, Drijvers y Bakker (2019) reportan que un enfoque informal basado en

simulación del muestreo, ayuda a estudiantes tanto universitarios como de bachillerato a desarrollar un razonamiento normativamente correcto sobre las distribuciones muestrales. Inzunza e Islas (2019b, p. 23) trabajaron con estudiantes universitarios, con el *software* Tinkerplots y utilizó datos con contextos reales para introducir los intervalos de confianza desde la perspectiva informal; reporta que estos recursos permitieron “un acercamiento empírico a conceptos abstractos como es el caso de las distribuciones muestrales, el margen de error y el nivel de confianza de un intervalo de estimación”.

Para el caso de las pruebas de hipótesis, un grupo particular de propuestas de enseñanza aboga por el uso de distribuciones muestrales empíricas como el modelo probabilístico que permita realizar un acercamiento informal al concepto (Rossman, 2008; Rossman; Chance, 2014; Batanero; Díaz, 2015; Case; Jacobbe, 2018). Sin embargo, algunos estudios han generado evidencias empíricas que respaldan modestamente las potencialidades de un acercamiento informal al concepto (Garfield; Delmas; Zieffler, 2012), en tanto se ha documentado que “un acercamiento basado en la simulación no elimina todas las dificultades con el aprendizaje de la inferencia” (Case; Jacobbe, 2018, p.10).

Distintos investigadores proponen indagar con mayor detalle los retos y dificultades que surgen en el enfoque informal a la inferencia estadística, tanto en el diseño didáctico como en el aprendizaje de los estudiantes (Rossman; Chance, 2014; Van Dijke-Droogers; Drijvers; Bakker, 2019). En el caso de las pruebas de hipótesis, Bakker *et al.* (2008) sugieren prestar especial atención a cómo los estudiantes razonan e integran su conocimiento contextual durante el proceso de aprendizaje; esta comunicación es un esfuerzo en esta dirección. En este trabajo, mostramos el proceso de diseño y evaluación de un primer ciclo de diseño de una trayectoria hipotética de aprendizaje (THA), orientada a introducir la prueba de hipótesis en el nivel universitario desde un enfoque informal; nuestros objetivos son: (1) diseñar una THA, enmarcada en el enfoque informal a la inferencia, para introducir la lógica de la prueba de hipótesis para el caso del modelo nulo de no diferencia; (2) evaluar el primer ciclo de la THA vía un experimento de diseño, con relación al tránsito entre perspectivas y niveles conceptuales involucrados en la prueba.

2 Marco de referencia

2.1 Aproximación informal a la prueba de hipótesis

Nuestra aproximación informal a la prueba de hipótesis se caracteriza por adoptar el

enfoque del *test de significación* propuesto por Ronald Fisher, cuyo interés último es rechazar determinada hipótesis nula con un cierto nivel de significación y sin identificar una hipótesis alternativa en concreto (Batanero; Díaz, 2006, 2015). En este enfoque no se considera el error tipo II y se asume que la prueba se realiza una sola vez.

El modelo de probabilidad que liga los datos de la muestra con un contexto más amplio (i.e., la distribución muestral del estadístico) se construye de manera computacional, vía el muestreo repetido utilizando simulaciones aleatorias computarizadas (Rossman, 2008; Batanero; Díaz, 2015). Con ello se privilegia la lógica frecuentista del enfoque de Fisher y se facilita, además, que los estudiantes se centren en comprender y apropiarse de los razonamientos que enmarcan al procedimiento de la prueba. En adelante, llamaremos DME a la distribución muestral (empírica) que se genera por esta vía. Adicionalmente, en nuestra visión, un modelo probabilístico simple y razonable para el alcance de los estudiantes es el modelo 50-50 o de no diferencia, en el cual se muestrea a partir de una población con resultados dicotómicos (éxito/fracaso) que poseen igual probabilidad de ocurrencia (.5 cada uno).

2.2 Perspectiva dual y multinivel acerca de la prueba de hipótesis

Case y Jacobbe (2018) argumentan que, al utilizar simulaciones para generar una DME, la realización de la prueba de hipótesis requiere la coordinación entre dos perspectivas, una hipotética y otra real, así como entre múltiples conceptos estadísticos y probabilísticos que operan en tres niveles de abstracción distintos (Figura 1):

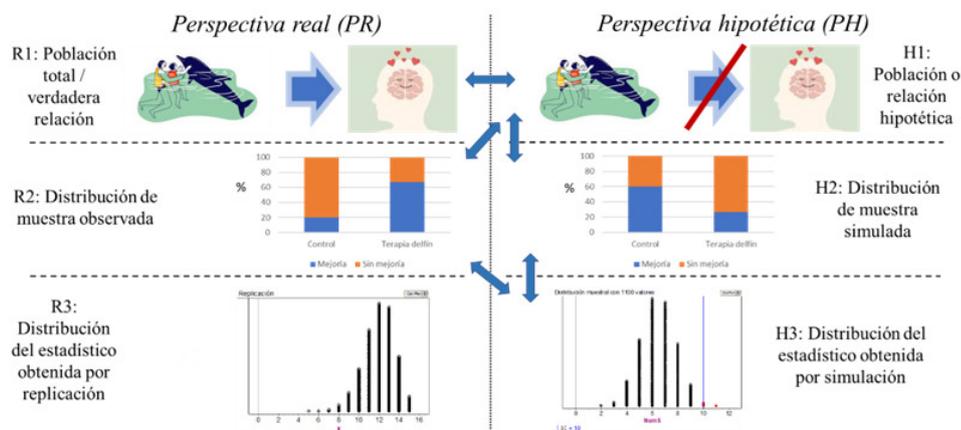


Figura 1 – Perspectivas y niveles conceptuales implicados en la prueba de hipótesis
Fuente: Adaptado de Case y Jacobbe (2018, p. 18)

En la Figura 1 se ilustran las perspectivas y niveles en el contexto del efecto benéfico que podría tener la terapia con delfines; aunque se observe que en cierto grupo experimental se presenta mayor número de pacientes con reducción de estrés y ansiedad (R2), es posible que la

terapia no tenga efecto benéfico alguno en los pacientes. Esta perspectiva se asume como la hipótesis nula (H1), y para operarla se considera el comportamiento de resultados muestrales generados por dicho modelo, transitando así a una perspectiva hipotética de la situación bajo estudio.

El resultado experimental mostrado en H2 se genera por un modelo probabilístico que se deriva de la hipótesis nula y que aproxima la variabilidad de los resultados muestrales, considerando únicamente la aleatoriedad del muestreo o la de una asignación aleatoria entre dos grupos. Haciendo uso de dispositivos físicos o computarizados, el proceso de simulación de resultados muestrales (empíricos) que parten del modelo de la hipótesis nula se repite una gran cantidad de veces, resultando en una DME mostrada en H3. Esta distribución es el concepto base para evaluar la fuerza del resultado experimental obtenido en la perspectiva o dimensión real de la situación bajo estudio (R2), en este caso, mediante la estimación e interpretación del p-valor. Si este valor resulta demasiado bajo, menor que un nivel de significancia previamente declarado, el modelo nulo debe ser rechazado.

