

# Conocimiento del Intervalo de Confianza por Futuros Profesores de Bachillerato

## Confidence Interval Knowledge by Future Teachers of Bachillerato

Carmen Batanero<sup>\*a</sup>, Maria del Mar López-Martín<sup>b</sup>

<sup>a</sup>Universidad de Granada, España.

<sup>b</sup>Universidad de Almería, España .

\*E-mail: [batanero@ugr.es](mailto:batanero@ugr.es)

---

### Resumen

El objetivo de este trabajo fue evaluar el conocimiento común y especializado del contenido de los futuros profesores españoles de bachillerato sobre el intervalo de confianza. Para lograr tal fin se propuso a un grupo de 73 futuros profesores un problema abierto sobre intervalo de confianza, similar a los propuestos en cursos anteriores a los estudiantes de Bachillerato en las pruebas de acceso a la universidad. Además, se les preguntó cómo explicarían a un estudiante el significado del intervalo de confianza y sus propiedades. Utilizando una metodología cualitativa de análisis de contenido, se estudia el planteamiento, cálculo, interpretación y definición del intervalo y las propiedades que recuerdan. Aunque la mayor parte de los participantes en el estudio plantea y calcula correctamente el intervalo, sólo el 28,8% de los futuros profesores lo interpreta correctamente. Se encuentran interpretaciones deterministas o propias a la metodología bayesiana, similares a las descritas en investigaciones con estudiantes. Otros o no proporcionan una interpretación. Pocos participantes fueron capaces de definir correctamente el intervalo de confianza cuando se les pide explicar a un estudiante su significado y recuerdan pocas propiedades del mismo. Se concluye la necesidad de reforzar el conocimiento del contenido sobre intervalo de confianza en estos futuros profesores.

**Palabras-clave:** Intervalo de Confianza. Conocimiento del Contenido. Futuros Profesores. Bachillerato. Evaluación.

### Abstract

*The aim of this research was to assess prospective high school teachers' common and specialized content knowledge of confidence intervals. To achieve this aim, we proposed to 73 prospective teachers an open problem similar to those proposed to high school students in the previous years at the entrance to university tests. Besides, participants were asked to describe how they would explain the concept and its properties to a student. Using the qualitative methodology of content analysis we study the setting, computation, interpretation and definition of confidence interval and its properties. Although most participants correctly set and computed the confidence interval, only 28.8% of them made a correct interpretation of the same. Deterministic and Bayesian interpretations, similar to those described in previous research with students arose or there was no interpretation. Few participants defined correctly the interval when asked how they would explain the topic to a student; additionally, only a few properties of the interval were remembered. We conclude the need to reinforce the content knowledge of confidence intervals in prospective teachers.*

**Keywords:** Confidence Interval. Content Knowledge. Prospective Teachers. High School. Assessment.

---

### 1 Introducción

La inferencia estadística tiene un papel destacado en las investigaciones en ciencias humanas y sociales y, en consecuencia, las orientaciones curriculares españolas para el Bachillerato de Humanidades y Ciencias Sociales (MECD, 2015) proponen este contenido (muestreo, teorema central del límite e intervalos de confianza). Igualmente, las pruebas de acceso a la universidad (PAU) de Matemáticas Aplicadas a las Ciencias Sociales propusieron en los últimos 15 años un problema (entre cuatro) sobre inferencia estadística, incluyendo generalmente uno sobre intervalo de confianza (López-Martín, Batanero, Díaz-Batanero y Gea, 2016). Una razón de la atención que actualmente recibe este procedimiento es que las recomendaciones metodológicas de asociaciones profesionales indican que en las publicaciones de estas áreas los contrastes de hipótesis han con los intervalos de confianza

(Wilkinson y TFSI, 1999; Yaremko, Harari, Harrison y Lynn, 2013). Sin embargo, este tema no es sencillo, como muestra la investigación previa (Castro Sotos, Vanhoof, Van den Nororgate y Onghena, 2007; Harradine, Batanero y Rossman, 2011).

La mejora de la comprensión y la aplicación de la inferencia implica la preparación adecuada de los profesores responsables de su enseñanza. Sin embargo, aunque algunas investigaciones han estado centradas en la comprensión de conceptos de inferencia por futuros profesores, el intervalo de confianza apenas ha sido tenido en cuenta. Consecuentemente, el objetivo de este trabajo es realizar un estudio exploratorio de evaluación del conocimiento matemático común y especializado sobre el intervalo de confianza estudiantes que se preparan como profesores de matemáticas de educación secundaria y bachillerato en España. A continuación,

presentamos los fundamentos del trabajo, el método y los resultados de la investigación.

## 2 Marco Teórico

Los intervalos de confianza fueron propuestos por Neyman (1934) para complementar la estimación puntual  $\hat{\theta}$  de un parámetro  $\theta$  con una medida de la exactitud de la estimación. El autor sugirió construir dicha medida con base en la desviación típica  $S_{\hat{\theta}}$  de la distribución muestral del estimador  $\hat{\theta}$ , en la forma:

$$\hat{\theta} - kS_{\hat{\theta}}; \hat{\theta} + kS_{\hat{\theta}}$$

donde  $k$  depende del coeficiente de confianza elegido por el investigador y cuyos extremos proporcionan los límites de la estimación del parámetro  $\theta$ . Olivo (2008) describe el intervalo de confianza de dos formas diferentes:

- Como rango de valores calculado con los datos de una muestra en el cual podría encontrarse el verdadero valor del parámetro, que depende del coeficiente de confianza.
- Como procedimiento de construcción de dicho rango de valores, a partir de un estadístico calculado en los datos de la muestra. En este procedimiento, el intervalo de confianza se particulariza dependiendo del parámetro a estimar (media, proporción, varianza, etc.) y las condiciones (tipo de distribución, qué se conoce de la misma, etc.).

Es importante recordar que, aunque el valor del parámetro es constante y desconocido, los extremos del intervalo varían aleatoriamente de muestra a muestra, puesto que los extremos del intervalo se calculan a partir de una estimación muestral, que es una variable aleatoria (Behar, 2001; Olivo, 2008). Por ejemplo, al considerar la media  $\mu$  de una población, si con cada posible muestra de tamaño  $n$  que pueda ser extraídas de la misma construimos el correspondiente intervalo, el  $(1 - \alpha) \times 100\%$  de estos intervalos contendrá el verdadero valor  $\mu$ . Es decir, el coeficiente de confianza nos da el porcentaje de intervalos calculados de la misma población con igual tamaño de muestra que cubrirían el valor del parámetro, pero no sabemos si el intervalo calculado en una muestra particular lo cubre o no (Cumming, Williams y Fidler, 2004; Fidler y Cumming, 2005).

Por otro lado, el nivel de confianza establece en alguna medida la amplitud del correspondiente intervalo de confianza. Aumentando el nivel de confianza (mayor certeza) se obtiene mayor amplitud (menor precisión); la amplitud también aumenta con la variabilidad de la población. Así pues, si se pretende aumentar la precisión disminuyendo la amplitud a la vez, es necesario ampliar el tamaño de la muestra.

