

DISEÑO DE UNA TRAYECTORIA HIPOTÉTICA DE APRENDIZAJE PARA ABORDAR LAS PRUEBAS DE HIPÓTESIS DESDE UN ACERCAMIENTO INFORMAL

DESING OF A HYPOTHETICAL LEARNING TRAJECTORY TO ADDRESS HYPOTHESIS TESTS FROM AN INFORMAL APPROACH

Eleazar Silvestre Castro, Ilseth Johana Leyva Zazueta, Maricela Armenta Castro
Universidad de Sonora, Colegio Americano de San Carlos. (México)
eleazar.silvestre@unison.mx, ilsethleyva@hotmail.com, maricela.armenta@unison.mx

Resumen:

El enfoque informal a la inferencia estadística puede contribuir significativamente al aprendizaje de técnicas básicas de la estadística inferencial. En este trabajo se presenta y analiza el primer ciclo de una trayectoria hipotética de aprendizaje que introduce la prueba de hipótesis desde un enfoque informal, dirigida a estudiantes del nivel universitario. La trayectoria constó de tres tareas, que se enfocan en introducir la noción de hipótesis nula, la manera de evaluarla, y la exploración del mecanismo de la prueba cuando se modifican dos de sus componentes. Con base en el análisis de las respuestas de un grupo de 39 estudiantes universitarios mexicanos ante las tareas, 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.

Palabras clave: prueba de hipótesis, enfoque informal a la inferencia, trayectoria hipotética de aprendizaje

Abstract:

The informal approach to statistical inference can significantly contribute to learning basic techniques of inferential statistics. This paper presents and analyzes the first cycle of a hypothetical learning trajectory that introduces hypothesis test from an informal approach, aimed at university students. The trajectory consisted of three tasks, which focus on introducing the notion of null hypothesis, the way to evaluate it, and the exploration of the test mechanism when two of its components are modified. Based on the analysis of the responses of a group of thirty-nine Mexican university students to the tasks, our results suggest that the trajectory favors the emergence and development of reasoning with the concept, but also that adopting a hypothetical perspective to carry out the test and its exploration constitutes a highly complex challenge.

Keywords: hypothesis test, informal approach to inference, learning hypothetical trajectory

■ Introducción

Muestreo e inferencia son consideradas ideas estadísticas fundamentales (Burril, 2019). Los conceptos de prueba de hipótesis (PH) e intervalos de confianza están presentes en prácticamente cualquier licenciatura que contemple una introducción a la inferencia estadística. Además de cumplir un rol importante en la formación estadística de diversos profesionistas, dado que constituyen herramientas clave para la investigación científica, también posibilitan el desarrollo de la capacidad de ser un ciudadano crítico a la información estadística encontrada en el día a día (Gal, 2002). No obstante, la investigación en educación estadística ha documentado una amplia gama de errores y dificultades, tanto por parte de estudiantes como de profesores, respecto al dominio de sus múltiples aspectos procedimentales y conceptuales.

En particular, muchas de las dificultades de aprendizaje relacionadas con la PH se enfocan en la interpretación incorrecta de sus distintos componentes. Harradine et al. (2011) señalan que a menudo estudiantes universitarios asumen la hipótesis nula como aquella que se desea probar, o bien a interpretar el nivel de significancia como la probabilidad de que la hipótesis nula sea verdadera cuando ha sido rechazada, e incluso también intercambiar el concepto de p-valor con el nivel de significancia. Así mismo, se ha documentado que la PH es vista como una prueba matemática formal que ofrece una prueba probabilística (i.e., bajo una perspectiva bayesiana) o determinista (i.e., que permite probar la veracidad o falsedad) acerca de alguna de las hipótesis involucradas (Lane-Getaz, 2017; Inzunza y Jiménez, 2013).

Para el caso de profesores de bachillerato y del nivel superior, se han reportado otras dificultades. Por ejemplo, Liu y Thompson (2009) evidenciaron que una muestra de profesores de bachillerato fue incapaz de asumir una perspectiva distributiva para evaluar la atipicidad de un resultado muestral, mientras que López-Martín et al. (2019) identificaron que profesores de secundaria y bachillerato en formación, dieron más peso al aspecto procedimental en la realización de las pruebas de hipótesis por parte de sus estudiantes, lo cual abona a una comprensión incipiente respecto a sus múltiples aspectos conceptuales.

