revista PARANAENSE DE EDUCAÇÃO MATEMÁTICA

# COMPRENSIÓN DEL INTERVALO DE CONFIANZA: UN ESTUDIO COMPARADO CON ESTUDIANTES UNIVERSITARIOS Y PREUNIVERSITARIOS

DOI: https://doi.org/10.33871/22385800.2020.9.19.52-73

Antonio Francisco Roldán López de Hierro[1]
Carmen Batanero[2]
Rocío Álvarez-Arroyo[3]

**Resumen:** Los intervalos de confianza se enseñan actualmente en España, tanto en los estudios preuniversitarios del Bachillerato de Ciencias Sociales como en la mayoría de grados universitarios, debido a su relevancia en la inferencia estadística. El objetivo de este estudio es el de comparar la comprensión mostrada por distintos tipos de estudiantes de las propiedades esenciales del intervalo de confianza, incluyendo la finalidad de la estimación y el efecto que sobre el intervalo tiene el cambio de la media, que no han sido tenidas en cuenta en la investigación previa. Para ello se analizan las respuestas de 58 estudiantes de Bachillerato, 57 de primer curso de Psicología y 37 de primer curso de Ingeniería a un cuestionario que consta de seis ítems de opción múltiple. Los resultados muestran la existencia de errores ya descritos en la investigación previa, relacionados con la interpretación de la definición del intervalo o de los factores que inciden en su amplitud. Además, se identifican nuevos errores asociados con la relación entre la media muestral y el intervalo, así como con la finalidad de la estimación por intervalos. Se informa de los errores que aparecen tanto con frecuencia similar en los tres grupos como de los que tienen más incidencia en uno de ellos con la finalidad de mejorar la enseñanza del tema.
**Palabras clave:** Comprensión de propiedades. Estudiantes de Bachillerato. Psicología. Ingeniería.

# UNDERSTANDING THE CONFIDENCE INTERVAL: A COMPARED STUDY WITH UNIVERSITY AND PRE-UNIVERSITY STUDENTS

**Abstract:** Due to its relevance in statistical inference, confidence intervals are currently taught in Spain in most university degrees and also in pre-university studies of Social Sciences Baccalaureate. The main aim of this study is comparing the understanding shown by different types of students of key properties of confidence intervals, including some properties that have not been taken into account in previous research such as the purpose of estimation and the effect of the mean on the interval. Having this objective in mind, the answers of 58 high school students, 57 students of the first course of psychology and 37 of the first course of engineering to a questionnaire consisting of six multiple-choice items have been analyzed. The results show the existence of errors already described in the previous research, related to the interpretation of the definition of confidence intervals or related to factors affecting their amplitude. Furthermore, new errors associated both with the relationship between the sample mean and the interval, as well as with the purpose of interval estimation are identified. We inform of errors that appear with similar frequency in the three groups and those with more incidence in one of them, with the aim of improving the teaching of the topic.
**Keywords:** Understanding of properties. High School Students. Psychology. Engineering.

[1] Profesor Titular de Universidad del Departamento de Estadística e Investigación Operativa de la Universidad de Granada (España). E-mail: aroldan@ugr.es - ORCID: https://orcid.org/0000-0002-6956-4328.
[2] Profesora Colaboradora Extraordinaria del Departamento de Didáctica de la Matemática de la Universidad de Granada (España). E-mail: batanero@ugr.es - ORCID: https://orcid.org/0000-0002-4189-7139.
[3] Profesora Ayudante Doctora del Departamento de Didáctica de la Matemática de la Universidad de Granada (España). E-mail: rocioaarroyo@ugr.es - ORCID: https://orcid.org/0000-0002-3201-8542.

## Introducción

El uso de intervalos de confianza en la investigación científica ha crecido en los últimos años debido a las recomendaciones realizadas para complementar con este procedimiento los contrastes de hipótesis (CUMMING; FIDLER; KALINOWSKI; LAI, 2012; THOMPSON, 2007; YAREMKO; HARARI: HARRISON; LYNN, 2013). Los anteriores autores también sugieren que este tema es sencillo para los estudiantes e investigadores, pero este supuesto se ve contradicho por los resultados de la investigación didáctica sobre el tema (BATANERO; DÍAZ-BATANERO; LOPEZ-MARTÍN; ROLDÁN-LÓPEZ DE HIERRO, en prensa; CASTRO; VANHOOF; NOORTGATEM; ONGHENA, 2007; OLIVO; BATANERO; DÍAZ, 2008). Estas investigaciones se han concentrado principalmente en la interpretación del intervalo de confianza por parte de investigadores o de estudiantes universitarios, con escasos estudios sobre la comprensión del intervalo de confianza por parte de los estudiantes españoles.

La importancia de que los estudiantes comprendan la filosofía del intervalo de confianza y sus propiedades más relevantes, además de su construcción, ha sido reconocida tanto en el último curso de Bachillerato de Ciencias Sociales (estudiantes de 17 años que entrarán en la universidad el curso siguiente), como en los cursos universitarios.

**Tabla 1:** Contenidos sobre intervalos de confianza estudiados por los estudiantes de la muestra.

| Ingeniería | Psicología | Bachillerato de Ciencias Sociales |
|---|---|---|
| Conceptos generales. Introducción al muestreo. Distribuciones en el muestreo en poblaciones normales. Estimación puntual. Propiedades de los estimadores. Estimación por intervalos de confianza. | Conceptos básicos. Distribuciones muestrales. Muestreo. Métodos de muestreo. Objetivo y conceptos básicos de la estimación. Estimación puntual. Métodos de construcción de estimadores. Propiedades de los estimadores. Estimación por intervalos. | Estimación por intervalos de confianza; relación entre confianza, error y tamaño muestral; intervalo de confianza para la media poblacional de una distribución normal con desviación típica conocida; intervalo de confianza para la media poblacional de una distribución de modelo desconocido y para la proporción en el caso de muestras grandes. |

Fuente: Elaboración propia (2020)

Así, en el currículo español para el último curso de Bachillerato de Ciencias Sociales (MECD, 2015) se incluyen los contenidos mostrados en la última columna de la Tabla 1. Estos alumnos tienen que realizar al finalizar este curso las pruebas de acceso a la universidad que incluyen habitualmente un problema de intervalos de confianza (LÓPEZ-MARTÍN; BATANERO; GEA; ARTEAGA, 2016). Los contenidos en los cursos universitarios de estadística son muy similares, en particular, en las carreras de Psicología e Ingeniería en la Universidad de Granada (ver la primera y segunda columna de la Tabla 1).

El objetivo de este trabajo es comparar la comprensión de las principales propiedades del intervalo de confianza entre tres tipos diferentes de estudiantes: dos clases de estudiantes universitarios (unos del grado en Ingeniería Informática y otros del grado de Psicología) y un grupo de estudiantes preuniversitarios.