En este trabajo analizamos el conocimiento de la construcción del intervalo de confianza, así como su interpretación de un grupo de futuros profesores españoles de Bachillerato. Nos basamos en el modelo del *Conocimiento Matemático para la Enseñanza* (MKT, Mathematical Knowledge for Teaching, Ball, Thames y Phelps, 2005), quienes diferencian entre conocimiento del contenido y conocimiento didáctico del contenido. Más concretamente,

nos centramos en dos componentes del conocimiento del contenido,

*Conocimiento Común del Contenido (CCK)* y *Conocimiento Especializado del Contenido (SCK)*.

El *Conocimiento Común del Contenido* se refiere al conocimiento puesto en juego por una persona para resolver problemas matemáticos, es decir, aquel conocimiento del tema que los profesores enseñarán a sus estudiantes. Puesto que la construcción e interpretación del intervalo de confianza se incluyen en el currículo español de este nivel educativo, estos conocimientos formarán parte del conocimiento común del contenido sobre el intervalo de confianza en la terminología de Ball, Thames y Phelps (2008).

El *Conocimiento Especializado del Contenido (SCK)*, indica un conocimiento especial del profesor que lo habilita para planificar y desarrollar secuencias de enseñanza del contenido. Dicho conocimiento está íntimamente relacionado con la práctica, según Ball, Thames y Phelps (2008). En nuestro estudio evaluamos este conocimiento pidiendo a los futuros profesores especificar la definición que darían y propiedades que enseñarían a sus futuros estudiantes, que serían parte del conocimiento especializado del contenido, según los anteriores autores.

## 3 Investigación Previa

El trabajo con intervalos de confianza implica la comprensión de una gran diversidad de conceptos, entre ellos los de población y muestra, parámetro y estadístico, distribución muestral del estadístico, y coeficiente de confianza, lo que explica la variedad de errores descritos en la investigación didáctica. Dichos trabajos no han estado centrados en estudiantes universitarios y profesionales que usan este procedimiento en sus investigaciones.

Behar (2001) evalúa la comprensión del intervalo de confianza de 41 profesionales dedicados a la estadística y estudiantes del último año de la licenciatura de estadística y 297 estudiantes universitarios de ingeniería en Colombia. Entre sus resultados, encuentra falta de comprensión de los distintos factores que influyen en la amplitud del intervalo, es decir, la variabilidad en la población, el tamaño de la muestra y el nivel de confianza. Un error muy extendido en su muestra fue pensar que los valores incluidos en el intervalo de confianza, se refieren a la variable aleatoria o al estadístico muestral utilizado como estimador y no al parámetro en estudio (también encontrado por Fidler, 2006). Otra interpretación incorrecta del coeficiente de confianza fue suponer que da la probabilidad de que el parámetro se encuentre en el intervalo; por ejemplo, dado un coeficiente de confianza del 95% y, habiéndose obtenido un intervalo  $(a \leq \mu \leq b)$  para la media  $\mu$  de una población, se piensa que, si se toman 100 muestras diferentes de la población, en 95 de ellas la media estará comprendida en el mismo intervalo, visualizándose los extremos del intervalo como valores fijos

(Morey, Hoekstra, Rouder, Lee y Wagenmakers, 2016). En esta línea, Garfield, delMas y Chance (2004) indican que algunos estudiantes universitarios de ciencias sociales interpretan el coeficiente de confianza del 95% como probabilidad de que el intervalo de confianza incluya la media muestral. Olivo, Batanero y Díaz (2008), que investigaron la comprensión de 252 estudiantes de ingeniería, sugieren que no se asocia el nivel de confianza con la frecuencia relativa a largo plazo con la que los intervalos generados aleatoriamente contengan al verdadero parámetro de la población. Los autores indican que hay una interpretación bayesiana del intervalo, y que el alumno confunde estadístico con parámetro.

Cumming y Fidler (2005) mostraron que los intervalos de confianza eran considerados como estadísticos descriptivos, ignorando su naturaleza inferencial. Además, Garfield, delMas y Chance (2004) encuentran estudiantes que piensan que un intervalo ancho significa siempre menos confianza, cuando esto depende de la variabilidad de la población y el tamaño de la muestra o bien que un intervalo de confianza estrecho es siempre preferible, sin tener en cuenta el coeficiente de confianza. Igualmente, Cumming y Fidler (2005) constaron, por un lado, la presencia de dificultades sobre las relaciones que se establecen entre los distintos conceptos que intervienen en los intervalos de confianza y, por otro lado, destacaron el hecho de considerar los extremos del intervalo de confianza como estadísticos descriptivos, ignorando su naturaleza inferencial.

Cumming, William y Fidler (2004) y Fidler y Cumming (2005) estudian sistemáticamente la creencia en la replicación de los intervalos de confianza por parte de los investigadores que han publicado artículos en los que usan dichos intervalos. Muchos de ellos conceden una alta probabilidad a que el parámetro caiga de nuevo en el intervalo de confianza original, cuando esta probabilidad es, en realidad, pequeña. También observan la creencia errónea de que los intervalos de confianza al 95% de medias independientes son significativamente diferentes cuando los dos intervalos se tocan justo extremo con extremo.

En otro trabajo, Belia, Fidler y Cumming (2005) pidieron a algunos investigadores que juzgaran cuando eran significativamente diferentes las medias de dos grupos independientes, a partir de la interpretación de los intervalos de confianza presentados gráficamente, observando que sólo unos cuantos sujetos dieron la respuesta correcta. Un grupo interpretó los resultados como si se refiriesen a una sola población, y otro resolvió el problema como si se tratase de dos medias relacionadas. Además, otros sujetos creían que dos intervalos de confianza, a un nivel de confianza del 95%, para la misma media poblacional, no se pueden solapar. También pensaban que cuando dos intervalos para medias no se solapan, ello implica una diferencia significativa en el nivel de 0,05 entre las medias, y que el solapamiento de los dos intervalos implica que no existe una diferencia significativa. Ninguna de estas creencias es cierta, pues depende de si las

muestras son independientes o relacionadas, del tamaño de la muestra y la magnitud del solapamiento.

Por otro lado, Hoekstra, Morey, Rouder y Wagenmakers (2014) encontraron que tanto estudiantes como investigadores en Psicología muestran dificultades similares en la interpretación de los intervalos de confianza, sugiriendo que tanto la educación recibida como la experiencia realizando investigación, no impiden las interpretaciones incorrectas en este tema.

Exceptuando el trabajo de Behar (2001), estas investigaciones han sido llevadas a cabo con estudiantes y no con futuros profesores. Nuestro trabajo pretende completar la investigación previa analizando el conocimiento del tema por parte de los futuros profesores españoles de Bachillerato, que estarán encargados de enseñarlo a los estudiantes de este nivel educativo. En lo que sigue se describe la metodología y resultados del estudio realizado.

#### 4 Metodología

La muestra participante en la investigación estuvo formada por 73 estudiantes que cursaban el Máster Universitario en Profesorado de Educación Secundaria Obligatoria y Bachillerato que es obligatorio para aquellos que quieren concursar a una plaza de profesor en España. De ellos 42 habían realizado el grado de Matemáticas o Estadística y el resto, grados de ciencias o ingenierías. Todos ellos habían cursado durante sus estudios de grado al menos una asignatura de estadística, incluyendo el componente de inferencia estadística. Además, el 57% de los participantes reconocieron tener experiencia de enseñanza de matemáticas en los niveles de educación secundaria y bachillerato.

