

**UN ANÁLISIS EXPLORATORIO DE LA COMPRESIÓN DEL
INTERVALO DE CONFIANZA POR ESTUDIANTES DE
BACHILLERATO**

Antonio Francisco Roldán López de Hierro



**UNIVERSIDAD
DE GRANADA**

Trabajo Fin de Máster en Didáctica de la Matemática

Universidad de Granada

Facultad de Ciencias de la Educación

Departamento de Didáctica de la Matemática

Dirigido por

Dra. Carmen Batanero Bernabeu

Septiembre de 2019

UN ANÁLISIS EXPLORATORIO DE LA COMPRENSIÓN DEL INTERVALO DE CONFIANZA POR ESTUDIANTES DE BACHILLERATO

Departamento de Didáctica de la Matemática

Septiembre de 2019



UNIVERSIDAD DE GRANADA

Memoria del TRABAJO FIN DE MÁSTER realizado bajo la dirección de la profesora Dra. Carmen Batanero Bernabeu y presentada por el alumno Antonio Francisco Roldán López de Hierro para optar al grado de Maestro en Didáctica de la Matemática.

Fdo.: Antonio Francisco Roldán López de Hierro

VºBº de la directora

Fdo.: Dra. Carmen Batanero Bernabeu

*A Carmen Batanero,
por un bienio de conocimiento científico y humano
como puerta abierta a un futuro de esperanza.*

ÍNDICE

Índice	I
Resumen / Abstract	III
Prefacio	IV

Capítulo I. Justificación del problema

1.1.- Introducción.....	1
1.2.- Importancia de la Inferencia Estadística.....	2
1.3.- Importancia de la estimación por intervalos.....	4
1.4.- La enseñanza de la estimación por intervalos en España.....	5
1.5.- El marco teórico	6
1.6.- La estimación por intervalos como objeto matemático	8
1.6.1.- El significado frecuencial: intervalo de confianza.....	9
1.6.2.- El significado bayesiano: intervalos de credibilidad.....	11
1.6.3.- El significado en remuestreo: intervalo de remuestreo	12
1.6.4.- Síntesis de significados	14
1.7.- El intervalo de confianza para la media de una población normal con desviación típica conocida.....	16
1.8.- Objetivos del trabajo.....	18

Capítulo II. Antecedentes

2.1.- Introducción.....	21
2.2.- Primeras investigaciones sobre el tema y errores de interpretación de los intervalos de confianza por parte de investigadores.....	22
2.3.- Comprensión mostrada por parte de los estudiantes del intervalo de confianza	24
2.4.- Comprensión por parte de futuros profesores	28

Capítulo III. Desarrollo de la investigación y análisis de los resultados

3.1.- Introducción.....	31
3.2.- Descripción de la muestra	31

3.3.- Descripción del cuestionario	32
3.4.- Resultados en los ítems de opción múltiple	37
3.4.1.- Resultados en el ítem 1. Definición de intervalo de confianza	37
3.4.2.- Resultados en el ítem 2. Relación entre ancho del intervalo y tamaño de la muestra	38
3.4.3.- Resultados en el ítem 3. Precisión y confianza	39
3.4.4.- Resultados en el ítem 4. Relación entre amplitud y varianza de la población	40
3.4.5.- Resultados en el ítem 5. Amplitud y posición de la media	41
3.4.6.- Resultados en el ítem 6. Uso del intervalo de confianza	42
3.4.7.- Síntesis de los resultados obtenidos en los ítems de opción múltiple	43
3.5.- Resultados en el problema abierto	46
3.5.1.- Cálculo del intervalo	47
3.5.2.- Errores más frecuentes en el cálculo del intervalo	50
Capítulo IV. Conclusiones del estudio	
4.1.- Conclusiones generales	53
4.2.- Conclusiones respecto de los objetivos planteados	55
4.3.- Propuestas de mejora para el futuro y desarrollo de nuevas investigaciones	56
Referencias	58
Anexo I. Puntuaciones individuales de los estudiantes en los ítems de opción múltiple	61
Anexo II. Sobre la dificultad al utilizar la notación $1 - \alpha$ para el nivel de confianza	62
Anexo III. Actividades de investigación realizadas durante la elaboración del Trabajo Fin de Máster	64

RESUMEN

Esta Memoria está dedicada a la descripción de un estudio exploratorio, desarrollado con estudiantes del segundo curso de Bachillerato, sobre su comprensión e interpretación de la noción de “*intervalo de confianza*”. Para ello, se han examinado las respuestas aportadas por el alumnado a un cuestionario con seis ítems de respuesta múltiple y una cuestión abierta. También se ha revisado y producido una síntesis de investigaciones previas sobre esta temática y se han identificado los principales conflictos semióticos latentes, utilizando como marco teórico el *Enfoque Ontosemiótico del Conocimiento e Instrucción Matemáticos*. Además, se incluyen aportaciones originales en este ámbito.

ABSTRACT

The main aim of this work is to describe an exploratory study on the understanding and interpretation of the notion of “*confidence interval*” by students of the second year of Baccalaureate. To achieve this goal, the answers provided by students to a questionnaire (including six items of multiple response and one open question) have been reviewed. Also, a synthesis of previous research on this subject has been carried out and the main latent semiotic conflicts have been identified using the theoretical framework provided by the *Onto-semiotic Approach to mathematical knowledge and instruction*. In addition, some new contributions to this area are included.

Prefacio

Mi interés por el tema elegido para el Trabajo Fin de Máster surge de mi experiencia como profesor de Educación Secundaria Obligatoria y Bachillerato durante 17 años, donde observé las dificultades de los estudiantes durante el estudio de intervalos de confianza, que ahora pretendo analizar de forma más sistemática. En las actuales pruebas de acceso a la universidad se suele incluir un problema de intervalo de confianza en casi todas las convocatorias. Así, una adecuada comprensión del tema es necesaria para avanzar en el estudio de la Inferencia Estadística durante los cursos superiores. Por tanto, la evaluación de las dificultades de los estudiantes permitirá a los profesores una mejor organización de la enseñanza y de la preparación para los exámenes de acceso.

La Memoria está organizada en cuatro capítulos.

- En el primer capítulo se describe la Inferencia Estadística junto con algunas de las razones que la dotan de una gran importancia en la actualidad, con especial atención a la enseñanza del tema de intervalos de confianza en España. Situándose en el marco teórico del *Enfoque Ontosemiótico del Conocimiento e Instrucción Matemáticos*, se hace un estudio de los posibles significados (frecuencial, bayesiano y en remuestreo) de la estimación por intervalos como metodología de estimación de parámetros desconocidos. También se fija la notación que se empleará y se describen los objetivos y las hipótesis de partida del trabajo.
- El segundo capítulo está dedicado al estudio de los antecedentes de este tema, mostrando las dificultades y los conflictos semióticos que han sido descritos en investigaciones sobre el tema en distintos grupos (estudiantes de diversos grados, de posgrado, profesores en formación, graduados en matemáticas y ciencias afines e incluso investigadores).

- El tercer capítulo incluye la descripción completa del estudio exploratorio realizado sobre la interpretación que hace el alumnado del segundo curso de Bachillerato acerca de cuestiones relacionadas con los intervalos de confianza. Para ello, se ha utilizado un cuestionario diseñado con seis ítems de opción múltiple y un problema (cuestión abierta) para que el alumnado pudiese expresar sus conocimientos sin verse restringido a cuatro posibles respuestas. Se describe la muestra y el contexto escolar en el que se ha realizado el estudio y se lleva a cabo un análisis detallado del cuestionario utilizado y de las respuestas aportadas por el alumnado. En este capítulo se muestran varias contribuciones originales fruto del estudio que se ha realizado.
- Teniendo en mente los objetivos que se pretendían alcanzar, el último capítulo está dedicado a la exposición de conclusiones y a la propuesta de mejoras que puedan llevarse a cabo en el futuro y que puedan dar lugar a nuevas investigaciones sobre esta temática.

Seguidamente, se relacionan las referencias utilizadas en el trabajo junto con varios anexos pertinentes que sirven para detallar algunos aspectos concretos del mismo. En el último anexo, puede consultarse las actividades formativas en las que he participado durante este periodo, varias de ellas orientadas asimismo al estudio de la interpretación que hace el alumnado de los intervalos de confianza.

Finalmente, deseo agradecer al alumnado y al profesorado de los dos institutos que han participado en este estudio por su buena disposición para contribuir al mismo y por su amabilidad en todo momento. Y a los compañeros del Departamento de Didáctica de la Matemática que me han ayudado en esta aventura.

Antonio Francisco Roldán López de Hierro.

Septiembre de 2019.

Nota: Para facilitar la lectura de la presente Memoria, en su elaboración se ha utilizado en ocasiones el masculino como género no marcado para referirse tanto al femenino como al masculino, práctica usual en el ámbito científico.

Capítulo 1.

Justificación del problema

1.1. Introducción

Uno de los problemas más importantes a los que se enfrenta cualquier docente es el de conseguir que el aprendizaje de su alumnado sea significativo, en el sentido de que se integre dentro de las estructuras cognitivas previas y pueda ser utilizado con posterioridad en cualquier contexto de la vida real en que pueda resultar de ayuda. De esta forma, una de las mayores dificultades a las que podemos hacer frente dentro del campo de la didáctica de la matemática es la de transmitir una adecuada interpretación de las diferentes nociones matemáticas. Por desgracia, el alumnado suele alcanzar una serie de habilidades de cálculo manual (operaciones, representación, derivación, etc.) que, en ocasiones, no se integran en una competencia más importante que el propio cálculo: interpretar adecuadamente los diferentes conceptos, procedimientos y resultados desarrollados durante la resolución de problemas. Esto origina graves deficiencias en el proceso de aprendizaje que se convertirán, posiblemente, en obstáculos insalvables cuando el estudio requiera de un mayor nivel de profundización.

Este Trabajo de Fin de Máster está dedicado al estudio de la comprensión del alumnado de los intervalos de confianza, tema de gran relevancia en la actualidad debido a sus importantes aplicaciones prácticas en la mayor parte de las ramas de las ciencias naturales y sociales. Aunque hemos recabado información procedente de alumnos de los primeros cursos universitarios de varios grados (para realizar estudios posteriores), este trabajo se centrará en la información procedente de alumnado del segundo curso de Bachillerato en Andalucía, pues este es el único nivel preuniversitario que incluye, entre sus contenidos, varios temas dedicados a la Inferencia Estadística (muestreo, estimación puntual y mediante intervalos, y, en los anteriores currículos, contraste de hipótesis) dentro de la asignatura “*Matemáticas aplicadas a las Ciencias Sociales IP*”.

A lo largo de nuestros años de experiencia como docentes tanto en centros de Educación Secundaria Obligatoria y Bachillerato como en la propia universidad, hemos detectado que el alumnado estudia la Inferencia Estadística de una forma muy superficial, dando prioridad al cálculo que realiza a través de fórmulas previamente memorizadas, lo que provoca dos inconvenientes: por un lado, los estudiantes no suelen conocer las razones fundamentales que originan los problemas que están tratando de resolver (por ejemplo, estimar un parámetro poblacional desconocido), que reducen a simples fórmulas y, por otro lado, son muchas veces incapaces de dar una interpretación adecuada a los resultados obtenidos. Por ejemplo, estamos interesados en detectar el porcentaje de alumnado de segundo curso de Bachillerato que afirma, una vez calculado el intervalo de confianza al 95%, que dicho intervalo posee una probabilidad del 95% de contener al verdadero valor del parámetro poblacional (error conocido como «*falacia fundamental de la confianza*»; véase Behar, 2001, y Olivo, 2008).

Para desarrollar esta labor, hemos realizado un estudio de evaluación con alumnado matriculado en la asignatura anteriormente mencionada en dos centros educativos, uno en la capital de Granada y otro en un pueblo de su periferia, y hemos analizado sus respuestas tanto a seis ítems de opción múltiple como a una cuestión abierta. Mostraremos los resultados en el tercer capítulo, junto con algunas conclusiones que se obtienen del presente estudio.

1.2. Importancia de la Inferencia Estadística

La *Inferencia Estadística* es el conjunto de métodos y técnicas que permiten inferir, a partir de la información empírica proporcionada por muestras aleatorias, el comportamiento global de una determinada población, a la vez que proporcionan un valor aproximado de la posibilidad de error en una serie de repeticiones del método medido en términos de probabilidad (Harradine, Batanero y Rossman, 2011). Coloquialmente, trata de describir las características básicas de un conjunto muy grande (posiblemente, inaccesible en su totalidad) a través del estudio de pequeñas porciones que se toman aleatoriamente y que se analizan en profundidad. Por tanto, parte de sus razonamientos son de tipo inductivo (que van de lo particular a lo general, es decir, de la muestra a la población), si bien también hace uso del razonamiento deductivo (por ejemplo, de la distribución muestral al valor del estadístico obtenido en una muestra particular).

Esta rama de la estadística está adquiriendo cada día mayor importancia en muy diversas disciplinas debido a la necesidad de inferir información sobre poblaciones

tremendamente grandes (considérese, por ejemplo, la información que se genera cada día en Internet) sobre las que hay que tomar decisiones basadas en hechos científicos. La didáctica, en general, no ha permanecido ajena a este auge, de manera que, coherentemente, en los últimos años han aparecido diversas publicaciones que sintetizan esta idea desde esta perspectiva (véanse, por ejemplo, Begué, Batanero, Ruiz y Gea, 2019; Ben-Zvi, Bakker y Makar, 2015; Castro Sotos, Vanhoof, Noortgate y Onghena, 2007; Harradine et al., 2011; y Tobías-Lara y Gómez-Blancarte, 2019). Además, está adquiriendo cada vez un mayor protagonismo en tareas de investigación, lo que provoca una mayor exigencia al docente a la hora de impartir sus contenidos a estudiantes muy distintos tanto por sus conocimientos previos como por sus capacidades y sus habilidades.

Si tuviéramos que resaltar una sola característica de la Inferencia Estadística, a la vez que una virtud, esta sería, sin duda, su carácter aplicado. Su esencia se basa en describir el mundo en que vivimos de la forma más correcta posible, admitiendo que el modelo que se obtenga siempre será aproximado, es decir, una representación simplificada de la realidad. En tanto que sus técnicas proponen la descripción de los fenómenos naturales, sociales y económicos que nos rodean, esta disciplina siempre será de gran interés no solo para científicos, sino para el público en general.

Muchos grados universitarios poseen, en sus primeros cursos, asignaturas que abordan la docencia de los contenidos de Inferencia Estadística dividiéndola, secuencialmente, en temas de muestreo, estimación (puntual y por intervalos) y contraste de hipótesis. Precisamente, esta organización es la misma que se emplea en el segundo curso de Bachillerato donde, a partir de razonamientos muy generales con nociones elementales de estadística, se abordan situaciones concretas en las que adquieren un papel relevante la distribución de partida, el tamaño de la muestra y el parámetro o los parámetros de interés. La profundidad y la amplitud con la que se tratan estos temas dependen, esencialmente, del grado cursado por el estudiante.

Si nos fijamos, de todas las asignaturas que estudiamos a lo largo de la educación preuniversitaria, una de las que probablemente aportan más contenidos comunes a prácticamente todos los grados universitarios (desde las matemáticas o la física, hasta los que tratan de describir el comportamiento humano, mucho más aleatorio e impredecible, como la didáctica o la psicología) es la estadística y, en particular, la importancia de la Inferencia Estadística queda patente en el hecho de que se trata de una de las partes comunes a casi todos los grados universitarios.

1.3. Importancia de la estimación por intervalos

Una de las tareas más importantes de la estadística es la de determinar, con la mayor precisión posible, la distribución de probabilidad de ciertas variables aleatorias de interés (especialmente físicas, sanitarias y socioeconómicas). En muchas ocasiones, dichas distribuciones quedan completamente determinadas mediante unos pocos parámetros poblacionales que permiten el conocimiento completo de la variable a efectos de poder calcular probabilidades asociadas (como es el caso de la distribución normal que queda determinada conociendo su media y su varianza), y nos proporcionan métodos científicos para la toma de decisiones en general. Sin embargo, lo usual es que dichos parámetros sean desconocidos, lo que plantea el problema de su estimación (Batanero y Borovcnik, 2016; Moore, McCabe, Alwan y Duckworth, 2016).

Aunque la estimación puntual puede ser de gran ayuda en situaciones concretas, es obvio que la estimación de parámetros poblacionales a través de intervalos de confianza es una técnica mucho más avanzada pues, por un lado, permite indicar un rango de valores entre los que se podría encontrar el verdadero valor (que siempre es desconocido) y, por otro, proporciona un valor real que sirve como medición del grado de incertidumbre asociado al proceso de construcción del intervalo (Cepeda-Cuervo et al., 2008).

En la actualidad, las críticas a las que se están enfrentando las investigaciones basadas únicamente en el cálculo de p-valores están provocando que se vuelva cada vez más al intervalo de confianza (véase Batanero, 2000; de la Fuente y Díaz-Batanero, 2004; Díaz-Batanero, Lozano-Rojas y Fernández-Calderón, 2019). En esta línea, algunas asociaciones profesionales como, por ejemplo, la *American Psychological Association*, están recomendando a los investigadores que, además de los propios contrastes, acompañen los mismos con los correspondientes intervalos de confianza que puedan aportar una visión más global de la decisión tomada (Wilkinson y TFSA, 1999; Yaremko, Harari, Harrison y Lynn, 2013). Dichos intervalos permiten a los interesados valorar de una forma más amplia la posibilidad de equivocarnos al tomar una decisión mediante un contraste de hipótesis y, por así decirlo, son capaces de medir la mayor o menor exactitud del proceso de estimación (Belia, Fidler, Williams y Cumming, 2005). Por todo ello, la estimación mediante intervalos de confianza vuelve a ser un tema de actualidad, y puede encontrarse en cualquier manual (universitario o preuniversitario) de Inferencia Estadística como uno de los contenidos más destacados.

1.4. La enseñanza de la estimación por intervalos en España

Como indicamos anteriormente, la legislación española establece el inicio del estudio de la estimación mediante intervalos de confianza durante el segundo curso del Bachillerato, dentro de la modalidad de Ciencias Sociales, como parte de la asignatura “*Matemáticas aplicadas a las Ciencias Sociales II*”. Los contenidos estadísticos de esta asignatura, además de la probabilidad, incluyen nociones de muestreo, estimación y, en el anterior currículo, de contraste de hipótesis. La estimación por intervalos se desarrolla con objeto de inferir información acerca de la media de una población normal, la diferencia de medias entre dos poblaciones normales independientes, la proporción de individuos que presentan una determinada característica o, incluso, la diferencia de proporciones. Resumimos en la Tabla 1.4.1 algunas de las orientaciones curriculares fijadas por el Ministerio de Educación, Cultura y Deporte en su disposición (MECD, 2015, p. 389), en cuanto a contenidos, criterios de evaluación y estándares de aprendizaje evaluables referentes a la estimación por intervalos.

Tabla 1.4.1. Orientaciones curriculares para la enseñanza de la estimación por intervalos en Bachillerato

Contenidos	Criterios de evaluación	Estándares de aprendizaje evaluables
<ul style="list-style-type: none"> • 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. 	<ul style="list-style-type: none"> • Describir procedimientos estadísticos que permiten estimar parámetros desconocidos de una población con una fiabilidad o un error prefijados, calculando el tamaño muestral necesario y construyendo el intervalo de confianza para la media de una población normal con desviación típica conocida y para la media y proporción poblacional cuando el tamaño muestral es suficientemente grande. 	<ul style="list-style-type: none"> • Construye, en contextos reales, un intervalo de confianza para la media poblacional de una distribución normal con desviación típica conocida. • Construye, en contextos reales, un intervalo de confianza para la media poblacional y para la proporción en el caso de muestras grandes. • Relaciona el error y la confianza de un intervalo de confianza con el tamaño muestral y calcula cada uno de estos tres elementos conocidos los otros dos y lo aplica en situaciones reales. • Utiliza las herramientas necesarias para estimar parámetros desconocidos de una población y presentar las inferencias obtenidas mediante un vocabulario y representaciones adecuadas. • Identifica y analiza los elementos de una ficha técnica en un estudio estadístico sencillo.

