

Un estudio exploratorio de dificultades de comprensión del intervalo de confianza

Eusebio Olivo y Carmen Batanero

Resumen

En este trabajo presentamos los resultados de un cuestionario orientado a evaluar la comprensión que tienen los estudiantes de ingeniería sobre los intervalos de confianza pasado a una muestra de 48 estudiantes. Los resultados apuntan a que el concepto podría ser comprendido por la mayoría de los estudiantes, aunque aparecen errores conceptuales y procedimentales, algunos no descritos en las investigaciones previas.

Abstract

We present results of a questionnaire directed to assess students' understanding of confidence intervals in a sample of 48 students of engineering. Results suggest that this concept might be easy to understand, although conceptual and procedural errors appear, some of them not previously described in research.

Introducción

El intervalo de confianza es un tema estudiado en todos los cursos universitarios de estadística e incluso en la educación secundaria; por ejemplo en España se incluye en el Bachillerato de Ciencias Sociales. Por otro lado, diferentes asociaciones profesionales, como la American Psychological Association (APA) o la American Educational Research Association (AERA), recomiendan su uso por parte de los investigadores, para complementar los contrastes de hipótesis y mejorar de este modo los errores denunciados en la práctica de la inferencia estadística (Morrison y Henkel, 1970; Vallecillos, 1994, Harlow, Mulaik y Steiger, 1997; Batanero, 2000). En este sentido se expresa Davies (1998), quien sugiere que los intervalos de confianza proporcionan la misma información que un valor p , pero añaden información sobre el tamaño de los efectos.

Este cambio metodológico, requiere asegurar que las dificultades sobre los tests de hipótesis no se repiten –o al menos no con tanta intensidad– en los intervalos de confianza, tema donde la investigación didáctica es todavía incipiente. Es por ello importante llevar a cabo estudios de evaluación sobre esta comprensión, que pudiera no ser inmediata, puesto que el intervalo de confianza se apoya en otros conocimientos previos y tiene un doble carácter, como concepto y como procedimiento. Considerado como concepto, y en el contexto de estimar un parámetro poblacional, un intervalo de confianza es un rango de valores (calculado a partir de los datos de una muestra) en el cual podría encontrarse el verdadero valor

del parámetro, junto con un coeficiente de confianza que indica el porcentaje de muestras tomadas en las mismas condiciones, en las cuales el intervalo cubriría el verdadero valor del parámetro.

Como procedimiento, da una regla general de construcción de dicho rango de valores a partir de un estadístico calculado en los datos de la muestra, para el parámetro correspondiente. Calculada la desviación típica del estadístico (error estándar) y obtenido un valor crítico correspondiente a la mitad del valor del coeficiente de confianza elegido, el producto del valor crítico por el error estándar se sumaría y restaría al valor del estadístico en la muestra, obteniendo así los límites del intervalo. Este procedimiento general se particulariza dependiendo del parámetro a estimar (media, proporción, varianza, etc.) y según las condiciones (tipo de distribución, qué se conoce de la misma, etc.), puesto que ellos determinan la distribución muestral del estadístico. Vemos, entonces, que la comprensión del intervalo de confianza requiere una serie de otros objetos matemáticos previos (tanto conceptos como procedimientos) como población y muestra, estadístico y parámetro, error estándar, y cálculo del mismo para diversos estadísticos, distribución muestral, valor crítico o uso de las tablas de diferentes distribuciones.

En este trabajo continuamos las investigaciones previas de Cumming, William y Fidler (2004), Behar (2001) y Terán (2006), presentando los resultados obtenidos en el estudio piloto de un cuestionario en proceso de elaboración. El objetivo principal de dicho instrumento de evaluación es hacer diagnóstico sobre la comprensión que tienen del concepto intervalos de confianza, los alumnos universitarios, aunque la mayoría de sus ítems podrían ser aplicados en el bachillerato en aquellas especialidades que incluyen el tema.

Antecedentes

Como hemos indicado, las investigaciones sobre comprensión del intervalo de confianza son pocas y, en su mayor parte, se centran en investigadores y no en estudiantes o bien no llegan a un análisis completo, como pretendemos realizar en nuestro estudio. Entre las investigaciones centradas en investigadores, encontramos las de Cumming, William y Fidler (2004), quienes estudian los errores de interpretación de intervalos de confianza, para el caso particular de la media. La mayoría de investigadores en su estudio esperaban (erróneamente) una alta probabilidad de replicación, esperando que en una nueva muestra la media caiga de nuevo en el intervalo de confianza original. Otra creencia errónea muy común en los investigadores (Schenker y Gentleman, 2001) es que los intervalos de confianza de dos medias de muestras independientes son sólo significativamente diferentes cuando se tocan justo extremo con extremo. También confunden el cálculo de intervalos de confianza para medias independientes y relacionadas.

Respecto a los trabajos con estudiantes universitarios, Fidler y Cumming (2005) indican que al estudiar la evidencia a favor de la hipótesis nula, el 44% de los estudiantes interpreta incorrectamente un *valor p* pequeño en un contraste de hipótesis, pero sólo el 18% interpreta incorrectamente los resultados cuando se

presentan mediante intervalos de confianza. Otros estudiantes en la investigación de Fidler y Cumming sólo consideran a los intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial o tienen ideas equivocadas sobre cómo los distintos conceptos que intervienen en los intervalos de confianza se relacionan entre sí. Por ejemplo, sólo el 16% de los estudiantes en la investigación citada pudo contestar correctamente a la relación entre ancho del intervalo y tamaño de la muestra.