La evaluación se implementó a lo largo de dos sesiones, de aproximadamente dos horas de duración como parte de un taller sobre inferencia estadística y dentro de una asignatura de innovación docente en matemáticas. La primera sesión estaba enfocada en desarrollar y evaluar el conocimiento matemático sobre contraste de hipótesis e intervalo de confianza resolviendo dos problemas de características similares a los problemas incluidos en las pruebas de acceso a la universidad en Andalucía. Los futuros profesores dispusieron de una hora para resolver cada uno de los problemas, pudiendo consultar sus apuntes sobre estadística e información en Internet y, la segunda parte de la sesión, se dedicó a corregir y discutir con los estudiantes las soluciones, una vez recogidos los cuestionarios. Con objeto de desarrollar su conocimiento didáctico, en la segunda sesión se les pidió resumir las dificultades que, a su juicio, podrían tener los estudiantes con los problemas propuestos. La tarea analizada en este trabajo se presenta a continuación.

**Tarea.** La esperanza media de vida en un estudio desarrollado por las Naciones Unidas es de 69,2 años y su desviación típica 10. Se ha tomado una muestra aleatoria de 16 países europeos y se ha obtenido una esperanza de vida media de 78 años. Suponiendo que la variable esperanza de

vida sigue la distribución normal con desviación típica 10, se pide:

- Determine un intervalo de confianza del 95% para la esperanza media de vida en los países europeos.
- ¿Cómo explicaría a un estudiante el significado de un intervalo de confianza y alguna de sus propiedades?

Mediante la respuesta a la primera parte de esta tarea se pretendía evaluar en qué medida los futuros profesores tenían un conocimiento común del contenido matemático suficiente del tema, de acuerdo a Ball, Thames y Phelps (2008). En concreto, se evaluarían los puntos siguientes: a) Planteamiento inicial, es decir, si se elige un intervalo que corresponda al enunciado del problema y se selecciona la distribución muestral y el estadístico adecuado; b) Cálculo correcto del intervalo: operaciones realizadas para obtener los extremos y su corrección; c) Interpretación del intervalo de confianza obtenido.

La segunda pregunta va dirigida a evaluar su conocimiento especializado del contenido según Ball, Thames y Phelps (2008). En concreto, sobre esta pregunta se analizarán dos aspectos: d) Definición que aportarían a sus estudiantes sobre el término en estudio; e) Propiedades que recuerdan y enseñarían a los estudiantes sobre los intervalos de confianza.

## 5 Análisis y Resultados

Cada participante en el estudio completó por escrito individualmente las tareas y posteriormente se procedió a realizar un análisis de contenido de las soluciones detalladas. Este método permite el estudio sistemático y la realización de inferencias sobre el contenido de documentos escritos, y es propio de la investigación cualitativa (Neuendorf, 2016). En nuestro caso, se realizó separadamente el análisis de cada uno de los apartados que se consideran en el trabajo. En base a las investigaciones previas relacionadas con el tema de estudio, se definieron a priori diversas categorías de análisis para cada una de estas variables y se crearon categorías nuevas cuando fue necesario. Mediante un proceso cíclico e inductivo, propio de la metodología cualitativa se refinaron las categorías y su codificación. Finalmente se resumieron los datos, como mostramos a continuación.

### 5.1 Planteamiento del intervalo de confianza

En primer lugar, analizamos la forma en que el estudiante plantea la construcción del intervalo de confianza. Dado que el valor de la media muestral, denotada por  $\bar{X}$ , varía de una muestra a otra, tratamos con una variable aleatoria (Schuyten, 1991), que se distribuirá según un modelo probabilístico determinado. Puesto que el enunciado del problema indica que la distribución de partida en la población es normal  $N(\mu, \sigma)$ , y en base a la propiedad reproductiva respecto a sus parámetros, la media muestral sigue una distribución normal, pues se define como suma de variables normales, salvo una constante ( $1/n$ ). Aplicando las propiedades de linealidad de la media y de la varianza (en este caso se puede aplicar, porque

las variables sumadas son independientes), se deduce que la media de la muestra sigue una distribución normal, es decir,  $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ . La construcción del intervalo de confianza para la media poblacional  $\mu$  se centra en encontrar los límites de un intervalo de valores  $(e_1, e_2)$ , con un nivel de confianza  $(1-\alpha)\%$  de tal forma que se verifique que  $P(e_1 < \mu < e_2) = 1-\alpha$ . En relación al *planteamiento*, se han encontrado las siguientes categorías de respuestas:

*C1. Respuesta correcta.* Se especifica que el objetivo es hallar los valores de  $e_1$  y  $e_2$  que cumplen que  $P(e_1 < \mu < e_2) = 0,95$  o se explicita de algún otro modo correcto la expresión del intervalo de confianza. Entre las respuestas más completas destacamos la dada por el participante BDM, quien, además de describir las características de las variables en estudio, realiza de forma precisa y correcta el desarrollo teórico que da como resultado la expresión del intervalo de confianza para la media de una población normal con varianza conocida.

Para hacer el intervalo de confianza de una variable que sigue una distribución normal cuya media es 69,2 y desviación típica 10, esto es,  $X \sim N(69,2; 10)$ , se utiliza el pivote  $T = (\bar{X} - \mu_0) / (\sigma/\sqrt{n}) \sim N(0, 1)$ . Como  $P(\lambda_1 < T < \lambda_2) = P(\lambda_1 < (\bar{X} - \mu_0) / (\sigma/\sqrt{n}) < \lambda_2) = P(X - \lambda_2 \sigma/\sqrt{n} < \mu_0 < X - \lambda_1 \sigma/\sqrt{n}) = P(\lambda_1' < \mu_0 < \lambda_2')$ . Siendo  $P(\lambda_1' < \mu_0 < \lambda_2') = \Phi(\lambda_2') - \Phi(\lambda_1')$  que, como tiene que ser un intervalo de confianza del 95%, entonces  $\Phi(\lambda_2') - \Phi(\lambda_1') = 1 - \alpha = 0,95$  ( $\lambda_1', \lambda_2' = (X - \lambda_2 \sigma/\sqrt{n}, X - \lambda_1 \sigma/\sqrt{n})$ ). Por tanto,  $\lambda_2 = -\lambda_1$  y  $\lambda_1 = z_{(\alpha/2)}$  con  $P(Z > z_{(\alpha/2)}) = \alpha/2$ . Entonces el intervalo de confianza del 95% será  $(\bar{X} - \lambda_2 \sigma/\sqrt{n}, \bar{X} - \lambda_1 \sigma/\sqrt{n})$  (Respuesta de BDM).

*C2. Explica lo que es el intervalo, pero no lo plantea.* Se describe verbalmente la forma en que se construye un intervalo de confianza, sin llegar a definir la expresión asociada a éste. Obsérvese que la respuesta dada por SS no deja entrever si el participante sabe plantearlo correctamente. Además, la expresión “varios pares de números” no deja claro si se refiere a un único intervalo. Otro error en la respuesta es suponer que  $\alpha$  es una probabilidad de error en la estimación, interpretación incorrecta similar a otras dadas sobre el nivel de significación de un contraste de hipótesis, por ejemplo, en Vallecillos (1994).

