INTERPRETACIONES ERRADAS DEL NIVEL DE CONFIANZA EN LOS INTERVALOS DE CONFIANZA Y ALGUNAS EXPLICACIONES PLAUSIBLES

Gabriel Yáñez Canal
Universidad Industrial de Santander. Colombia.
Roberto Behar Gutiérrez
Universidad del Valle. Colombia.

Resumen

Con base en investigaciones que muestran interpretaciones erradas tanto de profesores como de estudiantes respecto a los intervalos de confianza, en particular respecto al nivel de confianza, se presentan en este trabajo dos intentos de explicación de ellas. El primero es un análisis epistemológico de la construcción de los intervalos de confianza y el segundo la interpretación de entrevistas clínicas con un grupo de profesores en formación al finalizar su segundo curso de estadística básica. Los resultados apuntan a explicar las malas interpretaciones respecto a los intervalos de confianza por confusiones generadas en el proceso estadístico-algebraico que conduce a la construcción de los intervalos y a una indebida extrapolación de la probabilidad asociada a valores muestrales al nivel de confianza de que un intervalo contenga el valor de un parámetro.

Abstract

Based on research showing misinterpretations of both teachers and students about confidence intervals, particularly regarding the level of confidence, are presented in this paper two attempts to explain them. The first one is an epistemological analysis of the construction of confidence intervals, and the second the interpretation of clinical interviews with a group of pre-service teachers at the end of their second year of basic statistics. The results point to explain the misconceptions about confidence intervals because of the confusion generated in the statistical-algebraic process that leads to the construction of the intervals and undue extrapolation of the probability associated with sample values at the level of confidence that an interval contains the value of a parameter.

Yánez, 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.*

Presentación

Un intervalo de confianza (IC) es un rango de valores en el que, con cierto nivel de confianza, podría encontrarse el valor de cierto parámetro que se desea estimar. El nivel de confianza, es la probabilidad a priori de que el intervalo de confianza a calcular contenga al verdadero valor del parámetro. Se indica por 1-α y habitualmente se da en porcentaje (1-α) 100%. Hablamos de nivel de confianza y no de probabilidad ya que una vez extraída la muestra, el intervalo de confianza contendrá al verdadero valor del parámetro o no, lo que sabemos es que si repitiésemos el proceso con muchas muestras podríamos afirmar que el (1-α) % de los intervalos así construidos contendrían el verdadero valor del parámetro (Wikipedia).

La construcción de un intervalo de confianza se realiza a partir de un estadístico calculado con los datos de una muestra. Calculada la desviación típica del estadístico y obtenido un valor crítico asociado a su distribución correspondiente a la mitad del complemento del valor del coeficiente de confianza elegido (α/2), el producto del valor crítico por el error estándar se constituye en el radio del intervalo que está centrado en el valor del estadístico en la muestra obtenida. Este procedimiento general adopta formas particulares de acuerdo al parámetro que se desee estimar (media, proporción, varianza, etc.) y según el conocimiento que se tenga de la población de la cual se quiere estimar el parámetro.

La comprensión de los intervalos de confianza no está exenta de dificultades. Estudios como los de Behar (2001), Fidler (2005), Fidler y Cummings (2005), Bower (2003), Batanero y Olivo (2008) muestran que existen diversas concepciones erradas entre los estudiantes respecto a los IC. Desde el punto de vista conceptual existe la mala concepción, muy generalizada, de asumir los intervalos de confianza como un estadístico descriptivo, en el sentido de que

contienen valores posibles de la media muestral o son algo equivalente a un rango de valores poblacionales, y no como un estadístico inferencial que brinda información acerca del valor de un parámetro poblacional (Behar, 2001; Fidler, 2005; Olivo, 2008).