En respuesta a esta problemática, distintas investigaciones han evidenciado que un *enfoque informal* para la enseñanza de la prueba de hipótesis ofrece ventajas favorables y significativas para el aprendizaje del concepto (ver por ejemplo Zieffler et al., 2008). Batanero y Díaz (2015) argumentan que en este enfoque la PH podría plantearse desde la lógica de Fisher, que consiste en conceptualizarla como un procedimiento que permite refutar de manera empírica una determinada hipótesis nula. Rossman (2008) señala que este planteamiento es apropiado para introducir a los estudiantes novatos al mecanismo de la prueba y sugiere iniciar con el tratamiento de situaciones en las que la hipótesis nula es la de “no diferencia” o “modelo 50-50”. Además, en este enfoque se utilizan simulaciones computarizadas para generar la distribución muestral del estadístico de manera empírica, privilegiando así el aspecto más conceptual que procedimental de la PH.

Case y Jacobbe (2018), al igual que otras investigaciones (ver por ejemplo Inzunza e Islas, 2019), evidencian que el uso de este enfoque no previene ni resuelve los múltiples conflictos y dificultades asociadas con la enseñanza y aprendizaje de la PH, pero también que ofrece, en general, ventajas significativas para el aprendizaje de los estudiantes cuando se compara con el método de enseñanza formal y tradicionalista. En nuestra consideración, persiste la necesidad de profundizar en la comprensión de cómo estructurar una serie de actividades o tareas que, desde un enfoque informal, promuevan aprendizajes clave sobre la PH, así como profundizar en el cómo se desarrolla el aprendizaje de los estudiantes en el marco de dicho acercamiento didáctico.

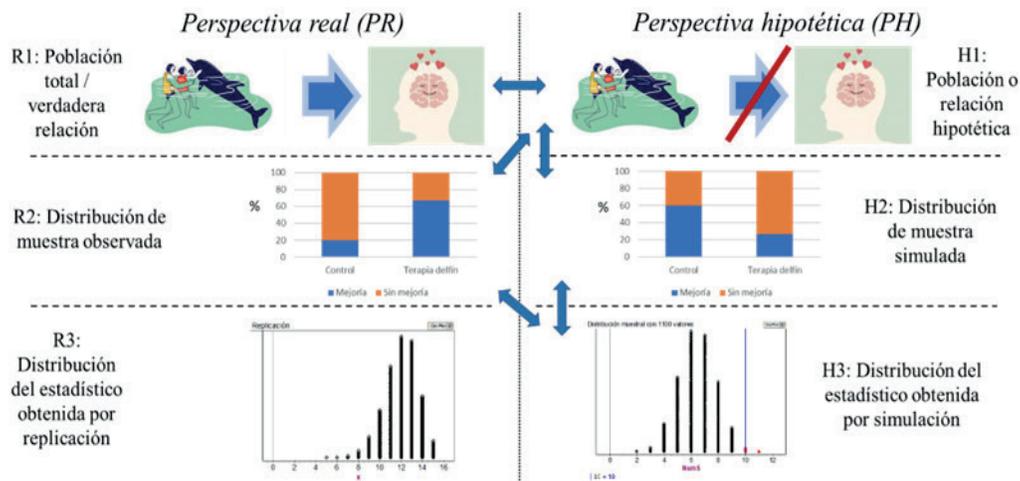
El presente trabajo es un esfuerzo en esta dirección, su objetivo general es presentar el proceso de diseño y evaluación de un primer ciclo de diseño de una trayectoria hipotética de aprendizaje cuyo objetivo es introducir la PH en el nivel universitario desde un enfoque informal. Para ello presentamos las consideraciones del diseño de la trayectoria, su proceso de implementación y las evidencias correspondientes que, en conjunto, nos llevan a reflexionar acerca de su funcionamiento y potencial modificación de componentes en aras de lograr un impacto significativo en el aprendizaje de los estudiantes acerca del concepto.

■ Marco conceptual

Trayectoria hipotética de aprendizaje (THA): su objetivo es acercar los conocimientos, intuiciones y propuestas matemáticas de los estudiantes con los objetivos de aprendizaje pretendidos por el profesor (Simon, 2014). Se compone de tres elementos relacionados entre sí: a) objetivos de aprendizaje, b) un conjunto de tareas o actividades cuyo objetivo es propiciar la actividad matemática de interés, y c) las hipótesis de aprendizaje (o de razonamiento) entendidas como una relación de acciones probables que tomará el estudiante durante la resolución de las tareas. La elaboración de una THA está basada en la comprensión inicial de los estudiantes involucrados y, dado su carácter hipotético, el profesor se involucra frecuentemente en la modificación de sus componentes. Una THA puede entenderse como un vehículo para apoyar tanto el diseño didáctico como la actividad docente y la investigación; en nuestro caso, utilizamos las THA como instrumento de diseño y de investigación dentro de un *experimento de diseño*. Utilizamos el mecanismo de la trayectoria para relacionar distintos elementos que la conforman, tales como los contenidos estadísticos involucrados, las situaciones problema que provocan su emergencia, la generación de tareas que desarrollan las situaciones problema, los materiales y las acciones necesarias para su implementación, y las expectativas de aprendizaje correspondientes para los estudiantes.