El intervalo de confianza es un par o varios pares de números entre los cuales se estima que estará cierto valor desconocido con una determinada probabilidad de acierto. Estos números determinan un intervalo, que se calcula a partir de datos de una muestra, y el valor desconocido es un parámetro poblacional. La probabilidad de éxito en la estimación se representa con  $1 - \alpha$  y se denomina “nivel de confianza”. En estas circunstancias,  $\alpha$  es el llamado “error aleatorio” o “nivel de significación”, esto es, una medida de las posibilidades de fallar en la estimación mediante tal intervalo (Respuesta de SS).

*C3. Plantea el intervalo usando la media poblacional.* Se describe la expresión del intervalo utilizando la media poblacional en lugar de la media muestral; es decir se utiliza el valor 69,2 (valor de la media poblacional) con el

fin de obtener los extremos del intervalo. Este tipo de error supone la confusión entre estadístico y parámetro, que ya ha sido detectado en investigaciones previas sobre inferencia estadística (Harradine et al., 2011; Vallecillos, 1994). Un ejemplo es la respuesta de MTF que construye un intervalo centrado en la media supuesta de la población. Por otro lado, el futuro profesor realiza una interpretación determinista al intervalo, pues no aparece ninguna expresión de confianza asociada al mismo. El estudiante no comprende la finalidad de la estimación de los parámetros, cuando indica que se puede comprobar si el parámetro pertenece al intervalo.

Una vez que tenemos  $(69,2-1,96 \times 10/\sqrt{16}; 69,2+1,96 \times 10/\sqrt{16})$  la fórmula sustituir y comprobar si el valor que queremos estimar pertenece al intervalo o no (Respuesta de MTF).

*C4. Planteamiento incorrecto por error en la fórmula del intervalo.* Cuando no se describe correctamente la expresión del intervalo, pues se olvida o confunde alguno de los términos que intervienen en la fórmula de cálculo. Obsérvese por ejemplo que PFS describe correctamente la distribución de la media muestral (denotada por  $X_m$ ) y señala que el intervalo de confianza es un intervalo donde se estima que estará la media poblacional. Sin embargo, en la expresión del intervalo introduce, en ambos extremos, la media de la muestra precedida de un signo negativo en lugar de un signo positivo

El intervalo de confianza para la media se determina sabiendo que  $X_m$  sigue una distribución que tipificada nos da la variable  $N(69,2; 2,5)$  que es una  $N(0;1)$ . El intervalo de confianza para  $\mu$  al 95% es  $[(-1,96 \times \sigma)/\sqrt{n}-X_m, (1,96 \times \sigma)/\sqrt{n}-X_m]$  (Respuesta de PFS).

La Tabla 1 recoge los resultados obtenidos sobre el planteamiento del intervalo, donde se observa que un alto porcentaje de los participantes plantean de una forma correcta la construcción del intervalo. Por consiguiente, se deduce un buen conocimiento matemático común de dicho planteamiento en esta muestra. Se han obtenido cinco repuestas imprecisas ya que los participantes describen verbalmente el proceso, pero no describen la expresión de cálculo del mismo. Respecto a los errores detectados, se han hallado, aunque con poca frecuencia, los citados por Harradine et al. (2011) y Vallecillos (1994) en relación a la confusión entre estadísticos muestrales y parámetros poblacionales y en el uso incorrecto de la expresión asociada al intervalo.

**Tabla 1 - Planteamiento del intervalo**

Código	Planteamiento	Frecuencia	Porcentaje
C1	Correcto	64	87,7
C2	Explica lo que es el intervalo, pero no lo plantea	5	6,9
C3	Plantea el intervalo usando la media poblacional	2	2,7
C4	Planteamiento incorrecto por error en la fórmula	2	2,7
Total		73	100,0

## 5.2 Cálculo del intervalo

Una vez planteado el intervalo, se deben identificar los datos del enunciado, para posteriormente obtener los valores numéricos que determinan sus extremos. Como ya se ha indicado, el cálculo del intervalo de confianza supone la identificación de los parámetros de una distribución muestral normal. Además, es necesario determinar los puntos críticos de tal manera que el área de cada cola, por encima y debajo de los mismos, sea igual a  $\alpha/2$ , es decir,  $P(-k < Z < k) = 1 - \alpha$  donde  $Z \sim N(0;1)$ . Realizando una lectura inversa de la tabla de la distribución estandarizada, vemos que los valores  $-k$  y  $k$  que cumplen dicha condición son los cuantiles  $Z_{(\alpha/2)}$  y  $Z_{((1-\alpha)/2)}$  respectivamente. En nuestro caso, dado que el nivel de confianza considerado es del 95%, se tiene que  $-k = Z_{0,025} = -1,96$  y  $k = Z_{0,975} = 1,96$ . Sustituyendo dichos valores en  $P(-k < Z < k) = 1 - \alpha$  y dado que  $Z = (X - \mu) / (\sigma/\sqrt{n}) \sim N(0;1)$ , se tiene, despejando el parámetro poblacional, la expresión  $(X - 1,96 \sigma/\sqrt{n}, X + 1,96 \sigma/\sqrt{n})$  que corresponde al intervalo de confianza para la media poblacional con varianza conocida. A partir de la información muestral ( $X = 78, n = 16$ ) y el valor de la desviación típica poblacional ( $\sigma = 5$ ) se deduce que el intervalo de confianza solicitado es  $(73,1; 82,9)$ . En vista del procedimiento seguido, se han definido las siguientes categorías para clasificar el cálculo realizado por los futuros profesores:

*C1. Cálculo correcto.* Se consideran aquellas respuestas en las que los participantes obtienen los valores de los extremos del intervalo pedido, realizando todos los pasos que se han descrito y sin cometer errores. Un ejemplo es la respuesta dada por AMC donde se observa la correcta obtención de los dos cuantiles en la distribución normal tipificada, elección adecuada de los valores de cada uno de los elementos que determinan la expresión del intervalo (media y varianza de la distribución muestral), así como valores críticos y cálculo correcto de los extremos.

La población sigue una distribución normal  $N(69,2;10)$ , y la muestra que hemos cogido es de 16 países. Solo nos queda saber qué es  $\alpha$ . Nos pide calcular el intervalo de confianza del 95%,  $\alpha$  es el nivel de significación, y se obtiene como 1 menos el nivel de confianza, por lo que  $\alpha = 0,05$ . El valor  $Z_{(\alpha/2)}$  lo obtenemos de la tabla de la  $N(0;1)$ , y vale 1,96. (Solo hay que buscar el valor de la variable para la probabilidad 0,0025, que es la misma que 0,975). Ya solo tenemos que sustituir en el intervalo los valores de la media,  $Z$ , la desviación típica y el número de países que hemos cogido para la muestra,  $n$ . Obteniendo el siguiente intervalo de confianza para la media:  $I = (78 - 1,96 \times 2,5; 78 + 1,96 \times 2,5) = (73,1; 82,9)$  (Respuesta de AMC).

