ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

Diseño y Evaluación de una Trayectoria Hipotética de Aprendizaje para Intervalos de Confianza basada en Simulación y Datos Reales

Design and Evaluation of a Hypothetical Learning Trajectory to Confidence Intervals based on Simulation and Real Data

Santiago Inzunza Cazares*

ORCID iD 0000-0003-4014-6031

Eldegar Islas Anguiano**

ORCID iD 0000-0002-8194-8922

Resumen

En el presente artículo se discute sobre el diseño de una trayectoria hipotética de aprendizaje para introducir los intervalos de confianza en un curso básico universitario, desde una perspectiva informal basada en datos de encuestas y simulación del muestreo. La trayectoria consta de cuatro actividades y fue evaluada como parte de un primer ciclo de mejora con un grupo de 11 estudiantes (19-21 años) de la carrera de estudios internacionales en una universidad mexicana. Los resultados se obtuvieron del análisis de las hojas de trabajo y los archivos del *software* entregados por los estudiantes al final de cada actividad, adicionalmente un conjunto de ítems seleccionados de la prueba AIRS (*Assessment Inferential Reasoning in Statistics*) fueron respondidos por los estudiantes en una evaluación final. Los resultados muestran que es posible razonar adecuadamente con conceptos complejos que subyacen a una inferencia estadística, utilizando datos con contextos reales y herramientas computacionales dinámicas e interactivas que permiten visualizar, en tiempo real, el muestreo y sus resultados. Sin embargo, algunos conceptos resultaron particularmente difíciles para el estudiantado, como la distinción entre población, muestra y distribución muestral de un estadístico, propiedades de las distribuciones muestrales e intervalos de confianza.

Palabras clave: Trayectorias de aprendizaje. Intervalos de confianza. Simulación computacional.

Abstract

This article discusses the design of a hypothetical learning trajectory to introduce the confidence intervals in a basic university course, from an informal perspective based on the use of data obtained from surveys and samples simulation. The trajectory consists of four activities and was evaluated as part of a first improvement cycle with a group of 11 students (19-21 years) of international studies career at a Mexican university. The results were obtained from the analysis of the worksheets and software files delivered by the students at the end

* 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). Dirección postal: Circuito de las Amapas 785-51, Residencial Amapas, Culiacán de Rosales, Sinaloa, México, CP: 80060. E-mail: sinzunza@uas.edu.mx.

** Maestra en Educación y estudiante de Doctorado en Educación por la Universidad Autónoma de Sinaloa (UAS). Dirección postal: Calle Universitarios Ote., Cd Universitaria, Culiacán de Rosales, Sinaloa, México, CP: 80010. E-mail: eldegarislas@uas.edu.mx.

of each activity, in addition a set of items selected from the AIRS test (Assessment Inferential Reasoning in Statistics), which were answered by the students in a final evaluation. The results show that it is possible to reason adequately with complex concepts that underlie a statistical inference, using data with real contexts and dynamic and interactive computer tools that allow real time visualization of the sampling results. However, some concepts were particularly difficult for the students, as the distinction between population, sample and sampling distribution, properties of sampling distributions and confidence intervals.

Keywords: Learning trajectories. Confidence intervals. Computer simulation.

1 Introducción

La tecnología computacional ha revolucionado las aplicaciones de la estadística, de la misma manera se vislumbra que lo puede hacer con su enseñanza (MOORE, 1990; BIEHLER et al., 2013; BATANERO; BOROVNICK, 2016). Esto puede ser posible a partir de una nueva generación de tecnologías educativas, con características dinámicas e interactivas, con gran poder de cálculo numérico y gráfico. Aunado a ello, el avance que ha tenido la investigación en educación estadística, en los años recientes, ha generado resultados que permiten tener mayor conocimiento sobre cómo los estudiantes aprenden los conceptos estadísticos y las dificultades que les presenta su comprensión, el razonamiento y el pensamiento estadístico.

La tecnología computacional puede apoyar en el aprendizaje de la probabilidad y la estadística mediante la automatización de cálculos y gráficas, la exploración de datos, visualización de conceptos abstractos, simulación de fenómenos aleatorios, en la investigación de problemas reales, y proporcionando herramientas de colaboración entre estudiantes (CHANCE et al., 2007). En este marco se han propuesto diversos modelos para la enseñanza de la estadística que hacen énfasis en el desarrollo del razonamiento y pensamiento estadístico de los estudiantes, los cuales tienen como elemento integrador el uso de datos reales, tecnologías digitales y ambientes constructivistas de aprendizaje, entre otras componentes pedagógicas (por ejemplo: COBB; MCCLAIN, 2004; PFANNKUCH; WILD; PARSONAGE, 2012).

En el caso específico de la inferencia estadística, Cobb (2007) reconoce que, además del poder de automatización de cálculos y gráficas que se le reconoce a la tecnología, constituye una herramienta valiosa para repensar la forma como se presentan las ideas centrales de la inferencia estadística, un área sumamente compleja para la mayoría de los estudiantes, cuya enseñanza tradicional ha estado basada en una serie de simplificaciones y enfocada en el desarrollo de habilidades matemáticas para uso de fórmulas y procedimientos, las cuales, muchas veces resultan insuficientes para que los estudiantes razonen y piensen

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

estadísticamente. De esta manera, "el currículo tradicional de inferencia necesita desechar la noción de la aproximación normal de la distribución muestral de la media como centro del currículo, y crear uno nuevo, cuyo centro sea el núcleo de la lógica de la inferencia" (COBB, 2007, p. 12).

En este contexto, Wild, Pfannkuch y Reagan (2011) discuten sobre la importancia de hacer accesible la inferencia estadística y desarrollar sus elementos precursores mucho más temprano que como se hace hasta ahora; presentan una propuesta de enseñanza para estudiantes de bachillerato y primer curso universitario basada en un enfoque altamente visual, con uso de tecnología. Identifican un conjunto mínimo de grandes ideas de la inferencia estadística (población, muestra, variabilidad muestral, distribución muestral, incertidumbre) para los estudiantes desde una visión holística de un ciclo investigativo. El propósito es desarrollar una ruta hacia la inferencia formal, buscando que las concepciones a las que se arriben tengan conexiones intuitivas con los métodos formales.

En los años recientes se han presentado algunas propuestas curriculares para la enseñanza de la estadística, basadas en el uso intensivo de diversas modalidades de tecnologías digitales en ambientes constructivistas, en las cuales se retoman algunas de las ideas anteriores. En el caso de Estados Unidos, la NCTM (2000, p. 324) incluye las siguientes recomendaciones para los grados 9-12:

> a) usar simulaciones para explorar al variabilidad de estadísticos muestrales desde una población conocida y construir distribuciones muestrales, b) comprender cómo un estadístico muestral refleja los valores de parámetros poblaciones y uso de distribuciones muestrales como base para la inferencia estadística.

En el caso de Nueva Zelanda, el currículo de Matemáticas y Estadística para el grado 7 establece: a) hacer inferencias desde encuestas y experimentos, b) hacer predicciones informales, interpolaciones y extrapolaciones, c) usar estadísticos muestrales para hacer estimaciones de parámetros poblacionales, d) reconocer el efecto del tamaño de muestra en la variabilidad de un estimador. Para el grado 8 establece: a) hacer inferencias desde encuestas y experimentos, b) determinar estimaciones e intervalos de confianza para medias, proporciones y diferencias, reconociendo la relevancia del teorema del limite central, c) usar métodos tales como el remuestreo y aleatorización para evaluar la fuerza de la evidencia (Ministry of Education, 2018).

Sin embargo, se requiere aún mucha investigación al respecto, pues los fundamentos de la inferencia estadística y sus métodos comprenden una diversidad de conceptos (muestreo, variabilidad muestral, distribuciones muestrales, intervalos de confianza, margen de error,

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

confiabilidad, hipótesis estadísticas, p-valor, significancia estadística, entre otros) que se caracterizan por ser complejos.

Tomando en consideración las ideas anteriormente expuestas y el planteamiento de Simon y Tzur (2004) sobre Trayectorias Hipotéticas de Aprendizaje (THA), en el presente trabajo nos hemos planteado el diseño de una THA con el propósito de promover el aprendizaje de la inferencia estadística y el desarrollo del razonamiento inferencial informal a través del uso de tecnologías digitales, simulando muestreo repetido de una población. A estos métodos, se les conoce en la literatura como métodos informales, en tanto no dependen de los métodos formales emanados de la teoría estadística y de la probabilidad. En particular nos hemos planteado los siguientes objetivos para la presente investigación:

1. Diseñar una THA, basada en un enfoque informal, para introducir las ideas centrales de la inferencia estadística y los intervalos de confianza a partir de datos reales que provienen de encuestas y simulación del muestreo.
2. Evaluar el primer ciclo de la THA mediante un experimento de enseñanza, en términos de desarrollo de razonamiento inferencial y comprensión de los conceptos que se involucran en una inferencia e intervalos de confianza.