Perspectiva dual y multinivel acerca de la PH: Case y Jacobbe (2018) argumentan que la realización de la prueba de hipótesis requiere la coordinación entre dos perspectivas, una hipotética y otra real, y entre múltiples conceptos estadísticos y probabilísticos que operan en tres niveles de abstracción:

Figura 1. Perspectivas y niveles implicados en la prueba de hipótesis.



Fuente: Adaptado de Case y Jacobbe (2018).

En la figura 1 se ilustran las perspectivas y niveles en el contexto del efecto benéfico que puede tener la terapia con delfines. Si se observa que un grupo experimental arroja mayor número de pacientes con reducción de estrés y ansiedad (R2), a raíz de participar en la terapia, no se puede garantizar que la verdadera relación entre el beneficio y la terapia sea de esta naturaleza (R1). Es posible que la terapia no tenga efecto benéfico alguno en los pacientes; esta perspectiva se asume como la hipótesis nula que se refutará mediante la prueba (H1).

Para evaluar dicha hipótesis nula, es necesario considerar el comportamiento de resultados muestrales (empíricos) generados por dicho modelo, transitando entonces a una perspectiva hipotética de la situación bajo estudio. El resultado experimental mostrado en H2 se produce por un modelo probabilístico que se deriva de la hipótesis nula, 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 generación de resultados muestrales (empíricos) generados por el modelo de la hipótesis nula se repite una gran cantidad de veces, resultando en una distribución del estadístico 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.

■ Método

Participantes y escenario del experimento: participaron 39 estudiantes de la Licenciatura en Finanzas de la Universidad de Sonora, quienes habían completado un curso de Estadística en el semestre previo y en el que se atendieron contenidos de estadística descriptiva y nociones de probabilidad. El experimento se realizó al inicio de su curso de Estadística Inferencial, llevado a cabo durante agosto a diciembre de 2021, y en modalidad virtual utilizando la plataforma Microsoft Teams. Para resolver las tareas, se organizó a los estudiantes aleatoriamente en equipos de tres personas cada uno; por motivos ajenos a nuestro control, para la realización de la tarea 2 se presentaron 12 de los 13 equipos.

Instrumentos y ejecución: como primera acción para el diseño de la trayectoria, utilizamos referencias educativas que consideramos pertinentes para identificar situaciones que permitieran desarrollar razonamientos clave sobre las PH. La tabla 1 muestra los razonamientos que se detonarían al realizar las tareas iniciales de la trayectoria:

Tabla 1. Referencias educativas para enmarcar la THA.

Tarea	Hipótesis general	Referencias
1	Partiendo de una situación inferencial de interés, en la que se trabaja con datos categóricos, los estudiantes utilizarán su conocimiento estadístico y contextual para evaluar la hipótesis nula de “no diferencia / no efecto / modelo 50-50”.	Partir de y utilizar conocimientos previos de estudiantes Zieffler et al. (2008). El enfoque de Fisher y el trabajo con datos categóricos son vía de acceso favorable a la PH (Rossman, 2008; Batanero y Díaz, 2015).
2	Analizando la misma situación, los estudiantes utilizarán la distribución muestral del estadístico que modeliza la hipótesis nula como vehículo para clasificar el resultado experimental como atípico, por lo cual la rechazarán.	La distribución muestral empírica es más accesible que su versión abstracta (Rossman, 2008; Batanero y Díaz, 2015; Case y Jacobbe, 2018). Se debe transitar apropiadamente entre perspectiva real e hipotética y entre niveles conceptuales (Case y Jacobbe, 2018).
3	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.	Aprovechamiento del acercamiento informal para identificar rol de elementos clave en la PH (Rossman, 2008; Batanero y Díaz, 2015).

Fuente: elaboración propia.

