

Comprensión del intervalo de confianza por estudiantes de Bachillerato

Antonio Francisco Roldán López de Hierro, Universidad de Granada (España)

Carmen Batanero, Universidad de Granada (España)

Rocío Álvarez-Arroyo, Universidad de Granada (España)

Comprensión del intervalo de confianza por estudiantes de Bachillerato

Resumen

El intervalo de confianza es un procedimiento básico en inferencia estadística y su estudio se incluye en las Matemáticas Aplicadas a las Ciencias Sociales II para Bachillerato. Además, en las pruebas de acceso a la universidad se propone con frecuencia un problema de este contenido. Con el objetivo de evaluar la comprensión del tema, en este trabajo se analizan las respuestas de 58 estudiantes de segundo curso del citado Bachillerato a un cuestionario que consta de seis ítems de opción múltiple y un problema abierto tomado de anteriores pruebas de acceso la universidad. Los resultados muestran una pobre comprensión del tema, con pocas respuestas correctas en los ítems de opción múltiple, que evalúan la comprensión conceptual, y con sólo un 40% de estudiantes que logran una resolución completa y correcta del problema.

Palabras clave. Intervalo de confianza; comprensión; estudiantes de Bachillerato.

High school students' understanding of confidence interval

Abstract

Confidence interval is a basic procedure in statistical inference and its study is included in Applied Mathematics II directed to Social Sciences high school. In addition, a problem related to this content is often posed in the university entrance tests. With the aim of assessing the understanding of the topic, we analyse the answers given by 58 second-year high school students of the aforementioned specialty to a questionnaire consisting of six multiple-choice items and an open problem taken from previous university entrance tests. The results show a poor understanding of the subject, with few correct answers in the multiple-choice items, which assess conceptual understanding, and with only 40% of students achieve a correct solution of the problem.

Keywords. Confidence interval; understanding; high school students.

1. Introducción

Un procedimiento de gran utilidad en la inferencia es el intervalo de confianza, cuyo uso se promueve actualmente debido a las críticas realizadas al contraste de hipótesis (Wilkinson y TFSA, 1999; Yaremko, Harari, Harrison y Lynn, 2013). A pesar del amplio uso, la investigación didáctica informa de dificultades asociadas a su comprensión y uso incorrecto en Batanero, Ruiz y Gea (2019) o en Castro Sotos, Vanhoof, Noortgatem y Onghena (2007). Esta importancia se ha reconocido en el currículo español para el Bachillerato de Ciencias Sociales (MECD, 2015), que propone los contenidos: estimación por intervalos de confianza; relación entre confianza, error y tamaño muestral; intervalo de confianza para la media de una distribución normal con desviación típica conocida, y para la media y proporción en grandes muestras. Además, las pruebas de acceso a la universidad incluyen a menudo un problema de intervalo de confianza (López-Martín, Batanero, Díaz-Batanero y Gea, 2016). Este contenido se

enseña en los cursos universitarios de estadística, con mayor amplitud. Cabe asegurar su comprensión por los estudiantes que realizan las pruebas de acceso, ya que el realizar los estudios elegidos depende de la calificación que consigan en ella. Sin embargo, no se tiene información al respecto, pues la investigación sobre la comprensión de intervalos de confianza se ha centrado en estudiantes universitarios o profesionales.

El objetivo de este estudio exploratorio es evaluar la comprensión de estudiantes de Bachillerato sobre los intervalos de confianza, con apoyo en el enfoque ontosemiótico y en las características del intervalo de confianza en estadística frecuencial.

2. Marco teórico

Nos basamos en el enfoque ontosemiótico del conocimiento y la instrucción matemática, EOS (Godino, Batanero y Font, 2019), que diferencia entre significado institucional y personal de los objetos matemáticos, donde el significado de cada noción matemática se basa en las prácticas realizadas por un individuo o institución en la resolución de problemas. En este marco se describe la *comprensión* como la coincidencia entre los significados institucional y personal del objeto matemático. Dicha comprensión crece con el aprendizaje y la experiencia y puede referirse a los diferentes elementos del significado del objeto (problemas, lenguaje, definiciones y propiedades, procedimientos y argumentos). En nuestro trabajo evaluamos la comprensión de la definición y propiedades básicas del intervalo de confianza, y del procedimiento seguido al resolver un problema. Nos basamos en la aproximación frecuencial a la inferencia estadística, que es la enseñada actualmente en Bachillerato.

2.1. Significado del intervalo de confianza en la metodología frecuencial de la inferencia estadística

Actualmente coexisten diferentes soluciones a la estimación por intervalos que se diferencian en la concepción de probabilidad que aplican, los conceptos, propiedades y lenguaje que utilizan, y el procedimiento de cálculo del intervalo (Olivo, 2008; Rivadulla, 1991). Desde el punto de vista del EOS, se entienden como significados institucionales diferenciados (usados por diferentes comunidades de estadísticos). Como los contrastes de hipótesis, la estimación por intervalos surge del problema de justificación del método inductivo en las ciencias empíricas, que basan sus teorías en la generalización de observaciones de datos muestrales de las poblaciones de interés (de la Fuente y Díaz, 2004). Al no poder aplicar el método deductivo, se recurre a la inferencia estadística para validar sus conclusiones, dando estimaciones que consideren la medida de la variabilidad del muestreo (Morey, Hoekstra, Rouder, Lee y Wagenmakers, 2016).