Garfield, delMas y Chance (1999) usando un software de simulación (Sampling SIM program) tratan de favorecer el aprendizaje de conceptos básicos de la estadística, entre ellos los intervalos de confianza. Los autores plantean un modelo en el cuál destacan los elementos conceptuales que el estudiante debería de entender acerca de los intervalos de confianza, las competencias básicas que el estudiante debería adquirir y algunas concepciones erróneas. También diseñan un conjunto de actividades, así como un pretest sobre los conocimientos requeridos antes de la enseñanza.

En base al trabajo anterior, Behar (2001) construye un cuestionario sobre comprensión del intervalo de confianza y contraste de hipótesis. Concluye, de las respuestas al cuestionario, que los alumnos no relacionan la influencia del ancho del intervalo sobre el nivel de confianza. También observa dificultad de comprensión de la definición del intervalo de confianza, porque se piensa que los valores que lo constituyen son los que toma la variable aleatoria que define la población, o valores del estadístico que se usa como estimador, lo cual podría deberse al propio procedimiento de construcción del intervalo, que se deduce de la distribución muestral del dicho estadístico. Los participantes en su estudio no asocian la confianza a un mecanismo aleatorio generador de intervalos, a partir de muestras aleatorias, ni el nivel de confianza, con la frecuencia relativa a la larga, de que los intervalos generados por tal mecanismo aleatorio incluyan al verdadero parámetro de la población. La utilidad de los intervalos de confianza para tomar decisiones sobre hipótesis, parece no ser comprendida; posiblemente por no considerar los valores del intervalo como un conjunto de valores plausibles del parámetro.

La investigación

Como hemos indicado, el fin de nuestra investigación es profundizar y continuar las anteriormente descritas. En concreto, pretendemos construir un cuestionario comprensivo de evaluación de las dificultades de comprensión, cálculo e interpretación de intervalos de confianza, incluyendo algunas propiedades o el cálculo de intervalos para algunos parámetros que no fueron considerados en las investigaciones citadas. Este cuestionario se ha construido siguiendo el modelo de metodológico de Díaz (Batanero y Díaz, 2006; Díaz y de la Fuente, 2007). El primer paso en la construcción del cuestionario fue llevar a cabo una definición semántica precisa de la variable objeto de medición. Puesto que la *comprensión del intervalo de confianza* es un constructo psicológico inobservable, es preciso llevar a cabo una definición detallada de los contenidos que serían evaluados. Ya que el estudio se centra en estudiantes de ingeniería, se llevó a cabo un análisis del contenido

relacionado con los intervalos de confianza en una muestra de 18 libros de estadística dirigidos la enseñanza en ingeniería. Del análisis se dedujo las propiedades y relaciones del intervalo de confianza con otros objetos matemáticos, así como los procedimientos de construcción de intervalos de confianza que serían objeto de evaluación. El estudio se ha recogido en Olivo (2006).

Seguidamente se procedió a la elaboración y depuración de ítems que pudieran evaluar cada uno de los contenidos fijados. También se adaptó su formato y redacción, para seguir las pautas marcadas en Osterlind (1989) sobre claridad del enunciado, número de distractores, formulación de la pregunta, etc y se realizaron pruebas de legibilidad con un grupo reducido de estudiantes. La colaboración de 10 expertos (profesores de estadística e investigadores en educación estadística) sirvió para seleccionar los ítems que finalmente formarían parte del cuestionario, mediante juicio de expertos. El cuestionario que analizaremos a continuación está compuesto por once ítems de opciones múltiples y cuatro problemas abiertos que en su mayoría fueron recopilados de diversas investigaciones. En este trabajo presentamos los resultados de un total de 48 estudiantes de ingeniería en el Instituto Tecnológico y de Estudios Superiores de Monterrey.

Ítems de opciones múltiples

En la Figura 1 se reproducen los once ítems de opciones múltiples. El ítem 1 (Cruise, Dudley y Thayer, 1984) corresponde a la definición de intervalo de confianza, tratando de detectar los siguientes sesgos: el intervalo es constante (distractor a), contiene siempre al parámetro (distractores b y d). El ítem 2 (Gardfield y delMas, 2004) evalúa la comprensión del efecto del tamaño de muestra sobre la precisión (ancho de intervalos) cuando se mantiene constante el coeficiente de confianza. Los ítems 3 y 4 (Cruise, Dudley y Thayer, 1984) evalúan la comprensión del efecto del coeficiente de confianza sobre el ancho del intervalos y el 5 (Behar, 2001) el efecto de la varianza sobre el ancho de los intervalos.

El ítem 6 (Behar, 2001) evalúa la comprensión de la variabilidad del intervalo en diferentes muestras y la interpretación de su resultado. El distractor b) da una interpretación contraria y el c) el error de que el intervalo se refiere a la variable y no al parámetro. Los ítems 7 y 8 evalúan el conocimiento procedimental en la construcción de intervalos de confianza para la media. El ítem 7 (Cruise, Dudley y Thayer, 1984) se refiere al caso de distribución normal con desviación típica conocida. El distractor a) incluye un error sobre la desviación típica del estimador (no dividir por el tamaño de la muestra); el d) no multiplica los límites por el valor crítico y el c) contiene los dos errores simultáneamente. El ítem 8 (Miller, Freund y Johnson, 1997) corresponde al caso de muestra grande, por lo que se puede usar la distribución normal. El distractor a) incluye un error en la determinación del valor crítico a partir de la tabla de la distribución normal, el c) error en la desviación típica al no dividir por n y el d) los dos errores simultáneamente.