La tarea 1 consistió 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 relacionen la evaluación de la hipótesis nula con su conocimiento estadístico y contextual de la situación bajo estudio. Al iniciarla, los estudiantes analizaron el experimento realizado por Hamlin et al. (2007), en el cual un grupo de psicólogos analizaron 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 con cual jugar. Dado que uno de los objetos mostró un comportamiento aversivo con el tercero y el otro no, 14 de los 16 infantes eligieron al segundo como juguete tras observar las interacciones. Tras analizar este planteamiento, se hicieron las siguientes preguntas a los estudiantes, para que las respondieran de manera libre y en equipo: 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 es incorrecta, ¿qué procedimiento o experimento harías para argumentarlo?

Así, las hipótesis de aprendizaje de la tarea 1 fueron, los estudiantes:

HA1. Rechazan la hipótesis nula con base en su conocimiento contextual de la situación.

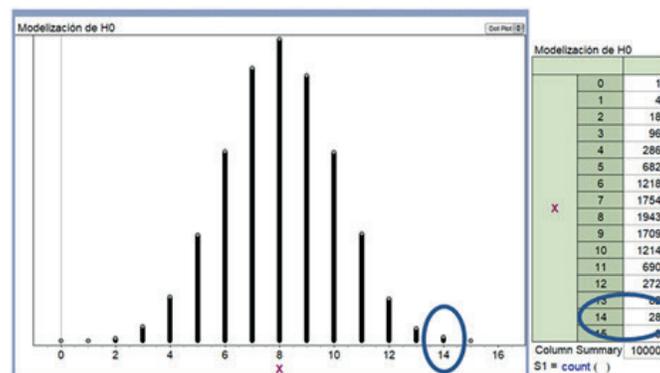
HA2. Rechazan la hipótesis nula utilizando un argumento inductivo (i.e., argumentando que, asumiéndola como verdadera, la proporción muestral es atípica de obtener).

HA3. Calculan la probabilidad de la proporción muestral vía un modelo binomial, asumiendo H_0 como verdadera, para justificar su rechazo.

La tarea 2 consistió en introducir la evaluación de la hipótesis nula con base en su distribución muestral empírica, por lo que su objetivo de aprendizaje fue el de tratar que los estudiantes relacionen su evaluación con su conocimiento estadístico (i.e., vía su modelización). En esta tarea, los estudiantes evaluaron nuevamente la hipótesis nula de no diferencia (los infantes del estudio tienen probabilidad de .5 de elegir al juguete ayudante), pero disponiendo de la distribución muestral empírica (con modelo probabilístico binomial de parámetros $n=16$ y $p=.5$) que la modeliza (ver figura 2).

La distribución fue generada por el profesor utilizando Fathom y se solicitó a los estudiantes, organizados en los mismos equipos, responder a los cuestionamientos de: con base en esta información, ¿qué puedes decir sobre la hipótesis nula “los niños eligen al juguete ayudante con probabilidad de .5”? Si la consideras incorrecta, ¿qué puede concluirse sobre ella tras realizar este procedimiento?

Figura 2. Distribución muestral empírica utilizada en la tarea 2.



Fuente: elaboración propia.

Así, las hipótesis de aprendizaje de la tarea 2 fueron, los estudiantes:

HA4. A la luz de la distribución muestral que modeliza la hipótesis nula, identifican que el dato experimental tiene baja frecuencia de aparición, lo cual lo hace atípico o muy improbable de ocurrir si la hipótesis nula fuese verdadera.

HA5. Rechazan la hipótesis de que la proporción de éxitos es .5, en oposición a aceptar que un evento atípico ha ocurrido únicamente por azar.

HA6. Asumen que la conclusión de la hipótesis nula es de carácter no determinista.

La tarea 3 consistió en la exploración y consolidación del mecanismo de la PH vía la modificación de sus componentes; su objetivo de aprendizaje fue provocar que los estudiantes exploren e identifiquen el papel del resultado experimental y el tamaño de muestra en el mecanismo de la PH. La tarea se realizó en el contexto de esta misma situación y se exploró qué sucedería con el mecanismo de la prueba y su resultado si se modificaran dos de sus componentes, a saber: el resultado experimental y el tamaño de muestra que se utilizaron en el estudio de Hamlin et al (2007).

Los estudiantes, organizados en los mismos equipos, dieron respuesta a: ¿qué pasaría con el procedimiento y resultado de la prueba si, al finalizar la etapa de experimentación, 10 de los 16 infantes hubieran elegido al juguete ayudante? ¿Y qué pasaría con el procedimiento y resultado de la prueba si hubieran participado 140 infantes y 84 eligieron al juguete ayudante?