De los seis elementos mostrados en la Figura 1, solo se dispone del resultado experimental H2, y aunque la hipótesis nula llegue a ser rechazada, no puede probarse con total veracidad que sea falsa o verdadera. Dado que en rara ocasión se trabaja con una distribución del estadístico producida por replicación (R3), es importante considerar que los estudiantes pueden confundirse con este y otros elementos implicados en la prueba, así como en la coordinación entre la perspectiva hipotética y real de la situación bajo estudio.

3 Método

3.1 Participantes y evidencias

El estudio se realizó con tres grupos completos de estudiantes de la Universidad de Sonora, México, provenientes de la Licenciatura en Finanzas, Licenciatura en Derecho, Licenciatura en Ciencias de la Comunicación y Licenciatura en Administración Pública. Al momento de implementar la trayectoria, los estudiantes de Finanzas se encontraban al inicio de su curso de Estadística Inferencial; para el resto de los estudiantes, las tareas se implementaron al final de su primer curso de Estadística; el primer autor de este reporte fungió como el profesor de la materia y conductor del experimento en todos los casos.

Los estudiantes de Finanzas habían completado un primer curso de Estadística, que comprende una introducción a la Estadística Descriptiva y la teoría de la probabilidad. El resto

de los estudiantes, dado que tenían antecedentes más limitados con la materia, trabajaron con el análisis exploratorio de datos, la modelación y la introducción a la inferencia estadística. En todos los casos se utilizó el *software* Fathom como herramienta para desarrollar conocimientos sobre distintos temas del curso; en particular, los estudiantes estaban familiarizados con la simulación del muestreo repetido para generar una DME al momento de la implementación de la trayectoria.

Las evidencias primarias de este trabajo se constituyen por las respuestas textuales de los estudiantes a las tareas de la trayectoria, que plasmaron en hojas de trabajo electrónicas. De manera complementaria, se recolectaron los archivos del *software* y se realizaron entrevistas no estructuradas a algunos estudiantes. Para desarrollar las tareas de la trayectoria, los estudiantes trabajaron en equipos de tres personas cada uno, con acceso a una computadora equipada con Fathom, durante siete sesiones de clase consecutivas de una hora cada una. Finalizada la implementación, se seleccionaron los equipos de estudiantes que se presentaron a todas las sesiones de clase. Debido a que la primera versión de la trayectoria se trabajó con los estudiantes de Finanzas, incluiremos sus evidencias solo en el inicio de la primera tarea (N=33 equipos en total); para las restantes se cuenta con las evidencias de los estudiantes de Derecho, Comunicación y Administración Pública (ajuste a N=20 equipos en total).

3.2 Instrumentos: diseño y presentación de la THA

Simon (2020) conceptualiza una THA como un modelo teórico para el diseño de la instrucción matemática; consiste en descripciones del pensamiento y aprendizaje de los estudiantes en un dominio matemático específico, enlazadas a una ruta conjeturada y plasmada en un conjunto de tareas instruccionales, diseñadas para el desarrollo de procesos mentales o acciones hipotéticas que favorezcan la progresión de niveles de pensamiento con relación a ese dominio matemático. Los componentes principales de una THA son: objetivos de aprendizaje para los estudiantes, actividades matemáticas diseñadas para promoverlos, y las hipótesis del proceso de aprendizaje (una predicción de cómo el pensamiento y aprendizaje de los estudiantes evolucionará en el contexto de las tareas).

Una THA puede constituir una guía para anticipar las prácticas colectivas o individuales que desarrollan los estudiantes y las maneras en las que razonan con artefactos y problemas diversos. En este trabajo utilizamos el constructo de la THA como un instrumento de diseño e investigación, tanto para relacionar empírica como teóricamente los elementos de la trayectoria de aprendizaje; nuestro punto de arranque para su diseño se constituye por referencias y

propuestas educativas en las que se introduce el concepto de la prueba de hipótesis desde un enfoque informal, que se muestran en el siguiente Cuadro 1:

Tarea	Hipótesis general	Referencias
1	Partiendo de una situación inferencial de interés, en la que se trabaja con datos categóricos, los estudiantes, inicialmente, utilizarán su conocimiento contextual para evaluar la hipótesis nula de “no diferencia / no efecto / modelo 50-50”. En un segundo momento, analizando la misma situación, los estudiantes utilizarán la DME que modeliza la hipótesis nula de no diferencia como vehículo para clasificar el resultado experimental como atípico, por lo cual la rechazarán. En un tercer momento, los estudiantes trasladarán estos razonamientos para analizar una situación inferencial distinta, en la que realizarán una prueba de hipótesis bajo el modelo 50-50 por cuenta propia.	- Partir de y utilizar conocimientos previos de estudiantes (Zieffler <i>et al.</i> , 2008) - El enfoque de Fisher y el trabajo con datos categóricos son vía de acceso favorable a la prueba de hipótesis (Rossman, 2008; Batanero; Díaz, 2015)
2	Los estudiantes realizarán cambios en el resultado experimental y repetirán la prueba, para identificar que éste afecta al cálculo del p-valor. Realizarán cambios en el tamaño de muestra y repetirán la prueba, para identificar que se altera la variabilidad de la distribución de probabilidad que modeliza la hipótesis nula. Identificarán que el resultado de la prueba puede alterarse al modificar alguno de estos elementos.	- Un acercamiento informal facilita identificar roles de elementos clave involucrados en la PH (Rossman, 2008; Batanero; Díaz, 2015)
3	Los estudiantes utilizarán los razonamientos desarrollados en las tareas previas para realizar una prueba de hipótesis en la que el modelo de la hipótesis nula difiera del modelo 50-50 o de no diferencia.	- Los estudiantes deben ser capaces, eventualmente, de investigar afirmaciones más allá del modelo 50-50 (Rossman, 2008)

Cuadro 1 – Vista general de hipótesis y su relación con referencias educativas, que enmarcan la THA
Fuente: Elaboración propia (2023)

Apoyándonos en Rossman (2008), la ruta general que sigue nuestra trayectoria parte de introducir la prueba de hipótesis tomando como modelo nulo el de no diferencia o 50-50, para después explorar el papel del tamaño de muestra y resultado experimental en el resultado de la misma, finalizando con la exploración de un modelo nulo diferente al de no diferencia.

La tarea 1 constó de tres momentos. El primero se enfocó en introducir la noción de hipótesis nula en función del modelo 50-50 y su procedimiento de evaluación; su objetivo de aprendizaje fue provocar que los estudiantes relacionaran la evaluación de la hipótesis nula con su conocimiento contextual de la situación bajo estudio.

Para comenzar la tarea, los estudiantes analizaron el experimento realizado por Hamlin *et al.* (2007), en el cual un grupo de psicólogos analizó si infantes preverbales son capaces de considerar a un sujeto como aversivo o simpático, basándose en la observación de su interacción con otros. En uno de los componentes del experimento, 16 infantes de diez meses de edad vieron *interactuar* por separado a dos objetos con un tercero, para después elegir alguno de los primeros como juguete. 14 de los 16 infantes eligieron al juguete que mostró un comportamiento empático hacia el tercer objeto. Tras analizar este planteamiento, los estudiantes respondieron de manera libre y en equipo a: A) Considerando lo sucedido en el estudio, ¿te parece que la hipótesis nula *los infantes no tienen preferencia, dan la misma*

probabilidad de elegir a cada juguete, es correcta? B) Y si consideras que es incorrecta, ¿qué procedimiento o experimento harías para argumentarlo?