Estos contenidos coinciden con los que luego serán requeridos a los estudiantes

durante la celebración de la *Prueba de Evaluación para el Acceso a la Universidad* (conocida como *PEvAU*), que permite comenzar sus estudios universitarios a los alumnos que la aprueban en las diversas comunidades, como es el caso de Andalucía. En concreto, la mencionada prueba ha incluido casi siempre, a lo largo de los últimos años, en el caso andaluz, una cuestión sobre intervalos de confianza, usualmente sobre la media o sobre la proporción (López-Martín, Batanero, Díaz-Batanero y Gea, 2016). Aunque hay una gran diversidad de cuestiones que se han propuesto, la mayoría de ellas versa sobre la determinación de un intervalo de confianza concreto, en un contexto aplicado, o sobre el cálculo del tamaño necesario para que una muestra cumpla cierta condición (usualmente, que tenga una amplitud inferior a una fijada de antemano). En menor medida aparecen preguntas sobre la interpretación del intervalo de confianza o sobre sus propiedades cuando se modifica uno de los factores que intervienen en su cálculo.

1.5. El marco teórico

Este Trabajo de Fin de Máster está basado en el marco teórico conocido como *Enfoque Ontosemiótico del Conocimiento* (abreviadamente, *EOS*), desarrollado en nuestra universidad (Godino, 2002; Godino, Batanero y Font, 2007; 2019). Por razones de extensión, no podemos describir aquí toda esta compleja teoría, pero sí damos a continuación algunas pinceladas acerca de los elementos que utilizaremos en el trabajo.

El EOS (Godino, Batanero y Font, 2007) es una teoría científica acerca de la forma en la que las personas e instituciones desarrollan o aprenden contenidos matemáticos. Parte de la idea de que la creación o asimilación de nociones matemáticas se basa en las prácticas concretas realizadas por los individuos dentro de una institución a lo largo de la fase de resolución de las tareas y problemas que les son propuestos (Godino, 2002). Una institución, según este autor, es un grupo de personas que se interesan por un mismo tipo de problemas y comparten las herramientas para su resolución. Así, puede considerarse la institución educativa en general, o bien puede considerarse una clase concreta como institución. De esta forma, el docente actúa tratando de acercar el *significado institucional* de cada una de las diferentes nociones matemáticas (entendido como las prácticas comúnmente acordadas y aceptadas dentro de la institución) al *significado personal* que cada estudiante construye sobre las mismas, el cual puede ser diferente al de la institución.

En el caso concreto del presente trabajo, consideramos problemas asociados al cálculo y a la interpretación del intervalo de confianza para la media poblacional. Por ello, dado que el significado de un objeto matemático viene determinado por el conjunto

de prácticas ligadas a la resolución de las situaciones problemáticas que lo originan, en nuestro caso, estamos interesados en indagar acerca de las prácticas relacionadas con toda clase de problemas que nacen en el contexto de la estimación por intervalos.

Teniendo en cuenta el carácter evaluador del presente trabajo, cabe mencionar que el EOS describe la evaluación de la *comprensión* mostrada por el alumno como el análisis de las coincidencias entre los significados institucional y personal sobre cada objeto matemático (Godino, 1996). Dicha comprensión supone que el estudiante va adquiriendo el significado institucional del objeto de forma progresiva. Se tienen en cuenta los aspectos del objeto cuya comprensión se quiere evaluar. En este sentido, trataremos de evaluar tanto la comprensión de la definición y propiedades básicas del intervalo de confianza como el razonamiento mostrado por los estudiantes al resolver una situación problemática relacionada con el intervalo de confianza para la media. Para este problema, determinaremos el nivel necesario para un alumno que ha alcanzado una adecuada comprensión de las nociones intervinientes (manifestado a través de una solución correcta y bien fundamentada), y también consideraremos algunas resoluciones como parcialmente correctas.

Según el EOS, el desarrollo de cualquier práctica matemática involucra, de forma explícita o implícita, diferentes objetos matemáticos que se clasifican de la siguiente forma (con objeto de analizar más profundamente dichas prácticas):

- *Situaciones-problemas* (contextos problemáticos de índole matemática o no que originan un determinado objeto). En el presente trabajo nos centraremos en problemas abiertos sobre el intervalo de confianza para la media.
- *Lenguajes* (términos, expresiones, notaciones, tablas, gráficos). Tendremos en cuenta nociones como media, desviación estándar, muestra, nivel de confianza, valor crítico, junto con sus correspondientes representaciones simbólicas, gráficas y/o tabulares.
- *Conceptos* (introducidos a través de definiciones o descripciones). En nuestra investigación tendremos en cuenta, por ejemplo, las nociones antes comentadas, así como la diferencia entre media de la población y media muestral, o las ideas de precisión, de amplitud y de valor crítico.
- *Proposiciones* (teoremas y propiedades que establecen relaciones entre los conceptos). Será de nuestro interés observar si el alumnado conoce la fórmula concreta de cálculo del intervalo de confianza para la media, expresándola mediante

los símbolos adecuados de forma teórica. Analizaremos la comprensión acerca de cómo afecta a la amplitud del intervalo de confianza la modificación de algunos de los factores que intervienen en su construcción (como el nivel de confianza, la desviación estándar o el tamaño de la muestra).

- *Procedimientos* (algoritmos, operaciones, estrategias). Nos fijaremos especialmente en las estrategias de resolución de la cuestión abierta junto con las operaciones involucradas para una correcta resolución.
- *Razonamientos* (argumentos usados para justificar los procedimientos y/o las soluciones de los problemas). En el presente trabajo nos fijaremos, por ejemplo, en la argumentación mostrada por el alumnado en el desarrollo del problema.

Finalmente, utilizaremos la idea de *conflicto semiótico*. Godino (2002) propuso la intervención de tres componentes en las funciones semióticas (o correspondencias) que tienen lugar durante el trabajo matemático: la expresión (el signo), el contenido (significado de tal signo, lo representado) y el criterio (o regla de correspondencia) que permite expresar la relación entre las dos componentes anteriores. Por ejemplo, utilizamos “ $z_{\alpha/2}$ ” como signo (expresión) para representar el valor crítico (contenido) asociado al nivel de confianza fijado, quedando establecido un vínculo (criterio) entre ellos que el alumnado debe conocer. Se produce un *conflicto semiótico* cuando existe una discrepancia entre el significado que el autor ha establecido de la función semiótica y el que manifiesta la persona encargada de su interpretación. Por ejemplo, un conflicto semiótico se produce cuando el alumnado considera que \bar{x} representa la media de la población. En el caso del intervalo de confianza, algunos conflictos semióticos se deben al hecho que, en la práctica estadística, existen varias metodologías que se confunden entre sí, y cada una de las cuáles representa un significado parcial del intervalo.

A continuación hacemos un resumen de las diferentes interpretaciones de los intervalos de confianza partiendo de un trabajo previo (Batanero, Díaz-Batanero, López-Martín y Roldán, en revisión), aunque en la enseñanza en Bachillerato y en la mayoría de los cursos universitarios sólo se muestra el significado frecuencial.

1.6. La estimación por intervalos como objeto matemático

La estimación por intervalos ha recibido diferentes soluciones matemáticas a lo largo de su historia, como se describe en Olivo (2007) y Rivadulla (1991) y, al igual que en el caso de los contrastes de hipótesis, ha estado ligada al problema de justificación del

método inductivo en las ciencias empíricas (de la Fuente y Díaz-Batanero, 2004). Dicho problema surge de la necesidad de mejorar el método científico y la lógica de la investigación en ciencias no deductivas (Díaz-Batanero, 2018; Mayo y Cox, 2006).

La dificultad de validar las teorías científicas obtenidas inductivamente a partir de datos empíricos derivó en una amplia discusión filosófica que contribuyó al desarrollo de la inferencia estadística y, al mismo tiempo, llevó a que se propusieran diferentes soluciones, tanto desde el punto de vista matemático como filosófico (Cabriá, 1994; de la Fuente y Díaz-Batanero, 2004; Lecoutre y Poitevineau, 2014).

La estimación por intervalos surge al estudiar un fenómeno aleatorio que viene caracterizado por una distribución de probabilidad que depende de uno o varios parámetros desconocidos. Puesto que es difícil recolectar los datos de la población completa, se toma una muestra aleatoria de valores independientes de la población, en donde se calcula un estadístico para dar un valor aproximado del parámetro (o parámetros). En la actualidad existe una variedad de aproximaciones a la estimación por intervalo que se diferencian en sus bases filosóficas y en el método de cálculo. No obstante, todas ellas intentan dar estimaciones que tengan en cuenta la medida de la variabilidad del muestreo, proporcionando un rango de valores para el parámetro (Morey et al., 2016). En lo que sigue nos restringimos al método frecuencial, la estimación bayesiana y el método de remuestreo. Cada uno constituye, de acuerdo al EOS, un significado parcial del intervalo de confianza pues varían algunos objetos matemáticos relacionados con los mismos.

1.6.1. El significado frecuencial: intervalo de confianza

Generalmente la estimación por intervalos se enseña siguiendo el método frecuencial, que fue iniciado por Neyman (1934; 1937). Según este autor, en primer lugar, fue necesario resolver el problema de *estimación*, identificando el estadístico muestral que permite aproximarse al parámetro. Además, es necesario valorar la *precisión matemática* de la estimación, dando unos límites de variación de la misma. Un supuesto importante en su teoría, que tiene una base frecuencial, es la ley de los grandes números. Por tanto, la probabilidad se considera, desde un punto de vista objetivo, como el límite al que tiende la frecuencia relativa de un suceso en una serie suficientemente amplia de repeticiones independientes del experimento (Díaz-Batanero, 2018).

La estimación del parámetro θ , que es un valor constante y desconocido, se realiza mediante un estimador $\hat{\theta}$, el cual es una variable aleatoria (que cambia de una muestra a

otra). El valor concreto obtenido en la muestra particular elegida se utiliza para determinar un único valor a partir de cada muestra aleatoria de la población. Así, para el caso de la media poblacional, se usa como estimador la media muestral.

Neyman (1934) sugiere calcular la desviación típica $S_{\hat{\theta}}$ del estimador $\hat{\theta}$ en la distribución muestral, es decir, la distribución de todos los posibles valores de $\hat{\theta}$ en todas las posibles muestras de igual tamaño formadas por elementos independientes de la población. Se debe expresar la precisión de la estimación en función de dicha desviación típica mediante un intervalo cuyos extremos indican los límites entre los que se presume se encuentre el valor verdadero del parámetro (Rivadulla, 1991):

$$EI = \hat{\theta} - k_1 S_{\hat{\theta}} \quad \text{y} \quad ES = \hat{\theta} + k_2 S_{\hat{\theta}}. \quad (1.1)$$

Neyman indica que se ha de calcular estos extremos teniendo en cuenta que la probabilidad de un error sea menor o igual que $1 - \alpha$, donde α es cualquier número $0 < \alpha < 1$, escogido anticipadamente y denominado *coeficiente de confianza* (Olivo, 1997). Además, sugiere los siguientes pasos (Mayo, 1981):

1. Elegir un valor del coeficiente de confianza $1 - \alpha$; con frecuencia se toma el valor 0,90; 0,95 o 0,99. Dicho valor determina las constantes k_1 y k_2 que se necesitan en el cálculo del intervalo. Si la distribución muestral es simétrica alrededor del estadístico $\hat{\theta}$, estas constantes son iguales y se determinan obteniendo los valores centrales de dicha distribución que incluyen una probabilidad $1 - \alpha$.
2. Tomar una muestra aleatoria x_1, x_2, \dots, x_n de valores independientes de la población, obteniendo a partir de ella el valor del estadístico muestral $\hat{\theta}$ y, como consecuencia, los extremos EI y ES del intervalo.
3. Suponer que el valor verdadero del parámetro se encuentra en el intervalo determinado, siendo $1 - \alpha$ la confianza en la estimación.

Mayo (1981) indica que la probabilidad $1 - \alpha$ se refiere a los extremos del intervalo, que son variables aleatorias, y que de ninguna manera representa la probabilidad de que el verdadero valor del parámetro θ esté situado dentro de tales límites. Es decir, dicha probabilidad se refiere al porcentaje de intervalos construidos a partir de diferentes muestras del mismo tamaño de la población que cubrirán el parámetro. Morey et al. (2016) sugieren la necesidad de diferenciar entre el procedimiento que genera el intervalo y el intervalo en sí mismo. Tenemos una confianza del $100(1 - \alpha)\%$ en que el procedimiento generará intervalos que, en dicho porcentaje, cubrirán el valor real del

parámetro, pero cada intervalo en sí mismo no es aleatorio, sino fijo, y puede, o no, cubrir el parámetro. Los autores denominan *falacia fundamental de la confianza* al hecho de interpretar dicha confianza como probabilidad de que el parámetro esté contenido en el intervalo.

1.6.2. El significado bayesiano: intervalos de credibilidad

La principal diferencia de la inferencia bayesiana con el enfoque frecuencial es que, mientras que en este último el parámetro θ es un valor desconocido pero constante, en la metodología bayesiana el parámetro se considera aleatorio y se caracteriza por una distribución *a priori* $p(\theta)$ (Bolstad, 2013; Gelman y Shalizi, 2012; Rivadulla 1991). Dicha distribución a priori representa el grado de credibilidad que el investigador concede a los diferentes posibles valores del parámetro. Por consiguiente, tiene una interpretación subjetiva, pues distintos sujetos podrían tener diferente conocimiento de la situación, de tal manera que su credibilidad en los diferentes valores del parámetro podría ser diferente.

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):

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (1.2)$$

En la expresión (2), $p(y|\theta)$ 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 la expresión (2) se puede reescribir como (3) considerando $K = 1/p(y)$:

$$p(\theta|y) = K p(y|\theta) p(\theta). \quad (1.3)$$

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 del contenido de su artículo publicado en Bayes (1970/1763); así, estrictamente hablando, puede considerarse que la inferencia bayesiana se inicia en esta fecha y es anterior a la frecuencial (Díaz, 2007; Rivadulla, 1991).

El método bayesiano ha recibido algunas críticas. Por un lado, no es posible asignar la distribución a priori y, por tanto, no es posible aplicar el teorema de Bayes en una investigación inicial donde se conoce poco del fenómeno. Sin embargo, al aplicarlo,

encontramos dos posturas diferentes (Díaz-Batanero, 2018; Rouanet; 1998):

- Algunos subjetivistas radicales finan la distribución a priori según su criterio y la conciben como una probabilidad subjetiva. Esta postura es criticada por asimilar “estadística bayesiana” y “estadística subjetiva”, provocando el rechazo al método bayesiano y quizás también a la enseñanza del mismo.
- Una postura moderada sugiere usar distribuciones iniciales uniformes que den a cualquier valor del parámetro el mismo peso, siendo “no informativas” (Cabriá, 1994; Lecoutre, 2006).

Dejando libertad al lector para adoptar su propio punto de vista, nos limitaremos a analizar la construcción de intervalos dentro de esta aproximación. 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 el intervalo C . Para el intervalo de credibilidad, $1 - \alpha$ se interpreta como la probabilidad de que el parámetro esté entre los valores dados, que se consideran fijos y no aleatorios (Cepeda-Cuervo et al., 2008).

1.6.3. El significado en remuestreo: intervalo de remuestreo

Algunas técnicas de inferencia más recientes, denominadas de *remuestreo*, se han desarrollado utilizando la facilidad, versatilidad y rapidez de los actuales ordenadores para producir simulaciones. La filosofía subyacente a los métodos de remuestreo es que toda la información requerida está en los datos que ya se han recogido. Por consiguiente, en lugar de basarse en modelos matemáticos teóricos de la distribución de la población, para aplicarlos sólo se utiliza la simulación de un proceso de muestreo aleatorio, aprovechando las capacidades tecnológicas descritas (Díaz-Batanero, 2018; Mooney, Duval y Duval, 1993).

Dentro del enfoque de remuestreo se incluye el método de *bootstrap*, conceptualizado por Efron (1979), cuya idea principal es extraer tantas muestras con reemplazamiento de la muestra original como sea posible y deducir, a partir de ellas, una distribución muestral aproximada. Para ello se calcula un estadístico de cada una de las diferentes re-muestras para obtener una distribución muestral empírica para el estadístico. La principal diferencia con el método frecuencial es que las muestras no se toman de la población, sino de una muestra inicial. La distribución resultante es similar a la distribución muestral pero, conceptualmente, es diferente; se denomina *distribución*

bootstrap y se utiliza para realizar inferencias sobre la población, incluido el cálculo de intervalos de confianza.

Otro método de remuestreo es el *Jackknife*, propuesto por Quenouille (1949) y refinado por Tukey (1956), y que se desarrolló originalmente para corregir el sesgo en los procedimientos de estimación. La principal diferencia con el *bootstrap* es que utiliza el muestreo sin reemplazamiento en lugar del muestreo con reemplazamiento. En este sentido, se eliminan sistemáticamente uno o unos pocos (hasta la mitad) de los elementos de la muestra original, dando lugar a nuevas submuestras que se utilizan para generar la distribución del estadístico.

En ambos métodos la construcción del intervalo de confianza se hace a partir de la distribución de remuestreo, siguiendo varios métodos. El más sencillo es obtener los límites de los valores centrales de la distribución muestral que dejan a cada lado la probabilidad $\alpha/2$ (método de percentiles). Estos límites obtenidos son próximos a los correspondientes en la inferencia frecuencial (Parsonage, Pfannkuch, Wild y Aloisio, 2016), pero la interpretación es muy diferente. Mientras que en la inferencia frecuencial el intervalo de confianza del 95% indica el porcentaje de muestras de la población que se espera contenga al parámetro, en el remuestreo, el intervalo del 95% se refiere a la precisión de las medidas repetidas del parámetro en el 95% de las muestras obtenidas por remuestreo (Borovnick, 2019). Es decir, la confianza se refiere a los límites en que esperamos varíen las futuras medias de las remuestras y no al parámetro en la población (Howell, n.d.).

Otro método sencillo se basa en la estimación de la desviación típica del estadístico, y el uso de la distribución normal (intervalo estándar). El lector puede encontrar varios métodos y una discusión sobre los mismos en Efron y Tibshirani (1986) y Engel (2010).

Algunos autores defienden que las técnicas de remuestreo tienen una formulación más general, además de ser más sencillas de comprender, porque se basan en menos conceptos que los métodos clásicos (Hesterberg, Monaghan, Moore, Clipson, y Epstein, 2003), aunque se exige la independencia de los elementos de la muestra original. Proporcionan, además, un procedimiento de estimación cuando no se dispone de información sobre las distribuciones muestrales, pero para parámetros clásicos como la media o la proporción el método frecuencial es más preciso.

La enseñanza del remuestreo se está extendiendo recientemente puesto que autores como Cobb (2007) proponen reemplazar la estadística inferencial tradicional por este enfoque que denominan inferencia informal y, además, simplificarlo aún más no haciendo

un estudio profundo de los conceptos subyacentes, mientras que otros autores, como Borovcnik (2019) sugieren que se utilice sólo como paso previo a la enseñanza de la inferencia frecuencial, para la que se dispone de mayor amplitud de métodos.