Es importante considerar que, al momento de enfrentarse a esta tarea, los estudiantes manipularon Fathom por cuenta propia, además de contar con el procedimiento institucionalizado para realizar la prueba (que incluye la manera de estimar el p-valor vía la distribución muestral empírica).

Así, las hipótesis de aprendizaje de la tarea 3 fueron, los estudiantes:

HA7. Identifican que solo modificar el resultado experimental afecta al cálculo del p-valor.

HA8. Reconocen que un resultado experimental considerado típico de obtener bajo la hipótesis nula implica que no puede rechazarse.

HA9. Identifican que, si el tamaño de muestra aumenta, la variabilidad de la distribución de probabilidad que modeliza la hipótesis nula se reduce.

HA10. Reconocen que el resultado de la prueba puede alterarse al modificar ya sea el resultado experimental o el tamaño de muestra.

Evidencias y análisis: para este reporte, la fuente primaria de evidencias fueron las respuestas de los estudiantes ante los cuestionamientos de las tres tareas. Dado que el experimento se desarrolló de manera virtual, los estudiantes utilizaron Excel para registrar sus respuestas y entregaron el documento electrónico correspondiente al final de cada sesión. Respecto a su análisis, cada autor de este reporte realizó primero una codificación abierta de las respuestas que luego fueron contrastados y refinados. Posteriormente, el análisis de los 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) que describimos en la sección previa. Finalmente, se realizó una valoración cualitativa para asignar una etiqueta (nula, débil, media o fuerte) al grado de presencia o cumplimiento de cada hipótesis de aprendizaje. Para la ilustración de los tipos de respuesta, en la siguiente sección, utilizamos respuestas que consideramos prototípicas de cada categoría.

■ Resultados

Tarea 1: 11 de los 13 equipos argumentaron que había (o no) relación causa-efecto entre la decisión de los infantes respecto a su exposición ante la interacción de los juguetes. Ocho equipos consideraron que la hipótesis nula “no es correcta” argumentando que los infantes eligieron al juguete ayudante por haberse comportado de manera empírica

o no aversiva. Esto es, los once equipos se ubicaron únicamente en la perspectiva real de la situación bajo estudio (R1 y R2), como es evidenciado por la respuesta del equipo 1 (E1), que argumentó que la hipótesis nula “es correcta” de la siguiente manera: “...son niños de 10 meses de edad 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”.

Los dos equipos restantes argumentaron desde una perspectiva hipotética, es decir, asumieron que si la hipótesis nula fuese verdadera, lo obtenido en el experimento distaría demasiado del valor esperado de acuerdo con dicha expectativa. Al hacerlo, realizaron un pase hacia la dimensión hipotética de la situación bajo estudio (R1, R2 y H1), pues consideraron que la relación que busca probarse en el estudio (los infantes tienen preferencia genuina por el juguete ayudante) no era verdadera y desde esa perspectiva juzgaron la fuerza del resultado experimental, aunque sin el cálculo de un p-valor. Esto se evidencia en la respuesta del equipo 13, que comentó “...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”.

Los diez equipos que rechazaron la hipótesis nula propusieron realizar una nueva experimentación. En las diferentes propuestas de los equipos se evidenció una intención de replicación del experimento de Hamlin et al. (2007), modificando algunas variables de la situación 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 (R1, R2 y R3), en donde la nueva experimentación se concibe como un tipo de replicación del experimento que brindaría resultados deterministas sobre la hipótesis nula. Por ejemplo, el equipo 7 mencionó:

E7: ...creemos que el tamaño de la muestra de los niños en el experimento es bastante pequeño, por lo tanto, para validar el dato, aumentaríamos el tamaño de la muestra..., por lo menos 2500...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 los cuales evidencien que esta hipótesis es falsa.

Con relación al cumplimiento de las hipótesis de aprendizaje de esta tarea, dado que ningún equipo utilizó un modelo binomial para modelizar la hipótesis nula de no diferencia, la hipótesis HA3 no recibió apoyo empírico alguno. En cambio, la mayoría de los equipos recurrieron a su conocimiento contextual enmarcado en la experimentación científica, privilegiando la perspectiva real de la situación, buscando replicar el experimento de Hamlin et al. (2007); esto nos motivó a considerar que la hipótesis HA1 recibió un grado alto de apoyo por parte de las evidencias. Finalmente, la hipótesis HA2 recibió apoyo bajo dado que solo dos equipos asumieron la perspectiva hipotética al utilizar un razonamiento inductivo tipo ‘modus tollens’.

