
EL PAPEL DEL BOOTSTRAP PARAMÉTRICO EN EL DESARROLLO DEL RAZONAMIENTO INFERENCIAL INFORMAL ASOCIADO A INTERVALOS DE CONFIANZA

Jonatan, López Cacia

jonatan1290@hotmail.com

Universidad Industrial de Santander (Colombia)

Gabriel, Yáñez Canal

gyanez@uis.edu.co

Universidad Industrial de Santander (Colombia)

Asunto: Desarrollo del pensamiento aleatorio

Temática: Estadística descriptiva e inferencial

RESUMEN

Comúnmente los estudiantes, como resultado de una enseñanza formal, consideran que el establecimiento de inferencias por medio de intervalos de confianza, se reduce a la aplicación de un algoritmo donde no es necesario comprender los conceptos que soportan el método. Por ello, consideramos necesario un enfoque diferente de enseñanza para la inferencia, cuyo objetivo es desarrollar el Razonamiento Inferencial Informal [RII] con el apoyo de simulaciones computacionales y sin necesidad de cálculos formales. Así, diseñamos y aplicamos una serie de actividades en Fathom involucrando el Bootstrap, con la intención de describir el papel de las simulaciones computacionales en el desarrollo del RII asociado a intervalos de confianza, con estudiantes de primer nivel universitario que no habían recibido hasta el momento instrucción formal de estadística.

244

PALABRAS CLAVE

Intervalos de confianza, Razonamiento inferencial informal, Bootstrap.

INTRODUCCIÓN

Los intervalos de confianza se constituyen en uno de los métodos de inferencia más enseñados en los cursos de Estadística, sin embargo, varias investigaciones han mostrado que los estudiantes logran una comprensión incompleta, o incluso nula, de la verdadera naturaleza del método y sus componentes, creyendo que el establecimiento de inferencias se limita a la simple aplicación de un algoritmo.

En el trabajo de Cumming, William y Fidler (2014) se muestra que es común que el estudiante subestime la variabilidad de la media muestral en las replicaciones, al creer que un intervalo del 95% de confianza captura el 95% de las medias de futuras replicaciones. Adicionalmente, a estos problemas de interpretación se les agregan confusiones con respecto a las relaciones entre algunos componentes de los intervalos de confianza. Por ejemplo, Behar (2001) encuentra que los estudiantes no perciben en un intervalo de confianza relaciones entre el nivel de confianza y el ancho del intervalo, al tiempo que desconocen la influencia del tamaño muestral sobre la variabilidad y por ende sobre el ancho del intervalo.

Por otra parte, Yáñez y Behar (2009) ponen en evidencia dificultades en estudiantes y profesores, para la comprensión de los conceptos básicos asociados a los intervalos de confianza, tales como población, parámetro, muestra, media, valor esperado, estimador, estadístico, variable aleatoria, distribución de probabilidad, distribución muestral, desviación estándar, z-valores, probabilidad frecuencial, nivel de confianza y precisión de la estimación. Si bien cada uno de estos conceptos tiene su propia complejidad, tener que integrarlos y relacionarlos exige un esfuerzo mayor (Yáñez y Behar, 2009).

Desde otro punto de vista, Rangel (2014) aporta una nueva concepción errónea a este listado, al encontrar que los estudiantes consideran que si el nivel de confianza es interpretado como la probabilidad de que el intervalo construido contenga la media poblacional, entonces se ganará precisión al aumentar esta probabilidad, porque estaría más seguro de que la media poblacional estará contenida allí, asumiendo erróneamente que se gana precisión al aumentar el nivel de confianza.

Para afrontar esta problemática han aparecido trabajos que sugieren un camino diferente para la enseñanza de la inferencia, un camino donde se permita al estudiante utilizar su propia experiencia con la variabilidad, su intuición y su sentido común en el momento de hacer inferencias, haciendo uso de lo que la comunidad internacional ha llamado el Razonamiento Inferencial Informal [RII] (Zieffler, Garfield, delMas & Reading, 2008).