**Fundamentos**

La estimación por intervalos ha recibido diferentes soluciones matemáticas a lo largo de su historia, como se describe en Olivo (2007) y Rivadulla (1991). Dicho intervalo puede ser construido en la actualidad con diferentes metodologías que no sólo se diferencian en los procedimientos matemáticos, sino en sus fundamentos filosóficos, aunque todas se basan en la variabilidad muestral al estimar un parámetro poblacional $\theta$ (MOREY et al., 2016).

En la metodología frecuencial, que es la habitualmente enseñada en España, dicho parámetro (por ejemplo, la media o la proporción de un cierto carácter en una población) se supone que es un valor constante, pero desconocido, que determina la distribución de la variable en la población. Para estimarlo se selecciona una muestra aleatoria de la misma a partir de la cual se calcula un estimador del parámetro $\hat{\theta}$ que puede variar de una muestra a otra. La precisión de la estimación, siguiendo a Neyman (1937), se expresa construyendo un intervalo de valores que incluye al estimador y toma la forma $[\,\hat{\theta} - k_1 S_{\hat{\theta}}\,, \hat{\theta} + k_2 S_{\hat{\theta}}\,]$, siendo $S_{\hat{\theta}}$ la desviación típica de dicho estimador.

Los valores $k_1$ y $k_2$ se calculan en función del coeficiente de confianza $1-\alpha$, siendo $0 < \alpha < 1$, escogido por el investigador, usualmente 0,95 o 0,99 (MAYO, 1981). Si la distribución muestral del estadístico es simétrica (por ejemplo, normal) estos valores coinciden.

Según Mayo (1981), el valor $1-\alpha$ (o *confianza en la estimación*) se refiere al porcentaje de intervalos construidos a partir de diferentes muestras independientes del mismo tamaño de la población que cubrirán el parámetro. Según este autor, no debe interpretarse esta confianza como la probabilidad de que el verdadero valor del parámetro $\theta$ esté situado dentro del intervalo construido. Morey et al. (2016) denominan *falacia fundamental de la confianza* al hecho de interpretar dicha confianza como la probabilidad de que el parámetro esté contenido en el intervalo, lo cual sería una interpretación bayesiana como veremos a continuación.

En inferencia bayesiana, el parámetro se considera aleatorio y se caracteriza por una distribución a priori $p(\theta)$ (GELMAN; SHALIZI, 2012) que refleja el grado de credibilidad que la persona que quiere construir el intervalo da a los valores posibles del parámetro, antes de recoger los datos. La probabilidad se interpreta de forma subjetiva; dependiendo de su conocimiento distintos individuos podrían proponer diferentes distribuciones a priori.

Dicha metodología bayesiana se desarrolla fundamentalmente a partir del teorema de Bayes, y su finalidad es calcular la distribución a posteriori del parámetro a partir del conocimiento de su distribución a priori y de los datos obtenidos experimentalmente $y$ (BOLSTAD, 2013):

$$p(\theta|y) = p(y|\theta)p(\theta)/p(y)$$

En esta expresión $p(y|\theta)$ es la verosimilitud de los datos observados para un valor dado del parámetro. Además, $p(y) = \int p(y|\theta)p(\theta)dy$, integrándose a través de todo el rango admisible de valores de $y$; por tanto, es un valor constante, de modo que podemos simplificar la expresión para obtener:

$$p(\theta|y) = Kp(y|\theta)p(\theta)$$

De esta forma, se puede decir que la distribución a posteriori del parámetro se obtiene multiplicando su distribución a priori por la función de verosimilitud. El teorema de Bayes se deduce de su artículo publicado en Bayes (1970-1763); por tanto, estrictamente puede considerarse que la inferencia bayesiana se inicia desde esta fecha y es anterior a la frecuencial (RIVADULLA, 1991). Una vez obtenida la distribución a posteriori, se utiliza dicha distribución para obtener intervalos de credibilidad para el parámetro. Elegido un nivel de credibilidad $1-\alpha$, $C$ es un intervalo de credibilidad al $100(1-\alpha)\%$ si $p(\theta|y) \geq 1-\ \alpha$ en dicho intervalo. Para el intervalo de credibilidad, $1-\alpha$ es la probabilidad de que el parámetro esté entre los valores dados, que se consideran fijos y no aleatorios (CEPEDA-CUERVO; AGUILAR; CERVANTES; CORRALES; DÍAZ; RODRÍGUEZ, 2008).

Otra técnica posible son los métodos de remuestreo, cuya filosofía está basada en que toda la información requerida sobre el parámetro está en los datos. En el método de *bootstrap*, conceptualizado por Efron (1979), la idea principal es extraer muchas muestras con reemplazamiento de una muestra original y deducir a partir de ellas una distribución muestral aproximada, utilizando el estadístico calculado en cada una de las diferentes remuestras. La distribución resultante es similar a la distribución muestral, pero conceptualmente diferente; se denomina distribución *bootstrap* y se utiliza para realizar inferencias sobre la población, incluido el cálculo de intervalos de confianza.

En todos los cursos participantes en el estudio se enseñó el intervalo de confianza siguiendo la metodología frecuencial, en forma simplificada y adaptando el contexto de los problemas presentados a la especialidad del estudiante. En los cursos universitarios se utilizaron paquetes estadísticos como ayuda al cálculo, mientras que los estudiantes de Bachillerato contaban con la calculadora para esta finalidad. Además, en este último curso se propusieron y resolvieron problemas planteados en las pruebas de acceso a la universidad, pues los alumnos se preparaban para dichas pruebas.

## Antecedentes

Una parte de investigaciones sobre la comprensión de los intervalos de confianza se han realizado tomando como sujetos a investigadores en psicología o educación, bien a partir de entrevistas o encuestas a los mismos, o mediante el análisis de su interpretación de intervalos de confianza en artículos publicados en revistas de investigación de impacto.

Uno de los estudios más citados fue el de Cumming, William y Fidler (2004), quienes entrevistaron a 263 autores de artículos publicados en revistas de impacto en los que habían utilizado intervalos de confianza para ver si entendían el concepto de replicación. Tras proporcionarles un intervalo de confianza para la media de una población, construido a un coeficiente de confianza del 95%, se les preguntó sobre los valores que esperaban de la media muestral si se repitiese el mismo experimento en un gran número de ocasiones. Más de tres cuartas partes de los investigadores participantes en el estudio respondieron que la media volvería a caer en el intervalo proporcionado en el 95% de las veces, lo que es una interpretación propia del método de remuestreo y no del método frecuencial de construcción de intervalos.

Otro estudio realizado, esta vez con investigadores y estudiantes, es el de Behar (2001), quien proporcionó un cuestionario a 47 investigadores y 297 estudiantes de ingeniería en Colombia. El autor indicó que el 29% de los investigadores y el 50% de estudiantes interpretaban el coeficiente de confianza como porcentaje de valores de la variable que cae en el intervalo. Además, el 40% de los investigadores y 50% de los estudiantes suponían que el nivel de confianza era la probabilidad de que el intervalo contuviese al verdadero valor del parámetro poblacional, lo que sería una interpretación bayesiana y no frecuencial del intervalo. Como se ha indicado, la interpretación frecuencial correcta (MAYO, 1981) es que algunos intervalos no contienen al verdadero valor del parámetro, pero al repetir muchas veces el experimento, el porcentaje de intervalos que contienen al parámetro poblacional se acerca al coeficiente de confianza. Otros errores fueron creer que al aumentar el coeficiente de confianza dejando invariantes los demás factores se obtienen intervalos más estrechos e incluso suponer que existe proporcionalidad directa entre la anchura del intervalo y el tamaño de la muestra.