Finalmente, hay que tener en cuenta que si la muestra inicial con la que se comienza el método no es extraída con un procedimiento que asegure su representatividad, o se usan muestras excesivamente pequeñas, se puede dar lugar a estimaciones sesgadas, algo que también sucede en las técnicas clásicas de estimación (Gil-Flores, 2005). Las técnicas no paramétricas podrían constituir una alternativa a la inferencia frecuencial cuando no se cumplen ciertos supuestos. Recordamos que los métodos no paramétricos generalmente requieren reducir las variables en sus niveles de medida –por ejemplo, puntuaciones numéricas a rangos–, perdiendo así parte de la información disponible en los datos (Gil-Flores, 2005). Esta reducción del tipo de escala de medida no es necesaria en el remuestreo.

1.6.4. Síntesis de significados

En las diferentes aproximaciones a la construcción de intervalos para la estimación de parámetros se utilizan diferentes objetos matemáticos como podemos apreciar en la Tabla 1.6.4.1. Por ello, no es indiferente el enseñar una técnica u otra, puesto que constituyen significados parciales y diferentes de la estimación por intervalos, siguiendo el enfoque ontosemiótico. Aunque el problema inicial en todos ellos era la acotación de valores de un parámetro, vemos que el mismo concepto de parámetro es diferente, pues en inferencia bayesiana se considera una variable aleatoria con una distribución y no una constante. La probabilidad de que el método de algún modo garantice un resultado es muy diferente pues, mientras que en la aproximación frecuencial y remuestreo se trata de una probabilidad frecuencial, en la inferencia bayesiana trabajamos con probabilidades subjetivas.

Además esta probabilidad se refiere a diversos objetos y da lugar a diferentes conceptos: a) confianza o probabilidad de que el método a la larga construya un intervalo donde se encuentre el parámetro; b) credibilidad o probabilidad de que el parámetro esté en el intervalo; y c) probabilidad de que las medias de otras muestras de remuestreo caigan en el intervalo. Como veremos en el estudio empírico, algunas interpretaciones del coeficiente de confianza de los estudiantes serían las dadas en la inferencia bayesiana o remuestreo, que posiblemente fuesen más sencillas conceptualmente que la frecuencial. En la Tabla 1.6.4.1 también podemos observar cómo varía el procedimiento cuando se

emplea cada uno de los diferentes métodos. En resumen, las configuraciones de objetos en las tres aproximaciones son diferentes.

Tabla 1.6.4.1. Algunos objetos matemáticos ligados a los diferentes significados de la estimación por intervalos

Significados parciales	Problema	Conceptos/Propiedades	Procedimientos
Inferencia frecuencial	<ul style="list-style-type: none"> • Acotar el valor del parámetro a partir del estadístico muestral fijando el porcentaje de muestras cuyo intervalo lo contiene. 	<ul style="list-style-type: none"> • Interpretación frecuencial de la probabilidad. • Ley de los grandes números. • Distribución muestral. • Intervalo de confianza. • Coeficiente de confianza. • Se puede conocer la probabilidad de que intervalos calculados con otras muestras de la población contengan al parámetro. 	<ul style="list-style-type: none"> • Fijar el nivel de confianza. • Determinar el valor crítico correspondiente. • Calcular los extremos del intervalo a partir de los percentiles de la distribución muestral teórica.
Inferencia Bayesiana	<ul style="list-style-type: none"> • Actualizar la distribución a priori del parámetro. • Acotar un intervalo de valores de la distribución a posteriori del parámetro, donde se encuentre el parámetro con cierta probabilidad. 	<ul style="list-style-type: none"> • Interpretación subjetiva de la probabilidad. • Distribución a priori y a posteriori. • El parámetro es una variable aleatoria. • Intervalo de credibilidad. • Credibilidad. • Se puede conocer la probabilidad de que el parámetro esté en el intervalo. 	<ul style="list-style-type: none"> • Asignar una distribución a priori. • Teorema de Bayes. • Actualizar la distribución del parámetro obteniendo la distribución a posteriori. • Calcular el intervalo a partir de los percentiles de la distribución a posteriori.
Remuestreo	<ul style="list-style-type: none"> • Acotar los posibles valores de un parámetro, estimando, mediante simulación la posible variabilidad en el muestreo. 	<ul style="list-style-type: none"> • Interpretación frecuencial de la probabilidad. • Distribución de remuestreo. • Se puede conocer la probabilidad de que en otras muestras la estimación del parámetro caiga en el intervalo. 	<ul style="list-style-type: none"> • Obtención de muestras a partir de una muestra original. • Obtención de la distribución muestral de remuestreo mediante simulación. • Calcular los extremos del intervalo utilizando los percentiles de la distribución muestral de remuestreo.

1.7. El intervalo de confianza para la media de una población normal con desviación típica conocida

Antes de describir los objetivos del presente trabajo, nos parece conveniente analizar, siquiera brevemente, el problema concreto que utilizaremos con nuestros estudiantes junto con los contenidos estadísticos que deseamos evaluar, lo cual nos permitirá, a la vez, introducir la notación que emplearemos.

Consideremos una población y supongamos que estamos interesados en una característica de sus individuos que viene modelizada por una variable aleatoria X que sigue una distribución normal de media μ y varianza σ^2 (equivalentemente, de desviación típica σ), a la cual denotaremos por $N(\mu, \sigma^2)$ (por ejemplo, X puede medir la estatura o el sueldo mensual de una persona de la población). El *problema de estimación* se plantea cuando deseamos conocer uno de los parámetros poblacionales que definen la función de densidad (por ejemplo, su media o su varianza). En la práctica, dicho parámetro suele ser desconocido por muy diversas causas: por ejemplo, debido al enorme tamaño de la población (no podemos acceder a todos los individuos de la población), debido a que la medición de la variable X es costosa en tiempo o en dinero, debido a que la característica X varía relativamente rápido con el tiempo (sueldo de trabajadores de un país, que son contratados o despedidos cada día), o, finalmente, debido a que la medición provoca la destrucción del individuo (medición del tiempo de duración de una bombilla).

El parámetro de mayor interés en una población normal es su media μ . La normativa andaluza actual establece que el alumnado de segundo curso de Bachillerato matriculado en la asignatura “*Matemáticas aplicadas a las Ciencias Sociales II*” estudiará la estimación mediante intervalos de confianza de la media poblacional μ cuando la varianza σ^2 es conocida. Ciertamente, el caso más interesante y frecuente en la práctica es aquél en el que tratamos de estimar la media poblacional cuando la varianza también es desconocida (parece raro conocer la varianza sin conocer previamente la media). Sin embargo, el caso planteado en la normativa actual tampoco es del todo extraño: aunque la media puede cambiar de unos estudios a otros (por ejemplo, realizados en estaciones del año diferentes), es usual asumir que la variabilidad de los datos (expresada a través de su varianza) permanece constante, y es conocida a partir de estudios previos.

Para estimar el valor de μ lo usual es tomar una muestra aleatoria simple X_1, X_2, \dots, X_n de la variable X , es decir, un conjunto de n variables independientes e idénticamente distribuidas según la misma distribución que X . Utilizando su

independencia y un simple cambio de variable en la función de densidad, es sencillo demostrar que la variable aleatoria *media muestral*, es decir,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

también sigue una distribución normal cuya media es igual a la media μ de la población y cuya varianza es $\sigma_{\bar{X}}^2 = \sigma^2/n$. Fijado un nivel de confianza $1 - \alpha$ (usualmente, del 90%, del 95% o del 99%), la estimación por intervalo de la media poblacional μ está basada en determinar un intervalo, centrado en la media muestral \bar{X} , de manera que la probabilidad de que dicha variable pertenezca a dicho intervalo sea, al menos, igual al nivel de confianza. Dicho de otra forma, estamos tratando de determinar un intervalo con la condición de que la distancia entre la media muestral y la media poblacional (desconocida) esté acotada con un cierto nivel de probabilidad.

$$P(|\bar{X} - \mu| \leq r) \geq 1 - \alpha.$$

Es muy importante observar que el procedimiento anterior nos permite construir intervalos que contendrán, o no, a la media poblacional. Nunca sabremos si el verdadero valor de la media poblacional pertenece, o no, al intervalo de confianza una vez construido este porque, desde el principio, asumimos que μ es desconocida. El nivel de confianza está depositado en el método de construcción de los intervalos de forma que si construimos muchísimos de ellos, asociados a muestras aleatorias independientes, estamos razonablemente seguros de que al menos el $100(1 - \alpha)\%$ contendrá al verdadero valor de la media poblacional (aun cuando tampoco sabremos cuáles de ellos son los que realmente la contienen).

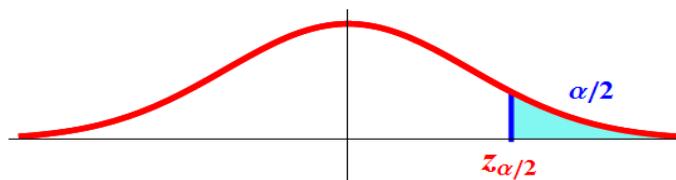


Figura 1.7.1. Representación gráfica de la propiedad que determina el valor crítico $z_{\alpha/2}$.

La determinación de la fórmula para obtener dicho intervalo está basada en la tipificación, la cual nos asegura que la variable aleatoria $Z = (\bar{X} - \mu)/\sigma_{\bar{X}}$ sigue una distribución normal estándar $N(0,1)$. Si denotamos por $z_{\alpha/2}$ al único valor real (positivo) que satisface la igualdad $P(Z > z_{\alpha/2}) = \alpha/2$ (al que llamaremos *valor crítico* al nivel de confianza $1 - \alpha$; véase la Figura 1.7.1), entonces:

$$\begin{aligned}
 1 - \alpha &= P\left(\left|\frac{\bar{X} - \mu}{\sigma_{\bar{X}}}\right| \leq z_{\alpha/2}\right) = P(|\bar{X} - \mu| \leq z_{\alpha/2} \sigma_{\bar{X}}) \\
 &= P\left(|\mu - \bar{X}| \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = P\left(-z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu - \bar{X} \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) \\
 &= P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right). \tag{1.4}
 \end{aligned}$$

En consecuencia, la fórmula que se suele emplear para estimar mediante un intervalo la media poblacional μ (que, recordemos, es un parámetro desconocido), con una confianza $1 - \alpha$, mediante una muestra de tamaño n que toma valores concretos x_1, x_2, \dots, x_n , es:

$$IC(\mu) = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right],$$

cuya notación se suele simplificar escribiendo:

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

siendo $\bar{x} = (x_1 + x_2 + \dots + x_n)/n$.

1.8. Objetivos del trabajo

El objetivo principal del presente trabajo es analizar la comprensión que muestra el alumnado de segundo de Bachillerato a lo largo de todo el proceso de estimación mediante intervalos de confianza, tanto de los procedimientos intermedios desarrollados durante el cálculo de dichos intervalos como de las nociones que intervienen en los mismos, prestando especial atención al significado personal final que el alumnado atribuye a los intervalos calculados. En concreto, y como ya hemos comentado, nos centraremos en el caso de la estimación de la media μ de una característica poblacional que sigue una distribución normal con varianza (o desviación estándar σ) conocida. Previamente, para poder interpretar nuestros resultados, hemos necesitado analizar los contenidos de la investigación previa como, por ejemplo, los diversos significados de la estimación por intervalos en las diferentes metodologías de la inferencia; se incluyen estos últimos para poder interpretar algunos errores citados en la investigación como conflictos semióticos consistentes en atribuir al intervalo de confianza y sus propiedades significados más próximos a las metodologías bayesianas y de remuestreo. Este objetivo general puede ser detallado en los siguientes objetivos parciales.

- O1.** Revisar y producir una síntesis de las investigaciones previas para poder situar respecto a ellas los resultados de nuestro trabajo. Este objetivo se aborda en el Capítulo 2.
- O2.** Analizar los significados parciales de la estimación por intervalos en diferentes metodologías de la inferencia. Se ha abordado en el Capítulo 1 y se usará posteriormente para interpretar algunos errores descritos en la investigación didáctica.
- O3.** Analizar la comprensión del alumnado del “*intervalo de confianza*” y cómo lo interpreta en casos prácticos concretos determinando, especialmente, si este considera que el parámetro poblacional debe estar incluido en dicho intervalo con un cierto nivel de probabilidad, es decir, si lo interpreta como intervalo de credibilidad o incluso como intervalo de remuestreo.

El resto de los objetivos se abordan en el Capítulo 3.

- O4.** Indagar el nivel de comprensión del alumnado de los diferentes factores que influyen en la fórmula del intervalo de confianza para la media de una población (a saber, la media muestral, el nivel de confianza, el valor crítico, la desviación estándar y el tamaño de la muestra), tanto de su significado concreto en el contexto del problema como de la forma en que su variación afecta a la amplitud y forma del propio intervalo de confianza.
- O5.** Detallar las estrategias que utilizan los estudiantes para resolver una cuestión abierta relativa a la determinación de un intervalo de confianza para la media, describiendo los pasos que parecen más razonables para su cálculo y el porcentaje de alumnos que siguen estos pasos. Estudiaremos también si el alumnado se apoya en una representación gráfica de la distribución normal para calcular el valor crítico asociado al problema.
- O6.** Identificar los principales conflictos semióticos latentes en las respuestas y estrategias de los estudiantes.
- O7.** Comparar todos estos resultados con los descritos en investigaciones anteriores, especialmente por Behar (2001), Olivo y Batanero (2007), Olivo (2008), Olivo, Batanero y Díaz (2008), Yáñez y Behar (2009) y por López-Martín, Batanero y Díaz-Batanero (2019a).

- O8.** Introducir nuevos interrogantes que puedan servir de base para futuros desarrollos didácticos sobre este tema. Este objetivo se aborda en las conclusiones.

Directamente relacionadas con los objetivos anteriores, y en base tanto a nuestra experiencia previa como a la literatura resumida en el Capítulo 2, describimos las siguientes hipótesis de partida.

- H1.** Se espera encontrar un buen nivel de conocimiento conceptual sobre la definición del intervalo y sus principales propiedades en los estudiantes de la muestra. Sin embargo, esperamos una escasa comprensión por parte del alumnado acerca de cómo influyen en la amplitud del intervalo de confianza los diversos factores intervinientes en la fórmula.
- H2.** Pensamos que algunos estudiantes de la muestra no comprenderán el problema planteado y la necesidad de resolverlo mediante intervalos de confianza. Sin embargo, se espera que la mayoría de los estudiantes desarrollen correctamente (o al menos de forma parcialmente correcta) los diversos pasos necesarios para calcular un intervalo de confianza. A priori, pensamos que el alumnado empleará la fórmula de cálculo del intervalo sin ningún tipo de razonamiento previo. Sin embargo, ante una cuestión abierta tan sencilla y frecuente en las pruebas de acceso a la universidad, consideramos que será elevado el porcentaje de alumnos que resuelvan dicho interrogante de una forma correcta o parcialmente correcta.
- H3.** No obstante lo anterior, esperamos que aparezcan en nuestros alumnos algunos conflictos semióticos descritos en la investigación previa entre los que se relacionarán en el Capítulo 2. Pensamos que la falacia fundamental de la confianza definida por Behar (2001) y Olivo (2008) está extendida entre el alumnado.

Finalmente, indicar que no esperamos encontrar porcentajes similares a los mostrados en estudios previos pues estos fueron realizados con estudiantes universitarios y nuestro caso se centra en alumnado preuniversitario. Pensamos también que seremos capaces de identificar nuevas aportaciones al estudio de los intervalos de confianza desde el punto de vista de la didáctica de la estadística.

Capítulo 2.

Antecedentes

2.1. Introducción

Como se ha indicado, la estimación por intervalos se enseña a los estudiantes desde diversos puntos de vista, predominando la concepción frecuencial, que es la que consideramos en esta Memoria. Si atendemos al número de nociones que se necesita comprender y manejar para demostrar un conocimiento apropiado del tema, la teoría de intervalos de confianza requiere de un menor esfuerzo por parte del alumnado que el campo de los contrastes de hipótesis. Ambas materias parten de los conceptos de población y muestra, hacen uso del proceso de muestreo y de las distribuciones muestrales y precisan de una cierta habilidad para el cálculo de probabilidades, prestando atención a la determinación de valores críticos utilizando tablas adecuadas de diferentes distribuciones de variables estadísticas (en el caso que nos ocupa, especialmente a partir de la distribución normal estándar). Sin embargo, como diferencia esencial, la estimación por intervalos obliga al alumnado a tener asimilada la noción de intervalo y de sus extremos, y a realizar una interpretación correcta de los intervalos obtenidos. Por el contrario, en el ámbito de los contrastes de hipótesis, también es imprescindible conocer otras muchas definiciones como hipótesis nula y alternativa, hipótesis simple y compuesta, nivel de significación, región crítica, región de aceptación, regla de decisión, tipos de error, potencia e incluso p-valor (Moore et al., 2016).

La menor cantidad de conceptos implicados supone una ventaja indiscutible de los intervalos de confianza frente a los contrastes de hipótesis lo que, añadido a su mayor sencillez pedagógica y conceptual, hace que muchos docentes se estén decantando por prestar más atención a este tema, y está llevando a muchos investigadores a incluirlos en sus estudios como complemento a los contrastes desarrollados (Cepeda-Cuervo et al., 2008; de la Fuente y Díaz-Batanero, 2004). Sin embargo, en su aplicación, surgen

multitud de dificultades tanto de cálculo como de interpretación. Este capítulo está dedicado a mostrar los antecedentes científicos de este campo de investigación en didáctica de la estadística, señalando sus avances y poniendo de manifiesto las dificultades descritas por diversos autores (tanto entre estudiantes de grado, de posgrado, como entre docentes, futuros docentes e incluso investigadores que hacen uso de la estimación por intervalos).

2.2. Primeras investigaciones sobre el tema y errores de interpretación de los intervalos de confianza por parte de investigadores

Los primeros estudios sobre la comprensión de los intervalos de confianza se llevaron a cabo mediante respuestas aportadas por investigadores que los utilizaban en sus trabajos de investigación. Estas respuestas fueron obtenidas de dos formas distintas: por un lado, a través de entrevistas o encuestas realizadas a los propios investigadores; por otro lado, también se utilizó el análisis de la interpretación de intervalos de confianza que hacían los propios investigadores en artículos publicados en revistas de investigación con impacto.

El primer estudio significativo en este tema fue desarrollado por Cumming, William y Fidler (2004), quienes se pusieron en contacto con 134 investigadores que habían publicado artículos de investigación en revistas internacionales en los cuales se utilizaban diversos métodos estadísticos, incluida la estimación por intervalos. Tras proporcionarles un intervalo de confianza para la media de una población construido a un nivel de confianza del 95%, se interrogó a dichos investigadores sobre los valores que ellos entendían que debían ser factibles para 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 (el 78%) respondieron que la media volvería a caer en el intervalo proporcionado en el 95% de las ocasiones (probabilidad de replicación). De esta forma, los encuestados hicieron una interpretación de los intervalos en el método de percentiles en remuestreo, o sea, concibieron el intervalo de confianza al 95% como un intervalo que contendrá a las medias muestrales con una probabilidad del 95%. Estes (1997) había indicado previamente que el porcentaje de investigadores que se dejaban llevar por esta idea era aún superior: del 83%.