*C2. Aproxima el valor de la distribución normal, obteniendo un intervalo aproximado.* Algunos futuros profesores realizan la mayor parte de los pasos descritos en la categoría *C1* correctamente, pero en lugar de utilizar los valores exactos de los cuantiles de la distribución normal, los aproximan, utilizando la propiedad de los intervalos de la distribución normal que contienen un cierto porcentaje de

casos, posiblemente por dificultad en la lectura de las tablas de la distribución normal. En el caso siguiente, el participante aproxima por 2 el valor de la distribución normal que comprende el 95% de los datos y, por ello, obtiene un valor aproximado para los extremos del intervalo.

En una distribución normal, el 95% de los datos están dentro de una desviación típica  $\sigma=2$ , por lo que  $P(-2 \leq z \leq 2)=0,95$ . Sabemos que  $z=(78-\mu)/2,5$ , por lo que, si sustituimos y despejamos las inecuaciones, obtenemos que el intervalo de confianza del 95% es el que va desde los valores de esperanza de vida de 73 hasta 83 años (Respuesta de MMR).

**C3. Pequeños errores de cálculo en las operaciones.** Se clasifican en esta categoría aquellas respuestas que realizan un planteamiento adecuado de la construcción del intervalo, pero en cambio comete errores en el cálculo de uno de los extremos. Un ejemplo lo tenemos en la respuesta de LAP que comete un error en el cálculo del extremo superior del intervalo, obteniendo 85,84 en lugar de 82,9:

Usamos la expresión del intervalo de confianza para una variable que sigue una ley normal:  $IC_{\mu} = (\bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n}, \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n})$ . Sabiendo que el valor crítico para un 95% de confianza es 1,96 sin más que sustituir y operar obtenemos  $IC_{\mu}=(73,1;85,84)$  que: (Respuesta de LAP).

**C4. Explica cómo se calcularía el intervalo de confianza, pero no realiza los cálculos.** Se incluye un planteamiento adecuado de la construcción del intervalo, pero sin llegar a obtener los valores que determinan sus extremos. Ejemplo de ello es la respuesta dada por JBL donde se evidencia que el participante tiene conocimientos sobre la construcción del intervalo, pero no llega a determinar los valores de los extremos del intervalo. Posiblemente ha tenido dificultad en identificar los datos del enunciado o no recuerda la expresión de la media o varianza de la distribución muestral.

En este caso, tenemos que buscar un número  $c$  tal que:  $P(X^- - c \leq \mu \leq X^+ + c) = 1 - \alpha$ . Para ello, tenemos que saber que si  $X \sim N(\mu, \sigma) \rightarrow Z = (X - \mu) / \sigma \sim N(0, 1)$ . Usando esto, tenemos que  $y$ , llamando  $P(-c / (\sigma/\sqrt{n}) \leq \mu \leq c / (\sigma/\sqrt{n})) = 1 - \alpha$  siendo  $c / (\sigma/\sqrt{n}) = z_{\alpha/2}$   $P(Z \geq z_{\alpha/2}) = \alpha/2$ . De esta manera, nuestro intervalo de confianza para la media quedaría como:  $I = (X^- - z_{\alpha/2} \cdot \sigma/\sqrt{n}, X^+ + z_{\alpha/2} \cdot \sigma/\sqrt{n})$  (Respuesta de JBL).

**C5. Utiliza la media poblacional.** Este tipo de error es consecuencia directa de utilizar la media poblacional en lugar de la media muestral en la expresión del intervalo de confianza. La confusión entre estadístico y parámetro ya ha sido detectada en investigaciones previas sobre inferencia estadística (Harradine et al., 2011; Vallecillos, 1994). A continuación, se muestra la respuesta dada por AML.

$IC = (\mu \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}); (69,2 \pm 1,96 \cdot 10/4); (67,5; 73,8)$ ; (Respuesta de AML).

En la Tabla 2 se presenta la frecuencia y porcentaje de futuros profesores cuyas respuestas han sido clasificadas en cada una de las categorías anteriores sobre la construcción del intervalo de confianza. Observamos que más del 60% de los participantes dan como resultado el intervalo de extremos (73,1;82,9) y un quinto explica cómo se obtiene el intervalo de

confianza, pero sin llegar a realizar cálculo alguno.

**Tabla 2 - Cálculos desarrollados**

Código	Respuesta	Frecuencia	Porcentaje
C1	Correcto	45	61,6
C2	Aproxima el valor de la distribución normal	5	6,9
C3	Pequeños errores de cálculo en las operaciones	4	5,5
C4	Explica cómo, pero no realiza los cálculos	15	20,5
C5	Utiliza la media poblacional	4	5,5
Total		73	100

En todo caso, son menos los que son capaces de calcular correctamente el intervalo de confianza que aquellos que lo plantean correctamente. Un 38,4% de participantes en el estudio comete algún tipo de error en el cálculo, lo que nos parece muy alto, teniendo en cuenta que se trata de futuros profesores que han de explicar el tema. Observamos que se ha duplicado, respecto a la información recogida en la Tabla 1, el número de futuros profesores que utilizan la media poblacional en lugar de la media muestral en el cálculo del intervalo, error descrito por Harradine et al. (2011) y Vallecillos (1994). Este hecho muestra un cierto problema, ya que una confusión de este tipo, en el ámbito de la inferencia estadística, supone falta de comprensión de uno de los principales objetivos para los que se desarrolla el entramado teórico de dicha materia. El conjunto de errores supone un 17,9% y se han dividido casi en la misma proporción entre los participantes. Alrededor de un 20% no son capaces de completar el cálculo del intervalo.

### 5.3 Interpretación del intervalo

Una vez determinados los extremos del intervalo, se espera que los estudiantes den una interpretación del significado del intervalo construido con un nivel de confianza del 95% similar a la siguiente: si se toman muestras aleatorias repetidas del mismo tamaño de la misma población y se calcula el intervalo de confianza para cada una de estas muestras, aproximadamente, el 95% de estos intervalos incluirá el parámetro poblacional. Después de seleccionar una muestra en particular y calcular el intervalo para ese parámetro, este intervalo puede o no cubrir el valor del parámetro; es una cuestión de incertidumbre. En base a esta interpretación se espera que los participantes consideren los extremos del intervalo como variables (Behar, 2001; Olivo, 2008). Una vez analizadas las respuestas de los futuros profesores, se han definido las categorías siguientes:

**C1. Interpretación correcta.** Se clasifican en esta categoría la interpretación dada anteriormente y las variaciones correctas de la misma. Asimismo, se han considerado las respuestas que señalan el porcentaje de intervalos calculados a partir de muestras de igual tamaño de la población que contiene al parámetro, es decir, indican que el parámetro está entre los límites calculados con una confianza del 95% (Behar, 2001). Como ejemplo de respuesta correcta se muestra la dada por CBF.

A la hora de explicarle al alumno que es un intervalo de confianza recalcaría que la probabilidad es referida a que el intervalo cubra al parámetro y no de que el parámetro esté en el intervalo (Respuesta de CBF).