El segundo momento de esta tarea consistió en, analizando la misma situación, introducir la evaluación de la hipótesis nula con base en su DME. Su objetivo de aprendizaje fue provocar que los estudiantes la utilizaran como modelo probabilístico para modelar la variabilidad muestral y a partir de ella juzgar la fuerza del resultado experimental contra el modelo nulo.

La distribución, con modelo probabilístico binomial de parámetros $n=16$ y $p=.5$, fue generada por el profesor vía Fathom (ver Figura 2, izquierda), y se solicitó a los estudiantes, organizados en los mismos equipos, responder a los cuestionamientos de: A) con base en esta información, ¿qué puedes decir sobre la hipótesis nula – *los niños eligen al juguete ayudante con probabilidad de .5?* B) Si la consideras incorrecta, ¿qué puede concluirse sobre ella tras realizar este procedimiento? En este reporte nos enfocaremos en las respuestas al inciso A).

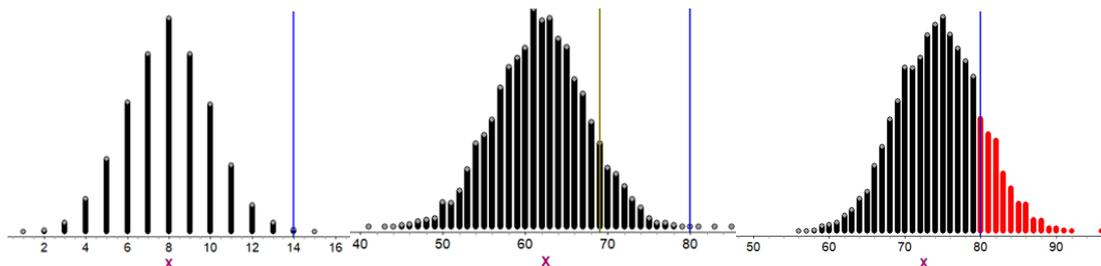


Figura 2 – Distribuciones utilizadas para resolver tareas de la THA

Fuente: Elaboración propia (2023)

Tras institucionalizar el procedimiento para realizar la prueba de hipótesis, se realizó el tercer momento de la tarea, que consistió en que los estudiantes la realizaran por cuenta propia pero en una situación de contexto distinto. Dado que en esta actividad preservamos el modelo nulo de no diferencia, su objetivo de aprendizaje fue que los estudiantes consolidaran una primera versión de la prueba bajo un enfoque informal.

Los estudiantes analizaron el experimento realizado por Onur Güntürkün (2003), que estaba enfocado en proveer evidencias que documenten la dominancia de efectuar tareas humanas diversas hacia el lado derecho. En su experimento, el psicólogo alemán observó a 124 parejas besarse en espacios públicos para registrar hacia qué lado inclinaron su cabeza; Onur documentó que 80 de las parejas se inclinaron a la derecha. Se pidió entonces a los estudiantes realizar una prueba de hipótesis con nivel de significancia de 5%, para evaluar la hipótesis nula de que las parejas tienen la misma probabilidad de inclinar su cabeza a cualquier lado (ver Figura 2, centro).

La tarea 2 consistió en la exploración del mecanismo de la prueba de hipótesis vía la

modificación de sus componentes; su objetivo de aprendizaje fue provocar que los estudiantes identificaran el papel del resultado experimental y el tamaño de muestra en el procedimiento y resultado de la prueba. La actividad ocurre en el contexto del experimento de Güntürkün; primero se pidió a los estudiantes realizar una prueba de hipótesis con nivel de significancia del 5%, para evaluar la hipótesis nula de no diferencia, considerando que se presentaron 69 parejas en la muestra inclinándose a la derecha (ver figura 2, centro). Luego se pidió a los estudiantes realizar una nueva prueba de hipótesis con nivel de significancia de 5% en el contexto del experimento de Hamlin y colaboradores (2007), también para evaluar la hipótesis nula de no diferencia, asumiendo que la muestra fue de 140 infantes y que 84 eligieron al juguete ayudante.

Finalmente, la tarea 3 consistió en la exploración de la prueba de hipótesis bajo un modelo nulo que difiere del de no diferencia o 50-50; su objetivo de aprendizaje fue provocar que los estudiantes consolidaran su noción de prueba de hipótesis al trasladar los razonamientos desarrollados en las tareas previas a una en la que la hipótesis nula deriva en una probabilidad de éxito diferente de .5.

Los estudiantes regresaron al experimento original de Güntürkün (80 de 124 parejas inclinándose a la derecha) y se les pidió que realizaran una prueba de hipótesis manteniendo un 5% como nivel de significancia, para responder a: ¿Puede considerarse a estos resultados como evidencia que más del 60% de las parejas, en general, se inclinan a la derecha cuando se besan? (ver Figura 2, derecha). De esta manera, los indicadores considerados como soporte de cada hipótesis general de aprendizaje se muestran en el siguiente Cuadro 2:

Tarea	Hipótesis de aprendizaje; los estudiantes:
1	HA ₁ . Rechazarán o no H_0 con base en su conocimiento contextual de la situación bajo estudio
	HA ₂ . Rechazarán H_0 utilizando un argumento inductivo (i.e., asumiéndola como verdadera, la proporción muestral sería atípica de obtener)
	HA ₃ . Utilizarán la DME que se deriva de H_0 para calcular la probabilidad de obtener el resultado experimental, para entonces justificar su rechazo
	HA ₄ . Realizarán la prueba de hipótesis bajo el modelo 50-50 por cuenta propia
2	HA ₅ . Identificarán que la modificación del resultado experimental afecta al cálculo del p-valor
	HA ₆ . Reconocerán que un resultado experimental considerado típico de obtener bajo la hipótesis nula implica que no puede rechazarse
	HA ₇ . Identificarán que la modificación del tamaño de muestra afecta el parámetro n de la distribución de probabilidad que modeliza la hipótesis nula
	HA ₈ . Reconocerán que el resultado de la prueba puede alterarse al modificar ya sea el resultado experimental o el tamaño de muestra
3	HA ₉ . Modificarán el parámetro p del modelo binomial para ajustar la nueva hipótesis nula
	HA ₁₀ . Repetirán el proceso de la prueba con una hipótesis nula diferente del modelo 50-50

Cuadro 2 – Indicadores considerados en las tareas de la trayectoria
Fuente: Elaboración propia (2023)

3.3 Procedimiento de análisis

Cada autor de este reporte realizó primero una codificación abierta de las respuestas de los estudiantes ante las tareas de la trayectoria, que luego fueron contrastados, depurados y refinados en una discusión conjunta. Posteriormente, el análisis de los nuevos códigos se interpretó según el tránsito entre las perspectivas real e hipotética y las relaciones conceptuales implicadas en el modelo de Case y Jacobbe (2018).