Figura 1. Ítems de opciones múltiples

Ítem 1. En un intervalo de confianza:

- a. De una muestra a otra, el intervalo es constante.
- b. Se especifica un rango de valores dentro de los cuales cae el parámetro con seguridad.
- c. **Indica un intervalo de posibles valores para el parámetro, y un porcentaje de intervalos que cubrirán, aproximadamente dicho valor, para el mismo tamaño de muestra.**
- d. Siempre contienen el parámetro poblacional.

Ítem 2. Dos muestras diferentes se toman de una población donde la media poblacional y la desviación estándar poblacional son desconocidas. La primera muestra tiene 25 datos, y la segunda muestra 64 datos. Se construye un intervalo de confianza de 95 % para cada muestra para estimar la media poblacional. ¿Qué intervalo de confianza esperaría que tenga mayor precisión?

- a. Espero que ambos intervalos de confianza tengan la misma precisión.
- b. **Espero que el intervalo de confianza basado en una muestra de 64 datos sea más preciso**
- c. Espero que el intervalo de confianza basado en la muestra de 25 datos sea más preciso.
- d. No puedo determinar cuál de los dos tendrá más precisión

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

- a. El intervalo de confianza no cambia.
- b. El intervalo de confianza será más ancho
- c. **El intervalo de confianza será más angosto**
- d. El cambio en el intervalo de confianza no es predecible.

Ítem 4. En un intervalo de confianza, el ancho del intervalo puede ser reducido:

- e. Disminuyendo el tamaño de la muestra.
- f. **Bajando el nivel de confianza (por ejemplo de 0.99 a 0.90)**
- g. Aumentando la magnitud de $\sigma_{\bar{x}}$.
- h. Un aumento en el tamaño de la población.

Ítem 5. ¿Cuál de las siguientes afirmaciones es verdadera?

- 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 no cambia.

Ítem 6. Un intervalo del 95% de confianza para la diferencia media en la producción de leche de una ganadería después de un tratamiento resultó ser (1.5- 3.5) Litros/vaca. ¿Cuál de las siguientes afirmaciones es verdadera?

- a. No sabemos el verdadero aumento medio en la producción, pero en el 95% de intervalos calculados con el mismo tamaño de muestra y población se incluye el aumento medio en la producción.
- b. **Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad de que el tratamiento no produce ningún efecto.**
- c. Debido a que el intervalo de confianza no contiene el cero, hay 95% de probabilidad que el verdadero aumento en la producción es 2.5 Litros/vaca.

Figura 1. Ítems de opciones múltiples (continuación)

Ítem 7. La media muestral de 100 observaciones en una prueba de matemáticas es 75. Encuentre el intervalo de confianza al 95% para la media de la población, asumiendo que $\sigma = 7$.

- a. (61.28, 88.72)
- b. **(73.63, 76.37)**
- c. (68, 82)
- d. (74.3, 75.7)

Ítem 8. Se han obtenido los siguientes datos de emisión diaria de óxidos de azufre, para una muestra de tamaño $n=100$, media: $\bar{x} = 18$ y cuasivarianza $s^2=36$. Elabore un intervalo de confianza de 95% para la verdadera emisión diaria promedio de óxidos de azufre.

- a. (17.016, 18.984)
- b. **(16.824, 19.176)**
- c. (6.24, 19.76)
- d. (8.16, 27.84)

Ítem 9. La distribución Muestral utilizada en la construcción de intervalos de confianza para la varianza en muestras pequeñas, tomadas de una población normal, es:

- a. Distribución t de Student
- b. **Distribución Ji-cuadrada**
- c. Distribución Normal
- d. Distribución F

Ítem 10. Al calcular un intervalo de confianza del 90% para la media poblacional, para un grupo de puntuaciones distribuido normalmente, usted pudiera usar un valor z de:

- a. 1.96
- b. **1.65**
- c. 0.90
- d. 1.29

Ítem 11. Considere el gráfico siguiente del rendimiento medio de cebada en 1980, 1984 y 1988 junto con un intervalo de 95% de confianza

Intervalos del 95% para la media

Año	N	Media	StDev	-----+-----+-----+-----+--
1980	6	184.00	2.61	(----*---)
1984	5	212.40	14.36	(-----*-----)
1988	5	182.40	1.82	(-----*-----)
				-----+-----+-----+-----+--
Desv. Típica conjunta = .19				180 195 210 225

¿Cuál de las siguientes afirmaciones es verdadera?:

- a. Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay buena evidencia que las medias de las muestras difieran.
- b. La estimación de la media de la población en 1980 es menos precisa que en 1988.
- c. Puesto que los intervalos de confianza para 1980 y 1984 no se solapan, hay poca evidencia que las medias de las poblaciones respectivas difieran.
- d. **Puesto que los intervalos de confianza para 1980 y 1988 tienen considerable solape, hay poca evidencia que las medias de las poblaciones difieran.**

El ítem 9, de elaboración propia, evalúa el conocimiento de la distribución muestral necesaria para calcular un intervalo de confianza para la varianza. El ítem 10 evalúa el conocimiento de determinación de valores críticos en la distribución del estadístico. El distractor a) incluye el error de considerar un coeficiente de confianza 95% en la determinación del valor crítico el c) confunde el coeficiente de confianza (90%) con el valor crítico y el d) hace un error al determinar el valor crítico acumulando el coeficiente .90 solo en la parte inferior de la distribución. Finalmente el ítem 11, tomado de Behar (2001) evalúa la interpretación de gráficos de intervalos de confianza para tomar una decisión sobre la diferencia de medias. Los distractores a) y c) dan la interpretación contraria (los intervalos han de coincidir para que exista diferencia); el b) evalúa la confusión sobre el concepto de precisión.