En un estudio posterior, Belia, Fidler y Cumming (2005) desarrollaron un experimento en el que diversos investigadores accedían a una página web interactiva en la que podían encontrar varios gráficos acerca de la estimación mediante intervalos de

confianza, construidos al 95%, para dos medias de poblaciones independientes. A la vista de dichos gráficos, solicitaron a los investigadores que situaran, en alguna posición del gráfico, una de las medias para que el contraste t de diferencia de medias resultase ser estadísticamente significativo. La mayoría de los participantes situaron dichas medias muy alejadas entre sí (de manera que hubiese gran seguridad de que eran distintas), por lo que los correspondientes intervalos de confianza resultaban ser obligatoriamente disjuntos. Como sabemos, esto es un error de interpretación pues el hecho de que el contraste de medias resulte ser estadísticamente significativo significa que cada media no pertenece al intervalo de confianza para la otra media, pero no es obligatorio que los intervalos sean completamente disjuntos. La respuesta correcta fue descrita por muy pocos participantes. Es de destacar que varios participantes indicaron que, para conseguir el fin descrito, los intervalos de confianza debían tocarse en sus extremos, es decir, debía coincidir el extremo superior de un intervalo con el extremo inferior del otro. En este sentido, un buen criterio para decidir cuándo una diferencia de medias es estadísticamente significativa fue aportado por Cumming y Finch (2005), quienes establecen lo siguiente:

Se pueden comparar dos medias independientes a nivel $\alpha \leq 0.05$ cuando el traslape de los intervalos de confianza al 95% no es más que alrededor de la mitad del margen promedio del error. Para $\alpha \leq 0.01$ cuando los dos intervalos de confianza no se traslapan... Esas relaciones son lo suficientemente exactas cuando ambas muestras son al menos de tamaño 10, y cuando los márgenes de error no difieran en más de un factor 2 (p. 176).

En su estudio, Belia, Fidler y Cumming (2005) también solicitaron a los participantes que respondiesen a la misma pregunta asumiendo que las muestras, en lugar de independientes eran apareadas (es decir, cada individuo proporciona dos datos, por ejemplo, antes y después de un cierto tratamiento médico). La mayor parte de las personas que respondieron a esta cuestión lo hicieron atendiendo a los mismos argumentos que en el caso de muestras independientes, si bien sabemos que la expresión del intervalo de confianza para la diferencia de medias entre muestras apareadas es diferente del caso con muestras independientes.

Además de las respuestas incorrectas anteriores, los mismos autores describen en su estudio otros errores manifestados por los investigadores encuestados: por una parte, no manifiestan entender la importancia de que el experimento aleatorio esté bien diseñado de cara a la interpretación posterior de los intervalos obtenidos; por otro lado, no distinguen el intervalo de confianza de los márgenes de error proporcionados por el error típico de estimación (es decir, por la desviación estándar del estadístico muestral involucrado).

2.3. Comprensión mostrada por parte de los estudiantes del intervalo de confianza

La mayor parte de los estudios realizados sobre la comprensión mostrada por los participantes acerca de la estimación por intervalos se ha llevado a cabo con estudiantes, tanto antes como después de recibir formación sobre los mencionados contenidos.

En un primer estudio, Behar (2001) analizó el grado de comprensión del tema de los intervalos de confianza mostrado por 47 expertos (profesionales dedicados a la estadística y/o a su enseñanza) y 297 estudiantes del grado de ingeniería. Este autor puso de manifiesto un grave error de interpretación pues tanto un 29% de los expertos como la mitad de los alumnos encuestados afirmaba que el nivel de confianza indicaba el porcentaje de datos poblacionales que caen dentro del intervalo de confianza, es decir, un intervalo al 95% de confianza debe contener al 95% de los valores plausibles que se pueden obtener en la población. Otra de las conclusiones de su estudio establecía que el 40% de los expertos y poco más de la mitad de los estudiantes realizaban una interpretación bayesiana del intervalo, afirmando que el nivel de confianza era una medida de la probabilidad de que el intervalo contenga al verdadero valor de la media poblacional. Finalmente, menos de la mitad de los expertos y el 65% de los estudiantes manifestaban una comprensión adecuada del hecho de que algunos intervalos no contienen al verdadero valor del parámetro pero, a la larga, cuando se repite muchas veces el experimento, el porcentaje de intervalos que contienen al parámetro poblacional tenderá a coincidir con el nivel de confianza. Visto al contrario, hay una importante proporción de encuestados en ambos grupos que no asocia la frecuencia relativa de intervalos que incluyen la media poblacional con el nivel de confianza. Es más, el alumnado cree que cuando el nivel de confianza es alto, se obtienen intervalos más estrechos (dejando invariantes los demás factores), e incluso establecen una relación equivocada de proporcionalidad directa entre la anchura del intervalo y el tamaño de la muestra.

Al respecto de este estudio, Yáñez y Behar (2009) relataron que las conclusiones obtenidas por Behar ponen de manifiesto las enormes dificultades expresadas por varias clases de personas interesadas en la estimación por intervalos desde el punto de vista de la comprensión de su significado, su contenido y la interpretación del nivel de confianza, así como la influencia de dicho nivel sobre la anchura del intervalo.

Otro estudio que cabe citar en este contexto de investigación fue el desarrollado por Fidler y Cumming (2005), quienes solicitaron a 55 estudiantes del grado en ciencias

ambientales (unos matriculados en el último curso del grado y otros, posgraduados) que elaborasen una interpretación adecuada de los resultados mostrados en artículos de investigación, publicados en revistas científicas, en los que intervenían intervalos de confianza y contraste de hipótesis. En particular, se les pidió que indicasen el grado de evidencia mostrado en las publicaciones a favor o en contra de la hipótesis nula, clasificándolo en fuerte, moderado o ambiguo. Por un lado, inicialmente, sólo el 18% de los participantes en el estudio proporcionó la respuesta correcta acerca de la interpretación de intervalo; por otro, tras recibir formación académica sobre estimación por intervalos, el alumnado mejoró en la tarea de interpretar el desarrollo y los resultados obtenidos en los contrastes de hipótesis considerados.

En el mismo estudio, los autores también trabajaron con 180 estudiantes de psicología, constatando algunos de los errores de interpretación que aparecían al manejar intervalos de confianza:

- un 38% de los estudiantes realizaban una interpretación del intervalo correspondiente al método percentil en remuestreo, es decir, imaginaban el intervalo como un conjunto de valores razonables para la media de la muestra;
- muchos alumnos veían los intervalos de confianza desde una perspectiva descriptiva en lugar de valorar su carácter inferencial;
- un 19% de participantes confundía las nociones de rango (o recorrido de la variable aleatoria) e intervalo de confianza.

En cuanto a la comprensión de los diferentes aspectos que pueden influir sobre los intervalos de confianza, un 20% manifestó que la anchura del intervalo aumentaría cuando creciese el tamaño muestral, un 29% consideró que, en tal caso, la anchura no variaría y, finalmente, un 36% no fue capaz de describir si existía, o no, relación entre la anchura y el tamaño de la muestra. De hecho, estas conclusiones hacían referencia a alumnado que había recibido previamente formación en el tema de la estimación por intervalos.

Morey et al. (2016) investigaron el posible origen de estos errores y pusieron de manifiesto la importancia de que el alumnado distinga entre qué información es conocida con anterioridad a la recogida de datos (por ejemplo, las propiedades del proceso de obtención de la fórmula del intervalo de confianza) y qué aspectos son conocidos tras el análisis de los mismos (que el intervalo calculado puede, o no, contener al verdadero valor). De hecho, cabe destacar que no tiene sentido calcular probabilidades a posteriori de la obtención del intervalo de confianza, por lo que esta teoría nunca ha considerado

tales probabilidades. En este sentido, Neyman (1937) se preguntaba:

Consideremos ahora el caso en el que se ha obtenido una muestra y el intervalo de confianza se ha dado. ¿Podemos decir que, en este caso particular, la probabilidad de que el valor verdadero (del parámetro) caiga en los límites es igual al X%? La respuesta es obviamente negativa.

Algunos de los participantes en el estudio, además de haber mostrado dificultades al establecer la relación existente entre el nivel de confianza y la anchura del intervalo, manifestaron una tendencia a considerarlos de forma conjunta, sin poder separarlos, es decir, negaban la posibilidad de alterar (subiendo o bajando) el nivel de confianza. Directamente relacionada con esta dificultad está el hecho, indicado por Morey et al. (2016), de preferir muy a menudo un intervalo más estrecho, sin observar que depende de muchos factores en juego. Finalmente, de su estudio se desprende que, en muchas ocasiones, no se valora suficientemente la técnica de estimación por intervalos en el ambiente de la toma de decisiones sobre ciertas hipótesis planteadas, lo cual puede deberse a la ausencia de interpretación del intervalo como una familia razonable de posibles valores del parámetro estudiado.

En una serie sucesiva de trabajos (Olivo, 2008; Olivo y Batanero, 2007; Olivo et al., 2008), Olivo estudió, utilizando un conjunto de alumnos de ingeniería, su comprensión acerca del intervalo de confianza. Para ello, basándose en una importante selección de libros de texto universitarios utilizados en la práctica por dicho alumnado, diseñó un amplio cuestionario (que puede consultarse en Olivo, 2008) en el que se recogían preguntas sobre diversos contenidos relacionados con este tema. Dicho cuestionario fue propuesto a una muestra de 48 estudiantes, y sus resultados fueron publicados en Olivo y Batanero (2007). Muchas de las cuestiones fueron correctamente respondidas por la mayor parte de participantes en el estudio. Sin embargo, también se detectaron muchas dificultades y errores conceptuales, que son descritos a continuación:

- la relación entre el tamaño de la muestra y la precisión del intervalo sólo es comprendida por el 25% de los alumnos;
- el 18% afirma que una variación en el nivel de confianza no afecta a la anchura del intervalo;
- el 21% relaciona el intervalo con posibles valores de la variable estudiada, en vez de con su media poblacional.

El 27% de los alumnos dio una interpretación bayesiana al intervalo, identificando el nivel de confianza con un cierto grado de credibilidad de que la media caiga dentro de

él. Además, cuando se pidió a las estudiantes que calculasen un intervalo de confianza, el 25% eligió mal el número de grados de libertad de la distribución t de Student y un 17% utilizó la tabla de la distribución normal cuando esta no era la apropiada.

En un trabajo con una muestra más amplia, Olivo et al. (2008) analizaron las respuestas proporcionadas por 252 estudiantes de ingeniería empleando otros ítems de su extenso cuestionario. Entre las dificultades encontradas, cabe citar las siguientes:

- un 33% manifiesta no entender la relación entre la anchura del intervalo para la media y los factores que la determinan (por ejemplo, el tamaño de muestra o el coeficiente de confianza);
- entre los estudiantes que habían determinado de forma correcta el intervalo de confianza para la media, un 32.1% daba una interpretación bayesiana al mismo como intervalo de credibilidad;
- hubo quienes mostraron confusión entre los conceptos de “parámetro” y de “estadístico” (un 25%), aportando una interpretación del intervalo con el método del percentil en remuestreo ya que consideraban que el cálculo del intervalo se llevaba a cabo para predecir los posibles valores que podía tomar la media poblacional.

Los autores también analizaron el aspecto procedimental de las respuestas aportadas a las cuestiones propuestas, encontrando los siguientes datos de interés:

- un 41.3% se equivocó al determinar el valor crítico;
- un 18.7% intercambió los grados de libertad de numerador y denominador durante el cálculo de un intervalo para el cociente de varianzas;
- un 11.5% del alumnado confundió las dos desviaciones típicas que intervenían en el cuestionario (muestral y poblacional).

Los autores interpretaron que existía un desconocimiento de la noción de “grados de libertad” ya que el 34.5% se equivocó al determinar el valor adecuado de estos al trabajar con un intervalo de confianza para la media.

Para concluir esta sección dedicada al trabajo con estudiantes universitarios, de máster y futuros investigadores, merece la pena comentar que Morey et al. (2016) resaltaron un alto porcentaje en todos los grupos estudiados de individuos que asignaban al nivel de confianza dos posibles interpretaciones: una, como la probabilidad de que la media pertenezca al intervalo, y otra, como el porcentaje de muestras en que resultaría el mismo intervalo al repetir sucesivos cálculos.

2.4. Comprensión por parte de futuros profesores

Un aspecto que nos parece esencial desde el punto de vista didáctico es la formación de los futuros profesores. Como ya hemos comentado, los contenidos relacionados con el cálculo e interpretación de los intervalos de confianza (en los casos más sencillos) forman parte y son evaluados tanto en una asignatura del Bachillerato de Ciencias Sociales como en la Prueba de Evaluación para el Acceso a la Universidad. Por ello, el hecho de que los profesores tengan un profundo conocimiento y una gran soltura sobre los aspectos que conforman este tema es un requisito imprescindible para garantizar la calidad y corrección de la enseñanza impartida. Dicho conocimiento no puede limitarse a lo puramente estadístico y matemático, sino que ha de verse condimentado con una variedad de recursos educativos, desde libros de texto hasta programas informáticos, pasando por toda clase de recursos tecnológicos. Y, además de lo anterior, es muy importante tanto conocer las dificultades que pueden presentarse como prever los errores más comunes que suelen cometer los alumnos.

Un estudio realizado al hilo de lo comentado fue llevado a cabo por Yáñez y Behar (2009) quienes propusieron algunas de las cuestiones diseñadas por Behar (2001) a 15 estudiantes que se estaban formando como profesores de matemáticas de Educación Secundaria en Colombia. Tras la aplicación del cuestionario, mantuvieron entrevistas con los participantes en el estudio. Estos alumnos habían adquirido previamente una formación (con ayuda de recursos tecnológicos) en estadística y, más en concreto, sobre la estimación por intervalos. Preguntados sobre si el intervalo de confianza debía contener a la media muestral, uno de ellos no negó esta afirmación, pero comentó que la proporción de medias muestrales contenidas en el intervalo de confianza podía ser distinta del nivel de confianza, y que tal circunstancia dependía fundamentalmente de la diferencia entre la media muestral y la poblacional.

En España, existe un Master Universitario de Profesorado que supone un requisito indispensable para los futuros profesores de Educación Secundaria Obligatoria y Bachillerato. Utilizando una muestra de 70 estudiantes de dicho máster universitario que previamente habían cursado el grado en matemáticas, López-Martín et al. (2019a) evaluaron el grado de conocimiento que se puede presuponer sobre este tema a los futuros profesores de matemáticas. Entre las cuestiones propuestas, se solicitó que los estudiantes interpretasen un intervalo de confianza que había sido construido con anterioridad. Entre las respuestas, sólo un 28% proporcionó una interpretación adecuada. Por el contrario, el

11% afirmó que el intervalo calculado debía contener, con seguridad, el verdadero valor del parámetro (visión determinista), error previamente indicado por Olivo (2008). Un 17.8% indicaron que la probabilidad de que el intervalo contuviese al valor concreto del parámetro debía coincidir con el nivel de confianza, con el agravante de interpretar los extremos de este como datos fijos, sin reconocer su carácter aleatorio (dependiendo de la muestra considerada), equivocación previamente indicada en Behar (2001) y Olivo (2008). Inclusive, un participante en el estudio manifestó que el intervalo hacía referencia a la media muestral en vez de a la media poblacional. Acerca de los resultados de este estudio, comentar también que un 41.1% no aportó ninguna interpretación (ni correcta ni incorrecta) sobre el significado del intervalo de confianza.

En otro estudio (López-Martín et al., 2019b), las mismas autoras, utilizando la misma muestra de 70 futuros profesores de matemáticas, analizaron el grado de conocimiento de los participantes acerca de las dificultades que podrían presentar sus propios estudiantes a la hora de resolver un problema sobre intervalos de confianza que había sido propuesto en las pruebas de acceso a la universidad y que había sido resuelto por ellos mismos con anterioridad. El estudio llevó a la descripción de un total de 195 posibles errores que los propios autores clasificaron en cuatro categorías: errores de planteamiento, errores conceptuales, errores procedimentales y errores de interpretación. En la Tabla 2.4.1 describimos algunos de los errores más frecuentemente citados como ejemplos de cada una de las categorías anteriores. Las autoras del estudio interpretaron que el hecho de que los futuros profesores fuesen capaces de describir una amplia variedad de posibles errores que pudiesen cometer sus estudiantes era un claro signo de un conocimiento adecuado sobre el tema tratado. Además, al haber identificado muchos más errores de las categorías procedimental y de interpretación que de las categorías conceptual o de planteamiento, los futuros profesores daban a entender que estas cuestiones podrían entrañar una mayor dificultad, lo cual es extraño porque, hoy en día, este aspecto está prácticamente resuelto mediante el uso de programas informáticos adecuados que versan sobre contenidos estadísticos y que están al alcance de los estudiantes en forma de software libre.

Para terminar esta sección debemos comentar que la mayor parte de los errores encontrados entre los futuros profesores ya habían sido descritos en investigaciones previas. Por un lado, Fidler y Cumming (2005) habían llamado la atención sobre los errores asociados a la interpretación del intervalo de confianza. También pusieron de manifiesto errores de tipificación y de lectura de la tabla de la distribución normal que

habían sido detectados en Ramos et al. (2009), junto con la confusión entre estadístico y parámetro (Harradine et al., 2011).

Tabla 2.4.1. Errores frecuentes clasificados por categorías.

Categoría	Descripción del error y frecuencia de aparición
Errores de planteamiento	<ul style="list-style-type: none"> • Identificación incorrecta de la distribución muestral (7.1%). • Identificación incorrecta del tipo de intervalo: unilateral o bilateral (5.7%).
Errores conceptuales	<ul style="list-style-type: none"> • Confusión entre estadístico y parámetro (28%). • Comprensión incorrecta del teorema central del límite (6%). • Interpretación incorrecta del nivel de confianza (5.7%).
Errores procedimentales	<ul style="list-style-type: none"> • Error en la identificación de datos (22.9%). • Error en la determinación del estadístico (4.3%). • Error en la tipificación (54.3%). • Lectura incorrecta de las tablas estadísticas (30%). • Error en el cálculo de los extremos del intervalo (14.3%). • Error en el uso de símbolos (7.1%).
Errores de interpretación	<ul style="list-style-type: none"> • Interpretación incorrecta del nivel de confianza (17.1%). • Consideración de los extremos del intervalo como fijos (44.3%). • Necesidad de que el parámetro esté dentro del intervalo (8.6%). • Interpretación incorrecta de factores que afectan al intervalo (10%).

Como propuesta para el futuro y tras detectar que los futuros profesores no son lo suficientemente precisos cuando describen los posibles errores que pueden encontrarse por parte de su alumnado, proponemos la conveniencia de aumentar su formación en estos aspectos, incluyendo actividades como las que les fueron propuestas, lo que les llevará, sin duda, a profundizar en su conocimiento acerca de este tema a través de una reflexión metodológica y didáctica.

Para finalizar este capítulo indicamos que no hemos encontrado trabajos completos sobre la comprensión del intervalo por estudiantes de Bachillerato, por lo que nuestros resultados, expuestos en el siguiente capítulo, aportarán información original sobre el tema.

Capítulo 3.

Desarrollo de la investigación y análisis de los resultados

3.1. Introducción

En este capítulo se describe la parte empírica del Trabajo Fin de Máster, que ha consistido en un estudio de evaluación realizado con un grupo de estudiantes de segundo curso de Bachillerato. El estudio es de tipo exploratorio porque, aunque contamos con antecedentes de investigación sobre comprensión del intervalo de confianza, ninguno se centra en estudiantes de Bachillerato.

Se ha utilizado una metodología mixta, incluyendo datos cuantitativos y cualitativos (León y Montero, 2003). Por un lado, contamos con ítems de opción múltiple, de los cuales indicamos las frecuencias con que el alumnado ha elegido cada una de las diferentes opciones de cada ítem, realizando un estudio de porcentajes y del total de respuestas correctas en los seis ítems. Además, hemos incluido en el estudio un problema abierto del que se realiza un análisis cualitativo de las respuestas mediante análisis de contenido (Krippendorf, 2013), una técnica que permite extraer datos de documentos mediante su estudio sistemático. A partir de la misma, mostraremos cómo se definieron las variables y categorías de análisis, clasificando los pasos que se deberían dar para una resolución correcta del ejercicio o atendiendo al tipo de error que el alumnado haya podido cometer.

En lo que sigue se describen las características de los estudiantes que han formado la muestra, el cuestionario utilizado y el análisis realizado de los datos. Seguidamente se exponen los resultados del estudio y las conclusiones que sacamos de dichos resultados.