Fidler y Cumming (2005) identificaron otros errores de comprensión del intervalo de confianza en una muestra de 180 estudiantes de psicología: el 38% de los participantes en el estudio interpretaban el intervalo como un conjunto de valores razonables para la media muestral, lo que es una interpretación del intervalo en remuestreo, y un 19% confundía el rango de la variable aleatoria con el intervalo de confianza. Por otro lado, un 20% pensaba que la anchura del intervalo aumentaría con el tamaño muestral, un 29% consideró que no variaría y un 36% no fue capaz de describir el tipo de relación entre la anchura y el tamaño de la muestra.

Olivo y Batanero (2007) estudiaron la comprensión del intervalo de confianza en 48 estudiantes de ingeniería teniendo en cuenta los contenidos de un análisis previo del tema en libros de texto universitarios utilizados por los estudiantes. Aunque la mayor parte de los estudiantes respondieron correctamente muchas preguntas, los autores también describieron los siguientes conflictos semióticos de algunos de ellos:

- El 27% de los alumnos dio una interpretación bayesiana al intervalo, identificando el coeficiente de confianza como la probabilidad de que el intervalo contenga al parámetro poblacional, mientras otro 21% interpretó que el intervalo contenía un porcentaje de valores de la variable, en vez de comprender que se trata de una estimación del parámetro.
- La relación entre el tamaño de la muestra y la precisión del intervalo sólo fue comprendida correctamente por el 25% de los alumnos, mientras que el 18% sugirió que el coeficiente de confianza no afectaba a la anchura del intervalo.

Olivo, Batanero y Díaz (2008) replicaron el estudio anterior con 252 estudiantes de ingeniería, el 25% de los cuales confundió el estadístico con el parámetro, pensando que el intervalo se construía para predecir los valores de la media muestral, lo que en realidad sería la interpretación propia de los intervalos construidos con el método de remuestreo. Otro 33% no comprendía la relación entre ancho del intervalo y tamaño de muestra o el coeficiente de confianza, y un 32% de estudiantes que habían calculado correctamente el intervalo le daban una interpretación bayesiana.

Estos errores también aparecen en ocasiones en los estudiantes universitarios que se preparan para ser profesores de matemáticas, como muestran López-Martín et al. (2019a), en un estudio con 73 futuros profesores de matemáticas de secundaria a los que se pidió construir e interpretar un intervalo de confianza. Mientras sólo un 28,8% de los participantes dio una interpretación correcta, el 11% afirmó que el intervalo debía contener, con seguridad, el verdadero valor del parámetro (visión determinista). Un 17,8% dio la interpretación bayesiana y el 41,1% no aportó ninguna interpretación (ni correcta ni incorrecta) sobre el significado del intervalo de confianza; finalmente un estudiante pensó que el intervalo servía para estimar el estadístico muestral y no el poblacional.

## Metodología

Esta investigación se llevó a cabo con una muestra intencional de estudiantes, por lo que nuestro estudio es de tipo exploratorio. De esta forma, no pretendemos extender las conclusiones obtenidas a una muestra o a una población más amplia sino explorar la información que nos proporcionan los individuos analizados.

La muestra estuvo formada por tres tipos diferentes de estudiantes. El primer grupo incluyó a 58 estudiantes preuniversitarios del último curso de Bachillerato de Ciencias Sociales, distribuidos en dos institutos diferentes: uno situado en la ciudad de Granada, con 24 estudiantes, y otro situado en la periferia de dicha ciudad, con 34 estudiantes. El segundo grupo consistió en 35 estudiantes universitarios del grado de Ingeniería, todos ellos matriculados en la Universidad de Granada, los cuales estaban cursando una asignatura de estadística en el primer curso del grado (en concreto, 23 de ellos estaban matriculados en el grado de Ingeniería de Telecomunicaciones y el resto en Ingeniería Informática). Finalmente, también se incluyó en el estudio un grupo de 57 estudiantes universitarios del grado de Psicología que estaban estudiando la asignatura "*Análisis de datos en Psicología 2*" en la Universidad de Huelva. Es conveniente aclarar que esta investigación se llevó a cabo después de que todos ellos hubiesen terminado el estudio de este tema.

Aunque los tres grupos de estudiantes son diversos, otra de las diferencias que debemos destacar es el tiempo que habían empleado en el estudio de la inferencia estadística. Los estudiantes de Bachillerato y los estudiantes de Ingeniería se enfrentaban por primera vez a este tema. En el primer grupo, la única asignatura preuniversitaria que aborda este tema se imparte precisamente en el Bachillerato de Ciencias Sociales. Por otro lado, los estudiantes de Ingeniería provenían del Bachillerato Tecnológico, que no incluye la inferencia, por lo que acababan de estudiarla por primera vez en el grado universitario. Sin embargo, los estudiantes del grado en Psicología habían pasado en su gran mayoría por el Bachillerato en Ciencias Sociales, donde se incluyen los intervalos de confianza, por lo que éste era el segundo curso que estudiaban inferencia. Todos los cuestionarios fueron completados a lo largo del mes de mayo de 2019, cuando todos los grupos de estudiantes habían finalizado el estudio de la inferencia y se preparaban para las pruebas de evaluación finales de sus correspondientes asignaturas. El cuestionario se completó dentro de la hora de clase, corrigiéndose en otra sesión posterior como una actividad que hiciera reflexionar a los participantes sobre su grado de comprensión sobre los intervalos de confianza.

***Instrumento***

El cuestionario utilizado se construyó expresamente para esta investigación y consta de seis ítems de opción múltiple, que incluimos en el apéndice final del presente trabajo. Los ítems 1 a 3 están adaptados de otros utilizados por Olivo (2008), el 4 de Behar (2001) y los ítems 5 y 6 son de elaboración propia.

El primer ítem analiza la comprensión de la definición del intervalo de confianza. La interpretación correcta de la definición viene dada en la opción *b*, donde se resalta el carácter aleatorio de los extremos del intervalo y el porcentaje de intervalos, calculados a partir de diferentes muestras del mismo tamaño, que cubren el valor del parámetro. El distractor *a* evalúa la confusión entre media muestral y poblacional (BEHAR, 2001), y es incorrecto porque el intervalo está centrado en la media muestral $\bar{x}$; el distractor *c* considera el conflicto producido al suponer constantes los extremos del intervalo, error identificado por López-Martín et al. (2019b), y el *d* la interpretación bayesiana del intervalo (BEHAR, 2001; OLIVO; BATANERO, 2007).