En la tabla 1 presentamos la frecuencia de respuestas a cada opción de cada ítem, la proporción de respuestas correctas y el índice de discriminación, definido como diferencia en proporción de aciertos entre el grupo que tiene puntuación superior e inferior en el total de la prueba (dividido el grupo en tres partes de acuerdo a su puntuación). Se considera discriminativo el ítem si la diferencia es mayor a 0.3.

Tabla 1. Resultados¹ en los ítems (n=48)

Ítem	Frecuencia				% correcto	Índice discriminación	Contenido del ítem
	A	B	C	D			
1	6	5	36	1	75	0.81	Definición de intervalo de confianza
2	2	12	31	3	25	-0.56	El ancho del intervalo disminuye cuando aumenta el tamaño de la muestra
3	9	3	34	2	70	0.88	El ancho del intervalo aumenta con el nivel de confianza
4	14	10	3	21	20	0.39	El ancho del intervalo aumenta con el nivel de confianza
5	4	14	14	16	29	-0.31	El ancho del intervalo aumenta con la varianza
6	31	6	10	--	64	0.01	Variación del intervalo en diferentes muestras
7	9	33	2	4	68	0.88	Construir un intervalo para la media en una muestra grande, σ conocida
8	5	33	7	2	68	0.63	Construir un intervalo para la media en una muestra grande, σ desconocida
9	17	27	3	1	56	0.91	Elegir un modelo de distribución muestral del estadístico
10	5	36	4	3	75	0.63	Determinar valores críticos en la distribución del estadístico
11	10	2	3	33	68	0.75	Interpretar intervalos de confianza en forma gráfica

¹ Soluciones correctas en negrita)

En general se observa un buen conocimiento del intervalo de confianza, con la mayoría de los ítems resueltos por más de la mitad de los alumnos en la muestra.

Las principales excepciones son los ítems: 2,4 y 5. En el ítem 2 sólo el 25% de los estudiantes conoce la relación correcta entre precisión y tamaño de muestra, coincidiendo con la investigación de Fidler y Cumming (2005) en cuya investigación sólo el 16% de los alumnos dieron la respuesta correcta en este ítem. En los ítems 4 y 5 sobre el mismo contenido (efecto de la varianza de la población sobre la anchura del intervalo), los resultados son consistentes. Las respuestas se reparten prácticamente al azar (los estudiantes piensan por igual que el aumento de varianza hace crecer, disminuir o deja igual el intervalo). Estos resultados coinciden de nuevo con los de Cumming y Fidler (2005) en cuya investigación el 20% de los estudiantes pensaban que el ancho del intervalo de confianza se incrementaría si se aumentara el tamaño de la muestra, un 29% que el ancho del intervalo de confianza no se vería afectado y el 36% de los estudiantes no sabía si existía alguna relación. Hubo también algunas dificultades en recordar la distribución Chi Cuadrada necesitada para la estimación de la varianza (ítem 9), donde el 35% de los estudiantes eligieron erróneamente la distribución T .

En menor proporción se presentan los siguientes errores: en distintas muestras se obtendrá el mismo intervalo (12% en ítem 1), el coeficiente de confianza no hace cambiar el intervalo (18% en el ítem 3), si el intervalo no contiene el cero, es indicio de falta de efecto (12% en ítem 6), el intervalo se refiere al valor de la variable y no al parámetro (21% en el ítem 6), error en el cálculo de la desviación típica del estimador (18% en el ítem 7 y 14% en el ítem 8), error en la determinación del valor crítico en la distribución del estadístico debido a una falla en la comprensión del nivel de confianza (10% en el ítem 10) y error en la interpretación gráfica, incluido en ello error en la comprensión de que la inferencia es sobre las medias poblacionales y no sobre las medias muestrales (20%, ítem 11). Los ítems en general discriminan a los estudiantes con alto y bajo rendimiento, excepto los ítems 2 (efecto del tamaño de la muestra sobre el ancho del intervalo), 5 (efecto de la varianza sobre la anchura del intervalo) y 6 (interpretación del intervalo de confianza).

Ítems abiertos

El cuestionario también contenía cuatro ítems abiertos, con objeto de evaluar las estrategias de resolución y argumentación de los estudiantes. A continuación analizamos el contenido de estos problemas y las posibles soluciones que se han puntuado en escala 0-2 atendiendo a su corrección.

Ítem abierto 1. Explica con tus propias palabras la diferencia entre una estimación puntual y un intervalo de confianza.