El significado del nivel de confianza que caracteriza al intervalo ofrece variadas interpretaciones que destacan la dificultad que tiene este concepto. En un estudio por e-mail realizado por Cumming, Williams & Fidler, (2004), se le solicitó a unos investigadores y autores de artículos científicos que estimaran 10 futuras medias muestrales a partir de una media muestral dada y un IC de 95%. El 78% (105 de 134) de los investigadores colocó 9 de las medias dentro del intervalo. Los investigadores dejaron claro que ellos habían intentado ajustar el número de futuras medias dentro de los intervalos al nivel de confianza (95%). Es decir, se interpreta el nivel de confianza como la probabilidad de que el IC contenga los valores de las medias muestrales. Behar (2001) realizó una investigación con 41 expertos (profesionales dedicados a la estadística y/o su enseñanza o estudiantes de último año de la licenciatura en estadística), y 297 estudiantes de diferentes ingenierías quienes respondieron un cuestionario en forma anónima. El 29% de los expertos y la mitad de los estudiantes asumieron el nivel de confianza como el porcentaje de datos poblacionales que están contenidos en el intervalo de confianza; asimismo alrededor del 40% de los expertos y poco más de la mitad de los estudiantes, interpretan la probabilidad como una medida de la certeza de que el IC contenga la media muestral y la media poblacional. Más de la mitad de los expertos y el 65% de los estudiantes no aceptan que a la larga, si se repite el muestreo muchas veces, un porcentaje igual al nivel de confianza de dichos intervalos atraparían el parámetro poblacional μ y que, por consiguiente, existirán algunos intervalos que no lo contienen.

Por los resultados obtenidos se puede afirmar que existen dificultades en el aspecto conceptual de los intervalos de confianza, es decir, sobre el significado mismo de los intervalos de confianza: ¿qué contienen en su interior? ¿Cómo se interpreta su nivel de confianza? El desconocimiento del significado del nivel de confianza se extiende hasta el punto de que una gran cantidad de expertos y de estudiantes no comprenden el efecto que éste tiene sobre la exactitud de la estimación por intervalos (Behar, 2001).

¿Cómo se pueden explicar las dificultades conceptuales que tienen las personas sobre los intervalos de confianza? es la pregunta que surge y que pretendemos responder parcialmente en este trabajo.

Una primera explicación se relaciona con la gran cantidad y complejidad de los conceptos básicos involucrados en los intervalos de confianza y de su necesaria integración. Conceptos tales como, población, parámetro, media, valor esperado, estimador, muestra, estadístico, variable aleatoria, distribución de probabilidad, distribución muestral, desviación estándar y error típico de estimación, distribución normal, distribución normal estándar, z-valores, probabilidad y probabilidad frecuencial, nivel de confianza, exactitud de una estimación, no solo poseen sus dificultades propias sino que también requieren un esfuerzo cognitivo mayor cuando se pretende relacionarlos e integrarlos.

En consecuencia, una comprensión completa de las razones de la dificultad y de las malas concepciones alrededor de los intervalos de confianza implica un estudio previo de las dificultades y concepciones que las personas poseen sobre todos los conceptos involucrados. Ahora bien, los únicos estudios que conocemos sobre las concepciones de las personas alrededor de los intervalos de confianza son descriptivos, adolecen de razones que puedan explicar la lógica que subyace en las respuestas de los sujetos que respondieron a los diferentes cuestionarios, tampoco

relacionan las malas concepciones para los intervalos de confianza con las que puedan existir con los conceptos constitutivos.

Para comenzar a llenar este vacío presentamos en este trabajo un análisis epistemológico de los intervalos de confianza y algunos resultados obtenidos en una investigación realizada con profesores en formación en la Universidad Industrial de Santander, Bucaramanga, Colombia.

El artículo está organizado en 4 secciones en las que se desarrollan los siguientes aspectos:

En la sección 2 realizamos un análisis matemático del procedimiento de construcción de los intervalos de confianza, resaltando algunos aspectos que muchas veces pasan desapercibidos por los docentes y que podrían explicar las malas interpretaciones que muchas personas poseen sobre los intervalos de confianza.

En la sección 3 transcribimos y analizamos apartes de una entrevista realizada con un profesor en formación que se realizó con el ánimo de identificar la lógica de sus respuestas.

En la sección 4 presentamos las conclusiones y algunas sugerencias didácticas para la enseñanza de los intervalos de confianza.

2. Estimación de la media μ por medio de un intervalo de confianza

Convencidos de que clarificar el significado de los objetos matemáticos es el primer paso para desentrañar las razones de las dificultades que implica su aprendizaje (Godino y Batanero, 1998), presentamos el contexto teórico de referencia para construir los intervalos de confianza, haciendo especial énfasis en los aspectos que creemos pueden generar concepciones erradas.

Presentamos el caso particular en que el parámetro a estimar es la media poblacional μ, con la esperanza de que nos dé luces para comprender mejor las razones de las malas concepciones relatadas en la presentación. Como el ánimo es solo mostrar el esquema general de construcción de los IC, lo abordamos en el caso especial de una población normal con varianza conocida. El caso más general de otro parámetro de cualquier población y con varianza desconocida, sólo implica algunos cambios que no afectan la esencia del proceso que aquí presentamos.