El ítem 2 estudia la comprensión de la relación entre el ancho del intervalo de confianza y el tamaño de la muestra, siendo la respuesta correcta la *b*, ya que el intervalo se obtiene sumando y restando a la media muestral un valor que depende inversamente de la raíz cuadrada del tamaño muestral, por lo que un aumento de éste en dos órdenes de magnitud implica una disminución en la anchura del intervalo a la décima parte. El distractor *a* evalúa la confusión entre confianza y amplitud, al igual que el *c*, pues la precisión del intervalo sí viene influenciada por el tamaño de muestra. La opción *d* contempla el error de suponer la relación opuesta entre tamaño de muestra y amplitud del intervalo.

El ítem 3 evalúa la comprensión de la relación entre el coeficiente de confianza y el ancho del intervalo. Cuanto mayor es la confianza, mayor es la anchura del intervalo, por lo que *c* y *d* son respuestas correctas y equivalentes, mientras que *a* y *b* son incorrectas.

El ítem 4 analiza la relación entre la anchura del intervalo y el valor de la varianza de la población y tiene dos opciones correctas (*b* y *d*). Puesto que la anchura del intervalo de confianza sigue la expresión $2z_{\alpha/2} \cdot \sigma/\sqrt{n}$ , si la desviación estándar ($\sigma$) disminuye, la anchura del intervalo de confianza también disminuye, y viceversa. Los distractores *a* y *c* evalúan la falta de comprensión de esta relación.

El ítem 5 es de elaboración propia y evalúa la comprensión de que el ancho del intervalo no varía cuando se traslada la media puesto que está centrado en ella. Las respuestas correctas son tanto *c* como *d*. La primera porque, al ser aleatorios los extremos, algunos intervalos no cubren el valor del parámetro. Con ella pretendemos observar si el alumnado cree que el verdadero valor del parámetro poblacional siempre está contenido en el intervalo de confianza cuando el nivel de confianza es muy alto, error señalado previamente por Olivo (2008). Y la respuesta *d* también es correcta, puesto que la anchura del intervalo de confianza no depende de la media muestral como puede comprobarse por su expresión matemática. El distractor *b* es claramente falso ya que, para obtener los extremos, siempre se suma y se resta una cantidad positiva a la media muestral, lo que obliga indiscutiblemente a que ésta no pueda ser uno de los extremos. El distractor *a* es la negación del distractor *d* y, por tanto, incorrecto.

Finalmente, en el ítem 6 pretendíamos analizar si el alumnado conoce el problema que se desea resolver, es decir, se trata de estimar, mediante un intervalo, un valor desconocido que resulta de interés para la descripción de una población. Por ello, la respuesta correcta es la *d*, ya que no hay problema de estimación cuando se conoce el valor poblacional. Las opciones *a* y *b* proporcionaban sendos intervalos para el parámetro poblacional, diferenciándose en que uno contiene al verdadero valor de la media poblacional y el otro no. Quienes eligen el intervalo que contenía al verdadero valor poblacional (distractor *a*), dan una interpretación al intervalo en remuestreo (describiéndolo como un conjunto de valores plausibles para el parámetro poblacional). El distractor *c* pretendía observar si el alumnado piensa que todos los intervalos de confianza contienen al verdadero valor del parámetro poblacional, error señalado tanto por Behar (2001) como por Olivo (2008).

En la Tabla 2 mostramos un resumen del contenido evaluado por el cuestionario que, como se puede observar, incluye un amplio abanico de propiedades del intervalo de confianza junto con algunos de los conflictos previamente detectados en la investigación.

**Tabla 2:** Contenido evaluado en los ítems.

| **Contenido** | I1 | I2 | I3 | I4 | I5 | I6 |
|---|---|---|---|---|---|---|
| Definición del intervalo de confianza | x | | | | | |
| Porcentaje de intervalos que cubre el valor del parámetro | x | | | | | |
| Extremos aleatorios en el intervalo | x | | | | | |
| Confusión entre media muestral y poblacional | x | | | | | |
| Considerar fijos los extremos del intervalo | x | | | | | |
| Confusión entre confianza y probabilidad (interpretación bayesiana) | x | | | | | |
| Relación del ancho del intervalo y tamaño de la muestra | | x | x | | | |
| Confusión entre confianza y amplitud | | x | x | | | |
| Relación entre precisión y confianza | | x | x | | | |
| Fórmula de la desviación típica de la distribución muestral | | x | | | | |
| Relación entre amplitud del intervalo y varianza de la población | | | | x | | |
| Relación del intervalo con la media muestral | | | | | x | |
| Algunos intervalos no cubren el parámetro | | | | | x | |
| El ancho del intervalo no depende de la media muestral | | | | | x | |
| El intervalo sirve para estimar valores desconocidos del parámetro | | | | | | x |
| Pensar que el intervalo siempre contiene al parámetro | | | | | | x |

Fuente: Elaboración propia (2020)

## Resultados

A continuación, se analizan los resultados obtenidos estudiando, en primer lugar, el número y porcentaje de estudiantes de cada grupo que dan la/s respuesta/s correcta/s a los diferentes ítems (recordemos que algunos tienen más de una opción correcta), así como los que eligen los distractores que suponen un error de interpretación. Además, compararemos los diferentes grupos con los resultados de las investigaciones citadas en los antecedentes.

### *Dificultad comparada de los ítems*

En la Tabla 3 mostramos las frecuencias absolutas de aciertos en cada ítem y presentamos tanto los correspondientes porcentajes de acierto como los porcentajes medios en el total de opciones correctas posibles (que suman 9). En dicha tabla se observa, de entrada, que el mayor porcentaje medio corresponde a los estudiantes de Psicología, quienes habían estudiado el tema por segunda vez.

**Tabla 3:** Frecuencia y porcentaje de respuestas correctas por ítem.

| | Ingeniería ($n = 35$) | | Psicología ($n = 57$) | | Bachillerato ($n = 58$) | | Porcentaje promedio |
|---|---|---|---|---|---|---|---|
| | Frecuencia | % | Frecuencia | % | Frecuencia | % | |
| Ítem 1, opción *b* | 25 | 71,4 | 28 | 49,1 | 32 | 55,2 | 58,6 |
| Ítem 2, opción *b* | 9 | 25,7 | 24 | 42,1 | 15 | 25,9 | 31,2 |
| Ítem 3, opción *c* | 19 | 54,3 | 28 | 49,1 | 13 | 22,4 | 41,9 |
| Ítem 3, opción *d* | 8 | 22,9 | 14 | 24,6 | 7 | 12,1 | 19,9 |
| Ítem 4. opción *b* | 17 | 48,6 | 45 | 78,9 | 14 | 24,1 | 50,5 |
| Ítem 4, opción *d* | 20 | 57,1 | 44 | 77,2 | 17 | 29,3 | 54,5 |
| Ítem 5, opción *c* | 29 | 82,9 | 44 | 77,2 | 21 | 36,2 | 65,4 |
| Ítem 5, opción *d* | 15 | 42,9 | 20 | 35,1 | 17 | 29,3 | 35,8 |
| Ítem 6, opción *d* | 14 | 40,0 | 16 | 28,1 | 21 | 36,2 | 34,8 |
| Promedio grupal | | 49,5 | | 51,3 | | 30,1 | |

Fuente: Elaboración propia (2020)

Al comparar los estudiantes de Ingeniería con los de Bachillerato, estos últimos obtienen un menor porcentaje medio de respuestas correctas, lo que atribuimos a su menor preparación en otras ramas de matemáticas, como el cálculo o el álgebra, que posiblemente les ocasione una mayor dificultad para comprender la definición y las propiedades del intervalo de confianza.