Este problema es de elaboración propia y evalúa el conocimiento de la *definición de intervalo de confianza*. Se ha puntuado bajo el siguiente criterio:

- *Definición correcta* (2 puntos; en total 43 casos). Un ejemplo es el siguiente: *“Una estimación puntual es un solo valor numérico utilizado para estimar el parámetro correspondiente de la población. En tanto que un intervalo de confianza es un intervalo aleatorio que consta de dos valores numéricos que,*

con un grado de confianza específico, se considera incluya el parámetro que se está estimando” (Alumno HV). El estudiante utiliza correctamente los conceptos estimación puntual, parámetro y población, intervalo de confianza, aleatorio y grado de confianza; comprende el carácter aleatorio de los límites del intervalo, y su definición incluye los elementos esenciales.

- *Explicación parcialmente correcta* (1 punto; 5 estudiantes). Cuando el alumno define correctamente solo una de los dos conceptos (1 caso) o las definiciones son incompletas (4 casos), como la siguiente: “Una estimación puntual es cuando tratamos de “adivinar un valor” directamente diciendo un solo número ej. 65. Un intervalo de confianza es cuando estimamos que la respuesta va a estar entre ciertos números ej. entre 62 y 67. Así tenemos más confianza de nuestra estimación” (alumno JJA). Este alumno da definiciones no muy rigurosas, sin hacer mención a la diferencia entre estadístico y parámetro o población y muestra, coeficiente de confianza o el carácter aleatorio de los límites del intervalo.
- No ha habido definiciones totalmente incorrectas en esta pregunta.

Ítem abierto 2. Construya un intervalo de confianza al 95% para la media de una población normal de desviación típica σ desconocida si en una muestra de tamaño 10, la media de la muestra es $\bar{x} = 25$ y la estimación de la desviación típica en la muestra es $s = 6$

Este problema tomado de Cruise, Dudley y Thayer (1984) evalúa el conocimiento procedimental en la construcción de intervalos de confianza para la media de una población normal con desviación típica no conocida. Se ha puntuado bajo el siguiente criterio:

- *Solución correcta* (2 puntos; 28 estudiantes): El estudiante determina el intervalo, empleando la distribución muestral correcta; encuentra correctamente el valor crítico y a partir de él los límites del intervalo. Puesto que el tamaño de la muestra es pequeño y σ es desconocida se debe usar la distribución T de student. Un ejemplo es el siguiente: “Sea μ la media. Se verifica $\bar{x} - t_{\alpha/2} \frac{S}{\sqrt{n}} < \mu < t_{\alpha/2} \frac{S}{\sqrt{n}}$, entonces $25 - t_{\alpha/2} \frac{6}{\sqrt{10}} < \mu < 25 + t_{\alpha/2} \frac{6}{\sqrt{10}}$, es decir, $25 - \frac{2.262 * 6}{\sqrt{10}} < \mu < 25 + \frac{2.262 * 6}{\sqrt{10}}$. Sustituyendo en las inecuaciones se llega al intervalo $20.708 < \mu < 29.292$ ” (Alumno JR). El estudiante usa correctamente las expresiones simbólicas, *media* distinguiendo la media de la población y la muestra y sabe que el intervalo se refiere a la de la población. El alumno recuerda la distribución muestral de la media para el caso de muestras pequeñas; al referirse al subíndice $\alpha/2$ manifiesta que comprende la relación del valor crítico con el nivel de confianza 95%. Recuerda las probabilidades necesarias para calcular el intervalo, sustituye los datos del problema en las inecuaciones anteriores y calcula los valores críticos en la distribución de T de Student, usando correctamente las tablas. Finalmente lleva a cabo varias operaciones algebraicas con inecuaciones para llegar finalmente al intervalo pedido.

- *Solución parcialmente correcta* (1 punto; 12 estudiantes). El estudiante plantea el intervalo con la distribución muestral correcta, pero hace algunos errores. Un caso es cuando confunde los grados de libertad de la distribución T . Otros alumnos cometen un error en la varianza del estimador, dividiendo por la raíz cuadrada del tamaño de muestra. Dos alumnos conocen todos los pasos la construcción del intervalo, pero no llegan a encontrar el valor crítico, por hacer un manejo incorrecto de las tablas de la distribución.
- *Respuesta incorrecta* (0 puntos, 8 estudiantes): No se responde o se responde en forma incorrecta. Los errores se producen porque confunde la distribución muestral (8 casos) usando la distribución normal en lugar de la T . En general estos estudiantes conocen el procedimiento general de construcción de intervalos, pues dividen la probabilidad total bajo la curva normal en tres partes, dejando el coeficiente de confianza (95% en el centro) y el nivel 5% dividido en dos partes a cada lado (procedimiento y propiedades). Calcula el valor crítico en la tabla de la normal estándar (procedimiento) usando las tablas (lenguaje) correctamente. Luego lleva a cabo correctamente varias operaciones algebraicas (procedimientos), pero no llega al intervalo correcto por usar una distribución incorrecta.