La idea es permitir al estudiante desarrollar mayor experiencia con la variabilidad, generando variadas muestras y analizando directamente la distribución del estimador. Para realizar esto se han propuesto las simulaciones computacionales como una buena alternativa, siendo la simulación Bootstrap la más sugerida cuando de intervalos de confianza se trata (Pfannkuch, Wild & Parsonage, 2012).

Dentro de esta línea de pensamiento, realizamos una investigación con el objetivo de determinar el papel de las simulaciones computacionales en el desarrollo del

Razonamiento Inferencial Informal asociado a Intervalos de confianza en estudiantes de primer nivel universitario, que no habían realizado ningún curso formal de Estadística.

MARCO DE REFERENCIA

El Razonamiento Inferencial Informal

Este trabajo comparte la idea de que una posible solución a la problemática del aprendizaje de los Intervalos de Confianza puede estar dada por un enfoque informal, donde se promueva el uso del sentido común y la verdadera comprensión de las ideas asociadas con un uso mínimo de cálculos formales. De tal manera, las actividades deben promover el uso y el desarrollo del RII que no es otra cosa que la forma en que los estudiantes utilizan y articulan su conocimiento informal de estadística para dar argumentos en apoyo a las inferencias sobre poblaciones desconocidas, a partir de muestras observadas (Zieffler, Garfield, delMas & Reading, 2008).

Así, para el planteamiento de las tareas, hemos tenido en cuenta el marco de referencia de Zieffler et al, (2008) el cual exige especial cuidado en que las situaciones propuestas desafíen al estudiante a:

1. Hacer afirmaciones o predicciones sobre la población basadas en la muestra, pero sin hacer uso de procedimientos o métodos formales.
2. Emplear e integrar sus conocimientos previos (formales e informales) en la medida en que estos conocimientos estén disponibles.
3. Articular argumentos basados en la evidencia para establecer juicios, afirmaciones y predicciones sobre las poblaciones a partir de muestras.

246

A su vez, las tareas están enfocadas en lograr una comprensión de los intervalos de confianza como proceso estocástico, para lo cual tomamos como referencia el marco propuesto por Pfankuch, Wild y Parsonage (2012) donde establecen que los estudiantes deben desarrollar, lo que ellos llaman, imágenes mentales que les van a permitir comprender más adelante el concepto formal de intervalo de confianza.

La primera parte del marco de referencia (Imagen 1) reseña la necesidad de que el estudiante reconozca que la situación requiere una forma estocástica de pensar, en otras palabras, el estudiante debe ser capaz de identificar si la situación requiere del establecimiento de inferencias y se da cuenta de la variabilidad presente en el muestreo, ya que cualquier conclusión que pueda establecer está lejos de ser una respuesta determinista. La segunda parte hace referencia a la concepción de la situación como un proceso estocástico, lo que implica la formación de imágenes mentales acerca de la necesidad de las repeticiones con diversas muestras para percibir la variación y la

distribución muestral lo que conlleva a la construcción de una imagen intuitiva de los intervalos de confianza y del efecto del tamaño muestral en su tamaño.

Reconocimiento de la situación:

- Imagen de la distribución de la población.
- Imagen de la relación de la distribución muestral y la distribución de la población.

Concibiendo la situación como un proceso estocástico

- Imagen de proceso repetitivo.
- Imagen de la franja muestreo-variación.
- Imagen del efecto del tamaño de la muestra.
- Imagen de la distribución muestral.
- Imagen de un intervalo de confianza intuitivo.
- Interpretación de un intervalo de confianza de forma empírica.