2 Fundamentos teóricos y conceptuales

2.1 Trayectorias Hipotéticas de Aprendizaje

De acuerdo con Simon y Tzur (2004, p. 93), una THA parte de los supuestos:

> 1 La generación de una THA está basada en la comprensión inicial de los estudiantes involucrados.
> 2 Una THA es un vehículo para planear el aprendizaje de conceptos matemáticos.
> 3 Las tareas matemáticas son herramientas para promover el aprendizaje de conceptos matemáticos, y son parte clave del proceso de enseñanza.
> 4 Debido a la naturaleza inherentemente hipotética e incierta de este proceso, el profesor está involucrado en la modificación de cada aspecto de la THA.

Una THA consta de tres componentes principales: el *objetivo de aprendizaje* para los estudiantes, las *tareas matemáticas* que se usarán para promover el aprendizaje de los estudiantes, y las *hipótesis sobre el proceso de aprendizaje* de los estudiantes. El logro de los objetivos para una actividad es parcial, pero considerando todas las actividades en conjunto se puede lograr la comprensión global que se pretende (SIMON, 2014).

El análisis conceptual de los intervalos de confianza, desde una perspectiva informal, ubica al muestreo y la variabilidad como conceptos centrales que se deben tener en cuenta en

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

el diseño de tareas de aprendizaje. Al respecto, Pfannkuch (2008, p. 1) señala: “cuando los estudiantes no son conscientes del muestreo, su razonamiento inferencial informal es limitado”. Dierdorp et al. (2016) identifican cinco conceptos que subyacen al muestreo y que son importantes para el razonamiento inferencial: aleatoriedad, tamaño de muestra, distribución, intervalos de confianza intuitivos y relación entre muestra y población. Una vez identificados los conceptos centrales como objetivos de aprendizaje, el siguiente paso consiste en el diseño de las tareas para promover el aprendizaje y razonamiento correcto sobre el bagaje conceptual que rodea a los intervalos de confianza. Esta no es una actividad trivial, pues se requiere un análisis profundo de las relaciones entre estos conceptos para enlazarlos en forma coherente y, así, generar una trayectoria de enseñanza que busca ser óptima en términos de razonamiento y compresión. El conocimiento del profesor del contenido estadístico, las potencialidades de la tecnología que desea utilizar y el contexto de las situaciones que se plantean son cruciales en el éxito de la THA.

2.2 Inferencia Estadística y Razonamiento Inferencial Informal

En un sentido convencional, “una inferencia estadística es un enunciado sobre una población o un proceso, el cual es generado a partir de una muestra y con un nivel explícito de confianza” (MAKAR; BAKKER; BEN-ZVI, 2011, p. 153). Makar y Rubin (2009) identifican tres características clave que forman una inferencia estadística: a) un enunciado de generalización *más allá de los datos*, b) uso de datos de una muestra aleatoria como evidencia para apoyar esta generalización y c) un lenguaje probabilístico que exprese alguna incertidumbre sobre la generalización.

Tradicionalmente, los métodos de inferencia estadística (estimación de parámetros y pruebas de hipótesis) son tema de estudio a partir del bachillerato, y su enseñanza se realiza, predominantemente, desde un enfoque formal que requiere buenos antecedentes de teoría estadística y probabilidad para su comprensión. Al respecto, la literatura en educación estadística (CASTRO SOTOS et al., 2007; GARFIELD; BEN-ZVI, 2008; SALDANHA; THOMPSON, 2014) muestra evidencia empírica sobre las dificultades que tienen los estudiantes con el aprendizaje de la inferencia estadística desde este enfoque, el cual ha mostrado, además, ser insuficiente para que desarrollen un razonamiento inferencial y una adecuada comprensión de las ideas y conceptos que se involucran en una inferencia.

En este contexto, en los años recientes ha cobrado auge una línea de investigación conocida como *Inferencia Estadística Informal*, con su correspondiente razonamiento

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

denominado *Razonamiento Inferencial Informal.* El calificativo de informal deriva de que no se utilizan métodos formales, basados en la teoría de la probabilidad (por ejemplo, variables aleatorias, distribuciones de probabilidad, teorema del límite central) para desarrollar un razonamiento inferencial adecuado y hacer inferencias estadísticas razonables.

Lo anterior ha generado la reflexión en muchos investigadores (MELETIOU-MAVROTHERIS, 2003; INZUNZA, 2010; PFANNKUCH, 2006; GIL; BEN-ZVI, 2011; ROSSMAN, 2008) sobre la viabilidad de introducir ideas informales de inferencia estadística desde la educación básica, incluso en el nivel universitario, con estudiantes con poca formación matemática, para ayudarlos comprender las ideas y conceptos centrales de la inferencia, y como paso previo al estudio de los métodos formales.

Otro elemento que abona a la viabilidad de un enfoque informal de la inferencia, lo aportan Bakker et al. (2008) al señalar que, en la sociedad actual, muchos empleados requieren obtener conclusiones más allá de los datos disponibles para hacer evaluaciones y tomar decisiones en su trabajo, y con frecuencia sin el uso de técnicas formales. Además, los ciudadanos con frecuencia requieren interpretar reportes publicados sobre encuestas de opinión y resultados de experimentos médicos, muchos de los cuales hablan de márgenes de error y confiabilidad (SALDANHA; THOMPSON, 2014). En este sentido, la inferencia estadística informal es bastante deseable antes que los estudiantes estén listos para aprender las técnicas inferenciales formales (MAKAR; BAKKER; BEN-ZVI, 2011).

De acuerdo con principios del enfoque constructivista del aprendizaje, el razonamiento informal juega un rol importante en el estudio de un tema en particular, pues este es punto de partida para el desarrollo de la comprensión formal. Para Rubin, Hammerman y Konold (2006) el razonamiento inferencial informal es el razonamiento que involucra ideas y relaciones como centro, variabilidad, tamaño de muestra y control de sesgo.

Zieffler et al., (2008) lo definen como la forma en la que los estudiantes usan su conocimiento estadístico informal para hacer argumentos para apoyar inferencias acerca de poblaciones basándose en muestras. El razonamiento inferencial informal "combina aspectos cognitivos de razonamiento sobre los datos y azar además e aspectos socioculturales que tienen lugar en el salón de clases y prácticas individuales, disposiciones y discurso" (BEN-ZVI; GIL; APEL, 2007, p. 3).

2.3 La tecnología en la enseñanza de la inferencia estadística

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

La interactividad y la multiplicidad de representaciones visuales dinámicas de las que disponen muchas herramientas de *software* que existen en la actualidad, así como el poder de simulación para extraer una gran cantidad de muestras de una población, casi de forma simultánea, pueden ayudar siempre que se combinen con un estrategia pedagógica adecuada, a que los estudiantes accedan a las grandes ideas de la inferencia estadística. De acuerdo con Wild, Pfannkuch y Reagan (2011), la tecnología proporciona excitantes posibilidades para cambiar el panorama de la educación estadística en las escuelas, de forma que podrían hacerla irreconocible. Para la educación estadística la tecnología es el último *cambio de jugador*, sus más grandes implicaciones vienen del hecho que nos permiten conceptualizarla en una forma que, previamente, no estaba disponible, potencialmente proporciona acceso a conceptos a edades mucho más tempranas de desarrollo.

Pea (1987) se refiere a este potencial de la tecnología computacional como *metáfora reorganizadora*, que cuando es utilizada adecuadamente tiene la capacidad para provocar cambios estructurales en el sistema cognitivo de los estudiantes, a través de una reorganización y transformación de las actividades que ellos realizan con las representaciones y sus transformaciones. Pea define a una herramienta cognitiva como cualquier medio que ayuda a trascender las limitaciones de la mente, en el pensamiento, en el aprendizaje y las actividades de resolución de problemas.

Particularmente, en el caso de las tecnologías educativas, constituyen una extraordinaria y potente herramienta cognitiva para aprender a pensar matemáticamente; con ellas se pueden operar no solo números, sino también símbolos, y permiten almacenar y manipular símbolos dinámicamente, favoreciendo interacciones con los usuarios en tiempo real. En este sentido, el aspecto representacional y de cálculo de la tecnología adquiere importancia en el desarrollo de imágenes visuales de los conceptos y sus conexiones, que les puede ayudar a los estudiantes a desarrollar imágenes mentales correctas y comprender, de forma adecuada, el proceso subyacente a una inferencia.