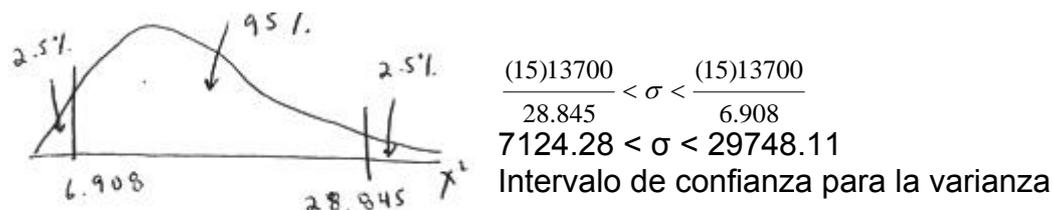
Ítem abierto 3. Sea σ^2 la varianza de la distribución de la tensión disruptiva. El valor calculado de la varianza muestral es $s^2=13700$, $n=16$. Calcular el intervalo de confianza de 95% para σ .

Este problema tomado de Devore (2005, pg. 310) evalúa el conocimiento del procedimiento de construcción de intervalos para la estimación de una varianza. Se ha puntuado bajo el siguiente criterio:

- *Solución correcta* (2 puntos; 20 casos): El estudiante determina el intervalo, empleando la distribución Chi cuadrado; determina correctamente el valor crítico y a partir de él los límites del intervalo. Un ejemplo es el siguiente (Alumno AC): “Sea σ^2 la varianza. Sabemos que se verifica $\frac{(n-1)S^2}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{(1-\alpha/2)}^2}$, entonces $\frac{(16-1)13700}{\chi_{\alpha/2}^2} < \sigma^2 < \frac{(16-1)13700}{\chi_{(1-\alpha/2)}^2}$, es decir, $\frac{(16-1)13700}{27.488} < \sigma^2 < \frac{(16-1)13700}{6.262}$. Sustituyendo en las inecuaciones se llega al intervalo $7975.9895 < \sigma^2 < 32816.991$; luego obteniendo raíz cuadrada a ambos lados de la desigualdad se obtiene el intervalo de confianza para la desviación estándar poblacional $89.308 < \sigma < 181.154$ ”. El estudiante diferencia la varianza de la población y la muestra (conceptos) y saber que el intervalo se refiere a la de la población (propiedad). Recuerda la distribución muestral de la varianza para el caso de muestras pequeñas (propiedad), en este caso la Chi cuadrado (concepto). Además se refiere a los valores críticos $\alpha/2$; $(1-\alpha/2)$ correspondientes al nivel de confianza 95% (conceptos y propiedades). Por otro lado debe recordar que las probabilidades necesarias para calcular el intervalo son precisamente estas, siendo $1-\alpha$ el coeficiente de confianza (propiedad). El alumno sustituye los datos del problema en las inecuaciones anteriores y calcula los valores críticos en la tabla del Chi cuadrado (procedimientos), usando correctamente las tablas (lenguaje). El

alumno lleva a cabo varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido.

- *Solución parcialmente correcta* (1 punto; 20 estudiantes). El estudiante plantea el intervalo con la distribución muestral correcta, pero hace algunos errores. Un caso es cuando confunde los grados de libertad de la distribución Chi cuadrado (13 alumnos), como en el ejemplo siguiente (Alumno EG), donde además hay una confusión de la notación:



El alumno introduce una representación gráfica de la distribución Chi cuadrado que ha identificado correctamente. Divide la probabilidad total bajo la curva en tres partes, dejando el coeficiente de confianza (95% en el centro) y el nivel 5% dividido en dos partes a cada lado (procedimiento y propiedades). Sin embargo tiene un conflicto al confundir los grados de libertad, tomando 16 en vez de 15. Calcula los valores críticos en la tabla del Chi cuadrado (procedimiento) y usa las tablas (lenguaje) correctamente, pero al ser los grados de libertad incorrectos el valor obtenido es incorrecto. Tiene también un conflicto de notación, al usar la terminología de la varianza con la correspondiente a la desviación típica. El alumno lleva a cabo correctamente varias operaciones algebraicas (procedimientos) con inecuaciones (concepto), para llegar finalmente al intervalo pedido. Pero, por la razón señalada el resultado es incorrecto. Otros alumnos puntuados como parcialmente correctos, los alumnos conocen la construcción del intervalo, pero no llegan a encontrar el valor crítico, por hacer una lectura incorrecta de las tablas de la distribución (4 casos) o bien hacen errores en los cálculos al no manejar las inecuaciones (2 casos) o bien sencillamente no completa los cálculos (1 caso).

- *Respuesta incorrecta* (0 puntos; 8 estudiantes): No se responde o se responde en forma incorrecta. Los errores se producen porque confunde la distribución muestral (5 casos) llegando a escribir la distribución T en el intervalo o recuerda levemente un límite de la fórmula del intervalo de confianza para la varianza (3 casos), como en el caso siguiente (Alumno

$$\text{MV): } \sigma^2 = \frac{(n-1)s^2}{\chi_{1-\alpha/2, v}^2} = \frac{(15)(13700)}{\chi_{1-\alpha/2, v}^2} \quad (1-\alpha)\% = 95\% \quad ; \quad \alpha = 1-95\%; \quad \alpha = 5\%$$

$$\frac{(15)13700}{\chi_{.025, 16}^2} = \frac{205500}{28.845} = 7124.284$$

El alumno recuerda solamente un límite de la fórmula del intervalo de confianza para la varianza. Sabe como identificar el valor α , que al dividir por dos, y usará posteriormente para obtener el valor crítico de la distribución chi cuadrada. Calcula el valor crítico en la tabla del Chi cuadrado (procedimiento) y usa las tablas (lenguaje) correctamente, pero

confunde los grados de libertad. Luego lleva a cabo correctamente varias operaciones algebraicas (procedimientos), pero no finaliza la solución.