*C2. Interpretación determinista.* Es la respuesta de aquellos participantes que indican que el parámetro se encuentra entre los límites del intervalo, dando una interpretación determinista del resultado. Pensamos que los participantes visualizan el intervalo de confianza como un estadístico descriptivo, como indican Cumming y Fidler (2005) y no aprecian la aleatoriedad de los extremos del intervalo (Olivo, 2008). Un ejemplo típico es la respuesta de DAM:

Lo que nos indica que el valor de la esperanza media de vida en los países europeos está comprendido entre 73 y 83 (Respuesta de DAM).

*C3. Interpretación del nivel de confianza en términos probabilísticos.* Se consideran aquellas respuestas en las que el nivel de confianza es interpretado en términos de probabilidad, es decir, como la probabilidad de que el parámetro poblacional esté dentro del intervalo obtenido, lo que es equivalente a dar una interpretación bayesiana y no frecuencial del intervalo. Esta última interpretación incorrecta fue encontrada en el estudio de Behar (2001), Garfield, delMas y Chance (2004) y Olivo (2008) e implica que se visualizan los extremos del intervalo como constantes (véase respuesta de LUG).

Por tanto, resulta que el intervalo de confianza es  $I=(73,1;82,9)$ , por lo que se puede concluir que la probabilidad de que el intervalo que se ha calculado contenga a la media de la población es de un 95% (Respuesta de LUG).

*C4. Intervalo que contiene a la media de la muestra.* También encontramos errores a la hora de interpretar el intervalo como aquel que contiene a la media de la muestra y no a la media de la población, véase la respuesta de JGM. Es de nuevo un caso donde se confunde estadístico y parámetro. Este tipo de error fue encontrado igualmente por Behar (2001), Garfield, delMas y Chance (2004) y Olivo (2008).

Una forma de explicar al alumno sería indicar que en el 95% de las muestras de países europeos que se elijan para cualquier estudio muestral, la media muestral de la esperanza de vida estará comprendida dentro del intervalo de confianza para los países europeos (Respuesta de JGM).

*C5. No interpreta.* Se han incluido aquellos casos en los que los futuros profesores no han interpretado los resultados obtenidos.

Los resultados recogidos en la Tabla 3 muestran las dificultades que tienen los futuros profesores a la hora de interpretar la información obtenida del intervalo de confianza. Observamos que aproximadamente dos quintas partes no han dado interpretación alguna y un tercio han tenido errores en la misma (categorías C2, C3 y C4). Estos resultados ponen de manifiesto las limitaciones que presentan, en relación a la comprensión del intervalo de confianza, las personas que serán las encargadas de preparar a los estudiantes de secundaria y bachillerato. Señalamos que, de los 21 participantes

clasificados en la primera categoría, siete han realizado una interpretación correcta y, salvo dos que definen correctamente el nivel de significación, el resto han sido algo imprecisos puesto que basan su interpretación en que el parámetro está entre los límites calculados con una confianza del 95%, sin interpretar en qué consiste esa confianza.

**Tabla 3.** Interpretación de los resultados obtenidos

Código	Respuesta	Frecuencia	Porcentaje
C1	Interpretación correcta	21	28,7
C2	Interpretación determinista	8	11,0
C3	Interpreta $1-\alpha$ en términos probabilísticos	13	17,8
C4	Intervalo que contiene a la media de la muestra	1	1,4
C5	No interpreta	30	41,1
<b>Total</b>		73	100

#### 5.4 Definición que enseñaría

En este apartado analizamos las respuestas sobre la definición de intervalo de confianza cuando los participantes son preguntados sobre cómo lo explicarían a los estudiantes. Se espera que los participantes definan un intervalo de confianza como un rango para un parámetro poblacional determinado por dos números a un nivel de confianza determinado. Las categorías definidas son las que se muestran a continuación.

*C1. Correcta.* Se han considerado correctas las respuestas en las que se menciona los límites del intervalo, el nivel de confianza y se señala que el intervalo construido es para un parámetro poblacional. Por ejemplo, se indica que un intervalo de confianza es un conjunto de valores plausibles del parámetro que pudiera haber generado el dato observado. Véase la respuesta de BMZ.

Un intervalo de confianza para un parámetro dado (por ejemplo, esperanza media de vida) son los límites entre los que se encontrará el verdadero valor del parámetro con un determinado valor de confianza (Respuesta de BMZ).

*C2. Define en términos de probabilidad.* Una definición errónea sobre el intervalo de confianza es aquella en la que el nivel de confianza es identificado como la probabilidad de que el intervalo contiene el valor poblacional, que, como hemos indicado anteriormente corresponde a una interpretación bayesiana del mismo. Cumming y Fidler (2005), señalan que esta definición es deficiente ya que para un intervalo de confianza dado no se puede hablar de probabilidades, puesto que, la probabilidad asociada a un intervalo de confianza, independientemente si incluye o no el verdadero valor, es 1 o 0. A modo de ejemplo se muestra la respuesta de MSM.

El intervalo de confianza de un  $x\%$  es un intervalo numérico en el cual sabemos que el parámetro poblacional desconocido debe encontrarse con una probabilidad de  $x/100$  (Respuesta de MSM).

*C3. No define.* Se codifican aquellos casos en los que los futuros profesores no definen el concepto de intervalo de confianza.

**Tabla 4** - Frecuencia y porcentaje de respuestas sobre definición

Código	Respuesta	Frecuencia	Porcentaje
C1	Correcta	11	15,1
C2	En términos de probabilidad	15	20,5
C3	No define	47	64,4
Total		73	100

De la información presentada en la Tabla 4 se observa que son pocos los futuros profesores que han sabido dar una definición para explicar a sus estudiantes qué es el intervalo de confianza, pese a que un porcentaje elevado plantea y calcula el intervalo de confianza (véase Tabla 1 y Tabla 2). Por tanto, se observa su falta de conocimiento especializado sobre el intervalo y, de nuevo, la dificultad de interpretación y/o comprensión por lo que se intuye que su aprendizaje es simplemente procedimental. Además, entre aquellos participantes que han dado una definición sobre el intervalo de confianza, más de la mitad ha considerado que el valor del parámetro está entre los extremos del intervalo con una probabilidad determinada, dando una interpretación bayesiana.

### 5.5 Propiedades que enseñaría

Se preguntó también a los futuros profesores qué propiedades más destacables de los intervalos de confianza enseñarían a sus estudiantes. Una vez analizadas las respuestas en relación a esta variable se han descrito las siguientes categorías.

*C1. Relación entre nivel de confianza y amplitud del intervalo.* Cuando el futuro profesor identifica que un aumento (disminución) del nivel de confianza supone un aumento (disminución) en la amplitud del intervalo. Véase por ejemplo la respuesta dada por ILB.

Si variamos el nivel de confianza,  $1-\alpha$ , el error de estimación y la longitud del intervalo varían en consecuencia. En nuestro caso, si el nivel de significación aumenta, el error aumenta y en consecuencia la longitud aumenta (Respuesta de ILB).