3 Metodología

3.1 El diseño de la THA

En el diseño de las actividades que integran la THA hemos tenido en cuenta los elementos definidos por Simon y Tzur (2004), así que, una vez identificados los conceptos que se involucran en un intervalo de confianza desde una perspectiva informal, se procedió a

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

definir los objetivos de aprendizaje para los estudiantes, las tareas estadísticas para promover el aprendizaje y las hipótesis de razonamiento sobre el proceso de aprendizaje. Primeramente, diseñamos un plan de clase para el profesor, y, a partir del plan, desarrollamos la hoja de trabajo para los estudiantes con las indicaciones para cada actividad.

En el diseño de la THA privilegiamos un enfoque conceptualmente holístico sobre un enfoque atomístico, en el sentido de Bakker y Derry (2011). Es decir, el contenido de cada concepto es articulado en sus relaciones con otros conceptos, y no como conceptos aislados parte por parte. En esta misma idea, la secuenciación de los conceptos a lo largo de las actividades sigue una trayectoria que tiene al muestreo como base conceptual y la relación que guarda con la aleatoriedad, tamaño de muestra, distribución muestral, intervalos de confianza y relación entre muestra y población, considerados clave en el desarrollo de un razonamiento inferencial informal.

La THA está compuesta por cuatro actividades y secuenciadas conceptualmente de la siguiente manera: poblaciones, muestras y variabilidad muestral, distribución muestral de una proporción, conexión entre distribución muestral e intervalo de confianza para proporción, aleatoriedad y la confiabilidad de un intervalo de confianza para proporción.

3.2 El *software* TinkerPlots

El *software* que hemos seleccionado para la investigación es TinkerPlots (KONOLD; MILLER, 2011). TinkerPlots se concibe como un *software* para la exploración dinámica de datos y modelación, el cual ha sido diseñado específicamente para la enseñanza de la probabilidad y estadística desde el nivel secundaria hasta el nivel universitario. Una pieza clave en su diseño es el dispositivo *Sampler,* a través del cual los usuarios, mediante los diversos componentes (ruletas, urnas, diagramas de barras etc.), pueden construir un modelo de la población, - al que llamaremos modelo de simulador - y correr simulaciones de eventos aleatorios (en este caso, el muestreo). En el caso de los intervalos de confianza, los usuarios pueden generar el modelo de la población, extraer la cantidad deseada de muestras y calcular a la vez el estadístico de interés, para generar una colección de estadísticos y formar la distribución muestral empírica, misma que puede ser visible en forma gráfica y tabular (ver Figura 1).

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

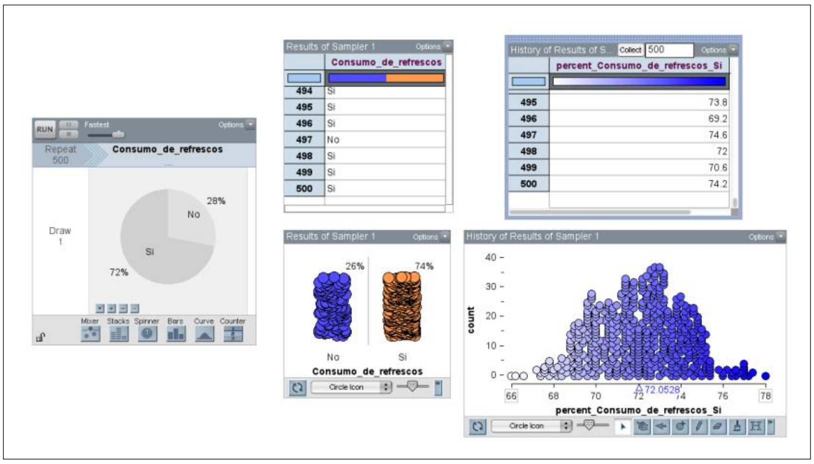

Figura 1 – Modelo de simulador para una población, resultados de un muestra (n=500) y distribución muestral empírica de una proporción para 500 muestras
Fuente: Elaboración propia

Siendo los estadísticos, estimadores puntuales del parámetro correspondiente (por ejemplo la proporción muestral es un estimador de la proporción poblacional), es posible agregar a la tabla de resultados del paso anterior, el cálculo de los demás elementos (error estándar y margen de error para una confiabilidad dada), que permiten construir de manera completa un intervalo de confianza para cada muestra seleccionada. Con el potencial del *software* podemos obtener no solo un intervalo, como es usual en una enseñanza de lápiz y papel, sino una gran cantidad de intervalos, con el propósito de explorar su amplitud, margen de error, cantidad de intervalos que capturan al parámetro – suponiendo que este es conocido – y el efecto que tiene sobre ellos, el tamaño de muestra y el nivel de confianza (ver Figura 2).

History of Results of Sampler 1 — Collect 999 — Options

	prop_LEGALI...	Error_Estandar	Margen_de_Error	Límite_Inferior	Límite_Superior	Resultado
993	0.765	0.0149906	0.0293816	0.735618	0.794382	Cae dentro
994	0.79375	0.0143052	0.0280382	0.765712	0.821788	Cae dentro
995	0.7525	0.0152579	0.0299056	0.722594	0.782406	Cae dentro
996	0.78	0.0146458	0.0287058	0.751294	0.808706	Cae dentro

Figura 2 – Hoja de cálculo con los elementos que forman un intervalo de confianza y el resultado *cae dentro* y *cae fuera* del intervalo
Fuente: Elaboración propia

3.3 Sujetos de estudio e instrumentos de recolección de información

La investigación se llevó a cabo con once estudiantes universitarios de estudios internacionales, que tomaban un curso básico de probabilidad y estadística. Los estudiantes tenían pocos antecedentes matemáticos en la materia, por lo que decidimos enfocar el curso hacia la modelación y simulación de eventos aleatorios y muestreo de poblaciones, utilizando el ambiente computacional que proporciona TinkerPlots. El investigador era el profesor de la materia y en el desarrollo de las actividades funcionó como un guía que solo participaba

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

cuando era necesario para que los estudiantes continuaran con la actividad. Como instrumentos de recolección de información se utilizaron hojas de trabajo para cada actividad, archivos del *software*, entrevistas con algunos estudiantes y un cuestionario de evaluación del razonamiento inferencial aplicado al final de la THA.

3.4 Descripción de las tareas de aprendizaje (actividades)

Las actividades tienen un contexto que, suponemos, es significativo para los estudiantes, pues se han tomado de estudios de opinión sobre temas de ciencias sociales publicados por empresas encuestadoras mexicanas. Cada actividad está basada en una pregunta de una encuesta, los resultados de la muestra que ha utilizado la encuestadora son el punto de partida para generar un modelo de simulador a partir del cual se inicia el proceso de selección de muestras.

Las primeras dos actividades tenían como contexto un estudio para estimar el consumo de refrescos (bebidas gaseosas) por la población mexicana. En una de las preguntas de la encuesta se obtuvo que el 72% de población consume refrescos con regularidad, el cual representa el parámetro para el modelo de simulador (P=0.72). La tercera actividad se diseñó a partir de una encuesta de opinión sobre la relación comercial entre México y Estados Unidos. Ante la pregunta sobre la importancia de Estados Unidos para México, el 79% de los encuestados respondió que era muy importante (P=0.79). La última actividad tenía como contexto una encuesta sobre la legalización del consumo de la mariguana en México, ante la pregunta si estaban de acuerdo con la legalización, el 77% de los encuestados se mostró en contra de la legalización (P=0.77).

La primera actividad inició con la lectura de un artículo muy explicativo para introducir las ideas de azar, tamaño de muestra y estimación, y algunos videos sobre poblaciones, muestreo e incertidumbre en inferencia estadística. Enseguida se procedió al análisis de los resultados de la encuesta y sus indicadores metodológicos (población objetivo, método de muestreo, tamaño de muestra, margen de error y nivel de confiabilidad). La actividad se enfocó en la generación del modelo de la población (simulador), selección de muestras en forma repetida como una primera aproximación a la variabilidad muestral. Después se incrementó el tamaño de muestra de ochocientas (tamaño de muestra utilizado por la encuestadora) a 1,500 personas con el fin de que los estudiantes observaran su efecto en la variabilidad del estadístico.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

La segunda actividad es una continuación de la primera, contemplaba la selección de una mayor cantidad de muestras (por lo general se utilizaron 1,000), identificar y calcular el resultado de interés en la muestra (proporción de personas que consumen refresco regularmente) para construir la distribución muestral empírica (para muestras de tamaño ochocientas y 1,500). El objetivo era identificar propiedades de una distribución muestral, como variabilidad, centro y forma, identificar valores muestrales poco probables de ocurrir e intervalos de confianza intuitivos para el parámetro.

La tercera actividad tenía el propósito de estimar el parámetro desconocido de la población (proporción de personas que responden que Estados Unidos si es muy importante para México, P=0.79). En este caso, los estudiantes no requerían construir el modelo para iniciar la simulación, pues ya estaba construido, pero oculto a la vista de los estudiantes con una opción que proporciona el *software*.