3.2. Descripción de la muestra

La muestra estuvo formada por 58 estudiantes que cursaban la asignatura de “*Matemáticas Aplicadas a las Ciencias Sociales II*”, matriculados en el segundo curso

del Bachillerato de Ciencias Sociales de dos institutos diferentes. Uno de ellos está situado en la ciudad de Granada, y participaron en él 24 alumnos, matriculados en dos grupos diferentes. Cuando sea necesario, nos referiremos a este instituto como “*instituto A*” y, manteniendo el debido anonimato hacia los participantes en el estudio, se etiquetaron sus cuestionarios con los códigos A1, A2,..., A24. El otro centro está situado en un pueblo del cinturón de Granada, es decir, en la periferia de la propia capital. En este centro, al que llamaremos “*instituto B*”, participaron 34 estudiantes, y sus cuestionarios se etiquetaron desde B1 hasta B34.

Como acabamos de poner de manifiesto, se trata de una muestra intencional, formada por todos los alumnos del segundo curso de Bachillerato, matriculados en la modalidad de Ciencias Sociales, de dos institutos que accedieron a participar en el estudio. Por consiguiente, como hemos comentado, debemos dejar claro desde el principio que nuestro estudio es de tipo exploratorio, de manera que no pretendemos extender las conclusiones obtenidas a una muestra o a una población más amplia.

Los cuestionarios fueron propuestos a lo largo del mes de mayo de 2019. En ese momento, los participantes en el estudio ya habían completado su formación correspondiente a las unidades didácticas de Inferencia Estadística y, más en concreto, sobre intervalos de confianza. Se acercaba el final del curso y los estudiantes se preparaban ya tanto para los exámenes finales como para la Prueba de Evaluación para el Acceso a la Universidad. Es por ello que consideramos que se trataba de un momento idóneo para su propuesta, durante una hora de clase, como una actividad que hiciera reflexionar a los participantes sobre su grado de comprensión del tema.

3.3. Descripción del cuestionario

El cuestionario utilizado se construyó para esta investigación partiendo del análisis de otros cuestionarios empleados en investigaciones previas. Puesto que se trataba de evaluar la comprensión del intervalo de confianza, que es un constructo inobservable (León y Montero, 2003), sus características debían ser inferidas a partir de las respuestas de los alumnos que forman la muestra, teniendo en cuenta que ellos mismos no son plenamente conscientes de su grado de conocimiento.

El cuestionario que construimos consta de seis ítems de opción múltiple y de un problema abierto. La decisión del número y del tipo de ítems se tomó, por un lado, porque los estudiantes de la muestra estaban acostumbrados a los dos formatos en los exámenes que llevan usualmente a cabo y, por otro, por el tiempo disponible para pasar el

cuestionario (que fue de una hora). Para seleccionar los ítems, en primer lugar se decidió la lista de contenidos que deseábamos evaluar y, a continuación, se consideraron tanto el libro de Cruise, Dudley y Thayer (1984) como las investigaciones previas de Behar (2001) y de Olivo (2008). De esta forma, se eligieron algunos ítems de estas referencias que se adaptaban a nuestros fines y, cuando fue necesario, se modificó su formato para adaptarlo al alumnado participante. Como pondremos de manifiesto, también se añadieron algunos ítems de elaboración propia.

Siguiendo recomendaciones de Yaremko, Harari, Harrison y Lynn (2013), se decidió utilizar el mismo número de posibles respuestas en los ítems de opción múltiple (en nuestro caso, cuatro), y se añadió un problema abierto pues considerábamos que nos iba a proporcionar una información más profunda acerca de la comprensión adquirida por los estudiantes de la muestra. A continuación se analiza el contenido de cada uno de los seis ítems que componen la primera parte del cuestionario. Se resaltan en negrita las respuestas correctas.

- Ítem 1.** Se calcula un intervalo de confianza del 90% para la media μ de una población. ¿Cuál o cuáles de las siguientes afirmaciones son verdaderas?
- A. Si se toman muchas muestras, la media muestral \bar{X} caerá dentro del intervalo de confianza aproximadamente el 90% de las veces.
 - B. **El intervalo de confianza es un intervalo de valores calculado a partir de los datos de la muestra. En el 90% de las muestras de una población, el intervalo calculado contiene a la media.**
 - C. El intervalo de confianza es un intervalo con extremos constantes, dentro del cual cae la media poblacional el 90% de las veces.
 - D. La probabilidad de que μ caiga dentro de un intervalo de confianza calculado a partir de una muestra es de 0.90.

Este ítem está adaptado de uno de Cruise, Dudley y Thayer (1984) y es muy similar a otro utilizado por Olivo (2008). El contenido evaluado es el conocimiento de la definición del intervalo de confianza. La solución correcta es el apartado B, donde se indica el carácter aleatorio de los extremos del intervalo y el porcentaje de intervalos, calculados a partir de diferentes muestras, que cubre el valor del parámetro.

El distractor (A) trata de evaluar la confusión, descrita por Behar (2001), que pueda existir entre el alumnado entre las nociones de media muestral y de poblacional, y es incorrecta porque la media muestral \bar{X} siempre cae dentro del intervalo (no el 90% de las veces), pues es obvio que el intervalo está centrado en la media muestral. El distractor (C) es incorrecto y trata de evaluar la confusión citada por López-Martín et al. (2019b) que consiste en suponer que los extremos del intervalo son constantes (es decir, han de ser siempre los mismos y no dependen de la muestra considerada). De hecho estos

extremos son variables aleatorias que dependen de la muestra. El apartado D describe la confusión entre confianza y probabilidad, dando una interpretación bayesiana al mismo (Behar, 2001; Olivo y Batanero, 2007), suponiendo que el coeficiente es la probabilidad de que el parámetro caiga en el intervalo. El intervalo no nos permite calcular la probabilidad de que este contenga la media de la población: la confianza no está depositada en el intervalo, sino en el método de construcción (Olivo et al., 2008).

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

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

Tomado también de Cruise, Dudley y Thayer (1984), este ítem evalúa la comprensión que posee el alumnado acerca de la relación entre el ancho del intervalo de confianza y el tamaño de la muestra. La solución correcta viene dada por el apartado B, porque el intervalo se obtiene sumando y restando a la media muestral un valor crítico (que depende de la confianza) multiplicado por la desviación típica de la distribución muestral de la media. Dicha desviación típica es igual a la de la población dividida por la raíz cuadrada del tamaño de la muestra, en este caso $\sqrt{1000} = \sqrt{10 \cdot 10^2} = 10\sqrt{10}$. Por consiguiente, con el segundo intervalo es 10 veces más estrecho ya que:

$$z_{\alpha/2} \frac{\sigma}{\sqrt{1000}} = z_{\alpha/2} \frac{\sigma}{10\sqrt{10}} = \frac{1}{10} \cdot \left(z_{\alpha/2} \frac{\sigma}{\sqrt{10}} \right)$$

(en realidad, la anchura del intervalo es el doble del número anterior)

El distractor A es incorrecto porque, en la segunda muestra, el nivel de confianza es el mismo: como ya hemos comentado, la confianza no está depositada en el intervalo, sino en el método de construcción de los intervalos (Olivo et al., 2008). Este distractor muestra la confusión entre confianza y amplitud. Igualmente es incorrecto el apartado C pues la precisión viene dada por la mitad del ancho del intervalo y, en la solución correcta hemos explicado que este ancho disminuye, luego la segunda muestra dará un intervalo más preciso. El apartado D igualmente muestra un error ya que en él se confunde la fórmula de la desviación típica muestral y, en vez de dividir por el tamaño, se multiplica.

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

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

Tomado de Cruise, Dudley y Thayer (1984), esta cuestión evalúa la comprensión de la relación entre el nivel de confianza y el ancho del intervalo. En este ítem tenemos dos respuestas correctas. Cuanto mayor es la confianza (por ejemplo, pensemos en el 99%), mayor es la anchura del intervalo de confianza. Por tanto, si bajamos del 90% al 80%, la confianza es menor, y la anchura del intervalo de confianza es menor. Por ello, C y D son respuestas correctas (al ser más pequeño el intervalo de confianza, tendremos más acotado el posible valor del parámetro poblacional μ). También puede interpretarse como la mitad de la anchura (es decir, el “radio” del intervalo), que lleva a la misma solución. En consecuencia, las otras dos opciones, A y B, son incorrectas. Obsérvese que la primera también trata la confusión entre confianza y precisión.

- Ítem 4.** Se calculan intervalos de confianza del 95% con muestras de 100 elementos. ¿Cuál o cuáles de las siguientes afirmaciones son verdaderas?
- A. Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza no cambia.
 - B. **Si la desviación estándar de la población disminuye, la anchura del intervalo de confianza disminuye.**
 - C. Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza disminuye.
 - D. **Si la desviación estándar de la población aumenta, la anchura del intervalo de confianza aumenta.**

Adaptado de Behar (2001), esta cuestión sirve para analizar la relación entre la anchura del intervalo y el valor de la varianza de la población. En este ítem hay también dos opciones correctas: B y D. La anchura del intervalo de confianza es $2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$. Por tanto, si la desviación estándar disminuye, la anchura del intervalo de confianza también disminuye (ídem si aumenta). Los distractores A y C evalúan la falta de comprensión de la relación entre la desviación estándar y la anchura del intervalo.

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

Este ítem es de construcción propia y evalúa la comprensión acerca del hecho de que el ancho del intervalo no varía cuando se traslada la media puesto que el intervalo está centrada en ella. Las respuestas correctas son tanto C como D. La respuesta C es verdadera porque, al ser aleatorios los extremos del intervalo, 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 sea muy alto, lo cual es un error señalado por Olivo (2008). La respuesta D también es correcta puesto que la anchura del intervalo de confianza es $2z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, expresión que no depende de la media muestral. 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 y, por tanto, esta no puede ser uno de los extremos. El distractor A es la negación del apartado D y, por tanto, es incorrecta.

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

Esta cuestión también es de construcción propia. Con ella, pretendemos determinar si el alumnado comprende que el intervalo de confianza sirve para estimar un valor desconocido de un parámetro. Por tanto, si este es conocido, no tiene ningún sentido aplicar esta técnica, lo que significa que la respuesta correcta es la D. Los distractores A y B son incorrectos porque no tiene sentido calcular un intervalo cuando se conoce la media de la población, y la opción C es falsa porque, en ocasiones, el intervalo de confianza no contiene al verdadero valor del parámetro.

La segunda parte del cuestionario está constituida por el siguiente problema abierto.

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

Este problema abierto se tomó de la prueba de acceso a la universidad en Andalucía en el año 2018. El alumnado de Bachillerato (e incluso el de primer curso de grado) podría resolver la cuestión planteada aplicando la fórmula de cálculo del intervalo de confianza para la media (con varianza conocida) o bien deducirla. Resumidamente, los pasos que se deben seguir son los siguientes:

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

Para resolver correctamente esta cuestión, el alumnado debe recordar la fórmula del intervalo de confianza para la media así como la de la desviación típica de la distribución muestral. Además, debe saber leer la tabla de la distribución normal $N(0,1)$ para deducir

de ella el valor crítico correspondiente. En la Tabla 3.3.1 mostramos un resumen del contenido evaluado por nuestro cuestionario.

Tabla. 3.3.1. Contenido evaluado por el cuestionario.

Contenido	I1	I2	I3	I4	I5	I6	P1
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	
Construcción de un intervalo de confianza							x
Fórmula del intervalo de confianza							x
Desviación típica de la distribución muestral							x
Lectura de valores críticos en la tabla de la distribución N(0,1)							x

3.4. Resultados en los ítems de opción múltiple

En la presente sección se presentan los resultados obtenidos en los ítems descritos en la sección anterior, indicando las frecuencias de aparición de cada apartado en las respuestas del alumnado y el porcentaje, respecto del total de 58 alumnos participantes, que eligieron cada distractor. Es importante aclarar que, puesto que algunos alumnos marcaron más de una opción, el porcentaje total de respuestas puede sumar más del 100%.

3.4.1. Resultados en el ítem 1. Definición de intervalo de confianza

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

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

- D. La probabilidad de que μ caiga dentro de un intervalo de confianza calculado a partir de una muestra es de 0.90.

En la Tabla 3.4.1.1 se presentan los resultados obtenidos en este ítem, en que la opción correcta fue elegida por más de la mitad de los alumnos, lo que puede ser interpretado como que, aparentemente, el alumnado ha comprendido la definición del intervalo de confianza. El error más frecuente fue elegir el distractor D, que consiste en confundir confianza y probabilidad, dando una interpretación bayesiana al intervalo de confianza. Este error fue descrito por Behar (2001) en la mitad de sus estudiantes y el 40% de los expertos, y también por Olivo y Batanero (2007) en el 27% de estudiantes de ingeniería. Los autores indican que no es posible calcular la probabilidad de que el intervalo de confianza contenga a la media de la población: la confianza no está depositada en el intervalo sino en el método de construcción de los intervalos (Olivo et al., 2008). Son pocos los estudiantes que responden utilizando los otros distractores.

Tabla 3.4.1.1. Resultados obtenidos en el ítem 1.

	Frecuencia	Porcentaje
A	1	1.7
B	32	55.2
C	2	3.4
D	22	37.9
En blanco	2	3.4

3.4.2. Resultados en el ítem 2. Relación entre ancho del intervalo y tamaño de la muestra

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

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

En la Tabla 3.4.2.1 se presentan los resultados al ítem. La opción correcta fue elegida únicamente por una cuarta parte del alumnado, lo que pone de manifiesto una importante deficiencia en la comprensión manifestada por el alumnado sobre el hecho de que aumentar el tamaño de la muestra, fijando todos los demás factores, repercute en una disminución de la anchura del intervalo. La respuesta mayoritaria fue la expresada por el distractor A, elegida por una tercera parte de los estudiantes. Este error puede estar asociado con el hecho de que el alumnado comprende la conveniencia de tener muestras de tamaños cada vez mayores pues parece ser que estas nos aportan más información que

las muestras pequeñas. Así, el alumnado ha traducido esta situación con la idea de tener una mayor confianza en el intervalo, sin atender al hecho de que el nivel de confianza no ha variado, lo que ya fue comentado por Morey et al. (2016), quienes indicaron que no se comprende que la amplitud también depende del coeficiente de confianza.

Tabla 3.4.2.1. Resultados obtenidos en el ítem 2.

	Frecuencia	Porcentaje
A	19	32.8
B	15	25.9
C	10	17.2
D	12	20.7
En blanco	2	3.4

Por otro lado, el distractor D ha sido elegido por la quinta parte del alumnado, posiblemente asociando la idea de que un aumento en el número de datos debe repercutir en un aumento en el ancho del intervalo de confianza (como si todos los datos debieran, de una forma u otra, engrosar la anchura del intervalo). Finalmente, una sexta parte de los encuestados ha respondido que el aumento de la muestra no conlleva una diferencia en la precisión del intervalo (distractor C), manifestando un desconocimiento o bien sobre la noción de precisión del intervalo o bien asociando esta palabra con la “*precisión*” de los números empleados, la cual, en principio, atendiendo al enunciado del ítem, pudiera no cambiar (Yáñez y Behar, 2009). Los errores de interpretación anteriores ya fueron encontrados por Fidler y Cumming (2005) en el 20% de estudiantes que pensaban que el ancho del intervalo de confianza se incrementaría si se aumentara el tamaño de la muestra, un 29% creía que no cambiaría el ancho y el 36% no supo si había o no relación entre dichos conceptos. Olivo y Batanero (2007) indican que, de los 48 estudiantes que participaron en su estudio, solo el 25% comprendía la relación entre la precisión y el tamaño de la muestra, si bien este porcentaje era del 67% entre los 252 estudiantes de ingeniería analizados en Olivo et al. (2008).

3.4.3. Resultados en el ítem 3. Precisión y confianza

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

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

En la Tabla 3.4.3.1 se presentan los resultados al ítem, donde las opciones correctas eran dos, C y D, pues una mayor precisión en el intervalo de confianza es sinónimo de

una disminución en su anchura. Sin embargo, casi dos terceras partes del alumnado ha asociado una reducción en el nivel de confianza con una reducción en la precisión, lo cual pone de manifiesto que los estudiantes asocian claramente estos conceptos (posiblemente se trate de un error derivado del lenguaje usual, pues depositamos más confianza en aquellos procesos que anuncian mayor precisión en sus resultados, de acuerdo a Yáñez y Behar, 2009). La segunda opción con mayor frecuencia es el apartado C que, a pesar de ser verdadero, ha sido seleccionado por menos de una cuarta parte del alumnado.

Tabla 3.4.3.1. Resultados obtenidos en el ítem 3.

	Frecuencia	Porcentaje
A	36	62.1
B	7	12.1
C	13	22.4
D	7	12.1
En blanco	1	1.7

Si descartamos que una disminución en el nivel de confianza pueda traducirse en un intervalo de igual anchura y precisión, la respuesta A es contraria a C y la respuesta B es la opuesta de D. Es por ello que resulta curioso que el alumnado no se haya dado cuenta de que esta cuestión debía tener dos respuestas correctas, por ser dos de ellas opuestas de las otras dos. Sólo cuatro alumnos han elegido dos posibles respuestas a la vez a este ítem y, en concreto, los cuatro se han equivocado eligiendo las opciones A y B.

Lo anterior también significa que ningún alumno ha escogido, a la vez, las respuestas correctas C y D. Investigaciones previas han mostrado la presencia de este error de interpretación. Por ejemplo, un 33% del alumnado de ingeniería encuestado en Olivo et al. (2008) no comprende la relación entre el ancho del intervalo de confianza y el coeficiente de confianza. Por otro lado, Yáñez y Behar (2009) pusieron de manifiesto numerosas dificultades entre las cuales estaba el efecto del nivel de confianza sobre la precisión de la estimación. Relacionado con la posible variación en el nivel de confianza, en Behar (2001) se describe cómo los estudiantes asumían que altos niveles de confianza llevan siempre a intervalos más estrechos cuando el resto de los datos no varía.

3.4.4. Resultados en el ítem 4. Relación entre amplitud y varianza de la población

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.**

En la Tabla 3.4.4.1 se presentan los resultados al ítem en que había dos posibles respuestas correctas, a saber, los apartados B y D, cuyos enunciados daban a entender su equivalencia (si entendemos que la anchura del intervalo debe aumentar cuando aumenta la desviación estándar, entonces también debe disminuir la primera cuando disminuye la segunda). Por ello, resulta extraño que no haya un mismo número de alumnos que hayan seleccionado, a la vez, estos apartados. Es más, aunque las frecuencias de los apartados B y D resultan ciertamente similares (el apartado B ha sido seleccionado 14 veces y el D, 17 veces), estas respuestas no provienen de los mismos estudiantes: sólo hay 5 alumnos que responden correctamente, a la vez, los apartados B y D (todos ellos del instituto A).

Tabla 3.4.4.1. Resultados obtenidos en el ítem 4.

	Frecuencia	Porcentaje
A	3	5.2
B	14	24.1
C	27	46.6
D	17	29.3
En blanco	2	3.4

Aparte de estos cinco alumnos, no hay ningún otro estudiante que se haya decantado por seleccionar, a la vez, dos posibles respuestas. La opción mayoritaria en este ítem ha sido el distractor C, el cual ha sido seleccionado por casi la mitad de los estudiantes. No está claro de dónde puede provenir este error pues es obvio que la anchura del intervalo de confianza es directamente proporcional a la desviación estándar de la población. Este error no está descrito en investigaciones anteriores (especialmente en el caso de que la varianza poblacional sea conocida, caso estudiado en el Bachillerato) y supone una aportación original de nuestro trabajo.