En nuestro trabajo utilizamos el significado frecuencial, debido a Neyman (1937) y apoyado en el significado frecuencial de la probabilidad¹, que supone constante y desconocido el parámetro poblacional θ . Se usa un estimador $\hat{\theta}$, que es una variable aleatoria, cuya distribución en todas las posibles muestras de la población es la distribución muestral. Neyman (1937) propuso calcular la desviación típica $S_{\hat{\theta}}$ en la distribución muestral y expresar la precisión de la estimación mediante un intervalo entre cuyos extremos (EI y ES) se presume se encuentra el parámetro:

$$EI = \hat{\theta} - k_1 S_{\hat{\theta}} \text{ y } ES = \hat{\theta} + k_2 S_{\hat{\theta}}$$

¹ La probabilidad de un suceso se estima a partir de su frecuencia relativa en un número suficientemente grande de repeticiones del experimento.

Neyman sugirió calcular estos extremos tomando una probabilidad de error menor o igual que $1 - \alpha$ (usualmente 0,95 o 0,99), llamada *coeficiente de confianza* (Olivo, 2008). Dicho valor determina k_1 y k_2 , usando los valores centrales de la distribución muestral que incluyen una probabilidad $1 - \alpha$. A partir de una muestra aleatoria de valores independientes de la población, se obtiene el estadístico muestral θ y los extremos del intervalo. Mayo (1981) recuerda que los extremos del intervalo son variables aleatorias, y $1 - \alpha$ no es la probabilidad de que el verdadero valor del parámetro θ esté situado dentro de sus límites.

Este error de interpretación es denominado por Morey et al. (2016) *falacia fundamental de la confianza*, es incorrecto en el significado frecuencial del intervalo, pero aceptable en el bayesiano, basado en el significado subjetivo de probabilidad², donde se habla de *intervalos de credibilidad*. En el significado bayesiano el parámetro es una variable aleatoria, cuya distribución de probabilidad asigna el investigador a priori. El intervalo de credibilidad se calcula combinando la distribución a priori con los datos de la muestra por medio del teorema de Bayes. Por tanto, el significado institucional de la estimación por intervalo depende de la metodología de inferencia que se utilice. En nuestro trabajo la comprensión del estudiante del intervalo de confianza es su grado de acuerdo con los elementos de la metodología frecuencial.

3. Antecedentes

Los estudios previos se han realizado con estudiantes universitarios y profesionales, pero informan sobre la comprensión que cabe esperar en Bachillerato. El primero fue desarrollado por Cumming, William y Fidler (2004), con 134 investigadores que utilizaban intervalos en sus trabajos, a los que dio un intervalo de confianza del 95% para la media poblacional, pidiéndoles valores factibles para la media muestral en nuevos experimentos. El 78% respondió que la media caería en el intervalo dado en el 95% de las veces (probabilidad de replicación). La relevancia de este estudio es su relación con la creencia errónea de que los extremos de los intervalos son constantes.

Behar (2001) analizó la comprensión del tema de 47 estadísticos y profesores y 297 estudiantes de ingeniería. Un 29% de los expertos y un 59% de los alumnos interpretaron el nivel de confianza como porcentaje de datos poblacionales que caen en el intervalo. El 40% de los expertos y más de la mitad de estudiantes dieron una interpretación bayesiana. El alumnado indicó que al crecer el nivel de confianza se obtienen intervalos más estrechos, e incluso establecen una relación de proporcionalidad directa entre anchura del intervalo y tamaño de la muestra.

Fidler y Cumming (2005) trabajaron con 180 estudiantes de psicología que, en general, apreciaban los intervalos de confianza desde una perspectiva descriptiva: un 38% lo imaginaban como un conjunto de valores razonables para la media muestral, y un 19% confundía rango e intervalo de confianza. Un 20% manifestó que la anchura del intervalo aumentaría con el tamaño muestral, un 29% consideró que la anchura no variaría y un 36% no fue capaz de describir la relación.

En sus trabajos (Olivo, 2008; Olivo, Batanero y Díaz, 2008), Olivo estudió la comprensión de 48 alumnos de ingeniería del intervalo de confianza, observando las siguientes dificultades: a) La relación entre el tamaño de la muestra y la precisión del intervalo sólo fue comprendida por el 25% de los alumnos; b) El 18% afirmó que una

² La probabilidad es un grado de creencia subjetivo que asigna quien juzga la probabilidad de un suceso.

variación en el nivel de confianza no afecta a la anchura; c) El 21% relacionó el intervalo con posibles valores de la variable estudiada; d) El 27% de los alumnos dio una interpretación bayesiana al intervalo. Además, al calcular un intervalo de confianza, el 25% eligió mal el número de grados de libertad de la distribución t de Student y un 17% utilizó la tabla de la distribución normal cuando no era apropiada.

López-Martín, Batanero y Gea (2019a) evaluaron pidiendo a 71 estudiantes del máster de formación del profesorado de matemáticas, interpretar un intervalo de confianza, obteniendo sólo un 28% de interpretaciones adecuadas. El 11% afirmó que el intervalo debía contener, con seguridad, el valor del parámetro (visión determinista). Un 17,8% interpretó los extremos como fijos y el nivel de confianza como probabilidad de que el intervalo contuviese al valor del parámetro. Incluso un participante manifestó que el intervalo hacía referencia a la media muestral en vez de a la media poblacional. Además, un 41,1% no aportó ninguna interpretación.

4. Metodología

4.1. Participantes y contexto

Participaron 58 estudiantes de *Matemáticas Aplicadas a las Ciencias Sociales II*, de segundo curso del Bachillerato en dos institutos. El *instituto A* está en la ciudad de Granada y participaron 24 alumnos de dos grupos diferentes, codificándolos como A1, A2..., A24. El *instituto B* está situado en la periferia y participaron 34 estudiantes de dos grupos, codificándolos desde B1 hasta B34. Es una muestra intencional formada por todos los alumnos del segundo curso de Bachillerato de Ciencias Sociales de los dos institutos. Por tanto, nuestro estudio es exploratorio y no pretendemos extender las conclusiones a una población más amplia. Los cuestionarios se completaron en mayo de 2019 durante una hora de clase. Los participantes habían finalizado su formación sobre inferencia estadística e intervalos de confianza y se preparaban para los exámenes finales y la Prueba de Evaluación para el Acceso a la Universidad en Andalucía.