La actividad se enfocó en construir la distribución muestral empírica con 1,000 muestras de tamaño ochocientas, y con base en ella se solicitaba realizar una estimación puntual de P, construir un intervalo intuitivo con una confianza aproximada de 90% y estimar visualmente el margen de error. Se exploró la relación entre el nivel de confiabilidad, tamaño de muestra y cantidad de muestras que caen dentro del intervalo de estimación. Finalmente, se verificó si el parámetro del modelo oculto fue capturado en el intervalo construido y con qué porcentaje de veces ocurría.

Utilizando el poder de la hoja de cálculo y visualización dinámica de TinkerPlots, la cuarta actividad se enfocó en calcular teóricamente intervalos de confianza para la proporción de personas que están en contra de legalización de la mariguana en México (P=0.77), a partir de las proporciones que se obtenían en cada muestra. Se consideró un tamaño de muestra de ochocientas personas y niveles de confiabilidad de 90% y 95%, con el propósito de explorar conceptos como aleatoriedad, nivel de confiabilidad y margen de error de un intervalo de confianza.

4 Resultados y discusión

4.1 La población y su modelo (simulador de muestras)

Para construir un modelo correctamente los estudiantes deben identificar, en el enunciado del problema, la información relevante para configurar el *Sampler*. En la parte central del *Sampler* se especifican los valores que representan características de la población,

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

el tamaño de la muestra a seleccionar se define mediante la opción *Repeat*, mientras que la opción *Draw* determina el número elementos a la vez que serán extraídos del dispositivo. Otras opciones de configuración del *Sampler*, que tienen efecto en la generación de resultados, es el *reemplazo* y *no reemplazo* cada vez que se realiza una selección aleatoria de elementos de la población.

En la actividad 1 el parámetro a representar era el 72% de las personas que respondieron consumían refrescos regularmente. La Figura 3 muestra dos tipos de modelos correctos que utilizaron los estudiantes para representar las características de la población. El primero de ellos está basado en una urna cuyo contenido (siete bolas N y dieciocho bolas S) representan el 28% de los que no consumen refresco y el 72% de los que, si, consumen**.** El segundo modelo está construido a partir de una ruleta en la que cada área representa el 28% de los *No consumidores* y el 72% de los *Si consumidores.*

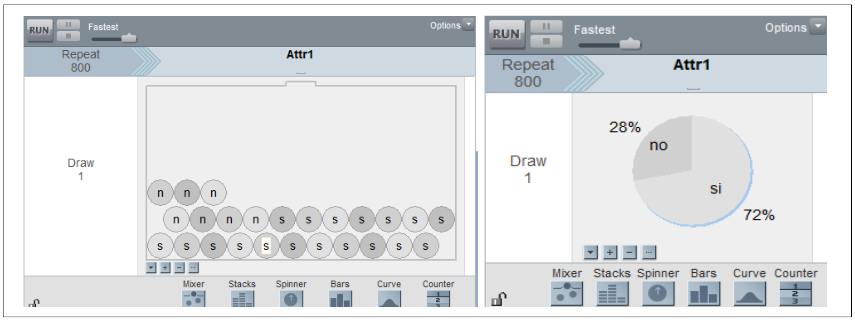

Figura 3 – Dos modelos de simulador correctos construidos por los estudiantes en la actividad 1
Fuente: Elaboración de los estudiantes

La construcción del modelo resultó sencilla para los estudiantes y expresaron correctamente el valor de P en el dispositivo seleccionado, plasmaron correctamente el tamaño de la muestra (n=800) y realizaron la primera corrida. El modelo de ruleta fue el preferido para esta actividad, pues resulta muy fácil identificar en dos secciones el porcentaje de respuestas que recibió cada opción de la pregunta planteada en la encuesta. El modelo de urna que se muestra a la izquierda de la Figura 3 es más complejo, pero igualmente representa las mismas características de la población.

La Figura 4 muestra los dos modelos que predominaron para representar el 77% de la población que está en contra de la legalización de la mariguana en México, que era la otra actividad que requería la construcción de un modelo.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

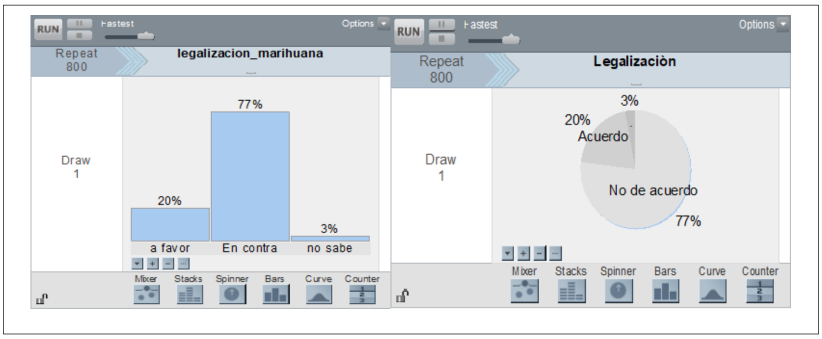

Figura 4 – Modelos de simulador construidos por los estudiantes en la actividad 4
Fuente: Elaboración de los estudiantes

Noll, Gebresenbet y Glover (2016) identifican a esta etapa como la primera fase de razonamiento en la modelación y simulación de un problema estadístico; para que el modelo permita una simulación precisa del problema, los estudiantes requieren configurar correctamente todos los elementos anteriores. En la investigación de Noll, Gebresenbet y Glover, esta etapa resultó complicada para muchos estudiantes, lo cual se debió a que la información de los problemas era mucho más compleja de extraer que en nuestro caso.

4.2 Selección de muestras y cálculo del estadístico de interés

Una vez que se ha construido un modelo de simulador, se procede a realizar la selección de una muestra, enseguida el *software* genera una tabla con los valores obtenidos y el usuario construye una gráfica con los resultados (ver Figura 5).

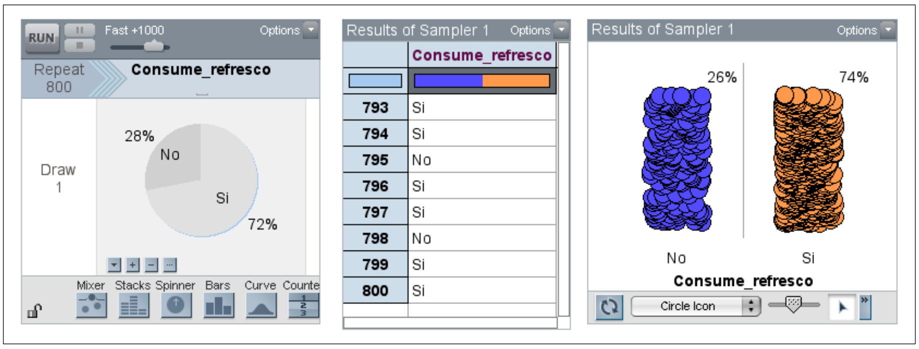

Figura 5 – Modelo y resultados de una muestra
Fuente: Elaboración de la estudiante Brisa

Una nueva corrida del simulador hace desaparecer la muestra anterior y genera nuevos resultados, de tal forma que es posible seleccionar una y otra muestra repetidas veces. Estas potencialidades del *software* nos parecen apropiadas para explorar la relación entre población, muestra, tamaño de muestra y variabilidad muestral, conceptos básicos a partir de los cuales se pueden construir razonamientos correctos sobre la inferencia estadística. En el contexto de la actividad 1, se indicó a los estudiantes que seleccionaran diez muestras de tamaño

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

ochocientos (la misma que utilizó la encuestadora) y registraran la proporción de consumidores de refresco, como variable de interés (ver Figura 6). Posteriormente, se hizo lo mismo con muestras de 1,500 personas. Los cuestionamientos se centraron sobre la variabilidad muestral, variación esperada de las proporciones muestrales, diferencia entre proporciones muestrales y proporción poblacional.

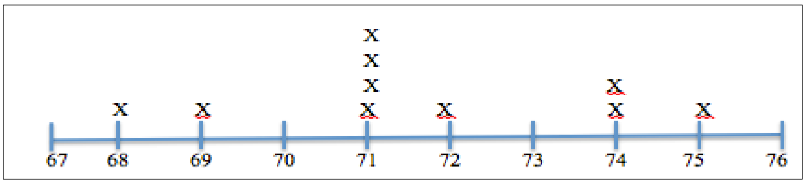

Figura 6 – Gráfica con los resultados de las proporciones de 10 muestras obtenidos por Brisa (proporciones redondeadas al entero más cercano)
Fuente: Elaboración de los estudiantes

La visualización de los valores cambiantes de las proporciones muestrales, y la comparación con el valor fijo de la proporción poblacional, hizo que los estudiantes identificaran que los resultados varían de una muestra a otra, generalmente alrededor del parámetro, lograron, además, identificar un rango razonable de variabilidad esperada. La mayoría estableció intervalos de 70% a 74%, de 71% a 73%, y dos estudiantes establecieron intervalos más amplios de 70% a 76%. Veamos algunas respuestas:

> *Brisa: Es ligeramente diferente porque la muestra es aleatoria, pero entra dentro del margen de error estimado.*
> *José Alfredo: Es diferente a causa del margen de error.*
> *Mariangel: Ligeramente diferente porque es una estimación con un margen de error.* (Cuestionario, 2017).