3.4.5. Resultados en el ítem 5. Amplitud y posición de la media

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

- A. El ancho del intervalo de confianza depende de la media muestral \bar{X} .
- B. La media muestral \bar{X} es un extremo del intervalo de confianza.
- C. **Cabe la posibilidad de que el verdadero valor de μ no esté contenido en el intervalo.**
- D. **El ancho del intervalo no depende de la media muestral.**

En la Tabla 3.4.5.1 se presentan los resultados al ítem, en el que había dos opciones correctas, que son los apartados C y D, pero sólo 7 alumnos (todos ellos del instituto B)

eligen conjuntamente las dos. Son más (36,2%) los que responden que el intervalo podría no contener a la media muestral que los que indican que el ancho no depende de la posición de la media.

Tabla 3.4.5.1. Resultados obtenidos en el ítem 5.

	Frecuencia	Porcentaje
A	29	50
B	16	27.6
C	21	36.2
D	17	29.3
En blanco	1	1.7

Respecto de los errores, la mitad piensa que el ancho del intervalo depende de la media muestral, lo que implica que no comprenden la fórmula de cálculo. Este error no se ha descrito por otros investigadores y constituye una aportación de nuestro trabajo. Está claro que el alumnado comprende que el intervalo de confianza depende de la media muestral, que es una parte esencial de su expresión, pero globalmente desconoce que el ancho de dicho intervalo no depende de la misma. Igualmente indican no comprender la fórmula quienes eligen el distractor B, pensando que la media está en un extremo. Por tanto, ni comprenden la fórmula ni el concepto de extremo del intervalo. Echándole un poco de imaginación, somos proclives a pensar que este error proviene de la expresión:

$$\left[\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right],$$

donde la media muestral está situada en un extremo (concretamente, a la izquierda) de la expresión algebraica empleada.

3.4.6. Resultados en el ítem 6. Uso del intervalo de confianza

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

- A. Un posible intervalo de confianza para μ es [6.25, 8.4].
- B. Un posible intervalo de confianza para μ es [6.75, 8.4].
- C. Cualquier intervalo de confianza para μ que calculemos, asociado a una muestra aleatoria cualquiera, debe contener al valor 6.5.
- D. **En este contexto, no tiene sentido determinar el intervalo de confianza ya que se conoce la altura media de todos los pinos.**

En la Tabla 3.4.3.1 se presentan los resultados al ítem. La opción correcta D fue elegida por el 36.2% del alumnado, resultado que nos parece muy pobre porque esta cuestión hace referencia directa a la justificación de los contenidos teóricos en sí mismos: determinamos un intervalo de confianza cuando deseamos hacer una estimación de un valor desconocido, de tal modo que si dicho valor es conocido, no tiene sentido llevar a

cabo un proceso (en ocasiones, costoso y complicado) para su estimación.

Sin embargo, la opción mayoritaria ha sido el distractor C, hecho que posiblemente tenga dos causas: por un lado, el alumnado centra su estudio en el proceso mecánico de cálculo del intervalo de confianza, sin prestar demasiada atención a su interpretación y al sentido que se atribuye a la necesidad de calcularlo (estimación de un parámetro desconocido); por otro lado, fruto de este proceso mecánico, el alumnado ha asimilado que la media muestral pertenece siempre al intervalo de confianza (de hecho, es su centro), pero confunde media muestral (\bar{X}) y media poblacional (μ) a pesar de haberse explicitado esta última en el enunciado. Error señalado tanto por Behar (2001) como por Olivo (2008).

Tabla 3.4.6.1. Resultados obtenidos en el ítem 6.

	Frecuencia	Porcentaje
A	10	17.2
B	5	8.6
C	23	39.7
D	21	36.2
En blanco	1	1.7

Los distractores A y B proporcionaban sendos posibles intervalos de confianza, de los cuales uno contenía al valor 6.5 y el otro no, si bien ninguno de ellos estaba centrado en dicho valor. A priori pensábamos que esta circunstancia (que ninguno estuviese centrado en el valor 6.5) disuadiría a cualquier estudiante que confundiese las nociones de media muestral y media poblacional. Este error de confusión entre los valores poblacionales y muestrales está presente en investigaciones previas como Olivo et al. (2008), donde los autores describen que un 25% de sus 252 estudiantes de ingeniería mostró confusión entre las ideas de estadístico y parámetro. Esta confusión no sólo tiene lugar con la media, sino que en el mismo estudio se pone de manifiesto que un 11.5% del alumnado confunde las desviaciones típicas de la población y de la distribución muestral.

Sin embargo, la principal razón por la que decidimos incluir este ítem en el cuestionario era para saber si comprendían que, conociendo la media de la población, calcular el intervalo de confianza carece de sentido, hecho del que no tenemos constancia en investigaciones previas y que constituye otra aportación novedosa al presente trabajo.

3.4.7. Síntesis de los resultados obtenidos en los ítems de opción múltiple

Para valorar globalmente la comprensión mostrada en la primera parte del cuestionario, se incluye a continuación un diagrama de barras en el que se describe el

número de respuestas correctas e incorrectas de cada uno de los 58 participantes en el estudio. Aunque sólo había seis ítems, tres de ellos poseían dos respuestas correctas por lo que, en total, podían seleccionarse nueve respuestas correctas. En la Figura 3.4.7.1 presentamos la distribución del número de respuestas correctas, donde ningún alumno consiguió darlas todas, o al menos ocho, siendo lo más frecuente tan sólo dos respuestas correctas. Muy pocos estudiantes lograron cinco o más respuestas correctas que sería un poco más que el valor medio esperado del número de respuestas correctas. Todo ello indica una gran dificultad conceptual del tema para estos estudiantes. Estos datos se desglosan con mayor detalle en el Anexo I.

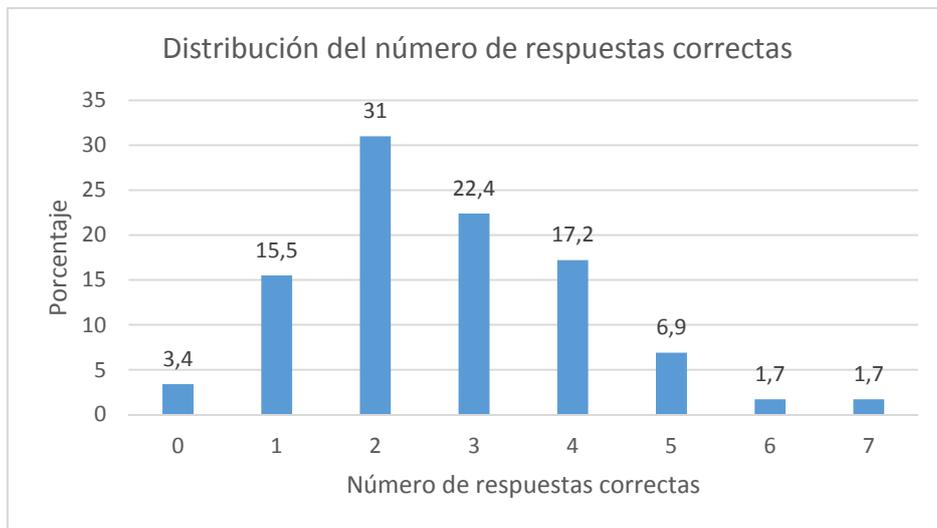


Figura 3.4.7.1. Distribución del número de respuestas correctas.

El número medio de respuestas correctas fue de 2.7, mientras que el 50% central de la distribución se situó entre 2 y 4 respuestas correctas (véase la la Figura 3.4.7.2).

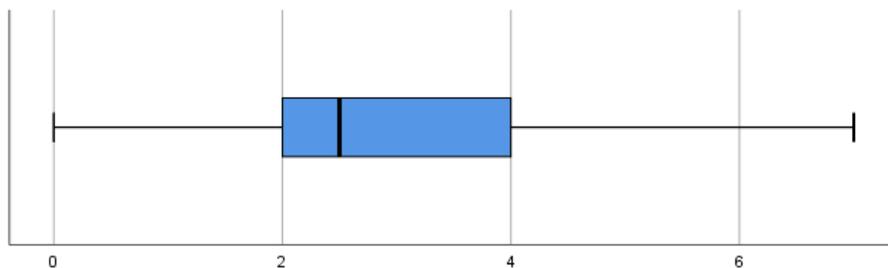


Figura 3.4.7.2. Diagrama de cajas y bigotes de la distribución del número de respuestas correctas (la línea vertical central representa la mediana, que fue de 2.5).

Para completar el análisis, a continuación describimos los principales conflictos semióticos observados en la elección de los distractores por parte de los estudiantes, que han sido los siguientes. En primer lugar, se describen una serie de conflictos identificados en investigaciones previas (casi todos ellos han sido descritos en López-Martín et al.,

2019a, como errores cometidos, y en López-Martín et al., 2019b, como posibles errores que puede cometer el alumnado, por lo que no los mencionamos en cada apartado).

- Considerar el intervalo de confianza con extremos constantes (3.4% en ítem 1), error citado por Behar (2001) y Olivo (2008).
- Confundir confianza con probabilidad de que el parámetro caiga en el intervalo (37.9% en el ítem 1; 39.7% en el ítem 6), dando una interpretación bayesiana al intervalo, error citado por Behar (2001), Morey et al. (2016) y Olivo (2008). Comparando con investigaciones previas sobre este tema, cabe citar que López-Martín et al. (2019a) encontraron un 11% de estudiantes de Máster que dieron una interpretación determinista del intervalo de confianza, suponiendo que el mismo incluía con seguridad el valor del parámetro.
- Considerar que una reducción en el nivel de confianza implica una reducción en la precisión (62.1% en el ítem 3), conflicto descrito en Yáñez y Behar, 2009.
- En la línea de los dos errores anteriores, ligar la precisión sólo a la confianza (17.2% en el ítem 2).
- Pensar que el ancho del intervalo aumenta cuando crece el tamaño de la muestra (20.7% en el ítem 2), error descrito en Fidler y Cumming (2005) entre el 20% de sus estudiantes de psicología (porcentaje muy parecido al nuestro).
- Considerar que el intervalo de confianza será más ancho cuando se reduce el nivel de confianza (12.1% en el ítem 3), conflicto descrito en Behar (2001) –en sentido contrario: cuando sube el nivel de confianza, baja la anchura del intervalo– y en Yáñez y Behar (2009). De hecho, la dificultad de entender la relación entre nivel de confianza y anchura del intervalo también fue expuesta en Morey et al. (2016), Olivo (2008) y Olivo y Batanero (2007).
- Confundir las nociones de estadístico y parámetro (17.2% en ítem 6), error señalado en Behar (2001), Harradine et al. (2011), Olivo (2008) y Olivo et al. (2008) en el 25% de sus estudiantes de ingeniería.

Además, hemos encontrado los siguientes, no descritos hasta ahora y que constituyen una aportación original de nuestro trabajo.

- Considerar que la media muestral se sitúa en un extremo del intervalo de confianza (27.6% en el ítem 5).

- No relacionar las ideas equivalentes de “intervalo más estrecho” con “intervalo más preciso” (en el ítem 3, ningún alumno ha elegido, a la vez, las dos opciones correctas y equivalentes).
- Suponer que algunos intervalos no contienen a la media muestral (1.7% en el ítem 1), lo que demuestra que el alumnado no comprende que el centro del intervalo se sitúa en dicha media.
- Para la misma confianza, suponer que el tamaño de muestra aumenta las posibilidades de que el parámetro caiga en el intervalo (32.8% en el ítem 2)
- Considerar que la anchura del intervalo de confianza depende de la media muestral (ítem 5, 50%).
- Suponer que la anchura del intervalo disminuye cuando aumenta la desviación estándar de la población (46.6% del ítem 4).
- En la línea del anterior, manifestar ausencia de comprensión acerca de cómo la desviación típica influye en la anchura del intervalo de confianza (51.8% del ítem 4).
- Ausencia de comprensión del hecho de que, conocida la media poblacional, no tiene sentido calcular el intervalo de confianza, es decir, no entender el verdadero objetivo de la Inferencia Estadística (en el ítem 6, el 63.8% no seleccionan la opción adecuada que permite inferir este conocimiento).
- En la línea del anterior, considerar que, toda vez conocido el valor de la media poblacional, el intervalo de confianza debe contener dicho valor (39.7% en el ítem 6).

3.5. Resultados en el problema abierto.

Desde el comienzo del planteamiento del cuestionario nos pareció interesante incluir un problema en el que el alumnado pudiese expresar abiertamente la forma en la que utiliza sus conocimientos para afrontar una situación práctica, alejándose del rigor establecido por una cantidad limitada de opciones posibles. De esta forma, teniendo en cuenta las pruebas de acceso a la universidad planteadas el curso anterior en Andalucía, seleccionamos el siguiente problema, básicamente idéntico a uno de los planteados en ellas.

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

En el apartado 3.3 hemos descrito una posible solución a los interrogantes que aquí se plantean. A partir de ella, para analizar las respuestas que han proporcionado los

estudiantes, nos hemos fijado en el proceso que deben seguir los estudiantes.

3.5.1. Cálculo del intervalo

En este apartado nos centramos en la parte del cálculo del intervalo de confianza. De acuerdo a la solución prevista, deberían haberse seguido los pasos que se describen a continuación, cada uno de los cuales constituye una práctica matemática que implica diferentes objetos.

- P1.** Recordar (o deducir) y escribir la expresión teórica de la fórmula de cálculo del intervalo de confianza para la media poblacional μ , con objeto de determinar los datos que intervendrán en su cálculo, y que viene dado por:

$$\left[\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad \text{o bien} \quad \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

- P2.** Determinar el valor de $1 - \alpha/2$ (en tablas de colas a la derecha, calcular $\alpha/2$), siendo α el nivel de significación, es decir, $1 - \alpha = 0.95$ el nivel de confianza.

$$\alpha/2 = 0.025.$$

- P3.** Representar adecuadamente la distribución normal estándar, donde se observe dicha distribución junto con una región centrada en cero conteniendo un área de 0.95, y dos colas laterales con áreas respectivas de 0.025 (en realidad, este paso no es estrictamente necesario, aunque pensamos que sí es conveniente).

- P4.** Hacer una lectura adecuada de la tabla de la distribución normal estándar, determinando el valor crítico $z_{\alpha/2}$, que tomaba el valor 1.96 al 95% de confianza.

$$z_{\alpha/2} = 1.96.$$

- P5.** Sustituir los valores de \bar{x} , $z_{\alpha/2}$, σ y n en la expresión teórica del intervalo de confianza, escribiendo este con los números adecuados.

$$\left[75 \pm 1.96 \frac{7}{\sqrt{100}} \right].$$

- P6.** Realizar correctamente las operaciones que se indican, determinando primeramente la cantidad que se ha de sumar y de restar (el radio del intervalo).

$$\left[75 \pm 1.96 \frac{7}{\sqrt{100}} \right] = [75 \pm 1.372] = [73.628, 76.372].$$

- P7.** Indicar, finalmente, la expresión del intervalo de confianza, especificando los números anteriores o bien redondeando adecuadamente. El intervalo estaba pensado para que no hiciese falta redondear, y pudiese ser obtenido directamente con la ayuda de la calculadora, que proporcionaba un intervalo cuyos extremos

venían dados por números de tres cifras decimales. Sin embargo, cabía la posibilidad de que el alumnado redondease dichos extremos con una o dos cifras decimales y se daba la circunstancia de que, en ambos casos, un extremo debía ser redondeado por defecto y el otro, por exceso, con lo que se obtendría uno de los resultados mostrados en la Tabla 3.5.1.1.

Tabla 3.5.1.1. Posibles soluciones según redondeo del intervalo.

Redondeo de los extremos del intervalo	Expresión final del intervalo
Sin redondeo	[73.628 , 76.372]
Redondeo a las centésimas	[73.63 , 76.37]
Redondeo a las décimas	[73.6 , 76.4]

En la Tabla 3.5.1.2 mostramos el análisis semiótico realizado de la solución al problema, descomponiendo la solución en prácticas matemáticas elementales y, para cada una de ellas, identificando los objetos y procesos matemáticos involucrados. De este modo se muestra la complejidad de la construcción del intervalo ya que se deben recordar muchos conceptos y propiedades diferentes, así como los símbolos asociados y los procedimientos de cálculo de los mismos.

Tabla 3.5.1.2. Análisis semiótico de la solución al problema.

Práctica matemática	Objetos y procesos involucrados
P1. Fórmula de cálculo del intervalo $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$	<ul style="list-style-type: none"> - Recordar los conceptos de intervalo y extremos. - Diferenciar la media de la muestra (\bar{x}) de la media de la población (μ) y recordar que el intervalo está centrado en la media muestral. - Recordar que la amplitud del intervalo viene dada por el valor crítico multiplicado por la varianza de la distribución muestral. - Recordar que la varianza de la distribución muestral es $\frac{\sigma}{\sqrt{n}}$. - Recordar que el valor crítico se obtiene a partir del nivel de confianza y la distribución normal estándar.
P2. Determinar el percentil requerido para leer la tabla	<ul style="list-style-type: none"> - Recordar los conceptos de nivel de confianza y valor crítico. - Determinar el valor $1 - \alpha/2$ (o $\alpha/2$ dependiendo del tipo de tabla) a partir del nivel de confianza.
P3. Representación gráfica	<ul style="list-style-type: none"> - Recordar la forma de la distribución normal y alguna de sus propiedades como simetría central y asíntota horizontal.
P4. Lectura de la tabla	<ul style="list-style-type: none"> - Lectura de la tabla, posiblemente con necesidad de interpolación.
P5. Sustitución de datos	<ul style="list-style-type: none"> - Identificar los datos necesarios en la fórmula del intervalo, recordando los conceptos requeridos.
P6. Cálculos	<ul style="list-style-type: none"> - Realizar los cálculos requeridos.
P7. Redondeo	<ul style="list-style-type: none"> - Recordar los criterios de redondeo y aplicarlos.

En la Tabla 3.5.1.3 describimos las frecuencias absolutas y los porcentajes totales observados en los diferentes pasos que consideramos imprescindibles para llevar a cabo una correcta resolución, mientras que en la Tabla 3.5.1.4 describimos los pasos que, aun

siendo convenientes, no son imprescindibles (hay estudiantes que han resuelto correctamente el problema sin llevar a cabo alguno de estos pasos).

Tabla 3.5.1.3. Frecuencia de estudiantes que realizan de los pasos imprescindibles para una correcta resolución de la cuestión abierta.

Descripción del proceso seguido	Frecuencia	Porcentaje
Determina $\alpha/2$	30	51.7
Determina $z_{\alpha/2}$	26	44.8
Escribe el intervalo de confianza con números	23	39.7
Realiza correctamente las operaciones	17	29.3
Indica la expresión final del intervalo de confianza:		
• Sin redondeo	14	24.1
• Redondeando con una o dos cifras decimales	3	5.2

La Tabla 3.5.1.3 pone de manifiesto que apenas la mitad del alumnado es capaz de determinar el valor de $\alpha/2$, cuestión que consideramos muy preocupante pues pone de manifiesto que el alumnado desconoce el significado más esencial del proceso que va a tratar de desarrollar. Entre los que determinan correctamente el valor de $\alpha/2$ abundan quienes llegan a encontrar el valor crítico $z_{\alpha/2}$, si bien hay dos alumnos que describen dicho valor crítico sin haber explicitado previamente el valor de $\alpha/2$ ni haber hecho un dibujo adecuado de la distribución normal.

Tabla 3.5.1.4. Frecuencia de estudiantes que realizan los pasos no estrictamente necesarios.