Para la ilustración de los tipos de respuesta, utilizaremos, en la siguiente sección, las que consideramos como prototípicas de cada categoría. Adicionalmente, tomamos en consideración la frecuencia de las respuestas que se ajustan a cada indicador como la base principal para asignarle un código, a saber: incipiente, media o fuerte, los cuales expresan la fuerza de nuestra evidencia a la hora de respaldar su cumplimiento en la implementación de cada tarea de la trayectoria. El indicador se codificó como incipiente si el porcentaje de respuestas esperadas fue menor de 40, como medio si el porcentaje fue de 40 y menor que 70, y como fuerte de 70 en adelante.

4 Resultados y discusión

4.1 Tarea 1: Evaluación de H_0 bajo un modelo 50-50, sin y con DME

En el primer momento de la tarea 1, 25 de los 33 equipos de estudiantes se ubicaron únicamente en la perspectiva real de la situación bajo estudio para evaluar la hipótesis nula de que los infantes del experimento no tienen preferencia por algún juguete. De estos equipos, 14 consideraron que la hipótesis nula era correcta, amparándose en la idea de que los infantes ignoraron o fallaron en identificar el comportamiento aversivo de un juguete hacia el otro durante el experimento. En este tipo de respuestas, los estudiantes se ubican en el nivel R1, pues ignoraron o despreciaron que el resultado experimental sugiere, inicialmente, la relación opuesta. Los 11 equipos restantes de este grupo consideraron que la hipótesis nula era incorrecta, y se ubicaron en los niveles R1 y R2 por utilizar la información del resultado experimental como apoyo directo a su argumento. Ejemplos de estos tipos de respuesta son:

E1: son niños de 10 meses y son curiosos al elegir al juguete, no demuestra si [la preferencia] es [por] lo simpático o agresivo... es razonable que elijan al azar, no por análisis previo.

E14: ...pensamos que un bebé de 10 meses elegirá al azar sin tomar en cuenta el experimento que se le ha realizado antes, una opción podría ser que al bebé le llame la atención el juguete y por ese motivo lo haya elegido.

E26: No [es correcta], creemos que si las figuras demostraron algún movimiento brusco de una figura hacia otra, en ese caso el cerebro de los menores sí pudiera percibir algún tipo de

estímulo, por eso salieron así los resultados del experimento.
(Trayectoria hipotética de aprendizaje, 2023).

Los ocho equipos restantes consideraron que la hipótesis nula de no preferencia era incorrecta y argumentaron desde una perspectiva hipotética, pues señalaron que si fuese verdadera, lo obtenido en el experimento distaría mucho del valor esperado que se deriva de esa hipótesis. Sus respuestas se ubicaron en las secciones R1, R2 y H1, pues consideraron que la relación que busca probarse en el estudio no era verdadera y desde esa perspectiva juzgaron la fuerza, aunque de manera cualitativa, del resultado experimental; por ejemplo:

E13: ...se puede observar que, en esa muestra de 16 niños, 14 eligieron el objeto ayudante, lo cual es una proporción de 0.875 y es demasiado alta si la hipótesis [nula] fuese verdadera.

E6: Es muy evidente la diferencia que hay entre las elecciones de ambos juguetes, ya que 14 de 16 fueron los que eligieron el juguete ayudante, si se hubiera elegido aleatoriamente como lo plantea la 'hipótesis nula', hubiera más equilibrio en cuanto a qué juguete eligieron.
(Trayectoria hipotética de aprendizaje, 2023).

Al proponer una manera de evidenciar que la hipótesis nula es incorrecta, los estudiantes coincidieron en replicar de una u otra forma el experimento de Hamlin y colaboradores, modificando algunas variables de la situación (por ejemplo, el tamaño de muestra) para así obtener una conclusión certera respecto a la naturaleza de la hipótesis nula. Tales equipos exhibieron adoptar una perspectiva real de la situación bajo estudio e inclusive, en algunos casos, se evidenció una intención de llegar a una acumulación de evidencias confirmatorias que derivaran en algo semejante a la distribución mostrada en la sección R3. Por ejemplo:

E7: ...creemos que el tamaño de la muestra de los niños en el experimento es bastante pequeña para validar el dato, aumentaríamos el tamaño de la muestra..., por lo menos 2,500...después analizaríamos cambiando factores que podrían afectar en la toma de decisiones, ya sea la figura, el color o el rol de los personajes en el experimento... podemos profundizar dentro del muestreo para así obtener datos acumulados que evidencien que la hipótesis es falsa.
(Trayectoria hipotética de aprendizaje, 2023).

En el segundo momento de la tarea se evaluó la hipótesis nula de no preferencia con base en su DME (ver figura 2, izquierda). 11 equipos de estudiantes mantuvieron la estrategia de argumentar a favor o en contra de la hipótesis nula con base en la existencia de una relación causa-efecto entre la decisión de los infantes y su exposición a la interacción de los juguetes. Dichos equipos ignoraron la DME a la mano y se mantuvieron únicamente en la perspectiva real de la situación (R1 y R2), como es evidenciado por las siguientes respuestas:

E8: Sí [es correcta], se puede deber a las preferencias individuales de cada niño.

E22: Seguimos pensando que la hipótesis nula es correcta, que [los infantes] eligen al azar y que la tabla y gráfico no están tan relacionadas con el experimento sino con probabilidad.
(Trayectoria hipotética de aprendizaje, 2023).

Un siguiente grupo de 10 equipos interpretó la DME como una especie de modelo de replicación del experimento con los infantes que evidencia de una u otra forma que la hipótesis nula es verdadera, y en ocasiones con dificultades para distinguir los elementos que conforman a la distribución simulada de los que conforman la muestra. Por lo tanto, los equipos adoptaron una perspectiva real que atraviesa las secciones R1, R2 y R3; por ejemplo:

E1: Es correcta [la hipótesis nula], un lado [de la distribución] es un juguete y el otro lado es otro juguete, se puede ver que casi tienen los mismos resultados.

E31: Opinamos que la hipótesis nula es correcta, puede verse claramente en las pruebas matemáticas y de cálculo que los niños muestran una tendencia en la elección de los juguetes. (Trayectoria hipotética de aprendizaje, 2023).

Los 12 quipos restantes utilizaron apropiadamente la DME como vehículo para evaluar, de manera cualitativa, el grado de atipicidad del resultado experimental, suponiendo que la hipótesis nula es correcta. Aunque sin realizar el cálculo del p-valor correspondiente, los estudiantes identificaron la baja frecuencia del resultado experimental y lo asociaron con una baja factibilidad si la hipótesis nula fuese verdadera, mostrando con ello un tránsito apropiado entre las perspectivas real e hipotética. Ejemplos de estas respuestas son:

E16: Nos podemos dar cuenta con la tabla [de frecuencias] que el [modelo] 50-50 no es correcto, ya que el resultado mostrado no está ni cerca del "8" como debería de ser si [la probabilidad de elegir a cualquier juguete] fuera un 50%.

E10: son 28 posibilidades [ocurrencias] de 10 mil [ensayos] para el $[X=]14$, el cual creemos poco probable...consideramos que [la decisión de los infantes] fue un acto de preferencia ya que teniendo estadísticas fue muy improbable de suceder. (Trayectoria hipotética de aprendizaje, 2023).