Imagen 1. Aspectos para la comprensión de los intervalos de confianza como un proceso estocástico

Fuente. Pfankuch, Wild y Parsonage (2012)

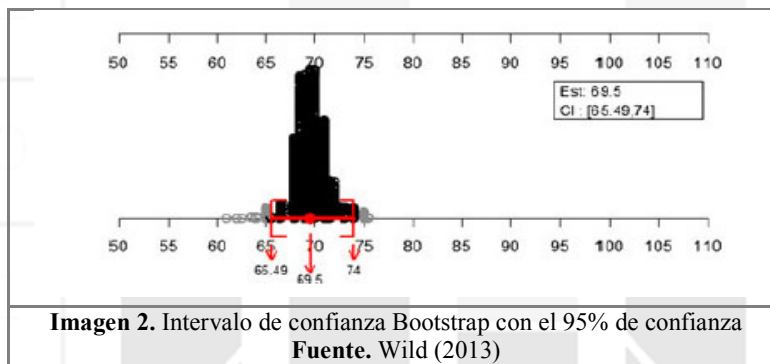
Para lograr la formación de estas imágenes mentales el estudiante debe entender primero que el método está considerando la toma de muestras de igual tamaño de manera aleatoria una gran cantidad de veces, es decir, se trata de un proceso repetitivo. El generar varias muestras implica que el estadístico de interés tendrá un valor diferente en cada muestra, sin embargo, esta variación presenta una tendencia, cierto comportamiento, que se concreta en la distribución muestral del estadístico cuya variación se hace menor si se aumenta el tamaño de la muestra. La secuencia de tareas que se implementan en esta investigación busca la interiorización de las ideas expresadas en el marco.

Intervalos Bootstrap

El Bootstrap fue introducido por Efron (1979) y básicamente es un método de simulación computacional intensivo, que consiste en la aplicación de un proceso de remuestreo para estimar la distribución muestral de un estimador basándose tan solo en una muestra disponible, bajo el supuesto de que ésta es un buen representante de la población (Efron, 2000). Se ha demostrado que el método funciona bien en muchas situaciones y ha sido aceptado por la comunidad estadística, incluso se considera que, en algunos casos, es mejor que la tradicional aproximación normal.

Básicamente el proceso para construir un intervalo de confianza Bootstrap:

1. De la única muestra disponible se toman datos aleatoriamente con reposición para generar una re-muestra del mismo tamaño que la original. Todos los datos de la muestra tienen la misma probabilidad de ser elegidos y un dato puede aparecer varias veces en una re-muestra.
2. El proceso descrito en el paso 1 se debe repetir una gran cantidad de veces, calculando el valor del estadístico de interés en cada una de las re-muestras.
3. Se construye la distribución de los valores del estadístico calculados en el paso 2, ésta se conoce como la distribución Bootstrap (Imagen 2).
4. Si se quiere construir un intervalo con un 95% de confianza se deben calcular los percentiles 2,5 y 97,5 de la distribución Bootstrap (Imagen 2).



DESARROLLO DEL TEMA

Aspectos metodológicos

Esta investigación se llevó a cabo con 20 estudiantes de primer semestre de ingeniería civil de la Universidad Industrial de Santander, de los cuales 4 fueron elegidos para ser analizados debido a su disposición, participación y facilidad de comunicación. Ninguno de los estudiantes había recibido instrucción formal en Estadística. Al grupo de participantes se le aplicó una prueba inicial con la intención de hacer un diagnóstico de su razonamiento inferencial informal y el estado de sus conocimientos sobre conceptos asociados a los Intervalos de Confianza como, la variabilidad, enfoque frecuencial de la probabilidad, ley de los grandes números e influencia del tamaño muestral.