Descripción del proceso seguido	Frecuencia	Porcentaje
Escribe la expresión teórica del intervalo de confianza	12	20.7
Realiza un dibujo adecuado	13	22.4

Es interesante observar que, entre los estudiantes que llegan a determinar correctamente el valor crítico $z_{\alpha/2}$, sólo tres de ellos no son capaces de escribir correctamente el intervalo de confianza con sus valores adecuados, uno porque no llega a determinar que el tamaño muestral es $n = 100$, y otros dos porque escriben mal la fórmula del intervalo de confianza. Esta criba lleva a que únicamente 23 alumnos escriban correctamente la expresión del intervalo de confianza empleando los datos adecuados (tras identificar en el enunciado los valores de las variables que intervienen), de los cuales seis se equivocan al realizar las operaciones necesarias. Finalmente, quienes han escrito el intervalo de confianza con números y han realizado correctamente las operaciones (17 estudiantes), han llegado a la solución correcta, habiendo 14 alumnos que la expresan con tres cifras decimales y tres alumnos que emplean correctamente el redondeo. Quiere esto

decir que únicamente el 29.3% del alumnado participante en el estudio ha llegado a la solución correcta de la cuestión abierta.

La Tabla 3.5.1.4 pone de manifiesto que poco más de una quinta parte del alumnado lleva a cabo dos tareas muy importantes de cara a la comprensión del proceso que se está llevando a cabo en la práctica: quienes escriben la expresión teórica del intervalo de confianza para la media y quienes se ayudan de una representación gráfica de la distribución normal para determinar el valor crítico que se ha de emplear. Si nos fijamos, de los 13 que hacen el dibujo correctamente, diez llegan a la resolución correcta del apartado, lo cual parece indicar que quienes dibujan tienen una visión global adecuada del proceso que están llevando a cabo. Es más, los tres que se equivocan lo hacen a través de errores que parecen razonablemente subsanables: uno no llega a encontrar n , otro emplea un valor equivocado para $z_{\alpha/2}$ y el último escribe el intervalo de confianza correctamente con sus números adecuados pero no lo calcula en la práctica.

La discusión anterior respecto de los pasos a seguir para una correcta resolución es pertinente de cara a determinar el número y el porcentaje de estudiantes que han resuelto la cuestión abierta de forma correcta, incorrecta o parcialmente correcta. Entendemos que una resolución correcta es aquella fruto de la cual se llega a una expresión final correcta del intervalo solicitado, habiendo redondeado sus extremos o no. Igualmente, entendemos que una resolución parcialmente correcta es aquella en la que el alumnado llega a escribir el intervalo de confianza con sus números apropiados y, sin embargo, por una u otra razón (posiblemente un fallo en las operaciones), no se llega a la expresión final del intervalo de confianza. En este sentido, en la Tabla 3.5.1.5 podemos encontrar los resultados finales donde aproximadamente el 40% de los estudiantes llega al menos a la solución parcialmente correcta.

Tabla 3.5.1.5. Resultados finales de la primera parte de la cuestión abierta.

Nivel de resolución	Frecuencia	Porcentaje
Resolución correcta	17	29.3
Resolución parcialmente correcta	6	10.4
Resolución incorrecta	17	29.3
En blanco	18	31

3.5.2. Errores más frecuentes en el cálculo del intervalo

Los errores que hemos podido detectar tras desarrollar el estudio han sido ya descritos en investigaciones anteriores y era razonable prever que se iban a presentar. En la Tabla 3.5.2.1 describimos los más frecuentes. En esta tabla indicamos porcentajes, pero

hay que tener en cuenta que algunos alumnos han cometido varios errores a la vez.

Tabla 3.5.2.1. Errores más frecuentes.

Descripción del proceso seguido	Frecuencia	Porcentaje
Determina incorrectamente $\alpha/2$	1	1.7
Determina incorrectamente $z_{\alpha/2}$	7	12.1
Error en alguna de las operaciones	5	8.6
Suma y resta antes de multiplicar	4	6.9
Redondea mal el extremo inferior del intervalo	5	8.6
Redondea mal el extremo superior del intervalo	4	6.9

Los errores que más llaman la atención son aquellos que, una vez producidos, indican objetivamente al alumnado que una parte del razonamiento es incorrecta, y que debería volver a replantearse los argumentos anteriores. En este sentido, es especialmente llamativo que el alumnado haya sumado y restado de la media muestral el valor crítico en la expresión $\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$ antes de multiplicar por la desviación típica de la distribución muestral, dando lugar al intervalo generado por $(\bar{x} \pm z_{\alpha/2}) \cdot \frac{\sigma}{\sqrt{n}}$. Al evaluar esta expresión algebraica, se obtiene el intervalo:

$$\left[(\bar{x} \pm z_{\alpha/2}) \cdot \frac{\sigma}{\sqrt{n}} \right] = \left[(75 \pm 1.96) \cdot \frac{7}{\sqrt{100}} \right] = [51.128, 53.872],$$

el cual es claramente incorrecto pues contiene a la media muestral $\bar{x} = 75$. Un ejemplo de este error puede observarse en la Figura 3.5.2.1. Obsérvese que el participante A21 también había determinado incorrectamente el valor crítico.

Figura 3.5.2.1. Respuesta donde A21 suma y resta antes de multiplicar.

El error más frecuente (siete alumnos, 12.1 %) es determinar incorrectamente el valor crítico $z_{\alpha/2}$ debido a lectura incorrecta de la tabla. En nuestra opinión, esta es una circunstancia extraña porque se había planteado determinar el valor crítico que más se usa en la práctica (correspondiente al 95% de confianza), donde los estudiantes debieran recordar que el valor crítico era 1.96 ya que lo deberían haber determinado multitud de veces en clase. Curiosamente, varios de estos estudiantes no habían confeccionado una representación gráfica, por lo que no podían apoyarse en la misma para hacer una lectura adecuada de la tabla de la distribución normal. En la Figura 3.5.2.2 mostramos algunos de estos errores: por ejemplo, donde el participante A21 había representado incorrectamente el nivel de confianza, confundiénolo con el correspondiente a la cola a

la izquierda, es decir, no comprende que se debe tomar el 95% central de los valores de la distribución normal. El alumno A24 confunde la representación simbólica del valor crítico ($z_{\alpha/2}$) con el número concreto que ha de buscar en la tabla para determinar dicho valor crítico (que, en la práctica, es la media aritmética entre el nivel de confianza y la unidad). Sin embargo, en vez de tomar 0.95 como nivel de confianza, emplea 0.75, tomada quizá de la media muestral $\bar{x} = 75$, indicando una proporción $75/100 = 0.75$. Este error le lleva a un valor equivocado del valor crítico. Además, el participante B29 resta los valores 0.95 (nivel de confianza) y 0.025 (cola derecha), por lo que determina un valor equivocado (0.925) para buscarlo en la tabla de la distribución normal.

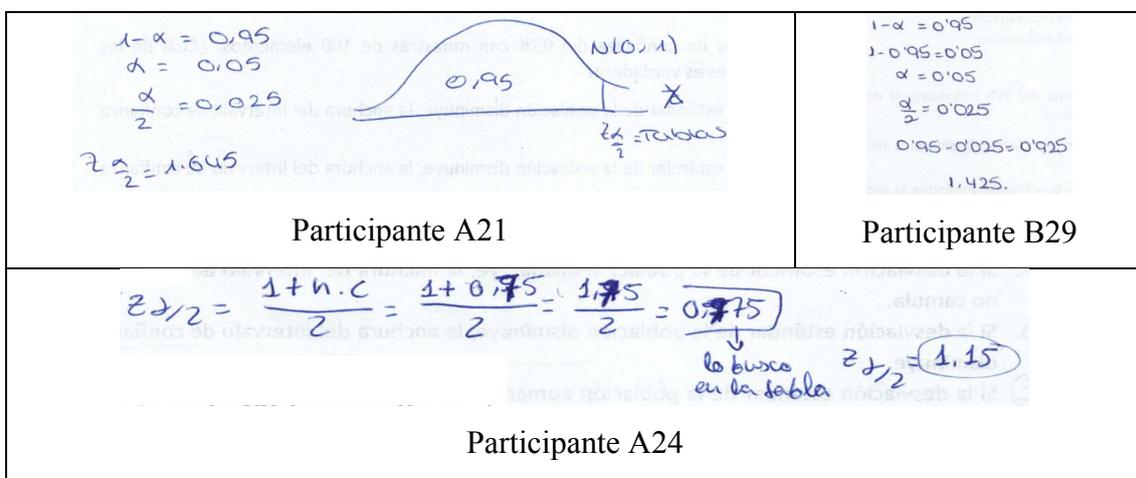


Figura 3.5.2.2. Algunos errores al determinar el valor crítico.

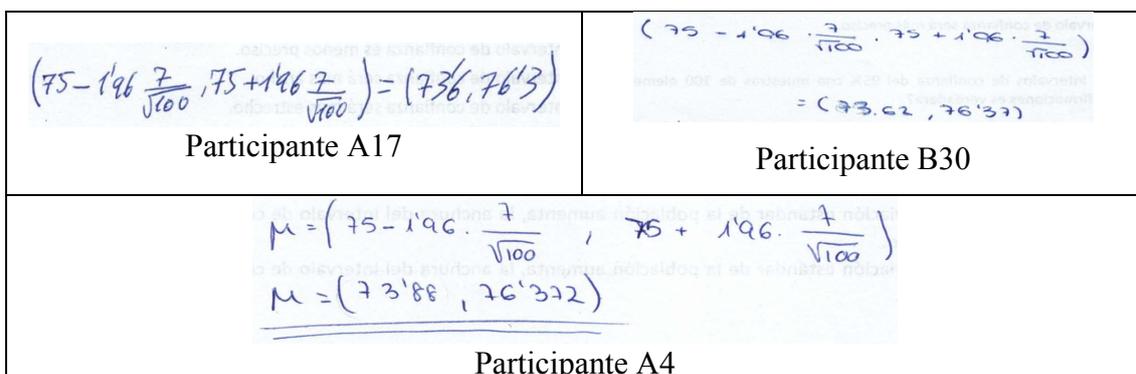


Figura 3.5.2.3. Algunos ejemplos de redondeo incorrecto en los extremos del intervalo.

Finalmente, con respecto a la cuestión del redondeo de los extremos del intervalo de confianza, lo que se observa es que el alumnado trunca el número que obtiene en la calculadora sin más miramiento, lo cual parece un error de cursos muy anteriores (primero o segundo de la E.S.O.), que no es propio de un alumno de segundo de Bachillerato.

Capítulo 4.

Conclusiones del estudio

4.1. Conclusiones generales

La conclusión general más importante que podemos extraer de nuestro estudio es la necesidad de incrementar el tiempo dedicado a la interpretación de los intervalos de confianza en Bachillerato. Al parecer, el alumnado reduce el tema al cálculo práctico del intervalo de confianza (en nuestro caso, para la media), sin atender a los aspectos que subyacen al mismo y que, casi siempre, son incluso más importantes que el cálculo. Si nos damos cuenta, el cálculo de intervalos de confianza es una cuestión ampliamente resuelta por la tecnología (cualquier programa estadístico, al alcance de todos mediante software libre, determina con exactitud dicho intervalo), que, a fin de cuentas, es un mero trámite dentro de los ambiciosos objetivos del tema. Por ello, el profesorado debería destinar más tiempo a transmitir nociones fundamentales de este tema como son las siguientes.

- (1) El intervalo de confianza es una técnica estadística que proporciona una respuesta aproximada, pero inexacta, a un problema de gran importancia: estimar un parámetro poblacional.
- (2) Por ello, en ocasiones, el intervalo de confianza no contiene al verdadero valor del parámetro poblacional. Eso sí: existe un cierto control sobre la proporción de intervalos, generados con muchas muestras aleatorias, que lo contienen.
- (3) Hay diversos factores que, incluso desde el más simple sentido común, influyen en el intervalo de confianza: por ejemplo, si tenemos una muestra de mayor tamaño, tenemos más información, lo que reduce la amplitud del intervalo de confianza; igualmente, cuanto mayor es el nivel de confianza, el proceso de cálculo debe generar intervalos más amplios pues la metodología empleada debe aumentar la proporción de intervalos que contienen al verdadero valor del parámetro.

Es importante apoyarse en las representaciones gráficas, comprendiendo su significado. La mayor parte del alumnado que realiza una representación gráfica adecuada calcula sin dificultad el valor crítico correcto.

Respecto a las hipótesis iniciales, se confirman en general, pues nuestro estudio mostró una serie de dificultades en la comprensión del intervalo de confianza por parte de los estudiantes que se preparan para realizar las pruebas de acceso a la universidad (donde, posiblemente, tengan que resolver un problema parecido al aquí propuesto), que reproducen las descritas en los antecedentes (según la hipótesis H2) y ha permitido identificar una serie de conflictos semióticos que surgen de manera natural al profundizar en este tema, unos descritos en trabajos previos y otros que constituyen una aportación original del presente trabajo (hipótesis H3).

Por otro lado, hemos observado que el alumnado comete errores impropios del curso en el que está. Los errores en el procedimiento de redondeo deberían estar más que superados y, en general, serían sencillos de subsanar si los estudiantes se fijaran con más atención en el resultado que proporciona la calculadora. No obstante, el error cometido por quienes suman y restan \bar{x} y $z_{\alpha/2}$ antes de multiplicar este valor crítico por la desviación típica muestral debería ser puesto de manifiesto en clase, incidiendo en el hecho de que lleva a un intervalo erróneo que no suele contener a la media muestral.

Aunque no era un objetivo específico del trabajo, hemos de comentar que la gran abundancia de errores cometidos por el alumnado (no sólo en cuestiones teóricas, sino también en prácticas) hubiese llevado a unos pobres resultados si se hubiese tratado de una evaluación de clase. Prueba de ello es que, en la primera parte, había nueve respuestas correctas, y el número medio de respuestas correctas ha sido de 2.7. En esta parte, como mostraremos en el Anexo I, sólo seis participantes (10.3%) han alcanzado, al menos, la mitad de respuestas correctas (cinco o más de ellas), y varios de ellos muestran algunas respuestas incorrectas. Es más, sólo hay dos alumnos que tengan, al menos, cinco respuestas correctas más que respuestas incorrectas. Por otro lado, la cuestión práctica, que posiblemente sea la más sencilla que pueda ser propuesta en las pruebas de acceso a la universidad, sólo ha sido resuelta por 17 de los 58 alumnos encuestados (un 29.3%). Al comienzo del estudio pensábamos que la mayoría de los estudiantes completarían correctamente (o al menos de forma parcialmente correcta) el problema (hipótesis H1). No ha sido así, lo cual es un poco sorprendente para nosotros.

De nuestro estudio no puede inferirse ninguna información acerca de la formación del profesorado en este tema. No obstante, siempre es adecuado (en primer lugar, aplicado

a quien escribe estas líneas) recordar la necesidad de formarse continuamente, tanto en los contenidos teóricos del tema como en las dificultades de aprendizaje que manifiestan una y otra vez nuestros estudiantes. Modestamente, esperamos que esta Memoria pueda contribuir a esta tarea entre el profesorado que se anime a leerla.

4.2. Conclusiones respecto de los objetivos planteados

Con respecto a los objetivos específicos (O1 a O8) que nos planteamos en la Sección 1.8 (página 18), podemos establecer las siguientes conclusiones.

- O1.** Hemos revisado y producido una síntesis de las investigaciones previas sobre la temática relacionada con el intervalo de confianza y su interpretación.
- O2.** Hemos analizado los significados parciales de la estimación por intervalos atendiendo a las diferentes metodologías de la inferencia.
- O3.** Hemos analizado la comprensión que manifiesta el alumnado acerca de la noción de “*intervalo de confianza*” y cómo este lo interpreta en casos prácticos concretos. Hemos constatado que una proporción importante del alumnado interpreta que la media poblacional debe estar incluida en el intervalo calculado con una probabilidad igual al nivel de confianza.
- O4.** Hemos estudiado el nivel de comprensión del alumnado de los diferentes factores que influyen en la fórmula del intervalo de confianza para la media, encontrando una proporción muy pequeña de alumnado que manifiesta conocer cómo el nivel de confianza o el tamaño de la muestra afecta a la amplitud del intervalo de confianza.
- O5.** Hemos detallado las estrategias que han empleado los estudiantes para resolver la cuestión abierta que propusimos, incluyendo una descripción del alumnado que se apoya en representaciones gráficas. Hemos descrito los pasos que han empleado para la resolución del problema. No hemos encontrado ningún estudiante que haya tratado de deducir razonadamente la fórmula del intervalo de confianza.
- O6.** Hemos identificado los principales conflictos semióticos latentes en las respuestas y estrategias que han aportado los estudiantes.
- O7.** Aunque las investigaciones previas que conocemos sobre este tema se han desarrollado con alumnado universitario, hemos hecho una comparación de los resultados obtenidos en nuestro estudio con los mostrados por otros autores.
- O8.** Aunque ya hemos introducido algunos interrogantes que nos han ido surgiendo de manera natural a lo largo del estudio, en la siguiente sección completamos esta

tarea, realizando algunas propuestas que pueden ayudar a mejorar el aprendizaje del alumnado y que puedan servir de base para futuras investigaciones.

4.3. Propuestas de mejora para el futuro y desarrollo de nuevas investigaciones

Es complicado hacer propuestas de mejora que sirvan para mejorar de manera efectiva los resultados obtenidos. No obstante, no podemos acabar esta Memoria sin aportar nuestro punto de vista sobre este asunto, y realizamos las siguientes propuestas quizá para ser llevadas a cabo en una próxima experimentación. Son las siguientes.

a) Utilizar hojas de cálculo para generar muestras aleatorias.

En nuestra opinión, una de las principales confusiones que encontramos en el alumnado está relacionada con el tipo de ejercicios que se proponen. En ellos, usualmente se trabaja con un único intervalo de confianza asociado a una única muestra aleatoria. Por ello, es complicado que el alumnado supere ciertas limitaciones que surgen cuando se trabaja así (es complicado entender la interpretación frecuencial del intervalo de confianza y el carácter aleatorio de sus extremos). Por ello, proponemos el uso de hojas de cálculo (o *applets* u otros recursos informáticos) con los que el alumnado sea capaz de simular cientos de muestras aleatorias y, asociadas a cada una de ellas, un intervalo de confianza distinto para cada caso. En esta línea de investigación se desarrollan los contenidos de varias contribuciones científicas descritas en el Anexo III.

b) Presentar el intervalo de confianza desde un punto de vista frecuencial en forma más explícita.

Una de las principales conclusiones que hemos extraído del estudio experimental desarrollado es que el alumnado cree, a ciencia cierta, que el verdadero valor del parámetro poblacional está siempre incluido en el intervalo de confianza calculado. Es necesario superar esta creencia con una enseñanza más explícita de la interpretación frecuencial del intervalo de confianza, acompañada de experimentos aleatorios que muestren que, en ocasiones, el verdadero valor no está contenido en ciertos intervalos.

c) Reescribir la propiedad esencial que determina el intervalo de confianza.