Los estudiantes justifican la diferencia con el parámetro mediante el margen de error, y, en el caso de Brisa, señala además la aleatoriedad de la muestra, lo cual es un razonamiento correcto. Antes de la actividad analizaron la metodología de la encuesta, y esta señala un margen de error máximo del 3%, por lo cual los estudiantes consideran que las diferencias con el parámetro están dentro de dicho margen. Los resultados muestran que desde la primer actividad algunos estudiantes desarrollaron un razonamiento intuitivo correcto sobre la variabilidad muestral y el error del muestreo.

4.3 Construcción de la distribución muestral empírica (acumulación de resultados del modelo)

Todas las actividades implicaban la construcción de una distribución muestral en forma empírica, a excepción de la primera en la que solo se exploró el muestreo y la

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

variabilidad en unas cuantas muestras. Una vez que se indica la característica de interés y la cantidad de muestras, el *software* genera una representación tabular de la distribución muestral empírica (ver Figura 7).

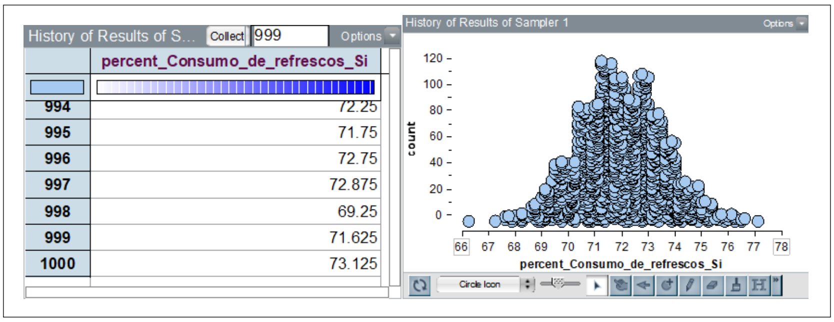

Figura 7 – Distribución muestral empírica para la proporción de *si* en el problema del consumo de refrescos (n=1000 muestras)
Fuente: Elaboración de los estudiantes

Siendo las distribuciones muestrales la base para realizar una inferencia estadística, nos propusimos que los estudiantes identificaran los elementos que las caracterizan, como es la forma, el centro, la variabilidad y el efecto del tamaño de muestra. A su vez nos interesaba que identificaran que muestras en las colas de la distribución tienen poca probabilidad de ocurrir (muestras inusuales). También, era nuestro propósito que determinaran intervalos intuitivos para el parámetro poblacional (P). En relación al significado de la distribución muestral algunos alumnos señalaron qué representa la variación que pueden tener las muestras y los porcentajes.

> *Brisa: Representan la variación de las muestras o los porcentajes.*
> *Frida: Representan la variación que pueden tener las muestras.*
> *Aglae: Los valores que aparecen en la gráfica representan la variación que existe entre los porcentajes.*
> *Mariangel: El porcentaje en cada muestra de personas que consumen refresco. Si el experimento se repite 800 veces, los 800 valores son ese porcentaje en cada uno de los intentos.*
> *Nayeli: Los valores que en efecto, el 72% de la población consume refresco.*
> *Jazmín: Representa un estimado de personas que toman y no toman refresco, dentro de una muestra de 1500.*
> (Cuestionario, 2017).

Mariangel es más explicita en su respuesta y señala que representa el porcentaje de personas que consume refresco en cada muestra seleccionada, lo cual es correcto. Algunos alumnos no respondieron esta pregunta, a causa de la dificultad que les representó. Las respuestas de Nayeli y Jazmín son totalmente incorrectas.

En cuanto a los intervalos construidos de manera intuitiva con un 90% de las proporciones, estos fueron obtenidos visualmente de la distribución muestral. Cuatro estudiantes estimaron un intervalo de 71 a 73, otros cuatro de 70 a 74 y tres estudiantes

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

formaron un intervalo de 71 a 74. En todos los casos los intervalos construidos contienen al parámetro, lo que refleja una idea correcta de la variabilidad muestral esperada.

La distribución muestral empírica presentada en forma gráfica permite visualizar intervalos donde concurren la mayoría de las muestras, pero, también, intervalos donde un resultado muestral es poco probable de ocurrir. Se les planteó que estimaran la probabilidad de que una muestra ocurriera a una distancia 6% por encima o debajo del 72%, es decir menos de 66% y más de 78% (ver Figura 7), lo cual fue identificado correctamente por los estudiantes, señalando que en las quinientas muestras seleccionadas no aparecían muestras tan alejadas del valor central (72%).

> *Frida: Nada probable, puesto que la gráfica muestra que en una recolección de 500 repeticiones del suceso esto no se presentó ni una sola vez.*
> *Nayeli: Es muy poco probable ya que ya se hizo la prueba en el programa y no aparecieron.*
> *Jazmín: Es muy poco probable ya que con esta simulación podemos comprobar que el parámetro varía solo alrededor de un 3% como máximo al correrlo múltiples veces, y un 6% sería un valor muy alejado, una anomalía.*
> (Cuestionario, 2017).

El siguiente paso de la actividad consistió en incrementar el tamaño de muestra a 1,500 personas y construir de nuevo la distribución muestral empíricamente, con el propósito de que los estudiantes identificaran el efecto del tamaño de muestra en los conceptos descritos anteriormente. Todos los estudiantes notaron que la diferencia entre el centro de la distribución muestral y el parámetro P disminuía considerablemente. Veamos algunas respuestas.

> *Frida: Presentan una diferencia muy mínima, casi nula, de 0.02%.*
> *Aglae: Es muy parecido, solo varía en un 0.0299%.*
> *Mariangel: Casi igual, aunque por alguna extraña razón la media se acercó más a la estimación de la encuestadora con una muestra de menor tamaño.*
> *Nayeli: Resulta casi idéntica con tan solo 0.165% de variación del 72%.*
> (Cuestionario, 2017).

Respecto a la probabilidad de que ocurra una muestra con 3% puntos por encima o por debajo del 72%, respondieron lo siguiente:

> *Frida: Es muy poco probable. Como se puede observar en la gráfica, solo se presentaron alrededor de 5 a 8 casos en una muestra de 500.*
> *Nayeli: Es poco probable, sin embargo no imposible ya que al realizar la simulación varias veces salieron hasta 69% y 76%.*
> *Carla: Es poco probable pues el resultado no se aleja mucho el 72%.*
> *Jazmín: Muy poco probable, ya que a. correr el simulador múltiples veces, se puede observar que no existe una variación tan grande entre las muestra.*
> (Cuestionario, 2017).

Al final de la actividad se pidió a los estudiantes que compararan las gráficas de ambas distribuciones muestrales y tomaran nota de la diferencia en forma, centro y variabilidad.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

Frida: Ambas son de forma acampanada, sin embargo en la primera (n=800) los valores se esparcen más a las orillas.
Sebastián: Si hay diferencia, más no es muy notable, ya que son muy similares.
Arlin: La primera (n=800) denota datos dispersos y la segunda como es más grande la muestra están más aglomerados en el centro.
Mariangel: En general la forma de la distribución es la misma, como normal.
Jazmín: Toman la misma forma normal.
Nayeli: En la segunda distribución hay muchos más valores en el centro.
Sebastián: Sí, hay más muestras y se juntan más en el intervalo de 72 %.
Arlin: Si, en la segunda hay más datos en el centro.
Frida: En la muestra de 800 se presentan casos más dispersos, más aislados del 72%.
Arlin: La dispersión de datos si es diferente en cuestión de cantidad.
Mariangel: Los resultados están un poco menos dispersos en la de 1,500.
(Cuestionario, 2017).

4.4 Relacionando las distribuciones muestrales con los intervalos de confianza

Las distribuciones muestrales contienen la información que se requiere para realizar una inferencia, por lo que en esta etapa nos propusimos que los estudiantes estimaran un intervalo de confianza para un parámetro desconocido, estableciendo una conexión directa con la distribución muestral. Esta actividad proviene de una encuesta sobre la relación entre Estados Unidos y México. El parámetro consiste en la proporción de mexicanos que considera que EU si, es importante para México. En esta actividad los estudiantes no requerían generar el modelo para iniciar el muestreo, ya que este estaba construido, pero oculto en el *Sampler*. El primer paso consistió en construir la distribución muestral, utilizando el mismo tamaño de muestra que la encuestadora (ochocientas personas).