En el tercer momento de la tarea, los estudiantes realizaron la prueba de hipótesis por cuenta propia en el contexto del experimento realizado por Güntürkün (2003). En esta ocasión, siete de los 20 equipos brindaron respuestas incorrectas. Dos equipos adoptaron una perspectiva hipotética y consideraron que el resultado experimental era atípico, ubicándose en las secciones R1, R2 y H1, al no acompañar su valoración con el cálculo del p-valor extraído de la DME correspondiente. Tres equipos presentaron dificultades para coordinar elementos que operan entre los niveles que conforman a la prueba, dado que simulaban de manera errónea la DME al utilizar la proporción muestral del experimento como la probabilidad de éxito en el modelo binomial, lo que llevó a que no rechazaran la hipótesis nula. Los otros dos equipos, aunque simulaban de manera correcta la DME, basaron su decisión de rechazar la hipótesis nula en sus conocimientos contextuales y utilizaron el resultado experimental como apoyo a su argumento (es decir, sin vínculo con la DME simulada), lo que ubica a sus respuestas en las secciones R1 y R2. Ejemplos de estos tipos de respuestas son, respectivamente:

E24: No se ajusta, si la hipótesis nula fuera correcta, habría menos parejas [en la muestra] que se inclinan a la derecha, este margen de diferencia es demasiado grande.

E27: Se comprueba que la hipótesis nula es verdadera porque como se puede ver, el gráfico es acampanado con centro en el 80.

E17: Creemos que hay muchos ejemplos de que la mayoría de las personas son diestras y por eso los resultados del experimento salen tan cargados hacia el 70%. (Trayectoria hipotética de aprendizaje, 2023).

Los 13 equipos restantes realizaron el procedimiento de la prueba, en general, de manera correcta y exhibiendo un tránsito apropiado entre los niveles y secciones que la conforman. En la mayoría de estos casos, los estudiantes comenzaron a integrar a su discurso el componente de la incertidumbre en la naturaleza de su conclusión, pues reconocieron que haber rechazado la hipótesis nula de no diferencia no garantizaba que fuese verdadera; por ejemplo:

E25: Observamos que los resultados obtenidos en el experimento son muy atípicos, estadísticamente lo esperado sería que las parejas que giran hacia la derecha oscilaran entre los 60 a 62, en la simulación de Fathom se utilizaron 5,000 casos y solamente 5 veces salieron los 80 casos y 2 veces 81, utilizando el p-valor serían 7/5,000 que nos da .0016, en porcentaje sería .16%, extremadamente por debajo del 5% establecido como α . Como discutimos en clase, este procedimiento no es para darnos un resultado absoluto, más bien es para darnos dirección y expresarla con medidas de probabilidad. (Trayectoria hipotética de aprendizaje, 2023).

Los indicadores de esta tarea fueron valorados de la siguiente forma (Cuadro 3):

Hipótesis de aprendizaje; los estudiantes:	Valoración
HA ₁ . Rechazarán o no H ₀ con base en su conocimiento contextual de la situación bajo estudio	Fuerte
HA ₂ . Rechazarán H ₀ utilizando un argumento inductivo (i.e., asumiéndola como verdadera, la proporción muestral sería atípica de obtener)	Incipiente
HA ₃ . Utilizarán la DME que se deriva de H ₀ para calcular la probabilidad de obtener el resultado experimental, para entonces justificar su rechazo	Incipiente
HA ₄ . Realizarán la prueba de hipótesis bajo el modelo 50-50 por cuenta propia	Media

Cuadro 3 – Valoración de indicadores en la tarea 1
Fuente: Elaboración propia (2023)

En esta tarea, solo uno de los indicadores fue plenamente observado. Cuando los estudiantes se confrontaron con la evaluación de la hipótesis, solo una quinta parte utilizó un razonamiento hipotético-inductivo para relacionar el resultado experimental con el modelo nulo de no diferencia. Aunque se obtuvo una mayor cantidad de respuestas correctas al evaluar la hipótesis nula con base en su DME (36%), desprenderse de una perspectiva real para analizar correctamente la distribución, resultó un gran reto para los estudiantes, inclusive considerando que ya tenían conocimientos previos acerca de cómo generarla e interpretarla. Los estudiantes mostraron, al inicio de esta tarea, las mayores dificultades para interpretar correctamente la DME que se deriva de la hipótesis nula. En particular, las dificultades para distinguir los elementos que la conforman, o interpretarla como un modelo de replicación, son patrones que

han sido bien documentados en otros estudios de referencia (Chance; Delmas; Garfield, 2004; Rossman; Chance, 2014; Case; Jacobbe, 2018).

De acuerdo con Bakker *et al.* (2008), estas dificultades pueden deberse a que nuestros estudiantes, al enfrentarse a una situación inferencial del mundo real, priorizaron de manera intuitiva las condiciones, acciones y argumentos que pueden darse dentro de ese contexto particular. Dado que “las pruebas estadísticas en situaciones educativas están mayormente enfocadas en probar la veracidad de las declaraciones” (Bakker *et al.*, 2008, p.140), también asumieron, en general, una postura relacionada con la experimentación científica que denota la creencia de que sus procedimientos ofrecen conclusiones certeras y libres de incertidumbre.

La investigación ha prestado poca atención a los motivos detrás de las acciones y dificultades conceptuales de los estudiantes cuando inician su proceso de aprendizaje sobre el tema de la prueba; en nuestro trabajo, encontramos que, aunque la perspectiva real influye de forma tan determinante en el razonamiento de nuestros estudiantes, se destaca que la primera tarea de la trayectoria finalizó con el 65% realizando correctamente el procedimiento de la prueba, lo cual sugiere que reconocer y atender este tipo de argumentos dentro del entorno escolar puede ayudarles a ser *disciplinados* (Bakker *et al.*, 2008) con la integración de la información contextual al procedimiento de la prueba (i.e., al principio y al final pero no durante).

4.2 Tarea 2: Exploración del papel del resultado experimental y el tamaño de muestra

Al inicio de esta tarea, los estudiantes realizaron una nueva prueba de hipótesis en el contexto del experimento de Güntürkün (2003), pero considerando que se presentaron 69 parejas en la muestra que se inclinaron a la derecha. El primer grupo de estudiantes, conformado por seis equipos, se caracterizó por brindar respuestas incorrectas a esta situación. Cuatro equipos recurrieron a la estrategia de argumentar desde una perspectiva hipotética, asumiendo primero que la hipótesis nula de no diferencia era verdadera y señalando que el resultado experimental no coincidía con el esperado, por lo cual debía ser rechazada. Además, en tres de estas respuestas se utilizó Fathom para realizar un proceso de simulación, pero únicamente para generar una muestra y tomando la proporción muestral como probabilidad de éxito.

Mientras que dichos equipos evidenciaron un tránsito incipiente entre la perspectiva real e hipotética y los niveles de abstracción de la prueba (R1, R2 y H1), los dos restantes generaron la DME de manera incorrecta, tomando la proporción muestral como probabilidad de éxito en el modelo binomial e interpretándola como si fuese un modelo de replicación (sección R3).