Sea X una variable aleatoria con distribución $N(\mu,\sigma)$ y tomemos una muestra aleatoria $(X_1, X_2, X_3, ..., X_n)$ de dicha población.

Se tiene que

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad (1)$$

El resultado (1), resalta que el proceso toma como base la *distribución* de las medias muestrales. Estandarizando la expresión (1) se obtiene la variable aleatoria Z que tiene distribución normal estándar:

$$Z = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad (2)$$

Con base en (2) se construye el intervalo de menor tamaño, centrado en 0 y con una probabilidad de *100(1-α)%:*

$$P\left\{-z_{1-\frac{\alpha}{2}} < \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right\} = 1-\alpha, \quad (3)$$

Donde $z_{1-\frac{\alpha}{2}}$ es tal que $\Phi\left(z_{1-\frac{\alpha}{2}}\right) = 1-\frac{\alpha}{2}$, siendo $\Phi(\cdot)$ la función de distribución acumulativa de probabilidad de la normal estándar.

De la expresión (3), despejando adecuadamente se obtienen sendos intervalos para $\bar{X}_n$ y μ:

$$P\left\{\mu - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \bar{X}_n \le \mu + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha \quad (4)$$

$$P\left\{\bar{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha \quad (5)$$

Como se observa, el procedimiento utilizado conduce a dos intervalos con la misma probabilidad: uno de ellos centrado en μ y que hace referencia a la probabilidad de encontrar valores $\bar{X}_n$ en ese intervalo (a cierta distancia del parámetro μ) y otro centrado en $\bar{X}_n$, que se refiere a la probabilidad de que el intervalo contenga al parámetro μ (a cierta distancia del valor $\bar{X}_n$). El primer intervalo, que lo notamos $\mu \pm \rho$ evidencia el significado del margen de error ρ, ya que al afirmar que $\bar{X}_n$ pertenece a ese intervalo se está afirmando que el error máximo, cuando se usa este valor como estimador de μ, es ρ. Cuando se conoce μ, este es el tipo de intervalo que hace referencia a valores posibles de la media muestral y que se genera cuando se estudian las distribuciones muestrales. El segundo intervalo, que lo notamos $\bar{X}_n \pm \rho$, es el intervalo que con *confianza 100(1-α)%* contiene al parámetro μ, es decir, es *el intervalo de confianza* asociado a la muestra considera. Es importante recalcar que no obstante que los dos intervalos se obtienen a partir de (3), no son ni mucho menos iguales ni tienen tampoco el mismo significado. Tal vez las interpretaciones que se hacen sobre los intervalos de confianza considerándolos como conjuntos de valores de medias muestrales, sean producto de confundir estos dos intervalos, de hecho estos intervalos no son disyuntos.

Desde el punto de vista del enfoque frecuencial de la probabilidad, la expresión (4) significa que a la larga, el *100(1-α)%* de los valores de las medias muestrales obtenidas tomando muestras de la población de tamaño *n*, caen en el intervalo centrado en μ y con radio $\rho = z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$, es decir en el intervalo $\mu \pm \rho$. La interpretación adecuada del nivel de confianza o probabilidad 1-α para la expresión (5), es que en el largo plazo, el 100(1-α)% de los intervalos centrados en $\bar{X}_n$ y con radio $\rho\left(\bar{X}_n \pm \rho\right)$ que se obtienen tomando diferentes muestras de tamaño *n* de la población, van a contener el parámetro μ. La interpretación complementaria, que es muy útil porque aborda el mismo significado desde orillas opuestas, es que en el largo plazo, el 100α% de estos intervalos no contienen el parámetro μ, es decir, puede suceder que el parámetro no se encuentre en el intervalo construido a partir de una media muestral.

3. La investigación

Se realizó una investigación con profesores de educación básica en formación en la Universidad Industrial de Santander, Bucaramanga (Colombia) los cuales tomaron dos cursos semestrales de estadística básica con una intensidad de cuatro horas semanales. El enfoque didáctico que se asumió fue el de resolución de problemas en un ambiente rico en tecnología: se utilizó el paquete Fathom para realizar simulaciones. Los temas que se desarrollaron durante los dos cursos fueron los siguientes: la probabilidad, enfoque clásico y frecuencial, la distribución binomial, la distribución normal, distribuciones muestrales, el Teorema Central del Límite, estimación puntual y por intervalos, pruebas de hipótesis. Durante los cursos se

realizaron varias evaluaciones tanto escritas como orales que se realizaron siguiendo la metodología de las entrevistas clínicas. En particular, al final del curso se realizaron entrevistas personales sobre los intervalos de confianza una vez los estudiantes habían respondido los cuestionarios de Behar (2001) y Olivo (2008).