4.2. Descripción del cuestionario

Se construyó el cuestionario partiendo de otros, descritos en Behar (2001), Cruise, Dudley y Thayer (1984) y Olivo (2008); consta de ítems de opción múltiple y un problema abierto. A continuación, se analiza el contenido de los seis primeros ítems, resaltando en negrita las respuestas correctas. Los ítems de Cruise et al. se validaron por Behar (2001) con 297 estudiantes y por Olivo (2008) con 252 estudiantes, mientras que los ítems de elaboración propia están pendientes de validación. Al inicio del cuestionario se indicó a los estudiantes que en cada ítem podía haber más de una respuesta correcta. Cada opción evalúa la comprensión de una o más definiciones o propiedades (Tabla 1).

Ítem 1. Se calcula un intervalo de confianza del 90% para la media μ de una población. ¿Cuál o cuáles de las siguientes afirmaciones son verdaderas?

- A. Si se toman muchas muestras, la media muestral \bar{x} caerá dentro del intervalo de confianza aproximadamente el 90% de las veces.
- B. El intervalo de confianza es un intervalo de valores calculado a partir de los datos de la muestra. En el 90% de las muestras de una población, el intervalo calculado contiene a la media.**
- C. El intervalo de confianza es un intervalo con extremos constantes, dentro del cual cae la media poblacional el 90% de las veces.
- D. La probabilidad de que μ caiga dentro de un intervalo de confianza calculado a partir de una muestra es de 0,90.

Este ítem está adaptado de Cruise et al. (1984) y es similar a otro utilizado por Olivo (2008). La solución correcta es la B, donde se indica el carácter aleatorio de los extremos del intervalo y el porcentaje de intervalos calculados a partir de diferentes muestras que cubre el valor del parámetro. El distractor A evalúa la confusión, descrita por Behar (2001), entre media muestral y poblacional, siendo incorrecta porque la media muestral \bar{x} siempre cae dentro del intervalo. El distractor C considera la confusión citada por López-Martín, Batanero y Gea (2019b) de suponer constantes los extremos del intervalo. El apartado D desvela la confusión entre confianza y probabilidad (interpretación bayesiana al intervalo; Behar, 2001; Olivo et al., 2008).

Ítem 2. Hemos calculado un intervalo de confianza al 90% utilizando el valor medio \bar{x} obtenido a partir de una muestra de 10 casos. Si incrementamos el tamaño de la muestra a 1000, y calculamos un segundo intervalo al 90% de confianza:

- A. Debemos tener más confianza de que μ caerá en nuestro segundo intervalo.
- B. Sabemos que el segundo intervalo será 10 veces más estrecho.**
- C. Espero que ambos intervalos de confianza tengan la misma precisión.
- D. El segundo intervalo de confianza es 10 veces más ancho que el primero.

Tomado de Cruise et al. (1984), evalúa la comprensión de la relación entre el ancho del intervalo de confianza y el tamaño de la muestra. La solución correcta es la B, pues el intervalo se obtiene sumando y restando a la media muestral un valor que depende inversamente de la raíz cuadrada del tamaño muestral. El distractor A, que evalúa la confusión entre confianza y amplitud, es incorrecto porque en la segunda muestra el nivel de confianza es el mismo. Igualmente es incorrecto el apartado C, pues la precisión del intervalo viene influenciada por el tamaño de muestra. El apartado D muestra el error de creer en la relación opuesta entre tamaño de muestra y amplitud del intervalo.

Ítem 3. Si, manteniendo todos los demás datos fijos, el nivel de confianza se reduce (por ejemplo, del 90% al 80%):

- A. El intervalo de confianza será menos preciso.
- B. El intervalo de confianza será más ancho.
- C. El intervalo de confianza será más estrecho.**
- D. El intervalo de confianza será más preciso.**

Tomado de Cruise et al. (1984), evalúa la comprensión de la relación entre el nivel de confianza, ancho del intervalo y precisión. Cuanto mayor es la confianza mayor es la anchura del intervalo de confianza y menor será la precisión. Por ello, C y D son respuestas correctas, mientras que los distractores A y B muestran una relación entre confianza y precisión opuesta a la real.

Ítem 4. Se calculan intervalos de confianza del 95% con muestras de 100 elementos. ¿Cuál o cuáles de las siguientes afirmaciones son verdaderas?

- A. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza no cambia.
- B. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza disminuye.**
- C. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza disminuye.
- D. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza aumenta.**

Adaptado de Behar (2001), esta cuestión analiza la relación entre la anchura del intervalo y varianza de la población. Hay dos opciones correctas: B y D. La anchura del

intervalo de confianza es $2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$. Por tanto, si la desviación estándar disminuye, la anchura del intervalo de confianza también disminuye. Los distractores A y C evalúan la falta de comprensión de esta relación.