Se diseñaron y aplicaron 5 actividades para ser trabajadas con Fathom (Finzer, 2005); cada estudiante dispuso de su computador y su guía de trabajo en cada sesión. Al finalizar cada actividad se hacía una discusión grupal donde podían participar todos los presentes. La actividad 1 buscaba estudiar el comportamiento de la proporción muestral, extrayendo muestras con reposición de una urna con composición conocida. La actividad 2 tuvo un sentido contrario a la primera actividad, en esta ocasión los estudiantes debían estimar la proporción de bolas de un color contenidas en una urna, extrayendo muestras del mismo tamaño (cada extracción con reposición). En la actividad 3 el estudiante debía determinar

por su propia cuenta, la media de una población a partir de muestras aleatorias del mismo tamaño. En la actividad 4 se hizo una comparación entre la distribución muestral de la media y la distribución Bootstrap. Finalmente, en la actividad 5, el estudiante debía enfrentar dos situaciones que requerían de la inferencia a partir de una sola muestra. La prueba final, tuvo como objetivo observar cuales de las concepciones erróneas sobre Intervalos de Confianza, que han sido reportadas como producto del enfoque formal, aparecían también en los estudiantes que hicieron parte de la investigación.

La información de las actividades fue recogida de 3 formas:

1. Escrita. Cada estudiante debía consignar detalladamente sus respuestas y razonamientos en su hoja de trabajo.
2. Oral. Durante el desarrollo de las actividades el investigador pasaba por el puesto de los estudiantes grabando audios con las respuestas de los a las preguntas realizadas por el profesor-investigador.
3. Video. Todos los estudiantes tenían instalado en sus computadores un programa que grabó la pantalla durante todas las actividades.

En cuanto al análisis de los resultados hicimos uso del marco de la Aproximación Instrumental propuesto por Artigue (2002).

Desarrollo de la propuesta

A continuación, se presenta un resumen general del desarrollo observado en uno de los estudiantes (Néstor) durante toda la investigación, basándonos en el análisis de sus respuestas y en el uso de Fathom.

Inicialmente Néstor no consideraba la variabilidad y realizaba estimaciones puntuales. Por ejemplo, a la pregunta: *Suponga que realizó 100 extracciones de una urna con bolas blancas y negras cuyas cantidades son desconocidas y se obtuvo 70 bolas blancas (y 30 bolas negras), es decir, 70% de los resultados fueron bolas blancas. ¿Qué conclusión sacaría respecto a la proporción de bolas blancas en la urna?* Néstor contestó: “probabilidad baja o alta, siempre va existir la posibilidad en que todas salgan de un mismo color ya que existen los dos colores y hay una probabilidad siempre para hacer esta extracción. Para confirmar dicho dato yo haría más repeticiones de la misma cantidad (100 veces) y el promedio me daría algo más cercano a la realidad”.

Observamos que en su primera oportunidad de hacer una estimación pensó en tomar varias muestras del mismo tamaño, calcular la proporción en cada una de ellas y luego calcular el promedio de esas proporciones, es decir, estimar el valor esperado de la proporción, esto como una forma de controlar la gran variabilidad que percibe. Lo anterior

sucedió en con una situación que involucraba variables discretas, pero cuando se trataba de variable continua, Néstor evidenció nociones de estimación por medio de intervalos, con afirmaciones del estilo de “alrededor de 1,62”.

Después de completar las tres primeras actividades, Néstor evidenció avances positivos en sus ideas de variabilidad, en la ley de los grandes números, en el enfoque frecuencial de la probabilidad, y en sus ideas de distribución muestral. Esto por su respuesta: *“Al hacer más intentos observo que cada 1000 veces en la medida hecha de las muestras tomadas de tamaño 50 el promedio me varía un poco los valores y siguen estando cercanos a 195,198 g, en el segundo fue 194,38 y en el tercero 194,97 g por tanto puedo decir que cuando es con decimales estos me varían un poco y la medida debería darla en un intervalo corto entre 194 y 196”*.