La expresión que dimos en la ecuación (1.4), a saber,

$$P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

entorpece más que ayuda para una adecuada interpretación del intervalo de confianza. Posiblemente, sea esta una de las principales razones por las que el alumnado cae en la

falacia fundamental de la confianza. Es necesario aclarar al alumnado que la expresión anterior resulta únicamente útil a la hora de determinar la expresión general para un posible intervalo de confianza para la media poblacional. De hecho, sería más conveniente presentarla en términos de distancia, asociada a la reiteración del experimento aleatorio (es decir, considerando muchas muestras independientes), como:

$$P\left(|\bar{X} - \mu| \leq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha,$$

que puede ser leída así: “la probabilidad de desarrollar un experimento aleatorio en el que las medias poblacional y muestral disten menos de $z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ es de $1 - \alpha$ ”.

d) Cambiar la notación para el nivel de confianza

Uno de los resultados obtenidos que más nos ha llamado la atención es la dificultad del alumnado para determinar el valor crítico. Utilizar la notación $1 - \alpha = 0.95$ puede llevar a la confusión de que $\alpha/2$ representa una cola a la derecha y la tabla de la distribución normal estándar que se utiliza es de colas a la izquierda (más información sobre este punto de vista en el Anexo II). Un simple cambio en la notación podría ayudar a paliar esta dificultad. Por ejemplo, si reemplazásemos la notación $1 - \alpha$ para el nivel de confianza por la letra γ (tercera letra del alfabeto griego), tendríamos que γ suele tomar un valor cercano a 1 (usualmente, 0.9, 0.95 ó 0.99), y el valor que tendríamos que buscar dentro de la tabla de la distribución normal estándar (de colas a la izquierda) para obtener el valor crítico asociado sería $(1 + \gamma)/2$, el cual es fácilmente interpretable como la media aritmética entre γ y 1. De hecho, algún alumno intentó realizar esta operación, pero se equivocó al elegir el nivel de confianza (véase la Figura 3.5.2.2, Participante A24).

e) Aclarar al alumnado la noción de “precisión” en este contexto.

Respecto al conocimiento conceptual, quizá sea este un buen momento para proponer que, en investigaciones futuras, se analice qué entiende el alumnado por “precisión”, y qué relaciones de significado establece entre la nociones de “precisión de un número” y “precisión del intervalo de confianza”.

f) Dedicar más tiempo en clase a la interpretación del intervalo de confianza.

Como conclusión general del estudio realizado, destacamos la necesidad de dedicar más tiempo en clase a la interpretación del intervalo de confianza, pues será una noción de enorme importancia en los estudios posteriores del alumnado.

REFERENCIAS

- Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. *Mathematical Thinking and Learning*, 2(1-2), 75-97.
- Batanero, C. y Borovcnik, M. (2016). *Statistics and probability in high school*. Rotterdam: Sense Publishers.
- Batanero, C., Díaz-Batanero, C., López-Martín, M. M. y Roldán López de Hierro, A. F. (En revisión). *Interval estimation: methodological approaches and understanding difficulties*.
- Bayes, T. (1970). 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* (Vol. 1, pp. 131-153). Londres: Griffin (trabajo original publicado en 1763).
- Begué, N., Batanero, C., Ruiz, K. y Gea, M.M. (2019). Understanding sampling: a summary of the research. *Beio*, 35(1), 49-78
- Behar, R. (2001). *Aportaciones para la mejora del proceso de enseñanza-aprendizaje de la estadística*. Tesis Doctoral. Universidad Politécnica de Cataluña.
- Belia, S., Fidler, F. y Cumming, G. (2005). Researchers misunderstand confidence intervals and standar error bars. *Psychological Methods*, 4, 389-396.
- Ben-Zvi, D., Bakker, A. y Makar, K. (2015). Learning to reason from samples. *Educational Studies in Mathematics*, 88(3), 291-303.
- Bolstad, W. (2013). *Introduction to Bayesian statistics*, 2ª ed. Nueva York: Wiley.
- Borovcnik, M. (2019). Informal and “informal” inference. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en www.ugr.es/local/fqm126/civeest.html
- Cabriá, S. (1994). *Filosofía de la estadística*. Valencia: Servicio de Publicaciones de la Universidad.
- Cruise, R., Dudley, R. y Thayer, J (1984). *A resource guide for introductory statistics*. Nueva York: Kendall/Hunt.
- Díaz Batanero, C. (2007). *Viabilidad de la enseñanza de la inferencia bayesiana en el análisis de datos en psicología*. Tesis doctoral. Universidad de Granada.
- Díaz-Batanero, C. (2018). *Proyecto docente*. Huelva: La autora.
- Díaz-Batanero, C., Lozano-Rojas, O. M., & Fernández-Calderón, F. (2019). La controversia sobre el contraste de hipótesis: Situación actual en psicología y recomendaciones didácticas. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística*. Disponible en www.ugr.es/local/fqm126/civeest.html
- 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 form research on statistical education. *Educational Research Review*, 2(2), 98-113.
- Cepeda-Cuervo, E., Aguilar, W., Cervantes, V., Corrales, M., Díaz, I. y Rodríguez, D. (2008). Intervalos de confianza e intervalos de credibilidad para una proporción. *Revista Colombiana de Estadística*, 31(2), 211-228.
- Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum. *Technology Innovations in Statistics Education*, 1(1), 1-15.
- Cumming, G., Williams, J. y Fidler, F. (2004). Replication, and researchers’ understanding of confidence intervals and standard error bars. *Understanding Statistics*, 18(3), 299-311. doi: 10.1111/j.1467-9280.2007.01881.x
- Gelman, A. y Shalizi, C.R. (2012). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1), 8-38.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1-26.

- Efron, B. y Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, 1, 54-75.
- Engel, J. (2010). On teaching bootstrap confidence intervals. En C. Reading (Ed.), *Proceedings of the Ninth International Conference on Teaching of Statistics*. Voorburg: International Statistical Institute.
- Fidler, F. y Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the 55th International Statistics Institute Session* CD-ROM. Sidney, Australia: International Statistical Institute.
- De la Fuente, E. I. y Díaz-Batanero, C. (2004). Controversias en el uso de la inferencia en la investigación experimental. *Metodología de las Ciencias del Comportamiento, Volumen especial 2004*, 161-167.
- Gil-Flores, J. (2005) Aplicación del método Bootstrap al contraste de hipótesis en la investigación educativa. *Revista de Educación*, 336, 251-265.
- Godino, J. D. (1996) Mathematical concepts, their meaning, and understanding. En L. Puig y A. Gutierrez (Eds.), *Proceedings of XX Conference of the International Group for the Psychology of Mathematics Education* (v.2, pp. 417-425). Universidad de Valencia.
- Godino, J. D. (2002). Un enfoque ontológico y semiótico de la cognición matemática. *Recherches en Didactique des Mathématiques*, 22 (2-3), 237-284.
- Godino, J. D. Batanero, C. y Font, V. (2007). The onto-semiotic approach to research in mathematics education. *ZDM. The International Journal on Mathematics Education*, 39(1-2), 127-135.
- Godino, J. D., Batanero, C. y Font, V. (2019). The onto-semiotic approach: Implications for the prescriptive character of didactics. *For the Learning of Mathematics*, 39(1), 38-43.
- Harradine, A., Batanero, C. y Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. En C. Batanero, G. Burrill y C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (pp. 235-246). Springer Netherlands.
- Hesterberg, T., Monaghan, S., Moore, D. S., Clipson, A., y Epstein, R. (2003). Bootstrap methods and permutation tests: En D. Moore, G. McCabe, W. Duckworth y L. Alwan (Eds.). *The practice of business statistics* (pp. 1-70). Nueva York: Freeman
- Howell, D. (n.d.). Resampling statistics: Randomization & Bootstrap. Statistical page Howell. Online: www.uvm.edu/~dhowell/StatPages/Resampling/Resampling.html.
- Krippendorff, K. (2013). *Content analysis: an introduction to its methodology*. London, Sage.
- León, O. G. y Montero, I. (2003). *Métodos de investigación en psicología y educación*. Madrid: McGraw-Hill.
- 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.
- Lecoutre, B. (2006). Training students and researchers in Bayesian methods for experimental data analysis. *Journal of Data Science*, 4(2), 207-232.
- Lecoutre, B. y Poitevineau, J. (2014). *The significance test controversy revisited*. Springer, Berlin, Heidelberg.
- López-Martín, M.M., Batanero, C. y Gea, M.M. (2019). Prospective high school teachers' interpretation of hypothesis tests and confidence intervals. Trabajo presentado en *CERME II*, Utrecht, Febrero, 2019.
- López-Martín, M.M., Batanero, C. y Gea, M.M. (2019). ¿Conocen los futuros profesores los errores de sus estudiantes en inferencia? *Bolema*, en prensa.
- Mayo, D. G. (1981). In defense of the Neyman-Pearson theory of statistics. *Philosophy of Science* 48, 269-280.
- Mayo, D. G. y Cox, D. R. (2006). Frequentist statistics as a theory of inductive inference. *IMS Lecture Notes-Monograph Series*, 49, 77-97.

- Ministerio de Educación, Cultura y Deporte, MECD (2015). *Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato*. Madrid: Autor.
- Mooney, C. Z., Duval, R. D. y Duval, R. (1993). *Bootstrapping: A nonparametric approach to statistical inference*. Londres: Sage.
- Moore, D. S., McCabe, G. P., Alwan, L. C., Craig, B. A., y Duckworth, W. M. (2016). *The practice of statistics for business and economics*. New York: Freeman.
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23(1), 103-123.
- Neyman, J. (1934). On the two different aspects of the representative method. *Journal of the Royal Statistical Society*, 97, 558-625.
- Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transaction of the Royal Society of London, series A, Mathematical and Physical Sciences* 236 (767), 33-380.
- Olivo, E. (2008). *Significados del intervalo de confianza en la enseñanza de la ingeniería en México*. Tesis Doctoral. Universidad de Granada.
- Olivo, E. y Batanero, C. (2007). Un estudio exploratorio de dificultades de comprensión del intervalo de confianza. *Unión* 12, 37-51.
- Olivo, E., Batanero, C. y Díaz, C. (2008). Dificultades de comprensión del intervalo de confianza en estudiantes universitarios. *Educación Matemática*, 20(3), 5-32.
- Parsonage, R., Pfannkuch, M., Wild, C. J. y Aloisio, K. (2016). Bootstrapping confidence intervals. In D. Ben-Zvi y K. Makar (Eds.), *The teaching and learning of statistics* (pp. 181-191). Springer, Cham.
- Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Mathematical Proceedings of the Cambridge Philosophical Society* 45(3), 483-484.
- Rivadulla, A. (1991). *Probabilidad e inferencia científica*. Barcelona: Anthropolos.
- Rossman, A. J. (2008). Reasoning about Informal Statistical Inference: One Statistician's View. *Statistics Education Research Journal*, 7(2).
- Rouanet, H. (1998). Statistics for researchers. En H. Rouanet et al. (Eds.), *New ways in statistical methodology* (pp. 1 – 28). Berna: Peter Lang.
- Scotto, M. G. y Tobías-Garcés, A. (2003). Interpretando correctamente en salud pública estimaciones puntuales, intervalos de confianza y contrastes de hipótesis. *Salud Pública de México*, 45, 506-511.
- Tukey, J. W. (1956). Bias and confidence in non-quite large samples. *Annals of Mathematical Statistics* 29, 614.
- Wilkinson, L. y Task Force on Statistical Inference (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594 – 604.
- Yáñez, G. y Behar, R. (2009). Interpretaciones erradas del nivel de confianza en los intervalos de confianza y algunas explicaciones plausibles. En M.J. González, M.T. González y J. Murillo (Eds.). *Investigación en Educación Matemática XIII*, Santander: SEIEM.
- 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.

Anexo I.

Puntuaciones individuales de los estudiantes en los ítems de opción múltiple

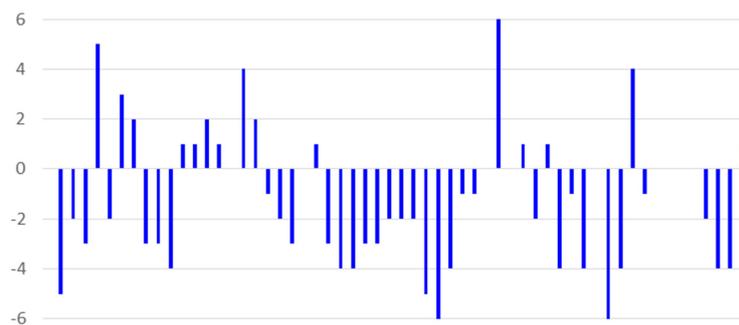
En la Figura A.1 se observa que predominan las respuestas incorrectas frente a las correctas. Es más, sólo 16 participantes (27.6%) tienen cuatro o más respuestas correctas (en muchas ocasiones, acompañadas de varias respuestas incorrectas). Teniendo en cuenta que había nueve posibles respuestas correctas, sólo seis participantes (10.3%) han alcanzado, al menos, la mitad de respuestas correctas (cinco o más de ellas). Evidentemente, este porcentaje es muy pobre, especialmente si se tiene en cuenta la diferencia entre respuestas correctas e incorrectas que se representa en la Figura A.2.

Obsérvese que en la Figura A.2 da la sensación de que faltan algunos datos ya que varios alumnos tuvieron la misma cantidad de respuestas correctas e incorrectas. Además, puede apreciarse claramente la prevalencia de errores frente a aciertos.

Figura A.1. Comparación del número de respuestas correctas e incorrectas de cada participante.



Figura A.2. Diferencia entre respuestas correctas e incorrectas de cada uno de los 58 participantes.



Anexo II

Sobre la dificultad al utilizar la notación $1 - \alpha$ para el nivel de confianza

Existen muchas formas de confeccionar una tabla para la distribución normal estándar: unas de colas a la izquierda, otras de colas a la derecha, e incluso indicando valores críticos en vez de probabilidades. Todas ellas son, en cierta forma, equivalentes. Esto nos puede llevar a pensar que cada docente tiene cierta libertad para elegir la tabla que mejor se adapte a la notación que utiliza en clase. Sin embargo, la realidad es bien distinta. Salvo en contadas excepciones, los alumnos del segundo curso de Bachillerato en Andalucía tienen bien presente que, al final del curso, si aprueban, se presentarán a las pruebas de evaluación para el acceso a la universidad. En concreto, está establecido que durante la prueba de “*Matemáticas aplicadas a las Ciencias Sociales IP*”, el alumnado podrá utilizar una tabla de la distribución normal estándar, y dicha tabla será suministrada por el evaluador junto con el examen de la asignatura. El modelo de tabla que se ha venido utilizando durante los últimos años es siempre el mismo y es conocido por el alumnado con antelación a dicha prueba. Por ello, el profesorado de la asignatura (que pudiera decantarse por otro tipo de tablas, por ejemplo una de valores críticos) siempre utiliza dicha tabla en clase con objeto de que su alumnado adquiera las habilidades necesarias para su manejo.

La tabla a la que hacemos referencia indica, para cada número real k , aproximadamente entre 0 y 4, tomado con dos cifras decimales, la probabilidad $P(Z \leq k)$, donde Z es cualquier variable con distribución normal estándar. Por ello, es una tabla de colas a la izquierda que resulta especialmente útil para calcular la probabilidad de que una variable normal tome valores en un intervalo real (después de tipificarla). Sin embargo, no es tan sencillo (al menos, no es inmediato), el cálculo de valores críticos pues no tiene claro cuál es el valor que ha de buscar en la tabla.

La notación $1 - \alpha$ para el nivel de confianza suele llevar al alumnado a determinar el valor de α debido, especialmente, a la inercia de saber que este es un dato muy importante en un contraste de hipótesis. Esta inercia también lleva al alumnado a calcular fácilmente el valor de $\alpha/2$. Sin embargo, ¿qué número se ha de buscar en la tabla para obtener el valor crítico? Hay una pequeña dificultad: $\alpha/2$ es el valor de la probabilidad acumulada en la cola a la derecha del valor crítico, pero la tabla que se utiliza en la práctica es de colas a la izquierda.

A pesar de que esta dificultad se resuelve tan fácil como calcular el valor de $1 -$

$\alpha/2$ y buscarlo dentro de la tabla, en la práctica, consideramos excesivo el porcentaje de alumnado que no consigue determinar con acierto el valor crítico $z_{\alpha/2}$ a pesar de poner de manifiesto que conoce el valor de $\alpha/2$. En concreto, en nuestro estudio, entre quienes intentan calcular el valor crítico (32 encuestados), hay 30 alumnos que indican el valor correcto de $\alpha/2$ antes de determinar el valor crítico, y 2 alumnos que tratan de calcular directamente el valor crítico sin indicar expresamente $\alpha/2$. Entre ellos, 6 se equivocan en el cálculo del valor crítico, lo supone un porcentaje de equivocación del 18.75% en una tarea tan sencilla.

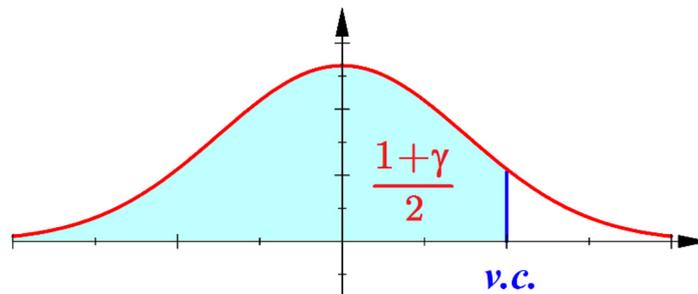
En nuestra opinión, esta dificultad nace del hecho de que el valor de $\alpha/2$ tiene gran importancia cuando se trata de resolver contrastes de hipótesis y es especialmente adecuado cuando se manejan tablas de colas a la derecha. Sin embargo, no es la notación más adecuada cuando se manejan tablas de colas a la izquierda.

En nuestra opinión, un simple cambio en la notación podría ayudar a paliar esta dificultad. Por ejemplo, si reemplazásemos la notación $1 - \alpha$ para el nivel de confianza por la letra γ (tercera letra del alfabeto griego), tendríamos que γ suele tomar un valor cercano a 1 (usualmente, 0.9, 0.95 ó 0.99), y el valor que tendríamos que buscar dentro de la tabla de la distribución normal estándar (de colas a la izquierda) para obtener el valor crítico asociado sería:

$$\frac{1 + \gamma}{2},$$

el cual es fácilmente interpretable como la media aritmética entre γ y 1. En tal caso, la representación gráfica de dicho valor quedaría como se puede observar en la Figura A.3, donde γ sería el nivel de confianza y *v.c.* el valor crítico asociado a dicho nivel de confianza.

Figura A.3. Representación gráfica de la cantidad $(1 + \gamma)/2$.



Anexo III.

Actividades de investigación realizadas durante la elaboración del Trabajo Fin de Máster

Publicaciones en revistas

- Batanero, C., Díaz-Batanero, C., López-Martín, M. M. y Roldán López de Hierro, A. F. (en revisión). *Interval estimation: methodological approaches and understanding difficulties*.
- Roldán López de Hierro, A. F. y Batanero, C. (en prensa). *¿Podemos confiar en el nivel de confianza del intervalo de confianza para la proporción?* Aceptado para su publicación en la revista *Suma*.
- Roldán López de Hierro, A. F., Batanero, C. y Beltrán-Pellicer, P. (2018). *El diagrama de árbol: un recurso intuitivo en Probabilidad y Combinatoria*. *Épsilon 100*, 49-63.

Contribuciones a congresos internacionales

- Roldán López de Hierro, A. F. (2019a). Uso de Excel para mejorar la enseñanza del intervalo de confianza para la proporción. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística (CIVEEST 2019)*. Disponible en <http://digibug.ugr.es/handle/10481/55055>
- Roldán López de Hierro, A. F., López-Martín, M. M. y Batanero, C. (2019). *An alternative method to compute confidence intervals for proportion*. 11th Congress of the European Society for Research in Mathematics Education (CERME 2019), Utrecht (the Netherlands).
- Roldán López de Hierro, A. F. y Roldán, C. (2019). El diagrama de árbol como instrumento de resolución de problemas de probabilidad en Bachillerato. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística (CIVEEST 2019)*. Disponible en <http://digibug.ugr.es/handle/10481/55056>

Contribuciones a congresos nacionales

- Roldán López de Hierro, A. F. (2019b). Un método computacional para mejorar la interpretación de intervalos de confianza para la proporción en Bachillerato y grados universitarios. Libro de resúmenes de las *IX Jornadas de Enseñanza y Aprendizaje de la Estadística y la Investigación Operativa (GENAEIO 2019)*, Granada. Disponible en https://www.ugr.es/~genaeio2019/abstract_book.pdf