- Ítem 5.** Se ha tomado una muestra de tamaño $n = 100$ para calcular el intervalo de confianza, al 95%, para la velocidad media de los coches que pasan por una carretera. Se sabe que la desviación típica poblacional de estas velocidades es de 20 km/h.
- A. El ancho del intervalo de confianza depende de la media muestral \bar{x} .
 - B. La media muestral \bar{x} es un extremo del intervalo de confianza.
 - C. Cabe la posibilidad de que el verdadero valor de μ no esté contenido en el intervalo.**
 - D. El ancho del intervalo no depende de la media muestral.**

Este ítem es de elaboración propia y evalúa la comprensión de que el ancho del intervalo no varía cuando se traslada la media ya que el intervalo está centrado en ella. La respuesta C es correcta porque, al ser aleatorios los extremos del intervalo, algunos intervalos no cubren el parámetro. La respuesta D también es correcta puesto que la expresión de la anchura del intervalo de confianza $\left(2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$ no depende de la media muestral. El distractor B es falso ya que, para obtener los extremos, siempre se suma y se resta una cantidad positiva a la media muestral y esta no puede ser uno de los extremos. El distractor A es la negación del apartado D y es incorrecto.

- Ítem 6.** Se sabe que la altura media μ de los pinos de una zona forestal muy amplia es de 6.5 metros. ¿Cuál de las siguientes afirmaciones es verdadera?
- A. Un posible intervalo de confianza para μ es [6.25, 8.4].
 - B. Un posible intervalo de confianza para μ es [6.75, 8.4].
 - C. Cualquier intervalo de confianza para μ que calculemos, asociado a una muestra aleatoria cualquiera, debe contener al valor 6.5.
 - D. En este contexto, no tiene sentido determinar el intervalo de confianza ya que se conoce la altura media de todos los pinos.**

Esta cuestión también es de construcción propia. Con ella pretendemos determinar si el alumnado comprende que el intervalo de confianza sirve para estimar un valor desconocido de un parámetro. Por tanto, si éste es conocido, no tiene sentido aplicarlo, lo que hace que la respuesta correcta sea la D. Por esa misma razón, los distractores A y B son incorrectos y la opción C es falsa porque, en ocasiones, el intervalo de confianza no contiene al verdadero valor del parámetro.

La segunda parte del cuestionario está constituida por el siguiente problema abierto tomado de la prueba de acceso a la universidad en Andalucía en el año 2018.

- Problema.** La media muestral de 100 observaciones en una prueba de matemáticas es de 75 puntos. Asumiendo que las calificaciones siguen una distribución normal y que $\sigma = 7$, encuentre el intervalo de confianza al 95% para la media de la población.

Para resolverlo se ha de deducir la fórmula de cálculo del intervalo de confianza para la media (con varianza conocida), siguiendo los pasos siguientes:

$$IC(\mu) = \left[\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] = \left[75 \pm 1.96 \frac{7}{\sqrt{100}} \right] = [73.628, 76.372]$$

Se deben deducir los datos del enunciado del problema y saber leer la tabla de la distribución normal $N(0,1)$ para deducir de ella el valor crítico (Z) correspondiente.

La Tabla 1 resume el contenido evaluado por el cuestionario, que como se observa evalúa definiciones, propiedades y el conocimiento sobre el método de construcción. En todos los ítems es necesario comprender la definición del intervalo de confianza.

Tabla 1. *Contenido evaluado por el cuestionario*

Contenido	I1	I2	I3	I4	I5	I6	Problema
Definición del intervalo de confianza	x	x	x	x	x	x	X
Porcentaje de intervalos que cubre el valor del parámetro	x						
Extremos aleatorios	x						
Confusión entre media muestral y poblacional	x						
Considerar fijos los extremos	x				x		
Confusión entre confianza y probabilidad (interpretación bayesiana)	x						
Relación del ancho del intervalo y tamaño muestral		x	x				
Confundir confianza y amplitud		x	x				
Relación entre precisión y confianza		x	x				
Fórmula de desviación típica de distribución muestral		x					
Relación entre amplitud del intervalo y varianza poblacional				x		x	
Relación del intervalo con la media muestral					x		
Algunos intervalos no cubren el parámetro					x		
Ancho del intervalo no depende de media muestral					x		
El intervalo sirve para estimar valores desconocidos del parámetro						x	
Pensar que el intervalo siempre contiene al parámetro						x	
Construcción de un intervalo de confianza							x
Fórmula del intervalo de confianza							x
Desviación típica de la distribución muestral							x
Leer valores críticos en la tabla de la $N(0,1)$							x

5. Resultados

5.1. Resultados en los ítems de opción múltiple

La Tabla 2 presenta el porcentaje de alumnos que elige cada opción en los ítems; como algunos marcan más de una opción, el porcentaje total de respuestas al mismo ítem puede sumar más del 100%. Se ha marcado en negrita las opciones correctas, que, salvo en el primer ítem, son elegidas por menos del 50% de la muestra.

Tabla 2. *Porcentaje de estudiantes que elige cada opción en cada ítem*

Ítem	Opción u opciones elegidas				
	Opción A	Opción B	Opción C	Opción D	En blanco
Ítem 1	1,7	55,2	3,4	37,9	3,4
Ítem 2	32,8	25,9	17,2	20,7	3,4
Ítem 3	62,1	12,1	22,4	12,1	1,7
Ítem 4	5,2	24,1	46,6	29,3	3,2
Ítem 5	50,0	27,6	36,2	29,3	1,7
Ítem 6	17,2	8,6	39,7	36,2	1,7

Los principales errores de comprensión deducidos de la elección de los distractores por parte de los estudiantes suponen que estos han adquirido un significado personal del concepto o propiedad analizada que no coincide con el significado institucional en estadística frecuencial. Primero describimos errores detectados en estudios previos:

- Considerar el intervalo de confianza con extremos constantes (3,4% en la opción C del ítem 1), citado por Behar (2001) y Olivo (2008). Implica la creencia en la probabilidad de replicación, descrita por Cumming et al. (2004).
- Confundir confianza con probabilidad de que el parámetro caiga en el intervalo (37,9% en opción D de ítem 1; 39,7% en distractor C de ítem 6), dando interpretación bayesiana al intervalo, error citado por Behar (2001), Morey et al. (2016) y Olivo (2008). Se interpreta el coeficiente de confianza como una probabilidad a posteriori, sólo calculable en el método bayesiano, que se basa en el significado subjetivo de la probabilidad y no en el frecuencial, propio de los intervalos de confianza.
- Considerar que el intervalo de confianza debe contener la media poblacional (39,7% en opción C de ítem 6), citado por López-Martín et al. (2019a). Subyace la interpretación determinista y la creencia en la replicación de los resultados (Cumming et al., 2004).
- Suponer que reducción del nivel de confianza implica reducción en la precisión (62,1% en opción A de ítem 3), confundiendo confianza y precisión, conflicto descrito en Yáñez y Behar (2009). Otra idea errónea es ligar la precisión solo a la confianza (17,2% en distractor C de ítem 2). Estas propiedades son esenciales para asegurar la precisión de las estimaciones.
- Pensar que el ancho del intervalo aumenta cuando crece el tamaño de la muestra (20,7% en ítem 2), error detectado por Fidler y Cumming (2005). Es necesario clarificar este punto para apreciar la importancia de un tamaño de muestra para asegurar, a la vez, precisión y confianza.
- Considerar que el intervalo de confianza será más ancho al reducir el nivel de confianza (12,1% en opción B de ítem 3), conflicto descrito en Behar (2001), Morey et al. (2016), Olivo (2008) y Yáñez y Behar (2009).
- Confundir estadístico y parámetro (17,2% en distractor A de ítem 6), señalado por Behar (2001), Harradine et al. (2011), Olivo (2008) y Olivo et al. (2008). Es crucial diferenciar ambos conceptos en el intervalo de confianza y contraste de hipótesis.

Además, hemos encontrado los siguientes errores, que constituyen una aportación original de nuestro trabajo:

- Situar la media muestral en un extremo del intervalo de confianza (27,6% en opción B de ítem 5) o suponer que algunos intervalos no contienen a la media muestral (1,7% en distractor A de ítem 1). Considerar que la anchura del intervalo de confianza depende de la media muestral (50%, opción A de ítem 5). Estos tres errores indican falta de comprensión del papel de la media muestral en la construcción del intervalo.
- No relacionar “intervalo más estrecho” con “intervalo más preciso” (en ítem 3, ningún alumno ha elegido ambas opciones correctas y equivalentes). Esto implica una falta de conocimiento de la idea de precisión.
- Para un mismo nivel de confianza, suponer que el tamaño de muestra aumenta las posibilidades de que el parámetro caiga dentro del intervalo (32,8% en opción A de ítem 2).
- Suponer que la anchura del intervalo disminuye cuando aumenta la desviación estándar de la población (46,6% en opción C de ítem 4). Implica una confusión en el papel que juegan la media y la desviación típica en la construcción del intervalo.
- Ausencia de comprensión de las situaciones en que tiene sentido calcular un intervalo de confianza (en ítem 6, el 63,8% no selecciona la opción adecuada).

Para valorar globalmente la comprensión conceptual, presentamos en la Figura 1 la distribución del número de respuestas correctas de cada estudiante en los ítems, tres de los cuales poseían dos respuestas correctas, por lo que podían seleccionarse un total de nueve respuestas correctas y la media teórica sería de 4,5. El número medio de respuestas fue de 2,7 y el 50% central de la distribución se situó entre 2 y 4 respuestas correctas. El máximo fue solo 7, siendo lo más frecuente tan sólo dos aciertos. Muy pocos estudiantes lograron 5 o más respuestas correctas. Todo ello indica una gran dificultad conceptual del tema para estos estudiantes de bachillerato.

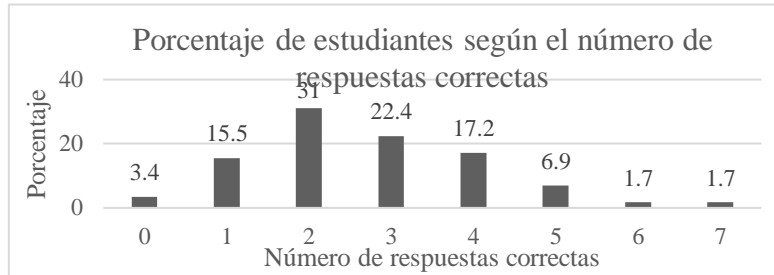


Figura 1. Distribución del número de respuestas correctas

5.2. Resultados en el problema abierto

En la Tabla 3 se clasifica el proceso de resolución según su grado de corrección, entendiendo por resolución correcta la que lleva a una expresión final correcta del intervalo, y parcialmente correcta si llega a plantear la expresión del intervalo de confianza, sustituyendo correctamente los datos, sin llegar a la expresión final. El 39,7% de los estudiantes llega al menos a la solución parcialmente correcta.

Tabla 3. Distribución de estudiantes, según resolución del problema

Nivel de resolución	Frecuencia	Porcentaje
Resolución correcta	17	29,3
Resolución parcialmente correcta	6	10,4
Resolución incorrecta	17	29,3
En blanco	18	31,0

La Tabla 4 muestra un análisis semiótico de la solución, descomponiendo la solución en prácticas matemáticas elementales e identificando los objetos y procesos matemáticos involucrados en cada una. Dicho análisis ilustra la gran cantidad de objetos matemáticos que se deben comprender y aplicar y, por tanto, la complejidad semiótica de la solución.