*C2. Relación entre tamaño muestral y amplitud del intervalo.* Partiendo de la expresión del error de estimación,  $Z_{(\alpha/2)} \times \sigma / \sqrt{n}$ , se puede establecer, además de la relación descrita en la categoría anterior, que un aumento del tamaño de la muestra (manteniendo constante el nivel de confianza) implica que el error de estimación disminuirá (Behar, 2001). Si nuestro objetivo es conseguir una mayor precisión en las estimaciones realizadas por medio de los intervalos de confianza, lo ideal es aumentar el tamaño muestral, ya que habitualmente no es recomendable utilizar un nivel de confianza inferior al 90%. Esta propiedad queda reflejada en la respuesta dada por JLS donde el participante relaciona un tamaño muestral elevado con una mejor determinación del valor del parámetro poblacional bajo estudio.

La longitud del intervalo viene determinada por la muestra que hemos tomado. Si el valor de  $n$  aumentara tendríamos más información y por tanto podremos ajustar más el intervalo, es decir, disminuye. [...]. Intuitivamente, si queremos estar más

seguros de dar con una franja de valores que incluya al valor real de la media, entonces cuanto más grande sea el intervalo más probabilidad hay de que englobemos en este el verdadero valor de la media (Respuesta de JLS).

*C3. No se sabe si el intervalo cubre el valor del parámetro.* Cuando se señala que la obtención de los extremos del intervalo de confianza supone el desconocimiento de si el verdadero valor del parámetro está o no en el intervalo obtenido, véase la respuesta de ACG. Esta categoría está estrechamente relacionada con la interpretación del nivel de confianza, dado que no se tiene la certeza que, de 100 intervalos construidos, ese intervalo sea uno de los 95 que contiene el verdadero valor del parámetro.

Por último, el alumno debe ser consciente de que no sabremos si el intervalo realmente contiene al valor buscado, debido a que siempre habrá probabilidad de que no lo contenga y por tanto no es exacto (Respuesta de ACG).

*C4. Sirve para realizar contrastes.* Se han considerado aquellos casos en los que se identifica una conexión entre los intervalos de confianza y los contrastes de hipótesis. Este hecho viene recogido también en las investigaciones llevadas a cabo por Cumming y Fidler (2005) donde los autores ponen de manifiesto como una ventaja de la construcción de los intervalos de confianza, la estrecha relación existente entre estos y la hipótesis nula del test de significación. Ejemplo de ello es la respuesta dada por BCM.

La propiedad fundamental es que sirve para hacer un contraste bilateral (Respuesta de BCM).

*C5. Carácter aleatorio de los extremos.* Para cualquier muestra aleatoria de tamaño  $n$ , extraída aleatoriamente de una población normal con varianza conocida, se verifica que la media muestral  $X$  es una variable aleatoria. Este hecho provoca que los extremos del intervalo varíen según lo haga la media muestral (Behar, 2001; Olivo, 2008). Este hecho pone de manifiesto el carácter aleatorio del intervalo, propiedad recogida en la respuesta de MVC.

Debido a la naturaleza aleatoria de cada muestra, es poco probable que dos muestras de una población generen intervalos de confianza idénticos. Cada muestra generará un intervalo de confianza. (Respuesta de MVC).

En la Tabla 5 se ha recopilado la frecuencia y porcentaje (sobre el número total de participantes) de cada una de las propiedades que han sido mencionadas por los futuros profesores. Queremos señalar que al igual que hay participantes que no describen ninguna propiedad fundamental del intervalo de confianza, otros de ellos indican más de una. De los 73 participantes solo 23 han nombrado al menos una de las propiedades categorizadas en este apartado, concretamente 8 han hecho referencia a una propiedad, 11 a dos propiedades y 4 a tres propiedades, por lo que, el número total asciende a 42 citas de propiedades. Observamos que la relación entre la amplitud con el nivel de significación y el tamaño de la muestra seleccionada son las más destacadas ya que aproximadamente un quinto de los participantes la mencionan.



Le sigue el carácter aleatorio que tienen los extremos del intervalo cuando se cambia de una a otra muestra. En menor medida se ha hecho referencia a la idea de si el intervalo cubre o no el valor del parámetro y a la relación existente entre los intervalos de confianza y los contrastes de hipótesis.

**Tabla 5** - Propiedades sobre I.C. señaladas por los participantes

Código	Respuesta	Frecuencia	Porcentaje
C1	Relación entre $(1 - \alpha)$ y amplitud del intervalo	14	19,2
C2	Relación entre $n$ y amplitud del intervalo	13	17,8
C3	No se sabe si el intervalo cubre el valor del parámetro	5	6,8
C4	Sirve para realizar contrastes	2	2,7
C5	Carácter aleatorio de los extremos	8	11,0
<b>Total</b>		42	57,5

Por otra parte, destacamos el hecho de que de los 73 participantes únicamente un tercio han mencionado al menos una propiedad en relación al intervalo de confianza y sólo el 5% tres propiedades. Estos resultados son altamente llamativos, pues implican un pobre conocimiento conceptual matemático del intervalo de confianza por parte de los participantes, que se une a su falta de conocimiento especializado del contenido, aunque como vimos su conocimiento procedimental fue mejor, ya que la mayoría lo plantea y calcula correctamente.

### 5.6 Síntesis de conocimientos sobre el intervalo de confianza

Para dar una medida global del conocimiento de los participantes, se analiza a continuación el número de apartados resueltos correctamente, puntuándolo con un punto por apartado. Recordemos que una correcta solución del mismo supone describir correctamente el planteamiento del problema para posteriormente realizar los cálculos necesarios con objeto de obtener los valores que determinan los extremos del intervalo. Además, es necesario que el participante de una interpretación correcta de los resultados del mismo. Se añade otro punto por la definición correcta, obteniéndose un máximo posible de cuatro puntos (véase Tabla 6). Se observa que únicamente un 29% han tenido al menos tres apartados correctos y seis han respondido adecuadamente a los cuatro apartados evaluados.

**Tabla 6** - Distribución del número de apartados correctos sobre el intervalo

Apartados Correctos	Frecuencia	Porcentaje	Fr. Acumulada	% Acumulado
0	9	12,3	9	12,3
1	14	19,2	23	31,5
2	29	39,7	52	71,2
3	15	20,6	67	91,8
4	6	8,2	73	100,0

Los participantes con uno o dos apartados correctos han

realizado un planteamiento adecuado del ítem, completándose, en su mayoría, con la obtención de los valores numéricos de los extremos del intervalo solicitado. Se han detectado mayores dificultades, por parte de los futuros profesores, a la hora de definir o interpretar el intervalo de confianza. Ello es debido a que muchos se han limitado a aplicar un procedimiento puramente mecánico con objeto de obtener los extremos del intervalo, sin mostrar una comprensión profunda. Por otro lado, es necesario enfatizar el hecho de que nueve participantes no han desarrollado correctamente ninguno de los apartados evaluados a lo largo del presente trabajo.

Así mismo, se observa un salto significativo entre 1 y 2 apartados correctos pues se produce un incremento de aproximadamente un 40%, aunque dicho crecimiento se ve más suavizado a medida que aumentan el número de apartados correctos. Estas conclusiones revelan las dificultades que presentan los futuros profesores cuando abordan problemas de construcción de intervalos de confianza, que completan las obtenidas sobre el número de propiedades fundamentales que recordaban.