Si realizamos una comparación global del porcentaje medio (uniendo los tres grupos) en cada respuesta correcta, observamos que la más sencilla (al menos, la que más acierto ha ocasionado entre el alumnado) ha sido la $5c$, que interpelaba al estudiante sobre la posibilidad de que la media poblacional no cayese dentro del intervalo. Sigue la opción $1b$, que corresponde a la definición correcta del intervalo (punto de vista frecuencial), y las $4d$ y $4b$ (relación de la anchura del intervalo con la desviación típica de la población). Todas estas propiedades son comprendidas por al menos el 50% de los participantes, aunque hay importantes diferencias: así, la $5c$ es comprendida sólo por el 36,2% de estudiantes de Bachillerato, mientras que una amplia proporción en los dos grupos universitarios la domina; igualmente ocurre con las propiedades $1b$, $4d$ y $4b$.

La propiedad más complicada de detectar por parte del alumnado fue la $3d$, que involucra la palabra “precisión”. En este sentido, hay que considerar que esta palabra puede tener diferentes significados según el contexto en que se utiliza. Hay contextos en los que la precisión de un número viene dada por la cantidad de cifras decimales significativas, pero en otros contextos, es más precisa una estimación que se acerca más al verdadero valor desconocido, aun teniendo menos cifras decimales. En el contexto de la inferencia estadística, el intervalo es más preciso cuando es más estrecho, pero el alumnado puede pensar que este término hace referencia al número de cifras decimales con que se trabaja. Sin duda, la detección de esta dificultas, superada únicamente por la quinta parte del alumnado, pone de manifiesto la necesidad de incidir este aspecto en el aula ordinaria. En menor medida también resultó complicado el apartado 2*b*, que hace referencia nuevamente a la amplitud del intervalo de confianza, analizando su variación cuando cambia el tamaño muestral.

***Errores identificados en los distractores de los ítems***

Observamos un alto porcentaje de estudiantes que muestran errores de comprensión y que se presentan en la Tabla 4. Algunos de ellos ya fueron detectados en investigaciones previas. El más frecuente consiste en considerar que cuando se reduce el coeficiente de confianza aumenta la anchura del intervalo (ítem 3, distractor *b*), citado por Yáñez y Behar (2009), así como en Morey et al. (2016), Olivo (2008) y Olivo y Batanero (2007). Otro error frecuente se presenta al considerar que el intervalo de confianza debe contener siempre a la media poblacional (ítem 6, distractor *c*), lo que supone dar una interpretación determinista al mismo, dificultad que encontraron López-Martín et al. (2019a) en su trabajo, aunque en menor porcentaje al nuestro (11% vs. 46,3%).

Sigue en frecuencia la interpretación bayesiana al intervalo de confianza (ítem 1, distractor *d*), es decir, estos estudiantes interpretan el coeficiente de confianza como una probabilidad a posteriori que sólo puede ser calculada en el método bayesiano. Este error, según Morey et al. (2016), constituye la "*falacia fundamental de la confianza*" y está ampliamente extendida entre los participantes de nuestro estudio. Olivo et al. (2008) indicaron que no es posible calcular la probabilidad de que el intervalo de confianza contenga al parámetro, pues la confianza no está depositada en el intervalo, sino en el método de construcción de los intervalos.

Tienen variada frecuencia los conflictos ligados a la relación de la amplitud y precisión con diversos factores, tales como pensar que la anchura del intervalo aumenta cuando crece el tamaño de la muestra (ítem 2, distractor *d*), error descrito en Fidler y Cumming (2005) entre el 20% de sus estudiantes de Psicología (porcentaje muy parecido al nuestro). Esta dificultad se observa también al ligar precisión y confianza (ítem 3, distractor *a*) o pensar que al cambiar la desviación típica no cambia el intervalo (ítem 4, distractor *a*), error encontrados también en Olivo (2008).

**Tabla 4:** Porcentajes de estudiantes que presentan diferentes errores.

| **Errores identificados en otras investigaciones** | **Ítem** | **Inge-niería** | **Psico-logía** | **Bachi-llerato** | **Pro-medio** |
|---|---|---|---|---|---|
| Reducir el coeficiente de confianza reduce la precisión | 3 (*a*) | 57,1 | 75,4 | 62,1 | 64,9 |
| El intervalo siempre contiene la media poblacional | 6 (*c*) | 34,3 | 64,9 | 39,7 | 46,3 |
| Interpretación bayesiana del intervalo | 1 (*d*) | 40,0 | 36,8 | 37,9 | 38,2 |
| Reducir la confianza aumenta el intervalo | 3 (*b*) | 28,6 | 43,9 | 12,1 | 28,2 |
| Aumentar la muestra aumenta la amplitud del intervalo | 2 (*d*) | 25,7 | 12,3 | 37,9 | 25,3 |
| Extremos del intervalo constantes | 1 (*c*) | 2,9 | 5,3 | 3,4 | 3,9 |
| Disminuir la desviación típica no cambia el intervalo | 4 (*a*) | 2,8 | 3,5 | 5,2 | 3,8 |
| **Nuevos errores identificados en el trabajo** | | | | | |
| La amplitud depende de la media muestral | 5 (*a*) | 51,4 | 49,1 | 50,0 | 50,2 |
| Aumentar la muestra aumenta siempre la confianza | 2 (*a*) | 37,1 | 43,9 | 32,8 | 37,9 |
| Disminuir la desviación típica aumenta el intervalo | 4 (*c*) | 28,6 | 15,8 | 46,6 | 30,3 |
| No comprende la finalidad de la estimación y calcula el intervalo cuando se conoce el parámetro | 6 (*a*) | 54,3 | 5,3 | 17,2 | 25,6 |
| Ligar la precisión solo a la confianza | 2 (*c*) | 20,0 | 1,8 | 17,2 | 13,0 |
| La media muestral es un extremo del intervalo | 5 (*b*) | 8,6 | 3,5 | 27,6 | 13,2 |
| El intervalo contiene un % de veces la media muestral | 1 (*a*) | 5,7 | 3,5 | 1,7 | 3,6 |

Fuente: Elaboración propia (2020).

Al mismo tiempo hay pocos estudiantes que consideren los extremos del intervalo constantes (ítem 1, distractor *c*) o que contenga la media muestral un porcentaje de veces (ítem 1, distractor *a*), en contra de lo obtenido por Behar (2001) y Olivo (2008).