Tabla 4. Análisis semiótico de la solución al problema

Práctica matemática	Objetos y procesos involucrados
P1. Fórmula de cálculo del intervalo $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	<ul style="list-style-type: none"> Comprender los conceptos de intervalo y extremos. Diferenciar media muestral \bar{x} y media poblacional μ y comprender que el intervalo está centrado en la media muestral. La amplitud del intervalo viene dada por el valor crítico multiplicado por la varianza de la distribución muestral. La varianza de la distribución muestral es $\frac{\sigma}{\sqrt{n}}$. Obtener el valor crítico y conocer la distribución normal estándar.
P2. Determinar el percentil requerido para leer la tabla	<ul style="list-style-type: none"> Aplicar los conceptos de nivel de confianza y valor crítico. Determinar el valor $1 - \alpha/2$ a partir del nivel de confianza.
P3. Representación gráfica	<ul style="list-style-type: none"> Conocer la forma de la distribución normal y propiedades

	de simetría central y asíntota horizontal.
P4. Lectura de la tabla	• Leer la tabla, posiblemente con necesidad de interpolación.
P5. Sustitución de datos	• Identificar los datos necesarios en la fórmula de cálculo, recordando los conceptos requeridos.
P6. Cálculos	• Realizar los cálculos.
P7. Redondeo	• Aplicar los criterios de redondeo.

La Tabla 5 presenta los porcentajes de estudiantes que llevaron a cabo los diferentes pasos requeridos para la solución.

Tabla 5. Frecuencia de estudiantes que realizan los pasos en una correcta resolución de problema

Descripción del proceso seguido	Frecuencia	Porcentaje
Determina el percentil requerido	30	51,7
Determina el valor crítico, leyendo correctamente la tabla	26	44,8
Sustituye correctamente los datos en expresión del intervalo	23	39,7
Realiza correctamente los cálculos	17	29,3
Da la expresión final del intervalo de confianza:		
Sin redondeo	14	24,1
Redondeando	3	5,2

La Tabla 5 muestra que apenas la mitad del alumnado determina el valor del percentil requerido para leer la tabla. Entre los que determinan correctamente el valor $\alpha/2$, abundan quienes llegan a encontrar el valor crítico $z_{\alpha/2}$, si bien dos alumnos dan dicho valor sin haber explicitado previamente el valor de $\alpha/2$ ni haber hecho una gráfica o esquema adecuado de la distribución normal que le ayude en dicho proceso.

Solo tres de los estudiantes que llegan a determinar correctamente el valor crítico escriben incorrectamente el intervalo de confianza, uno porque no llega a determinar el tamaño muestral $n = 100$ y los otros por error en la fórmula. Por tanto, sólo 23 de los 58 alumnos escriben correctamente la expresión del intervalo de confianza empleando los datos adecuados, de los cuales seis cometen errores al realizar las operaciones. Finalmente, 17 estudiantes han escrito la expresión del intervalo de confianza y han realizado correctamente las operaciones, dando la solución correcta, habiendo 14 alumnos que la expresan con tres cifras decimales y 3 alumnos que emplean correctamente el redondeo. En consecuencia, únicamente el 29,3% del alumnado participante en el estudio ha llegado a la solución correcta del problema planteado.

Poco más de una quinta parte del alumnado lleva a cabo dos tareas muy importantes de cara a la comprensión del proceso: a) escriben la fórmula de cálculo del intervalo de confianza para la media; y b) se ayudan de una representación gráfica de la distribución normal para determinar el valor crítico. De los 13 alumnos que hacen una representación gráfica correcta, 10 llegan a la resolución correcta, lo cual indica que quienes utilizan una representación gráfica tienen una mejor visión global del proceso. Es más, los 3 estudiantes que se equivocan lo hacen a través de errores que parecen razonablemente subsanables: uno no llega a encontrar el tamaño muestral; otro emplea un valor erróneo para $z_{\alpha/2}$ por no saber leer la tabla de la distribución normal; y otro escribe la expresión correcta del intervalo de confianza, pero no lo calcula.

En segundo lugar, se analizaron los errores cometidos en la resolución del problema, para analizar si coinciden con los descritos en investigaciones anteriores. Es llamativo que algunos alumnos hayan cometido un error puramente algebraico, pues suman y restan a la media muestral el valor crítico antes de multiplicar por la desviación típica

de la distribución muestral, dando lugar al intervalo erróneo $(\bar{x} \pm z_{\alpha/2}) \cdot \frac{\sigma}{\sqrt{n}}$ (véase un ejemplo en la Figura 2). Al aplicar esta expresión algebraica se obtiene el intervalo [51,12; 53,87], incorrecto ya que no contiene el valor de la media muestral $\bar{x} = 75$. Esto revela falta de comprensión del significado y de propiedades del intervalo de confianza.

$$\text{Intervalo de confianza} = \left(75 - 1.645 \cdot \frac{7}{\sqrt{100}}, 75 + 1.645 \cdot \frac{7}{\sqrt{100}} \right) = (51.348, 53.651)$$

Figura 2. Respuesta del estudiante A21

Los estudiantes comenten errores al determinar el valor crítico. Varios ejemplos de esto se muestran en la Figura 3.