Por razones de espacio, solamente se presenta el texto de la entrevista con Otoniel que creemos ayuda a interpretar las razones que pueden respaldar algunas de las concepciones erradas de los estudiantes sobre los IC.

Después de que Otoniel respondiera acertadamente las preguntas relacionadas con la construcción de intervalos de confianza, la conversación se encaminó a indagar las razones que tuvo para dar algunas respuestas en los cuestionarios de Behar (2001) y Olivo (2008).

En concreto la parte que aquí presentamos se relaciona con el ítem 1 del cuestionario de Olivo (2008), y más exactamente con la opción a) donde se afirma que "El intervalo de confianza del 50% para la media de una población μ es el rango dentro del cual caen el 50% de los valores de la media de la muestra $\bar{x}$".

OTONIEL: *No es cierto. Entonces, digamos, (*dibuja una curva normal con centro en μ) *que tal que el valor que haya sacado yo, este por acá* (dibuja $\overline{X}$ a la derecha de μ)*, el intervalo no sé, me quede aquí. Aquí está el 50% de posibles valores y por acá hay otro poco de posibles valores que voy a obtener cuando haga otra media muestral. Digamos que esta es la uno* (señala la $\overline{X}$ inicial) *y* (luego hace otra $\overline{X}$ alejada) *ésta es una que no está metida, entonces ya hay una que no está allá metida.*

PROFESOR: *La afirmación es que en este de aquí está el 50% de las* $\overline{X}$.

OTONIEL: *Entonces ya no están, porque…*

PROFESOR: *Usted dejó sólo una por fuera*.

OTONIEL: *¡¡¡No!!! Todo esto aquí son posibles valores* (señala el lado izquierdo del intervalo que había dibujado) *que va a tomar. Si la media de la muestra que tomé me hubiera dado exactamente la media de la población, entonces ¡¡¡ahí sí!!!*
PROFESOR: *Ahí si, ¿qué?*
OTONIEL: *Ahí sí sería lo que está diciendo la afirmación.*
PROFESOR: *¿Por qué?*
OTONIEL: *Por el teorema central del límite, esto tiene distribución normal* (dibuja la curva normal y $\overline{X}$ n debajo de la μ). *Entonces ahí sí sería cierto si son iguales* (dibuja un intervalo del 50%), *pero si estuviera aquí* (a la derecha, lejano de μ) *entonces el intervalo que yo haga dependiendo de eso, ya van a haber muchos valores que no van a estar metidos ahí.*

Este corto diálogo deja entrever varios entendimientos acerca de las distribuciones muestrales relacionadas con los intervalos de confianza que bien vale la pena resaltar:

- En primer lugar Otoniel no discute que el intervalo de confianza contenga medias muestrales, lo que él alega y es su razón para descartar esta opción es que no contiene exactamente el 50% sino que pueden ser menos y que eso va a depender de la posición de $\overline{X}_n$ respecto a μ.
- Parece ser que la argumentación de Otoniel reposa en la expresión (1) que define la distribución de la media muestral $\overline{X}_n$ como normal centrada en el parámetro μ y en el intervalo de confianza (5) que está centrado en el valor $\overline{X}_n$. Son los intervalos centrados en μ los que contienen el porcentaje de medias muestrales señalado, como bien lo dice Otoniel: "*Por el teorema*

central del límite, esto tiene distribución normal). *Entonces ahí sí sería cierto si son iguales …*"; cuando la media $\overline{X}_n$ no coincide con μ no se puede garantizar el porcentaje dado.

- El llamado de atención sobre la posible confusión que se podría presentar entre los dos intervalos definidos en las expresiones (5) y (6) parece ser que se concreta en el análisis de Otoniel cuando interpreta el nivel de confianza (50% para el caso) asociado al intervalo $\mu \pm \rho$ centrado en μ y lo extrapola para el intervalo $\overline{X}_n \pm \rho$, centrado en $\overline{X}_n$. El problema radica en que Otoniel asume que si bien en el primer intervalo se interpreta que el nivel de confianza o probabilidad hace referencia a los valores de $\overline{X}_n$ que se encuentran en él, para el segundo intervalo esta interpretación no tiene mayor sentido ya que en este caso el interés se centra sobre la pertenencia de μ a los intervalos centrados en $\overline{X}_n$.
- El hecho de que los dos intervalos contengan valores $\overline{X}_n$ puede hacer creer a los estudiantes que el nivel de confianza se refiere al porcentaje de valores $\overline{X}_n$ que caen en el intervalo de confianza (5).
- Asumir el nivel de confianza como un porcentaje de valores $\overline{X}_n$ que pertenecen al intervalo puede estar reflejando una concepción respecto a la distribución de probabilidad de una variable aleatoria, en la que el valor de probabilidad asociada a un intervalo se interpreta como el porcentaje de los valores de la variable que caen en ese intervalo. Este aspecto amerita mayor investigación.