Además, hemos identificado nuevos errores no descritos anteriormente, algunos de los cuales se dan en un alto porcentaje, como los siguientes:

- Considerar que la amplitud del intervalo depende de la media muestral (ítem 5, distractor *a*), lo que podría venir explicado por el hecho de que la media muestral interviene en la fórmula de cálculo.
- Suponer que el tamaño de muestra aumenta la confianza (ítem 2, distractor *a*); esta creencia puede ser debida a que en general el mayor tamaño de muestra proporciona menor variabilidad en la estimación y el alumno confunde la idea de confianza con la de variabilidad.
- Creer que al disminuir la desviación típica aumenta el intervalo (ítem 4, distractor *c*), lo que indica una falta de comprensión de la relación que guardan, visible incluso en la aritmética de la propia fórmula de cálculo.
- No relacionar las ideas equivalentes de “intervalo más estrecho” con “intervalo más

preciso" (en el ítem 3, ningún alumno eligió ambas opciones correctas y equivalentes) y suponer que la precisión está ligada solo a la confianza (ítem 2, distractor *c*).

- Con menor proporción, algunos estudiantes piensan que la media muestral se sitúa en un extremo del intervalo de confianza (ítem 5, distractor *b*).
- Suponer que algunos intervalos no contienen a la media muestral (distractor *a* del ítem 1), lo que demuestra que el alumnado no comprende que el centro del intervalo se sitúa en dicha media.

Los errores anteriores aparecen con frecuencias muy similares en los tres grupos de estudiantes, lo que indica que se transmiten en diferentes tipos de enseñanza. Es más frecuente en Bachillerato el error de pensar que el tamaño de muestra aumenta la amplitud del intervalo (también frecuente en Ingeniería), e incluso creer que la media puede ser un extremo del intervalo.

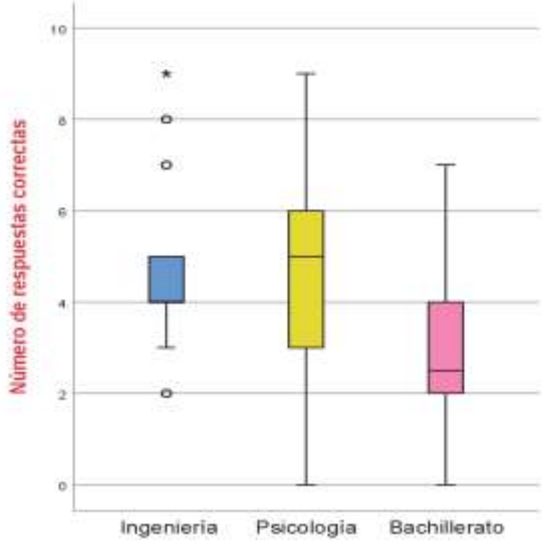
En Psicología tiene mayor aparición el pensar que reducir la confianza aumenta el intervalo (muy poco frecuente en Bachillerato), que el intervalo contiene siempre a la media poblacional, y que aumentar la muestra siempre aumenta la confianza.

En Ingeniería tiene más frecuencia el ligar la precisión a la confianza (también en Bachillerato) o pensar que la amplitud depende de la media muestral y, sobre todo, no comprender la finalidad de la estimación. En resumen, estos estudiantes de Ingeniería, aun cuando se les presupone una mayor formación matemática que la de los otros dos grupos, tienen más conocimiento de las propiedades del intervalo, pero confunden algunos conceptos y no muestran un mayor conocimiento de en qué situaciones debe aplicarse la estimación por intervalo.

***Número de respuestas correctas***

Con objeto de comparar los resultados globalmente en los tres grupos, se calificó cada una de las opciones correctas elegidas con 1 punto, obteniéndose una nueva variable "puntuación total" que puede variar teóricamente entre 0 y 9 puntos (puesto que hay seis cuestiones, pero tres de ellas tienen dos opciones correctas). y se representó gráficamente la distribución de las puntuaciones en cada grupo con gráficos de caja (Figura 1). En ella se observa un mejor desempeño de los estudiantes de Ingeniería, algunos de los cuáles alcanzan las puntuaciones 8 y 9, mientras que hay estudiantes de Bachillerato que no son capaces de identificar ninguna o sólo una respuesta correcta. Además, esta diferencia entre grupos también se evidencia al comparar la mediana, pues en los estudiantes de ingeniería es de 5 respuestas correctas (algo más de la mitad) mientras que la de los estudiantes de Bachillerato es escasamente de dos. Por otro lado, se puede observar el valor superior de los cuartiles en el grupo de ingeniería, donde el 50% de los estudiantes alcanza una puntuación de 4 a 5, mientras que el 50% de los estudiantes de Bachillerato alcanza un valor entre 2 y 4 respuestas correctas. Estos datos se completan con los estadísticos presentados en la Tabla 5.

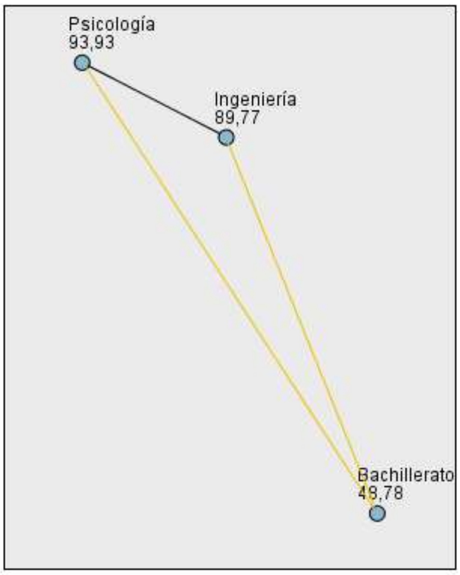
**Figura 1:** Distribución de puntuaciones en cada grupo.




Fuente: Elaboración propia (2020)

Dado que la variable que mide el número de respuestas correctas toma valores enteros entre 0 y 9, se realizó un test de normalidad de Shapiro-Wilk, el cual nos informó de que no podemos suponer que la variable sea normal. Al aplicar el test no paramétrico de Kruskal-Wallis de igualdad de medias encontramos un estadístico de contraste $W = 37{,}022$ y un *p*-valor $P < 0{,}001$, lo que indica que hay diferencias estadísticamente significativas entre las medias. En la Figura 2 podemos encontrar los correspondientes rangos promedio asociados a los diferentes grupos. Las comparaciones múltiples descritas en la Tabla 5 (véase también la Figura 2) nos permiten concluir que los estudiantes de Ingeniería y de Psicología ofrecen el mismo desempeño medio, mientras que el éxito de los estudiantes de Bachillerato en esta tarea es menor.