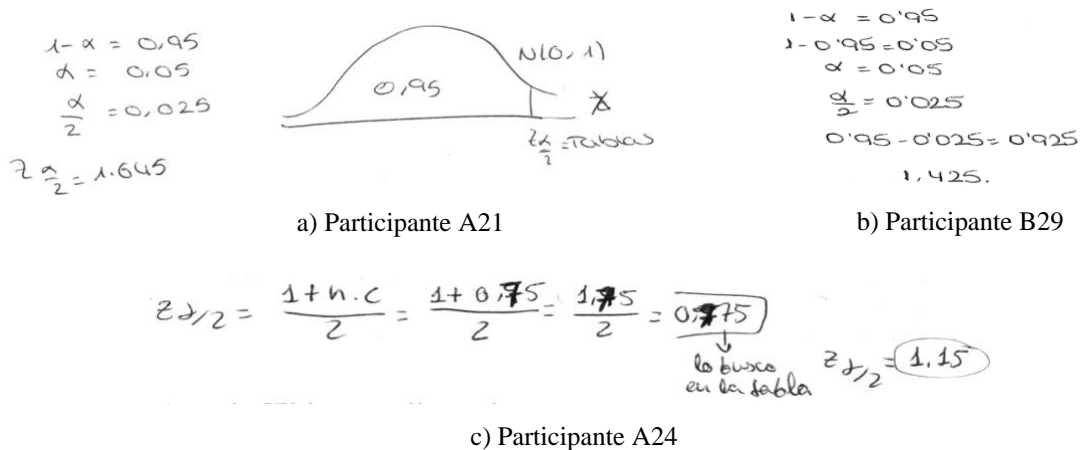


Figura 3. Respuestas erróneas de algunos estudiantes al determinar el valor crítico

A21 interpreta incorrectamente el nivel de confianza, confundiéndolo con el área correspondiente de la cola a la izquierda de la distribución normal (Figura 3^a). Es decir, no comprende que para calcular el intervalo de confianza se debe tomar el 95% central de los valores de la distribución normal. B29 (Figura 3b) resta los valores 0,95 (nivel de confianza) y 0,025 (cola derecha), por lo que determina un percentil erróneo (0,925) para buscar el valor crítico. A24 (Figura 3c) confunde el valor crítico $z_{\alpha/2}$ con el número concreto que ha de buscar en la tabla para determinarlo (que en la práctica es la media aritmética entre el nivel de confianza y la unidad). En vez de tomar 0,95 como nivel de confianza, emplea 0,75, lo que lleva a determinar un valor crítico incorrecto.

Se producen también errores al redondear los cálculos de los extremos del intervalo de confianza. Un 15% del alumnado trunca el número que obtiene en la calculadora sin utilizar ningún tipo de regla, lo cual parece un error propio de cursos anteriores. Estos errores aparecen en pocos estudiantes y se resumen en la Tabla 6. Al interpretar la tabla hay que considerar que algunos alumnos han cometido varios de esos errores a la vez.

Tabla 6. Errores más frecuentes

Descripción del proceso seguido	Frecuencia	Porcentaje
Suma y resta antes de multiplicar	4	6,9
Determina incorrectamente el valor crítico $z_{\alpha/2}$	7	12,1
Determina incorrectamente $\alpha/2$	1	1,7
Error en las operaciones	5	8,6
Redondea mal el extremo inferior del intervalo	5	8,6
Redondea mal el extremo superior del intervalo	4	6,9

El error más frecuente (12,1 %) es determinar incorrectamente el valor crítico $z_{\alpha/2}$ debido a lectura incorrecta de la tabla de la distribución normal, a pesar de que el 95% es el que más se usa en la práctica, que los estudiantes debieran recordar que corresponde a 1,96 al haberlo usado con mucha frecuencia en clase. Varios estudiantes que cometen este error no han confeccionado una representación gráfica, por lo que no pueden apoyarse en ella para hacer una lectura adecuada de la tabla.

6. Conclusiones

En este trabajo aportamos información nueva sobre la comprensión de estudiantes de Bachillerato de la definición del intervalo de confianza, sus propiedades principales y su cálculo. La principal conclusión es la dificultad que implica la comprensión del significado de este objeto en toda su complejidad para estos estudiantes. El significado personal mostrado por los estudiantes de la muestra contiene elementos que lo diferencian del significado institucional de los intervalos de confianza en la estadística frecuencial. Se incluyen interpretaciones deterministas, y se desconocen varias de las relaciones más importantes entre los elementos que intervienen en la construcción del intervalo. Una posibilidad es revisar el currículo de estadística para Bachillerato, y las pruebas de acceso, pues los estudiantes de Bachillerato cometen los mismos errores que los descritos en los trabajos con estudiantes universitarios y algunos nuevos más.

Debería cambiar la enseñanza teniendo en cuenta que el proyecto GAISE (Franklin et al., 2007) recomienda enfatizar el conocimiento conceptual sobre el procedimental y utilizar la tecnología como recurso de aprendizaje. El cálculo de intervalos de confianza es una cuestión resuelta por la tecnología, cuyo cálculo es un objetivo menor comparado con la necesidad de comprender fundamentos y propiedades del objeto matemático. Además, las simulaciones permiten aumentar la comprensión del significado y de propiedades del intervalo de confianza. Por ejemplo, el recurso disponible en <http://www.rossmanchance.com/applets/ConfSim.html> permite simular la extracción de n muestras para calcular los intervalos de confianza de la media o proporción cambiando parámetros de la población, tamaño de muestra, coeficiente de confianza y número de repeticiones. Estas simulaciones ayudarán también a superar la creencia de que los extremos del intervalo son fijos y la interpretación bayesiana del intervalo.

Otra recomendación del proyecto GAISE (Franklin et al., 2007) es utilizar la evaluación de los estudiantes para apoyar el aprendizaje. Esta ha sido la finalidad de nuestro trabajo, que puede interesar a los profesores que preparan a los estudiantes para las pruebas de acceso a la universidad, ya que no todos los profesores tienen consciencia de estos problemas (López-Martín et al., 2019). Dichas dificultades también se transmiten y pasan inadvertidas en los libros de texto (Gea, López-Martín y Roa, 2015).

Finalmente, reconocemos las limitaciones de este estudio exploratorio, que se piensa continuar con nuevas muestras de estudiantes y tareas complementarias.