Tarea 2: cuatro de los equipos ignoraron la distribución muestral empírica a la mano y abordaron la tarea de evaluar la hipótesis nula adoptando una perspectiva real de la situación (R1 y R2). Los equipos 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, como es evidenciado por la respuesta del equipo 7: “...diversos factores sí podrían influir al momento de la toma de decisiones...”; y la del equipo 8: “...sí [es correcta], se puede deber a las preferencias individuales de cada niño...”.

Los siguientes cuatro equipos, para realizar su evaluación de la hipótesis nula, interpretaron la distribución muestral empírica como una especie de modelo de replicación del experimento con los infantes. Por ejemplo, el equipo 1 declaró que “...un lado es un juguete y el otro lado es otro juguete, se puede ver que casi tienen los mismos resultados...”; mientras que el equipo 2 señaló que “...con base a las pruebas matemáticas y cálculo se puede ver que la hipótesis nula muestra un cambio en las elecciones de los niños...”. Así, estos equipos evidenciaron adoptar una perspectiva real de la situación que, en este caso, atraviesa los tres niveles conceptuales implicados en la prueba de hipótesis (R1, R2 y R3).

Los cuatro equipos restantes interpretaron la distribución muestral empírica como vehículo para evaluar cualitativamente la atipicidad del resultado experimental; aunque no realizaron la estimación del p-valor correspondiente, evidenciaron un tránsito apropiado entre las perspectivas real e hipotética de la prueba y a través

de sus niveles conceptuales. Por ejemplo, el equipo 10 argumentó que “son 28 posibilidades [ocurrencias] de 10 mil [ensayos] para el $[X=]14$, el cual creemos poco probable...consideramos que fue un acto de preferencia ya que teniendo estadísticas fue muy improbable de suceder”.

Con relación al cumplimiento de las hipótesis de aprendizaje de esta tarea, la hipótesis HA4 recibió apoyo empírico bajo dado que ocho de los 12 equipos ignoraron o interpretaron incorrectamente la distribución empírica que modeliza la hipótesis nula. Esto conllevó a considerar que la hipótesis HA5 recibió apoyo empírico medio-bajo, en virtud de que solo cuatro de los ocho equipos que rechazaron la hipótesis nula lo hicieron correctamente utilizando la distribución muestral empírica a la mano. Y respecto a la hipótesis HA6, su apoyo empírico fue catalogado como medio puesto que, de los equipos que rechazaron la hipótesis nula, cinco eligieron la opción “Con este procedimiento se infiere que los niños eligen al juguete ayudante con probabilidad mayor a .5”.

Tarea 3: seis de los 13 equipos recurrieron a la estrategia de argumentar desde una perspectiva hipotética, asumiendo que la hipótesis nula de no diferencia era verdadera. Argumentaron, en general, que el resultado experimental no coincidía con el esperado y por ello debía ser rechazada. Ninguno utilizó la distribución muestral empírica para calcular un p-valor, por lo que sus respuestas evidencian un tránsito incipiente entre la perspectiva hipotética y real y los niveles de abstracción de la prueba (R1, R2 y H1). Como ejemplo de estas respuestas, el equipo E3 mencionó que “no la aceptamos [la hipótesis nula de no diferencia,] porque si se fueran a guiar con él .5 sería un resultado [esperado] entre el 7, 8 y 9”.

Dos equipos realizaron el procedimiento de la prueba de manera correcta y concluyeron (correctamente) que la hipótesis nula no podía ser rechazada; pero al hacerlo, también evidenciaron una tendencia al sincretismo entre la perspectiva hipotética y la real al interpretar la distribución muestral empírica que la modeliza (i.e., sincretismo entre R3 y H3). Por ejemplo, el equipo 9 sugirió que para tener mayor certeza en su conclusión acerca de la hipótesis nula, necesitaría realizar una nueva experimentación en la que participase un mayor número de infantes (i.e., aumentar el tamaño de muestra):

E9: el [p-]valor que nos dio a nosotros está muy lejos del 0.05, ya que nos dio 0.2106 la suma de los valores [de la frecuencia relativa] del $[X=]10$ al 15, la parte del procedimiento a cambiar es la muestra, ya que siento que ocupamos más niños para tener un resultado más conciso y correcto.