**Figura 2:** Rangos promedio obtenidos en la prueba de Kruskal-Wallis.




Fuente: Elaboración propia (2020)

**Tabla 5:** Comparaciones múltiples por la prueba de Kruskal-Wallis.

| **Muestra 1 – Muestra 2** | **Prueba estadística (*W*)** | **Error típico** | **Significación** |
|---|---|---|---|
| Bachillerato – Ingeniería | 40,996 | 9,167 | < 0,001 |
| Bachillerato – Psicología | 45,154 | 7,988 | < 0,001 |
| Ingeniería – Psicología | - 4,158 | 9,197 | 0,651 |

Fuente: Elaboración propia (2020)

## Conclusiones

El estudio realizado informa de la comprensión de tres grupos diferentes de estudiantes sobre la definición del intervalo de confianza y sus principales propiedades, especialmente de cómo los distintos factores influyen en el mismo. El trabajo aporta información original, pues es el primero realizado con estudiantes españoles, ya que nuestras propias investigaciones con estudiantes fueron realizadas en otros países. Los resultados indican que el tema ofrece grandes dificultades a los estudiantes de la muestra, especialmente para los de Bachillerato, a los que le faltan herramientas matemáticas para mejorar su comprensión. No obstante, es de destacar que prácticamente todos los estudiantes, a pesar de su distinta preparación matemática, cometen los mismos errores de interpretación, lo que puede indicar que estos errores no derivan de los conocimientos matemáticos previos o de sus destrezas actuales, sino de la enseñanza recibida. Si queremos conseguir profesionales estadísticamente bien formados, nos debiéramos plantear la posibilidad de incidir sobre la formación actual de los docentes en este tema, que podría mejorar los resultados en el tema de la inferencia. Ello supondría dedicar menos tiempo al cálculo práctico de los intervalos de confianza (cuestión ampliamente superada con la tecnología actual) y más al significado último tanto de los intervalos calculados como del método general de cálculo de los mismos.

Desde nuestro punto de vista, la información descrita a lo largo del presente trabajo puede ser de utilidad a los docentes, quienes en muchos casos podrían no ser conscientes de que están creando limitaciones en la compresión del intervalo de confianza entre su alumnado e incluso algunos de ellos pueden tener dificultades con el tema (LÓPEZ-MARTÍN et al., 2019a). De hecho, los libros de texto que utilizamos en clase para ayudarnos en nuestro quehacer diario también están impregnados de interpretaciones erróneas de las fórmulas matemáticas y su significado (GEA; LÓPEZ-MARTÍN; ROA, 2015), lo cual redunda en que el problema se vaya haciendo cada vez más grande y se transmita de un nivel educativo a otro.

La diversidad de errores que aparecen durante nuestro trabajo nos lleva a destacar que la interpretación del intervalo de confianza y la consideración de sus propiedades esenciales es tan importante como su cálculo efectivo en la práctica. No podemos dotar a este objeto matemático de un significado real si no es a través de la comprensión de las situaciones en las que debe plantearse, y cómo los diferentes factores los determinan de una forma u otra.

Coherentemente con lo anterior y a la vista de los resultados obtenidos en cada una de las cuestiones, proponemos las siguientes recomendaciones para la mejora de los procesos de enseñanza y aprendizaje de la estimación a través de intervalos de confianza para el alumnado universitario y preuniversitario:

1) En primer lugar, nos parece conveniente profundizar en el problema de estimación que trata de resolverse a través del intervalo de confianza, que es obtener un valor aproximado del verdadero valor del parámetro de la población desconocido. Además, hay que ser conscientes de que la respuesta que se dé siempre será aproximada.
2) En segundo lugar, nos parece importante hacer un esfuerzo por atajar la *falacia fundamental de la confianza*. Ello solo se conseguirá si transmitimos al alumnado algunas de las interpretaciones que existen del intervalo de confianza y destacamos la correcta.
3) Desde el punto de vista práctico, es esencial el cálculo del intervalo de confianza, pero debería hacerse un esfuerzo por analizar cómo los diferentes factores influyen en el resultado final. Por ejemplo, después de completar un ejercicio en el que se calcule un intervalo de confianza asociado a una muestra, podría ser

conveniente preguntar a los estudiantes sobre lo que ocurriría si la muestra fuese 100 veces mayor, si la desviación típica fuese el doble o si la media muestral se trasladase una cierta cantidad. Con un poco de esfuerzo los resultados podrían ser significativamente mejores.

4) El uso de la tecnología mejoraría indudablemente la interpretación del intervalo de confianza. Por ejemplo, en muchas ocasiones construimos un único intervalo asociado a única muestra. Si nos apoyamos en la tecnología, podríamos generar multitud de muestras aleatorias independientes que nos llevarían a intervalos diferentes, de forma que las interpretaciones erróneas que a veces proporcionan los alumnos fallarían en este contexto multi-intervalar y los haría conscientes. Unas simples hojas de cálculo o el empleo de *applets* a través de Internet puede conducirnos, de manera natural, a la observación de que únicamente un cierto porcentaje de los intervalos generados contiene al verdadero valor del parámetro (cuando lo suponemos conocido y desarrollamos experimentos de simulación).
5) Finalmente, ponemos de manifiesto la importancia de utilizar la palabra "precisión" a través de su relación con la anchura del intervalo, destacando que no tiene nada que ver con los significados usuales en contextos donde se analiza la precisión de números o la precisión de una medida.

Como resultado de los argumentos anteriores, concluimos que merece la pena dedicar más tiempo a la correcta interpretación de los intervalos de confianza y de sus propiedades esenciales, lo que redundará en un mejor conocimiento de cara a las múltiples aplicaciones de este tema tanto en estudios posteriores como en informaciones que nos llegan a través de los medios de comunicación e Internet. Todo ello debe ser objeto de investigación en futuros trabajos, donde se siga la propuesta de enseñanza que acabamos de describir y se analice si los errores descritos mejoran en los estudiantes.

**Agradecimientos:** Proyectos Proyecto PID2019-105601GB-I00 (AEI) y TIN2017-89517-P (AEI, FEDER) y Grupos de Investigación FQM-365 y FQM-126 (Junta de Andalucía).

## Referencias

BATANERO, C.; DÍAZ-BATANERO, C.; LÓPEZ-MARTÍN, M. M.; ROLDÁN LÓPEZ DE HIERRO, A. F. (en prensa). Interval estimation: methodological approaches and understanding difficulties. **Boletín de Estadística e Investigación Operativa**.

BAYES, T. An essay towards solving a problem in the doctrine of chances. En E. S. Pearson y M. G. Kendall (Eds.), **Studies in the history of statistics and probability,** v. 1, p. 131-153, 1970. Londres: Griffin (trabajo original publicado en 1763).

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

BOLSTAD, W. **Introduction to Bayesian statistics**. 2ª ed., Nueva York, Wiley, 2013.

CASTRO SOTOS, A. E.; VANHOOF, S.; VAN DEN NORORGATE, W.; ONGHENA, P. Student´s misconceptions of statistical inference: A review of the empirical evidence form research on statistical education. **Educational Research Review**, v. 2, n. 2, p. 98-113, 2007.

CEPEDA-CUERVO, E.; AGUILAR, W.; CERVANTES, V.; CORRALES, M.; DÍAZ, I.; RODRÍGUEZ, D. Intervalos de confianza e intervalos de credibilidad para una proporción. **Revista Colombiana de Estadística**, v. 31, n. 2, p. 211-228, 2008.

CUMMING, G.; FIDLER, F.; KALINOWSKI, P.; LAI, J. The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis. **Australian Journal of Psychology**, v. 64, n. 3, p. 138-146, 2012.