Se pidió a los estudiantes realizaran una estimación puntual para el parámetro (P), para lo cual disponían de los resultados de 1,000 muestras con sus proporciones correspondientes. La estimación, en todos los casos, estuvo basada en la media de la distribución muestral, lo que evidencia que los estudiantes optaron por el valor con mayor probabilidad. Enseguida, se planteó que generaran un intervalo con una confiabilidad aproximada de 90%. Considerando el cálculo teórico del intervalo de confianza del 90% y tomando el valor de p=0.79 se tiene que el intervalo sería de 76.6 a 81.4. Un total de ocho estudiantes hicieron una estimación muy cercana a dicho intervalo, a partir de la visualización de la distribución muestral y utilizando la opción del *Medidor* para sombrear el intervalo (ver Figura 8).

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

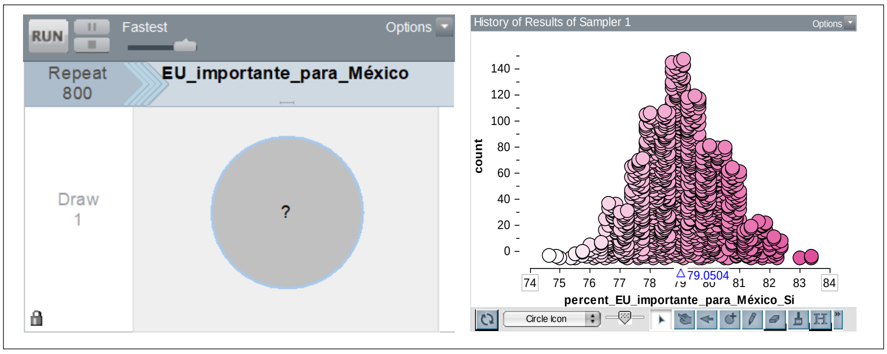

Figura 8 – Modelo oculto y distribución muestral empírica para muestras de 800 personas (1,000 muestras)
Fuente: Elaboración de los estudiantes

Algunos estudiantes no centraron el intervalo, tal es el caso de Nayeli que en la parte izquierda coloca el 2% y en la parte derecha un 7%, cuando debiera hacerlo con 5% de cada lado (ver Figura 9b). La distribución muestral también sirvió de base para estimar el margen de error. Un total de siete estudiantes proporcionaron un margen de error que oscila entre 2% y 3% obtenido de visualizar la distancia entre el centro de la distribución a una de los límites del intervalo, lo cual es consistente con los resultados que proporciona la encuestadora.

Arlin: El margen de error para un nivel de confianza del 90% es entre 2.13% y 2.24%
Brisa: El margen de error sería un aproximado de 2.5 %.
(Cuestionario, 2017).

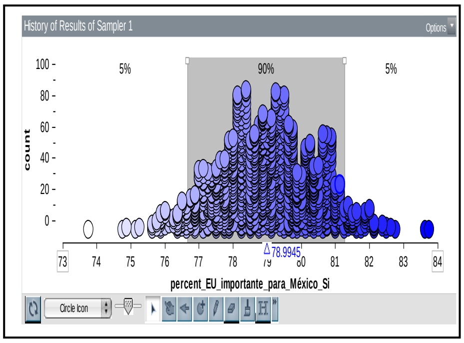

Figura 9a – Distribución muestral de Arlín

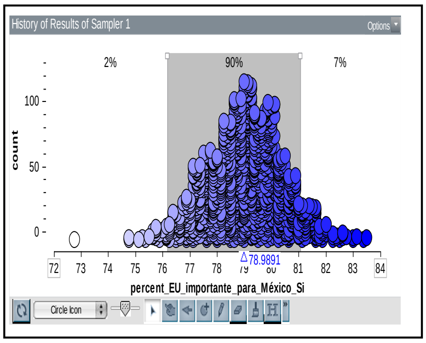

Figura 9b – Distribución muestral de Nayeli
Fuente: Elaboración de los estudiantes

Al considerar un nivel de confianza del 95%, en general, los estudiantes identifican un intervalo ligeramente más amplio que en el caso de 90% (ver Figura 10).

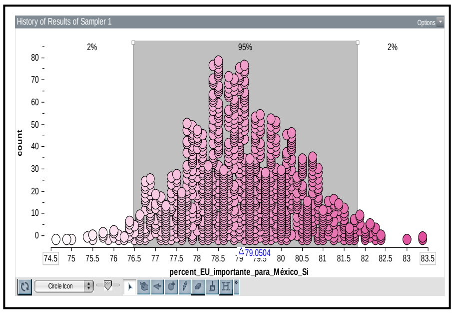

Figura 10 – Distribución muestral indicando con área sombreada de 90%
Fuente: Elaboración de los estudiantes

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

Arlin: Si se aumenta el nivel de confianza podemos observar que el margen de error aumenta.
Brisa: El margen de error aumenta proporcionalmente al porcentaje del nivel de confianza.
Frida: El margen de error aumenta.
Sebastian: Que tienes mayor precisión en el resultado.
(Cuestionario, 2017).

Y otros mencionan la cantidad de muestras, no amplitud ni precisión:

Jacobo: El margen error al variar daña el nivel de confianza del resultado.
Nayeli: Existen más muestras dentro del nivel de confianza, por lo que es menos probable el margen de error.
Jazmín: Su efecto es que se puede delimitar de manera más especifica un posible margen de error, y al ampliar se hace más estrecho el margen, ya que se consideran más muestras.
(Cuestionario, 2017).

4.5 Cálculo de intervalos de confianza y exploración de sus propiedades

La última actividad tenía como propósito explorar los conceptos de margen de error, nivel de confianza y el efecto del tamaño de muestra. El contexto de la actividad proviene de una encuesta sobre la despenalización del consumo de mariguana en México. La pregunta que hemos elegido para diseño de la actividad se refiere a la opinión sobre la legalización de la mariguana, en la cual el 77% responde en contra, el 20% a favor y 3% aún no sabe. Para la exploración se requería utilizar la hoja de cálculo del software para introducir las fórmulas de los elementos del intervalo para una proporción: $p \pm Z\sqrt{\frac{p(1-p)}{n}}$, la cual se explicó antes de utilizarla y se mostró un video donde se mostraba el cálculo de un intervalo. La simulación repetida del muestreo genera, de manera automática, la primera columna de la hoja de cálculo, el resto de las columnas requiere de la introducción de la fórmula correspondiente (ver Figura 11).

History of Results of Sampler 1 — Collect 999 — Options

	prop_LEGALI...	Error_Estandar	Margen_de_Error	Límite_Inferior	Límite_Superior	Resultado
993	0.765	0.0149906	0.0293816	0.735618	0.794382	Cae dentro
994	0.79375	0.0143052	0.0280382	0.765712	0.821788	Cae dentro
995	0.7525	0.0152579	0.0299056	0.722594	0.782406	Cae dentro
996	0.78	0.0146458	0.0287058	0.751294	0.808706	Cae dentro
997	0.765	0.0149906	0.0293816	0.735618	0.794382	Cae dentro
998	0.80375	0.0140417	0.0275218	0.776228	0.831272	Cae fuera
999	0.78625	0.014494	0.0284082	0.757842	0.814658	Cae dentro
1000	0.77625	0.0147346	0.0288797	0.74737	0.80513	Cae dentro

Figura 11 – Hoja de cálculo del *software* con los elementos de un intervalo de confianza
Fuente: Elaboración de los estudiantes

Una vez realizada la simulación de quinientas muestras, se solicitó la gráfica del margen de error, con la idea de explorar su distribución (ver Figura 12).

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

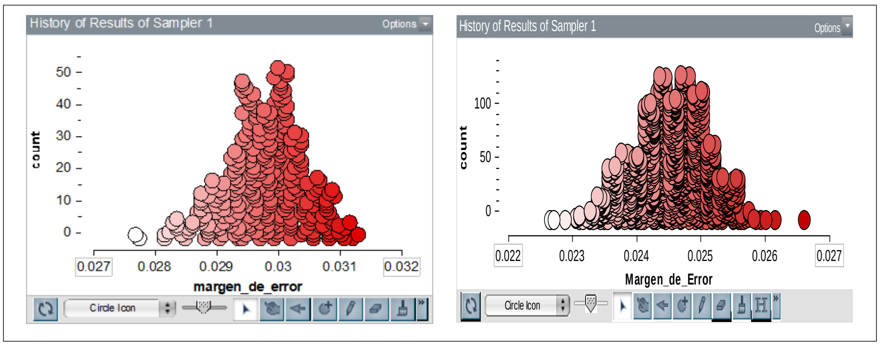

Figura 12 – Gráficas del margen de error construidas para 95% y 90% de confianza respectivamente
Fuente: Elaboración de la estudiante Frida

Explotar visualmente la gráfica anterior ayudó que muchos estudiantes comprendieran la variabilidad del margen de error, que, en algunos casos, puede ser casi cero, pero que en otros casos puede ser mayor al 3%, muy cercano al 3.5% reportado en la encuesta. Por ejemplo, Frida observó que para un nivel de confianza de 95%, el margen de error osciló entre 2.6% y 3.1% aproximadamente, pero cuando el nivel de confianza fue de 90%, el margen de error no sobrepasó el 2.7% para las quinientas muestras simuladas (ver Figura 12).