Agradecimientos

Proyectos TIN2017-89517-P y PID2019-105601GB-I00 y Grupos FQM-268 y FQM-126.

Referencias

Begué, N., Batanero, C., Ruiz, K. y Gea, M. M. (2019). Understanding sampling: A summary of the research. *Beio*, 35(1), 49-78.

- Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística*. Trabajo de Tesis Doctoral. Universidad Politécnica de Cataluña.
- Castro Sotos, A. E., Vanhoof, S., Van den Nororgate, W. y Onghena, P. (2007). Student's misconceptions of statistical inference: A review of the empirical evidence from research on statistical education. *Educational Research Review*, 2(2), 98-113.
- Cruise, R., Dudley, R. y Thayer, J. (1984). *A resource guide for introductory statistics*. Nueva York: Kendall/Hunt.
- Cumming, G., Williams, J. y Fidler, F. (2004). Replication, and researchers' understanding of confidence intervals and standard error bars. *Understanding Statistics*, 18(3), 299-311.
- Fidler, F. y Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings 55th International Statistics Institute Session CD-ROM*. Sidney, Australia: International Statistical Institute.
- De la Fuente, E. I. y Díaz, C. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento, Volumen especial 2004*, 161-167.
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M. y Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A Pre-K-12 curriculum framework*. Alexandria, VA: ASA.
- Gea, M. M., López-Martín, M. M. y Roa, R. (2015). Conflictos semióticos sobre la correlación y regresión en los libros de texto de Bachillerato. *Avances de Investigación en Educación Matemática*, 8, 29-49.
- Godino, J. D., Batanero, C y Font, V. (2019). The onto-semiotic approach: Implications for the prescriptive character of didactics. *For the Learning of Mathematics*, 39(1), 38-43.
- Harradine, A., Batanero, C. y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education* (pp. 235-246). Dordrecht, Holanda: Springer.
- López-Martín, M. M., Batanero, C., Díaz-Batanero, C. y Gea, M. (2016). La inferencia estadística en las Pruebas de Acceso a la Universidad en Andalucía, *Revista Paranaense de Educação Matemática*, 5(8), 33-59.
- López-Martín, M. M., Batanero, C. y Gea, M. M. (2019a). Prospective high school teachers' interpretation of hypothesis tests and confidence intervals. En U. T. Jankvist, M. Van den Heuvel-Panhuizen y M. Veldhuis (Eds.) (2019). *Proceedings XI Congress of the European Society for Research in Mathematics Education*. Utrecht, Holanda: ERME.
- López-Martín, M. M., Batanero, C. y Gea, M. M. (2019b). ¿Conocen los futuros profesores los errores de sus estudiantes en inferencia? *Bolema*, 33(64), 672-693.
- Mayo, D. G. (1981). In defense of the Neyman-Pearson theory of statistics. *Philosophy of Science* 48, 269-280.
- Ministerio de Educación, Cultura y Deporte (MECD) (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato*. Madrid: MECD.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D. y Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin &*

- Review, 23(1), 103-123.
- Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transaction of the Royal Society of London, series A, Mathematical and Physical Sciences* 236 (767), 33-380.
- Olivo, E. (2008). *Significados del intervalo de confianza en la enseñanza de la ingeniería en México*. Trabajo de Tesis Doctoral. Universidad de Granada.
- Olivo, E., Batanero, C. y Díaz, C. (2008). Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Educación Matemática*, 20(3), 5-32.
- Wilkinson, L., & Task Force on Statistical Inference (TFSI) (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Yáñez, G. y 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 XIII, Comunicaciones de los grupos de investigación*. XIII Simposio de la SEIEM. Santander: SEIEM. Disponible en https://seiem.es/docs/comunicaciones/GruposXIII/depc/Yanez_Behar_R.pdf
- Yaremko, R. M., Harari, H., Harrison, R. C. y Lynn, E. (2013). *Handbook of research and quantitative methods in psychology: For students and professionals*. Hilldale, NJ: Erlbaum.

Referencias de los autores

- Antonio Francisco Roldán López de Hierro, Universidad de Granada (España).
aroldan@ugr.es
- Carmen Batanero, Universidad de Granada (España). batanero@ugr.es
- Rocío Álvarez-Arroyo, Universidad de Granada (España). rocioarroyo@ugr.es

High school students' understanding of confidence intervals

Antonio Francisco Roldán López de Hierro, Universidad de Granada

Carmen Batanero, Universidad de Granada

Rocío Álvarez-Arroyo, Universidad de Granada

Due to its many applications, an important topic in statistics is confidence intervals. Many degrees include a subject that deals with this topic, which is also essential in research. On the one hand, current teaching pays attention to theoretical and practical aspects on confidence intervals, forgetting the importance of reflecting on their interpretation. On the other hand, prior literature on the understanding of the confidence intervals have focused on university students and researchers who used these intervals in their publications. However, research on the understanding of high school students is almost nonexistent. The objective of our work is to describe an exploratory study on the understanding and interpretation of confidence intervals by students of the second year of high school. In order to analyze this understanding and interpretation, we theoretically drew on the ontosemiotic perspective and designed a questionnaire using items from previous research as well as other original items. The first part of the questionnaire includes six multiple-choice items that evaluate conceptual understanding on this subject; while the second part is an open problem taken from the previous entrance to university tests, in which students use their knowledge in a practical case. The sample was constituted by 58 students of the social sciences modality in the Spanish system. In this paper we describe the proposed questions and examine these students' answers. Although some of the errors observed had already been shown in previous research, we also contribute to research in the field by documenting new errors. A major conclusion is the need to increase the lesson time devoted to interpret confidence intervals, and to reduce the lesson time for computational work, which is today solved by technology.