Ejemplos de los dos últimos tipos de respuestas son:

E15: Pensamos que la H_0 es incorrecta porque el psicólogo obtuvo 69 y deberían haber sido 62, aunque el margen ya no es tan grande como antes. Nuestra simulación comprueba que la hipótesis es falsa porque obtuvimos un resultado muy cercano al 69.

E20: La H_0 es correcta, [la simulación / gráfico] nos arroja los datos que desde un inicio nos planteó el problema, dándonos datos cercanos al 69 que había contabilizado el psicólogo. (Trayectoria hipotética de aprendizaje, 2023).

El segundo tipo de respuestas, también con seis equipos, fueron catalogadas como parcialmente correctas. Tres realizaron de manera correcta el procedimiento de la prueba, pero consideraron que la falla en el rechazo de la hipótesis nula constituía una prueba de su veracidad. Los otros tres equipos calcularon de manera incorrecta el p-valor, pues no consideraron los valores más extremos que el resultado experimental, lo que los llevó a rechazar la hipótesis nula; no obstante, reconocieron que el resultado experimental ya no era tan atípico si la hipótesis nula fuese verdadera. Ejemplos de estas respuestas son, respectivamente:

E29: Nosotros pensamos que ahora el 69 es un resultado muy probable ya está muy cerca del 62 esperado y por eso se obtiene un p-valor tan alto. En este caso se acepta la H_0 y con esto se comprueba que puede considerarse como verdadera.

E32: Aunque el 69 ya está muy cerca del 62, apareció 181 veces de las 5000 simulaciones que hicimos, entonces el p-valor quedó más chico que el nivel de significancia del 5% establecido en el problema. Consideramos que el resultado del experimento es muy poco probable de obtener y por eso rechazamos la H_0 de que las parejas se inclinan a cualquier lado. (Trayectoria hipotética de aprendizaje, 2023).

El último grupo de respuestas, brindadas por los ocho equipos restantes, mostró un tránsito correcto entre las perspectivas real e hipotética y los niveles de abstracción de la prueba. En la mayoría de estos equipos, los estudiantes incluyeron su reconocimiento del componente de la incertidumbre, declarando de una u otra manera que el fallo en rechazar la hipótesis nula no aseguraba probar su veracidad. Por ejemplo:

E21: No se mantienen las conclusiones, esto debido a que el nuevo número (69) se repite bastantes veces en comparación a las anteriores pruebas. La parte [del procedimiento] que hay que cambiar es la del p-valor y que en este caso fue de .1296, [calculado a partir] del intervalo $[X=]$ 69 en adelante. Consideramos que el resultado del experimento es mucho más probable que antes. Esto concuerda con la H_0 y por eso decidimos no rechazarla, aunque al final del día podría ser falsa pero [eso] ya no tendría que ver con nuestro procedimiento. (Trayectoria hipotética de aprendizaje, 2023).

Para finalizar esta tarea, los estudiantes regresaron al escenario del experimento realizado por Hamlin y colaboradores (2007) para realizar una nueva prueba de hipótesis, pero considerando ahora que se usó un tamaño de muestra de 140 infantes y que 84 eligieron al juguete ayudante. En esta ocasión se presentaron siete respuestas incorrectas, caracterizadas por adoptar solamente una perspectiva real de la situación bajo estudio y transitar por todos los

niveles de la prueba (R1, R2 y R3), al asumir que el planteamiento de la nueva situación correspondía, en realidad, a una nueva replicación del experimento. Los estudiantes tampoco generaron una DME para modelizar la hipótesis nula de no diferencia y algunos confundieron la probabilidad de éxito con el p-valor, como lo ejemplifica la respuesta del equipo 18:

E18: Con estos nuevos resultados podemos asegurar que la probabilidad de que [los infantes] elijan al ayudante es mayor que el .05%, con una muestra tan grande se comprueba que la H_0 es falsa y eso es lo más razonable porque los infantes muestran desde los 10 meses que tienen cierto tipo de inteligencia, no eligen al azar.
(Trayectoria hipotética de aprendizaje, 2023).

Cuatro equipos brindaron respuestas parcialmente correctas. Aunque evidenciaron un tránsito correcto por las perspectivas real e hipotética y los niveles conceptuales de la prueba, tres mantuvieron la estrategia de calcular el p-valor de manera incompleta (i.e., sin considerar resultados más extremos que el experimental) y uno más consideró que el rechazo de la hipótesis nula podía tomarse como prueba de su falsedad. Los nueve equipos restantes realizaron correctamente el procedimiento de la prueba, ajustando el tamaño de muestra en el modelo binomial y evidenciando un recorrido apropiado entre las perspectivas y niveles implicados en la prueba. Ejemplos de estas respuestas son, respectivamente:

E19: Se ajusta el tamaño de muestra en la binomial y podemos concluir que la H_0 es falsa porque lo obtenido en el experimento es muy poco probable de suceder. El resultado [p-valor] que nos arrojó Fathom fue de 28/3,000 y así se verifica que los niños se fijan en algo de los juguetes (su color, apariencia o comportamiento) que los hace elegir al ayudante.

E23: En este caso la H_0 se rechaza porque el p-valor nos salió más chiquito que el nivel de significancia del 5% ($[X \geq 84]$ con solo 18 de 2,000 [simulaciones]). O sea que estaría muy raro que se presentaran 84 niños eligiendo al ayudante si lo estuvieran haciendo al azar y por eso concluimos que lo más probable es que H_0 sea falsa, aunque no podemos probarlo con este procedimiento matemático estadístico.
(Trayectoria hipotética de aprendizaje, 2023).

Los indicadores de esta tarea se valoraron de la siguiente forma (Cuadro 4):

Hipótesis de aprendizaje; los estudiantes:	Valoración
HA ₅ . Identificarán que la modificación del resultado experimental afecta al cálculo del p-valor	Media
HA ₆ . Reconocerán que un resultado experimental considerado típico de obtener bajo la hipótesis nula implica que no puede rechazarse	Media
HA ₇ . Identificarán que la modificación del tamaño de muestra afecta el parámetro n de la distribución de probabilidad que modeliza la hipótesis nula	Media
HA ₈ . Reconocerán que el resultado de la prueba puede alterarse al modificar ya sea el resultado experimental o el tamaño de muestra	Media

Cuadro 4 – Valoración de indicadores en la tarea 2
Fuente: Elaboración propia (2023)

Los indicadores de esta tarea fueron parcialmente observados, en vista de que entre 40% y 45% de las respuestas fueron catalogadas como correctas en todos los casos. En esta tarea surgió la dificultad de estimar correctamente el p-valor a través de las simulaciones, que

también ha sido observada en Batanero *et al.* (2018); también se observaron las creencias reportadas en Batanero (2000) y López Martín, Batanero y Gea (2019), acerca de considerar que un fallo en rechazar la hipótesis nula implica aceptarla, así como la de considerar que la prueba ofrece una conclusión verídica acerca de la falsedad o veracidad de la hipótesis nula. Otra dificultad que se comparte con los resultados de Rossman y Chance (2014) y los de Case y Jacobe (2018), fue la de muestrear a partir del resultado experimental.