> *Frida: Si se disminuye el nivel de confianza se tiene un menor margen de error.*
> *Nayeli: Tienen un margen de error máximo de un 3.2%, muy cercano al de la encuesta.*
> *Aglae: El nivel de confianza aumenta el margen de error que se presenta en las graficas.*
> *Julio: Conforme se aumenta el margen de error la grafica se dispersa un poco más.*
> *Marielena: El efecto del nivel de confianza en el margen de error es que este se reduce al reducir el porcentaje de nivel de confianza.*
> (Cuestionario, 2017).

La otra componente que nos propusimos investigar era el significado de nivel de confianza. Para ello, se indicó que construyeran una gráfica con los resultados de la última columna. Se pretendía que los estudiantes relacionaran el nivel de confianza elegido (90% o 95%) con el porcentaje de muestras que capturan o no capturan el parámetro, para comprender el nivel de confianza como la cantidad de veces en un muestreo repetido que el intervalo de estimación contiene al parámetro (ver Figura 13).

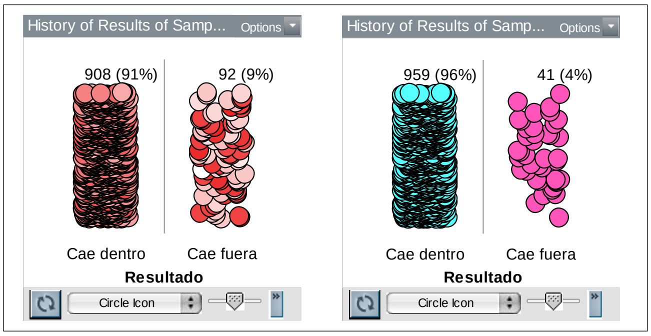

Figura 13 – Cantidad de muestras cuya proporción que cae dentro o fuera del intervalo de confianza
Fuente: Elaboración de los estudiantes

Algunas respuestas representativas se muestran a continuación:

Frida: Si el nivel de confiabilidad aumenta más muestras caen dentro del intervalo de confianza; si se disminuye, caen más muestras fuera de este.
Marielena: Mientras más grande es el nivel de confianza, más muestras entran dentro del intervalo de confianza.
Aglae: La relación que hay entre el nivel de confianza y el porcentaje de muestras, es que el nivel de confianza decide el porcentaje de muestras que están dentro del margen de confianza.
Julio: El nivel de confianza determina que tan probable es que caigan los resultados dentro de un intervalo, sí cae un mayor numero de muestras dentro del intervalo de confianza quiere decir que los datos son correctos o la probabilidad era acertada.
Nayeli: Son muy cercanos, ya que el porcentaje de muestras que caen dentro del intervalo de confianza es muy similar a este, incluso en ocasiones puede ser igual.
Arlin: Se puede analizar que entre menor es el porcentaje de confianza va a darse un menor margen de error ya que hay menos muestras dentro del parámetro, es decir caen mas muestras por fuera, en un nivel de confianza de 95% solo un 4% caen fuera mientras que en un 90% un 9% de las muestras caen fuera.
(Cuestionario, 2017).

4.6 Resultados del cuestionario AIRS

Una vez concluida la secuencia de actividades se evaluó el razonamiento inferencial desarrollado por los estudiantes respecto a los conceptos involucrados en la investigación. Para ello, utilizamos diecisiete ítems del cuestionario AIRS (*Assessment Inferential Reasoning in Statistics*) desarrollado por Park (2012). Los objetivos de aprendizaje, el razonamiento pretendido y el número de respuestas correctas se muestran en el Cuadro 1. Se obtuvo un promedio de diez reactivos correctos por estudiante con un rango de variación de siete a trece reactivos correctos.

Ítem	Objetivo de aprendizaje	Razonamiento pretendido	Correctas
1	Razonar sobre la incertidumbre al hacer una inferencia utilizando lenguaje probabilístico.	Se reporta un 70% de probabilidad de lluvia, el intervalo para la proporción poblacional de lluvia debería incluir al 70%.	5 (45%)
2	Comprender la naturaleza y comportamiento de la variabilidad muestral. Comprender la asociación entre tamaño de muestra y variabilidad muestral.	Las proporciones de dulces café en 10 muestras de 100 dulces serán más cercanas a la proporción media poblacional (0.5) que en 10 muestras de 10 dulces, debido a que muestras más pequeñas tienen más variabilidad.	4 (36%)
3.1	Razonar sobre el valor esperado de una muestra particular conocida la estructura poblacional.	Si una ruleta con cuatro letras (A, B, C, D) está equilibrada, el número de letras que aparecen sería igualmente probable, ya que hay cuatro posibilidades y cada una de las letras tiene una probabilidad de un cuarto. Entonces, es usual que aparezcan entre 2 o 3 letras B en 10 giros.	11 (100%)
3.2	Construir el modelo de simulador de la población en un contexto dado.	Si se asume que la ruleta está equilibrada, cada letra (A, B, C, D) tiene una probabilidad de un cuarto.	9 (82%)

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

3.3	Razonar sobre una muestra inusual en un contexto dado, conocida la distribución muestral para un tamaño dado de muestra.	Un resultado de 5B en 10 giros es inusual si la ruleta está equilibrada, porque en la distribución de 100 muestras hay únicamente 4 casos donde 5B o más sucedieron en 10 giros.	5 (45%)
3.4	Generalizar y obtener conclusiones de una población a partir de la distribución muestral.	La ruleta es más probable que esté equilibrada porque 2Bs y 3Bs aparecieron en la mayoría de las simulaciones.	2 (18%)
3.5	Razonar y articular sobre la relación entre tamaño de muestra y la forma de la distribución muestral	La distribución de la proporción de B obtenida de 100 muestras de 20 giros debe ser más angosta que en muestras de 10 giros, porque hay menos variabilidad en una muestra más grande.	8 (73%)
3.6	Hacer una conclusión de una población desde una muestra en asociación con el cambio en el tamaño de muestra	Debido a que 100 muestras de 20 giros tienen una distribución más angosta que una de una de 10 giros es menos probable obtener un resultado inusual con una ruleta equilibrada. Por lo tanto, 100 muestras de 20 giros pueden tener una evidencia más fuerte para apoyar que la ruleta no esté equilibrada.	7 (64%)
4.1	Razonar sobre una colección de datos a partir de casos individuales de un agregado.	Inválido. La afirmación se enfoca en algunos datos no en la tendencia general de los datos.	6 (55%)
4.2	Razonar sobre posibles diferencias entre dos poblaciones basados en las diferencias observadas entre dos muestras de datos.	Válido. Porque el promedio de tiempo para el grupo que tomó la nueva fórmula es menor que el promedio de la vieja fórmula.	7 (64%)
4.3	Razonar sobre posibles diferencias entre dos poblaciones basados en diferencias observadas entre dos muestras de datos.	Inválido. Aunque los tamaños de muestra son diferentes para los dos grupos, podemos obtener una conclusión debido a que ambos tamaños de muestra son grandes.	6 (55%)
5	Comprender la definición de distribución muestral. Comprender el papel de la distribución muestral.	Ya que se desea calcular el costo de una muestra de 25 libros, necesitamos la distribución muestral de todas las muestras de tamaño 25 de la población (universidad).	6 (55%)
6.1	Comprender la relación entre distribución de una muestra y distribución población.	Una sola muestra aleatoria de 500 valores puede ser representativa (parecida) de una población.	5 (45%)
6.2	Comprender la relación entre distribución muestral y distribución población.	Una distribución de 500 muestras se comporta de acuerdo con el teorema del límite central: normalmente distribuida, centrada en la media y con menos variabilidad.	4 (36%)
7	Comprender el efecto del tamaño de muestra en la distribución muestral. Comprender cómo el error de muestreo está presente y relacionado con una inferencia.	La muestra más grande y la media muestral más pequeña proporcionan la más fuerte evidencia al enunciado.	4 (36%)
8.3	Interpretar la relación entre	Si el nivel de confianza se	8 (73%)

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

	intervalo de confianza y margen de error.	incrementa, el margen de error se incrementa, por lo tanto, el rango de valores es más amplio.	
9	Comprender la importancia del muestreo aleatorio (reconocer un muestreo sesgado)	Es un muestreo sesgado porque la muestra no es representativa de la población.	6 (55%)

Cuadro 1 – Objetivos de aprendizaje, razonamiento pretendido y frecuencia de respuestas correctas
Fuente: Elaboración propia

5 Conclusiones

Los resultados de la evaluación de la THA muestran que es posible razonar adecuadamente con conceptos complejos que subyacen a una inferencia estadística desde una perspectiva informal, utilizando datos con contextos reales y herramientas computacionales dinámicas e interactivas que permiten visualizar el muestreo y sus resultados en tiempo real.