Conclusiones

Tanto el análisis epistemológico realizado alrededor del procedimiento de construcción de intervalos de confianza como los apartes de la entrevista con Otoniel destacan las complicaciones que poseen en sí mismos los intervalos de confianza.

Si bien los estudios previamente realizados y citados en la introducción a este trabajo destacan variados tipos de malas concepciones, este trabajo se interesó básicamente en la interpretación del nivel de confianza asociado a los IC y en tratar de explicar las razones que podrían respaldar las malas concepciones que sobre este aspecto existen.

Una primera explicación se basa en los diferentes significados que asume el concepto de probabilidad en los dos intervalos generados (4) y (5). El intervalo dado en (4) se traduce como el intervalo que con probabilidad 1-α contiene valores de la variable aleatoria $\bar{X}_n$, en tanto que el definido en (5), que es precisamente el intervalo de confianza, se traduce como el intervalo que con probabilidad 1-α contiene valores del parámetro buscado, μ. El problema surge porque si μ es una constante ¿cómo se puede explicar que se hable de una probabilidad asociada a él? La verdad debe hablarse de la probabilidad de que el intervalo (que sí tiene elementos aleatorios ya que depende del valor de $\bar{X}_n$) contenga a μ. La mala interpretación puede provenir de representaciones como la dada en (5). Tal vez sea mucho más conveniente escribirla de la siguiente forma:

$$P\left\{\left(\underbrace{\overline{X}_n - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}_{L_1}, \underbrace{\overline{X}_n + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}}_{L_2}\right) \text{ contenga a } \mu\right\} = 1-\alpha \quad (7)$$

Adicional a esta explicación lógico-semiótica como origen de algunas de las malas concepciones que presentan los estudiantes y profesores de estadística respecto a los intervalos de confianza, habría que añadir las incomprensiones que puede generar un proceso basado en el comportamiento de muchas muestras cuando solo se cuenta con una sola. Esta situación puede ser la razón para asumir el valor de probabilidad como una medida de certeza lejana a cualquier interpretación frecuencial. En este sentido, es recomendable adoptar el enfoque frecuencial de la probabilidad desde un comienzo y no limitarse a utilizarlo solamente para interpretar el nivel de confianza, como se acostumbra en muchos libros de texto.

De otro lado, la entrevista con Otoniel resalta la necesidad que existe de discutir ampliamente con los estudiantes todos los pasos realizados en el proceso de construcción de los intervalos de confianza para evitar que sean los estudiantes mismos los que tengan que construir sus propias justificaciones y cayendo, muy seguramente, en malas interpretaciones.

Referencias bibliográficas

Behar, R. (2001). Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística. Tesis doctoral no publicada. Universidad Politécnica de Catalunya. Barcelona, España.

Bower, D.M. (2003). Some Misconceptions about Confidence Intervals. *Six Sigma Forum-American Society for Quality,* July.

Cumming, G., Williams, J.,& Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. Understanding Statistics, 3, 299 – 311.

Fidler, F. (2005). From Statistical Significance to Effect Estimation. Tesis doctoral no publicada. Universidad La Trobe, Melbourne, Australia.

Fidler, F. & Cumming, G. (2005). Teaching Confidence Intervals: Problems and Potential Solutions. *Proceedings of the 55th International Statistics Institute*, Sydney, Australia.

Godino, J.D. & Batanero, C. (1998). Clarifying the meaning of mathematical objects as a priority area of research in Mathematics Education. En Sierpinska, A. & Kilpatrick (Eds.), Mathematics education as a research domain: A search for identity. P. 177-195. Dordrecht: Kluwer.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6(1), 59-98. Fidler, F. (2005). From Statistical Significance to Effect Estimation. Tesis doctoral no publicada. Universidad La Trobe, Melbourne, Australia.

Olivo, E., (2008). Significado de los intervalos de confianza para los estudiantes de ingeniería en México. Tesis doctoral no publicada. Universidad de Granada, Granada, España.