Los cinco equipos restantes realizaron correctamente el procedimiento de la prueba y concluyeron que no podía ser rechazada; realizaron un tránsito apropiado entre las perspectivas real e hipotética y los niveles de abstracción de la prueba, como lo evidencia la respuesta del equipo 5:

E5: ...hay que cambiar es la del cálculo del p valor, esta vez se realizará a partir del intervalo de 10 a 16 infantes. El nuevo p valor que se obtuvo es un resultado de 0.2204, podemos concluir en que la hipótesis nula debe ser aceptada ya que este p valor es mayor que el nivel de significancia ($\alpha=.05$). El tipo de error que se podría cometer sería el número 2, el cuál es aceptar la hipótesis como verdadera cuando en realidad los infantes sí tenían una preferencia genuina por el juguete ayudante y no había forma de saberlo.

Al explorar la influencia del cambio de tamaño de muestra en el mecanismo de la prueba, siete de los 13 equipos se caracterizaron por adoptar solamente una perspectiva real de la situación y transitar por todos sus niveles (R1, R2 y R3), al asumir que el planteamiento de la nueva situación bajo análisis podía ser considerada realmente como una replicación del experimento. Tampoco generaron una distribución muestral empírica para modelizar la hipótesis nula de no diferencia y confundieron el concepto de probabilidad con el de p-valor. Por ejemplo, la respuesta del equipo 4 fue:

E4: ...con estos dos experimentos nos podemos guiar en que la probabilidad de elección del juguete ayudante es mayor a 0.05%, esto tal vez porque los infantes aun teniendo 10 meses saben distinguir entre lo malo y lo bueno y la mayoría elige al juguete ayudante, por lo que se puede cambiar la hipótesis a que no es tanto al azar, sino que tenemos que tomar en cuenta que los infantes cuentan con cierto tipo de inteligencia y la mayoría elegirá la mejor opción.

Los seis equipos restantes realizaron correctamente el procedimiento de la prueba, evidenciando así un tránsito apropiado entre las perspectivas y niveles conceptuales involucrados. Por ejemplo, el equipo 8 respondió:

E8: ...se transforma la hipótesis nula en modelo probabilístico binomial ya que el tamaño pasa a ser $n=140$ & $p=0.5$ ". El p-valor queda como .0108 porque el $[X=]84$ [en adelante] apareció 216 de 20,000 simulaciones. Podemos determinar después de estos resultados que la probabilidad de que 84 infantes elijan al juguete ayudante es muy baja en comparación con los demás posibles resultados, por lo tanto, se debe rechazar la hipótesis nula en la cual se establece que .05 sea el nivel de significancia. Podríamos caer en el error tipo 1.

Con relación al cumplimiento de las hipótesis de aprendizaje de esta tarea, la hipótesis HA7 recibió apoyo empírico bajo puesto que cinco de los 13 equipos identificaron el cambio en el p-valor a raíz de alterar el resultado muestral. La hipótesis HA8 recibió apoyo empírico medio-bajo, dado que si bien la mayoría de los equipos tomó la decisión correcta de rechazar la hipótesis nula de no diferencia, la mitad lo hizo a partir de la estimación del p-valor, y en ocasiones con dificultades para interpretar correctamente la distribución muestral empírica. La hipótesis HA9 recibió apoyo empírico medio-bajo, en virtud de que seis de los 13 equipos identificaron el cambio en la distribución muestral empírica a raíz de modificar el tamaño de muestra. En general, poco menos de la mitad de los equipos fue capaz de identificar que un cambio en el resultado experimental o el tamaño de muestra sería suficiente para alterar el resultado de la prueba, lo que motivó a considerar que la hipótesis HA10 recibió apoyo empírico medio-bajo.

■ Discusión y conclusiones

Nuestros resultados sugieren que es factible provocar y desarrollar el razonamiento con la PH desde un enfoque informal. El contexto de la situación de Hamlin et al. (2007) fue accesible para los estudiantes y el trabajo en este contexto, junto al uso de simulaciones aleatorias, hizo visible el razonamiento de los estudiantes mientras daban sentido al mecanismo de la prueba. En general, el desempeño de casi la mitad de nuestros estudiantes se ubicó entre las categorías de respuestas correctas y parcialmente correctas. Considerando esto, nuestro trabajo aporta evidencias de que el enfoque informal para abordar la PH puede ser favorable para el aprendizaje de los estudiantes (Garfield et al., 2012; Case y Jacobbe, 2018). Si bien consideramos que nuestros estudiantes desarrollaron razonamientos importantes sobre el mecanismo de la prueba, en nuestro trabajo se destaca la alta complejidad que implica adoptar una perspectiva hipotética para el planteamiento de la hipótesis nula y su interpretación, además de coordinarse apropiadamente con la perspectiva real e hipotética durante el desarrollo de toda la prueba. Estos patrones han sido documentados en nuestras investigaciones de referencia (Lane-Getaz, 2017; Case y Jacobbe, 2018).