Allí repite tres veces el proceso de construcción de la aproximación de la distribución muestral de la media, observando el valor asociado al rectángulo de mayor altura en el histograma y termina estableciendo una estimación por medio de un intervalo. Con las actividades 4 y 5, Néstor ratificó sus buenas nociones: *“Solo nos dan una muestra de 20 personas, entonces saco con re-muestreo otros 20 casos y hago la medida, dije que dejaba el mismo tamaño de la muestra para conservar lo mismo, porque si hago 100 se me empezarían a repetir más valores, entonces aumentaría el error, en cambio con 20 tengo la misma probabilidad de repetir valores pero no aumentar el error, e hice la medida y saqué el promedio, cuando hace uno re-muestreo hay que ser más crítico con el rango que se va a dar entonces yo dije que el promedio estaba entre 173 y 177 ya que son los 5 picos más altos de la gráfica”*. Con lo que demuestra que comprende que se puede utilizar la distribución muestral para establecer inferencias sobre la población; dejó atrás las estimaciones puntuales comprendiendo la poca, realmente nula, probabilidad de acertar y encontró en el Bootstrap un apoyo para establecer estimaciones por intervalos.

Al final de la experiencia, Néstor evidenció no tener del todo clara la relación entre el ancho del intervalo y el nivel de confianza; *“entre más cerca estén los límites del intervalo a la media de la distribución muestral habrá mayor probabilidad de acertar”*, es decir, si el nivel de confianza aumenta, disminuye el tamaño del intervalo. Adicionalmente en algunos de sus argumentos no asocia el ancho del intervalo y la precisión de la estimación. A pesar de esto, en su respuesta final dice: *“Si aumento el intervalo disminuyo la precisión de mi estimación, ya que puede que no me vaya a equivocar dando un intervalo más grande, pero voy a tener más error en este, al tener más error voy a ser menos preciso dando mi estimación en la media poblacional”*. Planteando ideas correctas sobre las relaciones nombradas; considera la pérdida de precisión al aumentar el ancho del intervalo; y demuestra que concibe mayor probabilidad de acierto en un intervalo más ancho.

CONCLUSIONES

Con el trabajo realizado durante esta investigación, a medida que los procesos en Fathom propiciaban el desarrollo de buenas nociones sobre aquellos aspectos implícitos en el método formal. Logramos guiar a los estudiantes para que por sí mismos descubrieran la conveniencia de las estimaciones por medio de intervalos. Creemos que el trabajo permitió apreciar el potencial que la metodología implementada tiene para reducir la aparición de las concepciones erróneas sobre los Intervalos de Confianza que se encuentran reportadas en las investigaciones precedentes: varios estudiantes lograron percibir que al aumentar el tamaño de la muestra se reduce el ancho del intervalo y se gana precisión en la estimación, a su vez varios son conscientes que cuando se aumenta el ancho del intervalo sin variar todos los demás elementos, se aumenta la probabilidad de acertar.

Teniendo en cuenta el marco propuesto por Pfankuch, Wild y Parsonage (2012) para la comprensión de los Intervalos de Confianza como un proceso estocástico hemos logrado generar en los estudiantes:

- Imágenes sobre la relación entre la distribución muestral y la población, ya que durante las actividades en varias ocasiones por sí mismos simulaban la distribución muestral o la distribución Bootstrap según el caso, y las utilizaron como un apoyo en el establecimiento de sus inferencias.
- Imágenes que les permiten entender que el muestreo es un proceso repetitivo. Tanto en la construcción de la distribución Bootstrap como en la construcción de la distribución muestral, los estudiantes comprendían que se trataba de la simulación de un proceso aleatorio en que, según el caso, se toma una gran cantidad de muestras y re-muestras.
- Imagen de la franja de variabilidad en el muestreo. Inicialmente los estudiantes asociaban a la variabilidad muestral con un comportamiento descontrolado, pero a medida que iban transcurriendo las actividades empezaron a considerar que los valores de un estadístico, aunque son diferentes de una muestra a otra, presentan una tendencia en su comportamiento, es decir, responden a una distribución.
- Imagen del efecto del tamaño de la muestra. Los estudiantes desde la primera actividad pudieron notar que, al aumentar el tamaño de la muestra, la variabilidad en la distribución muestral se reducía, esta idea se evidenció en ellos hasta las últimas actividades y la mayoría la utilizó para el establecimiento de sus intervalos.
- Imagen de intervalo de confianza intuitivo. Como se ha dicho anteriormente, los estudiantes iniciaron su recorrido con una tendencia al establecimiento de inferencias por medio de valores puntuales y podríamos decir que finalizaron asociando a este tipo de estimaciones una probabilidad nula de acertar, incluso les