## 6 Discusión y Conclusiones

Los resultados obtenidos muestran que la mayoría de los participantes realizan un planteamiento adecuado y un cálculo correcto de los extremos del intervalo. Sin embargo, fueron pocos los que llegaron a interpretar este resultado en el contexto del problema, por lo que, aunque muestran cierto conocimiento común del contenido, este conocimiento es superficial, ya que, aunque realizan el cálculo, no saben interpretar su significado. Este último resultado es preocupante, pues se espera que el futuro profesor desarrolle en sus estudiantes la capacidad de interpretar estudios estadísticos. Así, por ejemplo, uno de los estándares de aprendizaje evaluables para el cuarto curso de la Educación Secundaria Obligatoria indica (MECD, 2015, p. 398 y p. 407): “Interpreta un estudio estadístico a partir de situaciones concretas cercanas al alumno”. Por tanto, es necesario que un futuro profesor sea capaz de contextualizar la información resultante de un procedimiento estadístico para ayudar al alumno a adquirir esta capacidad de interpretación.

Respecto al conocimiento especializado del contenido, la mayor parte de las propiedades que enseñarían estarían centradas en destacar la relación que existe entre la amplitud del intervalo con el tamaño muestral y el nivel de significación, además de mencionar el carácter aleatorio de los extremos del intervalo. Sin embargo, solamente un tercio ha reconocido alguna propiedad de los intervalos. Aun así, los resultados obtenidos en el presente trabajo mejoran a los encontrados por Olivo, Batanero y Díaz (2008) con ingenieros, ya que en este estudio solo el 12% indica que el intervalo disminuye con el tamaño de la muestra y el 10% que aumenta con el coeficiente de confianza.

Algunos de los errores detectados se han identificado en investigaciones previas con estudiantes Behar (2001), Cumming y Fidler (2005), Garfield, delMas y Chance (2004),

Harradine et al. (2011), Olivo, Batanero y Díaz (2008) y Vallecillos (1994), entre otros. Dado que el tema forma parte de la educación preuniversitaria y de las pruebas de acceso a la universidad en y de muchos grados universitarios en España es necesario asegurar que los futuros profesores posean un adecuado conocimiento sobre el tema ya que ellos son los encargados de transmitirlo en el aula. Además, dentro del ámbito científico, la asociación profesional American Psychological Association (APA, 2010) recomienda hacer uso de la información que aporta los intervalos de confianza con objeto de tener el “mejor informe” (p. 34). En particular, una correcta y clara transmisión, profundizando tanto en la interpretación como en la definición del mismo, puede garantizar un aprendizaje más significativo por parte de los estudiantes.

Los resultados obtenidos nos advierten de la necesidad de preparar al futuro profesorado de Educación Secundaria y Bachillerato en el tema de inferencia estadística, y más concretamente en el intervalo de confianza. Por tanto, se debe trabajar en la dirección correcta para que los futuros profesores adquieran un mejor conocimiento sobre el tema con el fin asegurar un proceso de enseñanza y aprendizaje de calidad. Una posibilidad sería incluir en los cursos de formación de profesores experiencias como la descrita por Dolor y Noll (2015) con aproximaciones informales a la inferencia. Es importante también reforzar su conocimiento de las distribuciones muestrales (Liu y Thompson, 2005).

Por supuesto, un conocimiento matemático, incluso especializado únicamente, es insuficiente para el éxito de la enseñanza y aprendizaje de un tema y sería necesario continuar la investigación planteada con el análisis del conocimiento didáctico sobre el contraste de hipótesis de estos futuros profesores. Pero dicho conocimiento didáctico no puede construirse sin un conocimiento matemático suficiente; de ello se deriva el interés de la investigación que hemos iniciado. Esperamos que esta problemática interese a formadores de profesores y a otros investigadores en educación estadística y conjuntamente podamos, en el futuro, asegurar una buena formación matemática y didáctica de los futuros profesores de Educación Secundaria y Bachillerato en el campo de la inferencia estadística.

### Agradecimientos

Proyecto PID2019-105601GB-I00 (AEI) y Grupo FQM126 (Junta de Andalucía).

### Referencias

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Ball, D., Thames, M. H. y Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407. doi: <https://doi.org/10.1177%2F0022487108324554>

Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística*. Tesis Doctoral. Universidad Politécnica de Cataluña.

Belia, S., Fidler, F. y Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, 4, 389-396. doi: <https://doi.org/10.1037/1082-989X.10.4.389>

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.

Cumming, G. y Fidler, F. (2005). Interval estimates for statistical communication: problems and possible solutions. Trabajo presentado en la *IASE Satellite Conference on Communication of Statistics*. Sydney: International Association for Statistical Education.

Cumming, G., Williams, J. y Fidler, F. (2004). Replication and researchers' understanding of confidence intervals and standard error bars. *Understanding statistics*, 3(4), 299-311. doi: [https://doi.org/10.1207/s15328031us0304\\_5](https://doi.org/10.1207/s15328031us0304_5)

Dolor, J. y Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal*, 14(1), 60-89.

Fidler, F. (2006). Should psychology abandon p-values and teach CIs instead? Evidence-based reforms in statistics education. En C. Reading (Ed.), *Proceedings of the Seventh International Conference on Teaching Statistics*. International Association for Statistical Education.

Fidler, F. y Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the 55th Session of the International Statistical Institute* (pp. 1-5). Voorburg: International Statistical Institute. Online: [iasweb.org/documents/papers/isi55/Fidler-Cumming.pdf](http://www.iasweb.org/documents/papers/isi55/Fidler-Cumming.pdf).

Garfield, J., delMas, R. y Chance, B. (2004). *Tools for teaching and assessing statistical inference*, [http://www.gen.umn.edu/research/stat\\_tools](http://www.gen.umn.edu/research/stat_tools).

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). Springer Netherlands.

Hoekstra, R., Morey, R. D., Rouder, J. N. y Wagenmakers, E. J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review*, 21(5), 1157-1164. doi: <http://dx.doi.org/10.3758/s13423-013-0572-3>

Liu, Y. y Thompson, P. (2005). Teachers' understandings of hypothesis testing. En S. Wilson (Ed.), *Proceedings of the 27th Annual Meeting of the International Group for the Psychology of Mathematics Education*. Vicksburg, VA. [Online: <http://pat-thompson.net/PDFversions/>]

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.

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: Autor.

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.

- Neuendorf, K. A. (2016). *The content analysis guidebook*. London: Sage.
- Neyman, J. (1934). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transaction of the Royal Society of Londres, series A, Mathematical and Physical Sciences*, 236(767), 333-380
- Olivo, E. (2008). *Significados del intervalo de confianza en la enseñanza de la ingeniería en México*. 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.
- Schuyten, G. (1991). Statistical thinking in psychology and education. En D. Vere-Jones (Ed.), *Proceeding of the Third International Conference on Teaching Statistics* (pp. 486-490). Voorburg, The Netherlands: International Statistical Institute.
- Vallecillos, A. (1994). *Estudio teórico experimental de errores y concepciones sobre el contraste de hipótesis en estudiantes universitarios*. Tesis doctoral. Universidad de Granada.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- 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.