Dos retos para el aprendizaje especialmente complejos se detonan, respectivamente, al momento de interpretar la distribución muestral empírica que modela la hipótesis nula o al modificar algún componente de la prueba. Respecto al primer caso, nuestros resultados están en línea con lo reportado por Garfield et al. (2012) y Lane-Getaz (2017), puesto que muchos de nuestros estudiantes interpretaron en diferentes ocasiones que la distribución muestral empírica era un modelo de replicación de la situación bajo estudio. Respecto al segundo caso, el cumplimiento medio-bajo de nuestras hipótesis de aprendizaje en la última tarea sugieren que los estudiantes tienen un dominio muy inestable de los elementos estadísticos y probabilísticos involucrados en la prueba. Esto se basa en que alterar el tamaño de muestra o el resultado experimental representó considerables dificultades para mantener una apropiada coordinación entre la perspectiva real e hipotética en la PH y los elementos que la conforman; este patrón ha sido poco nualmente reportado en nuestra literatura de referencia. Finalmente, consideramos que un rediseño de la trayectoria que podría ser fructífero de explorar sería, como tercera tarea, explorar otra situación distinta a la de Hamlin et al. (2007), lo cual permitiría que los estudiantes realicen la PH manteniendo la hipótesis nula como un modelo 50-50 o de no diferencia y así consolidar un primer esquema general de su mecanismo. Este esfuerzo de investigación tendrá que ser explorado en otra contribución.

■ Referencias bibliográficas

Batanero, C., y Díaz, C. (2015). Aproximación informal al contraste de hipótesis. En J. M. Contreras, C. Batanero, J. D. Godino, G.R. Cañadas, P. Arteaga, E. Molina, M.M. Gea y M.M. López (Eds.), *Didáctica de la*

- Estadística, Probabilidad y Combinatoria 2*, (pp. 207-214). Granada, 2015.
- Burril, G. (2019). Building concept image of fundamental ideas in statistics: The role of technology. En G. Burril y D. Ben-Zvi (Eds.), *Topics and trends in current statistics education research* (pp. 123-149). Springer. https://doi.org/10.1007/978-3-030-03472-6_6.
- Case, J., y Jacobbe, T. (2018). A framework to characterize student difficulties in learning inference from a simulation-based approach. *Statistics Education Research Journal*, 17(2), 9-29. <https://doi.org/10.52041/serj.v17i2.156>
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70, 1-51. <https://doi.org/10.2307/1403713>
- Garfield, J., delMas, R., y Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory, tertiary-level statistics course. *ZDM—The International Journal on Mathematics Education*, 44(7), 883-898. <https://doi.org/10.1007/s11858-012-0447-5>
- Hamlin, J. K., Wynn, K., y Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, 450, 557–559. <https://doi.org/10.1038/nature06288>
- Harradine, A., Batanero, C., y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In Batanero, C., Burrill, G. & Reading, C. (Eds.), *Teaching Statistics in School Mathematics: Challenges for Teaching and Teacher Education. A Joint ICMI/IASE Study* (pp. 235-246). Springer. https://doi.org/10.1007/978-94-007-1131-0_24
- 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. <https://doi.org/10.12802/relime.13.1622>
- Lane-Getaz, S. J. (2017). Is the p-value really dead? Assessing inference learning outcomes for social science students in an introductory statistics course. *Statistics Education Research Journal*, 17(1), 357-399.
- Liu, Y., y Thompson, P. W. (2009). Mathematics teachers' understandings of protohypothesis testing. *Pedagogies*, 4(2), 126-138. DOI:10.1080/15544800902741564
- López-Martín, M. D. M., Batanero, C. y Gea, M. M. (2019). ¿Conocen los futuros profesores los errores de sus estudiantes en la inferencia estadística?, *Bolema*, 33(64), 672-693. <http://dx.doi.org/10.1590/1980-4415v33n64a11>
- Rossman, A. (2008). Reasoning about informal statistical inference: One statistician's view. *Statistics Education Research Journal*, 7(2), 5-19. DOI:10.52041/serj.v7i2.467
- Simon, M. (2014). Hypothetical Learning Trajectories in Mathematics Education. En: Lerman, S. (Ed.), *Encyclopedia of Mathematics Education*, Springer. https://doi.org/10.1007/978-94-007-4978-8_72
- Zieffler, A., Garfield, J., Delmas, R. y Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58. <https://doi.org/10.52041/serj.v7i2.469>