Ítem abierto 4. La siguiente salida de computadora presenta una muestra simulada de una población normal con $\mu=130$ y $\sigma=10$. Luego se usó un comando para establecer un intervalo de confianza del 95% para μ .

118.69 144.22 138.82 125.93 136.19 148.73 133.39 134.18 127.26 145.34 121.96 125.43 141.23
150.021 126.89 118.49 137.22 136.78 119.52 125.62 134.86 116.41 134.31 138.05 140.82

Mean	Median	TrMean	StDev	95% Confidence Interval	
132.82	134.31	132.78	9.74	128.798	136.840

Escriba el intervalo de confianza que se obtuvo e interprete el resultado.

Este ítem es una adaptación de un problema tomado de Johnson y Kubly (2004, pg.363) y se usa para evaluar la competencia para interpretar intervalos de confianza obtenidos de un programa de ordenador. Las respuestas obtenidas se han puntuado en una escala 0-1-2 bajo el siguiente criterio:

- **Solución correcta** (2 puntos; 35 alumnos). El alumno reconoce la terminología asociada a los extremos del intervalo en la salida del ordenador. Además, da la interpretación correcta, que se refiere al porcentaje de intervalos que incluirían la media de la población, en las mismas condiciones. Un caso es el siguiente (Alumno RF): *“Las dos últimas columnas de la tabla en la salida nos indican los extremos del intervalo de confianza para la media de la población a un nivel de confianza de 95%. Simbólicamente podemos escribirlo: $128.7978 < \mu < 136.840$. La interpretación es que en el 95% de muestreos ejecutados en estas mismas condiciones estará contenido el verdadero valor de la media, aunque no sabemos si esta muestra particular lo contiene”*. El alumno reconoce los extremos del intervalo de confianza y usa los conceptos de intervalo, extremos, media de la población y nivel de confianza. Transforma adecuadamente el intervalo dado en términos numéricos en una tabla a expresión simbólica. Su interpretación correcta relaciona el coeficiente de confianza con el porcentaje de intervalos que cubre el parámetro, si el muestreo se ejecuta en las mismas condiciones. También indica que no sabemos si esta muestra particular lo contiene. (razonamiento)
- **Respuesta parcialmente correcta** (1 punto; 13 alumnos): Identifica el intervalo pero no lo interpreta o bien da una interpretación errónea, por ejemplo, que la probabilidad de que el intervalo dado incluya a la media es del 95%, lo cual sería una interpretación bayesiana y no frecuencial, como el caso siguiente (Alumno BT): *“ $128.7978 < \mu < 136.840$. Esto significa que tenemos un 95% de probabilidad de que el valor verdadero de la media poblacional (μ) se encuentre entre 128.79 y 136.84”*. El alumno identifica el intervalo y lo escribe simbólicamente en forma correcta, usando una notación correcta. Da sin embargo una interpretación bayesiana al intervalo, considerando los límites fijos (fija el intervalo) y el parámetro como aleatorio (hay una probabilidad de que la media esté en el intervalo particular). En realidad la interpretación clásica es que el parámetro (media) está fijo y los límites del intervalo varían

de un intervalo a otro; es por esto que sólo el 95% los cubren. No hubo respuestas incorrectas en este ítem.

Tabla 2. Resultados en los problemas abiertos

Respuesta	Ítem abierto 1		Ítem abierto 2		Ítem abierto 3		Ítem abierto 4	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje	Frecuencia	Porcentaje
Incorrecta	0	0.00	8	0.17	8	0.16	0	0.00
Parcialmente correcta	5	0.11	12	0.25	20	0.41	13	0.27
Correcta	43	0.89	28	0.58	20	0.41	35	0.72

En la Tabla 2 se muestran los resultados; los porcentajes de los alumnos que han contestado correctamente los cuatro ítems abiertos son superiores al 50%, excepto el ítem abierto 3, que evalúa la estimación de la varianza, donde sólo se alcanza el 41% de respuestas totalmente correctas. En este ítem un 41% de los estudiantes tienen una confusión en la obtención de los grados de libertad de la distribución correspondiente o bien hacen un mal manejo de las tablas y un 16% confunde la distribución muestral o sólo recuerda la fórmula de un límite del intervalo. Algunos estudiantes dieron definiciones incompletas o imprecisas en el ítem abierto 1, un 25% confunde los grados de libertad en la distribución T o la desviación típica del estimador y un 17% trata de usar la distribución normal en vez de la T en el ítem 2. En el ítem 4 un 27% de los estudiantes no interpretan el intervalo o dan una interpretación bayesiana del mismo.

Discusión

Aunque el estudio es exploratorio, debido al tamaño de muestra reducido, sus resultados sugieren una amplia variedad de posibles dificultades conceptuales, procedimentales e interpretativas de los estudiantes en relación al intervalo de confianza, que extienden los citados en investigaciones previas. Como ejemplo citamos la confusión de la distribución muestral, sus grados de libertad o la desviación típica del estadístico; olvido de la fórmula que da uno de los límites del intervalo en el caso de la estimación de la varianza o error en la determinación del valor crítico. Asimismo se repiten errores denunciados por otros autores, como falta de relación entre precisión y tamaño de muestra o entre ella, variación en los datos y coeficiente de confianza o la interpretación bayesiana de un intervalo de confianza, o pensar que se refiere al estadístico y no al parámetro.