CUMMING, G.; WILLIAMS, J.; FIDLER, F. Replication and researchers' understanding of confidence intervals and standard error bars. **Understanding Statistics**, n. 18, n. 3, p. 299-311, 2004.

EFRON, B. Bootstrap methods: Another look at the jackknife. **The Annals of Statistics**, v. 7, p. 1-26, 1979.

FIDLER, F.; CUMMING, G. Teaching confidence intervals: Problems and potential solutions. **Proceedings of the 55th International Statistics Institute** Session CD-ROM. Sidney, Australia: International Statistical Institute, 2005.

GEA, M. M.; LÓPEZ-MARTÍN, M. M.; ROA, R. Conflictos semióticos sobre la correlación y regresión en los libros de texto de Bachillerato. **Avances de Investigación en Educación Matemática**, v. 8, p. 29-49, 2015.

GELMAN, A.; SHALIZI, C. R. Philosophy and the practice of Bayesian statistics. **British Journal of Mathematical and Statistical Psychology**, v. 66, n. 1, p. 8-38, 2012.

GODINO, J. D. Mathematical concepts, their meaning, and understanding. In: PUIG, L.; GUTIÉRREZ, Á. (Eds.), **Proceedings of XX Conference of the International Group for the Psychology of Mathematics Education**, v. 2, p. 417-425, Universidad de Valencia, 1996.

GODINO, J. D. Un enfoque ontológico y semiótico de la cognición matemática. **Recherches en Didactique des Mathématiques**, v. 22, n. 2-3, p. 237-284, 2002.

GODINO, J. D.; BATANERO, C.; FONT, V. The onto-semiotic approach to research in mathematics education. ZDM. **The International Journal on Mathematics Education**, v. 39, n. 1-2, p. 127-135, 2007.

HARRADINE, A.; BATANERO, C.; ROSSMAN, A. Students and teachers' knowledge of sampling and inference. In: BATANERO, C.; BURRILL, G.; READING, C. (Eds.). **Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education, Springer**: Netherlands, p. 235-246, 2011.

LÓPEZ-MARTÍN, M. M.; BATANERO, C.; GEA, M. M.; ARTEAGA, P. Análisis de los problemas de inferencia propuestos en las Pruebas de Acceso a la Universidad en Andalucía, **Vidya**, v. 36, n. 2, p. 409-428, 2016.

LÓPEZ-MARTÍN, M. M.; BATANERO, C.; GEA, M. M. Prospective high school teachers' interpretation of hypothesis tests and confidence intervals. In: JANKVIST, U. T.; VAN DEN HEUVEL-PANHUIZEN, M.; VELDHUIS, M. (Eds.). **Proceedings of the Eleventh Congress of the European Society for Research in Mathematics Education**. Utrecht, the Netherlands: ERME. 2019a.

LÓPEZ-MARTÍN, M. M.; BATANERO, C.; GEA, M. M. ¿Conocen los futuros profesores los errores de sus estudiantes en inferencia? **Bolema**, v. 33, n. 64, p. 672-693, 2019b.

MAYO, D. G. In defense of the Neyman-Pearson theory of statistics. **Philosophy of Science**, v. 48, p. 269-280, 1981.

MECD, Ministerio de Educación, Cultura y Deporte. **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, 2015.

MOREY, R. D.; HOEKSTRA, R.; ROUDER, J. N.; LEE, M. D.; WAGENMAKERS, E.-J. The fallacy of placing confidence in confidence intervals. **Psychonomic Bulletin & Review**, v. 23, n. 1, p. 103-123, 2016.

NEYMAN, J. Outline of a theory of statistical estimation based on the classical theory of probability. **Philosofical Transaction of the Royal Society of London, series A, Mathematical and Physical Sciences**, v. 236, n. 767, p. 33-380, 1937.

OLIVO, E. **Significados del intervalo de confianza en la enseñanza de la ingeniería en México**. Tesis Doctoral, Universidad de Granada, 2008.

OLIVO, E.; BATANERO, C. Un estudio exploratorio de dificultades de comprensión del intervalo de confianza. **Unión**, v. 12, p. 37-51, 2007.

OLIVO, E.; BATANERO, C.; DÍAZ, C. Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. **Educación Matemática**, v. 20, n. 3, p. 5-32, 2008.

RIVADULLA, A. **Probabilidad e inferencia científica**. Barcelona: Anthropos, 1991.

THOMPSON, B. Effect sizes, confidence intervals, and confidence intervals for effect sizes. **Psychology in Schools**, v. 44, p. 423-432, 2007.

YÁÑEZ, G.; BEHAR, R. Interpretaciones erradas del coeficiente de confianza en los intervalos de confianza y algunas explicaciones plausibles. In: GONZÁLEZ, M. J.; GONZÁLEZ, M. T.; MURILLO, J. (Eds.). Investigación en Educación Matemática. Comunicaciones de los grupos de investigación. XIII **Simposio de la SEIEM**, Santander, 2009.

YAREMKO, R. M.; HARARI, H.; HARRISON, R. C.; LYNN, E. **Handbook of research and quantitative methods in psychology:** For students and professionals. Hilldale, NJ: Erlbaum, 2013.

## APÉNDICE

### Cuestionario

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

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

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

a. Debemos tener más confianza de que $\mu$ caerá en nuestro segundo intervalo.
**b. Sabemos que el segundo intervalo será 10 veces más estrecho.**
c. Espero que ambos intervalos de confianza tengan la misma precisión.
d. El segundo intervalo de confianza es 10 veces más ancho que el primero.

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

a. El intervalo de confianza será menos preciso.
b. El intervalo de confianza será más ancho.
**c. El intervalo de confianza será más estrecho.**
**d. El intervalo de confianza será más preciso.**

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

a. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza no cambia.
**b. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza disminuye.**
c. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza disminuye.
**d. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza aumenta.**

**Ítem 5.** Se ha tomado una muestra de tamaño $n = 100$ para calcular el intervalo de confianza, al 95%, para la velocidad media de los coches que pasan por una carretera. Se sabe que la desviación típica poblacional de estas velocidades es de 20 km/h.

a. El ancho del intervalo de confianza depende de la media muestral $\bar{x}$.
b. La media muestral $\bar{x}$ es un extremo del intervalo de confianza.
**c. Cabe la posibilidad de que el verdadero valor de $\mu$ no esté contenido en el intervalo.**
**d. El ancho del intervalo no depende de la media muestral.**

**Ítem 6.** Se sabe que la altura media $\mu$ de los pinos de una zona forestal muy amplia es de 6.5 metros. ¿Cuál de las siguientes afirmaciones es verdadera?

a. Un posible intervalo de confianza para $\mu$ es [6.25,8.4].
b. Un posible intervalo de confianza para $\mu$ es [6.75,8.4].
c. Cualquier intervalo de confianza para $\mu$ que calculemos, asociado a una muestra aleatoria cualquiera, debe contener al valor 6.5.
**d. En este contexto, no tiene sentido determinar el intervalo de confianza, ya que se conoce la altura media de todos los pinos.**

**Recebido em: 28 de março de 2020**
**Aprovado em: 26 de agosto de 2020**