El contexto de encuestas nos parece apropiado para introducir a los estudiantes al estudio de la inferencia estadística, y en particular a los intervalos de confianza, ya que en ellas se involucran conceptos como el margen de error, nivel de confianza, tipo de muestreo, tamaño de muestra, incertidumbre, entre otros, que pueden ser discutidos en su contexto con significado para los estudiantes.

El poder de simulación y visualización del *software* utilizado ha permitido 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.

Los resultados de la evaluación final muestran grupos de conceptos que resultaron más difíciles que otros para los estudiantes, como la distinción entre población, muestra y distribución muestral de un estadístico, propiedades de las distribuciones muestrales e intervalos de confianza en los cuales cinco o menos de los once estudiantes respondieron correctamente.

Sin embargo, hubo conceptos donde el porcentaje de correctas fue superior al 70% como fue el caso del valor esperado de una muestra, construcción del modelo, la relación entre intervalo de confianza y margen de error y relación entre tamaño de muestra y distribución muestral. De acuerdo con los principios de las THA, estos resultados servirán para un rediseño de actividad y modificación de la secuencia de actividades de la THA.

Algunas limitaciones que observamos en la THA son la falta actividades para formar una secuencia más completa e integral de los intervalos de confianza, que contemplen contextos más allá de las encuestas y que incluyan poblaciones grandes con datos reales,

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

diseño de la simulación con otras herramientas de *software* como hojas de cálculo y recursos en línea, así como el uso de trabajo colaborativo en equipos de estudiantes.

Referencias

BAKKER, A.; DERRY, J. Lessons from inferentialism for statistics education. **Mathematical Thinking and Learning**, London, v. 13, n. 1-2, p. 5-26, 2011.

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.; BOROVCNIK, M. **Statistics and Probability in High School**. First ed. Rotterdam: Sense Publishers, 2016.

BEN-ZVI, D.; GIL, E.; APEL, N. What is hidden beyond the data? Helping young students to reason and argue about some wider universe. **Proceedings of the 5th International Research Forum on Statistical Reasoning, Thinking and Literacy.** Warwick: University of Warwick, UK, 2007. p. 1-26.

BIEHLER, R. et al. Technology for Enhancing Statistical Reasoning at the School Level. In: CLEMENTS, M. et al. (Ed.). **Third International Handbook of Mathematics Education**. New York: Springer, 2013. p. 643-689.

CASTRO-SOTOS, A. E. et al. P. Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. **Educational Research Review.** Amsterdam, v. 2, n. 2, p. 98-113, 2007.

CHANCE, B. et al. The role of the technology in improving student learning of statistics. **Technology Innovations in Statistics Education**, Los Angeles, v. 1, n. 1, p. 1-26, 2007

COBB, G. The introductory statistics course: a Ptolemaic curriculum? **Technology Innovations in Statistics Education**, Los Angeles, v. 1, n. 1, p. 1-15, 2007

COBB, P.; MCCLAIN, K. Principles of Instructional Design for Supporting the Development of Students' Statistical Reasoning. In: BEN-ZVI, D.; GARFIELD, J. (Ed.). **The Challenge of Developing Statistical Literacy, Reasoning, and Thinking**. Dordrecht: Kluwer Academic Publishers, 2004. p. 375-396.

DIERDORP, A. et al. Supporting Students to Develop Concepts Underlying Sampling and to Shuttle Between Contextual and Statistical Spheres. In: BEN-ZVI, D.; MAKAR, K. (Ed.). **The Teaching and Learning of Statistics: International Perspectives.** New York: Springer, 2016. p. 37-50.

GARFIELD, J.; BEN-ZVI, D. **Developing Students' Statistical Reasoning**: Connecting Research and Teaching Practice. First ed. New York: Springer, 2008.

GIL, E.; BEN-ZVI, D. Explanations and context in the emergence of students' informal inferential reasoning. **Mathematical Thinking and Learning**, London, v. 13, n. 1-2, p. 87-108, 2011.

INZUNZA, S. Entornos virtuales de aprendizaje: un enfoque alternativa para la enseñanza de la inferencia estadística. **Revista Mexicana de Investigación Educativa**, Ciudad de México, v. 15, n. 45, p. 423-442, 2010.

KONOLD, C.; MILLER, C. **TinkerPlots Dynamic Data Exploration Software**. Versión 2.3.4, Emeryville, CA: Key Curriculum Press Technologies, 2011.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

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

MAKAR, K.; RUBIN, A. A framework for thinking about informal statistical inference. **Statistics Education Research Journal**, San Luis Obispo, v. 8, n. 1, p. 82-105, 2009.

MELETIOU-MAVROTHERIS, M. Technological Tools in the Introductory Statistics Classroom: Effects on Student Understanding of Inferential Statistics. **International Journal of Computers for Mathematical Learning**, New York, v. 8, n. 3, p. 265-297, 2003.

MINISTRY OF EDUCATION. **The New Zealand Curriculum: Mathematics and Statistics**. 2018. Recuperado de: http://nzcurriculum.tki.org.nz/The-New-Zealand-Curriculum/Mathematics-and-statistics.

MOORE, D. S. Uncertainty. In: STEEN, L. A. (Ed.). **On the Shoulders of Giants**: New Approaches to Numeracy. Washington: National Academy Press, 1990. p. 95-137.

NATIONAL COUNCIL OF TEACHERS OF MATHEMATICS. **Principles and Standards for School Mathematics**. First ed. Reston: NCTM, 2000.

NOLL, J.; GEBRESENBET, M.; GLOVER, E. D. Modeling and Simulation Approach to Informal Inference: Successes and Challenges. In: BEN-ZVI, D.; MAKAR, K. (Ed.). **The Teaching and Learning of Statistic**s: International Perspectives. NewYork: Springer, 2016. p. 139-150.

PARK, J. **Developing and Validating an Instrument to Measure College Students' Inferential Reasoning in Statistics**: **An Argument-Based Approach to Validation**. 304 páginas. Doctoral Dissertation in Mathematics Education. Universidad de Minnesota, 2012.

PEA, R. Cognitive technologies for mathematics education. In: Schoenfeld, A. (Ed.). **Cognitive Science and Mathematics Education**. New Jersey: Lawrence Erlbaum Associates, 1987. p. 89-122.

PFANNKUCH, M. Building sampling concepts for statistical inference: A case study. **Proceedings of the 11th International Congress on Mathematical Education.** Monterrey México. 2008. p. 1-8.

PFANNKUCH, M. Informal inferential reasoning. **Proceedings of the 7th International Conference on Teaching Statistics**, Salvador Bahía Brazil. Voorburg: International Statistical Institute 2006. p. 1-6.

PFANNKUCH, M.; WILD, CH.; PARSONAGE, R. A conceptual pathway to confidence intervals. **ZDM Mathematics Education**, Berlín, v. 44, n. 7, p. 899-911, 2012.

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.

RUBIN, A.; HAMMERMAN, J.; KONOLD, C. Exploring informal inference with interactive visualization software. **Proceedings of the 7th International Conference on Teaching Statistics**. Salvador Bahía**.** Voorburg: International Statistical Institute, 2006. p. 1-6.

SALDANHA, L.; THOMPSON, P. Conceptual issues in understanding the inner logic of statistical inference: Insights from two teaching experiments. **Journal of Mathematical Behavior**, Amsterdam, v. 35, p. 1-30, 2014.

SIMON, M. Hypothetical Learning Trajectories in Mathematics Education. In: Lerman. S. (Ed.). **Encyclopedia of Mathematics Education**, New York, Springer, 2014. p. 272-275.

ISSN 1980-4415
DOI: http://dx.doi.org/10.1590/1980-4415v33n63a01

SIMON, M.; TZUR, R. Explicating the role of Mathematics tasks in conceptual learning. **Mathematical thinking and learning**, London, v. 6, n. 2, p. 91-104, 2004.

WILD, CH.; PFANNKUCH, M.; REAGAN, M. Towards more accessible conceptions of statistical inference. **Journal of the Royal Statistical Society Series A**., London, v. 174, n. 2, p. 247-295, 2011.

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 27 de Junho de 2017.
Aprovado em 24 de Julho de 2018.