A diferencia de la tarea previa, las dificultades para distinguir los elementos que conforman la DME o interpretarla como modelo de replicación, se presentaron en mucho menor frecuencia. Aunque el cálculo incompleto del p-valor puede considerarse como una dificultad sutil, impidió que muchos estudiantes identificaran plenamente que el resultado de la prueba puede alterarse a raíz de modificar el resultado experimental o el tamaño de muestra. Enfatizar este aspecto de manera que sea razonable para los estudiantes dentro del contexto de la realización de la prueba, no es una tarea trivial dado que implica comprender la convergencia entre modelos de probabilidad (Rossman, 2008); este resultado no ha sido reportado en nuestra literatura de referencia.

4.3 Tarea 3: Exploración de la prueba de hipótesis más allá del modelo 50-50

En esta tarea, los estudiantes realizaron una nueva prueba de hipótesis en el contexto del experimento original de Güntürkün (2003), pero con el objetivo de refutar la hipótesis nula de que un máximo de 60% de parejas se inclinan a la derecha al besarse.

En esta ocasión, se presentaron 14 respuestas incorrectas. Ocho equipos se ubicaron únicamente en la perspectiva real para analizar esta situación, en las secciones R1 y R2, dado que argumentaron que como el resultado experimental (porcentaje muestral de 64%) superaba al porcentaje de 60% de parejas que se inclinan a la derecha, la hipótesis nula debía ser rechazada. Dos equipos también la rechazaron y fueron incapaces de generar la DME que modeliza la hipótesis nula, ubicándose en los niveles H1 y H2 de la perspectiva hipotética, puesto que recurrieron a la simulación de una o varias muestras tomando como probabilidad de éxito el resultado experimental.

Los cuatro equipos restantes generaron de manera correcta la DME, pero dos calcularon de manera incompleta el p-valor (lo que llevó a que rechazaran la hipótesis nula), en tanto que los otros dos interpretaron la distribución como modelo de replicación. Estos últimos equipos se ubicaron en la sección R3 pues interpretaron la DME como evidencia de que más del 60% de las parejas se inclinan a la derecha, y al notar que el resultado experimental no coincidía con

el valor esperado, tomaron la decisión de rechazar la hipótesis nula. Ejemplos del primer y último tipo de respuestas son:

E33: Pensamos que no se debe aceptar la hipótesis nula ya que se presenta más del 60 % [de parejas que se inclinan a la derecha en la muestra].

E24: Al realizar la probabilidad binomial en Fathom, obtenemos como media 74 parejas inclinándose a la derecha, por lo que podemos decir que más del 60% de las parejas se inclinan a la derecha. El experimento arrojó 80 parejas en la muestra que se inclinan a la derecha y eso no se ajusta con la hipótesis nula porque deberían haber sido 73 o 74 [parejas].
(Trayectoria hipotética de aprendizaje, 2023).

Solo Seis de los 20 equipos brindaron respuestas mayoritariamente correctas a esta situación, evidenciando en general un tránsito apropiado por las perspectivas y niveles implicados en la prueba. En las diferentes respuestas de estos equipos se consideró que la hipótesis nula debía ser “aceptada”, aclarando que esta conclusión no se extendía hasta asegurar que sea verdadera:

E18: Ahora obtuvimos un p-valor que es igual a 0.181 y es mayor que el nivel de significancia de 0.05, lo que significa que se acepta la hipótesis nula de “un máximo de 60% de las parejas se inclinan a la derecha”. De igual manera tenemos que considerar que podría haber falla, ya que la hipótesis nula se acepta cuando en realidad partimos de que era un supuesto y podría ser que menos del 60% de las parejas se inclinan a la derecha en general.
(Trayectoria hipotética de aprendizaje, 2023).

Los indicadores de esta tarea se valoraron de la siguiente forma (Cuadro 5):

Hipótesis de aprendizaje; los estudiantes:	Valoración
HA9. Modificarán el parámetro p del modelo binomial para ajustar la nueva hipótesis nula	Media
HA10. Repetirán el proceso de la prueba con una hipótesis nula diferente del modelo 50-50	Incipiente

Cuadro 5 – Valoración de indicadores en la tarea 3

Fuente: Elaboración propia (2023)

Esta tarea fue de mucha complejidad para los estudiantes, en vista de que un indicador fue parcialmente observado y el otro de manera incipiente. Por una parte, se destaca un 40% de respuestas que mostraron un retroceso, en términos de recurrir a una estrategia previa y fallida, dado que tales estudiantes se apoyaron en la heurística de juzgar la hipótesis nula de manera determinista (i.e., sin utilizar un razonamiento inductivo); en contraparte, en el 50% de las respuestas se modificó correctamente el parámetro p del modelo binomial, aunque en ocasiones se interpretó a la distribución como modelo de replicación o se calculó el p-valor de manera incompleta.

Este desempeño nos lleva a conjeturar que la exploración del modelo nulo de no diferencia puede constituir un antecedente importante pero, para varios estudiantes, resulta insuficiente a la hora de mantener un razonamiento inductivo que ayude a dar sentido a una hipótesis nula que va más allá del modelo 50-50. Si se desea que los estudiantes consoliden su

noción de la prueba como “...principalmente una manera formalizada del razonamiento estadístico inductivo” (Bakker *et al.*, 2008, p.139), sería prudente dedicar más esfuerzos a la interpretación de un modelo nulo en su versión generalizada. Este resultado no ha sido reportado en nuestras investigaciones de referencia.

5 Conclusiones

En este trabajo nos propusimos el diseño, implementación y evaluación de una THA orientada a introducir la prueba de hipótesis desde un enfoque informal. Utilizamos referencias educativas para identificar situaciones problema e hipótesis de aprendizaje que consideramos potencialmente útiles para introducir y desarrollar el concepto de la prueba en estudiantes universitarios que tienen bagaje matemático limitado. Al igual que en los estudios de Inzunza e Islas (2019a, 2019b) y Van Dijke-Droogers, Drijvers y Bakker (2019), nuestro experimento de enseñanza aporta evidencias de que los estudiantes universitarios son capaces de desarrollar conocimientos sobre la inferencia estadística, en este caso sobre las pruebas de hipótesis, cuando se introducen desde una aproximación informal.

Durante el proceso de implementación de la trayectoria, identificamos la presencia de varias dificultades de aprendizaje ya reportadas en nuestra literatura de referencia, tales como interpretar la DME como modelo de replicación (Rossman; Chance, 2014; Case; Jacobbe, 2018), confundir los elementos y distribuciones que operan entre los niveles conceptuales involucrados en la prueba (Inzunza e Islas, 2019a; Chance; Delmas; Garfield, 2004; Rossman; Chance, 2014), muestrear a partir del resultado experimental (Case; Jacobbe, 2018), o considerar que la prueba de hipótesis ofrece conclusiones certeras sobre la veracidad de alguna hipótesis (Batanero, 2000; López Martín; Batanero; Gea, 2019); pero también encontramos otras dificultades que han sido poco o nulamente exploradas en la literatura, como la tendencia a priorizar una lectura de la DME desde la perspectiva real, la renuencia a incorporar valores más extremos que el experimental al momento de estimar el p-valor, y la tremenda dificultad que surge para dar sentido a una hipótesis nula cuyo modelo difiere del de no diferencia o 50-50. Respecto a la dimensión didáctica y el impacto en el aprendizaje, consideramos que, en general, nuestro experimento también exhibe que muchas de las dificultades de aprendizaje mencionadas son factibles de atajarse dentro del entorno escolar, y que para ello es crucial prestar atención a cómo los estudiantes razonan e integran la información contextual a medida que exploran un sistema tan complejo de ideas y conceptos que provienen tanto de la estadística como de la probabilidad.