Algunos de estos errores pueden estar relacionadas con el hecho de que muchos estudiantes visualizan los Intervalos de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial (Cumming y Fidler, 2005). Pero también se observa una debilidad conceptual en la comprensión de la manera como se relacionan los distintos conceptos asociados con un intervalo de confianza (Behar, 2001). Tampoco parece ser comprendida la utilidad de los intervalos de

confianza para tomar decisiones sobre hipótesis, posiblemente por no considerar los valores del intervalo como un conjunto de valores plausibles del parámetro.

La importancia del intervalo de confianza y las recientes sugerencias de la necesidad de su uso en la investigación recomiendan tener en cuenta estos resultados y tratar de mejorar su enseñanza. Una posibilidad sería diseñar unidades didácticas basadas en la simulación que incrementa la relevancia del aprendizaje, pues los estudiantes exploran y experimentan el significado de los intervalos y el efecto del tamaño de muestra, varianza y coeficiente (Terán, 2006). Sería asimismo necesario relacionar este tema con el estudio de las distribuciones de probabilidad y hacer ver a los estudiantes la importancia de la distribución muestral en el cálculo de los intervalos de confianza.

Agradecimientos

El trabajo ha sido realizado dentro del Proyecto SEJ2007-60110. MEC/FEDER, España.

Bibliografía

- C. Batanero (2000): "Controversies around significance tests". *Journal of Mathematics Thinking and Learning*, 2(1-2), 75-98.
- C. Batanero y C. Díaz (2006). "Análisis del proceso de construcción de un cuestionario sobre probabilidad condicional". *Educação e Pesquisa*, 8(2), 197-223.
- R. Behar (2001): "Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística". Tesis Doctoral. Universidad Politécnica de Cataluña.
- R. Cruise, R. Dudley y J. Thayer (1984): *A resource guide for introductory statistics*. New York: Kendall/Hunt Publishing company
- G. Cumming y F. Fidler (2005): "Interval estimates for statistical communication: problems and possible solutions". *IASE/ISI Satellite*.
- G. Cumming y S. Finch (2005): "Inference by eye: Confidence intervals, and how to read pictures of data". *American Psychologist*, 60,170-180
- G. Cumming, J. Williams y F. Fidler (2004): "Replication, and researchers' understanding of confidence intervals and standard error bars". *Understanding Statistics*, 3, 299-311.
- H. Davies (1998): "What are confidence intervals?" On line: <http://www.evidence-based-medicine.co.uk>.
- C. Díaz, C. e I. de la Fuente, I. (2007). "Validación de un cuestionario de razonamiento probabilístico condicional". *REMA*, 12(1), 1-15.
- L. L. Harlow, S. A. Mulaik y J. H. Steiger (1997): "What if there were no significance tests?" Mahwah, NJ: Lawrence Erlbaum.
- I. Miller, J. Freund y R. Johnson (1997): *Probabilidad y estadística para ingenieros*. Quinta edición. México: Prentice Hall

- D. Montgomery y G. Runger (2004): Probabilidad y estadística aplicadas a la ingeniería. Segunda edición. México: Limusa.
- D.E. Morrison y R.E. Henkel (1970): "The significance test controversy". Chicago: Aldine.
- E. Olivo (2006): "Análisis de la presentación de intervalos de confianza en textos de estadística para ingenieros". Departamento de Didáctica de la Matemática. Universidad de Granada.
- S.J. Osterlind (1989): Constructing test ítems. Boston: Kluwer.
- T. Terán (2006): "Elements of meaning and its role in the interaction with a computacional program". En A. Rossman y B. Chance (Eds.), Proceedings of the Seventh International Conference on Teaching Statistics. Salvador (Bahia): IASE.
- A. Vallecillos (1994): "Estudio teórico - experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios". Tesis doctoral Departamento de Didáctica de la Matemática. Universidad de Granada.

Eusebio Olivo, Ingeniero Mecánico Administrador, ITESM, Campus Monterrey. Maestro en Ciencias con especialidad en Estadística. Universidad de Texas, Campus El Paso, USA. Diploma de Estudios Avanzados. Programa de Didáctica de la Matemática, Universidad de Granada. Profesor del Departamento de Matemáticas en el ITESM, Campus Monterrey. Ha publicado manuales dirigidos a estudiantes de probabilidad y estadística de las carreras de ingeniería y algunos trabajos de investigación en educación. Ha participado en el Comité Organizador de la XII Reunión de Investigación y Desarrollo Tecnológico del ITESM, Comité de Evaluación Premio Rómulo Garza y el Comité para estudiar el sistema de clasificación de profesores del ITESM, Campus Monterrey.

Carmen Batanero, Licenciada en Matemáticas en la Universidad Complutense de Madrid y Doctora en Matemáticas (Estadística) por la Universidad de Granada, España. Profesora de Didáctica de las Matemáticas en la Universidad de Granada. Ha publicado libros dirigidos al profesorado y artículos en diferentes revistas de educación matemática. Es miembro del Comité Ejecutivo de ICMI (International Commission on Mathematical Instruction) y fue Presidenta de IASE (International Association for Statistical Education). Ha coordinado la organización del VII Congreso Internacional sobre Enseñanza de la Estadística, ICOTS-7. Fue editora de la revista Statistics Education Research Journal.