fue más natural el uso de intervalos para establecer estimaciones a partir de una sola muestra. Por lo visto durante esta investigación, ellos consideran que una estimación por intervalo es un conjunto de valores que incluye el valor desconocido del parámetro poblacional. En el trabajo con el Bootstrap manifestaron la preocupación de que su intervalo no atrapara el valor del parámetro y deciden aumentar su tamaño, teniendo cuidado de no aumentarlo “mucho” ya que consideraban que de esa manera perderían precisión.

- Interpretación de un intervalo de confianza de forma empírica. Esto hace referencia a la interpretación frecuencial del nivel de confianza asociado al intervalo, lo cual no se profundizó durante nuestras actividades ya que hubiese requerido de un mayor tiempo, aunque sí en algunas discusiones se intentó guiar al estudiante hacia estas ideas. Sin embargo, consideramos que es posible diseñar actividades que permitan el estudio de este aspecto de una manera más profunda sin necesidad de formalismos y por medio de simulaciones computacionales en Fathom.

En general, se observó cambios positivos en los razonamientos informales de los participantes, cambios que terminaron influyendo en sus conocimientos sobre el establecimiento de inferencias. Los estudiantes iniciaron utilizando estimaciones puntuales o planteando intervalos intuitivos sin ninguna justificación, una percepción básica de la variabilidad en el muestreo, poca o nula noción de distribución muestral y algunas ideas intuitivas sobre la influencia del tamaño muestral que necesitaban moldearse más.

REFERENCIAS

- Artigue, M. (2002). Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. *International Journal of Computers for Mathematical Learning*, 7(3), 245.
- Behar, R. (2001). Aportaciones para la mejora del proceso de enseñanza aprendizaje de la estadística. Tesis doctoral no publicada, Universidad Politécnica de Catalunya, España.
- Cumming, G., Williams, J., & Fidler, F. (2004). Replication and researchers understanding of confidence intervals and standard error bars. *Understanding statistics*, 3(4), 299-311.
- Efron, B. (1979). The 1977 RIETZ lecture. *The Annals of Statistics*, 7(1), 1-26.
- Efron, B. (2000). The bootstrap and modern statistics. *Journal of the American Statistical Association*, 95(452), 1293-1296.
- Finzer, W. (2005). *Fathom Dynamic Data Software*. KCP technologies.
- Pfannkuch, M., Wild, C. J., & Parsonage, R. (2012). A conceptual pathway to confidence intervals. *ZDM*, 44(7), 899-911.

- Rangel, L. (2014) Estructuras y mecanismos mentales asociados a los Intervalos de Confianza: profesores de matemáticas en formación (Tesis de maestría). Universidad industrial de Santander. Colombia
- Wild, C. (2013). *ISR15: Accessible Conceptions of Statistical Inference: Pulling ourselves up by the bootstrap*. Recuperado de <https://www.stat.auckland.ac.nz/~wild/ISR-15/#Fig1>.
- Yáñez, G. & Behar, R. (2009). Interpretaciones erradas del nivel de confianza en los Intervalos de Confianza y algunas explicaciones plausibles. En M. J. González; M. T. González y J. Murillo (Eds.), *Investigación en Educación Matemática*. Comunicaciones de los grupos de investigación. XIII Simposio de la SEIEM. Santander.
- Zieffler, A., Garfield, J., delMas, R. & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58.