Como reflexiones finales, una limitación importante de nuestro trabajo es que nos restringimos a un análisis global de las respuestas de los estudiantes ante las tareas de la trayectoria, sin hacer un seguimiento individualizado o seccionado por las carreras de las que provienen. Considérese también que la trayectoria fue implementada con estudiantes universitarios ajenos a las áreas de ingeniería y ciencias exactas, por lo que su adaptación a otros contextos estudiantiles debe realizarse con cuidado. En particular, un punto de atención en acciones para el rediseño de la trayectoria es la consideración de las dificultades de los estudiantes al momento de explorar un modelo nulo diferente al de no diferencia.

Referencias

- BAKKER, A. *et al.* Statistical inference at work: Statistical process control as an example. **Statistics Education Research Journal**, San Luis Obispo, v. 7, n. 2, p. 131 - 146, 2008.
- BATANERO, C. Controversies around the role of statistical tests in experimental research. **Mathematical Thinking and Learning**, Queensland, v. 2, n. 1-2, p. 75 - 98, 2000.
- BATANERO, C; LÓPEZ-MARTÍN, M. M.; GEA, M.; ARTEAGA, P. Conocimiento del contraste de hipótesis por futuros profesores. **Publicaciones**, Granada, v. 48, n. 2, p. 73 - 95, 2018.
- BATANERO, C; DÍAZ, C. Methodological and Didactical Controversies around Statistical Inference. *In: JORNADAS DE ESTADÍSTICA*, 38., 2006, Paris. **Artículo de conferencia invitada**. Disponible en: <https://www.ugr.es/~batanero/pages/ARTICULOS/SFE.pdf>. Acceso en 19 jul 2024.
- BATANERO, C; DÍAZ, C. Aproximación informal al contraste de hipótesis. *In: JORNADAS VIRTUALES DE DIDÁCTICA DE LA ESTADÍSTICA, PROBABILIDAD Y COMBINATORIA*, 2., 2015, Granada. **Actas...** Granada: SEIEM, 2015, p. 135 - 144. Disponible en: <http://www.estadis.net/3/actas/Actas%20de%20las%20%20Jornadas%20Virtuales.pdf>. Acceso en: 7 may 2024.
- CASE, J.; JACOBBE, T. A framework to characterize student difficulties in learning inference from a simulation-based approach. **Statistics Education Research Journal**, San Luis Obispo, v. 17, n. 2, p. 9 - 29, 2018.
- CHANCE, B.; DELMAS, R.; GARFIELD, J. Reasoning about sampling distributions. En BEN-ZVI, D.; GARFIELD (eds.). **The Challenge of Developing Statistical Literacy, Reasoning, and Thinking**. Netherlands: Kluwer Academic Publishers, 2004. p. 295-323.
- GARFIELD, J.; DELMAS, R.; ZIEFFLER, A. Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. **ZDM**, Berlín, v. 44, n. 7, p. 883 - 898, 2012.
- GÜNTÜRKÜN, O. Adult persistence of head-turning asymmetry. **Nature**, v. 421, n. 6924, p. 711, 2003.
- HAMLIN, J.K.; WYNN, K.; BLOOM, P. Social evaluation by preverbal infants. **Nature**, v. 450, n. 7169, p. 557 - 559, 2007.
- HARRADINE, A.; BATANERO, C.; ROSSMAN, A. Students' and teachers' knowledge of sampling and inference. En: BATANERO, C.; BURRILL, G.; READING, C. (eds.). **Teaching statistics in**



school mathematics- Challenges for teaching and teacher education. A Joint ICMI/IASE Study. Nueva York: Springer, 2011. p. 235 - 246.

INZUNSA, S.; JIMÉNEZ, J.V. Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. **Revista Latinoamericana de Investigación en Matemática Educativa**, Ciudad de México, v. 16, n. 2, p. 179 - 211 , 2013.

INZUNSA, S.; ISLAS, E. Análisis de una trayectoria de aprendizaje para desarrollar razonamiento sobre muestras, variabilidad y distribuciones muestrales. **Educación Matemática**, Guadalajara, v. 33, n. 3, p. 203 - 230, 2019a.

INZUNSA, S.; ISLAS, E. Diseño y Evaluación de una Trayectoria Hipotética de Aprendizaje para Intervalos de Confianza basada en Simulación y Datos Reales. **Bolema**, Rio Claro, v. 33, n 63, p. 1 - 26, 2019b.

LANE-GETAZ, S. J. Is the p-value really dead? Assessing inference learning outcomes for social science students in an introductory statistics course. **Statistics Education Research Journal**, San Luis Obispo, v. 17, n. 1, p. 357 - 399, 2017.

LIU, Y.; THOMPSON, P. W. Mathematics teachers' understandings of proto-hypothesis testing. **Pedagogies**, Abigdom, v. 4, n. 2, p. 126 - 138, 2009.

LÓPEZ-MARTÍN, M. M.; BATANERO; C.; GEA, M. ¿Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística? **Bolema**, Rio Claro, v. 33, n. 64, p. 672 - 693, 2019.

MAKAR, K.; BAKKER, A.; BEN-ZVI, D. The Reasoning Behind Informal Statistical Inference. **Mathematical Thinking and Learning**, Londres, v. 13, n. 1-2, p. 152-173, 2011.

ROSSMAN, A. Reasoning about informal statistical inference: One statistician's view. **Statistics Education Research Journal**, San Luis Obispo, v. 7, n. 2, p. 5-19, 2008.

ROSSMAN, A.; CHANCE, B. Using simulation-based inference for learning introductory statistics. **Wiley Interdisciplinary Reviews: Computational Statistics**, Nueva Jersey, v. 6, n. 4, p. 211- 221, 2014.

SIMON, M. Hypothetical Learning Trajectories in Mathematics Education. En: LERMAN, S. (ed.). **Encyclopedia of Mathematics Education**. Nueva York: Springer, 2020. p. 354 - 357.

VALLECILLOS, A. Some empirical evidence on learning difficulties about testing hypotheses. **Bulletin of the International Statistical Institute**, Londres, v. 58, [s.n.], p. 201-204, 1999.

VAN DIJKE-DROOGERS, M.; DRIJVERS, P.; BAKKER, A. Repeated sampling with a black box to make informal statistical inference accessible. **Mathematical Thinking and Learning**, Londres, v. 21, n. 2, p. 1-23, 2019.

ZIEFFLER, A. *et al.* A framework to support research on informal inferential reasoning. **Statistics Education Research Journal**, San Luis Obispo, v. 7, n. 2, p. 40-58, 2008.

**Submetido em 01 de Julho de 2023.
Aprovado em 01 de Fevereiro de 2024.**